Analysis of a Simple Debugging Model

Adrian E. Raftery

Applied Statistics, Vol. 37, No. 1 (1988), 12-22.

Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281988%2937%3A1%3C12%3AAOASDM%3E2.0.CO%3B2-K

Appl. Statist. (1988)
37, *No.* 1, *pp.* 12–22

Analysis of a Simple Debugging Model

By ADRIAN E. RAFTERY†

University of Washington, USA

[Received August 1986. Revised August 1987]

SUMMARY

A system has an unknown number of faults. Each fault causes a failure of the system, and is then located and removed. The failure times are independent exponential random variables with common mean. A Bayesian analysis of this model is presented, with emphasis on the situation where vague prior knowledge is represented by limiting, improper, prior forms. This provides a test for reliability growth, estimates of the number of faults, an evaluation of current system reliability, a prediction of the time to full debugging, and a model checking procedure. Three examples are given.

Keywords: Bayes factor; Improper prior; Non-homogeneous Poisson process; Simulation-based model checking; Reliability growth; Software reliability

1. Introduction

Consider a system with an unknown number of faults N. Each fault causes a failure of the system, and is then located and removed. The times at which the N failures occur are assumed to be independent exponential random variables with common mean β^{-1}. Early analyses of this model were carried out by Bazovsky (1961, chap. 8) and Cozzolino (1968). It has been much studied in the software reliability literature, where it is often attributed to Jelinski and Moranda (1972).

Problems of interest include finding the probability that all the faults have been removed, estimating the number of remaining faults, evaluating the current reliability of the system, and predicting the time to full debugging. Another question is whether the system's failure rate is decreasing, as the model predicts. Littlewood and Verrall (1981) and Ascher and Feingold (1984, pp. 110–111) emphasised the need to test this assumption, and described software reliability data sets in which the failure rate increased over long periods of time.

My aim here is to develop methods which can provide solutions to such problems, as well as a framework for making decisions, such as when to stop debugging. My approach is Bayesian, with an emphasis on the situation where vague prior knowledge about the model parameters is represented by limiting, improper, prior forms.

Much previous research has focussed on point estimation of N (Blumenthal and Marcus, 1975; Joe and Reid, 1985; Watson and Blumenthal, 1980). My results suggest that, for this problem, point estimation is not very useful. Generally speaking, point estimation may be helpful in situations where, having seen the data, one can reasonably act as if one knew the unknown parameter exactly. Here, the posterior distribution of N tends to be highly asymmetric, and often quite diffuse, even when the data set

† *Address for correspondence*: Department of Statistics GN-22, University of Washington, Seattle, WA 98195, USA.

 0035–9254/88/37012 $2.00

is fairly large. Thus, one should look at the overall uncertainty contained in the complete posterior distribution.

Point estimation of N also presents difficult technical problems. For example, the *maximum likelihood estimator* (*MLE*) of N can be infinite with substantial probability. Indeed, Goudie and Goldie (1981), who studied the case where the observed number of failures is specified in advance, concluded that all standard non-Bayesian techniques are liable to fail. My approach does yield estimators of N; these are described and compared with other estimators in Section 3.

Forman and Singpurwalla (1977) proposed a stopping rule for debugging the system based on how close the observed likelihood is to a large-sample approximation; their aim was to ascertain whether the system had been fully debugged. Their data are reanalysed in Section 6. I hope that this paper provides a more precise answer to that question, as well as the basis for a more general stopping rule, which explicitly takes into account the costs associated with the various possible outcomes.

2. Testing for Reliability Growth

I assume that the system has been observed for the period $[0, T]$, during which n failures have occurred at times $t=(t_1, \ldots, t_n)$, where $n>1$. I consider the problem of comparing the model described in Section 1 with the constant rate Poisson process $M_0: \lambda(s)=\mu$, where $\lambda(s)$ is the rate of occurrence of failures at time s.

