Maximum Likelihood Estimation
For Multi-Risk Model

R. J. Herman and Rusi K. N. Patell
IBM Corporation, White Plains, New York

In this paper we discuss a probabilistic model of failure due to competing causes of failure operating on a unit during its lifetime. Following Cox we term this the Multi-Risk Model. We impose some well-known parametric forms on the underlying distributions of the model and obtain maximum likelihood estimates for the parameters of the distributions. The asymptotic variance-covariance matrix of the estimates is derived. We give examples using the exponential and Weibull distributions. The theory is applied to data published by Mendenhall and Hader, who studied a Single-Risk Model.

1. Introduction

Suppose a unit is exposed to several potential causes of failure during its installation lifetime. Let there be a finite number of independent causes of failure, labelled 1, 2, \ldots, J. We associate with each cause j a non-negative random variable \( X^{(j)} \) with a c.d.f. \( F^{(j)}(x) \). \( F^{(j)}(x) \) is called the underlying distribution of \( X^{(j)} \) since \( X^{(j)} \) is considered as the failure time that would be observed if all causes of failure except the jth were inoperative. The actual failure time observed will be the smallest of the \( X^{(j)} \)’s and will have as survivor function the product of the component survivor functions of the \( X^{(j)} \)’s. This is the fundamental relation in the Minimum Process. Following Cox [6], [7] we may refer to the Minimum Process discussed here as the Multi-Risk Model. Justification stems from the fact that an installed item during its installation lifetime is exposed to multiple causes of failure as against another model called the Single-Risk where each item is predestined for a specific cause of failure only. That is, there exists an unknown mixture of various types in the cohort. See Mendenhall and Hader [9] for a fuller discussion.

It is of interest to impose some well-known parametric forms on the underlying distributions \( F^{(j)}(x) \) and develop maximum likelihood estimates (m.l.e.s.) for the parameters of the distributions. The m.l.e.s. obtained are in a form analogous to one that would be obtained if for jth cause of failure we consider the problem as that of obtaining m.l.e.s. for progressively censored samples, defined by Cohen [3], [4], [5] from \( F^{(j)}(x) \). The censored units are those that fail due to any other cause except the jth one and those surviving beyond censoring point. This result is obvious if we compare the likelihood functions for the Multi-Risk Model with that for progressively censored samples.

Asymptotic variance-covariance matrices for the various cases are obtained.
readily. Hence we can make a joint confidence statement on the parameters of interest. Examples and an application are presented using the exponential and Weibull distributions.

2. Multi-Risk Model

A unit is exposed to several potential causes of failure during its installation lifetime. Let there be a finite number of independent causes of failure, labelled 1, 2, $\cdots$, $J$. We associate with cause $j$ a non-negative random variable $X^{(j)}$ with a continuous c.d.f. $F^{(j)}(x)$, $j = 1, \cdots, J$. Then the observed failure time is given by the random variable.

$$Z = \min (X^{(1)}, \cdots, X^{(J)}) \quad (2.1)$$

Let $N$ be a random index $j$ for which $Z = X^{(j)}$ then, we define

$$G^{(j)}(x) = \Pr \{\text{failure is due to cause } j \text{ and occurs on or before } x\}$$

The corresponding p.d.f. is denoted by $g^{(j)}(x)$. Furthermore, let

$$F(x) = \Pr \{\text{failure occurs on or before time } x\}$$

and the corresponding p.d.f. is $f(x)$. The survivor function of the r.v. $Z$ is defined as

