順位を基礎とする統計学の数学的方法のある種の体系化について

田　口　時　夫

(1964 年 2 月受付)

On a Plan of Mathematical Analysis of Empirical and Statistical Data, Mainly of Economical Ones

Tokio TAGUTI

This paper contains some mathematical methods for the statistical research and the analysis of statistical data. The methods are not probabilitic but analytic; the sampling design treated of in this paper, for example is based on the concrete nature of Lorentz, curve. Generally speaking I guess that the concept of "random", in statistics is to be reexamined by more actual dynamic point of view on the future stage of the statististical science, and I wish the unification of theories and practices.

Institute of Statistical Mathematics

はしがき

本篇は統計学の順位を基礎とする別個の数学的方法の理論と実践を貫ぬく全体系の一環をなすものであるが，特に基礎理論的部分をひとまず後日に譲り，実際的技術手段をとり上げたのはそれが直接的であり，当面の活用のためである.

もちろん両者は首尾一貫して緊密な関連の下にあるが,「非確率論的統計数学」(実際は後節のように確率の経験的規定を含むから超確率的といった方が適切かもしれない.）と題した.昭和38年度日本統計学会における発表がその初歩的概観を与えるであろう.

巻末に用意した資料は，ここでは具体的一例としてであり，本文の適用範囲はより広範に亙るか否か一先づは読者が賢察にまつことにする.

§1. 序　論

統計的諸集団の内部及び相互間に予想される，統計的因果関係の支配とその構造，機能に予想されるダイナミックな関係は，確率的観念がその形式によって全面的かつ画一的に統計観を与えるとする従来の安易な主張と矛盾し，経験データによる科学的再吟味が必要とされて来た.

こうしたいきさつは経験データの処理に関する諸方法，理論を自らの一手に掌握していると信じている数理統計学者には一見奇異な事実であるかも知れないが，皮相な大量観念をもとに遂に危険率の観念を脱却しえないそのトートロジイを反省すれば，こうした事情はまたいわば内面的論理的必然ともいえるのである．

はしがきに述べた如くこの課題の理論的一般的な検討は避けて，以下の各章においては専らランダム・サンプリング・メソッド，推定，検定の数理統計学的方法と技術にかわるべき方法を掲げ，最後に総括的な批判，検討を加えることにする．

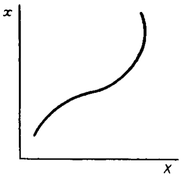
この新たな技術体系を一貫するものは統計学の主要課題を集団内部の分配構造とその機能の分析においたとき，その科学的分析方法に対応しているという一点においてある．

§2.　標本調査法に関する組織的な数学的対案

論理の一般的な運びから最も単純なケースから出発し，序々に複雑形態に進むことにする．この過程において方法自体は次第に一般性を獲得するので，最初の各段階の仮設は終局的には取除かれることを指向する．

全体的な仮設としては*1集団のサイズが限りなく大となる究極の関係おいて経験的カーブが一様に収斂するグレードに応ずる分配比率*2の理論的函数関係，つまり下位からの積算比率としての累計函数 $Y=\psi(x)$ あるいはその導函数にあたる*3順位に応ずる分配比率の存在とその函数関係 $y=\varphi(x)$ の存在とその正値性及び Taylor 展開可能性である．

ここに掲げる方法の具体的作業に関しては従来通りの方式を多々含むが誤差評価に関しては基本的に異なっている．



（一）　等間隔組織選択法

（1）　第一段階としてまず調査によって獲得すべき数量xの大小による序列は既知であるとする．これは例えばxと同一あるいは正反の序列を与える既存データの所在を前提とする．また例えばxと単調な関係を結ぶデータであってよい*4．従って更にxな有界変動な函数をなし具体的に二つの増加函数で示されるならばよい．

（2）　(1) による順位をもとに思考上それを充分大なる総数Nからの標本数nに応じて同数個 $2Nh$ を擁する階級に区分する．

したがって grade, rank に応ずる分配関係の縮図では，$(0,\xi_1)(\xi_1,\xi_2)\cdots(\xi_{i-1},\xi_i)\cdots(\xi_{n-1},1)$ が決定される．

ここでこの各階級のメディアンが標本値を与えるように

$$X_1=h,\ X_2=3h,\cdots\cdots,X_i=(2i-1)h,\ \cdots\cdots,X_n=1-h$$

註 *1, *3　グレード及びランクに関する定義は全体との関係で％で示すがこれは Kandall の定義と一致する．

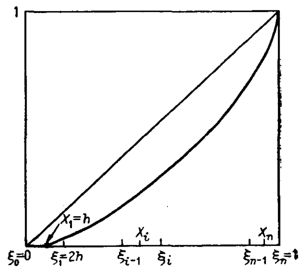
Generally, we may regard the grade of an individual as the proportion of individuals which lie below him. Ranks and grades are conected by a simple relation. In fact, if an individual is of rank k, there are $k-1$ individuals belows him (assuming that the ranking proceeds from the lowest variate value.). If we admit conventionally, that one-half of the admit conventionally, that one-half of the individual is to be regarded as lying to the left of the line of division which he makes, and one-half to the right, his grade g_k is given by　$g_k=(k-1)+\dfrac{1}{2}=k-\dfrac{1}{2}$.

