任意標本調査の母集団

関　弥三郎

は し が き

任意標本調査は第2次大戦後わが国に導入されて以来急速に普及し，官庁統計を始めとして広く社会統計調査に用いられるようになり，今日では多くの統計が任意標本調査によって作成されている。このような任意標本調査の普及は，標本調査一般のメリット（実地調査の費用，労力と集計の時間の節約）に，無作為抽出法のもつ合理性（標本構成の簡便さ，標本誤差の計算）が加わったためだけではなく，より重要なことは，標本があまり小さくない限り無作為抽出が全数調査によく近似する結果をもたらす理論的根拠をもっていることによるのである。とはいっても，任意標本調査の実際においては，このような無作為抽出法の効果や合理性を相殺するマイナスの要因が多数あり，無作為抽出によったからといって直ちに標本誤差が少ないとみなすことはできないから，無条件に任意標本調査を信用することは危険である。したがって，統計利用者にとっては任意標本調査が社会統計調査においてもつ長所と短所，限界を十分わきまえていることが，統計の正しい利用を期する上から非常に大切である。そのためには，任意標本調査の有限母集団=任意標本の数理的構造を知るだけではなく，それの社会統計調査の代用法としての問題点をよく理解していることが必要である。統計調査は多くの場合，特定の時間と場所において客観的に存在する社会集団現象の大きさと構成を記述するために行われ，社会集団現象を正確に認識するためには，それの構成要素をもれなく調べることが必要である。したが

って，統計調査は全数調査を原則とするのであるが，全数調査は多くの場合困難ないしは不可能であるので，一部分の要素のみを調べて全数調査の結果を推定せんとする標本調査が発達せしめられたのであり，そこへ無作為抽出法を採り入れたのが任意標本調査であるから，当然任意標本の数理以上の問題が存在する。したがって，任意標本調査は統計調査の代用法として体系付けることが必要であるが，一般にはその数理的構造の研究に重点が置かれており，したがって先に述べた統計利用者の要請には十分こたえ得ないことは明らかである。

次に注意しなければならないのは，社会統計では客観的に存在する社会集団現象の数的記述の場合と，偶然変動する社会現象を集団的に観察することによって一般性，規則性を見いださんとする場合とがあり，任意標本調査は前者の場合に適用されるということである。後者の場合，偶然的現象の集団観察の結果から一般性，規則性をあらわす値を知るのに，無作為抽出を仮定して無限母集団=任意標本の 確率論的図式を適用するのであるが， この図式は推測統計理論 (statistical inference または推計学 stochastics) として発達してきたものである。ここにいう任意標本調査と推測統計理論はいずれも，任意標本から母集団を推定する方法として統一的に論ずることができるのであり，数理統計学ではそうして説明されるのであるが，それぞれが適用される状況と条件が違うために別個のものとして考えることが，これらの方法の誤用，濫用を避ける上から必要であることを強調したい。第2次大戦後の任意標本調査の導入，推測統計理論の社会統計への適用に伴って，それに対する疑問，反対の意見が社会統計学者から提起され，激しい統計論争が展開されたのであるが，その中には統計調査の代用法としての任意標本調査と推測統計理論の違いを明確にするならば解消し得る問題点が多くあったと考える。

以上のように任意標本調査を統計調査の代用法として体系化し，またそれと推測統計理論の違いを明らかにするには，いろいろな方向からの考察が必要であろうが，その一つの道は母集団概念の考察であると思う。それは，母集団概念こそは社会統計と母集団=任意標本 の確率論的図式の接点であり，以上の問

題を明らかにする基礎となると考えられるからである。ところで，最近標本調査論争の問題点の整理と検討を行った木下滋氏の論文[1]において，任意標本調査の場合の母集団について述べられているところは，上述の問題意識からみるときは理解し難い点を含んでいる。そこで便宜上，木下氏の母集団概念を手掛かりとして任意標本調査の母集団の意義を考察し，それを推測統計理論の母集団と比較して両者の違いを明らかにせんとするのが本稿の目的である。

まず，木下氏の任意標本調査の場合の母集団概念の理解をみると，次のようである。すなわち，「母集団というものは，常にそこから確率的に実現した，あるいは，確率的に抽出された標本との関係でのみ意味をもつものであり，しかも，フィッシャーのいうように，統計家の想像の産物であったし，原理的には構成要素は無限とされていた。」として，母集団を調査対象として定義されたものすべての集りとする，津村善郎氏始め一般の母集団の定義に反対する。そして，吉田忠氏の『統計的方法の基礎』(1970年）において述べられた母集団概念に依拠して，「任意抽出の無限の繰返しによる個々の標本の標識の無限系列を想定して，これを母集団と呼ぶのが標本理論の論理にしたがった理解であり，こうしてはじめて確率変数たる母集団も定義され，母集団と標本の関係がもたらされる」とする[2]。したがって，木下氏によれば任意標本調査の場合でも有限母集団という考え方は誤りであり，仮説無限母集団を考えねばならないことになる。

