UNIVERSIDADE DE SÃO PAULO
CAMPUS DE SÃO CARLOS
INSTITUTO DE CIÊNCIAS MATEMÁTICAS DE SÃO CARLOS

I.C.M.S.C.

A Bayesian approach to reparametrization of the exponential distribution with type I censored data.

ACHCAR, JORGE ALBERTO

nº 59

Notas do ICMSC - USP

ISSN 0103-2577

A Bayesian approach to reparametri zation of the exponential distribu tion with type I censored data.

ACHCAR, JORGE ALBERTO

nº 59

São Carlos (SP)

1990

A BAYESIAN APPROAÇH TO REPARAMETRIZATION OF THE EXPONENTIAL DISTRIBUTION WITH TYPE I CENSORED DATA

JORGE ALBERTO ACHCAR
ICMSC-UNIVERSIDADE DE SÃO PAULO
CAIXA POSTAL 668
13.560, SÃO CARLOS, SP, BRASIL

S U M M A R Y

"The problem of good parametrization is very important, specially using asymptotical results. In this paper, we explore a Bayesian approach to find a good reparametrization considering an exponential distribution with type I censored data. We also compare our approach with another parametrization procedure proposed by Sprott (1973)".

1. INTRODUCTION

Usually, with censored survival data the statistician uses asymptotical results to get inferences of interest. For many parametrical models and depending on the censoring mechanism, one of these approximation results is given by the asymptotical normality of the maximum likelihood estimators. With this asymptotical distribution, we obtain confidence intervals and hypothesis tests of interest (see for example, Lawless, 1982). These normal approximations may not be very good in small-or moderate size samples. It can be shown that these approximations are usually poor with censored data unless the sample size is fairly large. These approximations may be improved considering a good reparametrization (see for example, Anscombe, 1964; Sprott, 1973).

In this paper, we present a Bayesian approach to choose a good parametrization considering type I censored data and an exponential distribution for the survival data. We also campare our approach with other reparametrization procedure proposed by Sprott (1973).

2. FORMAL STATEMENT OF THE PROBLEM

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

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

where $t \geq 0$.

Assuming type I censored data, the period of follow-up for the i^{th} individual is limited to a fixed value L_i. Then, the observed survival time of the i^{th} individual is $t_i = \min(T_i, L_i)$. Define δ_i such that $\delta_i = 0$ 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,

$$L(\theta) = \prod_{i=1}^{n} f^{\delta_i}(t_i;\theta)\, S^{1-\delta_i}(t_i;\theta) \tag{2}$$

where $S(t_i;\theta) = \exp\{-t_i/\theta\}$ is the survival function of the exponential distribution with density (1).

That is,

$$L(\theta) = \theta^{-r} \exp\{-\sum_{i=1}^{n} t_i/\theta\} \tag{3}$$

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

The second derivative of the logarithm of the likelihood function (3) is given by,

$$\frac{d^2 \log L}{d\theta^2} = \frac{r}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^{n} t_i \tag{4}$$

To find the Fisher information $I(\theta) = E\{-d^2 \log L/d\theta^2\}$, we observe that:

(i) $P\{\delta_i = 0\} = \exp\{-L_i/\theta\} = 1 - P\{\delta_i = 1\}$

(ii) $E\{t_i | \delta_i = 1\} = E\{T_i | T_i \le L_i\} = \int_0^{L_i} \frac{x\theta^{-1} e^{-x/\theta} dx}{(1 - e^{-L_i/\theta})} = \theta - \frac{L_i e^{-L_i/\theta}}{(1 - e^{-L_i/\theta})}$

That is,

$$E\{t_i\} = E\{t_i|\delta_i = 0\}\ P\{\delta_i = 0\} + E\{t_i|\delta_i = 1\}\ P\{\delta_i = 1\}$$

$$= L_i e^{-L_i/\theta} + (\theta - \frac{L_i e^{-L_i/\theta}}{1-e^{-L_i/\theta}})(1-e^{-L_i/\theta}) = \theta(1-e^{-L_i/\theta})$$

(iii) $E\{r\} = \sum_{i=1}^{n} E\{\delta_i\} = \sum_{i=1}^{n} (1-e^{-L_i/\theta}) = Q$

Thus, the Fisher information is given by,

$$I(\theta) = E\{-\frac{d^2 \log L}{d\theta^2}\} = \frac{Q}{\theta^2} \tag{5}$$

The maximum likelihood estimator $\hat{\theta} = T/r$, where $T = \sum_{i=1}^{n} t_i$ has an approximate normal distribution given by,

$$\frac{(\hat{\theta}-\theta)}{I^{-1/2}(\theta)} \overset{a}{\sim} N\{0,1\} \tag{6}$$

where $I(\theta) = Q/\theta^2$.

