UNIVERSIDADE DE SÃO PAULO

Use of approximate bayesian inference for software reliability

JORGE ALBERTO ACHCAR

Nº 18

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

Use of approximate bayesian inference for software reliability

JORGE ALBERTO ACHCAR

Nº 18

NOTAS DO ICMSC
Série Estatística

São Carlos
jun./95

Resumo

Apresentamos um breve resumo de alguns modelos para confiabilidade de software introduzidos na literatura. O uso de métodos Bayesianos é um modo apropriado de se obter inferencias para confiabilidade de software, pois especialistas em software, em geral tem opinião prévia sobre as falhas de um programa computacional. Exploramos o uso de métodos de aproximação para a obtenção de sumários a posteriori de interêsse considerando alguns modelos de confiabilidade de software.

Use of Approximate Bayesian Inference for Software Reliability

Jorge Alberto Achcar
Universidade de São Paulo
ICMSC-USP, C. Postal 668
13560-970, São Carlos, S.P., Brazil

Abstract

We present a brief review of some existing software reliability models introduced in the literature. The use of Bayesian methods is a suitable way to get inferences for software reliability, since software experts usually have prior opinion on the failures of a computer program. We explore the use of approximation methods to get posterior summaries of interest considering some software reliability models.

Keywords: software reliability models, Bayesian inference, approximate methods.

1 Introduction

Software reliability is the probability of a computer program to be free of error in operation during a specified period of time. The software failures are related to errors in syntax or logic (see for example, Singpurwalla and Wilson, 1994). Once such an error is found, it can be corrected and does not give rise to any more failures.

Software randomly that justify the use of stochastic models for software reliability is related to a program receiving many different inputs, each with its own path through the software and so capable of bringing different errors in light. These different inputs arrive to the software randomly, which implies in detection of errors in a random way.

The literature in statistics and especially in software engineering presents many different models for software reliability (see for example, Singpurwalla and Wilson, 1994; or Mazzuchi and Soyer, 1988).

Among these stochastic models, we have two different strategies:

Type I Strategy: Modelling times between successive failures of the software.

Type II Strategy: Modelling the number of failures of the software up to a given time.

The type I strategy models can be derived from considerations of the failure rates of the software (type I-1 strategy) or to define a stochastic relationship between successive failure times (type I-2 strategy).

A popular model of type I-1 strategy is introduced by Jelinski and Moranda (1972). Suppose thet the total number of bugs in the program is N, and suppose that each time the software fails, one bug is corrected.

The failure rate of the i-th time between failures T_i, is assumed to be a constant proportional to $N - i + 1$, which is the number of bugs remaining in the program.

Thus, the failure rate for T_i is given by,

$$r_{T_i}(t|N, \Lambda) = \Lambda(N - i + 1) \tag{1}$$

where $i = 1, 2, 3, \ldots, t \geq 0$, for some constant Λ.

This means that if N and Λ are Known, then T_i has an exponential density,

$$f(t_i|\lambda_i) = \lambda_i \; e^{-\lambda_i t_i} \tag{2}$$

where $\lambda_i = \Lambda(N - i + 1)$.

Considering $T_1, \ldots, T_n$, independent random variables with density (2), the maximum likelihood estimators for Λ and N satisfy the equations,

$$\sum_{i=1}^{n} \frac{1}{\widehat{N} - i + 1} = \frac{n}{\widehat{N} - \frac{1}{s_n} \sum_{j=1}^{n} (j-1) t_j} \tag{3}$$

where $\widehat{N} \geq n$, and

$$\widehat{\Lambda} = \frac{n}{\widehat{N} s_n - \sum_{i=1}^{n} (i-1) t_i} \quad \text{where} \quad s_n = \sum_{j=1}^{n} t_j.$$

Some modifications of JM model (1) are presented in the literature. Moranda (1975), supposed that the fixing of bugs that cause early failures in the system reduces the failure rate more than the fixing of bugs that occur later, because these early bugs are more likely to be the bigger ones. Thus, Moranda (1975) assume that the failure rate should remain constant for each T_i, but that it should be made to decrease geometrically in i after each failure, that is, for constants D and K,

$$r_{T_i}(t|D, K) = DK^{i-1} \tag{4}$$

where $t \geq 0, D > 0$ and $0 < K < 1$.

Goel and Okumoto (1978) propose a model similar to JM model, but assuming that there is a probability $p, 0 \leq p \leq 1$, of fixing a bug when it is encountered. Thus, the failure rate of T_i is given by

$$r_{T_i}(t \mid N, \Lambda, p) = \Lambda(N - p(i-1)) \tag{5}$$

When $p = 1$, we get the JM model (1).

Schick and Wolverton (1978) assume that the failure rate is proportional to the number of bugs remaining in the system and the time elapsed since the last failure. Thus,

$$r_{T_i}(t \mid N, \Lambda) = \Lambda(N - i + 1)t \tag{6}$$

Other models also are proposed in the literature following type $I-1$ strategy. An alternative to model time between failures is to define a stochastic relationship between successive failure times (type I-2 strategy). As an example, let $T_1, \ldots, T_i, \ldots$ be random variables denoting the length of time between successive failures of the software, with the relationship,

$$T_{i+1} = \rho T_i + \varepsilon_i \tag{7}$$

where $\rho \geq 0$ is a constant, and ε_i is an error term.

Among type II strategy models, Moranda (1975), proposes a Poisson process to describe the number of failures in each successive time period. Other type II strategy models are proposed by Goel and Okumoto (1979), Goel (1983) and Musa and Okumoto (1984).

2 The Use of Bayesian Methods for Software Reliability

The use of Bayesian methods is a suitable alternative to get inferences for software reliability. Usually software experts have prior opinion about software which could be incorporated as a prior distribution for the parameters of the model.

When we have difficulties to get analytical solutions for the Bayesian integrals required to find the posterior summaries of interest, we could use one among the different existing strategies: the use of numerical methods (see for example, Naylor and Smith, 1982); the use of approximation methods for integrals (see for example, Tierney and Kadane, 1986); or the use of Monte Carlo or Gibbs sampling methods (see for example, Kloek and Van Dijk, 1978; or Gelfand and Smith, 1990).

Some special Bayesian software reliability models are introduced in the literature. Littlewood and Verrall (1973) assume that the i-th time between failures has an exponential distribution with failure rate Λ_i and density,

$$f_{T_i}(t \mid \Lambda_i) = \Lambda_i e^{-\Lambda_i t} \tag{8}$$

where $t \geq 0, \Lambda_i > 0$ and that instead of Λ_i be considered decreasing with certainty, as is assumed in the JM model (1), they assumed the sequence of Λ's to be stochastically decreasing, that is, $P(\Lambda_{i+1} < \lambda) \geq P(\Lambda_i < \lambda)$, for $i = 1, 2, \ldots$, and $\lambda \geq 0$.

