

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

LOG-BURR XII REGRESSION MODELS WITH
CENSORED DATA

GIOVANA OLIVEIRA SILVA
VICENTE G. CANCHO
EDWIN M. M. ORTEGA
MAURICIO LIMA BARRETO

Nº 297

RELATÓRIOS TÉCNICOS



São Carlos – SP
Mai./2007

SYSNO	1597002
DATA	/ /
ICMC - SBAB	

Log-Burr XII Regression Models with Censored Data

Giovana Oliveira Silva *

Universidade de São Paulo

Edwin M. M. Ortega[†]

Universidade de São Paulo

Vicente G. Cancho. [‡]

Universidade de São Paulo

Mauricio Lima Barreto. [§]

Universidade Federal da Bahia

Abstract

In survival analysis applications, when the failure rate function has an unimodal shape, that is a common situation, the log-normal or log-logistic distributions are used. In this paper, a regression model based in the Burr XII distribution is proposed for modeling data what has a unimodal failure rate function. The Burr XII distribution has a advantage over the log-normal that the Burr XII survival function is written in closed form and the log-logistic distribution is a special case of the Burr XII distribution. Assuming censored data, we considered a classic analysis, a Bayesian analysis assuming no informative priors and jackknife estimator for the parameters of the model. The Bayesian approach is considered using Markov Chain Monte Carlo Methods with Metropolis-Hasting algorithms steps to obtain the posterior summaries of interest. Besides, we used the sensitivity analysis to detect influential or outlying observations and residual analysis is used to check assumptions in the model such as departures from the error assumptions. The relevance of the approach is illustrated with a real data set.

Keywords: Burr distribution; regression models; censored data; local influence; generalized leverage; residual analysis.

1 Introduction

We consider in this paper data set given by Instituto de Saúde Coletiva - Universidade Federal da Bahia. This data set was designed to evaluate the effect of vitamin A supplementation on recurrent diarrheal episodes in small children (see Barreto et al., 1994). We

*Address: ESALQ, Universidade de São Paulo, Piracicaba, Brasil. E-mail: giosilva@esalq.usp.br

[†]Address: ESALQ, University of São Paulo, Piracicaba, São Paulo, Brasil. E-mail: edwin@esalq.usp.br

[‡]Address: ICMC, Universidade de São Paulo, São Carlos, Brasil. E-mail: garibay@icmc.usp.br

[§]Address: ISC, Universidade Federal da Bahia, Salvador, Brasil. E-mail: mauricio@ufba.br

¹Address for correspondence: Departamento de Ciências Exatas, USP Av. Pádua Dias 11 - Caixa Postal 9, 13418-900 Piracicaba - São Paulo - Brazil.
e-mail: edwin@esalq.usp.br

aim to model the treatment effect on the time to the occurrence of diarrhea. Moreover, the censoring times are random.

In many applications there is qualitative information about the failure rate function shape, which can help with selecting a particular model. In this context, a device called the total time on test (TTT) plot (Aarset, 1987) is useful. The TTT plot is obtained by plotting $G(r/n) = [(\sum_{i=1}^r T_{i:n}) + (n-r)T_{r:n}]/(\sum_{i=1}^n T_{i:n})$, where $r = 1, \dots, n$ and $T_{i:n}$, $i = 1, \dots, n$, are the order statistics of the sample, against r/n (Mudholkar et al., 1996). For this data, the TTT plot indicates a unimodal failure rate function.

It is known that the log-normal distribution is a popular model for survival time when the failure rate function is unimodal and the log-logistic distribution is often used as an alternative to the log-normal. The main purpose of this paper is to present other distribution that can be viewed as a more useful and flexible alternative.

We proposed to use the Burr XII distribution in modelling survival time as a viable alternative to the log-normal. The Burr XII distribution has the advantages that your survival function can be written in closed form. Besides, the log-logistic distribution is a special case of the Burr XII distribution. The Burr XII distribution was used in reliability analysis by Zimmer et al. (1998) but in data set without covariates.

We considered a classic analysis for log-Burr XII regression model. The inferential part was carried out using the asymptotic distribution of the maximum likelihood estimators, which in situations when the sample is small, it might present difficult results to be justified. As an alternative for classic analysis, we explored the use of techniques of Chains Markov Monte Carlo (MCMC) Method to develop an Bayesian inference for the log-Burr XII regression model and it was also used the Jackknife estimator. In both cases, Bayesian and Jackknife, it isn't need using asymptotic distribution of the maximum likelihood estimators.

After modelling, it is important checking assumptions in the model and to conduct a robustness study to detect influential or extreme observation that can cause distortions on the results of the analysis.

The examination of residuals was used to check assumptions in the model. Numerous approaches have been proposed in the literature to detect influential or outlying observations. An efficient way to detect influential observations is the diagnostic analysis. Cook (1986) uses this idea to motivate his assessment of local influence. He suggest that more confidence can be put in a model which is relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook and Weisberg, 1982 and Xie and Wei, 2007) in which the effect is studied of completely removing cases from the analysis. This reasoning will form the basis for our global influence introduced in section 4.1 and in doing so it will be possible to determine which subjects might be influential for the analysis. On the other hand using case deletion all information from a single subject is deleted at once and therefore it is hard to tell whether that subject has some influence on a specific aspect of the model. A solution for the earlier problem can be found in a quite different paradigm being a local influence approach where one again investigates how the results of an analysis are changed under small perturbations of the model but where these perturbations can be specific interpretations. Also, some authors have investigated the assessment of local influence in survival analysis models: for instance, Pettit and Bin Daud (1989) investigate of local influence in proportional hazard regression models, Escobar and Meeker (1992) adapt local influence methods to

regression analysis with censoring, Ortega et al. (2003) consider the problem of assessing local influence in generalized log-gama regression models with censored observations, Leiva-Sánchez et al. (2006) investigate influence local in log-Birnbaum-Saunders regression models with censored data and more recently Ortega et al. (2006) derive curvature calculations under various perturbation schemes in log-exponentiated-Weibull regression model with censored data. We developed a similar methodology to detect influential subjects in log-Burr XII regression models with censored data, it is presented in section 4.2. Finally applied methodology the leverage generalized developed by Wei et al. (1998)

In Section 2 this article is considering a briefing study of the Burr XII distribution besides the inferential part of this model. In the section 3 we suggest a log-Burr XII regression model, in addition with the maximum likelihood estimators, Bayesian inference and the Jackknife estimator. In the section 4 we used several diagnostics measures considering three perturbation schemes, case-deletion and the generalized leverage in log-burr XII regression model with censored observations. We present residual from a fitted model using the Martingale residual proposed by Therneau et al. (1990) and we proposed a modified deviance residual for the log-Burr XII regression model in the section 5. Finally, in the section 6 the real data set is analyzed and the conclusion appears in section 7.

2 The Burr XII distribution

The Burr XII distribution, used in Zimmer et al. (1998), with parameters s , c and k considers that the life time T has a density function given by

$$f(t; s, k, c) = ck \left(1 + \left(\frac{t}{s}\right)^c\right)^{-(k-1)} \frac{t^{c-1}}{s^c} \quad (1)$$

where $k > 0$ and $s > 0$ are scale parameters and $c > 0$ is shape parameter. The survival function corresponding to the random variable T with Burr XII density is given by

$$S(t; s, k, c) = P(T \geq t) = \left(1 + \left(\frac{t}{s}\right)^c\right)^{-k}$$

The corresponding failure rate function has the following form

$$h(t; s, k, c) = \frac{ck \left(\frac{t}{s}\right)^{c-1}}{s \left(1 + \left(\frac{t}{s}\right)^c\right)}$$

2.1 Characterizing the failure rate function

According to Zimmer et al. (1998), the failure rate function of the Burr XII distribution can be decrease when $c \leq 1$ and when $c > 2$ the failure rate function reaches a maximum and the decreases, where the range of values in which the failure rate function is increasing can be manipulated using s . When c values between 1 and 2, the failure rate function can be made to be essentially constant over much of the range of the distribution, this

depends of s values. To study the shape of the failure rate function we have found its derivative that can be written as

$$h'(t; c, k, s) = \frac{ckt^{c-2}}{s^c \left(1 + \left(\frac{t}{s}\right)^c\right)^2} \left[c - 1 - \left(\frac{t}{s}\right)^c \right].$$

In order to study better this function one can note that two situations might be considered:

- $c \leq 1$

To any $t > 0$, $h'(t) < 0$ and therefore $h(t)$ is an decreasing function.

- $c > 1$

When $h'(t^*) = 0$ we have $c - 1 - \left(\frac{t^*}{s}\right)^c = 0$, hence the critic point is given by $t^* = s(c - 1)^{\frac{1}{c}}$. When $t < t^*$, $h'(t) > 0$, the failure rate function is increasing and when $t > t^*$, $h'(t) < 0$, the failure rate function is decreasing. Hence, t^* is an inflexion point and the failure rate function has unimodal shape property. Besides, $h(t) \rightarrow 0$ for $t \rightarrow 0$ or $t \rightarrow \infty$.

Figure 1 shows the plots of the failure rate function for some different parameter combinations.

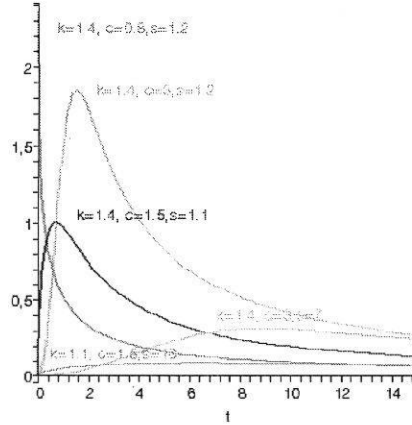


Figure 1: Plots of the failure rate function for Burr XII distribution.

