経　営	105
測　定	34

森林測定研究資料

9

（標本調査関係論文）

昭和37年3月

農林省林業試験場経営部

目　　次

1　有限母集団よりの抽出理論

On the theory of sampling from populations

I　近代的標本抽出理論の歴史的基礎

抽出単位と分析の単位が一致している母集団から、要素を無作為に抽出する理論は200年以上の昔 *Bernoulli* によって展開された。それより一世紀遅れて、サンプリングに層化を導入することによってえられる利益を評価する理論が Poisson によって示された。引続いて Lexis がそれまでの研究を体系づけ、要素の集落抽出（*sampling cluster of elements*）に対する理論的な基礎づけを行った[1]。*Bernoulli* および Poisson の有限母集団からの標本抽出理論に対する修正は Poisson の研究から約1世紀遅れて1926年に *Bowley* がとりまとめた。

Pearson, *Fisher* および他の人々の寄与に引続く標本調査理論の急速な進歩は、*Neyman* が1934年にその論文で *representative method* の二つの異なる側面について発表した[8]ことに刺戟されたものである。その論文のなかで彼は、標本内の全抽出単位数は一定であるという条件のもとにおいて、色々な層に対する抽出単位の最適割当の概念を含む調査資源の最適利用の基準を導入したのである。

もし標本が抽出されても、費用が標本に含まれる要素の数に全く支配されるのであれば、層内の要素の独立な無作為抽出を取扱

---

(1)　要素の集落の調査とは、1つ以上の要素を含む単位の抽出を意味する。集落抽出の例としては、個々の人または個々の世帯から構成される母集団特性を知ることを目的とする調査で *city block* または *county* を抽出単位とする場合である。これらの例で *city block* または *county* は要素の集落と考えられ、個人または世帯は要素と考えられる。

った Bernoulli および Poisson の古典的理論以上の、有限母集団理論の拡張、あるいは抽出単位の最適割当に対する拡張などで補正された理論を用いる必要性は少ないのである。しかし統計的研究で母集団要素の独立な無作為抽出を行なうことは、不可能でないまでも非常に屢々多額の費用を要するのである。実際の場合そのような標本調査では母集団の全要素を確認することのできるリストが必要であるが、この記録は存在しないことが多いか、または入手に非常に費用がかかる。そのような記録が利用できても、標本が非常に広範囲にわたっていると、要素を数えるための費用が極めて大となる。同様に標本計画（*sample design*）に他の制約があることも多い。例えば計算者の作業を限られた人数の監督者の厳重な管理下におくとか、限定された行政中心地で野外作業を行なわなければならない場合などである。集落抽出〔2, 3, 4, 5, 6, 7, 8, 10〕、副次抽出（*subsampling*）、二重抽出（*double sampling*）〔9〕のような抽出法は、利用できる資材の最も有効な使用を目的として、一方では設けられた管理的制限の範囲内で止まりながら他方ではこれらの資材および制限内で可能な最大限の情報を入手するために発展してきたものである。この点については Neyman〔8〕, Yates および Zacopany〔10〕, Cochran〔2〕, Mahalanobis〔7〕 および他の人々が重要な寄与をしている。

我々は上に引用した多くの進歩を簡単ではあるが統一的に一般的な副次抽出計画として説明することができる。この計画は層化母集団の要素の集落抽出および抽出された各集落要素の副次抽出を含むものである。この場合層内の一次抽出単位（*primary sampling unit*）のそれぞれにおける要素の数は同一である。

我々は母集団を L ケの層に分ち、i 番目の層はそれぞれ N_i ケの要素を有する M_i ケの一次抽出単位を含むものと仮定する。X_{ijk} は第 i 層の第 j 一次抽出単位の第 k 番目のある性質を示す値とする。そうして推定すべき性質は

$$(1)\quad \bar{X} = \sum_{i}^{L}\sum_{j}^{M_i}\sum_{k}^{N_i} X_{ijk} \Big/ \sum_{i}^{L} M_i N_i$$

であると仮定する。例えば $\bar{X}$ は与えられた街（*city*）の一世帯当りの平均収入である。X_{ijk} は第 i 区（*ward*）の第 j 通り（*city block*）、第 k 世帯の収入である。ここで世帯は副次抽出単位であり、通りは一次抽出単位、かつ層化は区によるものである。更に我々は第 i 層から m_i ケの一次単位を抽出し、この一次抽出単位のおのおのから n_i ケの単位を副次抽出するものとする。

標本からの $\bar{X}$ の"最良線型不偏推定値（*best linear unbiased estimate*）"〔8〕は

$$(2)\quad \bar{X}' = \sum_{i}^{L} \frac{M_i N_i}{m_i n_i} \sum_{j}^{m_i}\sum_{k}^{n_i} X_{ijk} \Big/ \sum_{i}^{L} M_i N_i$$

でありまた $\bar{X}'$ の分散は

$$(3)\quad \sigma_{\bar{X}'}^2 = \sum_{i}^{L} M_i^2 N_i^2 \left\{ \frac{M_i - m_i}{M_i - 1} \frac{\sum_{j}^{M_i} (\bar{X}_{ij} - \bar{X}_i)^2}{M_i m_i} + \frac{N_i - n_i}{N_i - 1} \frac{\sum_{j}^{M_i}\sum_{k}^{N_i} (X_{ijk} - \bar{X}_{ij})^2}{M_i N_i m_i n_i} \right\} \Big/ \left(\sum_{i}^{L} M_i N_i \right)^2$$

ここで $\bar{X}_{ij} = \sum_{k}^{N_i} X_{ijk} / N_i$, $\bar{X}_i = \sum_{j}^{M_i}\sum_{k}^{N_i} X_{ijk} / M_i N_i$

である。

これらの公式は更に費用の区別を考慮に入れなければ実際の抽出計画には役立たない。時には費用の関係を m_i および n_i の函数として正確に表わしうることがあるが、これは多くの場合、計画に含まれる種々の可能性の中から直観または理論を通じて妥当な決定を導くのに十分な近似を与えるものである。

もし我々に費用函数がわかっていれば、定められた費用に対し、また加えられるかもしれない他の制限のもとにおいて $\sigma_{\bar{X}'}^2$ を最小ならしめる m_i および n_i の値を定めることができる。この理論は色々な層ならびに層内の一次および二次抽出単位の抽出比の最適配分を決定するための基礎となるものである。

しかしこのような発展は標本計画の第1段階としてのみ考えられるべきである。我々は最適抽出計画が母数と費用因子のある種の函数であることを知っただけでは前に進むことはできない。それと同時に種々のサンプリングまたは推定に附随する費用の知識と、考察している特定の母集団内のある母数の相対的な大きさについての知識が必要である。

このように最近の多数の論文の取扱っているものは種々なタイプの抽出単位の単位間および単位内の分散および共分散の相対的な大きさ、ならびにそれに作用する費用函数のタイプについての研究である。この方面の研究は農業的な対象については *Department of Agriculture* が行なっており、また *Bureau of the Census* あるいは他の場所においても研究されている。

II 最近の研究動向

上に指摘した標本抽出法は、第1段階としての、例えば抽出法は集落抽出法、二重抽出法または副次抽出法を含むかどうかというような標本抽出方式（*system of sampling*）の定義、もしくはこれとともに層化および抽出単位の定義を含むものである。第2の段階は抽出単位の配分と抽出法を決定することである。

第1段階の抽出方式の定義は管理上の実行可能性と効率の面から行なわれる。しかし非実際的なリストの作成、またはすべての可能な方法を考え、その中からある基準で最良のものを一つ選ぶというようなことをする以外にはいつでも同じ方式を選び出せる簡単な方法は存在しない。しかし推定さるべき母集団特性の与えられた定義に対し、我々が最良線型不偏推定値のような"最良"についてのある種の基準を受け入れるならば、任意の費用あるいは規定された制限内で第2の段階に対する単一の解を与える簡単な方法がある。このような方法は、推定法および定義された抽出方式内におけるサンプリングの配分法の双方を我々に与えるものである。

"最良"の抽出方式を示すと同時に"最良"の推定法および抽出配分を示すための実際に応用できる理論はこれまで展開されていないが、改善された抽出方式と推定方式の選択については若干の進歩があった。我々は次の二つの方向の進歩が特に当を得たもののように思われる。

1 あるタイプの標本抽出方式で、より確実な結果を与えるための、全く一般に受けいれられる良好な推定値というものについての基準の修正（これらのいくつかは Sec.III で述べる）

2 抽出単位、層および抽出方式の他の面の決定を改善しうるある種の原則が現われる（このような原則を構成しようとする試みについては Sec.IV, V および VI で述べる）。

我々は主としてセンサスにおける最近の研究をまとめることにする。したがって密接な関係をもつ他の研究についても論及するであろう。我々が要約しようとする研究の大部分は、抽出単位が要素の集合で、その大きさが種々変化するという問題に関するものである。

III 良い推定値の基準に対する修正

Sec.I において提起した副次抽出の一般問題で与えられた推定値は"最良線型不偏推定値（*best linear unbiased estimate*）"の基準を満たすものである。我々の経験によれば、これは層内のある抽出単位中の要素の数が等しい母集団では、屢々最も有効（*efficient*）なものである。しかし抽出単位間の要素数が違っている場合でも、偏ってはいるが大体において最良線型不偏推定値よりも平均平方誤差[2]（*mean square error*）の小さい一般推定

---

2) $Ex=\hat{x}$ なるときの $E(x-\hat{x})^2$ を指すためこの論文では"平均平方誤差"と"分散"を同じ意味に用いる。Ex は推定さるべき母集団特性である。しかし $\sum x \neq \hat{x}$ のときは $E(x-\hat{x})^2$ はただ"平均平方誤差"だけを指すものとする。後者の場合は $E(x-\hat{x})^2=E(x-Ex)^2+E(x-\hat{x})^2$ となって平均平方誤差は x の分散＋偏りによる成動となる。

量（consistent estimate）が存在することが多い。

例えば集落が抽出単位のとき、ある特性の要素当りの平均値 $\bar{X} = \sum_i^M X_i / \sum_i^M N_i$ を推定したいとする。ここでMは母集団の抽出単位の数で、X_iはi番目の集落の全要素に対する特定の性質の合計値（aggregate value）である。N_iはこの集落の要素の数。X_iとN_iの同時分布は未知であるが、$\sum_i^M N_i = N$はわかっている。このような条件のもとでm個の集落を含む標本からの"最良線型不偏推定値"は$\frac{M}{m}\sum_i^m X_i/N$となることがわかる。しかし標本から$\sum_i^m X_i / \sum_i^m N_i$なる比推定を作れば多くの場合更に小さい平均平方誤差が与えられる。この推定値の偏りは通例無視できるものであり、一致性をも有するが、偏りのあることと非線型という理由で"最良不偏推定値"の基準にはあてはまらない。$\bar{X}$の最良線型不偏推定値にはNがわかっていなければならないから、この標本比は、Nが未知の場合でも使えるという点で優れている。

最近のcochran［3］の論文は、X_iのN_iへの線型回帰の最小二乗推定を用いて、一致性はあるが、偏りのある$\bar{X}$の多くの推定値を与えている。

これらの推定値の平均平方誤差は一般に上にあげた最良線型不偏推定値あるいは単純比推定値のそれよりも小さい。しかしこれらに対しては最良線型推定値と同じくNの知識が必要であり、その上推定の手続の一部として相当な計算を要する。

上に述べた偏りのある推定値はいずれも一致推定量であって、普通抽出単位の大きさが異なるような抽出方式では、最良線型不偏推定値よりも平均平方誤差が小さい。したがって、"最良線型不偏推定値"の基準に、非線型、一致推定量ではあるが平均平方誤差が最良線型不偏推定値よりも小さいような推定値を含むことにすれば、改善された標本推定値が与えられる。

Ⅳ　抽出方式の指定に関する改善

母集団構成に関する知識が少なく、その特定の母集団毎数についての知識が皆無の場合でさえ、抽出方式の指定の改善を通じて、抽出計画を大きく改善することができる（Sec.Ⅵをみよ）。

1　抽出単位の大きさ

最近の研究の多くは、圃場調査の抽出単位に要素の集落を用いるとき、考えられている費用のもとで集落の大きさをできるだけ小さくすることが望ましいことを示している［2, 5, 6, 7, 8］。しかしこの原則は副次抽出方式では必らずしもあてはまらないことに注意しなければならない。副次抽出を含む抽出方式では、一次抽出単位に大きい集落を用いると副次抽出を用いないで小さい集落を用いた場合よりもっと別の利点がある。その上大規模調査において屢々生ずる問題の1つは、調査を行なう場所数に管理上の厳重な制約があって、そのような条件のもとで調査計画をたてなければならないことである。このような制限がある場合によく用いられる方法は、限られた数のcounty などの行政単位を一次抽出単位とし、副次抽出単位にはblockや小さい地方的地域あるいは世帯をとるものである。このような事情のもとで標本内に含まれる一次抽出単位の数を一定に保つとすれば、実際に存在する行政的単位より大きい一次抽出単位を用いることによって抽出分散を減少させることができる。

副次抽出で大きい抽出単位を用いることの利点は、例えば、f回数の要素を含むもとの単位（original unit）を、それぞれしの2倍の大きさをもつ$1/2$の数の一次抽出単位に結びつけるような簡単な場合を考えれば明らかである。拡大された一次抽出単位間の分散は$\sigma_{1c}^2 = \frac{1}{2}\sigma_{1b}^2(1+f)$となる。ここで$\sigma_{1b}^2$はもとの一次抽出単位間の分散で、$f$は対にした単位間の相関係数である。この相関係数は対が確率的に作られている場合には無んどのに近い（正確には$-1/(M-1)$に等しい）。ここでMはもとの一次抽出単位の数である。したがってcounty 間の分散は少くとも略々半分になる。対にする単位は隣りあっていなけれ

ばならないという条件があれば、普通 f の値は0より大となろう。しかしなるべく違いの大きい隣接単位を選んでこのような単位の結合をつくれば、f は可能な限り小さくなり、ある場合にはこの極小値が負になることさえある。何れの場合でも、f の値が小さい程拡大単位を用いて生ずる抽出単位間の抽出分散（*sampling variance*）の減少は大きくなる。一方、一次抽出単位間の抽出分散はこのような結合を行なうことによって増大するが、その増加は僅少だから、全抽出分散は常に減少する（*Appendix Sec. 1*）。一次抽出単位の結合を大きくする場合の制約は、地域が増すにつれて副次抽出の費用が増大することによって生ずる。このような費用の増加は分散の減少と逆の仕らきをする。もし費用の制限がそうきびしくなければ、結合は一次抽出を全然用いずに副次抽出単位を分層から独立に抽出する方向に進められるであろう。

2' 一次抽出単位の大きさが異なる場合の副次抽出。副次抽出で大きさに比例する確率を用いること。

一次抽出単位の大きさが変化すると否とにかかわらず屢々用いられる副次抽出組織では、それぞれの層から層内の一次抽出単位の抽出される確率が同じになるようにして、1つまたはそれ以上の一次単位を抽出し、その中から定められた割合の副次抽出単位を抽出しなければならない。一次抽出単位の大きさが不同の場合には、抽出された一次抽出単位内の要素数につれて標本中に含まるべき副次抽出単位の数が変化するから、この抽出方式には若干の管理的な不利益が生ずる（この論文において用いる"抽出単位の大きさ（*size*）"とは、抽出単位内に含まれる要素の数を指すものである）。上記の抽出方式の不利益は、たとえ一次抽出単位が等確率で抽出されても、副次抽出が一定比率というよりはむしろ一定数であるように指定された第2の副次抽出方式の場合に生じたものであった。上記の二つの抽出方式より更に推奨できる第3の抽出方式は、一次抽出単位にその大きさに相応する抽出確率を与え、したがって一定数の副次抽出単位を副次抽出するものである。我々は3つの抽出方式の何れについても、各層からは唯一つの一次抽出単位を抽出するものと仮定する。この程度まで層化を行なうと、層化が不完全な場合よりはるかに抽出分散が小さくなる。比較のための単純化として、我々は更に副次抽出単位が分析単位であり、用いられる標本推定値が $\bar{X}' = \sum N_h \bar{X}'_h / \sum N_h$ なる形のものとする。ここで $\bar{X}'_h$ は推定している特性の h 層に対する標本平均であり、N_h はその層の大きさである。屢々用いられる推定値は、最初の二つの抽出方式では偏りがあるが、推奨された抽出方式では不偏である。しかし不偏推定値すなわち最初の二つの抽出方式に対する"最良"線型不偏推定値の平均平方誤差は、一般にこれらの比較で用いられる偏りのある推定値のものよりずっと大きいから、次の比較では考慮しなかった（*Sec. III. 脚注7*を見よ）。

前にあげた最初の二つの副次抽出方式は、何れも適当な大きさの一次抽出単位から副次抽出を行なえば、その効率は大体同じである。しかしどちらも推奨した抽出方式より普通平均平方誤差が大きい。最初の二つの抽出組織と推奨した計画との平均平方誤差の差は近似的に何れも

$$(4)\quad \frac{1}{N^2} \sum_h Q_h \bar{N}_h \sigma_h^2 \left[\sum_j f_{hj} \bar{N}_h - \sum_j f_{hj} N_{hj} \right]$$

で表わされる。ここで N_{hj} は h 番目の層内の j 番目の一次抽出単位内の要素数であり、$\bar{N}_h$ は一次抽出単位の平均の大きさである。Q_h は一次抽出単位の数、f_{hj} は j 番目の単位内要素間の級内相関係数（*intra-class correlation*）、σ_h^2 はこの層内の個々の要素間の分散で、L は層の数である（*Appendix Sec. 2* にあるこの差の展開をみよ）。

もし N_{hj} と f_{hj} の相関が負なら、N_{hj} と f_{hj} 間の平均の共分散（*covariance*）であるところのこの差は正になる。そうして

これは我々が社会統計および経済統計で屢々遭遇する大部分の実際的問題においてみられるものである（Sec. VI をみよ）。

推奨される計画において平均平方誤差が減少するのはその大抽出単位と小抽出単位間の抽出単位の配分が、他の二つのものより最適に近いことによるものである。もちろん他のものと同様に、大きさで一次抽出単位を層化し、最適配分を考えて色々な層に抽出をふりわけることもできる。しかしこれによって何か他の、しかもおそらく更に重要な層化の方法が犠牲になり、加えて大部分の実際問題で大抽出単位と小抽出単位間の最適配分は推測による他はないであろう。その上、普通層内の単位の大きさが同じ場合には大きさによる層化は不可能である。

推奨された抽出方式からえられる推定値は他の二つの推定値が通例偏りをもち、ある場合には非常な偏倚を示すにもかかわらず不偏であること（このことの証明は Appendix Sec 1 を、数値例については Sec. VII をみよ）。

大きさに比例した確率を用いても一次抽出単位間の抽出変動が減少するだけであるから、級内の抽出変動を減少させるとくには効果が少ない。したがって一次抽出単位間の平均平方誤差の全平均平方誤差に対する寄与が大きい場合には推奨された計画は他の二つのものより非常に有利である。

通例一次抽出単位の実際の大きさは不明であるが、これと直接な関係をもつ単位の数はわかるものである。たとえば都市 block, city, あるいは county 等の母集団は、標本がとられたときには我々にはわからないものである。しかし以前のセンサスにおけるこれらの母集団については知ることができる。これらの状況のもとで、一次抽出単位はすでに知られている（あるいは推定された）大きさに比例した確率で抽出される。しかしこのようなことが行なわれれば、二つの調査時点間の抽出単位の大きさの変化を考慮するため、この副次抽出は修正されなければならない。

もし実際の大きさが既知なら、ℎ-層で抽出された一次抽出単位の中からとる副次抽出単位の数は、

$$n_h = t_h N_h$$

である。そこで t_h はこの層に対して定められた抽出比であり、N_h は層内の全要素の数である。したがって抽出された一次抽出単位内の副次抽出率は $t_h N_h / N_{hj}$ である。N_{hj} は抽出された単位内の要素数である。一方もし単位の実際の大きさ N_{hj} と密接な関係にある大きさの尺度 P_{hj} だけしか利用できなくて、一次抽出単位の抽出確率を P_{hj} に比例するようにとった場合には、抽出された一次抽出単位内の副次抽出の率は $t_h P_h / P_{hi}$ に等しくなる。ここで P_h は層全体に対するこの尺度であり、P_{hi} は抽出された一次抽出単位の大きさの尺度である。大きさの尺度を用いたときの標本推定値の分散はこの論文のあとで述べる（E_8 (9) をみよ）。

3 副次抽出方式の場合一次層内で地域的副次層化を用いること。広い地域にわたる母集団から比較的小さい標本をとり、その作業を小数の中心地で行なわねばならないような場合には、一次層内（primary strata）で地域的副次層化（area substratification）とよばれる他の修正を行なうことが特に有用である。地域的副次層化について説明する前に少し準備が必要である。

