（８）'Developments of Statistical Inference: A meeting in honour of Professor Kei Takeuchi on the occasion of his 70th birthday' に関する研究報告

Large-deviation efficiency of first and second order

筑波大・数学 赤平昌文

1 はじめに

統計的推定において, Bahadur 有効性は大偏差確率の観点から導出された漸近有効性の概念の 1 つとしてよく知られている ([B71]). そこでは, 未知母数の一致推定量の裾確率の対数の下界が Kullback-Leibler (K-L) 情報量によって与えられ, また適当な正則条件の下では, Fisher 情報量によっても与えられる. そして, その下界を達成する推定量を Bahadur 有効であるといい, さらに K-L 情報量の漸近展開に基づく Bahadur 型の 2 次有効性についても論じられている ([F82], [A95]).

本論では, Bahadur 有効性とはやや異なる観点から, 1 次, 2 次の大偏差有効性について論じる. 実際, 漸近中央値不偏推定量の裾確率に対する 2 次の下界を求め, 具体的には正規分布, 両側指数分布の場合に最尤推定量の 1 次, 2 次の大偏差有効性について論じる.

2 設定

$X_1, X_2, \cdots, X_n, \cdots$ をたがいに独立に, いずれも (σ-有限測度 μ に関する) 確率密度関数 (p.d.f.)$f(x,\theta)$ をもつ分布に従う確率変数列とする. ただし, $\theta \in \Theta$ とし, Θ は $\boldsymbol{R}^1$ の開区間とする. いま, $\boldsymbol{X} := (X_1, \cdots, X_n)$ に基づく θ の推定量 $\hat{\theta}_n := \hat{\theta}_n(\boldsymbol{X})$ が, θ について局所一様に

$$P_{\theta,n}\{\hat{\theta}_n \leq \theta\} = 1/2 + o(1), \quad P_{\theta,n}\{\hat{\theta}_n \geq \theta\} = 1/2 + o(1) \quad (n \to \infty)$$

をみたすとき, $\hat{\theta}_n$ を漸近中央値不偏 (asymptotically median unbiased 略して AMU) であるという. ここで $\mathcal{A}$ を θ の AMU 推定量全体のクラスとする.

定義 ある AMU 推定量 $\hat{\theta}_n^* = \hat{\theta}_n^*(\boldsymbol{X})$ が存在して, 任意の $\hat{\theta}_n \in \mathcal{A}$, 任意の $\theta \in \Theta$, 任意の $a > 0$ について

$$\begin{aligned} P_{\theta,n}\{|\hat{\theta}_n - \theta| > a\} &\geq P_{\theta,n}\{|\hat{\theta}_n^* - \theta| > a\}\{1 + (b_1/n) + o(1/n)\} \\ &=: B(a,\theta)\{1 + (b_1/n) + o(1/n)\} \quad (n \to \infty) \end{aligned} \tag{1}$$

とする. ただし, b_1 はある定数とする. このとき, 不等式 (1) が $o(1)$ まで成り立てば, $\hat{\theta}_n^*$ を 1 次の大偏差有効 (large-deviation efficient 略して LDE) であるといい, さらに (1) が $o(1/n)$ まで成り立てば, $\hat{\theta}_n^*$ を 2 次の LDE であるという.

次に, LDE を考えるために, $\hat{\theta}_n (\in \mathcal{A})$ の両側裾確率の下界を求める.

3 1 次, 2 次の大偏差有効性

まず, θ_0 を Θ に任意に固定し, 仮説 $H : \theta = \theta_0 + a$, 対立仮説 $K : \theta = \theta_0$ の漸近水準 $1/2 + o(1)$ の最強力検定 $\phi^* = \phi^*(\boldsymbol{X})$ は

$$\overline{Z(\theta_0)} := \frac{1}{n}\sum_{j=1}^{n} Z_j(\theta_0) > c$$

の形の棄却域をもつ. ただし, $a > 0$ とし,

$$Z_j(\theta_0) := \log(f(X_j, \theta_0)/f(X_j, \theta_0 + a)) \quad (j = 1, \cdots, n)$$

とする. ここで, c は漸近水準 $1/2 + o(1)$ をみたすように決められる定数であるが, いまの場合,

$$c = E_{\theta_0+a}[Z_1(\theta_0)] + o(1) =: \mu + o(1)$$

とすればよい. さて $E_{\theta_0}(\phi^*) = P_{\theta_0}\{\overline{Z(\theta_0)} > c\}$ であるから

$$P_{\theta_0,n}\{\hat{\theta}_n - \theta_0 > a\} \geq 1 - P_{\theta_0,n}\{\overline{Z(\theta_0)} > c\} \tag{2}$$

になる．また，(2) において $\overline{Z(\theta_0)}$ の裾確率の漸近展開を得るためには，鞍部点近似 (saddlepoint approximation) を利用する ([J95])．いま，原点を含むある区間の任意の t について $M(t) := E_{\theta_0}[\exp\{tZ_1(\theta_0)\}]$，$K(t) := \log M(t)$ とし，方程式 $K_1'(t) = \mu$ の t の解を $\hat{t}(a)$ とする．

定理 θ の任意の AMU 推定量 $\hat{\theta}_n$，任意の $\theta \in \Theta$，任意の $a > 0$ について

$$\begin{aligned}&P_{\theta,n}\{\hat{\theta}_n - \theta > a\}\\ &\geq \frac{1}{\lambda}M^n(\hat{t})e^{-n\mu\hat{t}}\left[B_0(\lambda) + \frac{\mathrm{sgn}(\hat{t})}{\sqrt{n}}\frac{\zeta_3(\hat{t})}{6}B_3(\lambda) + \frac{1}{n}\left\{\frac{\zeta_4(\hat{t})}{24}B_4(\lambda) + \frac{\zeta_3^2(\hat{t})}{72}B_6(\lambda)\right\} + O\left(\frac{1}{n^2}\right)\right] \qquad (3)\end{aligned}$$

が成り立つ．ただし，$\lambda = \sqrt{n}\,|\hat{t}|\,\sqrt{K''(\hat{t})}$，$\zeta_3(t) := \kappa_{3,a}(Z_1)/\{K''(t)\}^{3/2}$，$\zeta_4(t) := \kappa_{4,a}(Z_1)/\{K''(t)\}^2$ とし，$\kappa_{3,a}(Z_1)$，$\kappa_{4,a}(Z_1)$ はそれぞれ，仮説 $H: \theta = \theta_0 + a$ の下での Z_1 の 3 次，4 次のキュムラントとする．また

$$\begin{aligned}B_0(\lambda) &:= \lambda e^{\lambda^2/2}\{1 - \Phi(\lambda)\} \approx \frac{1}{\sqrt{2\pi}\lambda} \quad (\lambda \to \infty),\\ B_3(\lambda) &:= -\left\{\lambda^3 B_0(\lambda) - \frac{1}{\sqrt{2\pi}}(\lambda^3 - \lambda)\right\} \approx -\frac{3}{\sqrt{2\pi}\lambda} \quad (\lambda \to \infty),\\ B_4(\lambda) &:= \lambda^4 B_0(\lambda) - \frac{1}{\sqrt{2\pi}}(\lambda^4 - \lambda^2) \approx \frac{3}{\sqrt{2\pi}} \quad (\lambda \to \infty),\\ B_6(\lambda) &:= \lambda^6 B_0(\lambda) - \frac{1}{\sqrt{2\pi}}(\lambda^6 - \lambda^4 + 3\lambda^2) \approx -\frac{15}{\sqrt{2\pi}} \quad (\lambda \to \infty)\end{aligned}$$

とし，Φ を $N(0,1)$ の累積分布関数とする．

上記と同様にして，$a < 0$ についても $\hat{\theta}_n(\in \mathcal{A})$ の裾確率 $P\{\hat{\theta} - \theta < a\}$ の下界を得ることができる．

4 例

例 1 (正規分布)．$N(\theta, 1)$ の p.d.f. を $f(x, \theta)$ とすると，$\mu := E_a(Z_1) = -a^2/2$，$K'(t) = (a^2/2)(1 + 2t)$ となるから $\hat{t} = -1$ になる．また，$\kappa_{3,a}(Z_1) = \kappa_{4,a}(Z_1) = 0$ となるから，(3) より

$$P_{\theta,a}\{|\hat{\theta}_n - \theta| > a\} \geq 2\{1 - \Phi(\sqrt{n}a)\}(1 + O(1/n^2))$$

になる．$\bar{X} := (1/n)\sum_{i=1}^n X_i$ はこの限界を達成するので，θ の 2 次 LDE になる．

例 2 (両側指数分布)．p.d.f. を $f(x, \theta) = (1/2)e^{-|x-\theta|}$ $(x \in \boldsymbol{R}^1; \theta \in \boldsymbol{R}^1)$ とする．このとき (3) より，任意の $\theta \in \boldsymbol{R}^1$，任意の $a > 0$ について

$$\frac{P_{\theta,n}\{|\hat{\theta}_n - \theta| > a\}}{2\{1 - \Phi(\sqrt{n}a)\}} \geq 1 - \frac{1}{2}\left(\frac{13}{4} + \frac{1}{a}\right)\frac{1}{n} + O\left(\frac{1}{n^2}\right) \quad (n \to \infty) \qquad (4)$$

になる．一方，n が奇数，すなわち $n = 2k - 1$ とすると，θ の最尤推定量は標本中央値 $X_{(k)}$ になる．また，任意の $\theta \in \boldsymbol{R}^1$，任意の $a > 0$ について

$$\frac{P_{\theta,n}\{|X_{(k)} - \theta| > a\}}{2\{1 - \Phi(\sqrt{n}a)\}} = 1 - \frac{1}{2}\left(\frac{3}{8} + \frac{1}{a}\right)\frac{1}{n} + O\left(\frac{1}{n^2}\right) \quad (n \to \infty) \qquad (5)$$

になる．よって，(4)，(5) から $X_{(k)}$ は 1 次の LDE であるが，2 次の LDE ではないことが分かる．

参考文献

[A95] Akahira, M. (1995). Statistical inference and large-deviation probability. (In Japanese). *Mathematical Science No. 380*, 56–62.

[B71] Bahadur (1971). *Some Limit Theorems in Statistics.* Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.

[F82] Fu, J. C. (1982). Large sample point estimation: A large deviation theory approach. *Ann. Statist.*, **10**, 762–771.

[J95] Jensen, J. L. (1995). *Saddlepoint Approximations.* Clarendon Press, Oxford.

同時構造方程式（計量経済）モデルの セミパラメトリック推定法の改善

国友直人 (東京大学経済学研究科)

1. 同時方程式モデルのセミ・パラメトリック推定

計量経済学において開発された同時方程式（構造方程式）モデルのパラメトリック推定法としては例えば制限情報最尤法や操作変数法などを挙げることができる。制限情報推定法（または単一方程式法）として知られているこれら推定量の小標本の性質については Anderson=Kunitomo=Sawa (1982) や Anderson=Kunitomo=Morimune (1986) 等の研究がある。その後の計量経済学の展開では Hansen の GMM 法として知られている一般化モーメント法（Godambe (1960) の推定方程式法と本質的に同等）が実用面でもよく用いられている。さらに Owen (1989) が提唱し、推定方程式法に Qin=Lawless (1994) が応用した経験尤度法（Empirical Likelihood Method、略して MEL 法）による推定法も注目されている。

2. 同時方程式の経験尤度推定と GMM 推定

簡単化の為に線形モデルを考えることにして、いま推定したい構造方程式を

$$y_{1i} = (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})\begin{pmatrix} \beta \\ \gamma \end{pmatrix} + u_i \quad (i = 1, \cdots, n) \tag{1}$$

としよう。ここで y_{1i} と $\mathbf{y}_{2i}$ はそれぞれ 1×1 と $G_1 \times 1$ の内生変数、$\mathbf{z}_{1i}$ は $K_1 \times 1$ 外生変数、θ は $p \times 1$ 未知母数ベクトル、$\{u_i\}$ は互いに独立な確率変数列（$E(u_i) = 0 \ (i = 1, \cdots, n)$）である。($\theta^{'} = (\beta^{'}, \gamma^{'})$ は $1 \times p \ (p = G_1 + K_1)$ ベクトルとする。）さらに変数ベクトル $\mathbf{z}_{1i}$ を含む $K \times 1$ 外生変数（あるいは操作変数）が直交条件 (orthogonal condition)

$$E(u_i \, \mathbf{z}_i) = 0 \quad (i = 1, \cdots, n) \tag{2}$$

を満足することを想定しよう。内生変数の全体 $\mathbf{y}_i^{'} = (\mathrm{y}_{1i}, \mathbf{y}_{2i}^{'})$ についての線型性を想定できる場合には、同時方程式の誘導形は

$$\mathbf{y}_i = \mathbf{\Pi}^{'} \mathbf{z}_i + \mathbf{v}_i \quad (i = 1, \cdots, n) \tag{3}$$

と書ける。ここで誤差ベクトル $\mathbf{v}_i^{'} = (v_{1i}, \mathbf{v}_{2i}^{'})$ は $E[\mathbf{v}_i] = 0$ を満たし、$\mathbf{\Pi}^{'}$ は $(1 + G_1) \times K$ の係数行列である。この同時方程式モデルにおける母係数ベクトル $\theta^{'} = (\beta^{'}, \gamma^{'})$ の MEL 推定量は

$$\begin{aligned} &[\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i \, y_{1i}] \\ = \ &[\sum_{i=1}^{n} \hat{p}_i \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}][\sum_{i=1}^{n} \hat{p}_i u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})]\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \end{aligned} \tag{4}$$

