

The Asymptotic Equivalence of the Fisher Information Matrices for Type I and Type II Censored Data from Location–Scale Families

Luis A. Escobar
Dept. of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803

William Q. Meeker
Dept. of Statistics
Iowa State University
Ames, IA 50011

February 17, 2000

Abstract

Type I and Type II censored data arise frequently in controlled laboratory studies concerning time to a particular event (e.g., death of an animal or failure of a physical device). Log-location-scale distributions (e.g., Weibull, lognormal, and loglogistic) are commonly used to model the resulting data. Maximum likelihood (ML) is generally used to obtain parameter estimates when the data are censored. The Fisher information matrix can be used to obtain large-sample approximate variances and covariances of the ML estimates or to estimate these variances and covariances from data. The derivations of the Fisher information matrix proceed differently for Type I (time censoring) and Type II (failure censoring) because the number of failures is random in Type I censoring, but length of the data collection period is random in Type II censoring. Under regularity conditions (met with the above-mentioned log-location-scale distributions), we outline the different derivations and show that the Fisher information matrices for Type I and Type II censoring are asymptotically equivalent.

Key Words: censoring, maximum likelihood, survival.

1 Introduction

1.1 Background

Censored data arise frequently in statistical studies and particularly when the response is the time to some event. Examples include survival time after a treatment is applied to a biological unit in a medical study or the failure time of a device in an engineering-evaluation study. Suppose that n units are put on test at the same time and monitored continuously until failure or the end of the test, whichever occurs first. The most common kinds of censoring for such controlled laboratory studies are Type I or “time” censoring (where a test is terminated after a specified amount of time, say y_c , has elapsed) and Type II or “failure” censoring (where a test is terminated after a specified number $r \leq n$ units have failed).

Special tools are needed to plan tests that are expected to result in censored data. Most criteria that have been used for planning such studies have been based on functions of the elements of the large-sample approximate variance-covariance matrix of the maximum likelihood estimators of the model parameters. Chapter 10 of Meeker and Escobar (1998) provides a number of different examples.

The large-sample approximate variance-covariance matrix of the maximum likelihood estimators is computed as a function of the Fisher information matrix. Although the derivations of the Fisher information matrices for Type I and Type II censoring proceed differently (because in the former case the number of failures is random and in the latter case, the length of the test is random), we show for location-scale distributions, under standard regularity conditions, that the information matrices obtained are asymptotically equivalent. The results extend directly to the commonly used log-location-scale distributions (e.g., the Weibull, lognormal, and loglogistic distributions).

Escobar and Meeker (1994) provide an algorithm for computing the elements of these matrices. Escobar and Meeker (1998) provide extensions to problems and models involving truncation and explanatory variables.

1.2 Review of previous related work

Bennett (1952) studied the asymptotic behavior of the Best Linear Unbiased Estimators based on order statistics (which he called “ideal linear estimators”) of the location and scale parameters of a continuous random variable. He showed that under certain regularity conditions, the asymptotic variance

of the ideal linear estimators is equal to the asymptotic variance of the maximum likelihood estimates of these parameters, implying that these linear estimators are asymptotically efficient. Bennett's development allows for multiple Type II censoring which includes single censoring and no censoring as special cases. For a review of Bennett's results see David (1981, Section 9.7). Chernoff, Gastwirth, and Johns (1967) independently developed and extended some of Bennett's results.

For more general distributions, Halperin (1952) and Bhattacharyya (1985) showed, using other sets of regularity conditions, that Maximum Likelihood Estimators of parameters from Type II censored samples are consistent, asymptotically normally distributed, and efficient.

1.3 Overview

Section 2 describes the location-scale family of distributions, gives a general expression for the loglikelihood, and sets the regularity conditions used for the results in this paper. Section 3 gives expressions for the Fisher information matrix for both Type I and Type II censoring. Section 4 demonstrates the asymptotic equivalence of the Fisher information matrices for these two different kinds of censoring. Section 5 concludes the paper with discussion of some extensions and some other comments. The Appendix contains some technical details.

