日本オペレーションズ・リサーチ学会
2004年秋季研究発表会

確率Gompertz差分方程式とそのパラメータ推定法

01206600 NTTサービスインテグレーション基盤研究所 ＊佐藤 大輔 SATOH Daisuke

1 はじめに

コンピュータネットワークの急速な普及により，コンピュータは我々の生活に必要不可欠な物になりつつある．これにより、ソフトウェアの信頼性はますます重要になってきている．開発側には、より短期に、より信頼性の高いソフトウェアの製造が要求される．このような状況では，ソフトウェアの信頼性評価は非常に重要である．

ソフトウェア開発のテスト工程で発見された総フォールト数がS字形成長曲線を示すとして，ロジスティック曲線やゴンペルツ曲線などが総フォールト数推定に用いられている．これらの曲線は微分方程式で記述されるものであるが，これを厳密解を持つ差分方程式を基にしたモデルに置き換え、この差分方程式を使用することにより，テスト工程初期のデータで従来よりも正確に推定できることが報告されている [1, 4, 5, 6]．さらに，この厳密解を持つ差分方程式によって多重共線性の解消など統計的な利点も指摘されている [2]．

しかしながら、これらの方程式は，決定論的であるため確率的な議論を行うことが不可能であった．logistic 方程式に確率則を導入した確率 logistic 差分方程式が提案され，n ステップでの累積フォールト数の分布を表現することが可能となった [3]．

本論では，実際のソフトウェア開発のテスト工程でより頻繁に観測される Gompertz 曲線モデルについて，その厳密解を持つ差分方程式 [1] に確率則を導入した確率 Gompertz 差分方程式を提案し，合わせて，そのパラメータ推定法を提案する．

2 Gompertz 差分方程式

Gompertz 方程式とは以下のような方程式であり、下に示す厳密解を持つ．

$$\frac{dG(t)}{dt} = (\log b)G(t)\log\frac{G(t)}{k}, \tag{1}$$

$$G(t) = ka^{b^t} \tag{2}$$

ここで，$k>0, 0<a<1, 0<b<1$ である．次に、Gompertz 方程式を厳密解を持つように差分した方程式 [1] とその解を示す．

$$G_{n+1} = G_n\left(\frac{G_n}{k}\right)^{\delta\log b}. \tag{3}$$

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

$$(k>0, 0<a<1, \frac{1}{e}<b^{\delta}<1) \tag{4}$$

これらの差分方程式を回帰式として用いることにより従来手法よりも早期に正確な推定値が得られる [1, 4]．

3 確率 Gompertz 差分方程式

厳密解を持つ確率 logistic 差分方程が提案され n ステップでの累積フォールト数の分布が得られている式 [3]．ここでは、厳密解を持つ Gompertz 差分方程式に確率則を導入することにより確率 Gompertz 差分方程式を提案し，最尤法によるパラメータ推定法を提案する．

独立同一分布に従う確率変数列を B_n として，式 (3) を拡張し，確率 Gompertz 差分方程式は，

$$G_{n+1} = G_n\left(\frac{G_n}{k}\right)^{\delta\log B_{n+1}} \tag{5}$$

と表される．厳密解は、

$$G_n = ka^{\prod_{i=0}^{n}(1+\delta\log B_i)} \tag{6}$$

となる．

次に，パラメータ推定法について考える．今、独立同一確率変数列 X_n が分布 $F(x)$, 確率密度関数 $f(x)$ に従っていると仮定し，

$$X_n = -\frac{1}{\delta\log B_n} - 1, n = 1, 2, \ldots, \tag{7}$$

とする．ここで

$$\mu = E[X_1] \tag{8}$$

とおく．パラメータ k は，次の式を最小にするものとする．

$$\sum_{n=0}^{N-1}\left(\log\frac{G_{n+1}}{G_n} + \frac{1}{\mu+1}(\log G_n - \log k)\right)^2 \tag{9}$$

これにより，総データ数を N とすると

$$\log k = \frac{(\mu+1)\sum_{n=0}^{N-1}\log\frac{G_{n+1}}{G_n} + \sum_{n=0}^{N-1}\log G_n}{N} \tag{10}$$

となる．一方，独立同一確率変数列 X_n における対数尤度 LL は，

$$LL = \sum_{n=0}^{N-1}\log f\left(\frac{\log k - \log G_{n+1}}{\log G_{n+1} - \log G_n}\right) \tag{11}$$

となり，式 (10) および

$$\frac{\partial}{\partial\mu}LL = 0 \tag{12}$$

からパラメータ k が決定し，式 (10) から μ が決定する．パラメータ a の推定値 $\hat{a}$ は，

$$\hat{a} = \exp\left(\frac{\sum_{n=0}^{N-1}\log\frac{G_n}{k}}{\sum_{n=0}^{N-1}(\frac{\mu}{\mu+1})^n}\right) \tag{13}$$

によって決定する．

$$0 < -\delta\log B_n < 1 \tag{14}$$

であるから，今、独立同一確率変数列 X_n が指数分布に従い，その分布関数 $F(x)$ を

$$F(x) = 1 - \exp\left(-\frac{x}{\mu}\right) \tag{15}$$

とする．対数尤度関数は，

$$\begin{aligned} LL &= \sum_{n=0}^{N-1}\log\left(\frac{1}{\mu}\exp\left(-\frac{1}{\mu}\frac{\log\frac{k}{G_{n+1}}}{\log\frac{G_{n+1}}{G_n}}\right)\right) \\ &= -N\log\mu \\ &\quad -\frac{1}{\mu}\left(\log k\sum_{n=0}^{N-1}\frac{1}{\log G_{n+1}-\log G_n}\right. \\ &\quad \left.-\sum_{n=0}^{N-1}\frac{\log G_{n+1}}{\log G_{n+1}-\log G_n}\right) \end{aligned} \tag{16}$$

となる．式 (10) を式 (16) に代入して，式 (12) から

$$\begin{aligned} \mu &= \frac{1}{N}\left\{\left(\frac{1}{N}\sum_{n=0}^{N-1}\log G_{n+1}\right)\times\right. \\ &\quad \left(\sum_{n=0}^{N-1}\frac{1}{\log G_{n+1}-\log G_n}\right) \\ &\quad \left.-\sum_{n=0}^{N-1}\frac{\log G_{n+1}}{\log G_{n+1}-\log G_n}\right\} \end{aligned} \tag{17}$$

を得る．パラメータ k は，式 (10) と式 (17) から得られる．式 (7) から

$$E[\delta\log B_n] = -\frac{1}{\mu+1}. \tag{18}$$

が得られる．

4 まとめ

本論では、厳密解を持つ Gompertz 差分方程式を拡張し、確率 Gompertz 差分方程式を提案し，その厳密解を示した．さらに，最尤法を使ったパラメータ推定法を提案し，指数分布を仮定して，実際にパラメータ推定を行った．これにより，Gompertz 曲線モデルは実データによく適合するものの方程式が決定論的であるため，最尤法による推定ができないという欠点を克服することができた．

参考文献

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

[2] D. Satoh: A discrete Bass model and its parameter estimation, Journal of the Operations Research Society of Japan, **44-1** (2001) 1–18.

[3] D. Satoh: A Discrete Stochastic Logistic Equation and a Software Reliability Growth Model, *Supplementary Proceedings of 13th ISSRE*, (IEEE Computer Society, Annapolis, November, 2002) 141–142.

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

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

[6] 山田, 井上, 佐藤, ソフトウェア信頼性評価のための差分方程式に基づく統計的データ解析, 日本応用数理学会論文誌, **12-2** (2002) 155–168.