数理解析研究所講究録
25巻 1967年 212-220

randomization design の理論について

竹内　啓

§1 randomization design と情報量

randomization designとは、一般に配置のわりつけにおいて何らかの確率化を導入したものをいう。しかしその中でplotへの割りつけにおける確率化と、因子の割りつけにおける確率化とは区別して考えねばならない。前者がR. A. Fisherの意味でのrandomizationである。ここでは後者を主として考える。なお両者をともに考慮したものとして宮川強氏の研究があることをのべておこう。

抽象的に表現すると、次のように表わされる。いま $\mathbf{y}$ を実験観測値のベクトル、X を配置行列、θ を母数のベクトル、$\mathbf{u}$ を誤差のベクトルとして、

$$\mathbf{y} = X\theta + \mathbf{u} \tag{1}$$

とする。いま X のとり得る範囲を $\mathcal{X}$ と表わす。

$\mathbf{u}$ に独立正規分布を仮定すると、よく知られているように、θ に関する情報行列は、

$$I(\theta) = (X'X)/\sigma^2 \tag{2}$$

で与えられる。そこで $I(\theta)$ を何らかの意味で大きくするように X をえらばねばならない。

いま X が確率的に変動するものとすると、(2)は X が与えら

れたときの條件付情報量を表わしている. 従って平均情報量

は $\quad I(\theta) = E(X'X)/\sigma^2 \qquad (3)$

となる. そこで(3)を何らかの意味で最大にするように

Xの分布をえらぶことが考えられる.

ところで, 一般に

$$E(X'X) - E(X')E(X) = E\{(X'-E(X'))(X-E(X))\} \geq 0$$

(非負定符号)

だから, (3)を最大にするようなXの分布は1点に集中するとすれば

$\mathfrak{X}$の端点になる. そうでなければその分布は1点に集中しな

い, すなわちそれは真の確率分布になるであろう.

$I(\theta)$の大きさを表わす尺度としては, 一般にその固有根の

対称, かつ個々の根に関して単調非減少な函数をとるのが普

通である. 例えば $\sum \lambda_i = \mathrm{tr}\, I(\theta)$ $\quad \prod \lambda_i = \det I(\theta)$

$\min \lambda_i$ 等.

Xを構成する横ベクトルは, 1回の実験における配置を表

わしている. そこで $X = (x_1, x_2, \cdots x_n)'$ と表わす. Xの可

能な範囲として, n = given, 各ベクトル $x_i \in C$ = given

の場合を考えよう. このとき

$$E(X'X) = \sum_{i=1}^{n} E(x_i x_i')$$

となるから, 情報量のみを問題にするときは, 各 x_i を C

の上の測度 μ_i に従って独立に定めるものと考えればよい.

そうすると

$$E(X'X)=\sum_{i=1}^{n}\int xx'\,d\mu_i(x)$$

となる．そこで更に　$\bar{\mu}=\sum\mu_i/n$　とおけば

$$E(X'X)=n\int xx'\,d\bar{\mu}(x)$$

となるから，結局 $x_1 \cdots x_n$ をすべて同じ分布 $\bar{\mu}$ に従って定めることにすればよい。そうして問題はこのような $\bar{\mu}$ をどうえらぶかということに帰着する。

実際 randomization を許さない場合でも，$x_1 \cdots x_n$ を定める代りに，これらの各点に $1/n$ づつの確率を与えるような確率測度を考えてもよい。逆に n が十分大ならば，任意の測度 μ を，いくつかの点に $1/n$ の倍数の確率を与えるような分布で近似することができ，そうしてそれは更にこれらの点を何回かくり返えして実験するような非確率的な配置と同等になる．このように非確率的な配置を確率的な配置におきかえることによって，最適配置の問題を考えるのが Kiefer の一連の論文の方針であった．コンパクトな集合 C の上の確率測度の集合は一般にコンパクトであるから，このように確率測度を導入することによって問題は著しく簡明になる．（ただし Kiefer 自身はこれを非確率的な配置の問題に対する近似解という観点から考えていることが多い）

