Parameter Estimation using EM algorithm for Constant-stress and Step-stress Accelerated Life Tests under Interval Monitoring

David Han* and Tianyu Bai*

Abstract
Accelerated life tests quickly produce information on the lifetime distribution of a test unit by running the tests at higher stress levels than normal operating conditions. Using a regression model, the lifetime parameter at the normal design stress is then estimated via extrapolation. Recently, the design optimization of accelerated life tests has been studied by many authors but the associated inference for the regression parameters has not been. In this work, the EM algorithm is used to determine the maximum likelihood estimates of the regression parameters for time constrained exponential failure data from the constant-stress and step-stress accelerated life tests under interval monitoring. It is demonstrated that the method is feasible as well as easy to implement. The proposed method is illustrated using a real engineering case study.

Key Words: accelerated life tests, constant-stress loading, EM algorithm, interval monitoring, maximum likelihood estimation, progressive Type-I censoring, step-stress loading

1. Introduction

Thanks to the innovative modern manufacturing process and technology, the product reliability is continuously increasing, resulting in substantially long lifetimes of products. This, in turn, makes it often difficult to obtain sufficient information about the failure time distribution of the products under the standard life tests at normal usage conditions. This practical barrier can be overcome by accelerated life tests (ALT) where the test units are subjected to higher stress levels than normal so that more failure data can be collected in a shorter period of time. Through applying more severe stresses, ALT enables rapid collection of information on the parameters of failure time distributions. Using a stress-response regression model, the lifetime at the normal operating stress can be estimated via extrapolation.

Recently, the parameter estimation for ALT models at individual stress levels along with its design optimization has been discussed by many authors; see, for instance, Miller and Nelson (1983), Bai et al. (1989), Nelson (1990), Meeker and Escobar (1998), Bagdonavicius and Nikulin (2002), Han et al. (2006), Balakrishnan and Han (2008, 2009), Han and Balakrishnan (2010), Han and Ng (2013), Han and Kundu (2015), and Han (2015). However, the associated inference for the regression parameters has not been studied in detail. In the literature, the estimation problem has been approached by different techniques including probability plots, method of moments, and maximum likelihood estimation (MLE). In particular, solving a series of likelihood equations is required to obtain the MLE computationally. Moreover, despite its effectiveness in time and cost savings, progressively censored sampling has not gained popularity in ALT, partly due to the complexity of its likelihood function; see Cohen (1963), Lawless (1982), Balakrishnan et al. (2010). The Newton-Raphson algorithm has been one of the standard methods to calculate the MLE of the model parameters, and to employ this algorithm, one needs to derive the second derivatives of the log-likelihood, which may be complicated under progressive censoring.

Our research in this paper is motivated by the following real case study. A three-level step-stress ALT was conducted under progressive Type-I censoring in order to assess

*Department of Management Science and Statistics, University of Texas at San Antonio, TX 78249

Table 1: Progressively Type-I censored dataset from $n = 30$ prototypes of a solar lighting device on a three-level step-stress ALT with $\tau_1 = 15$, $\tau_2 = 20$, and $\tau_3 = 25$

Failure Count at Temperature Level 1 ($x_1 = 0.1$)	Failure Count at Temperature Level 2 ($x_2 = 0.5$)	Failure Count at Temperature Level 3 ($x_3 = 0.9$)
$n_1 = 11$ $c_1 = 4$	$n_2 = 7$ $c_2 = 1$	$n_3 = 4$ $c_3 = 3$
$n_\oplus = 22,\ c_\oplus = 8$		

the reliability characteristics of a solar lighting device, whose dominant failure mode is controller failure. Here, temperature is the stress factor whose level was changed during the test in the range of 293K to 353K with the normal operating temperature at 293K. The standardized stress loading was $x_1 = 0.1$, $x_2 = 0.5$, and $x_3 = 0.9$. The stress change time points were $\tau_1 = 15$ (in hundred hours) and $\tau_2 = 20$ (in hundred hours) with the termination time point at $\tau_3 = 25$ (in hundred hours). The number of devices censored at $\tau_1 = 15$ and $\tau_2 = 20$ were $c_1 = 4$ and $c_2 = 1$, respectively. The dataset consists of total $n_\oplus = 22$ failure times from the initial sample size of $n = 30$ prototypes ($i.e.$, 26.7% right censoring). The observed number of failure times at each temperature is presented in Table 1 without exact failure times due to interval/group monitoring.