From figure 1, it can be seen that the failure rate function is an decreasing function when $c \leq 1$ and $h(t)$ is a unimodal-shaped function when $c > 1$.

2.2 Moments for the failure time

The q th moments for the failure time is given by:

$$E(T^q) = s^q k B\left[\frac{q}{c} + 1, k - \frac{q}{c}\right], \quad \text{if } ck > q$$

where $B(a, b)$ is the complete beta function (see Lawless (2003)).

2.3 Relation the other distributions

The log-logistic distribution is a special case of the Burr XII distribution. When $\frac{1}{s} = m$ and $k = 1$, Burr XII distribution is reduced to the log-logistic distribution where the survival function can be written as $S(t; k, s, c) = 1 - \frac{1}{1 + (tm)^c}$.

Besides, Rodriguez (1977) shows that the Burr coverage area in specific plane is occupied by various well-known and useful distributions, including the normal, log-normal, gamma, logistic and extreme value type I distributions.

2.4 Maximum likelihood estimation

We assume that the lifetime are independently distributed, and also independent from the censoring mechanism. Considering right-censored lifetime data, we observe $t_i = \min(T_i, C_i)$, where T_i is the lifetime and C_i is the censoring, both for i th individual, $i = 1, \dots, n$. Assuming that $t_1, t_2, \dots, t_n$ is a random sample of the random variable T with Burr XII distribution (1). The likelihood function of c, k and s corresponding to the observed sample is given by

$$L(c, k, s) = (kc)^r \prod_{i \in F} \left[\left(1 + \left(\frac{t_i}{s} \right)^c \right)^{-(k+1)} \frac{t_i^{c-1}}{s^c} \right] \prod_{i \in C} \left[\left(1 + \left(\frac{t_i}{s} \right)^c \right)^{-k} \right] \quad (2)$$

where r is observed number of failures, F denotes the set of uncensored observations and C denotes the set of censored observations. The log-likelihood function is given by:

$$\begin{aligned} l(c, k, s) = & r \log(k) + r \log(c) - (k+1) \sum_{i \in F} \log \left(1 + \left(\frac{t_i}{s} \right)^c \right) + \sum_{i \in F} \log \left(\frac{t_i^{c-1}}{s^c} \right) \\ & - k \sum_{i \in C} \log \left(1 + \left(\frac{t_i}{s} \right)^c \right) \end{aligned}$$

The maximum likelihood estimator $\hat{c}$, $\hat{k}$ and $\hat{s}$ of c, k and s are obtained by maximizing the log-likelihood, which results in solving the equations

$$\frac{\partial l(c, k, s)}{\partial c} = \frac{r}{c} - (k+1) \sum_{i \in F} \frac{\left(\frac{t_i}{s} \right)^c \log \left(\frac{t_i}{s} \right)}{\left(1 + \left(\frac{t_i}{s} \right)^c \right)} + \sum_{i \in F} \log \left(\frac{t_i}{s} \right) - k \sum_{i \in C} \frac{\left(\frac{t_i}{s} \right)^c \log \left(\frac{t_i}{s} \right)}{\left(1 + \left(\frac{t_i}{s} \right)^c \right)}$$

$$\frac{\partial l(c, k, s)}{\partial k} = \frac{r}{k} - \sum_{i \in F} \log \left(1 + \left(\frac{t_i}{s} \right)^c \right) - \sum_{i \in C} \log \left(1 + \left(\frac{t_i}{s} \right)^c \right)$$

$$\frac{\partial l(c, k, s)}{\partial s} = c(k+1) \sum_{i \in F} \frac{t_i^c s^{-(c+1)}}{\left(1 + \left(\frac{t_i}{s} \right)^c \right)} - \frac{rc}{s} + c \sum_{i \in C} \frac{t_i^c s^{-(c+1)}}{\left(1 + \left(\frac{t_i}{s} \right)^c \right)}$$

These equations cannot be solved analytically so that statistical software such as Ox or R can be used to solved them. In this paper, software Ox (MAXBFGS subroutine) is used to compute the maximum likelihood estimator (MLE) but reparametrization is necessary. It can be used the transformations $c = \frac{1}{\sigma}$ and $s = \exp(\mu)$.

3 Log-Burr XII regression models

3.1 Log-location-scale regression model

In many practical applications, lifetimes are affected by variables, which are referred to as explanatory variables or covariates, such as the cholesterol level, blood pressure and many others. So it is important to explore the relationship between the lifetime and explanatory variables, an approach based in regression model can be used.

The covariate vector is denoted by $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$ which is related to the responses $Y = \log(T)$ through a regression model.

Considering the transformations $c = \frac{1}{\sigma}$ and $s = \exp(\mu)$. Hence, it follows that the density function of Y can be written as

$$f(y; k, \sigma, \mu) = \frac{k}{\sigma} \left(1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right)^{-k} \exp\left(\frac{y - \mu}{\sigma}\right) \quad (3)$$

$-\infty < y < \infty$, where $k > 0$, $\sigma > 0$, and $-\infty < \mu < \infty$. And survival function is given by

$$S(y) = \left[1 + \exp\left(\frac{y - \mu}{\sigma}\right) \right]^{-(k+1)}.$$

Besides, we have the following important theorem.

Theorem 1: For the variable Y the moment generating function (mgf) is given by

$$M_Y(t) = ks^t B\left[\frac{t}{c} + 1, k - \frac{t}{c}\right], \quad \text{if } kc > t$$

where $B[a, b]$ is the complete beta function (proof given in appendix B).

Hence the mean of Y is given by

$$E(Y) = s + \sigma[\psi(1) - \psi(k)], \quad \text{if } kc > t$$

where $\psi(a)$ is the digamma function (see Lawless(2003)).

We can write the above model as a log-linear models

$$Y = \mu + \sigma Z$$

where the variable Z follows the density

$$f(z) = k(1 + \exp(z))^{-(k-1)} \exp(z), \quad \forall -\infty < z < \infty \quad \text{and} \quad k > 0. \quad (4)$$

Now, it is also considered that the scale parameter μ of the log Burr model depends on the matrix of explanatory variables X , this is, $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$. We also consider the regression model based on the log-Burr XII given in (6) relating the response Y and the covariate vector $\mathbf{x}$, so that the distribution $Y|\mathbf{x}$ can be represents as

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma Z_i, \quad i = 1, \dots, n, \quad (5)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\sigma > 0$ and $k > 0$ are unknown parameters, $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$ is the explanatory vector and Z follows the distribution in (4).

In this case, the survival function of $Y|\mathbf{x}$ is given by

$$S(y|\mathbf{x}) = \left[1 + \exp\left(\frac{y - \mathbf{x}^T \boldsymbol{\beta}}{\sigma}\right) \right]^{-k}.$$

3.2 Estimation by maximum likelihood

The corresponding values to the sample $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \dots, (y_n, \mathbf{x}_n)$ of n observations from the distribution (6) where y_i represents the logarithm of the survival time and $\mathbf{x}_i$ the covariate vector associated with the i -th individual, the log-likelihood function can be written as

$$l(\boldsymbol{\theta}) = r \log(k) - r \log(\sigma) + \sum_{i \in F} z_i - (k+1) \sum_{i \in F} \log(1 + \exp(z_i)) - k \sum_{i \in C} \log(1 + \exp(z_i)), \quad (6)$$

where r is the number of uncensored observation (failures) and $z_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}$. Maximum likelihood estimates for the parameter vector $\boldsymbol{\theta} = (k, \sigma, \boldsymbol{\beta}^T)^T$ can be obtained by maximizing the likelihood function. In this paper, the software Ox (MAXBFGS subroutine) (see Doornik, 1996) was used to compute maximum likelihood estimates (MLE). Covariance estimates for the maximum likelihood estimators $\hat{\boldsymbol{\theta}}$ can be obtained using the Hessian matrix. Confidence intervals and hypothesis testing can be conducted using the large sample distribution of the MLE which is a normal distribution with the covariance matrix as the inverse of the Fisher information since regularity conditions are satisfied. More specifically, the asymptotic covariance matrix is given by $\mathbf{I}^{-1}(\boldsymbol{\theta})$ with $\mathbf{I}(\boldsymbol{\theta}) = E[\ddot{\mathbf{L}}(\boldsymbol{\theta})]$ such that $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = -\left\{ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right\}$.

Since it is not possible to compute the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ due to the censored observations (censoring is random and noninformative), but it is possible to use the matrix of second derivatives of the log likelihood, $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$, evaluated at the MLE $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, which is consistent. The asymptotic normal approximation for $\hat{\boldsymbol{\theta}}$ may be expressed as $\hat{\boldsymbol{\theta}}^T \sim N_{(p+2)}\{\boldsymbol{\theta}^T; \ddot{\mathbf{L}}(\boldsymbol{\theta})^{-1}\}$ where $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ is the $(p+2)(p+2)$ observed information matrix, obtained from:

$$-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{kk} & \mathbf{L}_{k\sigma} & \mathbf{L}_{k\beta_j} \\ . & \mathbf{L}_{\sigma\sigma} & \mathbf{L}_{\sigma\beta_j} \\ . & . & \mathbf{L}_{\beta_j\beta_s} \end{pmatrix}$$

with the submatrices given in appendix A.

