

ソフトウェア信頼度成長モデルに関する統合モデルの 解析的パラメータ推定法

古 山 恒 夫^{†,☆}

ソフトウェア信頼度成長曲線に関する統合モデル^{1),2)}は、これまで提案された代表的なモデルをカバーするだけでなく、これまでモデル化されていなかった領域もカバーできる。そのため、このモデルを用いれば既存のソフトウェア信頼度成長モデルより高い精度で残存フォールト数を推定することができる。しかしながら、統合モデルを表す微分方程式の解の形式は、既存モデルの種別を表すパラメータ γ の値により飽和型モデルに限っても3つのグループに分かれる。そのため、最適なパラメータ群を推定するためには、それぞれのグループに対してパラメータの推定を試みる必要があった。また、最尤推定法によるパラメータ推定では、超越方程式を数値的に解く必要があることから、解の収束に時間がかかったり、解そのものが求まらない場合があるという問題もあった。本論文では、統合モデルを表す微分方程式の対数をとることにより、得られたデータ系列からこのモデルを表す微分方程式のパラメータを解析的に推定できることを示す。また、推定したパラメータをもとに具体的な推定曲線を求めるための手順を示す。実際の累積バグデータを用いて、本方式と最尤推定法による残存バグの推定誤差を比較した結果、その差は2%以下であった。

Analytical Parameter Estimation of the Manifold Growth Model Using Y-equation

TSUNEO FURUYAMA^{†,☆}

The manifold growth model that unifies existing software reliability growth models can cover a wide range of accumulated fault data including various types of data which are difficult to treat with existing models^{1),2)}. However, there are problems. For example, it sometimes takes a long time to solve transcendental equations to estimate parameters of the model by using maximum likelihood estimation, and the solution is difficult to obtain in some cases. This paper shows that the parameters of the differential equation that defines the manifold model can be analytically estimated for the given data by using a "Y-equation" derived from the differential equation. This paper also shows the concrete procedure for determining the most appropriate software reliability growth model, or the most appropriate shape of curve from Y-equation. Finally, the effectiveness of the Y-equation is shown by applying it to both ideal and actual data.

1. はじめに

信頼性の高いソフトウェアを予定期間内に開発するためには、残存フォールト数を高い精度で推定することが不可欠である。この課題を解決するために、これまで多くの方法が提案されてきたが、最も広く利用されているのは、試験工程で検出されるフォールトの累積値の傾向から残存フォールト数を推定する方法である。すなわち、時間対検出フォールト累積数の特性をもとにしたソフトウェア信頼度成長モデル (SRGM)

を利用する方法である。

SRGM はすでに Jelinski and Moranda³⁾, Musa⁴⁾, Littlewood⁵⁾, Goel and Okumoto⁶⁾, Yamada, et al.⁷⁾, Ohba⁸⁾, Tohma, et al.⁹⁾, Kanoun, et al.¹⁰⁾, Karunanithi, et al.¹¹⁾ など多くの研究者によって論じられている。しかし、これまで論じられているモデル^{3)~8)}は、指数形やS字形などの各モデル固有の定型的な特性を持った曲線をもとにしたものであり、利用できる範囲もそれぞれ限られている。そのため、あるモデルを固定的に使用すると、対象データの特性によっては残存フォールト数の推定精度が低くなることがある。実際に第一線のソフトウェア開発管理者は、適切なモデルを選択することが煩雑で難しいことから、特定のモデルを固定して利用することが多く、データ

[†] 日本電信電話株式会社ソフトウェア研究所
NTT Software Laboratories

[☆] 現在、東海大学開発工学部

Presently with Tokai University

によっては信頼性の予測が大きくずれることがあるというリスクをつねに抱えている。

モデルのパラメータの値によって曲線の特徴が変化するモデルもある⁸⁾が、モデルのパラメータの決定は難しく、またその値が決定されると定型の特性曲線になってしまうので、上記の問題が解決されたとはいえない。ニューラルネットによってモデルのパラメータを推定する方法も提案されている¹¹⁾が、この方法は学習パターンの集合のとり方によりモデルのパラメータが異なるという問題がある。

汎用モデルとして NHPP モデルや超幾何モデル⁹⁾が提案されている。しかし、これらのモデルは考え方の枠組みを与えるものであり、具体的な関数型は新たに仮定を設けたり、プロジェクトの特性や条件を考慮して個々に選択・決定しなければならない。すなわち、関数型を決定しなければならないという問題は依然として残されている。

これまで、モデルの提案だけでなく、それらのモデル間の優劣比較も議論されてきた。しかし、これまでの議論は、あるデータ群に対する各モデルの有効性と優劣を議論するものが多く、モデル間の相対的位置付けについての議論は行われていない。すなわち、モデルは個々に独立したものであり、モデル間の数学的な関係については論じられていない。