$$S(x) = \Pr \{Z > x\}$$

$$= \prod_{i=1}^{J} S^{(i)}(x), \quad S^{(i)}(x) = 1 - F^{(i)}(x) \quad (2.2)$$

On using the relation

$$f(x) = \frac{d}{dx} F(x) = -\frac{d}{dx} S(x)$$

we get

$$f(x) = \sum_{i} f^{(i)}(x) \prod_{k \neq i} S^{(k)}(x)$$

$$= \sum_{i} g^{(i)}(x). \quad (2.3)$$

Finally, we can write, since $F(x) = 1 - S(x)$

$$\prod_{i=1}^{J} S^{(i)}(x) = 1 - \sum_{i} G^{(i)}(x). \quad (2.4)$$

Now the age specific failure rate (hazard rate) is given by

$$\phi^{(i)}(x) = -\frac{d}{dx} \ln S^{(i)}(x)$$

$$= g^{(i)}(x) \left/ S(x) \right. \quad (2.5)$$

Hence, we have

$$f^{(i)}(x) = \phi^{(i)}(x) S^{(i)}(x);$$
that is
\[ f^{(i)}(x) = \frac{g^{(i)}(x)}{S(x)} \exp \left[ -\int_0^x \frac{g^{(i)}(x)}{S(x)} \, dx \right] \]
or
\[ F^{(i)}(x) = 1 - \exp \left[ -\int_0^x \frac{g^{(i)}(x)}{1 - \sum_i G^{(i)}(x)} \, dx \right]. \tag{2.6} \]

Thus the set of functions \( \{G^{(i)}(x)\} \) is related to the set \( \{F^{(i)}(x)\} \) by functional equations
\[ G^{(i)}(x) = \int_0^x f^{(i)}(s) \prod_{k \neq i} S^{(k)}(s) \, ds. \tag{2.7} \]
The solution of this set of equations is
\[ F^{(i)}(x) = 1 - \exp \left[ -\int_0^x \frac{g^{(i)}(x)}{1 - \sum G^{(i)}(x)} \, dx \right] \]
which is a bona-fide c.d.f. provided the above mentioned integral diverges. This requirement appears to have been overlooked in the literature. See [B], [7]. Berman [2] gives an elegant derivation of the above mentioned result.

3. Maximum Likelihood Estimation

Let us develop the maximum likelihood estimates of the parameters of the underlying distributions \( F^{(i)}(x) \). It will be shown that the estimation problem can be solved and interpreted in a simple fashion.

Without loss of generality let us consider the case of two causes of failure. Let the r.v.s. \( X^{(1)} \) and \( X^{(2)} \) have the c.d.f. \( F^{(1)}(x, \theta^{(1)}) \) and \( F^{(2)}(x, \theta^{(2)}) \) with corresponding p.d.f. \( f^{(1)}(x, \theta^{(1)}) \) and \( f^{(2)}(x, \theta^{(2)}) \) respectively, where \( \theta^{(1)} = (\theta_1^{(1)}, \ldots, \theta_k^{(1)}) \) and \( \theta^{(2)} = (\theta_1^{(2)}, \ldots, \theta_k^{(2)}) \) are corresponding vectors of parameters with no common elements. Distributions \( F^{(1)}(x, \theta^{(1)}) \) and \( F^{(2)}(x, \theta^{(2)}) \) need not belong to the same family.

Suppose \( N \) units are put on life test and the experiment is terminated at time \( x_T \). By fixed time \( x_T \), we observe \( n_1 \) failures at instants \( x_{1,1}^{(1)}, \ldots, x_{n_1}^{(1)} \) due to first cause, \( n_2 \) failures at instants \( x_{1,2}^{(2)}, \ldots, x_{n_2}^{(2)} \) due to second cause and \( N - (n_1 + n_2) \) units are still surviving. The likelihood function of the sample of Type I censoring is
\[ L = C \prod_{i=1}^{n_1} \left[ f^{(1)}(x_{i,1}^{(1)}, \theta^{(1)}) S^{(2)}(x_{i,1}^{(1)}, \theta^{(2)}) \right] \]
\[ \cdot \prod_{j=1}^{n_2} \left[ f^{(2)}(x_{j,2}^{(2)}, \theta^{(2)}) S^{(1)}(x_{j,2}^{(2)}, \theta^{(1)}) \right] \left[ S^{(1)}(x_T, \theta^{(1)}) S^{(2)}(x_T, \theta^{(2)}) \right]^{N-(n_1+n_2)} \] \tag{3.1}
where \( C \) is a constant, and \( S^{(1)}(x, \theta^{(1)}) = 1 - F^{(1)}(x, \theta^{(1)}) \). Let
\[ L^{(1)} = \prod_{i=1}^{n_1} f^{(1)}(x_{i,1}^{(1)}, \theta^{(1)}) \prod_{j=1}^{n_2} S^{(1)}(x_{j,2}^{(2)}, \theta^{(1)}) \left[ S^{(2)}(x_T, \theta^{(2)}) \right]^{N-(n_1+n_2)} \] \tag{3.2}
and

\[ L^{(2)} = \prod_{i=1}^{n} f^{(2)}(x^{(2)}_i, \theta^{(2)}_i) \prod_{i=1}^{n} s^{(2)}(x^{(2)}_i, \theta^{(2)}_i)[S^{(2)}(x^{(2)}_i, \theta^{(2)}_i)]^{n-1}. \]