地域的副次層化には、

(a) 標本調査を行なう全母集団を一次抽出単位となる地域に分割する。

(b) これらの単位を更に細分して数個の副次区域（sub-areas）に分割する。そうして

(c) 標本を抽出する前に各副次区域についてある程のまとまった統計的知識が利用できる。

ことが必要である。この副次的な地域について知らねばならない知識にはそれらの大きさ（人口、全母集団、全世帯（dwelling

unit）あるいは全農家）に対する妥当かつ良好な尺度およびこの地域の特性を示すたとえば農業地域（predominantly farm）であるか、非農業地域（predominantly non-farm）であるか、あるいは白人の多い地域（predominantly white）か有色人種の多い地域（predominantly colore）であるかなどの他の知識も含まれる。斉一なクラスに群わけした場合、副次区域は次に述べる副次層（sub-strata）を決定するためだけに用いられるもので、普通それと独立に定められる副次抽出単位を与えるものではない。前に述べたように層内の一次抽出単位ができるだけ斉一になるように、また層間で可能な限り調質的になるように一次抽出単位の定義およびその層への群わけを行なう。この場合各層からは一次単位を一つだけ抽出し、h層のi-j一次単位の選ばれる確率は P_{hij} に比例させるということが仮定される。ここで P_{hij} は一次単位の大きさの尺度で、これはその中に含まれる副次区域の大きさの尺度の和に等しい。また h層内で用いる全抽出率 t_h は全部の層に対して最適配分の考えをもととして決定されているということも仮定される。

そうすると地域的副次層化は次のようにして一次層内に導入できる。

(a) それぞれの一次層内の副次区域をそれらの特性をもととして副次層に群わけする（たとえば農業地域と非農業地域という二つの副次区域に群わけし、次にこれらの農場の平均的な大きさまたは世帯（dwelling unit）の平均地代によって更に分類する。そのようなとき、農業地域にあり、かつ地代の平均が指定された幅の中にあるような一次層内の副次区域は一つの副次層を構成する）。

(b) 一次層の各々から選ばれた一次単位内の副次区域は同じ副次層内に群わけする。

(c) 副次抽出単位は選ばれた一次単位内の各次層内で定められる。i-j一次抽出単位内に含まれる i 番目の副次層の中で定められる副次抽出単位の個数は M_{hij} で表わされる（種々の副次抽出単位、たとえば個人、農場、世帯またはその構成員、あるいは非常に小さい地域などとして定義することができる。副次抽出単位は選ばれた一次抽出単位の中でのみ定義しなければならない）。

(d) 選ばれた（j 番目の）一次抽出単位内の i 番目の副次層からとった標本に含まれる副次抽出単位の数は

$$m_{hij} = M_{hij}\, t_h\, P_{hi} / P_{hij}$$

である。ここで P_{hij} は j 番目の一次単位内にある i 番目の副次層の大きさを表わす。$P_{hi} = \sum P_{hij}$ は h 番目の一次層（Primary stratum）の i 番目の副次層に含まれる副次区域の大きさの和である。

このような配分法によって副次抽出を行なえば、選ばれた一次単位からとった標本は、その層から標本内にたまたま選び出された特定の一次単位を代表するというよりはむしろ層全体を可能な限り代表することになる。説明のため、1940年のセンサスからえられた副次区域内の人数をそれらの大きさを表わす尺度にとり、これらの副次区域を1940年の Decennial Census of Population で示された1940年におけるそれらの特性にもとすいて副次層に分類するものとしよう。そうすると上述の副次抽出の配分から、農業地域内の副次区域に居住する人口比率が30%なら、標本内で予想される1940年の人口の30%が農業地域内の副次区域からのものとなるように標本を抽出する。このとき選ばれた一次抽出単位内では1940年の人口中そのような地域に住んでいたのはおそらく15%にしかすぎないかも知れないがそれは問題外である。

(e) 推定すべき母集団特性は

$$(6)\quad X = \sum_{h}^{L} \sum_{i}^{S_h} \sum_{j}^{Q_h} \sum_{a}^{M_{hij}} X_{hija}$$

である。ここで X_{hijk} は j 番目の一次単位、i 番目の副次層 h 番目の副次抽出単位に含まれる全要素についてある特性の値を加え上げたものである。また S_h は副次層の個数、Q_h は h 番目の一次層内の一次単位の数であり L は一次層の数であるく X は米国の全労働者数でもよいし、また全農業労働者数等でもよい)。標本からの推定値は、

$$(7)\quad X' = \sum_{h}^{L} 1/t_h \sum_{i}^{S_h} \sum_{k}^{M_{hij}} X_{hijk}$$

である。j に対する和は不要である。何故なら h 番目の層から抽出される一次単位は 1 つだけだからである。これは一次層の段階だけで重みをつけた和を含む非常に単純な推定値である。もし t_h がすべて t に等しい、すなわち標本の抽出比を一定とすれば、推定値は標本内のある特性を有する全要素の数に抽出比の逆数 $1/t$ をかけただけのものとなる。

上に述べた副次抽出の概念と選った方法をとることもできるが、その場合でも標本推定値に正しい修正を加えれば地域的副次層化の拘束が維持される。このときには一次層ばかりでなく副次層の段階にも別の重みを導入しなければならない。

異質な一次抽出単位の定義やそれらを正しく層化すること、あるいは一次単位を選ぶ際に大きさに比例した確率を用いることなどは、地域的副次層化を利用するときには特に望ましい。もしこれらを導入しないときは、地域的副次層化を行なっても実質的な利益のえられる可能性は少なくなるであろう。一次層の定義は副次層の定義と関連させて行なうべきであり、また各一次単位も一次層内で定められるそれぞれの副次層を十分代表するものでなければならない。ここに見られた制約のために、定めうるはっきりした層の数は一次単位の異質性によって制限をうけるであろう。したがって農業地域と非農業地域に副次層化する場合には農業地域と非農業地域の両方が各単位内で代表されるように一次抽出単位を定めなければならない。この方法は副次層化を一層効果的にするばかりでなく、母集団のそのようなクラスの推定を別々に行なう場合標本の効率を改善するものである。実際の場合にこの手続を正確にあてはめられなければ、ある副次層を代表しないような抽出単位が標本内に入ってくることが多い。このような場合にとるべき方法の一つとしては特定の副次層を合併することであり、もう一つは標本からそのような一次単位を取除くことである。

一次層の数は抽出すべき一次単位の数によって制約を受けるから、地域的副次層化で十分抑制できる変動因について一次の段階(primary level)で層を作ることは不経済である。例えば副次層で農業地域と非農業地域を区別すべきものとすれば、一次単位を農家比率(percent farm、全母集団中の農業によって生計をたてている一次単位の比率)で多数の層に分類してしまってはならない。それは農家比率の変動をコントロールするのは副次層化だからである。一次の段階で農家比率のクラスの数を制限することによって、農家の型や非農家母集団の生産的特性、あるいは他の同様な基準をコントロールできる別な形の層化を利用することができる。

地域的副次層化は、要素の色々なクラスの各々から、標本内に含ませる要素数を定めるのにごく普通に用いられている方法——一次層の全部または選ばれた一次抽出単位の特定の特性と標本とを一致させるために、そのような割当が固定されていてもいなくても——と区別しなければならない。

与えられた数の、指定された色々な特性をもつ要素(人、居住単位、農場あるいは有权者等)をうるために、割当を定めたり、面接者または調査者に指令を与えたりする方法には根本的な欠点があって、一次層内の地域的副次層化では避けなければならない。普通そのような割当は既往の情報や大雑把な推定をもととしなければならないことから、母集団の変化した特性を正確に示すことはできない。一方地域的副次層化は既往の情報を利用して、色々なタイプの地域が標本内に正しく代表されることを保証するもの

である。指定された種々の特性についてとられる要素の数は過去の時点でなく現在の母集団から決定される。変化が急激な時期に過去の情報に基づいて一定の割当を用いるとますます著しい偏りを生ずることになる。地域を層化する際、利用できる過去の情報を用いることによってえられる利益は、地域の時期毎の特性間に数年にわたって高い相関があるという事実から生ずるものである。ある時点において農家比率の高かった地域（農業地域）は数年後でも農家比率は高いのが普通である。同じように、母集団に事実上非常に大きい変化が生じても、ある時点における一組の地域内の人口は普通数年後の人口と非常に高い相関をもっている。しかし地域的副次層化は変化が生じないという事実に頼るものでなく、変化が生じたならそれを評価するものである。変化が大部分の小地域の特性を全く変えてしまう程大きい場合でも、この方法は母集団特性の変化を明らかにする推定値を与えるであろうが、しかしそのような事情のもとではこの方法の効率は減少する。

V　上述の原則を具体化する副次抽出方式の期待値と分散

一次抽出単位の拡大、大きさの尺度に比例した確率をもって行なう一次抽出単位の抽出、および地域的副次層化の原則を具体化する抽出方式を以下において一層詳細に論ずることにする。これを便宜上指定された副次抽出方式とよぶことにする。

1　指定された副次抽出方式に対する総計の推定値の期待値

以下の公式中の和は特に断らない限りすべて母集団全体にわたるものとする。

(ア) 式で定義された X' の期待値は

$$EX' = \sum_{\alpha}\sum_{\beta}\sum_{i}\sum_{\alpha\beta}(1/t_{\alpha})(P_{\alpha i}/P_{\alpha})(m_{\alpha i\beta}/M_{\alpha i\beta})X_{\alpha i\beta\alpha}$$

である。(5) から $t_{\alpha} = m_{\alpha i\beta}P_{\alpha i\beta}/M_{\alpha i\beta}P_{\alpha i}$ だから

$$EX' = \sum_{\alpha}\sum_{\beta}\sum_{i}\sum_{\alpha\beta}(P_{\alpha i}/P_{\alpha i\beta})(P_{\alpha i}/P_{\alpha})X_{\alpha i\beta\alpha}$$

$$= \sum_{\alpha}P_{\alpha}\sum_{\beta}\sum_{\alpha\beta}(P_{\alpha i}/P_{\alpha})(P_{\alpha i\beta}/P_{\alpha})(X_{\alpha i\beta}/P_{\alpha i\beta}) = \sum P_{\alpha}R_{\alpha(A)}$$

ただし、

$$P_{\alpha} = \sum_{i}P_{\alpha i} = \sum_{\beta}P_{\alpha i\beta} \quad : \quad R_{\alpha(A)} = \sum_{\beta}(P_{\alpha i\beta}/P_{\alpha})R_{\alpha i\beta(A)} \quad :$$

$$R_{\alpha i\beta(A)} = \sum_{i}(P_{\alpha i}/P_{\alpha})R_{\alpha i\beta} \quad :$$

$$R_{\alpha i\beta} = \sum_{\alpha}X_{\alpha i\beta\alpha}/P_{\alpha i\beta} = X_{\alpha i\beta}/P_{\alpha i\beta}$$

この $R_{\alpha i\beta(A)}$ は第 β 一次単位に対する補正された比と考えられる。これは第 β 単位内における $R_{\alpha i\beta}$ の重みづき平均値である。そこでは層内のそれぞれの一次単位の $R_{\alpha i\beta}$ に対し同一組の重みが適用される。$R_{\alpha(A)}$ は調整された比の第 α 層内の平均である。よって

(8)　$EX' = X + \sum_{\alpha}P_{\alpha}(R_{\alpha(A)} - R_{\alpha})$

ここで　$R_{\alpha} = X_{\alpha}/P_{\alpha}$,　$X_{\alpha} = \sum_{i}\sum_{\beta}X_{\alpha i\beta}$

は、第 α 層内要素の指定された特性値の総計と層の大きさとの比で、(イ) で推定している母集団特性は　$X = \sum X_{\alpha} = \sum P_{\alpha}R_{\alpha}$ に等しい。(8) からわかるように、通例それは実際には僅かだが、X' は X の偏りのある推定値である。この偏り $\sum P_{\alpha}(R_{\alpha(A)} - R_{\alpha})$ は色々な一次層に対する偏りの和である。現実には多くの場合これらのあるものは僅かに正となり、あるものは僅かに負となるだろうから、全体の偏りは比較的小となるであろう。この偏りは地域的副次層化を用いないか、または標本推定値の式を正しく修正すれば生じなかったであろうが、この場合もまた Sec. III で述べた不偏推定値の代りに偏りのある推定値を代入した場合と同じく、僅かの偏りを導入することによって分散を事実上減少させうるのである。

地域的副次層化を行なったときの標本推定値 (ア) が不偏であるための十分条件（必要条件ではない）は、比 $P_{\alpha i\beta}/P_{\alpha i}$ が各副

次層内の R_{hij} と無相関であることである。これらの条件のもとでは、

$$\sum_j \frac{P_{hj}}{P_h}\frac{P_{hij}}{P_{hj}} R_{hij} = \sum_j \frac{P_{hj}}{P_h}\frac{P_{hij}}{P_{hj}} \sum_k \frac{P_{hk}}{P_h} R_{hik} = \frac{P_{hi}}{P_h}\sum_j \frac{P_{hj}}{P_h} R_{hj}$$

だから

$$R_h = \sum_i \sum_j \frac{R_{hi}}{P_h}\frac{P_{hij}}{P_{hi}} R_{hij} = \sum_i \sum_j \frac{P_{hi}}{P_h}\frac{P_{hj}}{P_h} P_{hij} = R_{h(A)}$$

例をあげれば1940年の人口を大きさの尺度とすると、標本推定値は、種々の副次層内の一次抽出単位に対する1940年の人口比率が対応する R_{hij} と無相関なら不偏である。前に注意したように、これらの条件は多くの実際問題において、特に第一次層化（primary stratification）が効果的に行なわれた場合には近似的に満足される。更にこの条件が近似的に満たされなくても、導びかれる偏りは非常に小さいものであろう（数値例は Sec. 7 をみよ）。

2　指定された副次抽出方式に対する総計の推定値の平均平方誤差

指定された副次抽出方式に対する X' の平均平方誤差の展開は附録 Sec. 2 をみよ。それによれば X' の平均平方誤差は

$$\sigma^2_{X'} = \sum_h \sum_i \sum_j P_{hi}^2 \frac{P_{hj}}{P_h}\frac{M_{hij}-m_{hij}}{M_{hij}-1}\frac{\sigma^2_{hij}}{m_{hij}\bar{P}^2_{hij}}$$

(9)

$$+ \sum_h P_h^2 \sum_j \frac{P_{hj}}{P_h}(R_{hj(A)}-R_{h(A)})^2 + \left[\sum_h P_h (R_{h(A)}-R_h)\right]^2$$

ただし $\sigma^2_{hij} = \sum_k (X_{hijk}-\bar{X}_{hij})^2 / M_{hij}$ は副次抽出単位に対する指定された特性値総計の副次層内における副次抽出単位間の分散であり、$\bar{P}_{hij} = P_{hij}/M_{hij}$ は第 $h-i-j$ 番目の地域内における副次抽出単位の平均の大きさである。

(9) の第1項は副次抽出単位間の分散の寄与を表わすもので、これは副次抽出単位を正しく定義することにより、また初論副次抽出比を増すことによって小さくできる。(9) の第2項は層内の一次抽出単位間の分散の寄与を示す。第3項は偏りの寄与であるが、前に指摘したように通例無視できる大きさであるから、この平均平方誤差と分散は近似的に等しくなる。

多くの副次抽出で全分散に最も大きく寄与するのは一次抽出単位間の分散である。そうしてこの論文で提案した修正が効果をあげるのは主としてこの寄与に対してである。地域的副次層化の効果は、上に与えた一次単位間の分散と、地域的副次層化を用いず、また計画の他の部面を変更しなかったときにえられる分散を比較すればわかる。この場合一次単位間の分散には補正された比 $R_{hj(A)}$ の分散でなく、比 $R_{hj} = \sum_i X_{hij} / P_{hj} = X_{hj}/P_{hj}$ の分散が含まれる。

R_{hj} の分散と第 h 一次層内の $R_{hj(A)}$ の分散の間の関係は、

$$\sigma^2_{R_{hj}} = \sigma^2_{R_{hj(A)}} + \sigma^2_{R_{hj}-R_{hj(A)}} + 2\rho\,\sigma_{R_{hj(A)}}\sigma_{R_{hj}-R_{hj(A)}} \tag{10}$$

で与えられる。ここで $\sigma^2_{R_{hj}-R_{hj(A)}}$ は補正した比と補正しない比の差の分散であり、ρ は補正された比と補正量との相関である。したがってもしこの相関が0に近いか、あるいは正なら副次層化を導入することによって利益が生れるが、しかしこの相関が負の大きい値をとるときには損失が生ずるであろう。本質的に、ρ が0に等しいかまたは0に近い値となるための条件は、標本推定値が不偏となるための条件と同じである。すなわち P_{hij}/P_{hj} の各層内の R_{hij} と無相関であるか、または僅かの相関しかもたないことである。(3)

X' の分散の中に現われるのは R_{hj} のものよりはむしろ $R_{hj(A)}$ の分散である。なぜなら、副次抽出単位の数はたまたま標本内にどの一次抽出単位が選ばれたかには関係なく、R_{hi} に比例し

(3)　実際に ρ が0に等しいための十分条件（必要条件ではない）は P_{hij}/P_{hi} がすべての層の対に対して、比 R_{hij} および cross-product $R_{hij}R_{hkj}$ の両方と無相関なことである。

て配分されているからである。$R_{Aj(A)}$ と同じく比 R_{Aj} は R_{Aij} の重みづき平均値と考えられるが、この重みは P_{Ai} でなく P_{Aij} に等しい。したがって一次単位ごとに変化する。ゆえに上に与えた分散の関係から、もし副次層が効果的に設けられ、また P_{Aij} が副次層の実際の大きさと高い相関をもつならば、すべての一次単位において一定の重みを用いた重みづき平均値は重みを変化させた場合のものより分散が相当小さくなる等である。これは多くの実際的な状況のもとにおいて明らかとなるものであるが、その説明はあとで(Sec. 7)をみよ)与えることにする。

3　指定された抽出方式に対する比推定の平均平方誤差

標本から比を推定する必要が生ずるのは二つの場合である。その一つはこの比が推定したい母集団特性である場合であり、その二は希望する総計についての改善された推定値を求めるために、利用できる追加情報を用いて既知の総計に標本から求めた比を適用する場合である。

比の推定値はたとえば、時点間の特性の変化を考えているようなときは最終結果として要求される。したがってもしある時点における農業労務者の総所得の推定値を Y' とし、X の同時点における対応する推定値を X' とすると、$r' = X'/Y'$ はこの期間内の農業労務者の総所得についての相対的変化の一つの推定値である。同じように失業者率のようなパーセンテイジの推定値は標本から求めた二つの確率変数の比からなっている。標本からえた比の推定値は、それが分子または分母の何れの推定値より信頼性が高いとき(このようなことは屢々ある)特に有用となる。

比の推定値は、X と高い相関をもつ X の特性の総計 Y が独立な出所から正確にわかっていて X と Y の推定値 X' と Y' が標本からえられるとき、指定された特性値総計の推定値を求める手段として利用できる。すなわち

(11)　$X'' = \{X'/Y'\}Y = r'Y$

は指定された特性総計の推定値である。もし連続してとられた標本で、X' と Y' の相関が十分高ければ、この比の推定値は前に(7)で与えた総計の単純推定値 X' より一層有効な X の推定を与えるであろう。しかし相関が低い場合には X' の方より信頼できる推定値であることが証明される。[4] 故に X' と Y' との間の相関が十分高い場合には、X' を推定するよりも X'' を求めることによって、X の推定に利用できる情報を一層適切に利用しうるであろう。

指定された副次抽出方式に対する比の推定値の適用を以下において考察しよう。

(a)　比の推定値とその平均平方誤差

母集団の比 $r = X/Y$ の推定値は

$$(12)\qquad r' = \frac{X'}{Y'} = \frac{\sum_{A}^{L} \frac{1}{f_A} \sum_{i}^{S_A} \sum_{j}^{I} \sum_{k}^{m_{Aij}} X_{Aijk}}{\sum_{A}^{L} \frac{1}{f_A} \sum_{i}^{S_A} \sum_{j}^{I} \sum_{k}^{m_{Aij}} Y_{Aijk}}$$

である。ここで X' は上の(7)で与えられまた Y' は X の特性値総計の同様な推定値である。r' の平均平方誤差は近似的に

$$\sigma_{r'}^2 = \frac{1}{Y^2}\Bigg\{ \sum_{A}\sum_{i}\sum_{j} P_{Ai}^2 \frac{P_{Aj}}{P_A} \frac{M_{Aij} - m_{Aij}}{M_{Aij} - 1} \frac{\sum_{k} Y_{Aijk}^2 (r_{Aijk} - r_{Aij})^2}{m_{Aij}\, M_{Aij}\, \bar{P}_{Aij}^2}$$

$$+ \sum_{A}\sum_{i}\sum_{j} P_{Ai}^2 \frac{P_{Aj}}{P_A} \frac{M_{Aij} - m_{Aij}}{M_{Aij} - 1} \sigma_{R_{Aij}:Y}^2 \frac{(\bar{Y}_{Aij} - \bar{Y})^2}{m_{Aij}\, \bar{P}_{Aij}^2}$$

$$+ \sum_{A} P_A^2 \sum_{j} \frac{P_{Aj}}{P_A} R_{Aj(A):Y}^2 (\bar{r}_{Aj(A)} - \bar{r}_{A(A)})^2$$

$$+ \sum_{A} P_A^2 (\bar{r}_{A(A)} - r)^2 \sum_{j} \frac{P_{Aj}}{P_A} (R_{Aj(A):Y} - R_{A(A):Y})^2 \Bigg\}$$

