COMPUTATIONAL STATISTICS & DATA ANALYSIS

ELSEVIER Computational Statistics & Data Analysis 39 (2002) 371–386

www.elsevier.com/locate/csda

Estimation of parameters from progressively censored data using EM algorithm

H.K.T. Ng[a], P.S. Chan[b,*], N. Balakrishnan[a]

[a]*Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1*

[b]*Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong*

Abstract

EM algorithm is used to determine the maximum likelihood estimates when the data are progressively Type II censored. The method is shown to be feasible and easy to implement. The asymptotic variances and covariances of the ML estimates are computed by means of the missing information principle. The methodology is illustrated with two popular models in lifetime analysis, the lognormal and Weibull lifetime distributions. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: EM algorithm; Maximum likelihood estimators; Type-II progressive censoring; Lifetime data; Asymptotic variances; Missing information

1. Introduction

A Type II censored sample is one for which only r smallest observations in a random sample of n items are observed ($1 \leqslant r \leqslant n$). Experiments involving Type II censoring are often used in life testing. Such tests are time- and cost-effective since it might take a very long time for all items to fail. A generalization of Type II censoring is progressive Type II censoring. In this case, the first failure in the sample is observed and a random sample of size R_1 is immediately drawn from the remaining $n-1$ unfailed items and removed from the test, leaving $n-1-R_1$ items in test. When the second item has failed, R_2 of the still unfailed items are removed, and so on. The experiment terminates after some prefixed series of repetitions of this procedure.

Although progressive Type II censored sampling is effective in time and money, it is not very popular in lifetime experiment. It may be due to the complicated calculation

* Corresponding author. Tel.: +852-26097920; fax: +852-26035188.

E-mail address: benchan@cuhk.edu.hk (P.S. Chan).

PII: S 0167-9473(01)00091-3

of the likelihood function (Lawless, 1982). Newton–Raphson algorithm is one of the standard methods to determine the maximum likelihood estimates of the parameters. To employ the algorithm, the second derivatives of the log-likelihood are required. Sometimes the calculations of the derivatives based on the progressively Type II censored samples are complicated. To avoid such computation, we propose to use the EM algorithm instead.

EM algorithm (Dempster et al., 1977; McLachlan and Krishnan, 1997) is a very powerful tool in handling the incomplete data problem. It is an iterative method by repeating to fill in the missing data with estimated values and to updating the parameter estimates. It is especially useful if the complete data set is easy to analyze.

Some early works on progressive censoring can be found in Cohen (1963), Mann (1971) and Thomas and Wilson (1972). Recently, Viveros and Balakrishnan (1994) proposed a conditional method of inference to derive the exact confidence intervals. Balasooriya et al. (2000) studied the progressively censored sampling plan for the Weibull distribution. A book dedicated completely to progressive censoring has been prepared recently by Balakrishnan and Aggarwala (2000).

This paper is organized as follows. Section 2 gives the details of the progressively Type II censored samples. Section 3 explains how the EM algorithm is used to determine the maximum likelihood estimates in a general setting. Section 4 describes how to obtain the variances and covariances as well as the standard errors of the maximum likelihood estimates. Two popular models in lifetime analysis, via, normal and extreme value distributions, are used to illustrate how the algorithm works, respectively, in Sections 5 and 6.

2. Progressively censored data

Suppose n independent units are placed on a life-test with the corresponding lifetimes $X_1, X_2, \ldots, X_n$ being identically distributed. We assume that X_i, $i = 1, 2, \ldots, n$ are i.i.d. with p.d.f. $f_X(x; \boldsymbol{\theta})$ and c.d.f. $F_X(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the vector of parameters. Prior to the experiment, a number $m < n$ is determined and the censoring scheme $(R_1, R_2, \ldots, R_m)$ with $R_j > 0$ and $\sum_{j=1}^{m} R_j + m = n$ is specified. During the experiment, jth failure is observed and immediately after the failure, R_j functioning items are randomly removed from the test.

Note that in the analysis of lifetime data, instead of working with the parametric model for X_i, it is often more convenient to work with the equivalent model for the log-lifetimes $W_i = \log X_i$, for $i = 1, 2, \ldots, n$. W_i, $i = 1, 2, \ldots, n$ are i.i.d. with p.d.f. $f_W(w; \boldsymbol{\theta})$ and c.d.f. $F_W(w; \boldsymbol{\theta})$. We denote the m completely observed (ordered) log-lifetimes by $Y_{j:m:n}$, $j = 1, 2, \ldots, m$. The likelihood function based on $Y_{j:m:n}$ is

$$L(\boldsymbol{\theta}) = c \prod_{j=1}^{m} f_W(y_{j:m:n}; \boldsymbol{\theta})[1 - F_W(y_{j:m:n}; \boldsymbol{\theta})]^{R_j}, \tag{2.1}$$

where

$$c = n(n - R_1 - 1) \cdots (n - R_1 - R_2 - \cdots - R_{m-1} - m + 1).$$

The maximum likelihood estimators are those values of $\boldsymbol{\theta}$ which maximize (2.1). In most cases, the estimators do not admit explicit expressions and some numerical procedures such as Newton–Raphson method have to be used to determine the estimators. Balakrishnan and Cohen (1991) and Balakrishnan and Aggarwala (2000) have discussed this problem extensively. To employ the Newton–Raphson method, the second derivatives of the log-likelihood function are required, which may sometimes be complicated. In addition, the Newton–Raphson method is not easy to extend to generalize to other forms of censored data.

