Technical note

On reliability modeling and software quality

by A. J. Watkins

This note continues the recent discussion on reliability modeling and its application in software development by Kan in a recent issue of the* IBM Systems Journal. *We focus on the initial stages of a reliability modeling process, as decisions here will often influence the later stages of an analysis.

This note is intended as a contribution to the discussion on reliability modeling and its application in software development, and aims to cover and discuss some of the themes introduced in the recent review paper by Kan.[1] These themes include the use of the two-parameter Weibull distribution as a fundamental part of the modeling process, the estimation from field data of the parameters thus introduced, the interpretation of these results and further assessment possible with the field data, and the potential consequence for subsequent stages of the modeling process.

Our intention is to concentrate on concepts and ideas and on strategies for implementing them; we eschew formulae linked to particular assumptions, since these details are widely available elsewhere. Moreover, our comments here are not intended to be exhaustive but, rather, to guide the interested reader to some recent and relevant references in the reliability and statistical literature; these, in turn, furnish further references that will be of value. As in Kan,[1] we assume that the reader is convinced of the merit of reliability modeling and is now concerned with the details of using such a process for analyzing observable field data to develop and manage software quality procedures. We cover a variety of philosophical, statistical, and computational issues, which may be grouped under the following four headings.

Data collection

It may be argued that the first stage in a reliability modeling process is to consider the circumstances under which the data to be analyzed are collected. For instance, some initial discussion on what constitutes a defect may be necessary before testing starts; Kan[1] also covers briefly the various measures of system use that can be recorded. The testing of a software system may then take the form of a number of users running the system, perhaps working through a series of tests, examples, or tutorials, and reporting some measure of time in use before encountering a system defect. This framework allows for the possibility that a session is concluded before a defect is encountered, so that the datum for analysis can then be regarded as a *censored* observation, with a value equal to the length of the testing session.

We should emphasize that the censored data play an important part in the subsequent analysis. Thus ignoring such information at this stage introduces a potentially large bias to the remainder of the analysis. Leech and Watkins,[2] for example,

give a more detailed account of the importance of including censored data in analyses; they also illustrate the value of a "soft" censored value of system use, relative to that of a "hard" value of time before encountering a defect.

Next, we should note that most statistically based analyses will assume, often implicitly, that the data for analysis are independent observations. A full definition and discussion of statistical independence are beyond the scope of this note, but we remark that, although this assumption is often valid for defect data, it may be brought into question by various system repair regimes. Finally, we should note that, in certain circumstances, it may prove difficult to obtain a full record of all test data for analysis; this is often the case, for example, when testing is performed at more than one site. In such cases, it is often possible to proceed with a partial sample of data, provided that this sample can reasonably be assumed to be representative of the whole population of defects; see, for example, Suzuki.[3] Note that this partial sampling may preclude the possibility of extrapolating the conclusions based on the analysis of data from one center assigned to test a particular part or module of a system.

Model identification and parameter estimation

The framework for reliability modeling outlined by Kan[1] emphasizes certain special cases of the two-parameter Weibull distribution. These cases are the exponential and Rayleigh distributions, in which the shape parameter m of the Weibull distribution takes the value 1 and 2, respectively. This emphasis on special cases reduces the number of parameters to be estimated from data, and an analytical formula for the maximum likelihood estimator of the scale parameter c can be written down. The use of maximum likelihood estimators here enjoys the support of a central pillar of the theory of statistical inference. They are, by definition, the values of the parameters most in accordance with the available observed data, and, as such, are a statistically consistent guide to the true, unknown, values of these parameters. The more general case, in which both parameters are estimated by the method of maximum likelihood, is considered by Kalbfleisch,[4] who outlines a method for reducing the problem to a numerical search for the maximum likelihood estimator of m. Thus, the calculation of such estimators is possible with some experience either in programming or in the use of general spreadsheet packages and need not require access to commercial statistical packages.

The discussion in Kan[1] is largely based on the two-parameter Weibull distribution. Somewhat more generally, the estimation of model parameters may be preceded by model identification procedures—for example, based on the calculation of hazard and cumulative hazard functions—or some checks that a chosen distribution provides an adequate fit to the available data. In certain cases, these considerations may be prompted and guided by previous analyses of similar field data. In other situations, it may be sensible to follow Kan and appeal to the widespread applicability of certain distributions, such as the Weibull, to provide some justification for a given choice.

The framework for reliability modeling outlined by Kan emphasizes certain special cases of the two-parameter Weibull distribution.

