Weibull Component Reliability-Prediction in the Presence of Masked Data

John S. Usher, Member IEEE
University of Louisville, Louisville

Key Words — **Component-reliability prediction, Maximum likelihood estimation, Masked data**

Summary & Conclusions — **Analysts are often interested in obtaining component reliabilities by analyzing system-life data. This is generally done by making a *series*-system assumption and applying a competing-risks model. These estimates are useful because they reflect component reliabilities after assembly into an operational system under true operating conditions. The fact that most new systems under development contain a large proportion of old technology also supports the approach. In practice, however, this type of analysis is often confounded by the problem of *masking* (the exact cause of system failure is unknown). This paper derives a likelihood function for the *masked*-data case and presents an iterative procedure (IMLEP) for finding maximum likelihood estimates and confidence intervals of Weibull component life-distribution parameters. The approach is illustrated with a simple numerical example.**

1. INTRODUCTION

Acronyms [1]

CI (symmetric) s-confidence interval
MLE maximum likelihood estimate
IMLEP iterative MLE procedure (in this paper).

Estimates of component reliability are often obtained through the analysis of system life data. The resulting estimates are extremely useful because they reflect the reliability of components after their assembly into an operational system, as opposed to estimates that are obtained from individual component-life tests which might not reflect true component operating conditions. The estimates account for the many degrading effects that might have been introduced during system manufacturing, assembly, distribution, and installation. The use of these component-reliability estimates then enables analysts to develop more-accurate reliability predictions of new configurations of these same component types.

Component reliability from system-life data is generally estimated by making a *series* [2] system assumption and applying a competing-risks model. The observable quantities of interest are system-life (failure or censoring time) and the exact component causing failure. In practice, however, this type of analysis is often confounded by the problem of *masking* (the exact cause of system failure is unknown). This occurs frequently in complex-systems data where the exact cause of failure can be isolated only to some subset of components, such as a circuit card containing many individual components. The resulting quantities observed are then the system-life, and partial information on the cause of failure.

Under the simplifying assumption that components have exponentially distributed life, Miyakawa [5] considers a 2-component *series* system and derives closed-form expressions for the MLE when some of the sample observations are *masked*. Under the same exponential assumption, [8] extends the Miyakawa results to a 3-component system. They find that, in all but several special cases, closed-form MLE are intractable; they propose a simple iterative solution procedure. Ref [3] extends & clarifies the derivation of the likelihood in the *masked* data case under the assumption that *masking* is s-independent of the true failure-cause. Ref [4] investigates the effect of s-dependency between the *masking* set and the true cause of failure. Doganaksoy [2] presents various means of finding CI for the 3-component, s-independent *masking*, exponential-life case. Baxter [1] proposes both MLE & Bayes procedures for estimating component-life distributions from block-censored data, under the assumption that prior knowledge of the component-failure propensities exists.

Little work, however, has been done thus far to find MLE & CI for Weibull components from *masked* data because of the intractable nature of the resulting likelihood function. Ref [7] present an iterative procedure, analogous to traditional probability plotting, for finding Weibull parameter estimates. However, the procedure is cumbersome, and the resulting estimates, though close to MLE, do not allow for direct evaluation of CI.

This paper develops a likelihood function for *masked*-data and presents an IMLEP for finding MLE & CI of Weibull component-life distribution parameters. The approach is illustrated with a simple numerical example.

Notation

n	number of s-identical *series*-systems observed
J	number of different component types in each system
θ_j, β_j	[scale, shape] Weibull parameter for type-j component life
Q_j	number of type-j components in the system
$f_j(t)$, $R_j(t)$	[pdf, Sf] for a type-j component life
$h_j(t)$	hazard rate for $f_j(t)$
T_i	time to failure of system i

[1]The singular & plural of an acronym are always spelled the same.

[2]The terms, *series & parallel* are used in their logic-diagram sense, irrespective of the schematic-diagram or physical-layout.