3. EM algorithm

The progressive right censoring model problem can be viewed as an incomplete data problem and then the EM algorithm is applicable to obtain the maximum likelihood estimators of the parameters. First of all, denote the observed and censored data by $\mathbf{Y}=(Y_{1:m:n}, Y_{2:m:n}, \ldots, Y_{m:m:n})$ and $\mathbf{Z}=(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_m)$, respectively, where $\mathbf{Z}_j$ is a $1 \times R_j$ vector with $\mathbf{Z}_j = (Z_{j1}, Z_{j2}, \ldots, Z_{jR_j})$, for $j = 1, 2, \ldots, m$. The censored data vector $\mathbf{Z}$ can be thought of as the missing data. Combine $\mathbf{Y}$ and $\mathbf{Z}$ to form $\mathbf{W}$ which is the complete data set. The corresponding log-likelihood function is denoted by $\lambda(\mathbf{W}, \boldsymbol{\theta})$. Then the E-step of the algorithm requires the computation of the conditional expectation

$$E[\lambda(\mathbf{W}, \boldsymbol{\theta}_{(h+1)}) \mid \mathbf{Y} = \mathbf{y}, \boldsymbol{\theta}_{(h)}], \tag{3.1}$$

which mainly involves the computation of the conditional expectation of functions of $\mathbf{Z}$ conditional on the observed values $\mathbf{Y}$ and the current value of the parameters. Therefore, in order to facilitate the EM algorithm, the conditional distribution of $\mathbf{Z}$, conditional on $\mathbf{Y}$ and the current value of the parameters, needs to be determined.

Theorem. *Given $Y_{1:m:n} = y_{1:m:n}, Y_{2:m:n} = y_{2:m:n}, \ldots, Y_{j:m:n} = y_{j:m:n}$, the conditional distribution of Z_{jk}, $k = 1, 2, \ldots, R_j$ is*

$$\begin{aligned} &f_{Z|Y}(z_j \mid Y_{1:m:n} = y_{1:m:n}, Y_{2:m:n} = y_{2:m:n}, Y_{j:m:n} = y_{j:m:n}) \\ &\quad = f_{Z|Y}(z_j \mid Y_{j:m:n} = y_{j:m:n}) = \frac{f_W(z_j)}{[1 - F_W(y_{j:m:n})]}, \quad z_j > y_{j:m:n}, \end{aligned} \tag{3.2}$$

and Z_{jk} and Z_{jl}, $k \neq l$, are conditionally independent given $Y_{j:m:n} = y_{j:m:n}$.

Proof. The joint probability that the complete sample being observed is

$$\prod_{j=1}^{m} \left[f_W(y_{j:m:n}) \prod_{k=1}^{R_j} f_W(z_{jk}) \right].$$

The probability that $Y_{1:m:n} = y_{1:m:n}, Y_{2:m:n} = y_{2:m:n}, \ldots, Y_{j:m:n} = y_{j:m:n}$ being observed is

$$\prod_{j=1}^{m} \{ f_W(y_{j:m:n}) [1 - F_W(y_{j:m:n})]^{R_j} \}.$$

Then the conditional probability Z_{jk}, $k = 1,2,\dots,R_j$ given $Y_{1:m:n} = y_{1:m:n}$, $Y_{2:m:n} = y_{2:m:n},\dots,Y_{j:m:n} = y_{j:m:n}$ is given by

$$f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y}) = \prod_{j=1}^{m}\prod_{k=1}^{R_j}\frac{f_W(z_{jk})}{1-F_W(y_{j:m:n})}.$$

Therefore, by the factorization theorem, Z_{jk} are conditionally independent and follow the truncated distribution from the left at $y_{j:m:n}$, $k = 1,2,\dots,R_j$ and $j = 1,2,\dots,m$. □

The theorem states that given $Y_{j:m:n} = y_{j:m:n}$, $\mathbf{Z}_j$ form a random sample from the truncated population and hence the expectations of functions of $\mathbf{Z}_j$ can be obtained.

In the M-step on the $(h+1)$th iteration of the EM algorithm, the value of $\boldsymbol{\theta}$ which maximizes $E[\lambda(\mathbf{W},\boldsymbol{\theta}_{(h+1)})\,|\,\mathbf{Y}=\mathbf{y},\boldsymbol{\theta}_{(h)}]$ will be used as the next estimate of $\boldsymbol{\theta}_{(h+1)}$. The MLE of $\boldsymbol{\theta}$ can be obtained by repeating the E- and M-step until convergence occurs. A reasonable starting value for $\boldsymbol{\theta}_{(0)}$ is the estimates of the parameters based on the "pseudo-complete" sample by replacing the censored observations $\mathbf{Z}_j$ by $Y_{j:m:n}$, $j = 1,2,\dots,m$.

4. Asymptotic variances and covariances of the ML estimates

Louis (1982) developed a procedure for extracting the observed information matrix when the EM algorithm is used in order to find maximum likelihood estimates in incomplete data problem. The idea of the procedure can be expressed by the Missing Information Principle (Louis, 1982; Tanner, 1993):

Observed information = Complete information − Missing information.

We can use this procedure to compute the variances and covariances of the ML estimates under progressive Type II right censoring. The observed information, complete information and missing information are denoted by $I_Y(\boldsymbol{\theta})$, $I_Z(\boldsymbol{\theta})$ and $I_{Z|Y}(\boldsymbol{\theta})$, respectively.

The complete information $I_W(\boldsymbol{\theta})$ is given by

$$I_W(\boldsymbol{\theta}) = -E\left[\frac{\partial^2\lambda(\mathbf{W},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right]. \tag{4.1}$$

Based on the conditional distribution in (3.2), the Fisher information matrix in one observation which is censored at the time of the jth failure can be computed as

$$I_{Z|Y}^{(j)}(\boldsymbol{\theta}) = E\left[\left(\frac{\partial\log f_{Z_j}(z_j|y_{j:m:n},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)^2\right] = -E\left[\frac{\partial^2\log f_{Z_j}(z_j|y_{j:m:n},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^2}\right]. \tag{4.2}$$

Then the expected information for conditional distribution of $\mathbf{Z}$ given $\mathbf{Y}$ (missing information) is

$$I_{Z|Y}(\boldsymbol{\theta}) = \sum_{j=1}^{m} R_j I_{Z|Y}^{(j)}(\boldsymbol{\theta}). \tag{4.3}$$

Therefore, the observed information is

$$I_Y(\boldsymbol{\theta}) = I_Z(\boldsymbol{\theta}) - I_{Z|Y}(\boldsymbol{\theta}). \tag{4.4}$$

The asymptotic covariance matrix of the ML estimate for $\boldsymbol{\theta}$ can be obtained by inverting the observed information matrix $I_Y(\hat{\boldsymbol{\theta}})$.