M. G. Kendall. The Advanced Theory of Statistics, Vol. 1, p. 42-45.

*2　グレードの分配函数はグラフ的には，ローレンツ・カーブを画く．X, $\psi(X)$, $\varphi(X)$ の数学的性質については屢々拙稿で触れたのでここでは省略するが本文中では屢々適用されている．

註 *4　この仮定は従来“相関”或は“順位相関”として層別法や回帰推定法に適用されて来た．

を等間隔組織的に抽出する．後に補正するように X_1 および X_nの位置については多少の変更があっても大した影響はない．



（3） 各階級 (ξ_{i-1},ξ_i) に関する総評価額は前頁の $\psi(X)$ の定義と N が充分大であるという仮定のもとに

$$N_m[\psi(\xi_i)-\psi(\xi_{i-1})]=N_m[\psi(X_i+h)-\psi(X_i-h)]$$

$\psi(X)$ の展開可能性から

$$=N_m\left[2h\psi'(X_i)+\frac{h^3}{3}\psi'''(X^i)+0(h^5)\right]$$

したがって総量の推計を標本 X_i の評価額 x_i を用いてインターバル $(2hN)\times$ 評価額 (x_i) によって行えば，これは第一項に等しい．したがって第二項は誤差を表示し，標本数 $n=\dfrac{1}{h}$ を増大させる時加速度的に減少する．すなわちインターバル（これはこの場合同時に抽出率 P に等しい）$I=P=2Nh=$ 一定として $0(h^2)$ であるが，これを隣接階級の標本値 x_{i-1},x_{i+1} を用いて概算すると

$$h^2|m\varphi''(X)|=h^2\left|\frac{d^2x}{dX_2}\right|=\frac{1}{4}(x_{i+1}+x_{i-1}-2x_i).$$

したがって誤差評価額は

$$|E_i|=\frac{I}{24}(x_{i+1}+x_{i-1}-2x_i)\quad i=2,\cdots\cdots,n-1$$

となる．

もし厳密を期して両端の階級部分を含むならば，最低 x_L 最高 x_H を標本に追加することにより各

$$|E_L|=\frac{Nh}{9}|x_2+2x_L-3x_1|,\qquad |E_H|=\frac{Nh}{9}|2x_H+x_{n-1}-3x_n|$$

を得る．もちろん x_L,x_H に代り x_1,x_n を用いて概算できる．

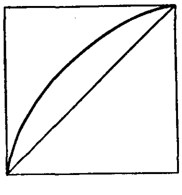
（4） 以上により集団の総量は

$$T=\sum_{i=1}^{n}T_i=2m\sum\left\{\varphi(X_i)+\frac{h^3mN}{3}\varphi''(X_i)+0(h^4)\right\}.$$

すなわち従来通り「抽出間隔 × 標本総額」で推定すれば総誤差は

$$|E|\leqq\frac{p}{24}\sum_{i=2}^{n-1}(x_{i+1}+x_{i-1}-2x_i)+|E_H|+|E_L|$$

$$=\frac{p}{24}(x_n-x_{n-1}+x_1-x_2)+|E_H|+|E_L|$$



である．すなわち高額部分と低額層とによって決定されるが格差が大なる時低額層は無視しうる．

（5） ここでは実際上の幾多の問題点を掲げる．即ち

（i） 実際の抽出に当っては高額層から抽出するが結果は上と同一である．高額層からの抽出は特に（二）の等倍率で抽出するとき不可避的である．

（ii） 順位を細部にわたり確認することは困難ないし徒労である時，調査後の数量により適宜補正し，あるいはクラス単位に一括抽出するかまたは二段階に分ち，第一次はあるクラスを単位に上記の方式を考察すればよい．高次のクラスは標本クラスのみを対象とする．

(iii) 一般に細部にわたり順位を変えぬ既存資料を求める必要はなく，全体的傾向として適切であればよい．細部の階級については層化ランダム抽出法のごとくその間に理論的に与えら

れた数を恣意的に抽出してもなお後節（三）のごとく誤差制限の手段を残している．

こうした意味での台帳は，例えば企業調査の場合，会社企業名鑑（総理府統計局）のごとく整備された資料を容易に見出しうるであろう．

（二）　非等間隔組織抽出法

（1）　等間隔抽出法は格差が大なる時，（一）の誤差評価諸式に認められるように限ずしも適切なものとはいえない．

したがってこれに換えるのに高額順に等倍率的に $n+1$ 個標本を抽出するとする．すなわち，標本は各

$$X_0=\frac{1}{N}\quad X_1=X_0r\ X_2=X_0r^2,\cdots\cdots,\ X_i=X_0r^{i-1},\cdots\cdots,\ X_n=X_0r^n=1$$

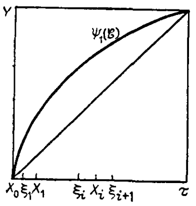
なる順位であり，これは高額より順位を対数比率（$r=\sqrt[n]{N}$）で表現した時の等間隔抽出に当る．すなわち前節の X に対し

$$\mathscr{X}=1+\frac{\log X}{\log N}$$

とすれば $\mathscr{X}_0=0\ \mathscr{X}_1=\log r=h',\cdots,\mathscr{X}_i=ih'\cdots,=1.$

