

Maximum Likelihood Analysis of Component Reliability Using Masked System Life-Test Data

John S. Usher, Member IEEE, Member ASQC
University of Louisville, Louisville
Thom J. Hodgson
North Carolina State University, Raleigh

Key Words — Masked data, Competing risk model, Maximum likelihood estimator.

Reader Aids —

Purpose: Describe and analyze a model

Special math needed for explanations: Elementary statistics

Special math needed to use results: Same

Results useful to: Reliability analysts and statisticians.

Summary & Conclusions — Life data from multi-component systems are often analyzed to estimate the reliability of each system component. Due to cost and diagnostic constraints, however, the exact cause of system failure might be unknown. We refer to such observations as being *masked*. Confronted with this real situation, we use a likelihood approach to exploit all the available information. We focus on a series system of three components, each with constant failure rate, and propose a simple numerical procedure for obtaining MLEs in the general case. We show that, under certain assumptions, closed-form solutions for the MLEs can be obtained. We consider that the cause of system failure can be isolated to some subset of components. This approach allows us to consider the full range of possible information on the cause of system failure. For example, if, for a particular observation, the subset contains a single component then the cause of failure is perfectly known. If all observations in the sample are of this type, then there is no masking and the model reduces to the standard competing risk model. In the other extreme, if, for a particular observation, the subset contains all of the components in the system, then the cause of system failure is completely unknown. If all of the masked observations in the sample are of this type, then the Miyakawa 2-component model applies.

The likelihood, while presented for complete data, can be extended to censoring. In the exponential case, censored system lifetimes serve only to increase the total time on test term in the likelihood expression. The general likelihood expression can be used with various component life distributions eg, Weibull, log-normal. However, closed-form MLEs will most certainly be intractable and numerical methods would be required.

1. INTRODUCTION

Life data from multi-component systems are often analyzed to estimate the reliability of each component. Such estimates can be extremely useful since they reflect

component reliability after assembly into an operational system. Under appropriate conditions, these estimates can be used to predict the reliability of a new configuration of components, eg, a new system.

In many cases, component reliability is estimated from system life-test data by making a "series" system assumption and applying a competing risk model [10]. This model has generally been applied under the assumption that the data consist of a system lifetime (failure time or censoring time) and an event indicator that denotes the cause of failure. Various studies have been done under this common assumption for both parametric [1, 5, 8] and nonparametric [3, 7] cases. However, in practice the cause of system failure and thus the indicator might be unobserved. Due to the high cost of detailed failure analysis as well as the lack of appropriate diagnostics, the cause of many failures is not fully investigated. The result is that we observe a system failure but the component causing failure might be unknown, see Usher [12] and Miyakawa [9]. Hereafter we refer to such observations as being *masked*. Because masked observations can represent an appreciable proportion of any set of system life-test or field data, it is essential that we exploit this additional, though limited, information.

Miyakawa [9] proposed both parametric and non-parametric estimators for a simple 2-component "series" system when the cause of failure can be unknown. In this case each system failure is known to be caused by component 1 or component 2 or else is completely unknown. In the parametric case he derives closed-form solutions for the MLEs when component lifetimes are exponentially distributed. He points out that these estimators can be used in the more general case of a J -component "series" system where J is some positive integer. That is, one simply considers a particular component as component-1 and all others (since they are in "series") as component-2. The cause of system failure is then either known to be component-1 or component-2 (one of the others in the system) or else is completely unknown.

The 2-component model when applied to the J -component case, however, implicitly assumes that when the cause of system failure is unknown, all J components can be suspected as being the true cause, ie, the cause of system failure is completely unknown. However, in many situations the exact cause of system failure, while still unknown, can at least be isolated to some subset of components. That is, many components might be eliminated as potential causes based upon such things as the mode of system failure or some brief diagnostic routine. For example, consider a large computer system that has failed. A repairman may immediately be able to isolate the cause of

failure down to a single circuit card (which may contain many components). In an effort to minimize downtime he may replace the entire circuit card and hence never know exactly which component caused the failure. While the cause of system failure is still unknown at the component level, the isolation of the smaller subset of possible causes provides additional information that can be useful in the estimation process. This motivates the need for a more general analysis of the problem.

