

# ロジスティック成長モデルのパラメータ推定

常 盤 洋 一

## 目 次

1. はじめに
2. ロジスティック成長モデルの導出
3. 既存のパラメータ推定法
4. 最小2乗法と逆関数による2段階推定法
5. テストデータによる検証
6. まとめ

## 1. はじめに

将来の人口などを予測するために、過去から現在までのデータの推移に直線や曲線を当てはめ、それを将来に延長するのがトレンドモデルである。その中でも、ロジスティック成長モデルは19世紀中ごろに Verhulst によって見出され、その後それとは無関係に、Pearl と Reed によって1920年に再び発見されて地域人口の予測によく用いられている。さらに、それ以外でも土地利用変化や新商品の普及率などの予測などにも用いられ、ポピュラーな存在になっている。ところが、現実のデータへの当てはめとなると、まだ簡易で統計学的な裏付けのはっきりしたパラメータ推定法が確立されていない。

問題は非線形関数であることと、1変数であるのにパラメータが3つある点にある。もちろんニュートン法などのように、コンピュータによる反復解法を使えば推定は可能であるし、それ以外に、これまでもいくつもの方法が考案されてきた。しかしパーソナル・コンピュータと表計算ソフトが普及した現在、初心者でも複雑なプログラムを組まずにパラメータの推定ができないか考えた。その研究結果をまとめたのが本論文である。ここで紹介する方法を用いれば、表計算ソフト(本研究では Excel)の回帰分析機能と、加

減乗除や対数計算だけでパラメータが推定でき、統計学的な裏付けもしっかりして精度も高い。

## 2. ロジスティック成長モデルの導出

ロジスティック成長モデルを述べる前に、その基となる指数成長モデルと修正指数成長モデルについて述べ、その後、ロジスティック成長モデルを導出する。

$t$  時点の人口を  $P_t$  とする。 $P_t$  は時間  $t$  の関数である。ここで人口の増加率が、一定値  $a$  である場合を考える。

$$\frac{P_{t+1} - P_t}{P_t} = a \quad (2.1)$$

このときの  $a$  を内的自然増加率という。 $a$  が 0 より大きければ人口は増加し、0 より小さければ減少する。この式を微分方程式で表すと

$$\frac{dP_t}{dt} = aP_t \quad (2.2)$$

となる。これを解くと

$$\begin{aligned} \frac{1}{P_t} dP_t &= a dt & \int \frac{1}{P_t} dP_t &= \int a dt & \ln P_t &= at + C \\ P_t &= e^{at+C} = e^C e^{at} \end{aligned} \quad (2.3)$$

となる。ゼロ時点の人口を  $P_0$  とすると

$$P_t = P_0 e^{at} \quad (2.4)$$

となる。これが Malthus が提示した指数成長モデルである (Pielou, 1969, p.8) (第 2-1 図)。大都市圏の近郊で人口が急増している場合や、過疎地で人口が逡減している場合に当てはまる。

次に地区内の市街地の拡大が進み、人口が飽和しかけている状態を考えてみよう。人口がこれ以上増加できない値、すなわち飽和値を  $P_\infty$  とする。モデルを微分方程式で表すと

$$\frac{dP_t}{dt} = a(P_\infty - P_t) \quad (2.5)$$

となる。すなわち人口  $P_t$  が飽和値  $P_\infty$  に近づくと、増加量が逡減していくと

するものである。これを解くと

$$\frac{1}{P_{\infty}-P_t}dP_t = a dt \quad \int \frac{1}{P_{\infty}-P_t}dP_t = \int a dt$$

$$-\ln(P_{\infty}-P_t) = at + C \quad P_{\infty}-P_t = e^{-at-C} = e^{-C}e^{-at} \quad (2.6)$$

となる。ゼロ時点の人口は  $P_0$  としたので

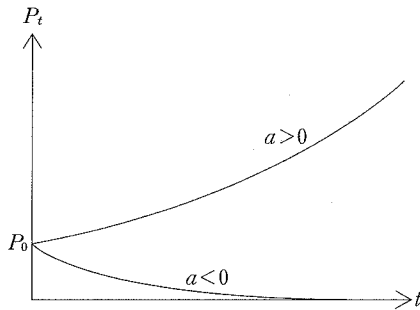
$$P_{\infty}-P_0 = e^{-C} \quad \therefore e^{-C} = P_{\infty}-P_0 \quad (2.7)$$

