

### 成長曲線の一般化と差分方程式による解析に関する研究

藤沢, 峻作 / Fujisawa, Shunsaku

---

(出版者 / Publisher)

法政大学大学院理工学・工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学・工学研究科編 / 法政大学大学院紀要. 理工学・工学研究科編

(巻 / Volume)

58

(開始ページ / Start Page)

1

(終了ページ / End Page)

6

(発行年 / Year)

2017-03-31

(URL)

<https://doi.org/10.15002/00014364>

# 成長曲線の一般化と差分方程式による解析に関する研究

A STUDY ON GENERALIZATION OF GROWTH CURVES AND ITS DISCRETIZATION

藤澤 峻作

Shunsaku FUJISAWA

指導教員 木村 光宏

法政大学大学院理工学研究科システム理工学専攻修士課程

In this study, we newly propose a generalized growth curve model of discrete type, which is based on the property of the Gompertz and logistic curve models. In particular, the bi-linearization technique is applied to the basic differential equation, and we discretize it to the difference equations so as to hold the property of integrable systems. These models can be used to predict the future behavior of the time series data by estimating the constant parameters included in the model. In the numerical examples, we analyze the actual data sets by using our model, and also we show 95% confidence intervals of the estimated growth curve via the bootstrap method.

**Key Words** : Gompertz curve, logistic curve, discretized growth model, bi-linearization method

## 1. はじめに

### (1) 研究背景

古くから、コンピュータウイルスの感染数、プログラムのバグ発見数の累積変化量など、当初は少なく途中で急激に増加し、終盤少なくなるような現象は多く見られている。この現象はグラフで表すと S 字曲線を描く。これを一般に成長曲線という。成長曲線を工学的に利用することを考えると、現象を観測し、成長曲線を当てはめ、将来予測をすることが挙げられる。成長曲線は元となる微分方程式で与えられているため、予測するという事は微分方程式のパラメータを推定することになる。一般には、時間軸が離散値のデータを取り扱うことが多いため [1]、微分方程式を差分方程式に置き換え、解を求めパラメータ推定を行うことも既往の研究では示されている。差分化を行う場合、微分方程式の性質を保つことができる「可積分差分」が多く用いられる。その差分方程式を用いることで、基礎方程式として解となる関数を求め、パラメータ推定などを行うことができる。

代表的な成長曲線に、ゴンベルツ曲線モデルとロジスティック曲線モデルがある。この二つのモデルを差分化し、それぞれのモデルを考察する研究は数多くある [1]-[4]。しかし、それぞれが適用例として挙げたデータセットに対してはうまく性能評価を發揮できるものの、それ以外のデータに対しては、うまく当てはまらないということがしばしば生じることが挙げられる。そのため、現状においてこの二つのモデルを含む、様々な成長曲線モデルを表すことができる離散型一般化成長曲線モデルを構築することが有効だと考えられる。

予測する上で、その予測値はより早期に、より正確であることが望まれる。しかし、この要求は非常に厳しく、満足することは非常に難しい。なぜなら、実測データが少なくとも変曲点を越えないとある程度正確な予測値は得られないといわれており、変曲点を越えたかどうか判定することは難しいという問題がある。そこで、文献 [5] ではソフトウェアのパ

グのデータセットを用いて、点推定ではなくブートストラップ法により区間推定を行い、予測を行うためのデータ数はどのくらい必要であるのかという研究が行われた。

### (2) 研究目的

本研究では、離散型一般化成長曲線モデル、すなわち離散型の少なくともゴンベルツ曲線とロジスティック曲線を一つの定数パラメータを導入することで統一的に表現できるモデルを、双線形化法を用いて構築する。一般に、非線形微分方程式の厳密解を求めること及び差分方程式の導出は、難しいと言われている。本研究では、これらを容易にするため、双線形化法を用いた。そして、実測データを用いた結果から、構築したモデルが、有用であるか検証・考察を行う。また、構築した成長曲線モデルの変曲点とブートストラップ法により導出した 95% 信頼区間を求め、点推定と区間推定の視点から考察を行い、予測するために必要な日数を検証する。

## 2. 従来知られている二つの成長曲線モデル

本研究では、離散型一般化成長曲線モデルを導出するにあたり、ゴンベルツ曲線モデルとロジスティック曲線モデルの線形式の性質に着目した。本章では線形式を導出する過程を示す。

