QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL, VOL. 5, 257–261 (1989)

AN ITERATIVE APPROACH FOR ESTIMATING COMPONENT RELIABILITY FROM MASKED SYSTEM LIFE DATA

JOHN S. USHER
Department of Industrial Engineering, University of Louisville, Louisville, KY 40292, U.S.A.

AND

FRANK M. GUESS
Department of Statistics, University of Tennessee, Knoxville, TN 37996

SUMMARY

Life data from systems of components are often analysed to estimate the reliability of the individual components. These estimates are useful since they reflect the reliability of the components under actual operating conditions. However, owing to the cost or time involved with failure analysis, the exact component causing system failure may be unknown or 'masked'. That is, the cause may only be isolated to some subset of the system's components. We present an iterative approach for obtaining component reliability estimates from such data for series systems. The approach is analogous to traditional probability plotting. That is, it involves the fitting of a parametric reliability function to a set of non-parametric reliability estimates (plotting points). We present a numerical example assuming Weibull component life distributions and a two-component series system. In this example we find estimates with only 4 per cent of the computation time required to find comparable MLEs.

KEY WORDS Competing risk model Masking Kaplan–Meier estimator reliability

1. INTRODUCTION

Engineers are often interested in estimating component reliability through the analysis of system life data. The resulting estimates are important since they reflect the true reliability of the components during actual system operation. These estimates account for reliability degradation induced by the effects of system assembly, interaction with other components, true operating conditions, etc. Once found, they can be used, for example, in evaluating the elimination of a particular failure mode, estimating the reliability of a new configuration of similar components, or assessing a potential design change.

The problem of estimating component reliability from system life data has usually been addressed by making a series-system assumption and applying a competing risks model. The literature on competing risks is abundant. General results and references can be found in References 1–4. In the competing-risks framework it is necessary to make the assumption that the data consist of a system life length (failure or consoring time) and an indicator of the cause of failure. However, owing to the high cost of detailed system failure analysis, especially in complex systems, the exact cause of failure may be unknown. For example, Usher[5] reports such a situation in the case of large computer systems. Upon failure, the cause may be isolated only to a circuit card. This circuit card is then viewed as a subsystem with relatively few components. In an attempt to repair the system quickly, the entire circuit card (subsystem) may be replaced and not investigated further. The result is an observed system failure time and partial information on which component caused the failure. We refer to such observations as being 'masked'.

Miyakawa[6] proposes useful estimators for a simple two-component series system when the cause of system failure may be unknown. In the parametric case he derives closed-form maximum likelihood estimators (MLEs) when the components have exponentially distributed life lengths. In the non-parametric case he suggests several estimators based upon the well-known product-limit estimator studied by Kaplan and Meier.[7] Guess, Usher and Hodgson[8] provide clarification of Miyakawa's likelihood approach. They develop a more general likelihood expression for an arbitrary J-component series system, which has analogous extensions to any type of coherent system. Usher and Hodgson[9] explore the use of the new likelihood expression in the case of a series sytem of $J = 3$ exponentially distributed components. They find that closed-form MLEs are algebraically intractable except under certain tight

0748–8017/89/040257–05$05.00

Received 1 November 1988
Revised 8 February 1989

restrictions. This result was obtained after attempting to use the SOLVE routine of MACSYMA,[10] a LISP-based symbolic manipulation program developed at Massachusetts Institute of Technology Laboratory for Computer Science. The program (executed on a VAX 11/750 with 15 MB of virtual memory) ran out of memory during execution and was unable to find a solution.

As a result of the problem's intractable nature (even in the exponential case) engineers are forced to use numerical procedures to obtain MLEs of component reliabilities from masked data. Although there are many non-linear numerical optimization algorithms available,[11] they can be time-consuming and cumbersome to apply. Most require good initial starting points to guarantee convergence. For a good discussion of convergence problems encountered in the use of the ML method in multiparameter situations, see Reference 12. In addition, the algorithms are generally computationally intensive. This can result in extremely long solution times for large problems in a real-world setting. In addition, the faster gradient methods, such as Newton–Raphson, require first and second partial derivatives of the likelihood function. For the masking case, these are not available in closed form, therefore requiring even more numerical computation.