§2 推定と検定

しかし問題は情報行列だけでは片づかない。

まず推定の問題から考えよう。θ の最小2乗推定量を

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

とすると、X が与えられたときの條件付分散は $\sigma^2(X'X)^{-1}$ だから、その分散は

$$V(\hat{\theta}) = \sigma^2 E(X'X)^{-1} \tag{3}$$

となる。ところで

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

$$= E\{X' - [E(X'X)^{-1}]^{-1}(X'X)^{-1}X'\}\{X - X(X'X)^{-1}[E(X'X)^{-1}]^{-1}\}$$

$$\geqq 0$$

だから、$V(\hat{\theta}) - I(\theta)^{-1} \geqq 0$

となる。そして等号は $(X'X)^{-1} \equiv \mathrm{const}$ すなわち $(X'X)$ がつねに一定の行列になるときにのみ成立する。

すなわちこの場合には、最小2乗推定量は有効推定量にならない。しかしこの場合最小2乗推定量以外の推定量も考えられる。例えば

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

とおくと、$\hat{\theta}^*$ は X が与えられたときには偏りを持つが、平均的には不偏になる。そしてその分散は

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

となり　この第２項はちょうど情報量の逆行列に一致している　すなわち $\hat{\theta}^*$ は $\theta=0$ における局所有効推定量になる。同様にして任意の $\theta=\theta_0$ における局所有効推定量は

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

という形で与えられる。

すなわち局所的には情報量から与えられる分散の下限を達成することができる。一般には分論(4)において第２項はかなり大きくなるであろうから $\hat{\theta}^*$ は $\hat{\theta}$ より悪い推定量になるであろう。しかし特別な場合　例えば $(X'X)$ が特異行列になるような場合には，最小２乗推定量 $\hat{\theta}$ は計算できないが $\hat{\theta}^*$ は計算され，その分散は有限になる。X の行の数が列の数より少ない場合，いわゆる oversaturated case はその一例である（Dempster，田口氏の確率対応法）。

また一般に，(3)，(4) および

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

から，$\hat{\theta}$ と $\hat{\theta}_0^*$ との結合推定量

$$\hat{\hat{\theta}} = A\hat{\theta} + (I - A)\hat{\theta}^*$$

を考えることができる。最適な係数行列 A は未知母数 θ，σ^2 をふくむものとして与えられるが　そこに $\hat{\theta}$ および $\hat{\sigma}^2 = (y - X\hat{\theta})'(y - X\hat{\theta})/(n-p)$ を代入することによって，一つの最適に近い結合推定量が導かれる。ただしこのような方式の正確な

効率についてはまだわかってはいない。

検定の問題については次のように考えられる。いま $\theta = 0$ という仮説を検定するとき、ふつうに行われる検定方式は

$$F = y'X(X'X)^{-1}X'y / p\hat{\sigma}^2 > F_\alpha$$

という形で与えられる。そしてその検出力は X given のとき、非心母数 $\phi_X = \theta'(X'X)\theta / p\sigma^2$ の函数として $\beta(\phi_X)$ という形に表わされるから、結局検出力は $E\{\beta(\phi_X)\}$ という形で表わされ、$\|\theta\|$ が小さいときには、これは

$$E\{\beta(\phi_X)\} \doteqdot \alpha + \beta_0 \theta' E(X'X)\theta / p\sigma^2$$
$$= \alpha + \beta_0 \theta' I(\theta)\theta / p$$

となる。ただし $\beta_0 = \beta'(0)$。

従って局所検出力は、同じ $E(X'X)$ を与える配置についてはすべて同一になる。すなわち randomized design においても検出力が小さくなることはない。(この点は Kifer が 1958 年の AMS の論文において指摘したところである)。

同じ仮説を $\hat{\theta}^*$ を用いて検定することもできる。すなわち

$$cF_1^* = \hat{\theta}^{*\prime} E(X'X)\hat{\theta}^* / (y - X\hat{\theta}^*)'(y - X\hat{\theta}^*) > cF_\alpha'$$

