

フィールド言語学にロジスティック回帰分析は寄与するか 横山詔一(国立国語研究所)・真田治子(埼玉学園大学)

キーワード：フィールド言語学，ロジスティック回帰分析，最尤推定法，学術用語の消長，外来語の増加過程

米国の高名な社会言語学者 Labov が打ち立てた変異理論では，多変量解析の手法としてロジスティック回帰分析がよく使われる。日本でも医学ではロジスティック回帰分析が多用されている。しかし，日本のフィールド言語学では「林の数量化理論Ⅲ類」が多変量解析の主流であり，ロジスティック回帰分析はあまり見かけない。その理由は，ロジスティック回帰分析に関する分かりやすい解説書が存在しないことにくわえて，日本のフィールド言語学における有効性を具体的に示した研究が数例しか存在しないからだと考えられる。本研究は，ロジスティック回帰分析の仕組みについて理解しやすい説明を行うとともに，フィールド言語学におけるデータ解析の実例を示した。

Application of logistic regression analysis to field linguistics

YOKOYAMA Shoichi*, & SANADA Haruko**

*The National Institute for Japanese Language, **Saitama Gakuen University

Key words: Socio-Linguistics, Logistic regression analysis, Maximum Likelihood Estimation

Logistic regression analysis is a common method of multivariate analysis in variation theory, which was established by Labov. However, it is not common in field linguistics in Japan. The purpose of the paper is to provide a general picture of logistic regression analysis, which plausibly contributes to field linguistics in Japanese. Logistic regression analysis is known as a useful method in various areas of studies and, for example, is commonly employed in medical research in Japan. However, it is not quite common in field linguistics in Japan, in which “Hayashi’s Quantification Theory 3” is quite popular. Unpopularity of logistic regression analysis in field linguistics in Japan may be attributable to the limited resources in Japanese. In particular, very few studies are available, showing its applicability to field linguistics in Japanese. An explicit path to application of logistic regression analysis should contribute to Japanese field linguistics, as found in medical research. The study demonstrates an example application of logistic regression analysis to field linguistics in Japanese, as well as it explains the essence of logistic regression analysis.

1. はじめに(横山)

フィールド言語学は「伊能忠敬」を連想させる要素を持っている。伊能忠敬は江戸時代に日本全国を歩き回って自分たちの歩幅をモノサシにして測量を行い，精度の高い日本地図を編纂した。フィールド言語学者も，いろいろな地域の人々の暮らしのなかに息づく言葉を収集・分析し，言語地図にまとめる。その仕事の光景が伊能忠敬のイメージと重なる。

フィールド言語学は，研究者が実際に野外を歩くかどうかはさておき，言語の変異・変化の実相を描き出すための学問である。本稿を嚆矢

とするこれからの一連の研究は，次の範囲を射程に収める。書き言葉については，学術用語の消長，外来語の増加過程，異体字選好の確率予測であり，話し言葉については，カナダ英語の語形交替を，順次取り上げていく。さらには医療場面における言語コミュニケーションの問題も含めて，種々の分野で観察される言語変化の記述と予測に「ロジスティック回帰分析(logistic regression analysis)」が有効であることを繰り返して説明していく。

ロジスティック回帰分析は日米の医学界できわめて盛んに利用されているが，社会言語学

の分野でも 1960 年代に米国の Labov [1] が言語データの解析に積極的に導入してきたという経緯がある。日本のフィールド言語学では「林の数量化理論Ⅲ類」が多変量解析の道具としてよく使われるが、松田 [2] によれば、Labov [1] が打ち立てた変異理論ではロジスティック回帰分析が主流だと言う。この重要な指摘に触発された横山 [3] は、次の感想を述べている。『1950 年代から国内では国立国語研究所による共通語化調査がフィールド言語学の典型的なパラダイムとして定着していた。これは日本の人文社会科学には珍しい国産パラダイムである。同じぐらいの時期に、統計数理研究所の林知己夫氏が独自の数量化理論を完成させた。これも統計学には希有な国産理論であった。当時の国立国語研究所と統計数理研究所は共同プロジェクトなどを通じて研究者の交流が盛んであり、フィールド言語学の国産パラダイムが統計学の国産理論と結びつくのは自然な成り行きであったと考えられる。この国産研究の成功は、海外の学界でも高く評価されている。』