The overall result is that MLEs, with their desirable statistical properties (invariance, asymptotic efficiency and normality, etc.), may be expensive to obtain. Industry experience suggests that many engineers are interested in simpler solution procedures that generate 'good' estimates at reasonable cost. The widespread use of graphical (probability plotting) techniques is a good example of this phenomenon. See Reference 13 for numerous examples of Weibull probability plots. In a sense, these graphical techniques gain the advantage of computational ease with the sacrifice of some statistical power. See References 1 and 14 for a review and discussion of graphical procedures and their properties. Also compare another graphical orientation to data analysis via the total time on test plots found in References 15 and 16.

In this paper, we address the problem of obtaining parametric estimates of component reliabilities from masked system life data. We propose an iterative solution procedure that is analogous to traditional probability plotting. That is, we first find non-parametric reliability estimates (plotting points) for each component in the system. These non-parametric estimates account for the effects of masking. To this, we then fit an assumed parametric distribution, e.g. Weibull. The fitting process is performed using least-squares, the 'graphical step'. The resulting parametric estimates are next used to obtain revised plotting points for the non-parametric reliability function. These are then used to obtain revised parametric estimates. The process is repeated iteratively until convergence is reached.

We present the procedure in Section 2. We illustrate its use with a numerical example in Section 3 for a two-component system under the assumption of Weibull component life-length distributions. In this example the procedure is compared to the method of maximum likelihood and found to obtain comparable estimates with only four per cent of the computation time. Some concluding remarks are given in Section 4.

2. THE MODEL

Consider that a sample of n J-component series systems are placed onto a life test. Let T_i denote the random life length of the ith system; $i = 1, 2, \ldots, n$. Let T_{ij} denote the random life length of the jth component in the ith system; $j = 1, 2, \ldots, J$. Note that $T_i = \min\{T_{i1}, T_{i2}, \ldots, T_{iJ}\}$. We assume that the T_{ij}s are independent random variables with $T_1, T_{2j}, \ldots, T_{nj}$ being identically distributed for each j. Let $F_j(t)$ and $\bar{F}_j(t)$ denote the distribution function and reliability function, respectively, for component j. We assume that $F_j(t)$ has density function $f_j(t)$ with parameter vector θ_j.

Upon system failure we also observe S_i, the subset of components known to contain the component causing system failure. Hence, $S_i \subset \{1, 2, \ldots, J\}$. Note that if S_i contains a single element j, then the cause of system failure is known to be component j. If S_i contains all possible elements, $1, 2, \ldots, J$, then the cause of failure is completely unknown. This subset approach allows us to consider the full range of possible information on the cause of system failure. In the case of complete data (no time censoring), the observed data is (T_1, S_1), (T_2, S_2), $\ldots$, (T_n, S_n).

Let $p_j(t_i)$ denote the conditional probability that component j is the true cause of system failure at time t_i. We make the reasonable assumption that each component has a smooth and continuous hazard function, $h_j(t)$, where

$$h_j(t) = \frac{f_j(t)}{\bar{F}_j(t)} \tag{1}$$

We then estimate $p_j(t_i)$ as

$$\hat{p}_j(t_i) = \frac{\hat{h}_j(t_i)}{\sum_{l \in S_i} \hat{h}_l(t_i)} \tag{2}$$

That is, $\hat{p}_j(t_i)$ represents the fractional contribution of component j's hazard rate to the total hazard rate of all components suspected as being the cause of failure. Note that when the cause is known to be component j, $S_i = \{j\}$, $\hat{p}_j(t_i) = 1$ and $\hat{p}_k(t_i) = 0$ for all $k \neq j$.