To improve the nornal approximation (6), Sprott (1973) considers a reparametrization ϕ such that $d^3 \log L(\phi)/d\phi^3|_{\hat{\phi}} = 0$, where $\hat{\phi}$ is the maximum of $L(\phi)$. This parametrization is given by $\phi = \theta^{-1/3}$. The maximum likelihood estimator $\hat{\phi} = \hat{\theta}^{-1/3}$ has an approximate normal distribution with mean $\phi = \theta^{-1/3}$ and asymptotical variance $\phi^2/9Q$ (see for example, Lawless, 1982).

Thus, $\phi_B \propto \ln\theta$. Therefore, we consider the reparametrization $\phi_B = \ln\theta$. In this reparametrization $\phi_B = \ln\theta$ (that is, $\theta = e^{\phi_B}$), the maximum likelihood estimator $\hat{\phi}_B = \ln\hat{\theta}$ has an approximate normal distribution with mean $\phi_B = \ln\theta$ and asymptotical variance Q^{-1}, where $Q = \sum_{i=1}^{n} (1 - e^{-L_i e^{-\phi_B}})$.

4. AN EXAMPLE

The data of table 1 represent the lifetimes of 10 pieces of equipment.

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

TABLE 1. Lifetimes of 10 Pieces of Equipment

From table 1, we have r=7, $T = \sum_{i=1}^{10} t_i = 308$ and the maximum likelihood estimator of θ is given by $\hat{\theta} = T/r = 44.0$. Since $\hat{Q} = \sum_{i=1}^{10} (1 - e^{-L_i/\hat{\theta}}) = 6.15$, we find the Fisher information $I(\hat{\theta}) = \hat{Q}/\hat{\theta}^2 = 0.00318$. Therefore, $I^{-1/2}(\hat{\theta}) = 17.7$ and

3. A BAYESIAN APPROACH TO REPARAMETRIZATION

The Jeffreys invariant prior for θ (see for example Box and Tiao, 1973) is given by,

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

$$\propto \theta^{-1}\{\sum_{i=1}^{n}(1-e^{-L_i/\theta})\}^{1/2} \tag{7}$$

Considering a reparametrization ϕ_B such that $E\{-d^2 \log L(\phi_B)/d\phi_B^2\} = \text{constant}$, the likelihood function for ϕ_B is well approximated by a normal density. This is equivalent to find a reparametrization ϕ_B with a locally uniform Jeffreys prior.

Therefore, we should find a reparametrization ϕ_B given by the differential equation,

$$\frac{d\phi_B}{d\theta} \propto \theta^{-1}\{\sum_{i=1}^{n}(1-e^{-L_i/\theta})\}^{1/2} \tag{8}$$

That is,

$$\phi_B \propto \int \theta^{-1}\{n-\sum_{i=1}^{n} e^{-L_i/\theta}\}^{1/2} d\theta \tag{9}$$

Considering $\sum_{i=1}^{n} e^{-L_i/\theta} \approx \int_0^{L_n} e^{-y/\theta} dy = \theta(1-e^{-L_n/\theta})$, we have:

$$\theta_B \propto \int \{n\theta^{-2} - \theta^{-1}(1-e^{-L_n/\theta})\}^{1/2} d\theta$$

With $1-e^{-L_n/\theta} \approx L_n/\theta$, we have:

$$\phi_B \propto \int \theta^{-1} d\theta \tag{10}$$

$(\hat{\theta}-\theta)/17.7 \stackrel{a}{=} N(0,1)$. Thus, a 95% confidence interval for θ is given by $9.3 \leq \theta \leq 78.7$. In figure 1, we have the graphs of the likelihood function for θ and the normal approximation for the maximum likelihood estimator $\hat{\theta}$. We observe a very poor approximation for the likelihood function for θ.