となる。よって

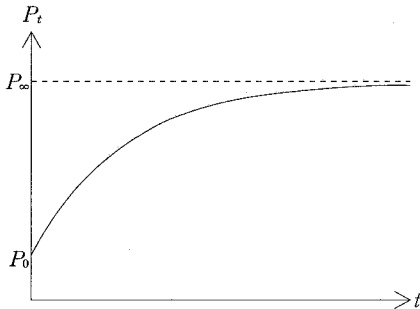
$$P_{\infty}-P_t = (P_{\infty}-P_0)e^{-at}$$

$$P_t = P_{\infty} - (P_{\infty}-P_0)e^{-at} \quad (2.8)$$

となる。これが修正指数成長モデルである (天野, 1982, p.15) (第 2-2



第 2-1 図 指数成長モデルの例



第 2-2 図 修正指数成長モデルの例

図)。

指数成長モデルは、急増する人口に対応し、修正指数成長モデルは、人口が飽和しかけている状態に対応している。これら2つのモデルを合体させたのが、先に述べた Verhulst-Pearl のロジスティック成長モデルである (Pielou, 1969, p.19; 鈴木, 1985, pp.47~51)。微分方程式で書くと次のようになる。

$$\frac{dP_t}{dt} = aP_t \left(1 - \frac{P_t}{P_\infty}\right) \quad (2.9)$$

はじめは少しずつ増加し、その後急増する。やがて直線に近づいた後、伸びが鈍化し、最後は  $P_t = P_\infty$  を漸近線として飽和に近づく。

式 (2.9) を解く。変形して

$$\begin{aligned} \frac{1}{P_t \left(1 - \frac{P_t}{P_\infty}\right)} dP_t &= a dt \\ \frac{P_\infty}{P_t(P_\infty - P_t)} dP_t &= a dt \end{aligned} \quad (2.10)$$

とする。ここで左辺を部分分数に分解する。

$$\frac{A}{P_t} + \frac{B}{P_\infty - P_t} = \frac{P_\infty}{P_t(P_\infty - P_t)} \quad (2.11)$$

とおくと、左辺は

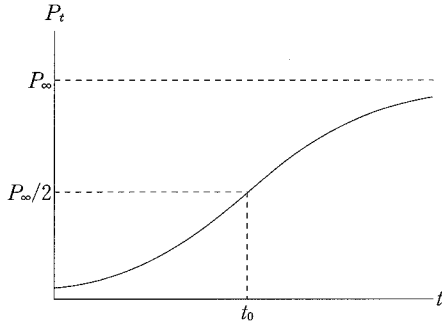
$$\frac{AP_\infty - AP_t + BP_t}{P_t(P_\infty - P_t)} = \frac{(B-A)P_t + AP_\infty}{P_t(P_\infty - P_t)} \quad (2.12)$$

となる。式 (2.11) は恒等式なので

$$B = A \quad A = 1 \quad \therefore B = 1 \quad (2.13)$$

と求まる。よって式 (2.10) は

$$\begin{aligned} \frac{1}{P_t} dP_t + \frac{1}{P_\infty - P_t} dP_t &= a dt \\ \int \frac{1}{P_t} dP_t + \int \frac{1}{P_\infty - P_t} dP_t &= \int a dt \\ \ln P_t - \ln(P_\infty - P_t) &= at + C \\ \ln\left(\frac{P_t}{P_\infty - P_t}\right) &= at + C \end{aligned}$$



第 2 - 3 図 ロジスティック成長モデルの例

$$\begin{aligned} \frac{P_t}{P_\infty - P_t} &= e^{at+c} \\ P_t &= P_\infty e^{at+c} - P_t e^{at+c} \\ (1 + e^{at+c})P_t &= P_\infty e^{at+c} \\ P_t &= \frac{P_\infty e^{at+c}}{1 + e^{at+c}} = \frac{P_\infty}{1 + e^{-at-c}} \end{aligned} \quad (2.14)$$

となる。

ロジスティック成長モデルでは、初期値  $P_0$  の代わりに中間時点  $t_0$  を用いる。すなわち  $t=t_0$  のときに  $P_t=P_\infty/2$  であるとする。したがって

$$\frac{P_\infty}{2} = \frac{P_\infty}{1 + e^{-at_0-c}} \quad \therefore -at_0 - C = 0 \quad C = -at_0 \quad (2.15)$$

となる。すると式 (2.14) は

$$P_t = \frac{P_\infty}{1 + e^{-a(t-t_0)}} \quad (2.16)$$

となる。これがロジスティック成長モデルである (第 2 - 3 図)。

### 3. 既存のパラメータ推定法

ロジスティック成長モデルのパラメータを推定するには、これまで多くの方法が提案されてきた (鈴木, 1985, pp.203~211)。しかし筆者が見る限り、それらは観測値  $P_t$  が飽和値  $P_\infty$  にかなり近づいた状況まで得られることを

前提としていて、中間時点  $t_0$  前後までしか観測値がない場合を想定していない。また、パラメータの推定方法も統計学的な裏付けがはっきりしていない。そこでそれらの方法を紹介するとともに、各方法の問題点を述べることにする。