Now, suppose that the $\hat{p}_j(t_i)$ have been obtained for all i (for j fixed). As an extension to the product limit estimator, consider the following non-parametric estimator for the reliability of component j:

$$\hat{\bar{F}}_j^*(t) = \prod_{t_i \le t} \left\{ \frac{n_i - \hat{p}_j(t_i)}{n_i} \right\} \quad (3)$$

where n_i denotes the number of systems still functioning at time t_i and $\hat{p}_j(t_i)$ is as given in (2). The term in parentheses represents the estimated conditional reliability of component j given survival to time t_i. We assume, for simplicity, that there are no ties between the t_is. Miyakawa[6] presents a similar estimator that uses a special case of (2), i.e. the cause of failure is assumed to be either known or completely unknown, see also Reference 17. In addition, Miyakawa presents another useful estimator where the $\hat{p}_j(t_i)$ is placed in the exponent. This estimator is shown to perform slightly better than (3) and to be consistent with the Kaplan–Meier estimator of the system reliability function. However, for savings in the computation cost of repeated exponentiation, we prefer the use of (3). In addition, we feel that (3) is simpler and has greater intuitive appeal. We point out that either estimator would be amenable to our proposed iterative approach.

When the cause of failure is known to be component j at time t_i we find that $\hat{p}_j(t_i) = 1$. This shows that (3) reduces to the standard product limit estimator in the complete data case. Also note that if component j is not the cause of failure at time t_i, i.e. $\hat{p}_j(t_i) = 0$, then the conditional reliability term (in parentheses) appropriately reduces to 1. That is, t_i simply acts as a censoring time for component j.

The values of $\{\hat{\bar{F}}_j^*(t)\}_i$ yield plotting points that provide parametric estimates either graphically or through the method of least squares. However, we must first estimate $p_j(t_i)$, $j = 1, 2, \ldots, J$, and $i = 1, 2, \ldots, n$, using (2). We see that this in turn requires estimates of $h_j(t_i)$, $j = 1, 2, \ldots, J$, and $i = 1, 2, \ldots, n$. These estimated hazard rates are functions of $\hat{\theta}_j$, $j = 1, 2, \ldots, J$. The parameter estimates, $\hat{\theta}_j$ are estimated by fitting the parametric reliability function, $\bar{F}_j$, to the non-parametric plotting points given by $(t_i, \hat{\bar{F}}_j^*(t))$, $i = 1, 2, \ldots, n$. This apparent closed-loop problem can be solved through the use of an iterative procedure as follows:

1. From the sample of n observations, select only those where the cause of failure is known. (We assume, of course, that at least two observations with known cause of failure are available.)
2. Find the initial non-parametric estimates, (plotting points), $\hat{\bar{F}}_j^*(t)$, at the selected failure times, using (3), i.e. using only the t_i with $S_i = \{k\}$ for some k, find the point $(t_i, \hat{\bar{F}}_j^*(t))$ with $\hat{p}_j(t_i) = 1$ if $S_i = \{j\}$ and $\hat{p}_j(t_i) = 0$ for $S_i = \{k\}$ and $k \ne j$.
3. Find $\hat{\theta}_j$ by fitting (graphically or by least squares) the parametric reliability function $\bar{F}_j$ to the non-parametric plotting points from step 2 (first iteration) or from step 6 (subsequent iterations).
4. Calculate $\hat{h}_j(t_i)$, $j = 1, 2, \ldots, J$, and $i = 1, 2, \ldots, n$ using $\hat{\theta}_j$ from step 3 and (1).
5. Find updated values of $\hat{p}_j(t_i)$, $j = 1, 2, \ldots, J$, and $i = 1, 2, \ldots, n$ using (2).
6. Find updated non-parametric estimates, $\hat{\bar{F}}_j^*(t)$, $j = 1, 2, \ldots, J$, and $i = 1, 2, \ldots, n$ using (3).
7. Go to step 3 and repeat until $\hat{\theta}_j$ converges, $j = 1, 2, \ldots, J$.

