On maximum likelihood estimation in method of complete series : An investigation by simulation

全系列法データの最尤法による分析について
：シミュレーションによる検討

岡 本 安 晴

近年のコンピュータの高性能化と低コスト化により数理モデルに基づく分析が数学的な解析以外の方法によって容易に行われるようになってきた。すなわち、コンピュータによるシミュレーションの可能なものが以前に比べて格段に増えてきている。例えば、種々の統計的性質を調べることは昔は数学的な解析によっていたが、今はシミュレーションを用いたbootstrap法（Diaconis & Efron, 1983）が用いられたりする。簡単な統計的性質の場合は、乱数を用いたシミュレーションによる分析がパーソナルコンピュータ上でも可能である。

統計的性質のパーソナルコンピュータ上のシミュレーションによる検討の例として、分散分析におけるサンプル数と検定力の関係を調べたものがあるので以下に示す。

サンプル数と検定力の関係について小牧（1992, p.150）は、サンプル数を増やしても効果のない変数が効果のある変数に変わるわけではないと論じている。これは帰無仮説に対する有意水準の考え方からすれば当然のことである。小牧は上の結論の前にϕが1である場合のサンプル数の増加に伴う検定力の変化をとりあげている。すなわち、ϕが1の場合サンプル数を増やしても検定力は高々30％程度の水準に達するだけだとしている。

ここで、ϕは次式で与えられているものである（Kirk, 1968, p.109）。

$$\phi = \frac{\sqrt{\sum_{j=1}^{k} \beta_j^2 / k}}{\sigma_\varepsilon / \sqrt{n}}$$

上式において、β_jは条件jにおける平均値、kは条件の数、σ_εは誤差の分散の平方根、nは各条件内のサンプル数である。

ϕが0でなく1であるということは、$\beta_j \neq 0$ということであり、これは条件差が無いという帰無仮説が成り立たないということである。従って、効果のない変数についての議論には不適当である。

さらに、ϕはサンプル数nの関数であるので、ϕを一定にしてnを大きくしたときの効果を論ずるのも問題がある。ϕが一定のままでnを大きくしていくとσ_εが一定であればβ_jは0になっていく。nの検定力に及ぼす効果を調べるときには、n以外のパラメー

タの値を一定にして、nのみの増加に伴う検定力の変化を見るべきである。この変化の様子はシミュレーションによって簡単に調べることができる。

$\sigma_\epsilon=1$、$k=3$、$\beta_1=0$、$\beta_2=-\beta_3$、$n=10$ のときの $\phi=1$ に対する β_2の値は $\beta_2 \fallingdotseq 0.3873$ である。この条件で分散分析のシミュレーションを行ってみた。すなわち、$\sigma_\epsilon=1$、$k=3$、$\beta_1=0$、$\beta_2=-\beta_3=0.3873$ として、5%水準での有意な結果が得られる割合がnの増加とともにどう変わるかを調べた。1つのnの値に対してシミュレーションを10000回行い、そのうちの有意な結果になったものの割合を求めたものが表1である。表1には$\beta_2=0$に対するものも載せられている。条件差がない場合（$\beta_2=0$）にはnの値によらず有意な結果の得られた割合は有意水準の5%に近い。しかし、条件差がある場合（$\beta_2=0.3873$）には、一条件当たりのサンプル数が100であればほぼ確実に(0.9994) 有意な結果が得られている。

表1　有意水準5%で有意になった割合

サンプル数	条件差	
	$\beta_1=\beta_2=\beta_3=0$	$\beta_1=0$, $\beta_2=-\beta_3=0.3873$
$n=10$	0.0484	0.2893
$n=100$	0.0483	0.9994

10000回のシミュレーションにおける割合。$\sigma_\epsilon=1$ である。

このようにシミュレーションによって統計の性質を簡単に調べることができるが、これは実験のデザインなどを考えるときに大いに役立つものである。

シミュレーションによる検討は、典型的な統計法以外のデータの分析法一般に対しても有効なことがある。次に、岡本 (1994) による全系列法データの最尤法による分析について検討を行う。