（2）　したがって $\mathscr{X}_i$ を挟む ξ_i', ξ'_{i+1}, の総量は N が充分大であれば，（一）と同様

$$T_i=N_m[\psi_1(\xi_{n+1})-\psi_1(\xi_i)]=m_m\left[h'\psi_1'(\mathscr{X}_i)+\frac{h'^3}{3!}\psi_1'''(\mathscr{X}_i)+0(h'^5)\right].$$



もし T_i に対する推定値として第一項をとれば

$$\hat{T}_i=N_mh'\psi_1'(\mathscr{X}_2)$$

$$h'=\frac{1}{n}\quad \psi_1'(\mathscr{X}_i)=\psi'(X)\frac{dX}{d\xi}=\frac{x}{m}X$$

あるいは

$\psi_1'(\mathscr{X}_i)\fallingdotseq N_m\psi'(X)\varDelta X_i$

＝隣接標本間の幅×標本値

あるいは $\dfrac{N}{n}xX=\dfrac{\text{絶対順位}}{\text{標本数}}$ ×標本値である．

誤差評価に関しては

$$|E_i|=\left|\frac{N_mh'^3}{3!}\psi'''(\mathscr{X}_i)\right|$$

$$\psi_1'''=\frac{md^2}{d\mathscr{X}^2}(xX)$$

したがって

$$|E_i|=\frac{Nh'}{24}(x_{i+1}X_{i+1}+x_{i-1}X_{i-1}-2\,x_iX_i)$$

$$\fallingdotseq\frac{N}{24}|\varDelta x_{i+1}X_{i+1}+\varDelta x_{i-1}X_{i-1}-2\,\varDelta x_iX_i|$$

$|\ |$ の内部は隣接部分との差額の合計になる．

総計については

$$\hat{T}=\sum x_i\varDelta X_i$$

として

$$|E_T|=\frac{Nh'}{24}\sum_{i=1}^{n-1}|E_i|+|E_0|+|E_n|$$

格差が充分大ならば

$$\sum_{i=1}^{n-1}|E_i|=\frac{N}{24}(\Delta x_n X_n-\Delta x_{n-1}X_{n-1}+\Delta x_1 X_1-\Delta x_i X_i).$$

（3）　ウェイト附撰択　一般的にもし抽出対象に分配法則 $Y=\psi(X)$ なる関係が明かならばローレンフカーブの単調性から

$$\mathcal{X}\fallingdotseq\psi^{-1}(X)$$

なる順位変換を施した後，等間隔抽出を適用するとすれば当然

$$\frac{dY}{d\mathcal{X}}\fallingdotseq 1,\qquad \frac{d^2Y}{d\mathcal{X}}\fallingdotseq 0$$

によって容易に効果を挙げることができる．前記（1）の system をサイズ・プロポーショナルであるとすればこの（二）は評価に関する一種の weight づけ撰択となる．

したがって分配法則についての認識を増し，近似度をたかめれば調査精度も上る．

この非等間隔の組織抽出は高額層をち密に，低額層を疎に選択するものであるから全体的に格差の大なる集団の調査に当ってさけられないが（データ参照）細部については（一）の（5）(ii)（iii）に掲げた問題を残しているのである．これを改めて次の（三）で一括推論する．

（三）　組織的抽出の簡便法

（一）（二）の方法は完全に実施するには極めて困難であるが，分配の概形を捉えて先大雑把な階級分類を行なう際クラス抽出，いわゆる「クラスターサンプリング」を行なって推計および誤差を計量する際には有効でありかつ幾多の調査の先例からして必要であろう．

ここではそうした操作の後各階級内から標本を選択する実用方式を考察したい．それは一般的には（一）（二）によって各階級内部の格差が減少しているにも拘わらず，もしこの内部の順位づけがあまりにも不完全である場合，あるいはランクが結果のランクと一致しない場合（一）（二）の推計を機械的に適用すれば可成りの誤差を齎すものと予想されるからである*.

もちろん各階級の標本は（一）（二）が可能であるとした際の理論的推理によって決定される．

（1）　各階級内の数量範囲（range すなわち最高-最低）R_i がサイズ N_i に比較して零細あるいは N_iR_i が階級間の（一）（二）による誤差に比して相対的に零細である場合．

この場合，この階級の総量並びに推定総量は $I_ix_{i(\max)}$（階級区間×最高額）と $I_ix_{i(\min)}$（階級区間×最低額）の中間に所在するであろう．したがって（一）にしたがって選択した標本から階級総量を推計した $\hat{T}_i$ につき最も単純には

$$|T_i-\hat{T}_i|<N_iR_i$$

この R_i は先験的あるいは標本値から推定できるとし．また間接的に隣接層と平均値の差をとり，これらの値の総合において決定することもできる．

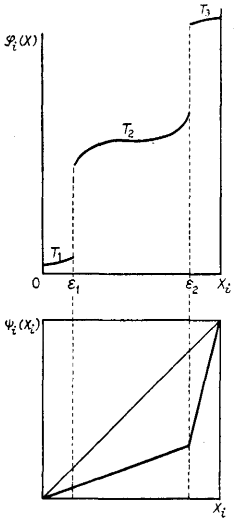
前頁の（二）にしたがう時はこれは

$$|T_i-\hat{T}_i|<N_iR_ix_iX_i$$

が（二）のそれぞれに追加される．

（一）（二）にしたがう階級分割が合理的かつ，零細ならば区間内を恣意的選択に任せてもこの方式によって充分であろう．また $\varphi_1(X)\leqq\varphi(X)\leqq\varphi_2(X)$ ならば $E\leqq N_m\int_0^1[\varphi_2(X)-\varphi_1(X)]dX.$

