EM algorithm for degradation data analysis

XU An-cha, TANG Yin-cai
(School of Finance and Statistics, East China Normal University, Shanghai 200241, China)

Abstract: Statistical inference procedures were proposed in assessing product reliability based on EM algorithm. The general formulae to estimate parameters were derived in the degradation model when its random effects follow an exponential family, and sensitivity of the choice of the distributions of the random effects was assessed via simulation. Two real data sets were analyzed.

Key words: degradation data; random effect; EM algorithm

CLC number: O213.2 Document code: A

Introduction

In some reliability studies, the main interest is to obtain the lifetime of products. Traditionally, this is addressed by estimating failure time distributions from failure data. With today’s high technology, however, many products are designed to work without failure for years or decades. Thus, some life tests, even accelerated life tests, result in few or no failures in a short life testing time. In such cases, it is difficult to access those high-reliability products with traditional statistical methods. Recently more interests and efforts turn to assess product reliability using degradation within a short period of time. Degradation data analysis has been developed in recent years. A general model for degradation data with a two-stage method of least squares has been suggested by Lu and Meeker[4] and illustrated using fatigue crack growth data. Meeker...
and Escobar[2] considered a maximum likelihood procedure and bootstraps method to obtain confidence interval of quantiles and reliability degree. Wu and Shao[3] proposed ordinary and weighted least squares procedures and established asymptotic properties of these estimators. Robinson and Crowder[4] proposed a full Bayesian method. More references on degradation analysis see also Yacout et al.[5], Lu et al.[6], Shiau and Lin[7], Chiao and Hamada[8], Chen and Zheng[9].

In this paper, we propose an alternative inference procedure: EM algorithm, in which the random parameters are treated as the latent variables. We augment the observed data with latent data so that the complete likelihood function is much simpler, so that estimating the parameters becomes plausible. This approach has the following advantages:

(1) for a relatively simple degradation path, explicit expression in the M step can be obtained;
(2) the approach can be more easily implemented than the usual maximum likelihood procedure because of the simpler complete likelihood, and assessing sensitivity of the choice of the distributions of the random parameters becomes convenient.

The paper unfolds as follows. The general model for repeated degradation measures is introduced in Section 1. In Section 2 the EM algorithm procedure is described. Two real data sets are analyzed and simulation studies are provided to demonstrate the efficiency of the procedure and to assess the sensitivity of the choice of the distribution of the random parameters in Section 3.

1 Degradation models

Let \( y \) denote an appropriate measure of the failure-causing of the product under consideration. The \( j \)th degradation measure to be made on the \( i \)th unit, or suitable transformation of it, is represented as

\[
y_{ij} = \eta(\alpha, \beta_i, t_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i,
\]

where \( \alpha \) denotes the vector of fixed-effect parameters; \( \beta_i \)s are i.i.d. vectors of random effects; \( \eta(\cdot) \) represents the actual level of degradation of the product under consideration and is a given continuous function in \( (\alpha, \beta_i) \); \( t_{ij} \) are the measurement time points and \( \epsilon_{ij} \) are the usual error terms. The errors \( \epsilon_{ij} \) are often assumed to be i.i.d and follow normal distribution \( N(0, \sigma^2) \) in which \( \sigma \) can be allowed to vary with time or degradation level, see Wakefield et al.[10] and Lu et al.[11]. In situations where the errors are not independent, one can embed a structured covariance model along with appropriate estimation methods. In this article, the errors are assumed to be independent.

For a degradation path model, it can be obtained based on physical or chemical mechanisms in some practical cases. See Lu et al.[11] and Meeker and Lurville[12]. Otherwise, because degradation is not reversible, suppose \( \eta \) as a given function of \( t \), is monotone. The component degrades over time and fails when \( \eta \) reaches a certain known or specified critical value \( \eta_c \). Therefore, theoretically, the failure time of unit \( i \), denoted by \( T_i \), is the solution of \( \eta(\alpha, \beta_i, t) = \eta_c \).