以上の木下氏の母集団概念の規定がもつ問題点は二つある。第１は，母集団を確率的な概念としてのみとらえ，客観的存在である調査対象の集りの意味の母集団を認めない点であり，第２は，確率変数たる母集団に関して有限母集団概念を否定して，仮想の無限母集団概念を任意標本調査にも持ち込む点である。次に，これらの問題を順次検討していこう。

1）木下滋「標本調査法の諸問題—標本調査法における母集団と標本の関係—」（京都大学『経済論叢』第116巻第３・４号，昭和50年）

2）同上，80—81ページ。

1　調査対象の集りとしての母集団

1）まず，木下氏が調査対象の集りとしての母集団概念を否定する点であるが，むしろ，理論的にも実際においても任意標本調査の場合の母集団は，それから標本が任意抽出される有限個の，調査対象として定義されたものすべての集りの意味に用いられるのが普通である。例えば，総理府統計局の家計調査や就業構造基本調査，労働力調査ではこの意味の母集団が用いられている1)。

そして，理論家による母集団の定義で注目すべきものをみると，デミングは Some Theory of Sampling (1950) において，母集団を調査単位 (ultimate unit) の集団としている。すなわち，「実際に統計的情報を必要とするのは，個人，世帯，農家，事務所，地域，工業製品等についてであって，これらのものを母集団の最終単位（調査単位）と呼ぶことにする。」2) そして，母集団を実践に適した形で定義し，また，調査単位より成る抽出単位の適切なリスト（枠）を用意することの重要性を説明している3)。しかし，他方では「数値的には母集団は N 個の有限数の集合，すなわち（ある特性に関する一引用者）N 個の計量 a_1，a_2，…，a_N から成立している。」とも述べている4)。したがって，デミン

1）家計調査では，調査対象は世帯であるが不適格世帯を除くことを説明した後，「そこで，母集団は正式には前述の不適格世帯を除外した全国の消費者世帯となるが，…」と述べており（総理府統計局『家計調査年報』昭和48年，388ページ），就業構造基本調査では，「この調査における母集団の範囲（結果数字の範囲）は，調査期日（昭和49年7月1日）現在において我が国の行政権の及ぶ地域内に常住する世帯及び15才以上の人口である。」と述べている（総理府統計局『就業構造基本調査報告』昭和49年，375ページ）。労働力調査では，邦語の説明には母集団の語は使われていないが，英文の説明の中で “Population coverage” の見出しの下に，“The universe of the survey is composed of all persons 15 years and over usually residing in the country, …” と述べている（総理府統計局『労働力調査報告』昭和50年年報，186ページ）。

2）Deming, W. E., Some Theory of Sampling, 1950, p. 77. 斉藤金一郎訳『標本調査の理論』昭和28年，81ページ。

3）Op. cit., pp. 31—33, 77—83. 邦訳書，35—37，81—87ページ。

グは母集団を多くの集団特性について調査される調査対象の集合とするとともに，また，一方向の集団特性について調査した結果である標識の集合（ただし，この段階ではまだ確率変数の意味は与えられていない）ともしているのであるが，前者は標本調査の実施上必要な抽出単位の決定の基礎となるものであるのに対して，後者は標本調査の結果の精度（標本誤差の程度）を確率論によって確かめるために前者を数学的に抽象したものにすぎないから，前者の方を基礎的なものとしなければならない。また，コクランは Sampling Techniques (2nd ed., 1963) において，「母集団という言葉は，そこからサンプルが選ばれる集合全体を表わすのに使われる。」と述べている。しかし，これだけではその集合の内容が調査対象なのか標識なのかは明瞭でないが，それに続いて農場の母集団からサンプリングするときには，調査員が現地で母集団に属するかどうかをためらうことなく決め得るように，農場を定義する必要があると説明していることから，母集団を調査対象の集合と考えていることは明らかである[5]。そして，母集団をそれについて情報が要求される母集団（目標母集団 target population）とサンプルのとられる母集団（抽出母集団 sampled population）とに分けて，両者は一致させるべきであるが「実行可能性や便宜性のため，抽出母集団は目標母集団よりも限定されて決められている。」場合があり，サンプルからひき出される結論は抽出母集団に適用するものであって，それがどの程度まで目標母集団にもあてはまるかということは，他のいろいろな情報によって判断しなければならない，と注意している[6]。このような目標母集団と抽出母集団の違いは，調査対象の抽出や実地調査の技術的可能性によって生ずるのであるから，この点からも母集団は調査対象の集りの意味で使われていることがわかる。

以上，デミングとコクランの母集団の規定をみたのであるが，それから知り得る重要な事柄は，調査対象の集りとしての母集団は，得られた抽出単位のリ

4）Op. cit., p. 85. 邦訳書，90ページ。

5）6）Cochran, W. G., Sampling Techniques, 2nd ed., 1963. 鈴木，高橋，脇本共訳『サンプリングの理論と方法』1，1972年，6ページ。

ストが妥当なものであるか否かを検討し，更に，調査せんとする集団と実際に標本を抽出した集団との間のギャップの有無を明らかにし，任意標本調査による統計の正確性を吟味する手掛かりを得るために必要であるということである。抽出単位のリストや実際に標本を抽出した集団の適切さを考える場合の基礎となるものは，調査目的から決められたすべての集団特性（個々の対象については標識として現われる）の規定を受けた調査対象の集団でなければならず，確率変動する数値の集りの意味での母集団は，特定の集団特性のみについて規定されたものであるから，この任に耐え得ないことは明らかであろう。このように，調査対象の集りとしての母集団は任意標本調査の実践上重要な役割を果たすのであるから，この意味の母集団を認めない立場には賛成し難い。