これらの問題を解決するひとつの手段として、上記のような代表的なモデルを包含する統合モデルが提案された^{1),2)}。このモデルは、これまで提案された多くのモデルをカバーできるだけでなく、これまでモデル化されていなかった領域もカバーする。そのため、このモデルを用いれば既存の SRGM より高い精度で残存フォールト数を推定することができる。しかしながら、統合モデルの解の形式は、飽和形モデルに限っても、統合モデルを表す微分方程式のパラメータ γ (既存モデルの種別を表すパラメータ) の値により3つのグループに分かれてしまう。そのため、最適なパラメータを推定するためには、それぞれのグループに対してパラメータの推定を試みる必要があった。また、最尤推定法によるパラメータ推定では、超越方程式を数値的に解く必要がある²⁾ことから解の収束に時間がかかったり、解そのものが求まらない場合があるという問題点も抱えていた。

本論文では、統合モデルを表す微分方程式の対数をとることにより、得られたデータ系列から統合モデルのパラメータを解析的に推定できることを示す。この方法では誤差の二乗和が最小になるようなパラメータを求めため、解そのものが求まらないことはまれで

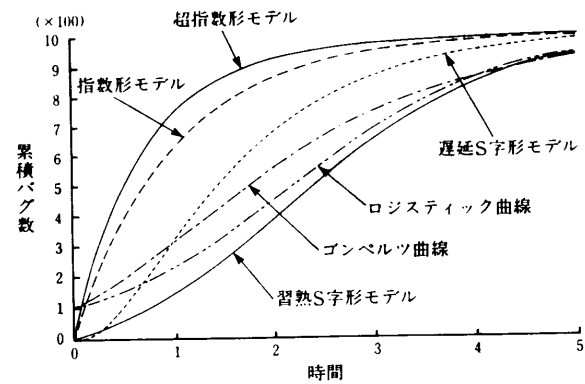


図1 代表的なソフトウェア信頼度成長モデルの例
Fig.1 Examples of representative software reliability growth models.

ある。これにより、与えられたデータの特性を容易にかつ確実に把握することができる。

2章で統合モデルの概要を紹介し、3章でパラメータの新しい推定法と、与えられたデータ列に適合する曲線の具体的な推定手順を示す。4章で理論曲線および実際の累積バグ曲線に対して本方法を適用した結果を示し、本方式が有効であることを示す。

2. 統合モデルの概要^{1),2)}

これまでに提案されているあるいは利用されている代表的な SRGM としては、指数形モデル、超指数形モデル、遅延S字形モデル、習熟S字形モデル、ギンベルツ曲線、ロジスティック曲線などがある (図1)。統合モデルはそれらのモデルあるいは曲線をカバーできるモデルである。

2.1 微分方程式

時刻 t における累積バグ数を y とする。統合モデルは次の微分方程式として定義される:

$$\frac{d(y + \delta)}{dt} \cdot (y + \delta)^{(\gamma-1)} = \alpha \cdot e^{-\beta t} \quad (1)$$

ここで、 $\alpha > 0$, $\beta \geq 0$, $\delta \geq 0$ である。 α は y 軸に関するスケールファクタ、 β は t 軸に関するスケールファクタである。 δ は $\gamma \leq 0$ のときに NHPP モデルの平均値関数の条件 ($t = 0$ で $y = 0$) を満たす解を持つために導入したパラメータで、この微分方程式の解となる関数の y 軸上の移動を意味する。

2.2 一般解

式 (1) の一般解は次のようになる。ただし、 y_0 は $t = 0$ のときの y の値である。

(A) $\beta > 0$, $\gamma \neq 0$ の場合

$$y = N(1 - a \cdot e^{-\beta t})^c - \delta$$

$$= N\{(1 - a \cdot e^{-\beta t})^c - (1 - a)^c\} + y_0 \quad (2)$$

ただし,

$$N = \left\{ \frac{\alpha \cdot \gamma + \beta(y_0 + \delta)^\gamma}{\beta} \right\}^{\frac{1}{\gamma}}, \quad (3)$$

$$a = \frac{\alpha \cdot \gamma}{\alpha \cdot \gamma + \beta(y_0 + \delta)^\gamma}, \quad (4)$$

$$b = \beta, \quad (5)$$

$$c = \frac{1}{\gamma}. \quad (6)$$

である。ここで、 N は (定数分 ($a = 1$ の場合は 0) を差し引いて) 試験開始時点に含まれるバグ総数、 b は残存フォールト 1 個あたりの検出率を表す。 a と c は成長曲線の形状 (たわみ) を決定するプロジェクトごとのフォールト検出特性と解釈できる。

