UNIVERSIDADE DE SÃO PAULO

Inference for the software reliability using asymmetric loss functions: A hierarchical Bayes approach

JOSEMAR RODRIGUES

Nº 21

N O T A S

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Inference for the software reliability using asymmetric loss functions: A hierarchical Bayes approach

JOSEMAR RODRIGUES

Nº 21

NOTAS DO ICMSC
Série Estatística

São Carlos

Resumo

A finalidade deste artigo é propor um modelo hierárquico Bayesiano e uma função de perda apropriada para fazer uma análise Bayesiana do total de erros no software o qual representamos por N. Demonstramos que o procedimento Bayesiano é mais estável que o estimador de máxima-verossimilhança e uma regra de parada para corrigir o software é sugerido via função de perda Linex. A função de densidade preditiva é obtida e dados reais sobre as falhas do software para ilustrar a utilidade da abordagem preditivista e que a função de perda quadrática pode não ser apropriada são analisadas.

Inference for the software reliability using asymmetric loss functions: A hierarchical Bayes approach

Josemar Rodrigues
ICMSC-USP-São Carlos-13560-970
Brasil

Abstract

The aim of this paper is to propose a hierarchical Bayes approach and an appropriate loss function to make a Bayesian analysis of the total number of software failures which is denoted by N. It is showed that the Bayes procedure is more stable than the maximum likelihood procedure and a stopping rule for debugging the software is suggested viá the Linex loss function. Also, the predictive density of the future failure time is obtained and some real life data on software failures to illustrate the usefulness of the predictive approach and that the squared loss function may be inappropriate are considered.

Key words: Software reliability; Bayesian statistics; Debugging; Stopping rule; Linex loss function.

1 Introduction

Let N be the unknown number of errors in a piece of computer software. An important problem in software reliability is the estimation of N from past data.

The most commonly used model for describing the stochastic failures of a software is that proposed by Jelinski and Moranda (1972) - henceforth J-M. The assumptions made by J-M are the following:

(i) When the software fails, the error causing the failure can be detected and removed without inserting any additional errors.
(ii) The failure rate of the software at any point is proportional to the residual numbers of faults in the program; the program begins life with N faults.
(iii) Each of the N faults contributes an equal amount Λ (unknown) to the failure rate.
(iv) Given N and λ,, the times between sucessive failures of the program , $T_1, T_2, \ldots, T_N$, are independent with the conditional density given by

$$f(t_i \mid N, \Lambda) = \Lambda(N - i + 1)\exp\{-\Lambda(N - i + 1)t_i\}, \tag{1}$$

with N and Λ being the unknown parameters.

There are some critical issues about the J-M model which are:
(i) the "bug counting" framework from which the model is derived is unrealistic (see Littlewood, 1981).
(ii) the maximum likelihood estimator (MLE) of N can often be misleading (see Forman and Singpurwalla (1977)).
(iii) the use of symmetric loss functions may be inappropriate to make inference for N, as has been recognized in the literature (see Zellner, 1986).
(iv) the stopping procedure for debugging the software proposed by Forman and Singpurwalla (1977) is empirical and needs a critical examination of the actual likelihood function at each stage of the procedure without considering an appropriate loss function.
(v) the squared loss function may be inappropriate for predicting future failure times, as showed in this paper via NTDS data given by Jelinski and Moranda (1972).

In this paper, we will discuss in details items (ii)-(v) of the five critical issues of the J-M model that we presented above.

These and other matters are developed as follows:

Motivated by Raftery (1988), in Section 2 we present a hierarchical Bayes approach to get information about N. Also, using the Table 1 introduced by Forman and Singpurwalla (1977) to study the instability of the m.l.e. of N, a comparison of the posterior mode with the m.l.e. is considered. In Section 3, we suggest the use of Linex function which seems us to be more appropriate for estimating the parameter N and a stopping rule for debugging the software based on the robustness of this asymmetric loss function is formulated. Finally, in Section 3 we use the NTDS data (Singpurwalla and

Wilson, (1994)) to show that the quadratic loss function is not appropriate to predict future failure times and we suggest the use of the relative loss function. This section is ended with some considerations about the use of asymmetric loss functions in software reliability.

