最有効確率抽出法と回帰関数比例確率抽出法*

統計数理研究所　田　口　時　夫

(1975年12月　受付)

On the Most Efficient Sampling and the Sampling with Proportional to Value of a Regression Function

Tokio Taguchi
(The Institute of Statistical Mathematics)

Let us consider the probability sampling proportional to a prositive one-valued function $\varphi(\boldsymbol{Y})$ of the given supplemental variable $\boldsymbol{Y}=(Y_1,\cdots,Y_k)$. Let X be the main variable and $(X^{(i)}, \boldsymbol{Y}^{(i)});i=1,\cdots,n^*$ be n-samples with this sanpling scheme.

Then, consider the unbiased estimate $\tilde{X}(\varphi)$ of the mean value $E(X)$ given by

$$\tilde{X}(\varphi)=\frac{1}{n}E\{\varphi(\boldsymbol{Y})\}\sum_{i=1}^{n}E\left\{\frac{X^{(i)}}{\varphi(\boldsymbol{Y}^{(i)})}\right\} \tag{1}$$

Then, the variance $\sigma^2_{\tilde{X}(\varphi)}$ of $\tilde{X}(\varphi)$ is given by

$$\sigma^2_{\tilde{X}(\varphi)}=\frac{1}{n}\left[E\{\varphi(\boldsymbol{Y})E\left\{\frac{X^2}{\varphi(\boldsymbol{Y})}\right\}-E^2(X)\right]. \tag{2}$$

Now, if there exists a minimizor $\varphi^{(0)}$ minimizing $\sigma^2_{\tilde{X}(\varphi)}$ in the set of all φ, then we define the most efficient probability sampling by the proportional to $\varphi^{(0)}(\boldsymbol{Y})$.

In particular, when the positive valued X is given by

$$\log X=\psi(\boldsymbol{Y})+\log\zeta, \tag{3}$$

where $\zeta \perp\!\!\!\perp Y_j; j=1\cdots k$ such $\varphi^{(0)}(\boldsymbol{Y})$ always exists and can be shown by

$$\varphi^{(0)}(\boldsymbol{Y})=e^{\psi(\boldsymbol{Y})}. \tag{4}$$

And then, the coefficient of variation $C^{(0)}$ of $\tilde{X}(\varphi^{(0)})$ satisfies

$$C^{(0)}=\frac{1}{\sqrt{n}}C_\zeta, \tag{5}$$

where C_ζ expresses the coefficient of variation of ζ.

On the otherhand, $\varphi^{(0)}(\boldsymbol{Y})$ is approximated by $e^{R(\boldsymbol{Y})}$, where $R(\boldsymbol{Y})$ express-es the least mean square linear regression of log X on the log Y_j or $E(\log X|\boldsymbol{Y})$.

For an example, when $(X,\boldsymbol{Y})$ has the $k+1$ dimensional log-normal distribution, $\varphi^{(0)}$ is of form

$$\varphi^{(0)}(\boldsymbol{Y})=\prod_{j=1}^{k}Y_j{}^{\beta_j}, \tag{6}$$

where the β_j are the least mean square linear regression coefficients of log X on the log Y_j.

As an another example, when the distribution function of (X,Y) belongs to the two dimensional Parets distribution of type 2 (see K.V. Mardia [4]) $\varphi^{(0)}(Y)$ can be approximated by a power function of Y, since

$$E\left\{\log\left(\frac{X}{a}\right)\Big|Y\right\}=(1-\beta)p^{-1}+q\,\beta p^{-1}\log\left(\frac{Y}{b}\right). \tag{7}$$

In the latter half of this paper, the result of data analysis based on the corporation statistics in Japan is exhibited as a real example.

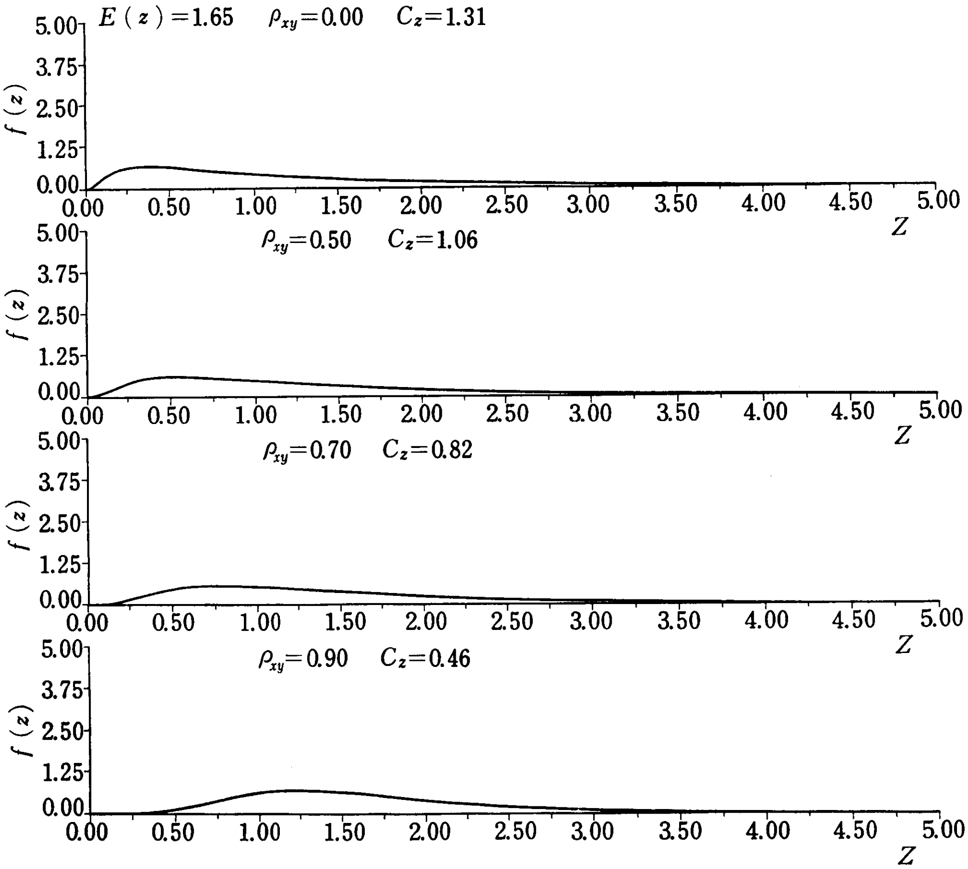
* 本稿は昭和50年度理論・計量経済学会に於て行った報告を理論的に再編し，整理したものである．猶以下の文中に於て回帰関数比例確率抽出法は回帰比例確率抽出法と略称されている。