Precision in parameter estimation

The calculation of point estimators of model parameters is rarely sufficient. It is usually important to have some guide to the precision, or lack thereof, in these estimates. One option is to calculate confidence intervals for estimates; however, as Kan[1] notes, this option is not always satisfactory, since the relevant, asymptotically valid, formulae do not necessarily provide an accurate guide to what happens in small samples.

One workable alternative, considered in Kalbfleisch,[4] is to base assessments of the precision in estimators around the definition of relative likelihood. This approach is valid for all admissible sample sizes and requires no further assumptions. For the two-parameter Weibull distribution, it allows us to draw contours of equal likelihood centered about the maximum likelihood estimator of a given sample; these contours contain all pair-

ings of (m, c) which are at least the specified percentage as likely as the maximum likelihood estimator. Watkins and Leech[5] discuss details of an algorithm for drawing such contours and consider the interpretations possible for points inside such contours. For instance, we note that these relative likelihood contours may be used to decide whether there is sufficient evidence to accept the hypothesis that the shape parameter is equal to some specified value.

Model extensions

The basic framework considered above can be extended in a number of ways, which may be of use when additional data are available. For instance, it is sometimes possible to classify defects, perhaps according to severity, with an ordinal range from relatively trivial to major. The recording of this classifying variable then allows the use of the framework of accelerated life testing—see, for example, Nelson[6]—to test whether there is any evidence, for instance, for allowing the scale parameters of the Weibull distribution to vary with level of defect. For such procedures, it may be possible to exploit certain algebraic simplifications to reduce the computational effort required; see, for example, Watkins,[7] for a further discussion of analyzing defect data under the assumption that the Weibull scale parameters follow a power-law model.

Conclusions

The brief discussion above has attempted to focus on a number of points raised by Kan.[1] It should, perhaps, now be emphasized here that this discussion is largely supportive of the views in that account, although we have also tried to consider extensions of, and alternatives to, the methods considered there.

We have been chiefly concerned with the initial stages of a reliability modeling process, since actions taken or opportunities missed here will often set the tone for the remainder of the analysis. This analysis, which aims to transform the raw data into usable results and evidence, is unlikely to follow the simple one-dimensional path envisaged by some, yet need not necessarily require the application of complex techniques. More realistically, in most analyses there will be stages at which two, or more, competing explanations need to be considered, and one may well require common sense, an awareness of the assumptions, powers, and limitations underlying statistical methods, and a modicum of perseverance and determination to see the analysis through its trickier moments.

To summarize: reliability modeling rarely follows the ideal or straightforward path from data collection through analysis to final recommendations; further, the number of reasons for deviation is potentially large. However, as Kan[1] emphasizes, the modeling process remains worthwhile, since the endeavors it represents will lead to the insights necessary to produce procedures for ensuring software quality.

Cited references

1. S. H. Kan, "Modeling and Software Development Quality," *IBM Systems Journal* **30**, No. 3, 351–362 (1991).
2. D. J. Leech and A. J. Watkins, "Some Factors Affecting Precision in Estimators of Parameters of Lifetime Distributions," *Reliability Engineering and System Safety* **29**, 249–260 (1990).
3. K. Suzuki, "Estimation of Lifetime Parameters from Incomplete Field Data," *Technometrics* **10**, 261–271 (1985).
4. J. G. Kalbfleisch, *Probability and Statistical Inference II*, Springer-Verlag, Inc., New York (1979).
5. A. J. Watkins and D. J. Leech, "Towards Automatic Assessment of Reliability for Data from a Weibull Distribution," *Reliability Engineering and System Safety* **24**, 343–350 (1989).
6. W. Nelson, *Accelerated Testing: Statistical Models, Data Analysis and Test Plans*, John Wiley & Sons, Inc., New York (1990).
7. A. J. Watkins, "On the Analysis of Accelerated Life-Testing Experiments," *IEEE Transactions on Reliability* **40**, 98–101 (1991).

Accepted for publication August 1, 1993.

Alan J. Watkins *European Business Management School, University College, Singleton Park, Swansea SA2 8PP, United Kingdom.* Dr. Watkins received B.Sc., M.Sc., and Ph.D. degrees from the University of Leeds in the United Kingdom. He teaches statistical inference at the University College of Swansea. His research interests include the analysis of reliability data. He is a Fellow of the Royal Statistical Society and has recently been associated with Scientific Software-Intercomp (U.K.) Ltd. in Egham, Surrey.

Reprint Order No. G321-5539.