数理解析研究所講究録
第 1802 巻 2012 年 133-138

Applying Daubechies wavelets to software failure rate estimation

Xiao Xiao and Tadashi Dohi
Graduate School of Engineering
Hiroshima University, Japan

1 Introduction

The quantitative assessment of software reliability is one of the main issues in software reliability engineering. In general, one needs several kinds of mathematical models to estimate the quantitative software reliability, which is defined as the probability that software system does not fail during a specified time period. Among a huge number of software reliability models (SRMs) [9, 13, 15, 20], non-homogeneous Poisson process (NHPP) based SRMs have gained much popularity in actual software testing phases to assess the software reliability, the number of remaining faults in software, the software release schedule, *etc.* A unique parameter to govern the probabilistic properties of NHPP-based SRMs is the *rate function.* Therefore, it is necessary to develop a method that can estimate the rate function with a high degree of accuracy, and this naturally results in trusted quantitative software reliability. According to this way of thinking, the software reliability assessment problems can be reduced to the estimation of the NHPP rate function.

Both parametric and nonparametric NHPP-based SRMs have been developed during the last three decades. The most classical parametric NHPP-based SRM is the Goel-Okumoto model [7] which was proposed in 1979. Other representative SRMs based on the NHPP have been also proposed in the literature [10, 14, 21]. In these parametric models, it is assumed that functional forms of the rate function are known with some unknown model parameters. This assumption inevitably leads to the fact that parametric estimation methods are able to follow only the average tendency of the cumulative number of software faults detected till certain testing date, but can not catch the microscopic behavior in each testing date. Additionally, the functional form of the rate function is not very well known or is completely unknown. This fact implies that the non-parametric methods without assuming the parametric form should be used to describe the software debugging phenomenon which is different in each testing phase.

Some frequentist approaches based on non-parametric statistics were introduced to estimate the quantitative software reliability [2, 6, 16]. These non-parametric models, in general, require high computation cost. The similar problem on the expensive computation cost appears in the Bayesian non-parametric estimation [5, 18]. In recent years the wavelet-based statistical methods have been well established especially in several areas such as non-parametric regression, probability density estimation, time series analysis, *etc* (see [1, 4, 12]). Among a lot of techniques which have been proposed to account for the Poisson intensity estimation problem, Xiao and Dohi [19] proposed a novel non-parametric estimation framework for the NHPP-based SRMs, where the Haar-wavelet-based techniques were applied to estimate the software failure rate. They treated with the fault count (group) data, where the number of software failures is recorded. This is known as an incomplete failure data since the exact time of each software failure is not recorded. Another type of software failure data is the one where the software failure time is observed and recorded. Kuhl and Bhairegond [8] proposed a Daubechies wavelet estimator for the NHPP rate function. They presented simulation-based performance evaluation for the wavelet procedure, and succeed in estimating three different types of rate functions. However, there are several mathematical difficulties when applying their procedure to the real software failure data analysis. In this paper, we consider the application of the Daubechies wavelet estimator in estimating the software failure rate from real project data. We give practical solutions to these mathematical difficulties in details.

The paper is organized as follows. In Section 2, we describe the basic modeling framework of NHPP-based SRMs. In Section 3, we introduce the Daubechies wavelet-based procedure employed in this paper. Section 4 is devoted to the solutions of some methematical difficultes occurred in analyzing real project data. We summarize the paper with comments in Section 5.

2 NHPP-based Software Reliability Modeling

Let $N(t)$ denote the number of software faults detected by testing time t, and be a stochastic point process in continuous time. We assume that one software failure is caused by one software fault and use the word "failure" as the same meaning with "fault" below. The stochastic process $\{N(t),\ t \geq 0\}$ is said to be a non-homogeneous Poisson process (NHPP) if the probability mass function at time t is given by