** 順序統計量の意味ではない。

1. 目的及び意義

従来，標本抽出法に於ては等確率抽出をもって原則とされているが，その場合は母集団分布の特性そのものによって直接その有効性が左右されることになる．此の点に於て Hansen（後掲 [1]）の提唱による不等確率抽出法は，第1図の対数正規母集団の例にみるような仮設的母集団からの等確率抽出と同等な効果をもつと考えられるから，その適用の如何によっては，飛躍的な有効性の向上が期待されるのである．然し乍ら不等確率抽出法に関する在来の研究や適用例は猶多分に模索的であり，必ずしも理論と実際を結ぶ包括的な体系を確立するに至らないかの観がある．

本稿は，次節に示す最有効確率抽出に関する基本定理をもとに，その具体的展開を試みたものであって，その一つの成果は補助変量による主変量の回帰分析を過程に含み，且両変量間の重相関比や重相関係数を有効性に関する主要指標とする回帰比例確率抽出法を得たことである．従って此の抽出法は従来の無作為抽出法に対してより有効性の向上，或いは小標本化を図るものであると共に，有意抽出法の一つの体系化ということも出来よう．更に此等の方法の展開は，従来の層化無作為抽出法の適用における層別法に対して有効な指針を与えることが期待出来る．



第1図 対数正規母集団をもとにした仮定的母集団 $f(z)=\Phi'\left(\dfrac{\log z-\frac{1}{2}\rho'^2_{xy}}{\sqrt{1-\rho'^2_{xy}}}\right)$; $Z=\dfrac{E(Y^{\beta'})}{Y^{\beta'}}X$ の相関係数による変化

2. 最有効確率抽出の定義と基本定理

一般に不等確率抽出法は，補助変量 Y が与えられた場合，その正値一価関数 $\varphi(Y)$ に比例した抽出確率による抽出法と考えることが出来よう．今此の抽出法によって得られた標本 $(X^{(i)};\ Y^{(i)})\ ;i=1,\cdots,n$ を用いて主変量 X の平均 $E(X)$ を推定する目的で，不偏推定量

$$\tilde{X}(\varphi)=\frac{1}{n}E\{\varphi(Y)\}\sum_{i=1}^{n}\frac{X^{(i)}}{\varphi(Y^{(i)})} \tag{1}$$

を用いることにすると，その分数 $\sigma^2_{\tilde{X}(\varphi)}$ は

$$\sigma^2_{\tilde{X}(\varphi)}=\frac{1}{n}\left[E\{\varphi(Y)\}E\left\{\frac{X^2}{\varphi(Y)}\right\}-E^2(X)\right] \tag{2}$$

である．(補論の証明参照)

扨，本稿で扱かう最有効確率抽出とは次の定義によるものとする．

定義 1.「φ の関数族に於て，$\sigma^2_{\tilde{X}(\varphi)}$ を最少ならしめる $\varphi^{(0)}$ が存在するとき $\varphi^{(0)}$ に比例する確率抽出を X の平均を推定する為に $\tilde{X}(\varphi)$ を用いた場合の最有効確率抽出という」．

此の場合容易に定理1が得られるが，その証明に先立って次の補助定理を証明しよう．

補助定理 「$\varphi(y)$ が y の正値関数であれば

$$[E\{e^{\psi(Y)}\}]^2\leqq E\{\varphi(Y)\}E\left[\varphi(Y)\left\{\frac{e^{\psi(Y)}}{\varphi(Y)}\right\}^2\right] \tag{3}$$

」

証明* 一般に $\phi(y)$ が連続な凸関数ならば $p(y)>0$ のとき，Jensen の不等式 (Hardy-Littlewood-Pólya [2] 参照) により

$$\phi\left(\int pf\,dy\Big/\int p\,dy\right)\leqq\int p\phi(f)\,dy\Big/\int p\,dy \tag{4}$$

が成立する．従って確率変数 Y に対して

$$\phi\{E(p\cdot f)/E(p)\}\leqq E\{p\cdot\phi(f)\}/E(p) \tag{5}$$

こゝで $\phi(f)=f^2$，$p(Y)=\varphi(Y)$，$f(Y)=\dfrac{e^{\psi(Y)}}{\varphi(Y)}$ とおくと上記不等式が得られる．

定理 1.「主変量 X が正値で，且

$$\log X=\psi(Y)+\log\zeta;\ \zeta \perp\!\!\!\perp Y, \tag{6}$$

で表わされるならば，$e^{\psi(Y)}$ 比例確率抽出が最有効抽出法を与える．又その場合 $\tilde{X}(\varphi^{(0)})$ の変動係数 $C^{(0)}$ は

$$C^{(0)}=\frac{1}{\sqrt{n}}\frac{\sigma_\zeta}{E(\zeta)}=\frac{1}{\sqrt{n}}C_\zeta \tag{7}$$

である．こゝで C_ζ は ζ の変動係数を表わす．

証明 (6) 式の条件下で，任意のφに対して $\sigma^2_{\tilde{X}(\varphi)}$ は (2) 式により

$$\begin{aligned}\sigma^2_{\tilde{X}(\varphi)}&=\frac{1}{n}\left[E\{\varphi(Y)\}E\left\{\frac{X^2}{\varphi(Y)}\right\}-E^2(X)\right]\\&=\frac{1}{n}\left[\mathrm{E}\{\varphi(Y)\}E\left[\varphi(Y)\left\{\frac{e^{\psi(Y)}}{\varphi(Y)}\right\}^2\right]E(\zeta^2)-E^2(X)\right]\end{aligned} \tag{8}$$

となるから，補助定理により

---

* (3) 式は直接 Schwarz の不等式により証明される事が東京大学の鈴木雪夫教授によって指摘された．

(9) $$\sigma^2_{\bar{X}(\varphi)} \geqq \frac{1}{n}\left[E^2\{e^{\psi(Y)}\}E(\zeta^2)-E^2(X)\right]$$

$$=\frac{1}{n}\left[E\{e^{\psi(Y)}\}E\{e^{\psi(Y)}\cdot\zeta^2\}-E^2(X)\right]$$

$$=\frac{1}{n}\left[E\{e^{\psi(Y)}\zeta E\left\{\frac{X^2}{e^{\psi(Y)}}\right\}-E^2(X)\right]=\sigma^2_{\bar{X}(e^{\psi})}$$