2）任意標本調査は，特定の時間と場所において客観的に存在する社会集団現象の記述的な統計調査の代用法として用いられるのであるが，統計調査の調査対象すべての集りは統計集団として規定されるので，次に，統計集団と母集団との関係を明確にすることが必要である。

客観的存在たる社会集団現象の統計調査の際には，調査せんとする社会集団現象に社会科学の理論や調査目的等から理論的，技術的限定を加えて，集団現象の要素の個別観察が可能なように規定し直した統計集団が構成されねばならず，この統計集団を数的に記述したものが統計である。客観的に存在する社会集団現象と統計集団との間には，このような理論的，技術的規定が加えられているために大なり小なりギャップが生じ，豊富な諸属性の総合としての社会集団現象が，限られた数種の集団性についてのみ規定された統計集団としてとらえられており，ここに社会の統計的認識の限界が存在する。そして，調査者の理論的立場や調査目的からする社会集団現象の理論的，技術的限定が，統計利用者の理論的立場や利用目的からみる時そのまま妥当するとは限らないのであって，ここに統計の信頼性が問題になる根拠がある。次に，統計集団の要素（調査対象）の実地調査，その結果の整理を経て統計が作られるのであるが，これらの過程で種々の理由により調査誤差が発生することから，統計の正確性の

問題が生ずる。そして，この実地調査，整理の過程で全数調査のもつ難点を回避するために任意標本調査が利用されるのである。

したがって，任意標本調査の出発点となるのは統計集団であって，統計集団を任意標本調査の立場から，それから標本を抽出する元の集団としてとらえ直したものが母集団であり[7]，コクランのいう目標母集団はこれに相当することになる。故に，母集団は単なる調査対象の集りではなく，統計集団としての規定を受けた調査対象の集団である。そして，これを数学的に抽象したものがデミングのN個の計量の集りとしての母集団であり，それは統計集団において規定された集団性の方向ごとに考えられることになる。この母集団の計量は一般的には歴史的な社会的事実を表わすものであって，確率変数とみなすことはできない。そして，たとえ社会集団現象が偶然変動する場合であっても，それを確率論の図式を適用し得るような確率変動とみなし得る場合は，きわめて少ないと考えられる。なお，任意標本調査による統計は社会の統計的認識の限界，統計の信頼性の問題を免れることはできず，更に標本の実地調査，整理の誤差をもつのであって，単に標本誤差の問題だけを考慮すればよいというものではなく，それらの問題を考察する基礎は母集団を統計集団としてとらえることによって与えられることがわかる。

このように母集団は統計集団そのものであるとすると，改めて母集団と言い替えずに統計集団のまま用いることが考えられる。しかし，統計集団は社会集団現象の統計調査を可能ならしめるために構成された概念であって，任意標本調査の立場からみるときに始めて標本を生み出す元の集団となるのであり，更に，それを抽出単位の集団に変換することが必要になり，この抽出単位の決定は主として標本抽出の技術的立場からなされるのであるから，任意標本調査の基本概念であることを明確にするために，母集団と言い替える方がよいと考える。

7）津村善郎『標本調査法』1956年，8ページ。森田優三『新統計概論』昭和49年，214ページ。

3）母集団から標本を抽出するためには抽出単位を定め，それのリストを作らなければならない。すなわち，母集団を抽出単位の集団に編成し直すことが必要であり，この抽出単位の集団が表わす母集団が実際に標本を取り出された母集団，すなわち，コクランの抽出母集団である。母集団を抽出単位の集団に直す場合（調査対象がそのまま抽出単位になり得るときもあるが，複数の調査対象を含む抽出単位を設定しなければならないことも多い）どの調査対象もどれか一つの抽出単位に含まれており，そして，唯一つの抽出単位に属していることが必要である[8)]。ところが，実際には抽出単位の集団が表わす抽出母集団と目標母集団との間に相違が生ずる場合がある。それは，抽出単位のリストの作成が容易でないので，他の目的で作成されたリストをそれに転用するために生ずることが多い。例えば，世論調査の場合は世論を構成する人々の集団が目標母集団であるが，有権者名簿をリストに用いて標本を抽出するために，抽出母集団は未成年者で世論の構成に参加し得る年令層を欠くことになり，調査の内容によっては無視し得なり影響を調査結果に及ぼすであろう。また，労働省の毎月勤労統計調査では，事業所統計調査による事業所リスト（抽出母集団）から標本事業所を抽出し，それを次の事業所統計調査まで3年間固定して調査を継続するのであるが，抽出後の経済の発展に伴う事業所集団（目標母集団）の変化のために抽出母集団と目標母集団の不一致が生じ，標本調査の結果に偏りを生ぜしめるのである[9)]。

