# 数量化理論とその応用例 (VI)*

林　　知　己　夫

(1961 年 8 月 受 付)

## Theory of Quantification and its Examples (VI)

Chikio HAYASHI

In the present paper, a method of generalization to multidimensional case of "$e_{ij}$"-type quantification (Ann. of I. S. M.) is described. Futhermere, the relation between the method of $e_{ij}$-type quantification and that of factor analysis and the relation between factor analysis and discrimination analysis is shown.

(1) A method of multidimensional quantification of "$e_{ij}$"-type problem is based on the euclidean distance between $i$ and $j$ in multidimensional space in stead of the square of difference between $i$ and $j$ in one dimension. We consider the measure $Q=-\sum_i\sum_j e_{ij}\left\{\sum_s^S \frac{({}^sx_i-{}^sx_j)^2}{\sigma^2{}_{{}^sx}}\right\}$, where $e_{ij}$ is numerical value to represent the relation between $i$ and $j$ and ${}^sx_i$ is to be given to the element $i$ in the $s$-th dimension; $i=1, 2, \cdots R$, $R$ being the size of objects, $s=1, \cdots S$, S being the number of dimension. It is reasonable for our purpose to require ${}^sx$'s, $s=1 \cdots S'$ to maximize $Q$, where the variances of ${}^sx$'s, $s=1, \cdots S$ being equal to 1, respectively. The solutions are given as the characteristic vectors corresponding to the characterestic roots $\lambda_1>\lambda_2>\cdots>\lambda_S$ (in the successive descending order of magnitude of characteristic roots) in the characteristic equation obtained in uni-dimensional case.

(2) Let $e_{ij}$ be the correlation coefficent between $i$ and $j$, the variances of all $i$ ($i=1, \cdot\cdot R$) being equal to 1. The characteristic equation in factor analysis (principal component method for variance-covariance matrix) is obtained when one term obtained by developing $Q$ is maximized with respect to $x$'s. Thus the structural relation between them is shown.

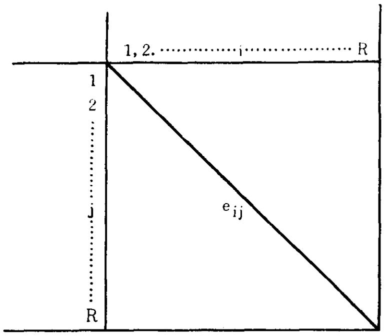
(3) Formally speaking, the statistical method of factor analysis is interpreted as a kind of discriminant analysis. The first factor is interpreted as the most efficient linear solution of discrimination. The second factor is as the most efficient, exept the first factor, linear solution, and so on. Considering from the point of view of multidimensional quantification, the formal principal component solution is obtained as the meaningful result of maximization of a measure, that is statistically reasonable, in multidimensional quantification.

Institute of Statistical Mathematics

* 文部省科学研究費による研究の一部である.

ソシオメトリーにおけるマトリックス表現から個々の要素を組分けするに際して，数量化の考へを用いることについては，すでに述べた*．ここでは，用いられている“$e_{ij}$ 型数量化”についてさらに検討したところを述べてみよう．なおサムプリングの問題については，Note on Sampling from a Sociometric Pattern, Annals of I. S. M., Vol. IX, 1957, にのべてあるが，同様のものを別角度から取扱ったものに，Leo. A. Goodman : Snow Ball Sampling, Ann. of M. S. Vol. 32, 1961, がある．

問題の定義を念のため，一度書いてみよう．ある集団（大きさ $R$）において，夫々×つづつのものの間の親近性をあらわす数量 $e_{ij}\gtreqless 0$, $i\neq j$ が与えられている場合，それらが，親近性のあるものが近く，それの少いものが相離れる様に分類することを考えるのである．このとき $e_{ij}$ が大なるほど親近性が強く，小なるほど，親近性が弱いとしておく．$e_{ii}$ は定義しないで，存在しないものとしておく．$i$ なるものに $x_i$ と言う数量を与えることによって，これを用いて分類を考えるのである．このため，$Q=-\sum\sum_{i\neq j}^{R} e_{ij}(x_i-x_j)^2$ なるものを測度とし，これを，$x$の分散，$(1/R)\sum x_i^2-\bar{x}^2$ $(\bar{x}=(1/R)\sum_i^R x_i)$, を一定，また $\bar{x}=0$, の条件の下に最大にする様に $x_i(i=1,\cdots R)$ を定めることを考えた．これについては，今述べないが，この変奏曲的応用として，態度数量化の一方法 II（彙報，第6巻，第2号）においてその一例を示した．なお，特性根として解を求めるのであるが，$e_{ij}$ が正負混合する場合，絶対値の最大根が求めるものとはならないので $e_{ij}+c<0$（$c$ はある常数）とする様に $c$ を任意にさだめても解が変らないことが示されれば一般に便利になる（$Q$ は常に正となるので最大根が求めるものになる）のである．任意の $c$ を加へてもよいことが，上記のものに示されている．