Then, we can write,

\[ L = CL^{(1)}L^{(2)}. \]  \hspace{1cm} (3.3)

Maximizing \( L \) with respect to \( \theta^{(1)} \) and \( \theta^{(2)} \) is equivalent to maximizing \( L^{(1)} \) and \( L^{(2)} \) with respect to \( \theta^{(1)} \) and \( \theta^{(2)} \) respectively. Thus we have reduced the joint maximum likelihood problem for the parameters of two distributions to two separate estimation problems for the parameters of each distribution. Of course, the total sample information is used in each problem. The relation

\[ \ln L = \ln C + \ln L^{(1)} + \ln L^{(2)} \]  \hspace{1cm} (3.4)

reduces the asymptotic variance-covariance matrix \( \Sigma \) of \( (\hat{\theta}^{(1)}, \hat{\theta}^{(2)}) \) to a block-diagonal matrix. For \( \Sigma \) is derived by inverting the information matrix, whose elements are the negatives of the expected values of the second order derivatives of the logarithm of the likelihood function. From the relation (3.4) it follows that

\[ \frac{\partial^2 \ln L}{\partial \theta^{(1)}_i \partial \theta^{(2)}_j} = 0 \quad i = 1, \ldots, k_1; \quad j = 1, \ldots, k_2 \]  \hspace{1cm} (3.5)

so that the information matrix is block-diagonal, that is, it can be written in the form

\[ I = \begin{bmatrix} I^{(1)} & 0 \\ 0 & I^{(2)} \end{bmatrix} \]  \hspace{1cm} (3.6)

where \( I^{(1)} \) and \( I^{(2)} \) are the information matrices for \( \hat{\theta}^{(1)} \) and \( \hat{\theta}^{(2)} \) respectively. Therefore, the asymptotic variance-covariance matrix can similarly be written in the form

\[ \Sigma = \begin{bmatrix} \Sigma^{(1)} & 0 \\ 0 & \Sigma^{(2)} \end{bmatrix} \]  \hspace{1cm} (3.7)

where \( \Sigma^{(1)} \) and \( \Sigma^{(2)} \) are the asymptotic variance-covariance matrices of \( \hat{\theta}^{(1)} \) and \( \hat{\theta}^{(2)} \) respectively. Notice that the estimates of the elements of \( \theta^{(1)} \) are uncorrelated with those of \( \theta^{(2)} \). The large sample 100(1 - \( \alpha \))% joint confidence region (ellipsoidal) for true \( (\theta^{(1)}, \theta^{(2)}) \) is

\[ ([\hat{\theta}^{(1)} - \theta^{(1)}]; [\hat{\theta}^{(2)} - \theta^{(2)}]) \Sigma^{-1} ([\hat{\theta}^{(1)} - \theta^{(1)}]; [\hat{\theta}^{(2)} - \theta^{(2)}])' \leq \chi^2_{k_1+k_2}(\alpha) \]  \hspace{1cm} (3.8)

where \( \chi^2_{k_1+k_2}(\alpha) \) is selected such that

\[ \Pr \{ x^2_{k_1+k_2} \geq \chi^2_{k_1+k_2}(\alpha) \} = \alpha \]

and

\[ \Sigma^{-1} = \mathcal{I} = \begin{bmatrix} I^{(1)} & 0 \\ 0 & I^{(2)} \end{bmatrix} \]  \hspace{1cm} (3.9)
The elements of $f^{(3)}$ are

$$
-\left(\frac{\partial^2 \ln L^{(1)}}{\partial \theta_1^{(1)} \partial \theta_1^{(i)}}\right)_j, \quad i, i' = 1, \cdots, k_1. \tag{3.10}
$$

