The heteroscedastic odd log-logistic generalized gamma regression model for censored data

Fábio Prataviera
Edwin M. M. Ortega
Gauss M. Cordeiro
Altemir da Silva Braga

VIII - Encontro dos Alunos da Pós Graduação em Estatística e Experimentação Agronômica

Piracicaba, november of 2018

Structure

Odd log-logistic family

- Following the same idea by Gleaton and Lynch (2006) - odd log-logistic-G ("OLL-G") family

$$F(t) = \int_0^{\frac{G(t;\boldsymbol{\xi})}{\bar{G}(t;\boldsymbol{\xi})}} \frac{\lambda\, x^{\lambda-1}}{(1+x^{\lambda})^2} dx = \frac{G(t;\boldsymbol{\xi})^{\lambda}}{G(t;\boldsymbol{\xi})^{\lambda} + \bar{G}(t;\boldsymbol{\xi})^{\lambda}}. \quad (1)$$

We can write

$$\lambda = \frac{\log\left[\frac{F(t)}{\bar{F}(t)}\right]}{\log\left[\frac{G(t)}{\bar{G}(t)}\right]} \quad \text{e} \quad \bar{G}(t;\boldsymbol{\xi}) = 1 - G(t;\boldsymbol{\xi}).$$

$$f(t) = \frac{\lambda g(t;\boldsymbol{\xi})\left\{G(t;\boldsymbol{\xi})[1-G(t;\boldsymbol{\xi})]\right\}^{\lambda-1}}{\left\{G(t;\boldsymbol{\xi})^{\lambda} + [1-G(t;\boldsymbol{\xi})]^{\lambda}\right\}^2}. \quad (2)$$

Generalized gamma distribution

- Stacy and Mihram (1965)

The generalized gamma (GG) model has probability density function (pdf) and cumulative distribution function (cdf) given by

$$g(t;\alpha,\tau,k) = \frac{|\tau|}{\alpha\Gamma(k)}\left(\frac{t}{\alpha}\right)^{\tau k-1}\exp\left\{-\left(\frac{t}{\alpha}\right)^{\tau}\right\},$$

and

$$G(t;\alpha,\tau,k) = \begin{cases} \gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right) = \frac{\gamma\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)}{\Gamma(k)}, & \text{if} \quad \tau > 0, \\ 1-\gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right) = 1-\frac{\gamma\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)}{\Gamma(k)}, & \text{if} \quad \tau < 0, \end{cases} \tag{3}$$

respectively, where $\tau \neq 0$ and $k > 0$ are the shape parameters, $\alpha > 0$ is the scale parameter.

Odd log-logistic generalized gamma distribution

- Density function

Further, the odd log-logistic generalized gamma (OLLGG) pdf (for $t > 0$), say $f(t) = f(t; \alpha, \tau, k, \lambda)$, becomes

$$f(t) = \frac{\lambda\,|\tau|\,(t/\alpha)^{\tau\,k-1}\exp[-(t/\alpha)^{\tau}]\{\gamma_1(k,(t/\alpha)^{\tau})[1-\gamma_1(k,(t/\alpha)^{\tau})]\}^{\lambda-1}}{\alpha\,\Gamma(k)\,\{\gamma_1^{\lambda}(k,(t/\alpha)^{\tau})+[1-\gamma_1(k,(t/\alpha)^{\tau})]^{\lambda}\}^2},$$

where τ is not zero and the other parameters are positive.

- Hazard function

The hrf of T can be constant, decreasing, increasing, upside-down bathtub (unimodal), bathtub and bimodal shaped. It is given by

$$h(t) = \begin{cases} \dfrac{\lambda\,\tau\,(t/\alpha)^{\tau\,k-1}\exp[-(t/\alpha)^{\tau}]\gamma_1^{\lambda-1}\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)}{\alpha\,\Gamma(k)\left\{\gamma_1^{\lambda}\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)+\left[1-\gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)\right]^{\lambda}\right\}\left[1-\gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)\right]}, & \text{if} \quad \tau > 0, \\[2ex] \dfrac{\lambda\,(-\tau)\,(t/\alpha)^{\tau\,k-1}\exp[-(t/\alpha)^{\tau}]\left[1-\gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)\right]^{\lambda-1}}{\alpha\,\Gamma(k)\left\{\gamma_1^{\lambda}\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)+\left[1-\gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)\right]^{\lambda}\right\}\gamma_1\left(k,\left(\frac{t}{\alpha}\right)^{\tau}\right)}, & \text{if} \quad \tau < 0. \end{cases} \tag{5}$$

Plots of the OLLGG density function

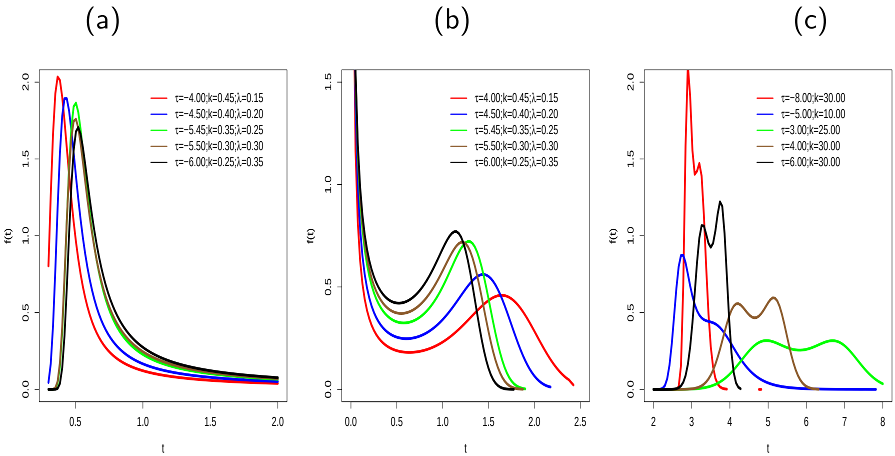

Figura: Plots of the OLLGG density function for some parameter values. (a) For some values of $\tau < 0$ with $\alpha = 1$ fixed. (b) For some values of $\tau > 0$ with $\alpha = 1$ fixed. (c) For some values of $\tau < 0$ and $\tau > 0$ with $\alpha = 2$ and $\lambda = 0.15$ fixed.