3. NUMERICAL EXAMPLE

To illustrate the approach, we now present a numerical example. We consider, for ease of exposition, a simple $J = 2$ component series system. We further assume that each component life length is Weibull distributed, where

$$\bar{F}_j(t) = \exp\left[-\left\{ \frac{t}{\theta_j} \right\}^{\beta_j} \right], \; j = 1, 2 \quad (4)$$

The data in Table I represent the life length and true cause of failure for a random sample of $n = 30$ systems. (The true cause of system failure was found by simply observing the minimum life length of the two components in the simulation.) The data were simulated, under Weibull parameters $\theta_1 = 15$,

Table I. Simulated Weibull data for a $J = 2$ component system with masking

System i	System time to failure T_i	Component causing failure	Masking set S_i
1	2·9536	2	{2}
2	3·6558	1	{1}
3	4·6899	2*	{1,2}
4	4·9116	1*	{1,2}
5	5·7715	2	{2}
6	5·8667	1*	{1,2}
7	5·9035	1*	{1,2}
8	5·9473	2	{2}
9	6·0083	2	{2}
10	6·7428	1*	{1,2}
11	6·8419	2	{2}
12	7·1702	1*	{1,2}
13	7·2764	2	{2}
14	7·3056	1*	{1,2}
15	7·3913	2	{2}
16	7·7065	1	{1}
17	7·7968	2*	{1,2}
18	8·6810	2*	{1,2}
19	8·8103	2	{2}
20	9·9686	1*	{1,2}
21	10·0562	1	{1}
22	10·5766	1*	{1,2}
23	10·9384	1*	{1,2}
24	11·9953	1	{1}
25	12·4611	1	{1}
26	13·7200	1	{1}
27	15·5686	2*	{1,2}
28	16·6299	1*	{1,2}
29	17·4406	2	{2}
30	20·1413	2	{2}

$\beta_1 = 2$ and $\theta_2 = 12$ and $\beta_2 = 2{\cdot}2$. We randomly masked approximately 50 per cent of the observations. These observations are denoted by the asterisk.

Under the assumption of no masking, i.e. perfect information on cause of failure, parameter estimates can be obtained through various techniques. For example, using a simple dichotomous search procedure,[4] the MLEs can be easily found as $\hat{\theta}_1 = 13{\cdot}397$, $\hat{\beta}_1 = 2{\cdot}417$, $\hat{\theta}_2 = 14{\cdot}322$ and $\hat{\beta}_2 = 2{\cdot}199$.

Next, we consider the case of masking. For comparative purposes we find the estimates, $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\beta}_1$, $\hat{\beta}_2$ using the method of maximum likelihood and our proposed iterative technique. Both procedures have been programmed in Advanced BASIC, compiled and run on an ITT-XTRA, 8088-based personal computer, running at 10 MHz. The stopping criterion for both procedures is chosen as $\epsilon_k \leq \epsilon$, where

$$\epsilon_k = \sum_{j=1}^{2} \left[\mid \theta_j^{(k)} - \theta_j^{(k-1)} \mid + \mid \beta_j^{(k)} - \beta_j^{(k-1)} \mid \right]$$

where $\theta^{(k)}$ denotes the parameter value at the kth iteration. For both procedures, we set $\epsilon = 0{\cdot}001$.

From Reference 18, and under the assumption of $J = 2$ Weibull components, the likelihood function, L, is given as

$$L = \prod_{i=1}^{n} \sum_{j \in S_i} \left\{ \left(\beta_j/\theta_j^{\beta_j}\right) t_i^{\beta_j - 1} \exp\left[-\left(\frac{t_i}{\theta_j}\right)^{\beta_j}\right] \prod_{\substack{l=1 \\ l \neq j}}^{J} exp \left[-\left(\frac{t_i}{\theta_l}\right)^{\beta_l}\right] \right\} \quad (5)$$

We seek the values of θ_1, θ_2, β_1, β_2 that maximize (5). Using the well-known Hooke–Heeves pattern search algorithm,[18] with starting points set at $\hat{\theta}_1 = 15$, $\hat{\beta} = 2$ and $\hat{\theta}_2 = 12$ and $\hat{\beta}_2 = 2{\cdot}2$, the MLEs are found to be $\hat{\theta}_1 = 16{\cdot}164$, $\hat{\beta}_1 = 2{\cdot}455$, $\hat{\theta}_2 = 13{\cdot}109$, $\hat{\beta}_2 = 1{\cdot}976$. The procedure converged in 13 improvement iterations and required 3·93 min of CPU time.