Because \( \eta \) is monotone in \( t \), the solution is unique and finite solution for any given \((\alpha, \beta_i)\).
Without loss of generality, we assume that eta is increasing. Besides, there is an important relation between the failure time distribution and the degradation distribution as follows.

\[ F_T(t; \theta) = \Pr\{T \leq t\} = \Pr\{\eta(\alpha, \beta, t) \geq \eta_c\} = 1 - F_q(t; \theta), \quad (1.2) \]

where \( \theta \) is a vector of parameters of the failure time distribution. In some simple cases, it is possible to get a closed-form expression for \( F_T(t; \theta) \). In general, however, such a closed-form will not exist, especially when \( \eta(\cdot) \) is nonlinear and there are more than one random parameter. Then it is necessary to evaluate \( F_T(t; \theta) \) with numerical methods.

### 2 EM algorithm procedure

We first give some assumptions and notations used in this section. Assume that the parameter vector \( \beta_i \) follows a \( k \)-dimensional exponential family, that is

\[ \beta_i \sim W(\lambda) \exp \left\{ \sum_{j=1}^{k} \lambda_j T(\beta_i) \right\} \cdot h(\beta_i), \quad i = 1, 2, \ldots, n, \quad (2.1) \]

where \( \lambda = (\lambda_1, \cdots, \lambda_k) \). Let \( y_i = (y_{i1}, \cdots, y_{im_i}) \), \( \phi = (\alpha, \sigma^2, \lambda) \), \( \theta = (\alpha, \sigma^2, \beta = (\beta_1, \cdots, \beta_n) \), \( y = (y_1, \cdots, y_n) \). As \( \epsilon_{ij} \) are normally distributed with mean 0 and variance \( \sigma^2 \), we obtain the joint distribution of \( (y_i, \beta_i) \):

\[ f(y_i, \beta_i | \phi) = f(y_i | \beta_i, \phi) f(\beta_i | \phi) = f(y_i | \beta_i, \theta) f(\beta_i | \lambda), \quad (2.2) \]

where \( f(y_i | \beta_i, \theta) \) is the density of multivariate normal distribution and \( f(\beta_i | \lambda) \) has the form of (2.1). The EM algorithm (Dempster et al.\cite{13}) is an iterative procedure where each iteration consists of an E (expectation) step and an M (maximization) step. The principle of the EM algorithm can be described as follows. The \((t+1)\)th E step computes the conditional expectation of the log-likelihood of the complete-data \((y, \beta)\) with respect to the distribution of the missing data \( \beta \) given the observed data \( y \) and the current estimated parameter value \( \phi^{(t)} \):

\[ Q(\phi | \phi^{(t)}) = \mathbb{E}[\log f(y, \beta)] | y, \phi^{(t)}]. \quad (2.3) \]

The \((t + 1)\)th M step then finds \( \phi^{(t+1)} \) as the maximization point of \( Q(\phi | \phi^{(t)}) \), so we have the inequality \( Q(\phi^{(t+1)} | \phi^{(t)}) \geq Q(\phi | \phi^{(t)}) \) for all \( \phi \) in the parameter space. This two-step process is repeated until convergence achieved.

#### 2.1 The E step

As described above, the E step computes \( Q(\phi | \phi^{(t)}) \). In the degradation model, it can be computed as follows.

\[ Q(\phi | \phi^{(t)}) = \mathbb{E}[\log f(y, \beta)] | y, \phi^{(t)}] = \sum_{i=1}^{n} \int \log f(y_i, \beta_i | \phi) f(\beta_i | y_i, \phi^{(t)}) d\beta_i 
= \sum_{i=1}^{n} \int \log f(y_i | \beta_i, \phi) f(\beta_i | y_i, \phi^{(t)}) d\beta_i + \sum_{i=1}^{n} \int \log f(\beta_i | \lambda) f(\beta_i | y_i, \phi^{(t)}) d\beta_i. \quad (2.4) \]
2.2 The M step