Plots of the OLLGG density function

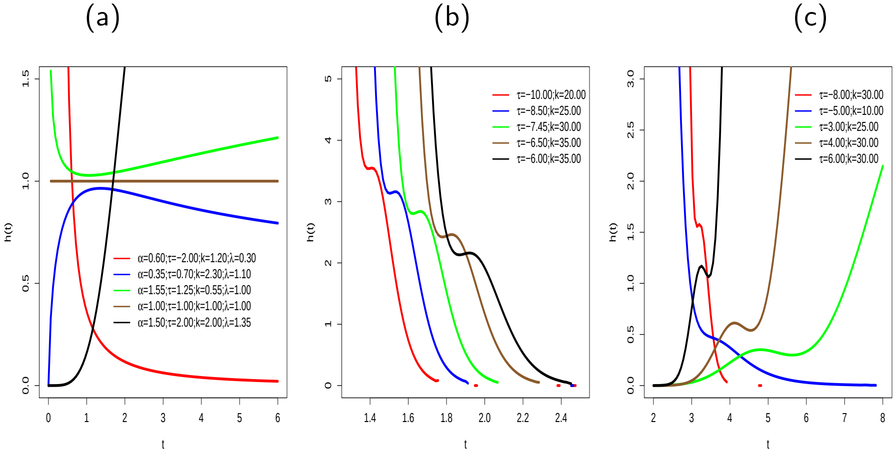

Figura: Plots of the OLLGG hrf for some parameter values. (a) For some values of $\tau < 0$ and $\tau > 0$, an k and α fixed. (b) For some values of $\tau < 0$, $\alpha = 1$, $\lambda = 0.15$ and k fixed. (c) For some values of $\tau < 0$ and $\tau > 0$, $\alpha = 2$, $\lambda = 0.15$

New models

Tabela: Some OLL-G sub-models for $\tau > 0$.

Distribution	α	τ	k	λ
OLL-Gamma	α	1	k	λ
OLL-Weibull	α	τ	1	λ
OLL-Exponential	α	1	1	λ
OLL-Chi-square	2	1	$\frac{n}{2}$	λ
OLL-Chi	$\sqrt{2}$	2	$\frac{n}{2}$	λ
OLL-Scaled Chi	$\sqrt{2}\sigma$	2	$\frac{n}{2}$	λ
OLL-Rayleigh	α	2	1	λ
OLL-Maxwell	α	2	$\frac{3}{2}$	λ
OLL-Folded normal	$\sqrt{2}$	2	$\frac{1}{2}$	λ
OLL-Reciprocal Circular normal	$\sqrt{2}$	2	1	λ
OLL-Spherical normal	$\sqrt{2}$	2	$\frac{3}{2}$	λ
OLL-Generalized half-normal	$2^{\frac{1}{2\gamma}}\theta$	2γ	$\frac{1}{2}$	λ
OLL-Half-normal	$2^{\frac{1}{2}}\theta$	2	$\frac{1}{2}$	λ

New models

Tabela: Some OLL-G sub-models for $\tau < 0$.

Distribution	α	τ	k	λ
OLL-Reciprocal Gamma	α	-1	k	λ
OLL-Reciprocal Weibull	α	-τ	1	λ
OLL-Reciprocal Exponential	α	-1	1	λ
OLL-Reciprocal Chi-square	2	-1	$\frac{n}{2}$	λ
OLL-Reciprocal Chi	$\sqrt{2}$	-2	$\frac{n}{2}$	λ
OLL-Reciprocal Scaled Chi	$\sqrt{2}\sigma$	-2	$\frac{n}{2}$	λ
OLL-Reciprocal Rayleigh	α	-2	1	λ
OLL-Reciprocal Maxwell	α	-2	$\frac{3}{2}$	λ
OLL-Reciprocal Folded normal	$\sqrt{2}$	-2	$\frac{1}{2}$	λ
OLL-Reciprocal Circular normal	$\sqrt{2}$	-2	1	λ
OLL-Reciprocal Spherical normal	$\sqrt{2}$	-2	$\frac{3}{2}$	λ
OLL-Reciprocal Generalized half-normal	$2^{\frac{1}{2\gamma}}\theta$	-2γ	$\frac{1}{2}$	λ
OLL-Reciprocal Half-normal	$2^{\frac{1}{2}}\theta$	-2	$\frac{1}{2}$	λ

Useful expansions for OLLGG distribution

First, we define the exponentiated-generalized gamma ("Exp-GG") distribution, say $W \sim \text{Exp}^c[G(\alpha, \tau, k)]$ with power parameter $c > 0$, if W has cdf and pdf given by

$$H_c(t) = G(t; \alpha, \tau, k)^c \text{ and } h_c(t) = \frac{c|\tau|}{\alpha\,\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau\,k-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] G(t; \alpha, \tau, k)^{c-1}$$

respectively.