One of the objectives of the study is to assess the stress-response regression model, and by expanding the scope, it is desired to determine the MLE of the regression parameters for interval monitored failure data under the general k-level constant-stress and step-stress ALT. Here, we propose to apply the expectation-maximization (EM) algorithm instead, which has been successfully applied under various problem settings; see, for example, Ng et al. (2002), Nandi and Dewan (2010), Balakrishnan and Ling (2012, 2013). The EM algorithm is a powerful technique in handling the incomplete data problem, and it is particularly useful when the augmented dataset is relatively easier to analyze; see Dempster et al. (1977), McLachlan and Krishnan (1997). It works by iterating the process of filling in the missing data with the estimated values and updating the parameter estimates until convergence. It is assumed here that the lifetimes of the products are exponentially distributed at each stress level, and the relationship between the mean lifetime parameter and stress level is log-linear. To explain the effect of changing stress in the step-stress ALT, the accelerated failure time (AFT) model is also adopted. The proposed method is demonstrated to be feasible as well as easy to implement for practitioners.

The rest of the paper is organized as follows. Section 2 presents the model descriptions and formulations for Type-I censored k-level constant-stress ALT and progressively Type-I censored k-level step-stress ALT. Section 3 describes how the EM algorithm is used to derive the MLE of the model parameters under the unified structure of the likelihoods. In Section 4, the proposed method is illustrated using the case study described above. Finally, Section 5 is devoted to some concluding remarks.

2. Model Formulations and MLE

Let us denote $s(t)$ to be the specified stress loading (a deterministic function of time) for the ALT under consideration. We define s_H to be the upper bound of stress level and s_U to be the normal use-stress level. The stress loading is then standardized as

$$x(t) = \frac{s(t) - s_U}{s_H - s_U}, \qquad t \geq 0$$

so that the range of $x(t)$ is between 0 and 1 inclusive. Now, let us define $0 \equiv x_0 \leq x_1 < x_2 < \cdots < x_k \leq 1$ to be the ordered k standardized stress levels used in the ALT. It is further assumed that the lifetime of a test unit follows an exponential distribution under any stress level x_i. Its probability density function (PDF) and cumulative distribution function (CDF) are given by

$$f_i(t) = \frac{1}{\theta_i} \exp\left(-\frac{t}{\theta_i} \right), \qquad 0 < t < \infty, \tag{1}$$

$$F_i(t) = 1 - S_i(t) = 1 - \exp\left(-\frac{t}{\theta_i} \right), \qquad 0 < t < \infty, \tag{2}$$

respectively. It is also assumed that under any stress level x_i, the mean time to failure (MTTF) of a test unit, θ_i, has a log-linear relationship with stress level, specified by

$$\log \theta_i = \alpha + \beta x_i, \tag{3}$$

where the regression parameters α and β are to be calibrated. For the accelerated exponential distribution, the log-linear link in (3) is a well-studied model. Its popularity is not only due to its simplicity but also due to the fact that several life-stress relationships built from physical principles are well represented by this log-linear link, which includes Arrhenius, Eyring, inverse power law, temperature-humidity, and temperature-non-thermal; see Miller and Nelson (1983).

Here we consider two popular classes of ALT: constant-stress and step-stress. In the constant-stress ALT, a test unit is subjected to a certain stress level and tested until the termination time point of the life test. In the (step-up) step-stress ALT, on the other hand, the stress levels are gradually increased at some predetermined time points during the test. The following subsections present the likelihoods and the MLE of α and β for general k-level constant-stress ALT and step-stress ALT under (progressively) Type-I censored interval monitoring. For simplicity, in this paper, we do not make notational distinctions between the random variables and their corresponding realizations. The usual conventions are also adopted, that is $\sum_{j=m}^{m-1} a_j \equiv 0$ and $\prod_{j=m}^{m-1} a_j \equiv 1$.

