An EM procedure for the estimation of lifetime distribution using degradation data

Zehua Chen ¹ & Shurong Zheng ²

¹ National University of Singapore.
² China Northeastern Normal University.

Short Title: EM-procedure for degradation analysis

Correspondence to:
Dr. Zehua Chen
Department of Statistics and Applied Probability
National University of Singapore
3 Science Drive 2
Singapore 117543

Tel: (65) 874-6307, Fax: (65) 872-3919
Email: stachenz@nus.edu.sg
Abstract

Commonly used methods for degradation analysis deduce the lifetime distribution of a product from the distribution of the random parameters in the degradation path model for the product. This approach requires the functional form of the distribution of the random parameters, which poses certain practical difficulties. In this paper, we develop an approach which makes inference directly on the lifetime distribution, avoiding assumptions on the distribution of the random parameters. Our approach is essentially an EM-procedure. Degradation data are used to derive an initial estimate of the lifetime distribution and predictive intervals of the individual lifetimes. Then an EM-algorithm incorporating the predictive intervals is iterated to obtain the final estimate of the lifetime distribution. The procedure has the following advantages: (i) the adequacy of the estimated lifetime distribution can be easily examined, (ii) the lifetime distribution has a closed form and (iii) the procedure is more easily implemented. A simulation study is reported to demonstrate the efficiency of the procedure. The procedure is illustrated by two real degradation data sets.

KEY WORDS: Degradation, EM-algorithm, lifetime distribution, reliability.
1 Introduction

Traditionally, reliability analysis is based on lifetime (time-to-failure) data. However, with today's high technologies, many products are designed to work without failure for years. For such highly reliable products, lifetime data can not be obtained in a reasonably short period. Because of the increasingly fierce competition of the international market, the research and development of a product does not allow a long period to conduct traditional life tests. Reliability analysis in such situation poses a great challenge. Degradation data analysis is one of the newly developed approaches for assessing the reliability of certain highly reliable products. Many product failures are caused by some degradation mechanisms such as fatigue, creep, cracks, wear, corrosion, oxidation, and weathering etc.. For a particular unit of such a product, the degradation measurements can be made either continuously or at discrete points in time. Such degradation measurements provide a rich source of information for assessing reliability and, if properly handled, can be used to make inference on the lifetime distribution of the product.

Let $y$ denote an appropriate measure of the failure-causing degradation of the product under consideration. The measurements of $y$ on a particular unit at specific time points can be modeled as

$$y_{ij} = (\mu_0; \mu_i; t_{ij}) + \epsilon_{ij}; \quad i = 1; \ldots; n; j = 1; \ldots; m_i;$$

where $\mu_0$ is a vector of constant parameters which is common for every individual units, $\mu_i$ is a vector of parameters particularly associated with the ith unit, $t_{ij}$ is the jth time point at which y is measured on the ith unit,
is a continuous function which describes the degradation mechanism over time, and $z_{ij}$ are random variables involving measurement errors and certain undetectable ambient environmental factors. The unit-to-unit variability is reflected by the vectors $\mu_i$ which hence are treated as random vectors. The time here is not necessarily real time. It could be operating time or some surrogate like mileage and loading cycles. It is assumed that $z_{ij}$s are independent of $\mu_i$s. In most of practical degradation tests, the $z_{ij}$s can be assumed as independent identically distributed white noises. Since degradation is irreversible, $\gamma$, as a function of $t$, is monotone. Without lose of generality, we assume $\gamma$ is an increasing function of $t$. Model (1) is referred to as the degradation path model or, simply, the path model.

In certain practical cases, the path model can be obtained based on physical or chemical mechanisms. For example, Carey and Tortorella (1988) employed a birth and death process for charge carriers and developed a path model for the degradation of MOS oxides. LuValle,Welsher and Mitchell (1986) and Luvalle,Welsher and Svoboda (1988) derived degradation models based on chemical kinetics and probabilistic considerations. Meeker and Luvalle (1995) described models for growth of failure-causing conducting filaments of chlorine-copper compounds in printed-circuit boards based on certain chemical mechanism. The well-known Paris-law model is used to describe the growth of fatigue cracks, see Meeker and Escobar (1998). A linear degradation path model is used to describe the degradation mechanism of semiconductor devices. Lu, Park and Yang (1997) provided some physical background for the linear degradation model. In the lack of the background on physical or chemical mechanisms, certain empirical models
might be developed to describe the degradation process. However, the path
model is not the focus of this paper. It is assumed in this paper that the
path model is already determined, i.e., the function $\dot{\gamma}$ in (1) is assumed to
be a known function of $\mu_0; \mu$ and $t$.

In degradation analysis, an unit is considered failed if its degradation
level reaches a fixed critical level, say $\dot{\gamma}_c$, the time at which this occurs is
considered as the failure time of the unit. Some authors refer such failure
as "soft failure". For some products, however, the definition of failure is
clear | the product stops working. The failure time does not necessarily
 correspond to a fixed degradation level. The true failure is referred to
as "hard failure". In degradation analysis, dealing with "hard failure" is
tremendously difficult. The approximation to the "hard failure" time by the
"soft failure" time makes it tractable to inference on the lifetime distribution
using degradation data while having no adverse impact from a practical
point of view. The lifetime of unit $i$ is then defined as the solution, $T_i$, of
the equation:

$$
\dot{\gamma}(\mu_0; \mu_i; T) = \dot{\gamma}_c;
$$

(2)

The purpose of degradation data analysis is to make inference on the dis-
tribution of $T_i$ by using the data $y_{ij}$.

