Statistica Sinica **8**(1998), 221-237

FISHER INFORMATION MATRICES WITH CENSORING, TRUNCATION, AND EXPLANATORY VARIABLES

Luis A. Escobar and William Q. Meeker

Louisiana State University and Iowa State University

Abstract: This paper shows how to compute the Fisher information matrix and the asymptotic covariance matrix for maximum likelihood estimators for a wide class of parametric models that include combinations of censoring, truncation, and explanatory variables. Although the models are based on underlying location-scale distributions, applications extend, for example, directly to the closely related and widely used Weibull and lognormal distributions. This paper unifies and generalizes a number of previously published results. The results are important for determining needed sample sizes and for otherwise planning statistical studies, especially in the areas of reliability and survival analysis where censoring and/or truncation are generally encountered.

Key words and phrases: Asymptotic variance-covariance, lognormal distribution, regression, reliability, survival analysis, test planning, weibull distribution.

1. Introduction

1.1. Motivation

When planning an experiment or other study that will involve the collection of data, it is generally important to evaluate the precision that the study will provide to estimate parameters and functions of parameters of interest. Such evaluations are essential for determining the test plan and the sample size to be used in the study. With complete data and an assumed normal distribution, standard methods found in text books on experimental design, survey design, and statistical methods provide the necessary information. For studies involving nonnormal distributions, truncation, and/or censored data, the necessary evaluations are more complicated. Typically exact theory is not available and asymptotic approximations, sometimes supplemented by simulations, are used in the planning stage.

1.2. Maximum likelihood estimation

Let $\mathcal{L}(\underline{\theta}) = \sum_{i=1}^{n} \mathcal{L}_i(\underline{\theta})$ denote the total log-likelihood for a specified model and test plan consisting of n independent but not necessarily identically distributed observations. Here $\mathcal{L}_i(\underline{\theta})$ is the contribution of the ith observation to the total log-likelihood. Let $\underline{g}(\underline{\theta})$ be a vector function $\underline{g}(\underline{\theta})$ of the parameters for which all the first derivatives with respect to the elements of $\underline{\theta}$ are continuous and let $\hat{\underline{\theta}}$ be the ML estimator of $\underline{\theta}$. Under standard regularity conditions (e.g., page 429 in Lehmann (1983) or Bhattacharyya (1985)) $\sqrt{n}[\underline{g}(\hat{\underline{\theta}}) - \underline{g}(\underline{\theta})]$ is asymptotically multivariate normal with mean zero and covariance matrix

$$\Sigma_{\hat{\underline{g}}} = \Big[\frac{\partial \underline{g}(\underline{\theta})}{\partial \underline{\theta}}\Big]^T \mathcal{I}^{-1} \Big[\frac{\partial \underline{g}(\underline{\theta})}{\partial \underline{\theta}}\Big], \tag{1}$$

where

$$\mathcal{I} = \lim_{n\to\infty} \Big\{\frac{1}{n} E\Big[-\frac{\partial^2 \mathcal{L}(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T}\Big]\Big\} = \lim_{n\to\infty} \Big\{\frac{1}{n} \sum_{i=1}^{n} E\Big[-\frac{\partial^2 \mathcal{L}_i(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T}\Big]\Big\} \tag{2}$$

and the expectation is with respect to the joint distribution of the data. Although the elements of the matrix $\mathcal{I}$ depend on $\underline{\theta}$, we suppress this in the notation. For a large class of model situations, including models with independent and identically distributed observations, $n\mathcal{I} = I_{\underline{\theta}}$ where $I_{\underline{\theta}}$ is the well known Fisher information matrix for $\underline{\theta}$

$$I_{\underline{\theta}} = E\Big[-\frac{\partial^2 \mathcal{L}(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T}\Big] = \sum_{i=1}^{n} E\Big[-\frac{\partial^2 \mathcal{L}_i(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}^T}\Big].$$

1.3. Applications

Although equation (1) can be evaluated at $\hat{\underline{\theta}}$ to estimate the covariance matrix, the common practice is to use the "local" estimate of $\Sigma_{\hat{\underline{g}}}$ that instead, estimates the expectations in $I_{\underline{\theta}}$ with the sample mean of the observed second derivatives over the n observations. Primary applications of the results in this paper, providing easy-to-compute expressions for expected information, will be for planning statistical studies involving censoring and truncation. The results are further extended to DOA (dead on arrival) and LFP (limited failure probability) models that have discrete atoms of probability at the extremes in the sample space.

Lawless (1982), page 171 and Nelson (1982), page 342, (1990), page 368 give formulas for computing the elements of $I_{\underline{\theta}}$ for some simple location-scale models involving censoring and discuss test planning applications. Some other specific applications include the following. Nelson and Meeker (1978) show how to find optimal regression designs for accelerated life tests. Meeter and Meeker (1994) find optimum accelerated life tests for a Weibull model in which the scale and shape parameters are functions of stress. David (1981), page 280 shows how to use the Fisher information matrix to determine coefficients for optimal asymptotic estimation of μ and σ using linear combinations of order statistics.

Meeker, Escobar, and Hill (1992) show how to find the minimum sample size needed to estimate the Weibull failure rate with specified precision in life tests with censored data. The algorithm by Escobar and Meeker (1994) facilitates these and other test-planning applications for the following location–scale distributions: smallest extreme value, largest extreme value, normal, and logistic distributions. Applications extend directly to the related Weibull, lognormal, and loglogistic distributions. For example, if a time to failure T follows a Weibull distribution, then $Y = \log(T)$ follows a smallest extreme value distribution.

This paper shows how to use the basic elements provided by the algorithm in Escobar and Meeker (1994) and the expressions in Meeker (1986) to compute $I_{\underline{\theta}}$ and thus $\Sigma_{\underline{\hat{\theta}}}$ for a wider range of models involving explanatory variables for both location and scale parameters, time, failure, and interval censoring, truncated distributions, and some special models involving a mixture of discrete and continuous parts. The methods apply to censoring on the right, the left, or on both sides. The censoring may be single or multiple at *fixed* point(s) or at *random* points (e.g. from competing risks).

1.4. Overview

In Section 2 we explain the basic quantities provided by the Escobar and Meeker (1994) algorithm and show how to use these quantities to compute $I_{\underline{\theta}}$ for models with a single distribution with observations that may be either left or right censored and for different kinds of censoring mechanisms (fixed or random). In Section 3 we extend these results to truncated distributions. Section 4 gives similar results for some special life models that include discrete, as well as continuous, components. In Section 5 we show how to extend applications to regression models that allow either or both of the parameters of a location-scale distribution to depend on explanatory variables. Section 6 provides additional generalizations while the Appendix contains derivations.