I assume that the sample space consists of systems, rather than of replications of the debugging process for the same system. N is thus a random variable, and I assume that it has a Poisson distribution. It then follows that the model is equivalent to a non-homogeneous Poisson process with rate function

$$M_1: \lambda(s)=\rho \exp(-\beta s), \tag{2.1}$$

where $\rho>0$ and $E[N]=\rho/\beta$ (Scholz, 1986). Non-Bayesian statistical analysis of this process has been considered by Cox and Lewis (1966), Lewis (1972), MacLean (1974), and Berman (1981).

The comparison of M_0 with M_1 is based on the *Bayes factor*, or ratio of posterior to prior odds for M_0 against M_1,

$$B_{01}=p(t\mid M_0)/p(t\mid M_1) \tag{2.2}$$

the ratio of the marginal likelihoods. In (2.2)

$$p(t\mid M_0)=\int_0^\infty p(t\mid \mu, M_0)p(\mu\mid M_0)\mathrm{d}\mu$$

$$p(t\mid M_1)=\int_0^\infty \int_0^\infty p(t\mid \rho, \beta, M_1)p(\rho, \beta\mid M_1)\mathrm{d}\rho\mathrm{d}\beta.$$

If the priors $p(\mu\mid M_0)$ and $p(\rho, \beta\mid M_1)$ are proper, (2.2) can be evaluated, if necessary using numerical integration.

I now develop an expression for B_{01} in the situation where vague prior knowledge is represented by limiting, improper, prior forms. I use the standard vague prior for μ

$$p(\mu\mid M_0)=c_0\mu^{-1} \tag{2.3}$$

(Jaynes, 1968). The likelihood for M_1 is

$$p(t \mid \rho, \beta, M_1) = \rho^n \exp\{-\beta S - \rho\beta^{-1}(1 - \exp(-\beta T))\},$$

where $S = \Sigma_{i=1}^n t_i$. This is an exponential family likelihood, for which a natural family of conjugate prior densities is

$$p(\rho, \beta \mid M_1) \propto \rho^{k_1} \exp\{-k_2\beta - k_3\rho\beta^{-1}(1 - \exp(-\beta T))\}. \tag{2.4}$$

Akman and Raftery (1986b) have shown that the unique prior of the form (2.4) for which B_{01} is invariant to scale changes in the time variable and independent of the stopping time T is

$$p(\rho, \beta \mid M_1) = c_1 \rho^{-2}. \tag{2.5}$$

However, the Bayes factor calculated using the improper priors (2.3) and (2.5) involves an arbitrary, undefined, multiplicative constant c_0/c_1. Akman and Raftery (1986b) have shown how this may be assigned using the minimal imaginary experiment idea of Spiegelhalter and Smith (1982). This consists of imagining that an experiment is performed which yields the smallest possible data set permitting a comparison of M_0 and M_1, and provides maximum possible support for M_0. It is then argued that the resulting Bayes factor should be only slightly greater than one. Raftery and Akman (1986) have applied this approach to the change-point Poisson process; their results may be compared with the non-Bayesian solution of Akman and Raftery (1986a). This approach has also been applied to log-linear models for contingency tables by Raftery (1986).

For the present problem, the appropriate minimal imaginary experiment consists of two failures occurring at the end of the observation period. Then the procedure outlined yields $c_0/c_1 = \pi^2/6 - 1 = 0.6449$, and

$$B_{01} = 0.6449(n-1)\left[\int_0^\infty \exp(-Ry)\{y/(1-\exp(-y))\}^{n-1} \mathrm{d}y\right]^{-1}, \tag{2.6}$$

where $R = S/T$. Strictly speaking, any value of B_{01} less than one indicates that the data provide evidence for reliability growth. However, as a rough order of magnitude interpretation, Jeffreys (1961, Appendix B) has suggested that the evidence should be regarded as strong only if $B_{01} < 10^{-1}$, and as decisive only if $B_{01} < 10^{-2}$.

3. Estimating the Number of Faults in the System

The framework developed in Section 2 is used. The results in this section and the next one are conditional on M_1. If the priors are proper, standard Bayesian inference is straightforward (Akman, 1985; Jewell, 1985; Langberg and Singpurwalla, 1985; Meinhold and Singpurwalla, 1983).