### (1) 既往の離散型ゴンベルツ曲線モデル

連続関数  $G(t)$  を時間  $t$  までの累積値として、ゴンベルツ曲線は微分方程式

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

で記述され、解は

$$G(t) = ka^{bt} \quad (k > 0, 0 < a < 1, 0 < b < 1), \quad (2)$$

である。ここで  $a, b, k$  は定数で  $k$  は飽和値、 $a$  は  $G(0) = ka$  の値、 $b$  は  $\log \frac{G(1)}{k} / \log a$  の値である。式 (1) は厳密解を持つような差分化 [4] ができる。その差分方程式は  $G_n$  を離散関数

として

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

であり、厳密解は

$$G_n = ka^{(1+\delta \log b)^n} (e^{-1} < b^\delta < 1), \quad (4)$$

である。これは  $n \rightarrow \infty, \delta \rightarrow 0$  のとき、式 (2) に一致する性質を持つ。ここで  $t = n\delta$  と置いている。式 (2) の両辺に対数をとると

$$\log G(t) = \log k + b^t \log a, \quad (5)$$

になる。式 (5) の両辺を微分すると

$$\log \frac{dG(t)}{dt} / G(t) = \log((\log a)(\log b)) + t \log b, \quad (6)$$

の線形式を導くことができる。式 (6) は未知パラメータの推定に利用される [1]。

## (2) 既往の離散型ロジスティック曲線モデル

連続関数  $L(t)$  を時間  $t$  までの累積値としてロジスティック曲線は微分方程式

$$\frac{dL(t)}{dt} = \frac{r}{k} L(t)(k - L(t)), \quad (7)$$

で記述され、 $L(0) = \frac{k}{1+m}$  として解くと、解は

$$L(t) = \frac{k}{1 + m \exp(-rt)}, \quad (8)$$

である。式 (7) は厳密解を持つような差分化 [6] ができる。その差分方程式は  $L_n$  を離散関数として

$$L_{n+1} - L_n = \delta \frac{r}{k} L_{n+1}(k - L_n), \quad (9)$$

であり、厳密解は

$$L_n = \frac{k}{1 + m(1 - \delta r)^n}, \quad (10)$$

である。これは  $n \rightarrow \infty, \delta \rightarrow 0$  のとき、式 (8) に一致する性質を持つ。ここで  $t = n\delta$  と置いている。式 (8) を微分し、 $L(t)$  で 2 回割った後で両辺に対数をとると、

$$\log \frac{dL(t)}{dt} / L(t)^2 = \log \frac{mr}{k} - rt, \quad (11)$$

線形式を導くことができる。式 (6) と同様、式 (11) を用いて未知パラメータの推定法が知られている [1]。

## 3. 離散型成長曲線モデルの一般化

本章では、既往の研究で示され、前章で述べたことを踏まえ、双線形化法を用いた離散型成長曲線モデルの一般化を行う。まず、ある連続関数  $H(t)$  について

$$\log \frac{dH(t)}{dt} / H(t)^d = A - Bt, \quad (12)$$

を考える [7]。ここで、 $A, d$  は実数、 $B$  は正の数とする。式 (12) は

$$\frac{dH(t)}{dt} = e^A e^{-Bt} H(t)^d, \quad (13)$$

となり、ベルヌーイ方程式の形となる。離散型成長曲線の一般化は式 (13) の差分化を考えればよい。なぜならば、式 (13) の微分方程式は、 $d = 0$  のときは指数形曲線モデル、 $d = 1$  のときはゴンベルツ曲線モデル、 $d = 2$  のときはロジスティック曲線モデルになるからである。ここで、指数形曲線とは

$$H(t) = H_0 + \frac{e^A}{B} (1 - e^{-Bt}), \quad (14)$$

の形で与えられ、ソフトウェア信頼性分野において、発見されるバグ数の累積値の時間変化を表すものとしてよく用いられる。

式 (13) に対して双線形化法による差分化 [8] を行う。以下に双線形化法を用いて、式 (13) の  $d = 1$  と  $d \neq 1$  の場合の差分方程式を示す。