8）Deming, op. cit., p. 77. 邦訳書，81ページ。

9）毎月勤労統計調査の全国甲調査では標本事業所を3年間固定し，新しい事業所統計調査の結果が得られるとそれに基いて抽出替えを行うという方法をとっているため，その間における規模30人以上の事業所の新設や30人未満から30人以上への事業所規模の上昇がとらえられないことになり，そのため30〜40人程度の標本事業所が相対的に少なくなり，その結果実際よりも労働者数は過少に，平均給与額は過大に推計される傾向が現われた。そして，この偏りは抽出替え直後は最も小さく，その後次第に大きくなるのである。そこで，この欠点を緩和するために6カ月ごとに標本事業所の補充を行っているのであるが，完全ではないので，3年ごとに行われる標本の抽出替えの際に結果数字にギャップが生ずることになる。常用雇用指数，労働時間指数，常用賃金

また，計画的に抽出母集団を目標母集団と違える場合もある。例えば，デミングは地域抽出法の場合，「かりに標本地域として選ばれた場合に極度の困難が予想されるような地域はあらかじめ母集団から除外しておくがよい。」として，「人口のまばらな地域で，それらの地域の人口を合計したものが母集団で定義している人口の2，3％に充たないことがわかっているような地域をすべて除外しておく方が，特にそれらの地域を調べる費用が非常に嵩むと予想されるならば，有利である。」と説明している[10)]。また，全国に散在しておりその把握が困難な多数の小規模業者を含む業界（目標母集団）を調査するような場合には，一定規模以上の業者ないしは同業者団体に加盟している業者に限定して抽出母集団とすることがある。更に，標本を得やすいことや費用の点等から，大学年令の全国の青年（目標母集団）に妥当する結果を得るために，特定大学の学生のみ（抽出母集団）から標本を選ぶ場合もこれに該当する[11)]。

このようにして目標母集団と抽出母集団との不一致が生ずるのであるが，標本調査の結果は当然抽出母集団に妥当するにもかかわらず，目標母集団を表わすものとして利用されるのであるから，このことは統計に重大な誤差を生ぜし

指数については，このギャップを補正して指数の時系列的連続性を保つために，3年前にさかのぼって指数の修正を行っているのであるが，毎勤の結果数字そのものはこのような修正を施さないためにギャップが残されていることに注意しなければならぬ（労働省統計情報部編『毎月勤労統計要覧』昭和50年，6，11―14ページ)。

10) Deming, op. cit., pp. 31―32, 邦訳書，35―36ページ。

11) Snedecor, G. W. and Cochran, W. G., Statistical Methods, 6th ed., 1967. 畑村，奥野，津村共訳『統計的方法』1972年，14―15ページ。なお，本書では目標母集団と抽出母集団が異なる場合として，人間の母集団の場合は標本として抽出された人の中には住所不明，病気，回答拒否等の者がいるから，統計的推論を行わんとする母集団は，標本のなかに抽出されたならば回答を与えるという人達の集りである，とみなさねばならず，実際に標本のとり出される母集団はもとの母集団より小さくなる，と述べているが（邦訳書，14ページ），このような実地調査の不可能な場合は全数調査（統計調査）でも起るのであって調査誤差の一つとされており，任意標本調査に特有のものでないから，目標母集団と抽出母集団の不一致の理由とするのは不適当であると考える。目標母集団と抽出母集団のズレは，前者を抽出単位の集団に変換することから生ずる場合に限るべきである。

めるおそれがある。この誤差は任意標本調査に特有の誤差であるが，確率論的な標本誤差とは別の性質のものであり，これを抽出母集団の誤差と呼ぶことにする。

4）統計調査の代用法としての任意標本調査では，単なる母数の推定以上に標本抽出という問題があるから，その基本概念として調査対象の集りとしての母集団が必要になるのであった。ところが，時間と場所を越えて妥当する一般的な値の推測の問題を取り扱う推測統計理論の場合は，調査対象は，例えば，稲のある品種の反当り収量の実験のときは栽培実験であり，また，出生性比の場合は出生児であるが，これらの対象は同じ条件の下で得られた無限のものが仮定されなければならないから，任意標本調査の場合のような，具体的に存在する全体の中から偶然的に抜き出すという意味での標本抽出は不可能であり，実際に得られた栽培実験や出生児が無限のものから偶然的に取り出されたと仮定するしか仕様がないのであるから，標本抽出の問題は存在しないのが普通である[12)]。あるのは偶然変動する結果を与える実験のやり方とか，現実界において生起する事象の偶然性の検討の問題であるが，それは調査対象全体の中から標本を選ぶのとは別の次元の問題である。したがって，推測統計理論の場合は調査対象の集りとしての母集団は考え得るにしても実践的意味がないから無用であり，確率変数たる母集団だけで十分である。

2　確率変数たる母集団

1）次に，確率変数たる母集団に関して有限母集団を否定して仮想の無限母集団とする点を考察しょう。その場合，木下氏は吉田氏の考えによっているので，問題点を明らかにするために吉田氏の任意標本調査の場合の母集団の規定

12）例えば，ある病気に対するある薬の効果を調べる場合，現存する N 人の患者の中から無作為に n 人を抽出して治療することによって，標本が重症者または軽症者に片よることなく薬の効果を正確に評価し得ると考えられるのであるが，それでも無限の患者からの無作為抽出としてはやはり仮定にすぎないであろう。