Thus, they consider a gamma distribution for Λ_i with shape parameter α and scale $\psi(i)$, where $\psi(i)$ is a monotonically increasing function of i,

$$\pi_{\Lambda_i}(\lambda \mid \alpha, \psi(i)) = \frac{[\psi(i)]^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\psi(i)\alpha}. \tag{9}$$

The function $\psi(i)$ is supposed to describe the quality of the programmer and the programming task, and is supposed completely known. Mazzuchi and Soyer (1988) propose the use of $\psi(i) = \beta_0 + \beta_1 i$.

Mazzuchi and Soyer (1988) consider a Bayes empirical model or hierarchical extension of Littlewood and Verrall model. As with the original model, they assume T_i to be exponentially distributed with scale parameter Λ_i (see (8)). To describe Λ_i, they assume two possible models:

Model A: The random variable Λ_i has a gamma distribution with parameters α and β. Also assume that α and β are independent random variables where α has a uniform distribution and β another gamma distribution. That is,

$$\pi_{\Lambda_i}(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \tag{10}$$

where $\lambda \geq 0$,

$$\pi(\alpha | v) = \frac{1}{v}, 0 \leq \alpha \leq v ,$$

and $\pi(\beta | a, b) = \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta}, \beta \geq 0, a > 0, b > 0; v$, a and b are known.

Model B: Assume Λ_i described in the same way as in Littlewood and Verrall, that is,

$$\pi_{\Lambda_i}(\lambda \mid \alpha, \psi(i)) = \frac{[\psi(i)]^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\psi(i)\alpha}, \tag{11}$$

where $\lambda \geq 0, \psi(i) = \beta_0 + \beta_1 \, i$,

$$\pi(\alpha \mid \omega) = \frac{1}{\omega}, 0 \leq \alpha \leq \omega$$

$$\pi(\beta_0 | a, b, \beta_1) = \frac{b^a}{\Gamma(a)} (\beta_0 + \beta_1)^{a-1} e^{-b(\beta_0+\beta_1)}, \beta_0 \geq -\beta_1, a > 0, b > 0 ,$$

$$\pi(\beta_1 | c, d) = \frac{d^c}{\Gamma(c)} \beta_1^{c-1} e^{-d\beta_1}, \beta_1 \geq 0, \; c > 0 \text{ and } d > 0.$$

3 A Bayesian Analysis for the Jelinski-Moranda Model

Considering $T_1, T_2 \ldots, T_n$ as a random sample of times between successive failures of the software under the JM model (1), the likelihood function for Λ and N is given by

$$L(\Lambda, N) = \Lambda^n A(N)\, exp\{-B(N)\Lambda\} \tag{12}$$

where $A(N) = \pi_{i=1}^{n}(N - i + 1)$ and $B(N) = \sum_{i=1}^{n}(N - i + 1)t_i$.

Observe that $N \geq n$.

Assuming prior independence between Λ and N with gamma distributions, a joint prior density for Λ and N is given by

$$\pi_1(\Lambda, N) \propto \Lambda^{a_1-1} N^{a_2-1} e^{-b_1\Lambda - b_2 N} \tag{13}$$

where $\Lambda > 0$ and $N > 0$; a_1, b_1, a_2 and b_2 are known constants based on the prior opinion of an expert.

Observe that $E(\Lambda) = a_1/b_1, var(\Lambda) = a_1/b_1^2, E(N) = a_2/b_2$ and $var(N) = a_2/b_2^2$.

When $a_1 = a_2 = b_1 = b_2 = 0$ we have a noninformative prior density given by,

$$\pi_2(\Lambda, N) \propto \frac{1}{\Lambda N} \tag{14}$$

where $\Lambda > 0$ and $N > 0$.

Considering the prior density (13), the joint posterior density for Λ and N is given by

$$\pi_1(\Lambda, N|t_1, \ldots, t_n) \propto \Lambda^{n+a_1-1} N^{a_2-1} A(N) exp\{-b_2 N - (b_1 + B(n))\,\Lambda\} \tag{15}$$

where $\Lambda > 0$ and $N \geq n$.

The marginal posterior density for N is given by

$$\pi_1(N|t_1, \ldots, t_n) \propto \frac{N^{a_2-1} A(N) e^{-b_2 N}}{\{b_1 + B(N)\}^{n+a_1}} \tag{16}$$

where $N \geq n$.

The marginal posterior density for Λ is given by

$$\pi_1(\Lambda|t_1,\ldots,t_n) \propto \Lambda^{n+a_1-1}\int_n^\infty e^{-nh_\Lambda(N)}d_N \qquad (17)$$

where $-nh_\Lambda(N) = (a_2-1)lnN + lnA(N) - b_2N - (b_1+B(N))\Lambda$.

A Laplace's approximate marginal posterior density for Λ (see (A.8) in appendix) is given by

$$\begin{aligned}\pi_1(\Lambda|t_1,\ldots,t_n) \;&\propto\; \Lambda^{n+a_1-1}\widehat{N}^{a_2-1}A(\widehat{N})\left\{1-\Phi\left(\tfrac{n-\widehat{N}}{\hat{\sigma}}\right)\right\}\times\\ &\times\; \frac{exp\left\{-b_2\widehat{N}-\left(b_1+B(\widehat{N})\right)\Lambda\right\}}{\left\{\frac{a_2-1}{\widehat{N}^2}+\sum_{i=1}^n\frac{1}{\left(\widehat{N}-i+1\right)^2}\right\}^{1/2}},\end{aligned} \qquad (18)$$

where $\Lambda \geq 0$, $\widehat{N}$ maximizes $-nh_\Lambda(N)$ for each value of Λ, $\hat{\sigma} = \left\{\frac{a_2-1}{\widehat{N}^2}+\sum_{i=1}^n\frac{1}{(\widehat{N}-i+1)^2}\right\}^{-1/2}$, and Φ denotes the distribution function of a standard normal distribution $N(0,1)$.