2.1 *k-level Step-stress ALT under Progressive Type-I Censoring*

Let us denote n_i to be the number of units failed at stress level x_i in time interval $[\tau_{i-1}, \tau_i))$ while c_i denotes the number of units censored at time τ_i, for $i = 1, 2, \ldots, k$. Also, let N_i denote the number of units operating and surviving on test at the beginning of stress level x_i, such that $N_i = n - \sum_{j=1}^{i-1} n_j - \sum_{j=1}^{i-1} c_j$. Then, a step-stress ALT under progressive Type-I censoring proceeds as follows. Initially, a total of $N_1 \equiv n$ test units are placed at stress level x_1 and they are tested until the first stress change time τ_1. At τ_1, c_1 functioning units are removed from the test in an arbitrary manner, and the stress level is subsequently changed to x_2. The test continues on $N_2 = n - n_1 - c_1$ remaining units until the second stress change time τ_2, at which c_2 live items are removed from the test and the stress level is changed to x_3, and so on. Finally, all the surviving items are withdrawn at time τ_k, thereby terminating the ALT. The number of surviving items at time τ_k is $c_k = n - \sum_{i=1}^{k} n_i - \sum_{i=1}^{k-1} c_i = N_k - n_k$ since $n \equiv \sum_{i=1}^{k}(n_i + c_i)$. The k-level step-stress ALT under conventional Type-I right censoring is a special case when no intermediate censoring takes place ($viz.$, $c_1 = c_2 = \cdots = c_{k-1} = 0$). This implies that the k-level step-stress ALT under complete sampling is also a special case if no right censoring takes place ($viz.$, $\tau_k = \infty$ and $n_k = N_k$).

For a non-constant stress loading, one needs an additional model in order to represent the effect of changing stresses. In many cases, the AFT model, also known as the additive accumulative damage model, was proven to be appropriate. For the exponential lifetime distribution, it also generalizes a number of well-applied models in reliability engineering, which includes the basic (linear) cumulative exposure model and the PH model. Now, under the AFT model along with the assumption of exponentiality, the PDF and CDF of a test unit are

$$f(t) = \left[\prod_{j=1}^{i-1} S_j(\Delta_j)\right] f_i(t-\tau_{i-1}) \quad \text{if} \begin{cases} \tau_{i-1} \le t \le \tau_i & \text{for } i = 1, 2, \ldots, k-1 \\ \tau_{k-1} \le t < \infty & \text{for } i = k \end{cases}, \tag{4}$$

$$F(t) = 1 - \left[\prod_{j=1}^{i-1} S_j(\Delta_j)\right] S_i(t-\tau_{i-1}) \quad \text{if} \begin{cases} \tau_{i-1} \le t \le \tau_i & \text{for } i = 1, 2, \ldots, k-1 \\ \tau_{k-1} \le t < \infty & \text{for } i = k \end{cases}, \tag{5}$$

where $\Delta_j = \tau_j - \tau_{j-1}$ is the step duration at stress level x_j, and $f_i(t)$ and $F_i(t)$ are as given in (1) and (2), respectively. Then, using (5), the joint probability mass function (JPMF) of $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$ is obtained as

$$\begin{aligned} f_J(\boldsymbol{n}) &= \prod_{i=1}^{k} \binom{N_i}{n_i} \big[F(\tau_i) - F(\tau_{i-1})\big]^{n_i} \big[1 - F(\tau_i)\big]^{c_i} \\ &= \left[\prod_{i=1}^{k} \binom{N_i}{n_i}\right] \left[\prod_{i=1}^{k} \left(1 - \exp\left(-\frac{\Delta_i}{\theta_i}\right)\right)^{n_i}\right] \exp\left(-\sum_{i=1}^{k} \frac{\Delta_i}{\theta_i}(N_i - n_i)\right). \end{aligned} \tag{6}$$

Upon using (6) and the log-linear link specified in (3), the log-likelihood function of (α, β) can be written as

$$l(\alpha, \beta) = \sum_{i=1}^{k} n_i \log\left(1 - e^{-\Delta_i \exp\left[-(\alpha+\beta x_i)\right]}\right) - \sum_{i=1}^{k} \Delta_i (N_i - n_i) \exp\left[-(\alpha+\beta x_i)\right]. \tag{7}$$

After differentiating (7) with respect to α and β, the MLE $\hat{\alpha}$ and $\hat{\beta}$ are derived by solving the following two equations simultaneously:

$$\sum_{i=1}^{k} n_i \Delta_i \frac{\exp\Big(-(\alpha+\beta x_i) - \Delta_i e^{-(\alpha+\beta x_i)}\Big)}{1 - e^{-\Delta_i \exp\left[-(\alpha+\beta x_i)\right]}} = \sum_{i=1}^{k} \Delta_i (N_i - n_i) \exp\big[-(\alpha+\beta x_i)\big], \tag{8}$$

$$\sum_{i=1}^{k} n_i x_i \Delta_i \frac{\exp\Big(-(\alpha+\beta x_i) - \Delta_i e^{-(\alpha+\beta x_i)}\Big)}{1 - e^{-\Delta_i \exp\left[-(\alpha+\beta x_i)\right]}} = \sum_{i=1}^{k} x_i \Delta_i (N_i - n_i) \exp\big[-(\alpha+\beta x_i)\big]. \tag{9}$$

