The Korean Journal of Applied Statistics (2008)
21(5), 835–843

Bayesian Inferences for Software Reliability Models Based on Beta-Mixture Mean Value Functions

Seungmin Nam[1] · Kiwoong Kim[2] · Sinsup Cho[3] · In-Kwon Yeo[4]

[1]Samsung Fire & Marine Insurance Co.; [2]Samsung Fire & Marine Insurance Co.;
[3]Dept. of Statistics, Seoul National University; [4]Dept. of Statistics, Sookmyung Women's University

(Received July 2008; accepted September 2008)

Abstract

In this paper, we investigate a Bayesian inference for software reliability models based on mean value functions which take the form of the mixture of beta distribution functions. The posterior simulation via the Markov chain Monte Carlo approach is used to produce estimates of posterior properties. Its applicability is illustrated with two real data sets. We compute the predictive distribution and the marginal likelihood of various models to compare the performance of them. The model comparison results show that the model based on the beta-mixture performs better than other models.

Keywords: Beta-mixture, MCMC, mean value function, nonhomogeneous Poisson processes.

1. Introduction

As the importance of computers is highly appreciated, software, one of the main components of computers, not only plays an important role but becomes more complicated. As a consequence, the errors inside software which are not detected can do serious harm. Theoretically, it is possible to make software error-free, but finding software faults is difficult as well as expensive.

The evaluation of software reliability is essential to produce software of good quality, quickly and efficiently. For example, with a good software reliability model, programmers can determine when to release their software package more easily and rationally.

Many statistical models have been developed to measure software reliability and utilized to estimate the failure intensity of software and predict the quality of software. Jelinski and Moranda (1972) model is the first model to be widely used. It assumes that the failure rate of the interfailure times is proportional to the number of errors remaining in the software code. Goel and Okumoto (1979) model is based on a nonhomogeneous Poisson process. It assumes that the expected number of

[1] Samsung Fire & Marine Insurance Co., Seoul 100-782, Korea.
[2] Samsung Fire & Marine Insurance Co., Seoul 100-782, Korea.
[3] Professor, Dept. of Statistics, Seoul National University, Seoul 151-742, Korea. E-mail: sinsup@snu.ac.kr
[4] Corresponding Author: Associate Professor, Dept. of Statistics, Sookmyung Women's University, Seoul 140-742, Korea. E-mail: inkwon@sm.ac.kr

failures in the interval is proportional to the product of the expected number of undetected failures and the length of the interval. The approach that uses a nonhomogeneous Poisson process for modeling the number of failures has been quite popular in recent years. Recent remarkable models are Kuo and Yang's (1996) general order statistics model and the model by Basu and Ebrahimi (2003). The general order statistics model assumes that the total number of failures is N and the observed epochs of failures are the first n order statistics taken from N *i.i.d.* observations with distribution supported in R^+. Basu and Ebrahimi (2003) propose Bayesian software reliability models based on martingale processes.They assume that the current failure rate is simply equal to the previous failure rate on average.

This paper is organized as follows. Section 2 introduces the beta-mixture model and its motivation. The Bayesian inferences are also developed. We illustrate real data analysis in two different cases in Section 3. Comparisons are also made with existing well-known software reliability models.

2. Beta-Mixture Model and Bayesian Inferences

Let $N(t)$ be the number of failures of the software observed during time $(0,t]$. Usually, $N(t)$ is modelled by a nonhomogeneous Poisson process with the mean value function $m(t) = E[N(t)]$ and the intensity function $\mu(t)$ which is the derivative of $m(t)$. The probability density function of $N(t)$ is given as

$$P(N(t) = n) = \frac{m(t)^n}{n!}e^{-m(t)}, \quad n = 0,\, 1,\, 2, \cdots . \tag{2.1}$$

Nonhomogeneous Poisson processes(NHPP) can be classified into two classes according to the limiting behavior of $m(t)$. The processes with $\lim_{t\to\infty} m(t) < \infty$ and $\lim_{t\to\infty} m(t) = \infty$ are called NHPP-I and NHPP-II, respectively. Members of NHPP-I are studied by Goel and Okumoto (1979), Goel (1983) and Ohba *et al.* (1982) and NHPP-II by Musa and Okumoto (1984), Duane (1964) and Cox and Lewis (1966). In this paper, we deal with the mean value function which can belong to either NHPP-I or NHPP-II according to the selection of the distribution function.

