数理解析研究所講究録
第 25 巻 1967 年 43-52

ブロック・デザインの無作為化について

日本大学生産工学部統計学教室　小川潤次郎

無作為化(Randomization)というのは,近代統計学における特色ある概念であって,分散分析(Analysis of variance)と共に実験計画法の二大支柱となっている訳である.

無作為化についての一般的叙述は例えば,その提唱者であるR.A.Fisherの古典的な名著[1]及び[2]又は,そのFisherの思想を解説した北川敏男[3][5]によって見られ度い.無作為化の数学的な取扱は最初B.L.Welch[10]とE.J.G.Pitman[11]によって与えられた.今日の言葉でいえば,彼等の取扱ったのは所謂"Fisher模型"の場合であった.

1935年にJ.Neyman et al [7]は乱塊法とラテン方格配置の無作為化を考察していわゆる技術誤差(Technical error)をも考慮しなければならない場合のあることを指摘した.此場合を吾々は"Neyman模型"と呼ぶことにする. Neyman模型に対する数学的取扱は1939年にM.D.McCarthy[6]によって試みられた.McCarthyは同一ブロック内のk個の観察値が,そのプロット誤差と技術誤差とが合併されたときk次元の正規分布に従うものとしてF統計量の分布を考察した.若しプロットと処理効果が交絡していなければF統計量の分布は近似的にF分布になるが,この交絡が在るときにはF統計量の分布はF分布とは大分異るものになるという結果を出した.無作為化されたプロット効果は離散的変量であるから,技術誤差が正規変量ならばそれらが互に独立な限り,その和が正規変量になることはない,このことはMcCarthy自身も知っていた訳である.

数学的には、Fisher模型はNeyman模型の特殊な場合となるので、Neyman模型の場合にブロック・デザインの無作為化をキチンと取扱って普通に用いられているF-検定が如何なる意味で正しいのかを吟味しておこうというのが此研究の目的である。

先ず完全ブロック配置 J. Ogawa [8]、ラテン方格配置 J. Ogawa [9]、BIB配置 J. Ogawa [10]、2-associate class の PBIB配置 J. Ogawa, S. Ikeda, M. Ogasawara [11]、m-associate class の PBIB配置 J. Ogawa, S. Ikeda, M. Ogasawara [12]、[13] であるが以上は Welch Pitman と同様モーメント法ともいうべき立場であったがこれをもう一層改良したのが S. Ikeda, J. Ogawa [15] である。

以下に於ては最も簡単な場合の完全ブロック配置について問題の焦点が何であるかを解説する。

1. ブロック・デザイン

実験単位(Experimental unit)又はプロット(Plot)いくつかの集りをブロック(Block)という。一ブロック内の実験単位の数をブロックの大きさ(size)という。以下我々が取扱うのはブロックの大きさが一定 k である b 個のブロックである。従って全体としては $n=bk$ 個の実験単位が与えられている。そこで実験によって比較さるべき処理(Treatment)の数は v であるとする。

何故にブロックが必要かというと、一ブロック内の実験単位の効果 —— これを単位効果又はプロット効果という —— は可成く一様にしたいのである。従ってブロックの大きさ k は実際的にはそんなに大きくとれないのである。例えば動物実験のような場合には遺伝的及び飼育条件が似ているから普通同腹(Same litter)のものをブロックとする。従って $k=2$ というよう

場合が大切になる O.Kempthorne [3].

$v=k$ の場合は各ブロック内で全部の処理が一回ずつ実験される. この場合を完全ブロック (Complete block) という. $v>k$ のとき不完全ブロック (Incomplete block) という. 実際的にはこの場合が多く, 且つ数学的には完全ブロックは不完全ブロックの極限的なケースである.

扨て, $n=bk$ 個の実験単位に1からn迄の通し番号をつけてf番目の実験単位における観察値をx_fで表わし, n次元ベクトル

$$\underline{x}' = (x_1, x_2, \cdots\cdots, x_n)$$

を観察値ベクトル (Observation vector) と呼ぶことにする.

以下話を簡単にする為に完全ブロックの場合を考察する Ogawa [1]. 完全ブロックの場合の分散分析を取扱うのであるが, より一般な不完全ブロックに対しても拡張出来る道具立てを用いることにする.