2. Fisher Information Matrix Elements for a Single Distribution Model

2.1. Model and assumptions

We assume that the random variable Y follows a location-scale distribution with cdf

$$G_Y(y; \underline{\theta}) = \Phi(z) \tag{3}$$

and pdf $d\,G_Y(y; \underline{\theta})/dy = \phi(z)/\sigma$, where $z = (y - \mu)/\sigma$, and the components of $\underline{\theta} = (\mu, \sigma)$ are, respectively, the location and scale parameters. Throughout the paper we assume that $\Phi(\cdot)$ and $\phi(\cdot)$ are the standardized (i.e., $\mu = 0$ and $\sigma = 1$) cdf and pdf, respectively, and that ϕ satisfies the regularity conditions given in Appendix A.

2.2. Time censored data

Let $\mathcal{L}_i(\underline{\theta})$ denote the natural logarithm of the likelihood of a single observation for Y. We assume that Y may be observed exactly or be censored on the left at (log) time y_l^c or on the right at (log) time y_r^c (log censoring times for the Weibull, lognormal, and loglogistic distributions). Appendix A.1 of Escobar and Meeker (1992) gives expressions for the likelihood and its first and second derivatives for uncensored observations as well as for left, right, and interval censored observations. Let $\underline{\theta} = (\mu, \sigma)$ and define the standardized censoring times

$$z_l^c = \frac{y_l^c - \mu}{\sigma}, \quad z_r^c = \frac{y_r^c - \mu}{\sigma}. \tag{4}$$

Then the elements of the Fisher information matrix, multiplied by σ^2 are

$$\begin{aligned} f_{11}(z_l^c, z_r^c) &= \sigma^2 E\Big[-\frac{\partial^2 \mathcal{L}_i(\underline{\theta})}{\partial \mu^2}\Big], \quad f_{12}(z_l^c, z_r^c) = \sigma^2 E\Big[-\frac{\partial^2 \mathcal{L}_i(\underline{\theta})}{\partial \mu \partial \sigma}\Big] \\ f_{22}(z_l^c, z_r^c) &= \sigma^2 E\Big[-\frac{\partial^2 \mathcal{L}_i(\underline{\theta})}{\partial \sigma^2}\Big]. \end{aligned} \tag{5}$$

Escobar and Meeker (1994) provide an algorithm to compute these elements. In general, these and other similar quantities can differ from observation-to-observation in a sample, but except for $\mathcal{L}_i(\underline{\theta})$, we generally suppress notation needed to show this dependence explicitly. The expectations in (5) are with respect to the joint distribution of the data which depends on the model and the censoring mechanisms associated with the data-generating process. This paper shows how the elements in (5) provide basic building blocks that can be used to obtain easy-to-compute expressions for the Fisher information matrix for a wide range of statistical models.

For singly time censored data (also known as Type I censored data) with fixed left and right censoring times the elements of $I_{\underline{\theta}} = n\mathcal{I}$ are obtained from

$$\mathcal{I}_{11} = \frac{1}{\sigma^2} f_{11}(z_l^c, z_r^c), \quad \mathcal{I}_{12} = \frac{1}{\sigma^2} f_{12}(z_l^c, z_r^c), \quad \mathcal{I}_{22} = \frac{1}{\sigma^2} f_{22}(z_l^c, z_r^c). \tag{6}$$

Note that $\Phi(z_l^c)$ and $1 - \Phi(z_r^c)$ are, respectively, the probabilities that the unit will be left and right censored. For multiply time-censored data (see Lawless (1982), page 34-36), the values of z_l^c or z_r^c can differ across the sample and the elements of $I_{\underline{\theta}}$ can be expressed as sums of the individual $\mathcal{I}_{jk}$'s.

2.3. Random censoring and competing risks

Here, for simplicity, we restrict the discussion to random right censored data; the extension to other random censored situations is straightforward. In survival analysis, random right censoring is often due to competing risks. (See, for example, Chapter 5 of Nelson (1982) or Chapter 10 of Lawless (1982).) We assume that the right random censoring point y_r^c has a continuous pdf $h(x)$ and that y_r^c and Y are statistically independent. Then, using conditional expectations, it follows that

$$\mathcal{I}_{jk} = \frac{1}{\sigma^2}\int_{-\infty}^{\infty} f_{jk}(-\infty, w)h(x)dx \quad (jk = 11, 12, 22),$$

where $w = (x-\mu)/\sigma$ and the $f_{jk}(-\infty, w)$ are defined in (5). David and Moeschberger (1978), Chapter 3 provide expressions for the elements of the Fisher information matrix for some particular models involving competing risks. This situation requires some mild extensions to the regularity conditions for the combination of distributions of y_r^c and Y. (See, for example, page 43 of Lawless (1982) for discussion and references.)

2.4. Failure censoring

For failure censoring (also known as Type II censoring), the smallest r_1 and the largest r_2 sample values are censored ($r_1 \geq 0$ and $r_2 \geq 0$, with $r_1 + r_2 \leq n$). The log-likelihood is given in terms of the order statistics which are not i.i.d., but the log-likelihood still has the general form given in Section 1.2 and $\mathcal{I}$ is defined by (2) with the understanding that $0 < \lim_{n\to\infty}(r_1/n) < \lim_{n\to\infty}(1 - r_2/n) < 1$. It is interesting that the elements of $\mathcal{I}$ are again given by (6) except that z_l^c and z_r^c are defined by $\Phi(z_l^c) = \lim_{n\to\infty}(r_1/n)$ and $\Phi(z_r^c) = \lim_{n\to\infty}(1 - r_2/n)$. The proof of this result is involved but it follows from Chernoff, Gastwirth, and Johns (1967), page 68 or Bhattacharyya (1985).

Harter and Moore (1966, 1967, 1968) give formulas and tabulate $\mathcal{I}$ for right and left failure censored samples from the normal, logistic, and extreme value distributions. For progressive failure censored data, where units are removed in stages determined by the times at which specified numbers of units have failed (see Lawless (1982), page 33), the elements of $I_{\underline{\theta}}$ can be expressed as sums of the individual $\mathcal{I}_{jk}$'s.

3. Fisher Information Matrix with Censored Data from Truncated Distributions

In this section we provide expressions for expected information based on censored data from truncated location-scale distributions.

3.1. Truncation model

We assume the same underlying location-scale distribution defined in (3). Now, however, we introduce left and right truncation points y_l^t and y_r^t ($y_l^t < y_r^t$)

and define the corresponding standardized left z_l^t and right z_r^t truncation quantiles as

$$z_l^t = \frac{y_l^t - \mu}{\sigma} < z_r^t = \frac{y_r^t - \mu}{\sigma}.$$

The standardized left and right censoring quantiles (z_l^c and z_r^c respectively) are as defined in (4) and we assume that $z_l^t \le z_l^c < z_r^c \le z_r^t$. The truncated cdf for $Y \mid Y \in (y_l^t, y_r^t)$ is

$$G_{Y|Y\in(y_l^t,y_r^t)}(y;\underline{\theta}) = \Pr\Big[Y \le y \mid Y \in (y_l^t, y_r^t)\Big] = \frac{\Phi(z) - \Phi(z_l^t)}{\Phi(z_r^t) - \Phi(z_l^t)}, y_l^t \le y \le y_r^t \quad (7)$$

and the corresponding truncated pdf is

$$\frac{d\, G_{Y|Y\in(y_l^t,y_r^t)}(y;\underline{\theta})}{dy} = \Big(\frac{1}{\sigma}\Big)\frac{\phi(z)}{\Phi(z_r^t) - \Phi(z_l^t)}, \quad y_l^t \le y \le y_r^t.$$

When $y_r^t \to \infty$ one gets the special case of only left truncated data. Similarly when $y_l^t \to -\infty$ the model yields the case of only right truncated data. When both $y_l^t \to -\infty$ and $y_r^t \to \infty$ equation (7) reduces to equation (3).