しかし、成功の光は影も作る。フィールド言語学をさらに飛躍させ、とかく記述研究にとどまりがちな現状から脱皮するには、「林の数量化理論Ⅲ類」などのほかに、確率論に立脚した新たな分析法を手中におさめる必要がある。言語変化の動向を数量的に予測する研究に向けて一歩踏み出すために、「確率予測の道具立て」を積極的に取り入れる時期が到来したのではないか。そのような研究環境整備が、記述研究や探索研究にくわえて仮説検証研究のさらなる活性化を促進すると期待できよう。

現に、話し言葉の機械認識における「隠れマルコフモデル (HMM)」や、書き言葉の「N-gram 解析」では、最尤推定法に基づく確率予測の手法が中心的な役割を果たしており、画期的な成果をおさめている。最尤推定法は DNA の塩基配列記号の解読にも利用されている。

先の松田 [2] は米国に留学し、Labov [1] の薫陶を受けた。彼は『1960 年代中ごろにアメリ

カで誕生した変異理論と日本のフィールド言語学の邂逅の一部始終は、国産パラダイムと外来パラダイムの邂逅例として非常に興味深い。ここから異なるパラダイム同士のあるべき姿について貴重な教訓が得られると思われる。』（横山が一部改変）と述べている。このような学問的系譜をふまえたうえで、今回は、書き言葉のデータについて検討する。本稿の前半部分は先行研究の紹介と、そのデータの再分析にあてる。後半部分は、ロジスティック回帰分析の理論を可能な限り高校文系数学までの範囲で解説していく。

2. 学術用語の消長：第 1 カーブ (横山・真田)

言語変化を記述する関数として Piotrowski [4] と Altmann [5] は 3 種類のカーブを提唱した。これに着目した真田 [6] は、明治期以降における学術用語の消長過程に 3 種類のカーブを当てはめる分析を行った。

言語変化は一般に「遅→速→速→遅」という過程をたどって伝播するとされる。Altmann [5] による 3 種類のカーブは S 字型をなし、他の関数では説明しにくい言語変化のデータ分布に適用しやすいところに特徴がある。言語変化の S 字カーブはロジスティック曲線で近似できると言われている。ロジスティック曲線 (シグモイド曲線) の例を図 1 に示す。

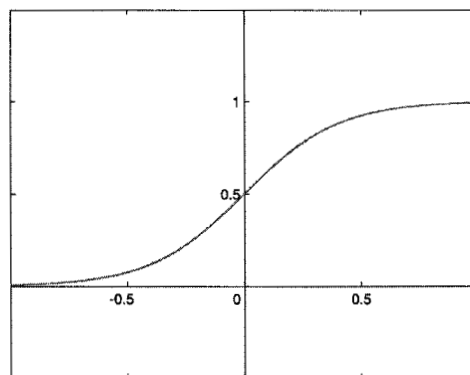


図 1 ロジスティック曲線の例
(縦軸上限値 1.0 未満, 下限値 0.0 より大)

この項は、真田【6】が紹介した第1のカーブと、医学分野等で盛んに使われているロジスティック回帰モデルの共通点を明らかにする。

第1のカーブは

$$p=1/\{1+A \times \exp(-K \times t)\} \quad [2.1]$$

$A=\exp(\log A)$ を代入して整理すると

$$\begin{aligned} p &= 1/\{1+\exp(\log A) \times \exp(-K \times t)\} \\ &= 1/\{1+\exp(\log A - K \times t)\} \\ &= 1/\{1+\exp(-K \times t + \log A)\} \end{aligned} \quad [2.2]$$

横山【7】【8】とYokoyama & Wada【9】は、文字研究にロジスティック回帰モデルを始めて導入した。日本語学を見渡すと、Hibiya【10】ならびにMatsuda【11】の一連の研究に続く試みになる。そのモデル式を式【2.3】に示す。

以下、 θ （シータ）は説明変数、 α （アルファ）は傾き、 β （ベータ）は定数を示す。

$$\begin{aligned} Z &= \alpha \times \theta + \beta \text{ として} \\ \log\{p/(1-p)\} &= Z \end{aligned} \quad [2.3]$$

