17. 最小自乗法に関するMarkoffの定理

所員 小川潤次郎

J. Neyman : On the different two aspects of the representative method, Journal of the Royal Statistical Society, Vol. XCVII, 1934 を読んだときその Appendix Note II に上記標題の所に証明なしに定理が述べてあり、又 stratified sampling の best linear estimate の variance の計算は省略して "Straight forward algebra gives ……" と云つてあるが、その計算は案外長いのでそれを述べて見よう。もつと簡単な方法があるなら御教示願へれば幸である。

或る母集団 π が n 個の strata $\pi_1, \pi_2, \ \ , \pi_n$ に分れて居り π_i の平均値 A_i, 標準偏差 σ_i とし, A_i は $s\ (<n)$ 個の unknown parameters $p_1, p_2, \cdots, p_s$ の linear forms であつて

$$A_i = a_{i1}p_1 + a_{i2}p_2 + \cdots + a_{is}p_s, \quad i=1,2,\cdots,n \qquad (1)$$

但し a's は known constants とする。このとき $b_1, b_2, \cdots, b_s$ を known constants として, π に関する parameter

$$\theta = b_1p_1 + b_2p_1 + \cdots + b_sp_s \qquad (2)$$

を estimate する問題を考へよう。

π の各 stratum π_i から at random に一つづつ sample を抽出してそれを

$$x_1, x_2, \cdots\cdots, x_n \qquad (3)$$

とするとき, 次の條件を満たす, x_i の linear function θ' を, Markoff に従つて, θ の "The best linear estimate" と呼ばう。

(i) $E(\theta') \doteqdot \theta$ (4)

(ii) $\sigma_{\theta'}^2 = E(\theta - \theta')^2 = minimum$ (5)

今

$$\theta' = \sum_{i=1}^{n} \lambda_i x_i \qquad (6)$$

とすれば (4) なる條件は

$$E(\theta') = \sum_{i=1}^{n} \lambda_i E(x_i) = \theta \qquad (7)$$

が $p_1, p_2, \cdots, p_s$ の如何に拘らず成立することであるから

$$\sum_{i=1}^{n} \lambda_i a_{ij} = b_j,\ j = 1, 2, \cdots, s \qquad (8)$$

と同じである。茲で x_i と x_j の間の相関係数を r_{ij} とすれば

$$\sigma_{\theta'}^2 = \sum_{i=1}^{n} \lambda_i^2 \sigma_i^2 + \sum_{i \neq j = 1}^{n} \lambda_i \lambda_j r_{ij} \sigma_i \sigma_j \qquad (9)$$

であるから "The best linear estimate" を求むるには，(8) なる條件の下に (9) を minimum ならしむる λ_i を求むればよいことになる。

特に $r_{ij} = 0$ で

$$\sigma_i^2 = \frac{\sigma^2}{P_i} \qquad (10)$$

但し P_i は known, σ は unknown ならば (9) は

$$\sigma_{\theta'}^2 = \sigma^2 \sum_{i=1}^{n} \frac{\lambda_i^2}{P_i} \qquad (11)$$

となるから (8) なる條件下に $\sum_{i=1}^{n} \frac{\lambda_i^2}{P_i}$ を minimum にする λ_i を求むればよいのであるが，この場合には次の Markoff の定理が成立つ。

Markoff の定理。

$r_{ij} = 0$, $\sigma_i^2 = \frac{\sigma^2}{P_i}$ の場合には θ の The best linear estimate

は (12) に於て p の代りに

$$S=\sum_{i=1}^{n}(x_i-a_{i1}q_1-a_{i2}q_2\cdots-a_{is}q_s)^2P_i \tag{12}$$

を minimum ならしむる如き x_i の函数

$$q_1^0,\ q_2^0,\ \cdots\cdots,\ q_s^0$$

を代入して得られる。

証明 (8) なる條件の下に

$$\sum\left(\frac{\lambda_i^2}{P_i}\right)$$

を minimum にする λ_i を求め、これを (6) に代入したものと定理に云ふ estimate が一致することを示せばよい。

$\varsigma_1,\varsigma_2,\cdots,\varsigma_s$ を未定常数として

$$Q\equiv\sum_{i=1}^{n}\frac{\lambda_i^2}{P_i}-\sum_{j=1}^{s}\varsigma_j\left(\sum_{i=1}^{n}\lambda_i a_{ij}-b_j\right) \tag{13}$$

