パネル調査結果分析のための一致指数

鈴木 達三，高橋 宏一

(1968 年 9 月受付)

On an Index of Consistency of Analyzing the Data from Multi-wave Panels.

Tatsuzo Suzuki, Koiti Takahasi

For studying the reliabilities of responses in social survey it will be good to use the data from multi-wave panels.

The data resulting from such tests are analyzed in datail to find relations of response variability to questionnaire. Then it will be convenient for us to express response variabilities by using an appropriate index. So, we shall propose to use the index defined below.

Let $P=(p_{ij})$ be a $k\times k$ symmetric matrix with non-negative elements, $\sum_{j=1}^{k} p_{ij}\neq 0$ for every i and $\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}=1$. We define the index $C(P)$ of consistency of the matrix P by

$$C(P)=\left(\sum_{i=1}^{k} p_{ii}-\sum_{i=1}^{k} p_i^2\right)\Big/\left(1-\sum_{i=1}^{k} p_i^2\right),\ \text{where } p_i=\sum_{j=1}^{k} p_{ij},\ i=1,2,\cdots k.$$

In § 1 some fundamental properties of the index, in § 2 the expressions of $C(P)$ for some special models and in § 3 the estimator of $C(P)$ are given. The applications of this index to practical data are given in [3].

The Institute of Statistical Mathematics

§ 0 まえがき

同一対象に対し，ある期間をおいて同一内容の調査を実施することにより，回答の信頼性を質問内容との関連に於いて研究しようとするとき，質問に対する回答の変動を表わす適当な指数が要求される．上記の研究目的のためには本質的，傾向的意見の変化が入りこまないように，比較的短期間に前後の調査をおこなう必要があり，その結果得られた資料では，質問ごとの前後調査のクロス集計は周辺分布はほゞ同じ且つ対角線に関しほゞ対称という様子を示すものが多かった([3] 参照)．そこで周辺分布が同一，対角線に関し対称なるクロス集計表に対し，一つの一致率を定義しその性質を調べた．さらにいくつかの特殊な場合について，その一致指数の意味，表現を求めた．

§ 1 一致指数の定義とその基本的性質

$P=(p_{ij})$ を $k\times k$ 行列で，$\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}=1$，$p_{ij}\geqq 0$，$p_{ij}=p_{ji}$，$0<\sum_{j=1}^{k} p_{ij}<1$，$(i, j=1, 2, \cdots, k)$，$k\geqq 2$ なるものとする．かゝる P の全体を $\boldsymbol{P}_k$，$\sum_{j=1}^{k} p_{ij}=p_i$ $(i=1,2,\cdots,k)$ とおく．このとき P の一致指数—$C(P)$ と表わす—を次のように定義する．

$$C(P)=\left(\sum_{i=1}^{k} p_{ii}-\sum_{i=1}^{k} p_i^2\right)\Big/\left(1-\sum_{i=1}^{k} p_i^2\right) \tag{1}$$

また

$$B(P)=1-C(P)=\left(1-\sum_{i=1}^{k} p_{ii}\right)\Big/\left(1-\sum_{i=1}^{k} p_i^2\right) \tag{2}$$

とおく. $\sum_{i=1}^{k} p_{ii}$ は P の主対角要素の和, $\sum_{i=1}^{k} p_{i}^2$ は周辺確率の平方和であるが後者は $p_{ij}^* = p_i \cdot p_j (i, j=1, 2, \cdots, k)$ を要素とする行列 $P^* \in \boldsymbol{P}_k$ の対角要素の和でもある. よって $T(P)=\sum_{i=1}^{k} p_{ii}$, $U(P)=1-T(P)=$非対角要素の和, $V(P)=1-\sum_{i=1}^{k} p_i^2$, とおけば,

$$
\begin{aligned}
(3) \qquad C(P) &= (V(P)-U(P))/V(P) = 1-(U(P)/V(P)) \\
&= (U(P^*)-U(P))/U(P^*) = 1-(U(P)/U(P^*)) \\
&= (T(P)-T(P^*))\,/\,(1-T(P^*))
\end{aligned}
$$

などと表わせる.

さて $C(P)$ の性質をいくつかまとめてみよう.

