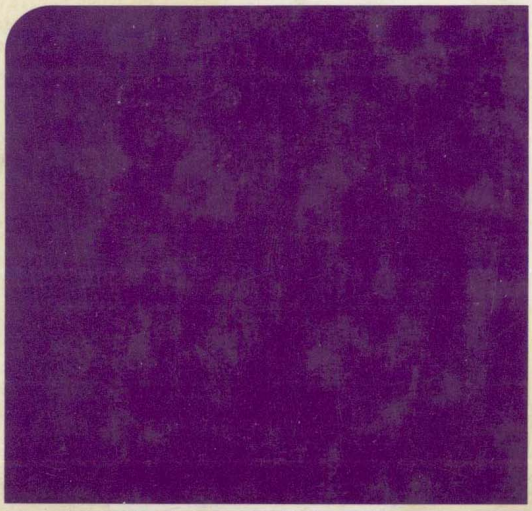
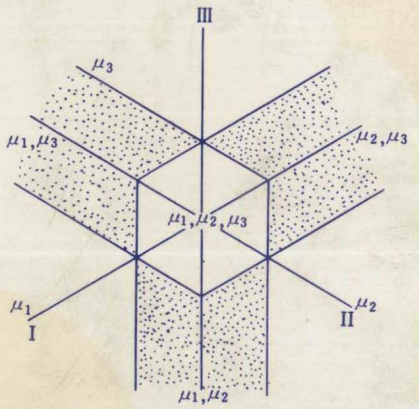
数理統計学の方法的基礎

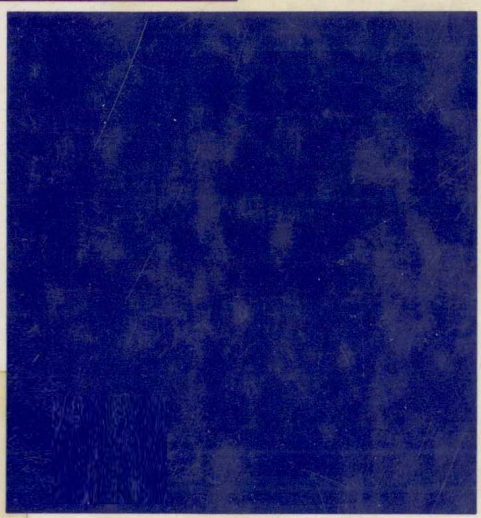
STUDIES IN SOME ASPECTS OF THEORETICAL FOUNDATIONS OF STATISTICAL DATA ANALYSIS

竹内　啓　著

東洋経済新報社







数理統計学の方法的基礎

STUDIES IN SOME ASPECTS OF THEORETICAL FOUNDATIONS OF STATISTICAL DATA ANALYSIS

竹内　啓 著

東洋経済新報社

著者紹介

1933年　東京に生まれる.
1956年　東京大学経済学部卒業.その後同大学大学院,同大学同学部助手を経て,1963年4月助教授となり現在に至る.
専　攻　数理統計学（経済学博士）
著　書　『数理統計学』『多変量解析の基礎』『計量経済学の研究』いずれも東洋経済新報社,『線型数学』培風館,『社会科学における数と量』東京大学出版会,『転機にたつ科学』『数学の世界』（共著）共に中央公論,など.
訳　書　フィッシャー著『統計的方法と科学的推論』岩波書店,ジョンストン著『計量経済学の方法』東洋経済新報社,など.
住　所　神奈川県鎌倉市山崎1488-9

数理統計学の方法的基礎

昭和48年7月16日発行

著　者　竹内啓（たけうち　けい）
発行者　宇梶洋司
発行所　東京都中央区日本橋本石町1の4　東洋経済新報社
郵便番号 103　電話東京(270)代表4111　振替口座東京6518

3033-3941-5214

はしがき

この本に「数理統計学の方法的基礎」という表題をつけたが，本書は体系的な展開を行なっているわけではない．より正確には「数理統計学の方法的基礎の諸側面に関する若干の問題についての研究 Studies in some aspects of theoretical foundations of statistical data analysis」とでもいうべきであろうと思うが，書名としては長すぎるので簡略化したしだいである．