3.2. Elements of the Fisher information matrix

When z_l^t, z_r^t, z_l^c and z_r^c are the same for all observations in a sample, the elements of $I_{\underline{\theta}} = n\mathcal{I}$ for samples with truncation and censoring are obtained from

$$\mathcal{I}_{11} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial \mu^2}\Big] = \Big(\frac{1}{\sigma^2}\Big)\frac{1}{\Phi(z_r^t) - \Phi(z_l^t)}\Bigg\{f_{11}(z_l^c, z_r^c) - \frac{\phi^2(z_r^c)}{1 - \Phi(z_r^c)} - \frac{\phi^2(z_l^c)}{\Phi(z_l^c)}$$
$$-\frac{[\phi(z_r^t) - \phi(z_l^t)]^2}{\Phi(z_r^t) - \Phi(z_l^t)} + \frac{[\phi(z_l^c) - \phi(z_l^t)]^2}{\Phi(z_l^c) - \Phi(z_l^t)} + \frac{[\phi(z_r^t) - \phi(z_r^c)]^2}{\Phi(z_r^t) - \Phi(z_r^c)}\Bigg\}$$

$$\mathcal{I}_{12} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial \mu \partial \sigma}\Big] = \Big(\frac{1}{\sigma^2}\Big)\frac{1}{\Phi(z_r^t) - \Phi(z_l^t)}\Bigg\{f_{12}(z_l^c, z_r^c) - z_r^c\frac{\phi^2(z_r^c)}{1 - \Phi(z_r^c)} - z_l^c\frac{\phi^2(z_l^c)}{\Phi(z_l^c)}$$
$$-\frac{[\phi(z_r^t) - \phi(z_l^t)][z_r^t\phi(z_r^t) - z_l^t\phi(z_l^t)]}{\Phi(z_r^t) - \Phi(z_l^t)}$$
$$+\frac{[\phi(z_l^c) - \phi(z_l^t)][z_l^c\phi(z_l^c) - z_l^t\phi(z_l^t)]}{\Phi(z_l^c) - \Phi(z_l^t)}$$
$$+\frac{[\phi(z_r^t) - \phi(z_r^c)][z_r^t\phi(z_r^t) - z_r^c\phi(z_r^c)]}{\Phi(z_r^t) - \Phi(z_r^c)}\Bigg\}$$

$$\mathcal{I}_{22} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial \sigma^2}\Big] = \Big(\frac{1}{\sigma^2}\Big)\frac{1}{\Phi(z_r^t) - \Phi(z_l^t)}\Bigg\{f_{22}(z_l^c, z_r^c) - (z_r^c)^2\frac{\phi^2(z_r^c)}{1 - \Phi(z_r^c)}$$
$$-(z_l^c)^2\frac{\phi^2(z_l^c)}{\Phi(z_l^c)} - \frac{[z_r^t\phi(z_r^t) - z_l^t\phi(z_l^t)]^2}{\Phi(z_r^t) - \Phi(z_l^t)}$$

$$+\frac{[z_l^c\phi(z_l^c) - z_l^t\phi(z_l^t)]^2}{\Phi(z_l^c) - \Phi(z_l^t)} + \frac{[z_r^t\phi(z_r^t) - z_r^c\phi(z_r^c)]^2}{\Phi(z_r^t) - \Phi(z_r^c)}\Bigg\}.$$

The derivation is given in Appendix B. As in Section 2, when the values of z_l^t, z_r^t, z_l^c or z_r^c differ in a sample, the elements of $I_{\underline{\theta}}$ can be expressed as sums of the individual $\mathcal{I}_{jk}$'s.

3.3. Special cases

There are a number of special cases of the expressions in Section 3.2 and these are easily obtained, depending on the desired case, by allowing,

- If $z_r^t \to \infty$ there is no right truncation.
- If $z_l^t \to -\infty$ there is no left truncation.
- If $z_r^c \uparrow z_r^t$ there is no right censoring.
- If $z_l^c \downarrow z_l^t$ there is no left censoring.

The expressions for these special cases are obtained directly from the equations above by noting that, under the regularity conditions given in Section (2.1), $z^2\phi^2(z)/\Phi(z)$ and $z^2\phi^2(z)/[1-\Phi(z)]$ approach 0 as $z \to \pm\infty$ and $[\phi(z_1) - \phi(z_2)]^2/[\Phi(z_1)-\Phi(z_2)]$ and $[z_1\phi(z_1) - z_2\phi(z_2)]^2/[\Phi(z_1)-\Phi(z_2)]$ approach 0 as $|z_1 - z_2| \to 0$.

4. Fisher Information Matrix for the DOA and LFP Models

In this section we provide expressions for expected information for models that have atoms of probability at one or both ends of the sample space of the response. These expressions again depend on the elements in Equation (6). The "Dead on Arrival" (DOA) model has a proportion of units that have already failed at time 0 (or $-\infty$ on the log scale), putting an atom of probability at that point. The "Limited Failure Population" (LFP) model has a proportion of units that will never fail, putting an atom of probability at ∞.

4.1. The DOA/LFP model

When the DOA and LFP models arise, the response is typically time T and usually we would have $Y = \log(T)$. For compactness we combine the DOA and LFP models and consider cases where Y follows a location-scale distribution and there is left censoring at y_l^c and right censoring at y_r^c. The DOA/LFP model has a cdf $G(y) = p_0 + p_1(1-p_0)\Phi(z)$ where $\underline{\theta} = (\mu, \sigma, p_0, p_1)^T$, $z = (y-\mu)/\sigma$, p_0 is the proportion of units "dead on arrival", and p_1 is the proportion of units that will eventually fail from the "good on arrival" population. The corresponding pdf of Y is

$$\frac{d\,G(y)}{dy} = \left(\frac{1}{\sigma}\right) p_1(1-p_0)\phi(z).$$

At time y_l^c (> 0) we cannot distinguish between "failures" that were DOA and those that failed between 0 and y_l^c. At time y_r^c we cannot distinguish between units that will and those that will not "fail" after y_r^c. The standardized censoring quantiles z_l^c and z_r^c are defined as in Section 2. Two special cases of the model are:

- The LFP model described by Meeker (1987) is obtained when $p_0 = 0$ giving $G(y) = p_1\Phi(z)$.
- The DOA model described in Nelson (1982), page 52 is obtained when $p_1 = 1$ giving $G(y) = p_0 + (1 - p_0)\Phi(z)$.