---

(4)　$r' = X'/Y'$ なる形の確率変数の比の分布は、近似的に $\sigma_{r'}^2 = r^2(V_X^2 + V_{Y'}^2 - 2\rho_{X'Y'}V_X V_{Y'})$ で表わされる。ここで V は添字で示された変量の変動係数で、$\rho_{X'Y'}$ は相関係数である。したがってもし $\rho_{X'Y'}$ が十分大きければ $V_{r'}^2$ は $V_{X'}^2$ より小さい。$\rho_{X'Y'}$ の大きさは X' と Y' の変動係数の相対的な大きさに依存せねばならない。

である。ここで

X_{hijk} = 第 $h-i-j$ 地域内の第 k 副次抽出単位内要素の、我々がその総計を推定しようとしている指定された特性値の総計

Y_{hijk} = 同じ副次抽出単位内の要素に対する第2の指定された特性値の総計で、これの母集団総計がわかっている。

$$Y_{hij} = \sum_k Y_{hijk}, \qquad Y_h = \sum_i \sum_j Y_{hij}$$

$\sigma^2_{hij:Y} = \dfrac{\sum(Y_{hijk} - \bar{Y}_{hij})^2}{M_{hij}}$ は第 $h-i-j$ 地域内の第2の特性に対する抽出単位間の分散で、また $\bar{Y}_{hij} = Y_{hij}/M_{hij}$ である。

$R_{hij(A):Y} = \sum \dfrac{P_{hi}}{P_h} \dfrac{r_{hij}}{P_{hij}}$ は Y_{hij} の補正された平均で、

$r_{hijk} = \dfrac{X_{hijk}}{Y_{hijk}}$, $r_{hij} = \dfrac{X_{hij}}{Y_{hij}}$ 等は添字で示された地域に対するXのYに対する比。また

$\bar{r}_{hij(A)} = \dfrac{R_{hij(A)}}{R_{hij(A):Y}}$, $\bar{r}_{h(A)} = \dfrac{R_{h(A)}}{R_{h(A):Y}}$ は添字で示されたXとYの補正された比の比である。

あとの記号は $\bar{X}'$ の期待値および分散を与えた前の諸節で定義したものと同じである。

(ハ)の最初と3番目の項は普通には主要な寄与を与える項である。第2項と第4項は標本が副次層および第一次層について層化されている場合でも、これらの各々に対する平均値間の変動にもとづく寄与を含んでいる。ある場合にはこれらの項の寄与が重大となることがある。層間の寄与は一次単位および副次抽出単位の特性Yの大きさが変動するためには生ずる。

比の平均平方誤差の公式は脚注(4)で与えたより一般的に用いられているものと近似的に等しい。普通には無視できる特定の項が(ハ)式で無視してもよければ、何れも近似であるこの二つの公式は等しくなる。このあとの方の公式は比の分散に対する標本設計の色々な面の効果を示すという利点がある。分散のこの近似公式の導びき方は、近似の正確さの説明とともに附録の Sec 3 で述べる。

(4) 総計の推定値とその平均平方誤差

前に指適したように、すべての要素に対する与えられた特性値の総計Xの二つの推定値は $\bar{X}'$(ア) および $\bar{X}''$(イ) である。$\bar{X}'$の平均平方誤差は(タ)で与えられ、$\bar{X}''$のそれは単に $Y^2\sigma^2_{\bar{Y}}$ である。この $\sigma^2_{\bar{Y}}$ は近似的に(ハ)で与えられる。

Xの推定値として $\bar{X}'$ を用いるか $\bar{X}''$ を用いるかということはもちろん第1の例ではYがわかっているかどうか、また第2の例では(タ)と(ハ)で与えられるそれぞれの平均平方誤差の相対的な大きさにかかっている。これらは調査している母集団の関係についての事前の知識で近似できるか、あるいは予備調査から推定できる。しかし副次層内の X_{hijk} と Y_{hijk} との間に正の相関があるとき、比の推定に必要な情報が利用できさえすれば比を用いることによる損失は僅かで、おそらくそれによって相当の利益をあげうると考えてよい。

(イ)のYが一次抽出単位の実際の大きさの総計で $\bar{Y}'$ がその推定値であるときは、(ア)の代りに(イ)を用いることが屢々必要となる。これはもし使用した大きさの尺度が実際の大きさの十分正確な尺度でなく、またこれとともに、実際の大きさが推定されている特性と高度の相関をもっていく、比の推定値を用いることにより一次抽出単位間の寄与と、一次抽出単位内分散の双方から利益がえられるようなときには特にそうであるが数値例についての Sec 例をみよ)。しかし大きさの尺度が実際の大きさに等しくすなわち $P_{hijk} = r_{hijk}$ と(ハ)の最後の2項は $\bar{X}'$ の分散(タ)に対する一次単位間の寄与と同じになって、比推定の影響を受けるのは一次単位内分散だけとなる。

実際の場合、Yが既知でXの推定値として $\bar{X}'$ の代りに $\bar{X}''$ を

用いうる際でも、X_{hijk} が少なくとも Y_{hijk} と平均値にかなり高い相関々係をもつことを確かめるのが無難である。ここで考えている相関は一次抽出単位内の副次層内のものである。この相関が低く、また副次抽出単位の大きさが相当大きく変動するなら、比推定の効率は単純総計の推定値よりもかなり低下するであろう。一方もし種々の副次層の大きさの尺度および一次抽出単位の大きさが実際の大きさの尺度に近いことが明らかであり、また大きさがあまり大きく変化しないように副次抽出単位を注意深く定義してあればこの二つの推定値は大体同じ効率を与えるものと思われる。

Ⅵ この論文で推奨した抽出原理の基礎となる母集団に屢々みられる自然的性質

現実の母集団の多くは次のような自然的な性質で特徴づけられる。

(i) 集落内の要素が指定された特性と正の相関をもつ

(ii) 多数の要素を含む集落は小数の要素を含む集落より内部の異質性が大きい

(iii) 集落の大きさが増すと相関をもった要素が導びかれる(たとえば人口または農業調査において、大きい集落は隣接地域の世帯または農場を含めることによって作られる)

この最初の性質は、抽出単位に大きい集落を用いたため効率に損失を生じたことを例証している多くの文献中で暗黙のうちに認められている。我々の経験では、オ2、オ3の性質は現実の母集団で全く普通に認められるもので、これは通例オ1の性質の成立する母集団と同じものである。

層内でこれらの自然的な性質が組合わせて現われると、この論文でこれまで扱ってきた次のような数学的な関係が導びかれる。

(a) 一次抽出単位の大きさ N_{hi} は単位内の級内相関 ρ_{hi} と負の相関をもつ

(b) N_{hi} と $N_{hi}\rho_{hi}$ は正の相関をもつ

(c) N_{hi} と σ_{hi}^2 は正の相関をもつ

(d) N_{hi} と σ_{hi}^2/N_{hi} は負の相関をもつ

この論文ではこれらの関係を用いていままで色々な方法間の選択を行なってきた。これらの関係はもちろん常に成立するものではないが、それらの[5]に示されている。これらの性質によって特徴づけられる母集団が屢々出現するということは、成立すると思われるこれらの性質や他の性質を一層効果的に利用する更に進んだ研究を正当化するものである。

Ⅶ この論文において記述した原理の現実の抽出問題に対する適用について

以下に大要を示す分析は、労働力および他の特性に対する国家月例標本(*monthly national sample*)の改訂において、色々な抽出手続から選択を行なうために実行されたものである。予算と管理との制限から、外業は国内に散在する限られた数の行政中心地で行なう必要があった。そうしてこれらから合衆国人口の 1/1,000 以下の標本を抽出すべきことが要求された。

(改訂すべき)もとの標本は一次抽出単位として1又2カウンティを用い、副次抽出単位として世帯または世帯の小集落を用いる普通の副次抽出計画であった。改訂された標本においては、それが管理上可能な限り、個々のカウンティより異質な一次単位を作るため隣接するカウンティを合併した。合衆国の3,000のカウンティから大体2,000個の一次抽出単位が作られた。カウンティの組合せ、一次層化、地域的副次層化、および大きさの尺度は1940年の *Decennial Census* のデータと利用することのできた最近のデータをもとにして決定された(5)。

---

(5) 用いた層化基準の概要を含む計画された改訂標本の完全な記述については(11)をみよ。この論文は指定された副次抽出方式の簡単な説明としても有用である。

この論文で提示した種々の原則の適用は、1940年と更に最近のデータにもとづいて層化された標本から1930年センサスにおける労働力の標識（characteristics）を推定することによって評価された。これは1930年から1940年にいたる10年間に起った大幅な変化のため幾つかの方法に特に厳格な検討になっている。

この節で大要を述べる分析は主として三つの抽出原則、すなわち、

(1) 拡大された一次単位（Sec. Ⅳ－1をみよ）

(2) 大きさの尺度に比例した確率を用いる一次単位の抽出（Sec. Ⅳ－2をみよ）

(3) 地域的副次層化（Sec. Ⅳ－3をみよ）

の導入に好適な状況のもとで獲得しうる利益を取扱うものである。

これとともに、他の標本推定公式を用いたときの効果を示すための幾つかの比較をおこなった。計算は最近労働月報に含まれている6つの主要項目について行なわれた。すなわち男子および女子労働者の総数、農業労働者の男子と女子の総数、および非農業労働者の男子と女子の総数である。他の色々な抽出方式との比較は一次層化の基準と標本内に抽出する人員の期待数の両方を一定にしておいて行なった。

下に与えた利得のパーセントは平均平方誤差に対する一次単位間の寄与（これには偏りの寄与を含む）の減少を示したものである。(6) 特に断らない限り使用した標本推定値は(ア)によるものである。

1 拡大した一次単位を導入することによってえられる利得

拡大された一次単位の使用による利得は、個々のカウンティを一次単位とする抽出計画の平均平方誤差と、カウンティの組

(6) 一次単位内分散の全平均平方誤差に対する寄与はすべての例において比較的小であり、また事実上色々な原則を導入しても変化しない。

合せを一次単位とする計画の平均平方誤差を比較して計算される。何れの計画においても一次単位は等確率で抽出し、地域的副次層化は用いなかった。この比較では、限られた数の層および上に与えた労働力の2つの項目、すなわち男子および女子労働者の総数についてのみ予備計算を完了した。拡大した一次単位を導入することによってえられた抽出誤差の減少は男女労働者総数に対しては4.8%、女子労働者総数については2.6%と推定された。

2 大きさの尺度に比例した確率を導入することによってえられる利得

確率を大きさの尺度に比例させる抽出原理の採用によってえられる利益は、単位を等確率で抽出する計画でえられる平均平方誤差と、その大きさに比例した確率で単位を抽出する計画でのものとを比較することによって計算される。どちらの計画でも一次単位はカウンティの組合わせで、いずれも地域的副次抽出は用いなかった。推定された利得のパーセントは次の通りである。

全労働者		農業労働者		非農業労働者	
男	女	男	女	男	女
5.0	8	27	6	19	21

この利益は抽出分散の減少と、抽出単位が等確率で抽出される場合に生ずる偏りが除去されることの両方を反映している。(7)

3 地域的副次層化を導入することによってえられる利益

地域的副次層化の原理を用いることによってえられる利益は、

(7) 前に述べたように上の比較を行なった2つの計画では推定値(ア)が用いられた。この推定値は一次単位をその大きさに比例した確率で抽出する計画では不偏であるが等確率で抽出する計画では偏りをもつ。しかし後者の計画では通例最良線型不偏推定値よりも偏りのある推定値の方がより有効である。6つの労働力の項目の場合、最良線型不偏推定値は偏りのある推定値の平均平方誤差の数倍の大きさの分散をもつ。

地域的副次層化を用いない計画でえられる平均平方誤差と、地域的副次層化を導入した計画のそれを比較すれば計算できる。これらのいずれの計画においても一次単位はカウンティの組合せであり、これらはその大きさの尺度に比例した確率で抽出された。推定された利益の%は次の通りである。

全労働者		農業労働者		非農業労働者	
男	女	男	女	男	女
6	31	46	51	32	22

4　上記の原理を単一の副次抽方式（指定された副次抽出方式）にすべて組入れた場合にえられる利益

三つの原理をすべて用いることによってえられる利益は、指定された副次抽出方式（これには三つの原理がすべて用いられている）に対する平均平方誤差と、これらの原理を全く用いない抽出方式の平均平方誤差を比較することによって計算される。指定された副次抽出方式においては、一次抽出単位はカウンティの組合せであり、これはその大きさの尺度に比例した確率をもって抽出され、また地域的副次層化が用いられた。もう一つの抽出方式では、一次単位は個々のカウンティで、抽出は等確率でありまた地域的副次層化は用いていない。この比較のための予備的な計算は6つの労働力の項目中の二つ、すなわち男子労働者の総数、女子労働者総数のみについて行なうことができた。推定された利益は、男子労働者については76%、女子労働者については53%であった。

最後の二つの原理、すなわち大きさに比例した確率での抽出と、地域的副次層化を同時に用いてえられる利益を知るための計算は6項目のすべてについて適用できる。これらの利益の大きさを知る方法は、両方の計画の一次単位がカウンティであることを除いては上述のものと同じである。推定された利益の%は次の通りである。

全労働者		農業労働者		非農業労働者	
男	女	男	女	男	女
54	37	88	54	45	39

指定された副次抽出方式と、いま比較を行なった一方の抽出方式はともに偏りをもつ計画であるが、しかし指定された方式での偏りは明らかに後者の偏りより小さい。たとえば全男子労働者数の推定に指定された抽出方式を用いたときの偏りは、真の全男子労働者数の1/2%より小さかったのに対して、同じ母集団特性の推定にもう一つの計画を用いたときの偏りは1/4%以上であった。

5　指定された抽出方式で用いる推定値の選択

指定された抽出方式に対して与えられた単純推定値(7)は、回帰の手法によって改善することができる（Sec.IV参照）。しかしそのような手法を用いるには多くの計算が必要だから、実際には利用できないことが多い。しかしSec.Vの最後に述べたように　もし全母集団の知識のようなある独立な情報が利用できれば、(12)の形の簡単な比の推定値を用いることによって(7)以上の利益の導びかれることが多い。一次抽出単位の大きさの尺度と実際の大きさとの相関がそう高くなく、これとともに実際の大きさが推定している特性の値と高度の相関をもつときには比の推定値の利用が特に望ましいものとなろう。1930年の全男子労働者数を推定するために示された労働力の項目について行なった小規模の抽出試験では、比の推定値(12)に対する一次単位間の分散、一次単位内の分散は何れも単純推定値(7)のものの約1/2であった。労働力の残りの5つの特性の推定では、比推定を用いても僅かの効果しかえられなかった。男子の全雇傭人口の分散が減少したのは、1930年以降の転居が1930年と1940年の大きさ（size）の間の相関を弱め、更に男子の労働者数が全母集団と高度の相関をもっていたことから生じたものである。他の5項目の分散では、他の項目に対する実際の

大きさとの相関がそれ程高くなかったから、このような減少はみられなかった。

6　最後にあたっての注意

いまえられた利益は、上に列挙した抽出原理を適用することによってえられたものである。これらの原理を適用できるような状況は好ましいものであるが、しかしこれは実際に屢々実現するものである。この原理の効果は、研究している母集団のもつ特定の属性によって左右される。拡大された一単位が小さい単位より更に内部的に異質なときは、常に拡大単位を用いることが望ましい。Sec. II で述べた一般的な母集団で単位の大きさが相当大きく変動しているときは、一次単位の抽出を大きさに比例した確率で行なうことが望ましい。地域的副次層化の利用は大きい一次抽出単位の用いられる抽出の場合に限られる。副次抽出を用い、一次単位は大きいが、大きさが変動しており、また標本内に含まれる一次単位の数が費用や管理上の条件で制限される場合には、上の三つの原理を同時に適用することによって最も大きい利益がえられる。Sec. II で説明したタイプの推定値は、この論文で指摘した以外の多数の具体的条件のもとで有効であろう。

謝　辞

多くの有益な注意を与えられ、またSec. III に要約した分析の計画と指導に当られた Harold Nisselson 氏に深く感謝するものである。また我々は原稿を通読され、多くの有益な注意を与えられた W. G. Cochran, W. Edwards Deming, L. R. Frankel, William G. Madow, および Frederick F. Stephan 氏に厚くお礼を申し上げる。この研究は J. C. Capt, Director and Philip M. Hauser, Assistant Director の一般的な指示のもとに、種々の標本調査の展開に関して Bureau of the Census（国勢調査局）において実行されたものである。

付　録

1　一次単位の合併が抽出分散に与える影響

標本平均を

$$\bar{X}_i' = \sum_{j}^{g} \sum_{k}^{n} X_{jk} / g \cdot n$$

とおく。ただし一次単位はもとの単位で X_{jk} は第 j 一次単位の第 k 要素の値、また g は標本内の一次単位の数、n は g 個の一次単位のおのおのから抽出された要素の数である。

$\bar{X}_i'$ の分散は

$$(14)\quad \sigma^2_{\bar{X}_i'} = \frac{N-n}{(N-1)ng}\,\sigma^2_{iw} + \frac{Q-g}{(Q-1)g}\,\sigma^2_{ib}$$

である。ここで Q は母集団におけるもとの一次単位の数、N はもとの一次単位内の要素の数で、$\sigma^2_{iw} = \sum\sum (X_{jk} - \bar{X}_j)^2/QN$ は $\bar{X}_j = \sum_k X_{jk}/N$ なるもとの一次単位内の分散である。また $\sigma^2_{ib} = \sum (\bar{X}_j - \bar{X})^2/Q$ は $\bar{X} = \sum \bar{X}_j/Q$ なるもとの一次単位間の分散である。

$$(15)\quad \sigma^2 = \sum\sum (X_{jk} - \bar{X})^2/QN = \sigma^2_{iw} + \sigma^2_{ib}$$

であるから、

$$(16)\quad \sigma^2_{ib} = \sigma^2\{1 + \rho_i(N-1)\}/N$$

ここで $\rho_i = \left[\sigma^2_{ib} - \frac{\sigma^2_{iw}}{N-1}\right]\frac{1}{\sigma^2}$ はもとの単位内の要素間の級内相関[2]である。

(15) と (16) から

$$(17)\quad \sigma^2_{iw} = \frac{N-1}{N}\,\sigma^2(1-\rho_i)$$

ゆえに

$$(18)\quad \sigma^2_{\bar{X}_i'} = \frac{N-n}{n}\cdot\frac{\sigma^2}{ng}(1-\rho_i) + \frac{Q-g}{(Q-1)}\cdot\frac{\sigma^2}{N}\{1+\rho_i(N-1)\}$$

同様に $\bar{X}_2'$ の分散は

$$(19)\quad \sigma^2_{\bar{X}_2'} = \frac{CN-n}{CN}\cdot\frac{\sigma^2}{ng}(1-\rho_2) + \frac{Q-gC}{(Q-c)g}\cdot\frac{\sigma^2}{CN}\{1+\rho_2(CN-1)\}$$

---

(2)　級内相関の定義および性質は R. A. Fisher の研究者のための統計的方法、および [5] をみよ。

ここで $\bar{X}'_2$ は拡大された一次単位に対する平均値で、ρ_2 は拡大された一次単位内の要素間の級内相関で、C は各拡大単位を作るために統合されたもとの単位数である。そうすると、

(20) $$\sigma^2_{\bar{x}'_1} - \sigma^2_{\bar{x}'_2} = \frac{\sigma^2}{gN}\left\{\frac{(g-1)(C-1)}{(Q-1)(Q-C)} + \rho_1 a_1 - \rho_2 a_2\right\}$$