In contrast, our proposed iterative procedure converged in 10 iterations, after only 0·16 min of CPU time to yield $\hat{\theta}_1 = 14{\cdot}677$, $\hat{\beta}_1 = 2{\cdot}651$, $\hat{\theta}_2 = 11.153$, $\hat{\beta}_2 = 2{\cdot}530$. The parameter estimates are summarized in Table II. A plot of true system reliability and the estimated reliability functions are given in Figure 1. These reliabilities are found as a simple product of the Weibull component reliabilities. From the plot we can see that both the iterative method and the MLE method provide good estimates of system reliability.

4. CONCLUSIONS

We have presented an iterative approach for obtaining parametric estimates of component reliability from masked system life data. The major advantage of the approach is its ability to yield

Table II. Component parameter estimates

	θ_1	β_1	θ_2	β_2	CPU* (min)
Population	15·000	2·000	12·000	2·200	N/A
MLE	16·164	2·455	13·109	1·976	3·93
Iterative	14·677	2·651	11·153	2·530	0·16

*Executed on ITT-XTRA, 8088 based PC, running at 10 MHz.

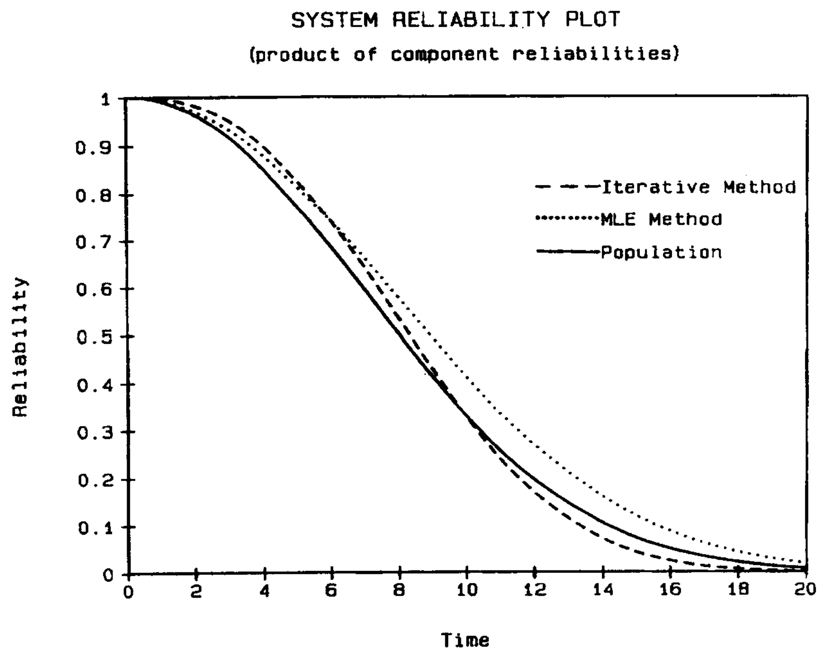

Figure 1. System reliability plot (product of component reliabilities): ----- iterative method; ····· MLE method; ——— population

good estimates with much less computation than the method of maximum likelihood. For our small numerical example we find good Weibull parameter estimates with only 4 per cent of the computation time required to find comparable MLEs. This advantage would become even more important in the analysis of complex systems of many components. For such systems, the method of maximum likelihood would require the maximization of a non-linear function over a large parameter space. For example, a system of 100 components, each with Weibull life lengths, would require the maximization of a function with 200 parameters. The numerical techniques required to perform such a task would be computationally intensive, cumbersome to apply and require good starting points to guarantee convergence. If MLEs are required, our iterative approach could also be used as an inexpensive means of obtaining good starting points. As such, our approach can serve as a useful alternative, or addition, to the ML procedure for obtaining component reliability from masked system life test data.