3.3 A Bayesian analysis

The use of the Bayesian method besides being an alternative analysis, it allows the incorporation of previous knowledge of the parameters through informative priori densities. When there is not this information one considers noninformative priori. In the Bayesian approach, the referent information to the model parameters is obtained through posterior marginal distribution. In this way it appears two difficulties. The first it refers to attainment marginal posterior distribution and the second to the calculation of the interest moments. In both cases are necessary integral resolutions that many times do not present analytical solution. In this paper we have used the simulation method of Markov Chain Monte Carlo, such as the Gibbs sampler and Metropolis-Hasting algorithm.

Consider the Burr XII distribution (1), censored data and the likelihood function (2) for k , c and s . For a Bayesian analysis, we assume the following priori densities for k , s and c

- $k \sim \Gamma(a_1, b_1)$, a_1 and b_1 known;
- $s \sim \Gamma(a_2, b_2)$, a_2 and b_2 known;
- $c \sim \Gamma(a_3, b_3)$, a_3 and b_3 known;

where $\Gamma(a_i, b_i)$ denotes a gamma distribution with mean $\frac{a_i}{b_i}$, variance $\frac{a_i}{b_i^2}$ and density function given by

$$f(v; a_i, b_i) = \frac{b_i^{a_i} v^{a_i-1} \exp\{-vb_i\}}{\Gamma(a_i)}$$

where $v > 0$, $a_i > 0$ and $b_i > 0$.

In the special case where $a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 0$, the noninformative case follows, and it assumed independence among the parameters the priori densities for k , s and c is written as

$$\pi(k, s, c) \propto \frac{1}{ksc}$$

We further assume independence among the parameters k, s and c . The joint posteriori distributions for k, s and c is given by,

$$\begin{aligned} \pi(k, s, c|D) \propto & k^{a_1-1} \exp\{-kb_1\} s^{a_2-1} \exp\{-sb_2\} c^{a_3-1} \exp\{-cb_3\} \\ & \left(\frac{kc}{s^c}\right)^r \prod_{i \in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)} \right] \prod_{i \in F} t_i^{c-1} \prod_{i \in C} \left[\left(1 + \left(\frac{t_i^c}{s}\right)\right)^{-k} \right] \end{aligned}$$

where D denotes the data sets.

It can be shown that the conditional posteriori densities are given by

$$\pi(k|s, c, D) \propto k^{a_1+r-1} \exp\{-kb_1\} \prod_{i \in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)} \right] \prod_{i \in C} \left[\left(1 + \left(\frac{t_i^c}{s}\right)\right)^{-k} \right]$$

$$\pi(s|k, c, D) \propto s^{a_2-cr-1} \exp\{-sb_2\} \prod_{i \in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)} \right] \prod_{i \in C} \left[\left(1 + \left(\frac{t_i^c}{s}\right)\right)^{-k} \right]$$

$$\pi(c|k, s, D) \propto c^{a_3-1} \exp\{-cb_3\} c^r s^{-cr} \left(\frac{kc}{s^c}\right)^r \prod_{i \in F} \left[\left(1 + \left(\frac{t_i}{s}\right)^c\right)^{-(k+1)} \right] \prod_{i \in F} t_i^{c-1} \prod_{i \in C} \left[\left(1 + \left(\frac{t_i^c}{s}\right)\right)^{-k} \right]$$

Observe that we need to use the Metropolis-Hastings algorithm to generate the variables k , s and c from the respective conditional posteriori densities since their forms are somewhat complex.

For Bayesian inference, considering model (5), assume the following priori densities for σ , k and β^T :

- $k \sim \Gamma(c_1, d_1)$, c_1 and d_1 known;

- $\sigma \sim \text{Inverse}\Gamma(c_2, d_2)$, c_2 and d_2 known;
- $\beta_j \sim N(\mu_{0j}, \sigma_{0j}^2)$, μ_{0j} and σ_{0j}^2 known, $j=0, \dots, p$.

Noninformative priors, assume independence among the parameters, follows by considering $c_1 = c_2 = d_1 = d_2 = 0$ and σ_{0j}^2 large.

We again assume independence among the parameters. The joint posteriori distribution for σ , k and β is given by:

$$\pi(\sigma, k, \beta^T | D) \propto k^{c_1-1} \exp\{-kd_1\} \sigma^{-(c_2+1)} \exp\left\{-\frac{d_2}{\sigma}\right\} \exp\left\{-\frac{1}{2} \sum_{j=0}^p \left(\frac{\beta_j - \mu_{0j}}{\sigma_{0j}}\right)^2\right\} \\ \left(\frac{k}{\sigma}\right)^r \exp\left\{\sum_{i \in F} z_i\right\} \prod_{i \in F} [(1 + \exp\{z_i\})^{-(k+1)}] \prod_{i \in C} [(1 + \exp\{z_i\})^{-k}]$$

where $z_i = \frac{y_i - \mathbf{x}_i^T \beta}{\sigma}$.

It can be show that the conditional marginal distributions are given by:

$$\pi(k | \sigma, \beta^T, D) \propto k^{c_1+r-1} \exp\{-kd_1\} \exp \prod_{i \in F} [(1 + \exp\{z_i\})^{-(k+1)}] \prod_{i \in C} [(1 + \exp\{z_i\})^{-k}]$$

$$\pi(\sigma | k, \beta^T, D) \propto \sigma^{-c_2-r-1} \exp\left\{-\frac{d_2}{\sigma}\right\} \exp\left\{\sum_{i \in F} z_i\right\} \prod_{i \in F} [(1 + \exp\{z_i\})^{-(k+1)}] \\ \prod_{i \in C} [(1 + \exp\{z_i\})^{-k}]$$

$$\pi(\beta_j | k, \sigma, \beta_{-j}, D) \propto \exp\left\{-\frac{1}{2} \sum_{j=0}^p \left(\frac{\beta_j - \mu_{0j}}{\sigma_{0j}}\right)^2\right\} \\ \exp\left\{\sum_{i \in F} z_i\right\} \prod_{i \in F} [(1 + \exp\{z_i\})^{-(k+1)}] \prod_{i \in C} [(1 + \exp\{z_i\})^{-k}]$$

Observe that we need to use the Metropolis Hastings algorithm to generate from the posteriori conditional distributions of k , σ and β_j ($j=0, \dots, p$).

3.4 The Jackknife Estimator for the model

The idea the jackknifing is to transform the problem of estimating any population parameter into the problem of estimating a population mean. So, what is done when estimating a mean value is realized in this method but from an unusual point of view. In this paper, we used this method as an alternative method to estimate the population parameter.

Suppose that $T_1, T_2, \dots, T_n$ is a random sample of n values and the sample mean is given by

$$\bar{T} = \sum_{i=1}^n \frac{T_i}{n}$$

and is used to estimate the mean of the population.

Now, it is calculated the sample mean with the l^{th} observation missed out,

$$\bar{T}_{-l} = \frac{\sum_{i=1}^n T_i - T_l}{n-1}$$

Then of two expressions above is obtain

$$T_l = n\hat{T} - (n-1)\bar{T}_{-l}. \quad (7)$$

In a general situation, consider that θ is a parameter estimated by $\hat{E}(T_1, T_2, \dots, T_n)$ and for ease of notation drop $(T_1, T_2, \dots, T_n)$. Finally, it is calculated $\hat{E}_{-l}$ what is obtained with the T_l observation missed out. It follows, by equation (7) that pseudo-values can be calculated

$$\hat{E}_l^* = n\hat{E} - (n-1)\hat{E}_{-l}, \quad l = 1, \dots, n$$

The average of the pseudo-values is given by

$$\hat{E}^* = \frac{\sum_{l=1}^n \hat{E}_l^*}{n}$$

that is the jackknife estimate of θ .

Manly (1997) suggests that an approximate $100(1 - \alpha)\%$ confidence interval for θ is given by $\hat{E}^* \pm t_{\alpha/2, n-1} s / \sqrt{n}$, where $t_{\alpha/2, n-1}$ is the value that is exceeded with probability $\alpha/2$ for the t distribution with $(n-1)$ degrees of freedom and the jackknife estimator had the effect of removing bias of order $1/n$.

The jackknife estimator calculations for the log-Burr XII regression model are realized to k , σ and β_j ($j=0, \dots, p$) and confidence intervals are calculated separately to each parameter.

4 Sensitivity analysis

4.1 Global influence

A first tool to perform sensitivity analysis as stated before is by means of global influence starting from case-deletion. Case-deletion is a common approach to study the effect of dropping the i th case from the data set. The case-deletion model for the model (5) is given by

$$Y_l = \mathbf{x}_l^T \boldsymbol{\beta} + \sigma Z_l, \quad l = 1, 2, \dots, n, \quad l \neq i. \quad (8)$$

In the following, a quantity with subscript " (i) " means the original quantity with the i th case deleted. For the model (8), the log-likelihood function of $\boldsymbol{\theta}$ is denoted by $l_{(i)}(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_{(i)} = (\hat{k}_{(i)}, \hat{\sigma}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}^T)^T$ be the ML estimate de $\boldsymbol{\theta}$ from $l_{(i)}(\boldsymbol{\theta})$. To assess the influence of the i th case on the ML estimate $\hat{\boldsymbol{\theta}} = (\hat{k}, \hat{\sigma}, \hat{\boldsymbol{\beta}})^T$, the basic idea is to compare the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$. If deletion of a case seriously influences the estimates, more attention should be paid to that case. Hence, if $\hat{\boldsymbol{\theta}}_{(i)}$ is far from $\hat{\boldsymbol{\theta}}$, then i th case is regarded as an

influential observation. A first measure the influence global is defined as the standardized norm of $\hat{\theta}_{(i)} - \hat{\theta}$ (generalized Cook distance)

$$GD_i(\theta) = (\hat{\theta}_{(i)} - \hat{\theta})^T [\ddot{L}(\theta)]^{-1} (\hat{\theta}_{(i)} - \hat{\theta})$$

Other alternative is to assess the values $GD_i(\beta)$ and $GD_i(k, \sigma)$, such values reveal the impact of i th case on the estimates of β and (k, σ) , respectively. Another popular measure of the difference between $\hat{\theta}_{(i)}$ and $\hat{\theta}$ is the likelihood distance

$$LD_i(\theta) = 2\{l(\hat{\theta}) - l(\hat{\theta}_{(i)})\}$$

Besides, we can also compute $\beta_j - \beta_{j(i)}$ ($j = 1, 2, \dots, p$) to see the difference between $\hat{\beta}$ and $\hat{\beta}_{(i)}$. Alternative global influence measures are possible. One could think of the behavior of a test statistics, such as a Wald test for covariate or censoring effect, under a case deletion scheme.