2 A hierarchical Bayes Model to estimate N

Bayesian inference for the unknown parameters N and Λ envolves assigning a prior distribution to the pair (N, Λ), and using the data and the Bayes theorem to obtain the posterior distribution. In this section we are concerned about the parameter N. A difficulty with Bayesian analysis to estimate the parameter N has been the absence of a sufficiently flexible family of prior distributions, mainly due the fact that N is an integer. To avoid this problem, as in Raftery's paper (1988), we assume that N has a Poisson model distribution. Since the resulting hyperparameters are continuous-valued, the following hierarchical Bayes model can be formulated:

$$\begin{array}{ll} (i)- & T_i \mid N, \Lambda \sim \exp\{\Lambda(N-i+1)\}, \quad i=1,\ldots,n, \\ (ii)- & N \mid \mu \sim \text{Poisson}(\mu), \\ (iii)- & \mu \sim \text{Gamma}[a,b], \quad \Lambda \sim \text{Gamma}[c,d], \end{array} \tag{2}$$

where μ and Λ are supposed to be independent. This model provides a kind of "unification" of many alternative models in software reliability (see Langberg and Singpurwalla, (1985)) because each alternative model is obtained by assigning particular prior distributions to the parameters N and Λ.

With the purpose to compare our Bayes procedure with the m.l.e., we only consider the noninformative case, that is, we take $a = b = c = d = 0$.

Let $t_1, \ldots, t_n$ be the observed times between failures, and let $t^{(n)} = (t_1, \ldots, t_n)$. Then given the observed data $t_{(n)}$ and noninformative priors , the likelihood function for $N \quad (N \geq n)$ and Λ is

$$L(N, \Lambda \mid t^{(n)}) = \Lambda^n \Pi_{i=1}^n [N-i+1] \exp\{-\Lambda s(N-a)\}, \tag{3}$$

where $s = \sum_{i=1}^n t_i$ and $a = \sum_{i=1}^n (i-1)t_i/s$.

The joint posterior density of (N, Λ, μ) for $N \geq n$ is

$$p(N, \Lambda, \mu \mid t^{(n)}) \propto \frac{e^{-\mu}\mu^{N-1}}{N!} \Pi_{i=1}^n [N-i+1] \Lambda^{n-1} \exp\{-\Lambda s(N-a)\}. \tag{4}$$

Integrating over μ and Λ, we obtain the marginal posterior density of N as

$$p(N \mid t^{(n)}) \propto \frac{\Pi_{i=1}^{n}[N-i+1]}{N(N-a)^n}, \qquad (5)$$

for $N \geq n$.

If $n << N$, Forman and Singpurwalla (1977) showed that m.l.e. of N may be very unstable with respect to a. They concluded that as a becomes large, i.e., when the times between failures in the latter stages of testing are greater than those during the earlier stages, the m.l.e of N tends to n. If a is small, the m.l.e. of N tends to become large, i.e., the m.l.e. of N is unstable for small values of a. This situation could provide estimatives of N that are highly misleading.

Let us denote the posterior mode of N by $\hat{N}_{\text{MOD}}$ and the m.l.e. of N by $\hat{N}_{\text{MLE}}$. In Table 1, taking $n = 5$, we compare the behavior of $\hat{N}_{\text{MOD}}$ and $\hat{N}_{\text{MLE}}$.

Table 1: Dependency of $\hat{N}_{\text{MOD}}$ and $\hat{N}_{\text{MLE}}$ on a

	n=5	
a	$\hat{N}_{\text{MLE}}$	$\hat{N}_{\text{MOD}}$
4.0	5.0	5.0
2.8	5.07	5.0
2.3	8.83	5.50
2.2	12.1	5.70
2.06	35.4	6.02
2.03	60.7	6.14
2.02	102.0	6.14
2.005	401.0	6.14
2.002	1,005.0	6.14
2.001	1,962.0	6.18
2.000	35,288.0	6.18

Table 1 demonstrates very clearly that for small values of a , the Bayes estimator $\hat{N}_{\text{MOD}}$ is very stable when comparing with the estimator $\hat{N}_{\text{MLE}}$.