We present a general method for estimating component reliabilities from masked system life-test data. In sections 2 and 3 we develop a generalized likelihood function for a J -component "series" system. In sections 4 and 5 we explore its use for the special case when $J = 3$ and component lifetimes are exponentially distributed. We find that, except under certain assumptions, closed-form MLEs are intractable and thus we propose a simple numerical solution procedure. Section 6 gives several numerical examples to illustrate the use of the method.

2. NOTATION & ASSUMPTIONS

Notation

T_{ij}	random lifelength of component j in system i , $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, J$
X_i	random lifelength of system i ; $X_i = \min\{T_{i1}, T_{i2}, \dots, T_{iJ}\}$
$f_j(t)$, $\bar{F}_j(t)$	pdf, Sf, of component j lifelength with parameter vector θ_j
S_i	subset of components known to contain the component causing system failure at time X_i , $S_i \subset \{1, 2, \dots, J\}$
L	likelihood function
\mathcal{L}	log-likelihood, $\mathcal{L} = \ln(L)$

Other, standard notation is given in "Information for Readers & Authors" at the rear of each issue.

Assumptions

1. T_{ij} , $i = 1, 2, \dots, n$, $j = 1, 2, \dots, J$ are s -independent random variables with T_{1j} , T_{2j} , ..., T_{nj} being identically distributed for each j .
2. The observable quantities for system i are (x_i, S_i) . There are n such systems.
3. The system is 1-out-of- J :F.
4. Masking occurs s -independently of the cause of failure.

3. THE LIKELIHOOD FUNCTION

The conditional pdf of the lifelength X_i for component j given that all other components survive is:

$$f_j(x_i) \prod_{\substack{l=1 \\ l \neq j}}^J \bar{F}_l(x_i).$$

Hence the likelihood function for a complete (uncensored) sample of n systems is [4]:

$$L = \prod_{i=1}^n \left(\sum_{j \in S_i} \left\{ f_j(x_i) \prod_{\substack{l=1 \\ l \neq j}}^J \bar{F}_l(x_i) \right\} \right). \quad (3.1)$$

The values of the parameter vector θ_j that maximize L yield the MLEs denoted as $\hat{\theta}_j$. These can be found by solving the set of likelihood equations:

$$\frac{\partial L}{\partial \theta_j} = 0 \quad j = 1, 2, \dots, J.$$

Usually such solutions cannot be obtained in closed form, and numerical procedures are required. However, we show that when component lifetimes are exponentially distributed, certain special-case closed forms can be obtained.

4. THE EXPONENTIAL CASE

When each component has a constant failure rate λ_j , the likelihood function is, from (3.1):

$$L = \prod_{i=1}^n \left(\sum_{j \in S_i} \left\{ \lambda_j \exp(-\lambda_j x_i) \prod_{\substack{l=1 \\ l \neq j}}^J \exp(-\lambda_l x_i) \right\} \right). \quad (4.1)$$

We seek the values of λ_j , $j = 1, 2, \dots, J$ which maximize L , or equivalently the log-likelihood \mathcal{L} .

2-Component System

Consider first the simplest case of $J = 2$ where the cause of failure can be unknown. Suppose n such systems are placed onto a life test. Let n_1 and n_2 denote the number of system failures for which the cause of failure is known to be components 1 and 2 respectively; ie, n_1 and n_2 denote the number of observations where $S_i = \{1\}$ and $S_i = \{2\}$ respectively, $i = 1, 2, \dots, n$. Let n_{12} denote the number of masked observations, ie, where $S_i = \{1, 2\}$. From (4.1), the log-likelihood and the resulting likelihood equations are:

$$\begin{aligned} x_{\text{TTT}} &\equiv \sum_{i=1}^n x_i, \text{ total time on test} \\ \mathcal{L} &= -x_{\text{TTT}} (\lambda_1 + \lambda_2) \\ &\quad + n_1 \ln(\lambda_1) + n_2 \ln(\lambda_2) + n_{12} \ln(\lambda_1 + \lambda_2) \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_1} &= 0 = -x_{\text{TTT}} + \frac{n_1}{\lambda_1} + \frac{n_{12}}{\lambda_1 + \lambda_2} \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= 0 = -x_{\text{TTT}} + \frac{n_2}{\lambda_2} + \frac{n_{12}}{\lambda_1 + \lambda_2} \end{aligned}$$