4.2 Local influence

As a second tool for sensitivity analysis the local influence method will now be described for log-Burr XII regression models with censored data. Local influence calculation can be carried out in the model (12). If the likelihood displacement $LD(\omega) = 2\{l(\hat{\theta}) - l(\hat{\theta}_\omega)\}$ is used, where $\hat{\theta}_\omega$ denotes the MLE under the perturbed model, the normal curvature for θ at the direction $\mathbf{d}$, $\|\mathbf{d}\| = 1$, is given by $C_d(\theta) = 2|\mathbf{d}^T \Delta^T \ddot{L}(\theta)^{-1} \Delta \mathbf{d}|$, where Δ is a $(p+2) \times n$ matrix that depends on the perturbation scheme and whose elements are given by $\Delta_{ji} = \partial^2 l(\theta|\omega) / \partial \theta_j \partial \omega_i$, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p+2$ evaluated at $\hat{\theta}$ and ω_0 , where ω_0 is the no perturbation vector (see Cook, 1986). For the log-Burr XII model the elements of $-\ddot{L}(\hat{\theta})$ are given in appendix A. We can calculate the normal curvatures $C_d(\theta)$, $C_d(k)$, $C_d(\sigma)$ and $C_d(\beta)$ to perform various index plots, for instance, the index plot of $\mathbf{d}_{max}$ the eigenvector corresponding to $C_{d_{max}}$ the largest eigenvalue of the matrix $\mathbf{B} = \Delta^T \ddot{L}(\theta)^{-1} \Delta$ and the index plots of $C_{d_i}(\theta)$, $C_{d_i}(k)$, $C_{d_i}(\sigma)$ and $C_{d_i}(\beta)$ named total local influence (see, for example, Lesaffre & Verbeke, 1998), where $\mathbf{d}_i$ denotes an $n \times 1$ vector of zeros with one at the i th position. Thus, the curvature at the direction $\mathbf{d}_i$ assumes the form $C_i = 2|\Delta_i^T \ddot{L}(\theta)^{-1} \Delta_i|$ where Δ_i^T denotes the i th row of Δ . It is usual to point out those cases such that

$$C_i \geq 2\bar{C}, \quad \bar{C} = \frac{1}{n} \sum_{i=1}^n C_i.$$

4.3 Curvature calculations

Next, we calculate, for three perturbation schemes, the matrix

$$\Delta = (\Delta_{ji})_{(p+2) \times n} = \left(\frac{\partial^2 l(\theta|\omega)}{\partial \theta_i \omega_j} \right)_{(p+2) \times n}, \quad j = 1, 2, \dots, p+2 \quad \text{and} \quad i = 1, 2, \dots, n,$$

considering the model defined in (5) and its log-likelihood function given by (6).

4.3.1 Case-weights perturbation

Consider the vector of weights $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$.

In this case the log-likelihood function takes the form

$$l(\theta|\omega) = [\log(k) - \log(\sigma)] \sum_{i \in F} \omega_i + \sum_{i \in F} \omega_i z_i + (k+1) \sum_{i \in F} \omega_i \log[1 + \exp\{z_i\}] - \sum_{i \in C} \omega_i \log[1 + \exp\{z_i\}]$$

where $0 \leq \omega_i \leq 1$ and $\omega = (1, \dots, 1)^T$. Let us denote $\Delta = (\Delta_1, \dots, \Delta_{p+2})^T$.

Then the elements of vector Δ_1 take the form

$$\Delta_{1i} = \begin{cases} \hat{k}^{-1} + \log[1 + \exp\{\hat{z}_i\}] & \text{if } i \in F \\ \log[1 + \exp\{\hat{z}_i\}] & \text{if } i \in C \end{cases}$$

On the other hand, the elements of vector Δ_2 can be shown to be given by

$$\Delta_{2i} = \begin{cases} -\hat{\sigma}^{-1} \{1 + \hat{z}_i + (\hat{k} + 1) \hat{z}_i \exp\{\hat{z}_i\} [1 + \exp\{z_i\}]^{-1}\} & \text{if } i \in F \\ \hat{k} \hat{\sigma}^{-1} \hat{z}_i \exp\{\hat{z}_i\} [1 + \exp\{z_i\}]^{-1} & \text{if } i \in C \end{cases}$$

The elements of vector Δ_j , for $j = 3, \dots, p+2$, may be expressed as

$$\Delta_{ji} = \begin{cases} -x_{ij} \hat{\sigma}^{-1} \{1 + (\hat{k} + 1) \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1}\} & \text{if } i \in F \\ x_{ij} \hat{k} \hat{\sigma}^{-1} \exp\{z_i\} [1 + \exp\{z_i\}]^{-1} & \text{if } i \in C \end{cases}$$

4.3.2 Response perturbation

We will consider here that each y_i is perturbed as $y_{iw} = y_i + \omega_i S_y$, where S_y is a scale factor that may be estimated standard deviation of Y and $\omega_i \in \mathbf{R}$.

Here the perturbed log-likelihood function becomes expressed as

$$l(\theta|\omega) = r[\log(k) - \log(\sigma)] + \sum_{i \in F} z_i^* - (k+1) \sum_{i \in F} \log[1 + \exp\{z_i^*\}] - k \sum_{i \in C} \log[1 + \exp\{z_i^*\}]$$

where $z_i^* = \frac{(y_i + \omega_i S_y) - x_i^T \beta}{\sigma}$. In addition, the elements of the vector Δ_1 take the form

$$\Delta_{1i} = \begin{cases} -S_y \hat{\sigma}^{-1} \hat{z}_i [1 + \exp\{z_i\}]^{-1} & \text{if } i \in F \\ -S_y \hat{\sigma}^{-1} \hat{z}_i [1 + \exp\{z_i\}]^{-1} & \text{if } i \in C \end{cases}$$

On the other hand, the elements of vector Δ_2 can be shown to be given by

$$\Delta_{2i} = \begin{cases} -S_y \sigma^{-2} \left\{ 1 - (\hat{k} + 1) \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} (\hat{z}_i [1 + \exp\{\hat{z}_i\}]^{-1} + 1) \right\} & \text{if } i \in F \\ S_y \hat{k} \hat{\sigma}^{-2} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} \left\{ \hat{z}_i [1 + \exp\{\hat{z}_i\}]^{-1} + 1 \right\} & \text{if } i \in C \end{cases}$$

The elements of vector Δ_j , for $j = 3, \dots, p+2$, may be expressed as

$$\Delta_{ji} = \begin{cases} x_{ij} S_y (\hat{k} + 1) \hat{\sigma}^{-2} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-2} & \text{if } i \in F \\ x_{ij} S_y \hat{k} \hat{\sigma}^{-2} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-2} & \text{if } i \in C \end{cases}$$

4.3.3 Explanatory variable perturbation

Consider now an additive perturbation on a particular continuous explanatory variable, namely X_t , by making $x_{it\omega} = x_{it} + \omega_i S_t$, where S_t is a scaled factor, $\omega_i \in \mathbf{R}$. This perturbation scheme leads to the following expressions for the log-likelihood function and for the elements of the matrix Δ :

In this case the log-likelihood function takes the form

$$l(\theta|\omega) = r[\log(k) - \log(\sigma)] + \sum_{i \in F} z_i^* - (k+1) \sum_{i \in F} \log[1 + \exp\{z_i^*\}] - k \sum_{i \in C} \log[1 + \exp\{z_i^*\}]$$

where $z_i^* = \frac{y_i - \mathbf{x}_i^* \mathbf{T} \beta}{\sigma}$ and $\mathbf{x}_i^* \mathbf{T} = \beta_1 + \beta_2 x_{i2} + \dots + \beta_t (x_{it} + \omega_i S_t) + \dots + \beta_p x_{ip}$.

In addition, the elements of the vector Δ_1 are expressed as

$$\Delta_{1i} = \begin{cases} S_x \hat{\beta}_t \hat{\sigma}^{-1} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} & \text{if } i \in F \\ S_x \hat{\beta}_t \hat{\sigma}^{-1} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} & \text{if } i \in C, \end{cases}$$

the elements of the vector Δ_2 are expressed as

$$\Delta_{2i} = \begin{cases} \hat{\beta}_t S_x \hat{\sigma}^{-2} \left\{ 1 - (k+1) \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} (1 + z_i [1 + \exp\{\hat{z}_i\}]^{-1}) \right\} & \text{if } i \in F \\ -\hat{\beta}_t \hat{k} S_x \hat{\sigma}^{-2} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} (1 + \hat{z}_i [1 + \exp\{\hat{z}_i\}]^{-1}) & \text{if } i \in C, \end{cases}$$

the elements of the vector Δ_j , for $j = 3, \dots, p+2$ and $j \neq t$, take the forms

$$\Delta_{ji} = \begin{cases} -x_{ij} S_x \beta_t (\hat{k} + 1) \sigma^{-2} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-2} & \text{if } i \in F \\ -x_{ij} S_x \beta_t \hat{k} \sigma^{-2} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-2} & \text{if } i \in C, \end{cases}$$

the elements of the vector Δ_t are given by

$$\Delta_{ti} = \begin{cases} S_x \hat{\sigma}^{-1} + (\hat{k} + 1) S_x \sigma^{-1} \exp\{z_i\} [1 + \exp\{z_i\}]^{-1} [x_{it} \hat{\beta}_t - 1] & \text{if } i \in F \\ \hat{k} S_x \sigma^{-1} \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1} [x_{it} \hat{\beta}_t - 1] & \text{if } i \in C \end{cases}$$

4.4 Generalized Leverage

Let $l(\theta)$ denote the log-likelihood function from the postulated model in equation (5), $\hat{\theta}$ the MLE of θ and μ the expectation of $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$, then, $\hat{\mathbf{y}} = \mu(\hat{\theta})$ will be the predicted response vector.