The predictive density for a future observation also could be obtained by using Laplace's method. In the special case with N known, the predictive densite for a future failure time is given by

$$f_1(t_{n+1}|t_1,\ldots,t_n) \propto \int_0^\infty f(t_{n+1}|\Lambda)\pi_1(\Lambda|t_1,\ldots,t_n)d\Lambda \qquad (19)$$

where $f(t_{n+1}|\Lambda) \propto \Lambda e^{-\Lambda(N-n)t_{n+1}}$ and $\pi_1(\Lambda|t_1,\ldots,t_n) \propto \Lambda^{n+a_1-1}exp\left\{-\left(b_1+B(N)\right)\Lambda\right\}$.

That is,

$$f_1\left(t_{n+1}|t_1,\ldots,t_n\right) \propto \left\{(N-n)t_{n+1}+b_1+B(N)\right\}^{-(n+a_1+1)} \qquad (20)$$

where $t_{n+1} \geq 0$.

4 A Bayesian Analysis for the Moranda Model

Considering a random sample of size n of times between successive failures of the software under the Moranda model (4), the likelihood function for D and K is given by

$$L(D,K) = D^n K^{\frac{n(n-1)}{2}} e^{-D\sum_{i=1}^{n} K^{i-1}t_i} \tag{21}$$

Assuming prior independence between D and K with a gamma distribution for D and a beta distribution for K, a joint prior density for D and K is given by

$$\pi(D,K) \propto D^{a_3-1}e^{-b_3 D}K^{a_4-1}(1-K)^{b_4-1}, \tag{22}$$

where $D > 0; 0 < k < 1; a_3, b_3, a_4$ and b_4 are known constants.

The choice of a_3, b_3, a_4 and b_4 is based on the prior opinion of an expert. Observe that $E(K) = a_4/(a_4+b_4)$ and $var(K) = a_4 b_4/[(a_4+b_4+1)(a_4+b_4)^2]$.

With prior (22), the joint posterior density for D and K is given by

$$\begin{aligned} \pi(D,K|t_1,\ldots,t_n) \quad &\propto \quad D^{n+a_3-1}K^{a_4+\frac{n(n-1)}{2}-1}(1-K)^{b_4-1} \times \\ &\times \quad e^{-\left(b_3+\sum_{i=1}^{n} K^{i-1}t_i\right)D} \end{aligned} \tag{23}$$

where $D > 0$ and $0 < K < 1$.

The marginal posterior density for K is (from (23)) given by

$$\begin{aligned} \pi(K|t_1,\ldots,t_n) \quad &\propto \quad K^{a_4+\frac{n(n-1)}{2}-1}(1-K)^{b_4-1} \times \\ &\times \quad \int_0^\infty e^{-nh_k(D)}dD\ , \end{aligned} \tag{24}$$

where $-nh_k(D) = (n+a_3-1)lnD - \left(b_3 + \sum_{i=1}^{n} K^{i-1}t_i\right) D$.

An approximate marginal posterior density for K (see appendix) is given by,

$$\pi(K|t_1,\ldots,t_n) \propto \frac{K^{a_4+\frac{n(n-1)}{2}-1}(1-K)^{b_4-1}}{\left(b_3+\sum_{i=1}^{n} K^{i-t}t_i\right)^{n+a_3}} \tag{25}$$

where $0 < K < 1$.

The marginal posterior density for D is (from (23)) given by

$$\pi(D|t_1,\ldots,t_n) \propto D^{n+a_3-1}e^{-b_3D}\int_0^1 e^{-nh_D(K)}dK \quad (26)$$

where

$$\begin{aligned}-nh_D(K) &= \left(a_4+\frac{n(n-1)}{2}-1\right)lnK+(b_4-1)\,ln(1-K)- \\ &- D\textstyle\sum_{i=1}^n K^{i-1}t_i.\end{aligned}$$

A Laplace's approximate marginal posterior density (see (A.8) in appendix) for D is given by

$$\begin{aligned}\pi(D|t_1,\ldots,t_n) &\propto D^{n+a_3-1}e^{-b_3D}\widehat{\sigma}\;e^{-nh_D(\widehat{k})}\times \\ &\times\left\{\Phi\left(\frac{1-\widehat{K}}{\widehat{\sigma}}\right)-\Phi\left(-\frac{\widehat{K}}{\widehat{\sigma}}\right)\right\},\end{aligned} \quad (27)$$

where $\widehat{K}$ maximizes $-nh_D(K)$ for each value of D and

$$\widehat{\sigma}=\left\{\frac{\left(a_4+\frac{n(n-1)}{2}-1\right)}{\widehat{K}^2}+\frac{(b_4-1)}{\left(1-\widehat{K}\right)^2}+D\sum_{i=1}^n(i-1)(i-2)\widehat{K}^{i-3}t_i\right\}^{-1/2}.$$

Assuming K known, the predictive density for a future failure time is given by

$$f(t_{n+1}|t_1,\ldots,t_n)\propto\left\{b_3+\sum_{i=1}^n K^{i-1}t_i+K^nt_{n+1}\right\}^{-(n+a_3+1)} \quad (28)$$

where $t_{n+1}\geq 0$.

5 The Littlewood-Verrall Model

From (8) and (9), with $\psi(i)$ completely known, the posterior density for α is given by

$$\pi(\alpha|t_1,\ldots,t_n) \propto f(t_1,\ldots,t_n|\alpha)\,\pi_0(\alpha) \tag{29}$$

where $\pi_0(\alpha)$ is a prior density for α.

The likelihood function for α is given (see Littlewood and Verrall, 1973) by,

$$f(t_1,\ldots,t_n|\alpha) = \prod_{i=1}^{n}\int_0^\infty f(t_i|\lambda)\,\pi_i(\lambda|\alpha,\psi(i))\,d\lambda \tag{30}$$

where π_i is the prior density (9) and

$$f(t_i|\lambda) = \lambda e^{-\lambda t_i}, t_i \geq 0 \ .$$

The posterior density for α is given by,

$$\pi(\alpha|t_1,\ldots,t_n) \propto \left\{\prod_{i=1}^{n}\int_0^\infty f(t_i|\lambda)\,\pi_i(\lambda|\alpha,\psi(i))\,d\lambda\right\}\pi_0(\alpha)\ . \tag{31}$$

Assuming a noninformative prior $\pi_0(\alpha) \propto 1/\alpha, \alpha > 0$, the posterior density for α is given by,

$$\pi(\alpha|t_1,\ldots,t_n) = \frac{\gamma^n}{\Gamma(n)}\,\alpha^{n-1}e^{-\gamma\alpha} \tag{32}$$

where $\alpha > 0$ and $\gamma = \sum_{i=1}^{n} ln\left(1+\frac{t_i}{\psi(i)}\right)$.