(イ) $C(P) \leqq 1$

(ロ) $p_{ij}=p_i \cdot p_j (i, j=1,2,\cdots,k)$ なら $C(P)=0$

(ハ) $k=2$ なら (ロ) の逆も成立

$k \geqq 3$ のときは (ロ) の逆はいえない. 例えば $k=3$, $p_{11}=p_{23}=p_{32}=1/3$, 他は0なる P は $\boldsymbol{P}_3$ に属し, 且つ $T(P)=1/3$ で, $U(P)=2/3$, $V(P)=2/3$, $C(P)=0$ で (ロ) の条件はみたさない.

(ニ) $C(P)=1$ なる必要十分条件は $T(P)=1$

(ホ) $p_{ii}/p_i \geqq p_{ji}/p_j \ (i, j=1, 2, \cdots, k)$ なら $C(P) \geqq 0$

(ヘ) $p_{ii} \geq p_{ij} \ (i, j=1, 2, \cdots, k)$ なら $C(P) \geqq 0$

(ホ) と (ヘ) の証明

$$
\begin{aligned}
\sum_{i=1}^{k} p_{ii} - \sum_{i=1}^{k} p_i^2 &= \sum_{i=1}^{k} p_{ii}\Big(\sum_{j=1}^{k} p_j\Big) - \sum_{i=1}^{k} p_i\Big(\sum_{j=1}^{k} p_{ij}\Big) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j \,(p_{ii}/p_i - p_{ij}/p_j) \\
&= \Big[\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j \{(p_{ii}/p_i - p_{ij}/p_j) + (p_{jj}/p_j - p_{ij}/p_i)\}\Big]/2 \\
&= \Big[\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j \{(p_{ii}-p_{ij})/p_i + (p_{jj}-p_{ij})/p_j\}\Big]/2
\end{aligned}
$$

よりわかる.

(ト) 単純一致率 (主対角要素の和) と k が固定されているときは, $C(P)$ は $\sum_{i=1}^{k} p_i^2$ あるいは $p_1, \cdots, p_k$ の分散, すなわち $\left\{\sum_{i=1}^{k} (p_i - 1/k)^2\right\}/k$ の単調減少関数である.

(チ) 周辺確率 $p_1, \cdots, p_k$ あるいは $\sum_{i=1}^{k} p_i^2$ が固定されているときは $C(P)$ は $T(P)$ の単調増加一次関数である.

(リ) $\operatorname*{Inf}_{P \in P_k} C(P) = -1$

証明: $k=2$ のときは $p_{12}=p_{21}=1/2$, $p_{11}=p_{22}=0$ とすることによって実際に $C(P)$ が -1 になりそれが最少値であることは容易にわかる. したがって $\operatorname*{Inf}_{P \in \boldsymbol{P}_k} C(P) \leqq -1$ はいえる.

さて,

$C(P)+1 = \Big(\sum_{i=1}^{k} p_{ii} - 2\sum_{i=1}^{k} p_i^2 + 1\Big)\Big/\Big(1-\sum_{i=1}^{k} p_i^2\Big)$ の分子は $\sum_{i=1}^{k} p_{ii} + \sum_{i=1}^{k} p_i(1-2p_i)$ である. もしすべての i について $p_i \leq 1/2$ なら $C(P)+1 \geq 0$ すなわち $C(P) \geqq -1$ がいえる. そこで $p_1 > 1/2$ とする. 勿論 $p_2, \cdots, p_k$ は 1/2 よりすべて小である. $p_1 = 1/2 + \delta$,

$0 < \delta < 1/2$ とすると $p_1 = 1/2 + \delta = p_{11} + (p_{12} + \cdots + p_{1k}) \leqq p_{11} + (p_2 + \cdots + p_k) = p_{11} + (1/2-\delta)$，したがって $p_{11} \geqq 2\delta$ でなければならない．故に $\sum_{i=1}^{k} p_{ii} + \sum_{i=1}^{k} p_i(1-2p_i) \geqq 2\delta + \sum_{i=2}^{k} p_i(1-2p_i) - \delta(1+2\delta) = \delta(1-2\delta) + \sum_{i=2}^{k} p_i(1-2p_i) > 0$ 証明終．