処理のインシデンス・ベクトルを次のように定義する

$$\zeta_{\alpha f} = \begin{cases} 1, & \text{若し}\alpha\text{番目の処理が}f\text{番目のプロットに施されれば;} \\ 0, & \text{然らざれば.} \end{cases}$$

としてn次元ベクトル

$$\underline{\zeta}_\alpha' = (\zeta_{\alpha 1}, \zeta_{\alpha 2}, \cdots\cdots, \zeta_{\alpha n}), \quad \alpha = 1, 2, \cdots, v$$

が処理インシデンス・ベクトルで, これを並べた $n \times v$ 行列

$$\Phi = \| \underline{\zeta}_1\ \underline{\zeta}_2\ \cdots\ \underline{\zeta}_v \|$$

が処理インシデンス行列である.

$$\underline{\zeta}_\alpha' \cdot \underline{\zeta}_\beta = r_\alpha \delta_{\alpha\beta}, \quad \text{但し}\ r_\alpha = b$$

であるから, $\Phi'\Phi = rI_v = bI_k$ であることは明らかであろう.

同様にしてブロックのインシデンス・ベクトルとインシデンス・行列を定義する.

$$\eta_{af} = \begin{cases} 1, & \text{若し } f \text{ 番目のプロットが } a \text{ 番目のブロックに属すれば;} \\ 0, & \text{然らざれば.} \end{cases}$$

として, インシデンス・ベクトルは

$$\eta'_a = (\eta_{a1}, \eta_{a2}, \cdots\cdots, \eta_{an}),\ a = 1, 2, \cdots, b$$

インシデンス行列は

$$\Psi = \| \eta_1 \eta_2 \cdots \eta_b \|$$

$$\Psi' \Psi = k I_b.$$

さて

$$n_{\alpha a} = \zeta'_\alpha \eta_a = \sum_{f=1}^{n} \zeta_{\alpha f} \eta_{af},\ \alpha = 1, \cdots, v;\ a = 1, \cdots, b,$$

とおくと α 番目の処理が a 番目のブロックに入っているときは1然らざるときは0である. 完全ブロックのときはつねに $n_{\alpha a} = 1$ である. それは処理とブロックとのインシデンスを表わすので, $v \times b$ 行列

$$N = \| n_{\alpha a} \|_{\substack{\alpha = 1, \cdots, v \\ a = 1, \cdots, b}}$$

をデザインのインシデンス行列 (Incidence matrix) というのである.

更に吾々はデザインの関係行列 (Relationship matrices) を次のように定義する.

(1) 恒等関係: I_n n次の単位行列.

(2) ユニバーサル関係: G_n: その要素がすべて1であるn次正方行列

(3) 処理関係: $T = \Phi\Phi' = \| t_{fg} \|$

$$t_{fg} = \sum_{\alpha=1}^{v} \zeta_{\alpha f} \zeta_{\alpha g} = \begin{cases} f \text{ プロットと } g \text{ プロットが同一の処理を受ければ } 1, \\ \text{然らざれば } 0. \end{cases}$$

(4) ブロック関係: $B=\Psi'\Psi=\|b_{fg}\|$

$$b_{fg}=\sum_{\alpha=1}^{b}\eta_{\alpha f}\eta_{\alpha g}$$

で、fプロットとgプロットが同一ブロックにあれば $b_{fg}=1$ 然らざるときは $b_{fg}=0$ である。

ところで上の I, B, T, G の間には次の掛算表が成立つ。

I	B	T	G
B	kB	G	kG
T	G	bT	bG
G	kG	bG	nG

従って $\frac{1}{n}G$, $\frac{1}{k}B$, $\frac{1}{b}T$ は冪等行列であって、且つ

$$I=\frac{1}{n}G+\left(\frac{1}{k}B-\frac{1}{n}G\right)+\left(\frac{1}{b}T-\frac{1}{n}G\right)+\left(I-\frac{1}{k}B-\frac{1}{b}T+\frac{1}{n}G\right)$$

