普及率の動的変化を記述する階層ロジスティックモデル

田代徹, 皆川紘恵, 千葉路子

お茶の水女子大学理学部物理学科

概要

流行がいかに社会の中で広まるかについて，普及率の動的変化を通じて考える．今までの普及率に関する研究として，ロジスティック方程式を用いたものがある．しかし，どのような人と人とのやりとりの結果に，ロジスティック方程式で記述される流行の伝搬があるのかは，必ずしも自明ではない．そこで本研究ではまずこのコミュニケーションを明らかにする．そしてその不自然さを指摘し，何人の流行を取り入れている人に出会ったかという記憶が重要な要素であると考え，この記憶を取り入れることで拡張したロジスティック方程式，階層ロジスティック方程式を構築する．そしてこのモデルの応用例として，iPod の売り上げデータを採用し，ロジスティックモデルよりも精度よく実データを再現できることを示す.

Hierarchical logistic model to describe the dynamical behavior of penetration rates

Tohru Tashiro, Hiroe Minagawa and Michiko Chiba

Department of Physics, Ochanomizu University

Abstract

We study how fashion diffuses in our society through dynamical behavior of a penetration rate. The logistic equation has been used in order to describe the behavior. However, what kind of human communication results in such a behavior? The answer to this question is not self-evident. Therefore, we clarify the communication, which turns out to be not natural. We think a memory, how many people who already possess it a person who does not process it yet met, is a more important factor, and then we construct the hierarchical logistic equation by incorporating it into the logistic equation. As an application, we apply this model to iPod sales data, and find that this model can approximate the data much better than the logistic equation.

1　はじめに

流行とは社会の中でどのように広まっていくのだろうか．社会の構成員である私たちが意図しない形で流行は広まっていく．ごく少数の人間がひょっとしたら流行を起こそうと意図しているかもしれないが，社会を構成す大部分の人間は流行を広げようとは考えていない．出現した当初は真新しいものでも，いつの間にか私たちの身の回りに当たり前の存在になっていたりする．時にして私たちの生活から消え失せてしまう場合もある．

我々は，流行の発生を生み出す人間同士の相互作用（コミュニケーション）とは何なのかを探りたい．科学の問題としてがこの社会現象をモデリングする上では，客観的な流行の指標となる（定量的な）デー

タが無くてはならない．ここでは，普及率を用いることにする．これは次のような普遍性を有している．一般に普及率は，初期の段階では非常にゆっくりと成長し，その後成長率は最大になり最後には成長は緩和する．すなわち俗にいうS字曲線を描く．この動的変化を解析するために，ロジスティック方程式の解，ロジスティック関数が今まで使われてきた．

よく知られているように，ロジスティック方程式はVerhulstによって，人口増加を記述する方程式として提案された．技術革新の普及にロジスティック曲線を導入したのはGrilichesである[1]．その後数多くの技術の広がりにロジスティック方程式は使われてきている．

つまりたくさんの普及率の動的変化はロジスティック方程式で説明することができるのである．では，この背後にはどういった人と人とのコミュニケーションがあるのだろうか．この問いは決して自明なものではない．そこで，本論文では先ずそのコミュニケーションを明らかにする．我々の研究によると，ロジスティック方程式を用いたモデル（ロジスティックモデル）は次のようなコミュニケーションを前提としていることが明らかになった:「あるモノ[1]を所持している（受け入れている）人にまだそれを所持していない（受け入れいていない）人が出会うと，すぐに所持しだす（受け入れる)」．つまり人々がまねをしていくことでそのモノとは社会の中で普及していくと，ロジスティックモデルでは考えられている．

これは確かに流行が伝搬していく様子を的確にとらえていると言える．しかしここで新たな疑問が浮かんでくる:「我々はそんなに簡単に人から影響をうけるだろうか？」．この問いが本論文の真の出発点となる．そこで我々はより自然なコミュニケーション，今まで何人そのモノを所持していた人に会ったかという記憶，をとり入れたモデルを考案する．

さらにはこの新しいモデルを実際のデータに適用する．本論文ではiPodの売り上げデータを採用することにする．その結果従来のモデル（ロジスティックモデル）よりも我々のモデルの方が良く実データを再現することを示す.