Since the mean value function is strictly increasing, it can be modelled by a distribution function; see Mallick and Gelfand (1994). A mixture model provides a dense class of distribution functions which relies on standard distributions as functional bases and is used to approximate the true distribution function as follows:

$$F(x) \simeq \hat{F}(x) = \sum_{i=1}^{L} \omega_i F_i(x \,|\, \theta_i),$$

where L is the number of mixands and $\omega_i \geq 0$ and $\sum_{i=1}^{L} \omega_i = 1$. As argued by Diaconis and Ylvisaker (1985), an unknown distribution function can be modelled by a mixture of beta distribution functions.

Let $B(u; c, d)$ be the beta distribution function with parameters c and d evaluated at u. Kim *et al.* (2006) propose the mean value function based on the beta-mixture

$$m(t) = G^{-1}\left[\sum_{l=1}^{L} \omega_l B\{F_0(t);\, \sigma l,\, \sigma(L+1-l)\}\right], \tag{2.2}$$

where F_0 denotes a centering distribution function and G^{-1} is the inverse function of a distribution function G which may be indexed by unknown parameters or given constants $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)$, usually $p = 1$ or 2. The functions F_0 and G can be specified by users. Note that if G^{-1} is upper bounded, then $m(t)$ is NHPP-I, otherwise NHPP-II. The prior information about the data may be helpful in choosing F_0 and G and in our experience, the uniform distribution is sufficient for many cases.

There are two types of failure data in software reliability models, interval-domain data and time-domain data. Interval-domain data, also called grouped data or error count data, are composed of the number of failures n_i during a fixed time interval $(0, t_i]$, $i = 1, 2, \dots, m$ and $0 < t_1 < \cdots < t_m$ and the joint probability density function is given as

$$f(\boldsymbol{n};\, \theta) = \exp\{-m_\theta(t_m)\} \prod_{i=1}^{m} \frac{m_\theta(t_{i-1},\, t_i)^{n_i - n_{i-1}}}{(n_i - n_{i-1})!}, \tag{2.3}$$

where $n_0 = 0$, $t_0 = 0$ and $m_\theta(t_{i-1}, t_i) = m_\theta(t_i) - m_\theta(t_{i-1})$. Time-domain data, also called interfailure data or failure times data, consist of ordered epochs $0 < t_1 < \cdots < t_m$, sometimes, with a terminal time T and the corresponding joint probability density function is written as

$$f(\boldsymbol{t};\, \theta) = \exp\{-m_\theta(t_m)\} \prod_{i=1}^{m} \mu_\theta(t_i), \tag{2.4}$$

where $m_\theta(t_m)$ is replaced by $m_\theta(T)$ for the time truncated model.

Kim *et al.* (2006) sketch out the EM-algorithm for estimating the parameters and apply it to the error count data of Tohma *et al.* (1991). In this paper, we present the Bayesian inference using Gibbs sampling which provides an easy implementation for the inferences of the parameters. Let D be the observed data and let $\pi(\theta)$ be a joint prior density of θ. Then, the implementation of Bayesian inferences is based on the posterior density of θ, $\pi(\theta \mid D) \propto f(D;\, \theta)\pi(\theta)$. For the selection of L and the model comparison, a criterion such as the predictive distribution and the Bayes factor can be used.

3. Applications

3.1. Error count model

Software failures data based on 111 observations was reported by Tohma *et al.* (1991) to test the error count model. The program consists of about 200 modules and the modules have, on average, 1000 lines of a high-level computer language. For these data, F_0 and G were assumed to be uniform$(0, 115)$ and uniform$(0, \lambda)$, respectively. We may assume either the mixture weight to be random or the parameters of the beta densities to be random. Given L, it is mathematically simpler to work with the mixture weight; see Gelfand and Mallick (1995). For that reason, we fix the beta density parameters and assume that mixture weights are random. The parameter $\sigma = 1$ is fixed and the number of mixands $L = 3$, 4 and 5 are considered. Then, the corresponding mean value function is $m_\theta(t) = \lambda A_\omega(t)$, where $A_\omega(t) = \sum_{l=1}^{L} \omega_l B\{F_0(t);\, l,\, L+1-l\}$ and in the case of $L = 3$,

$$A_\omega(t) = \omega_1 \left[1 - \{1 - F_0(t)\}^3\right] + \omega_2 F_0^2(t)\,\{3 - 2F_0(t)\} + F_0^3(t).$$

We assume that λ and ω are independent and

$$\pi(\lambda) \sim \text{Gamma}(1,\, 0.001), \qquad \pi(\omega) \sim \text{Dirichlet}(1, \dots, 1).$$

Note that the prior for λ is the diffuse prior and the Dirichlet prior for ω is the vague prior.