数理統計学の理論として，いわゆる Neyman-Pearson 理論と呼ばれるものが，現在では一応体系化され，教科書などでも説かれていることは，改めていうまでもない．これに対して，以前から R. A. Fisher の批判があり，最近ではいわゆる Neo-Bayesian 学派の人びとにより根本的に異なる考え方が提起されているが，いまや古典的ともいえる Neyman-Pearson 理論は，基本的な枠組においては依然として有効性を失ってはいないと筆者は考えている．

しかしながら，教科書的な Neyman-Pearson 理論には，いろいろな不十分な点，あるいは欠落があることも否定できない．それは，たとえばモデルが現実において必ずしも正確でないという事実の評価と，それをいかに推測理論に反映させるかの問題，あるいは推測の形式として推定や検定，あるいは区間推定というような形以外のタイプを考えること，あるいはデータの作り方と推測

の過程をどのように結びつけて考えるかの問題，等々である．これらの点において，伝統的な理論ではうまく取り扱われていない問題があり，それらを基礎から考え直すことによって，数理統計学の理論体系を，ある意味で古典的な理論の延長の上により豊かに肉づけすることが可能であると思われる．

この本は，以上のような方向へ向かっての筆者のいくつかの試みをまとめたものである．

内容的には，これまでいくつかの機会に発表したものを含んでいる．特に第2章は『季刊経済学論集』38巻1972年(東京大学経済学会)，第3章および第4章は「Robust Estimation についてⅠ，Ⅱ」という標題で『応用統計学』2巻1972年，3巻1973年(応用統計学編集委員会)に発表したもののほとんどそのままの再録であることをお断りしておかねばならない．また他の章についても，第7章の有限母集団に関する推定理論については1960～1965年ころに，また第9章の randomized design の理論については1960～1963年ころに一連の論文を発表したことがあり，ここでは改めてそれらの主要な内容をとりまとめ，かつその意味づけを中心として論じた．

なお，この本のとりまとめにあたっては，東洋経済新報社の丸山常喜氏に筆者の前著(『計量経済学の研究』など)にひきつづいてお世話になった．ここに記してお礼を申しあげたい．

1973年5月

竹内　啓

目　　次

1

統計的推測理論の方法的諸問題

1.1 統計的推測理論の問題点

統計的推測理論 theory of statistical inference と呼ばれるものも，その正確な呼び方はともあれ，実質的にはすでに長い歴史をもっているといえよう．わが国ではときに推計学あるいは推測統計学の名で呼ばれ，R. A. Fisher によれば，精密標本論 exact sampling theory といわれる統計量の厳密な標本分布に基礎をおく検定理論にしても，すでに 60 年以上の歴史をもっている．また統計的推定論にいたっては，150 年以上も前の C. F. Gauss の最小 2 乗法にまでさかのぼることができる．

しかし，その中には依然として一つの論理的困難が含まれ，ほとんどすべての統計学者が納得するような形で，それを解決することは近い将来においては望みえないように思われる．それは 19 世紀においては，P. S. de Laplace の権威によって裏付けられた“Bayes の規則”をめぐる問題点であり，1930 年代からは，R. A. Fisher と J. Neyman-E. S. Pearson との論争に現われ，また最近では，J. Savage 以後の Neo-Bayesian 学派とその批判者との間に争われているところである．この章において，まずその問題点についての解釈を述べて，その困難なところを明らかにするとともに，筆者自身の立場を述べたいと思う．

この問題が，ある意味では完全な解決に到達することはできないようなものであるにしても，そのことは統計的推測理論を現実に適用することを妨げるものではないし，またその適用範囲の広さを広げ，あるいはその性質についての理解を深めることを不可能にするものではないと思う．さしあたっては，一つの作業仮説を設けることによって，その難点を回避するとともに，現実の統計データの処理において，有効性を高めるように統計的推測理論の内容をより豊かにすることが必要である．次章以降には，この方向に向かってのいくつかの試みについて述べよう．