### 3-1. ホテリングの方法

はじめにホテリングによる推定方法を述べる (Davis, 1941, pp.251~255; Hoteling, 1927; 森田, 1955, pp.20~25)。式 (2.9) を離散型の式にすれば、次のようになる。

$$P_{t+1} - P_t = aP_t \left(1 - \frac{P_t}{P_\infty}\right) \quad (3.1)$$

左辺を増加率とすると

$$\frac{P_{t+1} - P_t}{P_t} = a - \frac{a}{P_\infty} P_t \quad (3.2)$$

となるので

$$a \Rightarrow A \quad -\frac{a}{P_\infty} \Rightarrow B$$

と置き換えると

$$\frac{P_{t+1} - P_t}{P_t} = A + BP_t \quad (3.3)$$

となる。ここで

$$\begin{cases} a = A \\ P_\infty = -a/B \end{cases} \quad (3.4)$$

である。

次に

$$m = \left(\frac{P_\infty}{P_t} - 1\right) e^{at} = e^{at_0} \quad (3.5)$$

とおいて、 $m$  を求め  $t_0$  を推定する。ただし観測値  $P_t$  は、必ずしも曲線上にある値といえないため、ある観測値 ( $P_t, t$ ) を用いて得た  $m$  と、その他の観測値を用いて得た  $m$  とは必ずしも一致しない。したがって、いくつかの  $P_t$  を用いて2個以上の  $m$  を見いだし、それらの  $m$  の算術平均  $\bar{m}$  を  $m$  の値とする

(森田, 1955, pp.20~25)。

ここで、増加率を  $R_t$  とおく。

$$R_t = \frac{P_{t+1} - P_t}{P_t} \quad (3.6)$$

そして

$$\begin{cases} \sum_{t=0}^n R_t = nA + B \sum_{t=0}^n P_t \\ \sum_{t=0}^n R_t P_t = A \sum_{t=0}^n P_t + B \sum_{t=0}^n P_t^2 \end{cases} \quad (3.7)$$

とおいて、この連立方程式から  $A$  と  $B$  を求め、式(3.4)によって  $a$  と  $P_\infty$  を求める。最後に、式 (3.5) を用いておのおのの  $t$  について  $m_t$  を計算する。

$$m_t = \left( \frac{P_\infty}{P_t} - 1 \right) e^{at} \quad (3.8)$$

そして

$$\bar{m} = \sum_{t=0}^n m_t \quad (3.9)$$

によって平均値  $\bar{m}$  を求め、式 (3.5) によって

$$\begin{aligned} \ln m &= at_0 \\ t_0 &= \frac{\ln m}{a} \end{aligned} \quad (3.10)$$

と  $t_0$  を求める。

この方法では、 $A$  と  $B$  の式 (3.7) による求め方には疑問が残る。2つの式は、単に1次式と2次式にただけである。2つの連立方程式は、本当に独立であろうか。また、 $\bar{m}$  が最適値であるかについても、統計学的な裏付けについて立証がなされていない。

### 3-2. 3群法 (ユールの第1法)

ユール (Eells, 1925) は、ロジスティック成長モデルのパラメータを推定するために、3種の方法を述べている。それを順に紹介する (中川, 1942)。

まず最初の方法である3群法は、 $n$  時点からなる時系列を3等分し、 $m$  時点からなる3組の時系列を作る。

$$\begin{cases} P_1, P_2, \dots, P_m \\ P_{m+1}, P_{m+2}, \dots, P_{2m} \\ P_{2m+1}, P_{2m+2}, \dots, P_{3m} \end{cases}$$

次に、各観測値の逆数について

$$S_1 = \sum_{t=1}^m \frac{1}{P_t} \quad S_2 = \sum_{t=m+1}^{2m} \frac{1}{P_t} \quad S_3 = \sum_{t=2m+1}^{3m} \frac{1}{P_t} \quad (3.11)$$

で定義される  $S_1, S_2, S_3$  を計算する。ここで式 (2.16) を

$$\frac{1}{P_t} = a + bc^t \quad (3.12)$$

と置き換える。すると

$$\begin{cases} S_1 = ma + b \frac{c(c^m - 1)}{c - 1} \\ S_2 = ma + bc^m \frac{c(c^m - 1)}{c - 1} \\ S_3 = ma + bc^{2m} \frac{c(c^m - 1)}{c - 1} \end{cases} \quad (3.13)$$

となる。ここで

$$\begin{cases} S_2 - S_1 = b(c^m - 1) \frac{c(c^m - 1)}{c - 1} \\ S_3 - S_2 = bc^m(c^m - 1) \frac{c(c^m - 1)}{c - 1} \end{cases} \quad (3.14)$$