The approaches for degradation data analysis found in the literature,
which are commonly used, share the following features. The inference on
the lifetime distribution is induced from the inference on the distribution of
the random parameters $\mu$. A functional form, say $\hat{\gamma}(\mu; \xi)$ where $\xi$ is a vector
of unknown constant parameters, is assumed for the distribution function
of $\mu$. The degradation data are used to estimate $\zeta$. Once the estimate of $\zeta$, say $\hat{\zeta}$, is obtained, the estimated distribution $\frac{1}{2}(\mu; \hat{\zeta})$ together with relation (2) are used to deduce the lifetime distribution. Usually, simulations are needed in this step since the lifetime distribution induced by $\frac{1}{2}(\mu; \zeta)$ does not have a closed form. Lu and Meeker (1993) developed a two-stage least squares method for the estimation of $\zeta$. Wu and Shao (1999) considered a direct ordinary or weighted least squares procedure. Meeker and Escobar (1998) proposed the maximum likelihood estimation of $\zeta$. Robinson and Crowder (2000) explored a Bayesian approach. For other important references on degradation analysis, we mention, to name but a few, Boulanger and Escobar (1994), Tseng, Hamada and Chiao (1995), Lu, Meeker and Escobar (1996), Cox (1999), Shiau and Lin (1999), Bagdonavicius and Nikulin (2000), and Chiao and Hamada (2001).

In this paper, we develop an approach from a different principle. Instead of making inference on the lifetime distribution through the inference on the distribution of the random parameters, we propose a direct inference on the lifetime distribution. To this end, assumptions on the functional form of the lifetime distribution need to be made, however, the needs for making assumptions on the distribution of the random parameters are avoided. Our approach is essentially an EM-procedure. The degradation data are used to derive an initial estimate of the lifetime distribution and predictive intervals of the individual lifetimes for the EM-procedure. Then the EM-procedure is iterated to obtain the final estimate of the lifetime distribution. The EM-procedure, which makes full use of the information in the degradation data, turns out to be very efficient. Besides, there are at least the following
advantages: (i) the distributional assumption can be easily examined and modified since lifetime is one-dimensional, (ii) the lifetime distribution can be represented by a few parameters since it has a closed form and (iii) the procedure is more easily implemented since simulations necessary for the commonly used approaches to go from the distribution of the random parameters to the lifetime distribution are no longer needed.

In Section 2, we give a further discussion about the motivation for direct inference on the lifetime distribution. In Section 3, the EM-procedure is developed in detail. In Section 4, we present some simulation studies which demonstrate the efficiency of the EM-procedure. In Section 5, the EM-procedure is illustrated by two real data sets.

2 Motivation

As mentioned in the introduction, for the commonly used degradation data analysis approaches, an assumption on the functional form of the distribution of the random parameters needs to be made. For the EM-procedure we are going to develop, an assumption on the functional form of the lifetime distribution needs to be made. Conceptually, an assumption on the distribution of the random parameters corresponds to an assumption on the lifetime distribution, and vice versa. However, practically, the difficulties of making assumptions on the distribution of the random parameters are quite different from those of making assumptions on the lifetime distribution. We elaborate on this in the following.

In most of practical problems, the random parameters enter into path
model (1) in unfamiliar forms such that little information about their distribution is known. It poses a great challenge in practice to make a reasonable assumption on the distribution of the random parameters. It is suggested by several authors, for example Lu and Meeker (1993), that the distribution of the random parameters can be assumed as normal, if not, a transformation of the original parameters can be made so that the transformed parameters follow a normal distribution. Though this strategy works in certain cases, there are some difficulties which can not be easily overcome. First, when the dimension of the parameters is higher than 1, it is not easy to examine the validity of the normality assumption. To find an appropriate transformation is not a trivial task. Second, more seriously, not every random vector can be transformed into a normal vector.

Another dilemma with the assumption on the distribution of the random parameters is that not every distribution of the random parameters induces a proper lifetime distribution. The following example was considered in the literature by many authors. Let the path model be given by

\[ \dot{(\mu; t)} = \mu_1 - \mu_2 t; \]

where \((\mu_1; \mu_2)^T\) is assumed to follow a bivariate normal distribution with mean vector and covariance matrix given by

\[
\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}; \quad \mathbf{\Sigma} = \begin{pmatrix} \xi_1^2 & \frac{\eta_{12}}{\xi_1 \xi_2} \\ \frac{\eta_{12}}{\xi_1 \xi_2} & \xi_2^2 \end{pmatrix};
\]

Under the constrains that \(b_1 > \dot{c}; \dot{b}_2 > 0\) and \(P \{\mu_1 > \dot{c}\} \approx 1; P \{\mu_2 > 0\} \approx 1\), Lu and Meeker (1993) derived an approximate expression for the induced lifetime distribution as follows.

