UNIVERSIDADE DE SÃO PAULO

An useful reparametrization for the survival function considering an exponential regression model

JORGE ALBERTO ACHCAR
VERA LÚCIA DAMACENO

Nº 7

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

An useful reparametrization for the survival function considering an exponential regression model

JORGE ALBERTO ACHCAR

VERA LÚCIA DAMACENO

Nº 7

N O T A S D O I C M S C

Série Estatística

São Carlos

Dez. /1993

An Useful Reparametrization for the Survival Function Considering an Exponential Regression Model

Jorge Alberto Achcar*
Vera Lúcia Damaceno†
Universidade de São Paulo, Brasil

Abstract

Considering an exponential regression model for lifetime data with one covariate x proposed by Feigl and Zelen (1965), we explore a modified form of Guerrero and Johnson (1982) transformation $\phi_{GJ}(\lambda)$ for proportions, to get very accurate inferences on the survival function at time t_0 and with a specified value x_0 for the covariate. We use the standardized third derivative of the logarithm of the likelihood function to find the appropriate value for λ in the transformation $\phi_{GJ}(\lambda)$. We also use the t-plot of Hills and Smith (1993) to verify the adequability of the proposed reparametrization.

Keywords: Exponential regression model, reparametrization, accurate inferences.

*ICMSC - USP, São Carlos, S.P., C.P. 668, 13560-970, São Carlos, S.P., Brasil.
†DEMAT - UFMA, São Luis, MA, Brasil

1 Introduction

Usually, in many lifetime data applications, we have dependence of lifetime on concomitant variables, and it is important to consider the relationship of lifetime to other factors. One way to do this is through regression models. In this situation, the literature presents many nonparametric regression models (see for example, Buckley and James, 1979; Cox, 1972; Koul, Susarla and Van Ryzin, 1981. or Schmee and Hahn, 1979), and many parametric regression models (see for example. Feigl and Zelen, 1965; or Zippin and Armitage, 1966).

Among these regression models, two families of models are very popular, especially in medical applications: the proportional hazards models for the lifetimes T and the location-scale models for the logarithms of T (see for example, Lawless, 1982).

A special case which belongs to these two classes of regression models is given by the Feigl and Zelen (1965) model assuming an exponential distribution for T.

Considering the Feigl and Zelen regression model and if the parameter of interest is given by the survival function $R(t_0) = P(T > t_0)$ at a specified time t_0, usually the statistician considers asymptotical results for the maximum likelihood estimator of $R(t_0)$ or in the Bayesian approach, numerical and approximate methods to get posterior moments or posterior densities for $R(t_0)$ (see for example, Naylor and Smith. 1982; Tierney and Kadane, 1986; or Tierney, Kass and Kadane, 1989).

Usually, the accuracy of these numerical or approximate methods depend on appropriate reparametrizations, especially for small sample sizes, or when $R(t_0)$ is close to zero or one. To improve these approximate results. we could search for an appropriate transformation of $R(t_0)$ to improve the "normality" of the likelihood function or posterior density (see for example, Anscombe, 1964: or Sprott, 1973; 1980).

In this paper, we explore a transformation for proportions proposed by Guerrero and Johnson (1982), to get accurate inference results for $R(t_0)$, considering the classical or Bayesian approaches.

We illustrate the proposed methodology considering a lifetime data set introduced by Feigl and Zelen (1965).

2 The Feigl and Zelen Regression Model

Let $T_1, T_2, \ldots, T_n$ be the survival times of a sample of size n assumed to be independent identically distributed random variables with an exponential distribution with density,

$$f(t;\alpha_x) = \frac{1}{\alpha_x} exp\left\{-\frac{t}{\alpha_x}\right\} \qquad (1)$$

where $t > 0, \alpha_x = \theta_1 e^{\theta_2 x}$ and x is a covariate associated to t (see Feigl and Zelen, 1965).

The survival function at time t_0 is given by

$$R(t_0) = P(T > t_0) = exp\left\{ - \frac{t_0}{\theta_1 e^{\theta_2 x_0}} \right\} \tag{2}$$

where t_0 is fixed and x_0 is a specified value for the covariate.