4.2. Elements of the Fisher information matrix

When z_l^c and z_r^c are the same for all observations, the elements of the information matrix of $\underline{\theta} = (\mu, \sigma, p_0, p_1)$, $I_{\underline{\theta}} = n\mathcal{I}$ are obtained from

$$\mathcal{I} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial\underline{\theta}\partial\underline{\theta}^T}\Big] = \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{pmatrix},$$

where

$$B_{11} = \frac{p_1(1-p_0)}{\sigma^2}\Big[\begin{pmatrix} f_{11}(z_l^c, z_r^c) & f_{12}(z_l^c, z_r^c) \\ f_{12}(z_l^c, z_r^c) & f_{22}(z_l^c, z_r^c) \end{pmatrix} - \frac{p_0\phi^2(z_l^c)}{\Phi(z_l^c)G(z_l^c)}\begin{pmatrix} 1 & z_l^c \\ z_l^c & (z_l^c)^2 \end{pmatrix}$$
$$- \frac{(1-p_1)\phi^2(z_r^c)}{[1-\Phi(z_r^c)][1-p_1\Phi(z_r^c)]}\begin{pmatrix} 1 & z_r^c \\ z_r^c & (z_r^c)^2 \end{pmatrix}\Big]$$

$$B_{12} = -\Big(\frac{1}{\sigma}\Big)\begin{bmatrix} \dfrac{p_1\phi(z_l^c)}{G(z_l^c)} & (1-p_0)\Big(\dfrac{\phi(z_r^c)}{1-p_1\Phi(z_r^c)} - \dfrac{p_0\phi(z_l^c)}{G(z_l^c)}\Big) \\ \dfrac{z_l^c p_1\phi(z_l^c)}{G(z_l^c)} & (1-p_0)\Big(\dfrac{z_r^c\phi(z_r^c)}{1-p_1\Phi(z_r^c)} - \dfrac{z_l^c p_0\phi(z_l^c)}{G(z_l^c)}\Big) \end{bmatrix}$$

$$B_{22} = \begin{bmatrix} \dfrac{1-p_1\Phi(z_l^c)}{(1-p_0)G(z_l^c)} & \dfrac{\Phi(z_l^c)}{G(z_l^c)} \\ \dfrac{\Phi(z_l^c)}{G(z_l^c)} & \dfrac{(1-p_0)}{p_1}\Big(\dfrac{\Phi(z_r^c)}{1-p_1\Phi(z_r^c)} - \dfrac{p_0\Phi(z_l^c)}{G(z_l^c)}\Big) \end{bmatrix}.$$

The derivation is given in Appendix C. As in Section 2, when the values of z_l^c or z_r^c differ in a sample, the elements of $I_{\underline{\theta}}$ can be expressed as sums of the individual $\mathcal{I}_{jk}$'s.

4.3. Special cases

The LFP model: In this case $p_0 = 0$, $\underline{\theta} = (\mu, \sigma, p_1)^T$, and the 3×3 Fisher information matrix is made up of B_{11} (with $p_0 = 0$), the second column of B_{12}

and the (2,2) element of B_{22}, giving

$$\mathcal{I}=\left[\begin{array}{c|c} \frac{p_1}{\sigma^2}\Big\{\begin{pmatrix} f_{11}(z_l^c,z_r^c) & f_{12}(z_l^c,z_r^c)\\ f_{12}(z_l^c,z_r^c) & f_{22}(z_l^c,z_r^c)\end{pmatrix}-\frac{(1-p_1)\phi^2(z_r^c)}{[1-\Phi(z_r^c)][1-p_1\Phi(z_r^c)]}\begin{pmatrix} 1 & z_r^c\\ z_r^c & (z_r^c)^2\end{pmatrix}\Big\} & \frac{1}{\sigma}\begin{pmatrix}\frac{-\phi(z_r^c)}{1-p_1\Phi(z_r^c)}\\ \frac{-z_r^c\phi(z_r^c)}{1-p_1\Phi(z_r^c)}\end{pmatrix}\\ \hline \text{symmetric} & \frac{1}{p_1}\Big(\frac{\Phi(z_r^c)}{1-p_1\Phi(z_r^c)}\Big)\end{array}\right].$$

This Fisher information matrix is for the LFP model with right and left censored data. It contains as a particular case the Fisher information matrix for the LFP model with right censored data given by Meeker (1987). Observe, that the left censoring point enters into $\mathcal{I}$ only through the f_{jk}'s. It is interesting that the left censoring affects only the (μ,σ) entries but it does not affect any entry involving p_1.

The DOA model: In this case $p_1=1$, $\underline{\theta}=(\mu,\sigma,p_0)^T$, and the 3×3 Fisher information matrix is made up of B_{11} (with $p_0=1$), the first column of B_{12} and the $(1,1)$ element of B_{22}, giving

$$\mathcal{I}=\left[\begin{array}{c|c} \frac{1-p_0}{\sigma^2}\Big\{\begin{pmatrix} f_{11}(z_l^c,z_r^c) & f_{12}(z_l^c,z_r^c)\\ f_{12}(z_l^c,z_r^c) & f_{22}(z_l^c,z_r^c)\end{pmatrix}-\frac{p_0\phi^2(z_l^c)}{\Phi(z_l^c)G(z_l^c)}\begin{pmatrix} 1 & z_l^c\\ z_l^c & (z_l^c)^2\end{pmatrix}\Big\} & \frac{1}{\sigma}\begin{pmatrix}\frac{-\phi(z_l^c)}{G(z_l^c)}\\ \frac{-z_l^c\phi(z_l^c)}{G(z_l^c)}\end{pmatrix}\\ \hline \text{symmetric} & \frac{1}{1-p_0}\Big(\frac{1-\Phi(z_l^c)}{G(z_l^c)}\Big)\end{array}\right].$$

This Fisher information matrix is for the DOA model with left and right censoring and it contains as a particular case the Fisher information matrix for the DOA model with left censored data.

5. Fisher Information Matrix for Location-Scale Regression Models

In this section we provide expressions for expected information based on censored data with regression models.