$p/(1-p)$ は「オッズ (odds)」、 $\log\{p/(1-p)\}$ は「ロジット (logit)」という。

式【2.3】を変形すると

$$p=\exp(Z)/\{1+\exp(Z)\} \quad [2.4]$$

$$=1/\{1+\exp(-Z)\} \quad [2.5]$$

$-Z=-\alpha \times \theta - \beta$ なので

$$p=1/\{1+\exp(-\alpha \times \theta - \beta)\}$$

ここで $\alpha=K$ 、 $\theta=t$ 、 $-\beta=\log A$ とおくと

$$p=1/\{1+\exp(-K \times t + \log A)\}$$

これは式【2.2】と一致し、第1のカーブはロジスティック回帰モデルと同じであることが証明された。これは、後述するテスト理論との関

連を念頭におくと、次のように変形できる。

$$\begin{aligned} p &= 1/\{1+\exp(-K \times t + \log A)\} \\ &= 1/\{1+\exp[-K \times (t - \log A/K)]\} \end{aligned} \quad [2.6]$$

式【2.6】の $\log A/K$ は変曲点にあたる。以下、第1のカーブは式【2.3】を意識して式【2.7】で表す。

$$\log\{p/(1-p)\} = \alpha \times \theta + \beta \quad [2.7]$$

3. 外来語の増加過程（横山・真田）

真田【6】は「第2、第3のカーブ」も紹介している。この項は、第2のカーブとロジスティック回帰モデルの関係について検討する。

第2のカーブは

$$p=C/\{1+A \times \exp(-K \times t)\} \quad [3.1]$$

第1のカーブと同様、 $A=\exp(\log A)$ を代入して整理すると

$$p=C/\{1+\exp(-K \times t + \log A)\} \quad [3.2]$$

ここで $K=\alpha$ 、 $t=\theta$ 、 $\log A=-\beta$ とおくと

$$p=C/\{1+\exp(-\alpha \times \theta - \beta)\}$$

先に $-\alpha \times \theta - \beta = -Z$ とおいたので

$$p=C/\{1+\exp(-Z)\}$$

右辺の分母を両辺に掛けて

$$p+p \times \exp(-Z)=C$$

$$p \times \exp(-Z)=C-p$$

$$\exp(-Z)=(C-p)/p$$

$$1/\exp(Z)=(C-p)/p$$

分子と分母を両辺で逆転させて

$$\exp(Z)=p/(C-p)$$

$$Z=\log\{p/(C-p)\}$$

$$\log\{p/(C-p)\}=Z$$

$$\log\{p/(C-p)\}=\alpha \times \theta + \beta \quad [3.3]$$

第1, 2カーブの数式を示しておく。

■第1のカーブ

$$p=1/\{1+A \times \exp(-K \times t)\} \quad [2.1]$$

または

$$p=1/\{1+\exp(-K \times t + \log A)\} \quad [2.2]$$

あるいは

$$\log\{p/(1-p)\} = \alpha \times \theta + \beta \quad [2.7]$$

ただし, $\alpha=K, \theta=t, -\beta=\log A$

注) 式 [2.7] は, 目的変数がロジットになっており, ロジスティック回帰モデルそのものである。

■第2のカーブ

$$p=C/\{1+A \times \exp(-K \times t)\} \quad [3.1]$$

または

$$p=C/\{1+\exp(-K \times t + \log A)\} \quad [3.2]$$

あるいは

$$\log\{p/(C-p)\} = \alpha \times \theta + \beta \quad [3.3]$$

ただし, $\alpha=K, \theta=t, -\beta=\log A$

さて, 橋本【12】は, 明治期以降の新聞社説に登場した外来語の増加過程に関するデータを収集し, その変化を第2のカーブに当てはめた。本研究は, 同じデータに対して第1のカーブ(ロジスティック回帰モデル)を最小自乗法と最尤推定法で当てはめる試みを行った。その結果, 以下が明らかになった。