5. Lognormal lifetime data

Lognormal distribution is a commonly used lifetime distribution model in lifetime data analysis since the logarithm of the lifetime variables are normally distributed. Because of the well-known properties of the normal distribution and because it is a location-scale model, the log-lifetimes will be used in the analysis.

The log-likelihood function based on the log-lifetimes $\mathbf{W}$ is

$$\begin{aligned}\lambda(\mathbf{W};\mu,\sigma) &= \text{constant} - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(w_i-\mu)^2 \\ &= \text{constant} - n\log\sigma - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{j:m:n}-\mu)^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{m}\sum_{k=1}^{R_j}(Z_{jk}-\mu)^2.\end{aligned} \tag{5.1}$$

5.1. The algorithm

In the E-step, one requires to compute

$$\begin{aligned}&-n\log\sigma - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{j:m:n}-\mu)^2 \\ &\quad - \frac{1}{2\sigma^2}\sum_{j=1}^{m}\sum_{k=1}^{R_j}[E(Z_{jk}^2|Z_{jk}>y_{j:m:n}) - 2\mu E(Z_{jk}|Z_{jk}>y_{j:m:n}) + \mu^2].\end{aligned}$$

As a result of the Theorem in Section 3, the conditional distribution of Z_{jk} given $Y_{j:m:n} = y_{j:m:n}$ follows a truncated normal distribution with left truncation at $y_{j:m:n}$.

The first and second moments of Z_{jk} given $Z_{jk} > y_{j:m:n}$ can be found in Cohen (1991) as

$$E(Z_{jk}\,|\,Z_{jk} > y_{j:m:n},\mu,\sigma) = \sigma Q_j + \mu, \tag{5.2}$$

$$E(Z_{jk}^2\,|\,Z_{jk} > y_{j:m:n},\mu,\sigma) = \sigma^2(1+\xi_j Q_j) + 2\sigma\mu Q_j + \mu^2, \tag{5.3}$$

where

$$\xi_j = \frac{y_{j:m:n} - \mu}{\sigma},$$

$$Q_j = \frac{\phi(\xi_j)}{1 - \Phi(\xi_j)} \quad \text{is the hazard function of standard normal distribution.}$$

From the usual results for complete data maximum likelihood estimation for normal distribution, the explicit formulas for the MLE of μ and σ are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} w_i,$$

$$\hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(w_i - \hat{\mu})^2\right]^{1/2}.$$

Therefore, in the $(h+1)$th iteration of the EM algorithm, the value of $\mu_{(h+1)}$ and $\sigma_{(h+1)}$ are computed by the following formulas:

$$\hat{\mu}_{(h+1)} = \frac{1}{n}\left[\sum_{j=1}^{m} y_{j:m:n} + \sum_{j=1}^{m} R_j E(Z_j \mid Z_j > y_{j:m:n}, \mu_{(h)}, \sigma_{(h)})\right], \tag{5.4}$$

$$\hat{\sigma}_{(h+1)} = \left\{\frac{1}{n}\left[\sum_{j=1}^{m} y_{j:m:n}^2 + \sum_{j=1}^{m} R_j E(Z_j^2 \mid Z_j > y_{j:m:n}, \mu_{(h+1)}, \sigma_{(h)})\right] - \hat{\mu}_{(h+1)}^2\right\}^{1/2} \tag{5.5}$$

5.2. Asymptotic variances and covariance

If the log-lifetime data follows normal distribution, it is well known that the information matrix based on the complete data is

$$I_W(\boldsymbol{\theta}) = \frac{n}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

with $\boldsymbol{\theta} = (\mu\ \sigma)'$.

The logarithm of the truncated normal p.d.f. is

$$\log f_{Z_j}(z_j \mid Z_j > y_{j:m:n}, \mu, \sigma)$$
$$= \text{constant} - \log\sigma - \log[1 - F_W(y_{j:m:n})] - \frac{1}{2\sigma^2}(z_j - \mu)^2. \tag{5.6}$$

Differentiation of (5.6) with respect to μ and σ yields

$$\frac{\partial \log f_{Z_j}}{\partial \mu} = \frac{1}{\sigma}\left[\frac{z_j - \mu}{\sigma} - Q_j\right], \tag{5.7}$$

$$\frac{\partial \log f_{Z_j}}{\partial \sigma} = \frac{1}{\sigma}\left[\left(\frac{z_j - \mu}{\sigma}\right)^2 - \left(1 + \xi_j Q_j\right)\right]. \tag{5.8}$$

It is easily shown that

$$E[(Z_j - \mu) \mid Z_j > y_{j:m:n}, \mu, \sigma] = \sigma Q_j, \tag{5.9}$$

$$E[(Z_j - \mu)^2 \mid Z_j > y_{j:m:n}, \mu, \sigma] = \sigma^2(1 + \xi_j Q_j), \tag{5.10}$$

$$E[(Z_j - \mu)^3 \mid Z_j > y_{j:m:n}, \mu, \sigma] = \sigma^3(2 + \xi_j^2), \tag{5.11}$$

$$E[(Z_j - \mu)^4 \mid Z_j > y_{j:m:n}, \mu, \sigma] = \sigma^4[3(1 + \xi_j Q_j) + \xi_j^3 Q_j]. \tag{5.12}$$

Using (5.9)–(5.12), the Fisher information matrix based on one observation which is censored at the time of the jth failure can be computed by straightforward substitution. Then corresponding entries in (4.2) are

$$E\left[\left(\frac{\partial \log f_{Z_j}}{\partial \mu}\right)^2\right] = \frac{1}{\sigma^2}[1 + \xi_j Q_j - Q_j^2],$$
$$E\left[\left(\frac{\partial \log f_{Z_j}}{\partial \sigma}\right)^2\right] = \frac{1}{\sigma^2}[2 + \xi_j Q_j(1 - \xi_j Q_j + \xi_j^2)],$$
$$E\left[\left(\frac{\partial \log f_{Z_j}}{\partial \mu}\right)\left(\frac{\partial \log f_{Z_j}}{\partial \sigma}\right)\right] = \frac{1}{\sigma^2}[Q_j + \xi_j Q_j(\xi_j - Q_j)].$$

Thus, the expected information for conditional distribution of $\mathbf{Z}$ given $\mathbf{Y}$ can be obtained using (4.3) and hence $I_Y(\boldsymbol{\theta})$. Inverting $I_Y(\boldsymbol{\theta})$ yields the variance–covariance matrix of $\hat{\boldsymbol{\theta}} = (\hat{\mu}\ \hat{\sigma})'$.