With this model, the likelihood function for θ_1 and θ_2 is given by

$$L(\theta_1, \theta_2) = \theta_1^{-n} exp\left\{ -n\bar{x}\theta_2 - \frac{1}{\theta_1} \sum_{i=1}^{n} t_i e^{-\theta_2 x_i} \right\} \tag{3}$$

where $n\bar{x} = \sum_{i=1}^{n} x_i$.

The maximum likelihood estimator for θ_1 is given by

$$\hat{\theta}_1 = \left(\sum_{i=1}^{n} t_i e^{-\hat{\theta}_2 x_i} \right) / n \ , \tag{4}$$

and the maximum likelihood estimator for θ_2 satisfies the equation,

$$n\bar{x} \sum_{i=1}^{n} t_i e^{-\hat{\theta}_2 x_i} \ = \ \sum_{i=1}^{n} t_i x_i e^{-\hat{\theta}_2 x_i} \ . \tag{5}$$

which is obtained by using an iterative procedure (for example, the Newton-Raphson method).

Inferences on θ_1 and θ_2, usually are obtained by using the asymptotical normality of the likelihood function,

$$\left(\hat{\theta}_1, \hat{\theta}_2 \right) \overset{a}{\sim} N\left\{ (\theta_1, \theta_2), I^{-1} \right\} \ , \tag{6}$$

where I is the Fisher information matrix.

For inferences on the survival function at time t_0 with a fixed value x_0 for the covariate, we could consider the "delta method", (see for example, Miller, 1981) to get from (6) an asymptotical normal distribution for $\hat{R}(t_0) = exp\left\{ - \frac{t_0}{\hat{\theta}_1 e^{\hat{\theta}_2 x_0}} \right\}$

3 A Reparametrization for $R(t_0)$

To improve the accuracy of the approximate inferences for $R = R(t_0)$ considering the classical or Bayesian approaches, we explore a transformation for proportions introduced by Guerrero and Johnson (1982), given by

$$\phi^*_{GJ}(\lambda) = \left\{ \left(\frac{R}{1-R} \right)^{\lambda} - 1 \right\} / \lambda \ . \tag{7}$$

For a given value of λ, we consider a modified form of the transformation (7) given by

$$\phi_{GJ}(\lambda) = \left(\frac{R}{1-R}\right)^{\lambda} - 1 \,, \tag{8}$$

which should not produce different results as considering (7).

The transformation (8) has an inverse given by

$$R = \frac{(\phi_{GJ}+1)^{1/\lambda}}{1+(\phi_{GJ}+1)^{1/\lambda}} \,. \tag{9}$$

Assuming the exponential regression model (1) with θ_2 known, the logarithm of the likelihood function for $\phi_{GJ}(\lambda)$ is (from (3)) given by,

$$l(\phi_{GJ}) \propto n\, lnB(\phi_{GJ}) - \frac{B(\phi_{GJ})}{t_0} \sum_{i=1}^{n} t_i e^{-\theta_2(x_i - x_0)} \,, \tag{10}$$

where $B(\phi_{GJ}) = ln\left[1+(\phi_{GJ}+1)^{-1/\lambda}\right]$ and x_0 is the covariate value associated to the survival function $R(t_0)$.

observe that, in parametrization (8), we should find an appropriate value for λ that gives close "normality" for the likelihood function or posterior density for ϕ_{GJ}. One way to do this, is to search for a value of λ such that the third derivative of the logarithm of the likelihood function $l(\phi_{GJ})$, locally at the maximum likelihood estimator $\hat{\phi}_{GJ}(\lambda)$ in a standardized form.

$$STD\left(\hat{\phi}_{GJ}(\lambda)\right) = \left|l'''\left(\hat{\phi}_{GJ}(\lambda)\right)\left(-l''\left(\hat{\phi}_{GJ}(\lambda)\right)\right)^{-3/2}\right| \tag{11}$$

is equal to zero (see for example, Kass and Slate, 1992; or Sprott, 1973).