*註　従ってこの（三）はランダム・サンプリング・メソッド適用に関する別個の考察と受取れるが「ランダム・サンプリング・メソッド」の方式を公式的に適用することを意味しない．結果に組織的順位抽出法が直接の効果をもたない場合である．



（2）　R_i が大あるいは未知である場合

しかしながら，階級分けによる格差の低下はあくまで一般的にであり個々については大きく変動する結果 Range が過大になる可能性が生ずる．

階級分類の基本方針（二）(3) にしたがうとそれは主に高低両局の微小範囲 T_{1i}, T_{3i} に帰因する．

もし T_{2i} の部分の range R_{i2} が小であるならば両極端のサンプルを除外し，例えば（一）の場合 $\hat{T}_i=\dfrac{N_i}{n_i}\sum_j x_{ij}$ で推計すれば

$$|T_{2i}-\hat{T}_{2i}|\leqq R_{i2}N_i$$

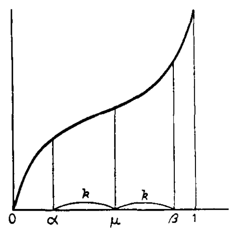
$$T_{1i}\fallingdotseq\psi(0)\varepsilon_1,\quad T_{2i}\fallingdotseq\psi(1)\varepsilon_2$$

もし ε が充分小であれば当然これらは無視しうるのである．標本が特にこれらの両端に集中しているとみられる時は ε が小なる限り例外的であり，また常識的にさらに標本を追加するか，事前に検討を重ねて層を分割するであろうから，それほど問題にならぬかも知れぬが，そうした場合の配慮を含めて，次に中央値乃至最頻値を用いる推計を考察しよう．　なお，図中にある T は，煩雑を避けるため i の添字を略してある．

（3）　(2) の $\varepsilon_1, \varepsilon_2$ の部分が比較的大である場合

μ をある階級内の任意の一点とし，

$$\mu-\alpha=\beta-\mu=k$$



なる範囲で考えるとき

$$r=\varphi(\beta)-\varphi(\alpha)=2k\varphi'(\mu)+\frac{k^3}{3}\varphi'''(\mu)+0(k^5)$$

註　ここで μ が最頻値であれば γ が最小である事が分る．

今この内部で $e_1=2kr$ とすれば

$$e_1=4k^2\varphi'(\mu)+0(k^3)$$

さらにこの範囲からはずれた部分につき

$$e_2=\int_0^\alpha\varphi(X)dX+\int_\beta^1\varphi(X)dX=1-\{\psi(\beta)-\psi(\alpha)\}$$

$$=1-2\left\{k\varphi(\mu)+\frac{k^3}{3!}\varphi''(\mu)+0(k^5)\right\}$$

今　$$e=e_1+e_2=1-2k\varphi(\mu)+4k^2\varphi'(\mu)+0(k^3)$$

として μ を固定した時 e が最小となる条件をもとめると

$$\frac{de}{dk}=-24(\mu)+8k\varphi'(\mu)=0$$

$$\therefore\ k_\mu=\frac{\varphi(\mu)}{4\varphi'(\mu)}$$

また　$$r_\mu=2k_\mu\varphi'(\mu)=\frac{\varphi(\mu)}{2}$$

$$e_{\mu.\min.}\fallingdotseq1-2k\varphi(\mu)+4k_\mu{}^2\varphi'(\mu)$$

$$=1-\frac{\{\varphi(\mu)\}^2}{2\varphi'(\mu)}+\frac{\{\varphi(\mu)\}^2}{\varphi'(\mu)}$$

$$=1-\frac{\{\varphi(\mu)\}^2}{4\varphi'(\mu)}$$

ここで $\dfrac{\varphi'(\mu)}{\varphi(\mu)}$ は，μ における評価額の順位弾力性 $\varepsilon_\mu=\dfrac{d\log x}{d\mu}$ を示すからもし ε_μ あるいはその密度が既知ならばある中心部に近い標本値 x_μ をきめると $x_\mu\pm\dfrac{R}{2}$ の範囲すなわち，$\dfrac{3}{4}x_\mu\leqq x_i\leqq\dfrac{5}{4}x_\mu$ につき $x_\mu\times 2k\dfrac{N}{n}$ あるいは $N\dfrac{2k}{n}\sum x_i=\dfrac{N}{nm}\times\dfrac{x_\mu\times x_\mu \text{の密度}}{2}\times\sum x_i=\dfrac{N}{nm}\cdot\dfrac{1}{2E_\mu}$ で推計すれば，この階級内の総額の前節の立論に基づく誤差は最小となり

$$E_{\mu.\min.}=Nme_{\mu.\min.}=Nm\left[1-\frac{x_\mu}{m}\cdot\frac{1}{4\varepsilon_\mu}\right]$$
$$=Nm-\frac{Nx_\mu}{4\varepsilon_\mu}\doteqdot\text{推計量}-\frac{Nx_\mu}{4\varepsilon_\mu}$$

で支えられる．亦 μ をモードに近くとれば効果があり平均値にとれば $e=1-\dfrac{4\varphi'(\mu)}{1}$ となる．

μ 従って x_μ は一般に任意であるが，

$$2\varphi'^2(\mu)=\varphi(\mu)\varphi''(\mu)$$

すなわち順位弾力性 $=\dfrac{1}{2}\times$ 弾力性の変化率を充していればEは最小となり，一方 ε_μ は，標本によってある程度の推測を下しうる*.