をみると次のようである。すなわち，客観的に存在する社会集団から統計集団が抽象された場合，「統計集団に対して無限にくり返され，かつランダムな結果を生み出す操作=無作為抽出が加えられて始めて，母集団分布が構成される。そこで抽出方法に応じていろいろな母集団分布が与えられる。」まず，統計集団から1個の単位を無作為に抜き出しその標識をみることを無限に繰り返すと，1次元確率変数 X がとる値の無限系列が得られ，それを X のとり得る値について整理すると，統計集団の構成に等しい確率分布をもつ無限母集団が構成される。そして，一般に統計集団から n 個の単位を引き続き無作為に抜き出してその標識をみるという非復元抽出を無限に繰り返すとき，n 次元確率変数 $(X_1,\ X_2,\ \cdots,\ X_n)$ がとる値の無限系列が得られ，それを $(X_1,\ X_2,\ \cdots,\ X_n)$ がとり得る値について整理するとき得られる確率分布をもつ n 次元母集団が構成される[1)]。

要するに，吉田氏によれば N 個の単位の統計集団から無限回の非復元無作為抽出を繰り返すときに現われる，無限の確率変動する数値の集りが母集団であり，無限回の繰り返しは単なる仮定であるから，母集団は仮想の無限母集団となるのである。しかし，これに対して母集団の確率分布を規定するためには必ずしも無限回抽出を仮定する必要はなく，組合せ理論を用いて N 個の単位から n 個の単位を取り出す場合の理論的に可能な取り出し方の数を考えることによっても，母集団確率分布を得ることができるのではなかろうか，という疑問が生ずる。次に，この点を吟味してみよう。

その前に，二つの確率の定義の仕方を知らねばならない。まず第1は古典的な確率の定義であって，ある試行の結果として現われ得る場合が全部で n 個あり，そのいずれか一つは必ず起り二つ以上同時に現われることは決してなく，そして，いずれが現われるかは同様に確からしいとすると，事象Aが r 個の場合に現われるならば，事象Aが起る確率を $\frac{r}{n}$ とするのである[2)]。しかし，この

1）吉田忠『統計的方法の基礎』1970年，101—103ページ。
2）河田竜夫『確率と統計』昭和26年，3ページ。

ような確率の定義は，試行の結果起り得る場合の数を先験的に知ることができるときでなければ適用し得ず，また，どの場合の出現も同様に確からしいとの前提が容易に承認し得ないから，実際問題に用いることはほとんど不可能である。そこで，一般に用いられるのが頻度確率の定義である。すなわち，同一条件の下で試行または実験，観察を n 回繰り返し，その結果のうちで事象Aが r 回起きたとすると，事象Aの相対度数 $\frac{r}{n}$ は試行回数 n が大きくなるにしたがってある一定の値に近づく傾向を示すことが経験的に確認された場合，n を無限に大きくしたときに相対度数 $\frac{r}{n}$ がとると想像される値を事象Aの確率とするのである[3)]。

さて，古典的な確率の定義はほとんどの場合実用し得ないのであるが，偶然遊戯の領域では完全に適用することができる。今，つぼの中に同じ大きさ，重さ，形の球が10個入っており，そのうち7個が赤球で3個が白球であるとすると，つぼの中の球をよく混ぜ合わせてから1個取り出すとき，それが赤球である確率は $\frac{7}{10}$ であるといえるであろう。ここでは，試行の結果起り得る場合が10個で，そのうち7個が赤球であり，かつ，どの球も取り出される可能性が等しいと考えられるからである。もし，与えられた球が物理的に均一でなく手触りで赤，白を識別し得るおそれがあるときは，10個の球に番号をつけ乱数表を使って取り出す球を決めるならば，10個の球の抽出の可能性は等しいという条件を満たすことができるであろう。

任意標本調査は丁度このつぼの中から球を取り出す例と同じであって，したがって，古典的な確率の定義をそのまま適用することができる。今，N 個の単位よりなる統計集団から，N 個の単位に通し番号をつけ乱数表で1個の番号を決めて，その番号の単位を取り出すというやり方で1個の単位を抽出するとき，起り得る場合は N 個であり，かつ，どの単位も抽出される可能性は相等しいと考えられるから，古典的な確率の定義によって，任意の1個の単位が抽

3）宮沢光一『近代数理統計学通論』昭和29年，4－5ページ。

出される確率は$\frac{1}{N}$といえる。そして，統計集団において特定の標識A_i（$i=1$，2，…，k）をもつ単位をN_i個とすると，その標識の単位の抽出確率は$\frac{N_i}{N}$とすることができる。したがって，この場合頻度確率の定義によって，統計集団のN個の単位から1個の単位を無作為抽出することを無限に繰り返すと仮定して，特定の標識の単位の相対度数の極限値をもって，その標識の単位の抽出確率とする必要はないであろう。