At the maximum likelihood estimator for $\phi_{GJ}(\lambda)$ with θ_2 known.

$$\hat{\phi}_{GJ}(\lambda) = \left\{exp\left[\frac{nt_0}{\sum_{i=1}^{n} t_i e^{-\theta_2(x_i-x_0)}}\right] - 1\right\}^{-\lambda} - 1 \,, \tag{12}$$

we have from (11),

$$STD\left(\hat{\phi}_{GJ}(\lambda)\right) = \left|n^{-1/2}\left(2 - \frac{3B\left(\hat{\phi}_{GJ}(\lambda)\right)B''\left(\hat{\phi}_{GJ}(\lambda)\right)}{\left[B'\left(\hat{\phi}_{GJ}(\lambda)\right)\right]^2}\right)\right| \tag{13}$$

Assuming both ϕ_{GJ} and θ_2 unknown, we could search for the appropriate value for λ such that the "profile" likelihood function for $\phi_{GJ}(\lambda)$ (see for example, Barndorff-Nielsen, 1983) is close to "normality".

We also could check the adequability of the proposed reparametrization by using the t-plot (see Hills and Smith, 1993) $T(\phi_{GJ})$ against some values of ϕ_{GJ}, where

$$T(\phi_{GJ}) = sgn\left(\phi_{GJ} - \hat{\phi}_{GJ}\right)\left\{-2l(\phi_{GL}) + 2l\left(\hat{\phi}_{GJ}\right)\right\}^{1/2} \tag{14}$$

and $\hat{\phi}_{GJ}$ is the maximum likelihood estimator for $\phi_{GJ}(\lambda)$.

4 An Example

Consider the leukemia survival data introduced by Feigl and Zelen (1965) considering a concomitant variable WBC, the white blood cell count of a patient. They consider the model introduced in section 2, where x is the natural logarithm of the white blood cell count measured in units of 10000.

Thus θ_1 represents the mean survival time of a patient with a white blood cell count of 10000 and θ_2 represents the approximate percentage change in mean survival time corresponding to an one percent increase in the white blood cell count. Their sample consisted of patients classified as AG positive or AG negative based on examination of the leukemia cells.

In table 1, we have the data of $n = 17$ positive patients. We want to find inferences for the two year ($t_0 = 96$ weeks) survival probability of patients with a white blood cell count of 50000 (see (2)) given by $R = g(\theta_1, \theta_2) = exp\left\{-\frac{96}{\theta_1 5^{\theta_2}}\right\}$.

$WBC/_{10000}$	Survival times (Weeks)
0.230	65
0.075	156
0.430	100
0.260	134
0.600	16
1.050	108
1.000	121
1.700	4
0.540	39
0.700	143
0.940	56
3.200	26
3.500	22
10.000	1
10.000	1
5.200	5
10.000	65

Table 1. Feigl and Zelen survival time data (AG positive)

The logarithm of the likelihood function for R and θ_2, obtained from (3), is given by

$$l(R,\theta_2) \propto 17ln(-lnR) + 23.6052\,\theta_2 + \\ + \frac{lnR}{96}\sum_{i=1}^{17} t_i\,(5c^{-x_i})^{\theta_2}. \tag{15}$$

The maximum likelihood estimators for R and θ_2 are given by $\hat{R} = 0.02555$ and $\hat{\theta}_2 = -0.48183$, respectively. Considering the asymptotical normality of the maximum likelihood estimators for R and θ_2 based on the observed Fisher information matrix, approximate 95% confidence intervals for θ_2 and R are given by (- 0.72316 ; - 0.23634) and (- 0.01899 ; 0.07009), respectively.

In figure 1, we observe that the contour plot for the likelihood function of R and θ_2 does not present an eliptical form, which indicates bad "normality" for the likelihood function.