2 Model and Assumptions

Assume that the random variable Y follows a location-scale distribution with cdf $G(y; \boldsymbol{\theta}) = \Phi(z)$, where $z = (y - \mu)/\sigma$, $\boldsymbol{\theta} = (\mu, \sigma)$, $-\infty < \mu < \infty$ is the location parameter, $\sigma > 0$ is the scale parameter, and Φ is a standardized cdf (i.e., $\mu = 0$ and $\sigma = 1$). Then the pdf for Y is $g(y; \boldsymbol{\theta}) = \phi(z)/\sigma$, where ϕ , the derivative of Φ , is the corresponding standardized pdf. To simplify the notation, we use $G(y) = G(y; \boldsymbol{\theta})$ and $g(y) = g(y; \boldsymbol{\theta})$.

For n observations consisting of exact failures (i.e., not censored) and right censored (at time y_c) observations, the log-likelihood can be written as (see details for Singly Type I censored and Type II censored data below)

$$\mathcal{L} = \mathcal{L}(\boldsymbol{\theta}) = \mathcal{C} + \sum_{i=1}^n \delta_i \log[g(y_i)] + \log[1 - G(y_c)] \sum_{i=1}^n (1 - \delta_i)$$

where \mathcal{C} is a constant that does not depend of $\boldsymbol{\theta}$ and

$$\delta_j = \begin{cases} 1 & \text{if } y_j \leq y_c \\ 0 & \text{if } y_j > y_c. \end{cases}$$

The censoring time y_c is fixed for Type I censoring and random for Type II censoring (see details below). The Fisher information matrix is defined by

$$\text{E} \left[\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^T} \right) \right]$$

where E is the expectation operator and T is a vector (matrix) transpose.

Through out the development, we assume that the following regularity conditions are satisfied

1. $\phi(z) > 0$ for all $-\infty < z < \infty$.
2. $\lim_{z \rightarrow \pm\infty} z^2 \phi'(z) = 0$, where ϕ' is the first derivative of ϕ .
3. The second derivative ϕ'' is continuous.
4. The expectations

$$\text{E} \left[-\frac{\partial^2 \log(\phi(z))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]$$

are all finite.

Chernoff, Gastwirth, and Johns (1967, page 67) state that, for complete samples, regularity Condition 2 above and the existence of ϕ'' imply that

$$\text{E} \left[\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^T} \right) \right] = -\text{E} \left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right].$$

Also, using simple algebra and L'Hospital's rule, it is easy to see that the regularity Condition 2 is equivalent to

$$\begin{aligned} \lim_{z \rightarrow -\infty} \frac{z^2 \phi^2(z)}{\Phi(z)} &= 0 \\ \lim_{z \rightarrow +\infty} \frac{z^2 \phi^2(z)}{1 - \Phi(z)} &= 0. \end{aligned} \tag{1}$$

The regularity conditions in equation (1) are used by David (1981, page 278) and Bennett (1952, page 63) to assure the asymptotic efficiency of the linear estimators of μ and σ based on order statistics from Type II censored samples.

3 Fisher Information Matrices

3.1 Singly Type I censored data

For a sample of size n , consisting of exact (i.e., not censored) and right censored observations at y_c , the data are the failure times y_i for units that fail ($y_i < y_c$) and the number of units that exceeded the censoring time y_c . The contribution to the log-likelihood from observation j is

$$\mathcal{L}_j^I = \delta_j \log[g(y_j)] + (1 - \delta_j) \log[1 - G(y_c)]$$

and the total sample log-likelihood for a singly Type I censored sample is

$$\mathcal{L}^I = \mathcal{L}^I(\boldsymbol{\theta}) = \sum_{j=1}^n \mathcal{L}_j^I.$$

Note that $\boldsymbol{\theta} = (\theta_1, \theta_2)$ with $\theta_1 = \mu$ and $\theta_2 = \sigma$. Under the regularity conditions, the Fisher information matrix is