In a similar fashion we estimate the elements of $f^{(2)}$. The functions $L^{(1)}$ and $L^{(2)}$ carry a simple physical interpretation. $L^{(1)}$ is the likelihood function of a progressively censored sample, in which failures are observed at $x_1^{(i)}, \cdots, x_n^{(i)}$, one unit is removed from further observation (censored) at each of the times $x_i^{(1)}, \cdots, x_i^{(n)}$, and the remaining $(N - n_1 - n_2)$ units are censored at time $x_T$. This is intuitively reasonable, since, from the definition of the Multi-Risk Model, the fact that a unit failed due to the second cause at time $x_i^{(1)}$ means that it would have failed due to the first cause at some (unobserved) time greater than $x_i^{(1)}$. The interpretation of $L^{(2)}$ is similar, with the role of cause one and two reversed.

Two summarize, we have shown that the joint maximum likelihood problem for two causes of failure can be reduced to two separate maximum likelihood estimation problems for progressively censored samples from each failure distribution and that the estimates of parameters of one distribution are asymptotically uncorrelated with those of the other. Known results from [3], [4] for progressively censored samples can be applied immediately. We shall illustrate the method with the exponential and Weibull distributions.

4. Examples

Example 1: Underlying distribution belongs to Exponential Family.

For the first cause, let

$$
f^{(1)}(x, \theta^{(1)}) = \frac{1}{\theta^{(1)}} \exp \left( -\frac{x}{\theta^{(1)}} \right) ; \xi^{(1)}(x, \theta^{(1)}) = \exp \left( -\frac{x}{\theta^{(1)}} \right), \quad x \geq 0, \theta^{(1)} > 0
$$

Then

$$
-\ln L^{(1)} = \sum_1^N \left( \frac{x_i^{(1)}}{\theta^{(1)}} - \ln \theta^{(1)} \right) + \ln \theta^{(1)} + \frac{N - n_1 - n_2}{\theta^{(1)}} x_T
$$

The m.l.e. of $\theta^{(1)}$ is

$$
\hat{\theta}^{(1)} = \frac{\sum x_i^{(1)} + \sum x_j^{(2)} + (N - n_1 - n_2)x_T}{n_1}
$$

where $\sum_i^*$ indicates summation over the entire sample including the $x_i^{(1)}$s, $x_j^{(2)}$s, and the $(N - n_1 - n_2)$ observations censored at time $x_T$ assigned the value $x_i = x_T$. The variance-covariance matrix $\Sigma^{(1)}$ reduces to the single term

$$
\text{Var} (\hat{\theta}^{(1)}) = \frac{\theta^{(1)}_{\text{m.l.e.}}}{n_1}
$$

(4.3)
Let us define

$$\theta^{(1)} = \left( \sum x_i + \frac{n_1}{n_1 + n_2} (N - n_1 - n_2) x_T \right) / n_1$$

(4.4)

and similarly for the second cause

$$\theta^{(2)} = \left( \sum x_i^{(2)} + \frac{n_2}{n_1 + n_2} (N - n_1 - n_2) x_T \right) / n_2$$

(4.5)

$$\bar{\theta}^{(1)}$$ can be viewed as a usual estimate of $$\theta^{(1)}$$, using the censored sample

$$(x_1^{(1)}, \ldots, x_n^{(1)})$$ from $$F^{(1)}(x, \theta^{(1)})$$ adjusted for first cause items surviving in

$$(N - n_1 - n_2)$$ remaining items, and ignoring sample information from the

second cause. A similar interpretation for $$\theta^{(2)}$$ holds. Then

$$\theta^{(1)} = \theta^{(1)} + \frac{n_1}{n_1} \theta^{(2)}$$

(4.6)

$$\bar{\theta}^{(2)} = \bar{\theta}^{(2)} + \frac{n_2}{n_2} \bar{\theta}^{(1)}$$

(4.7)

If $$n_1 = 0$$, the estimating equation gives no reasonable estimate of $$\theta^{(1)}$$. This

do not present any practical problems of interpretation for it is reasonable to

conclude that either $$\theta^{(1)}$$ is very large or the corresponding cause of failure is

inoperative. In real life situations we expect to have the sample size $$n_1 + n_2$$

and the point of censoring $$x_T$$ large enough so that the probability of encountering

this difficulty will be quite small.