(1) パラメータの値に大きな開きはない

第1と第2で推定されたパラメータの数値に微妙な違いがあるものの, 大差はない。

<第1のカーブ>

最小自乗法: $K=0.046, \log A=3.760$

$$\log\{p/(1-p)\} = 0.046 \times t - 3.760$$

最尤推定法: $K=0.041, \log A=3.406$

$$\log\{p/(1-p)\} = 0.041 \times t - 3.406$$

<第2のカーブ>

本研究では式 [3.3] でパラメータを推定した。橋本【12】によると $C=0.8288$ である。数学的には橋本【12】と式 [3.3] は同じである。

最小自乗法: $K=0.049, \log A=3.639$

$$\log\{p/(0.8288-p)\} = 0.049 \times t - 3.639$$

注) 第2カーブの最尤推定法はツールなし。

(2) 決定係数もほぼ同じ

カーブから得た p の理論値を実測値に当てはめた決定係数(説明率)を示す。第1, 2の違いは小さい。計算法によっては第1の方が微妙によい場合さえある。

<第1のカーブ>

最小自乗法 $r^2=0.834$, 最尤推定法 $r^2=0.829$

<第2のカーブ>

最小自乗法 $r^2=0.831$

(3) 理論値の相関はほぼ+1

第1, 2のカーブによる理論値の相関を計算したところ, $r=0.999$ (ほぼ+1) になった。

橋本【12】のパラメータによる第2カーブ理論値と, 本研究が計算した第1カーブ理論値について, それぞれの理論値が実測データにどのくらい当てはまるかを図2に示す。以上の結果から, このデータにおいては, 説明率やパラメータの値で第1, 2のカーブに微差しかないことが明らかになった。より単純なモデルを採用するなら, 第1カーブの式 [2.7] になる。

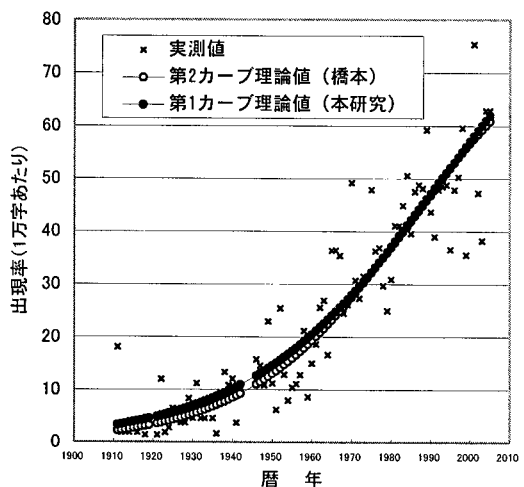


図2 第2カーブと第1カーブの比較

4. 再び学術用語の消長：第2カーブ（横山・真田）

真田【6】は、「学術漢語」が外国語辞書に採用された語数をデータとして、その時系列的な増加状況を第2カーブで分析した。学術漢語は1,847語、時系列は辞書の出版年をあてた。

$C=0.75$ として、本研究では式〔3.3〕によりパラメータを推定した。

<第2のカーブ>

最小自乗法： $K=0.035$, $\log A=2.807$

$$\log\{p/(0.75-p)\} = 0.035 \times t - 2.807$$

p の理論値を得る式は

$$p = 0.75 / \{1 + \exp(-0.035 \times t + 2.807)\}$$

実測値に当てはめた決定係数（説明率）は $r^2=0.876$

さて、このデータに対して、2006年11月30日に横山がSPSSの操作ミスをおかし、偶然に「オッズを目的変数とする回帰分析」を以下のように行った。

$Z = \alpha \times \theta + \beta$, ただし $\alpha = K$, $\theta = t$, $\beta = -\log A$ として

$$p/(1-p) = Z \quad [4.1]$$

左辺の分母を両辺に掛けて

$$p = Z - pZ$$

$$p + pZ = Z$$

$$p(1+Z) = Z$$

$$p = Z / (1+Z) \quad [4.2]$$

この式〔4.2〕は、式〔2.4〕から \exp を除いた形になっていることに注意されたい。

また、式〔4.1〕の両辺を対数変換すると $\log\{p/(1-p)\} = \log Z$ なので、式〔4.2〕は次のような表現もできる（式の導出は省略）。

$$p = 1 / \{1 + \exp(-\log Z)\} \quad [4.3]$$

これは式〔2.2〕と似た形をしていることから、以下では「第1カーブの新種（以下、第1カーブ亜種という）」として議論を進める。式〔4.1〕を用いてパラメータを推定した結果を示す。