Instead, had we employed Newton–Raphson method for finding the MLEs numerically, we would have solved the equations

$$\frac{\partial \log L(\mu, \sigma)}{\partial \mu} = \frac{1}{\sigma}\left\{\sum_{j=1}^{m} \xi_j + \sum_{j=1}^{m} R_j Q_j\right\} = 0,$$
$$\frac{\partial \log L(\mu, \sigma)}{\partial \sigma} = -\frac{m}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{m} \xi_j^2 + \frac{1}{\sigma}\sum_{j=1}^{m} R_j \xi_j Q_j = 0$$

by using the second-order derivative forms

$$\frac{\partial^2 \log L(\mu, \sigma)}{\partial \mu^2} = -\frac{m}{\sigma^2} + \frac{1}{\sigma^2}\sum_{j=1}^{m} R_j \left\{\frac{\xi_j \phi(\xi_j) - \xi_j \phi(\xi_j)\Phi(\xi_j) - \phi^2(\xi_j)}{[1 - \Phi(\xi_j)]^2}\right\},$$
$$\frac{\partial^2 \log L(\mu, \sigma)}{\partial \sigma^2} = \frac{m}{\sigma^2} - \frac{3}{\sigma^2}\sum_{j=1}^{m} \xi_j^2 - \frac{2}{\sigma^2}\sum_{j=1}^{m} R_j Q_j,$$
$$\frac{\partial^2 \log L(\mu, \sigma)}{\partial \mu\, \partial \sigma} = \frac{1}{\sigma^2}\sum_{j=1}^{m} \{2\xi_j + R_j Q_j + R_j \xi_j Q_j - \xi_j^2 Q_j\}.$$

Using the Newton–Raphson method, Balakrishnan et al. (2002a) have recently discussed construction of confidence intervals for μ and σ. It is important to mention here that Mi and Balakrishnan (2002) have shown that the MLEs of μ and σ do exist

Table 1
Simulated progressively censored sample from standard normal distribution

j	1	2	3	4	5	6	7
$y_{j:m:n}$	−3.1538	−0.84064	−0.79798	−0.65705	−0.58301	−0.12642	−0.1145
R_j	1	2	0	1	2	0	3

and are unique in this case (by making use of the fact that the normal/lognormal density is log-concave). Due to this result, it is then evident that the EM algorithm and the Newton–Raphson method would converge to the same values.

5.3. Illustrative example

To illustrate the method presented in this section, a progressively Type II right censored sample of size $m=7$ are generated from standard normal distribution. The data are presented in Table 1 below.

A computer program written in FORTRAN is used to execute the EM algorithm. Since the mean and standard deviation of the 7 observed sample points equal to -0.89620 and 0.96127, respectively, thus we can put $\mu_{(0)}=-0.89620$ and $\sigma_{(0)}=0.96127$ as the starting values of the EM algorithm. After a few EM iterations, the estimates converge to $\mu_{(\infty)}=-0.10071$ and $\sigma_{(\infty)}=1.14316$. To compare the convergence rate of EM algorithm with that of Newton–Raphson method, we used the same initial values and fixed the level of accuracy at 5×10^{-5}. The EM algorithm took 27 iterations while the Newton–Raphson method took 4 iterations to converge to the same values.

From these data, we have

$$I_W(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 12.24340 & 0 \\ 0 & 24.48681 \end{pmatrix},$$

$$I_{Z|Y}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 3.47011 & 4.81360 \\ 4.81360 & 12.04041 \end{pmatrix},$$

$$I_Y(\hat{\boldsymbol{\theta}}) = I_W(\hat{\boldsymbol{\theta}}) - I_{Z|Y}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 8.77330 & -4.81360 \\ -4.81360 & 12.44640 \end{pmatrix}.$$

The covariance matrix of $(\hat{\mu}\ \hat{\sigma})$ is

$$I_Y^{-1}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 0.14468 & 0.05596 \\ 0.05596 & 0.10199 \end{pmatrix}.$$

6. Weibull lifetime data

Weibull distribution is another commonly used lifetime distribution model. Similar to the lognormal case, it is more convenient to work with the log-lifetime which is distributed as extreme value.

The log-likelihood function based on the complete log-lifetime $\mathbf{W}$ is

$$\lambda(\mathbf{W};\mu,\sigma) = -n\log\sigma + \sum_{i=1}^{n}\left(\frac{w_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left(\frac{w_i-\mu}{\sigma}\right)$$

$$= -n\log\sigma + \sum_{j=1}^{m}\left(\frac{y_{j:m:n}-\mu}{\sigma}\right) - \sum_{j=1}^{n}\exp\left(\frac{y_{j:m:n}-\mu}{\sigma}\right)$$

$$+\sum_{j=1}^{m}\sum_{k=1}^{R_j}\left(\frac{z_{jk}-\mu}{\sigma}\right) - \sum_{j=1}^{m}\sum_{k=1}^{R_j}\exp\left(\frac{z_{jk}-\mu}{\sigma}\right). \tag{6.1}$$

6.1. The algorithm

Using the Theorem in Section 3, the conditional distribution of Z_j given $Y_{j:m:n}=y_{j:m:n}$ is a truncated extreme value distribution with left truncation at $y_{j:m:n}$ have p.d.f.

$$f_Z(z_j \mid Z_j > y_{j:m:n},\mu,\sigma) = \frac{\exp[\exp(\xi_j)]}{\sigma}\exp\left[\left(\frac{z_j-\mu}{\sigma}\right) - \exp\left(\frac{z_j-\mu}{\sigma}\right)\right],$$

$$y_{j:m:n} < z_j < \infty, \tag{6.2}$$

where $\xi_j = (y_{j:m:n}-\mu)/\sigma$.