と表現される。（ここで $\hat{p}_i$ は経験確率の推定値、$u_i(\hat{\theta})$ は残差の推定値である。）他方、GMM 推定量は

$$\begin{aligned} &[\frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}][\frac{1}{n}\sum_{i=1}^{n} u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i \, y_{1i}] \\ = \ &[\frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{1i} \end{pmatrix} \mathbf{z}_i^{'}][\frac{1}{n}\sum_{i=1}^{n} u_i(\hat{\theta})^2 \mathbf{z}_i \mathbf{z}_i^{'}]^{-1}[\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i (\mathbf{y}_{2i}^{'}, \mathbf{z}_{1i}^{'})]\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \end{aligned} \tag{5}$$

と表される。

3. 経験尤度法の改善

MEL 推定量と GMM 推定量を含む推定量のクラス（修正 MEL 推定法）を導入する。ここで経験確率 $\{p_i(i = 1, \cdots, n)\}$ の推定値を

(6) $$\hat{p}_i^* = \frac{1}{n[1+\delta\, \lambda' \mathbf{z}_i u_i(\hat{\theta})]}$$

とする。ここでδは正に一定値 ($0 \leq \delta \leq 1$) であり、λはMEL推定法に表れるラグランジュ乗数ベクトルである。ここで過剰識別度を$L = K - p$とすると、特に$\delta = \frac{L-1}{L}$ とおいて構成した推定量（修正MEL推定量）について考察した。

このとき一定の仮定の下で漸近的な意味で修正MEL推定量はMEL推定量を改善し、特にLが大きいときには同時にGMM推定量も顕著に改善することが分かった。なお、既に得られた研究結果の一部分については Kunitomo (2002), Kunitomo=Matsushita (2003a,b) に説明した。

4. 高次の効率性との関係

通常のパラメトリック・モデルの枠組みでは構造方程式の推定法に関する高次の漸近効率についてはより一般的な枠組みを用いて Takeuchi=Morimune (1985) が示した結果が重要である。

セミパラメトリック法に関しては最近、Newey=Smith (2001) は Phanzagl= Wefelmeyer (1978) の多項分布に関する議論を利用して、推定方程式の経験尤度推定法が漸近的に高次有効性を持つことを示したと主張している。我々の漸近展開の結果からは、この主張は必ずしも一般的に成り立たないのではないかとの結果が導かれている。この問題については我々の結果の方が正しいという確証はまだないが、セミ・パラメトリック推定の高次の効率についての問題としてみると、なお未解決な重要な問題を提起しているように思われる。

5. 参考文献

Akahira, M. and Tekeuchi, K. (1981), *Asymptotic Efficiency of Statistical Estimators : concepts and Higher Order Efficiency*, Springer.

Anderson, T.W., N. Kunitomo, and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, Vol. 50-4, 1009-1027.

Anderson, T.W., N. Kunitomo, and K. Morimune (1986), "Comparing Single Equation Estimators in a Simultaneous Equation System," *Econometric Theory*, Vol. 2, 1-32.

Godambe, V.P. (1960), "An Optimum Property of Regular Maximum Likelihood Equation," *Annals of Mathematical Statistics*, Vol. 31, 1208-1211.

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, Vol. 50, 1029-1054.

Kunitomo, N. (2002), "Improving Small Sample Properties of the Empirical Likelihoöd Estimation," Unpublished Manuscript.

Kunitomo, N. and Y. Matsushita (2003a), "Finite Sample Distributions of the Empirical Likelihood Estimator and the GMM Estimator," Discussion Paper CIRJE-F-200, Faculty of Economics, University of Tokyo.

Kunitomo, N. and Y. Matsushita (2003b), "Asymptotic Expansions of the Distributions of Estimators in Linear Simultaeous Equation Models," Discussion Paper CIRJE-F-237, Faculty of Economics, University of Tokyo.

Newey, W. K. and R. Smith (2001), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimator," Unpublished Manuscript.

Owen, A. (1989), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, Vol. 22, 300-325.

Phanzagl and Wefelmeyer (1978),"A Their Order Optimum Properties of Maximum Likelihood Estimator," *Journal of Multiariate Analysis*, 8, 1-29.

Qin, J. and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, Vol. 22, 300-325.

Takeuchi, K. and Morimune, K. (1985) "Higher Order Efficiency of the Extended Maximum Likelihood Estimator," *Econometrica*, Vol. 50-4, 1009-1027.

Non-regular Estimation Theory for Piecewise Continuous Spectral Densities

by Masanobu TANIGUCHI
Waseda University, Japan *

Abstract

For a class of Gaussian stationary processes, the spectral density $f_\theta(\lambda)$, $\theta = (\tau', \eta')'$, is assumed to be a piecewise continuous function, where τ describes the discontinuity points, and the gain is smmothly parameterized by η. Although estimating the parameter θ is a very fundamental problem, there has been no systematic asymptotic estimation theory for this problem. This paper develops the systematic asymptotic estimation theory for piecewise continuous spectra based on the likelihood ratio for contiguous parameters. It is shown that the log-likelihood ratio is not locally asymptotic normal (LAN). Two estimators for θ, i.e., the maximum likelihood estimator $\widehat{\theta}_{ML}$ and Bayes estimator $\widehat{\theta}_B$ are introduced. Then the asymptotic distributions of $\widehat{\theta}_{ML}$ and $\widehat{\theta}_B$ are dervied and shown to be nonnormal. Furthermore we observe that $\widehat{\theta}_B$ is asymptotically efficient, but $\widehat{\theta}_{ML}$ is not so. Also various versions of step spectra are considered.

1 Introduction

For independent observations from a uniform distribution, for example, uniform distribution on the interval $[\theta, \theta+1]$, estimating θ is one of the most fundamental problems in statistics. The systematic asymptotic estimation theory has been established by use of the likelihood ratio for contiguous parameters (e.g., Ibragimov and Has'minskii (1981)). However, in time series settings, there has been no systematic theory for such problems. Concretely, suppose that $\{X_t\}$ is a Gaussian stationary process whose spectral density $f_\theta(\lambda), \theta = (\tau, \eta)$, is a step function, and assume that τ describes the location of step, and the gain is parameterized by η smoothly. Spectra of this type appeared in signal processing literature (e.g., Papoulis(1984)) and in econometric literature. Especially, the permanent income hypothesis (e.g., Geweke and Singleton(1981)) is written as a spectral density of step function (see Corbae et al(2002, p.1074)). Furthermore, more generally, suppose that $f_\theta(\lambda), \theta = (\tau', \eta')'$, is piecewise continuous

Mathematics Subject Classification (2000): 62F12, 62M15, 62M99
Key words or phrases: Piecewise continuous spectra, likelihood ratio, non-regular estimation, Maximum likelihood estimator, Bayes estimator, Asymptotic efficiency

with respect to λ, where $\tau = (\tau_1, ..., \tau_p)'$ describes the discontinuous points, and $\eta = (\eta_1, ..., \eta_q)'$ is a smooth parameter for the gain. We can find such spectra in many fields. For example, an environmental time series called the Southern Oscillation Index has the spectral density of piecewise continuous type (see Shumway and Stoffer(2000, p.244)). However, the systematic asymptotic theory for these spectra has been quite barren. Therefore the purpose of this paper is to develop the modern asymptotic theory for piecewise continuous spectral densities. It is shown that the asymptotics of the likelihood ratio between contiguous hypotheses and estimators are greatly different from those for regular spectral models.

This paper is organized as follows. Section 2 deals with a Gaussian stationary process with piecewise continuous spectral density $f_\theta(\lambda), \theta = (\tau', \eta')'$. Then the asymptotics of the likelihood ratio between θ and $\theta_n = \theta + (n^{-1}\Delta^{\tau'}, n^{-1/2}\Delta^{\eta'})'$ are elucidated. It is shown that this family is not locally asymptotic normal (LAN). Section 3 introduces two estimators of θ, i.e., the maximum likelihood estimator (MLE) $\widehat{\theta}_{ML}$ and Bayes estimator $\widehat{\theta}_B$. Based on the asymptotics of the likelihood ratio, the asymptotic distributions of $\widehat{\theta}_{ML}$ and $\widehat{\theta}_B$ are given and shown to be nonnormal. Also the results imply that $\widehat{\theta}_B$ is asymptotically efficient, but $\widehat{\theta}_{ML}$ is not so, which makes a sharp contrast to the ordinary asymptotic theory for smooth spectra (see Taniguchi and Kakizawa (2000)). Section 4 provides various spectra of step function type. For them, the asymptotics of the likelihood ratios are illuminated. Then the similar results to those in Section 3 follow.

As for notations, we denote the Euclidean norml of a matrix A by $\|A\|_E$, the spectral norm of A by $\|A\|$, τ^+ and τ^- are the right and left limits, respectively, and $f_n \xrightarrow{L_2} f$ implies $\int |f_n(\lambda) - f(\lambda)|^2 \, \mathrm{d}\lambda \longrightarrow 0$ as $n \to \infty$.

On nonparametric and semiparametric testing of multivariate time series

矢島美寛 東大・経済

松田安昌　新潟大・経済

1. 序

本稿では、スペクトル密度行列に関する制約条件に対してのノンパラメトリックあるいはセミパラメトリックな検定を統一的に扱う数学的定式化およびスペクトル密度行列の推定量に基づく新たな検定統計量を提案し、その漸近的性質を導く。

2. モデルと帰無仮説

$\boldsymbol{Z}_t = (Z_{1t}, \ldots, Z_{rt})'$を$r$次元正規定常過程とし、そのスペクトル密度行列を$f(\lambda)$とする。$g(\theta, y)$は、$r \times r$行列値関数で、その引数は$v_1$次元ベクトル、$\theta = (\theta_1, \ldots, \theta_{v_1})'$、および$v_2(v_2 < r^2)$次元ベクトル、$y = (y_1, \ldots, y_{v_2})'$からなる。このとき帰無仮説の下で、$f(\lambda)$はつぎの方程式

$$(1) \qquad f(\lambda) = g(\theta, R\mathrm{vec}(f(\lambda))),$$

をみたすとする。ここでRは$v_2 \times r^2$の「選択行列」とし、すなわち各成分はすべて1か0であり、各行ではちょうど1個の成分のみ1、各列では高々1個の成分のみ1である。したがって帰無仮説の下ではスペクトル密度行列はv_1個のパラメータと、r^2より少ないv_2個の自分自身の成分で表現可能なことを意味する。$v_1 = 0$のときはノンパラメトリック検定、$v_1 > 0$のときはセミパラメトリック検定になる。

応用例としては分離型モデルの検定、独立性の検定、条件付き独立性の検定、自己共分散関数の同等性の検定、自己相関関数の同等性の検定がある。

3. 検定統計量

観測値$\boldsymbol{Z}_1, \ldots \boldsymbol{Z}_n$が与えられたとき、フーリエ周波数を$\lambda_j = \frac{2\pi j}{n} (j = 1, 2, \ldots, [n/2])$とし、$f_t = f(\lambda_t)$の推定量を平滑化ピリオドグラム

$$\hat{f}_{U,t} = \frac{1}{m+1} \sum_{j=-m/2}^{m/2} I_{Z,t+j}, \ t = 1, 2, \ldots, [n/2],$$

によって定義する。ここで$I_{Z,j}$はフーリエ周波数λ_jにおけるピリオドグラム行列とする。一方帰無仮説(1)の下での推定量を$\hat{f}_{R,t} = g(\hat{\theta}_n, R\mathrm{vec}(\hat{f}_{U,t})),\ t = 1, \ldots, [n/2]$とする。ここで$\hat{\theta}_n = (\hat{\theta}_{1n}, \ldots, \hat{\theta}_{v_1 n})'$は$\theta$の推定量である。

いま$K(A)$を複素$r \times r$行列Aを引数とする非負の値をとる関数で、0となるのはAが単位行列に等しいときのみとする。そして統計量$T_n = \sum_{t=1}^{[n/2]} K(M_t)$, を導入する。ここで$M_t = (m_{ab,t}) = \hat{f}_{U,t} \hat{f}_{R,t}^{-1}$とする.