統計的推測理論の基本的な枠組が，次のようなものであることについては，さしあたって問題はないであろう．

まず，現実のデータ $x=(x_1, \cdots, x_n)$ は，ある確率変数 $X=(X_1, \cdots, X_n)$ の一つの実現値であると考えられる．ところでXの分布については，少なくともその一部は未知である．したがってわれわれはXの可能な分布の集合

$$\mathscr{P}=\{P_\omega : \omega \in \Omega\}$$

を想定する．ここでωはただ分布に付けられたラベルであって，それ以上の意味はないと考えておく．

ところで，われわれがデータから知りたいと思うのは，未知の確率分布 P_w によって決定されるある量である．それはある場合には一つ，あるいはいくつかの定数からなるベクトルθであって，なんらかの意味で客観的な対象の性質を反映するものである．またある場合には，将来のデータの値であるかもしれない．狭い意味の推測 inference とは，前者の場合であるとして，後者すなわち予測 prediction とは区別することにしよう．すなわち，推測とはωの実数値（あるいは実ベクトル値）関数 $\theta=\theta(\omega)$ について，なんらかの判断を下すことであると考える．判断を下す形式はいろいろな形が考えられるから，可能な結論の集合をDとしよう．そこで問題は，与えられたデータに対応してDの要素を選ぶことである．

いま最も簡単な場合として，ある物理量θをn回操り返し観測して，観測値

$x_1, \cdots, x_n$ を得たとしよう．ここで $x_1, \cdots, x_n$ はそれぞれ互いに独立に平均 θ，分散 σ^2 の正規分布に従う確率変数 $X_1, \cdots, X_n$ の実現値であると仮定されるものとしよう．

いま，正規分布の仮定については問題はないものとすると，θ の推定量としては，

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$$

が最もよいものであることはまず異論のないところである．すなわち，$X_1, \cdots, X_n$ を確率変数として考えたとき，それから確率変数 $\bar{X}$ は，平均が真の値 θ に等しく，かつ同じような性質をもつ $X_1, \cdots, X_n$ から計算される確率変数(推定量)の中で分散最小である．したがって，たとえば $X_{med}=X_1, \cdots, X_n$ の中央値とすると，

$$E(\bar{X}-\theta)^2<E(X_{med}-\theta)^2$$

である．実際，より詳しく次のことがいえる．すなわち，ε を任意の正数とすると，

$$\Pr\{|\bar{X}-\theta|>\varepsilon\}<\Pr\{|X_{med}-\theta|>\varepsilon\}$$

となる．すなわち，$\bar{X}$ の誤差が一定の値をこえる確率は，X_{med} の誤差が同じ値をこえる確率よりつねに大きい．このことから，$\bar{X}$ が X_{med} より“よい”推定量であることには疑問の余地がないであろう．

しかしこのことは，得られた値 $x_1, \cdots, x_n$ から計算した平均値 $\bar{x}=\sum x_i/n$ および中央値 $x_{med}=x_1, \cdots, x_n$ について，

$$|\bar{x}-\theta|<|x_{med}-\theta|$$

となっていることを保証するものではない．$\bar{X}$ および X_{med} の確率分布については，

$$\Pr\{|\bar{X}-\theta|<|X_{med}-\theta|\}>\frac{1}{2}$$

であるが，具体的な二つの値 $\bar{x}$ および x_{med} については $|\bar{x}-\theta|$ は $|x_{med}-\theta|$ より小さいか大きいかのどちらかであるが，どちらであると断言することはでき

ない．

ここで，われわれは特定の場合について $\bar{x}$ と x_{med} とどちらがよいかをいうことはできないとする立場がある．ただ推定値を計算する“ルール”として，$\overline{X}$ と X_{med} とを比較すると，前者を採用すれば，何回も操り返し測定を行なう場合に，平均的には前者のほうが誤差が小さくなる．そういう意味で，問題にすることができるのは，具体的な推定値ではなく，推定値を求める“ルール”あるいは計算方式なのである，というのが Neyman の主張であり，彼はデータから真の値へという inductive inference なるものはありえず，ただデータに基づく行動方式としての inductive behaviour があるだけと論じたのである．ここで同じ対象について，同じ方式で推定値を操り返し計算すると想定することは無意味である，との主張に対しては，計算方式はそれぞれ違う対象に対して適用されてもよい，それでも平均的な性質についての結論は変わらないというのが Neyman の反論であった．

