A Discrete Stochastic Logistic Equation and a Software Reliability Growth Model

Daisuke Satoh
NTT Service Integration Laboratories
3-9-11 Midori-cho Musashino-shi,
Tokyo 180-8585 Japan
satoh.daisuke@lab.ntt.co.jp

Abstract

This paper proposes a discrete stochastic logistic equation that has an exact solution and describes an Software reliability growth model based on this equation. This model conserves the property of the deterministic ones and furthermore yields the distribution of the estimates along with the estimates themselves when an assumption is made. We evaluate the assumption by using actual data.

1 Introduction

Software reliability growth models (SRGMs) that produce S-shaped curves are empirically known to be a way of representing cumulative numbers of software failures or software faults uncovered over time. The logistic curve provides one of simplest of such models. The logistic curve model has been observed in testing software systems [2, 3].

Satoh and Yamada have described SRGMs based on discrete analogs of a logistic equation that have exact solutions. They yield accurate estimates of parameters, even with small amounts of input data. These models, however, are deterministic equations, so they do not yield a distribution of the estimates.

We propose discrete stochastic logistic equations that have exact solutions and describe SRGMs based on these equations. These models yield a distribution of the estimates along with the estimates themselves.

2 A discrete logistic equation

A logistic curve model is described by

$$\frac{dL(t)}{dt} = \frac{a}{k}L(t)\left(k - L(t)\right) \tag{1}$$

where $L(t)$ is the cumulative number of software failures that have occurred in testing up to time t.

A solution of Eq. (1) is given by

$$L(t) = \frac{k}{1 + m\exp(-at)} \tag{2}$$

where $k > 0$, $m > 0$, and $a > 0$. Parameter k represents the total number of potential software failures, i.e., the number of failures that will occurred over an infinitely long period or the initial number of faults in the software system.

We show a discrete equation and its exact solution [1] below:

$$L_{n+1} - L_n = \delta\frac{a}{k}L_n(k - L_{n+1}), \tag{3}$$

$$L_n = \frac{k}{1 + m\left(\frac{1}{1+\delta a}\right)^n}. \tag{4}$$

SRGM based on this discrete equation yields more accurate estimates of parameters, even with small amounts of input data, than the conventional model [4, 5, 6, 7].

3 A discrete stochastic logistic equation

We propose the following form of discrete stochastic logistic equation:

$$L_{n+1} - L_n = \delta\frac{A_{n+1}}{k}L_n(k - L_{n+1}), \tag{5}$$

where $\{A_n : n = 1, 2, \ldots\}$ is a sequence of independent and identically distributed (i.i.d.) random variables. Its exact solution is described by

$$L_n = \frac{k}{1 + m\prod_{j=0}^{n}\left(\frac{1}{1+\delta A_j}\right)}. \tag{6}$$

We suppose that $\{X_j : j = 1, 2, \ldots\}$ in eq. (6):

$$X_j = \frac{1}{1 + \delta A_j} \tag{7}$$

has an i.i.d. power-function distribution. We consider the probability $P\{L_n > \underline{l}\}$ where

$$\underline{l} = \frac{k}{1 + m\underline{x}}. \tag{8}$$

Then, $P\{L_n > \underline{l}\}$ is described as follows,

$$\begin{aligned} & P\{L_n > \underline{l}\} & (9)\\ = & P\{\prod_{j=1}^{n} X_j < \underline{x}\} & (10)\\ = & P\{\sum_{j=1}^{n} Y_j > -\log \underline{x}\} & (11)\\ = & (\exp(\gamma \log \underline{x})) \sum_{j=0}^{n-1} \frac{(-\gamma \log \underline{x})^j}{j!}, & (12)\end{aligned}$$

where $Y_j = -\log X_j, j = 1, 2, \ldots$ have i.i.d. exponential distribution and $\sum_{j=1}^{n} Y_j$ is a random variable that follows Erlang distribution. Therefore, the proposed equation enables us to obtain a distribution for the estimate at step n.

4 Distribution of actual data

The assumption of power-function distribution stated in the previous section was evaluated on the same actual data as had been used in an earlier work [6]. We used the last value in the data series as the value of k. The distribution of X_j is shown in Fig. 1. Figure 1 indicates that X_j has a power-function distribution except at the tail, where the small amount of data leads to deviation from the power-function distribution.

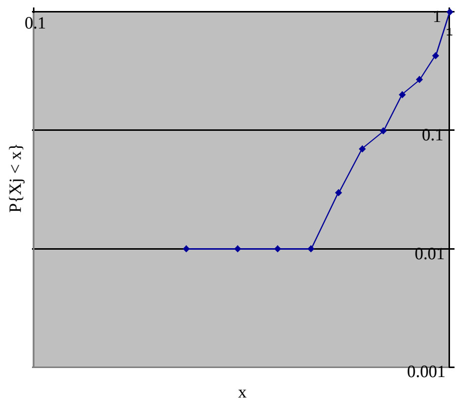

Figure 1: Distribution of the actual data

5 Conclusion

A discrete stochastic logistic equation and an SRGM based on this equation have been proposed. The equation has an exact solution and enables us to obtain a distribution of the estimate for any step, which cannot be obtained by the discrete deterministic models. The stochastic model was evaluated by using actual data.

Acknowledgement

The author is grateful to Professor Emeritus Ryogo Hirota and Professor Daisuke Takahashi at Waseda University for valuable advice. He thanks Mr. Kazunori Kumagai at NTT for careful reading of this paper.

References

[1] R. Hirota: Lecture on difference equations – from continuous values to discrete values (in Japanese), Saiensu-sha, Tokyo, 2000.

[2] K. Ohmori, and E. Shinohara: Predictive precision analysis of undiscovered errors, *Technical Report of IEICE* , **SSE98-190** (1999) 25–30.

[3] T. Sakamaki: Software reliability – Software reliability forecast for quality management. *Technical Report of IEICE* , **R-81-8** (1981), 17–24.

[4] D. Satoh, A discrete Gompertz equation and a software reliability growth model, *IEICE Trans.*, **E83-D-7** (2000), 1508–1513.

[5] D. Satoh and S. Yamada, Discrete equations and software reliability growth models, *Proceedings of the 12th ISSRE*, (IEEE Computer Society, Hong Kong, November, 2001) 176–184.

[6] D. Satoh and S. Yamada: Parameter Estimation of Discrete Logistic Curve Models for Software Reliability Assessment, *Japan J. of Industrial and Applied Mathematics*, **19-1** (2002) 39–53.

[7] S. Yamada, S. Inoue, and D. Satoh: Statistical Data Analysis Modeling Based on Difference Equations for Software Reliability Assessment (in Japanese), *Transactions of the Japan Society for Industrial and Applied Mathematics*, **12-2** (2002) 155–168.