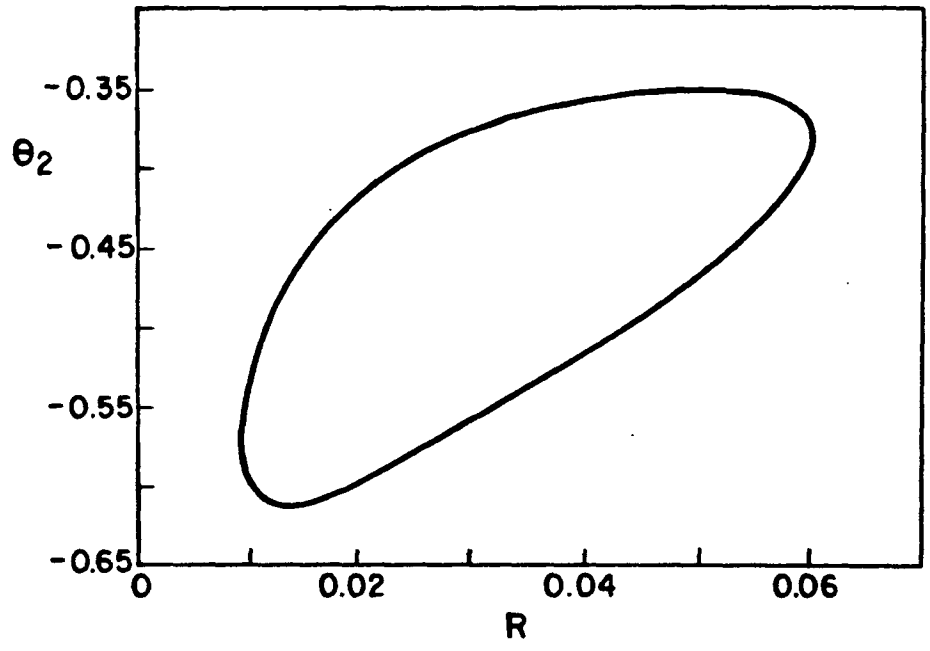


Figure 1: Contour plot for the likelihood function of R and θ_2.

In figure 2, we have the plot of the likelihood function for R assuming $\theta_2 = -0.48183$ known. We observe a very asymmetrical curve, which indicates a bad "normality" for the likelihood function.

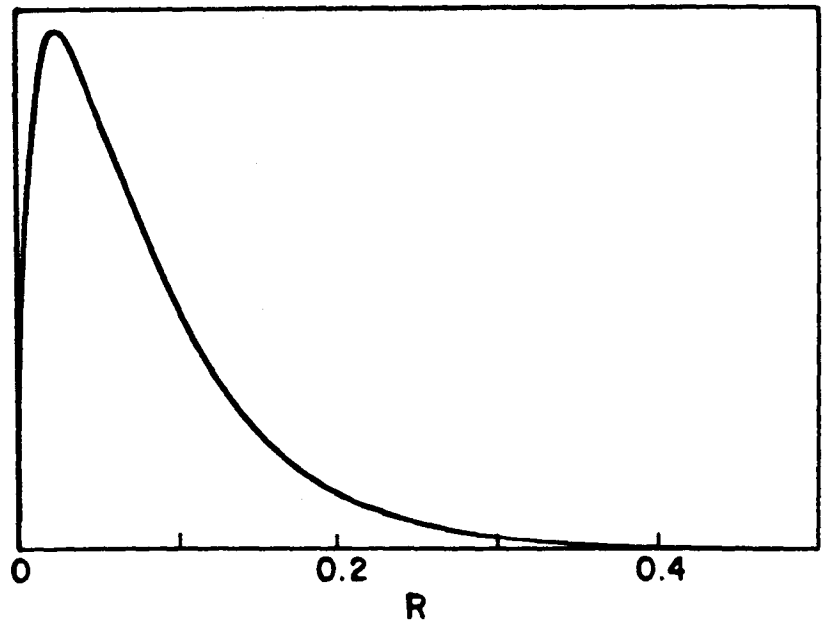


Figure 2: Likelihood function for $R(96)$ assuming θ_2 known

Considering the transformation (8) and assuming $\theta_2 = -0.48183$ known, the value of λ such that $STD(\widehat{\phi}_{GJ}) = 0$ (see (13)), is given by $\lambda = 0.1516$, which corresponds to $\widehat{\phi}_{GJ}(0.1516) = -0.4242$.

In figure 3, we observe a good "normality" for the likelihood function considering the reparametrization $\phi_{GJ}(0.1516)$.