ただし

$$a_1 = \frac{(Q-g)(N-1)}{Q-1} - \frac{N-n}{n}$$

および

$$a_2 = \frac{(Q-Cg)(CN-1)}{(Q-C)C} - \frac{CN-n}{Cn}$$

$$a_1 - a_2 = \frac{(C-1)(g-1)(QN-1)}{(Q-1)(Q-C)} \geqslant 0$$

かつ

$$\frac{(g-1)(C-1)}{(Q-1)(Q-C)} \geqslant 0$$

だから $\rho_1 > \rho_2$ なる限り拡大された一次単位を用いることによって利益がえられる。

ただし ρ_1 および ρ_2 はいずれも正である。

2　一次単位の大きさが等しくない場合の他の副次抽出方式の分散の比較

Sec. IV－2 で比較を行なった標本推定値の分散間の差の公式(21)の誘導を述べる。

我々は問題を限定して、標本に各層からただ一つの抽出単位が選ばれているという簡単な場合を考えることにする。

(21) $$\bar{X}' = \sum N_h \bar{X}'_h / N$$

は比較すべき三つの計画の各々について用いられる標本推定値とする。ここで

$$\bar{X}'_h = \bar{X}_{hj} = \sum_{\alpha}^{n_{hj}} X_{hj\alpha} / n_{hj}$$

であり、$X_{hj\alpha}$ は h 層の第 j 一次単位の第 α 要素の値である。L は層の数、n_{hj} は第 h 層の第 j 一次単位から標本内に抽出される要素数で N_{hj} はこれに対応する全要素数である。

$N_h = \sum_{j}^{Q_h} N_{hj}$ で Q_h は h 層内の一次単位の数、また $N = \sum_{h}^{L} N_h$ である。層内の副次抽出が前に述べた最初の副次抽出方式のように一定の割合 C なら、上の推定値の n_{hj} は CN_{hj} に等しい。もし推奨された抽出方式および前に述べた第2の抽出方式の場合のように、層内の副次抽出が一定数であれば、n_{hj} は

$$\bar{n}_h = C \sum_j N_{hj} / Q_h = C \bar{N}_h$$

に等しい。

我々は第1の計画に対する標本推定値を $\bar{X}'_1$ で、また第2の計画を $\bar{X}'_2$ で、また推奨された計画のものを $\bar{X}'_3$ と書くことにする。

最初の二つの計画の標本推定値の期待値 $\bar{X}'_1$ と $\bar{X}'_2$ は等しく、

$$E\bar{X}'_1 = E\bar{X}'_2 = \bar{\bar{X}} = \frac{1}{N}\sum_h \frac{N_h}{Q_h}\sum_j \frac{n_{hj}}{N_{hj}}\sum_{\alpha} \frac{X_{hj\alpha}}{n_{hj}}$$

$$= \frac{1}{N}\sum_h \frac{N_h}{Q_h}\sum_j \bar{X}_{hj}$$

である。ここで $\bar{X}_{hj} = \sum_{\alpha} X_{hj\alpha} / N_{hj}$ したがって一般に $\bar{\bar{X}}$ は $\sum_{h,j} X_{hj\alpha} / \sum_{h,j} N_{hj} = \bar{\bar{X}}$ だから $\bar{X}'_1$ および $\bar{X}'_2$ は $\bar{\bar{X}}$ の偏りのある推定値である。

一次単位がその大きさに比例した確率で抽出され、また層内から一定個数の単位をとる推奨された計画では、標本推定値の期待値は、

(22) $$E\bar{X}'_3 = \frac{1}{N}\sum_h \sum_j N_h \frac{N_{hj}}{N_h}\frac{n_h}{N_{hj}}\frac{X_{hj}}{\bar{X}_h} = \frac{1}{N}\sum_h \sum_j X_{hj} = \bar{\bar{X}}$$

だから、推奨された計画に対する推定値は不偏である。

$\bar{X}'_1$ の平均平方誤差は、

(23) $$\sigma^2_{\bar{x}'_1} = \frac{1}{N^2}\sum_h \frac{N_h^2}{Q_h}\left[\sum_j \frac{N_{hj}-n_{hj}}{(N_{hj}-1)n_{hj}}\sigma^2_{hj} + \sum_j (\bar{X}_{hj} - \bar{X}_h)^2\right]$$

$$+ (\bar{\bar{X}} - \bar{X})^2 - \frac{1}{N^2}\sum_h N_h^2 (\bar{X}_h - \bar{X}_h)^2$$

である。ここで $\sigma_{hj}^2 = \sum_k (X_{hjk} - \bar{X}_{hj})^2 / N_{hj}$ は第 h 層の，第 j 一次抽出単位内の要素間分散である。

(23) の大括弧内の第1項は一次単位内分散の寄与であり，第2項は近似的に一次単位間の平均平方誤差を表わし，残りの項のこの近似の誤差を与える。$\bar{X}_2'$ の平均平方誤差もこれと同じ式で表わせるが，ただそのときには n_{hj} を n_h でおきかえねばならない。

$\sigma_{\bar{X}_1}^2$ と $\sigma_{\bar{X}_2}^2$ の差は，

$$(24)\quad \sigma_{\bar{X}_1}^2 - \sigma_{\bar{X}_2}^2 = \frac{1}{CN^2} \sum_h \frac{N_h^2}{Q_h} \sum_j \sigma_{hj}^2 \frac{N_{hj}}{N_{hj}-1} \left(\frac{1}{N_{hj}} - \frac{1}{\bar{N}_h} \right)$$

である。

これは実際の場合には殆んどいつもそうであるように，σ_{hj}^2/N_{hj} が N_{hj} と負の相関をもつときには正となる（Sec. VI をみよ）。したがって普通 $\sigma_{\bar{X}_1}^2$ は $\sigma_{\bar{X}_2}^2$ より大きいから，推奨された抽出方式が，上述の最初の二つの計画のいずれより有効（efficient）であることを示すには $\sigma_{\bar{X}_2}^2$ と $\sigma_{\bar{X}_3}^2$ を比較すれば十分である。

推奨された計画の分散は，

$$(25)\quad \sigma_{\bar{X}_3}^2 = \frac{1}{N^2} \sum_h N_h^2 \left[\sum_j \frac{N_{hj}}{N_h} \frac{N_{hj} - \bar{n}_h}{N_{hj}-1} \frac{\sigma_{hj}^2}{\bar{n}_h} + \sum_j \frac{N_{hj}}{N_h} (\bar{X}_{hj} - \bar{X}_h)^2 \right]$$

である。$\bar{X}_2'$ の平均平方誤差を $\bar{X}_3'$ の分散と比較するため，第 j 一次単位内の要素間の級内相関を

$$\rho_{hj} = \frac{1}{\sigma_h^2} \left[(\bar{X}_{hj} - \bar{X}_h)^2 - \frac{\sigma_{hj}^2}{N_{hj}-1} \right]$$

と定義する。ここで σ_h^2 は第 h 層内のすべての要素間の分散である。(23) の大括弧の外の項は平均平方誤差への寄与が正か，または無視できる位の大きさなのでこの比較では考慮に入れなかった。このとき，

$$(26)\quad \sigma_{\bar{X}_2}^2 - \sigma_{\bar{X}_3}^2 = \frac{1}{N^2} \sum_h \frac{N_h^2}{Q_h} \left\{ \sum_j \frac{N_{hj}}{N_{hj}-1} \frac{\sigma_{hj}^2}{\bar{n}_h} \left(1 - \frac{N_{hj}}{N_h} \right) + \sigma_h^2 \sum_j \rho_{hj} \left(1 - \frac{N_{hj}}{N_h} \right) \right\}$$

この差の第2項は Sec. IV-2 で，第1項を無視したときの近似的な差として与えたものである。この項の相対的な大きさを調べるため，

$$(27)\quad \frac{N_{hj}}{N_{hj}-1} \sigma_{hj}^2 = \sigma_h^2 (1 - \delta_{hj})$$

とおく。このとき，

$$(28)\quad \sigma_{\bar{X}_2}^2 - \sigma_{\bar{X}_3}^2 = \frac{1}{N^2} \sum_h \frac{N_h^2}{Q_h} \sigma_h^2 \left\{ \frac{1}{\bar{n}_h} \sum_j \delta_{hj} \left(\frac{N_{hj}}{N_h} - 1 \right) - \sum_j \rho_{hj} \left(\frac{N_{hj}}{N_h} - 1 \right) \right\}$$

Sec. VI で示した一般的な母集団のクラスでは δ_{hj} と N_{hj} との間の共分散および ρ_{hj} と N_{hj} との共分散も負になる。更にこのクラスに属する母集団での多くの実際的問題では，この二つの共分散の大きさは近似的に等しくなる。そのような場合，(27) の第1項は第2項の $\frac{1}{\bar{n}_h}$ 倍となるから，$\bar{n}_h > 1$ なるときは常に第2項より小となり，また十分大きい $\bar{n}_h$ の値については更に小となる。たとえば，一定の大きさの集落で特定の性質をもつ要素の条件付確率がすべての大きさの集落について等しい場合，このような種々の大きさの集落からなる母集団ではこの二つの共分散の大きさは非常に近い値となる。実際問題は近似的にこのような状況にあることが多い。更に δ_{hj} と N_{hj} の共分散が ρ_{hj} と N_{hj} とのものの数倍たとえば5倍になっている場合でも，第2項は $\bar{n}_h > 5$ なる値では常に第1項より大きくなる。

推奨された抽出方式の使用によってえられる利益を Sec. VII で幾つかあげておいた。そこで結果を示した若干の項目については，この利益はかなり大きいものであった。

3 分散公式 (17) および (7) の誘導

確率変数の比の平均平方誤差は一般に Taylor 展開で近似される。

x' と y' が確率変数で $Y' > 0$，また $x'/y' = r'$ によって推定され

各母集団特性を Y とする。

このとき

(29) $$E\left\{\frac{X'}{Y'}-r\right\}^2 = E\frac{Y^2}{(EY')^2}\left(\frac{X'}{Y'}-r\right)^2+E\left(1-\frac{Y'}{(EY')^2}\right)^2\left(\frac{X'}{Y'}-r\right)^2$$

が成立する。(29)の右辺の第1項は Taylor 展開で与えられる平均平方誤差の第1近似で第2項はこの近似の誤差項である。

等式(13)および特別な場合として(9)は次のようにして導びかれる。

(30) $$E(r'-r)^2 = E\left\{\frac{\sum_h^L \frac{1}{t_h}\sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} X_{hijk}}{\sum_h^L \frac{1}{t_h}\sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} Y_{hijk}} - r\right\}^2$$

$\psi_{hijk} = Y_{hijk}(r_{hijk}-r)$ および

$$Y' = \sum_h^L \frac{1}{t_h}\sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} Y_{hijk}$$

とする。このとき、

(31) $$\theta = \sum_h^L \frac{1}{t_h}\sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} \psi_{hijk} \Big/ E\left(\sum_h^L \frac{1}{t_h}\sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} Y_{hijk}\right)$$
$$= \frac{Y'}{EY'}\left(\frac{X'}{Y'}-r\right)$$

とおくと

$$E\theta^2 = EY'^{-2}(r'-r)^2/(EY')^2$$

は平均平方誤差の第一近似である。EY' は EX'(8) と同じ方法で計算できるから、$E\theta^2$ の分子 $EY'^{-2}(r'-r)^2$ を求めればよい。いま

$$EY'^2(r'-r)^2 = E\left[\sum_h \frac{1}{t_h}\sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} \psi_{hijk}\right]^2$$
$$= E\sum_h \frac{1}{t_h^2}\psi_h'^2 + E\sum_{\substack{h,g\\ h\neq g}} \frac{\psi_h'}{t_h}\frac{\psi_g'}{t_g}$$

である。ここで

$$\psi_h' = \sum_i^{S_h}\sum_j^{1}\sum_k^{m_{hij}} \psi_{hijk} = \sum_i^{S_h} \psi_{hi}' \quad .$$

$$E\sum \frac{1}{t_h^2}\psi_h'^2 = E\sum_h\sum_i \psi_{hi}'^2/t_h^2 + E\sum_h \sum_{\substack{i,r\\ i\neq r}} \psi_{hi}'\psi_{hr}'/t_h^2$$

であるから、

(32) $$EY'^2(r'-r)^2 = E\sum_h \frac{1}{t_h^2}\sum \psi_{hi}'^2 + E\sum_h \frac{1}{t_h^2}\sum_{\substack{i,r\\ i\neq r}} \psi_{hi}'\psi_{hr}' + E\sum_{\substack{h,g\\ h\neq g}} \frac{\psi_h'}{t_h}\frac{\psi_g'}{t_g}$$

である。

(32)の右辺の第1項は、

(33) $$E\sum_{h,i} \psi_{hi}'^2 \frac{1}{t_h^2} = \sum_{h,i,j} \frac{1}{t_h^2}\frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}\frac{M_{hij}-m_{hij}}{M_{hij}-1}\sum_k \psi_{hijk}^2$$
$$+ \sum_{h,i,j} \frac{1}{t_h^2}\frac{P_{hj}^2}{P_h}\frac{m_{hij}}{M_{hij}}\frac{m_{hij}-1}{M_{hij}-1}\left(\sum_k \psi_{hijk}\right)^2$$

(32)の第2項は、

(34) $$E\sum_h \frac{1}{t_h^2}\sum_{\substack{i,r\\ i\neq r}} \psi_{hi}'\psi_{hr}' = \sum_{h,j} \frac{1}{t_h^2}\left(\sum_i \frac{m_{hij}}{M_{hij}}\psi_{hij}\right)^2$$
$$- \sum_{h,j} \frac{1}{t_h^2}\frac{P_{hj}^2}{P_h}\sum_i \frac{m_{hij}}{M_{hij}}\psi_{hij}^2$$

である。ただし $\psi_{hij} = \sum_k \psi_{hijk}$ である。また(32)の第3項は、

(35) $$E\sum_{\substack{h,g\\ h\neq g}} \frac{\psi_h'}{t_h}\frac{\psi_g'}{t_g} = \left[\sum_{h,j} \frac{1}{t_h}\frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}\psi_{hij}\right]^2$$
$$- \sum \frac{1}{t_h^2}\left[\sum_i \frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}\psi_{hij}\right]^2$$

である。したがって

$$EY'^2(r'-r)^2 = (33)+(34)+(35)$$

から、ψ_{hij} の代りに $Y_{hijk}(r_{hijk}-r)$ を代入すると、

(36) $$EY'^2(r'-r)^2 = \sum \frac{1}{t_h^2}\left[\sum_{i,j} \frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}\frac{M_{hij}-m_{hij}}{M_{hij}-1}\sum Y_{hijk}^2(r_{hijk}-r)^2\right.$$
$$+ \sum_{i,j} \frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}\frac{m_{hij}-1}{M_{hij}-1}\left[\sum_k Y_{hijk}(r_{hijk}-r)\right]^2$$
$$\left.+ \sum_j \frac{P_{hj}}{P_h}\left[\sum_i \frac{m_{hij}}{M_{hij}} Y_{hij}(r_{hij}-r)\right]^2 - \sum_{i,j} \frac{P_{hj}}{P_h}\frac{m_{hij}^2}{M_{hij}^2}\right.$$

$$-Y_{hij}^2(\bar{Y}_{hij}-\bar{Y})^2-\Big\{\sum_{ij}\frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}(\bar{Y}_{hij}-\bar{Y})Y_{hij}\Big\}^2\Big]$$

$$+\Big[\sum_{hij}\frac{1}{t_h}\frac{P_{hj}}{P_h}\frac{m_{hij}}{M_{hij}}Y_{hij}(\bar{Y}_{hij}-\bar{Y})\Big]^2$$

となる。

(36)の第1項の $(\bar{Y}_{hijk}-\bar{Y})^2$ に $(\bar{Y}_{hijk}-\bar{Y}_{hij}+\bar{Y}_{hij}-\bar{Y})^2$ を代入し、第1、第2、第4項の $1/t_h$ を $P_{hi}M_{hij}/P_{hij}m_{hij}$ でおきかえるとこれら三つの項は、

(37) $$\sum_{hijk}\frac{P_{hj}}{P_h}\frac{M_{hij}}{m_{hij}}\frac{P_{hi}^2}{P_{hij}}F_{hij}Y_{hijk}^2(\bar{Y}_{hijk}-\bar{Y}_{hij})^2$$

$$+2\sum_{hijk}\frac{P_{hj}}{P_h}\frac{M_{hij}}{m_{hij}}\frac{P_{hi}^2}{P_{hij}}F_{hij}Y_{hijk}^2(\bar{Y}_{hijk}-\bar{Y}_{hij})(\bar{Y}_{hij}-\bar{Y})$$

$$+\sum_{hij}\frac{P_{hj}}{P_h}\frac{M_{hij}}{m_{hij}}\frac{P_{hi}^2}{P_{hij}^2}F_{hij}(\bar{Y}_{hij}-\bar{Y})^2\Big[\sum_k\bar{Y}_{hijk}^2-\frac{Y_{hij}^2}{M_{hij}}\Big]$$

ここで $F_{hij}=(M_{hij}-m_{hij})/(M_{hij}-1)$ で $\bar{Y}_{hij}=\sum_k X_{hijk}/\sum_k Y_{hijk}$ である。

(36)の第3、第5、第4項の $\frac{1}{t_h}$ に適当な値を代入すると、これらの項の和は、

(38) $$\sum_{hj}\frac{P_{hj}}{P_h}\Big[\sum_i\frac{P_{hi}}{P_{hij}}Y_{hij}(\bar{Y}_{hij}-\bar{Y})\Big]^2-\sum_h\Big[\sum_{ij}\frac{P_{hj}}{P_h}\frac{P_{hi}}{P_{hij}}Y_{hij}(\bar{Y}_{hij}-\bar{Y})\Big]^2$$

$$+\Big[\sum_{hij}\frac{P_{hj}}{P_h}\frac{P_{hi}}{P_{hij}}Y_{hij}(\bar{Y}_{hij}-\bar{Y})\Big]^2$$

となる。

いま、

(39) $$\sum_i\frac{P_{hi}}{P_{hij}}Y_{hij}(\bar{Y}_{hij}-\bar{Y})=\sum_i P_{hi}\Big(\frac{X_{hij}}{P_{hij}}-\frac{Y_{hij}}{P_{hij}}\bar{Y}\Big)$$

$$=P_h(R_{hj(A)}-\bar{Y}R_{hj(A):Y})$$

$$=P_hR_{hj(A):Y}(\bar{\bar{Y}}_{hj(A)}-\bar{Y})$$

ただし、$\bar{\bar{Y}}_{hj(A)}=R_{hj(A)}/R_{hj(A):Y}$

また、

(40) $$\sum_{ij}\frac{P_{hj}}{P_h}\frac{P_{hi}}{P_{hij}}Y_{hij}(\bar{Y}_{hij}-\bar{Y})=\sum_j P_{hj}(R_{hj(A)}-\bar{Y}R_{hj(A):Y})$$

$$=P_h(R_{h(A)}-\bar{Y}R_{h(A):Y})$$

$$=P_hR_{h(A):Y}(\bar{\bar{Y}}_{h(A)}-\bar{Y})$$

ただし $\bar{\bar{Y}}_{h(A)}=R_{h(A)}/R_{h(A):Y}$ である。

(38)に(39)と(40)を代入すると、

(41) $$\sum_{hj}(P_{hj}/P_h)^2R_{hj(A):Y}^2(\bar{\bar{Y}}_{hj(A)}-\bar{Y})^2-\sum_h P_h^2R_{h(A):Y}^2(\bar{\bar{Y}}_{h(A)}-\bar{Y})^2$$

$$+\Big[\sum_h P_hR_{h(A):Y}(\bar{\bar{Y}}_{h(A)}-\bar{\bar{Y}})\Big]^2$$

(41)の第1項の $(\bar{\bar{Y}}_{hj(A)}-\bar{Y})^2$ の代りに $(\bar{\bar{Y}}_{hj}-\bar{\bar{Y}}_{h(A)}+\bar{\bar{Y}}_{h(A)}-\bar{Y})^2$ を代入して展開すると、(41)は、