(B) $\beta > 0$, $\gamma = 0$ の場合

$$y = N \exp(-k e^{-\beta t}) - \delta$$

$$= N \{ \exp(-k e^{-\beta t}) - \exp(-k) \} + y_0 \quad (7)$$

ただし,

$$N = (y_0 + \delta) \exp\left(\frac{\alpha}{\beta}\right), \quad (8)$$

$$k = \frac{\alpha}{\beta}, \quad (9)$$

$$b = \beta. \quad (10)$$

である。ここで、 N と b はそれぞれ式 (2) の N と b と、 k は式 (2) の a および c と同じ意味の特性である。なお、 k は $\gamma \rightarrow 0$ ($c \rightarrow \infty$, $a \rightarrow 0$) のときの c と a の積と等しい。

一般解は、数式上は $\gamma > 0$ の場合と $\gamma < 0$ の場合でまったく同じ形式になるので、まとめて取り扱うことができる。しかし、一般解のべき乗の指数 c は、式 (6) より $\gamma > 0$ では正であるが $\gamma < 0$ では負となるため数値計算上は連続にはならず、実際に数値計算をするにあたってはそれぞれ別々のものと考えなければならない。すなわち、 $\gamma < 0$ の場合の一般解は式 (2) とは別の次のような式として取り扱って計算した方がよい。

$$y = \frac{N}{(1 + \Psi e^{-\beta t})^{cc}} - \delta$$

$$= N \left\{ \frac{1}{(1 + \Psi e^{-\beta t})^{cc}} - \frac{1}{(1 + \Psi)^{cc}} \right\} + y_0. \quad (11)$$

ただし、 $\psi = -a > 0$, $cc = -c > 0$ である。

このように、統合モデルの一般解は、 $\beta > 0$ に限っ

ても実質上は 3 つの形式に分かれる。

3. パラメータ推定法

統合モデルのパラメータの推定方法としては、これまで最尤推定法を用いる方法が提案されていた²⁾。しかし、最尤推定法では超越方程式の解を数値的に求める必要があるため、計算の手間が大きく、またときとして解そのものが求まらないことがある。また、 $y_0 \neq 0$ のデータの場合には、 $y_0 = 0$ となるように y 軸方向にグラフを平行移動してからパラメータを推定するなどの手段を講じる必要もあった。

本章では、統合モデルで $\delta = 0$ の場合は解のパラメータを、対数誤差の二乗和最小となる方法を用いて解析的に推定することができることを示す。

3.1 Y 方程式

$\delta = 0$ の場合、式 (1) は次のようになる。

$$y' \cdot y^{\gamma-1} = \alpha e^{-\beta t}. \quad (12)$$

この式の両辺の対数をとると、次の式が得られる。

$$\ln(y') + (\gamma - 1) \ln(y) = \ln(\alpha) - \beta t. \quad (13)$$

ここで

$$Y_1 = t, \quad (14)$$

$$Y_2 = \ln(y), \quad (15)$$

$$Y_3 = \ln(y'), \quad (16)$$

と変数変換をして式を整理すると、

$$\beta Y_1 + (\gamma - 1) Y_2 + Y_3 - \ln(\alpha) = 0 \quad (17)$$

が得られる。式 (17) を Y 方程式と呼ぶことにする。

式 (17) が線形となることから、 $i = 0, 1, \dots$ に対する (t_i, y_i, y'_i) のデータの組をそれぞれ (Y_{1i}, Y_{2i}, Y_{3i}) の組に変換すると、 Y_1, Y_2, Y_3 を軸とする 3 次元グラフ上ではその軌跡は 1 つの平面上に存在することを示している。一方、式 (17) の導出過程の逆をたどることにより、Y 方程式が線形となる方程式は統合モデルの解となることが分かる。

3.2 従来の SRGM の Y 方程式

これまでに提案された代表的な SRGM の Y 方程式は、次のいずれかの方法によって求めることができる。

- モデルの一般式を微分して両辺の対数をとる。
- 統合モデルの微分方程式のパラメータ^{1), 2)}を式 (17) に代入する。

これまでに提案された主な SRGM に対する Y 方程式を表 1 に示す。指数形モデル、ゴンペルツ曲線、ロジスティック曲線、ワイブル過程モデルの場合は、統合モデルのパラメータ δ が 0 となることから、式 (17) より線形の Y 方程式が得られる。しかし、習熟 S 字形モデルでは δ が 0 とはならないため、線形の Y 方

表1 従来の代表的SRGMとそのY方程式
Table 1 Existing SRGM and their Y-equation.