It is noted that unlike progressive Type-II censoring well studied in the literature, prefixing the progressive Type-I censoring scheme $\boldsymbol{c} = (c_1, c_2, \ldots, c_{k-1})$ comes with an inherent mathematical issue since there is a positive probability that all the units could fail before reaching the last stress level x_k, resulting in an early termination of the life test as well as failing to fully implement $\boldsymbol{c}$; see Balakrishnan and Han (2009), Balakrishnan et al. (2010). For this reason, the assumption of a large sample is usually required so that the planned number of units can be withdrawn at the end of each stress level. Under this assumption, analysis of progressively Type-I censored data is conducted in an approximate/asymptotic manner. In a reliability test, nevertheless, the sample size is usually small and severe censoring might be expected because of budgetary and/or facility constraints. Consequently, the assumption of a large sample cannot be satisfied in such conditions, and the progressive censoring scheme has to be modified so that its feasibility can be guaranteed.

One simple modification is to decide on a fixed proportion of unfailed items to be censored at the end of each stress level x_i, denoted by $\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*, \ldots, \pi_{k-1}^*)$ with $0 \leq \pi_i^* < 1$. Then, the actual number of items censored at the end of x_i is determined by $c_i = \Upsilon((N_i - n_i)\pi_i^*)$ where $\Upsilon(\cdot)$ is a discretizing function of choice, transforming its argument to a non-negative integer. It could be $round(\cdot)$, $floor(\cdot)$, $ceiling(\cdot)$, and $trunc(\cdot)$, for instance. This modification essentially allows the life test to terminate before reaching the last stress level x_k. Since the number of live units at the end of each stress level before censoring takes place is random, the actual censoring scheme $\mathbf{c}$ is also random via this modification. Another modification which can be entertained in practice is first to decide on a fixed number of items to be censored at the end of each stress level x_i, say $\mathbf{c}^* = (c_1^*, c_2^*, \ldots, c_{k-1}^*)$ with $c_i^* \geq 0$ and $\sum_{i=1}^{k-1} c_i^* < n$. Then, the actual number of items censored at the end of x_i is determined to be $c_i = \min\left\{c_i^*, N_i - n_i\right\}$. If the number of remaining units at any censoring time point is less than or equal to the prefixed number of units to be censored at that point, all the remaining units are withdrawn and the life test is terminated. Hence, this modification also allows the life test to terminate earlier than scheduled whenever there are insufficient live units remaining on the test. Since the number of surviving units at the end of each stress level before censoring takes place is random, the actual censoring scheme $\mathbf{c}$ is essentially random as well.

2.2 *k-level Constant-stress ALT under Type-I Censoring*

For illustrative simplicity, let us consider the procedure of a constant-stress ALT under Type-I censoring. A constant-stress ALT under progressive Type-I censoring can be described in a similar manner like in the previous subsection by introducing a set of time points for intermediate censoring. For $i = 1, 2, \ldots, k$, we allocate N_i units on test at stress level x_i such that $\sum_{i=1}^{k} N_i = n$. The allocated units are then tested until the termination time Δ_i at which point all the surviving items are withdrawn and the life test is terminated. Like before, let us denote n_i to be the number of units failed at stress level x_i in time interval $[0, \Delta_i)$. Then, $N_i - n_i$ denotes the number of units censored at time Δ_i. If there is no right censoring ($viz.$, $\Delta_i = \infty$ and $n_i = N_i$), this obviously corresponds to the k-level constant-stress ALT under complete sampling as a special case.

Then, using (2), the JPMF function of $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$ is obtained as

$$\begin{aligned} f_J(\boldsymbol{n}) &= \prod_{i=1}^{k} \binom{N_i}{n_i} \left[F_i(\Delta_i)\right]^{n_i} \left[1 - F_i(\Delta_i)\right]^{N_i - n_i} \\ &= \left[\prod_{i=1}^{k} \binom{N_i}{n_i}\right] \left[\prod_{i=1}^{k} \left(1 - \exp\left(-\frac{\Delta_i}{\theta_i}\right)\right)^{n_i}\right] \exp\left(-\sum_{i=1}^{k} \frac{\Delta_i}{\theta_i}(N_i - n_i)\right), \end{aligned}$$

which is identical to (6). Using (6) and the log-linear link in (3), the log-likelihood function of (α, β) can be written as in (7) and as a result of this unified likelihood structure, we obtain the MLE $\hat{\alpha}$ and $\hat{\beta}$ as simultaneous solutions to (8) and (9).