§2 いくつかのモデルに於ける一致指数の表現

2.1 (Ξ, μ) を確率測度空間とし，各 $\xi \in \Xi$ に対し $\sum_{i=1}^{k} p_i(\xi) = 1$, $p_i(\xi) \geqq 0\ (i=1,2,\cdots,k)$ なる $p_1(\xi), \cdots, p_k(\xi)$ が対応しているものとする．

(4) $$p_{ij} = \int_{\Xi} p_i(\xi) p_j(\xi) d\mu(\xi), \quad (i, j = 1, 2, \cdots, k)$$

と定義する．この p_{ij} を要素とする行列 P は $\left(\int_{\Xi} p_i(\xi) d\mu(\xi) = 0\right.$ なる i が存在する場合を除けば$\left.\right)$ $\boldsymbol{P}_k$ に属する．さて

(5) $$p_{ii} = \int_{\Xi} p_i^2(\xi)\, d\mu(\xi), \quad p_i = \int p_i(\xi) d\mu(\xi), \quad i = 1, 2, \cdots, k$$

であるから，$(p_1(\xi), \cdots, p_k(\xi))$ を k 次元確率ベクトルとして $(Q_1, \cdots, Q_k)$ と書くなら，

(6) $$p_{ii} - p_i^2 = \mathrm{Var}\, Q_i \geqq 0$$

したがって $\sum_{i=1}^{k} p_{ii} - \sum_{i=1}^{k} p_i^2 = \sum_{i=1}^{k} \mathrm{Var}\, Q_i \geqq 0$，すなわち $C(P) \geqq 0$ である．$\sum_{i=1}^{k} \mathrm{Var}\, Q_i = b$, $\sum_{i=1}^{k} \int_{\Xi} p_i(\xi)\ (1 - p_i(\xi)) d\mu\ (\xi) = 1 - \sum_{i=1}^{k} p_{ii} = w$ とおくと

(7) $$C(P) = b/(b+w)$$

である．$(p_1(\xi), \cdots, p_k(\xi))$ を生起確率ベクトルとする試行回数1の多項分布を，分布 μ により混合した混合分布にしたがう確率ベクトルを $(X_1, \cdots, X_k)$ とする．このとき，

(8) $$\begin{aligned} \mathrm{Var}\, X_i &= \int_{\Xi} p_i(\xi)\,(1-p_i(\xi))\, d\mu(\xi) + \int_{\Xi} (p_i(\xi) - p_i)^2\, d\mu(\xi) \\ &= (p_i - p_{ii}) + (p_{ii} - p_i^2) = p_i(1-p_i) \end{aligned}$$

したがって

(9) $$C(P) = \sum_{i=1}^{k} \mathrm{Var}\, Q_i \Big/ \sum_{i=1}^{k} \mathrm{Var}\, X_i$$

とも表現できる．

2.2 $(p_1, \cdots, p_k)$ を $\sum_{i=1}^{k} p_i = 1$, $0 < p_i < 1\ (i = 1, \cdots, k)$ なるものとする．

また，$0 < \varepsilon < 1$ とする．そこで

(10) $$p_{ij} = \varepsilon p_i p_j\ (i \neq j), \quad p_{ii} = p_i\{(1-\varepsilon) + \varepsilon p_i\}, \quad i = 1, 2, \cdots, k$$

と定義すると，この p_{ij} を要素とする行列 P は $\boldsymbol{P}_k$ に属し

(11) $$C(P) = 1 - \varepsilon$$

である．

2.3 $P_1, \cdots, P_N$ を $0 \leqq P_j \leqq 1$, $(j = 1, 2, \cdots, N)$ とする．$\bar{P} = \left(\sum_{j=1}^{N} P_j\right) \Big/ N$ とおく．

X を生起確率 P_j，試行回数1の二項分布を $j = 1, 2, \cdots, N$ について等確率 $1/N$ で混合した分布にしたがう確率変数とすると，

(12) $$\mathrm{Var}\, X = \bar{P}(1-\bar{P}) = \sum_{j=1}^{N} P_j(1-P_j) \Big/ N + \sum_{j=1}^{N} (P_j - \bar{P})^2 \Big/ N.$$

