ILL: 13425240

Call Number: QA276.A1 P35

Location: General

Maxcost: $21.25IFM

DateReq: 10/18/2005 ☐ Yes

Date Rec: 10/19/2005 ☐ No

☐ Conditional

Borrower: ORE

Affiliation: OCLC Western, GWLA (BTP), ORBIS

Request Type:

Source: ILLiad

LenderString: *IWA,CRL,REB,UIU,WAU

OCLC Number: 14174022

DueDate:

Verified: <TN:156849> OCLC 1012-9367

Email: valley.ill@oregonstate.edu

Fax: 541-737-1328

Billing Notes:

Title: Pakistan journal of statistics.

Uniform Title:

Author:

Edition:

Imprint: Lahore : Social Sciences Research Centre, University of the Punjab, [1985?-

Article: Achcar, J. A.: Some aspects of reparametrization for statistical models

Vol: 10

No.:

Pages: 597--616

Date: 1994

Dissertation:

Borrowing Notes:

ShipTo: Library-ILL/Oregon State University/121 The Valley Library/Cross Streets Jefferson Way & Waldo Pl/Corvallis, OR 97331-4501

Ship Via: ARIEL: OSU-ILL.library.orst.edu when possible

ShipVia: ARIEL: OSU-I

Return To:

Interlibrary Loan
Room 198B Parks Library
Iowa State University
Ames, IA 50011-2140

Ship To:

Library-ILL
Oregon State University
121 The Valley Library
Cross Streets Jefferson Way & Waldo Pl
Corvallis, OR 97331-4501

NeedBy: 11/17/2005

Borrower: ORE

ILL: **13425240**

Lender: IWA

Req Date: **10/18/2005**

OCLC #: 14174022

Patron: **Bulatov, Yaroslav**

Author:

Title: **Pakistan journal of statistics.**

Article: Achcar, J. A.: Some aspects of reparametrization for statistical models

Vol.: 10

No.:

Date: 1994

Pages: 597--616

Verified: <TN:156849> OCLC 1012-9367

Maxcost: $21.25IFM

Due Date:

Lending Notes:

Bor Notes:

EK 10/19

Pak. J. Statist.
1994 Vol 10(3), pp 597-616

SOME ASPECTS OF REPARAMETRIZATION FOR STATISTICAL MODELS

J. A. Achcar
Universidade de São Paulo
ICMSC-USP, C. Postal 668
13560-970, São Carlos, S.P., Brazil

(Received: Jan, 1993 Accepted: Feb, 1994)

ABSTRACT

We show in some selected applications, the effect of parametrization in the accuracy of asymptotic results based on the normality of the likelihood function. The normality of the likelihood function or posterior density of interest can be improved by appropriate choice of parametrization. We also use some diagnostic of nonnormality recently proposed by Kass and Slate (1992).

KEY WORDS

Reparametrization, normality of likelihood function or posterior density, accuracy of asymptotics.

1. INTRODUCTION

The performance of numerical techniques or the accuracy of asymptotical results based on the limiting normality of maximum likelihood estimators (see for example, Kass and Slate, 1992) usually could be affected by a choice of appropriate parametrization. In the Bayesian approach, the integration methods can be improved by parameter transformation in any of the existing procedures. For example, using numerical methods (see for example, Naylor and Smith, 1982) or using approximation methods (see for example, Tierney and Kadane, 1986).

One important problem to all statisticians is to find an one-to-one transformation from θ to ϕ such that the likelihood function or posterior density of ϕ is better behaved than in the parametrization θ.

Parametrization is a topic of extensive research in the last years. For example, the orthogonal parametrization, proposed by Cox and Reid (1987); the parametrization to have "normality" of the likelihood function (see for example, Anscombe, 1964; Sprott, 1973, 1980); or the parametrization to have "normality" of posterior densities (see Kass and Slate, 1992). To improve the "normality" of the likelihood function, we could consider:

(i) The third derivatives of the logarithm of the likelihood function locally in the maximum likelihood estimators are close to zero (see Anscombe, 1964).

(ii) The expected values of the second derivatives of the logarithm of the likelihood function should be constant (see Sprott, 1973, 1980). In the Bayesian context, this means a parametrization with a locally uniform Jeffreys prior density (see for example, Box and Tiao, 1973).

Other approaches exist for determining the "closeness to normality" of the likelihood functions or posterior distributions of a particular set of parameters. These include curvature measures (see for example, Bates and Watts, 1988) or graphical methods which can be used for both assessing a proposed parametrization and choosing a new improved parametrization (see Hills and Smith, 1993).

In this paper, we present some selected applications, where the effect of reparametrization to improve normality of the likelihood function or the posterior density is evaluated by checking density plots, contours plots (in two-dimensional problems) or using some diagnostics to nonnormality of likelihood functions or posterior densities proposed by Kass and Slate (1992), especially for high dimensional problems.

2. A USEFUL REPARAMETRIZATION

First, let us consider the one-parameter case, where we want to find an one-to-one transformation from θ to ϕ that yields approximate normality for the likelihood function. That is, we search for a reparametrization ϕ such that the expected value of the second derivative of the logarithm of the likelihood function is constant (see Sprott, 1973). With this parametrization ϕ, we have a constant Fisher information, that is, in the Bayesian context, we have a locally uniform Jeffreys prior density for ϕ (see for example, Box and Tiao, 1973).

In this one-to-one transformation, we have,

$$\pi(\theta) \propto \text{constant} \ \left|\frac{d\phi}{d\theta}\right| \ . \qquad (2.1)$$

Since the Jeffreys noninformative prior density in the original parametrization θ is given by

$$\pi(\theta) \propto \{I(\theta)\}^{1/2} \qquad (2.2)$$

where $I(\theta)$ is the Fisher information, we should find ϕ by solving the differential equation

$$\phi \propto \int \{I(\theta)\}^{1/2}\, d\theta \qquad (2.3)$$

In the two-parameter case, we search for an one-to-one transformation from (θ_1, θ_2) to (ϕ_1, ϕ_2) such that in the parametrization ϕ_1 and ϕ_2 we have a constant Fisher information matrix. In the Bayesian context, we have a locally uniform Jeffreys prior density for ϕ_1 and ϕ_2.

Observe that, in the original parametrization θ_1 and θ_2, the Jeffreys prior density is given by,

$$\pi(\theta_1, \theta_2) \propto \{det I(\theta_1, \theta_2)\}^{1/2} \qquad (2.4)$$

where $I(\theta_1, \theta_2)$ is the Fisher information matrix.