4. 理論的結果

定理1

ある仮定の下で、帰無仮説が正しいとき、ある定数η, σが存在して、$(n/m)^{1/2}(T_n - (n/m)\eta)/\sigma$は$n \to \infty$のとき標準正規分布に収束する。

したがって$\hat{\eta}_n$, $\hat{\sigma}_n$をη, σの一致推定量かつ$\hat{\eta}_n - \eta = o_p((m/n)^{1/2})$とすれば、

$$\hat{T}_n = \sqrt{\frac{m}{n}} (T_n - \frac{n}{m} \hat{\eta}_n) / \hat{\sigma}_n$$

は帰無仮説が正しいとき、標準正規分布に収束する。

定理2

ある仮定の下で、帰無仮説が正しくないとき、$\hat{T}_n$は$(nm)^{1/2}$のオーダーで正の無限大に発散する。

したがって通常のパラメトリック・モデルの検定における、対立仮説のもとでの発散のオーダー$n^{1/2}$より、スピードが速くなる。

非ガウス多変量指数型分布族

椿　広計，岩崎正和
筑波大学大学院ビジネス科学研究科

竹内　啓先生の古希を心からご祝福申し上げます。

0.　はじめに

竹内　啓先生の講義・統計輪講並びに先生との対話の中には忘れられないものが多々ある。本題ではないが、田口玄一先生、G.E.P　Box との懇談会、R.A.Fisher の業績を振り返った講義（特に実験計画法）など印象に残っている。

大学院生時代、先生がリタイアされたら Kendall and Stuart を超えた大著「統計学原典」を執筆すると仰るのを聴いて、一体どんな本になるのだろうかと，わくわくしながら聴いていた。考えてみれば竹内先生は生涯超Ａ級現役なのだろうから、これはどうも難しい話なのかもしれない。ビジネス科学という，統計自身よりは，その対象を統計科学化することに関心がある今日この頃ではあるが、統計の「原典」がどんなものなのかについては，大いに気になる。

さて、1986 年の秋だったと思うのだが、椿の一般化線形モデルに関する学位論文予備審査をお願いに行ったときの、先生の指導も忘れられない：

「椿君、GLIM は統計学のタイホだよ」

「先生、タイホって何ですか？」

「退歩は、進歩の反対だよ」

というものである。小生は、これを聴いて、学位論文に通常の指数型分布族に従わない確率分布に対して GLIM でどのようにモデルを拡張しつつ、分布形情報を回復するかといった章を追加しなければばならなくなったのである。

その後、1990 年代を中心に世間では Liang and Zeger(1986)の一般化推定方程式（Generalized Estimating Equation, GEE）が，個体間，個体内相関を柔軟に取り扱える技法として計量生物学分野を中心に応用上大きなインパクトを与えた．特に、それまでの接近が質的多変量データでも多変量正規モデルの枠組みに押し込めていたものを、一般化線形モデル（Generalized Linear Models, GLIM）の枠組みで取り扱えるようになったものとして，評価されている。同時期、計量経済・金融分野では，既に類似技法として，Hansen(1980)の一般化モーメント法（Generalized Method of Moments, GMM）が一定の地歩を築いていた．先生が、「原典」の中でこれらの方法論をどう評価するかは大いに興味のあるところである。

1．最小二乗推定とモーメント法

竹内先生には、「退歩」と叱られつつも、統計学における正規性の仮定ないしは指数型分布族の仮定の意義について復習したい。これも、考えてみれば竹内先生の講義（現象と行動の統計数理、新耀社）の中で習ったことである。また、1985 年ころ計測自動制御学会主催のリモートセンシングシンポジウムにおいて招待講演をお願いしたとき、「多変量正規性の仮定の本質というのは、回

帰関係の線形性と外れ値がないことだけ」と発言されたのは、安岡善文先生（当時国立公害研究所、現在、東大生産技術研究所教授）をはじめ一部聴衆に竹内先生でなければできない発言としてインパクトを与えた。

さて、ここで強調したいのは、「方便」としての分布形の仮定である。真宗の安田理深氏が「世俗にうそも方便というが、うそが方便でなく、うそをまことにするのが方便である。」と述べているが、これは、「科学法則が、実用便宜のために人間が作ったもの」とする K.Pearson の「科学の文法」ないしは E. Mach の考え方と同一歩調と思われる。

Gauss の最小二乗推定量と Gauss 分布の特徴づけには，密接な関係がある（Gauss,1805, 1820）. Gauss は、当時，確率モデルの母数推定に使われていた Most Probable Method（現在の最尤推定）が最小二乗推定（LSE）と一致する統計モデルとして Gauss 分布を導いたのである．しかも，LSE の標準誤差は、変量が Gauss 分布に従わなくても，線形性，独立性，等分散性の仮定の下で一定かつ線形不偏推定量の中では分散を最小にすることを示したのである（Gauss－Markov 定理）. Gauss 以後，LSE ユーザーにとては，モデルの線形性，誤差の共分散行列の無相関かつ等分散性の仮定に比べれば，分布形の仮定は「擬似的」なものに過ぎなくなった.

一方，Galton（1877）は Quetelet の影響を受け Gauss 分布を擬似的なものではなく，真の生物測定個体間変動の近似モデル（Law of Deviations）として利用した．生物測定値の変動が Gauss 分布で近似できるようなものではないことを指摘したのが，Pearson(1894)であり、Weldon のナポリ蟹の前頭部幅と体長の比のデータが、Gauss 分布の混合状態を観察したものとして近似するのが良いとした。この論文は，混合分布モデルの識別可能性の検討や母数推定方式としての「モーメント法」，すなわちデータ $\mathbf{Y}=(Y_j)$, $(j=1,\ldots,n)$の i 次の標本モーメントと関心のある母数ベクトル θ の関数としての母モーメントに関する方程式、

$$\hat{\mu}_i(\mathbf{Y})=\mu_i(\boldsymbol{\theta}),\ (i=1,\ldots m,\ \dim\theta=m),$$

すなわちモーメント関係に基づく推定方程式を用いて推定を行うことが提唱された論文として大きな価値を持つ．しかし，皮肉にも，この論文は「以下では Gauss 分布を正常な分布（Normal Curve）と呼ぶことにする」と記述し，その後，Gauss 分布を正常分布と考える方便を嘘にしてしまう、きっかけとなり晩年 Pearson は後悔することとなる。

その後、Pearson は 4 次までのモーメント関係に基づく分布族（Pearson 系）の分類に着手した. が，Pearson 系は指数分布族とは限らないので，実はモーメント法との相性は良くない．また，Fisher の最尤法を基調とする推定論における有効性概念や小標本論の発展によって，モーメント法は批判され，Gauss の提起した操作的意味での誤差分布の想定という考え方も閑却される．一方，Fisher(1935)は，明確に推定方程式あるいは Pivotal Quantity の概念を導入し，これは推定方程式論の前駆として位置付けられる.

2．　推定方程式の効率と最も弱い意味での頑健性

Godambe(1960)は，不偏な推定方程式に対して「効率」概念を導入し，最良推定方程式の理論を確立した．Bhapker(1972), Morton(1981)は，それを多母数の推定問題に拡張した．特に，不偏な推定関数 $\mathbf{g}=(g_i(\mathbf{Y},\theta))$, $j=1,\ldots,p>m$ が与えられたとき，最良推定方程式として，

$$\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}}\mathrm{Cov}(\mathbf{g})^{-1}\mathbf{g}=\mathbf{0}$$

が導かれている．Godambe は，モーメント法の効率の低さと最尤法の適用範囲の限界といった弱点を克服する接近として最良推定方程式の理論を構成したとしているが，どちらかというとある性質を共有する統計モデルのクラス内での「効率」を強く意識した体系と言わざるを得ない．この推定方程式論と流れを同じにするのが，Hansen (1980) の一般化モーメント法 (GMM) であり，こちらは推定方程式のような停留方程式ではなく，推定関数ベクトルの L_2 ノルムをその共分散行列の逆行列を計量として最小化

$$\min_{\theta} \mathbf{g}^T \mathrm{Cov}(\mathbf{g})^{-1} \mathbf{g}$$

を行うように母数の推定値を定めるものであり，一種の最小カイ二乗推定量である．上記の計量を用いるのが，2 次形式最小型の推定量の中で推定量の漸近共分散行列を最小にするという意味で最良であることも示されている．

さて一方，Huber(1964)の M 推定量は，明確に最尤法を統計モデルの誤想定のもとで利用する，すなわち統計モデルの「擬似性」を意識しており，標準誤差の評価方法（今日の Information Sandwich）や Min-Max 原理に基づく擬似統計モデル想定として，「最小情報量分布(Least Informative Distribution)」の概念を導入している．GEE を特徴付ける技法の一つである，Information Sandwich は共分散構造誤想定への標準対処として White(1980)の方法と呼ばれ，計量経済分野でも日常的操作となっている．

これに関係して，椿が学位論文作成にあたり注目したのが、Wedderburn(1974)の考え方である。彼は，Nelder and Wedderburn(1972)の GLIM における散らばり母数を持つ指数型分布族の尤度想定は，「擬似尤度（Quasi-likelihood）」と位置付けられる事を主張し，その分布が、分散関数所与の下で指数型分布族が期待値母数に関する「最小情報量分布(Least Informative Distribution)」であることを示した．Wedderburn の擬似スコア関数

$$\mathrm{Q}(Y,\mu)=\frac{Y-\mu}{\mathrm{V}(\mu)}$$

は分散関数 $\mathrm{Var}[Y]=\mathrm{V}(\mu)$，$\mu=\mathrm{E}[Y]$の想定が正しければ，分布形によらず期待値構造に関する母数推定の漸近分散共分散行列が等しくなり（一種のミニマックス性），漸近最良線形推定方程式推定量となることも意味する．つまり，LSE の自然な発展が GLIM の Quasi MLE ということになる．また，Wedderburn の擬似尤度は，分散関数所与の分布族のスコア関数を $\Sigma_{i=1,\ldots n}\, g(\mu_i)(Y_i-\mu_i)$といった一次結合で情報量損失の意味で最良近似したものに相当する(椿,1988, Heyde, 1997)．なお，計量経済分野の議論で注目すべきなのが Gourieroux et al.(1984)の Pseudo likelihood である．すなわち，分布系の想定が誤っていても推定量の一致性は満たされる尤度のクラスとして自然指数型分布族を特定したものである．素朴なモーメント法が，最尤法と一致する分布として指数型分布族を特徴付けたものである．なお，White の方法は、この種の分布形誤想定の下でも議論されている。

3．　多変量擬似尤度とその問題:可積分共分散構造：2 変量ガンマ分布は構築可能か？

McCullagh(1983)は，Wedderburn の擬似スコア関数を

$$\mathbf{Q}(\mathbf{Y},\mu)=\Sigma^{-1}(\mathbf{Y}-\boldsymbol{\mu})$$

と多変量に拡張した．ここで，μ=E[**Y**], Σ=Cov[**Y**]であり，共分散関数Σはμの関数である．擬似尤度あるいは最良推定方程式論では，Cov(**Q**)=- $\mathrm{E}\left[\frac{\partial \mathbf{Q}}{\partial \boldsymbol{\theta}}\right]$という関係（Fisher 情報行列が満たすべき性質）が，共分散関数の誤想定が無い限り性質するというのが，効率ないしは前記ミニマックス性の成立に本質的である．GEE は，

$$\Sigma=\mathrm{diag}[\mathrm{Var}(Y_i)]^{1/2}\mathrm{R}(\rho)\ \mathrm{diag}[\mathrm{Var}(Y_i)]^{1/2}$$

という作業相関構造(Working Correlation) R(ρ)を持つクラスに限定することで，White の方法を適用可能にした点を除けば，McCullagh の擬似尤度そのものである．それにも関わらず McCullagh が GEE に批判的なのは，McCullagh and Nelder(1989)で指摘しているように、一変量の場合と異なり，共分散関数を想定したときに，擬似スコア関数は構成できても，可積分条件が満たされなければ、対応する Dispersive 指数型擬似尤度が存在しないからであり、GEE で用いられている共分散構造の大半はこの状況なのである。

一方、GEE の代わりに Quadratic 指数型分布族に基づく推論を提唱している論文も多い.特に，やはり Gourieroux et. al. (1984)に端を発する Quadratic 指数型分布族, その二値データへの適用と考えられる Zhao and Prentice (1990)らの方法もある。一方，Morris(1982)の 2 次分散関数を持つ指数分布族を多変量に拡張しようとしても，GEE が想定する特定の作業相関構造から導かれる指数分布族が存在するとは考えられない(Kotz et al.,2000，山本，椿，1992).

山本、椿(1992)は、McCullagh and Nelder(1989)が指摘した可積分条件が、次の偏微分方程式、

$$\frac{\partial F}{\partial \mu_i}+F\frac{\partial F}{\partial \mu_j}=0$$

ただし、$F(\mu_i,\mu_j)=\dfrac{Cov(Y_i,Y_j)}{Var(Y_i)}$，E[$Y_i$]=$\mu_i$である。更に、この偏微分方程式の完全解は、

$$F=G(F\mu_i-\mu_j)$$

となることも示した。ここで、G は微分可能な任意関数である。

命題 1：相関係数一定の多変量指数型分布族は、F が定数となるか、$c\mu_j/\mu_i$と表されるかの何れかである。

証明：$\rho^2=F(\mu_i,\ \mu_j)F(\mu_j,\mu_i)=G(F(\mu_i,\ \mu_j)\ \ \mu_i-\mu_j)\ G(F(\mu_j,\ \mu_i)\ \ \mu_j-\mu_i)$を満たすので、

$$\rho^2=G(F(\mu_i,\ \mu_j)\ \ \mu_i-\mu_j)\ G(\mu_j/F(\mu_i,\ \mu_j)-\mu_i) \qquad (1)$$

となり、G、F は定数関数か、G は定数倍、$F(\mu_i,\ \mu_j)=c\mu_j/\mu_i$となる。

従って，分散関数がμ^2に比例する指数型分布族，すなわちガンマ分布は，自然な共分散関数

$\rho\,\mu_i\mu_j$ を持ちつつ相関係数一定を実現できる多変量正規分布以外の唯一の指数型分布族ではないかと予想された．ところが，山本，椿(1992)は,単純な２変量ガンマ型共分散関数,

$$\mathrm{Cov}(\mathbf{Y}) = \phi\ \Sigma = \phi\ \mathrm{diag}(\mu_1{}^2, \mu_2{}^2)^{1/2} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\ \mathrm{diag}(\mu_1{}^2, \mu_2{}^2)^{1/2}$$

は可積分条件を満たさないことを示すと共に，これに積分因子 $(\mu_1\mu_2)^{-\frac{\rho}{1+\rho}}$ を乗じれば可積分となることを示した．この場合，積分因子は，通常の GEE に対しては「重み」として作用することに注意されたい．さらに，Iwasaki and Tsubaki (2002)は，この推定方程式に対応する自然指数型分布が実在すること，すなわちこの推定方程式は擬似統計モデルとして利用可能なことを示した．