$$\begin{aligned} I_{\boldsymbol{\theta}}^I &= \text{E} \left[\left(\frac{\partial \mathcal{L}^I}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \mathcal{L}^I}{\partial \boldsymbol{\theta}^T} \right) \right] \\ &= \text{E} \begin{bmatrix} \frac{\partial \mathcal{L}^I}{\partial \mu} & \frac{\partial \mathcal{L}^I}{\partial \mu} & \frac{\partial \mathcal{L}^I}{\partial \mu} & \frac{\partial \mathcal{L}^I}{\partial \sigma} \\ \frac{\partial \mathcal{L}^I}{\partial \sigma} & \frac{\partial \mathcal{L}^I}{\partial \mu} & \frac{\partial \mathcal{L}^I}{\partial \sigma} & \frac{\partial \mathcal{L}^I}{\partial \sigma} \end{bmatrix} = -\text{E} \left[\frac{\partial^2 \mathcal{L}^I}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \end{aligned}$$

where E is the expectation operator; justification for the last equality is given in Sections A.1 and A.2 of the Appendix. Thus, the information matrix is

$$I_{\boldsymbol{\theta}}^I = -n \left[\int_{-\infty}^{y_c} \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) dy + [1 - G(y_c)] \frac{\partial^2 \log[1 - G(y_c)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]. \quad (2)$$

Making the change of variables $z = (y - \mu)/\sigma$, the information matrix can be expressed as function of the standardized censoring time $\zeta_c = (y_c - \mu)/\sigma$ as follows:

$$I_{\boldsymbol{\theta}}^I = -n \left[\int_{-\infty}^{\zeta_c} \frac{\phi(z)}{\sigma} \frac{\partial^2 \log[(1/\sigma)\phi(z)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} dz + [1 - \Phi(\zeta_c)] \frac{\partial^2 \log[1 - \Phi(\zeta_c)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right].$$

The Fisher information per observation is $\mathcal{I}^I(\boldsymbol{\theta}) = I_{\boldsymbol{\theta}}^I/n$.

3.2 Type II censored data

In this case we observe the $k = n - r$ smallest order statistics from a random sample of size n where n and r are fixed. Then the joint probability density function of the ordered observations, $\mathbf{y} = (y_{(1)}, \dots, y_{(k)})^T$ is

$$p(\mathbf{y}; \boldsymbol{\theta}) = \begin{cases} \left(\frac{n!}{r!}\right) [1 - G(y_{(k)})]^r \prod_{i=1}^k [g(y_{(i)})] & \text{if } y_{(1)} < \dots < y_{(k)} \\ 0 & \text{otherwise} \end{cases}$$

The log-likelihood for the Type II sample is

$$\mathcal{L}^{II} = \mathcal{L}^{II}(\boldsymbol{\theta}) = \log\left(\frac{n!}{r!}\right) + \sum_{i=1}^k \log[g(y_{(i)})] + r \log[1 - G(y_{(k)})]. \quad (3)$$

The Fisher information matrix, under the regularity conditions, is

$$\begin{aligned} I_{\boldsymbol{\theta}}^{II} &= \text{E} \left[\left(\frac{\partial \mathcal{L}^{II}}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \mathcal{L}^{II}}{\partial \boldsymbol{\theta}^T} \right) \right] = -\text{E} \left[\frac{\partial^2 \mathcal{L}^{II}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \\ &= - \left[\int \frac{\partial^2 \mathcal{L}^{II}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} p(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{y} \right] \end{aligned}$$

where the range of all integrals, unless otherwise specified, is over the k dimensional space $(-\infty, \infty) \times \dots \times (-\infty, \infty)$ for the multiple integrals and over the real line $(-\infty, \infty)$ for the single integrals. Then from (3) and using the marginal distribution of the order statistics to compute the expectation, and the notation $S(y) = 1 - G(y; \boldsymbol{\theta})$, we get

$$\begin{aligned} I_{\boldsymbol{\theta}}^{II} &= - \left[\sum_{i=1}^k \int \frac{n!}{(i-1)!(n-i)!} \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) G^{i-1}(y) S^{n-i}(y) dy \right. \\ &\quad \left. + \int \frac{n!}{(k-1)!(r-1)!} \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) G^{k-1}(y) S^r(y) dy \right] \\ &= -n \left[\int \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) \sum_{i=1}^k \frac{(n-1)!}{(i-1)!(n-i)!} G^{i-1}(y) S^{n-i}(y) dy \right. \\ &\quad \left. + \int \frac{(n-1)!}{(k-1)!(r-1)!} \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) G^{k-1}(y) S^r(y) dy \right] \\ &= -n \left[\int \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) [1 - H_{(k:n-1)}(y)] dy \right. \\ &\quad \left. + \int \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) h_{(k:n-1)}(y) dy \right] \quad (4) \end{aligned}$$