3. The EM Algorithm for MLE

As shown above, the MLE $\hat{\alpha}$ and $\hat{\beta}$ do not allow explicit expressions and hence, necessitate numerical procedures such as the Newton-Raphson method for estimation. To employ the Newton-Raphson method, however, the second derivatives of the log-likelihood function are required and this may not be convenient to extend or generalize to other censoring schemes. To ease or avoid this issue, we propose to use the EM algorithm to estimate the MLE of the model parameters by treating the unobserved exact failure times as missing data. First, at stress level x_i, let $y_{i,l}$ denote the l-th ordered failure time of n_i failed units for $l = 1, 2, \ldots, n_i$, under both step-stress ALT and constant-stress ALT. Then, the set of unobserved (missing) failure times is defined by $\boldsymbol{y} = (\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_k})$ with $\boldsymbol{y_i} = (y_{i,1}, y_{i,2}, \ldots, y_{i,n_i})$ while $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$ defined in the previous section denotes the observed failure count data. Using this augmented data, the following subsections present the likelihoods for general k-level constant-stress ALT and step-stress ALT under (progressively) Type-I censored continuous monitoring, and explain the procedure to obtain the MLE of α and β using the EM algorithm.

3.1 *k-level Step-stress ALT under Progressive Type-I Censoring*

Using (4) and (5), the distribution function of the continuously-monitored data $(\boldsymbol{y}, \boldsymbol{n})$ is derived as

$$f_J(\boldsymbol{y}, \boldsymbol{n}) = \left[\prod_{i=1}^{k} \frac{N_i!}{(N_i - n_i)!}\right] \left[\prod_{i=1}^{k} \theta_i^{-n_i}\right] \exp\left(-\sum_{i=1}^{k} \frac{U_i}{\theta_i}\right), \tag{10}$$

where

$$U_i = \sum_{l=1}^{n_i} (y_{i,l} - \tau_{i-1}) + (N_i - n_i)\Delta_i \tag{11}$$

for $i = 1, 2, \ldots, k$. Under the continuous monitoring, U_i in (11) is known as the *Total Time on Test* statistic at stress level x_i. Using (10) and the log-linear relationship given in (3), the log-likelihood function of (α, β) is written as

$$l_C(\alpha, \beta) = -\alpha \sum_{i=1}^{k} n_i - \beta \sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} U_i \exp\big[-(\alpha + \beta x_i)\big]. \tag{12}$$

Now, the E-step of the algorithm requires the computation of the conditional expectation of (12) given the observed data $\boldsymbol{n}$ along with the current estimate of the parameters $\big(\alpha^{(h)}, \beta^{(h)}\big)$ at the h-th iteration. By denoting $\big(\alpha^{(h+1)}, \beta^{(h+1)}\big)$ as the next estimate of the parameters, this conditional expectation is written as

$$\begin{aligned} E\Big[l_c\big(\alpha^{(h+1)}, \beta^{(h+1)}\big) \quad \Big| \quad \boldsymbol{n}; \alpha^{(h)}, \beta^{(h)}\Big] &= -\alpha^{(h+1)} \sum_{i=1}^{k} n_i - \beta^{(h+1)} \sum_{i=1}^{k} n_i x_i \\ &\quad - \sum_{i=1}^{k} u_i^{(h)} \exp\Big[-\big(\alpha^{(h+1)} + \beta^{(h+1)} x_i\big)\Big], \end{aligned} \tag{13}$$

where

$$u_i^{(h)} = E\big[U_i \big| \boldsymbol{n}; \alpha^{(h)}, \beta^{(h)}\big] = \sum_{l=1}^{n_i} E\big[y_{i,l} - \tau_{i-1} \big| n_i; \alpha^{(h)}, \beta^{(h)}\big] + (N_i - n_i)\Delta_i. \tag{14}$$