REFERENCES

1. J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1982.
2. H.A. David and M.L. Moeschberger, *Theory of Competing Risks*, Griffin, London, 1978.
3. J.D. Kalbfleisch and R.L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley, New York, 1980.
4. W. Nelson, *Applied Life Data Analysis*, Wiley, New York, 1982.
5. J.S. Usher, 'Estimating component reliabilities from incomplete accelerated life test data', *Unpublished Ph.D. Dissertation*, Department of Industrial Engineering, North Carolina State University, 1987.
6. M. Miyakawa, 'Analysis of incomplete data in competing risks model', *IEEE Trans. Reliab.*, **R-33**, (4), 293–296 (1984).
7. E.L. Kaplan and P. Meier, 'Nonparametric estimation from incomplete observations', *J. Am. Stat. Asoc.*, **53**, 457–481 (1958).
8. F.M. Guess, J.S. Usher and T. J. Hodgson, 'Estimating system and component reliabilities under partial information on the cause of failure', *Technical Report No. 136*, Department of Statistics, University of South Carolina, 1987.
9. J.S. Usher and T.J. Hodgson, 'Maximum likelihood analysis of component reliability using masked system life test data', *IEEE Trans. Reliab.*, **37**, (5), 550–555 (1988).
10. Symbolics Incorporated, *MACSYMA Reference Manual: Version Ten*, Volume 1, Massachusetts Institute of Technology, Cambridge, 1984.
11. G.V. Reklaitis, A. Ravindran and K.M. Ragsdell, *Engineering Optimization: Methods and Applications*, Wiley, New York, 1983.
12. N. Mantel and M. Myers, 'Problems of convergence of maximum likelihood iterative procedures in multiparameter situations', *J. Am. Stat. Assoc.*, **66**, (335), 484–491 (1971).
13. F. Jensen and N.E. Petersen, *Burn-In: An Engineering Approach to the Design and Analysis of Burn-In Procedures*, Wiley, New York, 1982.
14. D.R. Cox, 'Some remarks on the role in statistics of graphical procedures', *Appl. Stat.*, **27**, 4–9 (1978).
15. B. Bergman and B. Klefsjö, 'Burn-in models and TTT-transforms', *Qual. Reliab. Eng.*, **1**, 125–130 (1985).
16. B. Klefsjö, 'TTT-plotting: a tool for both theoretical and practical problems', to appear in *International Research Conference in Reliability Proceedings*, Columbia, Missouri, 1988.
17. B.W. Turnbull, 'Nonparametric estimation of a survivorship function with doubly censored data', *J. Am. Stat. Assoc.*, **69**, 169–173 (1974).
18. R. Hooke and T.A. Jeeves, 'Direct search solution of numerical and statistical problems', *J. Assoc. Comp. Machines*, **8**, 212–229 (1961).

Authors' biographies

John S. Usher is an Assistant Professor of Industrial Engineering at the University of Louisville in Louisville, Kentucky. He received his BS and Master of Engineering in IE from the University of Louisville in 1980–1981. He earned his Ph.D. in IE from North Carolina State University in 1987. While working on this degree he was employed by IBM as a reliability engineer. His current research interests are in the areas of quality and reliability especially as applied to the design and manufacture of electronic systems. He is a member of ASQC, IEEE, IIE, ORSA and Tau Beta Pi Honorary Engineering Society.

Frank M. Guess is an Associate Professor of Statistics at the University of Tennessee. His research for this paper was supported and performed while he was an Assistant Professor at the University of South Carolina's Department of Statistics. He received his Ph.D. in Statistics from Florida State University in 1984. Also, in 1984, he was selected for the Ralph A. Bradley Award for outstanding achievement for a Ph.D. student from Florida State University's Department of Statistics. He has published on reliability, censored data and non-parametric statistics. He is a member of the American Statistical Association and the Institute of Mathematical Statistics, and is an associate member of the Institute of Electrical and Electronics Engineers. Dr. Guess has had research grants from the Air Force Office of Scientific Research, the Army Research Office and IBM.