Second, we obtain an expansion for $F(t)$ in (2) using a power series for $G(t; \alpha, \tau, k)^{\lambda}$ ($\lambda > 0$ real)

$$G(t; \alpha, \tau, k)^{\lambda} = \sum_{j=0}^{\infty} a_j\, G(t; \alpha, \tau, k)^j, \tag{6}$$

where

$$a_j = a_j(\lambda) = \sum_{j=u}^{\infty} (-1)^{u+j} \binom{\lambda}{u} \binom{u}{j}.$$

Useful expansions for OLLGG distribution

For any real $\lambda > 0$, we use generalized binomial expansion to obtain

$$[1 - G(t;\alpha,\tau,k)]^{\lambda} = \sum_{j=0}^{\infty}(-1)^j \binom{\lambda}{j} G(t;\alpha,\tau,k)^j. \tag{7}$$

Inserting (6) and (7) in equation for cdf gives the following expressions

$$F(t) = \begin{cases} \frac{\sum_{j=0}^{\infty} a_j\, G(t;\alpha,\tau,k)^j}{\sum_{j=0}^{\infty}\, b_j\, G(t;\alpha,\tau,k)^j}, & \text{if} \quad \tau > 0, \\ \frac{\sum_{j=0}^{\infty} a_j^*\, G(t;\alpha,\tau,k)^j}{\sum_{j=0}^{\infty}\, b_j\, G(t;\alpha,\tau,k)^j}, & \text{if} \quad \tau < 0, \end{cases}$$

where $a_j^* = (-1)^j \binom{\lambda}{j}$ and $b_j = a_j + (-1)^j \binom{\lambda}{j}$ for $j \geq 0$.

Thus, for $\tau < 0$, the ratio of the two power series can be expressed as

$$F(t) = \sum_{j=0}^{\infty} c_j \, G(t;\alpha,\tau,k)^j, \tag{8}$$

where $c_0 = a_0^*/b_0$ and the coefficients c_j's (for $j \geq 1$) are determined from the recurrence equation

$$c_j = b_0^{-1} \left(a_j^* - \sum_{r=1}^{j} b_r \, a_{j-r}^* \right).$$

By differentiating (8), the pdf of T follows as

$$f(t) = \sum_{j=0}^{\infty} c_{j+1} \, h_{j+1}(t), \tag{9}$$

where $h_{j+1}(t)$ (for $j \geq 0$) is the Exp-GG density function with power parameter $j+1$ given by

$$h_{j+1}(t) = \frac{(j+1)|\tau|}{\alpha\,\Gamma(k)} \left(\frac{t}{\alpha}\right)^{\tau k - 1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\tau}\right] \gamma_1(k,(t/\alpha)^{\tau})^j. \tag{10}$$

- If the random variable T follows the OLLGG density function (4), we define $Y = \log(T)$. Setting $k = q^{-2}$, $\tau = (\sigma\sqrt{k})^{-1}$ and $\alpha = \exp\left[\mu - \tau^{-1}\log(k)\right]$, the density function of Y can be expressed as ($y \in \mathbb{R}$)

$$
\begin{aligned}
f(y) \quad = \quad & \frac{\lambda\,|q|(q^{-2})^{q^{-2}}}{\sigma\,\Gamma(q^{-2})} \exp\left\{q^{-1}\left(\frac{y-\mu}{\sigma}\right) - q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right\} \times \\
& \left\{\gamma_1\left(q^{-2}, q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)\left[1-\gamma_1\left(q^{-2}, q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)\right]\right\}^{\lambda-1} \times \\
& \left\{\gamma_1^{\lambda}\left(q^{-2}, q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right) + \left[1-\gamma_1\left(q^{-2}, q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)\right]^{\lambda}\right\}^{-2},
\end{aligned}
$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda > 0$ and q is different from zero. We consider an extended form including the case $q = 0$ (Lawless, 2003).

The LOLLGG distribution

- Thus, the density of Y can be expressed as

$$f(y)=\begin{cases}\frac{\lambda\,|q|(q^{-2})^{q^{-2}}}{\sigma\,\Gamma(q^{-2})}\exp\left\{q^{-1}\left(\frac{y-\mu}{\sigma}\right)-q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\times \\ \left\{\gamma_1\left(q^{-2},q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)\left[1-\gamma_1\left(q^{-2},q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)\right]\right\}^{\lambda-1}\times \\ \left\{\gamma_1^{\lambda}\left(q^{-2},q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)+\left[1-\gamma_1\left(q^{-2},q^{-2}\exp\left[q\left(\frac{y-\mu}{\sigma}\right)\right]\right)\right]^{\lambda}\right\}^{-2}, & \text{if}\quad q\neq 0, \\ \frac{\lambda}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\left[1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\lambda-1}\left\{\Phi^{\lambda}\left(\frac{y-\mu}{\sigma}\right)+\left[1-\Phi\left(\frac{y-\mu}{\sigma}\right)\right]^{\lambda}\right\}^{-2}, & \text{if}\quad q=0,\end{cases} \tag{11}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative functions, respectively.

Plots for LOLLGG distribution

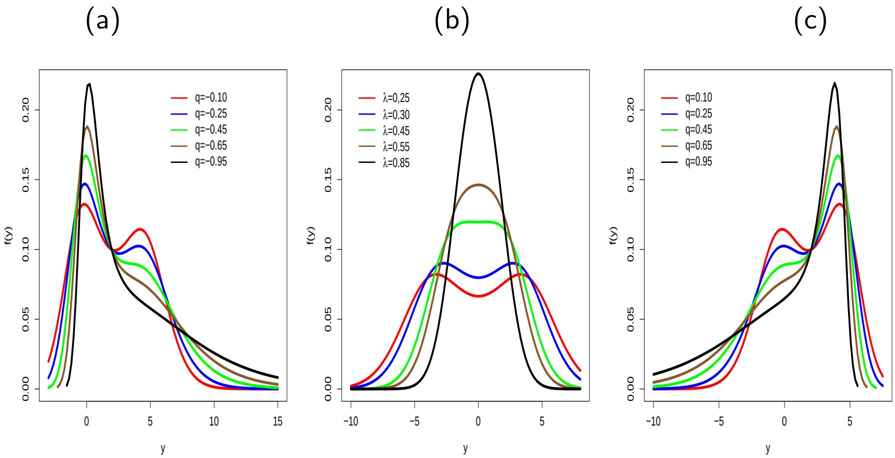

Figura: Plots of the LOLLGG density function for some parameter values. (a) For $q < 0$ fixed and $\mu = 2$, $\sigma = 1$ and $\lambda = 0.25$. (b) For $q = 0$ fixed and $\mu = 0$, $\sigma = 1.5$ and varying λ. (c) For $q > 0$ fixed and $\mu = 2$, $\sigma = 1$ and $\lambda = 0.25$.