In principle, the M step is carried out by solving the score equations
\[
\frac{\partial}{\partial \phi} Q(\phi|\phi^{(t)}) = \frac{\partial}{\partial \phi} \mathbb{E}[\log \{f(y, \beta)\}|y, \phi^{(t)}] = 0.
\] (2.5)

From (2.4) and (2.5) can be put into three equations:
\[
\sum_{i=1}^{n} \frac{\partial}{\partial \lambda} \mathbb{E}[\log \{f(y_i|\beta)\}|y, \phi^{(t)}] = 0,
\]
\[
\sum_{i=1}^{n} \frac{\partial}{\partial \sigma_i^2} \mathbb{E}[\log \{f(y_i|\beta, \phi)\}|y, \phi^{(t)}] = 0,
\]
\[
\sum_{i=1}^{n} \frac{\partial}{\partial \alpha} \mathbb{E}[\log \{f(y_i|\beta, \phi)\}|y, \phi^{(t)}] = 0.
\]
(2.6) (2.7) (2.8)

Using (2.1) and allowing differentiation under the integral sign, (2.6) can be simplified as
\[
\sum_{i=1}^{n} \mathbb{E} \left[ \frac{1}{W(\lambda)} \frac{\partial W(\lambda)}{\partial \lambda} + T(\beta_i)|y, \phi^{(t)} \right] = 0
\]
\[
\Rightarrow \frac{1}{W(\lambda)} \frac{\partial W(\lambda)}{\partial \lambda} = -\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[T(\beta_i)|y, \phi^{(t)}],
\] (2.9)

where \(T(\beta_i) = (T_1(\beta_i), \ldots, T_n(\beta_i))\). For an explicit form of \(W(\lambda), \lambda^{(t+1)}\) can be obtained from (2.9). In a same way, an explicit solution to (2.7) is given by
\[
\sigma_i^{2(t+1)} = \frac{1}{k} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \mathbb{E}[(y_{ij} - \eta(\alpha, \beta_i, t_{ij}))^2|y, \phi^{(t)}],
\]
(2.10)

where \(k = \sum_{i=1}^{n} m_i\). Note that (2.8) cannot be solved in closed form unless \(\eta(\cdot)\) is polynomial of \(\alpha\). We consider using iteration of Newton’s method to update \(\alpha^{(t)}\):
\[
\alpha^{(t+1)} = \alpha^{(t)} - [H(\beta_i, \alpha^{(t)}) + cI]^{-1}S(\beta_i, \alpha^{(t)}),
\] (2.11)
where \(S(\cdot) = \sum_{i=1}^{n} \mathbb{E} \left[ \frac{\partial \log \{f(y_i|\beta, \phi)\}}{\partial \alpha} \right]|y, \phi^{(t)}\), \(H(\cdot) = \frac{\partial S(\beta_i, \alpha^{(t)})}{\partial \alpha}\), and \(c \geq 0\) is some small constant chosen to ensure that \(H(\cdot)\) is invertible. When \(H(\cdot)\) is positive definite, \(c = 0\); otherwise, \(c > 0\) is slightly larger than the magnitude of the most negative eigenvalue of \(H(\cdot)\).

2.3 Monte Carlo EM algorithm

In the case of non-exponential family of random effects, the expectation usually cannot give rise to a closed form in the E step. An alternative is to replace the E step with its Monte Carlo (MC) approximations constructed using a sample, say \(\beta^{(1)}, \ldots, \beta^{(M)}\), from \(f(\beta|y, \phi)\), then
\[
\mathbb{E}[u(\beta)|y, \phi] \approx \frac{1}{M} \sum_{m=1}^{M} u(y_i, \beta^{(m)}),
\]
where \(u(\cdot)\) is an any function. To obtain a satisfactory accuracy, the MC sample size \(M\) needs to be large enough. For example, to obtain two decimal digits of accuracy, \(M \geq 10\,000\) is required.
Besides, directly sampling from \( f(\beta | y_i, \phi) \) is often difficult or impossible. Sampling methods such as importance sampling or Markov chain Monte Carlo (MCMC) techniques have been used to alleviate such difficulty. For more effective sampling methods, such as ARS and ARMS, see Gilks and Wild\(^{[14]}\), Gilks et al.\(^{[15]}\).