(42) $$\sum_{hj}P_h^2\frac{P_{hj}}{P_h}R_{hj(A):Y}^2(\bar{\bar{Y}}_{hj(A)}-\bar{\bar{Y}}_{h(A)})^2$$

$$+2\sum_{hj}P_h^2\frac{P_{hj}}{P_h}R_{hj(A):Y}^2(\bar{\bar{Y}}_{hj(A)}-\bar{\bar{Y}}_{h(A)})(\bar{\bar{Y}}_{h(A)}-\bar{Y})$$

$$+\sum_h P_h^2(\bar{\bar{Y}}_{h(A)}-\bar{Y})^2\Big[\sum_j\frac{P_{hj}}{P_h}R_{hj(A):Y}^2-\bar{R}_{h(A):Y}^2\Big]$$

$$+\Big[\sum_h P_hR_{h(A):Y}(\bar{\bar{Y}}_{h(A)}-\bar{Y})\Big]^2$$

となる。

仲えに $(EY')^2E\theta^2=$(37)+(42) だから、

(43) $$(EY')^2E\theta^2=\sum_{hij}F_{hi}^2\frac{P_{hj}}{P_h}\frac{M_{hij}-m_{hij}}{M_{hij}-1}\frac{\sum_k Y_{hijk}^2(\bar{Y}_{hijk}-\bar{Y}_{hij})^2}{m_{hij}M_{hij}\bar{P}_{hij}^2}$$

$$+2\sum_{hij}P_{hi}^2\frac{P_{hj}}{P_h}\frac{M_{hij}-m_{hij}}{M_{hij}-1}\frac{\sum_k Y_{hijk}^2(\bar{Y}_{hijk}-\bar{Y}_{hij})(\bar{Y}_{hij}-\bar{Y})}{m_{hij}M_{hij}\bar{P}_{hij}^2}$$

$$+\sum_{hij}P_{hi}^2\frac{P_{hj}}{P_h}\frac{M_{hij}-m_{hij}}{M_{hij}-1}\sigma_{hij:Y}^2\frac{(\bar{Y}_{hij}-\bar{Y})^2}{m_{hij}\bar{P}_{hij}^2}$$

$$+\sum_{hj}P_h^2(P_{hi}/P_h)^2R_{hj(A):Y}^2(\bar{\bar{Y}}_{hj(A)}-\bar{\bar{Y}}_{h(A)})^2$$

$$+2\sum P_h^2(P_{hj}/P_h)R_{hj(A):Y}^2(\bar{\bar{Y}}_{hj(A)}-\bar{\bar{Y}}_{h(A)})(\bar{\bar{Y}}_{h(A)}-\bar{Y})$$

$$+\sum_{hj}P_h^2(\bar{\bar{Y}}_{h(A)}-\bar{Y})^2(R_{hj}/P_h)(R_{hj(A):Y}-R_{h(A):Y})^2$$

$$+\sum_k P_k R_{k(A):Y}(\bar{r}_{k(A)}-r)\Big]^2$$

ここで $\sigma^2_{hij:Y}=\sum_{k}^{M_{hij}}(Y_{hijk}-\bar{Y}_{hij})^2/M_{hij}$ かつ

$$\bar{Y}_{hij}=\sum_{k}^{M_{hij}}Y_{hijk}/M_{hij}=Y_{hij}/M_{hij}$$

$E(r'-r)$ に対する近似式は (43) を $(EY')^2$ で割ってえられる。第2, 第5, 第7項は殆んどの母集団で無視できる大きさだから、これを無視すると (13) がえられる。

x' の分散は (43) で単に Y_{hijk} のところに $\bar{P}_{hij}/P$ を代入すればえられる。これは下に述べる考察から導びかれるものである。

$r'=x'/Y'$ で、x' は r' の分子だから、分母 Y' が、繰返し抽出 (repeated sampling) で恒等的に1に等しい場合、$\sigma^2_{x'}$ は $\sigma^2_{r'}$ で与えられる。

(5) から

$$\frac{1}{t_k}=\frac{M_{hij}P_{hi}}{m_{hij}P_{hij}}=\frac{P_{hi}}{m_{hij}\bar{P}_{hij}}$$

だから $\sum_k^{L}\sum_r^{S_h}\sum_j^{I}\sum_k^{m_{hij}}\frac{P_{hi}}{m_{hij}P_{hij}}Y_{hijk}$ に等しい r' の分母は、Y_{hijk} が $\bar{P}_{hij}/P$ に等しくとられている場合、繰返えし抽出 (repeated sampling) において恒等的に1に等しい。ただし、$P=\sum P_k$ である。

x' の平均平方誤差に対する公式 (9) はもちろん誤差項が

$$E\{Y'^2/(EY')^2\}\{r'-r\}^2=0$$

だから正確なものである。x' は確率変数の比から推定されるのでないから、$\sigma^2_{x'}$ は (27) を使わずに直接もっと簡単な方法で求めることができる。

(29) から $E(r'-r)^2$ に対する近似式の誤差項 (43)/$(EY')^2$ は $E(1-\frac{Y'^2}{(EY')^2})\times\{r'+r\}^2$ で与えられる。これは個々の観測値の簡単な函数としては表わせないが、しかし求めうる最大および最小値を与えるものとして有用である。

r' の分散の上限、下限を求める方法は、x' と Y' の同時分布 (joint distribution) とかかわりなく成立する次の不等式から容易にえられる。

$$(44)\quad \frac{EY'^2}{Y^2_{max}}(r'-r)^2\leqslant E(r'-r)^2\leqslant\frac{EY'^2}{Y^2_{min}}(r'-r)^2$$

ここで Y_{max} は簡単に各層の最大の Y'_k を選ぶか、あるいは推定するかしてえられる。Y_{min} (Y' の最小値) についても同様である。(44) を計算すると、

$$(45)\quad \frac{(EY')^2E\theta^2}{Y^2_{max}}\leqslant E(r'-r)\leqslant\frac{(EY')^2E\theta^2}{Y^2_{min}}$$

となる。ここで $(EY')^2E\theta^2$ は (43) で与えられる。(45) は層内の Y の変動性が限られているような抽出方式では、$E\theta^2$ の正確さの評定すべき指標となるであろう。しかし層化を用いず、Y の変動性が制限されていないような他の計画では、(45) の与える限界では広すぎて役にたたないであろう。

M. H. Hansen and W. N. Hurwitz
(A. M. S. Vol. 14, 1943 pp 333～362)

2　系統的抽出の理論について（I）

On the theory of Systematic Sampling

1　緒　言

標本設計理論の必要性は説明するまでもない。政府ならびに私企業のいづれにおいても、政策ないし行動（*operating*）の決定は標本の知識に頼っている。政府や企業においては益々標本調査の理論を利用する傾向がある。(1)

不幸なことに標本調査理論と実際との間にはまだかなりの隔りがある。これらの原因は、一方では標本調査理論の与えうる実際的寄与に関係のある行政官（管理者）の無知があり、また一方では有用なサンプリング設計の価値を認めさせるような抽出理論が欠けているからでもある。

今迄にもまた現在でも、理論と実際との一致を計る多くの努力が払われている。(2) 管理者と抽出者（*Sampler*）は互いにうまく調整し合っている。しかし経験によってその理論が発展すれば今までに理論の組立てられているどのデザインよりも優れていることが証明されると思われるにもかかわらず、理論の解明されていないデザインが依然として存在するのである。おそらく今日における抽出理論の主要な欠陥は、N個の要素の母集団から、n要素からなる完全に無作為な標本をとった方が良いか、あるいは、i要素から始めて、i, $i+k$, $i+2k$, ……, $i+(n-1)k$ なる要素を含む系統的標本（出発点iは無作為に選びNは大体knになるようにする）をとる方がよいかをきめる統計的手段のないことである。(3) 理論のない系統的抽出法と系統的抽出より悪い結果の出やすい無作為抽出法との二者択一を迫られていることによって統計家のおちいっているジレンマは系統的標本をとるか無作為標本をとるかというこの問題に関係するものである。

この論文の目的は系統的抽出の満足な理論を与えることによってこの不一致を解決しようとするものである。次節において系統的抽出理論の我々の研究の最初の部分を提出する。この研究が、単一要素の抽出理論と、要素の集落（*cluster*）の抽出理論の双方にわたるものであるとしても、この論文での単位は要素の集落でなく単一の要素でなければならない。後者の問題はあとの論文で取扱う。我々は層化しない母集団と層化した母集団の双方から与えられた系統的抽出の理論を与えることにする。推定値の平均値および分散に対する公式を導いた。無作為抽出計画と層化無作為抽出計画との比較を行なった。更に分散の推定値と標本の"最適"な大きさおよび配分（*allocation*）を導いた。分析の基礎的な部分は、母集団の分散や、系列相関あるいは *serial variance* の知識から系統的標本にもとづく推定値の分散が推定できるという

---

(1)　標本抽出理論の知識を有する統計家の必要性は現在数ヶ所の大学において与えられているサンプリングの課程を生み出すことになった。

(2)　応用の分野での研究だけでなく、標本抽出技術について助言を与えるということをも含めた責任の、近年における立場の変化を考えるだけで十分である。

(3)　この論文では$N=kn$と仮定する。この仮定をはずしても一般性という点ではあまり得るところはないが、その場合にはかなり細かい論議が必要となるであろう。Nが丁度knでない場合その出発点がすべて等確率で選ばれるような系統的抽出の手続きは偏りをもっている。しかしこの偏りは普通問題にする必要のない程小さい。もしNが既知なら系統的標本の可能な大きさに比例させた抽出を行なうことによってこの偏りを除去できることが知られている。

(4)　我々が系統的抽出手続を定義するとき、ある系統的抽出手続はN個の中からn個をとるC_n^N個の選び方の中の多数を除外するような無作為抽出手続である。

との説明にある。[5] 基本的な結果は

a. 系列相関次正の和（positive sum）なら系統的標本は無作為標本より悪い。

b. 系列相関の和が近似的に0なら系統的標本は大体無作為標本と同等で

c. 系列相関が負の和（negative sum）のときには系統的標本は無作為標本より良い

2 有限母集団の利用

この論文では期待値の計算の際、極限分布が使える程母集団が大きい場合でも有限母集団から抽出を行なったと仮定する。これは数学的には扱々しい方の問題である。同様な結果は正しく定義された多次重正規分布を仮定し、条件付確率を用いてもうることができよう。しかし現実的な観点からは、有限母集団の利用を導く幾つかの因子が存在する。我々が最も普通に抽出を行なうのは、変換の法則（law of transformation）が知られていないか、あるいは数学的に表示されないような実在する（existing）母集団である。[6] 結局我々が抽出を行なう正規とか、その他の特定の分布の概念および条件付確率の利用ということは、実際問題に関する我々の思考の対象外のものである。

一方我々が母集団は有限でなければならないと考えて、有限母集団から乱数表を用いて標本を抽出するなら、我々は現実的な問題において漠然とした形の数学を用いているにすぎないことになる。更に乱数を選ぶことによって繰返し可能な実験を行なうことができるが、これは統計的管理の状態にあるというこんな我々は

(5) 特に限定しないで"母集団の分散"というときは、母集団のN要素からなる無作為標本の分散を意味するものとする。

(6) 言い換えれば、我々の母集団は時間の上で統計的なコントロールを受けていない状態にある。

知っている。

普通の無作為抽出の理論では、色々な標本平均値を求めるための可能な標本の数は十分大きいから、十分大きい母集団および標本での標本平均値は、近似的に正規分布すると仮定できる。しかし系統的抽出においては、可能な標本平均値の数は普通非常に小さく、母集団や標本が大きい場合でもそれが正規分布すると仮定することは難かしい。結局系統的抽出の平均値および分散に対する我々の解釈においては、母集団要素は確率変数の単一観測値の結果であり、その分布は要素毎に変動するという考えをとらざるを得ない。したがって我々が次に行なう解釈は条件付確率についてのものとなるが、もし母集団と標本が十分大きいなら、我々は可能な各標本平均値の算術平均は正規分布に従うという仮定を設けることができる。

適当な正規多変量母集団の仮定のもとにおける系統的抽出の理論は後に述べるであろう。

3 定義

N個の要素 $x_1, \cdots, x_N$ からなる有限母集団から標本抽出を行なうと仮定する。標本設計とは、これらのN個の要素を（重なることもあり重ならないこともあるが）n個のクラスに分割する方法と、各クラスが定められた確率で抽出されるようにして、n個の中から1つのクラスを抽き出す方法を組合せた手続きを意味する。与えられた標本設計（sample design）と組合せた抽出手続は、標本設計において定められた方法に従って、n個のクラスから1つを選び出す作業である。標本は抽出手続によって定められた特別なクラスである。

無作為抽出手続とは、標本設計でn個のクラスができたら、これらのどの1つを抽出する確率も $1/n$ となるようにする抽出方法である。

無作為抽出手続と組合わされる任意の標本設計は無作為抽出設

計（random sampling design）である。用いられている無作為でない抽出手続の1つは、その中のクラスに大きさ（size）と数を結びつけ、与えられたクラスの標本となる確率がその大きさに等しくするような手続である。(7) 勿論その他の無作為でない抽出手続も用いられている。

N個の要素のうちn要素を選ぶ単純（無制限）無作為抽出は、Nからnを選ぶ可能なC_n^N通りのクラスがあって、その各々の標本となる確率が$1/C_n^N$になっているような抽出設計である。これと結合される抽出手続は、各クラスを番号し、$i=1, \cdots\cdots, C_n^N$と同一視し、乱数表を用いて番号iを選ぶものである。同様に無作為抽出手続はN個の要素を$j=1, \cdots\cdots, N$の番号と同じと考えて、乱数表から番号jを選び次に乱数表から別の番号jを選ぶ。このようにして1～Nの中からn個の数が繰返えしなしに乱数表から抽かれるまでこの手続を続けることでもある。これらの整数と結合された要素は無作為標本である。この二つの手続が同等であることは容易にわかる。無制限でない無作為抽出手続は制限付（restricted）抽出計画とよばれる。制限付無作為抽出計画には色々な種類があるが、その中で我々が系統的抽出とよぶものは唯一つしかない。

系統的抽出計画とはN個の要素をn個のクラス、S_1, S_2 …… S_nに分割し、S_iの一つを選ぶのに無作為抽出手続を用いる計画である。ここでS_iはx_i, x_{i+n}, …… $x_{i+(n-1)n}$である。したがって系統的抽出は一種の集落抽出計画（cluster sampling design）であることが容易にわかる。すぐわかるように、系統的抽出計画において導入された集落抽出の新らしい側面は、級内相関係数の値および、標本の大きさが変るときのこの値の変化を求めるために母集団における要素の順序の知識が用いられるということである。

無作為抽出と無作為でない抽出手続の場合と同じように、抽出計画には無作為と系統的抽出計画の組合せたものが含まれることもある。

これらの標本を抽出した母集団は層化されていることもあり層化されていないこともある。また抽出単位も単一の要素のときもあるし要素の集落（cluster）であることもある。

(7) この問題の論議については "On the theory of sampling from finite populations" なる題名の（M. H. Hansen and W. N. Hurwitz, A. M. S. Vol. 14, 1943）をみよ。

4 抽出計画を選ぶ際の基礎

望ましい推定値をうるために構成できる多くの抽出計画の中から、人は管理上の（行政的な）考慮、費用および抽出誤差等をもとにしてどれを用いるかをきめるであろう。抽出誤差の尺度としては極限分布の理論や最良線型不偏推定値の理論をもとにして、推定された特性に対する標本推定量の標準偏差を用いるのが普通になっている。この論文では以後このやり方を用いるのであるが、ここで注意すべきことは多くの抽出計画を組立てた場合、そのうちの或るものについては極限分布の理論が成立しないので、標準誤差を用いること自体が分析の結果以上の問題となるおそれがあるということである。この危険は、系統的抽出計画で生ずるものであるかこれについては後に研究することにする。

偏り（bias）、一致性（consistency）および効率（efficiency）などは用いられた抽出計画および推定函数の性質であって、実際にえられた特定の標本の性質でないということを注意しなければならない。標本にもとずくいかなる推定値も、おそらく推定された特性とは違ってくるであろう。このような食い違いの大きさがどれ位かということを示すのが統計解析の役目である。

5 記　号

記号Pに適当な添字をつけて母集団および層などの副次母集団

を表わすことにする。

層の数を L と書き　また i 層内の要素の数を N_i で表わす。標本の大きさは n に適当な添字をつけて表わす。

抽出計画によって定義されたときの母集団のある特別なクラスは S に添字をつけて示す。

6　層化しない系統的抽出（抽出単位は一つの要素からなる）

この節で用いた添字のとる値の範囲については附録Aに示してある。

母集団 P は N 個の要素 $x_1, \cdots\cdots, x_N$ からなっているとする。P の算術平均 $\bar{x}$ の推定値を求めたい。

$N = kn$ とし，クラス S_i が n 個の要素 $x_i, x_{i+k}, \cdots\cdots, x_{i+(n-1)k}$ からなるものとする。したがって，大きさ n の標本から $\bar{x}$ を推定するための系統的抽出計画には k 個のクラス $S_1, S_2, \cdots\cdots, S_k$ があり　また S_i がこの抽出手続で選ばれる確率が $1/k$ とならなければならないという要求がある。

$\bar{x}_i$ は S_i の要素の算術平均値，すなわち $n\bar{x}_i = x_i + x_{i+k} + \cdots + x_{i+(n-1)k}$ とし，また $\hat{\bar{x}}$ は標本平均値，すなわちもし S_i が抽出手続きで抽出されるなら

$$\hat{\bar{x}} = \bar{x}_i$$

である。

系統的抽出を取扱う場合，我々は屢々系列相関および附随する *serial variance* の定義として *circular* および非-*circular* な定義の両方を用いることにする。

我々は，$h > kn$ なら　$x_h = x_{h-kn}$　と仮定する。これは *circular* な定義に用いる。

$$kn\sigma^2 = \sum_{\nu}(x_\nu - \bar{x})^2$$

および

$$kn C_{h\mu} = \sum_{\nu}(x_\nu - \bar{x})(x_{\nu+h\mu} - \bar{x})$$

とする。

このとき系列相関係数 $\rho_{h\mu}$ の *circular* な定義は

$$\sigma^2 \rho_{h\mu} = C_{h\mu}$$

である。この式は $\sigma_{\bar{x}}^2$ の公式を簡単にするため $\rho_{hn/2}$ を

$$2\sigma^2 \rho_{hn/2} = C_{hn/2}$$

で定義したときは n が奇数のときだけ用いる。同様に *serial variance* $s_{h\mu}$ を

$$kn\, s_{h\mu} = \sum_{\nu}(x_\nu - x_{\nu+h\mu})^2$$

で定義すれば，*serial variance* の *circular* な定義を用いたことになる。

ゆえに *serial variance* の比 $v_{h\mu}$ の *circular* な定義は

$$\sigma^2 v_{h\mu} = s_{h\mu}$$

である。これは $v_{hn/2}$ を

$$2\sigma^2 v_{hn/2} = s_{hn/2}$$

で定義した場合は n が奇数のときだけに用いることにする。

系列相関および *serial variance* の非-*circular* な定義は

(1)　$$k(n-\delta) C'_{h\delta} = \sum_{j}(x_j - \bar{x})(x_{j+h\delta} - \bar{x})$$

$$\sigma^2 \rho_{h\delta} = C'_{h\delta}$$

$$k(n-\delta) s'_{h\delta} = \sum_{j}(x_j - x_{j+h\delta})^2$$

および

$$\sigma^2 v_{h\delta} = s'_{h\delta}$$

で与えられる。

級内相関係数 $\bar{\rho}_h$ は式

$$\sigma^2 \bar{\rho}_h = \sum_{\mu}(x_\mu - \bar{x})(x_\nu - \bar{x})$$

で定義される。

この場合無作為化の手続（*random process*）は，まず S_i を無作為に選び，次にこの選ばれた S_i から無作為に n つの x を選ぶことによって実行される。そうすると，

$$k\sigma_{\bar{x}}^2 = \sum_{i}(\bar{x}_i - \bar{x})^2$$

で、また

(2) $\sigma^2\bar{\rho}_k = (n/n-1)\sigma_{\bar{x}}^2 - (1/n-1)\sigma^2$

であるから

(3) $\sigma_{\bar{x}}^2 = \frac{1}{n}(1+(n-1)\bar{\rho}_k)$

が成立する。

(7) から級内相関係数は

$$\bar{\rho}_k = \frac{2}{n(n-1)}\sum_\delta (n-\delta)\rho'_{k\delta}$$

$$= \frac{2}{n-1}\sum_\mu \rho_{k\mu}$$

で、結局、n が奇数のとき $\bar{\rho}_k$ は $\rho_{k\mu}$ の算術平均となるが、n が偶数のときは $\bar{\rho}_k$ は $\rho_{k\mu}$ の $n/n(n-1)$ 倍の算術平均値となることがわかる。

定理(2): 系統的抽出計画による推定値 $\bar{\bar{x}}$ は $\bar{x}$ の不偏推定量で、その分散 $\sigma_{\bar{\bar{x}}}^2$ は

(4) $\sigma_{\bar{\bar{x}}}^2 = \sigma^2\{1-\frac{1}{n^2}\sum_\delta (n-\delta)v'_{k\delta}\}$

$$= \sigma^2(1-\frac{1}{n}\sum_\mu v_{k\mu})$$

$$= \frac{\sigma^2}{n}(1+2\sum_\mu \bar{\rho}_{k\mu})$$

$$= \frac{\sigma^2}{n}\{1+(n-1)\bar{\rho}_k\}$$

である。

証明 期待値、$\bar{x}_i$、$\bar{\bar{x}}$ および系統的抽出計画の定義から、$\bar{\bar{x}}$ はそのとりうる値が $\bar{x}_1, \cdots\cdots, \bar{x}_k$ で $\bar{\bar{x}} = \bar{x}_i$ となる確率は $\frac{1}{k}$ となっているような変量である。ゆえに、

(5) $k\mathcal{E}\bar{\bar{x}} = \bar{x}_1 + \cdots\cdots + \bar{x}_k$

で、$\bar{x}_i$ の値を (5) に代入すると、$\mathcal{E}\bar{\bar{x}} = \bar{x}$ すなわち $\bar{\bar{x}}$ が $\bar{x}$ の不偏推定量であることがわかる。

$\mathcal{E}\bar{\bar{x}}$ が計算されれば $\sigma_{\bar{\bar{x}}}^2$ を評価するため $\mathcal{E}\bar{\bar{x}}^2$ を計算しなければならない。期待値の定義から