\[
\overline{F}_T(t) = \int_{\{ b_1 \sqrt{2\xi_1 \xi_2 t + \xi_2^2 t^2} = b_2 \}^{b_1} } \sim \frac{(b_1 i - \dot{c} = b_2)}{[\xi_1^2 i + 2\xi_1 \xi_2 t + \xi_2^2 t^2]} \cdot t \approx 0;
\]
However, as pointed out by Zhuang (1994), the approximation $F_T$ is not a proper distribution function. It is not difficult to see that

$$
F_T(+1) = \lim_{t \to 1} F_T(t) = \frac{\hat{b}_2 - \hat{c}_2}{\hat{d}_2} < 1; \\
F_T(0) = \lim_{t \to 0} F_T(t) = \frac{\hat{b}_1 - \hat{c}_1}{\hat{d}_1} > 0.
$$

Furthermore, $F_T(t)$ is not even an increasing function in certain range of $t$. Zhuang (1994) also considered the exact distribution of the induced lifetime in this model and found that the expected lifetime is infinity. This example is not unique. Other examples can also be found in Lu and Meeker (1993) and Zhuang (1994). This dilemma complicates matters further in making assumptions on the distribution of the random parameters.

On the contrast, we are in a much better position to make assumptions on the lifetime distribution. First, we have a rich repertoire of distribution families such as Gamma, Weibull, Log-normal distributions etc. which can be used to model the lifetime of manufactured products. In certain practical cases, people usually have certain idea about the type of the lifetime distribution from past experiences. On the other hand, since lifetime is one-dimensional, it is easy to check the validity of a tentative assumption and finally lead to an appropriate assumption about the lifetime distribution.

The above discussion motivates us to promote approaches which avoid assumptions on the distribution of the random parameters and are based on assumptions about the lifetime distribution itself. The EM-procedure is developed in this spirit.
3 The EM-procedure

Let the probability density function and the cumulative distribution function of the lifetime be denoted by \( f(t; \theta) \) and \( F(t; \theta) \) respectively, where \( \theta \) is the vector of unknown parameters. Before the EM-procedure is called, we need to provide an initial estimate of \( F(t; \theta) \). This is done through the so-called pseudo-lifetimes obtained from the degradation data. We start our discussion on the pseudo-lifetimes, then consider how to choose \( F(t; \theta) \) based on the pseudo-lifetimes and finally describe the EM-procedure in detail.

(i) The pseudo-lifetimes and their properties. We make some slightly less general assumptions than expressed in path model (1). Assume that \( t_{ij} = t_j \) for all \( i \) and \( j \) and \( m_i \neq m \), that is, the degradation measurements made on each individual unit are at the same time points for the same number of times. Assume that \( \mu_i; i = 1; \ldots; n \), are i.i.d. and independent of \( \xi_{ij} \)'s and that \( \xi_{ij} \)'s are i.i.d. with mean zero and variance \( \varphi^2 \).

First, we solve the following minimization problem:

\[
\min_{\mu_0} \sum_{i=1}^{n} \sum_{j=1}^{m} \left[ y_{ij} - \xi_{ij}(\mu_0, \mu_i; t_j) \right]^2
\]

with respect to \( \mu_0 \) and \( \mu_i; i = 1; \ldots; n \): Let \( \hat{\mu}_0; \hat{\mu}_1; \ldots; \hat{\mu}_n \) denote the minimizer. Then, for each \( i \), solve

\[
\xi_{ij}(\hat{\mu}_0, \hat{\mu}_i; t) = 0
\]

Denote the solution by \( T_i^n \) which is referred to as the pseudo-lifetime of the \( i \)th unit.
In what follows we discuss some properties of the pseudo-lifetimes. We start with the properties of the conditional least square estimates $\hat{\mu}_i$s. Denote the dimension of $\mu_0$ by $q$ and the dimension of $\mu_i; i \in \mathbb{N}_0$ by $p$. Let $\mu = (\mu_0^T, \mu_1^T, \ldots, \mu_n^T)^T$ and $\hat{\mu}$ denote the vector of the corresponding least square estimates. Let $\hat{\mu} = (\hat{\mu}_0^T, \hat{\mu}_1^T, \ldots, \hat{\mu}_n^T)^T$

Let $\hat{\mu} = (\hat{\mu}_0^T, \hat{\mu}_1^T, \ldots, \hat{\mu}_n^T)^T$

\[
\frac{\partial^2}{\partial \mu_0^T} = \begin{bmatrix}
\frac{\partial^2}{\partial \mu_0^T} \frac{\partial}{\partial \mu_1^T} & \cdots & \frac{\partial^2}{\partial \mu_0^T} \frac{\partial}{\partial \mu_n^T}
\end{bmatrix}
\]

and

\[
\frac{\partial^2}{\partial \mu_i^T} = \begin{bmatrix}
\frac{\partial^2}{\partial \mu_i^T} \frac{\partial}{\partial \mu_0^T} & \cdots & \frac{\partial^2}{\partial \mu_i^T} \frac{\partial}{\partial \mu_n^T}
\end{bmatrix}
\]

Let

\[
G(\mu) = \begin{bmatrix}
\frac{\partial^2}{\partial \mu_0^T} & \cdots & \frac{\partial^2}{\partial \mu_0^T} \\
\frac{\partial^2}{\partial \mu_1^T} & \cdots & \frac{\partial^2}{\partial \mu_1^T} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2}{\partial \mu_n^T} & \cdots & \frac{\partial^2}{\partial \mu_n^T}
\end{bmatrix}
\]

The following are some classical results on non-linear least square estimates which can be found in standard text books.

