

プロット・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つて、分散分析 (Analysis of variance) と共に実験計画法の大支柱となつてゐる訳である。

無作為化について一般の叙述は勿論、その提唱者である R. A. Fisher の古典的のり居居 [1] 及び [2] 又此の Fisher の思想を解説した北川敏男 [3] [4] によつて見られ度い。無作為化の教學的取扱は最初 B. L. Welch [5] と E. J. G. Pitman [6] によつて與へられた。今日言葉で云ふと後者の取扱つたのは所謂 "Fisher 模型" の場合であつた。

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へて、いわゆる技術誤差 (Technical error) をも考慮しなければならぬ場合があることを指摘した。此場合を今々は "Neyman 模型" と呼ぶこととする。Neyman 模型に對する教學的取扱は 1939 年に M. D. McCarthy [8] によつて試みられた。McCarthy は同一プロット内の各個の觀察値を、そのプロット誤差と技術誤差とを合併されたところ次元の正規分布に従うものとして F 統計量の分布を考へた。若しプロットと処理効果と交絡してゐれば、F 統計量の分布は近似即ち F 分布になり、この交絡が存在しないとすれば F 統計量の分布は F 分布とハズれるもう大いである。結果を出した。無作為化されたプロット効果は離散型変量であるから、技術誤差と正規変量とを合併するに依り、その和が正規変量であることは、このことは McCarthy 自身も知らなかつた訳である。

数学的には Fisher 模型は Neyman 模型の特殊な場合といふので、Neyman 模型の場合にプロット・デザインが無作為にもオキタンと取扱つて普通に用ゐられてゐる F-検定が、如何に正確で正しい結果を吟味してゐるのか、とこの研究の目的である。

先づ完全プロット配置 J. Ogawa [1], ~~ラテン方格配置~~ 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用ゐる場合であつたが、これをより一般化させたのが S. Ikeda, J. Ogawa [15] である。

以下に於ては最も簡単な場合の完全プロットから、10 段の案を出してあるものと取扱ふ。

1. プロット・デザイン

実験単位 (Experimental unit) 又はプロット (Plot) といふものをプロット (Block) といふ。一プロット内の実験単位の数をプロットサイズ (size) といふ。以上をそれぞれ取りつて、プロットサイズを一定とすれば、何個のプロットであるか、従つて全体として $n = b \times r$ の実験単位が得られる。この実験単位を比較するべき処理 (Treatment) の数 t であるとする。

何故にプロットが必要といふのは、一プロット内の実験単位の結果——これを単位効果又はプロット効果といふ——はすべて一様にならなければならない。従つてプロットサイズは実験的には、そんなに大きくめとれなかつた。例へば動物実験のいふ場合には、遺伝的及び飼育条件が似てゐる、普通同腹 (Same litter) のものをプロットとする。従つて $b=2$ といふ。

場合が大切なのは D. Kempthorne [3].

$U=k$ の場合は各プロット内に全部の処理を一回ずつ実験し、この場合を完全プロット (Complete block) とする。 $U>k$ のとき不完全プロット (Incomplete block) とする。実際的にはこの場合が多い。且つ数学的には、完全プロットは不完全プロットを極限的リケースである。

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

$$x' = (x_1, x_2, \dots, x_n)$$

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

以下では簡単に u 層の完全プロットの場合を考察する Ogawa [1]。完全プロットの場合の分散分析を取扱うのであるが、一般に不完全プロットに対しても拡張出来る道具立てを用いることになる。

処理のインジケンスベクトルを次のように定義する

$$s_{aj} = \begin{cases} 1, & \text{もし } a \text{ 番目の処理が } j \text{ 番目のプロットに施される場合,} \\ 0, & \text{施されない場合.} \end{cases}$$