全系列法データ分析法の比較

全系列法データの分析法として岡本 (1994) は次のような Thurstone (1927) 流の確率モデルに基づいたものを試みている。

標準刺激 S_S、および比較刺激 S_i（$i=-N$、$\cdots N$）の感覚強度が平均 μ_S、および μ_i の確率変数で表わされるとする。但し、$S_S=S_0$ である。判断は3件法、例えば「重い」「同じ」「軽い」で行われるとし、「同じ」判断に対応してカテゴリー判断の基準（境界）値 C を設ける。これは、信号検出理論 (Green & Swets, 1966) における判断の基準値

(criterion) の概念を3件法に適用したものであり、感覚の弁別力を表す閾値 (threshold) とは区別されるものである。

Thurstone (1927) のモデルでは感覚量は正規分布に従うとされるが、岡本 (1994) では計算速度を上げるため感覚量の差の累積分布関数を次の logistic 分布で与えている。

$$\phi(x;\mu) = \phi_0(x-\mu)$$

ここで

$$\phi_0(x) = 1/(1+\exp(-s\cdot x), \quad s>0$$

このとき、順位誤差を無視すれば、例えば S_i の方が S_S より「重い」と判断される確率 $P(S_i > S_S)$ は次のように与えられる。

$$P(S_i > S_S) = 1 - \phi(C;\mu_i - \mu_S)$$

S_i と S_S との比較の確率が与えられれば、データに対する尤度を最大にする値として各パラメータの値が決まる。

パラメータの値が決まれば、閾値は75％の弁別確率に対応する刺激値の差 (Falmagne, 1986) として求めることができる。上の logistic 分布に基づくモデルの場合の閾値 δ は次のように与えられる (岡本, 1994)

$$\delta = \frac{log\,3}{s} \tag{1}$$

Saito (1992) では C (Saitoでは ε で表記) を閾値としているが、これは Psychophysics の考え方からすれば誤りである。

本学の文学部心理学専攻生を対象とする初級実験においては、全系列法データから閾値を求める方法として次のような方法が示されている (心理学実験 (I) マニュアル (改訂版), 1990)。

隣り合う比較刺激間の差を h で表わす。

$$h = S_i - S_{i-1}$$

各比較刺激 S_i と標準刺激 S_S との比較が n 回行われるとし、S_i が S_S より「重い」と判断された回数を g_i、「軽い」と判断された回数を k_i で表わす。比較刺激の最大のものは S_N で表されるが、このとき上弁別閾 ΔS_o は次式で与えられるとされている。

$$\Delta S_o = \left[\left(S_N + \frac{h}{2}\right) - \frac{\sum g_i}{n}\cdot h\right] - S_S$$

同様に下弁別閾ΔS_μを求める式は次のように示されている。

$$\Delta S_u = S_S - \left[\left(S_{-N} - \frac{h}{2} \right) + \frac{\sum k_i}{n} \cdot h \right]$$

ここで、S_{-N}は比較刺激のうち最小のものである。

ΔS_oとΔS_uから平均弁別閾ΔSが算出されている。

$$\Delta S = \frac{\Delta S_o + \Delta S_u}{2} \tag{2}$$

(2)式より弁別閾を求めるとき、被験者の判断の基準の要因の影響はどのようであるのか。この問題を、岡本（1994）の最尤法による求め方との比較においてシミュレーションを用いて検討する。

岡本（1994）に報告されている実験データの分析結果を参考にしてシミュレーションのためのパラメータを以下のように設定する。

$$\mu_S = 60$$
$$\mu_i = 60 + 3 \times i; \qquad i = -7, \cdots, 7$$

S_SおよびS_iの感覚量は平均μ_Sおよびμ_i、分散$\sigma^2 = 25$の正規分布に従うものとする。各S_iとS_Sとの比較は$n = 8$回行われる。順位誤差は無いものとする。S_iとS_Sの感覚量の差の絶対値がD_Cより小さいとき「同じ」と判断されるが、大きい場合には差の正負に応じて「重い」、「軽い」の判断が選ばれるとする。判断の基準D_Cの値を$D_C = 1$，2，3と変えて、それぞれの値に対して生成されるデータから式(1)及び式(2)よって閾値を算出し、比較検討を行う。1つのD_Cの値に対して、全系列法データの生成とそのデータから式(1)および式(2)によって閾値を算出するということを100回繰り返し、それぞれの100回の算出値から中央値を求めたものが表2に示されている。