**Lemma 1** Under certain regularity conditions,

(a) Conditioning on $\mu$, the conditional estimate $\hat{\mu}$ is strongly consistent, i.e., $\hat{\mu} \to \mu$ almost surely as $m \to 1$.

(b) The conditional estimate $\hat{\mu}$ is asymptotically normally distributed with mean $\mu$ and covariance matrix

\[
\text{Var} (\hat{\mu}_0^T) = \frac{1}{m} [G(\mu)^T G(\mu)]^{-1}.
\]
(c) Let
\[ \tilde{\gamma}^2 = \frac{1}{nm \cdot np} \sum_{i=1}^{n} \sum_{j=1}^{m} [y_{ij} - (\hat{\mu}_0; \hat{\mu}_i; t_j)]^2. \]

The \( \tilde{\gamma}^2 \) defined above is a consistent estimate of \( \gamma^2. \)

Now let us turn to the pseudo-lifetimes. The equation
\[ \gamma_{ij}(\mu_0; \mu_i; t) = \gamma_c \]
determines its solution \( T_i \) as an implicit function of \( \mu_0 \) and \( \mu_i \), say, \( T_i = t(\mu_0; \mu_i). \) The properties of \( \gamma_{ij}(\mu_0; \mu_i; t) \) such as continuity and differentiability can be carried over to the implicit function \( t(\mu_0; \mu_i). \) Denote
\[ \frac{\partial}{\partial \mu_0} = \frac{\partial t(\mu_0; \mu_i)}{\partial \mu_0}, \quad \frac{\partial}{\partial \mu_i} = \frac{\partial t(\mu_0; \mu_i)}{\partial \mu_i}. \]

Let \( H_0; H_1 \) and \( H_{0i} \) be, respectively, the variances of \( \hat{\mu}_0, \hat{\mu}_i \) and the covariance between \( \hat{\mu}_0 \) and \( \hat{\mu}_i \) after dividing by \( \gamma^2. \) By Taylor expansion, we have
\[ t(\hat{\mu}_0; \hat{\mu}_i) = t(\mu_0; \mu_i) + \frac{\partial}{\partial \mu_0}(\hat{\mu}_0 - \mu_0) + \frac{\partial}{\partial \mu_i}(\hat{\mu}_i - \mu_i) + O_p\left(\frac{1}{m}\right); \quad (3) \]

The above Taylor expansion and Lemma 1 imply the following results.

Theorem 1 (a) A \( 100(1 - \alpha) \% \) predictive interval of \( T_i \) is given by
\[ T_i \pm z_{\alpha/2} \frac{3}{2} \frac{\gamma^2}{\text{Var}(T_i)} \]
where \( z_{\alpha/2} \) denotes the \( (1 - \alpha/2) \)th quantile of the standard normal distribution and
\[ \gamma^2(\mu) = \frac{\text{Var}(T_i)}{\text{Var}(T_i) + \frac{1}{m}}. \]

(b) \( E T_i = E T_i + O\left(\frac{1}{m}\right); \)

(c) \( \text{Var}(T_i) = \text{Var}(T_i) + \frac{3}{2} \text{Var}(\gamma^2(\mu) + O\left(\frac{1}{m}\right)). \)
Remark: If there are no common constant parameters in path model (1) then

\[ \frac{\partial^2}{\partial \mu_i^2} \left( \frac{\partial^2}{\partial \mu_j^2} \right) = \frac{1}{\theta^2} \left( \frac{\partial^2}{\partial \mu_i^2} \right)_j \frac{1}{\theta^2} \left( \frac{\partial^2}{\partial \mu_i^2} \right)_i. \]

(ii) The moment estimates and the choice of lifetime distribution. Let \( \mu_T \) and \( \theta^2_T \) denote the mean and variance of the lifetime respectively. It is obvious from Theorem 1 that asymptotically unbiased estimates of \( \mu_T \) and \( \theta^2_T \) can be obtained as

\[ \hat{\mu}_T = \frac{1}{n} \sum_{i=1}^{n} T_i; \quad (4) \]

\[ \hat{\theta}_T^2 = \hat{\theta}_T^2 T_i \hat{\mu}_T; \quad (5) \]

where

\[ \hat{\theta}_T = \frac{1}{n} \sum_{i=1}^{n} (T_i - \hat{\mu}_T)^2 \]

and

\[ \hat{\theta}_T (\hat{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\theta^2} \left( \frac{\partial^2}{\partial \mu_i^2} \right)_i \]

