A measure to distinguish between a logistic curve model and a Gompertz curve model

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

Abstract— **Many software reliability growth models have been proposed and analyzed for measuring the growth in software reliability. However, far fewer comparison criteria have been proposed than the number of reliability models. We propose a measure to determine whether the logistic curve model or Gompertz curve model is more appropriate for analyzing a data set. The two models are the simplest models for estimating an S-shaped software reliability growth process, which is often observed in actual projects. The two models used in the proposed measure are discrete ones. One is based on a discrete equation proposed by Morishita and the other is based on a discrete equation that is proposed in this paper. These discrete equations have exact solutions. These two models reproduce the values of the parameters perfectly when exact solutions are used as an input data. Estimated parameters are independent of time scale. The two models enable us to accurately estimate parameters in the early testing phase with actual data.**

I. INTRODUCTION

Software reliability assessment is a key technology for reducing software costs and producing highly reliable software. Many software reliability growth models have been proposed and analyzed for measuring software reliability growth. Recently, discrete analogs of software reliability growth models have been proposed [3], [4]. These provide accurate parameter estimates using the data available during the early testing phase.

However, software reliability growth models that only yield accurate parameter estimates in the early testing phase do not fulfill the requirements of software engineers and managers because such models do not provide grounds for selecting the most appropriate model. Far fewer comparison criteria have been proposed than the number of reliability models.

In this paper, we propose a measure to distinguish between a logistic curve model and a Gompertz curve model, which are simple S-shaped models. We also propose a new discrete Gompertz curve model to use it in the measure.

II. DISCRETE LOGISTIC CURVE MODEL

Morishita[2] proposed a discrete logistic equation that has an exact solution:

$$L_{n+1} - L_n = \delta \frac{\alpha}{k} L_{n+1}(k - L_n), \quad (1)$$

$$L_n = \frac{k}{1 + m(1-\delta\alpha)^n}. \quad (2)$$

Equation (1) converges to a logistic equation:

$$\frac{dL}{dt} = \frac{\alpha}{k} L(k-L), \quad (3)$$

as $\delta \to 0$. The solution of Eq. (1) converges to a solution of Eq. (3), too.

A regression equation is obtained from Eq. (1) as follows:

$$Y_n = A + BL_{n+1}. \quad (4)$$

Y_n can be expressed as

$$Y_n = \frac{L_{n+1}}{L_n}. \quad (5)$$

The discrete model is independent of the constant time interval δ and reproduces the values of the parameters correctly when an exact solution is used as input data[4]. Moreover, the model enables us to estimate parameters accurately in the early test phase with actual data[4].

III. NEW DISCRETE GOMPERTZ CURVE MODEL

The author[3] proposed a discrete Gompertz equation that has an exact solution. The discrete equation is

$$G_{n+1} = G_n \left(\frac{G_n}{k}\right)^{\delta \log b}. \quad (6)$$

An exact solution of Eq. (6) is

$$G_n = ka^{(1+\delta \log b)^n}. \quad (7)$$

The model based on the discrete equation reproduces the values of the parameters perfectly when exact solutions are used as an input data. Estimated parameters are independent of time scale. The model enables us to estimate parameters accurately in the early testing phase with actual data.

We propose a new discrete Gompertz equation as a measure to determine whether a logistic curve model (LCM) or Gompertz curve model (GCM) is more appropriate. It is

$$G_{n+1} = G_n \left(\frac{G_{n+1}}{k}\right)^{\delta \log b}. \quad (8)$$

Equation (8) has an exact solution:

$$G_n = ka^{(1-\delta \log b)^{-n}}, \quad (9)$$

where $k > 0, 0 < a < 1, 0 < b < 1$.

Copyright 2003 Chillarege Press

Equation (8) converges to a Gompertz equation:

$$\frac{dG}{dt} = (\log b) G \log \frac{G}{k}, \tag{10}$$

as $\delta \longrightarrow 0$. The solution of Eq. (8) converges to a solution of Eq. (10), too.

Equation (9) satisfies the following property:

$$G_n \longrightarrow k \quad (n \longrightarrow \infty), \tag{11}$$

for any value of time-interval δ. The discrete Gompertz equation proposed by the author[3] also satisfies Eq. (11).