where

$$\begin{aligned} H_{(k:n-1)}(y) &= 1 - \sum_{i=1}^k \frac{(n-1)!}{(i-1)!(n-i)!} G^{i-1}(y) S^{n-i}(y) \\ &= \sum_{j=k}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} G^j(y) S^{n-1-j}(y) \end{aligned}$$

and

$$h_{(k:n-1)}(y) = \frac{(n-1)!}{(k-1)!(r-1)!} g(y) G^{k-1}(y) S^{r-1}(y)$$

are, respectively, the cdf and pdf for the k th order statistic in a sample of size $(n-1)$ from $G(y)$.

The Fisher information per observation is

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta}) &= - \left[\int \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) [1 - H_{(k:n-1)}(y)] dy \right. \\ &\quad \left. + \int \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) h_{(k:n-1)}(y) dy \right]. \end{aligned}$$

4 Asymptotic Equivalence of the Information Matrices

Here we consider the asymptotic behavior in Type II censored samples where the sample size n and the number of observed failures k increase such that $k/n \rightarrow p_c$, $0 < p_c \leq 1$. We will show that the limiting amount of information per unit in this process is equivalent to the per unit information from a Type I censored sample with censoring time $y_c = G^{-1}(p_c)$. Formally,

Result 1 *Let y_c be a fixed censoring time such that $p_c = G(y_c)$ with $0 < p_c \leq 1$. Then*

$$\lim_{\substack{n \rightarrow \infty \\ (k/n) \rightarrow p_c}} \mathcal{I}^{II}(\boldsymbol{\theta}) = \mathcal{I}^I(\boldsymbol{\theta}).$$

Proof: We give the proof when $0 < p_c < 1$; the proof when $p_c = 1$ (complete data) is similar. First, observe that $H_{k:n-1}$ converges in distribution to the degenerate distribution $\Psi(y)$ which puts probability one at y_c , i.e.,

$$\lim_{\substack{n \rightarrow \infty \\ (k/n) \rightarrow p_c}} H_{k:n-1}(y) = \Psi(y) = \begin{cases} 1 & \text{if } y \geq y_c \\ 0 & \text{if } y < y_c \end{cases}.$$

This holds because, under the regularity conditions in Section 2, the “sample quantile” $Y_{(k)}$ [with $(k/n) \rightarrow p_c$] is a consistent estimator of the population quantile $G^{-1}(p_c)$ (see Wilks 1962 page 272 for a proof of this result).

Because $H_{k:n-1}(y)$ is a cdf, the matrix elements satisfy the inequalities,

$$\left[\frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) [1 - H_{(k:n-1)}(y)] \right] \leq \left[\left| \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right| g(y) \right].$$

Also, from the regularity Condition 4 in Section 2 the elements of the Fisher Information matrix are finite, i.e., element-wise we have

$$\left[\int \left| \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right| g(y) dy \right] < \infty.$$

Consequently, by the dominated convergence theorem, element-wise

$$\begin{aligned} & \left[\lim_{\substack{n \rightarrow \infty \\ (k/n) \rightarrow p_c}} \int \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) [1 - H_{(k:n-1)}(y)] dy \right] \\ &= \left[\int \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) \lim_{\substack{n \rightarrow \infty \\ (k/n) \rightarrow p_c}} [1 - H_{(k:n-1)}(y)] dy \right] \\ &= \left[\int \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) [1 - \Psi(y)] dy \right] \\ &= \left[\int_{-\infty}^{y_c} \frac{\partial^2 \log[g(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} g(y) dy \right]. \end{aligned} \quad (5)$$

Now, using the definition of Riemann-Stieltjes integrals it follows that

$$\left[\int \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) h_{(k:n-1)}(y) dy \right] = \left[\int \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) dH_{(k:n-1)} \right].$$

Section A.3 of the Appendix shows that the elements of the matrix

$$\left[\frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) \right]$$

are continuous and bounded functions. Then using the fact that $H_{k:n-1}$ converges in distribution to Ψ and Theorem A of Serfling (1980, page 16), we have the following result for the matrix elements

$$\begin{aligned} \left[\lim_{\substack{n \rightarrow \infty \\ (k/n) \rightarrow p_c}} \int \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) dH_{(k:n-1)} \right] &= \left[\int \frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) d\Psi \right] \\ &= \left[S(y_c) \frac{\partial^2 \log[S(y_c)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]. \end{aligned} \quad (6)$$

Result 1 follows from (2), (4), (5), and (6).