Next, we choose \( F(t; \theta) \) from our repertoire of lifetime distribution families. Suppose we have \( l \) families, \( G_1(t; \theta); G_2(t; \theta); \ldots; G_l(t; \theta) \), in our repertoire. For example, \( G_1(t; \theta) \) could be the Weibull distribution family, \( G_2(t; \theta) \) could be the Gamma distribution family, and so on. For each of the families, let the parameters of the family be estimated by the method-of-moments using \( \hat{\mu}_T \) and \( \hat{\theta}_T^2 \) as the estimates of \( \mu_T \) and \( \theta^2_T \) respectively. Then we compute and compare the likelihoods

\[ L(G_k) = \prod_{i=1}^{n} g_k(T_i^\theta, \hat{\theta}); k = 1; \ldots; l; \]

where \( g_k \) is the density function corresponding to \( G_k \) and \( \hat{\theta} \) is the method-of-moments estimate of \( \theta \). Finally, we take \( G_k(t; \theta) \) which has the largest
likelihood as $F(t; \hat{\theta})$. The initial estimate of the lifetime distribution is then taken as $F(t; \hat{\theta})$.

(iii) The EM-procedure. Let $t_{j(i)}$ denote the maximum of the $t_j$'s which are less than or equal to $T_i^\mu$. Let

$$T_i^L = \max t_{j(i)}: T_i^\mu \leq \mu t_{j(i)} + \mu z_{\hat{\theta}=2} \sqrt{\frac{3}{2} \hat{\mu}} g,$$

$$T_i^U = \min t_{j(i)+1}: T_i^\mu + z_{\hat{\theta}=2} \sqrt{\frac{3}{2} \hat{\mu}} g,$$

where $t_{j(i)+1} = 1$ if $T_i^\mu > t_m$. The interval $[T_i^L; T_i^U]$ is a modified 100(1-@)% predictive interval of $T_i$.

The EM-procedure alternates between E-steps and M-steps. In an E-step, each individual lifetime $T_i$ is predicted by its conditional expectation under the most recently estimated lifetime distribution given that it falls into the predictive interval $[T_i^L; T_i^U]$. The M-step that follows estimates $\hat{\theta}$ by the maximum likelihood estimation based on the predicted lifetimes obtained in the previous E-step. The procedure can be implemented by the following algorithm.

The Algorithm

Let $\hat{\theta}^{(1)}$ be the initial estimate of $\theta$. For $k = 1, 2, \ldots$, take $F_k(t) = F(t; \hat{\theta}^{(k)})$, do

E-step For $i = 1, \ldots, n$, predict $T_i$ by $T_i^\mu$, the conditional expectation of $T_i$ under $F_k$ given that $T_i \in [T_i^L; T_i^U]$, i.e.,

$$T_i = \frac{R_{T_i} \cdot t_f(t; \hat{\theta}^{(k)}) dt}{F_k(T_i^U) - F_k(T_i^L)}.$$

13
M-step. Maximize $\prod_{i=1}^{P} \log(f(T_i; \theta))$ with respect to $\theta$ to obtain the updated MLE $\theta^{(k+1)}$.

Repeat until convergence occurs.

The choice of the confidence level $\alpha$ for the predictive intervals in the EM-procedure needs to be addressed. The $\alpha$ can not be taken too small or too large. If $\alpha$ is too small, the predictive intervals might be too wide to provide any useful information. On the other hand, if $\alpha$ is too large, the confidence level of the predictive intervals is low. Some balance must be stricken. To put the issue in an asymptotic context, $\alpha$ should be chosen such that, as $m \to \infty$, both $\alpha$ and the lengths of the predictive intervals approach zero. In a finite sample context, as a guideline, $\alpha$ should be chosen as small as possible while the ordering of the resultant predictive intervals remains the same as the ordering of the pseudo-lifetimes in the following sense. Suppose $T_i^\alpha < T_j^\alpha$. Let $[T_i^{L}; T_i^{U}]$ and $[T_j^{L}; T_j^{U}]$ be the predictive intervals of $T_i$ and $T_j$ respectively. The same ordering of the predictive intervals and the pseudo-lifetimes means that $T_i^{L} < T_j^{L}$ and $T_i^{U} < T_j^{U}$ and, at least in one of the inequalities, the strict inequality holds.

4 Simulation studies

We report in this section a simulation study to demonstrate the efficiency of the EM-procedure developed in the previous section. In the simulation study, the following degradation model is considered:

$$y_{ij} = \log(1 + \mu_{ij}) + \epsilon_{ij}; i = 1; \ldots; 50; j = 1; \ldots; m_i.$$
where \( t_j = 0.3j \) and \( T_s \) denotes the stopping time. The critical level of degradation is set at \( \gamma_c = 8 \). The lifetime distribution is taken as \( \text{Lognormal}(\gamma; \frac{3}{2}) \) where \( \gamma = 2 \) and \( \frac{3}{2} = 1 \). Three stopping times are considered: \( T_s = 4.5; 7.5 \) and \( 12.3 \), which roughly correspond, in turn, to the 0.3th, 0.5th and 0.7th quantiles of the log-normal distribution. The corresponding \( m_{T_s} \) are 15, 25 and 41. The degradation data for each \( i \) are generated as follows. First, the random number \( T_i \) is generated from \( \text{Lognormal}(2,1) \), then the corresponding \( \mu_i \) is computed as \( \mu_i = (e^{\gamma} - 1)T_i \) and, finally, \( y_{ij}'s \) are generated according to the degradation model with \( \epsilon_{ij}'s \) generated as i.i.d. from a normal distribution with mean zero and variance \( \frac{3}{2} \). Here \( \frac{3}{2} \) is taken as 0.01. For each \( T_s \), 5000 copies of degradation data are generated. For each copy of data, the EM-procedure is applied to obtain the estimates of \( \gamma \) and \( \frac{3}{2} \), and the log-normal distribution with the estimated \( \gamma \) and \( \frac{3}{2} \) is used to compute the 0.5th, 0.7th and 0.9th quantiles of the lifetime distribution. The confidence level of the predictive intervals is taken as 90% in all the cases. The average and the mean squared error over the 5000 copies are computed for each of these estimates. The same quantities are also computed for the initial method-of-moments estimates. These quantities are presented in Table 1. In the table, MM and EM stand for \"Method-of-Moments\" and \"EM-procedure\" respectively. The numbers in the parentheses of the first column are the true values of the parameters.

It can be seen from Table 1 that (i) the EM-procedure provides approximately unbiased estimates for both \( \gamma \) and \( \frac{3}{2} \) and (ii) the procedure is very efficient. It should be noted that the Cramer-Rao lower bounds for the variances of unbiased estimates of \( \gamma \) and \( \frac{3}{2} \) based on full uncensored lifetime
Table 1: Results of the simulation study

<table>
<thead>
<tr>
<th>Method</th>
<th>mm</th>
<th>em</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{p}$ (2)</td>
<td>2.094</td>
<td>2.030</td>
</tr>
<tr>
<td>MSE $\hat{p}$</td>
<td>0.797</td>
<td>0.942</td>
</tr>
<tr>
<td>MSE $\hat{q}$</td>
<td>0.070</td>
<td>0.024</td>
</tr>
<tr>
<td>$\Phi_5$ (7.389)</td>
<td>8.326</td>
<td>7.704</td>
</tr>
<tr>
<td>$\Phi_7$ (12.483)</td>
<td>13.083</td>
<td>12.804</td>
</tr>
<tr>
<td>$\Phi_9$ (26.617)</td>
<td>25.749</td>
<td>26.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>mm</th>
<th>em</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{q}$ (1)</td>
<td>2.081</td>
<td>2.002</td>
</tr>
<tr>
<td>MSE $\hat{p}$</td>
<td>0.851</td>
<td>0.956</td>
</tr>
<tr>
<td>MSE $\hat{q}$</td>
<td>0.061</td>
<td>0.022</td>
</tr>
<tr>
<td>$\Phi_5$ (7.389)</td>
<td>8.286</td>
<td>7.501</td>
</tr>
<tr>
<td>$\Phi_7$ (12.483)</td>
<td>13.509</td>
<td>12.386</td>
</tr>
<tr>
<td>$\Phi_9$ (26.617)</td>
<td>27.872</td>
<td>26.121</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>mm</th>
<th>em</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{q}$ (2)</td>
<td>2.077</td>
<td>2.001</td>
</tr>
<tr>
<td>MSE $\hat{p}$</td>
<td>0.908</td>
<td>0.977</td>
</tr>
<tr>
<td>MSE $\hat{q}$</td>
<td>0.047</td>
<td>0.021</td>
</tr>
<tr>
<td>$\Phi_5$ (7.389)</td>
<td>8.158</td>
<td>7.477</td>
</tr>
<tr>
<td>$\Phi_7$ (12.483)</td>
<td>13.488</td>
<td>12.541</td>
</tr>
<tr>
<td>$\Phi_9$ (26.617)</td>
<td>28.531</td>
<td>26.584</td>
</tr>
</tbody>
</table>

Data are given by 0:02 and 0:04 respectively. With the EM-procedure, even though the stopping time is at the 0.3th quantile of the lifetime and only about a third of the individual items fail until the stopping time, the mean square errors of the estimates of $\hat{p}$ and $\hat{q}$ are comparable with the Cramer-Rao lower bounds. As expected, the EM-procedure improves greatly the initial method-of-moments estimates since it makes use of not only the information provided by the initial estimates of the moments but also the important information contained in the predictive intervals.

5 Analysis of real degradation data

We illustrate the EM-procedure in this section by two real data sets: (i) the fatigue-crack-growth data which was originally introduced by Hudak et al. (1978), and (ii) the CG36A transistor degradation data presented in Zhuang (1994).