(6) $k\mathcal{E}\bar{\bar{x}}^2 = \bar{x}_1^2 + \cdots\cdots + \bar{x}_k^2$

$\bar{x}_i$ の値を (6) に代入すると

(7) $n^2k\mathcal{E}\bar{\bar{x}}^2 = \sum_{i,\alpha,\gamma} x_{i+(\alpha-1)k}\ x_{i+(\gamma-1)k}$

ゆえに附録Bの Lemma 6 の $f(u)$ を u とおくと、分散の定義から

$$\sigma_{\bar{\bar{x}}}^2 = (\frac{1}{kn})\sum_\nu (x_\nu - \bar{\bar{x}})^2 - (\frac{2}{kn^2})\sum_{\delta,i}(x_i - x_{i+k\delta})^2$$

$$= \sigma^2 - \frac{1}{n^2}\sum_\delta (n-\delta)d'_{k\delta}$$

また Lemma 8 の $f(u)$ を u でおきかえると

$$\sigma_{\bar{\bar{x}}}^2 = (\frac{1}{kn^2})\sum_\nu (x_\nu - \bar{\bar{x}})^2 + (\frac{2}{kn^2})\sum_{\delta,i}(x_i - \bar{\bar{x}})(x_{i+k\delta} - \bar{\bar{x}})$$

$$= (\frac{1}{n})\sigma^2 + \frac{2}{n^2}\sum_\delta (n-\delta)c'_{k\delta}$$

いま附録Bの Lemma 9 で $f(x_i, x_{i+k\delta})$ を $(x_i - x_{i+k\delta})^2$ でおきかえると

$$\sigma_{\bar{\bar{x}}}^2 = \sigma^2 - (\frac{1}{n})\sum_\mu d_{k\mu}$$

で、$f(x_i, x_{i+k\delta})$ を $(x_i - \bar{\bar{x}})(x_{i+k\delta} - \bar{\bar{x}})$ でおきかえれば

$$\sigma_{\bar{\bar{x}}}^2 = \frac{1}{n}(\sigma^2 + 2\sum_\mu c_{k\mu})$$

よって結局

$$\sigma_{\bar{\bar{x}}}^2 = \sigma^2(1-\frac{1}{n}\sum_\mu v_{k\mu})$$

で、また

$$\sigma_{\bar{\bar{x}}}^2 = \frac{\sigma^2}{n}(1+2\sum_\mu \rho_{k\mu})$$

---

(8) 定理1の幾分わかりやすい証明は、(2), (3) を用いて ρ_k を代入することによってえられるが、筆者の考えではこれは以下に述べるもの程有益なものでない。附録Bの Lemma はそれ自身有限母集団よりの抽出理論において重要であることは勿論である。

そうろ。

7 ρ'_{hs}, ρ_{hl} および $\sigma_{\bar{x}}^2$ のとりうる値

ここで我々は変動の型が色々違った場合に、ρ'_{hs} および $\sigma_{\bar{x}}^2$ の値がどのようになるかを簡単に調べてみよう。

いま、

$$\sigma^2 \rho'_{hs} = \frac{1}{n(n-s)} \sum_i (x_i - \bar{x})(x_{i+hs} - \bar{x})$$

である。$x_i = x_{i+hs}$, $s = 1, \cdots\cdots, n-1$; $i = 1, \cdots\cdots, n$ とすると、

$$\sum_v (x_v - \bar{x})^2 = n \sum_i (x_i - \bar{x})^2$$

および

$$\sum_i (x_i - \bar{x})(x_{i+hs} - \bar{x}) = (n-s) \sum_i (x_i - \bar{x})^2$$

である。これを代入すると、

$$\rho'_{hs} = 1, \quad \sigma_{\bar{x}}^2 = \sigma^2$$

が求められる。

$\sigma_{\bar{x}}^2$ に対するこの結果は直観的に明らかである。なぜなら変動はすべてとりうる標本間のものであるから、任意の特定の系統的標本は単一観測値と同等になるからである。

一方 $x_{hs+\alpha} = x_{hs+\beta}$, $\alpha, \beta = 1, \cdots\cdots n$; $s = 1, \cdots\cdots, n-1$ とおいてみよう。このとき任意の i, $i = 1, \cdots\cdots, n$ に対して

$$\sum_v (x_v - \bar{x})^2 = n \sum (x_{i+(\alpha-1)n} - \bar{x})^2$$

および

$$\sum_i (x_i - \bar{x})(x_{i+hs} - \bar{x}) = n \sum_\lambda (x_{i+(\alpha-1)n} - \bar{x})(x_{i+(\alpha+s-1)n} - \bar{x})$$

が成立する。更に、

$$0 = \Big[\sum_\alpha (x_{i+(\alpha-1)n} - \bar{x})\Big]^2 = \sum_\alpha (x_{i+(\alpha-1)n} - \bar{x})^2 + 2\sum_{\lambda s} (x_{i+(\alpha-1)n} - \bar{x})(x_{i+(\alpha+s-1)n} - \bar{x})$$

である。よって

$$2 \sum_s \frac{n-s}{n} \rho'_{hs} = -1, \quad \sigma_{\bar{x}}^2 = 0$$

である。

任意の特定の $\rho'_{hs} = -1$ であるような例を作ることができるが、このような場合には残りの ρ'_{hs} はすべて0となる。よく知られているように $\bar{r}_h$ の最小値は $-1/(n-1)$ である。

最後にZの添字が無作為に定められているときの ρ'_{hs} および $\sigma_{\bar{x}}^2$ の期待値を考えよう。これらの値は、

$$\rho'_{hs} = -1/(nn-1)$$

および

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{nn-n}{nn-1}\right)$$

となる。

系統的抽出を実際に用いる際は、添字が無作為にさめられているといえるような分布にしたがうZを取扱うことはまずないといってよい。一般に、Zにはある種の基本的なトレンド（傾向）があるということを期待できる論理的な理由が存在するのである。したがって、その情報を入手できることも多いし、またそうでなければ小標本からそれを求めて、これらの基礎のもとにZの添字が無作為になっていると仮定した場合とは違った何らかのやり方を採用することができるのである。

8 パラメーターの推定

今までえた系統的標本の平均値の分散の公式は母集団に対する公式である。これらの値は母集団のすべての要素の値に関係する。しかし抽出手続のテストが可能な場合でも、分母集団を研究するための資料を利用できることは殆んどない。結局、問題は標本から母集団分散と系列相関を推定することが可能かどうかを調べることになってくる。単一の S_i から導びかれた分散および相関係数には偏りがあり、一致性（consistency）のないことが示される

が、一方 S_i の一つ以上の標本から不偏あるいは一致性のある推定値を構成することが可能である。これらの推定値の抽出分散は将来の研究にまたねばならない。

我々は§6においてしたように S_i を1つだけ抽出するのでなく、g 個の S_i を無作為に抽出すると仮定する。このとき標本は S_β 内のすべての要素からなる。標本平均 $\hat{\bar{x}}$ は、我々の標本クラスの添字が $i_1, \cdots\cdots i_g$ であるとき、

$$g\hat{\bar{x}} = \sum_\beta \bar{x}_\beta$$

で定義される。そうすると $\hat{\bar{x}}$ が不偏であることが容易にわかる。更に我々はこの抽出手続が n 要素の g 組を無作為に抽出するものと考えることもできるから、

$$\sigma_{\hat{\bar{x}}}^2 = \frac{n-g}{n-1}\,\frac{1}{g}\,\sigma_x^2$$

が導びかれる。我々は§6[9]において $\sigma_{\bar{x}}^2$ を評価しておいた。

$$n\sigma_{\bar{x}}^2 = \sum_i (\bar{x}_i - \bar{\bar{x}})^2$$

であるから $\sigma_{\bar{x}}^2$ を Δ_g^2 で推定することを考えよう。

$$g\Delta_g^2 = \sum_\beta (\bar{x}_\beta - \bar{\bar{x}})^2$$

である。

いま　$E\hat{\bar{x}}^2 = \sigma_{\bar{x}}^2 + \bar{\bar{x}}^2$

かつ　$E\sum_\beta \bar{x}_\beta^2 = \frac{g}{n}\sum_i \bar{x}_i^2 = g(\sigma_{\bar{x}}^2 + \bar{\bar{x}}^2)$

---

(9) 我々がなぜこれらの S_i を系統的にでなく無作為に抽出したかについては不思議に思うかもしれない。もし S_i を系統的に抽出すれば、これは標本要素間の間隔がより小さい単一の系統的標本をとることと同等となる。その上ここで考えたような抽出分散の不偏推定値を導びくことはできない。

であるから

$$E\Delta_g^2 = \sigma_{\bar{x}}^2$$

で、結局 Δ_g^2 は $\sigma_{\bar{x}}^2$ の不偏推定値であることがわかる。更に、

$$E\Delta_g^2 = \frac{g(n-1)}{n-g}\,\sigma_x^2$$

ここで我々は方向をかえて $P_{k\mu}$ および σ^2 の推定を考えよう。

$$gn\hat{\Delta}_g^2 = \sum_{\beta,\alpha}(x_{\beta+(\alpha-1)k} - \hat{\bar{x}})^2$$

および

$$gn\hat{C}_{k\mu g} = \sum_{\beta,\alpha}(x_{\beta+(\alpha-1)k} - \hat{\bar{x}})(x_{\beta+(\alpha+\mu-1)k} - \hat{\bar{x}})$$

とおく。このとき

$$E\hat{\Delta}_g^2 = \sigma^2 - \sigma_{\bar{x}}^2$$

および

$$E\hat{C}_{k\mu g} = C_{k\mu} - \sigma_{\bar{x}}^2$$

と示すことができる。よって

$$E\Delta_g^2 + \Delta_g^2\left(\frac{n-g}{g(n-1)}\right) = \sigma^2$$

および

$$E\hat{C}_{k\mu g} + \Delta_g^2\left(\frac{n-g}{g(n-1)}\right) = C_{k\mu}$$

よって

$$\hat{C}_{k\mu g} + \Delta_g^2\left(\frac{n-g}{g(n-1)}\right) = r_{k\mu}\left[\hat{\Delta}_g^2 + \Delta_g^2\left(\frac{n-g}{g(n-1)}\right)\right]$$

で定義される $P_{k\mu}$ の推定値 $r_{k\mu}$ には偏りがある。しかし多くの場合この偏りは小であろう。初論九が偶数で $\mu = \frac{n}{2}$ なら前に定義したように $r_{k\mu/2}$ を推定するには $r_{n/2}$ を2倍にする。

もう一つの方法は

$$gn_w\hat{\Delta}_g^2 \sim \sum_{\beta,\alpha}(x_{\beta+(\alpha-1)k} - \bar{\bar{x}}_\beta)^2$$

および

$$g n_w \hat{C}_{k\mu g} = \sum_{\beta\alpha} (x_{\beta+(\alpha-1)k} - \bar{x}_\beta)(x_{\beta+(\alpha+\mu-1)} - \bar{x}_\beta)$$

を考えるものである．このとき

$$E_w \hat{A}_g^2 = \sigma^2 - \sigma_{\bar{x}}^2$$

$$E_w \hat{C}_{k\mu g} = C_{k\mu} - \sigma_{\bar{x}}^2$$

および

$$E(w \hat{A}_g^2 + A_g^2) = \sigma^2$$

$$E(w \hat{C}_{k\mu g} + A_g^2) = C_{k\mu}$$

が導びかれる．よって $\rho_{k\mu}$ のもう一つの推定値は

$$w \hat{C}_{k\mu g} + A_g^2 = w r_{k\mu}[w \hat{S}_g^2 + A_g^2]$$

なる式で定義される．$g=1$ のときは $A_g^2 = 0$ だから標本から σ^2, $C_{k\mu}$ および $\sigma_{\bar{x}}^2$ の不偏推定値を与えることは不可能である．しかし

$$\frac{1-r_{k\mu}}{1-r_{k\mu'}} = \frac{\hat{S}_g^2 - \hat{C}_{k\mu g}}{\hat{A}_g^2 - \hat{C}_{k\mu' g}}$$

で　$E[\hat{A}_g - \hat{C}_{k\mu g}] = \sigma^2 - C_{k\mu}$

だから近似的に

$$\frac{1-\rho_{k\mu}}{1-\rho_{k\mu'}} = E\frac{1-r_{k\mu}}{1-r_{k\mu'}}$$

が導びかれる．

同様な方程式は $w r_{k\mu}$ に対しても成立する．

$\rho'_{k\delta}$ を推定すると"クラス内"の定義は簡単にできる．

$$k(n-\delta)_w C'_{k\delta g} = \sum_{i,\lambda} (x_{i+(\lambda-1)k} - 1\delta \bar{x}_i)(x_{i+(\lambda+\delta-1)} - 2\delta \bar{x}_i)$$

とおく．

ここで

$$(n-\delta)_1 \bar{x}_i = \sum_\lambda x_{i+(\lambda-1)k}$$

$$(n-\delta)_2 \bar{x}_i = \sum_\lambda x_{i+(\lambda+\delta-1)k}$$

また

$$g(n-\delta)_w \hat{C}_{k\delta g} = \sum_{\beta,\lambda} (x_{\beta+(\lambda-1)k} - 1\delta \bar{x}_\beta)(x_{\beta+(\lambda+\delta-1)k} - 2\delta \bar{x}_\beta)$$

とする．更に

$$k_g C'_{k\delta g} = \sum_i (1\delta \bar{x}_i - \bar{x})(2\delta \bar{x}_i - \bar{x})$$

および

$$g_g \hat{C}_{k\delta g} = \sum_\beta (1\delta \bar{x}_\beta - \bar{x})(2\delta \bar{x}_\beta - \bar{x})$$

とおくと

$$C'_{k\delta} = w C'_{k\delta g} + g C'_{k\delta g}$$

および

$$E(w \hat{C}'_{k\delta g} + g \hat{C}'_{k\delta g}) = C'_{k\delta}$$

である．

よって $\rho'_{k\delta}$ の推定値として $r'_{k\delta}$ をうる．ここで

$$w \hat{C}'_{k\delta g} + g \hat{C}'_{k\delta g} = r'_{k\delta}(w A_g^2 + A_g^2)$$

$\bar{x}$ が既知のときには

$$E \sum_{\beta\alpha} (x_{\beta+(\alpha-1)k} - \bar{x})(x_{\beta+(\alpha+\mu-1)k} - \bar{x}) = g n C_{k\mu}$$

$$E \sum_{\beta,\lambda} (x_{\beta+(\lambda-1)k} - \bar{x})(x_{\beta+(\lambda+\delta-1)k} - \bar{x}) = g(n-\delta) C'_{k\delta}$$

および $E \sum_{\beta\alpha} (x_{\beta+(\alpha-1)k} - \bar{x})^2 = g n \sigma^2$

だから $\rho_{k\mu}$, $\rho'_{k\delta}$ および σ^2 の単純な推定値が容易に求まる．

ゆえに予備調査で $\bar{x}$ が既知なら単一の標本からでも $\sigma_{\bar{x}}^2$ のパラメータを推定することが可能である．

9　標本の大きさが変ったときの分散の変化

系統的抽出計画の分散を無作為標本の分散および系列相関係数で表わすということの主な理由は

1　無作為抽出と他の抽出計画の比較をおこなうため

2　系統的と無作為計画の効率の差が何に原因するかを分析し

やすくするため

3　色々な大きさの標本に対する分散推定値を簡単に求めるため

である。この節においてはこれらの理由の三番目を取扱う。p'_{hs} による分析は容易だから、我々は $p_{h\mu}$ のみを論ずることにする。

問題は h の函数 $\bar{p}_h$ を推定することにある。h の一つの値について $p_{h\mu}$ が計算されているとき、どのようにすれば h のすべての値に対する $\bar{p}_h$ の推定ができるかを示すため、まず σ^2 が h と無関係なことから、我々の考察を $C_{h\mu}$ に限ってよいことに注意する。§6において、$C_{h\mu}$ を

$$hn\,C_{h\mu}=\sum_{\nu}(x_\nu-\bar{x})(x_{\nu+h\mu}-\bar{x})$$

で定義した。

したがって $h'\neq h$ で $h'n'=hn=N$ なる h' について $C_{h'\mu'}$ を評価したければ $C_{h'\mu'}=C_{h\mu}$ である。したがって $h'\mu'=h\mu$ なら任意の h' および μ' について

$$C_{h'\mu'}=C_{h'\mu'/h}$$

をうる。この場合 μ を $h'\mu'/h$ でおきかえている。

この方法は $h'<h$ ならある種の内挿をおこなうことになるが、もし $p_{h\mu}$ が μ に対してプロットされれば、これは図上で行なえることが多い。しかし普通 h' の値が $h'>h$ となるように h をとるのがよい。

ある場合には相関々数を作れることがある。例えば x_ν をりの多項式で表現できれば、p'_{hs} はよの多項式で表わせる。このことからもし x_ν が滑らかな傾向に従って変るなら、p'_{hs} もまた滑らかな傾向に従って変化する。したがって内挿が可能である。この問題については更に検討が必要である。

10　層化系統抽出

サムプリングを実行する場合には層化母集団を取扱うことが多い。層化母集団にもとずく推定値の分散は普通層間の変動性を含まない。結局、母集団がうまく層化されていれば、大きさ n の標本から求めた推定値の変動は、層を考えずにとられた大きさ n の無作為標本の推定値より非常に変動が小さいのが普通である。ここで我々は層化された母集団からの系統的抽出の理論を考えることにする。

母集団 P が L 個の層 $P_1,\cdots\cdots,P_L$ からなっていて、その α 番目には N_α 個の要素 $x_{\alpha1},\cdots\cdots,x_{\alpha N_\alpha}$ が含まれるものとする。要求されているのは P の算術平均値 $\bar{x}$ の推定である。P_α の算術平均を $\bar{x}_\alpha$ と書く。$N_\alpha=h_\alpha n_\alpha$ とする。

我々は可能な二つの場合を考える。その一つは、異なる場所でも標本抽出に同じ手順を用いてよいという管理上の簡単さのために屢々用いられるものである。この節の結果から、この方法を用いてよい場合が指示されるであろう。

抽出方式 I ―― $h_1=h_2=\cdots\cdots=h_L=h$ で、抽出方式は $1,\cdots,h$ の中から一つの整数を無作為に選ぶものであるとする。これらの各整数の抽出確率は $1/h$ になっている。このとき選ばれた整数がたとえば i なら P_α の標本は $x_{\alpha i},x_{\alpha,i+h},\cdots\cdots,x_{\alpha,i+(n_\alpha-1)h}$ からなる。したがってこの抽出方式をとることにより、実際の標本となる確率が何れも $1/h$ であるような丁度 h 個の可能な標本、すなわち $S_1,\cdots\cdots,S_h$ が存在することになる。

抽出方式 II ―― この抽出方式はそれぞれの α について $1,\cdots\cdots,h_\alpha$ の整数の一つを、各整数の選ばれる確率が $1/h_\alpha$ となるようにして無作為に抽出するものである。故にこの抽出方式を適用したとき、実際に標本となる確率がそれぞれ $1/h_1,\cdots\cdots,1/h_L$ であるような可能な標本が丁度 $h_1,\cdots\cdots,h_L$ 個あることになる。

層化抽出には勿論他の抽出方式もある。しかし上にあげた二つは、系統抽出に含まれるものを除けば実際的問題の殆んど全部を網羅している。後者については後の論文で取扱うことにする。更にこれらの方式について導びかれた結論から、他の層化抽出方式

に対する結論を推測することも可能である。

S_{ai} を要素 χ_{ai}, χ_{ai+k}, ……, $\chi_{ai+(n_a-1)k}$ のクラスとする。まず抽出方式Ⅰを考察する。P_a から大きさ n_a の系統的標本を抽出しなければならない。

可能な標本は S_1, ……, S_k である。ここで S_i は S_{1i}, ……, S_{Li} 内のすべての要素からなる。S_{ai} 内の要素の算術平均値を $\tilde{\chi}_{ai}$ と書く。P_a からとられた標本の算術平均値を $\tilde{\chi}_a$, 標本平均値を $\tilde{\chi}$ と書く。ここで

$$N\tilde{\chi} = N_1\tilde{\chi}_1 + \cdots\cdots + N_L\tilde{\chi}_L$$

である。そうすると

$$N\mathcal{E}\tilde{\chi} = \sum_a N_a \mathcal{E}\tilde{\chi}_a = \sum_a N_a \frac{1}{k}\sum_i \tilde{\chi}_{ai} = N\tilde{\chi}$$

附録Cから

$$\sigma_{\tilde{\chi}}^2 = \frac{1}{N^2}\sum_{a,b} N_a N_b \sigma_{\tilde{\chi}_a\tilde{\chi}_b}$$

がえられる。ここで

$$\sigma_{\tilde{\chi}_a\tilde{\chi}_b} = \mathcal{E}(\tilde{\chi}_a - \bar{\chi}_a)(\tilde{\chi}_b - \bar{\chi}_b)$$

$$= \frac{1}{k}\sum_i (\tilde{\chi}_{ai} - \bar{\chi}_a)(\tilde{\chi}_{bi} - \bar{\chi}_b)$$