In Figure 1, we illustrate the behavior of the logarithm of the profile likelihood, $L(N)$, (see Forman and Singpurwalla,(1977)) and the posterior density of N for $n = 2$ and $a = 0.52$. We note in Figure 1 that when

$a = 0.52$ and $n = 2$, the maximum likelihood estimator of N, $\hat{N}_{\text{MLE}}$, is not finite, however, the Bayes estimator is finite, i.e., $\hat{N}_{\text{MOD}} = 2$.

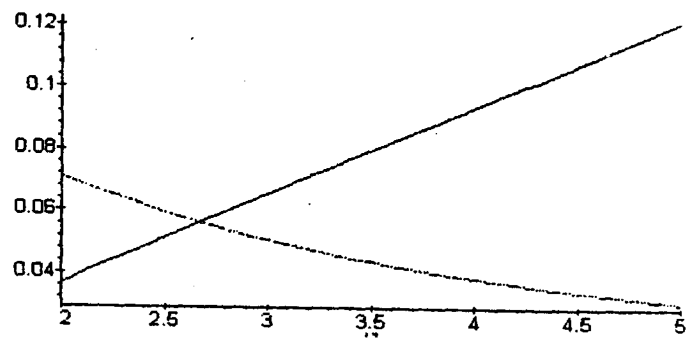

Figure 1: $L(N)$ (solid line) and $p(N \mid t^{(2)})$ (dot line) for $n = 2$ and $a = 0.52$.

3 Bayes estimates under asymmetric loss

In the estimation of N use of symmetric loss function may be inappropriate as has been recognized in the literature - see, for example, Zellner, (1986). It seems to us that underestimate the parameter N results in more serious consequence than overestimate N. To take into account this problem, we suggest in this paper the use of the LINEX function (Zellner, (1986)) which is formulated as follows:

$$L(\Delta) = b[\exp^{c\Delta} - c\Delta - 1], \quad c > 0, b > 0, \tag{6}$$

where $\Delta = N - \hat{N}$. It is seen that, for $c = 1$, the function is quite asymmetric with underestimation being more costly than overestimation. If $c > 0$ and $\Delta > 0$, loss grows exponentially as Δ grows whereas when $\Delta < 0$, loss grows approximately linearly. Thus positive and negative errors of equal magnitude give rise to quite different losses when the asymmetric loss function (6) is employed. For small values of c, the function is almost symmetric and not far from a squared loss function (see Varian (1975) for more details). It is not difficult to see that the Bayes estimator that minimizes the posterior expectation of the LINEX loss function with respect to $p(N \mid t^{(n)})$ is given by

$$\hat{N}_{\text{LINEX}} = -\frac{1}{c} ln\{\sum_{N\geq n} e^{-cN} p(N \mid t^{(n)})\} \quad c > 0. \tag{7}$$

Also, it can be demonstrated that:

$$
\begin{aligned}
&1. \quad E(N \mid t^{(n)}) \leq \hat{N}_{\text{LINEX}}, \\
&2. \quad \hat{N}_{\text{LINEX}} \text{ is a decreasing function of } c \quad (c > 0).
\end{aligned} \tag{8}
$$

For the standard debugging procedure, Forman and Singpurwalla (1977) proposed an empirical stopping rule based upon the relative likelihood function for the data, generated by simulation. Here motivated by (8) we propose the following sequencial stopping rule for debugging the software :

1. Given that n errors are detected and $c = c_o$, compute

   $$d(n, c_o) = E(N \mid t^{(n)}) - \hat{N}_{\text{LINEX}},$$

2. If $d(n, c_o)$ is closed to zero, stop testing and accept the software and use the posterior mean or the optimal Bayes estimator under the LINEX function to estimate the parameter N. If not, observe another failure time t_{n+1} and go to step 1.

This stopping rule provides a kind of robustness with respect to the loss functions, i.e., we do not need to be worried about symmetric or asymmetric loss functions to estimate the parameter N. Also, it seems that this procedure is more realistic than the stopping rule suggested by Forman and Singpurwalla (1977) in the sense that we have a numerical measure to decide to stop or not debugging the software.

To illustrate the performance of our proposed procedure to stop testing the software, we consider the data obtained during the debugging of a data reduction program called the F 11-D program. This program consists of " approximately 3-4 thousand" Fortran statements. As can be seen from Table 2, a total of 107 errors were detected after debugging the program for a total of 226.11 seconds of CPU time. Several strategies for distributing the errors in each time interval are available to a user. In our case, we adopt the strategy in which the times between errors in each interval is the quotient of the interval length (CPU time) to the number of errors in the interval. For more details about this data we suggest to read Moranda (1975) or Forman and Singpurwalla (1977).