$S_{i,j}$ number of type-j components suspected of causing failure in system i
L likelihood function
$\mathcal{L}$ ln(L)
$\mathcal{I}(\cdot)$ indicator function: $\mathcal{I}(\text{True})=1$, $\mathcal{I}(\text{False})=0$
δ_i $\mathcal{I}$(system-i fails)
$\hat{}$ implies: MLE
$^{(m)}$ implies: iteration m
$\sum_i$ sum over all i from 1 to n
$\sum_j$ sum over all j from 1 to J.

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

2. MASKED-DATA LIKELIHOOD

Assumptions

1. System failure times, T_i, are s-independent.
2. Type-j components have i.i.d. Weibull distributed lives. The Weibull assumption is not necessary, but is used in this paper.
3. *Masking* is s-independent of the cause of failure.
4. Either the system fails, or test time is censored.

2.1 Likelihood Function

A sample of n s-identical systems, each containing J types of components in *series*, is placed on test for some time. For each system, a failure-time or censoring-time is observed. For systems that fail, an attempt is made to isolate the cause of failure. Depending on the diagnostic effort/luck, the failure-cause information could consist of:

- exact component causing failure,
- a subset of components, one of which is the cause of failure.

Thus the observed quantities on each system are: t_i, δ_i, $S_{i,j}$. There are 3 cases:

- $\delta_i=0$. $S_{i,j}=0$ for all j. No components are suspected of causing failure, because system i did not fail during the period of observation.
- $\delta_i=1$, and the failure-cause is known to be a type-j component. $S_{i,j}=1$ and $S_{i,k}=0$ for $k\neq j$.
- $\delta_i=1$, and the failure-cause is isolated to a subset of components. $S_{i,j} \leq Q_j$, for $1\leq j\leq J$. ◄

Use results from [7] (along with some algebra):

$$\mathcal{L}_i = \delta_i \cdot \ln\left(\sum_j S_{i,j}\cdot h_j(t_i)\right) + \sum_j Q_j\cdot \ln(R_j(t_i)). \quad (1)$$

If $\delta_i=0$, then $S_{i,j}=0$ over all j; and there is no need for the δ_i. Therefore, δ_i is dropped.

$$\mathcal{L} = \sum_i \left[\ln\left(\sum_j S_{i,j}\cdot h_j(t_i)\right) + \sum_j Q_j\cdot \ln(R_j(t_i))\right]. \quad (2)$$

From assumption #2,

$$h_j(t) = (\beta_j/\theta_j)\cdot(t/\theta_j)^{\beta_j-1},$$

$$R_j(t) = \exp(-(t/\theta_j)^{\beta_j}),$$

$$\mathcal{L} = \sum_i \left[\ln\left(\sum_j S_{i,j}\cdot(\beta_j/\theta_j)\cdot(t_i/\theta_j)^{\beta_j-1}\right) - \sum_j Q_j\cdot(t_i/\theta_j)^{\beta_j}\right]. \quad (3)$$

2.2 Maximizing $\mathcal{L}$

When no *masking* is present in the system-life data, $\mathcal{L}$ can be maximized separably over each j. Failures due to one component simply yield censoring times for the other non-failed components in the system, which in effect creates J sets of s-independent component-life data. In the Weibull case, the problem of maximizing over (θ_j,β_j) for all j, can be performed with just 2 parameters at a time, rather than over all parameters simultaneously. In such case, the likelihood function for each of the J component types is:

$$\mathcal{L} = \sum_i [\ln(S_{i,j}\cdot(\beta_j/\theta_j)\cdot(t_i/\theta_j)^{\beta_j-1}) - Q_j\cdot(t_i/\theta_j)^{\beta_j}]. \quad (4)$$

The numerical methods required for maximizing (4) with respect to (θ_j,β_j) are well known and straightforward.

When *masked* data are present in the sample, $\mathcal{L}$ is no longer separable, and the $\mathcal{L}_j$ take on a different form. From (3),

$$\mathcal{L}_j = \sum_i \left[\ln\left(\sum_{j'} S_{i,j'}\cdot(\beta_{j'}/\theta_{j'})\cdot(t_i/\theta_{j'})^{\beta_{j'}-1}\right) - Q_j\cdot(t_i/\theta_j)^{\beta_j}\right]. \quad (5)$$

The first term in the main sum forces the maximization to proceed over all component parameters simultaneously. For realistic systems with many component types, this can be very difficult.