が成立する．これは $e^{\psi(Y)}$ 比例確率抽出が最有効抽出であることを示す．又此の時

(10) $$C^{(0)}=\frac{1}{\sqrt{n}}\frac{\sigma_{\bar{X}(e^{\psi})}}{E(X)}=\frac{1}{\sqrt{n}}\sqrt{\frac{E(\zeta^2)}{E^2(\zeta)}-1}$$

$$=\frac{1}{\sqrt{n}}\frac{\sigma_\zeta}{E(\zeta)}=\frac{1}{\sqrt{n}}C_\zeta$$

となり，定理が証明される．

此の定理はベクトル値の補助変量 $\boldsymbol{Y}=(Y_1,\cdots,Y_k)$ の場合に拡張されると共に対数関数より一般の関数にも拡張される*，此処では，上記の定理の結果を具体的に実現する方向について考察を進めよう．

定義 2.「$\log X$ の $\log Y_1,\cdots,\log Y_k$ に関する回帰関数を $R(\boldsymbol{Y})$ としたとき，$e^{R(\boldsymbol{Y})}$ に比例する抽出確率を用いる抽出法を回帰比例確率抽出法** と定義する．特に回帰関数が回帰係数 $\beta_1,\cdots,\beta_k$ をもつ最少二乗線型回帰の場合は，$\prod_{j=1}^{k} Y_j^{\beta j}$ に比例する確率で抽出することになるが，これを特に区別する場合は，最少二乗対数線形回帰比例確率抽出法ということにする．」

回帰比例確率抽出法は，そのまゝ最有効確率抽出法となることは稀であっても，その近似を与える可能性は少なくないように思われる．

定理 2.「最少二乗対数線形回帰比例確率抽出法は，有効確率抽出法の一つの近似となりうる．又 $\log X$ の $\boldsymbol{Y}$ に関する他の回帰形式

(11) $$R(\boldsymbol{Y})=E(\log X\,|\,\boldsymbol{Y})$$

は，最有効確率抽出法の他の近似となりうる．その際其等の有効性は，それぞれ $\log X$ の $\log Y_1,\cdots,\log Y_k$ に関する重相関係数 $\rho'_{x(1,\cdots,k)}$ 及び $\log X$ の $Y_1,\cdots,Y_k$ に関する重相関比 $\eta'_{x(1,\cdots,k)}$ によって判定される．」

証明 $\log X$ の $\log Y_1,\cdots,Y_k$ に関する最少二乗線形回帰の残差を $\log\zeta_a$ とすれば，一般に ζ_a は $Y_j; j=1,\cdots,k$ と無相関であり，これは $\zeta_a \perp\!\!\!\perp Y_j$ の必要条件であるから，定理1により，これは最有効確率抽出法の一つの近似となりうる．又 $\boldsymbol{Y}$ の任意の正値一価関数 $\varphi(\boldsymbol{Y})$ に対して

(12) $$X=\zeta_\varphi\,\varphi(\boldsymbol{Y})$$

とすれば $\sigma^2_{X(\varphi)}$ は (2) 式により

(13) $$\sigma^2_{X(\varphi)}=\frac{1}{n}\left[E\{\varphi(\boldsymbol{Y})\}E\{\zeta_\varphi^2\,\varphi(\boldsymbol{Y})\}-E^2\{\zeta_\varphi\,\varphi(\boldsymbol{Y})\}\right]$$

と表わされる，従って ζ_φ が1に近く且つ変動が小であれば，つまり $\log\zeta_\varphi$ が0に近く且つその分散が小であるような φ が最有効条件に近いと云える．今 $\varphi(\boldsymbol{Y})$ として $e^{E(\log X|\boldsymbol{Y})}$ を用いた時の ζ_φ を特に ζ_e で表わすと，

(14) $$E(\log\zeta_e)=0,\quad V(\log\zeta_e)=\min_{\varphi} V(\log\zeta_\varphi)$$

* 関数の一般化は不連続性を伴ない，特異なものとなる．

** これは規模比例確率抽出法と関聯させた仮称である．

が成立つから，上記の条件に近く従って一つの近似を与えうる．

扨，(13) 式の関係は，一方で規模最少二乗対数線影回帰や $e^{E(\log X|Y)}$ 比例確率抽出法を適用する場合，その有効性が，ζ_a や ζ_e 従って回帰残差 $\log\zeta_a$ や $\log\zeta_e$ の分散によって判定されることを示している．それは結局，それぞれ $\rho'_{x(1,\cdots,k)}$ や $\eta'_{x(1,\cdots,k)}$ が一つの判定基準となることに帰着し，定理が成立する．

定理 2 は，又回帰比例確率抽出に関する問題に関しては，回帰推定等の場合と対照的に正値分布及び対数尺度が基準となることを示すものであり，その結果次のように歪み型分布族の中に典型的例題を見出すことが出来る．

例題 1.「(X, Y) が $k+1$ 次元対数正規分布をもつと仮定する．此の場合最有効確率抽出法は規模最少二乗対数線型回帰抽出法である．従って此の抽出法による $\hat{X}(\varphi^{(0)})$ の変動係数 $C^{(0)}$ は，

$$C^{(0)} = \frac{1}{\sqrt{n}} \sqrt{e^{\{1-\rho'^2_{x(1k)}\}\sigma_x'^2} - 1} \tag{15}$$