Table 3.1. Posterior summaries of parameters

		λ	ω
$L=3$	Mean	482.07	(0.98, 0.01, 0.01)
	S.D.	21.82	(0.01, 0.01, 0.01)
$L=4$	Mean	482.47	(0.77, 0.21, 0.01, 0.01)
	S.D.	21.96	(0.06, 0.06, 0.01, 0.01)
$L=5$	Mean	482.70	(0.42, 0.54, 0.02, 0.01, 0.01)
	S.D.	21.86	(0.06, 0.06, 0.01, 0.01, 0.01)

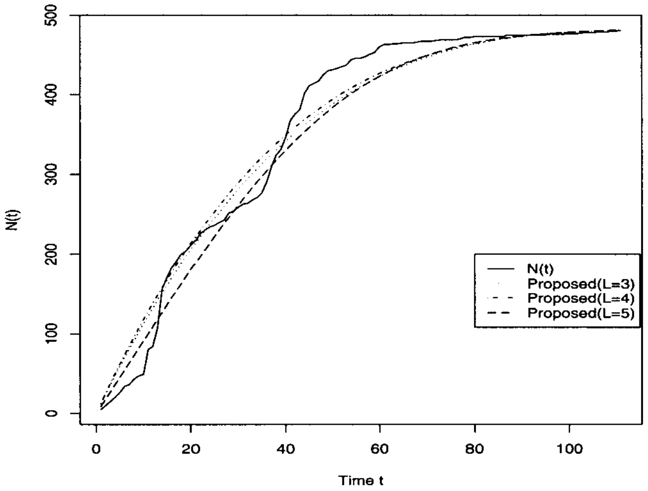

Figure 3.1. Cumulative number of failure times and the estimated mean value functions

Let n_i be the number of failures in the time interval $(t_{i-1}, t_i]$, $i = 1, \ldots, m$ and $B_\omega(t_{i-1}, t_i) = A_\omega(t_i) - A_\omega(t_{i-1})$. Then, the full posterior density is given as

$$\pi(\omega, \lambda \,|\, D) \propto \lambda^{\sum_{i=1}^m n_i} \left\{ \prod_{i=1}^m B_\omega(t_{i-1}, t_i)^{n_i} \right\} \exp\left\{ -\lambda A_\omega(t_m) \right\} \pi(\omega)\pi(\lambda)$$

and the conditional posterior densities for the Gibbs algorithm are

$$\pi(\lambda \,|\, \omega, D) = \text{Gamma}\left(\sum_{i=1}^m n_i + 1, A_\omega(t_m) + 0.001 \right),$$

$$\pi(\omega \,|\, \lambda, D) \propto \left\{ \prod_{i=1}^m B_\omega(t_{i-1}, t_i)^{n_i} \right\} \exp\{-\lambda A_\omega(t_m)\}.$$

To remove the effect of the initial values and reduce correlations among Gibbs samples, we used the following method: For each initial value for ω(9 cases), we iterate 35,000 times and drop the first 5,000 samples and chose every tenth iteration so that samples of size 3,000 are obtained. We monitor the convergence of the Gibbs samplers using the Gelman and Rubin (1992) method that uses the analysis of variance technique and all estimates of interest are acceptable in our settings. Table 3.1 summaries the estimated posterior means and standard deviation of λ and ω and Figure