を求める。ここで

$$\begin{cases} S_2 - S_1 = d_1 \\ S_3 - S_2 = d_2 \end{cases} \quad (3.15)$$

とすれば

$$d_2/d_1 = c^m \quad (3.16)$$

となるので

$$c = \sqrt[m]{d_2/d_1} \quad (3.17)$$

と求めることができる。

次に式 (3.14) から

$$b = d_1 \frac{c - 1}{c(c^m - 1)^2} \quad (3.18)$$



を得る。最後に、式 (3.13) の第 1 番目の式と式 (3.18) から

$$a = \frac{1}{m} \left( S_1 - \frac{d_1}{c^m - 1} \right) \quad (3.19)$$

を得る (館, 1960, pp.314~317; 錦織, 1960, pp.207~210)。

現実の問題としては、数多くの時点で観測値が得られることは少なく、3 分割すると、それぞれの時点ではわずかな観測値しかなくなる。また、観測値が飽和値に近いところまであればよいのであるが、中間時点  $t_0$  付近までの観測値しかない場合、この方法では無理がある。

### 3-3. 選点法 (ユールの第 2 法)

選点法は 3 点法とも言われている。式 (2.16) を変形して

$$P_t = \frac{P_\infty}{1 + e^{(\beta-t)/\alpha}} \quad (3.20)$$

として  $\alpha$ ,  $\beta$ ,  $P_\infty$  を推定する。

時系列の観測値  $P_1, P_2, \dots, P_n$  が与えられたとき、最初の時点の  $t$  を「0」、最後の時点をも「2」、そして、その中央の時点をも「1」として、その各時点の値を  $P(0)$ ,  $P(1)$ ,  $P(2)$  とする。このとき

$$\begin{cases} \frac{1}{P(0)} = \frac{1}{P_\infty} (1 + e^{\beta/\alpha}) \\ \frac{1}{P(1)} = \frac{1}{P_\infty} (1 + e^{(\beta-1)/\alpha}) \\ \frac{1}{P(2)} = \frac{1}{P_\infty} (1 + e^{(\beta-2)/\alpha}) \end{cases} \quad (3.21)$$

が得られる。ここで

$$\begin{cases} d_1 = \frac{1}{P(0)} - \frac{1}{P(1)} \\ d_2 = \frac{1}{P(1)} - \frac{1}{P(2)} \end{cases} \quad (3.22)$$

とすれば、式 (3.21) と式 (3.22) から

$$e^{1/\alpha} = \frac{d_1}{d_2} \quad (3.23)$$

となる。したがって

$$\alpha = \frac{1}{\ln d_1 - \ln d_2} \quad (3.24)$$

と求められ

$$\frac{1}{P_\infty} e^{\beta/\alpha} = \frac{d_1^2}{d_1 - d_2} \quad (3.25)$$

であるので

$$\beta = \alpha \left\{ \ln P_\infty + \ln \left( \frac{d_1^2}{d_1 - d_2} \right) \right\} \quad (3.26)$$

と変形できる。さらに

$$\frac{1}{P(0)} = \frac{1}{P_\infty} + \frac{1}{P_\infty} e^{\beta/\alpha} = \frac{1}{P_\infty} + \frac{d_1^2}{d_1 - d_2} \quad (3.27)$$

となるので

$$P_\infty = \frac{1}{\frac{1}{P(0)} - \frac{d_1^2}{d_1 - d_2}} \quad (3.28)$$

と求まり、この  $P_\infty$  を式 (3.26) に代入することによって  $\beta$  が求まる。

数式は単純であるが、わずか3つの観測値しか用いておらず、すべての観測値から得られる情報は使われていない。したがって、これがすべての観測値を代表するパラメータであるとはいえない。

### 3-4. ユールの第3法

ユールが第3に挙げた方法、すなわちユールの第3法も式 (3.20) から  $\alpha$ ,  $\beta$ ,  $P_\infty$  を推定する方法である。

時間  $t$  から時間  $t+1$  までの  $P_t$  の増加率  $R_t$  は

$$R_t = \frac{P_{t+1} - P_t}{P_t} = \frac{e^{(\beta-t-1)/\alpha}}{1 + e^{(\beta-t-1)/\alpha}} (e^{1/\alpha} - 1) \quad (3.29)$$

によって示される。一方、式 (3.20) から  $P_{t+1}$  は

$$P_{t+1} = \frac{P_\infty}{1 + e^{(\beta-t-1)/\alpha}} \quad (3.30)$$

によって示されるから

$$R_t = \left( 1 - \frac{P_{t+1}}{P_\infty} \right) (e^{1/\alpha} - 1) \quad (3.32)$$