The heteroscedastic LOLLGG regression model

- If $Y \sim LOLLGG(\mu, \sigma, q, \lambda)$, the density function of the standardized random variable $Z = (Y - \mu)/\sigma$ is given by

$$f(z) = \begin{cases} \frac{\lambda\,|q|(q^{-2})^{q^{-2}}}{\Gamma(q^{-2})} \exp\left\{q^{-1}z - q^{-2}\exp\left(qz\right)\right\} \times \\ \left\{\gamma_1\left(q^{-2}, q^{-2}\exp(qz)\right)\left[1-\gamma_1\left(q^{-2}, q^{-2}\exp(qz)\right)\right]\right\}^{\lambda-1} \times \\ \left\{\gamma_1^{\lambda}\left(q^{-2}, q^{-2}\exp(qz)\right) + \left[1-\gamma_1\left(q^{-2}, q^{-2}\exp(qz)\right)\right]^{\lambda}\right\}^{-2}, & \text{if} \quad q \neq 0, \\ \\ \lambda\,\phi\left(z\right)\left\{\Phi\left(z\right)\left[1-\Phi\left(z\right)\right]\right\}^{\lambda-1}\left\{\Phi^{\lambda}\left(z\right) + \left[1-\Phi\left(z\right)\right]^{\lambda}\right\}^{-2}, & \text{if} \quad q = 0. \end{cases} \qquad (12)$$

We write $Z \sim \text{LOLLGG}(0, 1, q, \lambda)$.

The heteroscedastic LOLLGG regression model

- In order to introduce a regression structure in the class of models (12),

$$y_i = \mu_i + \sigma_i \, z_i, \quad i = 1, \ldots, n, \tag{13}$$

where the random error z_i has density function (12), μ_i and σ_i are parameterized as

$$\mu_i = \mu_i(\boldsymbol{\beta}_1), \qquad \sigma_i = \sigma_i(\boldsymbol{\beta}_2),$$

where $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1p_1})^T$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2p_2})^T$.

- Null hypothesis $H_0 : \beta_{21} = \beta_{22} = \ldots = \beta_{2p_2} = 0$

Maximum Likelihood Estimation

- The log-likelihood function for the vector of parameters $\boldsymbol{\theta} = (q, \lambda, \boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$ from model (13) can be partitioned as

$$
l(\boldsymbol{\theta}) = \begin{cases}
r \log\left[\dfrac{\lambda\, q\left(q^{-2}\right)^{q^{-2}}}{\sigma_i \Gamma(q^{-2})}\right] + q^{-1}\sum\limits_{i\in F} z_i - q^{-2}\sum\limits_{i\in F}\exp(q\, z_i) + (\lambda-1)\sum\limits_{i\in F}\log[u_i(1-u_i)] - & \\
2\sum\limits_{i\in F}\log[u_i^{\lambda} + (1-u_i)^{\lambda}] + \lambda\sum\limits_{i\in C}\log(1-u_i) + \sum\limits_{i\in C}\log[u_i^{\lambda}+(1-u_i)^{\lambda}], & \text{if} \quad q>0, \\
r \log\left[\dfrac{-\lambda\, q\left(q^{-2}\right)^{q^{-2}}}{\sigma_i \Gamma(q^{-2})}\right] + q^{-1}\sum\limits_{i\in F} z_i - q^{-2}\sum\limits_{i\in F}\exp(q\, z_i) + (\lambda-1)\sum\limits_{i\in F}\log[u_i(1-u_i)] - & \\
2\sum\limits_{i\in F}\log[u_i^{\lambda} + (1-u_i)^{\lambda}] + \lambda\sum\limits_{i\in C}\log(u_i) + \sum\limits_{i\in C}\log[u_i^{\lambda}+(1-u_i)^{\lambda}], & \text{if} \quad q<0, \\
r \log\left(\dfrac{\lambda}{\sigma_i}\right) + \sum\limits_{i\in F}\log[\phi(z_i)] + (\lambda-1)\sum\limits_{i\in F}\log\{\Phi(z_i)[1-\Phi(z_i)]\} + \lambda\sum\limits_{i\in C}\log[1-\Phi(z_i)] - & \\
2\sum\limits_{i\in F}\log\{\Phi^{\lambda}(z_i) + [1-\Phi(z_i)]^{\lambda}\} + \sum\limits_{i\in C}\log\{\Phi^{\lambda}(z_i)+[1-\Phi(z_i)]^{\lambda}\}, & \text{if} \quad q=0,
\end{cases}
\tag{14}
$$

where r is the number of uncensored observations (failures) and

$$
u_i = \gamma_1\left(q^{-2}, q^{-2}\exp(q\, z_i)\right) \qquad \text{and} \qquad z_i = \frac{y_i - \mu_i}{\sigma_i}.
$$

Simulation study

The log-lifetimes denoted by $\log(t_1), \cdots, \log(t_n)$ are generated from the LOLLGG regression model (13), where $\mu_i = \beta_{10} + \beta_{11}\, x_i$, $\sigma_i = \exp(\beta_{20} + \beta_{21}\, x_i)$ and x_i is generated from a uniform distribution in the interval $(0, 1)$. Thus, we consider two scenarios for the simulations with sample sizes n =50, n=150 and n=350, and censoring percentages approximately equal to 0%, 10% and 30%.
The survival times are generated considering the random censoring mechanism as follows:

i. Generate $x_i \sim$ uniforme (0,1).

ii. Generate $c \sim$ uniforme (0,ν), where ν denotes the proportion of censored observations.

iii. Generate $z \sim$ LOLLGG(0,1,q,λ), the values from the density (12).

iv. Write $y^* = \beta_{10} + \beta_{11} x_i + \sigma_i z$.

v. Set $y = min(y^*, c)$.

vi. Create a vector δ of dimension n which receives 1's if $(y^* <= c)$ and zero otherwise.