Hansen et al. [1] は

$$(13)\qquad \left\{\sum_{j=1}^{N} P_j(1-P_j)\Big/N\right\}\Big/\{\bar{P}(1-\bar{P})\}$$

を不一致性の指数として提案している．いま $(X_1,\ X_2)$ を生起確率 P_j，試行回数1の二項分布の2個の直積を $j=1,2,\cdots,N$ について等確率 $1/N$ で混合した混合分布にしたがう確率ベクトルとし，$p_{ij}=Pr\{X_1=2-i,\ X_2=2-j\}$，$(i,\ j=1,\ 2)$ とすれば，p_{ij} を要素とする 2×2 の行列 P は $\boldsymbol{P}_2$ に属し，$\sum_{i=1}^{2} p_{ii}=\sum_{j=1}^{N}(P_j{}^2+(1-P_j)^2)\Big/N$，$\sum_{i=1}^{2} p_i{}^2=\bar{P}^2+(1-\bar{P})^2$ であるから

$$(14)\qquad C(P)=1-\left\{\sum_{j=1}^{N} P_j(1-P_j)\Big/N\right\}\Big/\{\bar{P}(1-\bar{P})\}$$

となり1から (13) の不一致指数をひいたものである．

2.4 $(\rho_1,\cdots,\rho_k)$，$(a_{i1},\cdots,a_{ik})$，$(i=1,2,\cdots,k)$ を

$$0<\rho_i<1,\ \sum_{i=1}^{k}\rho_i=1,\ 0\leq a_{ij}\leq 1,\ \sum_{j=1}^{k}a_{ij}=1,\ \ (i,\ j=1,2,\cdots,k)$$

とする．$p_{ij}=\sum_{l=1}^{k}\rho_l\, a_{li}\, a_{lj}$，$(i,\ j=1,\ 2,\cdots,\ k)$ とおくと p_{ij} を要素とする行列 P は $\boldsymbol{P}_k$ に属する．このとき

$$(15)\qquad C(P)=\sum_{i=1}^{k}\left\{\sum_{l=1}^{k}\rho_l\, a_{li}{}^2-\left(\sum_{l=1}^{k}\rho_l\, a_{li}\right)^2\right\}\Big/\left\{1-\sum_{i=1}^{k}\left(\sum_{l=1}^{k}\rho_l\, a_{li}\right)^2\right\}$$

とくに $k=2$ なら $\rho_1=\rho$，$a_{11}=\alpha$，$a_{22}=\beta$ とおいて，

$$(16)\qquad \sum_{i=1}^{2}p_{ii}-\sum_{i=1}^{2}p_i{}^2=2\rho(1-\rho)(\alpha+\beta-1)^2$$

$$(17)\qquad 1-\sum_{i=1}^{2}p_i{}^2=2\{\rho(\alpha+\beta-1)+(1-\beta)\}\{(1-\rho)(\alpha+\beta-1)+(1-\alpha)\}$$

$$(18)\qquad C(P)=\rho(1-\rho)(\alpha+\beta-1)^2/[\{\rho(\alpha+\beta-1)+(1-\beta)\}\{(1-\rho)(\alpha+\beta-1)+(1-\alpha)\}]$$

2.5 一つの equilibrium alternating renewal process を考える ([2] 第7章参照)．Type 1 成分の寿命密度函数を $f_1(x)$，平均を μ_1，Type 2 成分のそれらを $f_2(x)$，μ_2 とする．時刻0で Type i，時刻 t で Type j $(i,\ j=1,\ 2)$ が使用されている確率を $p_{ij}(t)$ とおくとき $p_{ij}(t)$ を要素とする行列 $P(t)$，$t\geqq 0$ は $\boldsymbol{P}_2$ に属する．そして

$$(19)\qquad C(P(t))=(\mu_1+\mu_2)\omega(t)/(\mu_1\mu_2)$$

である．ここで $\omega^*(s)=\mu_1\mu_2/\{(\mu_1+\mu_2)s\}-\{1-f_1{}^*(s)\}\{1-f_2{}^*(s)\}/\{s^2(1-f_1{}^*(s)f_2{}^*(s))\}$，$\omega^*(s)$，$f_1{}^*(s)$，$f_2{}^*(s)$ はそれぞれ $\omega(t)$，$f_1(t)$，$f_2(t)$ のラプラス変換である．とくに，$f_1(x)$，$f_2(x)$ が指数分布なら $\omega(t)=\mu_1\mu_2\exp\{-(\mu_1+\mu_2)t/(\mu_1\mu_2)\}/(\mu_1+\mu_2)$ であるから