### (1) $d = 1$ の場合

まず双線形形式を求める。新しい変数  $f, g$  を導入し  $H(t)$  を以下のように変換する。

$$\log H = \frac{g}{f}. \quad (15)$$

この変換により式 (13) は

$$\frac{dg}{dt} f - g \frac{df}{dt} = e^A e^{-Bt} f^2, \quad (16)$$

となる。これが双線形の微分方程式である。式 (16) のように二つの関数  $f, g$  の比として表すと  $f, g$  の選び方はただ一つではなく、 $f$  と  $g$  を  $h$  倍しても値は変わらない。この  $h$  をゲージという。 $h$  を変えても値が変わらないことをゲージ不変性 (Gauge Invariance) という。このゲージ不変性を双線形の微分方程式に対して要請する。このゲージ不変性を利用すると厳密解を求めたり、差分化を行うのが容易になる。ここで、双線型方程式のゲージ不変性を確かめる。式 (16) をゲージ変換

$$f(t) \rightarrow f(t)h(t), \quad (17)$$

$$g(t) \rightarrow f(t)h(t), \quad (18)$$

すると

$$\left( \frac{dg}{dt} h + g \frac{dh}{dt} \right) f h - g h \left( \frac{df}{dt} h + f \frac{dh}{dt} \right) = e^A e^{-Bt} (f h)^2, \quad (19)$$

となり、両辺を  $h^2$  で割ると

$$\frac{dg}{dt} f - g \frac{df}{dt} = e^A e^{-Bt} f^2, \quad (20)$$

となる。これは式 (16) と同じである。つまり、式 (16) はゲージ不変性をみたしている。

さて、式 (16) を次式に変形すると、

$$\left[ \frac{dg}{dt} - e^A e^{-Bt} f \right] f - g \frac{df}{dt} = 0, \quad (21)$$

となるので、次の二つの線形微分方程式に分離される。

$$\frac{df}{dt} = 0, \quad (22)$$

$$\frac{dg}{dt} = e^A e^{-Bt} f, \quad (23)$$

式 (13) の解は  $H(0) = H_0$  とすると、式 (22), (23) により、

$$H(t) = H_0 \exp \left[ \frac{e^A}{B} (1 - e^{-Bt}) \right], \quad (24)$$

となる。これは、式 (2) のゴンベルツ曲線と同じになる。式 (22), (23) の左辺の導関数を前進差分に置き換える。次の二つの線形差分方程式を得る。

$$f(t + \delta) - f(t) = 0, \quad (25)$$

$$g(t + \delta) - g(t) = \delta e^A e^{-Bt} f(t + \delta). \quad (26)$$

$t = n\delta$  と置き,  $f(t), g(t)$  をそれぞれ  $f_n, g_n$  と表すと, 解はそれぞれ

$$f_n = f_0, \quad (27)$$

$$g_n = g_0 + \delta e^A f_0 \sum_{k=0}^{n-1} e^{-Bk\delta}, \quad (28)$$

と表される. よって厳密解  $H_n(n = 0, 1, 2, \dots)$  は,

$$H_n = H_0 \exp \left[ \delta e^A \frac{1 - e^{-Bn\delta}}{1 - e^{-B\delta}} \right] \quad (29)$$

となる. 式 (29) は  $n \rightarrow \infty, \delta \rightarrow 0$  のとき, 式 (24) に一致する.

## (2) $d \neq 1$ の場合

まず双線形形式を求める. 新しい変数  $f, g$  とパラメータ  $\alpha$  を導入し  $H(t)$  を変換する.

$$H = \left( \frac{g}{f} \right)^\alpha. \quad (30)$$

パラメータ  $\alpha$  は式 (13) が双線形に変換されるように後で決定する. この変換により式 (12) は

$$\alpha \frac{\frac{dg}{dt} f - g \frac{df}{dt}}{f^2} = e^A e^{-Bt} \left( \frac{g}{f} \right)^{\alpha(d-1)+1}, \quad (31)$$

となる. ここでパラメータ  $\alpha$  を  $\alpha = -\frac{1}{d-1}$  と選ぶと, 式 (31) は双線形形式

$$\frac{dg}{dt} f - g \frac{df}{dt} = -(d-1)e^A e^{-Bt} f^2, \quad (32)$$

となる. この双線形形式のゲージ不変性を確かめる. ゲージ変換

$$f(t) \rightarrow f(t)h(t), \quad (33)$$

$$g(t) \rightarrow g(t)h(t), \quad (34)$$

により,

$$\left( \frac{dg}{dt} h + g \frac{dh}{dt} \right) f h - g h \left( \frac{df}{dt} h + f \frac{dh}{dt} \right) = -(d-1)e^A e^{-Bt} (f h)^2 \quad (35)$$

となり, 両辺を  $h^2$  で割ると

$$\frac{dg}{dt} f - g \frac{df}{dt} = -(d-1)e^A e^{-Bt} f^2, \quad (36)$$

となる. これは式 (32) と同じである. 式 (32) はゲージ不変性をみたしている.