The main idea behind the concept of leverage (see, for instance, Cook and Weisberg, 1982; Wei et al., 1998) is that of evaluating the influence of y_i on its own predicted value. This influence may well be represented by the derivative $\frac{\partial \hat{y}_i}{\partial y_i}$ that equals h_{ii} is the i -th principal diagonal element of the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ and $\mathbf{X}$ is the model matrix. Extensions to more general regression models have been given, for instance, by St. Laurent and Cook (1992), and Wei, et al. (1998) and Paula (1999), when θ is restricted with inequalities. Hence, it follows from Wei et al. (1998) that the $n \times n$ matrix $(\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{y}})$ of generalized leverage may be expressed as:

$$\mathbf{GL}(\hat{\theta}) = \left\{ \mathbf{D}_{\theta} [\ddot{\mathbf{L}}(\theta)]^{-1} \ddot{\mathbf{L}}_{\theta \mathbf{y}} \right\}$$

evaluated at $\theta = \hat{\theta}$ and where $D_{\theta} = \left(\frac{\partial[E(Y_i)]}{\partial k}, \frac{\partial[E(Y_i)]}{\partial \sigma}, x_{ij} \right)$ and

$$\ddot{L}_{\theta_y} = \frac{\partial^2 l(\theta)}{\partial \theta \partial y^T} = \left(\ddot{L}_{ky_i}, \ddot{L}_{\sigma y_i}, \ddot{L}_{\beta_j y_i} \right)^T$$

with

$$\ddot{L}_{ky_i} = \begin{cases} -\hat{\sigma}^{-1} \hat{h}_i & \text{if } i \in F \\ -\hat{\sigma}^{-1} \exp\{\hat{h}_i\} & \text{if } i \in C, \end{cases}$$

$$\ddot{L}_{\sigma y_i} = \begin{cases} \hat{\sigma}^{-2} \left\{ -1 + (\hat{k} + 1) \hat{h}_i [1 + \hat{z}_i + \exp\{\hat{z}_i\}] [1 + \exp\{\hat{z}_i\}]^{-1} \right\} & \text{if } i \in F \\ \hat{\sigma}^{-2} \hat{k} \hat{h}_i [1 + \hat{z}_i + \exp\{\hat{z}_i\}] [1 + \exp\{\hat{z}_i\}]^{-1} & \text{if } i \in C, \end{cases}$$

$$\ddot{L}_{\beta_j y_i} = \begin{cases} x_{ij} \hat{\sigma}^{-2} (\hat{k} + 1) \hat{h}_i [1 + \exp\{\hat{z}_i\}]^{-1} & \text{if } i \in F \\ x_{ij} \hat{\sigma}^{-2} \hat{k} \hat{h}_i [1 + \exp\{\hat{z}_i\}]^{-1} & \text{if } i \in C, \end{cases}$$

where $\hat{h}_i = \exp\{\hat{z}_i\} [1 + \exp\{\hat{z}_i\}]^{-1}$.

5 Residual analysis

In order to study departures from the error assumption as well as presence of outliers we will consider the deviance residual proposed by Barlow and Prentice (1988) (see also Therneau et al., 1990) and Martingale-type residual.

5.1 Martingale-type residual

This residual was introduced in counting process and can be written in log-Burr XII regression models as

$$r_{M_i} = \begin{cases} -\hat{k} \log(1 + \exp\{\hat{z}_i\}) & \text{if } i \in F \\ 1 - \hat{k} \log(1 + \exp\{\hat{z}_i\}) & \text{if } i \in C \end{cases}$$

where $\hat{z}_i = \frac{y_i - \hat{\mu}}{\hat{\sigma}}$. Due to the skewness distributional form of r_{M_i} , it has maximum value +1 and minimum value $-\infty$, transformations to achieve a more normal shaped form would be more appropriate for residual analysis.

5.2 Deviance residual

Another possibility is to use the deviance residual (see, for instance, definition in McCullagh and Nelder, 1989, section 2.4) that has been largely applied in generalized linear models (GLMs). Various authors have investigated the use of deviance residuals in GLMs (see, for instance, Williams, 1987; Hinkley et al., 1991; Paula 1995) as well as in other regression models (see, for example, Fahrmeir and Tutz, 1994). In log-Burr XII regression models the residual deviance is expressed here as

$$r_{D_i} = \begin{cases} - \left[-2 \left[-\hat{k} \log(1 + \exp\{\hat{z}_i\}) + \log(1 + \hat{k} \log(1 + \exp\{\hat{z}_i\})) \right] \right]^{\frac{1}{2}} & \text{if } i \in F \\ \text{sign} \left[1 - \hat{k} \log(1 + \exp\{\hat{z}_i\}) \right] \left[-2 + 2\hat{k} \log(1 + \exp\{\hat{z}_i\}) \right]^{\frac{1}{2}} & \text{if } i \in C. \end{cases}$$

5.3 Modified Deviance Residual

We proposed a change in the deviance residual and can be written as

$$r_{MD_i} = \delta_i + r_{D_i}$$

where $\delta_i = 0$ denotes censored observation, $\delta_i = 1$ uncensored and r_{D_i} is deviance residual that is defined in Section 5.2.

In the log-Burr XII regression models the modified residual deviance is given by

$$r_{MD_i} = \begin{cases} 1 - \left[-2 \left[-\hat{k} \log(1 + \exp\{\hat{z}_i\}) + \log(1 + \hat{k} \log(1 + \exp\{\hat{z}_i\})) \right] \right]^{\frac{1}{2}} & \text{if } i \in F \\ \text{sign} \left[1 - \hat{k} \log(1 + \exp\{\hat{z}_i\}) \right] \left[-2 + 2\hat{k} \log(1 + \exp\{\hat{z}_i\}) \right]^{\frac{1}{2}} & \text{if } i \in C. \end{cases}$$

5.4 Impact of the detected influential observations

To reveal the impact of the detected influential observations, we estimate the parameters again without the influential observations. Let $\hat{\theta}$ and $\hat{\theta}^0$ be the maximum likelihood estimates of the models that are obtained from the data sets with and without the influential observations, respectively. Lee, Lu and Song (2006) define the following two quantities to measure the difference between $\hat{\theta}$ and $\hat{\theta}^0$:

$$\text{TRC} = \sum_{i=1}^{n_p} \left| \frac{\hat{\theta}_i - \hat{\theta}_i^0}{\hat{\theta}_i} \right| \quad \text{and} \quad \text{MRC} = \max_i \left| \frac{\hat{\theta}_i - \hat{\theta}_i^0}{\hat{\theta}_i} \right|$$

where TRC is total relative changes, MRC maximum relative changes and $n_p = 6$ is the number of parameters, and likelihood displacement: $LD_I(\theta) = 2\{l(\hat{\theta}) - l(\hat{\theta}_{(I)})\}$, where $\hat{\theta}_{(I)}$ denotes MLE of θ after the set (I) of influential observations has been removed (see, Cook, Peña and Weisberg, 1988).

Now, the same number of the influential observation are randomly selected from the non influential observations and TRC, MRC and LD_I are again calculated. After this, the results can be compared if there is difference between them the observations are influential.

6 Application

We provide an application of the results derived in the previous sections using real data. The required numerical evaluations were implemented using the program Ox (see Doornik, 1996).

6.1 Application Vitamin A data

We illustrate the proposed model using data from a randomized community trial that was designed to evaluate the effect of vitamin A supplementation on diarrheal episodes in 1,207 pre-school age children, aged 6-48 months at baseline, who were assigned to receive either placebo or vitamin A in a small city in the Northeast of Brazil between December 1990 and December 1991.

The vitamin A dosage was 100,000 IU for children younger than 12 months and 200,000 IU for older children, which is the high dosage guideline established by the World Health Organization (WHO) for the prevention of vitamin A deficiency.

The total time was defined as the time from the first dose of vitamin A until the occurrence of an episode of diarrhea. An episode of diarrhea was defined as a sequence of days with diarrhea and a day with diarrhea was defined when 3 or more liquid or semi-liquid motions were reported in a 24 hour period. The information on the occurrence of diarrhea collected at each visit corresponds to a recall period of 48-72 hours. The number of liquid and semi-liquid motions per 24 hours was recorded.

The covariates considered in the models are:

- x_{i1} : age at baseline (in months);
- x_{i2} : treatment (0 = placebo, 1 = vitamin A);
- x_{i3} : gender (0 = girl, 1 = boy).