と 1 から n 次元ベクトル

$$s_a' = (s_{a1}, s_{a2}, \dots, s_{an}), \quad a=1, 2, \dots, v$$

が処理インジケンスベクトルで、これを並べた $n \times v$ 行列

$$\Phi = \| s_1, s_2, \dots, s_v \|$$

が処理インジケンス行列にある。

$$s_a' s_p = r_a \delta_{ap}, \quad \text{但し } r_a = b$$

であるから、 $\Phi \Phi' = r I_v = b I_v$ であること11電算機に55)。

同様にプロット \$a\$ 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。

$$\eta_a = \begin{cases} 1, & \text{もし i 番目のプロットが a 番目のプロットに属すれば;} \\ 0, & \text{然らざれば;} \end{cases}$$

よって、インシデンスベクトルは

$$\eta_a = (\eta_{a1}, \eta_{a2}, \dots, \eta_{a6}), \quad a=1, 2, \dots, b,$$

インシデンス行列は

$$\Psi = \|\eta_1, \eta_2, \dots, \eta_b\|$$

$$\Psi^T \Psi = \delta_{bb}.$$

さて

$$n_{da} = \xi_a^T \eta_a = \sum_{f=1}^n \xi_{af} \eta_{af}, \quad d=1, \dots, v; \quad a=1, \dots, b,$$

とみると \$a\$ 番目の処理が \$a\$ 番目のプロットに入っているのは 1 回、
入っていないのは 0 である。完全プロットならば、つまり \$n_{da}=1\$ である。処理とプロットとのインシデンスをあらわすので、\$v \times b\$ 行列

$$N = \|\eta_{da}\|_{\substack{d=1, \dots, v \\ a=1, \dots, b}}$$

を処理デザイン・インシデンス行列 (Incidence matrix) と呼ぶ。

更に各々はデザインの関係行列 (Relationship matrices) も次のように定義する。

(1) 恒等関係: I_n n 次の単位行列。

(2) エーバール関係: G_n その要素が全て 1 による \$n\$ 次正交行列。

(3) 処理関係: $T = \Phi \Phi^T = \|t_{ij}\|$

$$t_{ij} = \sum_{a=1}^v \xi_{ai} \xi_{aj} = \begin{cases} \text{プロット i と j が同一の処理に属すれば;} \\ \text{然らざれば 0.} \end{cases}$$

(4) フロツフ関係: $B = \Psi\Psi' = \|b_{fg}\|$

$$b_{fg} = \sum_{a=1}^b \gamma_{af} \gamma_{ag}$$

又、フロツフとフロツフの積 - フロツフに等しいから $b_{ff} = 1$ となる。

また $b_{fg} = 0$ となる。

とこで上の I, B, T, G の間に成る関係式を成す。

$$I \quad B \quad T \quad G$$

$$B \quad kB \quad G \quad kG$$

$$T \quad G \quad bT \quad bG$$

$$G \quad kG \quad bG \quad bG$$

従つて $\frac{1}{n}G, \frac{1}{k}B, \frac{1}{b}T$ は 固有値行列となる。

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

は単位行列 I の互に直交する固有値行列の和である。

$\mathbf{j}' = (1, 1, \dots, 1)$ とする。

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

又

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

とかくと

$$\text{処理率} \text{の} \mathbf{S}_f' = \mathbf{x}' \left(\frac{1}{b}T - \frac{1}{n}G \right) \mathbf{x} = \frac{1}{b} \sum_{a=1}^b T_a^2 - \frac{1}{n} \bar{x}^2$$

$$\text{そのうち平方和: } S_b^2 = X' \left(\frac{1}{k} B - \frac{1}{n} G \right) X = \frac{1}{k} \sum_{i=1}^k B_i^2 - n \bar{X}^2$$

$$\text{誤差平方和: } S_c^2 = X' \left(I - \frac{1}{k} B - \frac{1}{n} G + \frac{1}{n} G \right) X = \sum_{i=1}^n x_i^2 - n \bar{X}^2 - S_b^2 - S_a^2$$

よって、分散比統計量は

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

よって、吾々の統計量 F の標本分布は同型である。

2. 無作為化を施す前の F 統計量の分布

Neyman 模型の一般的形式は

$$x_f = \gamma + \sum_{\alpha=1}^u \gamma_{\alpha f} \tau_{\alpha} + \sum_{\alpha=1}^b \gamma_{\alpha f} \beta_{\alpha} + \sum_{\alpha=1}^u (\gamma_{\alpha f} \pi_{\alpha f} + e_f), \quad f=1, 2, \dots, n$$