2　まねする集団での普及率の変化

この節では，ロジスティック曲線で記述される普及率の動的変化が，どのような人と人との相互作用によってもたらされるかを明らかにする．

N人の集団を考えよう．この集団に対して次のルールを定める: i) 初めに，これから流行するであろうモノを何人かがもっている ii) まだそのモノをもっていない人 (non-adopter）がすでにもっている人 (adopter) に出会うとすぐにそれを所持しだす iii) non-adopterは複数のadopterに影響を受けないし，adopterは複数のnon-adopterに影響を与えない iv) adopterは所持したモノを手放さない.

この様な集団の挙動をより定量的に記述するために，$n \times n$ $(> N)$ 個の正方格子上に配置しよう．そこで人々は離散的な時間間隔（Δt）の間に隣の格子に等確率で移動するものとする．iii) のルールは，adopterとnon-adopterが1つの格子をシェアするのは必ず1人づつとなることを要請している．すると，adopterもしくはnon-adopterに出会う確率は，それぞれの人口に比例すると考えるのは極々自然なことである．

ここで，i番目の時間ステップでのadopter，non-adopterの人口をそれぞれP_i，Q_iとしよう．その結果，i番目のステップでadopter，non-adopterに出会う確率はそれぞれP_i/n^2，Q_i/n^2となる．ゆえに $(P_i/n^2) \times Q_i$ 人のnon-adopterは次のステップでadopterになるので，次の漸化式が得られる：

$$P_{i+1} = P_i + \frac{P_i}{n^2} Q_i \ , \ Q_{i+1} = Q_i - \frac{P_i}{n^2} Q_i \ . \tag{1}$$

この漸化式は，Ref. [2] において，別の見方で導出されているが，ここでの導出方法の方が，前提としている人と人とのコミュニケーションを明確に理解することができる．

ここで次のような置き換えをおこなう：$P(t) = P(i \cdot \Delta t) \equiv P_i$, $Q(t) = Q(i \cdot \Delta t) \equiv Q_i$．そして時間と空間の連続化を $\Delta t \to 0$，$n \to \infty$ とすることで行う．ただしその際 $n^2 \Delta t$ を有限に保ち，この値を a/N と

[1]これは商品のような実在の物だけでなく，流行語のような実在しない人間の行動様式も含めることにする．

する．すると先の漸化式は以下の微分方程式になる．

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \frac{a}{N}\left\{N - P(t)\right\}P(t), \tag{2}$$

ただし $P(t)+Q(t)=N$ を使った．これはまさにロジスティック方程式である．普及率 $p(t)\equiv P(t)/N$ を使うのであれば以下の方程式になる．

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = a\left\{1 - p(t)\right\}p(t), \tag{3}$$

3　階層ロジスティック方程式

今まで普及率の動的変化に関する研究とは，adopter に non-adopter が出会うと影響を受けてまねをし，すぐにそのモノを持ち始める，というコミュニケーションを前提としていたことが，前節でわかった．ここで一つ疑問が生まれる：「我々はそんなに簡単に人から影響を受けるのだろうか」．これまでの研究では我々の行動の中で大事な要素が考えられていなかった．それは，今まで何人の adopter に出会ったかという記憶である．もちろんこれから流行するであろうモノが琴線に触れ，即座に手に入れる（まねをする）人もいるだろう．しかし，たとえそのモノにそれ程興味がなくとも，多くの所持している人に出会うことで，それが流行しつつあることを感じ取り，流行に乗り遅れまいとして[2]ついには所持しだす．この様な人々の行動も存在し，その蓄積の結果，普及率は変化すると私たちは考える．そこで先のまねするランダムウォーカーの集団に対して，non-adopter に内部構造（階層構造）を与える．つまり，残り μ 人に会ったらそのモノを所持しだす，i 番目のステップにおける non-adopter の人数を Q_i^μ とする．この μ を remaining adopters numbrer（RAN）と呼ぶことにする．それ以外のルール変えない．つまり non-adopter の種類は増えたが，彼ら同士の相互作用は考えない．すると，RAN が μ の人が adopter に出会えば，次のステップでは RAN は $\mu-1$ になる．RAN の最大が m の集団に対しては，漸化式はこう変わる．