（4） 然るに（3）の方法による時誤差は猶過大であるからこれを補うために

$0\leqq X\leqq\alpha$ につき $\varphi(\mu-k)$ を

$\beta\leqq X\leqq 1$ 部分は $\varphi(\mu+k)$ によって推計するならば

$$e_2=1-2k\varphi(\mu)-(1-\mu-k)\left\{\varphi(\mu)+k\varphi'(\mu)+\frac{k^2}{2}\varphi''(\mu)\right\}$$
$$-(\mu-k)\left\{\varphi(\mu)-k\varphi'(\mu)+\frac{k^2}{2}\varphi''(\mu)\right\}+0(k^3)$$

$$e=1-\varphi(\mu)+k\varphi'(\mu)(-1+2\mu)+k^2\left\{4\varphi'(\mu)+\frac{1}{2}\varphi'(\mu)\right\}+0(k^3)$$

$\dfrac{d\rho}{dk}=0$ のとき

$$\varphi'(\mu)(2\mu-1)+2k\left\{4\varphi'(\mu)+\frac{1}{2}\varphi''(\mu)\right\}=0$$

$$k=\frac{(1-2\mu)\varphi'(\mu)}{8\varphi'(\mu)+\varphi''(\mu)}\quad \text{例えば指数型ならば}\quad k<\frac{1}{8}$$

この時

$$e_{\mu.\min.}=1-\varphi(\mu)-\frac{1}{2}\frac{\varphi'(\mu)(1-2\mu)^2}{8\varphi'(\mu)+\varphi''(\mu)}$$

μ が平均に近ければ　$\doteqdot-\dfrac{1}{2}\dfrac{1-2X_m}{8mf_m-\dfrac{f'_m}{f_m}}$

モードのとき　$=\dfrac{1-2X_{\text{mod.}}}{16mf_{\text{mod.}}}$

メディアンならば　$=1-\dfrac{X_{\text{mod.}}}{m}$

---

* 注　ここで集団について弾力性値の如く或種の情報を必要とするが，この点はランダム・サンプリングの分散についても同様である。

μ の適正値は

$$\frac{de_{\mu.\min.}}{d\mu}=\varphi'(\mu)+2k\varphi'(\mu)+(-1+2\mu)\varphi''(\mu)=0$$

で与えられる.

又推計値は例えば平均の場合

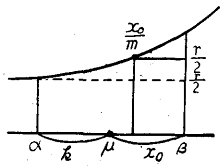
$$\hat{m}=m\{2k\varphi(\mu)+\alpha\varphi(\alpha)+\beta\varphi(\beta)\}$$
$$=m\left\{\varphi(\mu)+k\varphi'(\mu)(2\mu-1)+\frac{k^2}{2}\varphi''(\mu)\right\}$$

さきの k の値を用いると

$$\hat{m}\fallingdotseq m\varphi(\mu)-\frac{\varphi'^2(\mu)\,(2\mu-1)^2\{16\varphi'(\mu)+\varphi''(\mu)\}}{2\{8\varphi'(\mu)+\varphi''(\mu)\}^2}$$
$$R=2mk\varphi'(\mu)=\frac{m(1-2\mu)\varphi'(\mu)}{4\varphi'(\mu)+\dfrac{\varphi''(\mu)}{2}}$$

（5） 処で以上の（3）（4）においてより厳密に考えるならば実際上は x_0 及びそれを狭む集計範囲 $\pm\frac{r}{2}$ をさきに決定するのであるからそれを上式に適用するならば

$$\mu\fallingdotseq\varphi^{-1}\left(\frac{x_0}{m}\right)+\frac{r^2}{8}\varphi^{-1''}\left(\frac{x_0}{m}\right)$$
$$k\fallingdotseq r\varphi^{-1'}\left(\frac{x_0}{m}\right)$$



として,

$$\frac{\partial e}{\partial r}=0\quad\frac{\partial e}{\partial x_0}=0$$

等によって r, x_0 を決定すべきであるがここでは省略する.

又, $\varphi(X)$ を多項式近似して推計する方法も可能であるが省略する.

(四) ここでは実際に即して以上の (一) (二) (三) を取纏めてみる.

(一) (二) は, 既存データにより目的数量を支配する全体的関係の存在を予想し, 従来のいわゆる「層別法」の効果をたかめることを指示する.

(三) は, 準備段階での順位と, 最終的順位とのズレを補正するとともに, ランダム・サンプリング・メソッドを含む各層内での抽出および集計作業の簡約法を示すものである.

もちろん以上の方法はその概略にふれたのみであり, その形式や, より細部にわたる検討は続けられねばならない, ただ以上をもってしても分配の分析と調査の企画に一貫性同時性をもたせる意味である程度の示唆を斉すものと思う.