が得られる。

ここで

$$R_t \Rightarrow Y \quad e^{1/\alpha} - 1 \Rightarrow A \quad -\frac{e^{1/\alpha} - 1}{P_\infty} \Rightarrow B \quad P_{t+1} \Rightarrow X$$

と置き換えると

$$Y = A + BX \quad (3.33)$$

となる。 $Y$  と  $X$  は観測値から求まるので、単回帰分析で回帰係数  $A$  と  $B$  を求めることができる。すると

$$\begin{cases} \alpha = -\ln(A+1) \\ P_\infty = -A/B \end{cases} \quad (3.34)$$

と求まる。

$\beta$  を求めるには、等比級数の性質を用いる。時点  $t=0, 1, \dots, n$  で  $P_t$  が観測されているとすると

$$\sum_{t=0}^{n-1} e^{-t/\alpha} = \frac{1 - e^{-n/\alpha}}{1 - e^{-1/\alpha}} \quad (3.35)$$

であるので

$$\sum_{t=0}^{n-1} \frac{1}{R_t} = \frac{n}{P_\infty} + \frac{C}{P_\infty} e^{\beta/\alpha} \quad (3.36)$$

となる。ただし

$$C = \frac{1 - e^{-n/\alpha}}{1 - e^{-1/\alpha}} \quad (3.37)$$

である。式 (3.34) で得たパラメータ  $\alpha$  と  $P_\infty$  の値、および式 (3.36) を用いれば  $\beta$  が求まる。

観測値  $P_t$  が、誤差項がなく曲線上に並んでいれば、この方法でよい。しかし、実際には誤差項が存在するため等比級数になっているという保証はなく、 $\beta$  が最適値であるかどうかは疑問が残る。しかも計算が複雑すぎる。

#### 4. 最小2乗法と逆関数による2段階推定法

前章で、既存のパラメータの推定方法を紹介し、それぞれの問題点を指摘した。統計学的に有効と思われるのは、残差平方和を最小にする最小2乗法であると考えられる。それを適用するには、まず  $\alpha$  と  $P_\infty$  を単回帰分析によって推

定する。簡単なのは、ホテリングの方法の式 (3.3) で単回帰分析を行うことであろう。しかし実際に実験してみると、真の値に対して大きな誤差が出た。これはユールの第3法にもいえることである。そのわけは、式 (2.9) を離散型の式にするとときに問題があるからである。

$P_t$  における傾きは  $P_{t+1}-P_t$  ではない。真に傾きが  $P_{t+1}-P_t$  となるのは、ロールの定理により  $t$  と  $t+1$  の間にある。しかし式 (2.9) は微分方程式である。そこで右辺の  $P_t$  を、 $P_{t+1}$  と  $P_t$  の中間値  $(P_{t+1}+P_t)/2$  に置き換えて離散型の式を導出した。すなわち

$$P_{t+1}-P_t = a \frac{P_{t+1}+P_t}{2} \left( 1 - \frac{1}{P_\infty} \cdot \frac{P_{t+1}+P_t}{2} \right) \quad (4.1)$$

とし、これを变形して

$$\frac{2(P_{t+1}-P_t)}{P_{t+1}+P_t} = a - \frac{a}{P_\infty} \cdot \frac{P_{t+1}+P_t}{2} \quad (4.2)$$

とする。ここで

$$\frac{2(P_{t+1}-P_t)}{P_{t+1}+P_t} \Rightarrow Y \quad a \Rightarrow A \quad -\frac{a}{P_\infty} \Rightarrow B \quad \frac{P_{t+1}+P_t}{2} \Rightarrow X$$

と置き換えると

$$Y = A + BX \quad (4.3)$$

となる。 $P_{t+1}$  と  $P_t$  は観測値であるので、単回帰分析で回帰係数  $A$  と  $B$  を求めることができる。すると

$$\begin{cases} a = A \\ P_\infty = -a/B \end{cases} \quad (4.4)$$

と求まる。

ここで、求めたパラメータ  $a$  と  $P_\infty$  の推定区間を得るため、それぞれの標準誤差を求める (廣松・常盤・他, 1988, pp.143~148)。観測値から求めた  $X$  と  $Y$  を、それぞれ  $X_t$ ,  $Y_t$ , ( $t=0, 1, \dots, n$ ) とする。そして  $Y_t$  の推定値を  $\hat{Y}_t$  とする。このとき観測値と推定値の差である残差  $e_t$  は