$$P_{i+1} = P_i + \frac{P_i}{n^2}Q_i^1\ ,\ Q_{i+1}^1 = Q_i^1 - \frac{P_i}{n^2}Q_i^1 + \frac{P_i}{n^2}Q_i^2\ ,\ \ldots\ ,\ Q_{i+1}^m = Q_i^m - \frac{P_i}{n^2}Q_i^m \tag{4}$$

これは先と同様の空間と時間の連続化によって以下の拡張されたロジスティック方程式になる．

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \frac{a}{N}Q^1(t)P(t)\ ,\ \frac{\mathrm{d}Q^1(t)}{\mathrm{d}t} = -\frac{a}{N}P(t)Q^1(t) + \frac{a}{N}P(t)Q^2(t)\ ,\ \ldots\ ,\ \frac{\mathrm{d}Q^m(t)}{\mathrm{d}t} = -\frac{a}{N}P(t)Q^m(t) \tag{5}$$

ただし $Q^\mu(t) = Q^\mu(i\cdot\Delta t)\equiv Q_i^\mu$ とした．この方程式を階層ロジスティック方程式と呼ぶことにする．もちろん $P(t)+Q^1(t)+\cdots+Q^m(t)=N$ は常に満たされる．

Eq. (5) の最後の方程式は形式的に解くことが出来，その解は $Q^m(t) = Q^m(0)\exp\left[-a\int_0^t \mathrm{d}t' P(t')\right]$ となる．もし $Q^m(0)=0$ ならば，$Q^m(t)$ はつねに 0 となる．したがって Q^{m-1} に対する Q^m の寄与はなくなる．さらに Q^{m-1} も同様に解くことが出来るので，もし $Q^m(0)=Q^{m-1}(0)=\cdots=Q^2(0)=0$ であるならば，常に $Q^m(t)=Q^{m-1}(t)=\cdots=Q^2(t)=0$ が成り立つ．これはつまり階層ロジスティック方程式は通常のロジスティック方程式を包含していることを意味する．もし普及率など全人口に対する割合 $p(t)\equiv P(t)/N$，$q^\mu(t)\equiv Q^\mu(t)/N$ を使えば，階層ロジスティック方程式は

$$\frac{\mathrm{d}p(t)}{\mathrm{d}t} = aq^1(t)p(t)\quad \frac{\mathrm{d}q^1(t)}{\mathrm{d}t} = -ap(t)q^1(t) + ap(t)q^2(t)\ ,\ \cdots\ ,\ \frac{\mathrm{d}q^m(t)}{\mathrm{d}t} = -ap(t)q^m(t) \tag{6}$$

となる．

4　iPod の売り上げデータのフィッティング

それでは階層ロジスティック方程式を使って実際のデータのフィッティングに応用してみよう．ここでは iPod の売り上げデータを採用する．これは Apple 社のホームページ（http://www.apple.com/）から入手することが出来る．そこでは四半期ごとの売り上げが公表されている．我々は 2001 年 11 月から 2006 年 5 月までのデータを採用することとする[3]．

[2]あるいは，みんなが持っているからそのモノは信頼できるだろうという安心感が芽生えて．

[3]2012 年 11 月現在，2012 年 8 月までのデータが公開されている．2006 年以降，毎年 11 月の売り上げが突出しているが，これは明らかにクリスマスプレゼントとしての購入が反映されている．当然我々のモデルがこういった消費者の動向を記述できるわけは

表 1: SARE と SAE の積を最小にするパラメータ，そのパラメータでの SARE と SAE の積と SARE.