It follows from (2.5) that

$$p(N, \beta) = \int_0^\infty p(N \mid \rho, \beta) p(\rho, \beta) \mathrm{d}\rho \propto \{N(N-1)\}^{-1}\beta^{-1}. \tag{3.1}$$

Also,

$$p(t \mid N, \beta) = \beta^n \exp\{-\beta T(R + N - n)\} N!/(N-n)! \tag{3.2}$$

Combining (3.1) with (3.2), and integrating over β yields the posterior distribution of the number of remaining faults $M = N - n$,

$$p(M \mid t) \propto (M + R)^{-n} \prod_{i=1}^{n-2} (M + i). \tag{3.3}$$

The probability that the system has been fully debugged is simply $P[M = 0 \mid t]$. Interval estimates of N, such as ***highest posterior density*** (***HPD***) regions, or one-sided intervals, may readily be found from (3.3).

In many applications, estimation of N is an intermediate step in the solution of other problems. However, if a point estimator of N is required, it may be obtained from (3.3) by combining it with an appropriate loss function. The posterior mode, $\hat{N}_{mod}$, is the estimator which corresponds to a zero-one loss function, so that, if the appropriate loss function is bounded, $\hat{N}_{mod}$ may well be a good approximation. The posterior median, $\hat{N}_{med}$ (found by linear interpolation), is an estimator which corresponds to one unbounded loss function, and is also a useful summary of the posterior distribution.

Other point estimators of N which are always finite include Blumenthal and Marcus's (1975) modified maximum likelihood estimator N^*, and Joe and Reid's (1985) harmonic mean estimator $\tilde{N}$. Watson and Blumenthal (1980) considered three other estimators, but their performance in a simulation study was very similar to that of N^*, so I do not consider them further here.

The four estimators, $\hat{N}_{mod}$, $\hat{N}_{med}$, N^*, and $\tilde{N}$, were compared in a small simulation study whose results are summarised in Table 1. β was fixed at 1.0, and T was set equal to $-\log(1 - Q)$, where N and Q were fixed at the values shown. Q is thus the probability of a randomly chosen bug causing the system to fail before time T. The results are conditional on $n > 1$.

The most striking feature of Table 1 is how badly all four estimators performed; none did much better than an estimator which is identically equal to n. Also, no one estimator was uniformly better than any other. For $Q = 0.9$, corresponding to the situation where the system is close to being fully debugged, $\hat{N}_{mod}$ performed best, while for $Q = 0.25$ $\tilde{N}$ performed best. These results suggest that it would be better to report the full posterior distribution (3.3), or some of its salient characteristics, than any one point estimator. Example 1, in Section 6, illustrates this point empirically.

4. Estimating System Reliability and Time to Final Debugging

The reliability of the system is the probability that it operates without failure for a further, specified, period of length x, say. This is equal to $P[X > x \mid t]$, where

TABLE 1

Root mean squared error of point estimators of N†

			Root Mean Squared Error of Estimator			
Q	N	$E[M]$	$\hat{N}_{mod}$	$\hat{N}_{med}$	N^*	$\tilde{N}$
0.9	10	1.0	2.1	3.8	2.3	4.3
0.9	100	10.0	11.3	16.1	11.8	15.9
0.25	10	7.5	7.2	6.6	7.1	5.9
0.25	100	75.0	63.1	46.4	57.3	45.8

† 200 simulations with $n > 1$ for each value of (Q, N).