式 (32) を次式に変形すると, 双線形微分方程式

$$\left[ \frac{dg}{dt} + (d-1)e^A e^{-Bt} f \right] f - g \frac{df}{dt} = 0, \quad (37)$$

となるので, 次の二つの線形微分方程式に分離される.

$$\frac{df}{dt} = 0, \quad (38)$$

$$\frac{dg}{dt} = -(d-1)e^A e^{-Bt} f, \quad (39)$$

式 (13) の解は  $H(0) = H_0$  とすると, 式 (38), (39) により,

$$H(t) = \left( \frac{H_0^{(d-1)}}{1 - \frac{(d-1)e^A}{B}(1 - e^{-Bt})H_0^{(d-1)}} \right)^{\frac{1}{d-1}}, \quad (40)$$

となる. 式 (38), (39) の左辺の導関数を前進差分に置き換える. 次の二つの線形差分方程式を得る.

$$f(t+\delta) - f(t) = 0, \quad (41)$$

$$g(t+\delta) - g(t) = -(d-1)\delta e^A e^{-Bt} f(t+\delta), \quad (42)$$

$t = n\delta$  と置き,  $f(t), g(t)$  をそれぞれ  $f_n, g_n$  と表すと, 解はそれぞれ

$$f_n = f_0, \quad (43)$$

$$g_n = g_0 - (d-1)\delta e^A f_0 \sum_{k=0}^{n-1} e^{-Bk\delta}, \quad (44)$$

と表される. よって厳密解  $H_n(n = 0, 1, 2, \dots)$  は,

$$H_n = \frac{1}{\left( \frac{1}{H_0^{(d-1)}} - (d-1)\delta e^A \frac{1 - e^{-Bn\delta}}{1 - e^{-B\delta}} \right)^{\frac{1}{d-1}}}, \quad (45)$$

となる. 式 (45) は  $n \rightarrow \infty, \delta \rightarrow 0$  のとき, 式 (40) に一致する.

以下では,  $d = 0, d = 2$  の場合にそれぞれ指数形成長曲線モデル, ロジスティック成長曲線モデルとなることをそれぞれ示す.

### a) $d = 0$ の場合

式 (45) に  $d = 0$  を代入すると, 厳密解  $H_n$  は,

$$H_n = H_0 + \delta e^A \frac{1 - e^{-Bn\delta}}{1 - e^{-B\delta}} \quad (46)$$

となる. 式 (46) は  $n \rightarrow \infty, \delta \rightarrow 0$  のとき, 式 (14) に一致する.

### b) $d = 2$ の場合

式 (45) に  $d = 2$  を代入すると, 厳密解  $H_n$  は

$$H_n = \frac{H_0}{1 - \delta e^A \frac{1 - e^{-Bn\delta}}{1 - e^{-B\delta}} H_0}, \quad (47)$$

となる. 式 (47) は  $n \rightarrow \infty, \delta \rightarrow 0$  のとき, 連続型のロジスティック曲線に一致する.

以上のように, パラメータ  $d$  の導入によってゴンペルツ曲線とロジスティック曲線を統一的に表現することができる連続関数  $H(t)$  と, それが満たす微分方程式を作った上で, 双線形化法により差分化し, 更にそれを解くことで厳密解をもつ離散型一般化成長曲線モデルを構築することができた. 既往の研究ではゴンペルツ曲線とロジスティック曲線がそれぞれ個別に満たす微分方程式を出発点として, それぞれの離散型方程式を構築していたが, 本研究の方法は, それぞれ両成長曲線が満たす1つの微分方程式を用いたところに成果があると言える. このことは, 1つの関数が満たす微分方程式は無数に存在するという事実の1つの証左でもある.