であり，第 1 図に示すような効果を与える．こゝで $\sigma_x'^2$ は $\log X$ の分散を表わす．(T. Taguchi [7] 参照)」

次の例は，単純無作為抽出では推定不能とされる Pareto 分布の場合であるのが第 1 種 Pareto 分布に関しては既に拙稿 [6] に於て詳論した．

例題 2.「(X, Y) が 2 次元第 2 種 Pareto 分布 (K. V. Mardia [4] 参照) をもつ場合は，

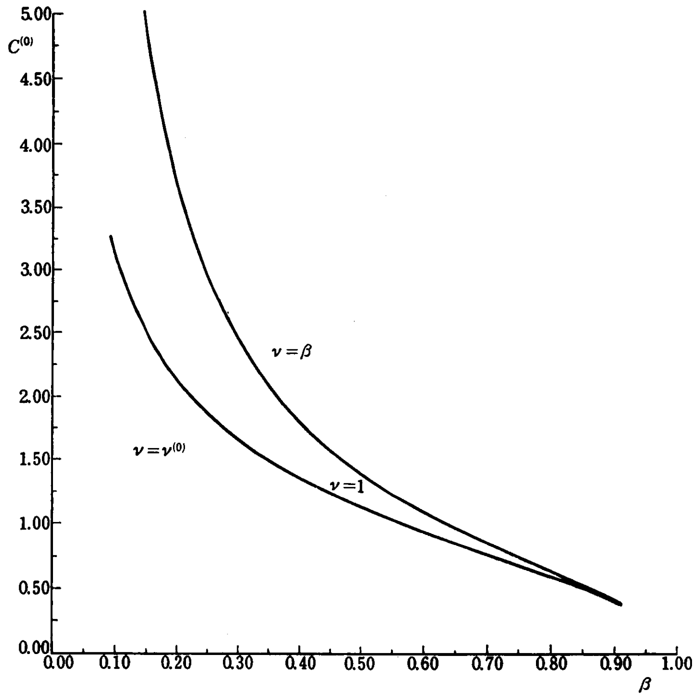
$$E\left\{\log\left(\frac{X}{a}\right)\middle| Y\right\} = (1-\beta)\,p^{-1} + q\,\beta\,p^{-1}\left\{\log\left(\frac{Y}{a}\right)\right\} \tag{16}$$

(同上 [4] 参照) が得られるから $\varphi(Y)=Y^{\nu}$ の関数のうちから最有効な $\nu^{(0)}$ を決定すれば，$Y^{\nu^{(0)}}$ 比例確率抽出は，最有効確率抽出法の近似的方法を与えることになる．その結果は母集団のパラメータ β に応じて第 1 表及第 2 図のような変化を示すものとなる．」

第 1 表　Mardia 第 2 種 Pareto 母集団
(母集団パラメータについて $a_x=a_y=2$ とした場合)

β	$\nu^{(0)}$	$\hat{X}$ の変動係数 (但し $n=1$)			
		$\nu=\nu^{(0)}$	$\nu=\beta'$	$\nu=1$	$\nu=0$
0.1	0.98321	3.09214	7.21682	3.09261	∞
0.2	0.97459	2.11488	3.64575	2.11570	—
0.3	0.97149	1.65204	2.42690	1.65293	—
0.4	0.97227	1.35187	1.79182	1.35265	—
0.5	0.97576	1.12489	1.38587	1.12546	—
0.6	0.98100	0.93506	1.08952	0.93541	—
0.7	0.98710	0.76274	0.84943	0.76291	—
0.8	0.99313	0.59241	0.63461	0.59246	—
0.9	0.99794	0.40154	0.41525	0.40154	∞

扨以上の例題が示す歪み型分布は実際上は経済諸量間に屢々認められる関係である．従って規模回帰確率抽出法の効果を実測する目的で，第 4 節に於て大蔵省法人企業統計調査の結果を用いて解析することにするが，その為には次節の実際的配慮が必要である．



第2図　各抽出法に対する $C^{(0)}$ の近似値の変化
(第2種 Pareto に於て $n=1$ とした場合)

3.　規模回帰抽出法の一般化について

本節に於ては，$\log X$ の $\log Y_1$, $\cdots\cdots$, $\log Y_k$ に関する最少二乗回帰によって得られる R ($\boldsymbol{Y}$) を中心に考察する．又，此の方式の典型は例題1に示した対数正規母集団であるから，それに関して得られる結果をもって方式の基準とする事が出来る．

問題は，実際の場合統計調査の主変量はベクトル $\boldsymbol{X}=(X_1,\ \cdots\cdots\ X_h)$ で与えられる事によって生ずる．すなわち此の場合各 X_t に対して得られる β'_{t1}, $\cdots\cdots$, β'_{tk} から総合的にみて有効な β'_{02}, $\cdots\cdots$, β'_{0k} を見出す必要があるからである*.

今

$$\beta'_{0j}=\bar{\beta}_j=\frac{1}{h}\sum_{t=1}^{h}\beta_{tj};\ j=1,\ \cdots,\ k \tag{17}$$

とする，これは，各 X_h に対して例題1の関係が成立する場合 $C^{(0)}=\sqrt{\prod_{t=1}^{h}(nC_t{}^{(0)}+1)}$ を最少化する条件によって得られる結果と一致し，この条件は更に $\rho'_{xt(1\cdots k)}$ が充分1に近い時

$$C^{(0)}\sim\bar{C}^{(0)}=\frac{1}{h}\sum_{t=1}^{k}C_t{}^{(0)} \tag{18}$$

となる．従って，β'_{0j} を (17) によって決定することは平均的に X_τ を最有効にする方式といえるであらう．

他方，$C_t{}^{(0)}$ 相互間の変動を少なくする為に，(18) 式の代りに

* β_{tj}' の t に関する変動が著しい場合は調査体系を二本だて，或いは三本だてにすることも考えられる．

$$w_t = \frac{\sigma'_{x_t}\sqrt{1-\rho^{\sigma'^2}_{xt(1,\cdots k,)}}}{\sum_{t=1}^{h}\sigma'_{xt}\sqrt{1-\rho^{\sigma'^2}_{xt(1,\cdots,k)}}} \tag{19}$$

で加重した平均を最少化することを考えると，$\rho'_{xt(1,\cdots,k)}$ が1に近い時

$$\beta'_{0j} = \frac{\sum_{t=1}^{h} w_t \beta'_{tj}}{h} \tag{20}$$

とすることが出来る．

更に此れ迄の方法は Y_1, $\cdots$, Y_k を既定の変量としたのであるが，実際問題としてその撰択，組合せは其等が $\boldsymbol{X}$ の構造を実質的に規定するものである事と同時に，調査の事前資料として利用しうる事等の条件を考慮しても，必ずしも一義的ではない．こゝに一般の多次元解析の場合と共通した課題が見出せるが，次節の分析は，従来用いられている諸補助変量に対して比較原理として

$$\bar{\rho}'_{x(1\cdots k)} = \frac{1}{h}\sum_{t=1}^{h}\rho'_{xt(1\cdots k)} \tag{21}$$

を用い得るかについての検討も加えた．

4. データー解析の実際

後掲第3, 4, 5の諸表は，大蔵省法人企業統計調査*（昭和48年度4〜6月期分）の調査資料を用い，第2表**に示すアウトプット形式に基づいて算定した結果を示すものである．此れによって規模最少二乗線型回帰確率抽出法に関して次に列記する諸点を観察する事で出来る***．