$$e_t = Y_t - \hat{Y}_t \quad (4.5)$$

となる。すると、残差から求まる標準誤差  $s$  は

$$s = \sqrt{\frac{1}{n-1} \sum_{t=0}^n e_t^2} \quad (4.6)$$

となる。

これを用いると、回帰係数  $A, B$  の分散  $Var(A)$  と  $Var(B)$  は

$$VAR(A) = s^2 \left\{ \frac{1}{n+1} + \frac{\bar{X}^2}{\sum_{t=0}^n (X_t - \bar{X})^2} \right\} \quad (4.7)$$

$$VAR(B) = \frac{s^2}{\sum_{t=0}^n (X_t - \bar{X})^2} \quad (4.8)$$

となる。なお  $\bar{X}$  は  $X_t$  の平均値である。このとき  $A, B$  の標準誤差  $SE(A)$  と  $SE(B)$  は

$$SE(A) = \sqrt{VAR(A)} \quad SE(B) = \sqrt{VAR(B)} \quad (4.9)$$

と定義される。ここで、それぞれの誤差の分布は正規分布であると仮定されているので、有意水準を  $\alpha$  としたときの推定区間は正規分布の確率分布関数から得ることができる。これらは Excel の回帰分析で得られる。そこで  $A$  と  $B$  の推定区間を

$$A_{\min} < A < A_{\max} \quad B_{\min} < B < B_{\max} \quad (4.10)$$

とする。

これらより  $a$  の推定区間は

$$A_{\min} < a < A_{\max} \quad (4.11)$$

となり、 $P_\infty$  の推定区間は

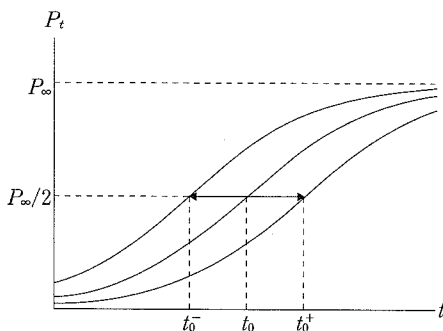
$$-\frac{A_{\min}}{B_{\max}} < P_\infty < -\frac{A_{\max}}{B_{\min}} \quad (4.12)$$

となる。

次に  $t_0$  は、第 4-1 図のようにロジスティック曲線を左右に動かすパラメータであり、式 (2.9) に現れていないことから  $a$  や  $P_\infty$  と独立している。従来、最小 2 乗法では目的変数の観測値と推定値の差の 2 乗和を最小化するという手法がとられていた。しかし  $t_0$  については、逆関数にして、説明変数の観測値と推定値の差の 2 乗和を最小にするという手法をとる。

そこで式 (2.14) の途中から  $t$  についての逆関数を求めると

$$\ln\left(\frac{P_t}{P_\infty - P_t}\right) = at - at_0$$



第4-1図  $t_0$  によるロジスティック曲線の左右の移動

$$t = \frac{1}{a} \ln \left( \frac{P_t}{P_\infty - P_t} \right) + t_0 \quad (4.13)$$

となる。この式に誤差項  $u_t$  を含めて

$$t = \frac{1}{a} \ln \left( \frac{P_t}{P_\infty - P_t} \right) + t_0 + u_t \quad (4.14)$$

とする。すると誤差項の2乗和  $Q$  は、時点  $t$  を  $t=0, 1, 2, \dots, n$  として

$$Q = \sum_{t=0}^n u_t^2 = \sum_{t=0}^n \left\{ t - \frac{1}{a} \ln \left( \frac{P_t}{P_\infty - P_t} \right) - t_0 \right\}^2 \quad (4.15)$$

となる。 $Q$  は  $t_0$  の関数なので、これを  $Q(t_0)$  と表すことにする。

$Q(t_0)$  は単純な1変数の2次関数なので、下に凸な2次曲線になり、最小となる点では傾きがゼロとなる。すなわち

$$\frac{d}{dt_0} Q(t_0) = 0 \quad (4.16)$$

である。この微分を解くと

$$-2 \sum_{t=0}^n \left\{ t - \frac{1}{a} \ln \left( \frac{P_t}{P_\infty - P_t} \right) - t_0 \right\} = 0 \quad (4.17)$$

となるので、式 (4.17) を  $t_0$  について解く。

$$\sum_{t=0}^n t - \frac{1}{a} \sum_{t=0}^n \ln \left( \frac{P_t}{P_\infty - P_t} \right) - (n+1)t_0 = 0$$

$$t_0 = \frac{1}{n+1} \left\{ \frac{1}{2} n(n+1) - \frac{1}{a} \sum_{t=0}^n \ln \left( \frac{P_t}{P_\infty - P_t} \right) \right\}$$

$$= \frac{n}{2} - \frac{1}{a(n+1)} \sum_{t=0}^n \ln \left( \frac{P_t}{P_\infty - P_t} \right) \quad (4.18)$$