The marginal posterior density for λ_{n+1} is given by,

$$\pi(\lambda_{n+1}|t_1,\ldots,t_n) \propto \int_0^\infty \pi_{n+1}(\lambda|\alpha,\psi(n+1))\,\pi(\alpha|t_1,\ldots,t_n)\,d\alpha \tag{33}$$

That is,

$$\pi(\lambda_{n+1}|t_1,\ldots,t_n) \propto \frac{e^{-\psi(n+1)\lambda_{n+1}}}{\lambda_{n+1}}\int_0^\infty e^{-nh_{\lambda_{n+1}}(\alpha)}d\alpha, \tag{34}$$

where $-nh_{\lambda_{n+1}}(\alpha) = (n-1)ln\alpha - \alpha A(\lambda_{n+1}) - ln\Gamma(\alpha)$ and $A(\lambda_{n+1}) = \gamma - ln\lambda_{n+1} - ln\psi(n+1)$.

A Laplace's approximate marginal posterior density for λ_{n+1} is given by,

$$\pi(\lambda_{n+1}|t_1,\ldots,t_n) \propto \frac{e^{-\psi(n+1)\lambda_{n+1}}\widehat{\alpha}^{n-1}e^{-\widehat{\alpha}A(\lambda_{n+1})}}{\lambda_{n+1}\Gamma(\widehat{\alpha})\left\{\frac{n-1}{\widehat{\alpha}^2}+\xi^{(1)}(\widehat{\alpha})\right\}^{1/2}}, \tag{35}$$

where $\lambda_{n+1} > 0, \widehat{\alpha}$ maximizes $-nh_{\lambda_{n+1}}(\alpha)$ for each value or λ_{n+1} and $\xi^{(1)}(\alpha) = d^2 ln\Gamma(\alpha)/d\alpha^2$ (a trigamma function).

The predictive density for a future failure time T_{n+1} is given by

$$\begin{aligned} f(t_{n+1}|t_1,\ldots,t_n) &= \int_0^\infty f(t_{n+1}|\lambda)\,\pi(\lambda_{n+1}|t_1,\ldots,t_n)\,d\lambda \\ &= \int_0^\infty (t_{n+1}|\lambda)\left\{\int_0^\infty \pi_{n+1}(\lambda|\alpha,\psi(n+1))\,\pi(\alpha|t_1,\ldots,t_n)\,d\alpha\right\}d\lambda \end{aligned} \tag{36}$$

(see Littlewood and Verrall, 1973).

Changing the order of integration in (36), we get

$$f(t_{n+1}|t_1,\ldots,t_n) = \frac{n\gamma^n\left\{\gamma + ln\left(1+\frac{t_{n+1}}{\psi(n+1)}\right)\right\}^{-(n+1)}}{(t_{n+1}+\psi(n+1))} \tag{37}$$

where $t_{n+1} \geq 0$ and γ is given in (32).

When we do not have $\psi(i)$ completely known, we could consider a parametric family $\psi(\underset{\sim}{\beta}, i)$, where $\underset{\sim}{\beta}$ is a vector of unknown parameters.

In this case, we also assume an exponential density (8) for the i-th time between failures and we consider Λ_i with a gamma density,

$$\pi_i\left(\lambda|\alpha,\underset{\sim}{\beta}\right) = \frac{\left[\psi\left(\underset{\sim}{\beta},i\right)\right]^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\psi\left(\underset{\sim}{\beta},i\right)\lambda} \tag{38}$$

where $\lambda > 0$.

The joint posterior density for α and $\underset{\sim}{\beta}$ (see (31)) is given by

$$
\begin{aligned}
\pi\left(\alpha, \underset{\sim}{\beta} | t_1, \ldots, t_n\right) &= \\
&= \alpha^n \left\{ \prod_{i=1}^n \left(\frac{\psi\left(\underset{\sim}{\beta}, i\right)}{t_i + \psi\left(\underset{\sim}{\beta}, i\right)} \right)^{\alpha} \left(\frac{1}{t_i + \psi\left(\underset{\sim}{\beta}, i\right)} \right) \right\} \pi_0\left(\alpha, \underset{\sim}{\beta}\right)
\end{aligned}
\tag{39}
$$

Assuming $\psi\left(\underset{\sim}{\beta}, i\right) = \beta_0 + \beta_1 i$, we have,

$$
\begin{aligned}
\pi(\alpha, \beta_0, \beta_1 | t_1, \ldots, t_n) &\propto \\
&\propto \alpha^n \left\{ \prod_{i=1}^n \left(\frac{\beta_0 + \beta_1 i}{t_i + \beta_0 + \beta_1 i} \right)^{\alpha} \left(\frac{1}{t_i + \beta_0 + \beta_1 i} \right) \right\} \pi_0(\alpha, \beta_0, \beta_1)
\end{aligned}
\tag{40}
$$

Assuming a noninformative prior density $\pi_0(\alpha, \beta_0, \beta_1) \propto 1/(\alpha\beta_0\beta_1)$, the joint posterior density for α, β_0 and β_1 is given by

$$
\pi(\alpha, \beta_0, \beta_1 | t_1, \ldots, t_n) \propto \frac{\alpha^{n-1} B(\beta_0, \beta_1) e^{-\alpha A(\beta_0, \beta_1)}}{\beta_0 \beta_1}
\tag{41}
$$

where

$$
\alpha, \beta_0, \beta_1 > 0, B(\beta_0, \beta_1) = \prod_{i=1}^n \left(\frac{1}{t_i + \beta_0 + \beta_1 i} \right)
$$

and

$$
A(\beta_0, \beta_1) = \sum_{i=1}^n \ln \left(\frac{t_i + \beta_0 + \beta_1 i}{\beta_0 + \beta_1 i} \right) .
$$

Integrating out α in (41), we get the joint marginal posterior density for β_0 and β_1,

$$
\pi(\beta_0, \beta_1 | t_1, \ldots, t_n) \propto \frac{B(\beta_0, \beta_1)}{\beta_0 \beta_1 [A(\beta_0, \beta_1)]^n} .
\tag{42}
$$

To get the marginal posterior densities for β_0, β_1 and λ_{n+1}, or the predictive density for a future observation T_{n+1}, we should use a numerical or approximation method (see appendix).

6 An Example

In table 1, we have a data set introduced by Jelinski and Moranda (1972). The data consists of the number of days between the 26 failures that ocurred during the production phase of a software (NTDS data-Naval tactical data system).