註 $R$ は要素の数，一般に $e_{ij}\neq e_{ji}$

以上の所では，1次元的な分類を考えたのであるが，ここでは，それを多次元空間の中において分類する事を考えよう．まず簡単のため2次元の数量を与えて，2次元的空間の中に分類する事を試みよう．これが出来れば次元をあげることは全く同様であらためて書く必要もない．$i$ に $(x_i, y_i)$, $j$ に $(x_j, y_j)$ なる数量を与えるものとする．$i$ なるものと $j$ なるものとの間の距離としてユークリッドの距離，$(x_i-x_j)^2+(y_i-y_j)^2$ を与えるものとする．

しかしここで，$x$ と $y$ の寸法を考える必要がある．そこで寸法として $\sigma_x$, $\sigma_y$ を単位に測るものとする．そうして，全体の測度として

$$G=-\sum_i\sum_j e_{ij}\left\{\frac{(x_i-x_j)^2}{\sigma_x^2}+\frac{(y_i-y_j)^2}{\sigma_y^2}\right\}$$

を考え，これを一種の直交条件 $(HX, Y)=0$ の下に最大にする事を考える（$H$ は1次元の時得られる特性方程式に現われる対称マトリックス，$X$, $Y$ は $x$, $y$ に応ずるヴェクトル，$(\cdot\cdot)$ は内積とする）．

---

* C. Hayashi : On the Prediction of Phenomena from Qualitative data and the Quantification of Qualitative Data from the Mathematico-Statistical Point of view, Ann. of I. S. M. Vol. III, 1952, なほ，前の論文中 $Q_{ij}=\sum_{\substack{k\\k\neq i\\k\neq j}}^{R}\frac{1}{(e_{ik}-e_{jk})^2}$ の様な形のものは $Q_{ij}=\frac{1}{\sum_{\substack{k\\k\neq i\\k\neq j}}^{R}(e_{ik}-e_{jk})^2}$ の様な形の誤植のため訂正する．

このとき，これを最大にすべき $X$, $Y$ は1次元の時と同様に

$$HX=\lambda X,\ \lambda=-\frac{\sum\sum e_{ij}(x_i-x_j)^2}{\sigma_x^2} \qquad (*)$$

$$HY=\mu Y,\ \mu^2=-\frac{\sum\sum e_{ij}(y_i-y_j)^2}{\sigma_y^2} \qquad (\ast\ast)$$

を満足すべきことを知る．したがって $X$ と $Y$ とは直交し $(X,\ Y)=0$ となる．したがって解の数を頭におけば，解は，$X\neq Y$ とすれば一般に $N(N-1)$ 個ある．$G=\lambda+\mu$ である．そこで $H$ に関する特性方程式の最大の特性根を $\lambda_1$, 次のを $\lambda_2$ とし，例えば $HX=\lambda_1 X$, $HY=\lambda_2 Y$ とすれば $G=\lambda_1+\lambda_2$ となる．これが最大となることは明らかである．$H$ が対称であることからみて $X$ と $Y$ とが直交する事は明らかである．

なお，$x$ と $y$ との関係をやや一般化し，距離を $\alpha^2\frac{(x_i-x_j)^2}{\sigma_x^2}+\beta^2\frac{(y_i-y_j)^2}{\sigma_y^2}$ で与えるものとしても全く同様に，$X$, $Y$ を求めることができる．但し $\alpha$, $\beta$は常数．