The analysis of the fatigue-crack-growth data. The fatigue-crack-growth data contains the degradation measurements on 21 test units. The crack
length of each unit was recorded once every 0.01 million cycles over a period of 0.12 million cycles. The initial crack length of every unit is the same as 0.9 inches. A unit is considered to have failed if the crack length exceeds 1.6 inches. The path model for the fatigue-crack-growth is derived from the Paris Law in material science and is given below:

\[ y_{ij} = \frac{1}{\mu_{2i}} \log(1 + 0.9^{\frac{\mu_{1i}}{\mu_{2i}^2}} \mu_{1i} \mu_{2i} t_{ij}) + \varepsilon_{ij}; \]

\[ i = 1; \ldots; 21; j = 1; \ldots; m_i. \]

The fatigue-crack-growth data has been analyzed by many authors using different methods. Lu and Meeker (1993) analyzed the data using their two-stage least squares method. Wu and Shao analyzed the data directly using ordinary and weighted least squares methods by treating the path model as a mixed effect regression model. The maximum likelihood method and a Bayesian method for the analysis of the data were explored by Robinson and Crowder (2000). Robinson and Crowder also made comparisons among Lu-Meeker’s two-stage method, the maximum likelihood method and the Bayesian method. All these methods provided similar conclusions in this application. The common feature of these methods is that an assumption on the distribution of the random parameters \( \mu_1 \) and \( \mu_2 \) was made.

Here, we re-analyze the fatigue-crack-growth data using our EM-procedure. First, the pseudo-lifetimes, the initial estimates of \( 1_T \) and \( 3_T \) are computed. The initial estimates are \( 1_T = 0.1242, 3_T = 0.00056 \). The log-likelihoods with parameters estimated by the method-of-moments are computed for the following distributions: Log-normal, Gamma and Weibull. Their values are 50.1182, 49.9117 and 48.4077 respectively. Since the log-likelihood of the
Log-normal distribution is the largest, the Log-normal distribution is taken as the assumed lifetime distribution. The adequacy of the Log-normal distribution is checked by the Normal probability plot of the log pseudo-lifetimes which is given in Figure 1. The EM-procedure is carried out with the initially estimated distribution Lognormal($\hat{\mu}(1); \hat{\sigma}^2(1)$) where $\hat{\mu}(1) = -2:1033$ and $\hat{\sigma}^2(1) = 0.03538$. In the EM-procedure, a 60% confidence level is used for each of the predictive intervals. The final estimates of the two parameters from the EM-procedure are $\hat{\mu} = -2:0990$ and $\hat{\sigma}^2 = 0.03258$. A point-wise $100(1 - \alpha)\%$ confidence interval for $F(t)$ is constructed as

$$[\hat{F} - z_{\alpha/2} \hat{\sigma}(2\hat{\mu} + 1); \hat{F} + z_{\alpha/2} \hat{\sigma}(2\hat{\mu} + 1)]$$
Figure 2: The EM-estimate, the MLE based on the observed lifetimes and the Lu-Meeker’s estimate with the point-wise 95% confidence limits associated with the first two estimates for the fatigue-crack-growth example.

In their Table 2, Lu and Meeker (1993) also provided the observed real lifetimes of the 21 test units. To facilitate comparisons, the EM-estimate, the maximum likelihood estimate using the full observed lifetime data and the assumed Lognormal distribution and the Lu-Meeker’s estimate together with the point-wise 90% confidence intervals associated with the first two estimates are plotted in Figure 2. The empirical distribution of the ob-

where \( \Phi(z) \) is the distribution function of the standard normal distribution and

\[
\Phi(z) = \frac{\log(t)}{\theta}.
\]

In their Table 2, Lu and Meeker (1993) also provided the observed real lifetimes of the 21 test units. To facilitate comparisons, the EM-estimate, the maximum likelihood estimate using the full observed lifetime data and the assumed Lognormal distribution and the Lu-Meeker's estimate together with the point-wise 90% confidence intervals associated with the first two estimates are plotted in Figure 2. The empirical distribution of the ob-

...
served lifetimes is imposed on the figure in the form of little circles. The Lu-Meeker's estimate is taken as the representative for the other methods mentioned previously. For computational reasons, the point-wise confidence intervals of Lu-Meeker's method are not reproduced. From these plots, we can make the following conclusions. The EM-estimate using the degradation data provides very similar inference to the MLE using the full lifetime data under the Lognormal assumption for the lifetime distribution. The Lu-Meeker's estimate seems closer to the empirical distribution of the observed lifetimes and has a heavier tail than both the EM-estimate and the MLE. Though there are slight differences among these estimates, essentially, all the estimates including those compared by Robinson and Crowder (2000) provide almost the same inference on the lifetime distribution. However, it is worth emphasizing that the EM-procedure has the following advantages in this application. The EM-estimate of the lifetime distribution has a closed form such that all the reliability information can be summarized in the two estimated parameters. For the other methods, since the inference on the lifetime distribution is induced from the inference on the distribution of the random parameters, simulations are needed to obtain the estimate of the lifetime distribution. The computation involved in the EM-procedure is negligible. But, the amount of computation with the other methods is a real burden. For example, as reported in Lu and Meeker (1993), to construct the point-wise confidence intervals at the failure times of the 21 test units using the two-stage least squares estimate, the computation takes about 144 hours using program written in S-Plus and run on a DEC 5000/200 workstation.
The analysis of the CG36A transistor data. The transistor data was presented in the Ph. D dissertation of Zhuang (1994). The original data contains the degradation measurements on a physical characteristic $Z$ of 100 units in a degradation reliability test. The characteristic $Z$ was measured for each test unit at time $t_1 = 0$, $t_2 = 1$, $t_3 = 2$, $t_4 = 10$, $t_5 = 30$, $t_6 = 100$, $t_7 = 250$, $t_8 = 500$ and $t_9 = 1000$ hours. The characteristic increases with time. The failure of a unit is determined when the increment of $Z$ reaches 30% of its value at time $t_1 = 0$. Hence, the degradation variable $Y$ is defined as

$$Y = \frac{Z_t - Z_0}{Z_0}$$

where $Z_0$ is the value of $Z$ at time $t_1 = 0$. The critical degradation level is then $\gamma_c = 0.3$.