5 Extensions and Comments

The result in Section 4 extends naturally to the following situations:

- Log-location-scale distributions (e.g., Weibull and lognormal random variables T) where the distribution of $Y = \log(T)$ satisfies the regularity conditions. This follows from *Result 1* and the fact that if γ are the parameters of the distribution of T . Then

$$\begin{aligned}\mathcal{I}^I(\gamma) &= A\mathcal{I}^I(\boldsymbol{\theta})A^T \\ \mathcal{I}^{II}(\gamma) &= A\mathcal{I}^{II}(\boldsymbol{\theta})A^T\end{aligned}$$

where A is a matrix of constants.

- The result holds when the distribution is just location or just scale and it satisfies the regularity conditions. For example, the result applies to an exponential distribution with a single scale parameter.
- The result can also be shown to hold for combinations of singly left and singly right censored data from location-scale (log-location-scale) families.

Finally, Bhattacharyya (1985) demonstrates the asymptotic normality and uniform strong convergence of a class of functions that arise in the context of estimating parameters with Type II censored samples from an arbitrary multiparameter family, not necessarily location scale, that satisfy a set of regularity conditions. Similar, but less general results are given in Halperin (1952). The results in this paper can be extended to the general setting considered in those two articles.

Acknowledgments

We thank professors H.A. David and V.N. Nair for very useful discussion, comments, and suggestions that greatly simplified the presentation in the paper.

APPENDIX

A Computing Fisher Matrices Using Second Partial Derivatives

Here we show that under the regularity conditions in Section 2 that

$$\mathbb{E} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^T} \right] = -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]. \quad (7)$$

A.1 Type I Data – Case

The log-likelihood for an single observation y is

$$\mathcal{L} = \delta \log[g(y)] + (1 - \delta) \log[S(y_c)]$$

where δ is the indicator functions of y being in the intervals $(-\infty, y_c]$. Starting with the identity

$$\int_{-\infty}^{y_c} g(y) dy + S(y_c) = 1$$

and taking derivatives on both sides with respect to $\boldsymbol{\theta}$, we get

$$\left[\frac{\partial}{\partial \boldsymbol{\theta}^T} \left\{ \int_{-\infty}^{y_c} g(y) dy + S(y_c) \right\} \right] = \mathbf{0}$$

which implies

$$\left[\int_{-\infty}^{y_c} \frac{\partial g(y)}{\partial \boldsymbol{\theta}^T} dy + \frac{\partial S(y_c)}{\partial \boldsymbol{\theta}^T} \right] = \mathbf{0}. \quad (8)$$

The interchange of the derivative and the integral is justified because the regularity conditions imply that the first partials of g with respect to μ and σ are continuous (Apostol 1957, page 442). Equation (8) is equivalent to $\mathbb{E} [\partial \mathcal{L} / \partial \boldsymbol{\theta}^T] = \mathbf{0}$. Taking partial derivatives of (8) with respect to $\boldsymbol{\theta}$ and using the continuity of g'' gives

$$\left[\int_{-\infty}^{y_c} \frac{\partial^2 g(y)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} dy + \frac{\partial^2 S(y_c)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] = \mathbf{0}$$

which is equivalent to $\mathbb{E} [M(y; \boldsymbol{\theta})] = \mathbf{0}$, where

$$M(y; \boldsymbol{\theta}) = \left(\frac{\delta}{g(y)} \right) \left[\frac{\partial^2 g(y)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] + \left(\frac{1 - \delta}{S(y_c)} \right) \left[\frac{\partial^2 S(y_c)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right].$$

Direct manipulations give

$$\mathbb{E} \left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] = -\mathbb{E} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^T} \right] + \mathbb{E} [M(\mathbf{y}; \boldsymbol{\theta})].$$

Because the last term is $\mathbf{0}$,

$$\mathbb{E} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^T} \right] = -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right], \quad (9)$$

which is the desired result.