は単位行列 I の互に直交する冪等行列の和への分解となる。

$\mathbf{j}'=(1,1,\ldots\ldots,1)$ とすれば

$$\frac{1}{n}G\mathbf{x}=\bar{x}\cdot\mathbf{j} \quad 但し \quad \bar{x}=\frac{1}{n}\sum_{f=1}^{n}x_f,$$

又

$$\underline{B}=\begin{bmatrix}B_1\\B_2\\\vdots\\\vdots\\B_b\end{bmatrix},\quad B_i=\sum_{f\in i\text{-th block}}x_f\quad;\quad \underline{T}=\begin{bmatrix}T_1\\T_2\\\vdots\\\vdots\\T_t\end{bmatrix},\quad T_\alpha=\sum_{\{f:\delta_{\alpha f}=1\}}x_f$$

とおくと

処理平方和: $S_t^2=\mathbf{x}'\left(\frac{1}{b}T-\frac{1}{n}G\right)\mathbf{x}=\frac{1}{b}\sum_{\alpha=1}^{t}T_\alpha^2-n\bar{x}^2$,

ブロック平方和：$S_b^2 = x'(\frac{1}{k}B - \frac{1}{n}G)x = \frac{1}{k}\sum_{a=1}^{b} B_a^2 - n\bar{x}^2$

誤差平方和；$S_e^2 = x'(I - \frac{1}{k}B - \frac{1}{b}T + \frac{1}{n}G)x = \sum_{f=1}^{n} x_f^2 - n\bar{x}^2 - S_t^2 - S_b^2$.

となる。分散比統計量は

$$F = \frac{(b-1)(k-1)}{k-1}\ \frac{x'(\frac{1}{b}T - \frac{1}{n}G)x}{x'(I - \frac{1}{k}B - \frac{1}{b}T + \frac{1}{n}G)x}$$

となる。吾々は此統計量Fの標本分布を問題にする。

2. 無作為化を施さない前のF統計量の分布

Neyman模型の一般的なものは

$$x_f = \gamma + \sum_{\alpha=1}^{v}\zeta_{\alpha f}\tau_\alpha + \sum_{a=1}^{b}\eta_{af}\beta_a + \sum_{\alpha=1}^{v}\zeta_{\alpha f}\pi_{\alpha f} + e_f ,\quad f=1,2,\cdots,n$$

但しここで；γは一般平均，$\tau' = (\tau_1, \tau_2, \cdots, \tau_v)$は処理効果，$\beta' = (\beta_1, \beta_2, \cdots, \beta_b)$はブロック効果で $\sum_\alpha \tau_\alpha = \sum_a \beta_a = 0$ とする。更に

$$\pi_\alpha' = (\pi_{\alpha 1}\ \pi_{\alpha 2}\ \cdots\ \pi_{\alpha n}),\quad \alpha=1,2,\cdots,v$$

は処理αに対するプロット効果であって

$$\Psi'\pi_\alpha = 0,\quad \alpha=1,2,\cdots,v$$

とする。$e' = (e_1, e_2, \cdots, e_n)$は技術誤差(technical error)であって$N(0, \sigma^2 I)$に従うものとする。

以下では処理効果とプロットの間に交絡のない場合のみ取扱うのである。そのときは模型は

$$x_f = \gamma + \sum_{\alpha=1}^{v}\zeta_{\alpha f}\tau_\alpha + \sum_{a=1}^{b}\eta_{af}\beta_a + \pi_f + e_f$$

又は行列記号では

$$x = \gamma j + \Phi\tau + \Psi\beta + \pi + e$$

となり之を謂ゆるFisher模型である。

吾々の検定したい仮説は

$$H_0:\ \underline{\tau}=\underline{0}$$

で，帰無仮説 H_0 のもとでは

$$S_t^2=\underline{x}'\left(\frac{1}{b}T-\frac{1}{n}G\right)\underline{x}+2\underline{x}'\left(\frac{1}{b}T-\frac{1}{n}G\right)\underline{e}+\underline{e}'\left(\frac{1}{b}T-\frac{1}{n}G\right)\underline{e}$$