From (2.2), the combined survival function

$$S(x, \theta) = S^{(1)}(x, \theta^{(1)}) S^{(2)}(x, \theta^{(2)})$$

(4.8)

is also of exponential form, with parameter

$$1/\theta = 1/\theta^{(1)} + 1/\theta^{(2)}$$

(4.9)

The m.l.e. of $$\theta$$ derived with Type I censoring and ignoring causes of failure is

$$\hat{\theta} = \left( \sum x_i^{(1)} + \sum x_i^{(2)} + (N - n_1 - n_2) x_T / (n_1 + n_2) \right)$$

$$= (n_1 \bar{\theta}^{(1)} + n_2 \bar{\theta}^{(2)}) / (n_1 + n_2)$$

(4.10)

It can be shown that

$$1/\hat{\theta} = 1/\hat{\theta}^{(1)} + 1/\hat{\theta}^{(2)}$$

(4.11)

which agrees with (4.8). This result does not in general hold for other distribution

families.

Example 2: Underlying distribution belongs to Weibull Family.

Let r.v. $$X^{(1)}$$ follow Weibull distribution, then

$$F^{(1)}(x, \nu^{(1)}, \theta^{(1)}) = 1 - \exp \left[ -\frac{x}{\theta^{(1)}} \right]$$

$$x \geq 0, \ \nu^{(1)} > 0, \ \theta^{(1)} > 0.$$
where we use the parametric form chosen by Cohen [4] for ease in deriving the m.l.e.s. For notational simplicity, we shall suppress superscripts referring to the parameters \( \nu \) and \( \theta \). Cohen [4] shows that \( \nu \) is the solution of

\[
\sum_{i} \nu^i x_i \ln x_i - \frac{1}{\nu} = \frac{1}{n} \sum_{i} \nu^i \ln x_i^{(i)}
\]

(4.10)

where \( \sum^\nu \) has the meaning given in (4.2).

Equation (4.10) can be solved iteratively. It has been our experience that the first term in (4.10) varies very slowly with \( \nu \). Therefore, if we substitute an initial approximation for \( \nu \) in the first term and solve for \( \nu \) recursively, the convergence is generally very rapid. An initial approximation to \( \nu \) can be readily found by plotting a non-parametric estimate of \( F^{(i)}(x) \) on Weibull probability paper and estimating the slope \( \nu \) graphically.

Once \( \nu \) has been determined, then

\[
\hat{\theta} = \frac{\left( \sum^\nu x_i^\nu \right)}{n}
\]

(4.11)

Cohen [4] also gives the following results for the large sample estimates of the elements of information matrix of \((\nu, \hat{\theta})\).

\[
\begin{align*}
-\left( \frac{\partial^2 \ln L}{\partial \nu^2} \right)_{i, \hat{\theta}} &= \frac{n_i}{\nu^i} + \frac{1}{\nu} \sum^\nu x_i^i (\ln x_i)^2 \\
-\left( \frac{\partial^2 \ln L}{\partial \nu \partial \hat{\theta}} \right)_{i, \hat{\theta}} &= -\frac{1}{\nu^i} \sum^\nu x_i^i \ln x_i \\
-\left( \frac{\partial^2 \ln L}{\partial \theta^2} \right)_{i, \hat{\theta}} &= \frac{n_i}{\hat{\theta}^2}
\end{align*}
\]

(4.12)

It is sometimes convenient and even meaningful to use another form of the Weibull distribution, namely

\[
P^{(1)}(x, \beta, \alpha) = 1 - \exp \left[ -\left( \frac{x}{\alpha} \right)^\beta \right]. \quad x > 0, \quad \beta > 0, \quad \alpha > 0,
\]

where \( \alpha \) is the value of \( x \) for which

\[
S^{(1)}(x, \beta, \alpha) = \frac{1}{e}, \quad \text{regardless of } \beta.
\]

\((\alpha, \beta)\) is related to \((\hat{\theta}, \nu)\) by

\[
\begin{align*}
\nu &= \beta \\
\hat{\theta} &= \alpha^\beta.
\end{align*}
\]

(4.13)