命題 2(Iwasaki and Tsubaki, 2002)
共分散関数

$$\mathrm{Cov}(\mathbf{Y}) = \phi\ (\mu_1\mu_2)^{-\frac{\rho}{1+\rho}}\,\mathrm{diag}(\mu_1{}^2, \mu_2{}^2)^{1/2} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\ \mathrm{diag}(\mu_1{}^2, \mu_2{}^2)^{1/2} \qquad (2)$$

に対応する２変量指数型分布族の密度関数をディラックの δ 関数を用いて表示すれば，(3)のようになる．

$$f_{\mathrm{Y}}(y_1, y_2) = \delta(y_1)\delta(y_2)\exp(-\lambda) + \sum_{j=1}^{\infty}\left[\frac{\lambda^j}{j!}\exp(-\lambda)\right]p_j(y_1)p_j(y_2) \qquad (3)$$

ただし，$p_j(y_i) = \dfrac{y_i^{\frac{\rho j}{1-\rho}-1} e^{-\phi y_i/\theta_i}}{\left(\phi/\theta_i\right)^{\frac{\rho j}{1-\rho}}\Gamma\left(\frac{\rho j}{1-\rho}\right)}$ であり（i=1,2），θ_i は，この２変量指数型分布族の自然母数である．また，$\lambda = \dfrac{1-\rho^2}{\phi\rho}\left\{\dfrac{\theta_1\theta_2}{(1+\rho)^2}\right\}^{\frac{\rho}{\rho-1}}$ である．

備考）確率分布(3)は，ρ=0 を除いては，**Y=0** に有限な確率を持つ分布であり，独立な２変量ガンマ分布が同一のポアソン変量によって複合された分布である．また，$\rho\to 0$ の極限では，独立な２変量ガンマ分布に収束する．

命題２の証明は，次の２つの補題を通じてなされる．

補題１　共分散関数(2)に対応する指数型分布族の対数尤度

$$\log\ f_Y(y;\theta) = \frac{y_1\theta_1 + y_2\theta_2 - b(\theta_1, \theta_2, \rho)}{\phi} + c(y_1, y_2, \rho, \phi)$$

が存在するとすれば，そのキュムラント関数$b(\theta_1,\theta_2,\rho)$は，$\dfrac{1-\rho^2}{\rho}\left[\left\{\dfrac{\theta_1\theta_2}{(1+\rho)^2}\right\}^{\frac{\rho}{\rho-1}}-1\right]$となる．

証明は，共分散関数とキュムラント関数の関数関係を偏微分方程式で表現すれば良い．

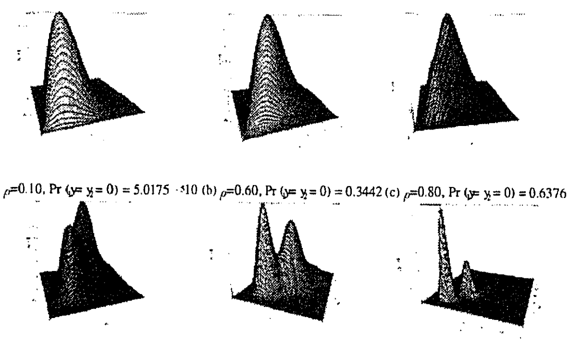
補題2　キュムラント関数$b(\theta_1,\theta_2,\rho)=\dfrac{1-\rho^2}{\rho}\left[\left\{\dfrac{\theta_1\theta_2}{(1+\rho)^2}\right\}^{\frac{\rho}{\rho-1}}-1\right]$に確率分布が対応するとすれば，そのモーメント母関数は，$\mathrm{M}_{\mathrm{Y}}(\mathbf{t})=\exp(-\lambda)\sum_{j=0}^{\infty}\dfrac{\lambda^j}{j!}(1-\dfrac{\phi}{\theta_1}t_1)^{-\frac{\rho j}{1-\rho}}(1-\dfrac{\phi}{\theta_2}t_2)^{-\frac{\rho j}{1-\rho}}$と表される（記法は，命題2と同様）．

このモーメント母関数が命題2で示した分布に対応することは直ちに分かる．

系1　分布(3)の周辺分布は，ガンマ分布のポアソン混合分布であり，これは，Tweedy 分布と呼ばれる指数型分布族であり，この分布の分散関数は$\mathrm{V}(\mu)=\mu^{2-\rho}$となる．

備考）Tweedy 分布は，分散関数から眺めるとガンマ分布（$\mathrm{V}(\mu)=\mu^2$）とポアソン分布($\mathrm{V}(\mu)=\mu$)とを繋ぐ役割を果たしている．

参考までに分布(3)に従う確率変量の密度関数（これ以外に原点に確率が）図1に示す．



(a) ρ=0.10, Pr (y_1= y_2= 0) = 5.0175 ×10 (b) ρ=0.60, Pr (y_1= y_2= 0) = 0.3442 (c) ρ=0.80, Pr (y_1= y_2= 0) = 0.6376

(d) ρ=0.85, Pr (y_1= y_2= 0) = 0.7215 (e) ρ=0.90, Pr (y_1= y_2= 0) = 0.8097 (f) ρ=0.95, Pr (y_1= y_2= 0) = 0.9025

図1　$\mu_1=\mu_2=1$，$\phi=1$の場合の密度関数(3)

分布(3)は，離散変量による複合で非正規変量の相関を設計できることを示唆する一方，分布(3)自体をポアソン変量を複合して相関のある負の二項変量を設計するのに利用することも可能である(Iwasaki and Tsubaki,2003).こうして，構成されたBivariate Negative Binomial分布は，当然のことながら，独立な負の二項分布のポアソン混合分布となり，従来提唱されている多変量負の二項分布よりは，相関構造が柔軟である．

4.　相関構造に関する再考察

命題1を眺めれば，相関係数一定以外に，興味のある相関構造を導出できることが分かる．実際，

$$F(\mu_i,\ \mu_j)=\frac{\varphi(\mu_j)+\mu_j}{\mu_i}$$

と置くことで，相関関数の二乗が$H(\mu_i)\,H(\mu_j)$と変数分離形になることが分かる．このとき，

$$H(\mu_i)=1+\frac{\varphi(\mu_i)}{\mu_i}$$

となり，$G(\varphi(\mu_i))=1+\dfrac{\varphi(\mu_i)}{\mu_i}$を満たすような$G$が存在すれば，この相関構造が可積分となることが分かる．ということは，上手に確率変数の定義域を設定できれば，可積分な相関構造は意外と広いクラスになり得るということと考えられる．

例：多項分布の場合

多項確率の場合，上記の条件は次のようにH，ϕは，それぞれ

$$H(p_i)=\frac{p_i}{1-p_i},\quad \varphi(p_i)=\frac{2p_i-1}{p_i(1-p_i)}$$

となる．従って，$G(\varphi)=\dfrac{\varphi+\sqrt{\varphi^2+4}}{2}$ととることができる．

参考文献

Bhapker, J. C. (1972) On a measure of efficiency in an estimating equation, *Sankhya*, **A34**, 467-472..

Fisher, R. A. (1935) The fiducial argument in statistical inference, *Ann. Eugenics*, **6**, 391-395.

Galton, F. (1877) Typical laws of heredity, *Nature*.492-495,512-514,532-533.

Gauss, C. F.(1809) *Theoria motus corporum coelestum*, *Werke*, **7**.

Gauss, C. F. (1823) Combinationes erroibus minimis obnoxiae, *Werke*, 4, 1-108.

Godambe, V. P. (1960) An optimum property of maximum likeihood estimation, *Ann. Math. Statist.*, **31**, 1208-1211.

Gourieroux, C., Monfort, A. and Trognon, A. (1984) Pseudo maximum likelihood methods: Theory, *Econometrica*, **52**, 681-700.

Hansen, L.(1982) Large sample properties of generalized methods of moments estimators, *Econometrica*, **50**, 1029-1054.

Heyde, C. C. (1997) *Quasi-Likelihood And Its Application*, Springer.

Huber, P. J. (1964) Robust estimation of location parameter, *Ann. Math. Statist.*, **35**, 73-101.

Iwasaki, M. and Tsubaki, H. (2002) Bivariate gamma type quasi-likelihood with constant correlation structure, working paper..

Iwasaki,M. and Tsubaki, H.(2003) Bivariate Negative Binomial Models for Environmental Count Data with Constantly Correlated Covariance structure, Proc. ENBIS 2003 + ISIS3, Barcelona.

Kots, S., Balakrishnan, N. and Johnson, N. L. (2000) *Continuous Multivariate Distributions*, Vol.1, 2nd ed., Wiley.

Liang, K. Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13-22

McCullagh, P. (1983) Quasi-likelihood functions, *Ann. Statist.*, **11**, 59-67.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd ed., Chapman and Hall.

Morris, C. M. (1982) Natural exponential family with quadratic variance functions, *Ann. Statist.*, 10, 65-80.

Morton, R. (1981) Optimal estimating equations with applications to insect development times, *Austr. J. Statist.*, 204-213.

Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized Linear Models, *J. R. Statist. Soc.*, **A135**, 370-384.

Pearson, K. (1894) Contribution to the Mathematical Theory of Evolution, *Phil. Trans. Royal Soc. London*, **A185**, 71-110.

Wedderburn, R. M. W. (1974) Quasi-likelihood function, generalized linear models and the Gauss-Newton method, *Biometrika*, **61**, 439-447.

White, H. (1980) A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Economerica*, **48**, 817-838.

Zhao, L. P. and Prentice, R. L. (1990) Correlated binary regression using a quadratic exponential model, *Biometrika*, **77**, 642-648.

椿広計(1988) 一般線形模型の問題点と擬似尤度の一般化，応用統計学，**17**,1-12.

山本渉，椿広計(1992)一般線形モデルの拡張，応用統計学会 14 回シンポジウム,21-24.

Validity of the expected Euler characteristic heuristic

Dept. of Statistics, Stanford Univ.　Jonathan Taylor
東京大学大学院情報理工学系研究科　竹村 彰通
Faculty of Industrial Engineering and Management, Technion,　Robert Adler

$\widehat{M}$ を C^3 級多様体，$M \subset \widehat{M}$ を $\widehat{M}$ に埋め込まれた部分的に滑らかな多様体，$\widehat{M}$ 上の C^2 確率場を $\widehat{f}(x)$, $x \in \widehat{M}$，そして $\widehat{f}$ の M への制限を $f(x)$, $x \in M$, とする．ここでは $f(x)$ の最大値の裾確率

$$\mathrm{P}\left(\sup_{x\in M} f(x) \geq u\right) \tag{1}$$

を近似するオイラー標数法の正当化を考える．$\widehat{f}$ が期待値 0 分散 1 のガウス場の場合には，(1) 式のオイラー標数法による近似は

$$\widehat{\mathrm{P}}\left(\sup_{x\in M} f(x) \geq u\right) = \sum_{j=0}^{\dim M} \mathcal{L}_j(M)(2\pi)^{-(j+1)/2} \int_u^\infty H_j(r)\, \mathrm{e}^{-r^2/2}\, dr \tag{2}$$

と表される．ただし H_j は j 次の Hermite 多項式，$\mathcal{L}_j(M)$ は f の共分散関数から M に誘導されたリーマン計量に基づく “intrinsic volume” あるいは “Lipschitz-Killing curvature” とよばれる M の幾何量である ([8])．$\widehat{f}$ が期待値 0 分散 1 のガウス場でかつ有限項数の Karhunen-Loève 展開を持つ場合は [6] で調べられている．この場合には写像 $\varphi : \widehat{M} \to S(\mathbb{R}^n)$ と n 次元標準正規確率ベクトル $\xi(\omega) \sim N(0, I_{n\times n})$ が存在して

$$\widehat{f}(x,\omega) = \langle \varphi(x), \xi(\omega) \rangle_{\mathbb{R}^n} = \sum_{j=1}^{n} \xi_j(\omega)\varphi_j(x)$$

と書ける．ただし $S(\mathbb{R}^n) \subset \mathbb{R}^n$ は $\mathbb{R}^n$ の単位球である．チューブ法を用いた議論により [6] ではオイラー標数法の誤差の上界が

$$C(M) \int_{u/\cos\theta_c(M)}^\infty w^{n-1} \mathrm{e}^{-w^2/2}\, dw \tag{3}$$

で与えられることを示した．ただし $\theta_c(M)$ は M の臨界半径 ([7] [3]) である．

別のアプローチとし，$\widehat{f}$ が等方的 (isotropic) なガウス場の場合には，M に関する適当な正則条件 ([2] の Theorem 4.5.2) のもとで，Piterbarg [4] は “double-sum” 法を用いることによりオイラー標数法の誤差の上界が

$$Ce^{-\alpha u^2/2}, \quad \exists \alpha > 1, \tag{4}$$

と表されることを示した．ただし α は明示的には与えられていない．

これらの上界は (2) 式の近似式の各項と比較して，指数オーダーでより小さいものであり，オイラー標数法による近似が良好であることを示している．しかしながら以上の結果は，Karhunen-Loève 展開が有限項である場合や，ガウス場が等方的な場合など，限られた設定でのオイラー標数法の正当化を与えるものであつた．(4) 式に関しては等方性の仮定はきついし，また M に関する正則条件 ([2] の Definition 4.5.1) も面倒な形をしている．(3) 式に関しては，任意のガウス場はその再生核を通じて無限級数の形の Karhunen-Loève 展開を有することは知られてい

る (cf. [1]) が，(3) 式の誤差評価を $n \to \infty$ の極限操作に用いることはできない．実は，上の二つのケースは互いに排反であって，任意の有界な領域 $M \subset \mathbb{R}^k$ に制限された等方的な確率場は決して有限項の Karhunen-Loève 展開を持たないことを示すことができる ([5]).