Therefore, we should search for an one-to-one transformation of (θ_1, θ_2) to (ϕ_1, ϕ_2) such that,

$$\left\{\frac{\partial\phi_1}{\partial\theta_1}\frac{\partial\phi_2}{\partial\theta_2} - \frac{\partial\phi_1}{\partial\theta_2}\frac{\partial\phi_2}{\partial\theta_1}\right\} \propto \text{constant} \ \{\det I(\theta_1, \theta_2)\}^{1/2} \qquad (2.5)$$

with

$$E\left(-\frac{\partial^2\ell}{\partial\phi_1^2}\right) = \text{constant} \ ,$$

$$E\left(-\frac{\partial^2\ell}{\partial\phi_2^2}\right) = \text{constant} \quad and \quad E\left(-\frac{\partial^2\ell}{\partial\phi_1\,\partial\phi_2}\right) = \text{constant} \ ,$$

where ℓ is the logarithm of the likelihood function for ϕ_1 and ϕ_2 .

In a similar way, we could find a reparametrization in the multiparameter case, but in general we have great difficulties to solve a similar differential equation (2.5) to find an appropriate reparametrization. In practical work, we usually try different parametrizations and using the diagnostic measures to nonnormality proposed by Kass and Slate (1992), we could decide by the best parametrization in any multiparameter model.

3. SOME DIAGNOSTIC MEASURES OF NONNORMALITY

3.1 One-parameter Case Kass and Slate (1992) proposes some diagnostic measures of nonnormality of posterior densities or likelihood functions. One of these diagnostic measures to check the nonnormality of the likelihood function in a parametrization θ is given by the standardized third derivative,

$$\mathrm{STD} = \left| \ell'''(\hat{\theta}) \cdot \left[-\ell''(\hat{\theta}) \right]^{-3/2} \right| \qquad 3.1$$

where $\ell(\theta)$ is the logarithm of the likelihood function and $\hat{\theta}$ is the maximum likelihood estimator (see Sprott, 1973).

The importance of the standardized third derivative givrn by the STD in (3.1) is that it is invariant to affine transformations of the parameters.

In a similar way, we could substitute ℓ by ℓ_p in (6) and $\hat{\theta}$ by $\tilde{\theta}$, where ℓ_p is the logarithm of the posterior density for θ, and $\tilde{\theta}$ is the mode of the posterior density for θ.

Other diagnostic measurres of nonnormality are given by interval probabilities or the Kullback-Leibler number (see for example, Kass and Slate, 1992).

3.2 Multiparameter Case In the two-parameter case, we usually could check the normality of the likelihood function or posterior density with the aid of contour plots. In situations with three or more parameters, we could use third derivatives summaries for diagnostic of global assessment of normality (see Kass and Slate, 1992), given by

$$m^2\overline{B}^2 \doteq \sum_{i,j,k,\ell,m,n} b_{ij} b_{\ell mn} b_{kn} d_{ijk} d_{\ell mn} \qquad (3.2)$$

where b_{ij} are the elements of the inverse of the information matrix, $d_{ijk} = \partial^3 \ell\left(\hat{\underset{\sim}{\theta}}\right) / \partial\theta_i\, \partial\theta_j\, \partial\theta_k$, for a $m$dimensional parameter $\underset{\sim}{\theta}$ and the summation are over all indices. Observe that this measure reduces to the square of the standardized third derivative in the one-parameter case (3.1).

In the same way, we could substitute ℓ by ℓ_p and $\widehat{\theta}$ by $\widetilde{\theta}$ in (8), where ℓ_p is the logarithm of the posterior density for θ and $\widetilde{\theta}$ is the mode of the posterior density.

4. SOME SELECTED APPLICATIONS

4.1 Reparametrization of the Exponential Distribution with Type I Censored Data Let $T_1, T_2, \ldots, T_n$ be a random sample of size n

representing the survival times of n individuals with an exponential distribution with density

$$f(t;\theta) = \frac{1}{\theta} e^{-t/\theta} \tag{4.1}$$

where $t > 0$.

Assuming that the period of follow-up for the i^{th} individual is limited to a fixed value L_i, that is, we have type I censoring data, the observed survival time of the i^{th} individual is $t_i = \min(T_i, L_i)$.

Defining $\delta_i = 0$ if $t_i < T_i$ (a censored observation) and $\delta_i = 1$ if $t_i = T_i$ (an observed failure), the likelihood function for θ is given by

$$L(\theta) = \prod_{i=1}^{n} f^{\delta_i}(t_i,\theta) S^{1-\delta_i}(t_i,\theta) \tag{4.2}$$

where $S(t_i,\theta) = \exp(-t_i/\theta)$ is the survival function of the exponential distribution.

That is,

$$L(\theta) = \theta^{-r} \exp\left\{ -\sum_{i=1}^{n} t_i/\theta \right\} , \tag{4.3}$$

where $r = \sum_{i=1}^{n} \delta_i$ is the observed number of failures.

The Fisher information is given by

$$I(\theta) = E\left\{ -\frac{d^2 \log L}{d\theta^2} \right\} = \frac{Q}{\theta^2} \tag{4.4}$$

where $Q = \sum_{i=1}^{n} (1 - e^{-L_i/\theta})$ (see for example, Lawless, 1982).

The maximum likelihood estimator for θ is given by $\widehat{\theta} = T/r$, where $T = \sum_{i=1}^{n} t_i$. Usually, we get inferences on θ based on the asymptotic normality of $\widehat{\theta}$, $\widehat{\theta} \overset{a}{\sim} N\{\theta; I^{-1}(\theta)\}$, where $I(\theta) = Q\big/\theta^2$.

To improve the accuracy of this asymptotic distribution, we could consider different parametrizations. For example, we could consider the parametrization $\phi_1 = \theta^{-1/3}$, where $\ell'''(\widehat{\phi}_1) = 0$, $\ell(\phi_1)$ is the logarithm of the likelihood function and $\widehat{\phi}_1$ is the maximum likelihood estimator. In this parametrization, we have,

$$\widehat{\phi}_1 \overset{a}{\sim} N\left\{\phi_1, \phi_1^2\big/9Q\right\} . \tag{4.5}$$

Another reparametrization is given by $\phi_2 = \ln\theta$. This parametrization is a variance-stabilizing transformation (see Kass and Slate, 1992; or Achcar, 1990), so that the Jeffreys prior density is uniform on ϕ_2 for uncensored data (approximate for censored data). In this parametrization, we have

$$\widehat{\phi}_2 \overset{a}{\sim} N\{\phi_2, Q^{-1}\} \tag{4.6}$$

where $Q = \sum_{i=1}^{n} \left(1 - e^{-L_i e^{-\phi_2}}\right)$.