これにより  $t_0$  が求まる。

## 5. テストデータによる検証

第4章で示した方法が、どの程度の精度を有するかを検証する。検証方法は、あらかじめパラメータを定めたロジスティック成長曲線に、乱数によって誤差項を加え、そのテストデータでパラメータの推定を行い吟味する。

はじめに中心極限定理を使って、標準正規分布に従う乱数、標準正規乱数を Excel で作成する（廣松・常盤・他, 1988, pp.109~112）。中心極限定理は、期待値  $\mu$ 、分散  $\sigma^2$  の任意の分布から抽出された大きさ  $n$  の標本  $(X_1, X_2, \dots, X_n)$  があるとき、

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (5.1)$$

で示される平均  $\bar{X}$  について

$$\frac{(\bar{X} - \mu)}{\sigma/\sqrt{n}} \quad (5.2)$$

を  $n \rightarrow \infty$  としたときの極限分布が、期待値 0、分散 1 の標準正規分布  $N(0, 1)$  となる定理である。

そこで、100個の標準正規乱数を作るため、 $n$  を 20 として RAND 関数を使って 100 行 20 列の  $(0, 1)$  一様乱数  $\hat{x}_{ij} (i=1, 2, \dots, 100; j=1, 2, \dots, 20)$  を作る。この一様乱数の期待値  $\mu$  は  $1/2$ 、分散  $\sigma^2$  は  $1/12$  となる。すると標準正規乱数  $x_i (i=1, 2, \dots, 100)$  は

$$x_i \sqrt{12 \cdot 20} \left( \frac{1}{20} \sum_{j=1}^{20} \hat{x}_{ij} - \frac{1}{2} \right) \quad (5.3)$$

として得ることができる。

この乱数を使ってテストデータを作るのであるが、ロジスティック成長モデルは非負の値をとり、指数関数を主体としたモデルである。そこでテストデータ  $P_t^*$  は、各時点において変動係数（標準偏差／平均） $c$  を一定とした条件で、乱数を使って与える形式で作成する。

$t$  時点における真の値を  $P_t$  とし、乱数による  $t$  時点の誤差項を  $u_t$  とする。 $u_t$  は期待値 0、分散  $\sigma_t^2$  の正規分布に従うとする。するとテストデータの値  $P_t^*$  は

$$P_t^* = P_t + u_t \quad (5.4)$$

となる。 $P_t^*$  の期待値を  $\mu_t^*$ 、分散を  $\sigma_t^{*2}$  とすると、変動係数  $c$  は標準偏差を  $\sigma_t^*$  として

$$c = \sigma_t^* / \mu_t^* \quad (5.5)$$

となる。この  $c$  を一定の値に設定して  $P_t^*$  を求める。

仮定から  $\mu_t^*$  は  $P_t$  に等しく、また  $P_t^*$  の標準偏差  $\sigma_t^*$  は  $u_t$  の標準偏差  $\sigma_t$  に等しいので、式 (5.5) は

$$\sigma_t = cP_t \quad (5.6)$$

となる。一方、先に作成した標準正規乱数から取り出した値を  $x_t$  とすると、正規分布の性質から

$$x_t = \frac{P_t^* - P_t}{\sigma_t} \quad (5.7)$$

という関係が成立する。したがって式 (5.6) と式 (5.7) から

$$\begin{aligned} x_t(cP_t) &= P_t^* - P_t \\ P_t^* &= P_t + cx_tP_t \end{aligned} \quad (5.8)$$

となる。式 (5.8) の右辺第 2 項が  $u_t$  である。これにより、テストデータ  $P_t^*$  を各時点で作成することができる。

ロジスティック成長モデルを用いる目的のうちで重要なのは、現在までの観測値から、飽和値  $P_\infty$  を推定することである。そこで、テストデータは  $t=0$  から  $t=(2/3)t_0$  までのものと、 $t=0$  から  $t=(4/3)t_0$  までのものの 2 つの期間で作成する。 $t_0=9$  とすると、それぞれの期間は (0,6) と (0,12) となる。ただし、単回帰分析の目的変数が増加率であるため、時点数はそれより 1 つ多く、それぞれ 8 時点と 14 時点とした。一方、変動係数は 1% と 5% の 2 つとした。10% まで上げてみたが、石村のいうカオス状態(石村・Richert, 2002, pp.200~203)に近くなり、ロジスティック曲線との乖離が著しいため 5% までにとどめた。作成したテストデータと、Excel による推定結果を第 5-1 表に示す。