### (3) 変曲点

連続のモデルの場合, 接線の傾きが増加から減少 (または減少から増加) に変わった点に変曲点である. 一般には, 二階導関数が0になる点を求めることで, 変曲点が求まる. 一方, 離散のモデルの場合は接線の傾きを求めるためには二階差分を行わなければならないので, 連続のモデルのように簡単に変曲点を求めることができない. ただし, 本研究で構築したモデルは飽和型モデルであるため, 変曲点は接線の傾きが増加から減少に変わった点のみなので, 前進差分と後退差分の大きさが変わる点  $j$  が変曲点とすると, 以下の2式を満たす.

$$\left. \begin{aligned} \hat{H}_{j+1} + \hat{H}_{j-1} &< 2\hat{H}_j \\ \hat{H}_j + \hat{H}_{j-2} &> 2\hat{H}_{j-1} \end{aligned} \right\}. \quad (48)$$

#### 4. パラメータ推定

$n$  組のデータ  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  について,  $X_i$  と  $Y_i$  の間に以下の, 一次の関数関係があるとする.

$$Y_i = C + DX_i, (i = 1, 2, \dots, n). \quad (49)$$

この係数  $C, D$  の値を  $n$  組のデータ  $(X_i, Y_i)$  から最小二乗法により, 推定するため次のように考える. 推定量  $\hat{C}, \hat{D}$  はそれぞれ,

$$\hat{C} = \frac{\sum_{k=1}^n Y_k - \hat{D} \sum_{k=1}^n X_k}{n}, \quad (50)$$

$$\hat{D} = \frac{\sum_{k=1}^n X_k Y_k - (\sum_{k=1}^n X_k)(\sum_{k=1}^n Y_k)/n}{\sum_{k=1}^n X_k^2 - (\sum_{k=1}^n X_k)^2/n}, \quad (51)$$

のように表される. 予測を行うために, 3章で述べた差分方程式から得られる離散の回帰式を用いる. ここで,  $\delta = 1$  とする.

##### (1) $d < 1$ の場合

式 (45) を前進差分することによって, 式 (52) を得る.

$$\log\left(\frac{1}{H_n^{(d-1)}} - \frac{1}{H_{n-1}^{(d-1)}}\right) = \log(1-d) + A - B(n-1). \quad (52)$$

ここで,

$$Y_n = \log\left(\frac{1}{H_n^{(d-1)}} - \frac{1}{H_{n-1}^{(d-1)}}\right), \quad (53)$$

$$X_n = n-1, \quad (54)$$

$$C = \log(1-d) + A, \quad (55)$$

$$D = -B, \quad (56)$$

である.

##### (2) $d = 1$ の場合

式 (29) を前進差分することによって, 式 (57) を得る.

$$\log(\log H_n - \log H_{n-1}) = A - B(n-1). \quad (57)$$

ここで,

$$Y_n = \log(\log H_n - \log H_{n-1}), \quad (58)$$

$$X_n = n-1, \quad (59)$$

$$C = A, \quad (60)$$

$$D = -B, \quad (61)$$

である.

##### (3) $d > 1$ の場合

式 (45) を後退差分することによって, 式 (62) を得る.

$$\log\left(\frac{1}{H_{n-1}^{(d-1)}} - \frac{1}{H_n^{(d-1)}}\right) = \log(d-1) + A - B(n-1). \quad (62)$$

ここで,

$$Y_n = \log\left(\frac{1}{H_{n-1}^{(d-1)}} - \frac{1}{H_n^{(d-1)}}\right), \quad (63)$$

$$X_n = n-1, \quad (64)$$

$$C = \log(d-1) + A, \quad (65)$$

$$D = -B, \quad (66)$$

である.

#### 5. ブートストラップ法に基づく成長曲線の 95%信頼区間の導出

実測データのバラつきを評価するためのブートストラップ標本を得る方法について述べる. 単回帰の式 (49) はこの方法において簡潔な手段を提供することができ, 以下にそのアルゴリズムを示す.

**Step1** はじめの未知パラメータ  $\hat{C}, \hat{D}$  を推定する.

**Step2** 回帰直線から  $\omega_i$  (本研究では等分散であると仮定する) を以下に従って求める.

$$\omega_i = \hat{Y}_i - (\hat{C}_0 + \hat{D}_0 X_i), (i = 1, 2, \dots, n). \quad (67)$$

**Step3** 繰り返しの回数を与える  $K$  を決め, 繰り返し番号を表す  $p$  を  $p = 1$  とする ( $p = 1, 2, \dots, K$ ).

**Step4**  $n$  個の残差  $\omega_n$  から等確率でかつ重複を許して,  $\omega_*$  の値を  $n$  回取り出し, 以下の式によって  $Y_{p_i}$  を求める.

$$Y_{p_i} = \omega_* + (\hat{C}_0 + \hat{D}_0 X_i), (i = 1, 2, \dots, n). \quad (68)$$

これにより新しいブートストラップ標本が以下のように得られる.