In summary, our MCEM algorithm consists of five steps as follows.

1. Give initial values of \( M \) and \( \phi \).
2. Generate \( \beta^{(1)}_i, \ldots, \beta^{(M)}_i \) from \( f(\beta_i | y_i, \phi) \) for \( i = 1, \ldots, n \).
3. EM step: compute \( \phi^{(t+1)} \) according to (2.9), (2.10) and (2.11) by replacing the conditional expectations with their corresponding MC approximations.
4. Assess MC error: at the \( (t+1) \)th EM iteration, obtain the confidence region of \( \phi^{(t+1)} \).
   - If \( \phi^{(t)} \) lies in this confidence region, then
     a. Set \( M \rightarrow M + \lfloor M/a \rfloor \) for some \( a > 0 \). (see Booth and Hobert,\(^{[16]}\)).
     b. Obtain \( \beta^{(1)}_i, \ldots, \beta^{(M)}_i \) from \( f(\beta_i | y_i, \phi) \) for \( i = 1, \ldots, n \).
5. Repeat steps (3) and (4) until convergence achieved.

### 2.4 Estimation of standard errors

Let \( \hat{\phi} \) be the estimates of \( \phi \) at convergence. Using Louis’s method (Louis\(^{[17]}\)), the observed information matrix of \( \hat{\phi} \) can be estimated by

\[
I(\hat{\phi}) = \left( \sum_{i=1}^{n} \frac{\partial^2 \log f(y_i | \phi)}{\partial \phi \partial \phi^T} \right)_{\hat{\phi}},
\]

where

\[
\frac{\partial^2 \log f(y_i | \phi)}{\partial \phi \partial \phi^T} \bigg|_{\hat{\phi}} = E \left[ \frac{\partial^2 \log f(y_i | \beta_i, \phi)}{\partial \phi \partial \phi^T} | y_i, \hat{\phi} \right] + \text{Var} \left[ \frac{\partial \log f(y_i | \beta_i, \phi)}{\partial \phi} \right] y_i, \hat{\phi}.
\]

The asymptotic variance-covariance matrix of \( \hat{\phi} \) is then given by the inverse of the estimated \( I(\hat{\phi}) \). For the MCEM algorithm, we can replace the conditional expectations with their corresponding Monte Carlo approximations to obtain asymptotic variance-covariance matrix of \( \hat{\phi} \).

Besides, The log-likelihood function estimated by the EM algorithm and the MCEM algorithm are approximated, respectively by

\[
l_{\text{EM}}(\hat{\phi}) = \sum_{i=1}^{n} \log \left[ E \left( f(y_i | \beta_i, \hat{\phi}) \right) \right],
\]

\[
l_{\text{MCEM}}(\hat{\phi}) = \sum_{i=1}^{n} \log \left[ \frac{1}{M} \sum_{m=1}^{M} f(y_i | \beta^{(m)}_i, \hat{\phi}) \right].
\]

### 2.5 Asymptotic confidence intervals

We now return to the degradation data statistical analysis, i.e., using degradation data to make inference about reliability degree of the product and percentiles of the failure time distribution. In the case of lifetime distributions having an explicit expression, say \( F_T(t, \phi) \), by substituting \( \hat{\phi} \) for \( \phi \), we obtain an estimate \( \hat{F}_T(t) = F_T(t, \hat{\phi}) \), and an approximating \( (1 - \alpha)\% \) confidence interval for \( F_T(t, \phi) \) can be constructed as follows. Under some regularity conditions

\[
\]
where \( \hat{\phi} - \phi \sim N(0, I^{-1}(\phi)) \), we use Delta Theorem