但し γ は一般平均、 $\tau = (\tau_1, \tau_2, \dots, \tau_u)$ は処理効果、 $\beta = (\beta_1, \beta_2, \dots, \beta_b)$ は

そのうち効果 $\sum_{\alpha} \tau_{\alpha} = \sum_{\alpha} \beta_{\alpha} = 0$ とする。更に

$$\pi_{\alpha}' = (\pi_{\alpha 1}, \pi_{\alpha 2}, \dots, \pi_{\alpha n}), \quad \alpha=1, 2, \dots, u$$

は処理 α に対するプロット効果であって

$$\Psi' \pi_{\alpha} = 0, \quad \alpha=1, 2, \dots, u$$

とする。 $e = (e_1, e_2, \dots, e_n)$ 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両方交絡がある場合を考えるので

より、そのモデルは

$$x_f = \gamma + \sum_{\alpha=1}^u \gamma_{\alpha f} \tau_{\alpha} + \sum_{\alpha=1}^b \gamma_{\alpha f} \beta_{\alpha} + \tau_f + e_f$$

又は行列記号では

$$X = \gamma j + \bar{\tau} I + \Psi \beta + \tau + e$$

と表す。これは F 検定モデルである。

音々の検定しえり仮説は

$$H_0: I = 2$$

て、帰無仮説 H_0 が正しい

$$J_0^2 = \pi \left(\frac{1}{b} T - \frac{1}{n} G \right) \pi + 2\pi \left(\frac{1}{b} T - \frac{1}{n} G \right) \varepsilon + \varepsilon' \left(\frac{1}{b} T - \frac{1}{n} G \right) \varepsilon$$

と Q の又帰無仮説と同様に誤差平方和は

$$J_0^2 = \pi \left(1 - \frac{1}{b} T - \frac{1}{n} B + \frac{1}{n} G \right) \pi + 2\pi \left(1 - \frac{1}{b} T - \frac{1}{n} B + \frac{1}{n} G \right) \varepsilon + \varepsilon' \left(1 - \frac{1}{b} T - \frac{1}{n} B + \frac{1}{n} G \right) \varepsilon$$

と Q の又 $\chi_1^2 = J_0^2 / \sigma^2$ は自由度 $k-1$ の中央分布

$$e^{-\frac{Q_1}{2\sigma^2}} \sum_{\mu=0}^{\infty} \frac{\left(\frac{Q_1}{2\sigma^2}\right)^{\mu}}{\mu!} \frac{\left(\frac{\chi_1^2}{2}\right)^{\frac{k-1}{2}+\mu-1}}{\Gamma\left(\frac{k-1}{2}+\mu\right)} e^{-\frac{\chi_1^2}{2}} d\left(\frac{\chi_1^2}{2}\right)$$

但し $Q_1 = \pi' \left(\frac{1}{b} T - \frac{1}{n} G \right) \pi = \frac{1}{b} \pi' T \pi$

又 $\chi_2^2 = J_0^2 / \sigma^2$ は自由度 $(b-1)(k-1)+v-1$ の中央分布

$$e^{-\frac{Q_2}{2\sigma^2}} \sum_{v=0}^{\infty} \frac{\left(\frac{Q_2}{2\sigma^2}\right)^v}{v!} \frac{\left(\frac{\chi_2^2}{2}\right)^{\frac{(b-1)(k-1)+v-1}{2}}}{\Gamma\left(\frac{(b-1)(k-1)+v}{2}\right)} e^{-\frac{\chi_2^2}{2}} d\left(\frac{\chi_2^2}{2}\right)$$