$$\Pr\{N(t) = n\} = \frac{\{\Lambda(t)\}^n}{n!} e^{-\Lambda(t)}, \quad (1)$$

$$\Lambda(t) = \int_0^t \lambda(x)dx, \quad (2)$$

where $\mathrm{E}[N(t)] = \Lambda(t)$ is called the mean value function of an NHPP, and means the expected cumulative number of software failures experienced by time t. In Eq. (2), $\lambda(t)$ is the rate function of NHPP, and implies the software failure rate at time t. Suppose that n software faults are detected by the present time. Let t_i $(i = 1,\ 2, \ldots,\ n)$ denote the time of the i-th software failure. Then, the likelihood function of the NHPP with n software failure time data is given by

$$\mathcal{LF} = \exp\{-\Lambda(t_n)\} \prod_{i=1}^{n} \lambda(t_i). \quad (3)$$

Taking the logarithm, we have the log likelihood $\mathcal{LLF}$ by

$$\mathcal{LLF} = \prod_{i=1}^{n} \ln\lambda(t_i) - \Lambda(t_n). \quad (4)$$

For the parametric models, the commonly used estimation technique is the maximum likelihood (ML) estimation. The ML estimate of unknown model parameters, denoted as $\boldsymbol{\theta}$, is given by the solution of $\mathrm{argmax}_{\boldsymbol{\theta}} \mathcal{LLF}(\boldsymbol{\theta})$. In the following section, we introduce the Daubechies wavelet-based estimation procedure, which does not assume the functional form of the rate function.

3 Nonparametric Estimation Using Wavelet Estimator

3.1 Daubechies Wavelet

Wavelets consist of two basis functions, the *scaling function* $\phi(t)$ and the *wavelet function* $\psi(t)$, that work together to provide wavelet approximations. These functions are orthonormal bases of Hilbert space, so that any signals or data in this vector space can be represented by linear combinations of scaling function and wavelet function. Daubechies [3] defined a set of compactly supported wavelets, which gained much popularity in wavelet analysis thereafter.

The Daubechies scaling function and wavelet function are of the following forms.

$$\phi(t) = \sum_{i=0}^{n} h_i \phi(2t-i), \tag{5}$$

$$\psi(t) = \sum_{i=0}^{n} (-1)^i h_{n-i} \phi(2t-i), \tag{6}$$

where n is the support of $\phi(t)$ and $\psi(t)$, and coefficients h_i, $(i = 0,\ 1, \ldots,\ n)$ for different supports are given in [3]. From the above two equations we know, Daubechies scaling function and wavelet function are not defined in a closed analytic form. In fact the scaling function is calculated by solving a simultaneous equation with the defined coefficients h_i and initial value $\phi(0) = \phi(n) = 0$. For instance, the coefficients of Daubechies wavelets with support $n = 7$ are defined as

$$\begin{aligned}
h_0 &= 0.3258034,\\
h_1 &= 1.0109457,\\
h_2 &= 0.8922014,\\
h_3 &= 0.0395750,\\
h_4 &= 0.2645072,\\
h_5 &= 0.0436163,\\
h_6 &= 0.0465036,\\
h_7 &= 0.0149870.
\end{aligned}$$

First, the starting values of Daubechies scaling function, $\phi(1),\ \phi(2), \ldots,\ \phi(6)$, can be calculated by solving $\phi(t) = \sum_{i=0}^{7} h_i \phi(2t-i)$ $(t = 1,\ 2, \ldots,\ 6)$ under the condition $\sum_{t=0}^{7} \phi(t) = 1$ and $\phi(0) = \phi(7) = 0$. Second, the values of the Daubechies scaling function at other points in time interval [0, 7] can be calculated by Eq. (5) using the starting values and the coefficients h_i. For example, we have

$$\phi(0.5) = \sum_{i=0}^{7} h_i \phi(1-i) = h_0\phi(1) + h_1\phi(0). \tag{7}$$

A feature of Daubechies scaling function is that it only takes value at such a time point t when $t = a \cdot 2^b$ $(a,\ b \in \mathbf{Z})$. This kind of number is called a dyadic number if and only if, it is integral multiple of an integral power of 2 (see [11]).

Note that Daubechies wavelet is not defined in a closed analytic form but in a set of discrete values, therefore it is classified as discrete wavelet with the same as Haar wavelet. However, since if sufficient values of the Daubechies wavelet are calculated, a smooth scaling function can be obtained, and the Daubechies wavelet is effective in representing continuous function.