The closed-form solution is:

$$\hat{\lambda}_1 = \left(n_1 + n_{12} \frac{n_1}{n_1 + n_2} \right) / x_{TTT}$$

$$\hat{\lambda}_2 = \left(n_2 + n_{12} \frac{n_2}{n_1 + n_2} \right) / x_{TTT}$$

When the sample does not contain any masked observations the estimators reduce to the standard MLEs for the failure rate parameter λ_j , viz, the ratio of the total number of failures of component j to the total time on test.

3-Component System

More Notation

- n_i number of failures where $S_i = \{i\}$, $i = 1, 2, 3$.
 n_{12} , n_{13} , n_{23} number of partially-masked observations where $S_i = \{1, 2\}$, $S_i = \{1, 3\}$, and $S_i = \{2, 3\}$ respectively.
 n_{123} number of completely masked observations where $S_i = \{1, 2, 3\}$.

The log-likelihood and likelihood equations are:

$$\begin{aligned} \mathcal{L} = & -x_{TTT}(\lambda_1 + \lambda_2 + \lambda_3) + n_1 \ln(\lambda_1) + n_2 \ln(\lambda_2) \\ & + n_3 \ln(\lambda_3) + n_{12} \ln(\lambda_1 + \lambda_2) + n_{13} \ln(\lambda_1 + \lambda_3) \\ & + n_{23} \ln(\lambda_2 + \lambda_3) + n_{123} \ln(\lambda_1 + \lambda_2 + \lambda_3). \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_1} = 0 = & -x_{TTT} + \frac{n_1}{\lambda_1} + \frac{n_{12}}{\lambda_1 + \lambda_2} + \frac{n_{13}}{\lambda_1 + \lambda_3} \\ & + \frac{n_{123}}{\lambda_1 + \lambda_2 + \lambda_3} \end{aligned} \quad (4.2a)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_2} = 0 = & -\sum_{i=1}^n x_i + \frac{n_2}{\lambda_2} + \frac{n_{12}}{\lambda_1 + \lambda_2} + \frac{n_{23}}{\lambda_2 + \lambda_3} \\ & + \frac{n_{123}}{\lambda_1 + \lambda_2 + \lambda_3} \end{aligned} \quad (4.2b)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_3} = 0 = & -\sum_{i=1}^n x_i + \frac{n_3}{\lambda_3} + \frac{n_{13}}{\lambda_1 + \lambda_3} + \frac{n_{23}}{\lambda_2 + \lambda_3} \\ & + \frac{n_{123}}{\lambda_1 + \lambda_2 + \lambda_3} \end{aligned} \quad (4.2c)$$

Solution of these equations by hand is unwieldy. In an attempt to find a set of closed-form solutions, a computer-based approach was used. The equations were input to the SOLVE routine of MACSYMA [11], a LISP-based symbolic manipulation program developed at Massachusetts Institute of Technology Laboratory for Computer Science. The program was run on a VAX 11/750 with 15 megabytes of virtual memory in the Department of Statistics at North Carolina State University. The procedure, however, ran out of memory during execution and was unable to provide a solution. It is thus apparent that the solution is intractable and that numerical methods are required. One of the most widely used is Newton's method [2]. While generally considered one of the fastest, Newton's method and be very sensitive to the proper choice of starting values, and it can fail to converge to a solution. Other more robust, yet generally slower, methods are also available. Consider again the set of equations (4.2 a,b,c). A rearrangement of terms yields:

$$\begin{aligned} \lambda_1 = & \left(n_1 + n_{12} \frac{\lambda_1}{\lambda_1 + \lambda_2} + n_{13} \frac{\lambda_1}{\lambda_1 + \lambda_3} \right. \\ & \left. + n_{123} \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \right) / x_{TTT} \end{aligned} \quad (4.3a)$$