\[
\hat{F}_T(t, \hat{\phi}) - F_T(t, \phi) \sim N\left(0, \frac{\partial F}{\partial \phi} I^{-1}(\phi) \frac{\partial F}{\partial \phi} \right),
\]

where \( \frac{\partial F}{\partial \phi} \) denotes the first derivative of \( F_T(t, \phi) \) with respect to \( \phi \). Let

\[
Z_F = \frac{\hat{F}_T(t, \hat{\phi}) - F_T(t, \phi)}{\sqrt{\frac{\partial F}{\partial \phi} I^{-1}(\hat{\phi}) \frac{\partial F}{\partial \phi}}},
\]

where \( \frac{\partial F}{\partial \phi} \) denotes the value of \( \frac{\partial F}{\partial \phi} \) evaluated at \( \hat{\phi} \). Then an approximating \((1 - \alpha)\)% confidence interval for \( F_T(t, \phi) \) can be obtained by solving \(|Z_F| \leq z_{\alpha/2} \), where \( z_{\alpha/2} \) is the \((1 - \alpha/2)\)th quantile of the standard normal distribution.

Using the same method, we can construct a confidence interval for the \( p \)th quantile \( t_p \), which is the solution of \( F_T(t, \phi) = p \), say \( t_p = g(\phi) \). Hence, an approximate \((1 - \alpha)\)% confidence interval for \( t_p \) is

\[
\left( \hat{t}_p - z_{\alpha/2} \tilde{V}^{1/2}, \hat{t}_p + z_{\alpha/2} \tilde{V}^{1/2} \right),
\]

where \( \hat{t}_p = g(\hat{\phi}) \) and \( \tilde{V} = \left[ \frac{\partial g(\phi)}{\partial \phi} \right]^T I^{-1}(\hat{\phi}) \left[ \frac{\partial g(\phi)}{\partial \phi} \right] \).

However, as we mentioned in Section 1, the lifetime distribution do not have an explicit expression in some times. We can then use bootstrap method (see Meeker and Escobar\cite{2}) to obtain the confidence intervals of \( F_T(t, \phi) \) and \( t_p \).

### 3 Examples and simulations

We apply the proposed methods to two degradation data sets. For the first data set, we also carry out simulations to examine the performance of the proposed methods and the sensitivity of the choice of the distributions of random effects.

#### 3.1 Data analysis for CG36A transistor example

The CG36A transistor degradation data are presented in [18]. There are 100 Transistors in this experiment and each has repeated measurements of a physical characteristic \( C \) taken at nine different times: 0, 1, 2, 10, 30, 100, 250, 500, and 1 000 hours. The characteristic \( C \) increases with time. The failure of a unit occurs when the increment of \( C \) reaches 30% of its value at initial time, that is the critical degradation level \( \eta_c = 0.3 \). The degradation variable \( Y \) is the ratio \( (C - C_0)/C_0 \), where \( C_0 \) is the value of \( C \) at time \( t = 0 \).

After transformation of \( C \), we find that there are negative increment of \( C \) in 11 records, which is impossible in physics. This may due to mistakes when the data were collected. We delete these 11 observations from the data set with 89 left, which are plotted in Fig. 1.

We consider the following model:

\[
Y_{ij} = \beta_i t_j^\alpha + \epsilon_{ij}, i = 1, \cdots, 89; j = 1, \cdots, 9,
\]

where \( \alpha \) is a fixed effect parameter, and \( \epsilon_{ij} \sim N(0, \sigma^2) \). The random effects \( \beta_i \)'s are independently distributed as an exponential distribution with mean \( 1/\lambda \).
Following Section 2, the estimates of $\lambda$ and $\alpha$ are 0.7613 and 598.37, respectively, and their corresponding estimated asymptotic variance-covariance matrix is

$$\hat{I}^{-1} = \begin{pmatrix} 0.0042 & 0.1335 \\ 0.1335 & 28.863 \end{pmatrix}. \quad \text{The estimator of the standard deviation of measurement error is } \hat{\sigma}_e = 0.0036. \text{ Point estimates and 90% confidence intervals for some percentiles of the failure time distribution are listed in Tab. 1. Fig. 2 is the plot of the estimates and confidence intervals, connected by straight lines.}

