

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 のわりつけにおける確率化と、因子のわりつけにおける確率化とは区別して考えねばならない。前者は 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て宮川強氏の研究があることそのべておこう。

抽象的に表現すると、次のように表わされる。いま 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、 w を誤差のベクトルとして、

$$y = X\theta + w \quad (1)$$

とする。いま X のとり得る範囲を $\mathcal{X}$ と表わす。

w 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、

$$I(\theta) = (X'X)/\sigma^2 \quad (2)$$

で与えられる。そこで $I(\theta)$ を何らかの意味で大きくする、すなわち X を大きくせねばならない。

いま X が確率的に変動するものとする。 (2) は X が与えら

れたときの条件付情報量を表わしている。従って平均情報量は

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

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。

そこで、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)を最大にするように X の分布は ^{1 変に} 集中するとすれば X の端点になる。そこで X を 1 変に集中せしめ、すなわち X 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数、かつ個々の根に因して単調非減少の函数をとるのが普通である。例えば $\sum \lambda_i = \text{tr } I(\theta)$ $\prod \lambda_i = \det I(\theta)$

$\min \lambda_i$ 等。

X を構成する観測ベクトルは、1回の実験における配置を表わしている。そこで $X = (x_1, x_2, \dots, x_n)$ と表わす。 X の可能な範囲として、 $n = \text{given}$, 各ベクトル $x_i \in C = \text{given}$ 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、情報量のみに問題にするときは、各 x_i を C の上の測度 μ_i に従って独立に定めることが考えられよう。

そうすると

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

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

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

となるから、結局 $x_1, \dots, x_n$ とすべて同じ分布 $\bar{\mu}$ に従って定めることにすればよい。そして問題はここのように $\bar{\mu}$ をどう選ぶかという事に帰着する。

実際 randomization を許す場合でも、 $x_1, \dots, x_n$ と定める代りに、これらの各点に $1/n$ ずつの確率を与えるという確率測度と考えるとよい。逆に n が十分大ならば、任意の測度 μ をいくつかの点に $1/n$ の係数の確率を与えるという分布で近似することができ、そしてそれは更にこれらの点を何回かくり返して実験するより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 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)。

3.2 推定と検定

しかし問題は情報行列だけでは片づかない。

もう推定の問題から考えよう。 θ の最小二乗推定量を

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

とすると、 X が定まるとしてその条件付分散は $\sigma^2 (X'X)^{-1}$ となる。

よってその分散は

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

と取る。よって

$$\begin{aligned} \{I(\theta) - V(\hat{\theta})^{-1}\} \sigma^2 &= E(X'X) - \{E(X'X)^{-1}\}^{-1} \\ &= E\{X' - [E(X'X)]^{-1} (X'X)^{-1} X'\} \{X - X(X'X)^{-1} [E(X'X)]^{-1}\}' \\ &\geq 0 \end{aligned}$$

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

となる。よって等号は $(X'X)^{-1} \equiv \text{const}$ となる場合 $(X'X)$ 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、最小二乗推定量は有効推定量になる。しかしこの場合最小二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 が定まるとときには偏りを持たない。平均的には不偏になる。よってその分散は

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

となり、この第2項はちよつと情報量の逆行列に一致してゐる。
 丁度おち $\hat{\theta}^*$ 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

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

といふ形で望まらる。

丁度おち局所的には、情報量から望まらる分散の下限を達成することゝなる。一般には分散(4)において第2項はかたゝ大さくなるであらうから、 $\hat{\theta}^*$ は $\hat{\theta}$ より悪い推定量になるであらう。しかし特別な場合、例えば $(X'X)$ が対角行列になるような場合には、最小2乗推定量 $\hat{\theta}$ は計算できる。この $\hat{\theta}^*$ 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 である(Dempster, 田口氏の確率忘却法)。

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 $\hat{\theta}^*$ との結合推定量

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

を考へることゝなる。最適係数行列 A は未知母数 θ , σ^2 をふくむといふことゝなるから、ここに $\hat{\theta}$ 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つて、一つの最適に近い結合推定量が得られる。たゞしこのとき、方程式の正確な

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よう。いま $\theta = 0$ として仮

説を検定するとき、以下に行われる検定方程式は

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

という形で定えられる。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 $\beta(\phi_x)$ という形で

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、 $\|\theta\|$ の

小さいときには、これは

$$\begin{aligned} E\{\beta(\phi_x)\} &\doteq \alpha + \beta_0 \theta' E(X'X) \theta / p\sigma^2 \\ &= \alpha + \beta_0 \theta' I(\theta) \theta / p \end{aligned}$$

となり、さらに $\beta_0 = \beta'(0)$ 。

従って局所検出力は、同じ $E(X'X)$ を定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 Kiefer 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 θ^* を用いて検定することもできる。すなわち

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

或いは

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

というより簡便な検定方式。更に

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

といふことの推定方式を考へられ、仮定の下では W と Y とが独立
 にあるから W の要素は Y の分布の要素と分子の要素と
 との比として計算される。そして Y を用いて W の分布と通る分布
 分布例は β 分布で近似することができる。これを整理すれば、近似では
 F 分布に従ふような統計量が得られる。そして Y のより更に $\| \theta \|$ が
 小さいと推定能力は、非心 F 分布で近似することができる。しかしこ
 のようであるについて、まだ詳細はしきへられていない。

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
 関心対象である場合を考へよう。これに応じて

$$Y = X_1 \pi + X_2 \beta + u \quad (5)$$

とする。この関心対象は 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、 $Y - X_2 \tilde{\beta}$ と作ると

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

とする。ここで、 $E(v) = 0$ であるから (6) より

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

とすると $\hat{\pi}^{*}$ は β に関する誤差項に
 なる。これは (6) より得られるものの中で、 $\tilde{u}$ の分散行列に因る一般化最小二
 乗推定量であるから、 $\theta = 0$ のときには分散最小である。すなわち π を
 誤差項に属するものとする推定量の中で、これが一層最小分散不偏推定量
 になっている。この配置についてこのように考へる方が応用される

§3 randomization design の意義

いくつかの特殊な場合において、このように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) においてくわしく述べたことがあるので、ここには繰り返さぬ。しかしここで得られた結果から説明される若干の事象について結論的にまとめておこう。

(a) randomization design の drastic な価値。例えばいわゆる oversaturated case の場合のことは、よく自作人間をおとすからすうなものであつて、少なくとも Neyman-Pearson 理論の範囲から下る限り、理論的正統性に疑われるべきところはない。しかし実際的な観点から下る限り、推定の精度、検出力という点で、このタイプの design の効率は極めて優れている。いふかえれば 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いふこともいふことが出来る。

(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、簡単に美しい結果が得られる場合は確かにあると思ふ。例えば、選ばれるプロットの数と大きさに対して BIB を簡単な P BIB のように作れる場合、randomize することは有効である。ただしこの場合でも、なるべく balanced に近いタイプのものの中で randomization を行うことが望ましい。この場合プロット効果のみを誤差項に残すタイプの解析法を用いることがよい。

と思われる。もし母数の値がすべてゼロの大きさと思われるならば、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ためである。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。筆者が一連の研究で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る。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に取り扱いが不可欠であると思われる。