<第1カーブ亜種>

最小自乗法： $K=0.021$, $\log A=0.953$

$$p/(1-p) = 0.021 \times t - 0.953$$

p の理論値を得る式は

$$p = (0.021 \times t - 0.953) / (1 + 0.021 \times t - 0.953)$$

実測値に当てはめた決定係数（説明率）は $r^2=0.888$

真田【6】の第2カーブと、本研究の第1カーブ亜種は、実測データへの当てはまりの良さにほとんど違いはない。フィッティングに関する両者のありさまを図3に示す。

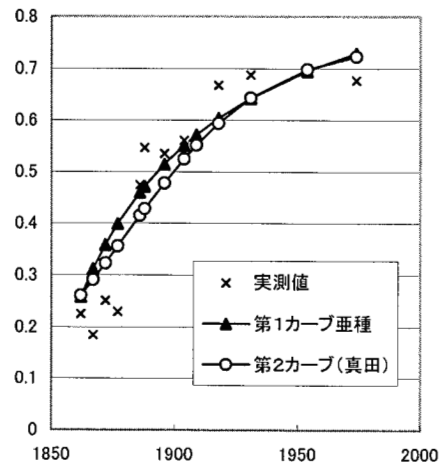


図3 第2カーブと第1カーブ亜種の比較
（横軸が出版年、縦軸は採用語数の割合）

5. ロジスティック回帰分析とは（横山）

これ以降の部分は、ロジスティック回帰分析の理論を可能な限り高校文系数学までの範囲で解説していく。書き言葉のN-gram解析では、最尤推定法を用いた確率予測の手法が大きな成

果をあげていることから、最尤推定法を思想を高校2年生（文系）にも理解できる形で提供したい。

(1) 尤度とは

ロジスティック回帰分析の説明に先立ち、「尤度（ゆうど：likelihood）」という考え方を導入する。尤度は確率を別の角度から見た概念である。

確率との違いを簡単な例で説明しよう。10枚のコイン投げをイメージされたい。直観的な理解を助けるために表1を掲出した。表中の数値は、コインの表が出る確率 λ （ラムダ）のときに、表が r 枚出る確率を示す。

表1 確率と尤度の関係

r	λ							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	...
0			0.03					
1			0.12					
2			0.23					
3	0.06	0.20	0.27	0.21	0.12	0.04	0.01	0.00
4			0.20					
5			0.10					
6			0.04					
7			0.01					
8			0.00					
9			0.00					
10			0.00					

■確率の世界

これから10枚のコインを投げる。コインの表が出る確率 $\lambda = 0.3$ の列を縦に見ると、表が0, 1, 2, ..., 10枚になる確率を一覧できる。縦の合計は1。

■尤度の世界

10枚のコインを投げた結果、表が3枚出たとする。 $r=3$ の行を横に見ると、コインの表が出る確率 λ によって表が3枚出る確率の変化を一覧できる。横の合計は1ではない。

以上をまとめると、次のようになる。

1. 確率は現状から未来を予測するための指標である。確率の総和は1となる。
2. 一方、尤度は現状から過去を推理するための指標といえる。尤度の総和は1ではない。

(2) 対数尤度とは

100枚のコインを投げる。それぞれのコインは等質だとする。表（オモテ）が出る確率を λ とすれば、表が70枚出る確率は

$${}_{100}C_{70} \lambda^{70} (1-\lambda)^{30}$$

いま赤のコインが100枚、青のコインが80枚あるとしよう。コインの表が出る確率は、赤が p_R 、青が p_B とする（注： p_R , p_B の p はここでは λ とすべきだが、 p の方が一般に確率を示す記号として「なじみ」があるため、以下このままとする）。例えば、赤100枚を投げて、表が70枚出る確率は

$${}_{100}C_{70} p_R^{70} (1-p_R)^{30}$$

青80枚を投げて、表が50枚出る確率は

$${}_{80}C_{50} p_B^{50} (1-p_B)^{30}$$

この赤100枚と青80枚を一度に投げたとき、赤の表が70枚、青の表が50枚となる確率 Ω （オメガ）は、

$$\Omega = {}_{100}C_{70} p_R^{70} (1-p_R)^{30} \times {}_{80}C_{50} p_B^{50} (1-p_B)^{30}$$

両辺の自然対数をとると

$$\begin{aligned} \log \Omega &= \log {}_{100}C_{70} + 70 \log p_R + 30 \log (1-p_R) \\ &\quad + \log {}_{80}C_{50} + 50 \log p_B + 30 \log (1-p_B) \end{aligned}$$