本論文の目的は一般の設定でオイラー標数法の正当化を与えることである．本論文の結果として例えば次のことが証明される．f が平均 0 分散 1 のガウス場であれば

$$\liminf_{u\to\infty} -u^{-2}\log\left(\left|\mathrm{P}\left(\sup_{x\in T} f(x)\geq u\right)-\widehat{\mathrm{P}}\left(\sup_{x\in T} f(x)\geq u\right)\right|\right)\geq \frac{1}{2}+\frac{1}{2\sigma_c^2(f)} \tag{5}$$

が成り立つ．ここで $\sigma_c^2(f)$ は，有限項の Karhunen-Loève 展開を持つ場合の M の臨界半径の概念を無限次元の場合に拡張したものである．

参考文献

[1] R. J. Adler. *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*, volume 12 of *IMS Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, Hayward, CA., 1990.

[2] R. J. Adler. On excursion sets, tube formulae, and maxima of random fields. *The Annals of Applied Probability*, 10(1):1–74, 2000.

[3] S. Johansen and I.M. Johnstone. Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.*, 18(2):652–684, 1990.

[4] V.I. Piterbarg. *Asymptotic methods in the theory of Gaussian processes and fields.* American Mathematical Society, Providence, RI, 1996. Translated from the Russian by V. V. Piterbarg, Revised by the author.

[5] A. Takemura and S. Kuriki. Some results on geometry of isotropic and spherically isotropic smooth gaussian fields. In preparation.

[6] A. Takemura and S. Kuriki. Maximum of Gaussian field on piecewise smooth domain: Equivalence of tube method and Euler characteristic method. *Ann. of Appl. Prob.*, 12(2):768–796, 2002.

[7] Akimichi Takemura and Satoshi Kuriki. Tail probability via the tube formula when the critical radius is zero. *Bernoulli*, 9(3):535–558, 2003.

[8] J. Taylor and R. J. Adler. Euler characteristics for Gaussian fields on manifolds. *Annals of Probability*, 31(2):533–563, 2003.

Inference on general unit-root cointegration and associated computational methods

Yuzo Hosoya* and Taro Takimoto†

Graduate School of Economics and Management, Tohoku University

The paper presents a practicable framework to deal with theoretical as well as computational aspects of statistical inference on general cointegration models. With general unit-root cointegrated processes which includ the vector ARMA model as a special instance, the paper proposes the Whittle-likelihood ratio test for cointegration rank and provides a limit theory useful for the asymptotics of that test. Section 2 derives the Granger representation theorem (Theorem 2.1) by employing the adjoint matrix of AR part of the generating mechanism of the process. The approach turns out to be useful for examine the root-condition of the DGP. In Section 3 the paper gives a central limit theory (Theorem 3.1) and an allied invariance principle (Theorem 3.2) based on a set of assumptions introduced by Hosoya and Taniguchi (1982,93) which requires the Martingale-difference and/or conditional homoscedasticity only asymptotically for the innovation process involved. The limit theory of the paper enables application of the cointegration rank test to a wider class of processes than those investigated in the literature so far. Section 4 considers the Whittle likelihood in the general set-up of the paper. Extending Johansen's asymptotic theory of the likelihood ratio (LR) test for cointegration rank, Section 5 presents Theorem 5.1 which shows that the log Whittle-likelihood ratio has asymptotically the distribution of the trace of a matrix composed of functionals of a multivariate Brownian motion and a Brownian motion as in Johansen (1995). Corollary 5.1 showing how the asymptotic result is modified by testing against the alternative hypothsis of trend stationarity. To derive the asymptotic distribution, the paper employs implicitly the asymptotics theory by Hosoya (1997) on the Whittle-likelihood based inference for stationary long-range dependent processes. Based on our limit theory, Section 6 proposes a feasible numerical method for evaluating the maximum Whittle likelihood estimator and the joint p-values by means of the idea of Hosoya (1989)'s generalized likelihood ratio test for hierarchical statistical models. Testing the cointegration rank is conducted by means of simultaneous comparison of a null-hypothesis rank against a nested set of alternative rank in which the full-rank explicitly involves a trend term. The most delicate issue involved in the numerical evaluation for finite samples is that the maximum of the Whittle likelihood is not necessarily attained inside the admissible parameter space where the root-conditions

*The paper is partially supported by Grant-in-Aid for Scientific Research (C)(2)15530136.

†Research Fellow of the Japan Society for the Promotion of Science, The Doctor program (The Secondary Course). The paper is partially supported by Grant-Aid for JSPS Fellows 15006983.

required by the company matrix of any model of given rank are satisfied. We proposed an adjustment method in case they are violated and also a numerical iteration procedure for maximization of the likelihood in each step of which the root requirements of estimated coefficients are implemented so that the estimate obtained in each stage of iteration is guaranteed to fall inside the admissible parameter space.

In Section 7 the empirical analysis is conducted in view of identifying the VARMA model pertinent to the trivariate U.S. interest-rate series on the basis of the paper's estimation and testing method. In comparison to Reinsel and Ahn (1992), Yap and Reinsel (1995), and Takimoto and Hosoya (2003), this paper investigated the same three U.S. series, the Federal Fund rate, 90-day and 1-year Treasury Bill rates, over the period 1960:1-1979:12 and it turns out that our simultaneous WLR test supports most the VAR trend-stationary model among the nested set of VARMA models in contrast to the existing literature all of which seem to indicate results more favourable to some reduced rank unit-root models rather than the constant-mean stationary model. The different outcome may possibly be due to our employment of the trend-stationary model as the full-rank alternative. Under the trend-stationary alternative hypothesis we conclude that these series are trend-stationary processes individually, but it seems to depend on the character of original data sets we use whether constant-mean or trend-stationary hypothsis is appropriate as the full-rank alternative. Recently, Andreou and Spanos (2003) discussed whether the macroeconomic series are trend or difference stationary by bringing in the concept of statistical adequacy which is concerned with the empirical validity of the probabilistic assumptions underlying a statistical model.

References.

Andreou, E. and Spanos, A. (2003). *Econometric Reviews*, 22, 217-37.

Hosoya, Y. (1989). *Journal of the Royal Statistical Society*, B, 51, 435-47.

Hosoya, Y. (1997). *The Annals of Statistics*, 25, 105-37.

Hosoya, Y. and Taniguchi, M. (1982). *The Annals of Statistics*, 10, 132-53: Correction, 1993, *The Annals of Statistics*, 21, 115-7.

Reinsel, G.C. and Ahn, S.K. (1992). *Journal of Time Series Analysis*, 13, 353-75.

Takimoto, T. and Hosoya, Y. (2003). Submitted to The Japanese Economic Review.

Yap, S.F. and Reinsel, G.C. (1995). *Journal of the American Statistical Association*, 90, 253-67.

準完全分離データの処理について

長崎大学・医歯薬学総合研究科　　柴田義貞

1　背景

表1は、チェルノブイリ原発事故後にベラルーシ、ロシア、ウクライナ3国のチェルノブイリ原発周辺地域で激増した小児甲状腺がんの主たる原因が事故で放出された放射性ヨウ素への内部被曝によることを証明するためにチェルノブイリ原発から半径 150 km 以内に位置する地域で行った検診結果を示している[1]。このデータの解析で経験したいくつかのことがらについて以下に紹介する。

表1．1998年2月から2000年末までに検診したチェルノブイリ周辺の子供21 601人における年齢・出生時期・性別の甲状腺がん有病率（症例数/検診数）

検診時年齢（歳）	群（生年月日）					
	Ⅰ（01/01/87-31/12/89）		Ⅱ（27/04/86-31/12/86）		Ⅲ（01/01/83-26/04/86）	
	男	女	男	女	男	女
8	0/67	0/75				
9	0/651	0/667				
10	0/1383	0/1287				
11	0/1523	0/1457	0/156	0/149	0/28	0/26
12	0/897	0/858	0/501	0/406	0/333	1/345
13	0/305	0/302	0/437	1/437	1/989	6/906
14			0/164	0/159	2/1449	6/1460
15					3/1339	9/1319
16					0/579	2/754
17					1/93	0/100
計	0/4826	0/4646	0/1258	1/1151	7/4810	24/4910

2　準完全分離データ

放射性ヨウ素への内部・外部被曝の程度は、Ⅲ群がもっとも大きく、次いでⅡ群であり、Ⅰ群は被曝していない。したがって、群は、放射性ヨウ素への被曝に関しては、順序尺度になっている。

Ⅰ群、Ⅲ群の検診者数は、男女ともほぼ同数で、5000人弱である。そして、放射性ヨウ

素に被曝していないⅠ群には甲状腺がんは1例もいなかったが、放射性ヨウ素に直接被曝したⅢ群には31例（男7例、女24例）の甲状腺がんが見つかった。また、胎内被曝のⅡ群2409のうち1人に甲状腺がんが見つかっている。そこで、表1を提示して、チェルノブイリ事故後に周辺地域で激増した甲状腺がんの原因は、明らかに、放射性ヨウ素への内部・外部被曝であると結論付けた論文を投稿したところ、3人の審査員（1人は統計家）から、統計解析を行うように、とくに統計家からは、p値を示すようにとの指摘を受けた。

説明変数は、性、検診時年齢、および群であるから、甲状腺がんの有病率をpとすると、標準的な解析は、次のようなロジスティックモデルに基づくものである。

$$\log [p/(1-p)] = \beta_0+\beta_1 s+\beta_2 a+\beta_3 x_1+\beta_4 x_2 \qquad (1)$$

ここに、s=0 (女)、1 (男)；a は検診時年齢（歳）；x_1=1（Ⅲ群）、0（その他）；x_2=1（Ⅱ群）、0（その他）であり、βは推定すべき未知母数である。

ところで、表1から分かるように、このデータは、モデル（1）の下では x_1、x_2 に関して準完全分離（quasicomplete separation）[2,3]であり、β_3、β_4 の最尤推定値は無限大となり求まらない。

そこで、群を表示する変数gを導入し、次のようなモデルの下で解析した。

$$\log [p/(1-p)] = \beta_0+\beta_1 s+\beta_2 a+\beta_3 g \qquad (2)$$

ただし、群が順序尺度であることから、g=0 (群Ⅰ)、g=1 (群Ⅱ)、g=2（群Ⅲ）とした。

公表論文に対してかなりの反響があったが、その中にモデル（2）に基づく解析は不適切で、したがって結論も疑わしいとするletterが翌月投稿され、編集者からの照会に応じて回答を投稿した。そこでは、α $(0 \leqq \alpha \leqq 1)$ について、g=0 (群Ⅰ)、g=α（群Ⅱ)、g=1（群Ⅲ）として解析した結果を紹介し、さらに、1991年から1996年までに行った12万人の検診結果に言及して、結論を訂正する必要のないことを述べた[4]。

3　結語

何事も仮説検定の結果をみないと判断できないという、治療困難な疾病が世界的に蔓延している。

参考文献

1. Shibata Y, Yamashita S, Masyakin VB, Panasyuk GD, Nagataki S: 15 years after Chernobyl: new evidence of thyroid cancer. Lancet 358, 1965-1966, 2001

2. Silvapulle MJ: On the existence of maximum likelihood estimators for the binomial response models. J R Statist Soc B 43, 310-313, 1981

3. Albert A, Anderson JA: On the existence of maximum likelihood estimates in logistic regression models. Biometrika 71, 1-10, 1984

4. Shibata Y: Thyroid cancer 15 years after Chernobyl. Lancet 359, 1947, 2002

交互作用の多重比較法とその様々な応用

明星大学・理工　広津 千尋

1. 因子の種類と交互作用

交互作用解析は様々な統計分析の鍵と言える．それは単に分散分析モデルに限らず，分割表解析や多項分布比較など広範な応用を持っている．それにもかかわらず，多くの場合単に無交互作用の検証を目的とした総括検定の行われることが多い．しかしながら，交互作用はそれを構成する因子の種類によって解析の目的や解析結果への対応の取り方が種々異なり，とても通り一遍の総括検定では済まされない．

2 元配置分散分析モデルでは，通常両方の因子が制御因子である場合を対象とし，最適水準組合せを求める手法のみが扱われている．しかしながら，例えば一方の因子が標示因子の場合，標示因子は最適選択の対象ではなく，標示因子の水準ごとに制御因子の最適水準を求めることが目的となる．それには標示因子の水準の多重比較が有用であり，2 元表の行あるいは列ごとの多重比較という手法が要請される．その方法は，制御因子同士の場合にも，最適組合せを単にセル平均の比較から決定する方法に替えて用いることができる．

次に，一方の因子が温度その他の使用環境，あるいは起炎菌，重症度のように実験室では同定できても実際の現場では同定できずノイズとして働く変動因子(誤差因子)である場合を考える．この場合は，変動因子のばらつきを超えて有意に優れる制御因子の水準を探す，あるいは変動因子の変化に対して安定した特性値を与える制御因子の水準を見つける等のことが目的となる．そのためにも制御因子の水準ごとのレスポンスの多重比較が有用な手法となる．