As a numerical illustration, consider the type I censoring data of table 1.

i	1	2	3	4	5	6	7	8	9	10
T_i	2	–	51	–	33	27	14	24	4	–
L_i	81	72	70	60	41	31	31	30	29	21

Table 1: Type I Censored Data

From table 1, we have $r = 7$, and $T = \sum_{i=1}^{10} t_i = 308$. The maximum likelihood estimator for θ is $\widehat{\theta} = 44.0$ and $\widehat{Q} = \sum_{i=1}^{10} \left(1 - e^{-L_i/\widehat{\theta}}\right) = 6.15$ (see (4.5)). A 95% approximate confidence interval for θ is given by (9.3;78.7). In figure 1, we observe that the normal approximation for the likelihood function is not appropriate for this data set.

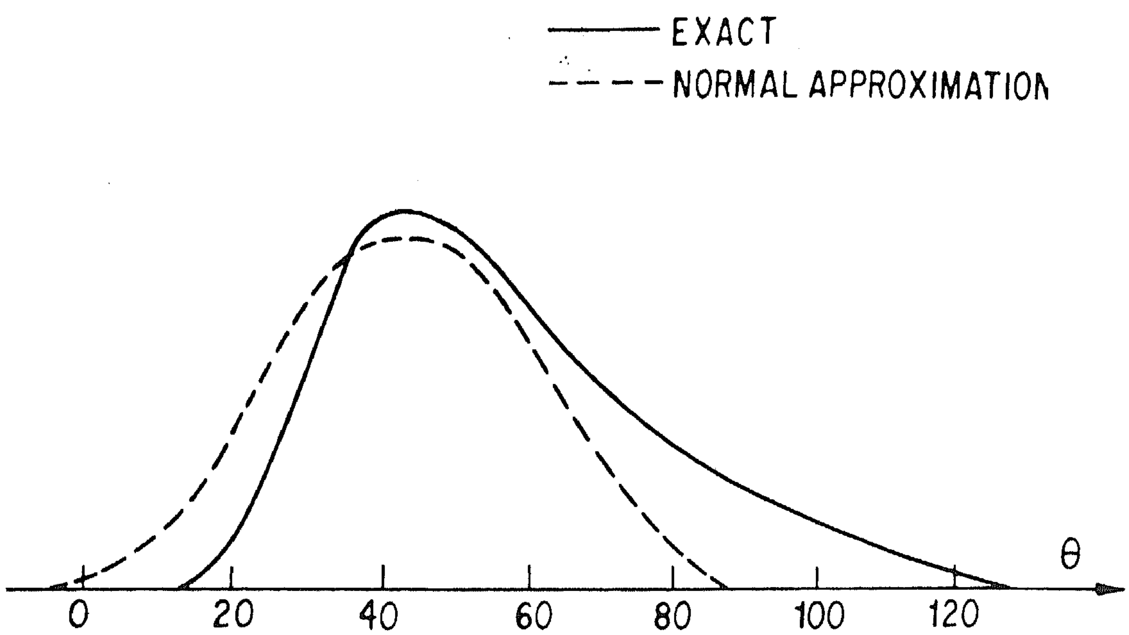

FIGURE 1: LIKELIHOOD FUNCTION FOR θ

Considering the parametrization $\phi_1 = \theta^{-1/3}$, the maximum likelihood estimator for ϕ_1 is given by $\hat{\phi}_1 = 0.2833$ and an approximate 95% confidence interval for ϕ_1 based on the asymptotical normality of $\hat{\phi}_1$ is given by (0.2088; 0.3577). From this, we get an approximate 95% confidence interval for θ given by (21.8 ; 109.9). We observe a very different confidence interval for θ in this parametrization. In figure 2, we observe a good normal approximation for the likelihood function for $\phi_1 = \theta^{-1/3}$.

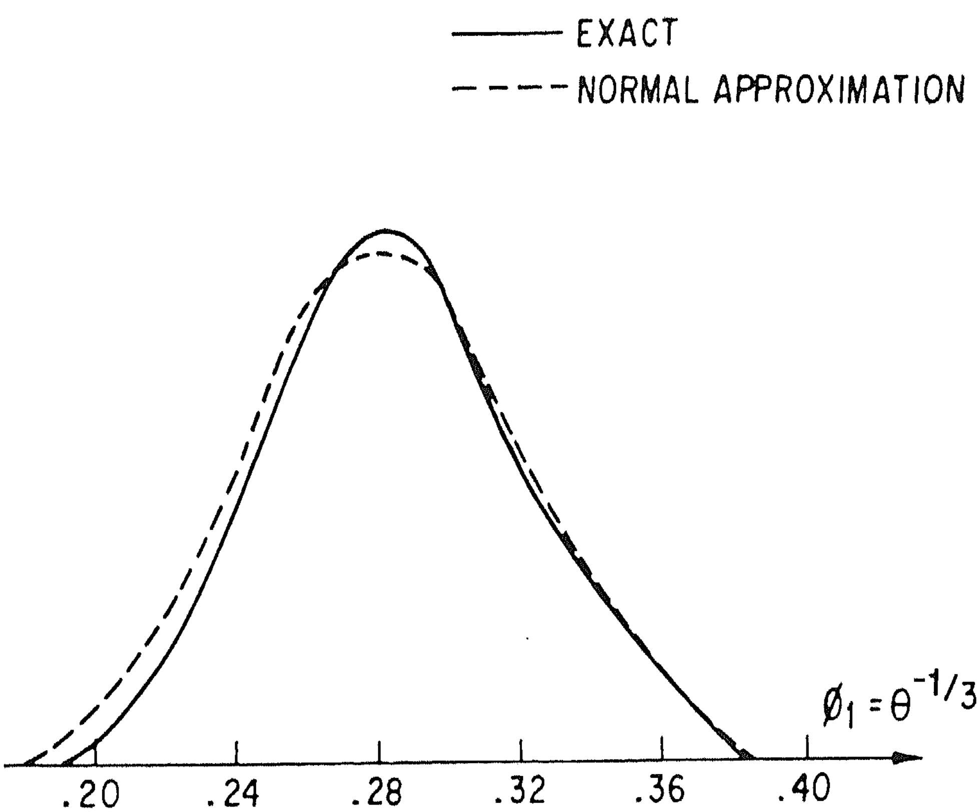

FIGURE 2: LIKELIHOOD FUNCTION FOR $\phi_1 = \theta^{-1/3}$

Considering the parametrization $\phi_2 = \ln\theta$, the maximum likelihood estimator for ϕ_2 is given by $\hat{\phi}_2 = 3.7842$ and an approximate 95% confidence interval for ϕ_2 is given by (2.9938 ; 4.5745). From this, we get an approximate 95% confidence interval for θ given by (19.9622 ; 96.9834). We observe similar confidence intervals for θ considering the parametrizations $\phi_1 = \theta^{-1/3}$ and $\phi_2 = \ln\theta$. In figure 3, we also observe a good normality for the likelihood function of $\phi_2 = \ln\theta$.