The estimation problem can either be solved directly or by transforming to the new set of parameters. In the latter case we find

\[
\hat{\beta} = \nu \quad \text{and} \quad \hat{\alpha} = \hat{\theta}^{1/\beta}.
\]

The new information matrix for \((\hat{\alpha}, \hat{\beta})\) has its elements given by

\[
-\left( \frac{\partial^2 \ln L}{\partial \hat{\alpha}^2} \right)_{\hat{\beta}, \hat{\beta}} = -\left( \frac{\partial^2 \ln L}{\partial \nu^2} \right)_{\nu, \nu} + 2 \hat{\theta} \ln \hat{\alpha} \left( \frac{\partial^2 \ln L}{\partial \nu \partial \hat{\theta}} \right)_{\nu, \hat{\theta}} + \left( \partial \ln \hat{\alpha} \right)^2 \left( \frac{\partial^2 \ln L}{\partial \theta^2} \right)_{\nu, \hat{\theta}}
\]
These relations can be derived easily by the change of variables method of calculus. See [8]. It is also sometimes of interest to consider the shape parameter $\beta$ and the mean $\mu$ as the underlying parameters. Then we have the transformation

$$\beta^* = \beta$$

$$\mu = a\Gamma\left[\frac{1}{\beta} + 1\right]$$

and information matrix of $(\beta^*, \mu)$ is given by

$$\left(-\frac{\partial^2 \ln L}{\partial \beta^* \partial \beta^*}\right)_{\beta^*, \mu} = \left(-\frac{\partial^2 \ln L}{\partial \beta^* \partial \beta^*}\right)_{\beta, \mu} + 2\left(\frac{\partial \alpha}{\partial \beta}\right)_{\beta, \mu} \left(-\frac{\partial \ln L}{\partial \alpha \partial \beta}\right)_{\beta, \mu} + \left(\frac{\partial \alpha}{\partial \beta}\right)_{\beta, \mu} \left(-\frac{\partial^2 \ln L}{\partial \alpha^2}\right)_{\beta, \mu}$$

and

$$\left(-\frac{\partial^2 \ln L}{\partial \mu \partial \beta}\right)_{\beta^*, \mu} = \left[-\left(\frac{\partial \alpha}{\partial \beta}\right)_{\beta, \mu} \left(-\frac{\partial^2 \ln L}{\partial \alpha \partial \beta}\right)_{\beta, \mu} + \left(-\frac{\partial^2 \ln L}{\partial \alpha \partial \beta}\right)_{\beta, \mu} \left(\frac{\partial \alpha}{\partial \mu}\right)_{\beta, \mu}\right]^{(4.15)}$$

where

$$\left(\frac{\partial \alpha}{\partial \mu}\right) = \frac{1}{\Gamma(\beta + 1)}$$

$$\left(\frac{\partial \alpha}{\partial \beta}\right) = \frac{\alpha}{\beta^2} \psi\left(\frac{1}{\beta} + 1\right)$$

$$\Gamma(s) = \int_0^\infty e^{-u} u^{s-1} du, \quad s > 0$$

and

$$\psi(s) = \frac{d}{ds} \ln \Gamma(s).$$

The gamma function $\Gamma(s)$ and the digamma function $\psi(s)$ are tabulated in [1].

Example 3: Underlying distributions belong to Weibull Family with a common shape parameter.

Let us assume that r.v.s $X^{(1)}$ and $X^{(2)}$ follow Weibull distributions with common shape parameter, that is,

$$F^{(1)}(x, \theta^{(1)}, \nu) = 1 - \exp\left[-\frac{x^\nu}{\theta^{(1)}}\right] \quad x \geq 0, \quad \nu > 0, \quad \theta^{(1)} > 0.$$ 

$$F^{(2)}(x, \theta^{(2)}, \nu) = 1 - \exp\left[-\frac{x^\nu}{\theta^{(2)}}\right] \quad x \geq 0, \quad \nu > 0, \quad \theta^{(2)} > 0.$$ 