It can be shown that given n_i, $\boldsymbol{y_i} = (y_{i,1}, y_{i,2}, \ldots, y_{i,n_i})$ are distributed jointly as order statistics from a random sample of size n_i from a left- and right-truncated distribution at τ_{i-1} and τ_i, respectively. Using (4) and (5), this truncated PDF is given by $f_{i;trLR}(t) = f(t)/\big(F(\tau_i) - F(\tau_{i-1})\big)$ for $\tau_{i-1} \le t \le \tau_i$, $i = 1, 2, \ldots, k$. After algebraic some manipulations, it is then observed that given n_i, $(y_{i,1} - \tau_{i-1}, y_{i,2} - \tau_{i-1}, \ldots, y_{i,n_i} - \tau_{i-1})$ are distributed jointly as order statistics from a random sample of size n_i from a right-truncated exponential distribution at Δ_i. Using (1) and (2), this truncated PDF is expressed as $f_{i;trR}(t) = f_i(t)/F_i(\Delta_i)$ for $0 \le t \le \Delta_i$, $i = 1, 2, \ldots, k$. Utilizing this distributional property, (14) is simplified as

$$\begin{aligned} u_i^{(h)} &= n_i \left[\theta_i - \Delta_i \frac{S_i(\Delta_i)}{F_i(\Delta_i)}\right]\Bigg|_{(\alpha,\beta)=\left(\alpha^{(h)},\beta^{(h)}\right)} + (N_i - n_i)\Delta_i \\ &= n_i \left[\exp\left(\alpha^{(h)} + \beta^{(h)} x_i\right) - \Delta_i \left(e^{\Delta_i \exp\left[-\left(\alpha^{(h)}+\beta^{(h)} x_i\right)\right]} - 1\right)^{-1}\right] + (N_i - n_i)\Delta_i. \end{aligned} \tag{15}$$

In the M-step of the algorithm, the next estimate of the parameters $\left(\alpha^{(h+1)}, \beta^{(h+1)}\right)$ is obtained by maximizing (13). Upon differentiating (13) with respect to $\alpha^{(h+1)}$ and $\beta^{(h+1)}$, the value of $\beta^{(h+1)}$ is obtained as the solution to

$$\left[\sum_{i=1}^{k} n_i\right]\left[\sum_{i=1}^{k} u_i^{(h)} x_i \exp\Big(-\beta^{(h+1)} x_i\Big)\right] = \left[\sum_{i=1}^{k} n_i x_i\right]\left[\sum_{i=1}^{k} u_i^{(h)} \exp\Big(-\beta^{(h+1)} x_i\Big)\right]. \tag{16}$$

Then, the value of $\alpha^{(h+1)}$ is obtained by plugging the value of $\beta^{(h+1)}$ from (16) into

$$\alpha^{(h+1)} = \log\left(\frac{\sum_{i=1}^{k} u_i^{(h)} \exp\big(-\beta^{(h+1)} x_i\big)}{\sum_{i=1}^{k} n_i}\right). \tag{17}$$

It is observed from (16) and (17) that at least one failure needs to be observed from at least two different stress levels to guarantee the existence of $\left(\alpha^{(h+1)}, \beta^{(h+1)}\right)$. Otherwise, the parameters cannot be estimated. Under this condition, the unique value of $\left(\alpha^{(h+1)}, \beta^{(h+1)}\right)$ exists and it can be determined by using a simple graphical method.

Eventually, the MLE of α and β are obtained by repeating the E- and M-steps iteratively until convergence occurs. For a reasonable starting value $\left(\alpha^{(0)}, \beta^{(0)}\right)$, the estimate of the parameters based on the *pseudo continuously-monitored* sample is suggested. This pseudo sample is obtained by replacing the unobserved failure times $\boldsymbol{y}$ with some pseudo failure times. For example, given the observed data $\boldsymbol{n}$, set $y_{i,l}^* = \tau_{i-1} + \Delta_i \frac{l}{n_i+1}$ for $l = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$ (*i.e.*, equally-spaced failure times). Using this set of pseudo failure times $\boldsymbol{y}^*$ along with the observed failure counts $\boldsymbol{n}$, U_i in (11) can be estimated. Then, the likelihood in (12) is maximized by solving (16) and (17) simultaneously with $u_i^{(h)}$ replaced by U_i. This solution can be used as the initial estimate of the parameters $\left(\alpha^{(0)}, \beta^{(0)}\right)$. Another suggested starting value is the least squares estimate (LSE) from (3). Han and Bai (2018) provided that the MLE of individual θ_i is obtained as $\hat{\theta}_i = \Delta_i\big[\log N_i - \log(N_i - n_i)\big]^{-1}$. Using this collection of pairs $(x_i, \log\hat{\theta}_i)$, a simple LSE of α and β can be obtained as

$$\beta^{(0)} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})\Big(\log\hat{\theta}_i - \overline{\log\hat{\theta}}\Big)}{\sum_{i=1}^{k}(x_i - \bar{x})^2} \qquad \text{and} \qquad \alpha^{(0)} = \overline{\log\hat{\theta}} - \beta^{(0)}\bar{x},$$

where $\bar{x} = \sum_{i=1}^{k} x_i/k$ and $\overline{\log \hat{\theta}} = \sum_{i=1}^{k} \log \hat{\theta}_i/k$. This estimate can be also used to initialize the algorithm.

