信頼性應用研究
제1권, 제2호, pp. 175-182, 2001

Interval estimation of mean value function using fuzzy approach

DaeKyung Kim

Division of Mathematics & Statistical Informatics, Chonbuk National University

Abstract

Recently, the quality of software has become a major issue. The statistical models used in making predictions about the quality of software are termed software reliability growth models (SRGM). However, the existing SRGMs have not been satisfactory in predicting software reliability behavior (Keiller and Miller(1991), Keiller and Littlewood(1984), Musa(1987)). In this paper, we present a fuzzy-based interval estimation of software errors (failures).

1. Introduction

Study on software reliability begins at the late 1960s. Software reliability has been the most extensively studies of all the quality attributes and is defined as the probability of failure software operation for a specified period of time in a specified environment (Musa et. Al.(1987)). High quality software is crucial both to software producers and users, as well as society in general, because the impact of failures of software can range from inconvenience(e.g. malfunctions of switching systems), economic loss(e.g. interceptions of banking systems) to life threatening (e.g. failures of flight systems, medical software).

A common procedure for determining software reliability is to fit an appropriate stochastic model to the available failure data and, based on this model, determine the current system reliability. The statistical models used in making predictions about the quality of software are termed software reliability growth models (SRGM). The first documented software reliability models were published by Jelinski and Moranda(1972) and Shooman(1972). Since then, there have been over forty models reported in the published literature (Brocklehurst, et.al.(1990)) used to determine these quality measures.

The accuracy of these models, however varies significantly (Keller, et.al.(1983)) and no single model is known to perform well in all contexts. A major difficulty in software reliability practice is to analyze the context in which reliability measurement is to take place. And the existing software reliability models have many assumptions. But these assumptions have to be modified in order to make a probabilistic software reliability models (PSRM) successful if probability assumption is adopted. However, the PSRM can become extremely complex and non-tractable mathematically if these assumptions are modified. As a result, there is no definite model that can be recommended unreservedly, and potential users are left in a dilemma as to which models to choose, which procedures to apply, and which predictions to trust, among the various competing options. So, the existing SRGMs have not been satisfactory in predicting software reliability behavior.

We may not estimate the reliability measures exactly because of model's incorrectness. In our circumstances, we need to estimate the interval of number of faults. But maximum likelihood estimation (MLE) method to estimate the unknown parameters in the software reliability models may not converge to a reasonable value (Hossain and Dahiya(1997)). Also we approximately obtain the MLE of the mean value function if n is very large(Yamada and Osaki (1985)). However, the assumption of large sample size does not always hold.

Because of the uniqueness of software, software reliability behavior is fuzzy in nature So fuzzy software reliability models (FSRMs) should be developed in place of PRSMs to remove the unpleasant fact that none of PRSMs has been proved to be mathematically tractable and universally applicable(Cai, et.al. (1991)).

Therefore, in this paper we apply fuzzy set theory to Goel-Okumoto (GO)(Goel and Okumoto (1979) software reliability model and present a fuzzy-based interval estimation of software errors (failures) and compare its results with current ones . We

choose the G-O SRGM because it is the most widely accepted model as well as logarithmic Poisson models (Musa et. al.(1987)). In many situations these models proved to be a good fit. Section 2 describes the fuzzy regression model. Section 3 estimates parameters using the linear programming and obtains the fuzzy interval estimation of mean value function. And we compare it with other software reliability model.

2. Fuzzy regression model

Fuzzy regression models have proposed by Takana et.al.(1988). The aim of a fuzzy regression model is to take all observed data into account in the fuzzy numbers that are estimated by the model. The fuzzy regression equation is written as follows.

$$Y_j = A_1 + A_2 x_{2j} + \Lambda + A_n x_{nj} = A x_j \tag{1}$$

$$j = 1,2,\Lambda ,m,$$

where the regression coefficients A_i's are triangular fuzzy numbers $A_i = (n_i, c_i)$ with center n_i and width c_i. And m denotes the numbers of observation. According to the extension principle the output of the fuzzy regression equation (1) can be written as follows.

$$Y_j = Ax_j = (n,c)x_j = (nx_j, c\,|\,x_j\,|)$$

A regression model with fuzzy coefficients can be expressed with a lower boundary $nx_j - c\,|\,x_j\,|$, center nx_j and upper boundary $nx_j + c\,|\,x_j\,|$. When a sample $(y_j, d_j)(j = 1,2,\Lambda ,m)$ with center y_j and width d_j is given as a fuzzy number (y_j, d_j), the membership function of fuzzy coefficients is set to $L(u)$, and the possibility of a sample being included in the regression model is α, the inclusion relation between the model and the data is written as follows.

$$\begin{aligned} y_j + L^{-1}(\alpha)d_j &\le nx_j + L^{-1}(\alpha)c\,|\,x_j\,| \\ y_j - L^{-1}(\alpha)d_j &\ge nx_j - L^{-1}(\alpha)c\,|\,x_j\,| \end{aligned} \qquad (2)$$

The model can be solved by linear programming.