3. ITERATIVE MLE APPROACH

To solve the problem of obtaining MLE of Weibull parameters for each component-type in a *series* system, using *masked* system-life data, we propose an IMLEP, along with a method for obtaining asymptotic (approximate) CI on the estimates.

3.1 IMLEP

Section 2 shows that the MLE of the component-life distribution parameters in the non-*masked* data case can be found separably — by maximizing the likelihood function 1 component-type at a time. IMLEP works on the same principle for the *masked* data. The procedure:

- treats *masked* observations as censoring times;
- finds a starting point MLE for each component type;
- maximizes the likelihood function for each component type separably — using the MLE from previous iterations when *masked* data are encountered.

In this way the procedure maximizes the likelihood, 1 component-type (2 parameters) at a time, rather than over all component-types simultaneously. This greatly reduces the computational burden, thus yielding MLE quickly & efficiently.

Algorithm

1. Create a *revised* data-set where all *masked* system-failures are treated as censoring times; *ie*, if system-i failed but the cause is unknown, then treat that observation as a censoring time.

2. Find the starting point (iteration 0), $\hat{\theta}_j^{(0)}$, $\hat{\beta}_j^{(0)}$ for all j from the non-*masked* data set, by maximizing (4) over all j separably. This is a common case of MLE of component reliability from non-*masked* system-life data.

3. Find improved estimates, $\hat{\theta}_j^{(m)}$, $\hat{\beta}_j^{(m)}$, for all j by maximizing the $\mathcal{L}_j$ from (5) with respect to component type-j. For the observations where the failure-cause is *masked*, use $\hat{\theta}_j^{(m-1)}$, $\hat{\beta}_j^{(m-1)}$ for $k \neq j$. That is, for the other components *masked* with component type j, use the parameter estimates from the previous iteration.

4. Compute $\mathcal{L}^{(m)}$ using (3).

5. Check to see if the procedure has converged (one way is to compare the likelihood values from consecutive iterations). If not converged, go to step 3.

6. The answer is $\hat{\theta}_j^{(m)}$, $\hat{\beta}_j^{(m)}$. ◄

For the Weibull case, there must be at least 2 known-cause failures of a component type to obtain parameter estimates. If all data are *masked* for a component-type, the procedure fails and no estimates for that component type can be obtained.

3.2 *s*-Confidence Intervals

The MLE in section 3.1 are point estimates. Interval estimates should be computed. This can be done under certain conditions, usually met in practice, where the MLE are asymptotically *s*-normally distributed. In this situation, we need to estimate the variance of each parameter.

We compute the local estimate of the covariance matrix for the parameters (inverse of the Fisher information matrix), evaluated at $\hat{\theta}$, $\hat{\beta}$. The matrix elements, though lengthy & tedious to derive, can be found in closed form. Alternatively one can estimate them using a suitable numerical method. The diagonal terms of the variance-covariance matrix are the local estimates of the variance terms, $\mathrm{Var}\{\hat{\theta}_j\}$, $\mathrm{Var}\{\hat{\beta}_j\}$. To ensure that the CI cover only positive values for the parameters, we assume that the sample size is large enough that $\ln(\hat{\theta}_j)$, $\ln(\hat{\beta}_j)$ are approximately *s*-normally distributed. Under this assumption Nelson [6] gives approximate positive upper & lower α CI on θ_j, β_j.