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		

Table 1 - NTDS data

Considering the JM model (2) for this data set, the maximum likelihood estimators for the parameters are given by $\widehat{\Lambda} = 0.006904$ and $\widehat{N} = 31.1$. Using the standard asymptotical normality for the maximum likelihood estimators based on the Fisher information matrix (see for example, Lawless, 1982), approximate 95% confidence intervals for Λ and N are given by $0.0007 < \Lambda < 0.0131$ and $20.1225 < N < 42.0775$. Usually, the accuracy of these asymptotical inferences depend on the data set and on an appropriate parametrization (see Sprott, 1973).

For a Bayesian analysis of the JM model, assume that the software expert has a prior opinion, given by $E(N) = 30, var(N) = 9, E(\Lambda) = 0.010$ and $var(\Lambda) = 0.0005$. From this prior opinion, consider the joint prior density for Λ and N (13) with $a_1 = 0.2, b_1 = 20, a_2 = 100$ and $b_2 = 3.33$.

With this choice of prior, the joint posterior for Λ and N (see(15)) is given by

$$\pi_1(\Lambda, N|t_1, \ldots, t_{26}) \propto \Lambda^{25.2} N^{99} A(N) exp\{-3.33N - (20 + B(N))\Lambda\} , \qquad (43)$$

where $\Lambda > 0, N \geq 26, A(N) = \prod_{i=1}^{26}(N - i + 1)$ and

$$B(N) = \sum_{i=1}^{26}(N - i + 1)t_i.$$

The marginal posterior density for N (see(16)) is given by

$$\pi_1(N|t_1, \ldots, t_{26}) \propto \frac{N^{99} A(N) e^{-3.33N}}{\{20 + B(N)\}^{26.2}} \qquad (44)$$

where $N \geq 26$.

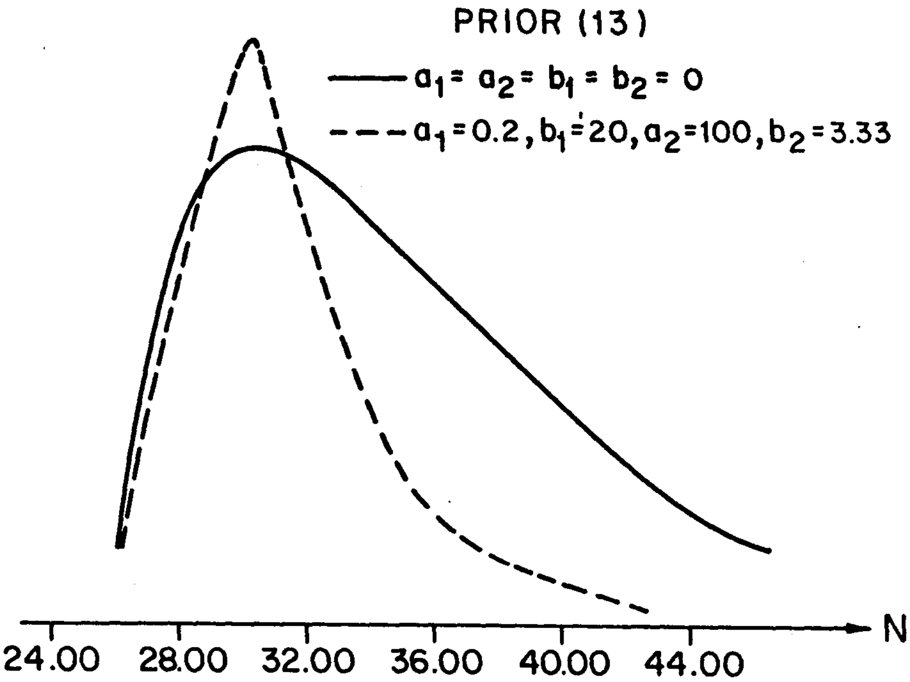

Figure 1 - Marginal Posterior Density for N

In figure 1, we have the graph of the marginal posterior density (45). The mode of (44) is $\widehat{N} = 30.1$. We also have in figure 1, the graph of the marginal posterior density for N considering the noninformative prior density (14). In this case, the mode of the marginal posterior density for N is given by $\widehat{N} = 30.3$ We observe in figure 1, better inference for N assuming the prior density (13), with $a_1 = 0.2, b_1 = 20, a_2 = 100$ and $b_2 = 3.33$.

From (43), an approximate marginal posterior density for Λ (see (18)) is given by

$$\pi_1(\Lambda|t_1,\ldots,t_{26}) \propto \Lambda^{25.2}\widehat{N}^{99}A\left(\widehat{N}\right)\widehat{\sigma}\left\{1-\Phi\left(\frac{26-\widehat{N}}{\widehat{\sigma}}\right)\right\} \tag{45}$$

where

$$\Lambda > 0, \widehat{\sigma} = \left\{-\frac{0.8}{\widehat{N}^2} + \sum_{i=1}^{26}\frac{1}{\left(\widehat{N}-i+1\right)^2}\right\}^{-1/2},$$

and $\widehat{N}$ maximizes $-0.8N + lnA(N) - 3.33N - (20 + B(N))\Lambda$, for each value of Λ.

In figure 2, we have the graph of (45). The mode of (45) is given by $\widehat{\Lambda} = 0.014$.

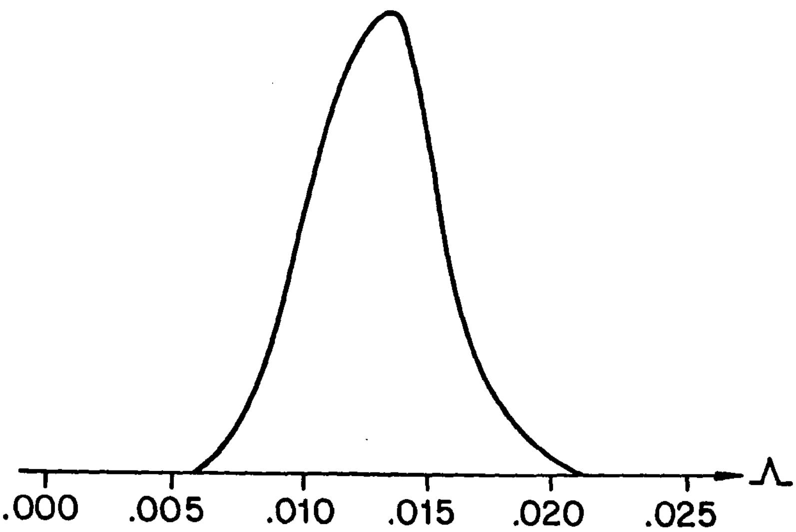

Figure 2 - Marginal Posterior Density for Λ.