これに対して，統計的方法は本来観測データからの帰納法にほかならないと考える Fisher は強く反対して，具体的な個々のデータについて何もいうことができないのでは無意味であると主張した．もし $\overline{X}$ のほうが X_{med} より確率的に誤差が小さいとするならば，その実現値である $\bar{x}$ は，x_{med} より“たぶん小さい誤差をもつだろう”と考えることは完全に合理的である，というのが Fisher の考え方であった．

Fisher の考え方の根本的な論理は，次のようなものであると思われる．いままで観測を行なう前の状態を考えてみよう．このときこれから n 回観測を行なって，その結果に基づいて $\overline{X}$ および X_{med} を計算するとすれば，$\overline{X}$ のほうが X_{med} より誤差が小さくなる確率が大きい．したがって，$\overline{X}$ のほうがたぶん“よりよい推定量”となるであろうと考えられる．このような考え方は，確率を厳密に頻度と解釈して，「これから観測される値について計算されるそれぞれ一つの $\overline{X}$ および X_{med} の値について確率を云々することはできない．それは同じことを何回も操り返して行なった際の頻度分布についてだけいえることである」という立場をとるのでないかぎり受け入れられるであろう．実は，

R. E. von Mises の確率の考え方は，このような意味で純粋の頻度説であり，Neyman も，ある場合にはそれに近いことを述べているが，しかし1回の試行におけるある結果の“確からしさ”の測度としての確率概念をまったく否定してしまうことは，統計的推測の基礎としての確率概念を著しく不十分なものにしてしまうであろう．実際，もし確率が頻度以外の何ものでもないならば，わざわざ確率ということばを用いる必要もないのであって，頻度に基づく確率概念にしても，それは“まれにしか起こらないことが起こる可能性は小さい”という形で，頻度を“確からしさ”に翻訳したものにほかならないのである．

ところで，次に実際に観測を行なって，$\bar{x}$ および x_{med} を計算したとき，それを平均的な状況から区別される特別な事情がなく，また θ の値は完全に未知であるとするならば，「$\bar{x}$ が x_{med} より小さい誤差をもつことは確からしい」ということを否定する何の根拠もないであろう．すなわち，現に与えられた特定の場合を，平均的な場合から区別する特定の理由がないならば，平均的な場合についての命題は，当面の場合にもあてはまると考えてよい．これが Fisher の論理である．

上に述べたことは，厳密にいえば Fisher の言明そのものではなく，ある意味では Fisher のいっていることを弱めて常識的な形にいいかえたものであるが，このかぎりでは多くの統計学者にとって受け入れられうるものであると思う．

そこで，以上のことを改めて定式化すれば次のようになる．

推定量の分布についての平均的な性質は，特定の推定値についても意味をもつ．ただし，その推定値を平均的な場合から区別する理由がない場合に限る．

ここで，のちの限定条件が重要である．もし，たとえば θ について何かはっきりしたことが事後的にいえるならば，平均的な性質を特定の場合にあてはめることは意味がなくなるかもしれない．たとえば $X_1, \cdots, X_n$ が区間 $[\theta-1, \theta+1]$ の間の一様分布に従うとき，θ の推定量として，

$$T=\frac{\min X_i+\max X_i}{2}$$

をとると，その分散は，

$$V(T)=\frac{2}{(n+1)(n+2)}$$

となり，また$\hat{\theta}$の確率分布も計算することができる．ところが実際に$x_1, \cdots, x_n$の値から

$$t=\frac{\min x_i+\max x_i}{2}$$

を計算するとき，$u=(\max x_i-\min x_i)/2$ とおくと，$t+u<\theta+1$，$t-u>\theta-1$だから，

$$t-(1-u)<\theta<t+1-u$$

となる．したがって，uが1に近ければ$|t-\theta|$は小さくなり，θの推定量としてのtの誤差は，平均的な分散から推定されるところよりも小さくなるであろう．逆にuが小さければ，tの誤差は大きいと考えられる．したがって，tの精度を表わすものとしては，Tの$U=u$が与えられたときの条件付き分散

$$V_\theta(T|U=u)=\frac{(1-u)^2}{3}$$

を用いるほうがよいであろう．

このような例が示すように，一般に，もし未知母数と無関係な分布に従う統計量(それをFisherは補助統計量 ancillary statistic と呼んだ)があれば，それについての条件付き分布を考えて，推測の問題はその条件付き分布に関して考えるべきだということになる．