或いは

$$cF_2^* = \hat{\theta}^{*\prime} X'X\hat{\theta}^* / (y - X\hat{\theta}^*)'(y - X\hat{\theta}^*) > cF_\alpha''$$

というような形の検定方式。更に

$$W = \hat{\theta}^{*\prime} E(X'X)\hat{\theta}^* / y'y = y'X\{E(X'X)\}^{-1}X'y / y'y > W_\alpha$$

というような検定方式を考えられる。仮説の下ではWとyyとが独立になるからWのモーメントはその分母のモーメントと分子のモーメントとの比として計算される。そうしてそれを用いてWの分布を適当な分布例えばB分布で近似することができる。これを変型すれば近似的にF分布に従うような統計量が得られる。そうしてそれから更に$\|\theta\|$が小さいときの検出力を、非心F分布で近似することができる。しかしこのような点について、まだ詳細はしらべられていない。

更にθが2つの組に分かれ　$\theta = \begin{pmatrix} \pi \\ \beta \end{pmatrix}$　と表わされて、πのみが関心の対象である場合を考えよう。これに応じて

$$y = X_1 \pi + X_2 \beta + u \qquad (5)$$

とする。いま関心の対象はπの方であるとする。このとき

$$\tilde{\beta} = E(X_2' X_2)^{-1} X_2' y$$

とおき、$y - X_2 \tilde{\beta}$　を作ると

$$y - X_2 \tilde{\beta} = (I - X_2 E(X_2' X_2)^{-1} X_2') X_1 \pi + v + \tilde{u} \qquad (6)$$

となる。ここで、$E(v) = 0$　であるから(6)から

$$\hat{\pi}^{**} = [X_1'(I - X_2 E(X_2' X_2)^{-1} X_2') X_1]^{-1} X_1' (y - X_2 \tilde{\beta})$$

とすると$\hat{\pi}^{**}$は誤差項には$\tilde{\beta}$はふくむがπはふくまないような推定量になる。これは(6)から得られるものの中で、$\tilde{u}$の分散行列に関する一般化最小2乗推定量であるから、$\beta = 0$のときには分散最小である。すなわちπを誤差項にふくまないような推定量の中で、これは局所最小分散不偏推定量になっている。ブロック配置についてこのような考え方が応用される

§3 randomization design の意義

いくつかの特殊な場合について §2のような理論から得られる具体的な結果については 私はかつて一連の論文（Report Stat. Appl. Res JUSE 1961）においてくわしくのべたことがあるので ここにはくり返さない。しかしここではそれらの結果から説明される若干の事実について結論的にまとめておこう。

(a) randomization design の drastic な性質 例えばいわゆる oversaturated case のようなものは それ自体人目をおどろかすようなものであっても、少くとも Neyman－Pearson理論の観点からする限り、理論的正統性を疑わるべきところはない。しかし実際的な観点からする限り、推定の精度、検出力というような点で このような design の効率は極めて疑わしい。いいかえれば randomization の導入によって、極めて有効な新しい手法が作られるとはいえないように思われる。

(b) しかし randomization の占める比重が小さいような場合には、この考え方を導入することによって、簡単で美しい結果が得られる場合は確かにあると思われる。例えば、與えられたブロックの数と大きさに対してBIBや簡単なPBIB のようなものが作れない場合、randomize することは有効であろう。ただしこの場合でも、なるべく balanced に近いようなものの中で randomization を行うことが望ましい。この場合ブロック効果のみを誤差項に残すような解析法を用いるのがよい

と思われる．もし母数の値がすべてかなり大きいと思われるならば，データを得たのちには，完全な条件付解析を行う必要があるかもしれない．しかしこのあたりの問題もよくはわかっていない．

(c) 要因計画においては，確率対応法は，非常にプリミティブな予備実験に用いる以外には，やはり危険が大きいであろう．しかし主効果以外の交互作用項の解析にrandomizationの考え方を導入することは有効であるかもしれない．筆者の一連の研究ではこの点に立ち入ることがなかったが，今後の一つの研究課題であろう．特に高次の交互作用の存在が疑われる場合にはこのような取り扱いが不可欠であると思われる．