A.2 Type II Data – Case

Now we justify (7) for Type II censored data. For simplicity, we write $p(\mathbf{y}) = p(\mathbf{y}; \boldsymbol{\theta})$, where the joint density $p(\mathbf{y}; \boldsymbol{\theta})$ of the data is given in Section 3.2. It is enough to show that

$$\begin{aligned} \left[\frac{\partial}{\partial \boldsymbol{\theta}^T} \int p(\mathbf{y}) d\mathbf{y} \right] &= \left[\int \frac{\partial p(\mathbf{y})}{\partial \boldsymbol{\theta}^T} d\mathbf{y} \right] = \mathbf{0} \\ \left[\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \int p(\mathbf{y}) d\mathbf{y} \right] &= \left[\int \frac{\partial^2 p(\mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} d\mathbf{y} \right] = \mathbf{0}. \end{aligned}$$

We give the proof when taking second partial derivatives with respect to μ . The other cases are similar. Simple computations give

$$\frac{\partial p(\mathbf{y})}{\partial \mu} = p(\mathbf{y}) \frac{(n-k)}{S_k} \frac{\partial S_k}{\partial \mu} + p(\mathbf{y}) \sum_{i=1}^k \frac{1}{g_i} \frac{\partial g_i}{\partial \mu}$$

where, $S_k = S(y_{(k)})$, $g_i = g(y_{(i)})$. Then using the marginal pdfs of the $1, \dots, k$ order statistics, we get

$$\begin{aligned} \int \frac{\partial p(\mathbf{y})}{\partial \mu} d\mathbf{y} &= \mathbb{E} \left(\frac{(n-k)}{S_k} \frac{\partial S_k}{\partial \mu} \right) + \mathbb{E} \left(\sum_{i=1}^k \frac{1}{g_i} \frac{\partial g_i}{\partial \mu} \right) \\ &= n \int g(y) g_{(k:n-1)}(y) dy + n \int g'(y) [G_{(k:n-1)}(y) - 1] dy \\ &= -n g(y) \Big|_{-\infty}^{+\infty} + n g(y) G_{(k:n-1)}(y) \Big|_{-\infty}^{+\infty} = 0 \end{aligned}$$

where $g'(y)$ is the derivative of $g(y)$, $G_{(k:n-1)}(y)$ is the cdf of the k th order statistic in a sample of size $(n-1)$ and $g_{(k:n-1)}(y)$ is the corresponding pdf.

Now we show that

$$\frac{\partial^2}{\partial \mu^2} \int p(\mathbf{y}; \boldsymbol{\theta}) d\mathbf{y} = \int \frac{\partial^2 p(\mathbf{y}; \boldsymbol{\theta})}{\partial \mu^2} d\mathbf{y} = 0$$

Direct computations yield

$$\begin{aligned} \frac{\partial^2 p(\mathbf{y})}{\partial \mu^2} &= (n-k)(n-k-1)p(\mathbf{y}) \frac{1}{S_k^2} \left(\frac{\partial S_k}{\partial \mu} \right)^2 + 2(n-k)p(\mathbf{y}) \frac{1}{S_k} \frac{\partial S_k}{\partial \mu} \sum_{i=1}^k \frac{1}{g_i} \frac{\partial g_i}{\partial \mu} \\ &+ 2p(\mathbf{y}) \sum_{i=2}^{k-1} \sum_{j=i+1}^k \frac{1}{g_i g_j} \frac{\partial g_i}{\partial \mu} \frac{\partial g_j}{\partial \mu} + (n-k)p(\mathbf{y}) \frac{1}{S_k} \frac{\partial^2 S_k}{\partial \mu^2} + p(\mathbf{y}) \sum_{i=1}^k \frac{1}{g_i} \frac{\partial^2 g_i}{\partial \mu^2}. \end{aligned}$$