5.1. Location/scale regression model

The assumed cdf for Y, allowing the location parameter μ and the scale parameter σ to depend on (a possibly different set of) explanatory variables, is

$$G_Y(y)=p_0+(1-p_0)\Phi\Big[\frac{y-\mu(\underline{x})}{\sigma(\underline{w})}\Big],$$

where $\mu(\underline{x})=\underline{x}^T\underline{\beta}$, $\log[\sigma(\underline{w})]=\underline{w}^T\underline{\gamma}$, $\underline{x}^T=(x_0,\ldots,x_r)$ and $\underline{w}^T=(w_0,\ldots,w_s)$ (with $w_0=1$) are vectors of explanatory variables, $\underline{\beta}^T=(\beta_0,\ldots,\beta_r)$ and $\underline{\gamma}^T=(\gamma_0,\ldots,\gamma_s)$ are vectors of unknown constants. We also assume, without loss of generality, that all of the explanatory variables have been standardized (i.e., $0 \le x_i \le 1$ and $0 \le w_j \le 1$ for all i, j). For convenience, we reparameterize the scale parameter $\sigma = \sigma(\underline{w})$ as follows. Let $\sigma_0 = \exp(\gamma_0)$ and $\sigma_j = \exp(\gamma_0 + \gamma_j)$, $j = 1, \ldots, s$. Then $\log(\sigma) = (2 - \underline{w}^T \underline{1}) \log(\sigma_0) + \sum_{j=1}^{s} w_j \log(\sigma_j)$, or equivalently,

$$\sigma = \sigma_0^{(2-\underline{w}^T\underline{1})} \prod_{j=1}^{s} \sigma_j^{w_j},$$

where $\underline{1}$ is a vector with all the components equal to 1. Observe that σ_0 is the scale parameter at $\underline{w}_0^T = (1, 0, \ldots, 0)$ and σ_j is the scale parameter at $\underline{w}_j^T = (1, 0, \ldots, 1, 0, \ldots, 0)$ where the second 1 is at the $j+1$ position. Let $\underline{\sigma}^T = (\sigma_0, \ldots, \sigma_s)$ and define the generic parameter vector $\underline{\theta}^T = (\underline{\beta}^T, \underline{\sigma}^T, p_0, p_1)$.

5.2. Elements of the Fisher information matrix

Let $\mathcal{L}_i(\underline{\theta}, \underline{x}, \underline{w})$ denote the contribution of the ith observation to the log-likelihood, possibly left or right censored. The Fisher information matrix for this observation is

$$\mathcal{I} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial \underline{\theta} \partial \underline{\theta}^T} \Big] = \begin{bmatrix} M B_{11} M^T & M B_{12} \\ (M B_{12})^T & B_{22} \end{bmatrix}, \tag{8}$$

where

$$M = \begin{bmatrix} \underline{x} & \underline{0} \\ \underline{0} & \rho \underline{\eta} \end{bmatrix}$$

with

$$\rho = \rho(\underline{w}) = \frac{\sigma}{\sigma_0} = \prod_{j=1}^{s} \upsilon_j^{w_j}, \quad \upsilon_j = \sigma_j / \sigma_0, \quad \underline{\eta} = \begin{bmatrix} 2 - \underline{w}^T \underline{1} \\ w_1/\upsilon_1 \\ \vdots \\ w_s/\upsilon_s \end{bmatrix}. \tag{9}$$

The derivation of these elements is contained in Appendix D. For observations that might also be truncated, one can simply substitute the values of $\sigma^2 \mathcal{I}_{jk}(z_l^c, z_r^c, z_l^t, z_r^t)$ from Section 3.2 for the $f_{jk}(z_l^c, z_r^c)$'s in (8).

5.3. Special cases

- When $p_0 = 0, p_1 = 1$, we get the simpler form

$$\mathcal{I} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial \underline{\theta} \partial \underline{\theta}^T} \Big] = \frac{1}{\sigma_0^2} \begin{bmatrix} \Big(\frac{1}{\rho^2}\Big) f_{11}(z_l^c, z_r^c)\, \underline{x}\, \underline{x}^T & \Big(\frac{1}{\rho}\Big) f_{12}(z_l^c, z_r^c)\, \underline{x}\, \underline{\eta}^T \\ \Big(\frac{1}{\rho}\Big) f_{12}(z_l^c, z_r^c)\, \underline{\eta}\, \underline{x}^T & f_{22}(z_l^c, z_r^c)\, \underline{\eta}\, \underline{\eta}^T \end{bmatrix}.$$

Other important special cases (still with $p_0 = 0, p_1 = 1$) that have been used previously in the literature are:

- When $r = s = 0$ then $\underline{x} = \underline{w} = 1$ (scalars), $\rho = 1$, and we get the expressions for the single distribution model given in Section 2.
- When $r = 1$, $s = 0$ then $\underline{x} = (1, x_1)^T$, $\underline{w} = 1$ (a scalar), $\rho = 1$, and we have the simple regression model used in Nelson and Kielpinski (1976), Nelson (1990) and elsewhere for planning single-factor accelerated life tests.
- When $r = k$, $s = 0$ then $\underline{x} = (1, x_1, \ldots, x_k)^T$, $\underline{w} = 1$ (a scalar), $\rho = 1$, and we have the model used in Escobar and Meeker (1995) for planning k-factor accelerated life tests.
- When $r = s = 1$ then $\underline{x} = \underline{w} = (1, x)^T$, $\rho = \sigma/\sigma_0$, and we get the special case model used in Meeter and Meeker (1994) for planning single-factor accelerated life tests with nonconstant location and scale parameters.

6. Further Generalizations and Applications

The general ideas described in this paper can be extended readily to other situations. In particular,

- The Escobar and Meeker (1994) algorithm assumes that uncensored data are observed exactly. When data are binned or when failures are discovered at inspection times, data will be interval censored. An algorithm, similar to the one given in Escobar and Meeker (1994), which specifies the standardized inspection quantiles, could be used in the general formulas derived in this paper. See Meeker (1986) for expressions for the basic elements of the Fisher information matrix for this case.
- The regression model used in Section 5 assumed linear and loglinear relationships, respectively, for μ and σ. For nonlinear regression relationships $\mu = \mu(\underline{\beta})$ and $\sigma = \sigma(\underline{\gamma})$, if one replaces $\underline{x}$ with $\partial\mu/\partial\underline{\beta}$ and $\underline{w}$ with $(1/\sigma)\partial\sigma/\partial\underline{\gamma}$, the expressions in this section allow straightforward computation of the asymptotic covariance matrix for the parameters of the nonlinear regression model.
- The results in this paper could also be used to obtain optimum test plans with censoring using a Bayesian-like prior as described in, for example, Chaloner and Larntz (1992). The basic idea is optimize by using the prior to average over needed "planning" values of model parameters.

Acknowledgements

We would like to thank Herbert A. David, Shuen-Lin Jeng, Michael Hamada, Wayne Nelson, two anonymous referees, and an associate editor for making helpful comments on earlier versions of this paper.

Appendix

A. Regularity Conditions

Let $z = (y - \mu)/\sigma$ and $\underline{\theta}^T = (\mu, \sigma)$, then the following regularity conditions are assumed to hold:

- $\phi(z) > 0$ for all $-\infty < z < \infty$.
- $\lim_{z\to\pm\infty} z^2\phi'(x) = 0$, where $\phi'(\cdot)$ is the first derivative of $\phi(\cdot)$.
- The second derivative $\phi''(\cdot)$ is continuous.
- The expected values $E\{-\partial^2 \log /\partial\underline{\theta}\ \partial\underline{\theta}^T\}$ are all finite.