	$m=1$ (logistic)	$m=2$	$m=3$	$m=4$
a	1.43	4.42	4.25	4.17
N [$\times 10^6$]	1.28×10^5	64.6	65.4	66.2
$p(0)$	9.26×10^{-7}	0.00207	0.00200	0.00189
$q^1(0)$	1.00	0.288	0.311	0.325
$q^2(0)$	–	0.710	0.675	0.657
$q^3(0)$	–	–	0.0125	0.00619
$q^4(0)$	–	–	–	0.00760
product	62.5	11.0	8.91	8.42
SARE	1.84	1.12	1.02	1.01

2001 年 11 月を時間の原点とし，売り上げを累積したデータを階層ロジスティック方程式でフィッティングした．フィッティングを行う際，残差の絶対値の和，あるいは，残差の 2 乗の和を最小にしてしまうと，売り上げの大きいところでのフィッティングを優先してしまい，小さいところでのフィッティングをないがしろになってしまう．そこで我々は誤差（残差）の絶対値 の和（SAE）と相対誤差の絶対値の和（SARE）の積を最小にするパラメータを採用することにする．結果を表 1 に示した．

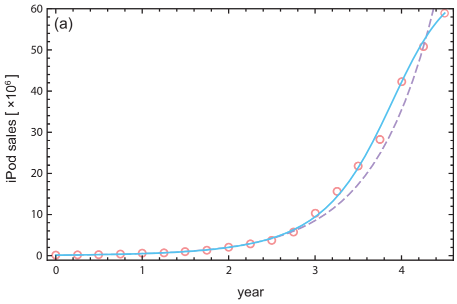

図 1: 2001 年 11 月を時間の原点とした iPod の累積売上げ（丸）とそれをフィッティングしたロジスティック関数（破線）と $m=4$ の階層ロジスティック方程式（実線）．

積と SARE は m の増加とともに減少している[4]．記憶を考慮した階層ロジスティックの方が，実データからのずれが小さくなったので，iPod を購入した全ての人たちは，直ぐに出会った adopter の真似をしたわけではない，と我々は推測する．$m=4$ での SARE は $m=1$ のときに比べておよそ半分近くになっている．言い換えれば，1 つのデータあたりの平均相対誤差は，$m=1$ では 9.7%だったものが，$m=4$ では 5.3%まで減少している．

$m=4$ の結果から，1 度目のピークを生み出した市場規模は約 6600 万人，影響を受けやすい人の割合（$q^1(0)$）は約 33%，慎重な人の割合（$q^2(0)$）は約 66%，更に慎重な人の割合（$q^3(0)+q^4(0)$）約 1%であることが予想される．比較のため，ロジスティック方程式と $m=4$ の階層ロジスティック方程式を図 1 に示した．

参考文献

[1] Griliches Z 1957 *Econometrica* **25** 501

[2] Mansfield E 1961 *Econometrica* **29** 741

ない．また周知の通り，発売以来 iPod は様々なバージョンアップを繰り返してきた．すると月日が経てば iPod を買い替える人も出てくるだろうが，この様な adopter の行動は我々のモデルには盛り込まれていない．そこで我々はこの様な効果がない（あるいは小さい）初めのおよそ 4 年半のデータを採用することにした．

[4]「フィッティングパラメータを増やしたのだから自明」という訳でもない．そもそも階層ロジスティック方程式の解は，ロジスティック関数を含んだあらゆる増加関数を表現していない．階層構造を持った結果，ロジスティック関数に比べて遅い成長の増加関数となる．従って（直ぐにまねをする集団の）早い成長の普及率は，ロジスティック関数の方がフィッティングの結果は良くなる．