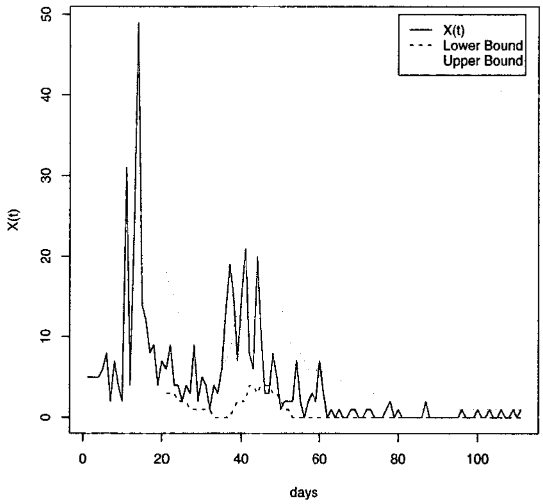

Figure 3.2. Cumulative number of failure times and the estimated mean value functions

3.1 shows the cumulative number of failures of the testing data and the estimated mean value function via the MCMC algorithm.

The predictive interval for the number of future failures is used to verify the model adequacy. The conditional predictive density for future failures n_{i+1} given $D_i = (n_1, \dots, n_i)$ is defined as

$$\begin{aligned} p(n_{i+1} \mid D_i) &= \int p(n_{i+1} \mid \theta,\, D_i)\pi(\theta \mid D_i)d\theta \\ &= \int \lambda^{n_{i+1}} \frac{B_\omega(t_i,\, t_{i+1})^{n_{i+1}}}{n_{i+1}!} \exp\{-\lambda B_\omega(t_i,\, t_{i+1})\}\pi(\theta \mid D_i)d\theta. \end{aligned} \tag{3.1}$$

The predictive probability $p(n_{i+1} \mid D_i)$ can be approximated by Monte Carlo integration. We found no remarkable difference among the approximated values of $p(n_{i+1} \mid D_i)$ for $L = 3$, 4 and 5. Figure 3.2 shows 95% predictive intervals for n_{i+1}, $i = 20, \dots, 111$, when $L = 3$. We see that the predictive interval covers most of the observed n_i and conclude that the model is adequate.

3.2. Failure times model

Jelinski and Moranda (1972) introduced data which are based on the trouble report for one of the larger modules of the Naval Tactical Data System(NTDS). The first 26 failures were found during the production phase and the remaining 5 failures were detected during the testing phase; see Goel and Okumoto (1979) and Mazzuchi and Soyer (1988). We assume that F_0 is uniform(0, 550) and G is uniform(0, λ). The number of mixands $L = 3$, 4 and 5 and $\sigma = 1$ are considered. Let $D_m = \{t_1, \dots, t_m\}$ be observed epochs of failures. Then, the likelihood function of this model is

$$L(\omega \mid D_m) = \left\{ \prod_{i=1}^{m} \lambda B_\omega^*(t_i) \right\} \exp\left\{ -\lambda A_\omega(t_m) \right\},$$

Table 3.2. Posterior summaries of parameters

		λ	ω
$L=3$	Mean	32.12	$(0.79, 0.10, 0.11)$
	S.D.	5.69	$(0.10, 0.09, 0.07)$
$L=4$	Mean	32.20	$(0.69, 0.13, 0.08, 0.10)$
	S.D.	5.72	$(0.11, 0.11, 0.06, 0.06)$
$L=5$	Mean	32.20	$(0.59, 0.16, 0.09, 0.07, 0.09)$
	S.D.	5.67	$(0.13, 0.13, 0.07, 0.06, 0.06)$

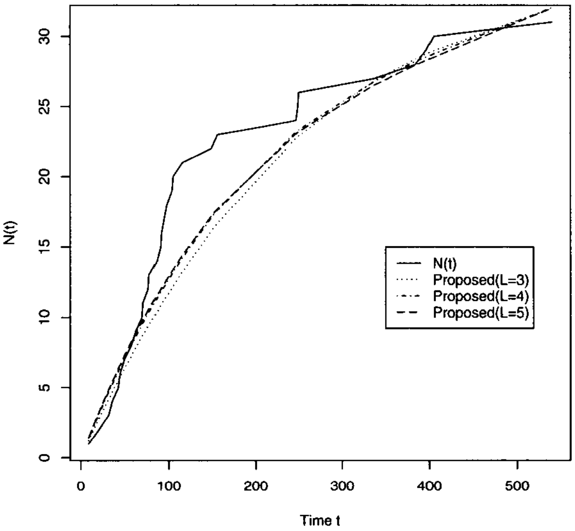

Figure 3.3. Cumulative number of failure times and the estimated mean value functions