第5-1表 テストデータと推定結果

時点	真の値	変動係数			
		1%		5%	
		設定時点		設定時点	
		0~6	0~12	0~6	0~12
		テストデータ			
		I	II	III	IV
0	6.30	6.18	6.33	6.43	6.99
1	8.32	8.19	8.19	8.35	8.26
2	10.91	10.80	10.71	11.13	10.85
3	14.19	14.22	14.36	12.62	13.55
4	18.24	18.35	18.50	19.02	16.84
5	23.15	23.21	23.49	22.75	25.63
6	28.91	28.75	29.52	26.71	28.59
7	35.43	35.51	34.98	35.24	37.38
8	42.56		42.24		40.92
9	50.00		50.90		54.20
10	57.44		58.16		60.19
11	64.57		64.51		63.17
12	71.09		70.11		77.58
13	76.85		76.15		72.67
重相関係数		0.9760	0.9713	0.1015	0.5636
内的自然増加率	真の値	0.3	...	...	...
	推定値	0.3049	0.3005	0.2609	0.2858
	誤差率	1.6%	0.2%	-13.0%	-4.7%
飽和値	真の値	100	...	...	...
	推定値	95.79	98.16	232.60	98.19
	誤差率	-4.2%	-1.8%	132.6%	-1.8%
中間時点	真の値	9	...	...	...
	推定値	8.533	9.175	12.787	9.154
	誤差率	-5.2%	1.9%	42.1%	1.7%

重相関係数は、回帰式 (4.3) について求めたものである。当然であるが、変動係数が1%のときは大きく、5%のときは小さい。とくにテストデータIIIでは非常に小さくなっていて、単回帰分析に耐えられない値である。これらについて内的自然増加率  $a$ 、飽和値  $P_{\infty}$ 、中間時点  $t_0$  をそれぞれ推定した。

表の誤差率は

$$\text{誤差率} = \frac{\text{推定値} - \text{真の値}}{\text{真の値}} \quad (5.9)$$

として計算したものである。

変動係数1%では非常に誤差が小さい。設定時点が0～6でも、真の値に極めて近い推定値を得ることができた。したがって、従来のように飽和に近づいた状況までのデータではなく、まだ増加の度合いが大きく、増加率がやや鈍った段階でも、この方法では各パラメータを推定することができるといえる。一方、変動係数5%では、テストデータⅢで誤差が非常に大きい。これは重相関係数が0.1しかないためで、致し方ないところである。しかしテストデータⅣでは、重相関係数が0.56しかないのに真の値に極めて近い推定値を得ている。このことは、重相関係数が統計的に有意であれば、たとえその値が小さくとも良い推定値が得られることを意味している。

## 6. まとめ

ロジスティック成長モデルのパラメータは、本稿で紹介したように、これまでにもいくつかの推定方法があった。しかしその論理を検討すると、それぞれ何らかの問題点を抱えており、推定するためのデータも飽和値まで近い点まで必要であった。本稿では、統計学的な裏付けがはっきりしていて、中間時点前後までしかデータがない状態でも可能なパラメータの推定方法を提示した。使ったソフトはExcelのみであり、VBAも用いていない。 $t=0 \sim (2/3)$ までの時点で、確率的な誤差項を含めたテストデータによる検証でも、高い精度で推定できることが証明された。ロジスティック成長モデルの発見から150年で、ようやく息を吹き込んだ気がする。この手法が多くの方々によって用いられれば幸いである。

## 参考文献

- 天野光三編 (1982) : 『計量都市計画－都市計画システムの手法と応用－』丸善。  
石村貞夫・Richert, S. (2002) : 『Excelでやさしく学ぶ時系列』東京図書。

- 鈴木啓祐 (1985) : 『人口分布の構造解析』 大明堂.
- 館 稔 (1960) : 『形式人口学』 古今書院.
- 中川友長 (1942) : 人口のロジスティック曲線について. 人口問題研究, vol.3, no.3, pp.1~14.
- 錦織理一郎 (1960) : 『統計学総論』 政文堂.
- 廣松 毅・田中明彦・常盤洋一・小暮 睦 (1988) : 『ロータス 1-2-3 による統計入門』 朝倉書店.
- 森田優三 (1955) : 『経済変動の統計分析法』 岩波書店.
- Davis, T.H.(1941) : The Analysis of Economic time Series. *The Cowles Commission for Research in Economics Monograph* No.6. Bloomington, Indiana, The Principia Press.
- Eells, W.C.(1925) : The Center of Population—A Prophecy and its Fulfilment. *The Scientific Monthly* 20, January 1925, pp.78~84.
- Hoteling, H.(1927) : Differential Equations Subject to Error and Population Estimates. *Journal of the American Statistical Association*, vol.22, pp.283~314p.
- Pielou, E.C.(1969) : *An Introduction to Mathematical Ecology*. John Wiley & Sons, Inc. (邦訳) 南雲仁一監訳・合田周平・藤村貞夫訳 (1974) : 『数理生態学』 産業図書.