The TTT plot that is in Figure (2) indicates an unimodal shaped failure rate function

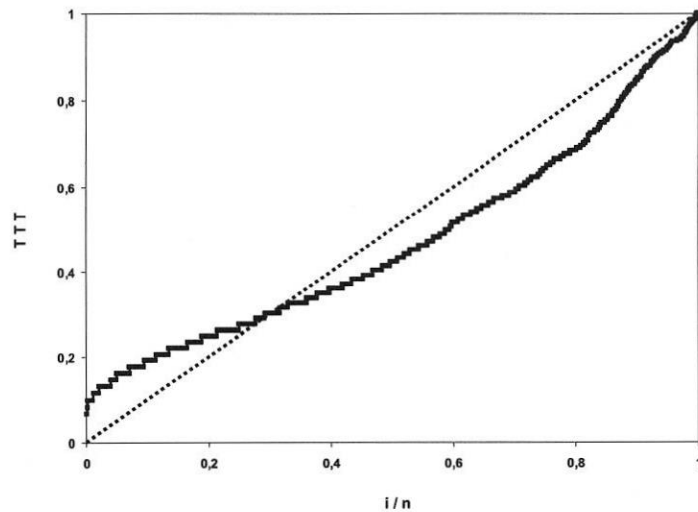


Figure 2: TTT-plot on Vitamin A data.

We present now results on fitting the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \sigma z_i$$

where the variable Y_i follows the log-Burr XII distribution given in (3), $i = 1, 2, \dots, 1207$.

6.1.1 Maximum likelihood estimation

To obtain the maximum likelihood estimates for the parameters in the model we use the subroutine MAXBFGS in Ox, whose results are given in the Table 1.

Table 1: Maximum likelihood estimates for the complete data set

Parameter	Estimate	SE	p-value
k	0,2764	0,038741	-
σ	0,3567	0,03053	-
β_0	2,2522	0,092627	0
β_1	0,0221	0,002866	< 0,01
β_2	0,0898	0,059989	0,1346
β_3	0,0441	0,059807	0,4601

We may observe that the variable x_1 is significant for the model.

6.1.2 Bayesian analysis

The following independent priors were considered to perform the Gibbs sampler. $\beta_j \sim (0, 1000)$ $j = 0, 1, 2, 3$, $\sigma \sim IG(0.01, 0.01)$ and $k \sim G(0.01, 0.01)$, so that we have a vague prior distribution. Considering these prior densities we generated two parallel independent runs of the Gibbs sampler chain with size 35000 for each parameter, disregarding the first 5000 iterations to eliminate the effect of the initial values and to avoid correlation problems, we considered a spacing of size 10, obtaining a sample of size 3000 from each chain. To monitor the convergence of the Gibbs samples we used the between and within sequence information, following the approach developed in Gelman and Rubin (1992) to obtain the potential scale reduction, $\hat{R}$. In all cases, these values were close to one, indicating the convergence of the chain. The histogram with the approximate posterior marginal density of the parameters are presented in Figure (3). In Table 2 we report posterior summaries for the parameters of the log-Burr regression model.

We may observe that the variable x_1 is significant for the model.

6.1.3 Jackknife estimator

In Table 3 we report the Jackknife estimatives for the parameters of the log-Burr XII regression model.

From table (3), we may observe that the variable x_1 is significant for the model when it is used the Jackknife estimator.

Table 2: Posterior summaries for the log-Burr regression model.

Parameter	Mean	Median	S.D.	2,5%	97.5%	$\hat{R}$
k	0.2853	0.2817	0.04085	0.2171	0.3763	1.001
σ	0.3628	0.3617	0.0308	0.3064	0.4271	1.009
β_0	2.2551	2.2534	0.0948	2.0693	2.4451	1.000
β_1	0.0224	0.0223	0.0028	0.0169	0.0281	1.002
β_2	0.0905	0.0904	0.0602	-0.0267	0.2098	1.006
β_3	0.0461	0.0458	0.0605	-0.0743	0.16428	1.004

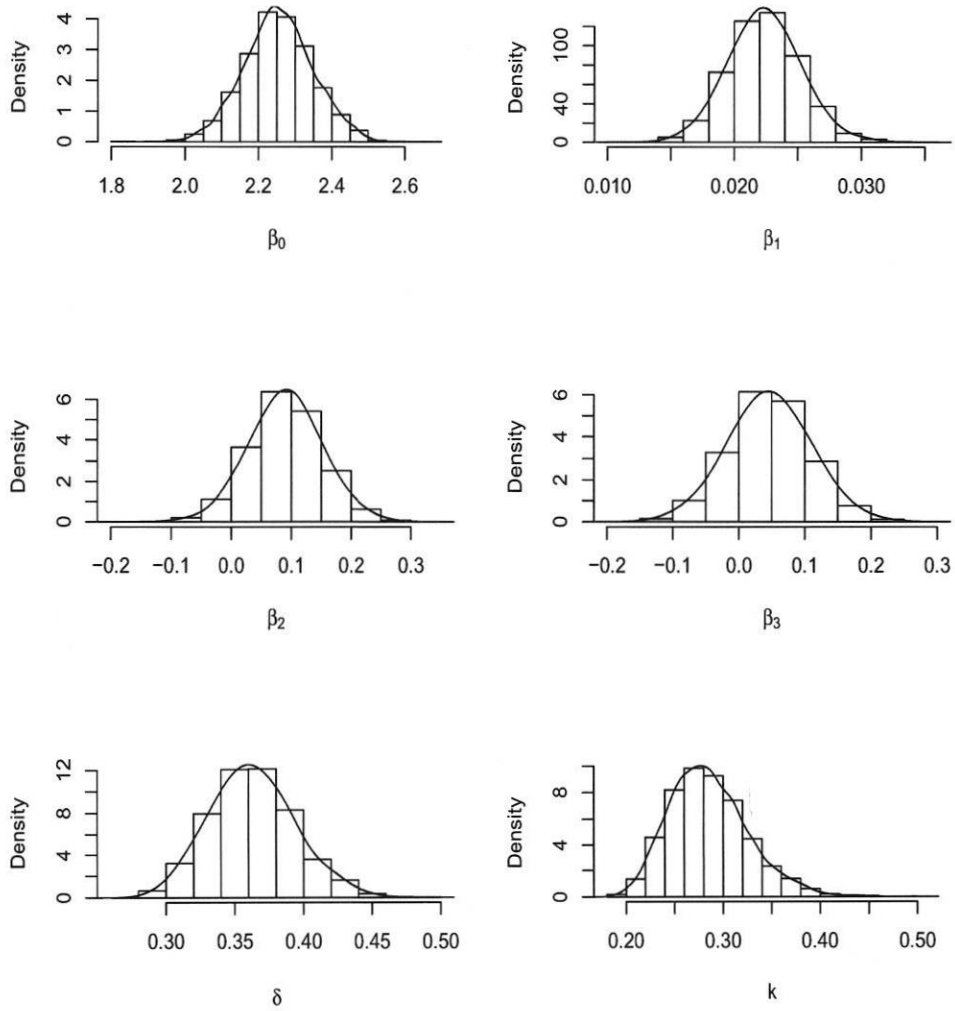


Figure 3: Approximate marginal posterior densities for β_0 , β_1 , β_2 , β_3 , σ and k

Table 3: Jackknife estimates for the complete data set

Parameter	Estimate	SE	95% Confidence Interval
k	0.26641	0.0477	(0.1728;0.3600)
σ	0.3599	0.0364	(0.2885;0.4313)
β_0	2.2464	0.0879	(2.0739;2.4189)
β_1	0.0255	0.0035	(0,0186;0.0324)
β_2	0.0921	0.0622	(-0.0299;0.2141)
β_3	0.0482	0.0616	(-0.0727;0.1691)

6.2 Global influence analysis

In this subsection, we use Ox to compute case-deletion measures $GD_i(\theta)$ and $LD_i(\theta)$ presented in sub-section 4.1. The results of such influence measures index plots are displayed in Figures 4.

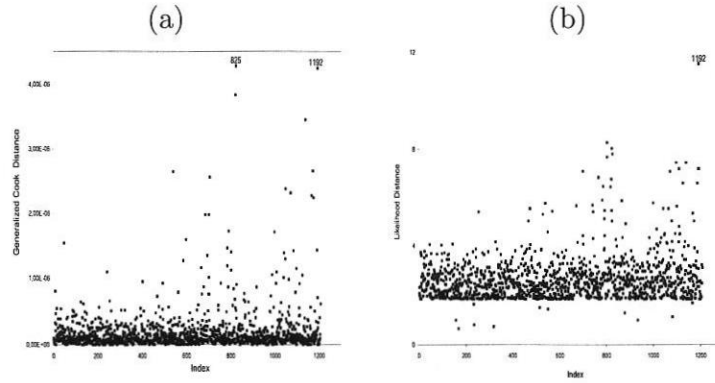


Figure 4: (a) Index plot of $GD_i(\theta)$ for θ (Generalized Cook's distance). (b) Index plot of $LD_i(\theta)$ for θ (Likelihood distance)

From figures we can see that cases 825 and 1192 are possible influential observations.

6.3 Local and total influence analysis

In this section, we will make an analysis of local influence for the data set using log-Burr XII regression models.

6.3.1 Case-weights perturbation

By applying the local influence theory developed in Section 4, where case-weight perturbation is used, value $C_{d_{max}} = 2.0230$ was obtained as maximum curvature. In Figure 5a, the graph of autovector corresponding to $C_{d_{max}}$ is presented, and total influence C_i is shown in figure 5b. Observation 1192 is the most distinguished in relation to the others.