次に，推測においては標本から得られる情報はすべて利用しなければならない．たとえば先の正規分布の例で，実は$x_1, \cdots, x_n$は与えられた観測値の一部であって，これ以外に$x_{n+1}, \cdots, x_N$という値が得られていたとすれば，データ全体から得られる最もよい推定値は，

$$\bar{x}_N=\frac{\sum_{i=1}^{N} x_i}{N}$$

であって，この値と比較すると，最初のn個の値から計算した

$$\bar{x}_n=\frac{\sum_{i=1}^{n} x_i}{n}$$

あるいは,

$$x_{med}=\mathrm{med}(x_1, \cdots, x_n)$$

は，平均的な場合より一定の方向に偏っていると考えられるかもしれない．たとえば，ここでもし，

$$|\bar{x}_n-\bar{x}|>|x_{med}-\bar{x}_N|$$

になっているならば，推定値としては，もはや $\bar{x}_n$ は x_{med} よりよいとは考えられなくなるであろう．

ただし，このようなことが生じるのは，$\bar{x}_n$, x_{med} の双方よりよい推定値 $\bar{x}_N$ が存在するからであって，したがって，いちばんよいと考えられる推定値に関しては，このようなことは生じないであろう．実際，二つの推定量 $\hat{\theta}_1$, $\hat{\theta}_2$ についてある集合 S が存在して，

$$\Pr\{|\hat{\theta}_1(X)-\theta|>\varepsilon|X\in S\}>\Pr\{|\hat{\theta}_2(X)-\theta|>\varepsilon|X\in S\}$$
$$\Pr\{|\hat{\theta}_1(X)-\theta|>\varepsilon|X\in \bar{S}\}<\Pr\{|\hat{\theta}_2(X)-\theta|>\varepsilon|X\in \bar{S}\}$$

となるならば,

$$\hat{\theta}_3(X)=\hat{\theta}_2(X) \qquad X\in S \text{ のとき}$$
$$=\hat{\theta}_1(X) \qquad X\in \bar{S} \text{ のとき}$$

とすれば，$\hat{\theta}_3$ は $\hat{\theta}_1$, $\hat{\theta}_2$ の双方よりよい推定量を与えるであろう．そして $\hat{\theta}_3$ は推定値としても，$\hat{\theta}_1$ あるいは $\hat{\theta}_2$ よりもよいと考えることができるであろう．

すなわち，上記の「ただし」以下の意味を定式化すれば，条件付き推測の原理 principle of conditionality inference および完全情報の原理 principle of full information の二つが導かれる．

しかし，このような原理は，数学的に厳密に定式化しようとすると困難にぶつかる．たとえば，条件付き推測については，次のような例を D. Basu が指摘している．

いま，$(X_i, Y_i)(i=1, \cdots, n)$ が互いに独立に平均 $(0, 0)$，分散 $(1, 1)$，およ

び相関係数ρの2次元正規分布に従うとすると，$(X_1, \cdots, X_n)$は母数と無関係な分布に従うから，上記の意味で補助統計量である．$(X_1, \cdots, X_n)$が与えられたとき，$(Y_1, \cdots, Y_n)$は条件付き分布，平均$(\rho X_1, \cdots, \rho X_n)$，分散$(1-\rho^2)$の正規分布に従う．したがって，このことを用いて$\rho$についての推測を考えることができる．

ところで，$(Y_1, \cdots, Y_n)$を同様に補助統計量と考えられるから，これが与えられたときのXの条件付き分布に関してρについての推測と考えることもできる．しかしもしここで両方の補助統計量をいっしょにしてしまうと，それは標本全体$(X_i, Y_i)(i=1, \cdots, n)$となって，その分布はもはや$\rho$に依存することになってしまうし，それについて条件付きで考えるということは無意味になる．