Table 2. Data on F 11-D Program

Interval number	Date	Number of errors:n	Cumulative number of errors	CPU time (seconds)
1	1/12	8	8	0.5
2	1/15	7	15	0.6
3	1/16	1	16	0.65
4	1/17	8	24	1.9
5	1/18	16	40	1.59
6	1/19	18	58	8.83
7	1/22	13	71	9.94
8	1/23	8	79	7.25
9	1/24	9	88	8.34
10	1/25	2	90	3.86
11	1/26	6	96	13.11
12	1/27	3	99	34.15
13	1/29	3	102	82.7
14	1/30	2	104	1.1
15	1/31	3	107	51.59
Total		107	226.11	

In Table 3, the sucessives values of $d(n, 10)$ and the posterior mean are computed at the end of each interval to decide to stop debugging the software. At the end of each interval, we compute $d(n, 10)$ to see the influence of the LINEX function on the estimative of N. For example, at the end of the interval number five we have a strong influence of the LINEX function with respect to posterior mean and our measure $d(40, 10)$ signals us to continue debugging the software. After the twelfth interval, our procedure shows a better agreement than we had before and giving us a good indication that the debugging process is closed to finish. Besides to get robustness with respect to the LINEX function after having observed $n = 99$, Forman and Singpurwalla (1977) showed for this data a very good agreement between the normal and the actual relative likelihood functions.

Table 3. Stopping rule for debugging the software using F11-D data.

n	$d(n,10)$	$E(N \mid t^{(n)})$	$\hat{N}_{MLE}$
8	99.64	124.30	6
15	108.25	140.81	16
16	31.51	140.82	17
24	32.59	61.55	24
40	163.13	226.42	43
58	12.04	70.74	60
71	7.72	79.18	73
79	7.51	86.98	81
88	8.69	97.24	90
90	7.08	97.55	92
96	4.46	99	99
99	0.66	99	100
102	0.03	102	102
104	0.08	104	104
107	0.05	107	107

4 The Predictive Bayesian Decision Approach

The aim of this section is to use the Bayesian decision approach introduced by Geisser (1982) to predict a future variable T_{n+1}. We shall examine some of the standard loss functions that are used in problems of parameter estimation and shall compare their usefulness to predict a future variable T_{n+1} based on the data developed by the Naval Tactical Data System (NTDS) (Jelinski and Moranda, (1972)). As in Geisser's paper (1982), let the predictive distribution of the future variable T_{n+1} given the current data $t^{(n)}$ as

$$f(t_{n+1} \mid t^{(n)}) = \sum_{N \geq N} \int_0^\infty f(t_{n+1} \mid N, \Lambda) p(N, \Lambda \mid t^{(n)}) d\Lambda, \tag{9}$$

where

$$\begin{aligned} f(t_{n+1} \mid N, \Lambda) &= \Lambda[N-n]\exp\{\Lambda[N-n]t_{n+1}\} \quad \text{and} \qquad (10) \\ p(N, \Lambda \mid t^{(n)}) &\propto \frac{\Lambda^{n-1}}{N}\Pi_{i=1}^{n}[N-i+1]\exp\{-\Lambda s(N - \frac{\sum_{i=1}^{n}(i-1)t_i}{\sum i = 1^n t_i})\}. \end{aligned}$$

Given a predictive loss function $L_p(\hat{T}_{n+1}, T_{n+1})$, the average predictive loss is

$$L_p(\hat{T}_{n+1} \mid t^{(n)}) = \int_0^\infty L_p(\hat{T}_{n+1}, t_{n+1}) f(t_{n+1} \mid t^{(n)}) dt_{n+1}, \tag{11}$$

and the optimal $\hat{T}_{n+1}$ is that T^*_{n+1} such that $L_p(\hat{T}_{n+1} \mid t^{(n)})$ is minimized with respect to $\hat{T}_{n+1}$. If we assume that a predictive squared loss function $L_p(\hat{T}_{n+1}, T_{n+1}) = (\hat{T}_{n+1} - T_{n+1})^2$ is appropriate then