さて、コインの「重心のずれ」が表の出る確率を下の数式で左右すると仮定しよう。

$$Z = a \times \text{重心のずれ} + b \quad [5.1]$$

赤コインの重心のずれを反映した関数を Z_R とおく（赤コインはすべて同じ重心のずれを持

つと仮定)。青コインも同様。

$$Z_R = a \times \text{赤の重心のずれ} + b$$

この関数 Z_R が赤コインの表が出る確率 p_R を (正規分布にしたがって) 左右するとしよう。それはロジスティック関数により、次のように表現できる。

表が出る確率

$$p_R = 1 / \{1 + \exp(-Z_R)\} \\ = \exp(Z_R) / \{1 + \exp(Z_R)\}$$

裏が出る確率

$$1 - p_R = \{1 + \exp(Z_R) - \exp(Z_R)\} / \{1 + \exp(Z_R)\} \\ = 1 / \{1 + \exp(Z_R)\}$$

それぞれ対数変換すると

$$\log p_R = \log \{ \exp(Z_R) / \{1 + \exp(Z_R)\} \} \\ = \log \exp(Z_R) - \log \{1 + \exp(Z_R)\} \\ = Z_R - \log \{1 + \exp(Z_R)\}$$

$$\log (1 - p_R) = \log \{ 1 / \{1 + \exp(Z_R)\} \} \\ = \log 1 - \log \{1 + \exp(Z_R)\} \\ = -\log \{1 + \exp(Z_R)\}$$

青コインの p_B , $1 - p_B$ も同様。

これを $\log \Omega$ に代入すると

$$\log \Omega = \\ \log_{100} C_{70} + 70 [Z_R - \log \{1 + \exp(Z_R)\}] - 30 \log \{1 + \exp(Z_R)\} \\ + \log_{80} C_{50} + 50 [Z_B - \log \{1 + \exp(Z_B)\}] + 30 \log \{1 + \exp(Z_B)\} \\ = 70 Z_R - 70 \log \{1 + \exp(Z_R)\} - 30 \log \{1 + \exp(Z_R)\} \\ + 50 Z_B - 50 \log \{1 + \exp(Z_B)\} - 30 \log \{1 + \exp(Z_B)\} \\ + \log_{100} C_{70} + \log_{80} C_{50} \\ = 70 Z_R - 100 \log \{1 + \exp(Z_R)\}$$

$$+ 50 Z_B - 80 \log \{1 + \exp(Z_B)\} \\ + \log_{100} C_{70} + \log_{80} C_{50}$$

すなわち

$$\log \Omega = 70 Z_R + 50 Z_B \\ - 100 \log \{1 + \exp(Z_R)\} - 80 \log \{1 + \exp(Z_B)\} \\ + \log_{100} C_{70} + \log_{80} C_{50}$$

この式に重心のずれデータを投入してみよう。

$$\log \Omega = 70 (a \times \text{赤の重心のずれ} + b) \\ + 50 (a \times \text{青の重心のずれ} + b) \\ - 100 \log \{1 + \exp(a \times \text{赤の重心のずれ} + b)\} \\ - 80 \log \{1 + \exp(a \times \text{青の重心のずれ} + b)\} \\ + \log_{100} C_{70} + \log_{80} C_{50} \quad [5.2]$$

上記の $\log \Omega$ を別の視点から眺めてみたい。コイン 100 枚と青コイン 80 枚を一度に投げて、表になった枚数を数えてみたら、赤が 70 枚、青が 50 枚という「結果」であった。これは誰にでも分かる。では、この結果をもたらした「原因」は何か。それは「重心のずれ」である。重心のずれが表の出る確率を変化させる式 [5.1] のパラメータを求めてみよう。

解法の道筋は次の通り。まず、赤コイン 100 枚と青コイン 80 枚を一度に投げたとき、赤の表が 70 枚、青の表が 50 枚となる確率は Ω である。確率 Ω を最大にする a^* と b^* が分かれば、それが表の出る確率を左右する数式として「一番ふさわしく、もっともらしい」と認定しても差し支えないだろう。