なお，前頁の条件として直交条件 $(X,\ Y)=0$ をおいても，一般に $H$ の特性根を $\varphi$ とすれば $(\varphi-\lambda)(\varphi-\mu)\geqq 0$ を満足すべきことから等号が成立すべきであるし，したがって上述の場合と同一の結果を得る．なお，$\lambda, \mu$ は (*), ($\ast\ast$) の右側の式で定義されるものである．

この様に考えれば何次元であっても同様であって，1次元の場合の特性方程式の大きい方からの特性根に相当する特性ベクトルを求めてゆけば，夫々の次元に応ずるベクトルを得ることになり，話は非常に簡単になる．

また，この様な関係は $G$ として $x$ に関する $G$ 即ち $G_x=-(\sum_i\sum_j e_{ij}(x_i-x_j)^2/\sigma_x^2)$ と $y$ に関する $G_y$ を求め $G=(-G_x)\cdot(-G_y)$ を考え，これを $X\neq Y$ で $X$ と $Y$ とが直交する $(X,\ Y)=0$ の条件の下で最大にすることを考えれば $G_x\neq G_y$ なる限り，上述のものと全く同様な結果を得る．測度として，前述の様な $G$ をとるのがよいか，今度の様な積の形の $G$ をとるのがよいかは問題の性質によるもので，後者の $G$ は分類の機能だけを考えての測度であり，前者には異った意味をもたせることも可能なのである*.

我々の $e_{ij}$ は親近性をあらわすものであれば何でもよかった．場合によっては $e_{ij}$ として，($e_{ii}$ が定義できるとしておく）何か $i$ と $j$ との相関係数 $r_{ij}$（または，$1-(1-r_{ij}^2)=r_{ij}^2$）を用いても差支えはない．この意味では我々の場合は一般的なものと言いうるのである．

いま1次元の場合の $G$ として

$$G'=\frac{\sum_i\sum_j^R e_{ij}x_i x_j}{\frac{1}{R}\sum_i^R x_i^2}$$

と考えてみよう．これは，我々の $G$ の分解としての一項としてあらわれてくるものである．前述の我々の場合，分母として $\sigma_x^2$ をとろうが $(1/N)\sum x_i^2$ をとろうが同じ結果となる事を考えて，

即ち

$$G=-\frac{\sum_i\sum_j^R e_{ij}(x_i-x_j)^2}{\frac{1}{R}\sum_i^R x_i^2}$$

---

* 林知己夫，高倉節子，橋爪浅治，梅津正熙，高橋正祺，病院における診療配置法，公衆衛生院報告　第9巻，第1号，1960，及び前掲彙報　第6巻，第2号．

$$=-\frac{\sum_i^R (e_{i.}+e_{.i})x_i^2}{\frac{1}{R}\sum_i^R x_i^2}+2\frac{\sum_i^R\sum_j e_{ij}x_ix_j}{\frac{1}{R}\sum_i^R x_i^2}$$

但し，　$e_{i.}=\sum_j e_{ij}$, $e_{.i}=\sum_j e_{ji}$ となる.

この第2項の変量のみを考えれば

$$G''=\frac{\sum_i\sum_j e_{ij}x_ix_j}{\sum_i x_i^2}$$

となるから，

これを最大にすることを考えればこれは $e_{ij}$ からつくられるマトリックスを $J$ とすると

$$JX=G''X$$

をみたすものであることがわかる. $e_{ij}=r_{ij}$ と考えれば，相関行列の主要素を求める式となり因子分析の一つの形となる（コミュアリティを1とした場合の特殊なもの，この因子分析としての妥当性の検討は大事なものであるが，ここでは触れない）即ち因子分析の一つの解釈となる.

この式が，我々の場合の一項目として入っているのである. 我々の前に述べた数量比の場合は，$G$ を大にしようとするのであるから第1項の $(e_{.i}+e_{i.})$ の大なるものほど $x_i$ 小さな値をとるべき機能をも付与しているものである.

いま $i$ なるもの数量を $\alpha_i$ とし，これが $R$ 個の数量的測定結果 $z_k(k=1, \cdots R)$ のウェイトつきの和を考えよう.

$$\alpha_i=\sum_k^R w_kz_k(i)$$

ここに $z_k(i)$ は $i$ なるものの $k$ 番目の測定で示される数量的測定結果とする. $i=1, 2, \cdots N$ とする.

また，$(1/N)\sum_i^N z_k(i)=0$ としておく.