を λ_i で偏微分して

$$\frac{\partial Q}{\partial\lambda_i}2\frac{\lambda_i}{P_i}-\varsigma_1a_{i1}-\varsigma_2a_{i2}\cdots-\varsigma_sa_{is}=0 \tag{14}$$

$$i=1,2,\cdots,n$$

故に

$$\lambda_i=\frac{1}{2}(a_{i1}\varsigma_1+a_{i2}\varsigma_2+\cdots+a_{is}\varsigma_s)P_i \tag{15}$$

$$i=1,2,\cdots,n$$

(15) を (8) に代入して

$$\left(\sum_{i=1}^{n}P_ia_{i1}a_{ij}\right)\varsigma_1+\cdots+\left(\sum_{i=1}^{n}P_ia_{is}a_{ij}\right)\varsigma=2b_j$$

$$j=1,2,\cdots,s \tag{16}$$

依って θ の linear estimate (6) は、(15)、(16) で定められる λ_i に対して best となるから、the best linear estimate θ' は

$$\theta'=\sum_{i=1}^{n}\lambda_ix_i=\frac{1}{2}\left(\sum_{i=1}^{n}P_ia_{i1}x_i\right)\varsigma_1+\cdots+\frac{1}{2}\left(\sum_{i=1}^{n}P_ia_{is}x_i\right)\varsigma_s \tag{17}$$

となる。又 (10) を minimum ならしむる q_i^0 は $\frac{\partial S}{\partial q_i^0}=0$ から求められるべきだから 0

$$\left(\sum_{i=1}^{n} P_i a_{ij} x_i\right) = \left(\sum_{i=1}^{n} P_i a_{i1} a_{ij}\right) g_1^0 + \cdots + \left(\sum_{i=1}^{n} P_i a_{is} a_{ij}\right) g_s^0 \qquad (18)$$

$$j = 1, 2, \cdots, s$$

から定められるから

$$\sum_{j=1}^{s} b_j g_j^0 = \frac{1}{2}\left(\sum_{i=1}^{n} P_i a_{i1} x_i\right)\zeta_1 + \cdots + \frac{1}{2}\left(\sum_{i=1}^{n} P_i a_{is} x_i\right)\zeta_s \qquad (19)$$

となって(18)(19)から証明は終了する。

次に母集団πがk個の strata

$$\pi_1, \pi_2, \cdots, \pi_k$$

に分れ，π_iの含む個体の数M_i，π_iの平均値$\bar{u}_i$，標準偏差σ_i 各個体の特性uに関して調べてあるものとしてi番目の stratum π_i の第j番目の個体のuの値をu_{ij}；次に標本としてπ_iより抽出される個体の数をm_iとしそのj番目の個体のuの値をx_{ij}とする

$$\theta = \sum_{i=1}^{k} M_i \bar{u}_i \qquad (20)$$

の the best linear estimate

$$\theta' = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \lambda_{ij} x_{ij} \qquad (21)$$

を求めよう。

その為には

$$E(\theta') = \theta \qquad (22)$$

及び

$$E(\theta' - \theta)^2 = minimum \qquad (23)$$

となる如くλ_{ij}を取ればよい。所で明かに

$$E(x_{ij}) = \bar{u}_i \quad (i = 1, 2, \cdots, k) \qquad (24)$$

であるから(22)式は次の如くなる。

$$\sum_{i=1}^{k}\left(\bar{u}_i\left(\sum_{j=1}^{m_i} \lambda_{ij} - M_i\right)\right) = 0 \qquad (25)$$

(25)式が$\bar{u}_i$の如何に拘らず成立つためには

$$\sum_{j=1}^{m_i} \lambda_{ij} = M_i \quad (i=1,2,\cdots,k) \tag{26}$$

(26) なる條件の下に $E(\theta'-\theta)^2$ を計算する。

$$\theta'-\theta = \sum_{i=1}^{k}\sum_{j=1}^{m_i}\lambda_{ij}x_{ij} - \sum_{i=1}^{k} M_i \overline{u}_i$$

$$= \sum_{i=1}^{k}\left(\sum_{j=1}^{m_i}\lambda_{ij}x_{ij} - \overline{u}_i\sum_{j=1}^{m_i}\lambda_{ij}\right) \quad \text{(26) に依る}$$

$$= \sum_{i=1}^{k}\left(\sum_{j=1}^{m_i}\lambda_{ij}(x_{ij}-\overline{x}_i) + \sum_{j=1}^{m_i}\lambda_{ij}(\overline{x}_i-\overline{u}_i)\right) \tag{27}$$