where $B^*_\omega(t) = d/dt \sum_{l=1}^{L} \omega_l B\{F_0(t);\, l,\, L+1-l\}$. Assuming λ and ω are independent and

$$\pi(\lambda) \sim \text{Gamma}(1, 0.001) \quad \text{and} \quad \pi(\omega) \sim \text{Dirichlet}(1, \ldots, 1),$$

the conditional posterior densities are written as

$$\begin{aligned} \pi(\lambda \,|\, \omega,\, D_m) &= \text{Gamma}(m+1,\, A_\omega(t_m) + 0.001); \\ \pi(\omega \,|\, \lambda,\, D_m) &\propto \left\{ \prod_{i=1}^{m} B^*_\omega(t_i) \right\} \exp\{-\lambda A_\omega(t_m)\}. \end{aligned}$$

Table 3.2 summarizes estimated posterior means and standard deviations for $L = 3$, 4 and 5 and Figure 3.3 shows $N(t)$ and estimated mean value functions. Since the data are obtained at two different phases, the change of the parameters in the mean value function may happen at that time.

The NTDS data have been analyzed by many authors including Jelinski and Moranda (1972), Goel and Okumoto (1979), Kuo and Yang (1995, 1996), Achcar *et al.* (1997) and Basu and Ebrahimi (2003). We compare the performance of other models with beta-mixture models. We use the first 26 failures as used in the most of the literature. For the failure times model, the conditional predictive

Table 3.3. Values for $c(l)$

Models	$c(l)$
Gen.Gamma Order Statistics Model	9.576×10^{-50}
Weibull Order Statistics Model	9.874×10^{-58}
Gamma Order Statistics Model	1.325×10^{-51}
Exponential Order Statistics Model	3.587×10^{-57}
Lognormal Order Statistics Model	1.130×10^{-51}
SA Basu *et al.* (2003)	3.007×10^{-39}
EA Basu *et al.* (2003)	1.052×10^{-39}
EV Basu *et al.* (2003)	2.021×10^{-40}
Beta mixture ($L = 3$)	4.413×10^{-37}
Beta mixture ($L = 4$)	7.153×10^{-37}
Beta mixture ($L = 5$)	1.694×10^{-36}

Table 3.4. Log-marginal likelihood in the NTDS data

Models	marginal likelihood
Jelinski-Moranda/Goel-Okumoto	−88.98
Littlewood-Verall (1973)	−93.13
Homogeneous Poisson process	−92.47
Musa-Okumoto	−96.87
Weibull order statistic	−97.02
Singpurwalla-Soyer (1985)	−91.74
SA Basu *et al.* (2003)	−84.54
EA Basu *et al.* (2003)	−85.19
EV Basu *et al.* (2003)	−90.94
Beta mixture ($L = 3$)	−84.99
Beta mixture ($L = 4$)	−84.65
Beta mixture ($L = 5$)	−83.51

density can be computed by

$$\begin{aligned} p(T_{i+1} \mid D_i) &= \int p(T_{i+1} \mid \theta,\, D_i)\pi(\theta \mid D_i)d\theta \\ &= \int \lambda B_\omega^*(T_{i+1}) \exp\left[-\lambda\left\{A_\omega^*(T_{i+1}) - A_\omega^*(t_i)\right\}\right] \pi(\theta \mid D_i)d\theta. \end{aligned}$$

Table 3.3 summarizes the values of $c(l) = \prod_{i=1}^{n-1} p(t_{i+1} \mid D_i)$, where l indexes the models. The results except the beta-mixture models are referred to Achcar *et al.* (1997) and Basu and Ebrahimi (2003). The model with maximum $c(l)$ is preferred. We observe that beta-mixture models are superior to others for the NTDS data.

As an another comparison criterion, we could consider the marginal likelihood. In general, the Bayes factor is the Bayesian model comparison criterion. The Bayes factor is defined as the ratio of marginal likelihood of the two models

$$B_{12} = \frac{p(t \mid M_1)}{p(t \mid M_2)},$$

where $p(t \mid M_i) = \int p(t \mid \theta,\, M_i)d\pi(\theta \mid M_i)$ is the marginal likelihood, $p(t \mid \theta,\, M_i)$ is the likelihood and $p(\theta \mid M_i)$ is the prior distribution under model M_i. We use the Harmonic mean estimator to estimate the marginal likelihood; see Basu and Ebrahimi (2003).