Tabela: MLEs and (MSEs) for the parameters of the LOLLGG distribution.

		scenario 1			
n	Parameters	Actual values	0%	10%	30%
	σ	0.50	0.4390 (0.0475)	0.4276 (0.0612)	0.4365 (0.0979)
	q	0.65	0.9301 (3.0041)	0.8699 (2.2938)	0.9462 (2.7652)
50	λ	0.50	0.5341 (0.2070)	0.5216 (0.2550)	0.5592 (0.3876)
	β_{10}	2.00	2.0719 (0.1428)	2.0826 (0.1603)	2.1062 (0.1835)
	β_{11}	3.00	2.9690 (0.1586)	2.9729 (0.1918)	2.9595 (0.2617)
	σ	0.50	0.4843 (0.0174)	0.4851 (0.0187)	0.4810 (0.0269)
	q	0.65	0.6942 (0.2934)	0.7172 (0.6194)	0.7407 (0.7758)
150	λ	0.50	0.5141 (0.0647)	0.5173 (0.0653)	0.5259 (0.1067)
	β_{10}	2.00	2.0177 (0.0403)	2.0226 (0.0405)	2.0327 (0.0579)
	β_{11}	3.00	3.0014 (0.0425)	2.9989 (0.0511)	2.9957 (0.0720)
	σ	0.50	0.4930 (0.0067)	0.4971 (0.0090)	0.4971 (0.0103)
	q	0.65	0.6610 (0.0350)	0.6668 (0.0323)	0.6831 (0.1681)
350	λ	0.50	0.5024 (0.0200)	0.5104 (0.0251)	0.5154 (0.0365)
	β_{10}	2.00	2.0130 (0.0131)	2.0132 (0.0128)	2.0126 (0.0213)
	β_{11}	3.00	2.9961 (0.0181)	2.9970 (0.0196)	2.9966 (0.0270)

Tabela: MLEs and (MSEs) for the parameters of the LOLLGG distribution.

scenario 2					
n	Parameters	Actual values	0%	10%	30%
	q	0.65	1.0073 (4.8963)	1.0095 (5.0658)	0.9346 (4.1024)
50	λ	0.50	0.5054 (0.2214)	0.5382 (0.6702)	0.6375 (1.6844)
	β_{10}	2.00	2.1526 (1.9449)	2.1909 (3.3273)	2.3141 (5.3634)
	β_{11}	3.00	3.1890 (4.3643)	3.1929 (6.7960)	3.2681 (12.1794)
	β_{20}	0.45	0.1436 (0.3471)	0.1332 (0.4069)	0.0523 (0.7274)
	β_{21}	0.50	0.5079 (0.1876)	0.5008 (0.1921)	0.5193 (0.2899)
	q	0.65	0.7499 (0.9678)	0.7494 (1.4065)	0.7665 (1.2918)
150	λ	0.50	0.5237 (0.0742)	0.5075 (0.0723)	0.5568 (0.3785)
	β_{10}	2.00	2.0497 (0.5382)	2.0420 (0.5174)	2.1279 (1.2047)
	β_{11}	3.00	3.1248 (1.2096)	3.0399 (1.3823)	3.1623 (3.2106)
	β_{20}	0.45	0.3868 (0.0866)	0.3615 (0.1081)	0.3350 (0.1994)
	β_{21}	0.50	0.4860 (0.0473)	0.4930 (0.0508)	0.5032 (0.0673)
	q	0.65	0.6533 (0.0242)	0.6406 (0.0224)	0.6673 (0.1122)
350	λ	0.50	0.4980 (0.0175)	0.4969 (0.0195)	0.5170 (0.0449)
	β_{10}	2.00	2.0306 (0.1579)	1.9997 (0.1597)	2.0213 (0.2189)
	β_{11}	3.00	3.0051 (0.4353)	3.0015 (0.4915)	3.0255 (0.6394)
	β_{20}	0.45	0.4162 (0.0331)	0.4192 (0.0378)	0.4195 (0.0590)
	β_{21}	0.50	0.5026 (0.0183)	0.4965 (0.0227)	0.5081 (0.0275)

Residual analysis

The deviance residuals for the LOLLGG regression model with censored data are given by

$$r_{D_i} = \text{sgn}(\widehat{r}_{M_i})\left\{-2\left[\widehat{r}_{M_i} + \delta_i \log\left(\delta_i - \widehat{r}_{M_i}\right)\right]\right\}^{1/2}, \qquad (15)$$

where

$$\widehat{r}_{M_i} = \begin{cases} \delta_i + \log\left\{\dfrac{\left\{1-\gamma_1\left(\widehat{q}^{-2},\widehat{q}^{-2}\exp(\widehat{q}\,\widehat{z}_i)\right)\right\}^{\widehat{\lambda}}}{\gamma_1^{\lambda}\left(\widehat{q}^{-2},\widehat{q}^{-2}\exp(\widehat{q}\,\widehat{z}_i)\right)+\left\{1-\gamma_1\left(\widehat{q}^{-2},\widehat{q}^{-2}\exp(\widehat{q}\,\widehat{z}_i)\right)\right\}^{\widehat{\lambda}}}\right\} & \text{if} \quad q > 0, \\ \delta_i + \log\left\{\dfrac{\gamma_1^{\lambda}\left(\widehat{q}^{-2},\widehat{q}^{-2}\exp(\widehat{q}\,\widehat{z}_i)\right)}{\gamma_1^{\lambda}\left(\widehat{q}^{-2},\widehat{q}^{-2}\exp(\widehat{q}\,\widehat{z}_i)\right)+\left\{1-\gamma_1\left(\widehat{q}^{-2},\widehat{q}^{-2}\exp(\widehat{q}\,\widehat{z}_i)\right)\right\}^{\widehat{\lambda}}}\right\} & \text{if} \quad q < 0, \\ \delta_i + \log\left\{\dfrac{[1-\Phi(\widehat{z}_i)]^{\widehat{\lambda}}}{\Phi^{\widehat{\lambda}}(\widehat{z}_i)+[1-\Phi(\widehat{z}_i)]^{\widehat{\lambda}}}\right\} & \text{if} \quad q = 0, \end{cases} \qquad (16)$$