但し $Q_2 = \pi \pi - Q_1$

よって F は F 分布とある

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

F 分布は普通の中心の F 分布と異なる、この F 分布とある

$$\frac{\Gamma\left(\frac{b(k-1)}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right)\Gamma\left(\frac{(b-1)(k-1)}{2}\right)} \left(\frac{F}{b-1}\right)^{\frac{k-1}{2}-1} \left(1+\frac{F}{b-1}\right)^{-\frac{b(k-1)}{2}} e^{-\frac{\pi F}{2\sigma^2}} \sum_{\mu, v=0}^{\infty} \frac{\left(\frac{Q_1}{2\sigma^2}\right)^{\mu}}{\mu!} \frac{\left(\frac{Q_2}{2\sigma^2}\right)^v}{v!} \left(\frac{F}{b-1}\right)^{\mu} \left(1+\frac{F}{b-1}\right)^{v} dF$$

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

これは Q_k に比例する。

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

Π は変量 π の行列で、無作為化によって処理のインシデンス行列 $T = \Phi \Pi'$ の変量 π の行列として、 $Q_k = \frac{1}{b} \Pi' A \Pi$ も確率変数となる。

先に無作為化による Permutation 分布によって、 Q_k の平均と分散を計算して、当では b が充分大きければ $Q_k / (\Pi' \Pi)$ の分布は B -分布

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

で近似出来ることと示す。この近似を用いると

$$E\left[\left(\frac{Q_k}{\Pi' \Pi}\right)^\mu \left(\frac{Q_k}{\Pi' \Pi}\right)^\nu\right] = \frac{\Gamma\left(\frac{b(k-1)}{2}\right) \Gamma\left(\frac{k-1}{2} + \mu\right) \Gamma\left(\frac{(b-1)(k-1)}{2} + \nu\right)}{\Gamma\left(\frac{b(k-1)}{2} + \mu + \nu\right) \Gamma\left(\frac{k-1}{2}\right) \Gamma\left(\frac{(b-1)(k-1)}{2}\right)}$$

となる。従って無作為化による Q_k の P -統計量の分布は近似的に中央 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_k の平均と分散の計算にある。完全 T -グラフの場合は Ogawa [4], ラテン方格の場合は Ogawa [7], BIBD の場合は Ogawa [10], 2-アソシエートクラスの PBIBD の場合は Ogawa, Ikeda, Ogasawara [11], m -アソシエートクラスの PBIBD の場合は Ogawa, Ikeda, Ogasawara [12] に計算されている。

モメント法を用いて、中心極限定理類似の方法で $Q_k / (\Pi' \Pi)$ の近似分布

7.4 (漸近分布と B -分布に $4/3$ として 2 の検定) 図表と計算 $2/3$
 9.2 Ikeda, Oyawa [15] 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] 北川敬男 (1948). 近代統計学の基盤, 統計教説第2巻 2.
- [5] 北川敬男 (1948). 統計学の組織 第12章, 白楊社
- [6] McCarthy, M. D. (1939). On the application of the t -test to randomized blocks, Ann. Math. Stat. 10.
- [7] Neyman, J. with the cooperation of K. Bierman and S. Kolaczynski (1935). Statistical problems in agricultural experimentation, J. Roy. Stat. Soc., Suppl. 2.
- [8] Oyawa, J. (1961). The effect of randomization on the analysis of randomized block design, Ann. Inst. Stat. Math. 13.
- [9] 北川敬男 (1962). ラテン方格組に $4/3$ の無作為化 (1) 21, 統計雑誌研究所彙報 10.
- [10] Oyawa, J. (1968). On the null-distribution of the F -statistic in a randomized balanced incomplete block design under the Neyman model, Ann. Math. Stat. 34.

- (17). 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. Memo Series 46*, 1-14.
- (18). 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 Belgrade, Yugoslavia, Sept. 1965.
- (19). 岡田 淳, 池田 史郎, 小笠原 基春: Incomplete block design & random-system における F -統計量の漸近分布について, 第34回日本統計学会年会報告
- (20). Ogawa, J. (1965). The theory of block designs, 早稲田大学理工学部大講義の Lecture note
- (21). 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. Memo Series 466*.

(Feb. 12, 1967)