2. 行（列）ごとの多重比較法定式化

繰返しのない 2 元配置モデル

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}, \qquad i = 1, \cdots, a; j = 1, \cdots, b$$

を考える．繰返しがある場合は，ここでの取り扱いの他に純粋な誤差分散推定量が得られるという違いがある．行ごとの多重比較の基となる交互作用要素を

$$\mathbf{L}(m; n) = \frac{1}{\sqrt{2}} P_b (\boldsymbol{\mu}_m - \boldsymbol{\mu}_n) \tag{1}$$

で定義する．$\boldsymbol{\mu}_i' = (\boldsymbol{\mu}_{i1}, \cdots; \boldsymbol{\mu}_{ib})$ は行の第 i 水準の応答ベクトルで，P_b'は$P_b P_b' = I_{b-1}$，$P_b' \mathbf{j}' = 0'$　$(\mathbf{j} = (11 \cdots 1)')$，を満たす(b−1)×b 直交行列である．従って $L(m;n)=0$ は，$\boldsymbol{\mu}_m - \boldsymbol{\mu}_n = k\mathbf{j}$，すなわち，第 m 水準と第 n 水準の応答が平行であることを意味する．(1) 式は水準をいくつかプールして群を構成した場合の群間交互作用 $L(G_m; G_n)$に自然に拡張される．この 2 群間交互作用要素はさらに多群間の交互作用要素

$$L(G_1; G_2; \cdots; G_m) = P_b'(\gamma_1 \boldsymbol{\mu}_1 + \gamma_2 \boldsymbol{\mu}_2 + \cdots + \gamma_a \boldsymbol{\mu}_a),$$
$$\gamma_i = \lambda_l \qquad if \qquad i \in G_l, \quad \sum q_l \lambda_l^2 = 0, \sum q_l \lambda_l^2 = 1 \tag{2}$$

に拡張される．ここで交互作用要素の大きさを表す統計量（群間の二乗距離）を次のように定義

する．まず，(1) 式に関しては，単にμ_iに $\mathbf{y}_i$を代入すればよい．次に (2) 式は未定係数λ_lを含んでいるので，

$$S(G_1;G_2;\cdots;G_m)=\max_{(\lambda_1\cdots\lambda_l)}\left\|P_b'(\gamma_1\mathbf{y}_1+\gamma_2\mathbf{y}_2+\cdots\gamma_n\mathbf{y}_n)\right\|^2 \tag{3}$$

を群間二乗距離とする (Hirotsu, 1991)．ここで，(3) 式の統計量は明らかに最大統計量

$$\chi^2_{\max}=\max_{\gamma}\left\|(\gamma'\otimes P_b')\mathbf{y}\right\|^2=\max_{\gamma}\left\{(\gamma_1\mathbf{y}_1+\cdots+\gamma_a\mathbf{y}_a)'P_bP_b'(\gamma_1\mathbf{y}_1+\cdots+\gamma_a\mathbf{y}_a)\right\} \tag{4}$$

で上から押えられる．しかるに$a\geq b$の時$\chi^2_{\max}$の帰無分布は Wishart 行列$\sigma^2W(I_{b-1},a-1)$の最大固有値の分布に等しい．そこで，行間二乗距離を基本に，Ward 法類のクラスタリングを進め，最後に一般化群間二乗距離を Wishart 分布で評価することにより，分類の有意性を判定することができる．実際には攪乱母数であるσ^2を消去するための統計量で除した統計量を用いる．

3. 単調性仮説の下での推測

行または列の水準に自然な順序がある場合に，2 章の考え方は様々に拡張される．例えば，列に自然な順序があり，第 i 行と i' 行の間で

$$\mu_{i1}-\mu_{i'1}\leqq\mu_{i2}-\mu_{i'2}\leqq\cdots\leqq\mu_{ib}-\mu_{i'b} \tag{5}$$

で示されるような差異（差が単調に拡大していく）をとくに検出したい場合がある．この場合，(5) 式は差分行列を用いて，$D_b'(\boldsymbol{\mu}_i-\boldsymbol{\mu}_{i'})\geqq 0$ と表されることから，(2)，(3)，(4) 式において P_b'を

$$P_b^{*}{}'=C(D_b'D_b)^{-1}D_b',\qquad C:\text{基準化定数の対角行列}$$

で置き換えた統計量が有用になる (Hirotsu, 1982, 1991)．一方，同じ設定で，列の群分けをしたい場合は

$$\max_j\left\|(P_a'\otimes\mathbf{p}_j')\mathbf{y}\right\|^2,\qquad \mathbf{p}_j:P_b^{*}\text{の第j列}$$

が提案される．これは列の順序に従ってどこで分割するのが最も大きな差異を示すかを探索する統計量になる．これらの考え方はさらに，凹凸の形状パターンの差異を抽出する目的で拡張され，具体的に血圧日内リズム解析へ応用される(Hirotsu et al., 2003).

さらに行と列の両方に順序がある場合は，

$$\max_i\left\|(\mathbf{p}_i'\otimes P_b^{*}{}')\mathbf{y}\right\|^2(\mathbf{p}_i:P_a^{*}\text{の第i列})\ ;\ \max_i\left\|(\mathbf{p}_i'\otimes\mathbf{b}')\mathbf{y}\right\|^2(\mathbf{b}:\text{スコアベクトル})\ ;\ \max_{i,j}\left\|(\mathbf{p}_i'\otimes\mathbf{p}_j')\mathbf{y}\right\|^2$$

のようないろいろな統計量が提案できる．これらの統計量は順序分割表にもそのまま適用でき，臨床試験データ解析をはじめいろいろ有用な場面がある．

参考文献

(1) Hirotsu, C. : Use of cumulative efficient scores for testing ordered alternatives in discrete models. Biometrika 69, 567-577. (1982)
(2) Hirotsu, C. : An approach to comparing treatments based on repeated measures. Biometrika 78, 583-594. (1991)
(3) Hirotsu, C., Ohta, E., Hirose, N. and Shimizu, K.: Profile analysis of 24-hours measurements of blood pressure. To appear in Biometrics. (2003)

Randomizationについて

竹内　啓

1

Randomizationは20世紀に発達した数理統計学の多くの手法の中でも，主要なものの一つということができよう．その一種と考えられる「くじ引き」は，ほとんど太古の昔から行われて来たから，その前史は古いといわねばならないが，統計学の方法の中に意識的に取り入れられたのは20世紀になってからであった．そうして20世紀に発達した統計的方法の中で，それはR.A.Fisherの実験計画法におけるrandomization，J.Neymanのrandom sampling，そしてコンピュータの発展とともに発達したMonte Carlo simulation，さらには最近のB.EfronによるBootstrap法に至るまで，いろいろな点で重要な位置を占めていることに注意すべきである．

しかし，randomizationの意味そのものについては，なおあまりすっきりしないものが残っているように思われる．私はこの問題についてほとんど50年にわたって，いろいろな方向からその時々に考えてきたことがあるので，以下それらについて概略を話して，今後の検討を待ちたいと思う．

2　randomizationの正当性

基本的な問題は「randomizationはなぜ正当な方法なのか」ということである．それはよりさかのぼれば「くじ引きはなぜ公平か」という問題に帰着する．これについて私はずっと前に一つの小文を書いたことがある．すなわち「くじ引き」は結果としては必ず不公平になる．もし結果も公平にすることが可能ならば「くじ引き」をする必要はない．くじ引きが行われるのは，分割不可能な少数のものを同等な権利を持つ大勢の人に分配するような「平等な分配」が不可能な場合である．その場合くじ引きの結果としては「当たった」人と「当たらなかった人」では不平等が生ずることになる．しかし「公正なくじ引き」であったと認められれば，当たらなかった人も「運が悪かった」としてあきらめるであろう．その場合結果は不平等になったとしても，事前には皆が「平等のチャンス」を与えられていたのだから，結果は平等でなくても「公平」あるいは「公正」であったとされるのである．しかし「平等はチャンス」というのは考えてみれば実体のないものである（ここで大数法則を引き合いに出して，同じようなくじを何度もくり返せば皆が平等に「当たる」ようになるといっても無意味である．同じことについての「くじ引き」はふつう一回しか行われないし，同じことが何回もあるならくじ引きなしでむしろ順番に配った方が結果は公平になる）．そうするとくじ引きの結果について，皆が納得するのは「くじを引く」（あるいはそれと同等の）手続きが「公正」だと認められているからにほかならない．もしそれが誰かが勝手に決めたことであったら，人々はそれを受け入れないであろう．とくに「手続き」の「公正」と，結果の

「公平」に乖離が生じているわけであり，そうして前者が保証されれば，1,2 は後者が維持されなくても，それを受け入れるということなのである．

これは一つの重要なポイントであると思う．ふつう手続きの公正ということと，その手続きを適用して生じた結果の公正ということは密接に結びついており，そうして前者は後者によって根拠づけられる．すなわち公正な結果を生み出すような手続きが公正なものであると考えられている．しかしながら，不特定多数回の場に適用される手続き，あるいはルールの公正ということと，個々の具体的な場における公正ということは必ずしも結びつかない場合は実は少なくない．法や規則が適用される場合,「そのように決まっているから」ということであれば，それを個々の場合に適用する場合の結果について争われることは稀である．その場合「法」や「規則」の決め方が妥当であれば，その適用の結果について，公正であることが明確でなくても，あるいは場合によっては若干不公正と思われても，受け入れられるのである．

randomization もこれと同じ意味を持つ．そこで問題になるのは，くじ引きの場合の公平性の代わりに「客観性」である．randomization は「客観的に偏りのない方法である」というのはその手続きが客観的であり,「偏りがない」というだけであって，結果はいずれにしても何らかの方向に偏るのである．しかし手続きが客観性を持つことによて，結果の偏りは容認される．そうしてその偏りの程度は「確率計算」によって保証されるが，それもまたくり返し同じ方法を適用した場合の比率を計算するという「ルール」によって求められたものであって，個々のケースにそのままあてはめられるものではない．

「推測のルール」という考え方に激しく反撥し，具体的に与えられた標本からの「推測」という考え方を強く主張した R.A.Fisher が，他方では randomization を推進したのは矛盾のようにも思われるが，Fisher の思想を理解する上での一つの問題点である．

さらにさかのぼって考えれば，Fisher の「仮説的無限母集団からのランダムサンプル」という考え方自体が，現実の観測値を「自然」が選び出すルールと想定したものであり，そこから結果の不均一性（つまり確率変動）が生ずるにもかかわらず，統計的推測の方法の「客観的妥当性」が保証されることになっているといってもよいのである．

ところで「くじ引き」が行われる場合「不確実性」つまり結果の不平等性自体を良しとする，あるいはその途中の「スリル」を楽しむ，純粋の賭博のような場合を除けば「手続きの公正」ということと「結果の不平等あるいは不公平」との間には，なるべくギャップが小さいことが望ましいであろう．すなわち可能な限り結果の「不公平」は少ない方がよいということになる．同様に完全に偏りのない標本を構成することが不可能な場合，randomization によって，平均的，あるいは確率的に偏りのない標本を作るとしても，それは結果としてもなるべく偏りの少ないものであることが望ましいであろう．

3　因子ランダム化実験

randomization に関連して，私が最初に「異常な」状況に直面したのは，田口玄一の「確率的対応法」を知ったときであった（1959 年夏三河田口での PSG サマーセミナー）．それは次のようなものである．2^n 回の実験ですべて 2 水準の k 個の因子について実験したい（この中には交互作用項もふくまれる）．そのとき $k < 2^n - 1$ ならば 2^n 直交表の各列に適当に k 個の因子を配分して実験を行えばよい．しかし $k \geq 2^n - 1$ の時は，いくつかの因子は交錯してしまうから，完全な実験は不可能になる．この場合 $2k < 2^n - 2$ ならば，次のようにす

ればよい．$k = p + q \quad p < 2^n - 1 \quad q < 2^n - 1$ として因子を p 個，q 個の2組に分けて，2^n 直交表を2つ用意して，一方の列に p 個の因子を，他方に q 個の因子を割りつける．2個の直交表の 2^n 個の行をランダムに組み合わせ，各行に対応する因子の組み合わせを実験して 2^n 個の観測値を得る．得られた結果について，第1組の因子の効果について推定，検定を行うときには，第2組の因子の効果はランダムな誤差とみなして，一括して扱い，通常の分散分析の方法を第一の直交表にもとづいて適用する．次に第2組の因子について解析するときは第1組の因子効果はランダムな誤差とみなして第2の直交表にもとづいて解析するというのである．これを最初に聞いた時，因子効果の自由度が標本の大きさを越える配置などというものは認められるはずがないと思ったが，実はしばらく計算をしている間にそれは論理的に全く正当であることがわかったのであった．そこでこれについて一連の論文を書くことになった．

状況を一般化してモデル化すれば次のようになる．y を実験観測値の n 次元ベクトル，θ を母数の p 次元ベクトル，u を誤差ベクトルとすると，θ と y は $n \times p$ 配置行列 design matrix$X$ によって

$$y = X\theta + u$$

と表される．ここで通例のように $E(u) = 0, \quad E(uu') = \sigma^2 I$ と仮定すると，$ranX = p < n$ ならば θ は最小2乗法によって

$$\hat{\theta}^* = (X'X)^{-1}X'y$$

によって推定され，また σ^2 は

$$\hat{\sigma}^2 = Y'(I - X(X'X)^{-1}X')y/(n-p)$$

と推定される．そうして θ について仮説検定や，区間推定も教科書通りに行うことができる．しかし $rankX = p > n$ の場合には θ の推定は不可能になる．