$X = t_{n+1} - T$ is the time to the next failure. Now, $X = \infty$ if $M = 0$, and $P[X > x \mid M, \beta] = \exp(-M\beta x)$ $(M \geqslant 1)$, so that

$$P[X > x \mid t] = P[M = 0 \mid t] + \sum_{M=1}^{\infty} \int_0^{\infty} \exp(-M\beta x) p(M, \beta \mid t) \mathrm{d}\beta$$
$$= P[M = 0 \mid t] + \sum_{M=1}^{\infty} p(M \mid t)\{1 + M(M+R)^{-1}(x/T)\}^{-n}.$$

$E[X \mid t]$ is always infinite, but we can calculate

$$E[X \mid t, M \geqslant 1] = T(n-1)^{-1}\{1 + R(1 - P[M = 0 \mid t])^{-1} \sum_{M=1}^{\infty} M^{-1} p(M \mid t)\}.$$

The time to final debugging of the system is $Z = t_N - T$. $Z = 0$ if $M = 0$, while if $M \geqslant 1$, Z is the maximum of M independent exponential random variables with mean β^{-1}. Thus

$$P[Z \leqslant z \mid t] = P[M = 0 \mid t] + \sum_{M=1}^{\infty} p(M \mid t) \sum_{k=0}^{M} (-1)^k \binom{M}{k} \{1 + kz/T(M+R)\}^{-n}$$

and

$$E[Z \mid t] = T(n-1)^{-1} \sum_{M=1}^{\infty} p(M \mid t)(M+R) \sum_{k=1}^{M} (-1)^{k+1} \binom{M}{k} k^{-1}.$$

5. Model Checking

Diagnostic checking of the model may be based on a comparison of the evolution over time of the cumulative number of failures with that predicted by the model. Discrepancies may be assessed by calculating the distribution of the cumulative number of failures under the model. I know of no exact distributional results which enable one to do this. Asymptotic approximations are not available because the total number of failures over all time is amost surely finite, so that the parameters cannot be consistently estimated. It is analytically possible, but computationally demanding, to calculate the posterior distribution of the cumulative number of failures.

I use an idea of Ripley (1977), and compare the observed evolution with those of several data sets simulated from the model. The observed curve, and the envelope of the simulated curves, are plotted on the same graph. Ripley (1977) used point estimates of the model parameters in his simulations, arguing that the effect of ignoring the imprecision of these estimates is small if the number of parameters is small compared to the number of data points. Here, the uncertainty about the parameters can be great, even when the number of data points is large. Thus Ripley's (1977) approach may, in this context, lead to simulated bands which are too narrow.

The present approach provides a simple way of incorporating uncertainty about the parameters. To simulate a data set, one first generates values of the parameters from their posterior distribution, and then proceeds as before. This seems quite a general prescription, applicable beyond the present context.

It can be simply implemented, as follows. First, generate $N = M + n$ from the posterior distribution of M in (3.3). Then, generate β from the conditional distribution $(\beta \mid M, t) \sim \text{Gamma}(n, T(R+M))$ obtained from (3.1) and (3.2). Next, generate N independent exponential random variables with common mean β^{-1}. Last, order those which are less than or equal to T, discarding those which are greater than T. I found it worthwhile to rescale the data in such a way that the period of observation is [0, 1], rather than [0, T]. The amount of computing time required is small. Examples are given in Section 6.

This procedure, like that of Ripley (1977), is for informal diagnostic checking, rather than formal goodness-of-fit testing. If it reveals inadequacies in the model, one should, in principle, build a more elaborate model, and compare it with the present one using ideas similar to those which underlay Section 2.

6. Examples

I now reanalyse three data sets consisting of the failure times of software during the production and testing phases. In each case, the purpose of collecting the data was to decide whether the software should be released for use. The precise criterion to be used is not fully defined. One might demand that, with high probability, the number of remaining bugs be small, or that the software function without failure for a certain period. Or one might adopt a full, decision-theoretic, approach, attaching a loss to each possible future outcome. The methods developed here provide ways of checking whether such criteria have been met. The results for the three examples are summarised in Table 2.