Assuming $N = 31$ known, the predictive density (see (20)) for a future failure T_{27} with $b_1 = 20$, is given by

$$f_1\left(t_{27}|t_1,\ldots,t_{26}\right) = \frac{1.7715e^{220}}{\left(3762+5t_{27}\right)^{27.2}} \tag{46}$$

where $t_{27} \geq 0$.

Using Simpson's rule, we get the posterior mean $E(T_{27}|t_1,\ldots,t_{26}) = 29.8492$. A numerically obtained 90% Bayesian interval for T_{27} is given by (1.475;91.120).

Considering the Moranda model (4), the maximum likelihood estimators for D and K are given by $\widehat{D} = 0.2112$ and $\widehat{K} = 0.951$.

For a Bayesian analysis of the Moranda model, let us assume that the prior opinion of a software expert is given by $E(K) = 0.92, var(K) = 0.001, E(D) = 0.22$ and $var(D) = 0.002$. Thus, consider the prior (22) with $a_3 = 24.2, b_3 = 110, a_4 = 66.8$ and $b_4 = 5.8$.

With this choice of prior, the joint posterior density dor D and K (see(23)) is given by

$$\pi(D,K|t_1,\ldots,t_{26}) \propto D^{49.2}K^{390.8}(1-K)^{4.8} \times \times \; exp\left\{-\left(110+\sum_{i=1}^{26} K^{i-1}t_i\right)D\right\} \tag{47}$$

where $D > 0$ and $0 < K < 1$.

An approximate marginal posterior density for K (see (25)) is given by,

$$\pi(K|t_1,\ldots,t_{26}) \propto \frac{K^{390.8}(1-K)^{4.8}}{\left\{110+\sum_{i=1}^{26} K^{i-1}t_i\right\}^{50.2}} \tag{48}$$

where $0 < K < 1$.

The mode of (48) is given by $\widehat{K} = 0.947$ (see figure 3).

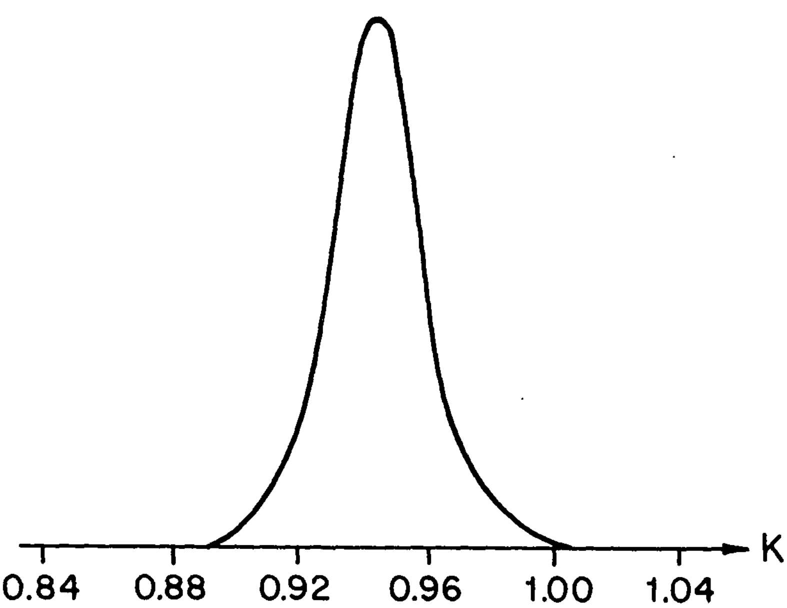

Figure 3 - Marginal Posterior Density for K

Assuming $K = 0.95$ known, the predictive density for a future observation T_{27} (see (28)) is given by

$$f(t_{27}|t_1,\ldots,t_{26}) = \frac{0.01519\; e^{280}}{(231.1+0.2635t_{27})^{51.2}} \tag{49}$$

where $t_{27} \geq 0$.

Using Simpson's rule, we get the predictive mean value $E(T_{27}|t_1,\ldots,t_{26}) = 17.8239$. A numerically obtained 90% Bayesian interval for T_{27} is given by $(0.9; 54.0)$.

Considering the Littlewood-Verrall model (see (8) and (9)) for the NTDS data of table 1, the posterior density for α (see (32)) assuming $\psi(i) = i$ is given by

$$\pi\left(\alpha|t_1,\ldots,t_{26}\right) = 3.0803\left(10^5\right)\alpha^{25}e^{-15.1346\alpha} \tag{50}$$

where $\alpha > 0$.

The mode of (50) is given by $\tilde{\alpha} = 1.65$ (see figure 4). The posterior mean for (50) is given by $E(\alpha|t_1,\ldots,t_{26}) = 1.718$

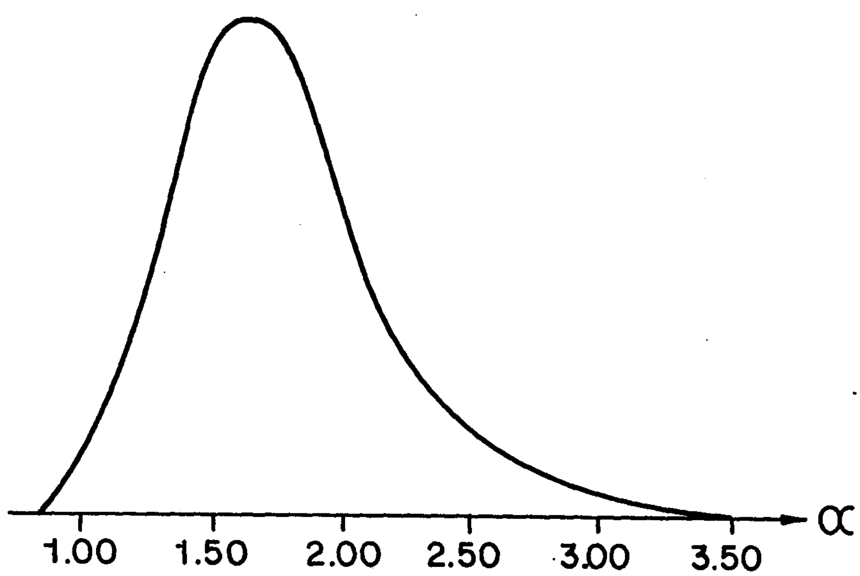

Figure 4 - Posterior density for α with $\psi(i) = i$

Also assuming $\psi(i) = i$, an approximate marginal posterior density for λ_{27} (see (35)) with $\xi^{(1)}(\alpha) \approx \frac{1}{\alpha}(1 + \frac{1}{2\alpha})$ (see Abramowitz e Stegun, 1972) is given by

$$\pi\left(\lambda_{27}|t_1,\ldots,t_{26}\right) \propto \frac{\lambda_{27}^{\hat{\alpha}-1}\hat{\alpha}^{26}e^{-11.84\hat{\alpha}-27\lambda_{27}}}{\Gamma\left(\hat{\alpha}\right)\left\{2\hat{\alpha}+51\right\}^{1/2}} \tag{51}$$

where $\lambda_{27} > 0$ and $\hat{\alpha}$ maximizes $-25ln\alpha - 11.839\alpha + \alpha ln\lambda_{27} - ln\Gamma(\alpha)$, for each value of λ_{27}. The mode of (51) is given by $\tilde{\lambda}_{27} = 0.024$ (see figure 5).