$$Z = a^* \times \text{重心のずれ} + b^* \quad [5.3]$$

重心のずれが原因なので、それを θ とおくと

$$Z = a^* \times \theta + b^* \quad [5.4]$$

このような観点で確率 Ω をみたのが尤度 (ゆうど) である。 $\log \Omega$ は計算の便のために対数変

換したもので、「対数尤度」という。

(3) 最尤推定法とは

対数尤度の最大値を与えるパラメータを推定するのが最尤推定法である。式〔5.2〕の例では、 a ならびに b で偏微分し、その連立方程式を解いて a^* と b^* を求めればよい。ただし、偏微分は高校文系数学の範囲を大きく逸脱するほか、その具体的な方法を熟知しなくてもフィールド言語学には差し障りがないので省略する。これは、言語学者が数量化理論Ⅲ類の固有値計算アルゴリズムを知る必要がないのと同じである。仮に、最尤推定法における偏微分の連立方程式が理解できたとしても、その連立方程式は数式の展開では解けないことが知られており、ニュートン・ラプソン法やEM法などの数値計算で近似解を得ることになる。そのアルゴリズムに立ち入るのは言語学の守備範囲を超えている、と言って大過ない。

重要なのは、最尤推定法を使う目的を理解することである。「対数尤度を最大化させるパラメータ（関数 Z の傾きや切片)を求めるのが最尤推定法だ」という言説のイメージが掴めさえすれば、普通は十分であろう。そのためには、尤度の思想を心に刻むことが肝要である。確率の計算は「原因→結果」の方向で予測するやり方である。表が出る確率 λ （原因）を知っていたら、100枚のコインのうち70枚が表になる確率が分かる。それに対して、尤度の計算は「結果→原因」の方向で推理するやり方である。100枚のコインのうち70枚が表だったという結果から、表が出る確率 λ （原因）がいくつだったかを推理する。その推理の「もっともらしさ」を示す程度の指標として確率 Ω を改めて見直したとき、 Ω は尤度と呼ばれる。

(4) θ と Z と正規分布の関係

ロジスティック回帰モデルでは、原因 θ の値が同じでも、あるときは成功し、別のときは失敗すると考える。つまり、成功するか失敗する

かは「確率的に」決定される。その確率を計算する流れなどを含めて、続きは別稿をなす。

6. 参考・引用文献

- [1] Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- [2] 松田謙次郎 (2006). 「変異理論と日本のフィールド言語学—邂逅と誤解の物語—」『日本のフィールド言語学—新たな学の創造に向けた富山からの提言—』, 中井精一・ダニエル＝ロング・松田謙次郎 (編), 真田信治 (監修), 3-16
- [3] 横山詔一 (2006). 「新刊寸感」『日本語学』, 25 巻, 11 月号, 88-89
- [4] Piotrovskaja, A.A., & Piotrovskij, R.G.(1974). *Mathematiceskie mode i v diachronii i tekstoobrazovanii (Statistika reci i avtomaticeskij analiz teksta)* Leningrad: Nauka, 361-400 [未見]
- [5] Altmann, G. (1983). A Law of Change in Language, *Historical Linguistics (Quantitative Linguistics. Vol.18)*, Studienverlag Dr. N. Brockmeyer (Bochum), 104-115
- [6] 真田治子 (2002) 『近代日本語における学術用語の成立と定着』, 絢文社
- [7] 横山詔一 (2006). 「異体字選好における単純接触効果と一般対応法則の関係」『計量国語学』, 25, 199-214
- [8] 横山詔一 (2006). 「文字認知の単位」『月刊言語』, 35 巻, 10 号, 36-43
- [9] Yokoyama, S. & Wada, Y. (2006). A logistic regression model of variant preference in Japanese kanji: an integration of mere exposure effect and the generalized matching law. *Glottometrics*, 12, 63-74.
- [10] Hibiya, J.(1988). A quantitative study of Tokyo Japanese. Doctoral Dissertation, Dept of Linguistics, University of Pennsylvania [未見]
- [11] Matsuda, K.(1993). Dissecting analogical leveling quantitatively: The case of the innovative potential suffix in Tokyo Japanese. *Language Variation and Change*, 5, 1-34
- [12] 橋本和佳 (2006). 「Logistic 曲線による外来語増加過程のモデル化——大正から平成までの社説を用いて——」『計量国語学』, 25, 293-308

【謝辞】カナダ Toronto 大学の Chambers 先生からは、2006年10月20日に国立国語研究所で開催された講演会でS字カーブについて議論する機会をいただいた。この講演会は国立国語研究所研究員の朝日祥之氏が企画し、通訳もつとめた。また、同志社大学の橋本和佳先生にはデータをご提供いただいた。さらに、国立国語研非常勤研究員の和田志子氏には英文要旨の作成を助けてもらった。以上の皆さまに深く感謝申し上げる。