$$T^*_{n+1} = E(T_{n+1} \mid t^{(n)}) = E_{(N,\Lambda)}[\frac{1}{(N-n)\Lambda}], \tag{12}$$

where $E_{(N,\Lambda)}$ is the espectation with respect to the posterior density $p(N, \Lambda \mid t_{(n)})$. If we assume that a predictive relative loss function $L_p(\hat{T}_{n+1}, T_{n+1}) = (T_{n+1}/\hat{T}_{n+1} - 1)^2$ is appropriate then optimal solution corresponding to this loss function is

$$T^*_{n+1} = \frac{E(T^2_{n+1} \mid t^{(n)})}{E(T_{n+1} \mid t^{(n)})}. \tag{13}$$

It can be demonstrated that under the hierarchical Bayesian model proposed in this paper, the predictive density of the time of the next failure , T_{n+1}, is given by

$$f(t_{n+1} \mid t^{(n)}) = \frac{n \sum_{N>n} \frac{\prod_{i=1}^{n}[N-i+1](N-n)}{N} [(N-n)t_{n+1} + s(N-a)]^{-(n+1)}}{\sum_{N>n} \frac{\prod_{i=1}^{n}}{N} [s(N-a)]^{-n}} \tag{14}$$

and the two first moments of this predictive density are given by

$$E(T_{n+1} \mid t^{(n)}) = \frac{1}{n-1} \frac{\sum_{N>n} \frac{\prod_{i=1}^{n}[N-i+1]}{N(N-n)} [s(N-a)]^{-(n-1)}}{\sum_{N>n} \frac{\prod_{i=1}^{n}[N-i+1][s(N-a)]^{-n}}{N}} \tag{15}$$

$$E(T^2_{n+1} \mid t^{(n)}) = \frac{2}{(n-1)(n-2)} \frac{\sum_{N>n} \frac{\prod_{i=1}^{n}[N-i+1]}{N(N-n)^2} [s(N-a)]^{-(n-2)}}{\sum_{N>n} \frac{\prod_{i=1}^{n}[N-i+1]}{N} [s(N-a)]^{-n}}.$$

To analyse the performance of the two standard loss functions mentioned before, we consider the NTDS data used for many authors when attempting to validate their models (see Singpurwalla and Wilson, 1994). Jelinsky and Moranda (1972) used this data which consists of the number of days between the 32 failures that ocurred during the production phase, first test phase and the user phase. In the next table we have the NTDS data introduced by Jelinski and Moranda (1972).

Table 4: NTDS Data

Production											
i	t_i	i	t_i	i	t_i	i	t_i	i	t_i	i	t_i
1	9	6	2	11	1	16	1	21	11	26	1
2	12	7	5	12	6	17	3	22	33		
3	11	8	8	13	1	18	3	23	7		
4	4	9	5	14	9	19	6	24	91		
5	7	10	7	15	4	20	1	25	2		
Test 1											
i	27	28	29	30	31						
t_i	87	47	12	9	135						
User											
i	32										
t_i	258										

Figure 2. shows the predictive density of T_{32} for the NTDS data with $n = 31$.

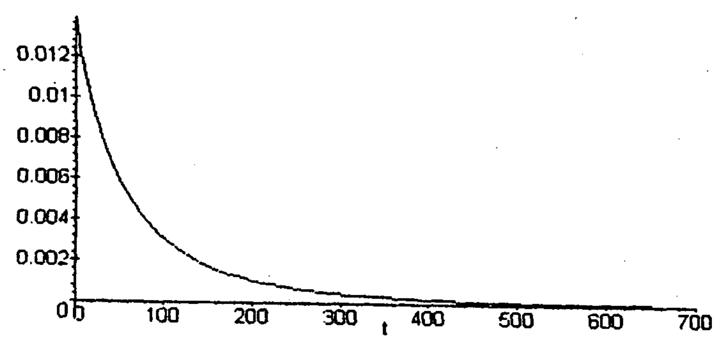

Figure 2. Predictive density of T_{32} for the NTDS data.

Table 5. Optimal predictors of T_{32} for the NTDS data based on the phases Production and Test 1.