These regularity conditions are sufficient to guarantee the correctness of the asymptotic covariance matrices presented in this paper and are satisfied by many location-scale families including the normal, smallest and largest extreme value, and logistic distributions.

B. Derivation of the Fisher Information Matrix for Censored Samples from Truncated Distributions

Let $\Delta(v,u) = \Phi(u) - \Phi(v), \Phi_L = \Phi(z_l^c), \Phi_R = \Phi(z_r^c), \phi_L = \phi(z_l^c), \phi_R = \phi(z_r^c)$. The log-likelihood contribution of one observation at y is

$$\begin{aligned}\mathcal{L}_i &= \Psi_{(z_l^t,z_l^c)} \log\Big[\frac{\Delta(z_l^t,z_l^c)}{\Delta(z_l^t,z_r^t)}\Big] + \Psi_{(z_l^c,z_r^c)} \log\Big[\frac{\phi(z)/\sigma}{\Delta(z_l^t,z_r^t)}\Big] + \Psi_{(z_r^c,z_r^t)} \log\Big[\frac{\Delta(z_r^c,z_r^t)}{\Delta(z_l^t,z_r^t)}\Big] \\ &= \Psi_{(z_l^t,z_l^c)} \log\Big[\Delta(z_l^t,z_l^c)\Big] + \Psi_{(z_l^c,z_r^c)} \log\Big[\phi(z)/\sigma\Big] + \Psi_{(z_r^c,z_r^t)} \log\Big[\Delta(z_r^c,z_r^t)\Big] \\ &\quad -\Psi_{(z_l^t,z_r^t)} \log\Big[\Delta(z_l^t,z_r^t)\Big], \end{aligned} \tag{10}$$

where Ψ_H denotes the indicator function of the set H.

To obtain the needed expectations with respect to the truncated distribution of Z, we use the following result. If W is a function of the random variable Z then

$$E\Big[W \mid Z \in (z_l^t, z_r^t)\Big] = \frac{1}{\Delta(z_l^t,z_r^t)} E\Big[\Psi_{(z_l^t,z_r^t)} W\Big], \tag{11}$$

where the second expectation is computed with respect to the unconditional distribution of Z. Then letting $\underline{\theta}$ be the vector $(\mu,\sigma)^T$, under the regularity conditions,

$$\begin{aligned}\mathcal{I} = E\Big[-\frac{\partial^2\mathcal{L}_i}{\partial\underline{\theta}\,\partial\underline{\theta}^T}\Big| Z \in (z_l^t,z_r^t)\Big] &= \frac{1}{\Delta(z_l^t,z_r^t)} E\Big[-\Psi_{(z_l^t,z_r^t)} \frac{\partial^2\mathcal{L}_i}{\partial\underline{\theta}\,\partial\underline{\theta}^T}\Big] \\ &= \frac{1}{\Delta(z_l^t,z_r^t)} E\Big[-\frac{\partial^2\mathcal{L}_i}{\partial\underline{\theta}\,\partial\underline{\theta}^T}\Big] \\ &= \frac{1}{\Delta(z_l^t,z_r^t)} E\Big[\Big(\frac{\partial\mathcal{L}_i}{\partial\underline{\theta}}\Big)^T\Big(\frac{\partial\mathcal{L}_i}{\partial\underline{\theta}}\Big)\Big]. \end{aligned} \tag{12}$$

The derivatives on the right-hand side (RHS) of (12) are obtained by taking term-wise expectations of gradient vector inner products. The expectation corresponding to the second term on the RHS of (10) can be obtained from the untruncated expectations (e.g. equations 9.3.7a, b, c in David (1981)) by subtracting out the terms corresponding to left and right censoring. Thus

$$E\Big[\Psi_{(z_l^c,z_r^c)}\Big(\frac{\partial}{\partial\underline{\theta}}\Big\{\log\Big[\frac{\phi(z)}{\sigma}\Big]\Big\}\Big)^T\Big(\frac{\partial}{\partial\underline{\theta}}\Big\{\log\Big[\frac{\phi(z)}{\sigma}\Big]\Big\}\Big)\Big]$$
$$=\Big(\frac{1}{\sigma^2}\Big)\begin{bmatrix} f_{11}(z_l^c,z_r^c) & f_{12}(z_l^c,z_r^c)\\ f_{12}(z_l^c,z_r^c) & f_{22}(z_l^c,z_r^c)\end{bmatrix}$$
$$-\frac{1}{\Phi_L}\Big[\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big]-\frac{1}{1-\Phi_R}\Big[\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big]. \tag{13}$$

For the other terms in (12), direct computations give

$$E\Big[\Psi_{(u,v)}\Big(\frac{\partial}{\partial\underline{\theta}}\Big\{\log[\Delta]\Big\}\Big)^T\Big(\frac{\partial}{\partial\underline{\theta}}\Big\{\log[\Delta]\Big\}\Big)\Big]=\frac{1}{\Delta}\Big[\frac{\partial\Delta}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Delta}{\partial\underline{\theta}}\Big] \tag{14}$$
$$E\Big[\Psi_{(z_l^c,z_r^c)}\frac{\partial}{\partial\underline{\theta}}\Big\{\log[\frac{\phi(z)}{\sigma}]\Big\}\Big]=\frac{\partial}{\partial\underline{\theta}}\Delta(z_l^c,z_r^c), \tag{15}$$

where, for simplicity, we use $\Delta=\Delta(u,v)$. Substitution into (12) gives

$$\mathcal{I}=\frac{1}{\Delta(z_l^t,z_r^t)}\Big\{\Big(\frac{1}{\sigma^2}\Big)\begin{bmatrix} f_{11}(z_l^c,z_r^c) & f_{12}(z_l^c,z_r^c)\\ f_{12}(z_l^c,z_r^c) & f_{22}(z_l^c,z_r^c)\end{bmatrix}-\frac{1}{\Phi_L}\Big[\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big]$$
$$-\frac{1}{1-\Phi_R}\Big[\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big]+\frac{1}{\Delta(z_l^t,z_l^c)}\Big[\frac{\partial\Delta(z_l^t,z_l^c)}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Delta(z_l^t,z_l^c)}{\partial\underline{\theta}}\Big]$$
$$+\frac{1}{\Delta(z_r^c,z_r^t)}\Big[\frac{\partial\Delta(z_r^c,z_r^t)}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Delta(z_r^c,z_r^t)}{\partial\underline{\theta}}\Big]-\frac{1}{\Delta(z_l^t,z_r^t)}\Big[\frac{\partial\Delta(z_l^t,z_r^t)}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Delta(z_l^t,z_r^t)}{\partial\underline{\theta}}\Big]\Big\}.$$

Then using

$$\Big[\frac{\partial\Delta(u,v)}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Delta(u,v)}{\partial\underline{\theta}}\Big]=\Big(\frac{1}{\sigma^2}\Big)\begin{bmatrix}[\phi(v)-\phi(u)]^2 & [v\phi(v)-u\phi(u)][\phi(v)-\phi(u)]\\ \text{symmetric} & [v\phi(v)-u\phi(u)]^2\end{bmatrix}$$

in $\mathcal{I}$ above gives the general form of the Fisher information matrix elements given in Section 3.2.