Table 1. Weibull Data for a 2-Component System
[*denotes masked observation
i = system
k = component that causes failure]

#i	T_i	#k	#i	T_i	#k
1	2.954	2	16	7.707	1
2	3.656	1	17	7.797	*
3	4.690	*	18	8.681	*
4	4.912	*	19	8.810	2
5	5.772	2	20	9.969	*
6	5.867	*	21	10.056	1
7	5.904	*	22	10.577	*
8	5.947	2	23	10.938	*
9	6.008	2	24	11.995	1
10	6.743	*	25	12.461	1
11	6.842	2	26	13.720	1
12	7.170	*	27	15.569	*
13	7.276	2	28	16.630	*
14	7.306	*	29	17.441	2
15	7.391	2	30	20.141	2

Table 2. Path of the Iterative Approach
[for the example data-set
Iteration #0 is the starting point]

Iteration	Component	$\hat{\beta}$	$\hat{\theta}$	$\mathcal{L}$ Value
0	1	2.676	19.278	-98.86739
	2	2.133	16.853	
1	1	2.496	14.029	-94.20467
	2	2.183	12.851	
2	1	2.539	15.249	-93.98109
	2	2.211	12.570	
3	1	2.539	15.368	-93.97768
	2	2.187	12.540	
4	1	2.543	15.368	-93.97766
	2	2.187	12.540	
5	1	2.543	15.371	-93.97765
	2	2.187	12.540	

4. NUMERICAL EXAMPLE[3]

The data are the simple set of 2-component *masked* data in [7]. Table 1 shows:

$n=30$, $J=2$, $Q_1=1$, $Q_2=1$;

14 out of the 30 observations are *masked*.

[3] The number of significant figures is not intended to imply any accuracy in the estimates, but to illustrate the arithmetic.

Using an exhaustive search method, [7] uses its procedure to find MLE in this simple 4-parameter case. Such a search, over all parameters simultaneously, would not be feasible for larger problems. The MLE in [7] are incorrect. The correct values are:

$$\hat{\beta}_1 = 2.543,\ \hat{\theta}_1 = 15.372;$$

$$\hat{\beta}_2 = 2.187,\ \hat{\theta}_2 = 12.540.$$

IMLEP, when applied to the same data set, converged, along the path shown in table 2, in 5 iterations. The procedure required only 12 seconds, on a 486-based, 50-MHz PC. Table 3 shows the final MLE and approximate 90% CI.

Table 3. MLE and *s*-Confidence Intervals
[for the example data-set
j = component]

j	$\hat{\beta}_j$	$\hat{\theta}_j$	90% CI for β_j	90% CI for θ_j
1	2.543	15.372	[1.732, 3.734]	[12.030, 19.642]
2	2.187	12.540	[1.623, 2.947]	[10.261, 15.327]

ACKNOWLEDGMENT

This research was supported, in part by AT&T Bell Laboratories, the Telecommunications Research Center at the University of Louisville, and the US Naval Surface Warfare Center at Crane. I am pleased to thank Dr. Ralph Evans, Dr. Michael Tortorella, Dr. Chun Kin Chan, and Dr. Lawrence Baxter, for their suggestions and thoughtful discussions regarding various aspects of this work.

REFERENCES

[1] L. Baxter, "Estimation subject to block censoring", *IEEE Trans. Reliability*, vol 44, 1995 Sep, pp 489-495.

[2] N. Doganaksoy, "Interval estimation from censored & masked system-failure data", *IEEE Trans. Reliability*, vol 40, 1991 Aug, pp 280-285.

[3] F.M. Guess, J.S. Usher, T.J. Hodgson, "Estimating system and component reliabilities under partial information on the cause of failure", *J. Statistical Planning & Inference*, vol 29, 1991, pp 75-85.

[4] D.K.J. Lin, F.M. Guess, "System life data analysis with dependent partial knowledge on the exact cause of system failure", *4th Int'l Research Conf. Reliability*, 1991 June; Columbia, Missouri.

[5] M. Miyakawa, "Analysis of incomplete data in competing risks model", *IEEE Trans. Reliability*, vol R-33, 1984 Oct, pp 293-296.

[6] W. Nelson, *Applied Life Data Analysis*, 1982; John Wiley & Sons.

[7] J.S. Usher, F.M. Guess, "An iterative approach for estimating component reliability from masked system life test data", *Quality & Reliability Eng'g Int'l*, vol 5, 1989, pp 257-261.

[8] J.S. Usher, T.J. Hodgson, "Maximum likelihood estimation of component reliability using masked system life test data", *IEEE Trans. Reliability*, vol 37, 1988 Dec, pp 550-555.