Example 1

The failure times of a piece of software developed as part of a large data system are shown in Table 3. They have been analysed by Jelinski and Moranda (1972) and

TABLE 2
Results for examples 1, 2 and 3

Example	n	R	$\log_{10} B_{01}$	$\hat{M}_{mod}$	$\hat{M}_{med}$	$P[M=0 \mid t]$	95% *HPDR*	$\hat{M}$	M^*	$\tilde{M}$	0.5 *LI*
1	31	8.4	−3.0	0	.9	.27	0 − 7	0	1	1	0 − 3
2	7	4.1	.7	2	9.8	.05	0 − 174	∞	4	10	2 − ∞
2	136	38.0	−16.0	6	6.5	.01	1 − 16	6	6	7	3 − 11
3	8	4.0	.6	1	7.4	.07	0 − 137	6	3	8	0 − ∞
3	24	8.4	−.2	2	4.4	.07	0 − 36	2	3	4	0 − 9
3	99	18.4	−27.2	0	.0	.57	0 − 2	0	0	1	0 − 1
3	107	13.3	−45.8	0	.0	.95	0	0	0	0	0

Note: $i-j$ denotes the set of integers from i to j inclusive. Notation: $\hat{M}_{mod} = \hat{N}_{mod} - n$; $\hat{M}_{med} = \hat{N}_{med} - n$; $\hat{M}$ is the *MLE* of M; $M^* = N^* - n$; $\tilde{M} = \tilde{N} - n$. 95% *HPDR* is the 95% *HPD* region from (3.2), while 0.5 LI is the 0.5 likelihood interval defined by Joe and Reid (1985).

TABLE 3
Data for example 1

9	7	5	1	3	11	2	12	16
12	2	7	9	3	33	1	9	35
11	5	1	4	6	7	87	135	
4	8	6	1	1	91	47	258	

Note: The data consist of intervals, in days, between successive failures, and are to be read down the columns. They are reproduced from Goel and Okumoto (1979, Table 1). There are 34 failures; the period of observation does not start with a failure but does end with one. The last three failures occurred after the software had been released for use.

Goel and Okumoto (1979). The software was released for use after the thirty-first failure; here, the data up to that point are analysed.

Fig. 1 shows the outcome, for these data, of the model checking exercise described in Section 5. The model appears to fit the data quite well. The Bayes factor B_{01}, at about 10^{-3}, indicated decisive evidence for reliability growth, but $P[M=0\,|\,t]$ was only 0.27, indicating that the system had probably not been fully debugged. Indeed, based on future data, three further failures later occurred. It seems likely that the software should not have been released when it was.

This example illustrates the inadequacy, and, indeed, misleading nature, of point estimators for this problem. All the point estimators of M lay between 0 and 1 inclusive, while the 95% *HPD* region was 0–7.

The techniques proposed here gave similar results to the likelihood analysis of Joe and Reid (1985). $\hat{M}_{med}$ and $\tilde{M}$ were very close. The 0.5 likelihood interval, proposed as an interval estimator by Joe and Reid (1985), had coverage probability close to 0.76, and was the same as the 76% *HPD* region based on (3.3).

Example 2

The failure times of a real-time command and control system are shown in Table 4. These data have been described and analysed by Musa (1975), Meinhold and Singpurwalla (1983), Goel (1985), and Okumoto (1985).

Fig. 2 shows that there were more failures than predicted by the model in the first one-twentieth of the observation period, during which 28 of the 136 failures occurred. This may be due to the presence of faults which were easier to detect than others. The violations of the model assumptions revealed by Fig. 2 are not too great, so that the results of fitting it may provide at least a rough guide to decision-making in this case.

Although the Bayes factor, $B_{01}=10^{-16}$, provides clear evidence for reliability growth, the probability that the system has been fully debugged is only 0.01, and the 95% *HPD* region for M includes values up to $M=16$ remaining faults. Further testing seems necessary. The present approach and the likelihood analysis of Joe and Reid (1985) gave results which were in close agreement.