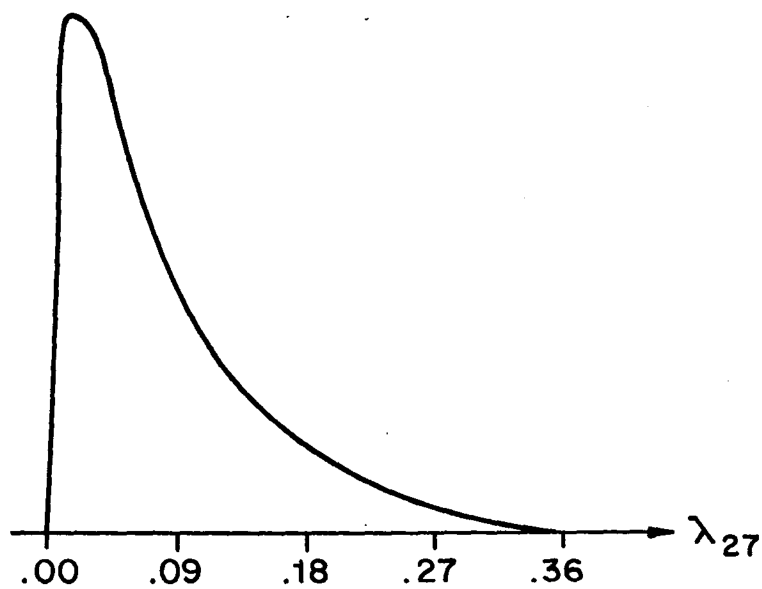

Figure 5 - Posterior density for λ_{27} with $\psi(i) = i$.

The predictive density for a future observation T_{27} (see (37) with $\psi(i) = i$) is given by

$$f\left(t_{27}|t_1, \ldots, t_{26}\right) = \frac{26(15.1346)^{26}}{(27+t_{27})\left\{15.1346 + ln\left(1+\frac{t_{27}}{27}\right)\right\}^{27}} \quad (52)$$

where $t_{27} \geq 0$.

A Laplace's approximation for the predictive mean of (52) is given by

$$\tilde{E}\left(T_{27}|t_1, \ldots, t_{26}\right) = 14.0305 \; .$$

Observe that we could consider any other choice for $\psi(i)$.

7 Concluding Remarks

The use of approximate Bayesian methods could be a suitable alternative to get accurate inferences for software reliability. We also could use Bayesian procedures to discriminate software reliability models (see for example, Gelfand and Dey, 1994; Geisser and Edy, 1979; Aitkin, 1991 or Berger and Pericchi, 1992).

A simple procedure to compare two models I and II (see for example, Box, 1980; or Mazzuchi and Soyer, 1998), is to consider the ratio.

$$R_1 = \prod_{i=1}^{n} \frac{f(t_i|t_1,\ldots,t_{i-1},I)}{f(t_i|t_1,\ldots,t_{i-1},II)} \tag{53}$$

where $f(t_i|t_1,\ldots,t_{i-1},I)$ and $f(t_i|t_1,\ldots,t_{i-1},II)$ are obtained by replacing T_i by its observed value t_i in the predictive distribution of T_i given $t_1,\ldots,t_{i-1}$ for models I and II, respectively.

If R_1 is greater than 1, then model I is preferable to model II; otherwise the reverse is true.

Equation (53) provides a global measure for comparing the two models. An alternative procedure is to use a local measure at each stage, given by

$$R_2 = \frac{f(t_i|t_1,\ldots,t_{i-1},I)}{f(t_i|t_1,\ldots,t_{i-1},II)} \tag{54}$$

As an illustrative example, let us assume the comparison of JM model (2) with Moranda model (4), for the data set of table 1. The ratio of predictive densities (46) and (49) is given by

$$\begin{aligned} R_2 &= \frac{f(t_{27}|t_1,\ldots,t_{26},JM)}{f(t_{27}|t_1,\ldots,t_{26},M)} \\ &= \frac{116.6228e^{-60}(231.1+0.2635t_{27})^{51.2}}{(3762+5t_{27})^{27.2}} \end{aligned} \tag{55}$$

If the next failure time is $T_{27} = 47$, we get $R_2 = 1.6955$, that is, the JM model is preferable to Moranda model, at this stage.

APPENDIX

Laplace's Method for Approximation of Integrals

Assuming h is a smooth function of an m-dimensional parameter θ with -h having a maximum at $\hat{\theta}$, Laplace's method approximates an integral of the form,

$$I_1 = \int f(\theta) exp\left[-nh(\theta)\right] d\theta \tag{A.1}$$

by expanding h and f in a Taylor series about $\hat{\theta}$ (see for example, Kass, Tierney and Kadane, 1990).

Considering first the case in which θ is one-dimensional, Laplace's method gives the approximation,

$$\hat{I}_1 \cong (2\pi)^{1/2}\sigma f\left(\hat{\theta}\right) exp\left\{-nh\left(\hat{\theta}\right)\right\} \tag{A.2}$$

where $\sigma = \left\{nh''\left(\hat{\theta}\right)\right\}^{-1/2}$.

In multiparameter case, with $\theta \in R^m$, we have

$$\hat{I}_1 \cong (2\pi)^{m/2}\left\{det\left(nD^2h\left(\hat{\theta}\right)\right)\right\}^{-1/2} f\left(\hat{\theta}\right) exp\left\{-nh\left(\hat{\theta}\right)\right\} \tag{A.3}$$

where $\hat{\theta}$ maximizes $-h(\theta)$ and $D^2h(\hat{\theta})$ is the Hessian matrix of h evaluated at $\hat{\theta}$.

A special case of Laplace's approximations is given with $f = 1$ (see Tierney and Kadane, 1986).

Laplace's method for approximation of integrals assumes that the main contribution to the relevant integrals comes from a peak in the function $e^{-nh(\theta)}$ defined in the entirely real line.