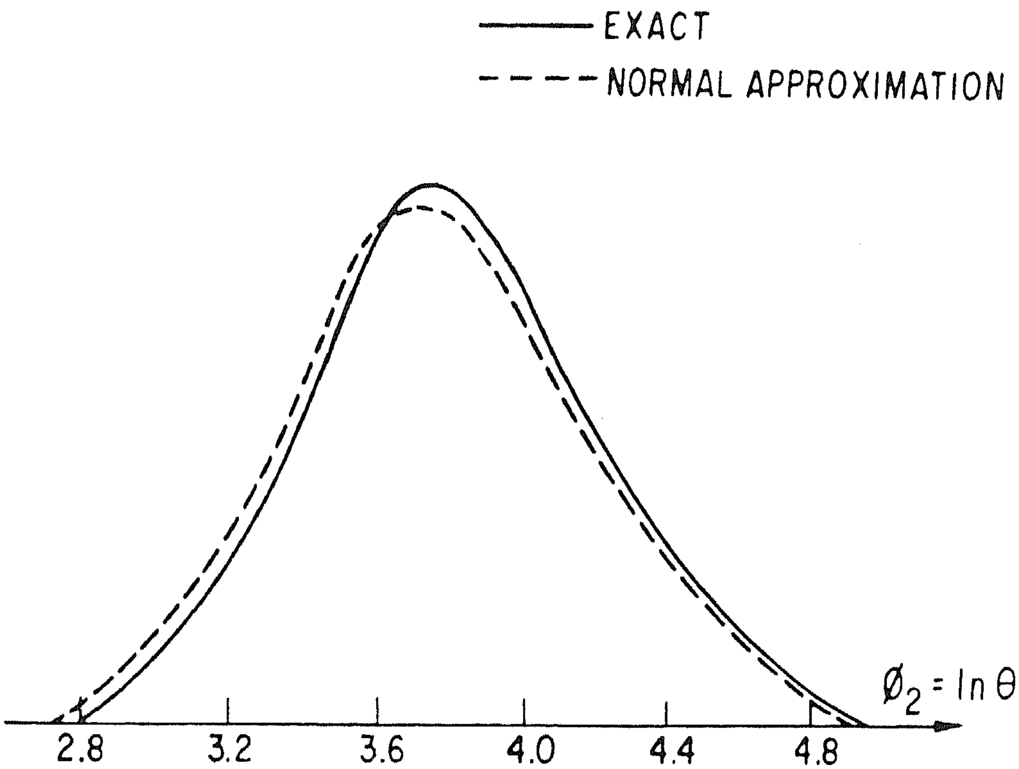

FIGURE 3: LIKELIHOOD FUNCTION FOR $\phi_2 = \ln \theta$

In fact, we can observe that the standardized third derivative of the logarithm of the likelihood function in the parametrization $\phi_2 = \ln \theta$ is small in comparison with the value in the parametrization θ (see table 2).

Parametrization	STD
θ	1.5118
$\phi_2 = \ln \theta$	0.3780
$\phi_1 = \theta^{-1/3}$	0.0000

Table 2: Standardized Third Derivatives of Log-likelihood Function

4.2 A Reparametrization for the Extreme Value Distribution

One of the most useful survival models for reliability data is given by the extreme value distribution for the logarithms of the life times of unitis submitted to a reliability experiment. This model corresponds to a Weibull distribution for the survival data.

Considering T a random variable representing the survival time of an unity, we assume the model,

$$y = \ln T = \mu + \sigma W \tag{4.7}$$

where the error term, W has an extreme value density $\exp\{w - e^w\}$, $-\infty < w < \infty$. In the Weibull form of the model, the random variable T has a Weibull density with parameters $\alpha = e^\mu$ and $\beta = 1/\sigma$.

Usually, inferences are based on the asymptotic normality of the maximum likelihood estimators for μ and σ , that cannot be accurate for small or moderate sample sizes.

Let $T_1, T_2, \ldots, T_n$ be the survival times of a random sample of size n of units submitted to a reliability experiment and assume that model (15) is appropriate. Thus, the logarithm of the likelihood function for μ and σ is given by

$$\ell(\mu, \sigma) = -n \ln \sigma + \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} e^{w_i} \tag{4.8}$$

where $w_i = (y_i - \mu)/\sigma$.

The Fisher information matrix for μ and σ is given (see for example, Lawless, 1982) by

$$I(\mu, \sigma) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2}(1 + \psi(1)) \\ \frac{n}{\sigma^2}(1 + \psi(1)) & \frac{n}{\sigma^2}\left(1 + \psi(2) + (\psi(2))^2\right) \end{pmatrix} \tag{4.9}$$

where $\psi(k)$ is the digamma function $d \ln T(k)/dk$.

To search for an appropriate reparametrization θ_1 and θ_2 to get better joint normality of the likelihood function, we can use (2.5). Since the Jeffreys prior for μ and σ is proportional to σ^{-2}, $-\infty < \mu < \infty$, $\sigma > 0$, we search for a parametrization θ_1 and θ_2 such that

$$E\left\{-\partial^2 \ell \Big/ \partial\theta_1^2\right\} = \text{constant}\ , \quad E\left\{-\partial^2 \ell \Big/ \partial\theta_2^2\right\} = \text{constant}\ ,$$

$E\left\{-\partial^2 \ell \Big/ \partial\theta_1 \partial\theta_2\right\} = \text{constant}$, and satisfies the differential equation,

$$\left\{\frac{\partial\theta_1}{\partial\mu}\frac{\partial\theta_2}{\partial\sigma} - \frac{\partial\theta_1}{\partial\sigma}\frac{\partial\theta_2}{\partial\mu}\right\} = \frac{\text{constant}}{\sigma^2} \tag{4.10}$$

For example, considering $\theta_1 = \ln\sigma$, we verify from (4.11), that θ_2 is given by the differential equation,

$$\frac{1}{\sigma}\frac{\partial\theta_2}{\partial\mu} = \frac{\text{constant}}{\sigma^2} \tag{4.11}$$

A solution for this equation is given by

$$\begin{aligned}\theta_2 &= \int \tfrac{d\mu}{\sigma} + c(\sigma) \\ &= \tfrac{\mu}{\sigma} + c(\sigma)\end{aligned} \tag{4.12}$$

where $c(\sigma)$ is an arbitrary function of σ .

Considering $c(\sigma) = a/\sigma$, where a is a constant, we have the reparametrization $\theta_1 = \ln\sigma$ and $\theta_2 = (\mu + a)/\sigma$. That is, $\sigma = e^{\theta_1}$ and $\mu = \theta_2 e^{\theta_1} - a$. The logarithm of the likelihood function for θ_1 and θ_2 is given by

$$\ell(\theta_1, \theta_2) = -n\theta_1 + \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} e^{w_i} \tag{4.13}$$

where $w_i = y_i\, e^{-\theta_1} - \theta_2 + a\, e^{-\theta_1}$.