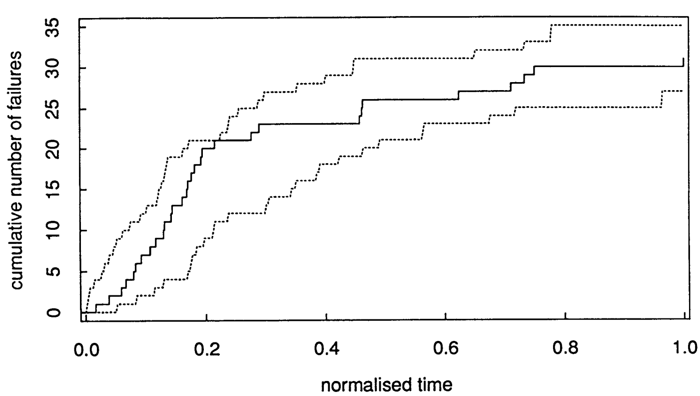

Fig. 1. Model checking for example 1. The solid line was calculated from the data. The broken lines constitute the envelope of 19 simulations from the model, as described in Section 5. The time axis was rescaled in such a way that the rescaled period of observation was [0, 1]

TABLE 4
Data for example 2

3	108	600	452	236	44	1783	865	447	371
30	88	15	255	31	129	860	1435	386	790
113	670	36	197	369	810	983	30	446	6150
81	120	4	193	748	290	707	143	122	3321
115	26	0	6	0	300	33	109	990	1045
9	114	8	79	232	529	868	0	948	648
2	325	227	816	330	281	724	3110	1082	5485
91	55	65	1351	365	160	2323	1247	22	1160
112	242	176	148	1222	828	2930	943	75	1864
15	68	58	21	543	1011	1461	700	482	4116
138	422	457	233	10	445	843	875	5509	
50	180	300	134	16	296	12	245	100	
77	10	97	357	529	1755	261	729	10	
24	1146	263	193	379	1064	1800	1897	1071	

Note: The data consist of intervals, in CPU seconds, between successive failures, and are to be read down the columns. They are reproduced from Meinhold and Singpurwalla (1983, Table 1).

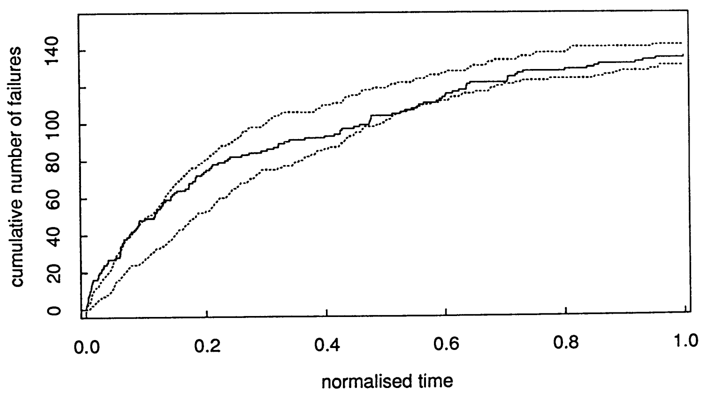

Fig. 2. Model checking for example 2.

An analysis after $n = 7$ failures is of interest because it reveals differences between the present approach and a likelihood analysis. The *MLE* of M was infinite, and the 0.5 likelihood interval was 2—∞, and had coverage probability less than 0.6, but posterior probability 0.88 from (3.3).

Meinhold and Singpurwalla (1983) also analysed the data after $n = 7$ failures. They suggested a Bayesian analysis with a proper prior for N which was Poisson with mean 50. This yielded a posterior distribution for M concentrated between 21 and 57; by comparison (3.3) yielded the 95% *HPD* region 0–155. In fact, 129 further failures occurred.

Example 3

Table 5 shows failure data for a data reduction program. These data were previously analysed by Forman and Singpurwalla (1977), using the same model that is considered here. The data were grouped, and, like them, I have assumed that the average time of occurrence within each group was at the centre of the time interval.

TABLE 5
Data for example 3

Interval number	*Length of interval*	*Number of failures*
1	.50	8
2	.60	7
3	.65	1
4	1.90	8
5	1.59	16
6	8.83	18
7	9.94	13
8	7.25	8
9	8.34	9
10	3.86	2
11	13.11	6
12	34.15	3
13	82.70	3
14	1.10	2
15	51.59	3

Note: The lengths of the intervals are given in CPU seconds. These data are from Forman and Singpurwalla (1977, Table 2).