C. Derivation of the Fisher Information Matrix for the DOA/LFP Model

Here, we use the additional notation, $G_L=G(z_l^c)$, $G_R=G(z_r^c)$, and $\underline{\theta}=(\mu,\sigma,p_0,p_1)^T$. The log-likelihood contribution of one observation at y is

$$\mathcal{L}_i=\Psi_{(-\infty,z_l^c)}\log(G_L)+\Psi_{(z_l^c,z_r^c)}\log[p_1(1-p_0)\phi(z)/\sigma]+\Psi_{(z_r^c,\infty)}\log(1-G_R)$$
$$=\Psi_{(-\infty,z_l^c)}\log(G_L)+\Psi_{(z_l^c,z_r^c)}\log[\phi(z)/\sigma]+\Psi_{(z_l^c,z_r^c)}\log[p_1(1-p_0)]$$
$$+\Psi_{(z_r^c,\infty)}\log(1-G_R). \tag{16}$$

In the computation of the Fisher matrix, we use the following result. Let W be a function of Y defined in a finite interval; then because of the form of the density of Y, it follows that

$$E_Y[W] = p_1(1-p_0)E[W], \tag{17}$$

where the first expectation is computed with respect to the density $d\,G_Y(y;\underline{\theta})/dy$ and the second expectation is computed with the density $\phi(z)/\sigma$.

Under the regularity conditions, and using (17), one gets the contribution to the Fisher information matrix for a single observation:

$$\begin{aligned}\mathcal{I} = E_Y\Big[-\frac{\partial^2\mathcal{L}_i}{\partial\underline{\theta}\,\partial\underline{\theta}^T}\Big] &= E_Y\Big[\Big(\frac{\partial\mathcal{L}_i}{\partial\underline{\theta}}\Big)^T\Big(\frac{\partial\mathcal{L}_i}{\partial\underline{\theta}}\Big)\Big]\\ &= p_1(1-p_0)E\Big[\Big(\frac{\partial\mathcal{L}_i}{\partial\underline{\theta}}\Big)^T\Big(\frac{\partial\mathcal{L}_i}{\partial\underline{\theta}}\Big)\Big]. \end{aligned}\tag{18}$$

The derivatives on the RHS of (18) are obtained from (16). With the exception of the term (19) below, the expected values in the RHS of (18) are easy to compute.

As in (13)

$$\begin{aligned}&E\Big[\Psi_{(z_l^c,z_r^c)}\Big(\frac{\partial}{\partial\underline{\theta}}\Big\{\log\Big[\frac{\phi(z)}{\sigma}\Big]\Big\}\Big)^T\Big(\frac{\partial}{\partial\underline{\theta}}\Big\{\log\Big[\frac{\phi(z)}{\sigma}\Big]\Big\}\Big)\Big] =\\ &\Big(\frac{1}{\sigma^2}\Big)\begin{bmatrix} f_{11}(z_l^c,z_r^c) & f_{12}(z_l^c,z_r^c) & O\\ f_{12}(z_l^c,z_r^c) & f_{22}(z_l^c,z_r^c) & O\\ O & O & O\end{bmatrix}\\ &-\frac{1}{\Phi_L}\Big(\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big)^T\Big(\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big) - \frac{1}{1-\Phi_R}\Big(\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big)^T\Big(\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big),\end{aligned}\tag{19}$$

where the O's indicate matrices of 0's appropriately dimensioned to complete this 4×4 matrix. Then after simplification, one gets

$$\begin{aligned}\mathcal{I} = p_1(1-p_0)\Big\{&\Big(\frac{1}{\sigma^2}\Big)\begin{bmatrix} f_{11}(z_l^c,z_r^c) & f_{12}(z_l^c,z_r^c) & O\\ f_{12}(z_l^c,z_r^c) & f_{22}(z_l^c,z_r^c) & O\\ O & O & O\end{bmatrix} - \frac{1}{\Phi_L}\Big[\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Phi_L}{\partial\underline{\theta}}\Big]\\ &-\frac{1}{1-\Phi_R}\Big[\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial\Phi_R}{\partial\underline{\theta}}\Big]\Big\}\\ &+\Big[\frac{\partial}{\partial\underline{\theta}}\Big\{\Phi_R-\Phi_L\Big\}\Big]^T\Big[\frac{\partial}{\partial\underline{\theta}}\Big\{p_1(1-p_0)\Big\}\Big] + \Big[\frac{\partial}{\partial\underline{\theta}}\Big\{p_1(1-p_0)\Big\}\Big]^T\Big[\frac{\partial}{\partial\underline{\theta}}\Big\{\Phi_R-\Phi_L\Big\}\Big]\\ &+\frac{\Phi_R-\Phi_L}{p_1(1-p_0)}\Big[\frac{\partial}{\partial\underline{\theta}}\Big\{p_1(1-p_0)\Big\}\Big]^T\Big[\frac{\partial}{\partial\underline{\theta}}\Big\{p_1(1-p_0)\Big\}\Big]\\ &+\frac{1}{G_L}\Big[\frac{\partial G_L}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial G_L}{\partial\underline{\theta}}\Big] + \frac{1}{1-G_R}\Big[\frac{\partial G_R}{\partial\underline{\theta}}\Big]^T\Big[\frac{\partial G_R}{\partial\underline{\theta}}\Big].\end{aligned}$$

Straightforward computations yield,

$$\Big[\frac{\partial \Phi_L}{\partial \underline{\theta}}\Big]^T = \Big(-\frac{\phi_L}{\sigma}\Big)\begin{bmatrix}1\\ z_l^c\\ 0\\ 0\end{bmatrix}, \quad \Big[\frac{\partial}{\partial \underline{\theta}}\Big\{p_1(1-p_0)\Big\}\Big]^T = \begin{bmatrix}0\\ 0\\ -p_1\\ 1-p_0\end{bmatrix}$$

$$\Big[\frac{\partial G_L}{\partial \underline{\theta}}\Big]^T = \begin{bmatrix} p_1(1-p_0)\Big(-\frac{\phi_L}{\sigma}\Big)\begin{Bmatrix}1\\ z_l^c\end{Bmatrix}\\ 1-p_1\Phi_L\\ (1-p_0)\Phi_L\end{bmatrix}$$

and similar expressions for

$$\Big[\frac{\partial \Phi_R}{\partial \underline{\theta}}\Big]^T \quad \text{and} \quad \Big[\frac{\partial G_R}{\partial \underline{\theta}}\Big]^T.$$

Using these expressions in $\mathcal{I}$ above, one gets the expressions for the B_{jk}'s given in Section 4.2.