Observe that if $a = -\widehat{\mu}$, we have $\widehat{\theta}_2 = 0$.

As a numerical illustration, consider the data of table 3 representing the voltage levels at which failures occurred when specimens were subjected to an increasing voltage stress in a laboratory experiment. The test involved 20 specimens and the failure voltages T_i were given in Kilovolts per millimeter (data in Lawless, 1982, page 189).

32.0	35.4	36.2	39.8	41.2
43.3	45.5	46.0	46.2	46.4
46.5	46.8	47.3	47.3	47.6
49.2	50.4	50.9	52.4	56.3

Table 3: Voltage Levels in Kilovolts/Millimeter

Assuming that the model (15) is appropriate, the maximum likelihood estimators for μ and σ are given by $\widehat{\mu} = 3.8666$ and $\widehat{\sigma} = 0.1066$. The third derivatives of the logarithm of the likelihood function $\ell(\mu, \sigma)$ given in (16), locally in the maximum likelihood estimates $\widehat{\mu}$ and $\widehat{\sigma}$ are given by $\partial^3\ell(\widehat{\mu}, \widehat{\sigma})/\partial\mu^3 = 16519.73$, $\partial^3\ell(\widehat{\mu}, \widehat{\sigma})/\partial\mu^2\partial\sigma = 40044.39$, $\partial^3\ell(\widehat{\mu}, \widehat{\sigma})/\partial\sigma^3 = 161667.91$ and $\partial^3\ell(\widehat{\mu}, \widehat{\sigma})/\partial\sigma^2\partial\mu = 42130.92$. We observe large values of the third derivatives of $\ell(\mu, \sigma)$ locally in $\widehat{\mu}$ and

$\hat{\sigma}$. In figure 4, we have contour plots, for the exact likelihood function considering $\ell(\mu, \sigma) = -1$ and for the normal approximation.

We observe that in this parametrization μ and σ , the normal approximation for the likelihood function is not good, and we can get bad inferences, especially for the scale parameter σ (see Achcar, 1991).

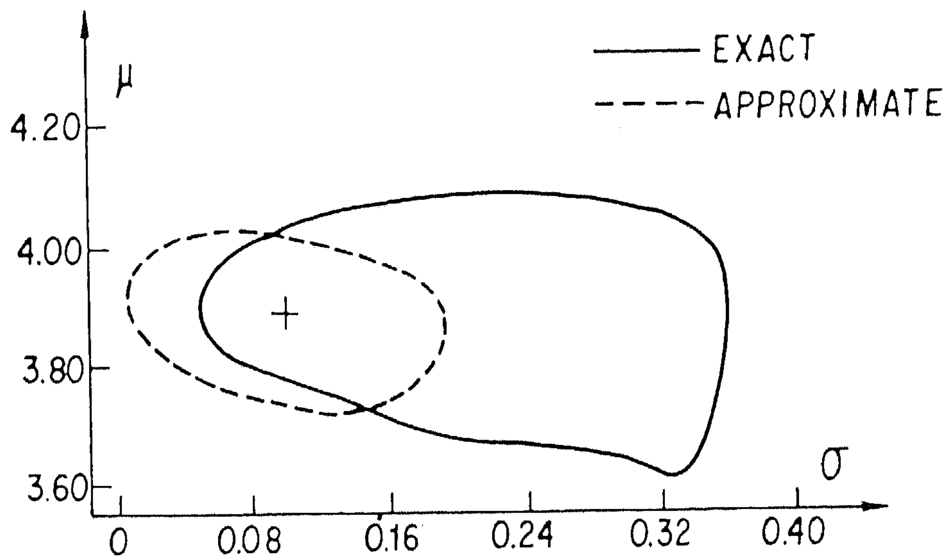

FIGURE 4: CONTOURS OF THE LIKELIHOOD FUNCTION FOR μ AND σ $(\ell(\mu, \sigma) = -1)$

Considering the data dependent reparametrization $\theta_1 = \ln \sigma$ and $\theta_2 = (\mu - 3.90)/\sigma$, the maximum likelihood estimators are given by $\hat{\theta}_1 = -2.2389$, and $\hat{\theta}_2 = 0$. From (21), we find $\partial^3 \ell(\hat{\theta}_1, \hat{\theta}_2)/\partial \theta_2^3 = 20.00$, $\partial^3 \ell(\hat{\theta}_1, \hat{\theta}_2)/\partial \theta_2^2 \partial \theta_1 = 8.4807$, $\partial^3 \ell(\hat{\theta}_1, \hat{\theta}_2)/\partial \theta_1^3 = 84.4866$ and $\partial^3 \ell(\hat{\theta}_1, \hat{\theta}_2)/\partial \theta_1^2 \partial \theta_2 = 25.5647$.

Considering $\ell(\theta_1, \theta_2) = -1$, we have in figure 5 the contour plots for the exact likelihood function for θ_1 and θ_2, and for the normal approximation. We observe very good normality for the likelihood function of θ_1 and θ_2.

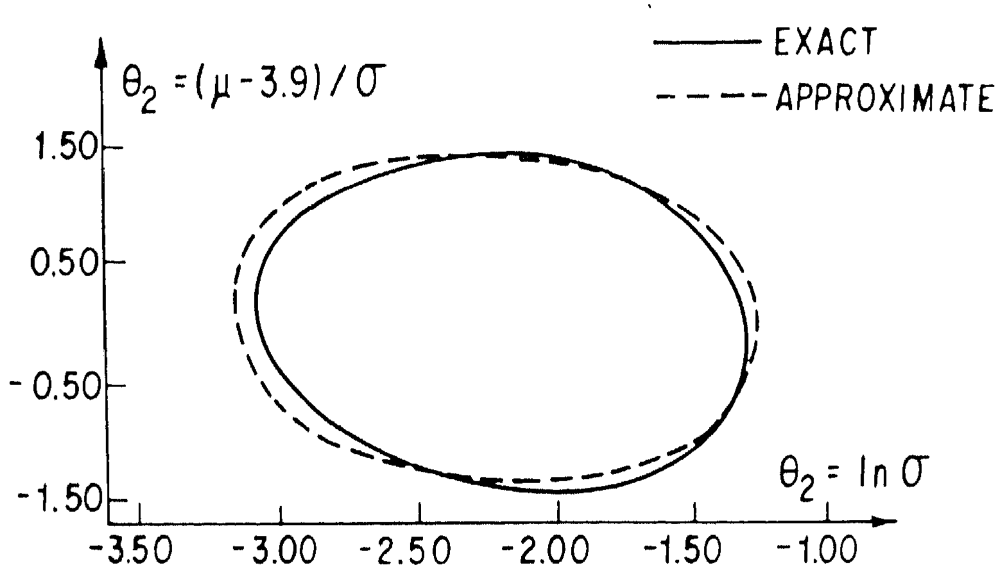

FIGURE 5: CONTOURS OF THE LIKELIHOOD FUNCTION FOR θ_1 AND θ_2 $(\ell(\theta_1, \theta_2) = -1)$

In table 4, we have the values of the standardized third derivatives summaries proposed by Kass and Slate (1992) to diagnostic the global assessment of normality. We observe a good improvement in the joint normality of the likelihood function for $\theta_1 = \ln\sigma$ and $\theta_2 = (\mu - 3.9)/\sigma$.