$$\{Y_{p_1}, Y_{p_2}, \dots, Y_{p_n}\}. \quad (69)$$

**Step5** 得られた  $n$  組のブートストラップ標本を用いて以下の回帰式に基づいて再び回帰分析を行い, 新しい回帰パラメータである  $C_p, D_p$  の推定値をそれぞれ算出し,  $\hat{C}_p, \hat{D}_p$  とする.

$$Y_{p_i} = C_p + D_p X_i, (i = 1, 2, \dots, n). \quad (70)$$

**Step6**  $p = p + 1$  とし,  $p < K$  ならば Step4 へ戻る.

**Step7** 終了

以上の処理を行うことによって,  $K$  組の推定されたパラメータ  $\hat{C}_p, \hat{D}_p$  を求めることができる.

##### (1) $H_n$ の 95%信頼区間の導出

$(\hat{C}, \hat{D})$  が下記のように 2 変数の正規分布に従うと仮定する.

$$\begin{pmatrix} \hat{C} \\ \hat{D} \end{pmatrix} \sim \text{BN}(\mu, \Sigma),$$

このとき,  $\mu$  及び  $\Sigma$  は下記のように求めることができる.

$$\Sigma = \begin{pmatrix} \text{Var}[\hat{C}] & \text{Cov}[\hat{C}, \hat{D}] \\ \text{Cov}[\hat{C}, \hat{D}] & \text{Var}[\hat{D}] \end{pmatrix}, \quad (71)$$

上記の式を用いて,  $H_n$  の 95%信頼区間は下記で求まる.

$$\hat{H}_n \pm 1.96 \sqrt{\text{Var}[\hat{H}_n]}. \quad (72)$$

このとき,  $\text{Var}[\hat{H}_n]$  は,

$$\text{Var}[\hat{H}_n] = \left( \frac{\partial H_n}{\partial C}, \frac{\partial H_n}{\partial D} \right) \times \Sigma \times \begin{pmatrix} \frac{\partial H_n}{\partial C} \\ \frac{\partial H_n}{\partial D} \end{pmatrix} \quad (73)$$

$(C \rightarrow \hat{C}, D \rightarrow \hat{D}),$

となり, 95% 信頼区間を導出することができる [9].

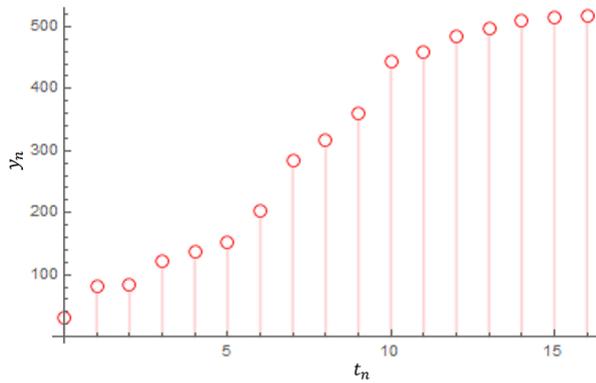


図1 実測データ  $(t_n, y_n)$  の振る舞い.

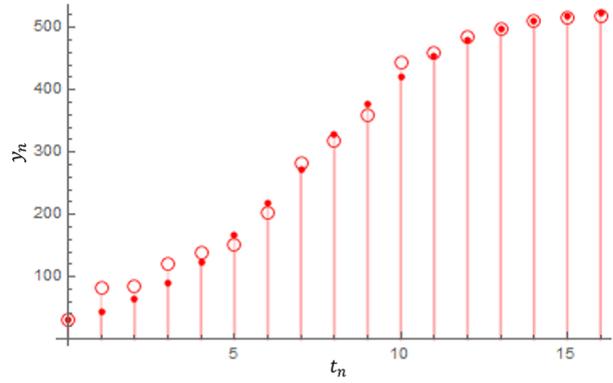


図2 推定値  $\hat{H}_t$  と実測データの比較.

表1 成長曲線の当てはめ分析によるパラメータの推定結果と誤差の二乗和.

	$\hat{A}$	$\hat{B}$	誤差の二乗和(対数値)
$d = 0$	3.616	0.013	9.861
$d = 1$	-0.662	0.186	9.054
$d = 2$	-4.595	0.404	8.433
$\hat{d} = 2.113$	-5.021	0.430	8.427

表2 パラメータ推定結果と分散分析表.