Table 3.4 summarizes the estimates of log-marginal likelihoods. The model having the higher marginal likelihood is preferred. We observe that beta-mixture models dominate other models. The values in Table 3.4 except beta-mixture models are referred to Basu and Ebrahimi (2003).

4. Conclusion

Software reliability models are used to monitor the faults of software in the testing phase. In this paper, we propose a Bayesian approach to a software reliability model using the beta-mixture. Beta-mixture model allows users to select F_0 and G flexibly so that the mean value function can have various shapes. To overcome the estimation problem with the density having multiple parameters, we apply Markov chain Monte Carlo technique. We show that the proposed models are applicable to two real data sets of different types. The predictive approaches are used for the model selection. In addition, the proposed model is compared with the existing software reliability models via the marginal likelihood and the predictive distribution.

References

Achcar, J. A., Dey, D. K. and Niverthi, M. (1997). *A Bayesian approach using nonhomogeneous Poisson process for software reliability models*, In *Frontiers in Reliability*, A.P. Basu, S. K. Basu and S. Mukhopadhyay,(Eds.), 1–18, World Scientific Publishing: Singapore.

Basu, S. and Ebrahimi, N. (2003). Bayesian software reliability models based on Martingale processes, *Technometrics*, **45**, 150–158.

Cox, D. R. and Lewis, P. A. (1966). *Statistical Analysis of Series of Events*, Chapman & Hall/CRC, London.

Diaconis, P. and Ylvisaker, D. (1985). *Quantifying prior opinion*, In *Bayesian Statistics 2*, J. M. Bernado, M. H. DeGroot, D. V. Lindley and A. F. M. Smith,(Eds.), 133–156, Amsterdam: North Holland.

Duane, J. T. (1964). Learning curve approach to reliability monitoring, *IEEE Transactions on Aerospace*, **AS-2**, 563–566.

Gelfand, A. E. and Mallick, B. K. (1995). Bayesian analysis of proportional hazards models built from monotone functions, *Biometrics*, **51**, 843–852.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–472.

Goel, A. L. (1983). *A Guide Book for Software reliability Assessment*, Technical Report, Rome air Development Center, Rome, New York.

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

Jelinski, Z. and Moranda, P. B. (1972). *Software reliability research*, In *Statistical Computer Performance Evaluation*, Freiburger, W. (Eds.), Academic Press, New York, 465–497.

Kim, D. K., Yeo, I. K. and Park, D. H. (2006). Nonhomogeneous Poisson processes based on Beta-mixtures in software reliability models, *Advanced Reliability Modeling II - Reliability Testing and Improvement*, 403–410.

Kuo, L. and Yang, T. Y. (1995). Bayesian computation of software reliability, *Journal of Computational and Graphical Statistics*, **4**, 65–82.

Kuo, L. and Yang, T. Y. (1996). Bayesian computation for nonhomogeneous Poisson processes in software reliability, *Journal of the American Statistical Association*, **91**, 763–772.

Littlewood, B. and Verall, J. L. (1973). A Bayesian reliability growth model for computer science, *Applied Statistics*, **22**, 332–346.

Mallick, B. K. and Gelfand, A. E. (1994). Generalized linear models with unknown link functions, *Biometrika*, **81**, 237–245.

Mazzuchi, T. A. and Soyer, R. (1988). A Bayes empirical-Bayes model for software reliability, *IEEE Transactions on Reliability*, **37**, 248–254.

Musa, J. D. and Okumoto, K. (1984). A logarithmic Poisson execution time model for software reliability measurement, In *Proceedings of the Seventh international conference on software engineering*, 230–238.

Ohba, M., Yamada, S., Takeda, K. and Osaki, S. (1982). S-shaped software reliability growth curve: How good is it?, COMPSAC '82, 38–44.

Singpurwalla, N. D. and Soyer, R. (1985). Assessing (software) reliability growth using a random coefficient autoregressive process and its ramifications, *IEEE Transactions on Software Engineering*, **11**, 1456–1464.

Tohma, Y., Yamano, H., Obha, M. and Jacoby, R. (1991). The estimation of parameters of the hypergeometric distribution and its application to the software reliability growth model, *IEEE Transactions on Software Engineering*, **17**, 483–489.