§3. 撰択推計および仮設の検討問題について（いわゆる検推定について）

(一) 筆者は今推定および検定に関する広範なジャンルのすべてにわたって, 提案を行なう訳にはいかないので, 差当り §1 (一) の体系に基づいて格差 (Gini 係数) を算出することを試みる*Φ. $(X)=\int\psi(X)dX$ とおき, その Taylor 展開可能性から略 Total の推計に準じて

$$G=1-2\sum_{i=1}^{n}\{\Phi(\xi_i)-\Phi(\xi_{i-1})\}$$

---

* 注　Kendall 上掲書 p. 43.

$$\text{Gini coef.}=\frac{\text{mean difference}}{2\times\text{mean value}}$$

$$=1-4\sum\left\{h\psi(X_i)+\frac{h^3}{3!}\psi''(X_i)+0(h^5)\right\}$$

ところで

$$\psi(X_i)=\int_0^{X_i}\varphi(X)dX=\sum_{j=1}^{i}\int_{\xi_{j-1}}^{\xi_j}\varphi(X)dX-\int_{X_i}^{\xi_i}\varphi(X)dX$$

$$=\sum_{j=1}^{i}\left\{2h\varphi(X_j)+\frac{2h^3}{3!}\varphi''(X_j)+0(h^5)\right\}-\left\{h\varphi(X_i)+\frac{h^2}{2!}\varphi'(X)+0(h^3)\right\}$$

したがって

$$1-4\sum h\psi(X_i)=1-4h\sum_{i=1}^{n}\left\{\sum_{j=1}^{i}2h\varphi(X_j)-h\varphi(X_i)\right\}$$

$$+4h\sum_{i=1}^{n}\left\{-\frac{h^2}{2!}\varphi'(X_i)-\sum_{j=1}^{i}\frac{2h^3}{3!}\varphi''(X_j)\right\}+0(h^3)$$

いま $\hat{G}=1-4h\sum_{i=1}^{n}\{\sum_{j=1}^{i}2h\varphi(X_j)-h\varphi(X_i)\}$ を G の推定に用いる時

$$\hat{G}=1-4h^2\sum_{i=1}^{n}\{2(n-i+1)\varphi(X_i)-\varphi(X_i)\}$$

$$=1-4h^2\sum_{i=1}^{n}(2n-2i+1)\varphi(X_i)$$

$$=1-\frac{8h}{m}\sum x_i-4\frac{h^2}{m}\sum_{i=1}^{n}(2i-1)x_i$$

$$=1-2\frac{\bar{x}}{m}+4\frac{h^2}{m}\sum_{i=1}^{n}(2i-1)x_i$$

故に近似的には

$$1-8\frac{h^2}{m}\sum(n-i)x_i=1-\frac{2I}{T}\sum\left(1-\frac{i}{h}\right)x_i$$

あるいは　$$-1+\frac{8h^2}{m}\sum_{i=1}^{n}ix_i=1+\frac{2I}{T}\sum_{i=1}^{n}\frac{i}{n}x_i$$

ここに $I=Nh$ 抽出間隔.

その誤差は

$$E=4h^3\sum_{i=1}^{n}\left\{-\frac{\varphi'}{2!}(X_i)+\sum_{j=1}^{i}\frac{2h}{3!}\varphi''(X_j)\right\}-\sum_{i=1}^{n}\frac{4h^3}{3!}\psi''(X_i)$$

$$=4h^3\sum_{i=1}^{n}\left\{\sum_{j=1}^{i}\frac{h}{3}\varphi''(X_j)-\frac{2}{3}\varphi'(X_i)\right\}$$

したがって

$$|E|\fallingdotseq\sum_{i=1}^{n}4h^3\varphi'(X_i)=\frac{h^2}{m}(x_n-x_1)$$

を用いることができよう.

推定に関してはなお

$$\lambda=\frac{X-Y_1}{X-Y_2}$$

が確定的である時回帰乃至比推定に当る方式を得ることができる.

（二） §1 によって危険率の伴わない推計を得ることに努めたのであるからもし予想の適否を判断したいとすれば例えば総量につき単に

$$|T-\hat{T}|<|E|$$

の成立の可否をみればよい.

以上の方法の目的はむしろ計画に寄与するデーターを得ることにあるのもその意味では(一)

の G の推計式に見られる G, N, T の例えば資本順位と個別の累進的分配関係を体制機構との関連の下にいかに理解利用するかが問題である．さらに立入って論ずると勿論そもそも獲得した統計資料がどのように無法則的或は非決定的といえるか否かが重要なのである．それは，資料獲得後に初めて偶然見出されるよりは，§1 における設計計画時に充分検討を重ねる必要があるのである．

したがって §1, §2 は，科学的分析の目的において相互に緊密な連繫を保つものと思われるので結びにかえて，次節に一の仮設を掲げてみたい．

§4.　統計調査，分析の科学的目的

一般的に統計資料の蒐集作製，分析に用いる数学的方法においては，近来単に「分散」の度合や「相関関係」の有無等による俗流化された方法により処理され勝ちな傾きがある．

例えば §1 に触れた調査法に関する基本的な立場はあくまでランダム・サンプリングに置かれており，それは当然確率論として抽象されるある状態条件を対象に仮定するものである．

