**International Journal of Pure and Applied Mathematics**
**Volume 17 No. 2 2004, 173-178**

# ON PRECISION IN MAXIMUM LIKELIHOOD ESTIMATES FOR THE WEIBULL DISTRIBUTION

A.J. Watkins
European Business Management School – EBMS
University of Wales Swansea
Haldane Building, Singleton Park, Swansea, SA2 8PP, U.K.
e-mail: a.watkins@swansea.ac.uk

**Abstract:** This note continues discussion on the assessment of precision in maximum likelihood estimates for the two parameter Weibull distribution. We consider methods based on the asymptotic normality of estimates, and an alternative derived directly from the likelihood.

**AMS Subject Classification:** 62F12, 62F25, 62N02
**Key Words:** fisher information, maximum likelihood estimation, Type I censoring, Weibull distribution

## 1. Introduction

This note considers various points arising from the contribution of Kahle [1], who outlines a framework for the estimation of precision in maximum likelihood estimates for the two parameter Weibull distribution. As there, the actual calculation of the maximum likelihood estimates is not, in itself, of primary interest, being well-documented – see, for instance, Kalbfleisch [2] – and we are more concerned with the subsequent assessment of precision in these estimates.

Received: September 14, 2004 

The structure of our discussion is as follows. We first consider the data available for analysis, summarise the large sample results at our disposal, and discuss the consequences of choosing between observed and expected Fisher information. We then apply these results to the Weibull distribution, and discuss the calculation of various contours associated with these results. We finally consider the relevance of these results to the distribution of the maximum likelihood estimators. For consistency, we adopt the same basic notation as Kahle [1], and write the probability density function of the Weibull distribution as

$$f = \frac{kt^{k-1}}{\vartheta} \exp\left\{-\left(\frac{t}{\vartheta}\right)^k\right\},$$

for $t \geq 0$, and $k, \vartheta > 0$, and write $\theta = (k, \vartheta)'$. We denote the true (unknown) values of $k, \vartheta$ by $k_0, \vartheta_0$, and write $\theta_0 = (k_0, \vartheta_0)'$, and denote the maximum likelihood estimates of $k, \vartheta$ by $\hat{k}, \hat{\vartheta}$, respectively, and write $\hat{\theta} = (\hat{k}, \hat{\vartheta})'$.

## 2. Asymptotic Theory

The large sample result which may be used to assess precision in the maximum likelihood estimates is

$$\hat{\theta}\,\tilde{}\,N\left(\theta_0, \{I_\theta(\theta_0)\}^{-1}\right), \tag{1}$$

where

$$I_\theta = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix} \equiv \begin{bmatrix} I_{kk} & I_{k\vartheta} \\ I_{k\vartheta} & I_{\vartheta\vartheta} \end{bmatrix}$$

is the expected Fisher information matrix, so that the large sample distribution of $\hat{\theta}$ is characterised by *ellipses* of constant density defined by

$$(\theta - \theta_0)' \, I_\theta(\theta_0) \, (\theta - \theta_0) = c,$$

for arbitrary non-negative $c$. Since, asymptotically,

$$\left(\hat{\theta} - \theta_0\right)' I_\theta(\theta_0) \left(\hat{\theta} - \theta_0\right) \tilde{}\,\chi^2_2,$$

it can be argued that the ellipse $(\theta - \theta_0)' \, I_\theta(\theta_0) \, (\theta - \theta_0) = c$ will contain the maximum likelihood estimator $\hat{\theta}$ with probability $1 - \exp(-c/2)$; see Mardia, Kent and Bibby [3] for an excellent summary of the merits of this argument. Further modifications to these results are based on the convergence of observed and expected Fisher information matrices, which allows us to replace $I_\theta$ by its *observed* counterpart $J_\theta$, and on noting that these results depend on $\theta_0$, which is unknown (except in simulation experiments), and must be replaced by the calculated estimates $\hat{\theta}$ in practice. Thus, we may obtain an approximate $100(1-\alpha)\%$ confidence region for $\theta_0$ by calculating the ellipse

$$\left(\theta - \hat{\theta}\right)' J_\theta(\hat{\theta}) \left(\theta - \hat{\theta}\right) = -2 \log_e \alpha, \tag{2}$$

in which $\alpha, \hat{\theta}$ are regarded as known. The above is consistent with the argument in Kahle [1] – and provides the necessary correction to his equation (6) – but is not consistent with the curve shown in Figure 3 there, which we would expect to show an ellipse.

## 3. An Alternative Approach

Kalbfleisch [2] oulines an alternative approach to the use of (1), based directly on the likelihood as a measure of consistency between the observed data and parameter values. This alternative has been considered by, *inter alia*, Kahle [1]. The basic idea may be simply stated: if $L(k, \vartheta)$ denotes the likelihood of data, then the relative likelihood

$$R(k, \vartheta) = \frac{L(k, \vartheta)}{L(\hat{k}, \hat{\vartheta})}$$

lies in $(0, 1]$ for all feasible $(k, \vartheta)$, and we may define an approximate $100(1-\alpha)\%$ confidence region for $\theta_0$ as the interior of the curve

$$\{(k, \vartheta) : R(k, \vartheta) = \alpha\}. \tag{3}$$

This definition can, as in Kahle [1], be recast in terms of the log-likelihood, and hence can be shown to be asymptotically equivalent with the ellipses discussed above. However, for finite samples, the curve (3) is non-elliptical; furthermore, points on it have to be found numerically.

When the two parameter estimates are of the same order, it is straightforward to produce the curve using standard numerical methods. However, when the data values (and hence $\vartheta$) are orders of magnitude greater than $k$, then we can employ a rescaling approach outlined in Watkins and Leech [5] in an attempt to avoid any loss of precision.