である。

$\sigma_{\tilde{\chi}_a\tilde{\chi}_b}$ の式はこれ以上簡単にはならないが、重要なことは、異なる層の対応する item 間に正の相関があれば、抽出誤差以外の考慮が重大でない限り、抽出方式Ⅰは用いない方がよい。しかし対応する item 間に負の相関があれば、抽出方式Ⅰの分散はⅡの分散より小さくなる。

次に抽出方式Ⅱを考えよう。ⅠとⅡの違いは、Ⅱでは各層内の標本抽出を別々に行なっているから、もし $a \neq b$ なら $\sigma_{\tilde{\chi}_a\tilde{\chi}_b} = 0$ となることである。故に方式Ⅱの場合の分散は

$$\sigma_{\tilde{\chi}}^2 = \frac{1}{N^2}\sum_a N^2 \sigma_{\tilde{\chi}_a}^2$$

となる。ここで $\sigma_{\tilde{\chi}_a}^2$ は次ぐ節で導びかれている。

11 系統的抽出と無作為抽出方式の効率の比較

どんな抽出方法の研究でも他の可能な抽出方法と比較をしなければ完成したとはいえない。よってこの節では、系統的抽出方式と無制限無作為および層化無作為抽出方式との比較を行なうことにする。

無作為および層化無作為抽出方式に対する平均値と分散はそれぞれ(′)および(″)で区別することにする。

そうすると

$$\sigma_{\tilde{\chi}}^2 / \sigma_{\tilde{\chi}'}^2 = (1 + 2\sum_{\mu} \rho_{k\mu})\left(\frac{kn-1}{kn-n}\right)$$

なることがわかるから、結局

$$\sum \rho_{k\mu} < -(n-1)/2(kn-1)$$

なら $\sigma_{\tilde{\chi}}^2 < \sigma_{\tilde{\chi}'}^2$ となる。これに比して大きければ $-(n-1)/2(kn-1)$ の近似として $-\frac{1}{2}k$ を使ってもよい。

もっと細かい比較を行なうには、母集団要素 χ_ν が ν のある函数で与えられるものと仮定し、その函数が

$$\chi_\nu = A_0 + A_1\nu + \cdots + A_h\nu^h$$

あるいは

$$\chi_\nu = B_0 + A_1 \sin\frac{2\pi\nu}{N} + B_1 \cos\frac{2\pi\nu}{N} + \cdots\cdots + A_h \sin\frac{2\pi h\nu}{N} + B_h \cos\frac{2\pi h\nu}{N}$$

のようなものであるとして、種々の可能な抽出方式の効果を、そのように仮定された χ_ν の分布に基いて研究するのが有用である。系統的抽出を用いる際には、母集団の要素を論理的なやり方で順序づけ、またこの順序を用いて標本を系統的に抽出できると仮定しているということに注意しなければならない。

我々はここで幾つかの可能性を考える。まず一つの層から1つ

の要素しか抽出しなければ層標本平均値の分散は抽出が無作為でも系統的でも同じだということを注意しておこう。一方オノの節から、母集団をL個の層にわけて、各層からi番目の要素をとることにより大きさLの系統的標本を作ったとすると、層化無作為標本平均値の分散は、系統的標本内の層標本平均値間の平均的相関が負であるか正であるかによって系統的標本の平均値の分散より大きくなったり小さくなったりすることがわかる。

ここで周期性をも母集団については系統的標本を用いてはならないという警告がどうして出てきたかを考えてみよう。周期がkなら系統的標本の層平均値間の相関は+1であるから無作為標本の方が優れている。しかし周期がkなら、系統的標本の方がおそらく分散が小さくなるということを証明しよう。

周期がkで、大きさnの隣接する二つの層内では常に

$$x_1 = x_{2k}, \quad x_2 = x_{2k-1}, \cdots\cdots, x_k = x_{k+1} \quad で$$

$$x_i - \bar{x} = -(x_{k+i} - \bar{x})$$

であると仮定する。このとき各層から1つの要素を抽出すると系統的標本平均値(この場合には個々の要素)間の相関は順番号の差が奇数なら−1で、偶数なら+1である。

n個の層のそれぞれの中の分散はσ_i^2である。ここで

$$k\sigma_i^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2$$

ある。層平均値間の分散は0である。よって$\sigma^2 = \sigma_i^2$。ゆえに$n = L$なるときの大きさnの無制限無作為標本平均値の分散は

$$\sigma_{\bar{x}}^2 = \frac{N-L}{N-1}\,\frac{\sigma_i^2}{L}$$

で、層化無作為標本平均値の分散は$\sigma_{\bar{x}'}^2$である。一方系統的抽出平均値の分散は

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{L^2}\sum_{i,j=1}^{L}(-1)^{i-j}(2-\delta_{ij})$$

である。ただし$i=j$なら$\delta_{ij}=1$、$i\neq j$なら$\delta_{ij}=0$である。

よってLが偶数なら$\sigma^2=0$であるがもしLが奇数なら$\sigma^2=(1/L^2)\sigma_i^2$である。

結局母集団の分布が周期性を有し、層の大きさが周期の半分なら系統的標本平均値の効率は層化無作為標本の効率より大きいということになる。kが周期に等しいかまたは周期の半分に等しい場合に成立するのと同様なことが、kが周期の半分の奇数倍または偶数倍に等しいときにもなり立つことに注意しなければならない。

しかし母集団の要素が直線分布に従うと仮定すれば事情は全く変ってくる。一般性を失なわずに、直線上の分布は$x_y = y$で与えられると仮定してよい。そうすると大きさnの無制限無作為標本については、標本平均を$\bar{x}'$と書くと

$$E\bar{x}' = \bar{x} = \frac{1}{2}(kn+1)$$

$$\sigma^2 = \frac{k^2n^2-1}{12}$$

および

$$\sigma_{\bar{x}'}^2 = \frac{(k-1)(kn+1)}{12}$$

そうろ。

層化無作為抽出計画では$N_1 = \cdots\cdots = N_L = \frac{CN}{n}$と仮定することでCは整数1, 2, ……, nのどれと同じでもよい。すなわち$L = \frac{n}{C}$、$n_1 = \cdots\cdots = n_L = C$とする。このとき

$$\sigma_{\bar{x}''}^2 = \frac{C^2(k-1)}{n^2(Ck-1)}\sum_{\alpha}\sigma_\alpha^2$$

ここで$\bar{x}''$は層化無作為標本の標本平均値である。もし第α層に$x_{(\alpha-1)\frac{CN}{n}+1}, \cdots\cdots, x_{\alpha\frac{CN}{n}}$が含まれていれば$\sigma_\alpha^2 = \frac{C^2k^2-1}{12}$および$\sigma_{\bar{x}''}^2 = \frac{C}{N}\cdot\frac{k-1}{Ck-1}\cdot\frac{C^2k^2-1}{12}$である。

結局

$$\sigma_{\bar{x}}^2 = \sigma^2 - \frac{1}{kn^2}\sum_{\delta}\sum_{j}(x_j - x_{j+k\delta})^2$$

$$= \sigma^2 - \frac{k^2(n^2-1)}{12}$$

$$= \frac{k^2-1}{12}$$

これをまとめると、

$$\sigma_{\bar{x}'}^2 = \frac{(k-1)(kn+1)}{12} = \frac{(k-1)(N+1)}{12}$$

$$\sigma_{\bar{x}''}^2 = \frac{c(k-1)(ck+1)}{12} = \frac{(k-1)(\frac{N}{L}+1)}{12L}$$

$$\sigma_{\bar{x}}^2 = \frac{k^2-1}{12}$$

のようになる。

$\sigma_{\bar{x}}^2$ および $\sigma_{\bar{x}''}^2$ のどちらも $\sigma_{\bar{x}'}^2$ より小さいことは明らかである。しかし

$$\sigma_{\bar{x}}^2 \Big/ \sigma_{\bar{x}'}^2 = \frac{L(k+1)}{\frac{N}{L}+1}$$

である。

$kn = N$, $cL = n$ だから $k = \frac{N}{cL}$。よって

$N > L(L-1)$ かつ $c \geq \frac{\frac{L}{1-\frac{L(L-1)}{N}}}{}$ なら

$$\sigma_{\bar{x}}^2 < \sigma_{\bar{x}''}^2$$

である。いかなる場合にも $\frac{N}{L} \geq c$。よって c が上記の値となるためには $\frac{N}{L} > \frac{NL}{N-L(L-1)}$ すなわち N は $2L^2-L$ 以上でなければならないことがわかる。よって $N > 2L^2-L$ および $c \geq \frac{\frac{L}{1-\frac{L(L-1)}{N}}}{}$ なら $\sigma_{\bar{x}}^2 \leq \sigma_{\bar{x}''}^2$ である。それ以外の場合は $\sigma_{\bar{x}}^2 > \sigma_{\bar{x}''}^2$ である。

この結果は二つの事実によってわかる。

(1) 各層から1つの要素を抽出していれば、層平均値間の平均的な相関が高いので、層内の分散が等しいにもかかわらず層化無作為標本平均値の効率が、系統的抽出の平均値の効率より大きくなってくる。

(2) 各層から2つ以上の要素が抽出されていると、系統的抽出平均の層内分散は層化無作為抽出平均の層内分散より小であって、層および層からとられた標本の大きさが十分大なら系統的標本の層内分散の減少は層平均値間の相関によるものより更に大となる。もちろん、直線的分布（straight line distribution）のときには、我々が用いたものより更に有効な層化無作為標本の定義の仕方がある。その上抽出手続もここに論じたものより一層有効なものが利用できる。しかしこの例は、実際に生ずる一般的問題のみならず、それらの解決に利用できる方法を指摘する上で有用である。

系統的抽出と層化無作為抽出の他の比較は、x_ν が二つの要素、すなわち傾向函数（trend function）および周期函数（periodic function）からなっていると考えてられる。したがって傾向（trend）からの偏差は周期函数となっている。

$x_\nu = \varphi_1(\nu) + \varphi_2(\nu)$ とする。ここで $\varphi(\nu)$ は傾向函数で $\varphi_2(\nu)$ は周期 $2k$, $N = 2kQ$ の周期函数である。

$\varphi_2(\nu) = y_\nu$ とする。このとき $y_j = y_{2k+j} = \cdots\cdots = y_{2k(Q-1)+j}$, $j = 1, 2, \cdots\cdots, 2k$ および $y_{2kQ+j} - \bar{y} = -(y_{2kQ+k+j} - \bar{y})$, $j = 1, \cdots\cdots, k$, $Q = 0, \cdots\cdots, Q-1$。この比較で我々が考える標本の大きさはすべて k の倍数であるから、必要な情報が全部一度でえられるように、分散、共分散を計算する。

$\varphi_1(\nu)$ の平均値を $\bar{\varphi}_1$, $\varphi_2(\nu)$ の平均値を $\bar{y}$ と書く。

そうすると $\bar{x} = \bar{\varphi}_1 + \bar{y}$ で

$$N\sigma^2 = \sum_\nu (x_\nu - \bar{x})^2$$

$$= \sum_\nu [\varphi_1(\nu) - \bar{\varphi}_1]^2 + \sum_\nu (y_\nu - \bar{y})^2 + 2\sum_\nu [\varphi_1(\nu) - \bar{\varphi}_1](y_\nu - \bar{y})$$

$$= \sum_{a,i} (\varphi_{1(a-1)k+i} - \bar{\varphi}_{1a})^2 + k\sum_a (\bar{\varphi}_{1a} - \bar{\varphi}_1)^2$$

$$+ \sum_{a,i} (y_{(a-1)k-i} - \bar{y}_a)^2 + k\sum_a (\bar{y}_a - \bar{y})^2$$

$$+\sum_{a,i}\left[\varphi_{i(a-1)k+i'}-\bar{\varphi}_{ia}\right]\left(y_{(a-1)k+i'}-\bar{y}_a\right)$$

$$+k\sum_a\left(\bar{\varphi}_{ia}-\varphi_i\right)\left(\bar{y}_a-\bar{y}\right)$$

ここで $a=1,\cdots\cdots,2Q$; $i'=1,\cdots\cdots,k$, $\bar{\varphi}_{ia}$ は $\varphi_{i(a-1)k+1},\cdots\cdots,\varphi_{i(ak)}$ の算術平均値で $\bar{y}_a$ は $y_{(a-1)k+1},\cdots\cdots,y_{ak}$ の算術平均値である。

y_v についての仮定から $\bar{y}_a=\bar{y}$ および $\sum_i(y_i-\bar{y}_a)^2$ はaのそれぞれの値について同一で、$\sum_{i=1}^{k}(y_i-\bar{y})^2$ に等しいことがわかる。

$y_i=y_{2k+i}=\cdots\cdots=y_{2k(Q-1)+i}$ および $y_{k+i}=y_{3k+i}=\cdots\cdots=y_{2k(Q-1)+k+i}$ であるから

$$\sum_v\left[\varphi_i(v)-\bar{\varphi}_i\right]\left(y_v-\bar{y}\right)=\sum_{i=1}^{k}\left(y_i-\bar{y}\right)\sum_{a=1}^{Q}\left[\varphi_{i(2a-2)k+i'}-\bar{\varphi}_{i2a-2}\right]$$

$$+\sum_{i=k+1}^{2k}\left(y_i-\bar{y}\right)\sum_{a=1}^{Q}\left[\varphi_{i(2a-2)k+i'}-\bar{\varphi}_{i2a-2}\right]$$

そうる。

また $y_i-\bar{y}=\left(y_{i+k}-\bar{y}\right)$ だから

$$\sum_v\left[\varphi_i(v)-\bar{\varphi}_i\right]\left(y_v-\bar{y}\right)=\sum_{i=1}^{k}\left(y_i-\bar{y}\right)\left\{\sum_{a=1}^{Q}\left[\varphi_i\left((2a-2)k+i'\right)\right.\right.$$

$$\left.\left.-\varphi_i\left((2a-1)k+i'\right)-\bar{\varphi}_{i2a-2}+\varphi_{i2a-1}\right]\right\}$$

かえられる。これは $\varphi_i(v)$ が直線であるかまたは $\varphi_i(v)$ がそれぞれ長さ $2k$ の直線の系列のときは0になる。

ここで $\varphi_i(v)=A+Bv$ とする。そうすると

$$\bar{\varphi}_{ia}=A+B\left[(a-1)k+\frac{k+1}{2}\right],$$

$$\varphi_i\left((a-1)k+i'\right)-\bar{\varphi}_{ia}=i'-\frac{k+1}{2},$$

$$\sum_i\left[\varphi_i\left((a-1)k+i'\right)-\bar{\varphi}_{ia}\right]^2=\frac{B^2k\left(k^2-1\right)}{12},$$

$$\sum_a\left(\bar{\varphi}_{ia}-\bar{\varphi}_i\right)^2=\frac{B^2k^2\left(4Q^2-1\right)(2Q)}{12}$$

$$\bar{\varphi}_i=A+B\frac{2k+1}{2},$$

$$\sum_v\left(\varphi_i(v)-\bar{\varphi}_i\right)^2=\frac{2kQB^2}{12}\left[4k^2Q^2-1\right]$$

よって $\sigma^2=\sigma_y^2+\frac{B^2}{12}\left[4k^2Q^2-1\right]$. ただし $k\sigma_y^2=\sum_i(y_i-\bar{y})^2$. また大きさ kn の単純無作為標本の平均値の分散は

$$\sigma_{\bar{y}}^2=\frac{N-n}{N-1}\frac{\sigma^2}{n}\quad\text{である。}$$

いま層の大きさが mk で、m は $2Q$ の因数すなわち $2Q/m=Lm$ としよう。そうすると Lm 個の各層内の分散は一定値すなわち σ_i^2 である。ただし m が偶数なら $\sigma_i^2=\sigma_y^2+\frac{B^2}{12}\left[k^2m^2-1\right]$。$m$ が奇数のときは Lm が偶数で、層の半分では層内分散は $\sigma_i^2+\frac{1}{km}\sum_{i=1}^{k}(y_i-\bar{y})(i'-\frac{k+1}{2})$ である。一方他の層における層内分散は $\sigma_i^2-\frac{1}{km}\sum_{i=1}^{k}(y_i-\bar{y})(i'-\frac{k+1}{2})$

そうすると Lm 個の層の各々から C 個の要素を無作為に抽出すると

$$\sigma_{\bar{y}}^2=\frac{1}{Lm}\left(\frac{mn-C}{mk-1}\right)\frac{\sigma_i^2}{C}$$

$$=\frac{1}{Lm}\left(\frac{mk-C}{mk-1}\right)\frac{1}{C}\left(\sigma_y^2+\frac{B^2}{12}\left[k^2m^2-1\right]\right)$$

なることがわかる。

系統的抽出の平均値の分散を求めるため $\sum_j(x_j-x_{j+k\delta})^2=k(n-\delta)s_{k\delta}'^2$ を計算してみよう。いま x_v を代入すると $k(n-\delta)s_{k\delta}'^2=\sum_j(y_j-y_{j+k\delta})^2-2Bk\delta\sum_j(y_j-y_{j+k\delta})+k(n-\delta)\delta^2k^2\delta^2$ なることがわかる。

よってkがkの偶数なら $\sum_j(y_j-y_{j+k\delta})=0$ となる。実にkがkの奇数倍ならjは $y_j=y_{j+\delta k}$ だから $\sum_j(y_j-y_{j+\delta k})^2=0$。結局kがkの奇数倍ならδは奇数となって $y_j-y_{j+\delta k}=2(y_j-\bar{y})$。一方δが偶数なら $y_j-y_{j+\delta k}=0$ となるから kがkの奇数倍でδが奇数のときは

$$\sum_i (y_i - y_{i+\delta k})^2 = 4\sum_i (y_i - \bar{y})^2$$

$$= 4\frac{k(n-\delta)}{n}\sum_i (y_i - \bar{y})^2$$

となる。

k が n の奇数倍なら n が偶数であることを注意しておく。

$$\sigma_{\bar{y}}^2 = \sigma^2 - \frac{1}{n^2}\sum_\delta (n-\delta)\Delta'_{k\delta}$$

だから $\sum_\delta (n-\delta)\Delta'_{k\delta}$ を評価しなければならない。いま k が n の偶数倍なら $(n-\delta)\Delta'_{k\delta} = (n-\delta)B^2k^2\delta^2$ および

$$\sum_\delta (n-\delta)\Delta'_{k\delta} = B^2k^2\{n\sum_\delta \delta^2 - \sum_\delta \delta^3\}$$

$$= B^2k^2\frac{n^2(n^2-1)}{12}$$

が成立する。よって もし k が n の偶数倍なら

$$\sigma_{\bar{y}}^2 = \sigma^2 - \frac{B^2k^2(n^2-1)}{12}$$

なることがわかる。

一方もし k が n の偶数倍で δ が奇数なら

$$(n-\delta)\Delta'_{k\delta} = (n-\delta)B^2k^2\delta^2 + 4(n-\delta)\sigma_{\bar{y}}^2$$

で、δ が偶数なら

$$(n-\delta)\Delta'_{k\delta} = (n-\delta)B^2k^2\delta^2$$

となる。

よって

$$\sum_\delta (n-\delta)\Delta'_{k\delta} = \frac{B^2k^2n^2(n^2-1)}{12} + n^2\sigma_{\bar{y}}^2$$

ゆえに k が n の奇数倍なら

$$\sigma_{\bar{y}}^2 = \sigma^2 - \frac{B^2k^2(n^2-1)}{12} - \sigma^2$$

なることがわかる。

また k が n の偶数倍のときは

$$\sigma_{\bar{y}}^2 = \sigma_{\bar{y}}^2 + \frac{B^2}{12}(k^2n^2-1) - \frac{B^2}{12}k^2(n^2-1)$$

$$= \sigma_{\bar{y}}^2 + \frac{B^2}{12}(k^2-1)$$

で、k が n の奇数倍なら

$$\sigma_{\bar{y}}^2 = \frac{B^2}{12}B^2(k^2-1)$$

となる。

このことから

$$c > \frac{L}{1 + \frac{12\sigma_{\bar{y}}^2}{B^2(km)(km-1)} - \frac{L(L-1)}{N}}$$

のときには系統的抽出が優れた結果を与える。

$\frac{N}{L} > c$ であるから c の解が存在するためには

$$N > 2L^2 - L - \frac{12\sigma_{\bar{y}}^2}{B^2}\left(\frac{L^2}{N-L}\right)$$

でなければならない。

12 摘　要

この論文では、分散の公式、可能な母数の値の検討、母数の推定、標本の大きさを変えたときの影響、および系統的抽出、単純無作為抽出、層化無作為抽出間の比較をも含めて、層化または非層化母集団に対する系統的抽出法の理論的基礎を提示した。この論文には、抽出単位が単一の要素からなっているような場合において必要とされる理論のみならず、更に系統的抽出を採用すべき条件についても若干の分析を行ないまた分散を計算するための公式が含まれている。