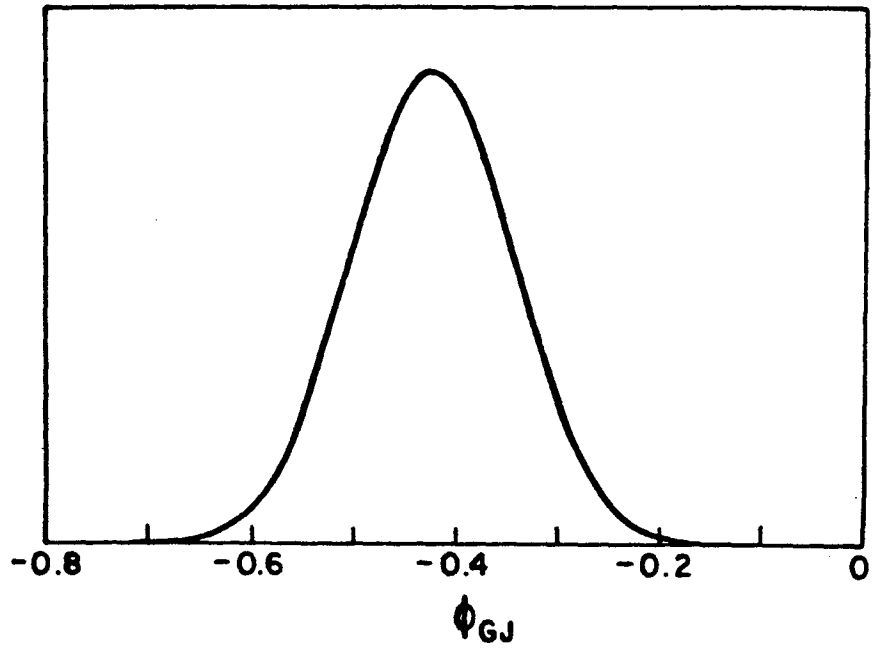


Figure 3: Likelihood function for $\phi_{GJ}(0.1516)$ assuming θ_2 known

We also could check the "normality" of the likelihood function in the parametrization $\phi_{GJ}(0.1516)$ considering the t-plot (see (14)) proposed by Hills and Smith (1993).

In figure 4, we observe a close linearity of the t-plot, indicating by the "normality" of the likelihood function for $\phi_{GJ}(0.1516)$. In the original parametrization, the t-plot is markedly curved (see figure 4), which indicates the "nonnormality" of the likelihood function for $R(96)$.

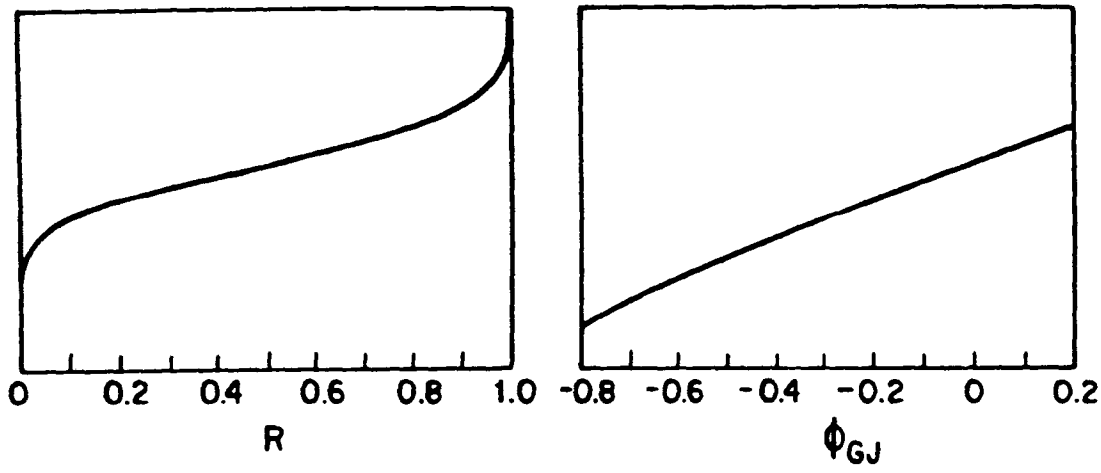


Figure 4: t-plot of Hills and Smith for $R(96)$ and $\phi_{GJ}(0.1516)$

Considering the asymptotical normality of the maximum likelihood estimator $\hat{\phi}_{GJ}(0.1516)$ with θ_2 known, an approximate 95% confidence interval for ϕ_{GJ} is given by $(-0.58037; -0.26806)$, which corresponds to a better 95% confidence interval for $R(96)$ given by (0.00324 ; 0.11321).