となり又帰無仮説に関係なく誤差平方和は

$$S_e^2=\underline{x}'\left(I-\frac{1}{b}T-\frac{1}{k}B+\frac{1}{n}G\right)\underline{x}+2\underline{x}'\left(I-\frac{1}{b}T-\frac{1}{k}B+\frac{1}{n}G\right)\underline{e}$$
$$+\underline{e}'\left(I-\frac{1}{b}T-\frac{1}{k}B+\frac{1}{n}G\right)\underline{e}$$

となるから，$\chi_1^2=S_t^2/\sigma^2$ は非心カイ自乗分布

$$e^{-\frac{\theta_1}{2\sigma^2}}\sum_{\mu=0}^{\infty}\frac{\left(\frac{\theta_1}{2\sigma^2}\right)^{\mu}}{\mu!}\frac{\left(\frac{\chi_1^2}{2}\right)^{\frac{k-1}{2}+\mu-1}}{\Gamma\left(\frac{k-1}{2}+\mu\right)}e^{-\frac{\chi_1^2}{2}}d\left(\frac{\chi_1^2}{2}\right)$$

但し $$\theta_1=\underline{x}'\left(\frac{1}{b}T-\frac{1}{n}G\right)\underline{x}=\frac{1}{b}\underline{x}'T\underline{x}.$$

又 $\chi_2^2=S_e^2/\sigma^2$ は非心カイ自乗分布で

$$e^{-\frac{\theta_2}{2\sigma^2}}\sum_{\nu=0}^{\infty}\frac{\left(\frac{\theta_2}{2\sigma^2}\right)^{\nu}}{\nu!}\frac{\left(\frac{\chi_2^2}{2}\right)^{\frac{(b-1)(k-1)}{2}+\nu-1}}{\Gamma\left(\frac{(b-1)(k-1)}{2}+\nu\right)}e^{-\frac{\chi_2^2}{2}}d\left(\frac{\chi_2^2}{2}\right)$$

但し $$\theta_2=\underline{x}'\underline{x}-\theta_1.$$

これらは互に独立であるから

$$F=(b-1)\frac{\chi_1^2}{\chi_2^2}$$

の分布は普通の中心的F-分布ではなく非心F-分布となる。

$$\frac{\Gamma\left(\frac{b(k-1)}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{(b-1)(k-1)}{2}\right)}\left(\frac{F}{b-1}\right)^{\frac{k-1}{2}-1}\left(1+\frac{F}{b-1}\right)^{-\frac{b(k-1)}{2}}\times e^{-\frac{\underline{x}'\underline{x}}{2\sigma^2}}\sum_{\mu,\nu=0}^{\infty}\frac{\left(\frac{\theta_1}{2\sigma^2}\right)^{\mu}\left(\frac{\theta_2}{2\sigma^2}\right)^{\nu}}{\mu!\,\nu!}\frac{\left(\frac{F}{b-1}\right)^{\mu}}{\left(1+\frac{F}{b-1}\right)^{\mu+\nu}}\times$$

$$\frac{\Gamma\left(\frac{b(k-1)}{2}+\mu+\nu\right)\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{(b-1)(k-1)}{2}\right)}{\Gamma\left(\frac{b(k-1)}{2}\right)\Gamma\left(\frac{k-1}{2}+\mu\right)\Gamma\left(\frac{(b-1)(k-1)}{2}+\nu\right)}d\left(\frac{F}{b-1}\right)$$

これは Q_1 Q_2 に関係する.

3. 無作為化による $Q_1=\frac{1}{b}\mathbf{x}'A\mathbf{x}$ の分布

$\mathbf{x}$ は変量ではないが無作為化によって処理のインシデンス行列 $A=\Phi\Phi'$ が変量となるから従って $Q_1=\frac{1}{b}\mathbf{x}'A\mathbf{x}$ も確率変数となる.