表2　2つの方法による閾値

基準値	最尤法		ΔS
	C	δ	
$D_C = 1$	0.95	4.42	0.94
$D_C = 2$	2.04	4.51	2.06
$D_C = 3$	3.08	4.45	3.00

それぞれのD_Cの値に対する100回のシミュレーションに基づく中央値。

式(2)より算出される閾値 ΔS は D_C の変化による影響を強く受けているが、式(1)から算出される最尤法の値 δ は判断の基準値 D_C が変化してもほとんど変わらない。最尤法においては D_C の変化は C の推定値の変化として現れている。

ちなみに、各刺激 S_C および S_i が独立に分散 σ^2 の正規分布に従うとき、その差の分散は $2\sigma^2$ である。このとき75%の点は $0.67\sqrt{2\sigma^2}$ であるので、$\sigma^2 = 25$ に対しては $0.67\sqrt{2\sigma^2} \fallingdotseq 4.7$ である。表2の最尤法が *logistic* 分布に基づくものであることを考慮すれば表2の δ の値とこの 4.7 という値はよく対応しているといえる。

結　　論

はじめに、コンピュータの高性能化と低コスト化によってシミュレーションが有力かつ実用的な方法として使える分野が多くなってきたことを指摘し、シミュレーションによる統計的性質の検討の簡単な1例として分散分析におけるサンプル数と検出力の関係を調べたものを報告した。

次に、本論である全系列法データの分析法のシミュレーションによる検討が行われた。感覚と判断の基準を区別する信号検出理論の考え方に従う岡本（1994）の最尤法による分析法では当然のことながら判断の基準が変化しても閾値の推定値には影響が現れなかった。これに対して、式(2)による閾値 ΔS は判断の基準の変化に強く影響された。これは ΔS の次の性質による。

S_i が S_S と「同じ」と判断された回数を e_i とおけば

$$\Delta S = S_N - S_{-N} + h - \frac{\Sigma g_i}{n} \cdot h - \frac{\Sigma k_i}{n} \cdot h$$

$$= \frac{\Sigma g_i + \Sigma e_i + \Sigma k_i}{n} \cdot h - \frac{\Sigma g_i}{n} \cdot h - \frac{\Sigma k_i}{n} \cdot h$$

$$= \frac{\Sigma e_i}{n} \cdot h$$

すなわち、ΔS は「同じ」判断の総数のみによって決まる値である。このため、判断の基準（厳しさ／緩さ）の影響を直接受けることになる。

引 用 文 献

Diaconis, P. & Efrom, B. (1983) Computer-intensive methods in statistics. *Scientific American,* May, 96-107.

Falmagne, J. C. (1986) Psychophysical measurement and theory. In K. R. Boff, L. Kaufman, & J.P.Thomas (Eds.), *Handbook of perception and human performance* : Vol.1. John Wiley and Sons.

Green, D. M., & Swets, J. A. (1966) *Signal detection theory and psychophysics.* New York・London・Sidney : John Wiley and Sons Inc.

Kirk, R. E. (1968) *Experimental design : Procedures for the behavioral sciences.* California : Btooks／Cole.

小牧純爾 (1992) 変動と一般性：個体型実験法の理解をめぐって。心理学評論, 35, 133-155.

岡本安晴 (1994) 全系列法データの最尤法による分析例。金沢大学文学部論集・行動科学科篇, 14, 63-71.

Saito, T. (1992) Measurement of the asymmetry observed in compatateve indgment. *Hokkaido Behavioral Science Report, Series M,* No.19, August

心理学実験（Ｉ）マニュアル（改訂版）(1990) 金沢大学文学部心理学研究室。

Thurstone, L.L. (1927) A law of comparative Judgment. *Psychological Review,* 34, 273-286.