D. Derivation of the Fisher Information Matrix for the Location/Scale Regression Model

Let $\mathcal{L}_i(\underline{\theta}, \underline{x}, \underline{w}) = L_i[g(\underline{\theta}, \underline{x}, \underline{w})]$, where $\underline{g}(\underline{\theta}, \underline{x}, \underline{w}) = [\mu(\underline{x}), \sigma(\underline{w}), p_0, p_1]^T$, and L_i denotes the ordinary log-likelihood function for one observation with parameters $(\mu(\underline{x}), \sigma(\underline{w}), p_0, p_1)$. Using a matrix form for the chain rule, we get

$$\frac{\partial \mathcal{L}_i}{\partial \underline{\theta}} = \Big[\frac{\partial L_i}{\partial \underline{g}}\Big]\Big[\frac{\partial \underline{g}}{\partial \underline{\theta}}\Big]$$

and

$$\frac{\partial^2 \mathcal{L}_i}{\partial \underline{\theta}\partial \underline{\theta}^T} = \Big[\frac{\partial \underline{g}}{\partial \underline{\theta}}\Big]^T\Big[\frac{\partial^2 L_i}{\partial \underline{g}\partial \underline{g}^T}\Big]\Big[\frac{\partial \underline{g}}{\partial \underline{\theta}}\Big] + \sum_{k=1}^{2}\Big[\frac{\partial L_i}{\partial g_k}\Big]\Big[\frac{\partial^2 g_k}{\partial \underline{\theta}\partial \underline{\theta}^T}\Big],$$

where $g_k = g_k(\underline{\theta})$ denotes the kth component of the $\underline{g}$ function (i.e., $g_1 = \mu(\underline{x})$, $g_2 = \sigma(\underline{w})$), etc. Taking expectations, we get

$$\mathcal{I} = E\Big[-\frac{\partial^2 \mathcal{L}_i}{\partial \underline{\theta}\partial \underline{\theta}^T}\Big] = \Big[\frac{\partial \underline{g}}{\partial \underline{\theta}}\Big]^T E\Big[-\frac{\partial^2 L_i}{\partial \underline{g}\partial \underline{g}^T}\Big]\Big[\frac{\partial \underline{g}}{\partial \underline{\theta}}\Big] \tag{20}$$

because

$$E\Big[\frac{\partial L_i}{\partial g_k}\Big] = 0, \ k = 1, 2.$$

Straightforward computations yield

$$\frac{\partial \mu}{\partial \underline{\beta}} = \underline{x}^T, \quad \frac{\partial \sigma}{\partial \sigma_0} = \Big(\frac{\sigma}{\sigma_0}\Big)(2 - \underline{w}^T\underline{1}), \quad \frac{\partial \sigma}{\partial \sigma_j} = \Big(\frac{\sigma}{\sigma_0}\Big)\Big(\frac{w_j}{\upsilon_j}\Big), \quad j = 1, \ldots, s,$$

where $\upsilon_j = \sigma_j/\sigma_0$. Thus, $\partial\sigma/\partial\underline{\sigma} = (\sigma/\sigma_0)\underline{\eta}^T$. As in Section 4.2,

$$E\Big[-\frac{\partial^2 L_i}{\partial \underline{g}\partial \underline{g}^T}\Big] = \begin{bmatrix} B_{11} & B_{12} \\ B_{12}^T & B_{22} \end{bmatrix}$$

and from above

$$\frac{\partial \underline{g}}{\partial \underline{\theta}} = \begin{bmatrix} \underline{x} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \rho\,\underline{\eta} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & 1 & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & 1 \end{bmatrix} = \begin{bmatrix} M & O \\ O & \mathcal{J} \end{bmatrix},$$

where $\mathcal{J}$ is a 2×2 identity matrix. Substituting into (20) gives (8), where $\underline{\eta}$ is defined in (9).

References

Bhattacharyya, G. K. (1985). The asymptotics of maximum likelihood and related estimators based on type II censored data. *J. Amer. Statist. Assoc.* **80**, 398-404.

Chaloner, K. and Larntz, K. (1992). Bayesian design for accelerated life testing. *J. Statist. Plann. Inference* **33**, 245-259.

Chernoff, H., Gastwirth, J. L. and Johns, Jr., M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist.* **38**, 52-72.

David, H. A. (1981). *Order Statistics*, 2nd Edition. John Wiley, New York.

David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks.* Griffin, London.

Escobar, L. A. and Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507-528.

Escobar, L. A. and Meeker, W. Q. (1994). Fisher information matrix for the extreme value, normal, and logistic distributions and censored data. *Appl. Statist.* **43**, 533-540.

Escobar, L. A. and Meeker, W. Q. (1995). Planning accelerated life test plans with two or more experimental factors. *Technometrics* **37**, 411-427.

Harter, H. L. and Moore, A. H. (1966). Iterative maximum–likelihood estimation of the parameters of normal populations from singly and doubly censored samples. *Biometrika* **53**, 205-213.

Harter, H. L. and Moore, A. H. (1967). Maximum–likelihood estimation, from censored samples, of the parameters of a logistic distribution. *J. Amer. Statist. Assoc.* **62**, 675-684.

Harter, H. L. and Moore, A. H. (1968). Maximum–likelihood estimation, from doubly censored samples, of the parameters of the first asymptotic distribution of extreme values. *J. Amer. Statist. Assoc.* **63**, 889-901.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* John Wiley, New York.

Lehmann, E. L. (1983). *Theory of Point Estimation.* John Wiley, New York.

Meeker, W. Q. (1986). Planning life tests in which units are inspected for failure. *IEEE Trans. Reliability* **R-35**, 571-578.

Meeker, W. Q. (1987). Limited failure population life tests: Application to integrated circuit reliability. *Technometrics* **29**, 51-65.

Meeker, W. Q., Escobar, L. A. and Hill, D. A. (1992). Sample sizes for estimating the weibull hazard function from censored samples. *IEEE Trans. Reliability* **41**, 133-138.

Meeter, C. A. and Meeker, W. Q. (1994). Optimum accelerated life tests with a nonconstant scale parameter. *Technometrics* **36**, 71-83.

Nelson, W. (1982). *Applied Life Data Analysis*. John Wiley, New York.

Nelson, W. (1990). *Accelerated Testing*. John Wiley, New York.

Nelson, W. and Kielpinski, T. (1976). Theory for optimum censored accelerated life tests for normal and lognormal distributions. *Technometrics* **18**, 105-114.

Nelson, W. and Meeker, W. Q. (1978). Theory for optimum accelerated censored life tests for weibull and extreme value distributions. *Technometrics* **20**, 171-177.

Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

E-mail: luis@stat.lsu.edu

Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

E-mail: wqmeeker@iastate.edu

(Received February 1995; accepted September 1996)