<table>
<thead>
<tr>
<th>p</th>
<th>$t_p$</th>
<th>lower limits</th>
<th>upper limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>216.24</td>
<td>206.56</td>
<td>225.92</td>
</tr>
<tr>
<td>0.10</td>
<td>305.53</td>
<td>291.01</td>
<td>320.04</td>
</tr>
<tr>
<td>0.20</td>
<td>489.06</td>
<td>462.44</td>
<td>515.69</td>
</tr>
<tr>
<td>0.30</td>
<td>716.06</td>
<td>671.72</td>
<td>760.40</td>
</tr>
<tr>
<td>0.40</td>
<td>1024.98</td>
<td>953.29</td>
<td>1096.66</td>
</tr>
<tr>
<td>0.50</td>
<td>1478.86</td>
<td>1362.07</td>
<td>1595.64</td>
</tr>
<tr>
<td>0.60</td>
<td>2208.21</td>
<td>2010.94</td>
<td>2405.47</td>
</tr>
<tr>
<td>0.70</td>
<td>3539.59</td>
<td>3178.21</td>
<td>3900.97</td>
</tr>
<tr>
<td>0.80</td>
<td>6554.01</td>
<td>5771.91</td>
<td>7336.09</td>
</tr>
<tr>
<td>0.90</td>
<td>17563.01</td>
<td>14968.62</td>
<td>20157.39</td>
</tr>
<tr>
<td>0.95</td>
<td>45210.06</td>
<td>37275.89</td>
<td>53144.22</td>
</tr>
</tbody>
</table>

3.2 Simulation

Here we demonstrate the efficiency of the EM algorithm, and examine the performance of estimation of $p$th quantile under misspecification of the random effect distribution.

We simulated 50 sample paths according to model (3.1). The true parameters ($\alpha, \sigma_e$) are equal to $(0.7613, 0.0036)$. The true distribution of random effects $\beta_i$ is assumed to be ExpM(598.37), Weibull(2,1/400) or Lognormal(–6.2,0.9). We denote the corresponding models as ExpM($\lambda$), WeiM($b, m$) and LNM($\mu, \sigma^2$), respectively. The critical degradation level $\eta_c = 0.3$. If the true model is ExpM($\lambda$), then the degradation data were simulated by generating a random exponential variable with mean $1/598.37$ and calculating the amount of degradation, at specified times, using this variable. So were WeiM($b, m$) and LNM($\mu, \sigma^2$) model.
We also present the model misspecification effect of the random effects distribution. For example, Under the true $\text{ExpM}(\lambda)$ model, to examine the performance of the EM algorithm estimate of $p$th quantile for misspecification of the random effect distribution, we consider the so-called ‘working’ $\text{WeiM}(b, m)$ and $\text{LNM}(\mu, \sigma^2)$ model.

All the results are calculated over 1 000 repetitions and presented in Tab. 2, where SSMSE denotes simulation square root of mean square error. In Tab. 2, we compute $t_p$ for $p = 0.1$, 0.25, 0.5, 0.75 and 0.9. T-M denotes true model. $t_p^{\text{ExpM}}$, $t_p^{\text{WeiM}}$ and $t_p^{\text{LNM}}$ denote the estimation of $t_p$ obtained by using ExpM, WeiM and LNM model, respectively.
Some features from Tab. 2 are listed below.

(1) The estimation of EM algorithm works well based on the true degradation model. For example, under ExpM model with a true value \( t_{0.9} = 17,563 \), the SM is 17,072 with SSMSE equal to 511.99. Similar results hold for other \( t_p \)s and other models.