are the martingale residuals. Here, δ_i is the censoring indicator, sign$(\cdot)$ is a function that leads to the values $+1$ if the argument is positive and -1 if the argument is negative and $\widehat{z}_i = (y_i - \widehat{\mu}_i)/\widehat{\sigma}_i$.

Residual analysis - scenario 1

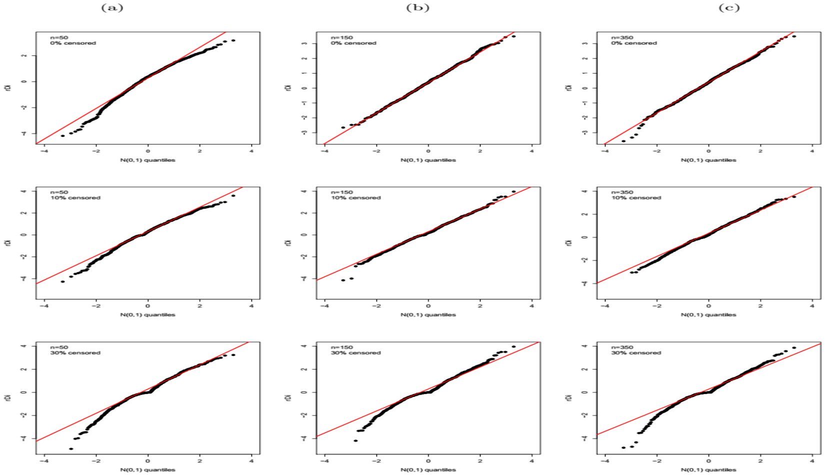

Figure 4: Normal probability plots for the residuals rD_i for scenario 1. (a) Sample size n=50. (b) Sample size n= 150. (c) Sample size n = 350. For percentages of censoring 0%, 10% and 30%.

Residual analysis - scenario 2

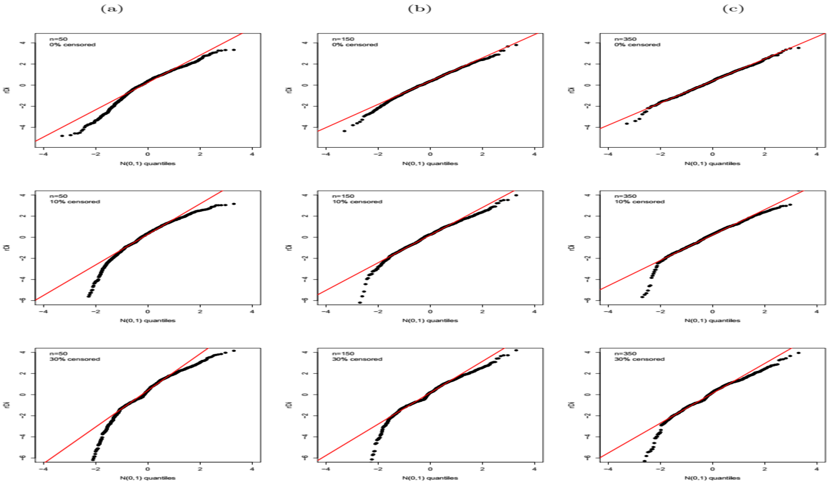

Figure 5: Normal probability plots for the residuals rD_i for scenario 2. (a) Sample size n=50. (b) Sample size n= 150. (c) Sample size n = 350. For percentages of censoring 0%, 10% and 30%.

Application

- In this application, we use a real data set available in the book of Lawless (2003, p. 335) to study the LOLLGG regression model in the presence of censoring. The model parameters are estimated by maximum likelihood using the *optim* subroutine in R.
- This data set presents an electrical insulation study (Stone, 1978) considering three voltage levels: $52.5kV$, $55kV$ and $57.5kV$ in which 20 samples were tested for each of the three levels. The variables considered in this study are t_i=time of failure in minutes of the electric insulation, $i = 1, \ldots, 60$ and the three voltage levels ($52.5kV$, $55kV$, $57.5kV$) are defined by dummy variables:$52.5kV$ ($x_{i1} = 1$ and $x_{i2} = 0$), $55kV$ ($x_{i1} = 0$ and $x_{i2} = 1$) and $57.5kV$ ($x_{i1} = 0$ and $x_{i2} = 0$).

- Descriptive analysis of the voltage data

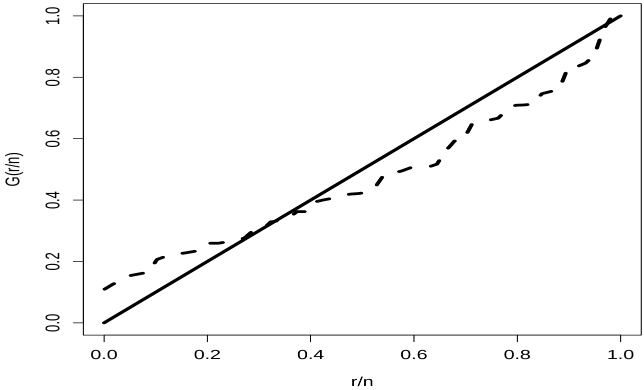

Figura: TTT-plot curve for voltage level data.