モデル (曲線)	一般式	統合モデルの微分方程式のパラメータ ^{1), 2)}				一次微分 y'	Y方程式
		α	β	γ	δ		
指数形	$y = N(1 - e^{-bt})$	Nb	b	1	0	$Nb e^{-bt}$	$Y_3 + bY_1 - \ln(Nb) = 0$
ゴンペルツ	$y = N \exp(-k e^{-bt})$ (注1)	kb	b	0	0	$kb e^{-bt} y$	$Y_3 - Y_2 + bY_1 - \ln(kb) = 0$
ロジスティック	$y = \frac{N}{1 + \psi e^{-bt}}$	$\frac{b\psi}{N}$	b	-1	0	$\frac{b\psi e^{-bt}}{N} \cdot y^2$	$Y_3 - 2Y_2 + bY_1 - \ln\left(\frac{b\psi}{N}\right) = 0$
習熟S字形	$y = N \frac{1 - e^{-bt}}{1 + \psi e^{-bt}}$ (注2)	$\frac{b\psi^2}{N(\psi + 1)}$	b	-1	$\frac{N}{\psi}$	$\frac{Nb(\psi + 1)e^{-bt}}{(1 + \psi e^{-bt})^2}$	—
遅延S字形	$y = N \{1 - (1 + bt)e^{-bt}\}$	$-2.2 N^{0.46} b$	$-0.78b$	-0.46	0	$Nb^2 t e^{-bt}$	$Y_3 - 0.54Y_2 + 0.78bY_1 - \ln(0.78Nb) = 0$ (近似式)
超指数形	$y = \sum_{i=1}^n N_i (1 - e^{-b_i t})$	$-\frac{N\beta}{\gamma}$	—	> 1	0	$\sum_{i=1}^n N_i b_i e^{-b_i t}$	(Y ₂ の係数が正となるY方程式。 ただしY ₃ =1とする)
ワイブル過程	$y = \lambda t^n$	$n\lambda^{\frac{1}{n}}$	0	$\frac{1}{n}$	0	$\lambda n \left(\frac{y}{\lambda}\right)^{\frac{n-1}{n}}$	$Y_3 + \frac{1-n}{n}Y_2 - \ln(n) - \frac{1}{n} \ln(\lambda) = 0$

(注1) 普通ゴンペルツ曲線は $y = N \cdot p^{q^t}$ で表わされるが、expを使った式に変形した方が扱い易いので、この式のように変形する。ここで、 $k = -\ln(p)$ 、 $b = -\ln(q)$ である。

(注2) $y = N \frac{\psi + 1}{\psi} \left(\frac{1}{1 + \psi \exp(-bt)} - \frac{1}{\psi + 1} \right)$ と変形することにより、ロジスティック曲線をy軸方向に移動したものと等価であることがわかる。

程式を得ることはできない。遅延S字形モデルと超指数形モデルはいずれも統合モデルの解ではないが、これらのモデルは $\delta = 0$ の場合の統合モデルの一般解で近似できるため、次のようにして近似的な線形のY方程式を得ることができる。

(1) 遅延S字形モデル

遅延S字形モデルの一般式

$$y = N \{1 - (1 + bt)e^{-bt}\} \quad (18)$$

は、統合モデルの解

$$y = N (1 - e^{-0.78bt})^{2.16} \quad (19)$$

で近似できることから、遅延S字形モデルのY方程式は次の式で近似できることが分かる。

$$Y_3 - 0.54Y_2 + 0.78bY_1 - \ln(0.78Nb) = 0. \quad (20)$$

(2) 超指数形モデル

超指数形モデル

$$y = \sum_{i=1}^n N_i (1 - e^{-b_i t}) \quad (21)$$

は、 $\gamma > 1$ となる統合モデルで近似できることから、Y₂の係数が正の値をとるY方程式で近似できる。ただし、Y₃の係数は1とする。

3.3 統合モデルの解のパラメータの推定手順

パラメータ推定の具体的な手順を以下に示す。

(手順1) $i = 0, 1, \dots, n$ に対応する (t_i, y_i, y'_i) のデータの組を式(14)、式(15)、式(16)を用いて (Y_{1i}, Y_{2i}, Y_{3i}) の組に変換する。

(手順2) 3組以上の (Y_{1i}, Y_{2i}, Y_{3i}) の組からY方程式の係数を求める。データが4組以上ある場合は、Y₃に関して誤差二乗和を最小とする回帰式を求める。なお、 γ のとりうる範囲が事実上(-1)から2の間であるため、Y₂に関して回帰式をとると、 γ が1の場合に回帰式を求めることができなくなることがある。そのためY₃に関する回帰式を求める。

(手順3) 統合モデルを表す微分方程式の各パラメータの推定値 $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ を回帰式の係数から式(17)を用いて計算する。 $\hat{\beta} \leq 0$ の場合は、飽和曲線とならない ($t \rightarrow \infty$ で $y \rightarrow \infty$) ので推定を中止する。

(手順4) $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ の値の妥当性をF検定を用いて検定する。すなわち、

$$F = \frac{S_R/2}{S_E/(n-3)} > F_{(2, n-3; 0.01)} \quad (22)$$

が成立すれば、パラメータの推定値は99%の信頼度で有意である。ただし、

$$S_R = \sum_{i=1}^n (\hat{Y}_{3i} - \bar{Y})^2 \quad (23)$$

$$S_E = \sum_{i=1}^n (Y_{3i} - \hat{Y}_{3i})^2 \quad (24)$$

であり、 \bar{Y} は Y_{3i} ($i = 0, 1, \dots, n$) の平均値、 \hat{Y}_{3i} は Y_{3i} の推定値である。