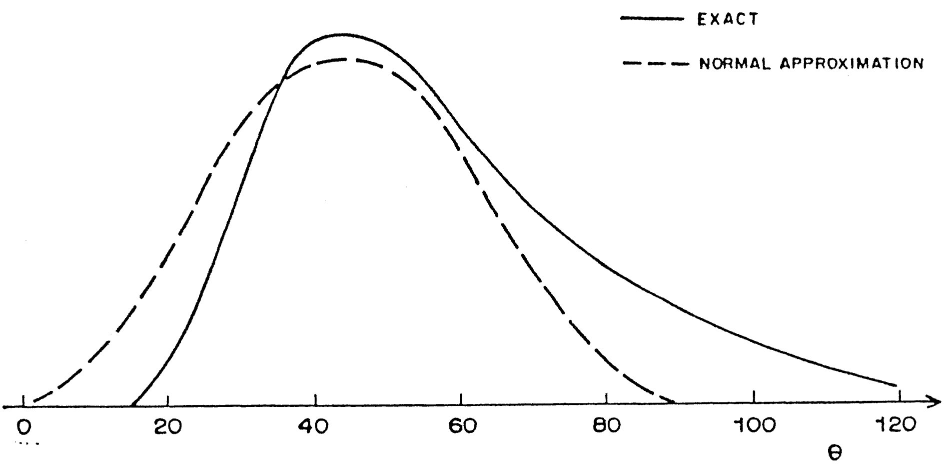

FIGURE 1. Likelihood Function For θ

Considering the parametrization $\phi=\theta^{-1/3}$ proposed by Sprott (1973), we have,

$$\frac{(\hat{\phi}-\phi)}{(\hat{\phi}^2/9\hat{Q})^{1/2}} \stackrel{a}{=} N(0,1) \tag{11}$$

where $\hat{\phi}=\hat{\theta}^{-1/3}=0.2833$ is the maximum likelihood estimator of ϕ.

Since $\hat{\phi}^2/9\hat{Q}=0.00144$, a 95% confidence interval for ϕ is given by $0.2088 \leq \phi \leq 0.3577$. From $\phi=\theta^{-1/3}$, we find a 95% confidence interval for θ given by $21.8 \leq \theta \leq 109.9$. We observe that this confidence interval is very different of the confidence interval in the original parametrization θ. In figure 2, we observe very good normal approximation for the likelihood function of $\phi=\theta^{-1/3}$.

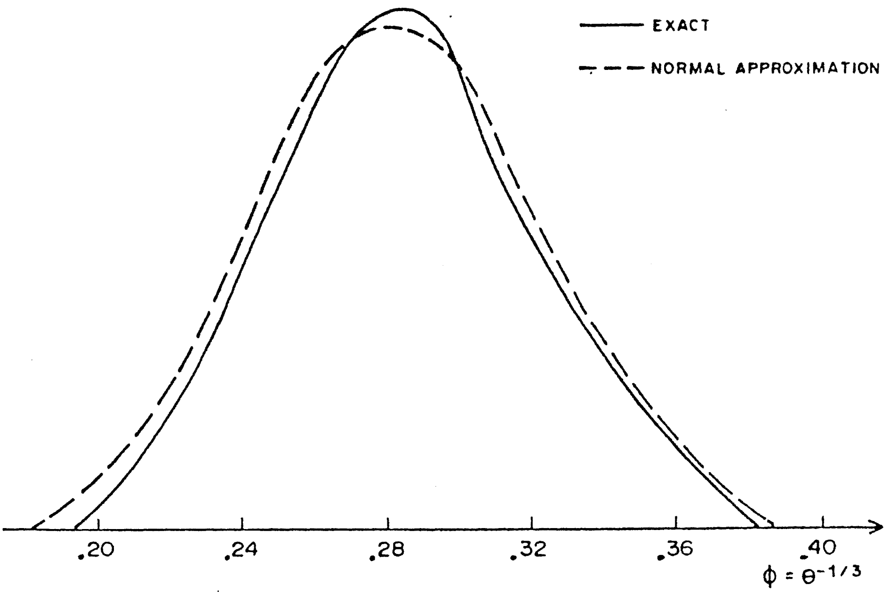

FIGURE 2. Likelihood Function For $\phi=\theta^{-1/3}$

Considering the parametrization $\phi_B=\ell n\theta$ using the Bayesian approach, the maximum likehood estimator $\hat{\phi}_B=\ell n\hat{\theta}=3.7842$ has a normal asymptotical distribution given by,

$$\frac{(\hat{\phi}_B-\phi_B)}{\hat{Q}^{-1/2}} \overset{a}{=} N\{0,1\} \tag{12}$$

Since $\hat{Q}^{-1}=0.1626$, we find a 95% confidence interval for ϕ_B given by $2.9938 \leq \phi_B \leq 4.5745$. From $\phi_B=\ell n\theta$, a 95% confidence interval for θ is given by $19.9622 \leq \theta \leq 96.9834$. We observe that this 95% approximate confidence interval for θ is close to the 95% approximate confidence interval for θ considering the parametrization $\phi=\theta^{-1/3}$. In figure 3, we observe very good normal approximation for the likehood function of ϕ_B.

In table 2, we have the percentage errors of the normal approximations considering the parametrizations θ, ϕ and ϕ_B. We observe small errors in the parametrizations ϕ and ϕ_B.