To evaluate the required conditional expectations in the E-step, the moment generating function for $(Z_j-\mu)/\sigma$ given $y_{j:m:n}$ is considered. It is given by

$$M_{\frac{Z_j-\mu}{\sigma}}(t) = E\left[e^{t((Z_j-\mu)/\sigma)}\right] = e^{e^{\xi_j}}\Gamma(t+1, e^{\xi_j})$$

$$= \Gamma(t+1)\left[e^{e^{\xi_j}} - \sum_{p=0}^{\infty}\frac{e^{(t+p+1)\xi_j}}{\Gamma(t+p+2)}\right],$$

where $\Gamma(a,x) = \int_x^\infty u^{a-1}e^{-u}\,du$ is the incomplete Gamma function and $\Gamma(a)=\Gamma(a,0)$ is the complete Gamma function.

Then the conditional expectations of interest are

$$E[Z_j \mid \xi_j,\mu,\sigma] = E_1\sigma + \mu, \tag{6.3}$$

$$E[e^{Z_j/\sigma} | \xi_j,\mu,\sigma] = e^{\mu/\sigma}[e^{\xi_j}+1], \tag{6.4}$$

$$E[Z_j e^{(Z_j/\sigma)} | \xi_j,\mu,\sigma] = e^{\mu/\sigma}[E_{2,j}\sigma + \mu(e^{\xi_j}+1)], \tag{6.5}$$

$$E[Z_j^2 e^{(Z_j/\sigma)} | \xi_j,\mu,\sigma] = e^{\mu/\sigma}[E_{3,j}\sigma^2 + 2\mu\sigma E_2 - \mu^2(e^{\xi_j}+1)], \tag{6.6}$$

where

$$E_{1,j} = E\left[\left.\frac{Z_j-\mu}{\sigma}\right| \xi_j,\mu,\sigma\right]$$

$$= \psi(1)e^{e^{\xi_j}} + \sum_{p=0}^{\infty}\frac{e^{(p+1)\xi_j}\psi(p+2)}{\Gamma(p+2)} - [\xi_j+\psi(1)]\sum_{p=0}^{\infty}\frac{e^{(p+1)\xi_j}}{\Gamma(p+2)},$$

$$E_{2,j} = E\left[\left(\frac{Z_j-\mu}{\sigma}\right)e^{((Z_j-\mu)/\sigma)}\middle|\,\xi_j,\mu,\sigma\right]$$

$$= \psi(2)e^{e^{\xi_j}} + \sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_j}\psi(p+3)}{\Gamma(p+3)} - [\xi_j+\psi(2)]\sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_j}}{\Gamma(p+3)},$$

$$E_{3,j} = E\left[\left(\frac{Z_j-\mu}{\sigma}\right)^2 e^{((Z_j-\mu)/\sigma)}\middle|\,\xi_j,\mu,\sigma\right]$$

$$= [\psi'(2)+\psi^2(2)]e^{e^{\xi_j}} - [\xi_j^2+2\xi_j\psi(2)+\psi'(2)+\psi^2(2)]\sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_j}}{\Gamma(p+3)}$$

$$+2[\xi_j+\psi(2)]\sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_j}\psi(p+3)}{\Gamma(p+3)} + \sum_{p=0}^{\infty}\frac{e^{(p+2)\xi_j}[\psi'(p+3)-\psi^2(p+3)]}{\Gamma(p+3)},$$

$\psi(\cdot)$ is the digamma function and $\psi'(\cdot)$ is the trigamma function. The value of gamma, digamma, and trigamma functions can be computed from the corresponding recursive formulae given in Abramowitz and Stegun (1964) in a straightforward manner.

Unlike the normal case, the maximum likelihood estimators of the parameters based on the complete data from the extreme value distribution cannot be solved explicitly. However, this problem has been well studied (see, for example, Lawless, 1982). The estimators can be obtained by solving the equation

$$\hat{\sigma} = \frac{\sum_{i=1}^{n} w_i \exp(w_i/\hat{\sigma})}{\sum_{i=1}^{n}\exp(w_i/\hat{\sigma})} - \frac{1}{n}\sum_{i=1}^{n} w_i$$

and then obtain $\hat{\mu}$ by the following equation:

$$\hat{\mu} = \hat{\sigma}\log\left[\frac{1}{n}\sum_{i=1}^{n}\exp\left(\frac{w_i}{\hat{\sigma}}\right)\right].$$

Thus, in the M-step of the $(h+1)$th iteration of the EM algorithm, the value of $\sigma_{(h+1)}$ is first obtained by solving the equation

$$\sigma_{(h+1)} = \frac{\sum_{j=1}^{m} y_{j:m:n}e^{(y_{j:m:n}/\sigma_{(h+1)})} + \sum_{j=1}^{m} R_j E[Z_j e^{(Z_j/\sigma)}|\xi_j,\mu_{(h)},\sigma_{(h)}]}{\sum_{j=1}^{m} e^{(y_{j:m:n}/\sigma_{(h+1)})} + \sum_{j=1}^{m} R_j E[e^{(Z_j/\sigma)}|\xi_j,\mu_{(h)},\sigma_{(h)}]}$$

$$-\frac{1}{n}\left[\sum_{j=1}^{m} y_{j:m:n} + \sum_{j=1}^{m} R_j E\left[Z_j\,|\,\xi_j,\mu_{(h)},\sigma_{(h)}\right]\right] \tag{6.7}$$

and then obtain $\mu_{(h+1)}$ by

$$\mu_{(h+1)} = \sigma_{(h+1)} \log \left\{ \frac{1}{n} \left[\sum_{j=1}^{m} \mathrm{e}^{y_{j:m:n}/\hat{\sigma}_{(h+1)}} + \sum_{j=1}^{m} R_j E(\mathrm{e}^{(Z_j/\sigma)} | \xi_j, \mu_{(h)}, \sigma_{(h)}) \right] \right\}.$$

6.2. Asymptotic variances and covariance

From the classical results on the extreme value distribution, the complete data information matrix is (see, for example, Stephens, 1977)

$$I_W \begin{pmatrix} \mu \\ \sigma \end{pmatrix} = \frac{n}{\sigma^2} \begin{pmatrix} 1 & 1-\gamma \\ 1-\gamma & c^2 \end{pmatrix}$$

where $\gamma = 0.577215665$ is the Euler's constant and c^2 is $\pi^2/6 + (1-\gamma)^2 = 1.823680661$.