$$\begin{aligned} \lambda_2 = & \left(n_2 + n_{12} \frac{\lambda_2}{\lambda_1 + \lambda_2} + n_{23} \frac{\lambda_2}{\lambda_2 + \lambda_3} \right. \\ & \left. + n_{123} \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \right) / x_{TTT} \end{aligned} \quad (4.3b)$$

$$\begin{aligned} \lambda_3 = & \left(n_3 + n_{13} \frac{\lambda_3}{\lambda_1 + \lambda_3} + n_{23} \frac{\lambda_3}{\lambda_2 + \lambda_3} \right. \\ & \left. + n_{123} \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \right) / x_{TTT} \end{aligned} \quad (4.3c)$$

These equations are of the form $\lambda = T(\lambda)$ where $\lambda \equiv [\lambda_1 \lambda_2 \lambda_3]^T$. A common technique used to approximate solutions to this type of equation set is to define an iteration process by the equation:

$$\lambda^{(i+1)} = T(\lambda^{(i)}) \quad (4.4)$$

That is, an initial approximation $\lambda^{(0)}$ is obtained (guessed), then $\lambda^{(1)} = T(\lambda^{(0)})$, $\lambda^{(2)} = T(\lambda^{(1)})$, ... form successive approximations to the solution of (4.4). This is Picard iteration. Jensen & Rowland [6] present general sufficient conditions for the convergence of the Picard method. Although we do not show such conditions are met here, the procedure, when applied to the exponential case, converges rapidly. It is extremely simple to implement and yields solutions even with poor starting values. The steps of the procedure are:

1. Choose a starting value $\hat{\lambda}^{(0)}$ for the initial iteration, $i = 0$.
2. Compute $\hat{\lambda}^{(i+1)} = T(\hat{\lambda}^{(i)})$.
3. Check stopping criterion such as:

$$\|\hat{\lambda}^{(i+1)} - \hat{\lambda}^{(i)}\| \leq \epsilon$$

where ϵ is some small predetermined value and:

$$\|\hat{\lambda}^{(i+1)} - \hat{\lambda}^{(i)}\|_1 = |\hat{\lambda}_1^{(i+1)} - \hat{\lambda}_1^{(i)}| + |\hat{\lambda}_2^{(i+1)} - \hat{\lambda}_2^{(i)}| + |\hat{\lambda}_3^{(i+1)} - \hat{\lambda}_3^{(i)}|$$

4. If the stopping criterion is satisfied, STOP. Otherwise, let $i = i + 1$ and GO TO (2).

5. SPECIAL CASE CLOSED-FORMS

The solution of the general case (4.2) is intractable and numerical solution procedures are required. However, under certain assumptions, useful closed-form MLEs can be obtained.

CASE 1: The cause of each system failure is either perfectly known or else completely unknown, ie, assume $n_{12}, n_{13}, n_{23} = 0$ and $n_1, n_2, n_3, n_{123} \geq 0$. The solution to (4.2) is:

$$\hat{\lambda}_1 = \left(n_1 + n_{123} \frac{n_1}{n_1 + n_2 + n_3} \right) / x_{TTT} \quad (5.1a)$$

$$\hat{\lambda}_2 = \left(n_2 + n_{123} \frac{n_2}{n_1 + n_2 + n_3} \right) / x_{TTT} \quad (5.1b)$$

$$\hat{\lambda}_3 = \left(n_3 + n_{123} \frac{n_3}{n_1 + n_2 + n_3} \right) / x_{TTT}. \quad (5.1c)$$

This set of solutions extends the Miyakawa 2-component model to the 3-component case.