Using the notations for observed data (y_j, x_j), $y_j = (y_j, d_j), x_j = [1, x_{2j}, \Lambda , x_{nj}](j = 1,2,\Lambda ,m)$, and for fuzzy coefficients (n,c), the regression model can then be recast as the following linear programming problem.

$$\min_{n,c} \sum_{j=1}^{m} c\,|\,x_j\,|$$

subject to

$$\begin{aligned} y_j + L^{-1}(\alpha)d_j &\le nx_j + L^{-1}(\alpha)c\,|\,x_j\,| \\ y_j - L^{-1}(\alpha)d_j &\ge nx_j - L^{-1}(\alpha)c\,|\,x_j\,| \end{aligned}$$

$$(j = 1,2,\Lambda ,m),\ c \ge 0$$

where $L(u)$ is the membership function for the fuzzy regression model and α is a possibility to which extent a sample is included in the model.

3. Numerical example

A software system is subject to failures at random times caused by errors present in the system. Let $\{N(t), t \geq 0\}$ be a counting process representing the cumulative number of failures by time t. And $m(t)$ represent the mean value function for the stochastic process such as

$$m(t) = E[N(t)], t \geq 0.$$

An attractive family of stochastic processes characterized by their mean value functions are the non homogenous Poisson processes (NHPP). Since this process is completely characterized by its mean value function, $m(t)$, what have developed are families of models with different mean value functions.

In this study, we consider Goel-Okumoto software reliability model (1979). Its model with fuzzy number is presented as follows.

$$m(t) = a(1 - e^{-bt}) \qquad (3)$$

where $a = [\underline{a}, \overline{a}]$ and $b = [\underline{b}, \overline{b}]$ are fuzzy numbers. Therefore, we must determine the linear fuzzy model that guarantees each actual y_j falls in the set of the fuzzy output.

$$\underline{m}(t) = [\underline{a}(1 - e^{-\underline{b}t}), \overline{a}(1 - e^{-\overline{b}t})]$$

where $a = [\underline{a}, \overline{a}]$ and $b = [\underline{b}, \overline{b}]$ are fuzzy numbers as before. We transform the equation (3) as follows.

$$y \equiv \log_e \frac{\{a - m(t)\}}{a} = -bt$$

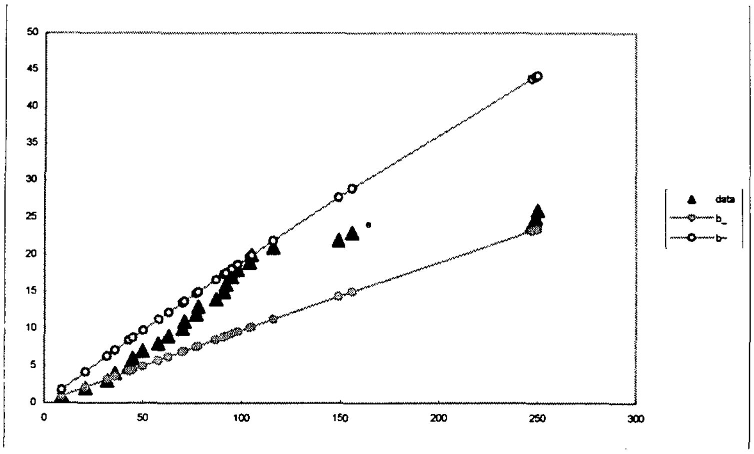
The software failure data to be analyzed in this section are taken from Jelinski and Moranda(1972). The data are originally from the U.S. Navy Fleet Computer Programming Center. We usually call it NTDS (Naval Tactical Data System) data. The times (days) between software failures are obtained . Here we will only use the twenty-six error which is found in the production phase. We solved a fuzzy regression analysis by excel. We first calculate $b = [\underline{b}, \overline{b}]$. This is calculated for each a value from 31 to 200. At $a = 200$, the interval $c\,|\,t\,|$ is minimum. The values of a and c that minimize $\sum_{j=1}^{26} c\,|\,t_j\,|$ are 0.00074 and 0.00027 respectively. And $\underline{b}$ and $\bar{b}$ are 0.0005 and 0.0010 respectively. A regression model with fuzzy coefficients i.e. $nt_j - c\,|\,t_j\,|$, nt_j and $nt_j + c\,|\,t_j\,|$ are as follows.