Then

$$\begin{aligned} \int \frac{\partial^2 p(\mathbf{y})}{\partial \mu^2} d\mathbf{y} &= (n-k)(n-k-1) \mathbb{E} \left[\frac{1}{S_k^2} \left(\frac{\partial S_k}{\partial \mu} \right)^2 \right] + 2(n-k) \mathbb{E} \left[\frac{1}{S_k} \frac{\partial S_k}{\partial \mu} \sum_{i=1}^k \frac{1}{g_i} \frac{\partial g_i}{\partial \mu} \right] \\ &+ 2 \sum_{i=2}^{k-1} \sum_{j=i+1}^k \mathbb{E} \left[\frac{1}{g_i g_j} \frac{\partial g_i}{\partial \mu} \frac{\partial g_j}{\partial \mu} \right] + (n-k) \mathbb{E} \left[\frac{1}{S_k} \frac{\partial^2 S_k}{\partial \mu^2} \right] + \sum_{i=1}^k \mathbb{E} \left[\frac{1}{g_i} \frac{\partial^2 g_i}{\partial \mu^2} \right]. \end{aligned}$$

Using the marginal and joint pdfs of the relevant order statistics, gives

$$\begin{aligned} \int \frac{\partial^2 p(\mathbf{y})}{\partial \mu^2} d\mathbf{y} &= n \int \frac{d}{dy} \{g'(y) - g(y)g_{(k:n-1)}(y) - g'(y)G_{(k:n-1)}(y)\} + \\ &n(n-1) \int \frac{d}{dy} \{g^2(y) - g^2(y)G_{(k-1:n-2)}(y)\} \end{aligned}$$

or

$$\begin{aligned} \int \frac{\partial^2 p(\mathbf{y})}{\partial \mu^2} d\mathbf{y} &= n \{g'(y) - g(y)g_{(k:n-1)}(y) - g'(y)G_{(k:n-1)}(y)\} \Big|_{-\infty}^{+\infty} + \\ &n(n-1) \{g^2(y) - g^2(y)G_{(k-1:n-2)}(y)\} \Big|_{-\infty}^{+\infty} \\ &= 0. \end{aligned}$$

A.3 Continuity and boundedness of second partial derivatives

We now show that the functions

$$\left[\frac{\partial^2 \log[S(y)]}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} S(y) \right]$$

are continuous and bounded. For the sake of illustration, we only consider the case of second partial derivatives with respect to σ . The approach is similar for the other combinations. Direct calculations give

$$\frac{\partial^2 \log[S(y)]}{\partial \sigma^2} S(y) = -\frac{1}{\sigma^2} \left[2z\phi(z) + z^2\phi'(z) + \frac{z^2\phi^2(z)}{1-\Phi(z)} \right]. \quad (10)$$

The continuity follows directly from the assumptions on the function $\phi(\cdot)$. From the regularity conditions, it follows that the right hand side of (10) tends to 0 when $z \rightarrow \pm\infty$. Then the function is bounded because it is continuous.

BIBLIOGRAPHY

- Apostol, T.M. (1957), *Mathematical Analysis*, Reading: Addison-Wesley.
- Bennett, C.A. (1952), "Asymptotic Properties of Ideal Linear Estimators," *Ph.D. Thesis*, University of Michigan.
- Bhattacharyya, G.K. (1985), "The Asymptotics of Maximum Likelihood and Related Estimators Based on Type II Censored Data," *Journal of the American Statistical Association* **80**, 398–404.
- 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," *Annals of Mathematical Statistics* **38**, 52–72.
- David, H.A. (1981), *Order Statistics – Second Edition*, New York: John Wiley and Sons.
- Escobar, L.A., and Meeker, W.Q. (1994), "Fisher Information Matrix for the Extreme Value, Normal, and Logistic Distributions and Censored Data," *Applied Statistics*, **43**, 533-540.
- Escobar, L.A., and Meeker, W.Q. (1998), "Fisher Information Matrices with Censoring, Truncation, and Explanatory Variables," *Statistica Sinica*, **8**, 221-237.
- Halperin, M. (1952), "Maximum Likelihood Estimation in Truncated Samples," *Annals of Mathematical Statistics*, **40**, 633-643.
- Meeker, W.Q., and Escobar, L.A. (1998), *Statistical Methods for Reliability Data*, New York: John Wiley and Sons.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*,
New York: John Wiley and Sons.

Wilks, S.S. (1962), *Mathematical Statistics*, New York: John Wiley and
Sons.