ところが X がランダムに決められる，すなわちそれが確率変動で，かつその分布が既知とすれば $p > n$ の場合でも推定が可能になる．

それは $z = X'y$ とすると

$$E(z) = E(X'X)\theta$$

となるが，$p \times p$ 行列 $E(X'X)$ の階数は（$X'X$ の階数はつねに n 以下であるが）p になるようにすることができる．

そのときに

$$\hat{\theta}_r = E(X'X)^{-1}y = E(X'X)^{-1}X'y \tag{1}$$

とすれば $E(\hat{\theta}_r) = \theta$ であるから $\hat{\theta}_r$ は確かに θ の不偏推定量（ベクトル）になっている．すなわち標本の大きさより多い母数をすべて推定することが可能である（？）．$\hat{\theta}_0$ の分散共分散行列も

$$V(\hat{\theta}_r) = E(X'X)^{-1}E(X'X\theta\theta'XX')E(X'X)^{-1} - \theta\theta' + \sigma^2 E(X'X)^{-1} \tag{2}$$

と表すことができる．このような推定量は $p < n$ の場合にも計算することができる．そのとき最小2乗推定量の分散共分散行列は

$$V(\hat{\theta}^*) = \sigma^2 E((X'X)^{-1})$$

となる．$\hat{\theta}_r$ と $\hat{\theta}^*$ の分散共分散行列を比較すると前者は θ をふくむ第1項だけ大きくなるが，逆に第2項は，つねに

$$E((X'X)^{-1}) \geq E(X'X)^{-1} \qquad (\geq \text{ は左辺}-\text{右辺が非負定符号になることを意味する})$$

であるから，後者の方が大きい．すなわち $(|\hat{\theta}|)$ が小さければ $\hat{\theta}_r$ はよい推定量になる．

さらに θ について，事前にそれがほぼ θ_0 に等しいと想定される場合には，$\theta - \theta_0$ を推定してそれに θ_0 を加えるという形で

$$\hat{\theta}_0 = E(X'X)^{-1}X'(y - X\theta_0) + \theta_0$$

という形の不偏推定量が得られる．この推定量の分散共分散行列は (2) において θ を $\theta - \theta_0$ でおきかえたものになる．そこで $\theta = \theta_0$ ならば $V(\hat{\theta}_0) = \sigma^2 E(X'X)^{-1}$ となる．

n が正規分布に従う場合には，θ に関する Fisher 情報量行列は $E(X'X)/\sigma^2$ になるから $V(\hat{\theta}_0)$ はその逆行列に等しく，したがって $\hat{\theta}_0$ は局所最良不偏推定量となる．すなわち $\hat{\theta}_0$ はある意味でよい推定量である（！）．

σ^2 の推定については

$$\begin{aligned} E\{\hat{\theta}_r E(X'X)\hat{\theta}_r\} &= traceE(X'X)E(\hat{\theta}_r\hat{\theta}_r') \\ &= trace\{E(X'X\theta\theta'X'X)E(X'X)^{-1} + \sigma^2 I\} \\ &= \theta' E(X'XE(X'X)^{-1}X'X)\theta + p\sigma^2 \end{aligned}$$

であり，また

$$E(y'y) = \theta' E(X'X)\theta + n\sigma^2$$

であるから，$E(X'XE(X'X)X'X) = kE(X'X)$ k は定数という形になっていれば，

$$\hat{\sigma}_r^2 = \frac{1}{n - p/k}(y'y - \frac{1}{k}\hat{\theta}_r' E(X'X)\hat{\theta}_r)$$

という形で推定することができる．一般の場合には σ^2 の不偏推定量は求められない．

4 Random design の有用性

$n < p$ の場合，いわゆる over-saturated design については，ちょうどその当時（1950年代末）Satterthwaite の提案があり，一時論議を巻き起こしたが，やがて田口氏自身も確率対応法についてあまりいわれなくなり，over-saturated design は，ほとんど受け入れられることなく終わってしまって，それとともに random design についての関心が消えてしまったのは残念であった．

というのは random design の価値は over-saturated design のような極端な場合に限るわけではないからである．over-saturated design も全く無用だとは考えられない．それは多数

の因子が存在する中で，有意な効果を持つものは少数であることがあらかじめわかっているが，その中のどれが有意であるかはわからないような場合には有用であると思われるからである．

それよりもむしろ，$p < n$ であるが $X'X$ を最適にすることが不可能な場合，X をランダムに選んで $E(X'X)$ を最適な形にすることは可能であり，しかも X のばらつきはなるべく小さくなるようにすれば，ほとんど最適に近い結果が得られると期待されるからである．

例えばもっとも簡明なのは $X'X = \sigma^2 I_p$ とすることが望ましい場合である（直交配置）．X に対する制限条件の下で，これが不可能な場合に $E(X'X) = \sigma^2 I_p$ とすることが可能な場合は少なくない．確率対応法はその一例であるが，田口氏の場合 X のばらつきが大きすぎるという欠点があるように思われる．

より実際的なのは混合水準の場合，例えば $p^m \times q^e$（p 水準の主効果因子 m 個，q 水準のもの e 個）の場合である．このとき $p^m \times q^e$ 個の行からなる混合水準直交表を作ることは，一般に困難である．このような場合には q^e 個の p^m 直交表を並べたものと，p^m 個の q^e 直交表を並べたものをそれぞれ用意し，その行をランダムに組み合わせればよい（確率対応法の一般化？）．

あるいは不完備ブロック配置において，v 個の品種を大きさ k の b 個のブロックで実験する場合，$br/v = r$ が整数ならば，一つの品種のくり返し数を r とすればよい．しかしこの場合でも完全な釣り合い型配置 balanced incomlete design が存在するためには

$$\lambda = r(k-1)/(v-1)$$

が整数でなければならないし，またその条件が満たされても BIB が存在するとは限らない．

このような場合に各品種のブロックへの割りつけをランダムに行って，任意の 2 つの品種の会合数（2 つが同じブロックに入る回数）の期待値が上記の λ に一致するようにすることが考えられる．このような randomBIB と呼ぶべきものを作ることは容易である．またその解析も BIB とほとんど同じように進めることができる．すなわち y を実験観測値の n ベクトル，u を一般平均，θ を品種効果を表すベクトル，τ をブロック効果を表す b ベクトル，u を実験誤差を表す n ベクトルとすれば（u については $E(uu') = \sigma^2 I$ また正規性を仮定しておく）．

$$y = u\mathbf{1}_n + T_\theta + B_\tau + u, \qquad E(u) = 0 \quad E(u^2) = \sigma^2 \tag{3}$$

と表すことができる．ただし $\mathbf{1}$ はすべての成分が 1 である n ベクトルを表す．また T，B のすべての要素は 0 または 1 であり，かつ

$$\begin{aligned} T'1_n &= r\mathbf{1}_v, \qquad T'T = rI_v, \\ B'1_b &= k\mathbf{1}_b, \qquad B'B = kI_b, \\ 1_v'\theta &= 0, \qquad 1_b'\tau = 0, \end{aligned}$$

を満たすとする．このとき最小 2 乗法にもとづく正規方程式は

$$\begin{aligned} &n\hat{u}' = \mathbf{1}_n' y = n\bar{y}, \\ &r\hat{u}\mathbf{1}_v + r\hat{\theta} + T'B\hat{\tau} = T'y, \\ &k\hat{u}\mathbf{1}_b + B'T\hat{\theta} + k\hat{\tau} = B'y. \end{aligned}$$

これから

$$\hat{\theta} + \frac{1}{r}T'B\hat{\tau} = \frac{1}{r}(T'y - \bar{y}\mathbf{1}_b), \tag{4}$$

$$\hat{\tau} + \frac{1}{k}B'T\hat{\theta} = \frac{1}{k}(B'y - \bar{y}\mathbf{1}_b), \tag{5}$$

$$(I - \frac{1}{kr}T'BB'T)\hat{\theta} = \frac{1}{r}T'(I - \frac{1}{k}BB')(y - \bar{y}\mathbf{1}) = Q, \tag{6}$$

を得る. この右辺 Q は adjusted yield vector と呼ばれるものである.　また

$$\Lambda = TB'B'T$$

は会合数行列である．BIB の場合は

$$\Lambda = (\gamma - \lambda)I + \lambda\mathbf{1}_v\mathbf{1}_v'$$

となっているから，$1_v'\hat{\theta} = 0$ に注意すれば $\{1 - (\gamma - \lambda)/kr\}\hat{\theta} = Q$
結局

$$\hat{\theta} = \frac{k(v-1)}{v(k-1)}Q \tag{7}$$

を得る．random BIB の場合にも

$$\hat{\theta} = (I - \frac{1}{kr}\Lambda)\theta + w$$

w は誤差ベクトルという形になるから，$E(\Lambda) = (r - \lambda)I + \lambda 1_v 1_v'$ に注意すれば，やはり (7) の形で θ の推定量を与えることができる. ただし今度は

$$\hat{\theta} = \theta - \frac{v-1}{rv(k-1)}(\Lambda - (r-\lambda)I - r1_v1_v')\theta + \frac{k(v-1)}{v(k-1)}w$$

となって，右辺の第 2 項の分だけ $\hat{\theta}$ の分散が大きくなる.
　このことから BIB と同様にして品種に対応する平方和 $\hat{Q}'Q$ を作ると，その期待値が

$$E(\hat{\theta}Q) = (v-1))\{(1+f)\sigma_\theta^2 + \sigma^2\},$$
$$\sigma_\theta^2 = (\theta'\theta)/(v-1),$$
$$f = \sigma_\lambda^2/v\bar{\lambda}^2, \qquad \sigma_\lambda^2 \text{ は } \Lambda \text{ の要素 } \lambda_{ii'},\quad i \neq i' \text{ の分散},$$

となることが導かれる. したがって誤差平方和は

$$S = y'y - y'B'By - \frac{1}{1+f}\hat{\theta}'Q$$

という形で求められ，$\hat{\sigma}^2$ の不偏分散は

$$\hat{\sigma}^2 = \frac{1}{b(k-1) - (v-1)/(1+f)}S$$

という形で求められる.

そこで分散分析は通常の BIB の場合と同様に進めることができることが示される．ただしそこで，品種間平方和 $\hat{\theta}Q$，およびその自由度 $v-1$ をともに $1+f$ で割ることが必要になる．f があまり大きくなければ，さらにここで通常の F 分布を用いて品種効果がないという仮説を検定することができる．

ところでこの場合，配置はランダムに選んだ後に，解析に当たっては配置を与えられたものとして条件付で考えることもできる．すなわち方程式 (4)(5) をそのまま解いて θ を求めればよい．現在ではこのような方程式を解く場合の計算上の困難は存在しないといってよいであろう．

実は上記の方法よりさらにさかのぼって (4)(5) 式の段階で BT について期待値を取ってしまうことも考えられる．すなわちランダム配置の解析法にはいくつかの段階のものが存在するのである．

5 Random 最適配置

J.Kiefer は 1958 年に"Non randomized optimality and randomized non optimality of symmetrical designs"という論文で，最適計画について論じ，その後しばらくし流行した最適計画の理論研究の口火を切ったのであったが，その際もっぱら注目され，Kiefer 自身も力を注いだのはこの論文のタイトルの前半の方で，後半の"randomized non-optimality"の部分はその後ほとんど関心を惹くことはなかったように思われる．

問題の本質を示すために最も簡単な例を挙げよう．k 通りの処理法に対して実験観測値は平均 θ_i，$i=1,\ldots,k$，分散 σ^2 の正規分布に従うとする．問題は N 回の実験を行って $\theta=\cdots=\theta_k$ を検定することであるとする．

そのとき最も普通の考え方は，k 通りの処理に対して，それぞれ N/k 回の実験を行いその結果を X_{ij} $\quad i=1,\ldots,k \quad j=1,\ldots,n$ として

$$F=\frac{n\sum(\bar{X}_i-\bar{X})^2/(k-1)}{\sum\sum(X_{ij}-\bar{X}_i)^2/(N-k)}$$

を F 分布を用いて検定することである．そうすると対立仮説の下での下の分布は非心度

$$\psi=n\sum(\theta_i-\bar{\theta})^2=N\sum(\theta_i-\bar{\theta})^2/k$$

自由度 $k-1$,$N-k$ の非心 F 分布になる．これは $\sum(\theta_1-\bar{\theta})^2=c$ となる対立仮説に対する検出力の下限を最大にする検定であり，また non-random な配置の中では同じ範囲で検出力の下限を最大にする方式であることが示される．これが"non-randomized optimality"の意味である．

ところがランダム化を許すとすれば次のようなことが可能になる．k 個の品種の中から m 個をランダムに選び，それぞれ N/m 回ずつ実験して F 検定を行えば，対立仮説の下で検定統計量の分布は，自由度 $m-1$,$N-m$, 非心度

$$\psi'=N\sum_i(\theta_j-\bar{\theta}')^2/m \qquad \bar{\theta}'=\sum\theta_j/m$$

の非心 F 分布となる．ただしここで和は選ばれた品種についてのみ取ることになる．ψ' は確率的に変動するが，その期待値は

$$E(\psi')=N(m-1)\sigma_\theta^2/m \qquad \sigma_\theta^2=\sum(\theta_i-\bar{\theta})^2/(k-1)$$