このようにして，すべての単位の抽出される確率が相等しいというやり方（無作為抽出法）で統計集団から1個の単位を抽出するという抽出操作を媒介として，統計集団のN個の単位に等しい確率$\frac{1}{N}$が与えられ，それぞれの単位の標識を表わすXは確率変数となり，無作為抽出された単位の標識を調べることはこの確率変数の実現値をみることであると解することができる。こうして$\frac{1}{N}$という同じ確率分布をもつ――これを標識について整理するときは統計集団の構成に等しい確率分布に従う――1次元確率変数Xが規定され，確率変数たる母集団が構成されるのである。ここでは，有限のN個の単位――確率論的な表現を借りれば標識空間（または標本空間）のN個の点――に確率$\frac{1}{N}$が与えられており，この意味において母集団は有限母集団といえる。

2）以上は1次元確率変数の場合であるが，一般にn次元確率変数の場合も同様に，古典的な確率の定義によって確率分布を求めることができる[4]。すなわち，統計集団のN個の単位からn個の単位を非復元抽出するときに得られるn個の単位の組の数は，N個の単位からn個の単位を取り出して並べる場合の並べ方の数

4）統計集団のN個の単位からn個の単位を無作為抽出して標本を構成する場合，復元抽出と非復元抽出の二つの方法があるが，任意標本調査の実際では非復元抽出のやり方が用いられ，復元抽出によることはほとんどない。それは非復元抽出による方が標本推定値の分散が小さく精度が高くなるからであって，例えば，標本平均$\bar{x}$の分散$\sigma_{\bar{x}}{}^2$は非復元抽出の場合は$\frac{N-n}{N-1}\frac{\sigma^2}{n}$，復元抽出の場合は$\frac{\sigma^2}{n}$であり，前者の方が$\frac{N-n}{N-1}$だけ小さい。したがって，ここでは非復元抽出の場合の母集団のみを考察する。

$$N(N-1)\cdots\cdots(N-n+1)={}_NP_n\equiv(N)_n \qquad (1)$$

として求められる。したがって，統計集団から非復元のやり方で取った n 個の単位の組を無作為抽出するとき，起り得る場合は $(N)_n$ 個であり，かつ，どの組も抽出される可能性は相等しいと考えられるから，任意の1個の組が抽出される確率は $\frac{1}{(N)_n}$ といえる[5)]。

この場合，$(N)_n$ 個の組を実際に作り，それに一連番号をつけて乱数表を用いて1個の組を抽出するという意味での無作為抽出は不可能であるが，N 個の単位から n 個の単位を，どの抽出段階においても無作為に非復元抽出することによって[6)]，それが実現されるのである。すなわち，非復元抽出であるから各段階での抽出結果はそれより先の抽出結果に依存し，したがって独立でなく，また，どの抽出段階においても無作為抽出が行われる（すなわち，どの単位も同一の抽出確率をもつ）のであるから，n 個の単位の組の抽出確率は確率の乗法定理によって求めることができる。まず1番目の単位は，N 個の中から無作為に抽出されるのであるからその確率は $\frac{1}{N}$ であり，2番目の単位は，残りの（$N-1$）個の中から無作為に抽出されるのであるからその確率は $\frac{1}{N-1}$ である。以下同様にして，最後の n 番目の単位は，残りの（$N-n+1$）個の中から無作為に抽出されるのであるからその確率は $\frac{1}{N-n+1}$ である。したがって，n 個の単位の組の抽出確率はこれらの確率の積

$$\frac{1}{N}\ \frac{1}{N-1}\cdots\cdots\frac{1}{N-n+1}=\frac{1}{(N)_n} \qquad (2)$$

であって，先に可能な n 個の単位の組の数から求めた確率と一致する。故に，

5）n 個の単位の並び方の順序は問題にしないで n 個の単位の組合せだけをみるときは，その数は $\frac{{}_NP_n}{n!}={}_NC_n\equiv\binom{N}{n}$ であるから，任意の n 個の単位の組が抽出される確率は $\frac{1}{\binom{N}{n}}$ となる。

6）実際には乱数表を用いて，同じ番号が繰り返し現われた場合にはそれをとばすことによって n 個の番号を決め，それらの番号の単位を次々に抽出していくことによって行われる。

n個の単位を非復元無作為抽出することによって，n個の単位の組を無作為抽出することができるのである。

次に，統計集団のN個の単位のうちN_1個が標識Aをもち，残りN_2（$=N-N_1$）個が標識Bをもつ場合，それから非復元無作為抽出したn個の単位のうち，標識Aの単位がr個（したがって，標識Bの単位が$n-r$個）ある組が現われる確率は次のようである。すなわち，n個の単位のうち標識Aの単位がr個ある組は全部で$\binom{n}{r}(N_1)_r(N_2)_{n-r}$あり[7]，他方，$n$個の単位の組は$(N)_n$あって，そのいずれも抽出の可能性は相等しいのであるから，古典的な確率の定義より，n個の単位のうち標識Aの単位がr個ある組が得られる確率は

$$\frac{\binom{n}{r}(N_1)_r\,(N_2)_{n-r}}{(N)_n}=\frac{\binom{N_1}{r}\binom{N_2}{n-r}}{\binom{N}{n}} \qquad (3)$$