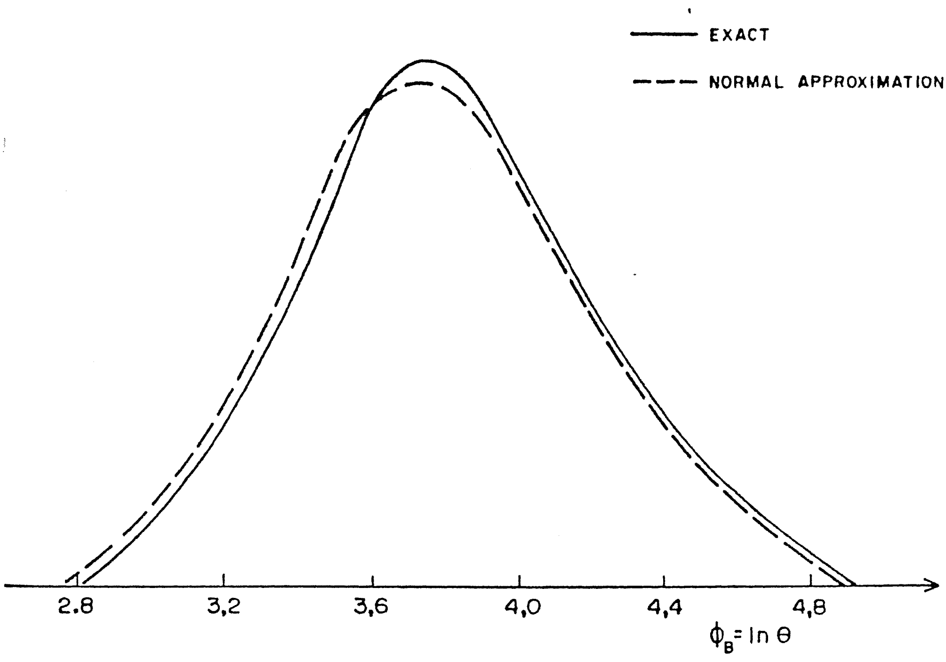

FIGURE 3. Likelihood Function For $\phi_B=\ell n\theta$

θ	EXACT $10^{16}L(\theta)$	NORMAL APPROXIMATION IN θ	NORMAL APPROXIMATION IN $\phi=\theta^{-1/3}$	NORMAL APPROXIMATION IN $\phi_B=\ell n\theta$
20	1.60	11.44 (86%)	2.31 (31%)	4.22 (62%)
30	15.90	20.92 (24%)	17.05 (6.8%)	18.20 (12.6%)
40	27.64	27.84 (0.73%)	27.74 (0.4%)	27.77 (0.5%)
44	28.56	28.56 (0%)	28.56 (0%)	28.56 (0%)
50	27.04	26.97 (0.24%)	27.21 (0.63%)	27.16 (0.45%)
60	21.06	19.02 (10.8%)	21.84 (3.55%)	21.25 (0.86%)
70	14.91	9.76 (52.75%)	16.12 (7.52%)	14.72 (1.3%)
80	10.15	3.64 (178%)	11.52 (12%)	9.52 (6.6%)
90	6.82	0.99 (588%)	8.16 (16%)	5.91 (15.4%)
100	4.60	0.20 (2243%)	5.80 (21%)	3.60 (28%)

TABLE 2. Likelihood Function For θ And Normal Approximations In Parametrizations θ, $\phi=\theta^{-1/3}$ And $\phi_B=\ell n\theta$. Percentage Errors Are Given In Parenteses

$$\left(\text{ERROR} = \frac{|\text{EXACT} - \text{APPROX.}|}{\text{APPROX.}} \times 100\right)$$

5. CONCLUSIONS

In both parametrizations $\phi=\theta^{-1/3}$ and $\phi_B=\ell n\theta$, we observe (see table 2) a very good improvement in the normal approximations for the likelihood functions. In the parametrization $\phi_B=\ell n\theta$ considering the Bayesian approach, we observe a better approximation for values $\theta>50$, where the effect of censoring in the likelihood function can be very large, specially with a large number of censored observations.

REFERENCES

ANSCOMBE,F.J. (1964). Normal Likelihood Functions, Ann. Inst. Stat. MATH. 16, 1-19.

BOX,G.E.P.; TIAO,G.C. (1973). Bayesian Inference In Statistical Analysis, Addison-Wesley.

LAWLESS,J.F. (1982). Statistical Models And Methods For Lifetime Data, John Wiley & Sons.

SPROTT,D.A. (1973). Normal Likelihoods And Relation To A Large Sample Theory Of Estimation, Biometrika, 60, 457-465.