手順5以下は $\hat{\gamma}$ の値により異なる。

(A) $\hat{\gamma} \neq 0$ の場合

(手順5) 統合モデルの一般解のパラメータ b と c の推定値 \hat{b} と \hat{c} を式(5)と式(6)を用いて計算する。

(手順6) 統合モデルの一般解のパラメータ N と a の推定値 \hat{N} と \hat{a} を求める。 \hat{N} と \hat{a} の求め方には、次の2通りの方法がある。ただし、 \hat{y}_0 を $t=0$ での y の推定値、 \hat{y}_n を最新時刻 $t=t_n$ での y の値(最新値)の推定値とする。

(手順6-a) 最新値一致条件 ($\hat{y}_n = y_n$) に基づく推定法: 式(3)と式(4)で $\delta = 0$ とおいた式、および

$$y_n = \hat{y}_n = N \left(1 - ae^{-bt_n}\right)^c \quad (25)$$

を、 N 、 a 、 y_0 について解くことにより \hat{N} と \hat{a} を求める。

$$\hat{N} = \left(\frac{\hat{\alpha} \cdot \hat{\gamma}}{\hat{\beta}} e^{-\hat{b}t_n} + \hat{y}_n^{\frac{1}{c}} \right)^{\frac{1}{1-c}}, \quad (26)$$

$$\hat{a} = \frac{\hat{\alpha} \cdot \hat{\gamma}}{\hat{\beta}} \frac{1}{\hat{N}^{\hat{\gamma}}}. \quad (27)$$

(手順6-b) 推定誤差二乗和最小に基づく推定法: 次の式で定義される $t = t_i$ における測定値 y_i とその推定値 \hat{y}_i の誤差二乗和 S が最小となるような y_0 の推定値を求める。

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - f(t_i; N(y_0), a(y_0), \hat{b}, \hat{c}))^2. \quad (28)$$

ここで関数 f は統合モデルの一般解である式(2) (あるいは式(11)) あるいは式(7)の右辺を表すものである。この段階では、 \hat{b} と \hat{c} の値が定まっているので式(2) (あるいは式(11)) と式(7)のいずれの式を用いるかは定まっている。また、この段階では N と a は y_0 のみの関数である。

式(28)を y_0 で微分した式は超越方程式となるので、数値計算で \hat{y}_0 を求めないといけないが、1変数であることと、 \hat{y}_0 の期待値が y_0 なので、 y_0 を中心に2分法などの既存の数値計算アルゴリズムで容易に式(28)を最小とする \hat{y}_0 を求めることができる。

\hat{y}_0 を式(3)と式(4)に代入して \hat{N} と \hat{a} を得る。

(B) $\hat{\gamma} = 0$ の場合

(手順5) 統合モデルの一般解のパラメータ b と k の推定値 \hat{b} と \hat{k} を式(9)および式(10)を用いて計算する。

(手順6) 統合モデルの一般解のパラメータ N の推定値 \hat{N} を求める。

(手順6-a) 最新値一致条件 ($\hat{y}_n = y_n$) に基づく推定法:

$$y_n = \hat{y}_n = N \exp\left(-\hat{k}e^{-\hat{b}t_n}\right) \quad (29)$$

よりを \hat{N} 求める。

$$\hat{N} = y_n \exp\left(\hat{k}e^{-\hat{b}t_n}\right). \quad (30)$$

(手順6-b) 推定誤差二乗和最小に基づく推定法: $t = t_i$ における測定値 y_i とその推定値 \hat{y}_i の誤差二乗和

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left(y_i - N \exp\left(-\hat{k}e^{-\hat{b}t_i}\right) \right)^2 \quad (31)$$

を最小とする N の推定値を次の式より求める。

$$\hat{N} = \frac{\sum_{i=1}^n y_i \exp\left(-\hat{k}e^{-\hat{b}t_i}\right)}{\sum_{i=1}^n \exp\left(-2\hat{k}e^{-2\hat{b}t_i}\right)}. \quad (32)$$

4. 残存バグ推定への応用

統合モデルのパラメータを最尤推定法により推定した場合と、Y方程式により解析的にパラメータを推定した場合を、理論データおよび実データを用いて比較する。推定精度の評価尺度としては、サンプル時系列データのうちの Endpoint (最後) のデータを途中の時系列までのサンプルデータを用いて、どの程度の誤差内で推定できるか (これを Endpoint 推定と呼ぶことにする) を用いる。また、比較モデルとして、特性の異なる指数形モデルと遅延S字形モデルも用いる。

Y方程式によりパラメータを推定する場合には、一次の微分値が必要であるが、累積バグデータからは微分値が得られないので、近似的に差分値を用いる。以下では、各時刻が等間隔の場合を扱い、一次差分として最も簡単な次の式を用いる。