We obtain a regression equation:

$$Y_n = A + B \log G_{n+1}, \tag{12}$$

from Eq. (8), where

$$Y_n = \log G_{n+1} - \log G_n. \tag{13}$$

By using Eq. (12), we can estimate parameters k, a, and b as

$$\hat{k} = \exp\left(-\frac{\hat{A}}{\hat{B}}\right), \tag{14}$$

$$\hat{a} = \exp\left(\frac{\sum_{n=1}^{N} \log \frac{G_n}{k}}{\sum_{n=1}^{N} (1 - \delta \log \hat{b})^{-n}}\right), \tag{15}$$

$$\hat{b} = \exp\left(\frac{\hat{B}}{\delta}\right), \tag{16}$$

where $\hat{a}$, $\hat{b}$, and $\hat{k}$ are estimates of a, b, and k, and $\hat{A}$ and $\hat{B}$ are estimates of A and B in Eq. (12) respectively.

The regression equation (12) gives the same estimates $\hat{a}$, $\hat{b}$, and $\hat{k}$ for any time-interval δ, which is the same property as the regression equation proposed by the author[3].

IV. PROPOSED MEASURE

We propose a measure that is the correlation coefficient between $f(x_{n+1})$ and $f(x_{n+1}/x_n)$, where

$$f(x) = \begin{cases} x & \text{if LCM is assumed,} \\ \log x & \text{if GCM is assumed,} \end{cases} \tag{17}$$

and a data set $x_1, x_2, \ldots, x_N$ is given. This measure is used to determine whether the LCM or GCM is more appropriate. If a data set satisfies an exact solution of one model, the value of the proposed measure is -1, and that of the measure using the other model is greater than -1.

Therefore, a model that shows a smaller value of the above measure is regarded as an appropriate model for analyzing a data set.

We verified the proposed measure using two actual data sets. For one data set[4], L, the LCM is appropriate. For the other data set[1], G, the GCM is appropriate. The results are shown in Figs. 1 and 2. The values of the proposed measure in both figures indicate the appropriate model, although they show opposite results in the very early test phase. The proposed measure is practical because the estimated values of parameters are unstable and not useful in the very early test phase.

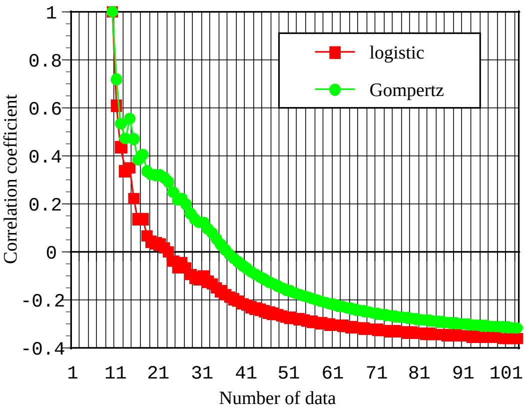

Fig. 1. Case of data set L.

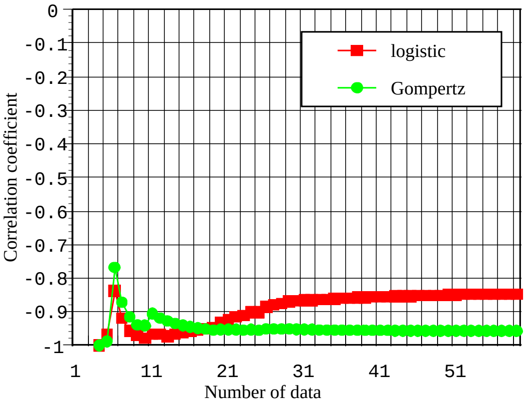

Fig. 2. Case of data set G.

V. CONCLUSION

We proposed a measure to determine whether the LCM or GCM is more appropriate. The GCM used in this measure is based on a new discrete Gompertz equation proposed in this paper. The new discrete Gompertz equation has the same properties as the discrete Gompertz equation proposed by the author. Using the new discrete GCM, the measure shows which model is appropriate for analyzing a data set.

REFERENCES

[1] T. Mitsuhashi: A Method of Software Quality Evaluation (in Japanese). JUSE Press, Tokyo, 1981.

[2] F. Morishita, The fitting of the logistic equation to the rate of increase of population density. *Research on Population Ecology*, **VII** (1965), 52–55.

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

[4] 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.

Copyright 2003 Chillarege Press