AUTHOR

Dr. John S. Usher, PE; Dept. of Industrial Eng'g; Univ. of Louisville; Louisville, Kentucky 40292 USA. Internet (e-mail): jsushe01@ulkyvm.louisville.edu

John S. Usher is an Associate Professor of Industrial Engineering at the University of Louisville. He received his BS (1980) and MEng (1981) in Industrial Engineering from the University of Louisville. He received his PhD (1987) in Industrial Engineering from North Carolina State University. While completing this degree he was employed by IBM in the Research Triangle Park as a Reliability Engineer. He has published in journals such as *IEEE Trans. Reliability*, *Quality & Reliability Engineering Int'l*, *Quality Engineering*, and *IIE Transactions*. He has conducted funded research for the US National Science Foundation, IBM, AT&T Bell Labs, US Navy, Quaker Oats Company, KFC, and the National Cooperative Highway Research Program, in system reliability and quality control. He is a member of IIE, ASQC, IEEE, and Tau Beta Pi, and is a Registered Professional Engineer in the State of Kentucky.

Manuscript received 1995 November 11.

Publisher Item Identifier S 0018-9529(96)04775-6

◄TR►

MANUSCRIPTS RECEIVED *MANUSCRIPTS RECEIVED* *MANUSCRIPTS RECEIVED* *MANUSCRIPTS RECEIVED*

"Reliability-based scheduling using supportability turn-around time" (D. Linton, S. Khajenoori, et al.), Dr. Darrell G. Linton, PE • Electrical & Computer Engineering Dept. • Univ. of Central Florida, Eng Bldg - Rm 407 • POBox 162450 • Orlando, Florida 32816-2450 • USA. (TR96-051)

"Design of reliability tests for Weibull distributions, indexed by scale & shape parameters" (I. Arizono, et al.), Dr. Ikuo Arizono • Dept. of Industrial Engineering • College of Engineering • Osaka Prefecture Univ. • Sakai, Osaka 593 • JAPAN. (TR96-052)

"A unified view of voting schemes" (B. Parhami), Dr. Behrooz Parhami, Professor • Dept. of Electrical & Computer Engineering • University of California • Santa Barbara, California 93106-9560 • USA. (TR96-053)

"Assessment of the distribution-type for VAX/VMS reboot data" (M. Buckley, D. Siewiorek), Dr. Michael F. Buckley • IBM TJW Res. Ctr. • 30 Saw Mill River Rd • Hawthorne, New York 10532 • USA. (TR96-054)

"Comment on: Bayes reliability growth model during a development testing program" (N. Jain), N. K. Jain • 48, Raman Hostel • Probyn Road, Timarpur • Delhi - 110 054 • INDIA. (TR96-055)

"The Makeham reliability-growth model" (T. Calvo, et al.), Dr. Tomasa Calvo • Dpto. de Matematicas e Informatica • Univ. de las Islas Baleares • Crta Valldemosa, km 7,5 • 07071 Palma de Mallorca • SPAIN. (TR96-056)

"Service-life prediction-models based on stochastic stress and resistance factors" (H. Kim, et al.), Dr. Hag-Min Kim • NREL • 1617 Cole Blvd • Golden, Colorado 80401 • USA. (TR96-057)

"Stochastic stress-strength model of reliability" (A. Zizek), Dr. Adolf Zizek • Bistra • Mestni trg 1 • 2250 Ptuj • SLOVENIA - Yu. (TR96-058)

"Transient solution of Markov models by combining adaptive & standard uniformization" (W. Sanders, et al.), Dr. William H. Sanders • Ctr. Reliable & High-Performance Computing • Coordinated Science Lab, Univ. Illinois • 1308 W. Main St • Urbana, Illinois 61801-2307 • USA. (TR96-059)

"The effect of assumptions on the probability of failure in mechanical stress/strength analysis" (D. Gond), Daniel Gond • Dept. Surete de Fonctionnement, D3E/BQ22 • MATRA Defense • 20-22, rue Dame Grange Rose (BP 150) • 78141 - Velizy cedex • FRANCE. (TR96-060)