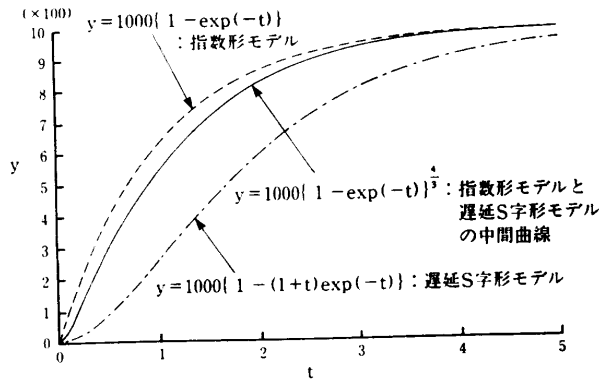


図2 指数形および遅延S字形モデルの曲線とそれらの中間曲線
Fig. 2 Curves of exponential, delayed S-shaped and their intermediate models.

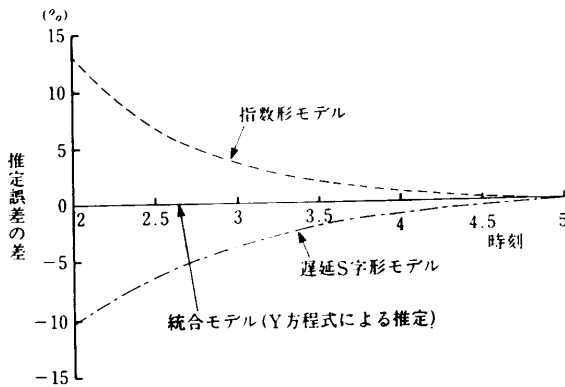


図3 中間曲線に対する Endpoint 推定誤差の推移
Fig. 3 Endpoint estimation derivation from the intermediate curve.

$$y'(t_i) = \frac{1}{2\Delta} \{y(t_{i+1}) - y(t_{i-1})\}. \quad (33)$$

ただし、 Δ は単位時間間隔である。以下では $\Delta = 1$ とするが、これにより一般性を失うことはない。

4.1 理論データによる推定精度の比較

指数形モデルと遅延S字形モデルの中間に位置する理論上（計算上）のデータ系列をY方程式で推定した場合の推定精度を確かめる。推定対象は式(2)で $N = 1000$, $a = 1$, $b = 1$, $c = 4/3$, $y_0 = 0$ とした理論曲線を $t = 0$ から5まで0.1きざみに y の値を求めた系列データである（図2）。

図3に統合モデルに基づいて最尤推定法で推定した場合と同じくY方程式で推定した場合の推定結果を示す。あわせて、指数形モデル、遅延S字形モデルで推定した結果も示す。Y方程式で推定した場合は、最尤推定法の場合と同様に推定誤差はない。

4.2 実データによる推定精度の比較

大規模ソフトウェア開発の試験工程で得られた時系

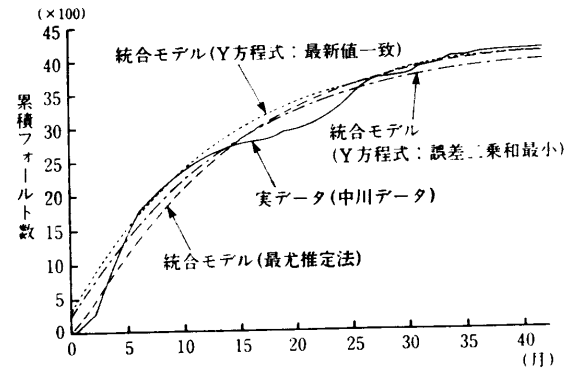


図4 実データ（中川データ）と Endpoint 推定曲線
Fig. 4 Nakagawa data with estimated curves.

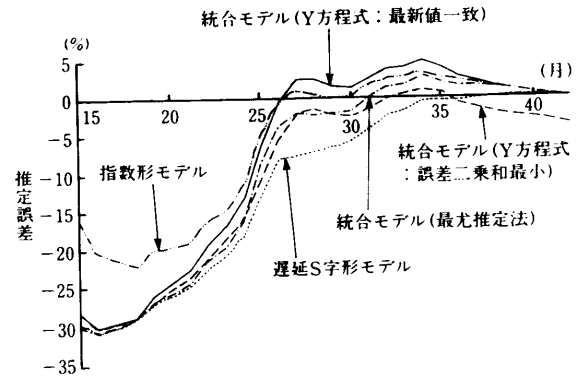


図5 Endpoint 推定誤差の推移
Fig. 5 Comparison of endpoint estimation error over time for Nakagawa data.