(Table 1. Fuzzy coefficient)

t_j	$nt_j - c\,|\,t_j\,|$	nt_j	$nt_j + c\,|\,t_j\,|$
9	0.004	0.008	0.009
21	0.010	0.016	0.021
32	0.015	0.024	0.032
36	0.017	0.027	0.036
43	0.020	0.032	0.043
45	0.021	0.033	0.045
50	0.024	0.037	0.050
58	0.027	0.043	0.058
63	0.030	0.047	0.063
70	0.033	0.052	0.070
71	0.034	0.053	0.071
77	0.036	0.057	0.077
78	0.037	0.058	0.078
87	0.041	0.064	0.087
91	0.043	0.067	0.091
92	0.043	0.068	0.092
95	0.045	0.070	0.095
98	0.046	0.073	0.098
104	0.049	0.077	0.104
105	0.050	0.078	0.105
116	0.055	0.086	0.116
149	0.070	0.110	0.150
156	0.074	0.115	0.157
247	0.117	0.183	0.248
249	0.118	0.184	0.250
250	0.118	0.185	0.251

<Table1> Continued

The fuzzy regression obtained in the above result is shown as Figure 1. The relation (2) between the model and the samples holds in the figure 1.



(Figure 1. Fuzzy regression model)

4. Conclusion

In this paper, we investigated a fuzzy-based interval estimation of software errors(failures). Because no single model is known to perform well in all contexts, we are interested in confidence interval of reliability characteristics. And the estimate of maximum likelihood estimation(MLE) of the unknown parameter in the software reliability models may not converge to a reasonable value. Also we approximately obtain the MLE of the mean value function if n is very large. However, the assumption of large sample size does not always hold. Therefore, we estimated interval of the mean value function using the fuzzy regression method.

References

[1] Brocklehurst, S., Chan, P.Y., Littlewood, B., and Shell, J.(1990), Recalibrating software reliability models, *IEEE Transaction of software Engineering*, Volume 16 No. 4, 458-470.

[2] Cai, K.Y., Wen, C.Y. and Zhang, M.L.,(1991), A critical review on software reliability modeling, *Reliability Engineering Systems Safety*, 32, 357-371.

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

[4] Hossain, S.A. and Dahiya, R.C.(1993), Estimating the parameters of a non-

homogenous poisson-process model for software reliability, *IEEE Transactions on Reliability*, 42, 4, 604-612.

[5] Jelinski, Z. and Moranda, P.B.(1972), Software reliability research, in Software Statistical Computer Performance Evaluation(W. Freiberger, ED.),Academic, New York,USA, 465-484.

[6] Keiller, P.A. and Littlewood, B.A.(1984), Adaptive software reliability modeling, 14th International Symposium on Fault-Tolerant Computing, Digest of papers, IEEE Computer Society Press, Washington, 108-113.

[7] Keliller,P.A. , Littlewood, B. , Miller, D.R. and Sofer, A.(1983), Comparison of software reliability predictions, in Dig. FTCS 13 (13th International Symposium Fault-tolerant Computation, 128-134.

[8] Keiller, P.A., and Miller, D.R.(1991), On the use and performance of software reliability growth models, in Reliability Engineering and System Safety, Special Issue on Software, Vol. 32, No 1&2, 95~117.

[9] Musa, John, et al. Iannino, A. and Okumoto, K(1987), *Software Reliability-Measurement, Predictions, Application*, MaGrall Hill, New York,

[10] Shooman, M.L.(1972), Probabilistic models for software reliability prediction, in Statistical Computer Performance Evaluation(W.Freiberger,Ed.), Academic, New York,USA, 465-484.

[11] Tanaka, H. and Watada, J.(1988), Possibilistic linear systems and their application to the linear regression model, *Fuzzy Sets and Systems*, 27, 275-289.

[12] Yamada, S and Osaki, S.(1985), Software reliability growth modeling : Models and applications, *IEEE Transactions on Reliability*, SE-11, 12 ,1431-1437.