(2) The misspecification effect is significant, which means that the choice of the random effect distribution is important for the precision of the estimation of \( p \)th quantile. However, if the true model is nested in the model we choose, the misspecification effect can be neglected. For example, under a true ExpM model, we choose WeiM model to obtain the estimation of \( p \)th quantile, and the bias is small. Besides, when the true model and the chose model are in close distance (Kullback-Leibler distance), like Weibull and Lognormal distribution, the misspecification effect is not so significant. Therefore, we use the log-likelihood function (2.14) to choose the random effect’s distribution. The distribution with the largest log-likelihood among all the candidate distributions is then chosen as the random effect distribution.

3.3 Data analysis for the metal fatigue example

Here we will discuss the example of fatigue-crack-growth data from Lu and Meeker[1], which is presented again in Fig. 3.

There are 21 test units in this experiment, each with the same initial crack length of 0.9 inches. We say a failure occurs when the crack length is beyond 1.6 inches as Lu and Meeker[1] did. The test continues until 0.12 million cycles, the termination time of the test. There are 13 measurement times in this test and only 13 units undergo the complete test and provide failure times. The path model for the fatigue-crack-growth is derived from the Paris Law in material science, and is given by

\[
y_{ij} = -\frac{1}{\theta_{2i}} \log \left( 1 - 0.9^{\theta_{2i} \theta_{1i} t_j} \right) + \epsilon_{ij},
\]

\( i = 1, \ldots, 21, \ j = 1, \ldots, m_i, \)

where \( y_{ij} = \log(\text{observed crack length at time } t_j/0.9) \), \( t_j \) is the measurement time (in million cycles), and the coefficients \((\theta_{1i}, \theta_{2i})\) are random effects. We assume that \((\theta_{1i}, \theta_{2i})\) have a
bivariate normal distribution \( N(\mu, \Sigma) \), where \( \mu \) and \( \Sigma \) are \( 2 \times 1 \) vector and \( 2 \times 2 \) matrix, respectively. Here we use MCEM algorithm to analyze this example. Our MCEM algorithm converges in 58 iterations in this example. The value of \( M \) increases from \( M = 50 \) at the start to \( M = 38,148 \) at the final iteration. Fig. 4 presents the history of the increase in \( M \) after each iteration of the MCEM algorithm. The estimations of \( \mu \) and \( \Sigma \) are \( \hat{\mu} = \begin{pmatrix} 3.684 \\ 1.572 \end{pmatrix} \) and \( \hat{\Sigma} = \begin{pmatrix} 0.523 & -0.117 \\ -0.117 & 0.079 \end{pmatrix} \). The estimation of standard deviation of the error term is \( \hat{\sigma}_e = 0.0061 \). These estimates approach the MLEs: \( \tilde{\mu} = \begin{pmatrix} 3.73 \\ 1.58 \end{pmatrix} \), \( \tilde{\Sigma} = \begin{pmatrix} 0.527 & -0.111 \\ -0.111 & 0.078 \end{pmatrix} \) and \( \tilde{\sigma}_e = 0.006 \), which were obtained by Robinson et al.\cite{4}. Point estimates and point-wise 90% confidence intervals for the percentiles of the failure time distribution are given in Fig. 5.

Fig. 4 The Monte Carlo sample size against the EM iteration number
Fig. 5 Point estimates and point-wise 90% confidence interval for the Metal Fatigue example

4 Conclusion

We have proposed two versions of the EM algorithm for maximum likelihood estimation in the nonlinear mixed effects model. These EM algorithms are simple and easy to implement. In the numerical examples, we find the estimates of lifetime is sensitive to the distributions of random effects and suggest using the log-likelihood function to choose a distribution for random effect. Analysis of the two real data sets shows that Em algorithm is comparable to the current maximum likelihood procedure.

[ References ]


It is easy to verify that $R(X) \subseteq R(A)$, $R(Y) \subseteq R(A^*)$ and $XY^*A^+XY^* = XY^*$. So by Corollary 2.3, 

$$
(A + XY^*)^+ = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2}
\end{pmatrix}, \quad (XY^*A^*)^+ = \begin{pmatrix}
0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & 0 & 0
\end{pmatrix}.
$$

[References]