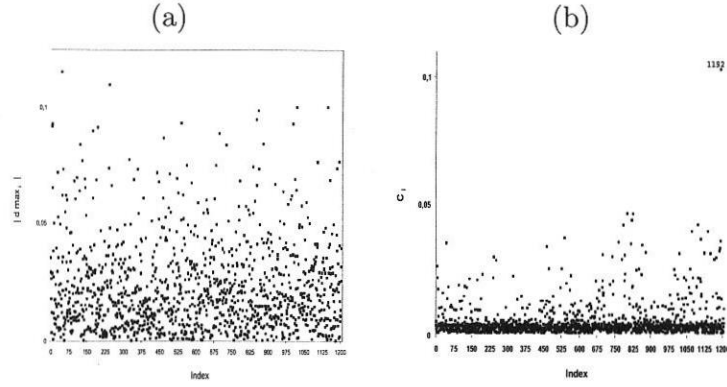


Figure 5: (a) Index plot of d_{max} for θ (case-weights perturbation). (b) Total local influence on the estimates θ (case-weights perturbation)

6.3.2 Response variable perturbation

Next, the influence of perturbations on the observed survival times will be analyzed. The value for the maximum curvature calculated was $C_{d_{max}} = 5.875$. The Figure 6a contains the graph for $|d_{max}|$ versus the observation index, shows that no point is on salient in relation to the others. The same applies to Figure 6b, which corresponds to total local influence (C_i).

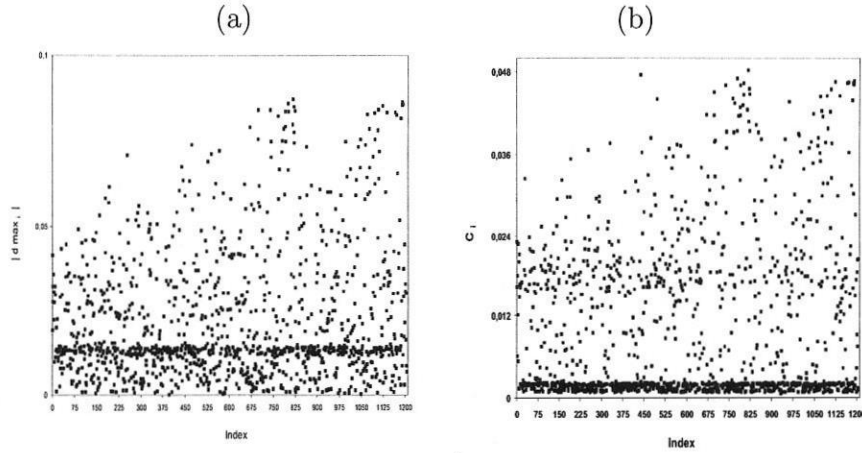


Figure 6: (a) Index plot of d_{max} for θ (response perturbation). (b) Total local influence on the estimates θ (response perturbation)

6.3.3 Explanatory variable perturbation

The perturbation of vector for covariable age (x_1) is investigated here. For perturbation of covariable age, value $C_{d_{max}} = 0.0086$ was obtained as maximum curvature. The respective graphs of $|d_{max}|$ as well as total local influence C_i against the observation index are

shown in Figures 7a and 7b, respectively. In these two graphs, we can't see no influential observation.

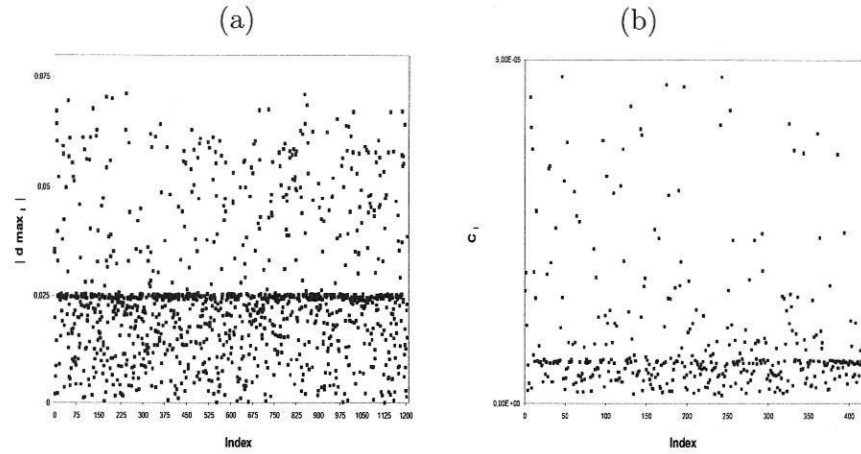


Figure 7: (a) Index plot of d_{max} for θ (age explanatory variable perturbation). (b) Total local influence on the estimates θ (age explanatory variable perturbation)

6.3.4 Generalized leverage Analysis

Figure 8 exhibits the index plot of $GL(\theta)$, using the model given in equation (12). The generalized leverage graph presented in figure 8 shows no point as possible leverage point. We can observe that all the observations well had been shaped.

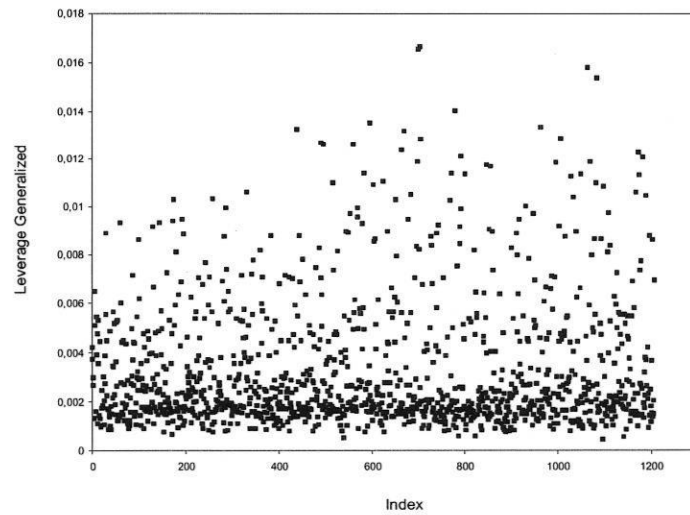


Figure 8: Index plot of generalized leverage on fitting log-Burr XII regression model for Vitamin A data.

6.4 Residual analysis

In order to detect possible outlying observations as well as departures from the assumptions of log-Burr XII model, we present in Figure 9 the graphics of r_{Mi} and r_{MDi} against the order observations.

By analyzing the martingale and deviance modified residual graph, a random behavior is observed for the data but the 1192 case is point salient in relation to the others.

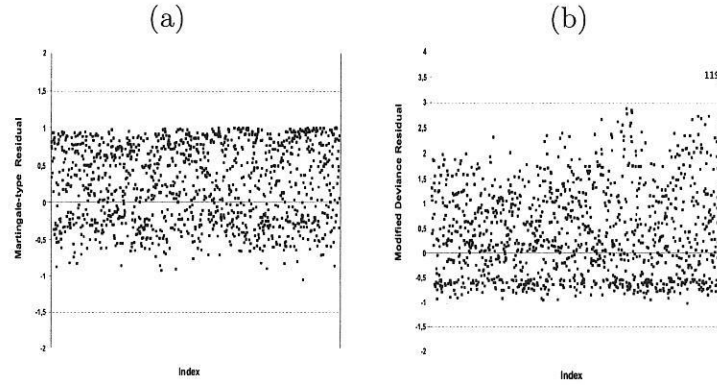


Figure 9: (a) Index plot of the Martingale-type residual (r_{Mi}). (b) Index plot of the modified deviance residual(r_{MDi}).

6.5 Impact of the detected influential observations

In concluding previous sections we can consider 1192 case as an possible influence point or outlier observation. The 1192 case has the lower time.

We find that $TRC = 0.2246$, $MRC = 0.1226$ and $LD_{(I)} = -5.58$. In order to compare the impact of the influential observations, we repeat the analysis by removing the same number (1 observation) randomly selected from the non influential observations. In this case, we find that $TRC = 0.1681$, $MRC = 0.1037$ and $LD_{(I)} = -2.53$. Hence, the results shows that 1192 case no cause strong impact in the estimative for the parameters.

6.6 Goodness of fitting

In order to assess if the model is appropriated, in Figure 10 the plot comparing the empirical distribution for the survival function and survival function estimated by the log-Burr XII regression model was introduced and there does not appear to be any suggestion in this that log Burr XII model is inappropriate.

7 Concluding Remarks

In this paper is proposed a log-Burr XII regression model with the presence of censored data as an alternative to model lifetime when the failure rate function presents unimodal

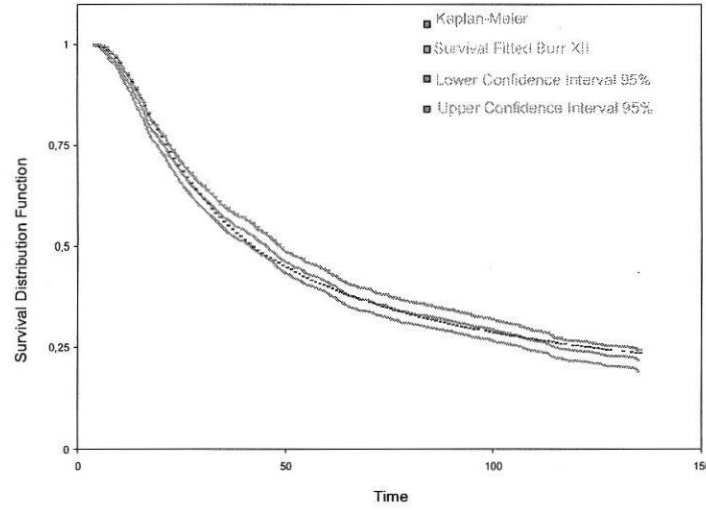


Figure 10: Estimated survival function on fitting the log-Burr XII distribution with empirical survival for Vitamin A data.

shape. We used the algorithm Quasi-Newton to obtain the estimators of maximum likelihood and to realize asymptotic tests for the parameters based on the asymptotic distribution of the maximum likelihood estimators. On the other hand, as an alternative analysis, the paper discusses the use of Markov Chain Monte Carlo methods as a reasonable way to get Bayesian inference and jackknife estimator for the log-Burr XII regression model. In the applications within a real data we observed that all estimation methods present similar results. We have discussed in this work applications of influence diagnostics in log-Burr XII regression model with censored data. Appropriate matrices for assessing global, local, total local influence as well as predictions on the fitted models under different perturbation schemes are obtained. Model fitting is also considered by using modified deviance residual and graphics of the survival function. The approach was applied to real data sets, which clearly indicates the usefulness of the approach.