ここで，$w_k$ なる $R$ 個のウェイトをきめる問題を考えよう. 測定は個々のものを弁別するために行うべきものであるから，測定のウェイトつきの線型和を出すのは，なお一層個々のものが弁別される様に考えることであり，このためウェイトを求めようとするのである. このためウェイト全体の寸法を一定とし，即ち $\sum_k^R w_k^2$ を一定にし，$\alpha_i$ の分散，$\sigma_\alpha^2=(1/N)\sum_i^N \alpha_i^2$：$\bar{\alpha}=(1/N)\sum\alpha_i=0$：が最大になる様に（個々のものの弁別が最大になることの一表現である），考えれば，これは

$$\frac{\sum_k\sum_l C_{kl}w_kw_l}{\sum_k^R w_k^2}$$

を最大にすることと等価になる. ここに $C_{kl}$ は $z_k$ と $z_e$ との共分散とする.

これを計算してみると

$$\sum_{l=1}^R C_{sl}w_l=\nu w_s \quad s=1, 2, \cdots R$$

なる特性方程式を解き特性根 $\nu$ の最大を求めることに帰著する. ウェイト $w$ として多次元ヴェクトルを与える場合は前述の様に夫々大きいものから順に考えた特性根に応ずる特性ヴェクトルを求める事が筋の立つ考え方であることが，この論述の後の部分で示される.

いま，ウェイトとして $w_s'=w_s/\sigma_{Zs}$：$\sigma_{zs}$ は $z_s$ の標準偏差：を考える. そして $w_s'$ について同様の式を書きあげてみると

$$\alpha_i=\sum_k^R w_k' z_k(i)$$

を考え，これを $\sum_k^R \sigma^2 z_k w_k'^2$ を一定の下に最大にする事を考える．これは

$$\frac{\sum_k\sum_l r_{kl} w_k w_l}{\sum_k^R w_k^2}$$

を最大にするウェイトを述めることになる．なお，これは $z_k$ の分散をすべて1とした場合に相当する．

この後者の考え方は，線型和を求めるためのウェイト $w_k'$ と一定にすべきもの $\sum_k^R \sigma^2 z_k w_k'^2$ とがそぐわず，このままでは作為的な感を与えている．また分散を一定にしておくとき，前の場合に一致し $C_{kl}=r_{kl}$ となるが，弁別のためには $z$ の分散を一定にする理由の妥当性を十分検討してかからねばならない．測定の性質からみて，弁別のためには，すなおに $C_{kl}$ を用いる方が望ましい場合もあるのである．この検討には，測定したものがたとえ数量化されたものであっても，これをアイテム・カテゴリーに分類し*，この質的パタンを用いて，この論述の最後にのべる数量化の方法によって，弁別の問題を考えるのが妥当である．この結果によって，弁別と言う観点からすれば夫々の項目（$z_k$ の）の分散がどうなるかが，明確に示されることになるのである．即ち $\sigma z_k$ と数量化された分散との関係が明確になるのである．（数量化の立場よりすれば，常に，アイテム・カテゴリー型の反応として，ものごとを捉えたいのであるが，計算の利便のために，$z_k$ と言う数量そのものをそのまま用いることもあるわけである）．

さて，今のべた場合（$z_k$ の分散は常に 1）を考えれば，ウェイトは，相関行列の主要素解を求めることに帰着する．即ちここでも，測定の綜合と個々のものの弁別と言う観点から因子分析の一つの解釈が得られたことになる．今まで述べたことを遡れば，これは我々の場合の$G$といかなる関係にあるかが了解せられよう．なお，しかし，これは $w$ に関する制限条件を上述の様な形にして，弁別効果を最大にするようにして得られたものに過ぎないので，制限条件を変化させる（これは問題の妥当性からきめるべきである）ならば当然の事ながら異った形のものとなる．したがってこれは，弁別による一つの解釈であって，常に弁別と主要素解とは対応するものではない．