となり，これを超幾何分布という[8]。そして，Nがnに比べて非常に大きいときは，超幾何分布は二項分布に近似するのである[9]。

なお，以上は統計集団がA，B二つの標識の部分集団よりなる場合であるが，一般に統計集団がk個の標識A_1，A_2，…，A_kの部分集団（それぞれの単位数をN_1，N_2，…，N_kとする）よりなるときは，n個の単位のうち標識A_1がr_1個，A_2がr_2個，…，A_kがr_k個（ただし，$r_1+r_2+\cdots+r_k=n$）ある組が得ら

7）これの証明は次のようである。すなわち，標識AをもつN_1個の中からr個取って並べるときの並べ方の数は$(N_1)_r$あり，また，標識BをもつN_2個の中から$n-r$個取り出して並べるときの並べ方の数は$(N_2)_{n-r}$だけある。そして，これらの標識Aの組，Bの組を一つずつ取って，それぞれの組の中の単位の順序を崩さないようにして並べる方法を考えると，n個並んだ箱の中からr個選んでその箱の中に標識Aの単位を元の順序のまま入れ，残り$n-r$個の箱の中に標識Bの単位を同様に元の順序のまま入れると一つの並べ方ができ，このような並べ方の数はn個の箱からr個取る方法の数$\binom{n}{r}$に等しい。したがって，n個の単位のうち標識Aの単位がr個ある組の数は全部で$\binom{n}{r}(N_1)_r(N_2)_{n-r}$である。

8）松下嘉米男『統計入門』1955年，30—31ページ。

9）同上，37—38ページ。

れる確率は

$$\frac{\binom{N_1}{r_1}\binom{N_2}{r_2}\cdots\binom{N_k}{r_k}}{\binom{N}{n}} \qquad (4)$$

であり，これは超幾何分布の拡張である[10)]。

このようにして，統計集団から非復元無作為抽出法により n 個の単位を抽出するという抽出操作を通して，n 個の単位の組のおのおのに等しい確率 $\frac{1}{(N)_n}$ が与えられ，各組の単位の標識を表わす $(X_1,\ X_2,\ \cdots,\ X_n)$ は確率変数となり，こうして $\frac{1}{(N)_n}$ という同一の確率分布をもつ——これを標識について整理するときは 超幾何分布ないしは それの 拡張に 従う—— n 次元確率変数 $(X_1,\ X_2,\ \cdots,\ X_n)$ が規定され，確率変数たる母集団が構成されることになる。ここにおいても，有限の $(N)_n$ 個の n 個の単位の組に確率 $\frac{1}{(N)_n}$ が与えられており，したがって，母集団は有限母集団であるといえる。

このような母集団の確率分布を吉田氏は頻度確率の定義にしたがって規定されるために，母集団は仮想の無限母集団になったと考えられる。ところが，吉田氏による標本平均 $\overline{X}$ の期待値 $E(\overline{X})$，分散 $V(\overline{X})$ の証明の過程をみると，母集団確率分布を，N 個の単位から n 個の単位を取り出す場合の理論的に可能な組の数から，古典的な確率の定義にしたがって規定する方法が用いられており，無限の結果という仮定は 何の役割も果していないのである[11)]。したがって，吉田氏にとっても無限母集団を仮定しなければならない必然性はないと思う。

3）次に，確率変数たる母集団について，任意標本調査の場合と推測統計理論の場合との違いをみよう。推測統計理論は農地実験を基盤に発達したのであるが，一般的にいってその対象は，事象発生のメカニズムが十分に解明されておらず個々の結果は偶然変動するが，同一条件の下で実験，観察を繰り返すと

10）同上，33ページ。

11）吉田忠，前掲書，103—106ページ。

その結果はある一定の値の周りに集中し，実験，観察の回数をふやすにしたがってその値に収束する傾向を示すという性質をもった事象，すなわち，ストカスチックな事象である。しかし，この場合何回の実験，観察を行えばその結果が一定の値に収束するかは知り得ないのであるから，これを理想化して無限の実験，観察の結果が始めて一定の値に収束すると仮定するのであり，無限の結果が収束する値がその事象の確率ないしは真値とみなされるのである。ここでは，想定された偶然変動する無限の実験，観察の結果が1次元確率変数 X のとる値の無限系列であり，それらの結果を種類の異なるものについて整理して得られる確率分布をもつ無限母集団が構成されるのである。したがって，この場合の母集団は本質的に無限母集団でなければならず全くの観念の創造物であり，また，確率分布は必然的に頻度確率の定義によって規定されていることは明らかであろう。これに対して，任意標本調査の場合は $(N)_n$ 個の n 個の単位の組を実際に作るのでなく作ったと仮定するだけであるが，統計集団の N 個の単位が存在している以上理論的には作ることが可能であるから，その意味において母集団は実在の有限母集団といえる。そして，同じく確率変数たる母集団といっても，任意標本調査の場合は無作為抽出という抽出操作を媒介として構成されるのであり，それの基礎にある統計集団の N 個の単位は何ら確率的な性格をもつ必要はないのに対して，推測統計理論の場合は実際に全体から無作為抽出が行われるわけではなく，ただ無作為抽出したと仮定するだけであるから，事象そのものが確率的な変動を示すことが必要である点で異なることも，容易にわかるであろう。