In figure 5, we observe an eliptical form for the joint likelihood function for $\phi_{GJ}(0.1516)$ and θ_2, which indicates good joint "normality" for the likelihood function.

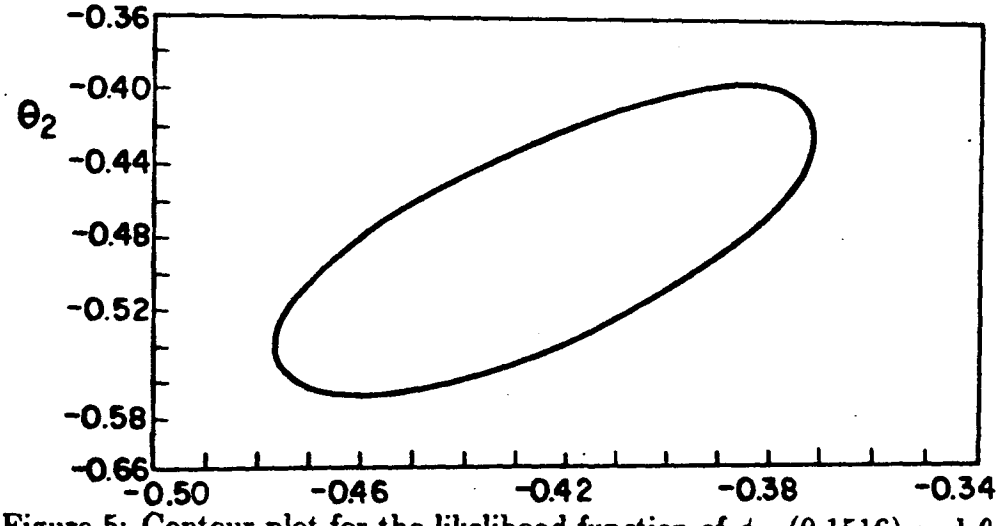


Figure 5: Contour plot for the likelihood function of $\phi_{GJ}(0.1516)$ and θ_2

5 Discussion and Conclusions

The use of reparametrization (8), could be of great practical interest, since we get very simple and accurate inference results for the survival function at time t_0 based on the asymptotical normality of the maximum likelihood estimators, as it was observed considering the Feigl and Zelen regression model. In the same way, we could consider the reparametrization (8) for the survival function at time to assuming other parametrical regression models for the lifetimes of the patients.

Assuming θ_2 unknown, we could consider the "profile" likelihood function for $\phi_{GJ}(\lambda)$, and find λ in a similar way as it was considered assuming θ_2 known. It is important, to point out that since,

$$E\left(\frac{-\partial^2 l}{\partial\phi_{GJ}\partial\theta_2}\right) = -\frac{nB'(\phi_{GJ})(\bar{x}-x_0)}{ln\left[1+(\phi_{GJ}+1)^{-1/\lambda}\right]}\,, \tag{16}$$

where $B(\phi_{GJ})$ is given in (10) and if we have a design such that $\bar{x} = x_0$, the Fisher information matrix is diagonal and asymptotically, the inferences on $\phi_{GJ}(\lambda)$ would not be affected by θ_2.

The proposed reparametrization also could be useful to find accurate Bayesian inferences for the parameters of interest, since we usually consider nemerical or approximation methods to find posterior moments or marginal posterior densities, when it is not possible to get exact analytical solutions (see for example, Naylor and Smith, 1982; Tierney and Kadane, 1986; Tierney, Kass and Kadane, 1989; or Leonard, Hsu and Tsui, 1989). These numerical or approximation methods, in general, depend on appropriate reparametrizations to find accurate results (see for example, Achcar and Smith, 1990).

We also get similar results considering censored lifetime data.