以上は一まづ1次元的に考え，これからみちびかれる特性方程式から，相関行列の主要素解を求めることとの関係をみたのである．いま，ここでは最初から，多次元的な $w_k$ として ${}^s w_k$ $(s=1, 2, \cdots S)$ を与えるものとし，${}^s\alpha_i=\sum_k^R {}^s w_k z_k(i)$，$i=1, 2, \cdots S$ を考えよう．このとき $z$ の平均を 0 としておく．即ち $(1/N)\sum_i z_k(i)=0$，$k=1, 2, \cdots R$．ここで多次元的な分散として一般化された分散

$$\begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1S} \\ \sigma_{21} & \sigma_{22} & \cdots\cdots & \sigma_{2S} \\ \\ \sigma_{S1} & \sigma_{S2} & \cdots\cdots & \sigma_{SS} \end{vmatrix}, \quad \sigma_{rs}=\frac{1}{N}\sum_i^N {}^r\alpha_i {}^s\alpha_i$$

を考えよう．いま簡単のため $S=2$ としておく．そこで従前の考え方を踏襲し，夫々 $\sum_k^R {}^s w_k^2$，$s=1$，2 を一定の下に一般化された分散を最大にする，即ち

$$\frac{\begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix}}{(\sum_k^R {}^1 w_k^2)(\sum_k^R {}^2 w_k^2)}=\lambda^2, \quad \sigma_{12}=\sigma_{21}$$

を最大にすることを考えよう．このことは，現在の所1次元の場合の一つの自然な拡張として，

---

* 数量化と予測の根本概念，彙報　第7巻，第1号．

考えてよいことと思われる（最善と言うのではない，現状の理論から言って一つの合理的な考え方であると思われる）．このいみで，現実的な妥当性をもつと認められよう．なお，このとき，さらに，条件として $\sigma_{12}=0$ をつけ加えることにする．一つの直交条件である．

$$\sigma_{rs}=\frac{1}{N}\sum_{i}^{N}{}^{r}\alpha_i{}^{s}\alpha_i=\frac{1}{N}\sum_{i}^{N}\left(\sum_{l}\sum_{m}{}^{1}w_l{}^{2}w_m z_l(i)z_m(i)\right)$$

$$=\sum_{l}\sum_{m}C_{lm}{}^{1}w_l{}^{2}w_m,\quad C_{lm}\text{ は共分散.}$$

$$\sigma_{11}=(C^1W,\ {}^1W),\quad \sigma_{12}=(C^2W,\ {}^2W)$$

$$\sigma_{12}=(C^1W,\ {}^2W)=({}^1W,\ C^2W),$$

$$\lambda^2=\frac{(C^1W,\ {}^1W)}{({}^1W,\ {}^1W)}\cdot\frac{(C^2W,\ {}^2W)}{({}^2W,\ {}^2W}$$

なお，$C$ は分散・共分散のマトリックス，${}^1W$, ${}^2W$ は夫々 ${}^1w_l$, ${}^2w_l$ $(l=1,\cdots R)$ をあらわすヴェクトルとする．また，$(A, B)$ はヴェクトル $A, B$ の内積をあらわす．我々の問題は $\lambda^2+\mu\{(C^1W,\ {}^2W)-0\}$，を最大にする様なヴェクトル ${}^1W$, ${}^2W$ を求めることになる．

これを書き下してみると，${}^1W$, ${}^2W$ は，

$$\sigma_{11}C^1W=\lambda^2({}^2W,\ {}^2W){}^1W+\mu({}^1W,\ {}^1W)({}^2W,\ {}^2W)C^2W \tag{1}$$

$$\sigma_{22}C^2W=\lambda^2({}^1W,\ {}^1W){}^2W+\mu({}^1W,\ {}^1W)({}^2W,\ {}^2W)C^1W \tag{2}$$

を満足しなくてはならない．

(1) と ${}^2W$ との内積，(2) と ${}^1W$ との内積をつくり，$(C^1W,\ {}^2W)=0$ $(C^2W,\ {}^1W)=0$ を考へに入れると夫々

$$0=\lambda^2({}^2W,\ {}^2W)({}^1W,\ {}^2W)+\mu({}^1W,\ {}^1W)({}^2W,\ {}^2W)(C^2W,\ {}^2W)$$

$$0=\lambda^2({}^1W,\ {}^1W)({}^2W,\ {}^1W)+\mu({}^1W,\ {}^1W)({}^2W,\ {}^2W)(C^1W,\ {}^1W)$$