Parametrization	$m^2\overline{B}^2$
(μ, σ)	5.1287
(θ_1, θ_2)	1.1146

Table 4: Standardized Third Derivatives Summaries in Parameterizations (μ, σ) and (θ_1, θ_2)

4.3 Reparametrization and Laplace Approximations for Posterior Moments Denoting the likelihood function for a parameter θ by $L(\theta)$ and the prior density for θ by $\pi(\theta)$, the posterior moment for selected real-valued functions $g(\theta)$ is given by

$$E\{g(\theta) \mid \text{data}\} = \frac{\int g(\theta)\pi(\theta)L(\theta)d\theta}{\int \pi(\theta)L(\theta)d\theta} \tag{4.14}$$

That is,

$$E\{g(\theta) \mid \text{data}\} = \frac{\int e^{-nh^*(\theta)}d\theta}{\int e^{-nh(\theta)}d\theta} \tag{4.15}$$

where $-nh = \ln\pi + \ln L(\theta)$, and $-nh^* = \ln g + \ln\pi + \ln L(\theta)$, assuming g is a positive function.

Laplace's approximation for the integrals in the numerator and denominator of (23) gives (see for example, Tierney and Kadane, 1986), the approximation,

$$\widehat{E}\{g(\theta) \mid \text{data}\} = \left(\frac{\sigma^*}{\sigma}\right)\exp\left\{-n\left[h^*(\widehat{\theta}^*) - h(\widehat{\theta})\right]\right\} \tag{4.16}$$

where $\widehat{\theta}^*$ maximizes $-h^*$, $\sigma^* = \left\{h^{*''}(\widehat{\theta}^*)\right\}^{-1/2}$, $\widehat{\theta}$ maximizes $-h$ and $\sigma = \left\{h''(\widehat{\theta})\right\}^{-1/2}$

The accuracy of Laplace's method usually depends on a good parametrization dependent upon both the data and the choice of prior to get accurate approximate posterior moments or marginal posterior densities (see for example, Achcar and Smith, 1990).

As a special illustration, consider a random variable Y with a binomial distribution $b(n, \theta)$. The likelihood function for θ is given by

$$L(\theta) \propto \theta^y(1-\theta)^{n-y} \quad (4.17$$

where $y = 0, 1, \ldots, n$.

The Jeffreys prior for θ is given by

$$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2} \tag{4.18}$$

where $0 < \theta < 1$ (see Box and Tiao, 1973).

Considering the reparametrization $\phi_1 = \sin^{-1}\sqrt{\theta}$ (that is, $\theta = \sin^2\phi_1$), we have a Jeffreys locally uniform prior for ϕ_1, $0 < \phi_1 < 90^o$. In table 5, we have Laplace's approximations for $E(\theta \mid \text{data})$ considering the original parametrization, the parametrization $\phi_1 = \sin^{-1}\sqrt{\theta}$ and also the usual logistic parametrization $\phi_2 = \ln\left(\theta \big/ (1-\theta)\right)$ We observe small percentage errors for Laplace's approximations of $E(\theta \mid \text{data})$ in the parametrizations ϕ_1 and ϕ_2 if compared with the available exact value of $E(\theta \mid \text{data})$.

y	EXACT	θ	$\phi_1 = \sin^{-1}\sqrt{\theta}$	$\phi_2 = \ln\left\{\theta \big/ (1-\theta)\right\}$
1	0.2500	0.2638(5.22%)	0.2446(2.22%)	0.2549(1.95%)
3	0.5833	0.5629(3.61%)	0.5797(0.62%)	0.5852(0.33%)
5	0.9166	0.8793(4.24%)	0.9129(0.41%)	0.9169(0.03%)

Table 5: Posterior Means $E(\theta \mid \text{data})$ Approximated by Laplace's Method ($n = 5$) (error in parenteses)

In table 6, we have the values of the standardized third derivatives of the logarithm of the posterior densities (with $n = 5$ and $y = 3$) considering the parametrizations θ, $\phi_1 = \sin^{-1}\sqrt{\theta}$ and $\phi_2 = \ln\left[\frac{\theta}{1-\theta}\right]$.

We observe better normality, of the posterior density in the parametrization $\phi_1 = \sin^{-1}\sqrt{\theta}$, with a locally uniform Jeffreys prior density.

	θ	$\phi_1 = \sin^{-1}\sqrt{\theta}$	$\phi_2 = \ln\left[\theta \big/ (1-\theta)\right]$
STD (6)	1.1002	0.0913	0.1380
Jeffreys Prior Density	$\theta^{-1/2}(1-\theta)^{-1/2}$ $0 \leq \theta \leq 1$	Locally Uniform $0 \leq \phi_1 \leq 90^o$	$\frac{e^{\phi_2/2}}{(1+e^{\phi_2})}$ $-\infty < \phi_2 < \infty$

Table 6: Standardized Third Derivatives of the Logarithm of Posterior Densities ($n = 5, y = 3$)

4.4 Reparametrization for a Bivariate Exponential Distribution Applied in Accelerated Life Testing

Let us assume that we have two failure times associated to each observational unit in an accelerated life test problem. That is, consider a two-component life times X and Y , J stress levels $V_1, \ldots, V_J$ and life tests are conducted at constant application of the selected stresses. Using this information, we get inferences about the component life times under normal stress condition given by V_0.

At a normal stress level V_0, assume that (X, Y) has a bivariate exponential distribution (BVED) of Block and Basu (1974) with parameters $\lambda_{10}, \lambda_{20}$ and λ_{30}. Also assume that under a stress level V_j, $j = 1, \ldots, J$, (X, Y) has the BVED with parameters $\lambda_{1j}, \lambda_{2j}$ and λ_{3j}, $j = 1, 2, \ldots, J$ and joint probability density function given by

$$f(x,y) = \begin{cases} \dfrac{\lambda_{1j}\lambda_j(\lambda_{2j}+\lambda_{3j})}{\lambda_{1j}+\lambda_{2j}} \exp\{-\lambda_{1j}x - (\lambda_{2j}+\lambda_{3j})y\} \\ \qquad\qquad\qquad\qquad \text{if} \quad x < y \\ \dfrac{\lambda_{2j}\lambda_j(\lambda_{1j}+\lambda_{3j})}{\lambda_{1j}+\lambda_{2j}} \exp\{-(\lambda_{1j}+\lambda_{3j})x - \lambda_{2j}y\} \\ \qquad\qquad\qquad\qquad \text{if} \quad x \geq y \end{cases} \tag{4.19}$$

where $\lambda_j = \lambda_{1j} + \lambda_{2j} + \lambda_{3j}$, $j = 0, 1, \ldots, J$.