そして，同一条件の下で実験，観察を n 回繰り返したときの結果を n 次元確率変数 $(X_1,\ X_2,\ \cdots,\ X_n)$ とするのであるが，同一条件の下で実験，観察を繰り返すことは，先の結果が後の結果に影響を及ぼさないことを意味するから，個々の結果 X_i $(i=1,\ 2,\ \cdots,\ n)$ は相互独立であり，いずれも同じ確率分布にしたがう1次元確率変数であることになる。このように，n 次元確率変数 $(X_1,\ X_2,\ \cdots,\ X_n)$ が同じ確率分布にしたがう相互独立な1次元確率変数 X の

n 個の組であるときは，1次元確率変数 X のみを母集団とし，n 次元確率変数（X_1, X_2, …, X_n）は同一母集団から独立に抽出された大きさ n の任意標本とするのである[12)]。（そして，n 次元確率変数—正確にはそれの関数である統計量—の確率分布を標本分布という。）ところが，任意標本調査の場合は n 個の単位を非復元抽出するために，n 次元確率変数（X_1, X_2, …, X_n）の要素 X_i は相互独立でなく，同一確率分布にしたがう1次元確率変数にならないから，n 次元確率変数（X_1, X_2, …, X_n）は同一母集団から独立に抽出された大きさ n の任意標本といえないことになる。そして，この n 個の単位の組が同じ確率分布 $\frac{1}{(N)_n}$ で抽出されるのであるから，n 次元確率変数（X_1, X_2, …, X_n）が母集団を構成し，非復元無作為抽出によって得られた n 個の単位の組は n 次元母集団確率変数の実現値であって，この母集団から取った大きさ1の任意標本とするのである[13)14)]。（したがって，母集団の確率分布と任意標本の確率分布は同じものになる。）

以上のように，任意標本調査の場合と推測統計理論の場合とでは確率変数たる母集団に違いがあり，それは任意標本調査と推測統計理論の適用対象の性格の違いによるものであることを明確に知らねばならぬ。

む　す　び

以上，2節にわたって任意標本調査の母集団の意義を考察して，調査対象の集りとしての母集団概念の積極的な役割を指摘し，次に，確率変数たる母集団が古典的な確率の定義に基くことを述べて，頻度確率の定義に基く推測統計理

12）宮沢光一，前掲書，17，39—40ページ。

13）小河原正己訳『S. S. ウィルクス数理統計学』上巻，昭和26年，124—125ページ，訳注(1)。

14）これは統計集団から n 個の単位を非復元抽出する場合であって，n 個の単位の復元抽出を行うときは，任意標本調査の母集団の規定はこれと違ったものになる。すなわち，復元抽出のときは毎回の抽出がN個の単位の中から行われるのであるから，その結果は相互独立で同一確率分布にしたがう1次元確率変数 X となる。したがって，n 次元確率変数（X_1, X_2, …, X_n）は推測統計理論の場合と同様に大きさ n の任意標本であり，1次元確率変数 X が母集団を構成することになる。

論の母集団との違いを明らかにしたのである。調査対象の集りとしての母集団は統計集団を構成する N 個の単位の集団であるが，確率変数たる母集団は N 個の単位から非復元抽出した n 個の単位の組全部の集りに対応する概念であった。したがって，非復元無作為抽出によって得られた 1 個の n 個の単位の組は，調査対象の集りとしては「大きさ n の標本」であるが，確率変数の実現値としてみるときは「大きさ 1 の任意標本」となるのである。これが母集団=任意標本の一般的定義による解釈であるが，一般にこのような区別はなされておらず，確率変数の実現値の場合でも「大きさ n の任意標本」というのが普通である[1)]。そして，n 次元確率変数（X_1, X_2, …, X_n）を母集団として規定することも少ないので，ここで述べたような推測統計理論の母集団との違いが明確でない場合が多い。

任意標本調査の母集団には以上のような二つの意義があるが，宮沢光一氏は『近代統計概論』(昭和31年）において両者を合せて母集団と規定しておられる。すなわち，「対象集団 Π と，単純ランダム抽出法を通して定まるこの確率変数（y_1, …, y_n）とを組にして（Π,（y_1, …, y_n)）を母集団，(y_1, …, y_n)を単純ランダム抽出における母確率変数とよぶことにする。」[2)] この宮沢氏の対象集団は調査対象の集りとしての母集団を意味し，母確率変数は確率変数たる母集団に相当する。しかし，調査対象の集りとしての母集団は標本調査の結果の非標本誤差の問題を考える基礎となり，確率変数たる母集団は標本誤差を算定する基礎となる等，両者を別々に考察することが必要であるので，それぞれの意義を明確に示す異なる名称を与えることが望ましい。そこで，調査対象の集りとしての母集団を，それから標本が取られる元の集団の意味で「母集団」と呼び，確率変数たる母集団はその本質である確率論的性格を明確に示すために「母集団確率分布（または母集団分布)」と呼ぶのがよいであろう。

1）小河原正己，前掲訳書，125ページ，訳注(1)
2）宮沢光一『近代統計概論』昭和31年，192ページ。