但し $\overline{x}_i = \dfrac{1}{m_i}\displaystyle\sum_{j=1}^{m_i} x_{ij}$ (28)

$$= \sum_{i=1}^{k}\left\{\sum_{j=1}^{m_i}(\lambda_{ij}-\lambda_i)x_{ij} + M_i(\overline{x}_i-\overline{u}_i)\right\}^2 \tag{29}$$

但し $\lambda_i = \dfrac{M_i}{m_i}$ (30)

$$E\left\{\sum_{j=1}^{m_i}(\lambda_{ij}-\lambda_i)x_{ij} + M_i(\overline{x}_i-\overline{u}_i)\right\}^2$$

$$= \sum_{j=1}^{m_i}(\lambda_{ij}-\lambda_i)^2 E x_{ij}^2 + M_i^2 E(\overline{x}_i-\overline{u}_i)^2 + 2M_i\sum_{j=1}^{m_i}(\lambda_{ij}-\lambda_i)$$

$$E x_{ij}(\overline{x}_i-\overline{u}_i) \tag{31}$$

所で

$$E(\overline{x}_i-\overline{u}_i)^2 = \frac{M_i-m_i}{m_i(M_i-1)}\sigma_i^2 \tag{32}$$

$$E x_{ij}^2 = \frac{1}{M_i}\sum_{j=1}^{M_i} x_{ij}^2 = \frac{M_i}{M_i-1}\sigma_i^2 \tag{33}$$

何者

$$\because \frac{M_i}{M_i-1}\sigma_i^2 = \frac{1}{M_i-1}\sum_{j=1}^{M_i}(x_{ij}-\overline{u}_i)^2$$

$$= \frac{1}{M_i-1}\,\frac{1}{M_i^2}\sum_{j=1}^{M_i}\left\{-x_{i1}-\cdots+(M_i-1)x_{ij}-\cdots-x_{iM_i}\right\}^2$$

$$= \frac{1}{(M_i-1)M_i^2} \sum_{j=1}^{M_i} \left\{ (M_i-1)^2 + M_i - 1 \right\} x_{ij}^2$$

$$= \frac{1}{M_i} \sum_{j=1}^{M_i} x_{ij}^2 = E x_{ij}^2 \qquad (34)$$

又 $\sum_{j=1}^{m_i} (\lambda_{ij} - \lambda_i) = 0$ であるから

$$E(\theta' - \theta)^2 = \sum_{i=1}^{k} E\left\{ \sum_{j=1}^{m_i} (\lambda_{ij} - \lambda_i) x_{ij} + M_i (\bar{x}_i - \bar{u}_i) \right\}^2$$

に代入して

$$\sigma_\theta^2 = \sum_{i=1}^{k} \left\{ \sigma_i^2 \left(m_i \frac{M_i - m_i}{M_i - 1} \lambda_i^2 + \frac{M_i}{M_i - 1} \sum_{j=1}^{m_i} (\lambda_{ij} - \lambda_i)^2 \right) \right\} \qquad (35)$$

従って(26) なる条件の下に (35) の $\sigma_{\theta'}^2$ を minimum ならしむる λ は $j = 1, 2, \cdots, m_i$ に対して

$$\lambda_{ij} = \lambda_i = \frac{M_i}{m_i} \qquad (36)$$

である従つて m_i 個の x_{ij} の平均値を $\bar{x}_i$ とすれば

$$\theta' = \sum_{i=1}^{k} M_i \bar{x}_i \qquad (37)$$

は θ の the best linear estimate であつてその variance は

$$\sigma_{\theta'}^2 = \sum_{i=1}^{k} M_i \frac{M_i - m_i}{m_i} \frac{M_i}{M_i - 1} \sigma_i^2 \qquad (38)$$

となる。

(1947, 5, 1)