In summary, the EM algorithm proceeds as follows:

Step 1. Initialize the iteration counter $h = 0$ and choose the initial value $\big(\alpha^{(0)}, \beta^{(0)}\big)$.

Step 2. Compute $u_i^{(h)}$ in (15) with the observed data $\boldsymbol{n}$.

Step 3. Compute $\beta^{(h+1)}$ by solving (16).

Step 4. Compute $\alpha^{(h+1)}$ in (17) with $\beta^{(h+1)}$ from **Step 3**.

Step 5. If $\big|\alpha^{(h+1)} - \alpha^{(h)}\big| < \epsilon$ and $\big|\beta^{(h+1)} - \beta^{(h)}\big| < \epsilon$, break the loop and present $\big(\alpha^{(h+1)}, \beta^{(h+1)}\big)$ as the MLE $(\hat{\alpha}, \hat{\beta})$.

Step 6. Otherwise, increment the counter $h = h + 1$ and go to **Step 2**.

3.2 *k-level Constant-stress ALT under Type-I Censoring*

Using (1) and (2), the distribution function of the continuously-monitored data $(\boldsymbol{y}, \boldsymbol{n})$ is obtained as in (10) where

$$U_i = \sum_{l=1}^{n_i} y_{i,l} + (N_i - n_i)\Delta_i. \tag{18}$$

Again, U_i in (18) is known as the *Total Time on Test* statistic at stress level x_i. Using (10) and the log-linear link function in (3), the log-likelihood function of (α, β) can be written as in (12). Then, the E-step of the algorithm requires the computation of the conditional expectation given in (13) where

$$u_i^{(h)} = E\big[U_i \big| \boldsymbol{n}; \alpha^{(h)}, \beta^{(h)}\big] = \sum_{l=1}^{n_i} E\big[y_{i,l} \big| n_i; \alpha^{(h)}, \beta^{(h)}\big] + (N_i - n_i)\Delta_i. \tag{19}$$

Similarly, it can be shown that given n_i, $\boldsymbol{y_i} = (y_{i,1}, y_{i,2}, \ldots, y_{i,n_i})$ are distributed jointly as order statistics from a random sample of size n_i from a right-truncated exponential distribution at Δ_i. Using (1) and (2), this truncated PDF is given by $f_{i;trR}(t) = f_i(t)/F_i(\Delta_i)$ for $0 \le t \le \Delta_i$, $i = 1, 2, \ldots, k$. This property simplifies (19) to (15). In the M-step of the algorithm, $\big(\alpha^{(h+1)}, \beta^{(h+1)}\big)$ is then obtained as the solution to (16) and (17) for maximizing (13). Eventually, the MLE of α and β are obtained by iterating the E- and M-steps until convergence. For an initial value $\big(\alpha^{(0)}, \beta^{(0)}\big)$ to start the algorithm, the parameter estimate based on the *pseudo continuously-monitored* sample or the LSE for (3) is suggested as described in the previous subsection.

4. Illustrative Example

Section 1 introduced the progressively Type-I censored three-level step-stress ALT under interval monitoring for assessing the reliability characteristics of a solar lighting device. The proposed EM algorithm discussed in Section 3 was applied to the failure count data presented in Table 1. In the initial assessment, Weibull distributions with a uniform shape parameter across different stress levels were preferred. However, the inference for the shape parameter could not reject a simpler exponential lifetime for the device at any constant temperature. Consistent with our model assumptions, the exponential distribution was fitted to the data with the log-linear parameter-stress link function in (3). The proposed estimation procedure produced $\hat{\alpha} = 3.6303$ and $\hat{\beta} = -2.3475$ after 9 iterations of the EM algorithm. The algorithm was initialized by using the LSE for (3), $\left(\alpha^{(0)}, \beta^{(0)}\right) = (3.5196, -2.1456)$ as the starting values. The final results are identical to the estimates from the Newton-Raphson method, and also agree well with the MLE $(\hat{\alpha}, \hat{\beta}) = (3.6597, -2.4131)$ under continuous monitoring (when the exact failure times are available).

