ロジスティック曲線モデルとゴンペルツ曲線モデルの判別法

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

成長曲線とは，人口の増加を記述した曲線であり，生態学の分野で提案されたものである．この曲線は，はじめゆっくりと増加するが途中から急激に増加し最終的にはある極限値に近づいていくもので，S字曲線を描く．この成長曲線は，生態学以外にも，様々な分野で顔を出す．農学，生命科学，マーケティング，ソフトウェア信頼性，コンピュータウィルスの感染数など多岐にわたっている．

あるデータが，どの成長曲線に適合するかは，経験則からある程度わかっているものもあるが，一般的にはほぼ飽和値まで観測してからでないと判定しづらい．より早期に適合するモデルを判定する指標が望まれる．ここでは，より早期に正確なパラメータ推定が可能である離散モデル [3, 4] を対象とし，成長曲線のなかでも，様々な分野で使われているロジスティック曲線モデルとゴンペルツ曲線モデルを対象とする．この2つのいずれのモデルがふさわしいかをより早期に判定する指標を提案する．

2 logistic 差分方程式

厳密解を持つ logistic 差分方程式 [2] は，

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

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

である．差分方程式 (1) から得られる回帰式は，

$$Y_n = A+BL_{n+1} \quad (3)$$

となる．ここで，Y_n は式 (4) で表される．

$$Y_n = \frac{L_{n+1}}{L_n} \quad (4)$$

3 Gompertz 差分方程式

厳密解を持つゴンペルツ差分方程式 [3] は，既に得られているが，ここで，新に別の差分方程式 (5) を提案する．

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

厳密解は，

$$G_n = ka^{(1-\delta\log b)^{-n}} \quad (6)$$

となる．ここで，$k>0, 0<a<1, 0<b<1$ である．

既に提案されているゴンペルツ差分方程式 [3] と同様に，式 (5) はどんな時間間隔 δ に対しても式 (7) を満たす．

$$G_n \to k \quad (n\to\infty) \quad (7)$$

式 (5) から回帰式を導くと

$$Y_n = A+B\log G_{n+1} \quad (8)$$

となる．ここで，

$$Y_n = \log G_{n+1}-\log G_n \quad (9)$$

である．式 (8) を使うことによって，パラメータ k, a, b は以下のように推定される．

$$\hat{k} = \exp\left(-\frac{\hat{A}}{\hat{B}}\right) \quad (10)$$

$$\hat{a} = \exp\left(\frac{\sum_{n=1}^{N}\log\frac{G_n}{\hat{k}}}{\sum_{n=1}^{N}(1-\delta\log\hat{b})^{-n}}\right) \quad (11)$$

$$\hat{b} = \exp\left(\frac{\hat{B}}{\delta}\right) \quad (12)$$

ここで，$\hat{a}$, $\hat{b}$, $\hat{k}$ はそれぞれ a, b, k の推定値であり，$\hat{A}$，$\hat{B}$ は式 (8) における A，B の推定値である．

文献 [3] におけるゴンペルツ差分方程式と同様に各パラメータ推定値は，時間間隔 δ の値に関わらず常に同じ推定値を与える．

4　提案指標

ロジスティック曲線モデルとゴンペルツ曲線モデルのうち，どちらがデータに適したモデルかを選択する指標として (i)x_{n+1} と x_{n+1}/x_n もしくは，(ii)$\log x_{n+1}$ と $\log(x_{n+1}/x_n)$ を 2 変数とする相関係数を提案する．ここで，x_n は，n ステップにおけるデータ値である．(i) の相関係数と (ii) の相関係数を比較し，(i) の相関係数の方が (ii) のそれより小さい値を示したとき，データは，ロジスティック曲線モデルにより適合しているとみなし，それ以外の場合は，ゴンペルツ曲線モデルにより適合しているとみなす．

次に提案指標の検証を行う．データが厳密解を満足する場合には，提案指標は式 (3), (8) より明らかに正しい判定をする．ここでは実データを用いて，提案指標の検証を行う．ゴンペルツ曲線モデルに適合するデータとして文献 [1] のデータを使用し，ロジスティック曲線モデルに適合するデータとして文献 [4] で使用したデータを用いた．結果を図 1, 2 に示す．各図ともに使用したデータ数に対する (i), (ii) の相関係数の値を示している．

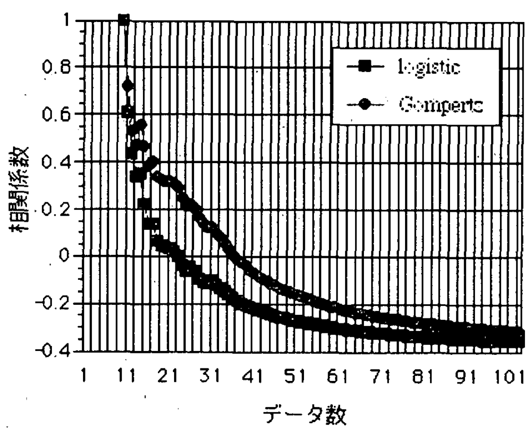

図 1: 文献 [4] で使用したデータの場合

5　まとめ

本論では，厳密解を持つ新しい Gompertz 差分方程式を提案し，データがゴンペルツ曲線モデル，ロジスティック曲線モデルのどちらに適合するかを判別する指標を提案した．この指標は，データ

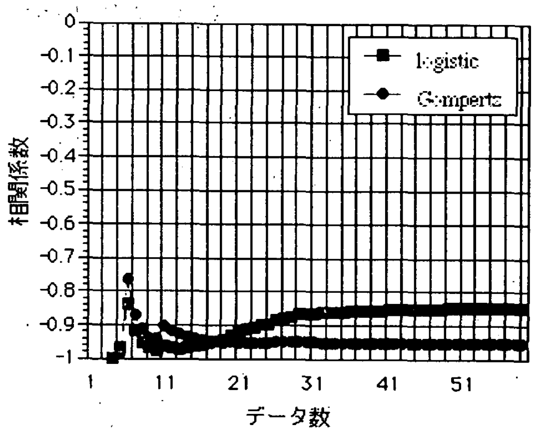

図 2: 文献 [1] のデータを使用したデータの場合

が厳密解を満足する場合には，正しい判定結果を与えることは明らかであり，実データに対しても，データ数が少ない段階から正確な判定をしている．

参考文献

[1] 三觜: ソフトウェアの品質評価法, 日科技連, 1981.

[2] F. Morishita, The fitting of the logistic equation to the rate of increase of population density. *Res. Popul. Ecol.*, **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.