

推 計 学 批 判 序 説

勝 矢 俊 一

A critical introduction to the statistical method of inference

SHUNICHI KATSUYA

近代推測統計方法が、推計学なる奇妙なる通称によって、各方面の学問領域に於て、実際統計に応用せられているが、多くはその適用を誤って、単に無意味なる数字計算を弄するのみならず、屢々不当なる結論を出すの結果を示している。

この原因を考察するに、推測統計法の数学的本質に対する無理解を主とするが、又所謂推計学者の無反省な誇大宣伝に幻惑されて、無批判に追従する者の多きにも帰せしめ得ると考えたい。

だから私は、今日学界に於けるこの誤用乱用は、推測統計法の本質を詳しく検討する事と、併せて推計学者の誤れる見解を指摘する事によって、多く避け得べきであろうと考える。本文の本旨もここに存する。

推測統計法の本質は、応用数学の一語に尽きるもので、統計に対する数学の一部門たる確率論の応用である。然しこの応用に当てる基本的の想定は、母集団と標本なる概念である、推測統計等の総ての計算、即ち推測、検定、分散分析及実験計画等のすべての思考の基本となるものは、母集団と標本との概念であって、この両者の関係に於て確率論が適用されると云う点にその特質があると云い得よう。

依って私は先づこの両者の検討より本論を展開したいと思う。

「母集団と標本」

母集団と標本なる概念は、これを一般的に云えば、全体と部分なる概念に包含せしめ得るものであるが、推測統計学に於ては、客観的妥当性を得る為に、次の如き条件が成り立つような論理的構造に基礎をおき、所謂局所推測をなすものであると説明する。

即ち現実の統計資料は、すべてその背後にある同一の標識を有する無限（又は有限）の集団より抽出された無作為標本と見做し得る事。この無作為標本をもととして母集団についての推測となる事。この現実の統計資料は確率変数の実現値と見なされること。

かゝる大前提によって確率論の適用を可能とし、母数推測のみならず、差の有意性の検定その他の複雑なる操作に漸次進んでいるのである。

かゝる基本的な思考形式は、これを一つの数学的模型と考ふる限りにおいては、数学自体としては当然認容すべき性質のものであろう。然しここで問題となるは、かゝる数学的模型によって実際の統計資料を扱わんとする際に、果してその条件が常に満たされるや否やの点にある。

この為には、母集団は必ず一定数値をもつ母数を持たねばならぬ。又一定の分布函数を有するものでなければならぬ。この条件を欠いてはかゝる数学的模型は適用せられないのである事は何人にも明であらう。

今実際の統計資料が、一乃至いくつかの標識によって集められたる数値の集団である事は云うまでもないが、この同じ標識を持つと云う無限の集合とは何を意味するか、時間空間を超越したる無限の集合を考ふる事は、想念としては可能であっても、これは統計学上の母集団、即ち一定の分布型や母数を有する集合として存在する事は不可能である。これは近似的にも成立しない。例えば身長とか体重の如き日常の計測量を実例として考えてみるも、人間の身長、哺乳動物の体重などと云う母集団は、推測統計の対象としては存在し得ないのは明かであらう。

然らば時空を超越した無限の集合とは何か、これは単なる概念に過ぎない。一般の概念が一定の算術平均や分散を持たぬ事は当然で、統計資料の属性より想念を拡充して概念化する事は可能であるが、母集団は作り得ない。推計学者は概念と母集団の区別を混同しているのである。

時間空間を規定せずば、有限個の母集団も考ふる事はできない。なぜならば、標本における標識以外の属性が、実際統計においては全然同一たる事は有り得ないから、異なる属性の配分を決定する方法はないからである。これは母集団の母数が不定であって一つの値を取り得ない場合といえるであらう。

時間空間を規定すれば、初めて有限の母集団として、母数や分布型の定った集団を求め得た事となる。（然しながらこの場合猶現在手にするこの部分に属する実際の統計資料が、この母集団よりの無作為抽出標本なりと必ずしも無条件には見做せないと云う別の問題が生ずるのであるが）

以上の事を要約すれば、実際の統計資料は常に母集団よりの無作為標本と見做し得るものではなく、一つの統計資料には、母集団の存在なきもの乃至不定なる場合多きを通常とし、存在する場合あるも、これは時間空間の規定ある特定の有限母集団に過ぎないと、かゝる結論になるのである。

この事は一面当然のことであって、母集団は数学的思惟形式の一つである純抽象的模型における観念であって、この言葉はその実在を意味するものでもなければこれを証明するものでもない。実在にこれと近似的に対応せしめ得るものありや否やは、統計学以前の知識がこれを吟味し判定するのであるから。

それでは今この特定有限の母集団とその意義について考察を進めよう。

一つの実際統計が成立する為には、時間と場所は不可欠の要素であるから、吾々は今一定の場所の一定の時間に於ける有限の統計量を考えよう。この場合当然算術平均や分散は一つの定値として計出される。今これを母集団とする場合、無作為標本との関係は如何であろう。

今この母集団が、正規分布型であるならば、これの無作為標本における母数（こゝでは算術平均又は比率を考えよ）の偏異は、確率変数として確率的計算は可能である。勿論実際統計のしかも時空に制限あるものに、数学的厳密な正規分布型の存在し得る筈もないが、今こゝではその近似を許し、又無作為標本の数を多くとると云う想定を許す事により（中心極限定理の成立を条件とし）計算を可能とするのであるが。

以上の如き場合に限り、吾々がこの無作為標本の一として、合理的な方法によってこの部分量を実際統計として手にするならば、この統計即ち標本より、この特定の母集団の母数を推測し得る事となるのである。