先づ此の抽出法の調査客体の産業及び規模に対する有効性を比較すると

(a) 第3表及び第4表の結果によると，各産業を通じて企業の規模が大きい階層程規模比例確率抽出法は単純無作為抽出法により有効であり，回帰比例確率抽出法は両者より更に有効である．然し乍ら規模が比較的大である階層に於ては，$k=1$ の時，β_1' は全体として1に近いから．規模比例確率抽出法を回帰比例確率抽出法の一つの近似法として用いる事が出来る．一般の k の値に対しても近似的に $\sum_{j=1}^{k}\beta_j' = 1$ が認められるが，此の性質の利用については別途慮する必要がある．

次に調査項目による有効性の差を検討すると

(b) $C\ (\prod_{j=1}^{k} Y_j^{\beta'_j})$, $C\ (\prod_{j=1}^{k} Y_j^{\beta'_j})$ 及び $\rho'_{x(1\cdots k)}$, の値は，3, 4, 5の各表とも項目によって変動することを示している．例えば固定資産については可成り有効であるが，借入金については有効性が低い等が共通した結果である．然し乍ら他の抽出諸方法の場合と比較すると，その差は少なく，此の方法の有効性を示している．

次に補助変量を変えた場合，又その数 k を増した時の有効性の変化を検討しよう．

(c) 補助変量として最も利用し易いのは一般に資本金であるが，各産業について比較した結果では資本金よりは従業員数の方が有効であり，従業員数よりは売上高或いは資産額の方が有効である等がいえる．

---

* 行政承統267号，蔵証第1446号．

** 此の形式は清水順子嬢によるアウトプット形式を縮約したものである．

*** 単純な規模比例確率抽出法については既に拙稿 [6] で検討した．

第2表 $k=1$ の場合のアウトプット形式

産 業 名________________

産業コード________________ 資本金階級________________ サンプル数________________

項目	平均	分散	変動係数	規模比例抽出分散	回帰比例抽出分散	平均回帰比例抽出分散	前項の変動係数	対数線形回帰係数		対数相関係数
(X)	$E(X)$	$V(X)$	$C(X)^{*1}$	$V_0(X)^{*2}$	$V_1(X)^{*3}$	$V_2(X)^{*4}$	$C_2(X)^{*5}$	B'_0	B'_1	$P'(XY)$
1. 資本金										
2. 売掛金										
3. 棚卸資金										
4. 固定資金										
5. 資金合計										
6. 買掛金										
7. 短期借入金										
8. 長期借入金										
9. 営業損益										
10. 売上高										
11. 従業員数										
									$\bar{B}'$	$\bar{R}'(XY)$

*1 $C(X)=\dfrac{\sqrt{V(X)}}{E(X)}$ *2 $V_0(X)=E(Y)\,E\left(\dfrac{X^2}{Y}\right)-E(X^2)$ *3 $V_1(X)=E(Y^{B_1})\,E\left(\dfrac{X^2}{Y^{B_1}}\right)-E^2(X)$

*4 $V_2(X)=E(Y^{\bar{B}_1})\,E\left(\dfrac{X^2}{Y^{\bar{B}_1}}\right)-E^2(X)$ *5 $C_2(X)=\dfrac{\sqrt{V_2(X)}}{E(X)}$

但し $\bar{B}_1=\dfrac{1}{11}\sum_{j=1}^{11}B_{1j}$, $\bar{R}=\dfrac{1}{11}\sum_{j=1}^{11}R_j$. つまり B_1, R の項目間平均を表わす．又，V(X) の計算は8桁までの計算によって行った．

第3表　補助変量として資本金を用いた場合の鉄鋼業の特性量

SANGYO TEKKOH
SANGYO CODE 31　　SHIHUNKIN KAIKYU 1億以上10億 SAMPLE数151

KOUMOKU (X)	E(X)	V(X)	C(X)	V0(X)	V1(X)	V2(X)	REG - CO C(X)	B0	B1	COR-CO R(XY)
1.SHIHONKIN	283.	39084.	0.69971	0.	0.	926.	0.10769	0.00000	1.00000	1.00000
2.URIKAKEKIN	1154.	2290437.	1.31168	1696727.	1575949.	1598303.	1.09572	2.50729	0.71679	0.33233
3.TANAOROSHISHISAN	596.	763743.	1.46697	387236.	388760.	396809.	1.05740	0.51782	0.93770	0.41974
4.KOTEISHISAN	1921.	6023534.	1.27777	5569903.	5216264.	5170360.	1.18382	2.45274	0.85617	0.61064
5.SHISANGOHKEI	4724.	28004208.	1.12027	21229088.	19629968.	19853648.	0.94326	3.82223	0.78020	0.57639
6.KAIKAKEKIN	1762.	3690952.	1.09038	3175426.	2760986.	2896830.	0.96598	3.38215	0.64198	0.31151
7.TANKISHAKUNYUKIN	430.	976738.	2.29873	470156.	475643.	483351.	1.61707	-1.78482	1.14096	0.31099
8.CHOKISHAKUNYUKIN	1065.	5271245.	2.15665	5740616.	5665811.	5404340.	2.18371	0.71680	0.96716	0.33171
9.EIGYOSONEKI	152.	63845.	1.65858	39849.	39873.	39854.	1.31041	-0.31727	0.82707	0.35735
10.URIAGEDAKA	1590.	2777132.	1.04791	2097195.	1899517.	1928502.	0.87324	2.67492	0.77944	0.46329
11.JYUGYOINSU	357.	161320.	1.12496	201342.	141207.	170887.	1.15783	2.70704	0.50015	0.31739
									0.83160	0.45739