References

ACHCAR, J.A.; SMITH, A.F.M. (1990). Aspects of Reparametrization in Aproximate Bayesian Inference. *Bayesian and Likelihood Methods in Statistics and Econometrics.* Nort-Holland, Elsevier Science Publishers B. V., Amsterdam, North - Holland.

ANSCOMBE, F.J. (1964). Normal Likelihood Functions,*Ann. Inst. Stat. Math.*, 6, 1-19.

BARNDORFF - NIELSEN, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator, Biometrika, 70, 343-365.

BUKLEY, J.; JAMES, I. (1979). Linear Regression With Censored Data. *Biometrika*, 66, 3. 429-436.

COX, D.R. (1972). Regression Models and Life Tables (With Discussion). *J. R. Stat. Soc., 27, 4-9.*

FEIGL, P.; ZELEN, M. (1965). Estimation of Exponential Survival Probabilities With Concomitant Information. *Biometrics*, 21, 826-837.

GUERRERO, V.M.; JOHNSON, R.A. (1982). Use of the Box-Cox Transformation With Binary Response Models, *Biometrika*, 69, 309-314.

HILLS, S.E.; SMITH, A.F.M. (1993). Diagnostics of Posterior Nonnormality in Bayesian Inference, *Biometrika*, 80, 1, pp, 61-74.

KASS, R.E.; SLATE, E.H. (1992). Reparametrization and Diagnostics of Posterior Nonnormality *in Bayesian Statistics* 4, Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. 289-306. Oxford University Press.

KOUL, H. SUSARLA, V.; RYZIN, J.V. (1981). Regression Analysis With Randomly Right-Censored Data, *The Annals of Statistics*, 9, 6, 1726-1285.

LAWLESS, J.F. (1982). Statistical models and methods for lifetime data. John Wiley & Sons.

LEONARD, T. ; HSU, J.S.J. ; TSUI, K.W. (1989). Bayesian Marginal Inference. *Journal of the American Statistical Association.* 84, 408, 1051-1058.

MILLER, R.G. (1981). *Survival Analysis.* New York, John Wiley & Sons.

NAYLOR, J.C. ; SMITH, A.F.M. (1982). Applications of a Method for the Efficient Computation of Posterior Distributions, *Applied Statistics*, 31. 214-225.

SCHMEE, J. ; HAHN, G.J. (1979). A Simple Method for Regression Analysis With Censored Data. *Technometrics*, 21, 4, 417-432.

SPROTT, D.A. (1973). Normal Likelihoods and Their Relation to Large Sample Theory of Estimation. *Biometrika*, 60, 30, 457-465.

SPROTT, D.A. (1980). Maximum Likelihood in Small Samples: Estimation in the Presence of Nuisance Parameters, *Biometrika*, 67, 515-523.

TIERNEY, L. ; KADANE, J.B. (1986). Accurate Aproximations for Posterior Moments and Marginal Densities, *Journal of The American Statistical Association*, 81, 82-86.

TIERNEY, L. ; KASS, R.E. ; KADANE, J.B. (1989). Aproximate Marginal Densities for Nonlinear Functions, *Biometrika*, 76, 425-433.

ZIPPIN, C. ; ARMITAGE, P. (1966). Use of Concomitant Variables and Incomplet Survival Informations in the Estimation of an Exponential Survival Parameter, *Biometrics*, 22, 665-672.

N O T A S D O I C M S C

Serie Estatistica

Nº 006/93 - RODRIGUES, J. - Bayesian estimation of a normal mean parameter using linex loss function and robustness considerations

Nº 005/93 - ACHCAR, J.A.; FOGO, J.C. - Some useful reparametrizations for the reliability function considering the exponential distribution and censored lifetime data

Nº 004/93 - LEME, E.C. - A m.v.u. estimator vs a m.l. estimator: a way of comparing their precisions by using the Pareto Distribution

Nº 003/93 - ACHCAR, J.A. - Some practical aspects of approximate Bayesian inference

Nº 002/93 - RODRIGUES, J.; BABA, M.Y. - Bayesian estimation of a simple regression model with measurement errors

Nº 001/93 - ACHCAR, J.A. - Some aspects of reparametrization for statistical model