故に母数の推測といわるものは、任意の実際統計資料を常に無作為標本として、母集団の母数推測を行い得るものでは決してなく、たゞ時空を規定したる有限母集団の、しかも分布型式に一定の制限あるものにおいて、相当数の無作為標本をとりたる場合、その母数の標本に於ける確率分布の計算可能なるを前提とし、その逆を行うものに過ぎないのであって、この逆を無制限に不当に拡充し得るものではないのである。

又この為には勿論標本はなるべく多数なるを要し、又予め母集団の分布型についても相当の知識を有する必要がある事は理解せらるゝであろう。

任意の実際の統計資料より算術平均を算出し、これに標準偏差の不偏推定量の何倍かを加減したる範囲を、何%の危険率における無限母集団の母数の推定なりと、無

雑作に考うるが如きは、全く無意味の数学公式計算に過ぎず、又母集団の分布型も不明なるを顧みず、小数の標本をもってするが如きも、甚だ危険にして不合理な操作といわざるを得ないのである。

以上に説いた如く、実際の統計資料より母集団の母数推測の可能なる場合は、特定の時空規定のある有限母集団に限らるゝとすれば、その実用価値やその意義は甚だ少きものとならざるを得ない。なぜならばかゝるものが実用上の要求を満たす為には、時間空間を異にしたる無数の同種のものが必要とするからである。

推計学者の考うる如き客観妥当性ある一般的数値を得るは、この推測にては成し得ぬ事となる。

吾々が一般的に用うる、或る集団に関する定値なるものは、推測統計法以外の知識や方法によって得られるものなのである。この点の詳説は略するが。

今吾々は、任意の標本あって其母集団を考へるのでなく、特定の母集団あって初めて標本あるを考へる事を明にしたが、それではできうる限り一般的普遍的なる値をうる為に、場所乃至時間をより広くとる事とし、この範囲の同一標識による統計量を母集団として考え、この無作為標本よりの推測によって、より普遍的の有用なる値をうる事が可能なりやの問題を更らに検討しよう。世論調査などもこの類に属すると云えよう。

時間と空間を広く延長するという事は、標識以外の属性において、多くの異質を含むに至るといふ弊害を免れ難い。これをその仮母集団とする事は、却って無意味な場合が多い。この為には層別化の問題が生ずるのである。

層別化そのものは推測統計法の問題ではなく、他の知識によってしか為し得ないものであり、しかもこの程度を如何に止むべきやは簡単に定め得べきものでない。

推測統計学に層別化の問題の生ずる事は、現実の統計資料が無限の母集団の無作為標本なりとの想定を、自ら破壊する反証として皮肉なるものといわざるを得ぬ。又他面、母集団を想定する場合に時間空間の延長や標識の単純化が、多くは無意味なるを示すに外ならないのである。

以上は推測結果の一般化なる問題が、単なる時間空間の延長なる操作によって解決せられないを明にしたのであるが、更らに母集団の確率変数たりやの問題がある。

無作為標本が確率変数の実現値なりとの想定の下に推測が行わるゝならば、母集団にその確率変数を求めねばならぬ。この検討は是非必要である。

一定の時間空間の拡がりやを有する一定の標識を共有する統計量の集合として、例えば医学生物学又は心理学的母集団の何かを想定した場合、この母集団のその時に於

ける分布が確率変数と見做しうるや否やは考える迄もなく明であろう。

確率変数たる為には、その個数を限りなく増加せしむる時一定値に確率収斂する事を要求せられる。今実際の母集団に於て、個数を限なく増加するには、時間空間を広くとってゆくより方法はない。この場合一定値に収斂すると確実に理論づけらるゝ事も、経験的に認めらるゝことも殆んどないのであって、現実の母集団に求めらるゝ数値は、厳密に云えば、皆これ時間空間に規定せられた記録的数値であって、数学的の確率変数ではないのである。それ故に無作為標本は確率変数の実現値と考え得ないのも当然であり、確率変数でないものの推測より一般的な数学的規則性を求める事は、多くは無意味と云うの外はない。

母集団の母数についても同様のことを云い得る。無限母集団の算術平均や比率も一定値でなければならぬ。この事は数を多くとればとるほどこの一定値に収斂する事を意味する。吾々が現実の時空を制限して有限母集団を想定した際、勿論一つの定った値を持つ算術平均や比率の存在する事は確実であるが、これが個数の増加によって絶えず一定値に近づくが如き実例は考え得ないのみならず、

その一定値の存在するやも不明であって、各時間空間において各異なる値をとるとか、永久に変動する不定の値なりとの認識だけに止めるが妥当と考えられる。これ等の数値よりいかなる概括をすべきや、いかなる通則を認むべきやは、これ全く推測統計法とは別の問題であって、他の知識経験を必要とするものである。

母数の算術平均を、測定法の真の値に類推し、母集団の分布を、偶然誤差や定誤差の観念を拡充してこれに当てはめて考える事は、簡単には認容されない。

この問題は又統計資料に対する確率論の適用可否とその限度の問題にまで発展する。又これは一般事象の数量化の可否とその意義の問題を究明するの必要を生ずる。

これ等の問題は更らに稿を改めて論ずる事とし、本序説は推測統計法の根本要素である母集団と標本との関係だけに止めたが、これにのみよっても、所謂母集団推測なるものが、いかに実用に意味の少き限定されたものかが解し得られようと思う。又この基礎の上に立つ、差の有意性の検定や分散分析なども、いかに実用上価値少なきものなるやも判断できようと思ふ。勿論この事も更らに詳細に述べる予定であるが。

(1958年6月30日 受理)