Also, consider the power rule model (see for example, Mann, Schafer and Singpurwalla, 1974), given by

$$\lambda_{ij} = c_i V_j^P \tag{4.20}$$

where $i = 1, 2, 3$; $j = 0, 1, \ldots, J$; c_1, c_2, c_3 and P are constants. The model (28) was also considered by Basu and Ebrahimi (1987).

Assuming that the data obtained for the J stress levels $V_1, V_2, \ldots, V_J$ are independent, the likelihood function for c_1, c_2, c_3 and P is given by

$$L(c_1,c_2,c_3,P) = \frac{c_1^r\, c_{23}^r\, c_2^{n-r}\, c_{13}^{n-r}\, c_{123}^n}{c_{12}^n} \times \\ \times \left\{\prod_{j=1}^J V_j^{2Pn_j}\right\} \exp\}-[c_1 S_x(P) + c_2 S_y(P) + c_3 T(P)]\} \tag{4.21}$$

where $r = \sum_{j\equiv 1}^J r_j$, $n = \sum_{j\equiv 1}^J n_j$, $S_x(P) = \sum_{j\equiv 1}^J n_j \overline{X}_j V_j^P$, $S_y(P) = \sum_{j\equiv 1}^J n_j \overline{Y}_j V_j^P$, $T(P) = \sum_{j\equiv 1}^J R_j V_j^P$, $r_j = \sum_{i=1}^{nj} \delta_{ij}$, $\delta_{ij} = 1$ if $X_{ij} < Y_{ij}$, $\delta_{ij} = 0$ if $X_{ij} \geq Y_{ij}$, $R_j = \sum_{i=1}^{nj} [Y_{ij}\delta_{ij} + (1-\delta_{ij})X_{ij}]$, $c_{12} = c_1 + c_2$, $c_{13} = c_1 + c_3$, $c_{23} = c_2 + c_3$, $c_{123} = c_1 + c_2 + c_3$ and n_j is the number of units at the beginning of each test with stress level V_j.

For inferences on $\underset{\sim}{\psi} = (c_1, c_2, c_3, P)$, or even functions of the parameters, we usually use the asymptotic normality of the maximum likelihood estimates given by

$$\left(\hat{\underset{\sim}{\psi}}\right) = \left(\hat{c}_1, \hat{c}_2, \hat{c}_3, \hat{P}\right) \overset{a}{\sim} N\left\{\underset{\sim}{\psi}; I_0^{-1}\right\} \tag{4.22}$$

where I_0 is the observed information matrix.

The approximation (4.23) may not be very good for small or moderate sample sizes. Therefore, we should try different reparametrizations to improve the global joint normality of the likelihood function.

As a numerical illustration, consider a generated data set (see table 7) considering three stress levels $V_1 = 1$, $V_2 = 2$ and $V_3 = 3$. At each stress level, 15 bivariate observations (X, Y) were generated from a BVED with density (27), and the power rule model (28). From table 7, we get $n_1 = n_2 = n_3 = 15$, $n = 45$, $r_1 = 4$, $r_2 = 3$, $r_3 = 2$ and $r = 9$.

$V_1 = 1$ (X_{1i}, Y_{1i})	$V_2 = 2$ (X_{2i}, Y_{2i})	$V_3 = 3$ (X_{3i}, Y_{3i})
(7.65,2.18)	(2.29,0.02)	(0.34,0.20)
(16.67,9.26)	(0.10,0.38)	(1.50,1.30)
(39.30,6.72)	(0.88,0.27)	(0.63,0.69)
(1.30,3.22)	(0.45,0.04)	(0.68,0.12)
(9.04,2.23)	(1.66,1.60)	(3.22,0.09)
(5.15,0.41)	(0.74,1.67)	(1.91,0.91)
(5.20,5.91)	(2.50,0.37)	(0.52,0.58)
(5.00,0.84)	(3.50,0.03)	(0.30,0.01)
(5.66,0.42)	(8.45,0.71)	(1.30,0.02)
(11.80,0.15)	(4.60,0.83)	(0.52,0.10)
(17.08,10.37)	(2.66,1.06)	(2.08,0.30)
(17.92,0.76)	(1.46,1.04)	(0.95,0.91)
(1.62,2.73)	(1.03,0.41)	(0.43,0.02)
(1.42,1.85)	(4.36,1.34)	(0.25,0.08)
(3.60,1.50)	(0.76,0.77)	(1.39, 0.08)

Table 7: Generated Bivariate Life Time Data

Using Newton-Raphson method, we find the maximum likelihood estimators $\hat{c}_1 = 0.0571$, $\hat{c}_2 = 0.2643$, $\hat{c}_3 = 0.0602$, and $\hat{P} = 2.03$. Approximate 95% confidence intervals for c_1, c_2, c_3, and P based on the observed information matrix are given by (0.0081,0.1060), (0.1322,0.3964), (-0.0061,0.1265) and (1.5791,2.4809), respectively.

Under the normal stress condition, we assume that $V_0 = 0.5$. Therefore, the maximum likelihood estimators for the parameters of the bivariate exponential distribution under normal condition are given by $\hat{\lambda}_{10} = 0.0139$, $\hat{\lambda}_{20} = 0.0647$, $\hat{\lambda}_{30} = 0.0147$.

The mean life times of the two components under the stress level j are given by

$$\begin{aligned} \mu_{1j} &= E(X) = \frac{(c_{123}c_{12} + c_2c_3)}{c_{123}c_{12}c_{13}} V_j^{-P} \\ \mu_{2j} &= E(Y) = \frac{(c_{123}c_{12} + c_1c_3)}{c_{123}c_{12}c_{23}} V_j^{-P} \end{aligned} \tag{4.23}$$

Thus, the maximum likelihood estimators for the mean life times under the normal condition are given by $\widehat{\mu}_{10} = 39.3340$ and $\widehat{\mu}_{20} = 12.9384$.