The logarithm of the p.d.f. given in (6.2) is

$$\log f_{Z_j}(z_j \,|\, Z_j > y_{j:m:n}, \mu, \sigma) = -\log \sigma + \mathrm{e}^{\xi_j} + \left(\frac{z_j - \mu}{\sigma} \right) - \mathrm{e}^{((z_j-\mu)/\sigma)}. \qquad (6.8)$$

The three second partial derivatives on (5.9) with respect to μ and σ are

$$\frac{\partial^2 \log f_{Z_j}}{\partial \mu^2} = \frac{1}{\sigma^2} [\mathrm{e}^{\xi_j} - \mathrm{e}^{((z_j-\mu)/\sigma)}],$$

$$\frac{\partial^2 \log f_{Z_j}}{\partial \sigma^2} = \frac{1}{\sigma^2} \left[\mathrm{e}^{\xi_j} + \xi_j \mathrm{e}^{\xi_j} - \mathrm{e}^{((z_j-\mu)/\sigma)} - \left(\frac{z_j - \mu}{\sigma} \right) \mathrm{e}^{((z_j-\mu)/\sigma)} \right],$$

$$\frac{\partial^2 \log f_{Z_j}}{\partial \mu \, \partial \sigma} = \frac{1}{\sigma^2} \left[\xi_j^2 \mathrm{e}^{\xi_j} + 2\xi_j \mathrm{e}^{\xi_j} - \left(\frac{z_j - \mu}{\sigma} \right)^2 \mathrm{e}^{((z_j-\mu)/\sigma)} \right.$$
$$\left. -2 \left(\frac{z_j - \mu}{\sigma} \right) \mathrm{e}^{((z_j-\mu)/\sigma)} + 2 \left(\frac{z_j - \mu}{\sigma} \right) + 1 \right].$$

We take the negative of the expected value of these three second partial derivatives by using the results given in (6.3)–(6.6), using which the Fisher information matrix in one observation which is censored at the time of the jth failure can be computed without any difficulty. Then the observed information can be obtained by formulas (4.3) and (4.4) from which the variance–covariance matrix of the maximum likelihood estimates $\hat{\boldsymbol{\theta}} = (\hat{\mu}\ \hat{\sigma})$ can be determined.

In order to assess the accuracy of the approximation of the variances and covariance of the MLEs determined from the information matrix computed through the EM algorithm described above, we have carried out a simulation study (based on 1000 simulations) for different choices of n, m and $(R_1, \ldots, R_m)$. Without loss of generality, we chose $\mu = 0$ and $\sigma = 1$. The simulated values of $Var(\hat{\mu})$, $Var(\hat{\sigma})$ and $Cov(\hat{\mu}, \hat{\sigma})$

Table 2
Variances and covariance of the MLEs

		Censoring	Simulated			From information		
n	m	Scheme	$Var(\hat{\mu})$	$Var(\hat{\sigma})$	$Cov(\hat{\mu},\hat{\sigma})$	$Var(\hat{\mu})$	$Var(\hat{\sigma})$	$Cov(\hat{\mu},\hat{\sigma})$
15	5	$(0,0,0,0,10)$	0.3676	0.1522	0.1549	0.3167	0.1507	0.1510
15	5	$(10,0,0,0,0)$	0.2194	0.0855	−0.0004	0.1874	0.0816	−0.0111
15	5	$(0,0,10,0,0)$	0.2558	0.0874	0.0550	0.1906	0.0750	0.0378
15	6	$(0,0,0,0,0,9)$	0.2544	0.1182	0.0984	0.2147	0.1220	0.0974
15	6	$(9,0,0,0,0,0)$	0.1751	0.0707	−0.0046	0.1543	0.0713	−0.0140
15	6	$(0,9,0,0,0,0)$	0.1785	0.0660	0.0119	0.1482	0.0620	0.0020
20	6	$(0,0,0,0,0,14)$	0.3320	0.1211	0.1390	0.2936	0.1256	0.1404
20	6	$(14,0,0,0,0,0)$	0.1754	0.0652	0.0001	0.1559	0.0661	−0.0092
20	6	$(0,14,0,0,0,0)$	0.1835	0.0593	0.0190	0.1518	0.0554	0.0084
20	10	$(0,\ldots,0,10)$	0.1302	0.0817	0.0477	0.1183	0.0811	0.0440
20	10	$(10,0,\ldots,0)$	0.1051	0.0508	−0.0107	0.1020	0.0502	−0.0146
25	5	$(0,0,0,0,20)$	0.5928	0.1567	0.2450	0.5421	0.1563	0.2428
25	5	$(20,0,0,0,0)$	0.2232	0.0739	0.0084	0.1922	0.0707	−0.0020
25	5	$(0,20,0,0,0)$	0.2535	0.0657	0.0418	0.1924	0.0561	0.0249
25	15	$(0,\ldots,0,10)$	0.0757	0.0530	0.0194	0.0679	0.0516	0.0167
25	15	$(10,0,\ldots,0)$	0.0713	0.0359	−0.0095	0.0679	0.0349	−0.0119
30	3	$(0,0,27)$	1.4423	0.1994	0.4612	1.2690	0.2047	0.4656
30	3	$(27,0,0)$	0.4693	0.0929	0.0730	0.3084	0.0772	0.0285
30	4	$(0,0,0,26)$	0.9209	0.1819	0.3450	0.9019	0.1823	0.3606
30	4	$(26,0,0,0)$	0.2876	0.0789	0.0177	0.2393	0.0720	0.0110
50	20	$(0,\ldots,0,30)$	0.0769	0.0444	0.0354	0.0763	0.0432	0.0346
50	20	$(30,0,\ldots,0)$	0.0511	0.0254	−0.0079	0.0521	0.0250	−0.0079
50	25	$(0,\ldots,0,25)$	0.0482	0.0330	0.0187	0.0485	0.0332	0.0181
50	25	$(25,0,\ldots,0)$	0.0397	0.0210	−0.0064	0.0417	0.0211	−0.0074

as well as the approximate values determined by averaging the corresponding values obtained from the information matrix are presented in Table 2 below.