In the original data, there are negative increments of $Z$ in 11 records, which is physically inexplicable. Therefore, we delete these 11 records in our analysis. For the remaining 89 records, the degradation variable $Y$ is plotted in Figure 3 (a). The following path model is assumed:

$$Y_{ij} = \mu_{ij} t_{ij}^{\gamma_{ij}} + \epsilon_{ij}$$

The residuals $\hat{\epsilon}_{ij}$ from fitting the path model for the 89 units are plotted in Figure 3 (b). The mean squared residuals equals 0.00021. The path model provides a very good description of the underlying degradation mechanism.

After computing the pseudo-lifetimes, it is found that one of retained units has an extremely large pseudo-lifetime which is far distanced from all the other units, see the boxplot in Figure 3 (c). We treat this unit as an outlier and exclude it from our analysis. The initial estimates of $\gamma_T$ and $\gamma_{\hat{\theta}}$ are
respectively $\gamma_T = 3015.67$ and $\gamma_T = 12551031$. Based on these estimates, the log-likelihoods are computed for the following distributions: Exponential, Weibull, Log-normal and Gamma. The values of the likelihoods are, in turn, 793019, 785473, 883844 and 783605. The log-likelihoods of the Gamma distribution and Weibull distribution are about the same and are significantly larger than the others. From the past field experiences with the same type of transistors, it is known that the Weibull distribution is appropriate to describe the lifetime distribution of the transistors. Therefore, though the log-likelihood of the Gamma distribution is slightly larger,
Figure 4: The EM-estimate and the point-wise 90% confidence intervals for the transistor data, the empirical distribution of the pseudo-lifetimes is imposed.

We take the Weibull distribution as the assumed distribution for the lifetime of the transistors. A Weibull probability plot for the pseudo-lifetimes is given in Figure 3 (d), which justifies the assumption of Weibull distribution. A 60% confidence level is used for the predictive intervals in the EM-procedure. The EM-estimates of the shape and scale parameters of the Weibull distribution are 0.7123 and 2356.72 respectively. The EM-estimate of $F(t)$ together with point-wise 90% confidence intervals are plotted in Figure 4. The confidence interval of $F(t)$ is constructed as

$$\left[ 1 - \exp\left( -\exp(\frac{L_1(t)}{\hat{\alpha}}) \right) ; 1 - \exp\left( -\exp(\frac{L_2(t)}{\hat{\alpha}}) \right) \right]$$
Table 2: Selected quantiles of the lifetime distribution of CG36A transistors

<table>
<thead>
<tr>
<th>Probability</th>
<th>Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>100</td>
</tr>
<tr>
<td>0.2</td>
<td>287</td>
</tr>
<tr>
<td>0.3</td>
<td>554</td>
</tr>
<tr>
<td>0.4</td>
<td>918</td>
</tr>
<tr>
<td>0.5</td>
<td>1409</td>
</tr>
<tr>
<td>0.6</td>
<td>2085</td>
</tr>
<tr>
<td>0.7</td>
<td>3058</td>
</tr>
<tr>
<td>0.8</td>
<td>4597</td>
</tr>
<tr>
<td>0.9</td>
<td>7600</td>
</tr>
</tbody>
</table>

where $L_1(t)$ and $L_2(t)$ are functions of $t$ such that

$$Pr^\ln(mA^2(l)=l) \cdot L_1(t) = \frac{1+\hat{\theta}}{2}; Pr^\ln(mA^2(l)=l) \cdot L_2(t) = \hat{\theta}:$$

Here $A^2(l)$ is a $A^2$ random variable with degrees of freedom $l$, and $m$ and $l$ are determined such that $\ln(mA^2(l)=l)$ has the same mean and variance as $W(t)$ given by

$$W(t) = \frac{\hat{\theta}^\frac{u}{b} i t^\hat{b}}{b};$$

where $u = \ln \hat{\theta}$, $b = 1=\hat{\tau}$, $\hat{\theta}$ and $\hat{\tau}$ being the scale parameter and shape parameter of the Weibull distribution respectively, $\hat{\theta}$ and $\hat{\tau}$ are the MLEs of $u$ and $b$ respectively. For more details about the confidence intervals for Weibull distribution, see Engelhardt and Bain (1977).

From the final estimated lifetime distribution, the mean and standard deviation of the lifetime of the CG36A transistors are 2935 and 4207 hours respectively. Some selected quantiles of the lifetime distribution are given in Table 2. It is also worth noting that the stopping time $t = 1000$ of the degradation test is the 0.42th quantile of the estimated lifetime distribution which is quite in line with the fact that, among the 88 test units considered,
39% of them failed before the stopping time.

REFERENCES