## 4. Data for Analysis

Here, it is sufficient to consider $n$ items entering service simultaneously. The data for analysis under Type I censoring, in which censoring takes place at a pre-specified time $T$, comprises a random number $d$ of times to failure $t_1, \ldots, t_d$, together with $n-d$ censored values each equal to $T$. The first example in Kahle [1] has $n = 100, T = 1, d = 20$, with observed times to failure given as

$$\begin{array}{cccccccccc} 0.77 & 0.50 & 0.73 & 0.83 & 0.32 & 0.97 & 0.31 & 0.96 & 0.54 & 0.87 \\ 0.98 & 0.94 & 0.42 & 0.59 & 0.97 & 0.72 & 0.59 & 0.79 & 0.35 & 0.72 \end{array};$$

we calculate the maximum likelihood estimates for this sample as $\hat{k} = 2.4588, \hat{\vartheta} = 1.8403$ (to four decimal places), which are consistent with the values in Figure 3 and Figure 5 in Kahle [1]. We then calculate the expected and observed Fisher informations at $(\hat{k}, \hat{\vartheta})$ as

$$I_\theta(\hat{\theta}) = \begin{bmatrix} 11.483 & 16.882 \\ 16.882 & 35.702 \end{bmatrix} \text{ and } J_\theta(\hat{\theta}) = \begin{bmatrix} 11.504 & 16.915 \\ 16.915 & 35.708 \end{bmatrix},$$

where entries are given to three decimal places; the reader should be aware that the second and third formulae in equation (12) in Kahle [1] for the observed information are valid only at the maximum likelihood estimator, and are based on simplifications derived from this restriction. Watkins and John [4] disucss the evaluation of elements in the expected information; the necessary functions are available in many software libraries. In this case, the agreement between $I_\theta(\hat{\theta})$ and $J_\theta(\hat{\theta})$ is very good, and the ellipses based on the two matrices essentially coincide; that for $J_\theta(\hat{\theta})$ with $\alpha = 0.1$ in (2) is shown in Figure 1, which also shows the curve (3) with $\alpha = 0.1$. This curve is consistent with that in Figure 5 in Kahle [1].

## 5. Conclusions

We first remark that the difference between the two ellipses based on observed and expected Fisher information will disappear as $n \to \infty$; we also note that elements in the observed matrix require $O(n)$ operations, which, in practice, represents only a slight increase in computational cost over that required to calculate the elements in the expected information.

Next, we note from Figure 1 that the two confidence regions based on (2) and (3) seem largely to overlap, with the relative likelihood contour almost

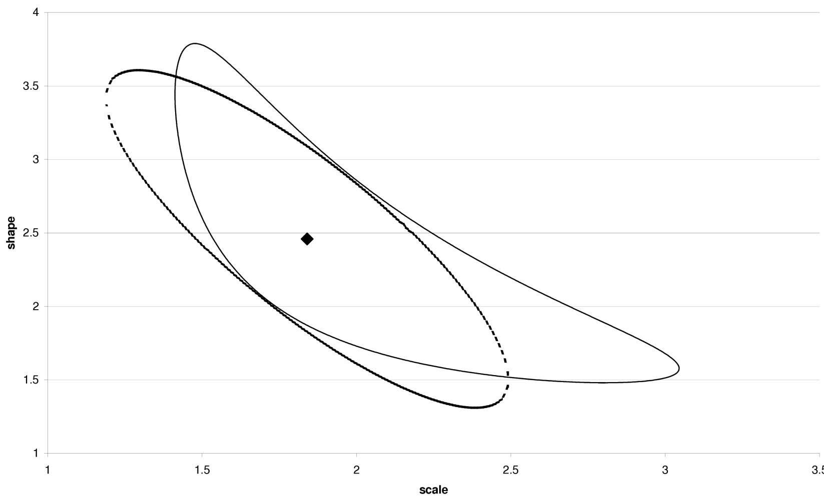


Figure 1: The maximum likelihood estimate (◊) together with 0.1 confidence regions based on asymptotic Normality (broken line) and relative likelihood (continuous line) for data from Kahle [1]

tangential to the ellipse; closer inspection regarding the last point shows a small distance between the curves. The extent of the overlap in general, and proximity between curves, will be discussed elsewhere, along with such related matters as the relative size of the two confidence regions. Further, as $n \to \infty$, we may expect the curve (3) to move towards the (shrinking) ellipse about the maximum likelihood estimates; there seems to be scope for further work on the rate of this convergence.

Finally, we remark that the non-elliptical nature of curves calculated from (3) seems to reflect more accurately the behaviour of the distributions of the maximum likelihood estimators for samples of small to moderate size; for further illustration of this, we may refer to the scattergrams in Watkins and Leech [6]. Again, there is scope for further work on the details of this link.

## References

[1] W. Kahle, Estimation of the parameters of the Weibull distribution for censored samples, *Metrika*, **44** (1996), 27-40.

[2] J.G. Kalbfleisch, *Probability and Statistical Inference II*, Springer-Verlag, New York (1979).

[3] K.V. Mardia, J.T. Kent, J.M. Bibby, *Multivariate Analysis*, Academic Press, New York (1979).

[4] A.J. Watkins, A.M. John, On the expected Fisher information for the Weibull distribution with type I censored data, *International Journal of Pure and Applied Mathematics*, **15** (2004), 401-412.

[5] A.J. Watkins, D.J. Leech, Towards automatic assessment of reliability for data from a Weibull distribution, *Reliability Engineering and System Safety,* **24** (1989), 343-350.

[6] A.J. Watkins, D.J. Leech, Methods for the continuous assessment of reliability of specialized equipment, *Quality and Reliability Engineering International,* **7** (1991), 377-391.