From this table, we observe that the approximate values determined from information matrix are quite close to the simulated values even for moderate values of m. Furthermore, we note that the approximation becomes quite accurate as m increases.

Instead, had we employed Newton–Raphson method for finding the MLEs numerically, we would have solved the equations

$$\frac{\partial \log L(\mu,\sigma)}{\partial \mu} = -\frac{m}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{m}(R_j+1)\mathrm{e}^{\xi_j} = 0,$$

$$\frac{\partial \log L(\mu,\sigma)}{\partial \sigma} = -\frac{m}{\sigma} - \frac{1}{\sigma}\sum_{j=1}^{m}\xi_j + \frac{1}{\sigma}\sum_{j=1}^{m}(R_j+1)\xi_j\mathrm{e}^{\xi_j} = 0$$

Table 3
Progressively censored sample presented by Viveros and Balakrishnan (1994)

j	1	2	3	4	5	6	7	8
$y_{j:m:n}$	−1.6608	−0.2485	−0.0409	0.2700	1.0224	1.5789	1.8718	1.9947
R_j	0	0	3	0	3	0	0	5

by using the second-order derivative forms

$$\frac{\partial^2 \log L(\mu,\sigma)}{\partial \mu^2} = -\frac{1}{\sigma^2}\sum_{j=1}^{m}(R_j+1)e^{\xi_j},$$

$$\frac{\partial^2 \log L(\mu,\sigma)}{\partial \sigma^2} = \frac{1}{\sigma^2}\left\{m + 2\sum_{j=1}^{m}\xi_j - 2\sum_{j=1}^{m}(R_j+1)\xi_j e^{\xi_j} - 2\sum_{j=1}^{m}(R_j+1)\xi_j^2 e^{\xi_j}\right\}.$$

It should be pointed out that the above second-order derivatives get used at every iteration in the Newton–Raphson method, but they only get used at the final stage of the EM algorithm while computing the information measure. Clearly, this is one advantage of the EM algorithm.

Though the Newton–Raphson method would yield exactly the same values for the MLEs and their asymptotic variances and covariance as the EM algorithm would result in, the convergence of the iterative process will be different than that of the EM algorithm. Using the Newton–Raphson iterative procedure, Balakrishnan et al. (2002b) have recently discussed construction of confidence intervals for μ and σ.

6.3. Illustrative example

A progressively censored sample generated from the log-times to breakdown data on insulating fluid tested at 34 KV by Viveros and Balakrishnan (1994) is used to demonstrate the above estimation procedure. The data are presented in Table 3.

For these data, we use EM algorithm with starting values $\mu_{(0)} = 1.4127$ and $\sigma_{(0)} = 0.7912$, which are the estimates of the parameters based on the observed data points. The EM algorithm converged to the values $\mu_{(\infty)}$=2.221960... and $\sigma_{(\infty)}$=1.0263807... . The values of $\mu_{(h)}$ and $\sigma_{(h)}$ are plotted against h and are presented in Figs. 1 and 2. Note that the dotted line indicate the values of $\mu_{(\infty)}$ and $\sigma_{(\infty)}$. The results agreed with the MLE of μ and σ ($\hat{\mu} = 2.222$ and $\hat{\sigma} = 1.026$) computed by Viveros and Balakrishnan (1994) via numerical maximization, and also with the values obtained from Newton–Raphson method. In order to compare the convergence rate of the EM algorithm with that of the Newton–Raphson method, same initial values were used and the level of accuracy was fixed at 5×10^{-5}. The EM algorithm took 151 iterations while the Newton–Raphson method took 37 iterations to converge to the same values.

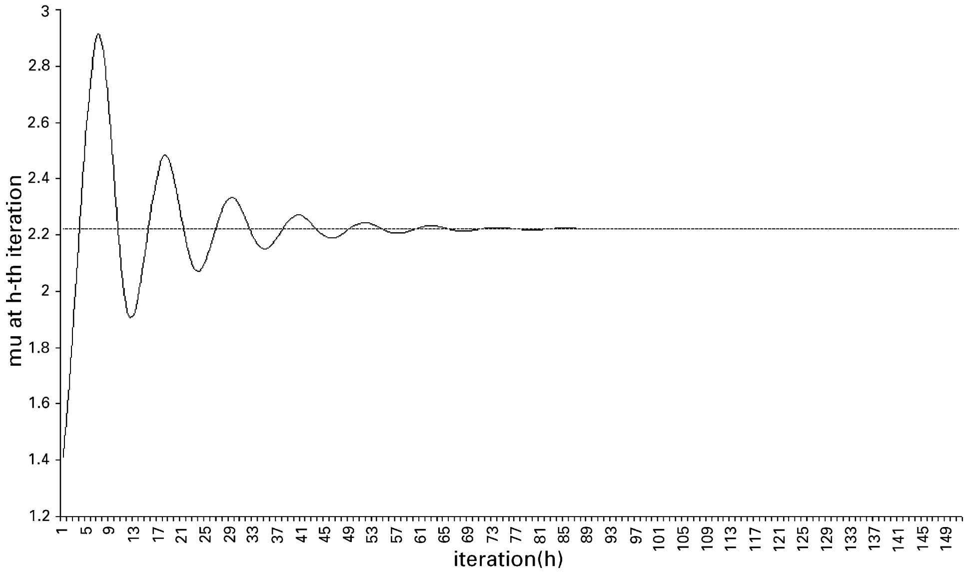

Fig. 1. Trace plot for $\mu_{(h)}$ under EM-iterations (Weibull lifetime data).