Fig. 3 shows that the model does not fit at all well. The cumulative number of failures lies outside the envelope for the first 71 failures, out of 107. This is followed by an almost failure-free period whose length is about half that of the entire observation period. In the last quarter of the observation period, the failure rate increases. This pattern may reflect both differences in the ease with which different faults could be found, and the introduction of new faults in the debugging process. Model failure is also shown by the fact that after $n = 99$ failures, the probability of eight or more faults remaining was less than 10^{-4}. Eight more failures did occur.

It is thus unwise to take at face value the conclusion from the model that the program has been fully debugged, which was also drawn by Forman and Singpurwalla

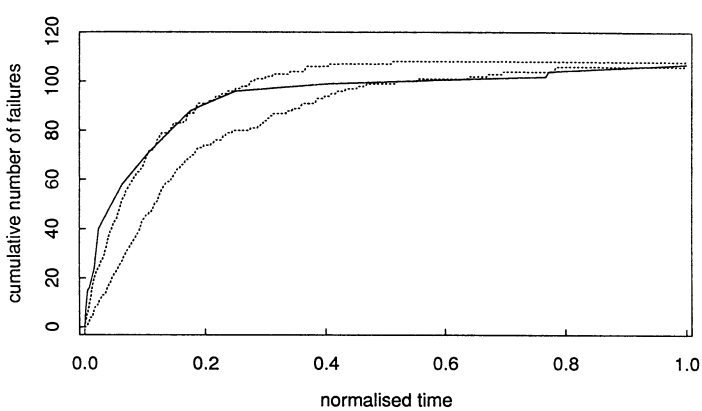

Fig. 3. Model checking for example 3. The solid line was calculated by linear interpolation, as follows. The points corresponding to the ends of the time intervals in Table 5 were plotted, and then joined by straight lines

(1983). One interesting feature of the analysis is that, after $n = 8$ failures, the procedure of Joe and Reid (1985) produced an interval estimate for M which included all its possible values, but whose coverage probability was less than 0.64.

7. Discussion

A Bayesian analysis of a simple and widely used debugging model has been presented. This provides a test for reliability growth, an estimation procedure for the number of remaining faults, an evaluation of current system reliability, and a prediction of the time to full debugging. It also yields an informal, simulation-based, model checking procedure. This extends Ripley's (1977) procedure to allow for the uncertainty associated with unknown parameters.

One conclusion is that point estimation of the number of faults, which has been the focus of much previous research, is difficult, inadequate, and potentially misleading. One must look at the overall uncertainty contained in the complete posterior distribution.

I have reanalysed three software reliability data sets. In the first example, the model fits the data; in the second, it appears to be slightly inadequate; in the third, it fits poorly. There are four key assumptions underlying the model, namely:

(i) the failure times are exponential;
(ii) the failure times have a common mean;
(iii) no additional faults are introduced by the debugging process;
(iv) the failure times are independent.

The realism of each of these assumptions is, of course, questionable; one hopes that they hold to a sufficient approximation for the model to yield useful results. This does seem to happen in example 1, and to a lesser extent in example 2. However, assumptions (i) and/or (ii) are clearly violated in example 3, as may also be assumptions (iii) and/or (iv). I am currently extending the methods proposed here to models in which assumptions (i) and (ii) are relaxed; this work will be reported elsewhere.

Acknowledgements

I am grateful to Fritz Scholz, Peter Guttorp, the Joint Editor (R. Mead), and two anonymous referees for very helpful comments on an earlier version of this paper. This work was supported by ONR contract N00014-84-C-0169.

References

Akman, V. E. (1985) Some aspects of statistical inference for stochastic point processes. Ph.D. Thesis, Trinity College, Dublin, Department of Statistics.

Akman, V. E. and Raftery, A. E. (1986a) Asymptotic inference for a change-point Poisson process. *Ann. Statist.*, **14**, 1583–1590.

——— (1986b) Bayes factors for non-homogeneous Poisson processes with vague prior information. *J.R. Statist. Soc.* B, **48**, 322–329.

Ascher, H. and Feingold, H. (1984) *Repairable Systems Reliability*. New York: Dekker.