5. Conclusion

In the literature, the estimation problem for ALT models under (progressive) censoring has been studied lightly if not at all. The Newton-Raphson method has traditionally been the choice of obtaining the MLE of the model parameters. Motivated by a real case study in reliability engineering, here we implemented the EM algorithm to determine the MLE of the regression parameters for (progressively) Type-I censored exponential failure data from two popular modes of ALT, namely the k-level constant-stress and step-stress ALT under interval monitoring. The method was proven to be feasible as well as easy to apply.

REFERENCES

Bagdonavicius, V., Nikulin, M. (2002). Accelerated Life Models: Modeling and Statistical Analysis. Boca Raton, FL: Chapman Hall.

Bai, D.S., Kim, M.S., Lee, S.H. (1989). Optimum simple step-stress accelerated life tests with censoring. IEEE Transactions on Reliability, 38, 528–532.

Balakrishnan, N., Han, D. (2008). Exact inference for a simple step-stress model with competing risks for failure from exponential distribution under Type-II censoring. Journal of Statistical Planning and Inference, 138, 4172–4186.

Balakrishnan, N., Han, D. (2009). Optimal step-stress testing for progressively Type-I censored data from exponential distribution. Journal of Statistical Planning and Inference, 139, 1782–1798.

Balakrishnan, N., Han, D., Iliopoulos, G. (2010). Exact inference for progressively Type-I censored exponential failure data. Metrika, 73, 335–358.

Balakrishnan, N., Ling, M.H. (2012). EM algorithm for one-shot device testing under the exponential distribution. Computational Statistics and Data Analysis, 56, 502–509.

Balakrishnan, N., Ling, M.H. (2013). Expectation maximization algorithm for one shot device accelerated life testing with Weibull lifetimes and variable parameters over stress. IEEE Transactions on Reliability, 62, 537–551.

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

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society - Series B, 39, 1–38.

Han, D. (2015). Time and cost constrained optimal designs of constant-stress and step-stress accelerated life tests. Reliability Engineering and System Safety, 140, 1–14.

Han, D., Bai, T. (2018). Interval estimation for progressively Type-I censored step-stress accelerated life test under interval monitoring (submitted).

Han, D., Balakrishnan, N. (2010). Inference for a simple step-stress model with competing risks for failure from the exponential distribution under time constraint. Computational Statistics and Data Analysis, 54, 2066–2081.

Han, D., Balakrishnan, N., Sen, A., Gouno, E. (2006). Corrections on Optimal step-stress test under progressive Type-I censoring. IEEE Transactions on Reliability, 55, 613–614.

Han, D., Kundu, D. (2015). Inference for a step-stress model with competing risks from the generalized exponential distribution under Type-I censoring. IEEE Transactions on Reliability, 64, 31–43.

Han, D., Ng, H.K.T. (2013). Comparison between constant-stress and step-stress accelerated life tests under time constraint. Naval Research Logistics, 60, 541–556.

Hu, C.H., Plante, R.D., Tang, J. (2013). Statistical equivalency and optimality of simple step-stress accelerated test plans for the exponential distribution. Naval Research Logistics, 60, 19–30.

Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. New York, NY: Wiley & Sons.

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

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. Journal of Royal Statistical Society - Series B, 44, 226–233.

McCool, J.I. (1980). Confidence limits for Weibull regression with censored data. IEEE Transactions on Reliability, 29, 145–150.

McLachlan, G.J., Krishnan, T. (1997). The EM Algorithm and Extensions. New York, NY: Wiley & Sons.

Meeker, W.Q., Escobar, L.A. (1998). Statistical Methods for Reliability Data. New York, NY: Wiley & Sons.

Miller, R., Nelson, W. (1983). Optimum simple step-stress plans for accelerated life testing. IEEE Transactions on Reliability, 32, 59–65.

Nandi, S., Dewan, I. (2010). An EM algorithm for estimating the parameters of bivariate Weibull distribution under random censoring. Computational Statistics and Data Analysis, 54, 1559–1569.

Nelson, W. (1990). Accelerated Testing: Statistical Models, Test Plans, and Data Analyses. New York, NY: Wiley & Sons.

Ng, H.K.T., Chan, P.S., Balakrishnan, N. (2002). Estimation of parameters from progressively censored data using EM algorithm. Computational Statistics and Data Analysis, 39, 371–386.

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