The likelihood function according to (3.1) is

\[ L = \text{const.} \left[ \left( \frac{\nu}{\theta^{(1)}} \right)^{n_1} \prod_i x_i^{(1)^{r-1}} \exp \left[ -\sum \frac{x_i^{(1)^r}}{\theta^{(1)}} \right] \right] \left[ \exp \left[ -\sum \frac{x_i^{(2)^r}}{\theta^{(2)}} \right] \right] \]

\[ \times \left[ \exp \left[ -\left( \frac{1}{\theta^{(1)}} + \frac{1}{\theta^{(2)}} \right) x_i^p \right] \right]^{N-(n_1+n_2)} \]

\[ \ln L = \ln \text{const.} + n_1 \ln \left( \frac{\nu}{\theta^{(1)}} \right) + (\nu - 1) \sum \ln x_i^{(1)} - \sum \frac{x_i^{(1)^r}}{\theta^{(1)}} - \sum \frac{x_i^{(1)^r}}{\theta^{(2)}} \]

\[ + n_2 \ln \left( \frac{\nu}{\theta^{(2)}} \right) + (\nu - 1) \sum \ln x_i^{(2)} - \sum \frac{x_i^{(2)^r}}{\theta^{(1)}} - \sum \frac{x_i^{(2)^r}}{\theta^{(2)}} \]

\[ - (N - n_1 - n_2) \left( \frac{1}{\theta^{(1)}} + \frac{1}{\theta^{(2)}} \right) x_i^p . \]

Setting

\[ \frac{\partial \ln L}{\partial \nu} = 0, \quad \frac{\partial \ln L}{\partial \theta^{(1)}} = 0 \quad \text{and} \quad \frac{\partial \ln L}{\partial \theta^{(2)}} = 0 \]

and after some simplification, we get, that m.l.e. \( \hat{\nu} \) is the solution of

\[ \sum x_i^{(1)} \ln x_i - \frac{1}{\nu} = \frac{1}{n_1 + n_2} \left[ \sum \ln x_i^{(1)} + \sum \ln x_i^{(2)} \right] \]

and

\[ \hat{\theta}^{(1)} = \frac{\sum x_i^{(1)}}{n_1} \]

\[ \hat{\theta}^{(2)} = \frac{\sum x_i^{(2)}}{n_2} \]

where \( \sum \) has the same meaning as before. The information matrix of \( (\hat{\nu}, \hat{\theta}^{(1)}, \hat{\theta}^{(2)}) \) can be shown readily to be

\[
\begin{bmatrix}
1 & (n_1 \cdot n_2) \cdot \left( \frac{1}{\hat{\theta}^{(1)}} \cdot \frac{1}{\hat{\theta}^{(2)}} \right) \sum x_i^{(1)} \ln x_i - \sum x_i^{(1)} \ln x_i - \sum x_i^{(2)} \ln x_i \\
- \sum x_i^{(1)} \ln x_i & \frac{n_1}{\hat{\theta}^{(1)}} & 0 \\
- \sum x_i^{(2)} \ln x_i & 0 & \frac{n_2}{\hat{\theta}^{(2)}}
\end{bmatrix}
\]

5. Application

Let us consider the illustration which led Cox [6] to formulate the Multi-Risk Model on which this paper is based. The data concern the failure-times of radio transmitter receivers, where the failures are classified into two types, those...
confirmed on arrival at the maintenance center (Type I) and those unconfirmed (Type II). Mendenhall and Hader [9] presented these data and applied a Single-Risk Model, in which the failure-time population was represented by a mixture of two exponential distributions, mixed in an unknown proportion. Cox studied their example critically, suggesting that the Multi-Risk Model might be more appropriate. For the sake of comparison, we assume underlying exponential distributions for two types of failures and give the m.l.e.s of the parameters.