SANGYO TEKKOH
SANGYO CODE 31　　SHIHUNKIN KAIKYU 10億以上　SAMPLE数 58

KOUMOKU (X)	E(X)	V(X)	C(X)	V0(X)	V1(X)	V2(X)	REG - CO C(X)	B0	B1	COR-CO R(XY)
1.SHIHONKIN	13878.	1315398900.	2.61344	0.	0.	310112.	0.04013	0.00000	1.00000	1.00000
2.URIKAKEKIN	18737.	1421666500.	2.01229	132368890.	102237690.	152182520.	0.65838	0.91562	0.94754	0.78368
3.TANAOROSHISHISAN	19573.	2299660200.	2.45006	49019904.	79685888.	56180736.	0.38295	-0.54385	1.09127	0.87111
4.KOTEISHISAN	80219.	50256130000.	2.79458	623542270.	688332800.	623255550.	0.31121	0.76645	1.08030	0.93083
5.SHISANGOHKEI	139402.	128663280000.	2.57311	1489989600.	1563463600.	1686110200.	0.29456	2.13722	1.01180	0.92247
6.KAIKAKEKIN	23251.	2151972800.	1.99515	172932090.	87693056.	203118840.	0.61296	1.77980	0.88184	0.80518
7.TANKISHAKUNYUKIN	16511.	1485168800.	2.33410	65930496.	257545210.	74250240.	0.52189	-2.90782	1.29178	0.61453
8.CHOKISHAKUNYUKIN	41782.	14310985000.	2.86313	293520380.	464369660.	302452730.	0.41623	-0.84531	1.16404	0.82159
9.EIGYOSONEKI	3392.	82986784.	2.68577	1213988.	1348460.	1353621.	0.34301	-1.90995	1.02686	0.72460
10.URIAGEDAKA	27187.	4309811200.	2.41476	118393340.	70260480.	142485240.	0.43907	1.58917	0.91018	0.85362
11.JYUGYOINSU	6022.	178449360.	2.21814	7673088.	4377392.	9085504.	0.50050	0.20530	0.89929	0.89093
									1.02772	0.83805

第4表　補助変量として従業員数を用いた場合の機械製造業の特性量

SANGYO KIKAI
SANGYO CODE 34　　SHIHUNKIN KAIKYU 1億以上10億 SAMPLE数318

KOUMOKU (X)	E(X)	V(X)	C(X)	V0(X)	V1(X)	V2(X)	REG - CO C(X)	B0	B1	COR-CU R(XY)
1.SHIHONKIN	262.	36859.	0.73245	103282.	29885.	61364.	0.94506	3.74094	0.27613	0.37714
2.URIKAKEKIN	1367.	2936871.	1.25376	3816121.	2997381.	3153067.	1.29909	2.27501	0.75226	0.54820
3.TANAOROSHISHISAN	859.	720577.	0.98875	258309.	253623.	257397.	0.59095	1.02256	0.91455	0.76306
4.KOTEISHISAN	968.	721612.	0.87722	320131.	282643.	273303.	0.53986	1.42807	0.87607	0.76300
5.SHISANGOHKEI	4246.	15172936.	0.91742	8929808.	7448336.	7638640.	0.65094	3.68641	0.74717	0.77521
6.KAIKAKEKIN	1323.	2320587.	1.15104	2228403.	1780474.	1872064.	1.03384	2.46469	0.73533	0.61802
7.TANKISHAKUNYUKIN	750.	1287732.	1.51259	1025459.	1092574.	913888.	1.27425	-0.82346	1.07129	0.44533
8.CHOKISHAKUNYUKIN	633.	807876.	1.42073	407199.	418048.	401765.	1.00190	-0.96590	1.07453	0.44150
9.EIGYOSONEKI	66.	17383.	2.00282	14903.	13069.	13544.	1.76789	-0.52407	0.62803	0.27903
10.URIAGEDAKA	1048.	990724.	0.95003	872465.	760467.	727839.	0.81429	1.54253	0.86619	0.72860
11.JYUGYOINSU	441.	108675.	0.74791	0.	0.	2910.	0.12238	0.00000	1.00000	1.00000
									0.81287	0.61264

SANGYO KIKAI
SANGYO CODE 34　　SHIHUNKIN KAIKYU 10億以上　SAMPLE数 78

KOUMOKU (X)	E(X)	V(X)	C(X)	V0(X)	V1(X)	V2(X)	REG - CO C(X)	B0	B1	COR-CU R(XY)
1.SHIHONKIN	3470.	15021167.	1.11680	3165976.	3094913.	2935191.	0.49367	2.53889	0.69117	0.77858
2.URIKAKEKIN	13621.	553719550.	1.72757	103236840.	104274920.	105351660.	0.75355	1.53746	0.96311	0.76599
3.TANAOROSHISHISAN	7524.	90924160.	1.26731	32534496.	30992496.	31538624.	0.74639	2.23622	0.81802	0.77077
4.KOTEISHISAN	11720.	427522810.	1.76419	45672464.	42040560.	51593248.	0.61286	0.75767	1.04457	0.88807
5.SHISANGOHKEI	41710.	3925649900.	1.50216	485323000.	510610430.	513205760.	0.54313	2.92406	0.94398	0.86944
6.KAIKAKEKIN	8422.	94426976.	1.15385	26678800.	24558704.	25341248.	0.59774	1.87908	0.87425	0.72640
7.TANKISHAKUNYUKIN	9183.	370962680.	2.09730	55326208.	65436000.	59464976.	0.83970	1.13218	0.87054	0.29142
8.CHOKISHAKUNYUKIN	5091.	82700496.	1.78646	24916960.	25085856.	25274448.	0.98760	-1.65640	1.13213	0.37209
9.EIGYOSONEKI	694.	1052759.	1.47768	387475.	379912.	379552.	0.88726	-1.94936	0.94348	0.34699
10.URIAGEDAKA	8141.	105006880.	1.29076	13590848.	13691648.	13905824.	0.45807	0.38980	1.05252	0.66387
11.JYUGYOINSU	2988.	9348752.	1.02598	0.	0.	24395.	0.05241	0.00000	1.00000	1.00000
									0.93943	0.67942

他方補助変量の数をますと有効性が増大する事は当然であるが，その検討には種々の計算を要するのでこゝではその一端として単に第5表を例示するに止める．

以上は調査によって同時的に得られる結果を検討したものであるが，実際は事前情報としての Y と事後の情報としての X との間には当然時間のズレが生じている．この点に関して，