しかし以上に述べたごとく，科学的性格をモットーとする場合，かかる仮設そのものの成立条件を理論的，実証的に確認せねばならない．これは同時に，以上の方法の前提とした $\psi(X)\varphi(X)$ なるものの存在と展開可能性という仮設にも触れることになる．実際上例えば先以上の諸論を否定する条件として，調査前に何等かの手段によって予想される順位 X に対応する調査結果 $y=\dfrac{x}{m}$ につき $\dfrac{dy}{dX}$ が存在しなかったり単純でない状態を考察することにする．しかしこのことをもって，統計的無法則的状態として反語的に確率論的仮設を設けるべきかについては極めて疑問である．むしろ経験確率的立場は特定の順位に関する，特定の評価額が $X\to 1$ の極限において特定の収斂するという特異な条件下にあるものといえるのではあるまいか．そしてこれらの何れでもない，例えば $\dfrac{dy}{dx}$ が 1 以外の一点で収斂したり，連続性も，収斂性ももたぬが，ある一定の範囲内に留まる等確率外的にして何等かの合法則性を予想させる場合すら考えられるのである．

例えば経済面においてパレート，チフ，ユールの法則を主張するケンダル* の論拠にも拘らず筆者は敢て「ランダムな状態，行為」の従来の想定を極力避け寧ろその段階性や対極的なマクロの力学系を予感するのである．統計学の対象の検討を通じて数理統計学の発展的解消を図ることこそ本来の目標であろう．なお，参考文献の提示は最小限に留めることを付記する．

付：実態調査に基づく参考資料

（昭和 30 年度国富調査法人企業資産に関する諸データ）

ここでは法人企業資産を調査することを目的として，本文 §1 の方法を適用する場合の参考資料である．これを産業別にみても類似の結果を得たことを付記しておく．

（1）　§1,（一）に仮定した順位の実際の対応を示す，ローレンツ・カーブである．ただし順位は資産額高額順によった．本文中 $Y=\psi(X)$ にあたるものである．

（2）　同じく §1,（一）において，準備段階における順位を資本金額によった場合の資産額との対応を実線で示したものである．

（3）　§1,（二）の方式にしたがい，(1) の順位を対数順位に変更した場合のグラフである．この場合グラフは均等線に極めて接近し，したがって等間隔抽出によるのがほぼ適切といえる．

註 *　M. G. Kendall; Natural Law in the Social Science, Journal of the Royal Statistial Society, Vol. 124, 1961.

なお (2) 図の点線は同様の趣旨の下に資本金額順位を変更したものであり，やはり効果ある結果を認めうる．

(4) および (5) は，各資産額および資本金順位による分配を示したものである．本文中 $y=\varphi(X)$ にあたる．

(6) および (7) は各資産額および資本金額順位による y の変動を本文 §2(二) (3) さらに重複対数によって追求したものである．

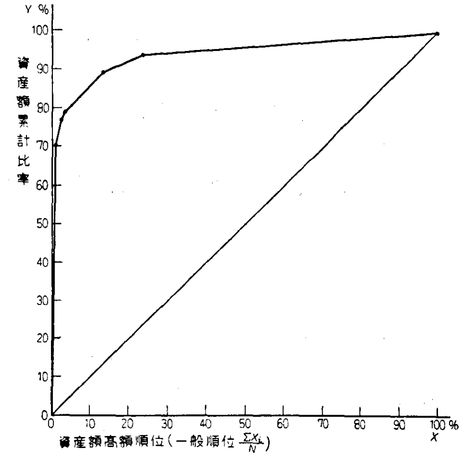
(8) 表は，順位による分析の例として産業順位表を作製した．最初の二欄および後の二欄の対応に特に興味を抱いた．

産業順位表

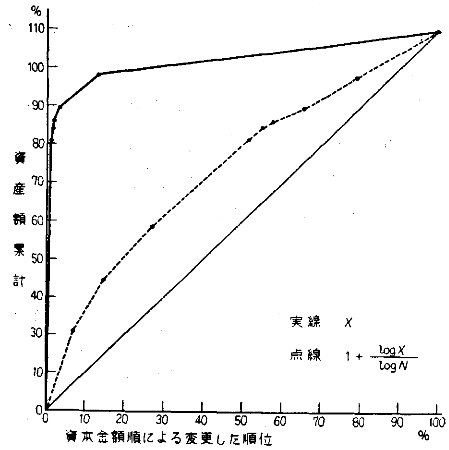
種別 / 産業名	資本金総額による順位	資産総額による順位	一法人当り平均資産額による順位	格差(ジニ係数)による順位
農林業	31	31	31	13
水産業	25	23	13	7
石炭鉱業	14	15	2	14
その他の鉱業	16	18	9	11
建設業	19	12	23	23
食料品製造業	8	8	21	22
繊維製品製造業	5	4	14	15
木材および木製品，家具製造業	18	11	25	25
紙および類似品製造業並びに出版印刷業	13	13	15	29
化学工業	4	5	11	8
ガラスおよび土石製品製造業	15	17	16	24
第一次金属製造業	6	6	10	4
金属製品製造業	26	25	26	20
機械工業	2	3	12	17
その他の製造業	30	28	28	31