CASE 2: The cause of each system failure is either known with certainty or can be isolated to a particular subset of components. Without loss of generality assume that components 1 and 2 belong to this subset. The data are of the form $n_{13}, n_{23}, n_{123} = 0$ and $n_1, n_2, n_3, n_{12} \geq 0$. The solution to (4.2) is:

$$\hat{\lambda}_1 = \left(n_1 + n_{12} \frac{n_1}{n_1 + n_2} \right) / x_{TTT} \quad (5.2a)$$

$$\hat{\lambda}_2 = \left(n_2 + n_{12} \frac{n_2}{n_1 + n_2} \right) / x_{TTT} \quad (5.2b)$$

$$\hat{\lambda}_3 = (n_3) / x_{TTT}. \quad (5.2c)$$

Component 3 is not masked, hence, $\hat{\lambda}_3$ is the standard MLE. This case can therefore be viewed as a 2-component system with masking, placed in "series" with another component.

CASE 3: The cause of each system failure is either known, isolated to a particular subset of components, or else is completely masked. Again without loss of generality, let components 1 and 2 belong to the subset. The data are of the form $n_{13}, n_{23} = 0$ and $n_1, n_2, n_3, n_{12}, n_{123} \geq 0$. The solution to (4.2) is:

$$\hat{\lambda}_1 = \left(n_1 + n_{12} \frac{n_1}{n_1 + n_2} + n_{123} \frac{n_1}{n_1 + n_2 + n_3} \right) / x_{TTT} \quad (5.3a)$$

$$\hat{\lambda}_2 = \left(n_2 + n_{12} \frac{n_2}{n_1 + n_2} + n_{123} \frac{n_2}{n_1 + n_2 + n_3} \right) / x_{TTT} \quad (5.3b)$$

$$\hat{\lambda}_3 = \left(n_3 + n_{123} \frac{n_3'}{n_1 + n_2 + n_3} \right) / x_{TTT} \quad (5.3c)$$

$$n_3' \equiv n_3 / \left(1 + \frac{n_{12}}{n_1 + n_2} \right).$$

6. NUMERICAL EXAMPLES

The 3-component "series"-system life-data presented in table 1 were simulated assuming exponentially distributed component lifetimes with parameters $\lambda_1 = 1$, $\lambda_2 = 1$ and $\lambda_3 = 1$. Columns 1-3 of table 1 give the system number, the time to failure of the system, and the true cause of system failure respectively. (The true cause of system failure was found simply by observing the minimum lifetime of the three components).

Under the assumption of perfect information (no masking) the MLE for each component can be found by the well-known relationship:

$$\hat{\lambda}_j = n_j / x_{TTT}, \quad j = 1, 2, 3$$

$n_1 = 8, n_2 = 12, n_3 = 10$ and $x_{TTT} = 10.136$. The resulting MLEs are $\hat{\lambda}_1 = 0.789, \hat{\lambda}_2 = 1.184$, and $\hat{\lambda}_3 = 0.987$.

To simulate the general effects of masking, approximately 30% of the observations were randomly masked. The information on the cause of system failure (S_i) is given in column 4 of table 1. The calculations are:

$$n_1 = 6 \quad n_{13} = 3 \quad n_{123} = 3$$

$$n_2 = 6 \quad n_{13} = 1$$

$$n_3 = 8 \quad n_{23} = 3 \quad \sum_{i=1}^{30} x_i = 10.13 = x_{TTT}$$

The data were analyzed using the numerical algorithm of section 4. Starting points were arbitrarily chosen as:

TABLE 1
Simulated System Life Data for a Three-Component System with Masking

#1 System i	#2 Time to Failure x_i	#3 Cause of Failure S_i	#4 General Masking S_i	#5 Case 1 S_i	#6 Case 2 S_i	#7 Case 3 S_i
1	.021	{2}	{2}	{2}	{2}	{2}
2	.038	{2}	{1,2}	{2}	{1,2}	{1,2}
3	.054	{3}	{3}	{3}	{3}	{3}
4	.066	{3}	{3}	{3}	{3}	{3}
5	.076	{2}	{1,2}	{2}	{1,2}	{1,2}
6	.078	{2}	{2,3}	{2}	{2}	{2}
7	.123	{3}	{3}	{3}	{3}	{3}
8	.130	{3}	{1,3}	{3}	{3}	{3}
9	.152	{2}	{1,2,3}	{1,2,3}	{2}	{1,2,3}
10	.159	{1}	{1}	{1}	{1}	{1}
11	.199	{3}	{3}	{3}	{3}	{3}
12	.201	{1}	{1}	{1}	{1}	{1}
13	.204	{1}	{1}	{1}	{1}	{1}
14	.215	{3}	{2,3}	{3}	{3}	{3}
15	.218	{2}	{1,2}	{2}	{1,2}	{1,2}
16	.281	{1}	{1}	{1}	{1}	{1}
17	.295	{2}	{2}	{2}	{2}	{2}
18	.310	{3}	{3}	{3}	{3}	{3}
19	.338	{3}	{3}	{3}	{3}	{3}
20	.341	{2}	{2}	{2}	{2}	{2}
21	.354	{1}	{1}	{1}	{1}	{1}
22	.358	{2}	{2}	{2}	{2}	{2}
23	.431	{1}	{1,2,3}	{1,2,3}	{1}	{1,2,3}
24	.457	{3}	{3}	{3}	{3}	{3}
25	.545	{1}	{1,2,3}	{1,2,3}	{1}	{1,2,3}
26	.569	{2}	{2}	{2}	{2}	{2}
27	.677	{3}	{3}	{3}	{3}	{3}
28	.818	{2}	{2}	{2}	{2}	{2}
29	.946	{2}	{2,3}	{2}	{2}	{2}
30	1.486	{1}	{1}	{1}	{1}	{1}

$$\hat{\lambda}_j^{(0)} = n_j / x_{TTT}, \quad j = 1, 2, 3.$$

As programmed in ADVANCED BASIC and run on an IBM-PC/AT, the procedure converged in only 7 iterations to $\hat{\lambda}_1 = 0.858$, $\hat{\lambda}_2 = 0.988$, $\hat{\lambda}_3 = 1.113$. The stopping criterion was set at $\epsilon = 10^{-4}$.

We now consider that the cause of system failure is perfectly known or else completely masked, (CASE 1, Section 5). To evaluate this situation with the same data set we replace the observations where $S_i = \{1, 2\}$, $S_i = \{1, 3\}$, and $S_i = \{2, 3\}$ with the true cause of system failure. The resulting values of S_i are given in column 5 of table 1. The calculations are:

$$\begin{aligned} n_1 &= 6 & n_{12} &= 0 & n_{123} &= 3 \\ n_2 &= 11 & n_{12} &= 0 & & \\ n_3 &= 10 & n_{23} &= 0 & x_{TTT} &= 10.136 \end{aligned}$$

$$\text{From (5.1) — } \hat{\lambda}_1 = 0.658, \hat{\lambda}_2 = 1.206, \hat{\lambda}_3 = 1.096.$$

Next, we consider that the cause of failure is either perfectly known or else isolated to the subset containing

components 1 and 2, (CASE 2, Section 5). To evaluate this situation we replace observations where $S_i = \{1, 3\}$, $S_i = \{2, 3\}$, or $S_i = \{1, 2, 3\}$ with the true cause of system failure. The resulting values of S_i are given in column 6 of table 1. The results are:

$$\begin{aligned} n_1 &= 8 & n_{12} &= 3 & n_{123} &= 0 \\ n_2 &= 9 & n_{13} &= 0 & & \\ n_3 &= 10 & n_{23} &= 0 & x_{TTT} &= 10.136 \end{aligned}$$

$$\text{From (5.2) — } \hat{\lambda}_1 = 0.929, \hat{\lambda}_2 = 1.045, \hat{\lambda}_3 = 0.987.$$

Finally, we consider that the cause of system failure is either perfectly known, isolated to the subset containing components 1 and 2, or else is completely unknown (CASE 3, Section 5). To evaluate this situation we replace observations where $S_i = \{1, 3\}$ or $S_i = \{2, 3\}$ with the true cause of system failure. The resulting values of S_i are given in column 7 of table 1. The calculations are:

$$\begin{aligned} n_1 &= 6 & n_{12} &= 3 & n_{123} &= 3 \\ n_2 &= 8 & n_{13} &= 0 & & \\ n_3 &= 10 & n_{23} &= 0 & x_{TTT} &= 10.136 \end{aligned}$$