(d)　主要量として当期の結果を用い，補助変量として前期分を用いて同様な計算を行った

第5表　補助変量として資本金のみ及び資本金と従業員数をいた場合の電気機械製造業の特性量

SANGYO　DENKI
SANGYO CODE 35　　SHIHONKIN KAIKYU 10億以上 SAMPLE数 91

KOUMOKU (X)	F(X)	V(X)	C(X)	V0(X)	V1(X)	V2(X)	REG - CO $C_2(X)$	B0'	B1'	CUR-CU R'(XY)
1.SHIHONKIN	7269.	330714880.	2.50168	0.	0.	378064.	0.08458	0.00000	1.00000	1.00000
2.URIKAKEKIN	19249.	2066684900.	2.36174	56140032.	52133120.	50432512.	0.36894	1.21512	0.96875	0.91573
3.TANAOROSHISHISAN	14541.	1265996200.	2.44692	24418096.	23883328.	23042560.	0.33012	0.72861	0.98945	0.92358
4.KOTEISHISAN	21088.	2456137900.	2.35014	108460280.	111035130.	102738940.	0.48066	0.89067	1.01372	0.92269
5.SHISANGOHKEI	72257.	29409419000.	2.37336	542253050.	527982590.	504152060.	0.31074	2.37263	0.98895	0.95224
6.KAIKAKEKIN	13883.	734957050.	1.95278	92506016.	63854624.	77811360.	0.63539	1.91721	0.85849	0.84668
7.TANKISHAKUNYUKIN	11116.	973282300.	2.80649	37787312.	59365520.	37542960.	0.55120	1.71000	0.73880	0.30549
8.CHOKISHAKUNYUKIN	5706.	226376990.	2.63679	16699344.	17165728.	15837552.	0.69743	-1.08921	1.01872	0.46925
9.EIGYOSONEKI	1707.	20333200.	2.64224	1851649.	2014525.	1691556.	0.76210	-2.38959	1.03932	0.51880
10.URIAGEDAKA	19533.	1910721200.	2.23779	135089660.	116187640.	120884730.	0.56287	1.72989	0.91322	0.87864
11.JYUGYOINSU	7234.	214922410.	2.02656	9200688.	4948624.	6744736.	0.35900	1.29886	0.86166	0.91417
									0.94464	0.78612

SANGYO　DENKI
SANGYO CODE 35　　SHIHONKIN KAIKYU [illegible] SAMPLE数 91

KOUMOKU (X)	E(X)	V(X)	C(X)	V0(X)	V1(X)	V2(X)	REG - CO $C_2(X)$	B0'	B1'	B2'	COR-CO R'(XY)
1.SHIHONKIN	7269.	330714880.	2.50168	1307824.	0.	1028928.	0.13954	0.00000	1.00000	0.00000	1.00000
2.URIKAKEKIN	19249.	2066684900.	2.36174	35932416.	36944128.	36499712.	0.31386	0.48905	0.48707	0.55900	0.93772
3.TANADROSHISHISAN	14541.	1265996200.	2.44692	15109024.	14851680.	15041744.	0.26672	0.00461	0.50915	0.55741	0.94472
4.KOTEISHISAN	21088.	2456137900.	2.35014	64643840.	61474304.	66314240.	0.38616	-0.04137	0.39541	0.71758	0.95584
5.SHISANGOHKEI	72257.	29409419000.	2.37336	249290750.	227811520.	252661760.	0.21998	1.56209	0.45124	0.62404	0.97953
6.KAIKAKEKIN	13883.	734957050.	1.95278	56115088.	45094768.	59217152.	0.55430	1.00773	0.25514	0.70022	0.88684
7.TANKISHAKUNYUKIN	11116.	973282300.	2.80649	36593280.	75361488.	36264752.	0.54173	2.70286	1.39747	-0.76441	0.32850
8.CHOKISHAKUNYUKIN	5706.	226376990.	2.63679	15665120.	18750112.	15888240.	0.69855	-0.57812	1.35779	-0.39350	0.47433
9.EIGYOSONEKI	1707.	20333200.	2.64224	1414400.	1917033.	1447538.	0.70499	-4.10744	-0.10029	1.32258	0.57686
10.URIAGEDAKA	19533.	1910721200.	2.23779	82769152.	70597376.	86173440.	0.47523	0.63575	0.18737	0.84238	0.93160
11.JYUGYOINSU	7234.	214922410.	2.02656	2134944.	0.	2667136.	0.22576	0.00000	0.00000	1.00000	1.00000
									0.54003	0.46957	

結果，特に大きな変化は認められなかった．

以上 (a)，(b)，(c)，(d) を通して此の方法は極めてロバストであり，適用に際して柔軟性をもつものであるといえよう．

5. 規模回規確率抽出法のメリットとデメリット

これまで規模回規確率抽出法について述べた事は有効性の量的増大に関する事柄であった．然し乍ら此の方式を客観的に評価する場合は，次のような幾つかの質的観点に照して行わるべきである．

先づ推定方式の変更による有効性の向上との比較であるが，

(a) 諸推定方式のうちで極めて有効な方式として回帰推定方式を挙げる事が出来るが，これは一般に正規型母集団とその周辺に関するものといえる．例えば例1の所得分布に対してその単純な最少二乗線型回帰による残差は，対数線形回帰の場合にべて甚だ大であり，且つ前節の資料によると項目による変動も激しいことが認められる．此の事は例2の所得分布に対してより決定的である．つまりこの分布に於ては $\alpha \leqq 2$ に於て分散が無限大となるから等確率抽出法に基づく，従来の如何なる推定方式によっても推定不能となるからである．

扨，前節までの分析は専ら母集団の平均の推定に関する有効性についてであった．然し乍らその結果は次のような母集団特性量の推定に関する場合にまで拡張されうる．すなわち

(b) 集中度の表現に用いられる Gini の平均差 Δ_x を推定する場合を考えよう．等確率抽出法の場合 Δ_x の不偏推定量は，例えば

$$\tilde{g} = \frac{1}{n\ (n-1)} \sum_{i,j=1,\cdots,n} \sum |X_i - X_j| \tag{22}$$

(Lomnicki [3] 参照) で与えられるが，母分散が大である時，多くの標本を必要とし計算量が増大する．つまり此の場合 $\tilde{g}$ の分散は

$$\text{(23)}\qquad \mathrm{var}\ (\tilde{g}) = \frac{1}{n\ (n-1)} \left\{ 4\ \sigma_x{}^2 + 4\ (n-2)\ \mathscr{F} - 2\ (2\,n-3)\ \varDelta_x{}^2 \right\}\ ;$$

$$\mathscr{F} = \underset{(Xi,Xj,Xk)}{E}\ (|Xi - Xj|\,|Xi - Xk|)$$

で与えられるからである.