Appendix A: Matrix of second derivatives $\ddot{\mathbf{L}}(\boldsymbol{\theta})$

Here we derive the necessary formulas to obtain the second order partial derivatives of the log-likelihood function. After some algebraic manipulations, we obtain

$$\mathbf{L}_{kk} = -\frac{r}{k^2}$$

$$\mathbf{L}_{k\sigma} = \sum_{i=1}^n \frac{z_i}{\sigma} h_i$$

$$\mathbf{L}_{k\beta_j} = \sum_{i=1}^n \frac{x_{ij}}{\sigma} h_i$$

$$\begin{aligned} \mathbf{L}_{\sigma\sigma} = & \frac{r}{\sigma^2} + \frac{2}{\sigma^2} \sum_{i \in F} z_i - 2(k+1) \sum_{i \in F} \frac{z_i}{\sigma^2} h_i - (k+1) \sum_{i \in F} \left(\frac{z_i}{\sigma} \right)^2 h_i + \\ & \sum_{i \in F} \left(\frac{z_i}{\sigma} h_i \right)^2 - 2k \sum_{i \in C} \frac{z_i}{\sigma^2} h_i - k \sum_{i \in C} \left(\frac{z_i}{\sigma} \right)^2 h_i + k \sum_{i=1}^n \left(\frac{z_i}{\sigma} h_i \right)^2 \end{aligned}$$

$$\begin{aligned} \mathbf{L}_{\sigma\beta_j} = & \sum_{i \in F} \frac{x_{ij}}{\sigma^2} - (k+1) \sum_{i \in F} \left(\frac{x_{ij}}{\sigma^2} + \frac{x_{ij} z_i}{\sigma^2} \right) h_i + \sum_{i \in F} \frac{x_{ij} z_i}{\sigma^2} h_i^2 \\ & - k \sum_{i \in C} \left(\frac{x_{ij}}{\sigma^2} + \frac{x_{ij} z_i}{\sigma^2} \right) h_i + k \sum_{i=1}^n \frac{x_{ij} z_i}{\sigma^2} h_i^2 \end{aligned}$$

$$\begin{aligned} \mathbf{L}_{\beta_j\beta_s} = & -(k+1) \sum_{i \in F} \frac{x_{ij} x_{is}}{\sigma^2} h_i + (k+1) \sum_{i \in F} \frac{x_{ij} x_{is}}{\sigma^2} h_i^2 - \\ & k \sum_{i \in C} \frac{x_{ij} x_{is}}{\sigma^2} h_i + k \sum_{i \in C} \frac{x_{ij} x_{is}}{\sigma^2} h_i^2 \end{aligned}$$

em que $j, s = 1, 2, \dots, p$, $h_i = \frac{\exp(z_i)}{1+\exp(z_i)}$ and $z_i = \frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}$

Appendix B: Proof of theorem 1

For log-Burr XII distribution (3), the moment generating function (mgf) is given by result in solving the equation

$$M_Y(T) = E(\exp(ty)) = \int_{-\infty}^{\infty} \exp\{ty\} k c \left(1 + \left(\frac{\exp\{y\}}{s}\right)^c\right)^{-k-1} \left(\frac{\exp\{y\}}{s}\right)^c dt$$

Let $u = \left(\frac{\exp\{y\}}{s}\right)^c$ then $du = c \left(\frac{\exp\{y\}}{s}\right)^c dy$. Hence

$$M_Y(T) = \int_{-\infty}^{\infty} \exp\{ty\} k (1+u)^{-k-1} du$$

Now make the univariate change of variable $v = \frac{1}{1+u}$ so that $dv = -(1+u)^{-2} du$ to obtain

$$M_Y(T) = \int_0^1 -s^t k 1 - v^{\frac{t}{c}} v^{-\frac{t}{c}+k-1} dv$$

$$ks^t B\left[\frac{t}{c} + 1, k - \frac{t}{c}\right], \quad \text{if } kc > t$$

where $B(a, b)$ is the complete beta function (see Lawless (1982)). To obtain the second identity we recognized the integrand as the kernel of the beta pdf.

References

- Aarset, M.V (1987) How to identify bathtub hazard rate, *IEEE Trans. Reliability*, **36** 106-108.
- Barlow, W. E., and Prentice, R. L. (1988). Residual for relative risk regression. *Biometrika*, **75**, 65-74.
- Barreto, M. L., Santos, L. M. P., Assis A. M. O., Araújo, M. P. N., Farenzena, G. G., Santos, P. A. B. e Fiaccone, R. L. (1994). Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. *Lancet*, **344**, 228-231.
- Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. New York: John Wiley.
- Collet D. (1994). *Modelling Survival Data in Medical Research*. Chapman and Hall: London.
- Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hill.

- Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19** 15-18.
- Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society*, **48** 2, 133-169.
- Cook, R. D., Peña, D. and Weisberg, S. (1988). The Likelihood Displacement: A Unifying Principle for Influence. *Communications in Statistics, Part Theory and Methods* **17**, 623-640.
- Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B*, **30** 248-275.
- Doornik, J. (1996). Ox: An Object-Oriented Matrix Programming Language. International Thomson Business Press.
- Escobar, L. A. and Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507-528.
- Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag: New York.
- Fleming, T. R. and Harrington, D. P. (1991). *Counting Process and Survival Analysis*. Wiley: New York.
- Fung, W. K. and Kwan, C. W. (1997). A note on local influence based on normal curvature. *Journal of the Royal Statistical Society B*, **59** 839-843.
- Galea, M., Paula, G. A. and Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician* **46**, 71-79.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*.
- Gu, H. and Fung, W. K. (1998). Assessing local influence in canonical analysis. *Annals of the Institute of Statistical Mathematics* **50**, 755-772.
- Hinkley, D. V., Reid, N. and Snell, E. J. (1991). *Statistical Theory and Modelling - In honour of Sir David Cox*. Chapman & Hall, London.
- Lawless, J. F. (2003). *Statistical Models and Methods for lifetime data*. Wiley: New York.
- Lee, S. Y., Lu, B. and Song, X. Y. (2006). Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational Statistics and Data Analysis*, **50** 1356-1377.
- Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54** 570-582.

- Leiva-Sanchez, V., Barros, M., Paula, G. A. and Galea, M. (2007). Influence Diagnostics in Log-Birnbaum-Saunders Regression Models with Censored Data. *Computational Statistics & Data Analysis*. In Press.
- Liu, S. Z. (2000). On local influence for elliptical linear models. *Statist. Papers* **41**, 211-224.
- Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd Edition. Chapman and Hall: London.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd Edition. Chapman and Hall: London.
- Mudholkar, G.S., Srivastava, D.k. and Kollia, G.D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, **91**, 1575-1583.
- Ortega, E. M. M., Bolfarine, H. and Paula G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis* **42**, 165-186.
- Ortega, E. M. M., Cancho, Vicente G. and Bolfarine, H. (2006). Influence diagnostics in exponentiated -Weibull regression models with censored data. *Statistics And Operations Research Transactions Sort*, **30**, 172-192.
- Paula, G. A. (1993). Assessing local influence in restricted regressions models. *Computational Statistics and Data Analysis* **16**, 63-79.
- Paula, G. A. (1995). Influence residuals in restricted generalized linear models. *J. Statist. Comput. Simulation* **51**, 63-79.
- Paula, G. A. (1999). Leverage in inequality constrained regression models. *The Statistician* **48**, 529-538.
- Paula, G. A., Sánchez, V. L., Galea, M. (2004). Influence Diagnostics in Log-Birnbaum-Saunders Regression Models. *Journal of Applied Statistics*, Inglaterra, v. 31, n. 9, p. 1049-1064.
- Paulino, C. D., Turkman, M A. A. e Murteira, B. (2003). *Estatística Bayesiana* Fundação Calouste Gulbenkian.
- Pettitt, A. N. and Bin Daud, I. (1989). Case-weight measures of influence for proportional hazards regression. *Applied Statistics* **38**, 51-67.
- Rodriguez, R. N. (1977). A Guide to the Burr Type XII distributions. *Biometrika*, **64**(1) 129-134.
- St. Laurent, R. T. and Cook, R. D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the American Statistical Association*, **87** 985-990.
- Therneau, T.M., Grambsch, P.M and Fleming, T.R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147-60.

- Thomas, W. and Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics* **32**, 59-65.
- Wei, B., Hu, Y. Q. and Fung, W. K. (1998). Generalized leverage and its applications. *Scandinavian Journal of statistics* **25**, 25-37.
- Williams, D.A. (1987). Generalized linear model diagnostic using the deviance and single case deletion. *Applied Statistics* **36**, 181-191.
- Xie, F. e Wei, B. (2007). Diagnostics analysis for log-Birnbaum-Saunders regression models. *Computational Statistics & Data Analysis*. In Press.
- Zimmer, W.J., Keats, J. B. e Wang F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology* **30**, 389-394.