In the parametrization $\theta_1 = \ln c_1$, $\theta_2 = \ln c_2$, $\theta_3 = \ln c_3$ and P , we have the maximum likelihood estimators $\widehat{\theta}_1 = -2.8629$, $\widehat{\theta}_2 = -1.3307$, $\widehat{\theta}_3 = -2.8101$, and $\widehat{\mathbb{P}} = 2.03$. Considering the asymptotical normality of the maximum likelihood estimators $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\widehat{\theta}_3$ and $\widehat{P}$, we obtain approximate 95% confidence intervals for θ_1, θ_2, θ_3 and $\mathbb{P}$ given by (-3.7218,-2.0040), (-1.8308,-0.8305), (-3.9143,-1.7058) and (1.5790,2.4809), respectively. From these intervals, we get approximate 95% confidence intervals for c_1, c_2 and c_3 given by (0.0242,0.1348), (0.1603,0.4358) and (0.0199, 0.1816), respectively. We observe different confidence intervals for c_1, c_2 and c_3 considering the two parametrizations $\underset{\sim}{\psi} = (c_1, c_2, c_3, P)$ and $\underset{\sim}{\theta} = (\theta_1, \theta_2, \theta_3, P)$.

In Table 8, we have the values of the third derivatives of the logarithm of the likelihood function locally in the maximum likelihood estimators considering the parametrizations $\underset{\sim}{\psi} = (c_1, c_2, c_3, P)$, $\underset{\sim}{\theta} = (\ln c_1, \ln c_2, \ln c_3, P)$ and also $\underset{\sim}{\xi} = (\ln c_1, \ln c_2, c_3, P)$. We observe large values in the parametrization c_1, c_2, c_3 and P. We also have in Table 8, the values of the third derivatives summaries (see (8)) proposed by Kass and Slate (1992).

We observe that parametrization $\underset{\sim}{\xi} = (\ln c_1, \ln c_2, c_3, P)$ gives better global assessment of the normal approximation, since the value of $m^2\overline{B}^2$ is smaller than in parametrizations $\underset{\sim}{\psi}$ and $\underset{\sim}{\theta}$.

	d_{ijk}		
(ijk)	$\underset{\sim}{\psi}$	$\underset{\sim}{\theta}$	$\underset{\sim}{\xi}$
111	140206	25.23	25.2
112	1091	0.97	1.0
113	46230	0.98	16.3
114	0	15.07	15.1
222	3335	37.80	37.8
221	1091	2.44	2.4
223	2146	2.75	45.7
224	0	22.49	22.5
333	46757	22.33	46757.1
331	46230	0.49	2639.8
332	2146	2.25	567.3
334	0	16.18	0.0
444	50	50.44	50.4
441	249	14.24	14.2
442	81	21.51	21.5
443	253	15.24	253.1
123	1620	1.47	24.4
124	0	0.00	0.0
134	0	0.00	0.0
234	0	0.00	0.0
$m^2\overline{B}^2$	1.8681 $\underset{\sim}{\psi} = (c_1, c_2, c_3, P)$		
	3.9712 $\underset{\sim}{\theta} = (lnc_1, lnc_2, lnc_3, P)$		
	0.5010 $\underset{\sim}{\xi} = (lnc_1, lnc_2, lnc_3, P)$		

Table 8: Third Derivatives of the Log-Likelihood Function

5. SOME CONCLUSIONS

Reparametrization of a statistical model is a very important problem to all statisticians. We showed in some applications of section 4, how the accuracy of asymptotical results could be improved by working with an appropriate reparametrization. The use of the results of section 2, could be a suitable way to search for an appropriate reparametrization, especially for small dimensional problems. For high dimensional problem, we could use the diagnostics proposed by Kass and Slate (1992) for each proposed parametrization.

ACKNOWLEDGEMENTS

The author acknowledges support from the British Council and the Brazilian institution CAPES to visit the department of statistics and modelling science, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow, G1 1XH, U.K. The author is also extremely grateful to the referees for their useful comments.

REFERENCES

(1) Achcar, J.A. (1990). *A Bayesian approach to reparametrization of the exponential distribution with type I censored data*, technical report Nº 59, ICMSC, Universidade de São Paulo, Brasil.

(2) Achcar, J.A. (1991). An useful reparametrization for the extreme value distribution, *Comp. Statist. Quar.*, 2, 113-125.

(3) Achcar, J.A.; SMITH, A.F.M. (1990). *Aspects of reparametrization in approximate Bayesian inference*, Essays in honor of G.A. Barnard (eds., S. Geisser, J.S. Hodges, S.J. Press, A. Zellner), Amsterdam, North Holland.

(4) Anscombe, F.J. (1964). Normal likelihood functions, *Ann. Inst. Stat. Math.*, 16, 1-19.

(5) Basu, A.P.; Ebrahimi, M. (1987), On a bivariate accelerated life test, *J. of Statist. Planning and Infer.*, 16, 297-304.

(6) Bates, D.M.; Watts, D.G. (1980), Relative curvature measures of non-linearity, *J. Roy. Statist. Soc.*, B, 42(1), 1-25.

(7) Bates, D.M.; Watts, D.G. (1988), Nonlinear regression analysis and its applications, New York: John Wiley & Sons.

(8) Block, H.; Basu, A.P. (1974). A continuous bivariate exponential extension, *J. Amer. Statist. Assoc.*, 69, 1031-1037.

(9) Box, G.E.P.; Tiao G.C. (1973), Bayesian Inference in Statistical Analysis, *Addison-Wesley*.

(10) Cox, D.R.; Reid, N. (1987), Parameter orthogonality and approximate conditional inference, *J. Roy. Statist. Soc.*, B, 49(1), 1-39.

(11) Hills, S.E.; Smith, A.F.M. (1993), Diagnostic plots for improved parametrization in Bayesian inference, *Biometrika*, 80, 1, 61-74.

(12) Holland, P.W. (1973), Covariance stabilizing transformations, *Ann. Statist.*, 1(1).

(13) Hougaard, P. (1982), Parametrizations of non-linear models, *J. Roy. Statist. Soc.*, B, 44(2), 244-252.

(14) Kass, R.E.; Slate, E.H. (1992), *Reparametrization and diagnostics of posterior nonnormality*. In Bayesian Statistics 4 (eds J.M. Bernardo, J. Berger, A.P. Dawid, and A.F.M. Smith), Oxford: Oxford University Press.

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

(16) Mann, N.R.; Schafer, R.E.; Singpurwalla, N.D. (1974), Methods for Statistical analysis of reliability and lifetime data, *New York: John Wiley & Sons.*

(17) Naylor, J.C.; Smith, A.F.M. (1982), Applications of a method for the efficient computation of posterior distributions, *Appl. Statist.*, 31, 214-225.

(18) Sprott, D.A. (1973), Normal likelihoods and their relation to large sample theory of estimation, *Biometrika*, 60, 457-465.

(19) Sprott, D.A. (1980), Maximum likelihood in small samples: estimation in the presence of nuisance parameters, *Biometrika*, 67, 515-523.

(20) Tierney, L.; Kadane, J.B. (1986), Accurate approximations for posterior moments and marginal densities, *J. Amer. Statist. Assoc.*, 81, 82-86.