3.2 Positive Basis Functions and Daubechies Wavelet Estimator

Walter and Shen [17] developed a positive wavelet estimator for estimating density functions. The Daubechies wavelet estimator for the rate function of an NHPP is based on this positive basis function. Let $\phi(t)$ be the scaling function having compact support. For $0 < r < 1$, a positive basis function is given by

$$P_r(t) = \sum_{j \in Z} r^{|j|} \phi(t-j), \tag{8}$$

where the constant value r is selected such that this positive basis developed is always greater than or equal to zero [17]. Kuhl and Bhairgond [8] constructed a wavelet estimator using this positive basis, which is given by the following form:

$$\hat{\lambda}_{m,k}(t) = \sum_{n=-k}^{k} \Big\{ \sum_{i=1}^{N} P_r(2^m t_i - n) \Big\} \Big(\frac{1-r}{1+r} \Big)^2 \times 2^m P_r(2^m t - n), \tag{9}$$

where t_i are the arrival times of an NHPP whose rate function is to be approximated, and N is the number of arrivals in the interval under consideration. The range for n is selected in such a way that the positive basis function $P_r(t)$ can translate through the entire range of arrival times, and the resolution m is selected based on the level of detail of the approximation desired. This wavelet estimator is used to approximate the rate function of an NHPP.

4 Set Up In Daubechies Wavelet Estimator

Kuhl and Bhairegond [8] presented a simulation-based performance evaluation for the above Daubechies wavelet estimator. However, there are several limitations of their procedure in the case of treating real software failure time data. We discuss these limitations in this section and give some suggestions in applying the Daubechies wavelet estimator to the real software failure time data.

4.1 Compact Support and Normalization

Daubechies [3] defines the coefficients h_i $(i = 0, 1, \ldots, n)$ for wavelets with different supports $n = 2, 3, \ldots, 10$. This means that the possible values of Daubechies wavelet estimator are in the time interval $[0, n]$. However, the real software failure time are not always observed in this interval. Therefore it is necessary to rescale the data into interval $[0, n]$. It should be noted that [data $\times$ (n/t_N)] does not work in this case. It is clear from Eq. (9) that t_i in $\hat{\lambda}_{m,k}(t)$ must be a dyadic number, otherwise, $P_r(2^m t_i - n)$ can not be defined. We consider such failure time data whose values are recorded in integer. This can be easily achieved by changing the unit of the data set to a smaller one. Since an integer is a dyadic number and an integer divided by 2^b $(b \in Z)$ is still a dyadic number, we suggest the following normalization rule:

[data $\times\ 2^{-b} \mid 2^b \leq t_N$].

In this way, software failure time data with arbitrary ending time can be analyzed with the Daubechies wavelet estimator.

4.2 Parameter Determination

As mentioned in Section 3.2, the range for n should be selected in such a way that the positive basis function $P_r(t)$ can translate through the entire range of arrival times. Note that the positive function $P_r(t)$ quickly decays to zero in both the positive and negative directions (see Figure 1), so we take the truncation for it from -7 to 8. The boundary is determined as -7 and 8 because the value of $P_r(t)$ outside the limit becomes negative. Walter and Shen [17] proved that there exists $0 < r < 1$ such that $P_r(t)$ satisfies $P_r(t) \leq 0$ $(t \in R)$, but this holds only when j in Eq. (8) takes all values in Z. This is difficult in computation so that we have to select an appropriate range for j. Since we consider the scaling function with support 7 in this paper, and the nature of $P_r(t)$ depends upon the nature of the scaling function,

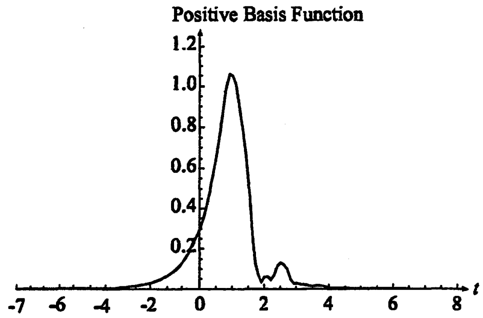

Figure 1: Positive basis function associated with Daubechies scaling function.

we set $j \in [-7,\ 7]$. We checked the value of $P_r(t)$ with different r (from 0.1 to 0.9), and found that $P_r(t)$ with $j \in [-7,\ 7]$ could always be positive from -7 to 8. As a result, k in Eq. (9) should be selected as

[Integer Part of $2^m t_N + 7$]

which ensures $2^m t_i - n$ is in the interval [-7, 8].