For definite integrals, we could get some simple extensions of Laplace's approximations.

As a special case, consider the approximation of the integral,

$$I_2 = \int_{-\infty}^{a} e^{-nh(\theta)} d\theta \tag{A.4}$$

where $\hat{\theta}$ which maximizes $-nh(\theta)$ is larger or close to the constant a. In this case, by expanding h in a Taylor series about $\hat{\theta}$, we have,

$$I_2 \cong e^{-nh(\hat{\theta})} \int_{-\infty}^{a} e^{-\frac{nh''\left(\hat{\theta}\right)}{2}(\theta-\hat{\theta})} d\theta \tag{A.5}$$

Since $\int_{-\infty}^{a} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \sqrt{2\pi}\sigma\Phi\left(\frac{a-\mu}{\sigma}\right)$, where Φ denotes the distribution function of a standard normal distribution $N(0,1)$, we get the approximation,

$$\widehat{I}_2 \cong \sqrt{2\pi}\widehat{\sigma}e^{-nh(\widehat{\theta})}\Phi\left(\frac{a-\widehat{\theta}}{\widehat{\sigma}}\right) \tag{A.6}$$

where $\widehat{\sigma} = \left\{nh''\left(\widehat{\theta}\right)\right\}^{-1/2}$

Observe that, if $(a-\widehat{\theta})/\widehat{\sigma} > 3$, Laplace's approximation (A.6) reduces to (A.2), since $\Phi\left(\frac{a-\widehat{\theta}}{\widehat{\sigma}}\right) \cong 1$. Also, observe that if $a = \widehat{\theta}$, that is, the maximum of $-h$ is at the boundary of the integration interval, $\Phi\left(\frac{a-\widehat{\theta}}{\widehat{\sigma}}\right) = 0.5$, and Laplace's approximation for I_2 is given by

$$\widehat{I}_2 \cong \sqrt{\frac{\pi}{2}}\,\widehat{\sigma}\,e^{-nh(\widehat{\theta})} \tag{A.7}$$

Similar results are obtained for other cases. For example,

$$\int_a^{\infty} e^{-nh(\theta)}d\theta \cong \sqrt{2\pi}\widehat{\sigma}\left\{1-\Phi\left(\frac{a-\widehat{\theta}}{\widehat{\sigma}}\right)\right\} \tag{A.8}$$

and,

$$\int_a^b e^{-nh(\theta)}d\theta \cong \sqrt{2\pi}\widehat{\sigma}\left\{\Phi\left(\frac{b-\widehat{\theta}}{\widehat{\sigma}}\right)-\Phi\left(\frac{a-\widehat{\theta}}{\widehat{\sigma}}\right)\right\},$$

where $\widehat{\theta}$ maximizes $-h(\theta)$ and $\sigma = \left\{nh''\left(\widehat{\theta}\right)\right\}^{-1/2}$.

References

Abramowitz, M.; Stegun, I. (1972) - Handbook of mathematical functions, Dover.

Aitkin, M. (1991) - Posterior factors. Journal of the Royal Statistical Society, B, 53, 111-142.

Berger, J.O.; Pericchi, L.R. (1993) - The intrinsic Bayes factor for model selection and prediction, technical report 93-43c, dep. of Statistics, Purdue University.

Box, G. (1980) - Sampling and Bayes' inference in scientific modelling and robustness, Journal of the Royal Statistical Society, A, 143, 382-430.

Geisser, S. ; Eddy, W. - A predictive approach to model selection, Journal of the American Statistical Association, 74, 153-160.

Gelfand, A.E.; Dey, D.K. (1994) - Bayesian model choice: asymptotics and exact calculations, Journal of the Royal Statistical Society, B, 56,3,501-514.

Gelfand, A.E.; Smith, A.F.M. (1990) - Sampling-based approaches to calculating marginal densities, Journal of the American Statistical Association, 85, 398-409.

Goel, A.L. (1983) - A guidebook for software reliability assessment, thechnical report RADC-TR-83-176.

Goel, A.L.; Okumoto, K. (1978) - An analysis of recurrent software failures ou a real-time control system, 496-500. Proc. ACM Ann. Tech. Conf., Washington D.C.

Goel, A.L.; Okumoto, K. (1979) - Time-dependent error detection rate model for software reliability and other performance measures, IEEE Transactions on Reliability, R-28, 206-211.

Jelinski, Z; Moranda, P. (1972) - Software reliability research, Statistical Computer Performance Evaluation, Ed. W. Freiberger, 465-484, New York: Academic.

Kass, R.E.; Tierney, L.; Kadane, J.B. (1990) - The validity of posterior expansions based on Laplace's method, in Essays in Honor of G.A. Barnard, ed. J. Hodges, Amsterdam: North-Holland, 473-488.

Kloek, T.; Van Dijk, H.K. (1978) - Bayesian estimates of equation system parameters: an application of integration by Monte Carlo, Econometrika, 46, 1-19.

Lawless, J.F. (1982) - Statistical models and methods for lifetime data. New York: John Wiley & Sons.

Littlewood, B.; Verrall, J.L. (1973) - A Bayesian reliability growth model for computer software, Applied Statistics, 22, 332-346.

Mazzuchi, T.A.; Soyer, R. (1988) - A Bayes Empirical-Bayes model for software reliability, IEEE Transactions on Reliability, R-37:2, 248-258.

Moranda, P.B. (1975) - Prediction of software reliability and its applications, Proceedings of the Annual Reliability and Maintainability Symposium, 327-332, Washington, D.C.

Musa, J.D.; Okumoto, K (1984) - A Logarithmic Poisson execution time model for software reliability measurement, Preceedings of the 7th International Conference of Software Engineering, Orlando, 230-237.

Naylor, J.C.; Smith, A.F.M. (1982) - Applications of a method for the efficient computation of posterior distributions, Applied Statistics, 31, 214-225.

Schick, G.J.; Wolverton, R.W. (1978) - Assessment of Software Reliability, Proc. Oper. Res., 395-422, Wirzberg- Wien: Physica-Verlag.

Singpurwalla, N.D.; Wilson, S.P. (1994) - Software Reliability Modeling, International Statistical Review, 62,3, 289-317.

Sprott, D.A. (1980) - Maximum likelihood in small samples: estimation in the presence of nuisance parameters, Biometrika, 67, 515-523.

Tierney, L.; Kadane, J.B. (1986) - Accurate approximations for posterior moments and marginal densities, Journal of the American Statistical Association, 81, 82-86.