列データ（中川データ^{1),2),12)}を図4に示す。統合モデルに基づいて最尤法で推定した場合と、同じくY方程式で推定した場合のEndpoint推定誤差の変化を図5に示す。あわせて指数形モデルと遅延S字形モデルで推定した結果も示す。また、 b , c の推定値の変動の比較を図6と図7に示す。特徴を次に要約する。

- (a) Y方程式による一般解のパラメータ b と c の推定値は、最尤推定法による推定値の良い近似を与えている。
- (b) Y方程式によるEndpoint推定についての推定誤差は、最尤推定法によって得られる推定誤差の良い近似を与えている。推定の前半から中盤にかけては ($t \leq 35$)、誤差二乗和最小条件が良い近似を与え、その差は2%以内である。後半になると ($36 < t \leq 42$) 最新値一致条件が良い近似を与える。

4.3 考察

(1) Y方程式の適用可能範囲

表2に最尤推定法とY方程式による推定法の推定

表2 パラメータ推定が可能な統合モデルの一般解の範囲
Table 2 Type of general solution of the manifold model whose parameters are capable to be estimated.

δ	0		+	
	y_0		+	
τ	0	+	0	+
+	$y = N(1 - e^{-bt})^c$ 指数形モデル* 指数形モデル* 遅延S字形モデル*	$y = N(1 - a \cdot e^{-bt})^c$	$y = N \{ (1 - a \cdot e^{-bt})^c - (1 - a)^c \}$	$y = N \{ (1 - a \cdot e^{-bt})^c - (1 - a)^c \} + y_0$
0	-	$y = N \exp(-k e^{-bt})$ ゴンベルツ曲線	$y = N \exp(-k e^{-bt}) - e^{-k}$	$y = N \exp(-k e^{-bt}) - e^{-k} + y_0$
-	-	$y = \frac{N}{1 + \psi e^{-bt}}$ ロジスティック曲線	$y = N \left(\frac{1}{1 + \psi e^{-bt}} - \frac{1}{\psi + 1} \right)$ 習熟S字形モデル	$y = N \left(\frac{1}{1 + \psi e^{-bt}} - \frac{1}{\psi + 1} \right) + y_0$

(凡例) : Y方程式を利用できる範囲 ($\delta=0$ の範囲)
 : 最尤推定法を利用できる範囲 ($y_0=0$ の範囲)
 * : 統合モデルで近似して推定

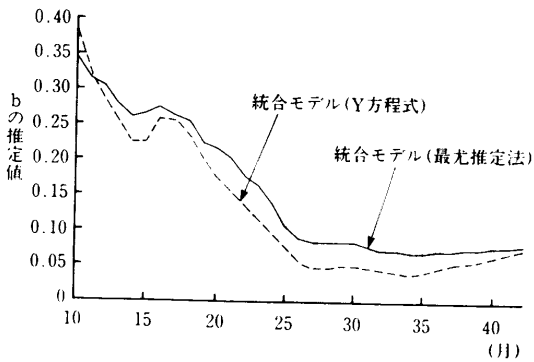


図6 パラメータbの推定値の比較
Fig. 6 Variation of estimated value of parameter b.

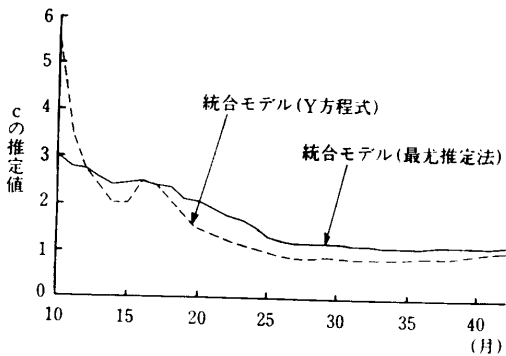


図7 パラメータcの推定値の比較
Fig. 7 Variation of estimated value of parameter c.

可能な統合モデルの一般解の範囲を示す。Y方程式では推定可能範囲がやや狭くなるが、習熟S字形モデルを除いてこれまでの代表的なSRGMをカバーしているので、実用上十分と考えられる。

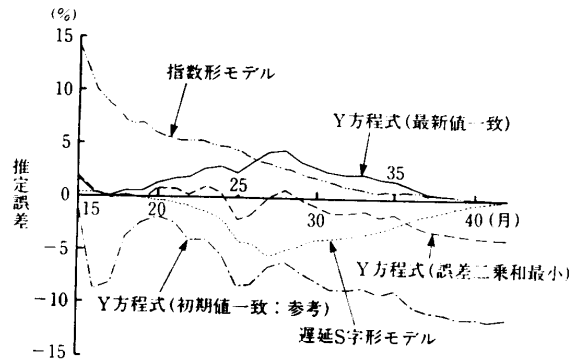


図8 最尤推定法に対するY方程式のEndpoint推定誤差の差
Fig. 8 Difference of endpoint estimation error between maximum likelihood estimation method and Y-equation method.