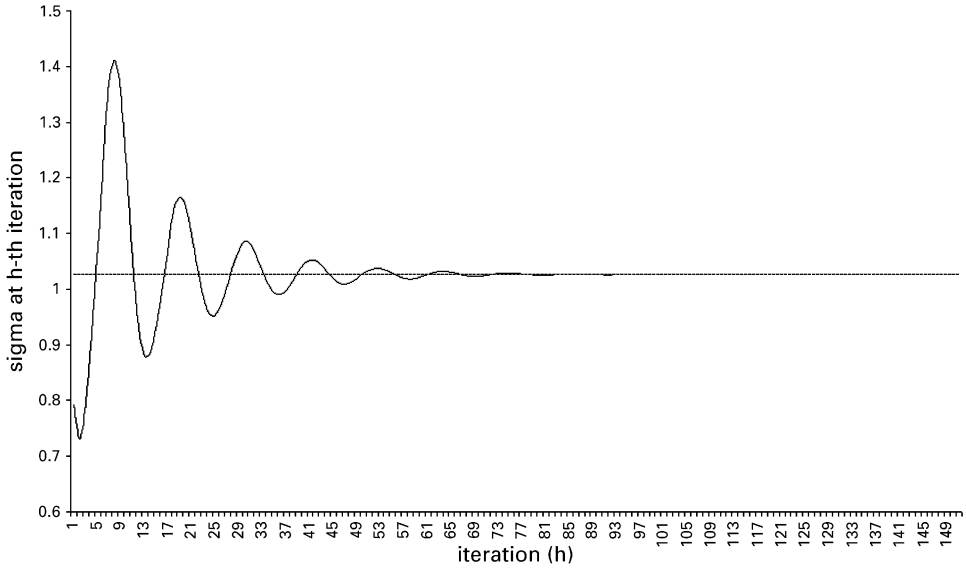

Fig. 2. Trace plot for $\sigma_{(h)}$ under EM-iterations (Weibull lifetime data).

From these data, we have

$$I_W(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 18.03585 & 7.62528 \\ 7.62527 & 32.89163 \end{pmatrix},$$

$$I_{Z|Y}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 10.44181 & 11.99591 \\ 11.99591 & 19.43904 \end{pmatrix},$$

$$I_Y(\hat{\boldsymbol{\theta}}) = I_W(\hat{\boldsymbol{\theta}}) - I_{Z|Y}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 7.59404 & -4.37064 \\ -4.37064 & 13.45259 \end{pmatrix}.$$

By inverting the $I_Y(\hat{\boldsymbol{\theta}})$, we have

$$Var(\hat{\mu}) = 0.16197, \quad Var(\hat{\sigma}) = 0.09143, \quad Cov(\hat{\mu}, \hat{\sigma}) = 0.05262.$$

7. Conclusion

Although the maximum likelihood estimation method based on the progressively censored data has been studied extensively, traditionally Newton–Raphson method was used to obtain the estimators. In this paper, the EM algorithm is proposed to solve the problem. Two popular lifetime models, normal and extreme-value distributions, have been used to demonstrate how the algorithm works. For the normal case, the subsequent guesses of the parameters are in explicit expression. This is a perfect application of the EM algorithm. Although this nice feature does not appear in the extreme-value case, the problem of obtaining the maximum likelihood estimates based on a complete sample has been studied extensively. The only problem in applying the EM algorithm is in evaluating the moments of the truncated distribution. If the moments of the truncated distribution and the estimation based on complete sample case can be handled, the EM algorithm can be applied.

As pointed out by Little and Rubin (1983), the EM algorithm will converge reliably but rather slowly (as compared to the Newton–Raphson method) when the amount of information in the missing data is relatively large. This has also been observed in the two examples considered in this paper. However, the EM algorithm possesses the advantages that (i) it gives a measure of information in the censored (missing) data in a natural way through the Missing Information Principle (which is not available in the Newton–Raphson method), and (ii) it can be generalized easily to other forms of censored data (such as Type I censoring).

Acknowledgements

We thank the co-editor, the associate editor and the referee for their insightful comments that led to substantive improvements of our article. The second author's research was supported by Direct Grant of the Faculty of Science of the Chinese University of Hong Kong 2060176.

References

Abramowitz, M., Stegun, I.A. (Ed.), 1964. Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables. Dover, New York.

Balakrishnan, N., Aggarwala, R., 2000. Progressive Censoring: Theory, Methods and Applications. Birkhäuser, Boston.

Balakrishnan, N., Cohen, A.C., 1991. Order Statistics and Inference: Estimation Methods. Academic Press, San Diego.

Balakrishnan, N., Kannan, N., Lin, C.T., Ng, H.K.T., 2002a. Point and interval estimation for the normal distribution based on progressive Type-II censored samples. IEEE Trans. Reliability, to appear.

Balakrishnan, N., Kannan, N., Lin, C.T., Wu, J.S., 2002b. Inference for the extreme value distribution under progressive Type-II censoring. Submitted for publication.

Balasooriya, U., Saw, S.L.C., Gadag, V., 2000. Progressively censored reliability sampling plans for the Weibull distribution. Technometrics 42, 160–168.

Cohen, A.C., 1963. Progressively censored samples in life testing. Technometrics 5, 327–339.

Cohen, A.C., 1991. Truncated and Censored Samples: Theory and Applications. Marcel Dekker, New York.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39, 1–38.

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

Little, R.J.A., Rubin, D.B., 1983. Incomplete data. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 4. Wiley, New York, pp. 46–53.

Louis, T.A., 1982. Finding the observed information matrix when using the EM algorithm. J. Roy. Statist. Soc. Ser. B 44, 226–233.

Mann, N.R., 1971. Best linear invariant estimation for Weibull parameters under progressive censoring. Technometrics 13, 521–533.

McLachlan, G.J., Krishnan, T., 1997. The EM Algorithm and Extensions. Wiley, New York.

Mi, J., Balakrishnan, N., 2002. Existence and uniqueness of the MLEs for normal distribution based on general progressively Type-II censored samples. Submitted for publication.

Stephens, M.A., 1977. Goodness of fit for the extreme value distribution. Biometrika 64, 583–588.

Tanner, M.A., 1993. Tools for Statistical Inference: Observed Data and Data Augmentation Methods, 2nd Edition. Springer, New York.

Thomas, D.R., Wilson, W.M., 1972. Linear order statistic estimation for the two-parameter Weibull and extreme value distributions from Type-II progressively censored samples. Technometrics 14, 679–691.

Viveros, R., Balakrishnan, N., 1994. Interval estimation of parameters of life from progressively censored data. Technometrics 36, 84–91.