然し乍ら回帰比例抽出法による平均差の推定に於ては

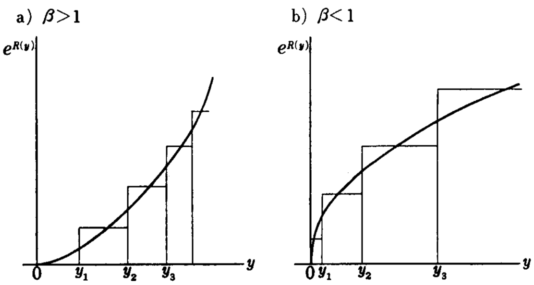
$$\text{(24)}\qquad \tilde{g}_R = \frac{1}{n(n-1)} \sum_{i,j=1,\cdots,k}\sum \frac{\mu_y{}^2}{Y_i Y_j}\ |X_i - X_j|$$

を不偏推定量とすることが出来，その分散は，(23) 式に於て $\sigma_x{}^2$ を $\sigma_x{}'^2$ で又 $\mathscr{F}$ の代りに $\mathscr{F}'$，すなわち

$$\text{(25)}\qquad \mathscr{F}' = \underset{(Xi,YiXj,Xk)}{E} \left\{ \frac{\mu_y}{Y_i}\ |X_i - X_j|\,|X_i - X_k| \right\}$$

を用いる事が出来る。その結果平均差の推定に於ても等確率抽出法よりも有効である事が予想される。

以上 (a)，(b) と逆に回帰比例確率抽出法のデメリットと思われる点は，その抽出作業が幾分複雑化されることにあると思われるが，その簡約化の試みは次のように層化無為作抽出法を合理化する結果になる. つまり



第 3 図　回帰比例確率抽出法の簡約法

(c)　回帰比例確率抽出法は，それをそのまゝ適用する場合には，母集団の各構成要素毎に weight として $e^{R(Y)}$ を計算せねばならず，大母集団に対して必ずしも実際的とはいえない. 然し乍ら第 3 図 a)，b) のように $e^{R(Y)}$ を階段関数で近似すれば，y_1，y_2，…等を分点とする通常の層化無作為抽出法に還元される. 此の事は，従来の経験的な層の形成や分点の決定を合理化するのに役立であらう. これ又一方で，従来の層化無作抽出法の体系を変更することなく分点の位置をかえるのみで有効性を増大させる可能性をもつものであり，体系の継続性を持たせる上で有益であらう.

より高度の簡約法は，折線による $e^{R(Y)}$ の近似を考慮することである. これは層化 $\alpha y+\beta$ 比例確率抽出法を与えることになり，y が充分大である階層に対しては，単純な規模比例確率抽出法が適当である事を理解させるのである.

尚，以上に於て対数線形回帰の形式そのものが，ある階層間に於て明らかに区別さるべきことが示される場合は，此の階層区分はより本質的な意味をもつものと理解することが出来よう.

本稿の結びにあたり，理論・計量経済学会に於て座長をされた和歌山大学の杉浦一平教授，

コメンタリをされた東京大学の鈴木雪夫教授に感謝するものである.
又，資料の利用について尽力された大蔵省証券局資本市場課の方々及び病魔をおして製表，製図に協力された清水順子嬢に厚くお礼する次第である.
更に，所長を初めとする統計数理研究所の諸氏の助言と批判に感謝するものである.

補　論

(1) 式の不偏性と (2) 式の成立の証明を示すと，標本 $(X^{(i)}, Y^{(i)})$; $i=1, \cdots n$ は相互に独立であるから，

$$\underset{\substack{(X^{(i)},Y^{(i)})\\ i=1\cdots,n}}{E}\{\tilde{X}(\varphi)\} = \underset{\substack{(X^{(i)},Y^{(i)})\\ i=1\cdots,n}}{E}\left[\frac{1}{n}E\{\varphi(Y)\}\sum_{i=1}^{n}\frac{X^{(i)}}{\varphi(Y^{(i)})}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\underset{(X^{(i)},Y^{(i)})}{E}\left[\frac{E\{\varphi(Y)\}}{\varphi(Y^{(i)})}X^{(i)}\right]$$

$$= \underset{(X,Y)}{E}\left[\frac{E\{\varphi(Y)\}}{\varphi(Y)}X\right] = E(X).$$

となり $\tilde{X}(\varphi)$ は不偏である．又これにより

$$\underset{\substack{(X^{(i)},Y^{(i)})\\ i=1,\cdots n}}{E}\{\tilde{X}(\varphi)-E(X)\}^2 = \frac{1}{n^2}\underset{\substack{(X^{(i)},Y^{(i)})\\ i=1,\cdots n}}{E}\left[E^2\{\varphi(Y)\}\sum_{i=1}^{n}\left\{\frac{X^{(i)}}{\varphi(Y^{(i)})}\right\}^2\right.$$

$$\left.-2E(X)\left\{E\{\varphi(Y)\}\cdot\sum_{i=1}^{n}\frac{X^{(i)}}{\varphi(Y^{(i)})}\right\}+nE^2(X)\right]$$

$$= \frac{E\{\varphi(Y)\}}{n^2}\sum_{i=1}^{n}\underset{(Y^{(i)},X^{(i)})}{E}\left[\frac{E\{\varphi(Y)\}}{\varphi(Y^{(i)})}\frac{X^{(i)2}}{\varphi(Y^{(i)})}\right]-\frac{1}{n}E^2(X)$$

$$= \frac{E\{\varphi(Y)\}}{n}\underset{(X,Y)}{E}\left[\frac{E\{\varphi(Y)\}}{\varphi(Y)}\frac{X^2}{\varphi(Y)}\right]-\frac{1}{n}E^2(X)$$

$$= \frac{1}{n}\left[E\{\varphi(Y)\}E\left\{\frac{X^2}{\varphi(Y)}\right\}-E^2(X)\right]$$

となり，(2) 式が成立する．

参　考　文　献

[1] Hansen, M. H and Hurwitz, W.N. 'A new sample of the population: sampling principles introduced into the Bureans' monthly reports on the labor force; U.S. Bureau of Census (1944).
[2] Hardy, G.H., Littlewood, J.E & Polya, G., *Inequalities*, Cambridge at the University Press (1952)
[3] Lomnicki, Z.A. The stamdard error of Gini's mean difference', Ann. Math. Statist. Vol. **23**, (1952)
[4] Mardia, K.V., 'Multivariate Pareto distribution', Ann. Math. Statist. Vol. 33 (1962)
[5] Nair, U.S., 'The standard error of Gini's mean difference', Biometrika, Vol. 28 (1936)
[6] 田口時夫，'日本における法人企業の模本抽出法について—ある不等確率抽出法における最適抽出確率に関する一考察—，統計数理研究所彙報，第22巻，第2号，(1975).
[7] Taguchi, T, 'On an optimal probability of selection, (to he appeared).