5 Conclusion

In this paper we have applied the Daubechies wavelet estimator to estimate the software failure rate function. We have given practical solutions to the mathematical difficulties in applying the procedure to the real software failure data. The practitioners are not requested to carry out troublesome procedures on model selection and to take care of computational efficiency such as the judgment of selecting initial value of parameters. In order to establish the credibility and the usefulness of the Daubechies wavelet estimation procedure in software failure data analysis, we will present a real data analysis to evaluate the goodness-of-fit performance of the Daubechies wavelet estimator in the future.

References

[1] A. Antoniadis, J. Bigot and T. Sapatinas, "Wavelet estimators in nonparametric regression: a comparative simulation study," *Journal of Statistical Software*, **6** (6), pp. 1–83 (2001).

[2] M. Barghout, B. Littlewood and A. Abdel-Ghaly, "A non-parametric order statistics software reliability model," *Software Testing, Verification and Reliability*, **8** (3), pp. 113–132 (1998).

[3] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Pennsylvania (1992).

[4] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, "Density estimation by wavelet thresholding," *Annals of Statistics*, **24** (2), pp. 508-539 (1996).

[5] M. A. El-Aroui and J. L. Soler, "A Bayes nonparametric framework for software-reliability analysis," *IEEE Transactions on Reliability*, **45** (4), pp. 652–660 (1996).

[6] A. Gandy and U. Jensen, "A non-parametric approach to software reliability," *Applied Stochastic Models in Business and Industry*, **20** (1), pp. 3–15 (2004).

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

[8] M. E. Kuhl, "Nonparametric estimation of nonhomogeneous Poisson processes using wavelets," *Proceedings of the 2000 Winter Simulation Conference*, pp. 562–571 (2000).

[9] M. R. Lyu (ed.), *Handbook of Software Reliability Engineering*, McGraw-Hill, New York (1996).

[10] J. D. Musa and K. Okumoto, "A logarithmic Poisson execution time model for software reliability measurement," *Proceedings of the 7th International Conference on Software Engineering* (ICSE' 84), pp. 230–238 (1984).

[11] Y. Nievergelt, *Wavelet Made Easy*, Birkhauser, Boston (1999).

[12] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press (2000).

[13] H. Pham, *Software Reliability*, Springer, Singapore (2000).

[14] H. Pham and X. Zhang, "NHPP software reliability and cost models with testing coverage," *European Journal of Operational Research*, **145**, pp. 443–454 (2003).

[15] N. D. Singpurwalla and S. P. Wilson, *Statistical Methods in Software Engineering: Reliability and Risk*, Springer-Verlag, New York (1999).

[16] A. Sofer and D. R. Miller, "A non-parametric software reliability growth model," *IEEE Transactions on Reliability*, **R-40** (3), pp. 329–337 (1991).

[17] G. G. Walter and X. Shen, "Positive Estimation with Wavelets," *Contemporary Mathematics*, **216**, pp. 63–79 (1998).

[18] S. P. Wilson and F. J. Samaniego, "Nonparametric analysis of the order-statistic model in software reliability," *IEEE Transactions on Software Engineering*, **33** (3), pp. 198–208 (2007).

[19] X. Xiao and T. Dohi, "Wavelet-based approach for estimating software reliability," *Proceedings of 20th International Symposium on Software Reliability Engineering* (ISSRE'09), pp. 11–20 (2009).

[20] M. Xie, *Software Reliability Modelling*, World Scientific, Singapore (1999).

[21] S. Yamada, M. Ohba and S. Osaki, "S-shaped reliability growth modeling for software error detection," *IEEE Transactions on Reliability*, **R-32**, pp. 475–478 (1983).