### Table 1

<table>
<thead>
<tr>
<th>Hrs. Type I</th>
<th>Type II Total</th>
<th>Hrs. Type I</th>
<th>Type II Total</th>
<th>Hrs. Type I</th>
<th>Type II Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 2 3 5</td>
<td>328 5 2 7</td>
<td>408 3 -- 3</td>
<td>16 4 7</td>
<td>716 1 7 3</td>
<td>416 7 -- 7</td>
</tr>
<tr>
<td>40 11 2 13</td>
<td>240 1 1 2</td>
<td>440 2 1 3</td>
<td>48 3 2 5</td>
<td>246 1 1 2</td>
<td>448 2 -- 2</td>
</tr>
<tr>
<td>64 2 2 4</td>
<td>260 -- -- 4</td>
<td>464 1 -- 1</td>
<td>72 6 3 9</td>
<td>264 2 -- 3</td>
<td>472 -- 2 2</td>
</tr>
<tr>
<td>104 4 6 10</td>
<td>290 4 -- 4</td>
<td>504 1 -- 1</td>
<td>112 7 5 12</td>
<td>304 2 3 5</td>
<td>512 1 1 2</td>
</tr>
<tr>
<td>120 4 3 7</td>
<td>320 2 2 4</td>
<td>528 2 -- 2</td>
<td>126 1 -- 1</td>
<td>328 4 -- 4</td>
<td>536 1 -- 1</td>
</tr>
<tr>
<td>144 6 2 8</td>
<td>352 1 -- 1</td>
<td>560 1 1 2</td>
<td>152 4 2 6</td>
<td>360 3 1 4</td>
<td>568 1 -- 1</td>
</tr>
<tr>
<td>176 5 3 8</td>
<td>384 3 -- 3</td>
<td>592 2 -- 2</td>
<td>184 6 3 5</td>
<td>392 4 2 6</td>
<td>600 2 -- 2</td>
</tr>
</tbody>
</table>

Source: Mendenhall and Hader [7].

From Mendenhall and Hader data, we have

\[ n_1 = 218, \quad n_2 = 107, \quad N = 369 \text{ and } x_T = 630 \text{ hrs.} \]
Then
\[ \bar{x}^{(1)} = \frac{\sum x_i^{(1)}}{n_1} = \frac{50,056}{218} = 229.6 \text{ hrs.} \]
\[ \bar{x}^{(2)} = \frac{\sum x_i^{(2)}}{n_2} = \frac{20,458}{107} = 191.2 \text{ hrs.} \]

and from Example 1, rewriting equation (4.4), we get
\[ \bar{v}^{(1)} = \bar{x}^{(1)} + \left( \frac{N - n}{n} \right) x_p = 314.9 \text{ hrs.} \]
\[ \bar{v}^{(2)} = \bar{x}^{(2)} + \left( \frac{N - n}{n} \right) x_p = 276.5 \text{ hrs.} \]

Finally, from equation (4.5), we get m.l.e.s
\[ \hat{v}^{(1)} = \bar{v}^{(1)} + \frac{n_2}{n_1} \hat{v}^{(2)} = 450.6 \text{ hrs.} \]
\[ \hat{v}^{(2)} = \bar{v}^{(2)} + \frac{n_1}{n_2} \hat{v}^{(1)} = 918.1 \text{ hrs.} \]

and from equation (4.7)
\[ \hat{\theta} = 302.3 \text{ hrs.} \]

Hence, the observed p.d.f., regardless of causes of failures, is
\[ f(x, \omega) = \frac{1}{302.3} \exp \left[ \frac{-x}{302.3} \right], \quad x \geq 0 \]

By relation (2.3), we can express the above mentioned p.d.f. as
\[ f(x, \hat{\theta}) = f^{(1)}(x, \hat{v}^{(1)}), S^{(1)}(x, \hat{v}^{(1)}) + f^{(2)}(x, \hat{v}^{(2)}), S^{(1)}(x, \hat{v}^{(1)}) \]
\[ = \frac{1}{450.6} \exp \left[ -\frac{x}{450.6} \right] \exp \left[ -\frac{x}{918.1} \right] \]
\[ + \frac{1}{918.1} \exp \left[ -\frac{x}{918.1} \right] \exp \left[ -\frac{x}{450.6} \right] \]
\[ = 0.6708f(x, \hat{\theta}) + 0.3292f(x, \hat{\theta}). \]

That is, the observed failure-time distribution can be apportioned into failure-time distributions attributed to Type I and Type II failures respectively.

6. Acknowledgments

The authors wish to thank Dr. G. W. Barclay and Mr. S. Fried for their interest and encouragement in the project; Dr. T. Tarson and other members of the DP Management Science Department for active systems support relating to construction of historical data files; and Mr. S. Cohen of S. B. C. for his untiring programming efforts. The authors are indebted to the referees for their comments and suggestions concerning the presentation of this paper.
REFERENCES