Loss functions	Est. mean time to next failure	Actual time to next failure
quadratic	101	258
relative	263	258

The estimated time obtained by Jelinski and Maranda to the detection of the next error in the user phase was about 239 days. The approach of this section yields a full solution to the decision-making problem, by the usual method of minimizing predictive posterior expected loss introduced by Geisser. As we can see in Table 5., the predictive quadratic loss function for NTDS data is not appropriate to predict the time to next failure based on the phases production and test 1. For this particular data, the relative loss function seems to be more reasonable. This example call our attention to an important problem which consists in choosing an appropriate loss function to predict future observations in software reliability. Another important aspect of our proposed stopping procedure is the robustness with respect to the shape parameter c of the LINEX function. It means that our Bayesian procedure to estimate N is not affected when replacing the squared loss function by the LINEX loss function. Also, as discussed by Forman and Singpurwalla (1977), for small values of $d(n, 10)$ we have the normality of the relative likelihood function.

References

Forman, E.H. and Singpurwalla, N.D. (1977). An empirical stopping rule for debugging and testing computer software. Journal of the American Statistical Association, 72,360, 750-757.

Geisser, S. (1982). Aspects of the predictive and estimative approaches in the determination of probabilities. Biometrics Supplement: Current topics in biostatistics and epidemiology, 75-85

Jelinski, Z. and Moranda, P.B. (1972). Software reliability research, in Statistical Computer Performance Evaluation, ed.W. Freiberger, New York and London: Academic Press,465-484.

Littlewood, B. (1981) A critique of the Jelinski-Moranda model for software reliability. Proc. Ann. Rel.and Maintainability Symp.,357-364

Langberger, N. and Singpurwalla, N.D. (1985). A unification of some software reliability models. Siam J.Sci. Stat. Comput., 6,3,781-790

Moranda, P.B. (1975) A comparison of software error-rate models. Texas Conference on Computing, 2A-6.1-2A-6.9.

Raftery, A.E. (1988). Inference for binomial N parameter: A hierarchical Bayes approach. Biometrika, 75,2,223-228.

Singpurwalla, N.D. and Wilson, S.P. (1994). Software reliability modeling. International Statistical Review, 62,3,289-317.

Varian, Hal R. (1975). A Bayesian approach to real estate assessment, in

Studies in Bayesian Econometrics and Statistician Honer of Leonard J. Savage. eds. Stephen E. Fienberger and Arnold Zellner, Amsterdam: North Holland, 195-208.
Zellner, A. (1986). Bayesian Estimation and prediction using asymmetric loss functions. Journal of the American Statistical Association, 81,394, 446-451.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

020/95 RODRIGUES, J.; BOLFARINE, H.; CORDEIRO, G.M. - Nonlinear quasi-bayesian theory and the inverse linear regression.

019/95 LEANDO, R.A.; ACHCAR, J.A.; - Generation of bivariate lifetime data assuming the Block & Basu exponential

018/95 ACHCAR, J.A.; - Use of approximate bayesian inference for software reliability

017/95 ACHCAR, J.A.; FAVORETTI, A.C. - Accurate inferences for the Michaelis-Menten model

016/95 RODRIGUES, J.; CHITTA, S.P. - Bayesian Analysis in M/M/1 queues using Sampling Methods

015/95 ACHCAR, J.A.; PEGORINM M.A. - Laplace's approximations for posterior expectations when the mode is not in the parameter space.

014/94 ACHCAR, J. A.; FOGO, J.C. - Accurate inferences for the reliability function considering accelerated life tests.

013/94 RODRIGUES, J. - Bayesian Solutions to a lass of selections problems using weighted loss functions.

012/94 ACHCAR, J.A.; DAMASCENO, V.L. - Extreme value models: an useful reparametrization for the survival function

011/94 ACHCAR, J.A.- Approximate bayesian analysis for non-normal hierarchical classification model.

010/94 RODRIGUES, J.; RODRIGUES, E.F. - Bayesian estimationin the study of tampered Random variables.

009/94 ACHCAR, J.A.; FOGO, J.C. - An useful reparametrization for the reliability in the Weibull case.

008/94 RODRIGUES, J.; LEITE, J.G. - A note on Bayesian analysis in M/M/1 queues from confidence intervals.