	$\hat{A}$	$\hat{B}$	$\hat{d}$		
	-7.138	0.445	2.439		
	DF	SSE	V	$F_0$	$R^2$
Model	1	67.523	67.523	74.726**	0.842
Error	14	12.65	0.903		
Total	15	80.173			

## 6. 数値例

成長曲線の一般化を行う以上、既往のモデルより予測精度が上回っている必要がある。本章では、実際に観測されたデータセットを用いて、予測精度の比較を行う。本研究では、既往の指数形曲線モデル、ゴンペルツ曲線モデルとロジスティック曲線モデルを用いて予測精度の比較を行った。成長曲線の当てはめ分析 [7] を行う際、以下の式を用いてパラメータ推定を行う。

$$\min_{d,A,B} \sum_{i=0}^n (H_{t_i} - y_i)^2. \quad (74)$$

さらに、構築したモデルの変曲点とブートストラップ法により導出した 95 % 信頼区間を求め、点推定と区間推定の視点から考察を行い、予測するために必要な日数を検証する。

### (1) 既往の成長曲線モデルとの比較

本節で扱うデータセット [10] は  $(t_n, y_n) (n = 0, 1, 2, \dots, 16)$  の形式で与えられており、 $t_n$  は時間 (日) で測られ、 $y_n$  は  $t_n$  までに発見されたソフトウェアフォールト数の累積値を使用する。まず、図 1 に  $(t_n, y_n)$  の振る舞いを示す。

#### a) 成長曲線の当てはめ分析

式 (74) により求められたパラメータ推定の結果と誤差の二乗和 (対数値) を表 1 に示す。表 1 により、構築したモデルの方が当てはまりの良いことが見て取れる。表 1 の  $d$  を推定した場合の値を構築したモデルに適用した結果と実測データを当てはめた結果を図 2 に示す。一方、AIC [11] について、 $d = 2$  の場合は 99.21、 $d$  を推定した場合は 101.92 となり、ほぼ同じ適合度である結果となった。

#### b) $H_n$ の 95% 信頼区間の導出

実測されたデータセット [10]  $(t_n, y_n) (n = 0, 1, 2, \dots, 16)$  に対して、4 章で述べた回帰式の導出方法を適用する。ここで、回帰式のパラメータ推定の結果と分散分析表の結果を表 2、回帰直線を図 3 に示す。

次に、ブートストラップ法により  $\hat{C}, \hat{D}$  を  $K = 1000$  個する。ここで、 $\hat{C} = \log(\hat{d} - 1) + \hat{A}$ 、 $\hat{D} = -\hat{B}$  であり、2 変数の正規分布と仮定する。それを踏まえ、信頼区間の導出を行う。式 (72) により信頼区間を導出することができ、式 (71) より  $\hat{\Sigma}$  は以下ようになった。

$$\hat{\Sigma} = \begin{pmatrix} 0.0214818 & 0.0196794 \\ 0.0196794 & 0.0023543 \end{pmatrix}, \quad (75)$$

式 (75) の値を式 (73) に適用すると図 4 のように、 $\hat{H}_n$  の 95 % 信頼区間を出力することができる。

式 (48) より変曲点は 8 日目であることがわかった。12 日目までのデータで逐次推定し、13 日目から 16 日目までの将来予測を行った結果、図 5 から見て取れるように、95% 信頼区間から外れ、変曲点から少し超えなければ正確な将来予測はできない結果となった。

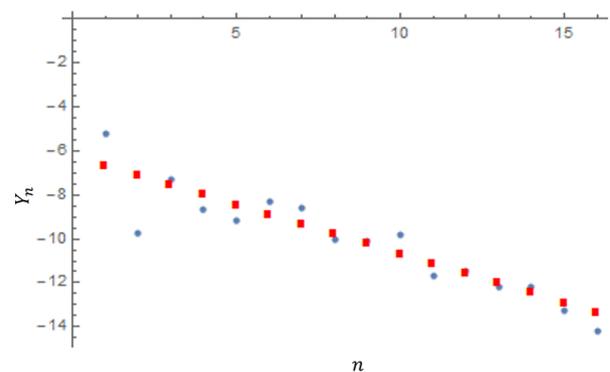


図3 回帰直線の推定結果.