この問題自体は，まず十分統計量を考えることによって解決される．すなわち，このとき十分統計量は$\sum X_i^2+\sum Y_i^2$および$\sum X_iY_i$の二つの量からなっているが，これらはともにρに依存する．また，十分統計量の関数として表わされるような補助統計量を構成することはできない．

しかし，このような形の問題は，つねに十分統計量を考えることによって解決できるとはかぎらない．たとえば，$(X_i, Y_i)(i=1, \cdots, n)$が相関のあるコーシー分布に従う場合，すなわち，

$$X_i=\rho U_i+(1-\rho)V_i$$
$$Y_i=(1-\rho)U_i+\rho V_i$$

となって，U_i, V_iは独立にコーシー分布に従う変数，$0<\rho<1$が実母数というような場合には，上記とまったく同じような状況が生じる．このとき，X_iについてのY_iの条件付き分布を考えるか，Y_iについてのX_iの条件付き分布を考えるかの二つの考え方が生じ，そのどちらかを選択する基準はないことになる．またこの場合，次元の小さい十分統計量も存在しないから，問題は正規分布の場合のようにしては解消しえない．

完全情報の原理については，よりいっそうあいまいな点が残る．この原理から導かれることとして，推測は十分統計量に基づかねばならないということがいえる．もし十分統計量が，未知母数と同じ次元をもつか，あるいは未知母数

と同じ次元をもつ量と補助統計量とからなる場合，すなわち，補助統計量についての条件付き分布に対する十分統計量が，未知母数と次元が等しい場合（このとき条件付き分布に対する十分統計量を Fisher は exhaustive statistic と呼んだ）には，問題は簡単であるが，そうでない場合には簡明な解が得られないことが多い．

たとえば，$X_1, \cdots, X_n, Y_1, \cdots, Y_m$ が，すべて互いに独立に平均 μ の正規分布に従い，かつ X_i の分散は σ_1^2，Y_j の分散は σ_2^2 であるとする．このとき，もし $\sigma_1^2=\sigma_2^2$ が既知ならば，μ の最もよい推定量は，

$$\frac{n\sigma_2^2}{m\sigma_1^2+n\sigma_2^2}\overline{X}+\frac{m\sigma_1^2}{m\sigma_1^2+n\sigma_2^2}\overline{Y}$$

で与えられる．もし σ_1^2, σ_2^2 が未知ならば，Y の代りにその推定値

$$\hat{\sigma}_1^2=\frac{\sum(X_i-\overline{X})^2}{n-1} \qquad \hat{\sigma}_2^2=\frac{\sum(Y_j-\overline{Y})^2}{m-1}$$

を代入すればよいと思われる．ところでこのとき，このようにして作られた推定量

$$\hat{\mu}^*=\frac{n\hat{\sigma}_2^2}{m\hat{\sigma}_1^2+n\hat{\sigma}_2^2}\overline{X}+\frac{m\hat{\sigma}_1^2}{m\hat{\sigma}_1^2+n\hat{\sigma}_2^2}\overline{Y}$$

は，任意に α を定めて構成した推定量

$$\hat{\mu}_\alpha=\alpha\overline{X}+(1-\alpha)\overline{Y}$$

より“よい”とみなしうるであろうか．特に推定値としても，実際に X_i, Y_j の実現値 x_i, y_i から計算された二つの値

$$\hat{\mu}^*=\frac{ns_2^2}{ms_1^2+ns_2^2}\bar{x}+\frac{ms_1^2}{ms_1^2+ns_2^2}\bar{y}$$

ただし，

$$s_1^2=\frac{\sum(x_i-\bar{x})^2}{n-1} \qquad s_2^2=\frac{\sum(y_i-\bar{y})^2}{m-1}$$

および，

$$\hat{\mu}_\alpha=\alpha\bar{x}+(1-\alpha)\bar{y}$$

について，$\hat{\mu}_\alpha$ より $\hat{\mu}^*$ のほうがよいといいうるであろうか．実際，$\hat{\mu}_\alpha$ はもし