先ず無作為化に基づく Permutation 分布によって Q_1 の平均と分散を計算して考えれば b が充分大きいならば $Q_1/(\mathbf{x}'\mathbf{x})$ の分布は B-分布

$$\frac{\Gamma\left(\frac{b(k-1)}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{(b-1)(k-1)}{2}\right)}x^{\frac{k-1}{2}-1}(1-x)^{\frac{(b-1)(k-1)}{2}-1}dx$$

で近似出来ることが示される. この近似を用いると

$$E\left[\left(\frac{Q_1}{\mathbf{x}'\mathbf{x}}\right)^{\mu}\left(\frac{Q_2}{\mathbf{x}'\mathbf{x}}\right)^{\nu}\right]=\frac{\Gamma\left(\frac{b(k-1)}{2}\right)\cdot\Gamma\left(\frac{k-1}{2}+\mu\right)\cdot\Gamma\left(\frac{(b-1)(k-1)}{2}+\nu\right)}{\Gamma\left(\frac{b(k-1)}{2}+\mu+\nu\right)\cdot\Gamma\left(\frac{k-1}{2}\right)\cdot\Gamma\left(\frac{(b-1)(k-1)}{2}\right)}$$

となるから, 結局無作為化される F-統計量の分布は近似的に中心的 F-分布

$$\frac{\Gamma\left(\frac{b(k-1)}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{(b-1)(k-1)}{2}\right)}\left(\frac{F}{b-1}\right)^{\frac{k-1}{2}-1}\left(1+\frac{F}{b-1}\right)^{-\frac{b(k-1)}{2}}d\left(\frac{F}{b-1}\right)$$

となるのである.

問題の焦点は Q_1 の平均と分散の計算にある. 完全ブロックの場合は Ogawa [8], ラテン方格の場合 Ogawa [9], BIBD の場合 Ogawa [10], 2-アソシエートクラスの PBIBD の場合 Ogawa, Ikeda, Ogasawara [11], m-アソシエートクラスの PBIBD の場合 Ogawa, Ikeda, Ogasawara [12] に計算されている.

モーメント法でなく中心極限定理類似の方法で $Q_1/(\mathbf{x}'\mathbf{x})$ の近似分布

でなく、漸近分布がβ-分布になることを示し更に検定の関数を計算したのが Ikeda, Ogawa [11] である.

文　　献

[1] Fisher, R. A. (1926). The arrangement of field experiments, J. Ministry of Agric. 33.

[2] Fisher, R.A. (1935). Design of Experiments, Oliver & Boyd.

[3]. Kempthorne, O. (1953). A class of experimental designs using blocks of two plots, Ann. Math. Stat. 24.

[4] 北川敏男 (1948). 近代統計学の基盤, 統計数理研究 2.

[5] 北川敏男 (1948). 統計学の認識 第12章, 白揚社

[6] McCarthy, M.D. (1939). On the application of the z-test to randomized blocks. Ann. Math. Stat. 10.

[7] Neyman, J. with the cooperation of K. Iwaszkiewicz and S. Kolodziejczyk (1935). Statistical problems in agricultural experimentation, J. Roy. Stat. Soc., Suppl. 2.

[8] Ogawa, J. (1961). The effect of randomization on the analysis of randomized block design, Ann. Inst. Stat. Math. 13.

[9] 小川潤次郎 (1962). ランダム化配置における無作為化について. 統計数理研究所彙報 10.

[10] Ogawa, J. (1963). On the null-distribution of the F-statistic in a randomized balanced incomplete block design under the Neyman model. Ann. Math. Stat. 34.

[1]. Ogawa, J., Ikeda, S., & Ogasawara, M. (1964). On the null-distribution of the F-statistic in a randomized partially balanced incomplete block design with 2-associate classes under the Neyman model. Inst. of Stat. UNC Mimeo Series No. 414

[2] Ogawa, J., Ikeda, S., & Ogasawara, M. (1965). On the null-distribution of the F-statistic for testing a 'partial' null-hypothesis in a randomized PBIBD with m-associate classes under the Neyman model. presented at the 35th session of the International Statistical Institute in Belgrad, Yugoslavia, Sept. 1965.

[3] 小川潤次郎, 池田貞雄, 小笠原正喜: Incomplete block design の random-ization における F-統計量の漸近分布について, 第31回日本統計学会年会講演

[4]. Ogawa, J. (1965). The theory of block designs, 早稲田大学理工学部大学院の Lecture note

[5]. Ogawa, J. & Ikeda, S. (1966). On the non-null distribution of the F-statistic for testing a partial null-hypothesis in a randomized PBIB design with m associate classes under the Neyman model. Inst. of Stat. UNC. Mimeo Series 466.

(Feb. 17, 1967)