- Descriptive analysis of the voltage data

Tabela: MLEs of the model parameters for voltage level data, the corresponding SEs (given in parentheses) and the AIC, CAIC and BIC statistics.

Model	α	τ	k	λ	AIC	CAIC	BIC
OLLGG	1154.1730	0.0544	1.2817	18.8330	862.9	864.5	871.3
OLLW	1310.4937	0.4846	1	2.3717	865.9	867.0	872.2
OLLE	1161.1952	1	1	1.0769	873.1	873.8	877.3
GG	1148.0888	0.9342	1.0474	1	875.1	876.2	881.4
Weibull	1146.9611	0.9484	1	1	873.2	874.0	877.4

- Descriptive analysis of the voltage data

Tabela: LR tests.

Models	Hypotheses	Statistic w	p-value
OLLGG vs OLLW	$H_0 : k = 1$ vs $H_1 : H_0$ is false	4.99	0.0252
OLLGG vs OLLE	$H_0 : \tau = k = 1$ vs $H_1 : H_0$ is false	14.19	0.0008
OLLGG vs GG	$H_0 : \lambda = 1$ vs $H_1 : H_0$ is false	14.15	0.0001
OLLGG vs Weibull	$H_0 : \lambda = k = 1$ vs $H_1 : H_0$ is false	14.31	0.0007

Application

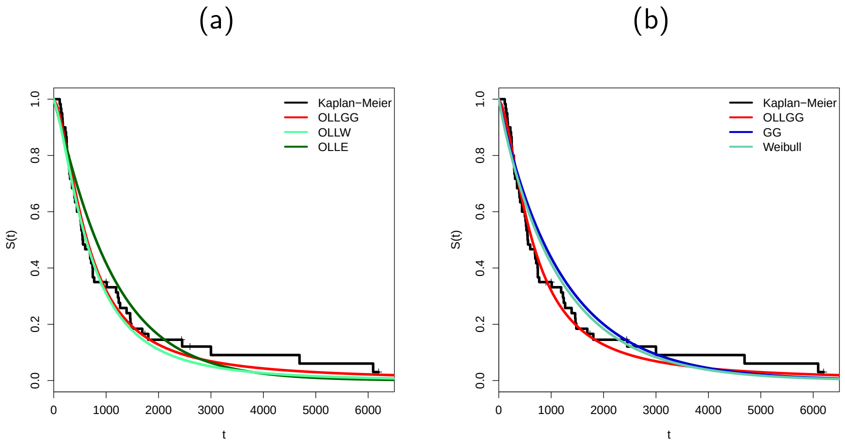

Figura: Estimated survival function with adjustment of the OLLGG distribution and some other models and empirical survival for the voltage level data. (a) OLLGG vs OLLW and OLLE. (b) OLLGG vs GG and Weibull.

Application

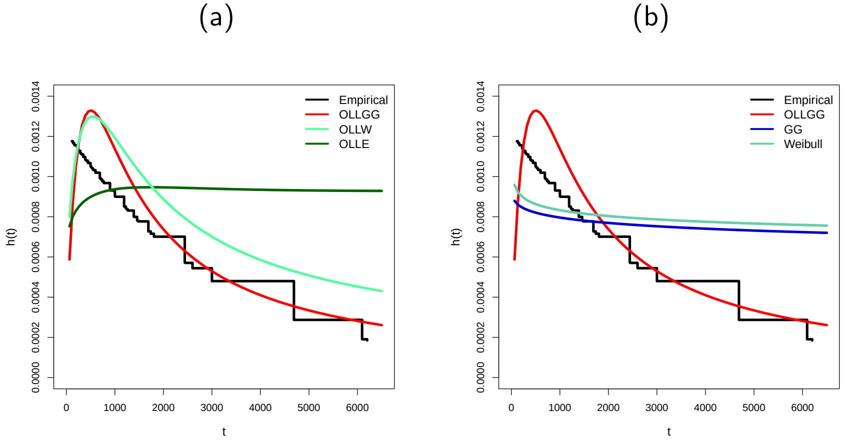

Figura: Estimated hrf with adjustment of the OLLGG distribution and some other models and empirical survival for the voltage level data. (a) OLLGG vs OLLW and OLLE. (b) OLLGG vs GG and Weibull.

- Heteroscedastic regression model

The heteroscedastic LOLLGG regression model for the voltage level data can be expressed as follows

$$y_i = \mu_i + \sigma_i z_i,$$

where $z_1, \cdots, z_{60}$ are independent random variables with density function (12) and the model parameters are defined by

$$\mu_i = \beta_{10} + \beta_{11}\, x_{i1} + \beta_{12}\, x_{i2} \quad \text{and} \quad \sigma_i = \exp(\beta_{20} + \beta_{21}\, x_{i1} + \beta_{22}\, x_{i2}).$$

The MLEs for the LOLLGG model are presented in Table 7. Thus, when establishing a significance level of 5%, we note that the voltage level is significant and should be used to model the location and scale.

Tabela: MLEs, SEs and p-values for the LOLLGG regression model fitted to the voltage level data.

Homoscedastic LOLLGG regression model				Heteroscedastic LOLLGG regression model			
Parameter	Estimate	SE	p-Value	Parameter	Estimate	SE	p-Value
σ	0.9493	0.8251	-	λ	1.9539	0.8453	-
λ	1.3466	2.3008	-	q	-8.9196	5.3776	-
q	-1.2900	2.3326	-	β_{10}	5.0220	0.5715	< 0.0001
β_{10}	5.6942	1.4660	0.0002	β_{11}	0.2822	0.4817	0.5602
β_{11}	0.7733	0.3306	0.0227	β_{12}	-0.5756	0.5842	0.3285
β_{12}	-0.1289	0.2904	0.6587	β_{20}	-1.7912	5.1280	0.7281
				β_{21}	0.5131	0.2436	0.0394
				β_{22}	0.5719	0.2732	0.0406

- The LR statistic for testing the null hypothesis $H_0 : \beta_{21} = \beta_{22} = 0$ is 6.8 (p-value 0.0333), which yields favorable indication toward to the heteroscedasticity in the voltage level data.