種別 / 産業名	資本金総額による順位	資産総額による順位	一法人当り平均資産額による順位	格差(ジニ係数)による順位
卸売業	3	1	20	10
百貨店業	23	21	7	2
その他の小売業	9	7	30	30
銀行および信託業	7	16	4	12
保険業	21	30	29	9
その他の金融保険業	12	20	27	18
不動産業	20	24	22	28
地方鉄道軌道業	10	14	3	27
道路運送業	27	22	18	16
水運業	17	9	8	5
倉庫業	28	27	19	19
その他の運輸業	24	26	17	26
通信業	29	29	6	3
電気業	1	2	1	6
ガス業	22	19	5	1
営利サービス業	11	10	24	12
全31産業				

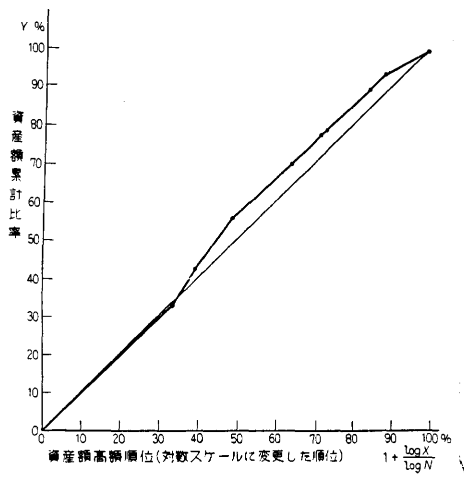
統計数理研究所



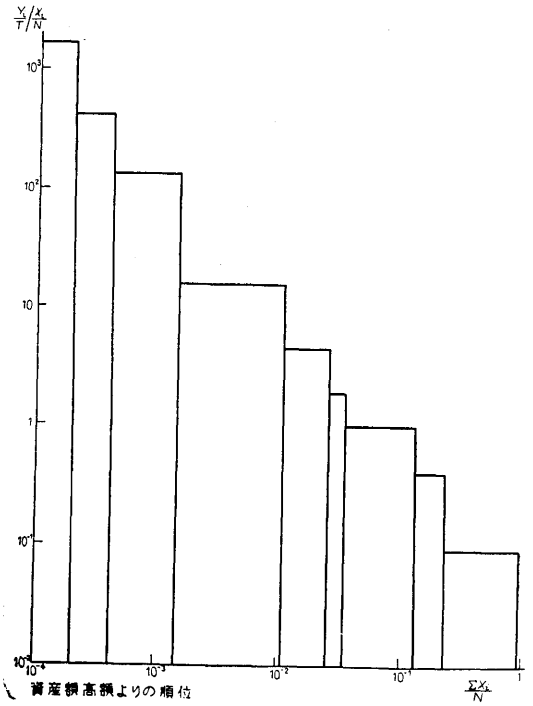
（1） 全 産 業



（2） 全 産 業

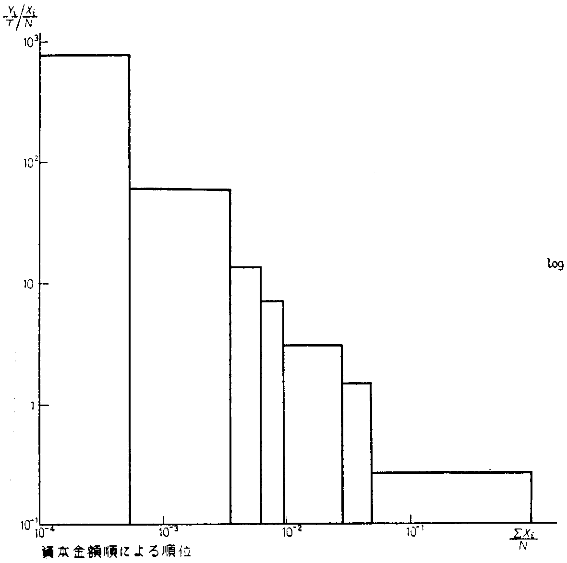


（3） 全 産 業

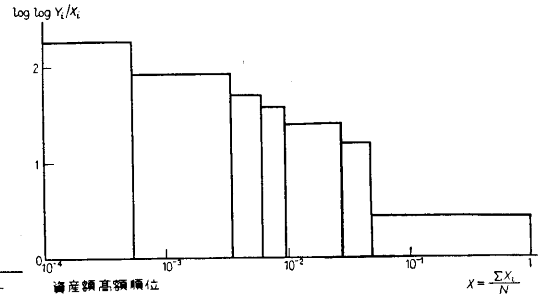


（4） 全 産 業

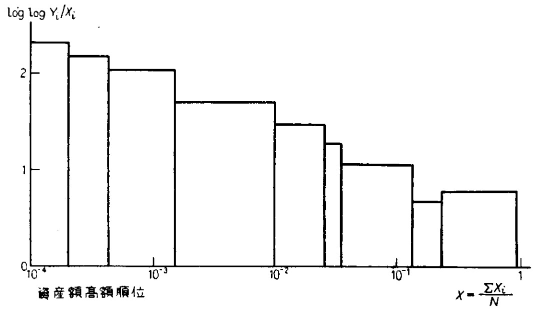
クラス別平均分配比率の順位による変化



（5） 全　産　業
クラス別平均分配比率の順位による変化



（6） 全　産　業 （資本金）



（7） 全　産　業
階級平均の重複対数