となる．

したがって

$$\mu\left\{\frac{(C^1W,\ {}^1W)}{({}^1W,\ {}^1W)}-\frac{(C^2W,\ {}^2W)}{({}^2W,\ {}^2W)}\right\}=0$$

$$\frac{(C^1W,\ {}^1W)}{({}^1W,\ {}^2W)}\fallingdotseq\frac{(C^2W,\ {}^2W)}{({}^2W,\ {}^2W)}\ \text{とすれば}\ \mu=0.$$

こうすると (1)，(2) から

$$C^1W=\left[\lambda^2\frac{({}^2W,\ {}^2W)}{\sigma_{11}}\right]{}^1W=\left[\frac{\lambda^2\sigma_{22}}{\sigma_{11}}\right]{}^1W=\nu_1{}^1W \tag{3}$$

$$C^2W=\left[\lambda^2\frac{({}^1W,\ {}^1W)}{\sigma_{22}}\right]{}^2W=\left[\frac{\lambda^2\sigma_{11}}{\sigma_{22}}\right]{}^2W=\nu_2{}^2W \tag{4}$$

をとけばよいことになる．

これから直ちにわかることは，(3) をみたす ${}^1W$ は一般に $R$ 個ある，また (4) をみたす ${}^2W$ は一般に $R$ 個あることである．${}^1W\fallingdotseq{}^2W$ であるから (3)，(4) をみたす解は $R(R-1)$ 個となる．なお (3)，(4) をみたすものは

$$CW=\varphi^2W \tag{5}$$

をみたすことがわかるし，$\nu_1\nu_2=\lambda^2$となることから，${}^1W$, ${}^2W$ は夫々 (5) の最大根 $\varphi_1{}^2$，その次の根 $\varphi_2{}^2$，に応ずるヴェクトルであることがわかる．そして最大の $\lambda^2$ は $\varphi_1{}^2\varphi_2{}^2$ となることがわかる．なお，このとき $({}^1W,\ {}^2W)=0$ となる．前提条件は $(C^1W,\ {}^2W)=0$ であるが $C^1W=\nu_1{}^1W$ から $(C^1W,\ {}^2W)=\nu_1(W,\ {}^2W)$ となり通常の直交条件となる．

以上の事は，1次元の数値を与えようとして求めた場合の特性方程式をその特性根の大きさの順の次々に求めたものと一致することを示したことになる．1次元の場合において次々に根

を求めるとこは，そのままでは意義がなく，最大根を求めることにのみ意味があったのであるが（求めるものを最大にするために，微分法を用いたに過ぎない，最大のみが我々にとって直接的な意味がある．それにも拘らずそれ以外の根を求める意味はどこにあるか．），ここで示したことは，$\lambda^2$ を最大にする様に考えた場合に，それが，夫々，1次元の場合の特性方程式の特性根の大きい順のものに一致し，したがって最大でない根を求めることの意味のあることが明瞭になったことである．言わば，分散・共分散行列の形式的な主要素解は，我々の作くった測度を最大にする場合の解に対応することになったわけである．即ち主要素解を求めると言う形式的な方法が——最大根のみを求める場合は我々の問題意識と一致する——一般多次元的数量化における最大問題と言う我々にとって妥当な問題意識に一致することが示されたことになる．なお，測度を変更すれば，一般的に言って当然の事ながら，主要素法の解は我々の最大問題の解と一致することはないのである．我々にとっては，形式的な立場を離れ，測度を問題に応じ妥当性を以て，合理的に思考し作成することが極めて大切なことなのである．ここに注意する必要がある．

なお $S$ が3以上の場合においても，同様の条件の数を $\binom{S}{2}$ だけ附加することによって，我々の求めるものは主要素法の次々の解となってあらわれてくることになる．

なお因子分析に対する上述の解釈，即ち個々のものを弁別しようとする機能——多次元空間の中の限られた範囲の中で近い関係のもの同志を近く，遠い関係のもの同志を離れて位置づけようとこころみる立場——は，質的パタンよりする因子分析的方法開発への根本的示唆を与えている．この考えの下に行われたものに関しては，堀明子，数量化理論の適用による聴取者層の分析，放送文化研究所報告，No. 5, 1960, 及び統計数理研究所国民性調査委員会，日本人の国民性，至誠堂，1961 を参照されたい．

ここで述べた考えを進めるに当って当研究所の高倉節子嬢の協力に俟つ所が多かった．ここに深く感謝の意を表する．

（統計数理研究所）