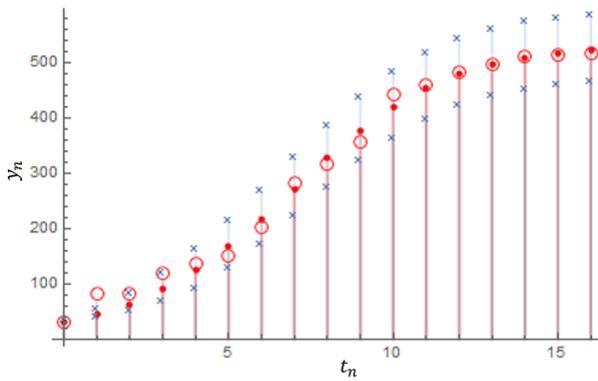


図4  $\hat{H}_n$  の95%信頼区間.

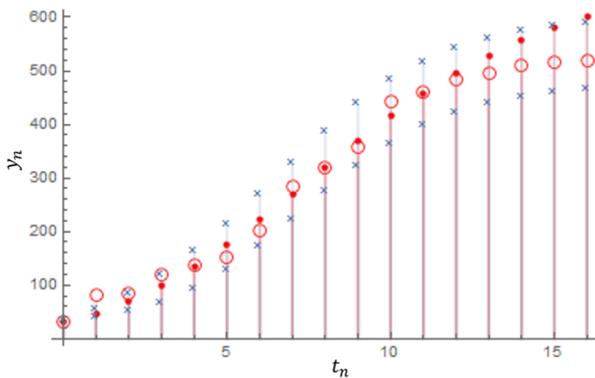


図5 12日目までのデータで逐次推定した場合の  $\hat{H}_n$  の95%信頼区間.

## (2) 考察

3つの既往のモデルと構築したモデルを用いて予測精度の比較を行った。比較を行った結果、実測値と予測値の誤差は構築したモデルの方が小さくなり、精度が良いことができた。

ブートストラップ法を用いて構築したモデルの95%信頼区間を求めた。その結果、変曲点までのデータだけでは正確な将来予測はできない結果となった。実際に将来予測をするために必要なデータ数は8割程度を用いなければならないと考えられる。

## 7. おわりに

本研究では、代表的な成長曲線モデルとして、ゴンペルツ曲線モデルとロジスティック曲線モデルを取り上げ、様々な

成長曲線モデルを表すことができる、離散型成長曲線モデルを構築した。また、一般化で得られるパラメータに対してブートストラップ法を用いることで信頼区間を導出し、予測するために必要な日数について議論した。成長曲線の当てはめ分析において、既往のモデルより実測値と予測値の誤差が小さいことを示せたことから、成長曲線モデルを一般化したことは有用であったと考えられる。

## 参考文献

- [1] 佐藤大輔, 「データからのゴンペルツ曲線とロジスティック曲線の判別」, 電子情報通信学会技術研究報告(信頼性), R2016-5, pp. 27-32, 2016.
- [2] D. Satoh, "A Discrete Gompertz Equation and a Software Reliability Growth Model", *IEICE Trans. & Syst.*, Vol. E83-D, No. 7, pp. 1508-1513, 2000.
- [3] 井上真二, 山田茂, 佐藤大輔, 「差分方程式に基づくソフトウェア信頼度成長モデルに関する考察」, 日本オペレーションズリサーチ学会秋季研究発表会予稿集, 1-D-4, pp. 66-67, 2001.
- [4] 佐藤大輔, 可積分差分による予測, 数理科学9月号, サイエンス社, pp. 9-14, 2003.
- [5] 羽生雅也, 「ブートストラップに基づくソフトウェア信頼性評価モデルに関する研究」, 法政大学大学院工学研究科修士論文, 2013.
- [6] M. Morishita, "The fitting of the logistic equation to the rate of increase of population density", *Res. Popul. Ecol.*, V II, pp. 52-55, 1965.
- [7] M. Kimura, "A Linearized Growth Curve Model for Software Reliability Data Analysis", Hoang Pham (Ed.), *Recent Advances in Reliability and Quality in Design*, Springer, pp. 275-290, 2008.
- [8] 広田良吾, 差分方程式講義一連続より離散へ, サイエンス社, 2000.
- [9] M. Kimura, A Study on Bootstrap Confidence Intervals of Software Reliability Measures Based on an Incomplete Gamma Function Model, *Advanced Reliability Modeling II Reliability Testing and Improvement*, pp. 419-426, 2006.
- [10] 芝田寛二, 「ソフトウェア製品の生産計画と工程管理」, 情報処理, Vol 21 No.10, pp. 1035-1042, 1980.
- [11] 鈴木義一郎, 情報基準による統計解析入門, 講談社, 1995.