$$(20)\qquad C(P(t))=\exp\{-(\mu_1+\mu_2)t/(\mu_1\mu_2)\}$$

2.6 $Q=(q_{ij})$，$i,\ j=1,2,\cdots,k$，$\sum_{j=1}^{k}q_{ij}=1$，$q_{ij}\geqq 0$，$i,\ j=1,\ 2,\cdots,\ k$ とし，さらに $(a_1,\ \cdots,\ a_k)$ を $\sum_{i=1}^{k}a_i=1$，$a_i>0$，$i=1,\ 2,\cdots,\ k$，なるものとし

$$(21)\qquad (a_1,\ \cdots,\ a_k)Q=(a_1,\ \cdots,\ a_k)$$

$$(22)\qquad Q=\Lambda D\Lambda'$$

$\Lambda=(\lambda_{ij})$ は $k\times k$ 直交行列，$D=(d_{ij})$ は $k\times k$ 対角行列とする．

$n=1,\ 2,\cdots$ に対し，

$$p_{ij}(n) = a_i q_{ij}^{(n)},\ i,\ j = 1,\ 2, \cdots, k \tag{23}$$

但し $q_{ij}^{(n)}$ は Q^n の (i, j) 要素，とおくと，$p_{ij}(n)$ を要素とする行列 $P(n)$ は $\boldsymbol{P}_k$ に属する．このときは，

$$C(P(n)) = \left\{ \sum_{i=1}^{k} \left(\sum_{l=1}^{k} a_i \lambda_{il}^2 d_l^n \right) - \sum_{i=1}^{k} a_i^2 \right\} \Big/ \left(1 - \sum_{i=1}^{k} a_i^2 \right) \tag{24}$$

である．

§ 3 一致指数の推定

$P \in \boldsymbol{P}_k$ とする．$\{n_{ij}\}$, $(i, j = 1, 2, \cdots, k)$ を試行回数 n，生起確率 $\{p_{ij}\}$ の多項分布からの標本とする．$\hat{p}_{ij} = (n_{ij} + n_{ji}) / (2n)$，$\hat{p}_i = \sum_{j=1}^{k} \hat{p}_{ij}$, $(i, j = 1, 2, \cdots, k)$ とし，

$$\hat{C} = \left(\sum_{i=1}^{k} \hat{p}_{ii} - \sum_{i=1}^{k} \hat{p}_i^2 \right) \Big/ \left(1 - \sum_{i=1}^{k} \hat{p}_i^2 \right) \tag{25}$$

とする．このとき，$(C(P)$，$U(P)$，$V(P)$ を C, U, V と略記して)

$$E(\hat{C}) = C - 2n^{-1} \left\{ 2V \left(1 - \sum_{i=1}^{k} p_i p_{ii} \right) - 4U \left(1 - \sum_{i=1}^{k} p_i^3 \right) - 2V^2 + 6UV - UV^2 \right\} V^{-3} + o(n^{-1}) \tag{26}$$

$$E(\hat{C} - C)^2 = n^{-1} V^{-4} \left\{ 4UV \left(1 - \sum_{i=1}^{k} p_i p_{ii} \right) - 4U^2 \left(1 - \sum_{i=1}^{k} p_i^3 \right) + 4U^2 V - U^2 V^2 - 3UV^2 \right\} + o(n^{-1}) \tag{27}$$

である．

統計数理研究所

参 考 文 献

[1] Hansen, M. H., Hurwitz, W. N. and Pritzker, L. (1964) : The estimation and interpretation of gross differences and the simple response variance, 「Contributions to Statistics Presented to Professor P. C. Mahalanobis on the Occasion of His 70th Birthday.」 Pergamon Press, Calcutta, India, 111-136.

[2] Cox, D. R. (1967) : 「Renewal Theory」, London : Methuen.

[3] 鈴木達三 (1968) : 面接調査における回答の安定性について，「統計数理研究所彙報」 第 16 巻 1 号，47-102