となる．

今Nが十分大きいと仮定すれば，$(m-1)F$がほぼ自由度$m-1$，非心度ψ'の非心X^2分布に従うことになり，

$$
\begin{aligned}
\beta &= P_r\{F > F_\alpha(m-1, N-k)\} \simeq P_r\{X^2(m-1,\psi') > X^2_\alpha(m-1)\} \\
&= E\{\sum_{k=0}^{\infty}\frac{(\psi'/2)^k}{k'}e^{-\psi'/2}P_r\{X^2(m-1+2k) > X^2_\alpha(m-1)\}
\end{aligned}
$$

という形に表される．ただしここで$X^2(m-1,\psi')$は非心X^2統計量$X^2(f)$，自由度fのX^2統計量$X^2_\alpha(m-1)$は自由度$m-1$のX^2分布の上側α点を表す．そうするとさらに

$$
\beta = \alpha + C_1E(\psi') + \frac{c_2}{2}E(\psi'^2) + \frac{c_3}{6}E(\psi'^3) + \ldots.
$$

という形で展開できるが，上式から第2項の係数c_1は

$$
C_1 = \frac{1}{m1(\frac{m}{2})}(\frac{y}{2})^{\frac{m}{2}-\frac{y}{2}}_{e} \qquad ただし\ y = X^2_\alpha(m-1)
$$

となる．ここで$E(\psi') = N(m-1)\sigma^2_\theta/m$を代入すると$c_1E(\psi')$は$m$が増加すると単調減少することがわかる．すなわち$\sigma^2_\theta$が小さいときには$m$が小さい方が検出力が大きくなる．これが「Randomized non-optimality」の意味するところである．σ^2_θが大きくなると$E(\psi')^2$以下の項が大きくなるのでこのことは成り立たなくなる．σ^2_θがある値より大きくなれば，$m = k$のときに最大になることは証明できる．

そこで興味ある問題は$m < k$の方が$m = k$より検出力が大きくなるのはσ^2_θがどの程度の範囲までかということであるが，この問題の数値的なチェックはまだ行われていない．

6 Randomizationの効率

Fisherがrandomized blockの考え方を導入したとき，多くの人の抵抗にあったようである．この問題をめぐって長年Fisherの理解者であった"Student" (Gosset)とも決裂してしまったとのことである．多くの人々が主張したことはブロックの中のプロットにいくつかの品種をランダムに配列するより，注意深くバランスさせた方が効率がよくなるということであった．これに対してFisherはそれでは有意性検定ができなくなるとのべただけでなく，ランダム化した方が効率がよい．すなわちプロット間のバラつきの品種間の収量差に与える影響が小さくなると主張したのであった．しかしこの後半部分は無理な主張である．ランダム化によって実は偶然に「偏った」配置になることは，実験のくり返し数が極めて大きい場合を除けば，実は起こり得ることなのである．そうして確率分布はそういう「偏った」場合もすべてふくめて計算されていることはいうまでもない．

多くの人に「現実の実験配置，あるいは標本抽出において乱数表を引いてみたら，明らかに偏った結果になったらどうするか」とたずねてみたことがあったが，大抵の人は乱数表を引きなおすと答えていた．しかしそれでは実は有意性の計算の前提が変わってしまうことになるが，そのことはとりあえず「忘れてしまう」のが経験豊かな実際家の態度であったと思う．私もそれが正しいと思う.「理論」にこだわって明らかに「偏った」結果を出す必要はない．

しかしそこで生ずる理論と実際のギャップをまじめに考えるとすれば，最初から明らかに偏っていると思われるものは除いて，残りからランダムに一つを選んで実行するのが正しいと思われる．そうするとそれに応じて確率分布や分散の公式も変えねばならなくなるかもしれないので，そのことを十分考慮しなければならない．実際正方形に近いブロックでラテン方式を採用するのは，このような偏りを避ける一つの工夫である．その際ラテン方格の中でランダム化を行わねばならないことはいうまでもない．しかしそこでR.A.Fisherはすべてのラテン方格をリストアップしてその中からランダムに一つを選ぶことに固執したが，しかし一つのラテン方格を固定してその文字と品種との対応，行番号，列番号と現実のブロックにおける行と列との対応をランダムにすれば，それで十分なはずである．もう一つこの場合誤差分散の公式が変わることをNeymanが指摘したのに対してFisherがそれを拒否し，それが2人の激しい感情的対立に発展してしまったこともあった．

実際のrandom samplingにおいて，しばしば名簿の中から系統抽出によって，スタート点だけをランダムに選び，後は一定の番号毎に標本を抽出する方法が取られることがある．これは実際にはごく少数の組の中からランダムに一つを選ぶことになるので，ランダム性の点では問題であるが，名簿の中では観測値の自己相関が存在するとすれば，系統抽出の抽出による誤差を小さくすることには役立つであろう．

7　有限母集団からのrandom sampling

oversaturated random designは有限母集団からのrandom samplingと密接な関係がある．というより後者は前者の一つであると考えることもできる．すなわち大きさNの母集団から大きさnの標本が観測される場合，母集団での値を$\xi_1,\dots,\xi_N$，それを要素にするベクトルを$\boldsymbol{\xi}$標本の値を$y_1,\dots,y_n$それを要素とするベクトルをyにすれば，観測誤差はないとして

$$y_i = \begin{cases} \sum x_{ij}\theta_j & (i=1,\dots,n \quad j=1,\dots,N) \\ x_{ij}=1 & (\text{第}\,i\,\text{標本が母集団の第}\,j\,\text{単位であるとき}) \\ 0 & (\text{そうでないとき}) \end{cases}$$

重複抽出はないとれば，すべてのiについて$\sum_{ji} x_{ij}=1$と表される．あるいは

$$\mathrm{y} = X\boldsymbol{\xi}$$

Xはランダムに変動する行列である．ここで$ranX = n < N$だから$X'X$は非正則であるが

$$E(X'X) = D$$

は対角線行列で，その第j対角要素は母集団の第j単位が標本にふくまれる確率p_iに等しくなる．したがって

$$\hat{\boldsymbol{\xi}} = E(X'X)^{-1}X'\mathrm{y}$$

が$\boldsymbol{\xi}$の不偏推定量になる．よりくわしく書けば

$$\hat{\xi}_j = \frac{1}{p_j}\sum_{i=1}^{n} x_{ij}yi$$

となり，これは第 j 単位が標本にふくまれれば ξ_j/p_j に，ふくまれなければ 0 に等しくなる．これは一見随分奇妙な推定量であるが，不偏推定量であることは容易に確かめられる．すなわち大きさ n の標本から N 個の母集団単位の値すべての不偏推定量が得られる．しかもこれがまともな推定量であることは，

$$\bar{\xi} = \frac{1}{n}\sum \xi_j$$

の推定量が

$$\bar{\xi} = \frac{1}{n}\mathbf{1}'\hat{\xi} = \frac{1}{n}\sum \theta_j = \frac{1}{n}\sum y_j = \bar{y}$$

となって，ごくふつうの推定量になることからも明らかである．

この考え方を拡張すると母集団母数 θ で $g(\xi_1,\ldots,\xi_N)$ が

$$g(\xi_1,\ldots,\xi_N) = \frac{1}{{}_N C_K}\sum_{j1,\ldots,j_k} h(\xi_{j_1},\ldots,\xi_{jk})$$

と表されるとき，$m > k$ ならば $h(\xi_{j1},\ldots,\xi_{jk})$ の不偏推定量が

$$h(\xi_{ji},\hat{\ldots},\xi_{jk}) = \begin{cases} \frac{1}{p(j_1,\cdots,j_k)} h(\xi_{j1},\ldots,\xi_{jk}) & (j_1,\ldots,j_k \text{ 番の単位がすべて標本にふくまれるとき}) \\ 0 & (\text{そうでないとき}) \end{cases}$$

ただし $p(j_1,\ldots,j_k)$ は $j_1,\ldots,j_k$ が標本にふくまれる確率で 0 でないとする

という形で求められるから，

$$\hat{\theta} = \frac{1}{{}_N C_K}\sum_{j_1,\cdots,j_k} \hat{h}(\xi_{j1},\ldots,\xi_{jk})$$

が不偏推定量になる．

実は有限母集団の母数が不偏推定可能であるてめの必要十分条件は，θ が上のような形に分解されることであることは容易に示される．

実は有限母集団の母数については，分散をつねにゼロにすることができるような特別の場合を除いて，一様最小分散不偏推定量は存在し得ない．そのことを示すのも容易である．例えば ξ の推定量についていえば，$\xi_j \quad j = 1,\ldots,N$ に対して，適当な contectured values$a_j \quad j = 1,\ldots,n$ を考え，a_j の平均値 $\bar{a}_j$ として $\bar{\xi} - \bar{a}$ の不偏推定量を考えれば，それは $\bar{y} - \bar{A}$（ただし $\bar{A}$ は標本に含まれた a の値の平均）という形で与えられるから，

$$\hat{\bar{\xi}}_a = \bar{y} - \bar{A} + \bar{a}$$

が $\bar{\xi}$ の不偏推定量になる．そうして $\hat{\bar{\xi}}_a$ の分散はたまたま $\xi_j = a_j \quad j = 1,\ldots,N$ であれば 0 になることは明らかであるから，任意の点での局所最小分散不偏推定量の分散はゼロになる．このような議論はすべての不偏推定可能な母数に関していえることも明らかである．

上記のように母集団での各単位の値 ξ_j に対して何らかの情報を与える量 a_j が存在するとき，それはいろいろな形で与えられる．上記の例は「差推定」と言うべきもであるが、このような補助情報を比推定や回帰推定の形で利用することは現実にもしばしば行われている．このような方法は対象の構造について特別のモデルを仮定するものでなく，いわゆる"model free"なものを考えられているが，しかしそこには問題がある．一つは通常用いられている比

推定量や，回帰推定量は不偏でない．不偏な比推定量や回帰推定量もいろいろ提案されているが，その多くは形が複雑であるし，またその中には不偏ではあっても分数が大きくなってしまうものも多いので，実際にはあまり用いられていない．しかし「不偏性」は model free な方法における「客観性」の唯一の基準であるから，軽々しく扱うべきではないと思う．第二にこのような補助情報を用いるいくつかの方法が提案されているとき，それらの間を選択をする客観的な基準が必ずしも存在しないことが問題である．ここで 2 つの推定量 $\hat{\theta}_1$, $\hat{\theta}_2$ が存在するとき，その分散の推定量を比較して，その大小によって優劣を決定すればよいと思われるかもしれない．しかしながらその分散の推定に当たっても，補助情報が用いられるとすれば，2 つの推定量が得られるはずであり，そのとき

$$\hat{V}_{\hat{\theta}_1}(1) < \hat{V}_{\hat{\theta}_2}(1), \qquad \hat{V}_{\hat{\theta}_1}(2) > \hat{V}_{\hat{\theta}_2}(2)$$

となることがあり得るからである．但し $\hat{V}_{\hat{\theta}_1}(1)$ は補助情報 1 を用いる $\hat{\theta}_1$ の分数の推定量である。そうすれば補助情報 1 を信じる人は，$\hat{\theta}_1$ がよいと主張し補助情報 2 を信じる人は $\hat{\theta}_2$ がよいと主張して，合意が成立しないであろう．

補助情報の利用は結局主観的な判断に依存することになるから，それが多く用いられるようになると randomization による「主観的判断から独立した客観的な方法」という magic は消えてしまう．

現実の統計調査は複雑化し，またその中の調査項目も互いに複雑な関連を持つようになっているから，そのような情報を利用することは結果の制度を精度を高めるために不可欠になっている．このような状況では「model free な客観性」というフィクションはやめて，対象となるデータの構造について明示的にモデルをどうにゅうして議論を進めるべきであると思う．そうして標本抽出においても，そのようなモデルは積極的に利用すべきであろう．

8 Monte Carlo 法

最後に Monte Carlo 法についてふれておこう．それは randomization がすべての方向にバランスするという性質を積極的に利用するものである．とくにそれは多次元積分（あるいはそれと同じことでああるが複雑な統計量の期待値）を計算するのに有効である．くり返す数を N とすると誤差が次元と無関係に $N^{-\frac{1}{2}}$ になるところが Monte Carlo 法の長所である．

しかし問題が「deterministic ならば random にできることが，どうして non-random にできないのだろうか？あるいは Monte Carlo 法，すなわち random 法は最も効率のよい方法といえるだろうか．

これについて一次元（時には二次元）の場合はシステマティックな方法の法が効率がよいことが知られているが，3 次元以上になればより高いオーダーの近似を得る方法が知られている．私はその方面のことはくわしくないので立ち入らないが，概していえば

$$I = \int f(x_1, \ldots, x_m) dx_1, \ldots, dx_n$$

を計算するとき，f の動きについて全く何も仮定されない．すなわち 2 点 $(x_1, \ldots, x_n)$ と $(x'_1, \ldots, x'_n)$ の間の関係（距離）と 2 点での値 $f(x_1, \ldots, x_n)$ と $f(x'_1, \ldots, x'_n)$ の差について，全く何も想定できないならば，おそらく単純な Monte Carlo 法が minimax 的な意味で最良であろう．しかし一般化は少なくとも区分的連続性，あるいは区分的にリプシッツ条件が成り立つことが

わかっている場合が少なくないので，そのような場合にはその情報を利用した方法を考えねばならない．それは複雑な確率構造を持つ Monte Carlo 法との形で与えられるかもしれないが，しかしそのような考え方を進めると最後にはシステマティックな方法との差が少なくなるであろう．

9　むすび

Randomization というものは，どうもつかみにくい概念である．未だに完全に理解できた感じはしない．それでも場合によって役立つものであることは疑いない．