Application - Goodness of fit

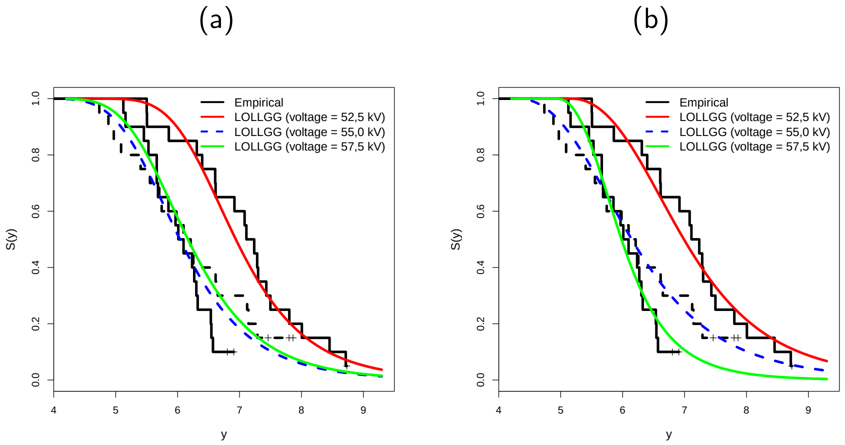

Figura: Estimated survival function for the voltage data. (a) Homoscedastic LOLLGG regression model. (b) Heteroscedastic LOLLGG regression model.

Application - Goodness of fit

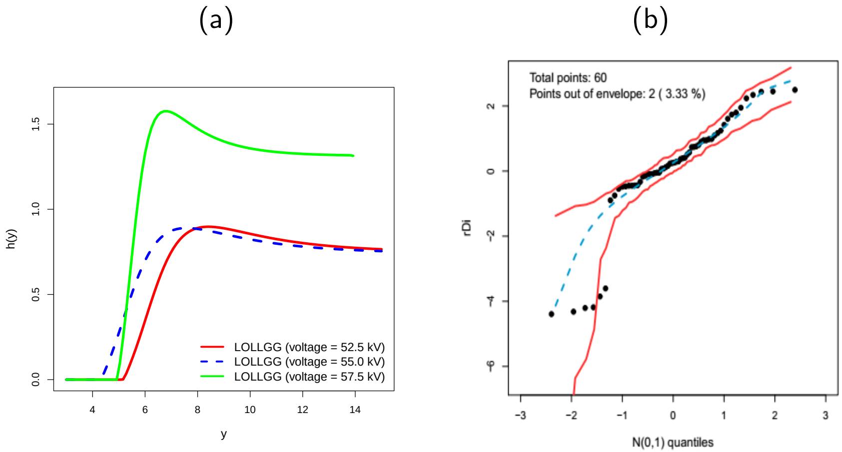

Figura: Estimated survival function for the voltage data. (a) Homoscedastic LOLLGG regression model. (b) Heteroscedastic LOLLGG regression model.

Concluding Remarks

- Through the studies, we have verified that the OLLGG distribution, in addition to allowing the modeling of data with bimodality, its risk function has very flexible forms, such as bathtub or U, increasing, decreasing and constant;
- Because we have distributions known as particular cases, such as Weibull, gamma and GG among others, this fact allows us to test the quality of fit with such models via likelihood ratio test;
- The new heteroscedastic regression model, which is very suitable for modeling censored and uncensored lifetime data.

Future research

- Semiparametric regression model;
- Competitive Risks;
- Regression model with random effect;
- Simulation study;
- Diagnostic;
- Applications.

References

Cordeiro, G.M., Alizadeh, Ozel, G., Hosseini, B., Ortega, E.M.M. and Altun, E. (2017). The generalized odd log-logistic family of distributions: properties, regression models and applications. *Journal of Statistical Computation and Simulation*, **87**, 908-932.

Cordeiro, G.M., Alizadeh, M., Tahir, M.H., Mansoor, M., Bourguignon, M. and Hamedani, G.G. (2016). The beta odd log-logistic generalized family of distributions. *Hacettepe Journal of Mathematics and Statistics*, **45**, 1175-1202.

Gleaton, J.U. and Lynch, J.D. (2006). On the distribution of the breaking strain of a bundle of brittle elastic fibers. *Advances in Applied Probability*, **36**, 98-115.

da Cruz, J.N., Ortega, E.M.M. and Cordeiro, G.M. (2016). The log-odd log-logistic Weibull regression model: Modelling, estimation, influence diagnostics and residual analysis. *Journal of Statistical Computation and Simulation*, **86**, 1516-1538.

References

Prataviera, F., Cordeiro, G.M., Suzuki, A.K. and Ortega, E.M.M. (2017). A new flexible gamma generalized model with properties, applications and Bayesian approach. *Biometrics & Biostatistics International Journal*, **6**, p. 174-193.

Prataviera, F., Ortega, E.M.M., Cordeiro, G.M. and Braga, A.D.S. (2018). The heteroscedastic odd log-logistic generalized gamma regression model for censored data. *Communications in Statistics-Simulation and Computation*, **47**, p. 1-25.

Stacy, E.W. (1962). A generalization of the gamma distribution. Annals of Mathematical Statistics, **33**, p. 1187-1192.

Stacy, E.W. and Mihram, G.A. (1965). Parameter estimation for a generalized gamma distribution. *Technometrics*, **7**, 349-358.

Thanks!!!