(2) Nとaの推定法 (図8)

推定時刻とEndpointまでの時間が小さい場合は最新値一致条件による方法の方が、推定誤差が小さくなり、最尤推定法による推定誤差の良い近似を与える。しかし、最新値一致条件の方法は最新値に敏感に反応するため、最新のデータが加わるたびに推定値が大きく変動することがある。

一方、誤差二乗和最小条件による方法は、Y方程式の最終値一致条件や最尤推定法と異なり、最新時刻がEndpointであるときのEndpoint推定が0とはならない。この方法は、測定データ全体を平均的に利用するため、推定時刻とEndpointまでの時間が大きい場合は一般に推定誤差が小さく、最新データの追加に対しても推定値は比較的安定している。

なお、この他に初期値一致条件 ($\hat{y}_0 = y_0$) に基づく推定法も存在するが、その推定法は最新のデータより初期段階のデータを重視することになるため、一般に上記2つの推定法より推定精度が低くなる。

5. むすび

(1) 統合モデルの解は $\gamma = 0$ で不連続となる3つのグループに分かれる。そのため、実際のパラメータ推定にあたってはそれら3つのグループのどれに属するかを試みなければならない。また、これまでの最尤推定方によるパラメータ推定では超越方程式を解く必要があり、解を求めるのに時間がかかったり、ときには解そのものの収束が困難であるという問題があった。しかし、Y方程式を用いれば、解析的に統合モデルのパラメータ α , β , γ を容易にかつ確実に推定できる。

(2) 本方式の有効性を検証するために、本方式と最尤推定法による推定誤差の比較を行った結果、理論データに対しては差はなく、実データでも2%以内の差であることが分かった。

参考文献

- 1) 古山恒夫, 中川 豊: ソフトウェア信頼度成長曲線に関する統合モデルと有効性の検証, 情報処理学会ソフトウェア工学研究会, Vol.97-10, pp.73-80 (1994).
- 2) Furuyama, T. and Nakagawa, Y.: A Manifold Growth Model that Unifies Software Reliability Growth Models, *Int. J. of Reliability, Quality and Safety Engineering*, Vol.1, No.2, pp.161-184 (1994).
- 3) Jelinski, Z. and Moranda, P.: Software Reliability Research, *Statistical Computer Performance Evaluation*, Freiburger, W. (Ed.), pp.465-484, Academic Press, New York (1972).
- 4) Musa, J.D.: A Theory of Software Reliability and Its Application, *IEEE Trans. Softw. Eng.*, Vol.SE-1, No.3, pp.312-327 (1975).
- 5) Littlewood, B.: Theories of Software Reliability: How Good Are They and How Can They Be Improved?, *IEEE Trans. Softw. Eng.*, Vol.SE-6, No.5, pp.489-500 (1980).
- 6) Goel, A.L. and Okumoto, K.: Time-Dependent Error-Detection Rate Model for Software

Reliability and Other Performance Measures, *IEEE Trans. Rel.*, Vol.R-28, No.3, pp.206-211 (1979).

- 7) Yamada, S., Ohba, M. and Osaki, S.: S-Shaped Reliability Growth Modeling for Software Error Detection, *IEEE Trans. Rel.*, Vol.R-32, No.5, pp.475-478 (1983).
- 8) Ohba, M.: Software Reliability Analysis Models, *IBM J. Res. Dev.*, Vol.28, No.4, pp.428-443 (1984).
- 9) Tohma, M., Tokunaga, K., Nagase, S. and Murata, Y.: Structural Approach to the Estimation of the Number of Residual Software Faults Based on the Hyper-geometric Distribution, *IEEE Trans. Softw. Eng.*, Vol.SE-15, No.3, pp.345-355 (1989).
- 10) Kanoun, K., Martini, M.R.B. and Souza, J.M.: A Method for Software Reliability Analysis and Prediction Application to the TROPICO-R Switching System, *IEEE Trans. Softw. Eng.*, Vol.SE-17, No.4, pp.334-344 (1991).
- 11) Karunanithi, N., Whitley, D. and Malaiya, Y.K.: Prediction of Software Reliability Using Connectionist Models, *IEEE Trans. Softw. Eng.*, Vol.SE-18, No.7, pp.563-574 (1992).
- 12) Nakagawa, Y. and Hanata, S.: An Error Complexity Model for Software Reliability Measurement, *Proc. 11th ICSE*, pp.230-236 (1989).

(平成8年2月19日受付)

(平成8年9月12日採録)

古山 恒夫 (正会員)



1945生。1968年東京大学工学部計数工学科卒業。1973年同大学院博士課程修了。同年日本電信電話公社入社。横須賀電気通信研究所で、拡張型言語, Ada, Common LISP

などの言語処理プログラムの研究実用化に従事。日本電信電話(株)ソフトウェア研究所でソフトウェアプロジェクト管理法, ソフトウェア品質保証法, ソフトウェア見積り法などの研究実用化に従事。平成8年東海大学開発工学部情報通信工学科教授。平成6年度山下記念研究賞受賞, IEEE, 電子情報通信学会, 日本ロボット学会各会員。工学博士。