Bazovsky, I. (1961) *Reliability Theory and Practice*. Englewood Cliffs: Prentice-Hall.

Berman, M. (1981) Inhomogeneous and modulated gamma processes. *Biometrika*, **68**, 143–152.

Blumenthal, S. and Marcus, R. (1975) Estimating population size with exponential failure. *J. Amer. Statist. Ass.*, **70**, 913–922.

Cox, D. R. and Lewis, P. A. W. (1966) *The Statistical Analysis of Series of Events*. London: Methuen.

Cozzolino, J. M. (1968) Probabilistic models of decreasing failure rate processes. *Nav. Res. Logist. Q.*, **15**, 361–374.

Forman, E. H. and Singpurwalla, N. D. (1977) An empirical stopping rule for debugging and testing computer software. *J. Amer. Statist. Ass.*, **72**, 750–757.

Goel, A. L. (1985) Software reliability models: assumptions, limitations, and applicability. *IEEE Trans. Software Engng*, **SE-11**, 1411–1423.

Goel, A. L. and Okumoto, K. (1979) Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE Trans. Reliabil.*, **R-28**, 206–211.

Goudie, I. B. J. and Goldie, C. M. (1981) Initial size estimation for the linear pure death process. *Biometrika*, **68**, 543–550.

Jaynes, E. T. (1968) Prior probabilities. *IEEE Trans. Syst. Sci. Cybernet.*, **SSC-4**, 227–241.

Jeffreys, H. (1961) *Theory of Probability*, 3rd edn. Oxford: Oxford University Press.

Jelinski, Z. and Moranda, P. B. (1972) Software reliability research. In *Statistical Computer Performance Evaluation* (W. Freiberger, ed.), pp. 465–484. London: Academic Press.

Jewell, W. S. (1985) Bayesian extensions to a basic model of software reliability. *IEEE Trans. Software Engng*, **SE-12**, 1465–1471.

Joe, H. and Reid, N. (1985) Estimating the number of faults in a system. *J. Amer. Statist. Ass.*, **80**, 222–226.

Langberg, N. and Singpurwalla, N. D. (1985) Unification of some software reliability models via the Bayesian approach. *SIAM J. Sci. Statist. Comput.*, **6**, 781–790.

Lewis, P. A. W. (1972) Recent results in the statistical analysis of univariate point processes. In *Stochastic Point Processes* (P. A. W. Lewis, ed.), pp. 1–54. New York: Wiley.

Littlewood, B. and Verrall, J. L. (1981) Likelihood function of a debugging model for computer software reliability. *IEEE Trans. Reliabil.*, **R-30**, 145–148.

MacLean, C. J. (1974) Estimation and testing of an exponential polynomial rate function within the non-stationary Poisson process. *Biometrika*, **61**, 81–85.

Meinhold, R. J. and Singpurwalla, N. D. (1983) Bayesian analysis of a commonly used model for describing software failures. *Statistician*, **32**, 168–173.

Musa, J. D. (1975) A theory of software reliability and its applications. *IEEE Trans. Software Engng*, **SE-1**, 312–327.

Okumoto, K. (1985) A statistical method for software quality control. *IEEE Trans. Software Engng*, **SE-11**, 1424–1431.

Raftery, A. E. (1986) A note on Bayes factors for log-linear contingency table models with vague prior information. *J. R. Statist. Soc.* B, **48**, 249–250.

Raftery, A. E. and Akman, V. E. (1986) Bayesian analysis of a Poisson process with a change-point. *Biometrika*, **73**, 85–89.

Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.

Scholz, F.-W. (1986) Software reliability modeling and analysis. *IEEE Trans. Software Engng*, **SE-12**, 25–31.

Spiegelhalter, D. J. and Smith, A. F. M. (1982) Bayes factors for linear and log-linear models with vague prior information *J. R. Statist. Soc.* B, **44**, 377–387.

Watson, D. A. and Blumenthal, S. (1980) Estimating the size of a truncated sample. *Commun. Statist. Theor. Meth.*, **A9**, 1535–1550.