このあとの論文においては、抽出単位が要素の集団である場合の系統的抽出の理論を提出するが、この理論は有限母集団からの抽出でなく、その各要素が正規分布するときの無限母集団からの抽出を仮定するものである。そうして更に系統的抽出理論の色々な部分ならびに実際を与えるであろう。

附録 A

特定の変数の示す値

変数の限界の繰返えしを避けるためこれらの範囲をこの附録に与える。

第 1 表　添字のとる値

文　字	この文字は 1 から下に示す値までのすべての整数をとるものとする
i	k
λ	$n-\delta$
j	$k(n-\delta)$
δ	$n-1$
ν, ν'	km
α	n
γ	n
μ, μ'	n が偶数なら $n/2$，　n が奇数なら $\frac{n-1}{2}$
a, b	L

文字 β は $i_1, i_2, \cdots\cdots, i_g$ なる値を仮定する，ただし $i_1, \cdots\cdots, i_g$ は k 個の整数 $1, \cdots\cdots, k$ のうちから選んだ g 個の数である。

附録 B

若干の有限和の範囲について

有限和を変換するときに生ずる困難は重積分の変換理論において生ずるもの，すなわち変換する変数の影響または和の範囲についてとる和の順序等に非常によく似ている。この論文で有用なことが証明された幾つかの *Lemma* を別々に求め一般的な形で与えることにする。

$f(u)$, $f(u,v)$ は u と v のすべての可能な値について有限な u v の函数と仮定する。

Lemma 1

$$\sum_{\substack{\alpha,\beta \\ \alpha<\beta}} f(x_{i+(\alpha-1)k},\ x_{i+(\beta-1)k}) = \sum_{\delta\lambda} f(x_{i+(\lambda-1)k},\ x_{i+(\lambda+\delta-1)k})$$

証明：　$\alpha=\lambda$，$\gamma=\lambda+\delta$ とする。$1\leqq\alpha<\gamma$ および $\gamma\leqq n$ だから δ の可能な値は $1, \cdots\cdots, (n-1)$ である。$\lambda=\gamma-\delta$ だから δ の各値について λ のとりうる値は $n-\delta$ である。また δ を一定としたとき λ の最大値は $\gamma=n$ なるとき決定される。これらの範囲に対して左辺の式の f の各項は右辺において唯一回だけしか現われない。更に右辺においてはこれ以外の項は出現しない。

Lemma 2

$$\sum_{i}\sum_{\lambda} f(x_{i+(\lambda-1)k},\ x_{i+(\lambda+\delta-1)k}) = \sum_{j} f(x_j,\ x_{j+k\delta})$$

証明：　$j=i+(\lambda-1)k$ とおく。このとき j は i および λ の単調増加函数である。j の最小値は $i=1$ のときに生ずる。そのとき $j=1$ である。j の最大値は $i=k$，$\lambda=n-\delta$ なるときに現われる。そのとき $j=(n-\delta)k$ である。これらの範囲に対して左辺の式の f の各項は右辺において唯一回しか現われない。更に右辺においてはこれ以外の項は出現しない。

Lemma 3

$$\sum_{\substack{i,\alpha,\beta \\ \alpha<\beta}} f(x_{i+(\alpha-1)k},\ x_{i+(\gamma-1)k}) = \sum_{\delta j} f(x_j,\ x_{j+\delta k})$$

証明：　まず *Lemma 1* を $\sum_{\alpha<\beta} f(x_{i+(\alpha-1)k},\ x_{i+(\gamma-1)k})$ に適用し次に与えられた式に *Lemma 2* を適用する。

Lemma 4

$$\sum_{\delta j}[f(x_j)+f(x_{j+k\delta})] = (n+1)\sum_{\nu} f(x_\nu)$$

証明：　$m=j+k\delta$ とおく。このとき δ を一定とすると m の最小値は $j=1$ なるときに生ずる。そのとき $m=k\delta+1$ である。δ の任意の一定値に対して m の最小値は $j=k(n-\delta)$ なるときに生ずる。その場合 $m=nk$ である。文字 m は $k\delta+1$ から nk までのすべての整数値をとるものとする。よって

$$\sum_{\delta j} f(x_j)+\sum_{\delta j} f(x_{j+k\delta}) = \sum_{\delta j} f(x_j)+\sum_{\delta m} f(x_m)$$

$\sum_{\delta m} f(x_m)$ を 1 から $n-1$ でなく $n-1$ から 1 までの δ につい

て加えあげると

$$\sum_{s,j} f(x_j) + \sum_{s,m} f(x_m) = \sum_{j=1}^{k(n-s)} f(x_j) + \sum_{m=k(n-s)+1}^{kn} f(x_m)$$
$$+ \cdots\cdots\cdots\cdots$$
$$+ \sum_{j=1}^{ks} f(x_j) + \sum_{m=ks+1}^{kn} f(x_m)$$

なることがわかる。ここで x_j の和は $\sum_{s,j} f(x_j)$ の項で x_m の和は $\sum_{s,m} f(x_m)$ の項である。しかるに

$$\sum_{j=1}^{k(n-s)} f(x_j) + \sum_{m=k(n-s)+1}^{kn} f(x_m) = \sum_{\nu} f(x_\nu)$$

であるから Lemma 4 が証明された。

Lemma 5

$$\sum_{i,\alpha,\gamma} f(x_{i+(\alpha-1)k})\, f(x_{i+(\gamma-1)k}) = A$$

とすると

$$A = n \sum_{\nu} [f(x_\nu)]^2 - \sum_{j,s} \{f(x_j) - f(x_{j+ks})\}^2$$

である。

証明：

$$A = \sum_{i,\alpha} [f(x_{i+(\alpha-1)k})]^2 + 2 \sum_{\substack{i,\alpha,\gamma \\ \alpha<\gamma}} f(x_{i+(\alpha-1)k}) f(x_{i+(\gamma-1)k})$$

Lemma 3 によって

$$2 \sum_{\substack{i,\alpha,\gamma \\ \alpha<\gamma}} f(x_{i+(\alpha-1)k}) f(x_{i+(\gamma-1)k}) = 2 \sum_{s,j} f(x_j) f(x_{j+sk})$$

また

$$2 f(x_j) f(x_{j+sk}) = f(x_j)^2 + f(x_{j+sk})^2 - \{f(x_j) - f(x_{j+sk})\}^2$$

が成立するから Lemma 4 5 用いれば証明が完結する。

Lemma 6　$kn\bar{f} = \sum_{\nu} f(x_\nu)$ とする。このとき

$$A\left(\frac{1}{kn^2}\right) - \bar{f}^2 = \left(\frac{1}{kn}\right) \sum_{\nu} [f(x_\nu) - \bar{f}] - \left(\frac{1}{kn^2}\right) \sum_{j,s} [f(x_j) - f(x_{j+ks})]^2$$

である。

証明：この Lemma は Lemma 5 の直接の結果である。

Lemma 7

$$A = \sum_{\nu} [f(x_\nu) - \bar{f}]^2 + 2 \sum_{j,s} [f(x_j) - \bar{f}][f(x_{j+ks}) - \bar{f}] + kn^2 \bar{f}^2$$

が成立する。

証明：Lemma 4 から

$$A = n \sum_{\nu} f(x_\nu)^2 - \sum_{j,s} \{[f(x_j) - \bar{f}]^2 + [f(x_{j+ks}) - \bar{f}]^2$$
$$+ 2 \sum_{j,s} [f(x_j) - \bar{f}][f(x_{j+ks} - \bar{f}]$$

が導びかれるから

$$A = n \sum_{\nu} [f(x_\nu) - \bar{f}]^2 + n^2 k \bar{f}^2 - (n-1) \sum_{\nu} [f(x_\nu) - \bar{f}]^2$$
$$+ 2 \sum_{j,s} [f(x_j) - \bar{f}][f(x_{j+ks}) - \bar{f}]$$

なることがわかる。このとき Lemma が証明される。

Lemma 8

$$A\left(\frac{1}{kn^2}\right) - \bar{f}^2 = \left(\frac{1}{kn^2}\right) \sum_{\nu} [f(x_\nu) - \bar{f}]^2 + \left(\frac{2}{kn^2}\right) \sum_{j,s} [f(x_j) - \bar{f}][f(x_{j+ks}) - \bar{f}]$$

この Lemma は Lemma 7 の直接の結果である。

Lemma 9　$k > kn$ なら x_k は x_{k-kn} に等しいものとする。

$f(u,v) = f(v,u)$　すなわち f は対称とする。

このとき

$$d_{ks} = \sum_{j} f(x_j, x_{j+ks})$$

とおけば

$$d_{ks} + d_{kn-s} = \sum_{\nu} f(x_\nu, x_{\nu+ks})$$

が成立する。

証明：明らかに

$$\sum_{\nu} f(x_\nu, x_{\nu+ks}) = d_{ks} + B$$

ここで

$$B = \sum_{g=k(n-s)+1} f(x_g, x_{g+ks})$$

いま $h = g - (n-s)k$ とおくと

$$B = \sum_{h=1}^{sk} f(x_{h+(n-s)k}, x_{h+kn})$$

$x_{k+kn} = x_k$, $f(x_{k+(n-s)k}, x_k) = f(x_k, x_{k+(n-s)k})$ だから
$B = d_{kn-s}$ となり Lemma が証明される。$f(u, v)$ の対称性は十分条件であるとともに必要条件でもあることに注意せよ。何となれば $f(x_v, x_{v+ks}) = x_v - x_{v+ks}$ ならこの定理は成立しない。

附録 C

層化抽出

母集団 P は L 個の層 $P_1, \cdots\cdots, P_L$ からなっているとする。P の算術平均を $\bar{x}$ とし、P_a の算術平均を $\bar{x}_a$ とする。$\bar{x}_a$ の標本推定値を $\tilde{x}_a$ とし、$\tilde{x} = \sum_a C_a \tilde{x}_a$ とおく。このとき $E\tilde{x} = \sum_a C_a A_a = A$。ここで $E\tilde{x}_a = A_a$ $\sigma_{\tilde{x}}^2$ は $\sigma_{\tilde{x}}^2 = E(\tilde{x} - \bar{x})^2$ で定義されるものとする。そうすると、$\sigma_{\tilde{x}}^2 = E(\tilde{x} - A)^2 + (A - \bar{x})^2$ となって、容易に $\sigma_{\tilde{x}}^2 = \sum_{a,b} C_a C_b \sigma_{\tilde{x}_a \tilde{x}_b} + (A - \bar{x})^2$ なることがわかる。ただし

$$\sigma_{\tilde{x}_a \tilde{x}_b} = E(\tilde{x}_a - A_a)(\tilde{x}_b - A_b)$$
$$(A - \bar{x})^2 = \left[\sum_a \left(C_a A_a - \frac{N_a}{N} \bar{x}_a\right)\right]^2$$

で、そし $NC_a = N_a$ なら

$$(A - \bar{x})^2 = \sum_{a,b} C_a C_b (A_a - \bar{x}_a)(A_b - \bar{x}_b)$$

かつ $\sigma_{\tilde{x}}^2 = \sum_{a,b} C_a C_b \sigma_{\tilde{x}_a \tilde{x}_b}$

である。ここで

$$\sigma_{\tilde{x}_a \tilde{x}_b}^2 = E(\tilde{x}_a - \bar{x}_a)(\tilde{x}_b - \bar{x}_b)$$

これらの公式は subclass の抽出が如何なる方法でなされるときにも成立する。そし $\tilde{x}$ が $\bar{x}$ の不偏推定値で $\tilde{x}_a$ が $\tilde{x}_b$ と独立なら普通の公式 $\sigma_{\tilde{x}}^2 = \sum_a C_a^2 \sigma_{\tilde{x}_a}^2$ が成立する。勿論 $\sigma_{\tilde{x}_a \tilde{x}_b}$ の公式は、無作為、有意、系統的および他の抽出手続のどれを採ったかによって変ってくる。

3 ある種の母集団に対する系統的標本と無作為標本の相対的精度.

Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of populations

1 要　約 (summary)

色々なサムプリングにおいて屢々出現する母集団のタイプは、その中の要素を辿わけしたとき、群が大きくなるにつれて群内分散が急速に増加してくるようなものである。この様な母集団の Class は、その要素に系列的な関係がある、即ち二つの要素間の相関は正で、それらの距離が増加するに従って単調に減少する組数。模型によって表わすことができる。この型の母集団について、各要素毎の系統的標本 (Systematic sample)、層当り1要素を抽出する層化無作為標本 (stratified random sample) 及び単純無作為標本 (random sample) の相対的な効率 (efficiency) を比較した。平均的には、層化無作為標本は少くとも単純無作為標本より常に精度が高く、その相対的効率は標本サイズの増加に伴って増加する単調増加箱数で表わすことができる。系統的標本の相対的な効率については何一つとして普遍性のある結果を与える事は出来ない。実際、ある抽出率では、系統的標本の方が層化無作為標本より精度の良いこともあるが、他の抽出率では逆に悪くなるような母集団の Class が存在するのである。しかしもしコレログラム (correlagram) が上に凹であれば、任意の標本サイズに対して、平均的には、層化無作為標本よりも系統的標本の精度が良い。コレログラムが1次型の指数型 (exponential) の場合について数つ分の数値が得られた。

2 緒　言

我々は要素 $x_1, x_2, \cdots\cdots x_{nk}$ からなる有限母集団を考える。こゝで

れ k は整数である。系統的標本を抽出するには、$x_1, \cdots, x_k$ から任意の一要素を選び、それから次々に k 番目毎の要素をとってゆけばよい。即ち、もし最初に選ばれた要素が x_i ならば、求める系統的標本には要素 x_i, x_{i+k}, $\cdots$ $x_{i+(n-1)k}$ が含まれる。このタイプの標本は

(i) 単純無作為標本あるいは層化無作為標本よりも抽出及び取扱いが簡単であること。

(ii) 標本が母集団全体にわたって一様に配置されるという事が直観的に示されること等の理由から、実際的に可成り数多く用いられる。しかし、この系統的標本にも、比較し得る単純無作為もしくは層化無作為標本との相対的精度について研究すべきことが数多く残されている。おそらく最も当を得た比較は系統的標本と層当り1個の要素を抽出する層化無作為標本とを比較することであろう。後者の場合母集団は n 個の層 $\{x_1, \cdots x_k\}$, $\{x_{k+1}, \cdots, x_{2k}\}$ $\cdots\cdots$ に分割され、この各層から無作為（random）かつ独立に一つの要素を抽出する。この型の標本は多くの点で系統的標本と似通っている。二つとも母集団を n 個宛の要素からなる n 個の層に分割し、各層から1つの要素を抽出する。その上どちらの標本と、少なくともその推定値が要素 x_i のどの様な母集団に対しても不偏（unbiased）であるという意味においては、平均値のまわりの抽出分散（sampling variance）の不偏推定法（unbiased estimate）のデータを与えることはできない。系統的標本の性質に関する最初の完全な研究は W. 今及び Madow (1) によって行なわれた。特にこれらの著者は、色々な型の有限母集団に対して、上述の様な系統的標本と層化無作為標本の精度の比較を行なった。母集団の要素が直線 $x_i = i$ の上にある場合層当り1要素の層化無作為標本の方が系統的標本より精度の高い事が示された。もし母集団が周期分布をすれば、n が周期の整数倍のときには層化無作為標本の方がずっと優れているが、n が周期の2分の1の奇数倍のときには系統的標本の方が優れている。

著者等は、母集団に傾向（trend）及び周期函数（periodic function）が同時に含まれている様なより複雑な場合についても考察した。この論文の目的は、色々なサムプリングで非常に多く見られる他のタイプの母集団について、同様な比較を行なうことである。この母集団は隣接する要素を群にまとめたとき、群の大きさが増すにつれて群内の要素間分散が急激に増加するようなものである。これまで長い間この型の母集団はブロック内のプロット間分散が、ブロックの大きさにともに増大するような圃場実験の研究に適用できるものと考えられてきた。n 回の均一性試験（uniformity trial）から得たデータを要約し Fairfield Smith [2] はこの考え方を証明し、増加率を推定することのできる実験的な関係を導いた。最近の広範な標本調査のいくつかに対する論文にも、同様な母集団が考察されている。このようにして Jessen は農場の母集団（farm population）のサムプリング法の研究で grid 内の農場間の分散は grid サイズの大きさの単調増加函数となるという法則を見出し、抽出単位（sampling unit）中に含まるべき最適の農場数を推定するために用いた。Mahalanobis [4] は大規模標本調査（large scale sample survey）の広範な研究の中で Fairfield Smith と同じ法則を独立に展開した。Hansen と Hurwitz [5] は、多くの現実的母集団の典型的なものでは、集落（cluster）内の分散が集落サイズの大きさとともに増大するということを指摘している。この他にも多くの参考文献をあげることができる。

2. 母集団の明確化

群内分散が群の大きさが大きくなるにつれて増加する状態を表わすには色々な数学的模型を設けることができる。例えば x_i に対して要素 x_i が規則的に変化している様な色々な母集団から抽出された要素 x_i を考えることができる。従って x_i は同一母集団に属するが、それ

には系列的相関がある（serial correlated）と考える事も出来る。簡単の為に，これからは x_i と x_{i+u} の間の系列相関は u のみに依存するある量 ρ_u であると仮定する。そうすると ρ_u が正で，u の単調減少函数であれば直観的に（後で証明する）要素 $x_i, x_{i+1}, \cdots, x_{i+k}$ からなる群内分散は k の単調増加函数となる事が予想される。この模型は我々の目的によくあてはまる様に思われる。というのは，多くの著者が分散増大現象の根拠としてあげているのは，x の間の相関が常に正となることであるからである。上記の特徴づけはある点に関して制約を受ける。普通の大きさの実際的な有限母集団で，ρ が厳密に単調であるという仮定は現実的でない様に思われる。一方コレログラムにはっきり下降の傾向（trend）が現われている場合でも，この傾向のまわりの個々のばらつきによってコレログラムが厳密には単調といえない事がある。更に適切であると思われるのは，有限母集団自体を単調な ρ をもつある無限母集団からの標本と考えることである。

私はこのような態度は上に引用した著者達の考えと一致していることを信ずるものである。論文を理解したところによれば，彼等は分散法則（variance law）が理想化された母集団において成立するものと考えている。このように，系統的標本と層化無作為標本との比較は，一つの有限母集団のみについて行なうのでなく，単調減少の ρ を有する無限母集団から抽出された有限母集団の平均について行なわれるのである。個々の有限母集団に対する結果は，それらの中のいが，期待値 ρ のまわりで変動するから，平均的な結果とはかけ離れたものになる。有限母集団が大となる従い，結果は平均的な結果に次第に接近するようになる。このことから要素 x_i，$i = 1, 2, \cdots\cdots, nk$，は

$$E(x_i) = \mu \qquad E(x_i - \mu)^2 = \sigma^2$$

$$E(x_i - \mu)(x_{i+u} - \mu) = \rho_u \sigma^2$$

こいでいくつなるときは常に $\rho_u > 0$, $\rho_v > 0$ なる母集団から抽出するものと仮定する。

4. 有用な幾つかの予備公式

特定の有限母集団の平均値を $\bar{X}$ とすると，分散分析で屢々有用である次の等式を容易に証明することができる。

(1) $(kn)\sum_{i=1}^{kn}(x_i - \bar{X})^2 = \sum_{i<j}^{kn}(x_i - x_j)^2$

(x_i, x_j) には $(kn)(kn-1)/2$ 個の対をとることができるから上式から

(2) $\sum_{i=1}^{kn}(x_i - \bar{X})^2 = \frac{(kn-1)}{2} E(x_i - x_j)^2 = \frac{(kn-1)}{2} E\{(x_i - \mu) - (x_j - \mu)\}^2$

こいで E は有限母集団の全体にわたってとるものとする。いま二次式を展開してすべての有限母集団について平均する。$(kn)(kn-1)/2$ 個の組合せ（combination）中には，j が i より 1 だけ多いものが $(kn-1)$ 個，j が i より 2 だけ多いものが $(kn-2)$ 個等々となっているから

(3) $E\sum_{i=1}^{kn}(x_i - \bar{x})^2 = (kn-1)\sigma^2\left\{1 - \frac{2}{(kn)(kn-1)}\sum_{u=1}^{kn-1}(kn-u)\rho_u\right\}$

これに対応して，k 個の連続した要素からなる 1 つの層内平方和の期待値をとるには，(3) の (kn) の代りに (k) を入れればよい。この結果は n 個の層の何れにおいても同じであるから

(4) E（層内の S, S）$= n(k-1)\sigma^2\left\{1 - \frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)\rho_u\right\}$

を得る。

公式 (3) で (kn) の代りに n，n の代りに (kn) を用いれば同様な計算から，特定の系統的標本内の平方和の期待値が得られる。

何となれば，標本内には n 個の要素があり，連続した要素間の相関は $\rho_1, \rho_2 \cdots\cdots$ でなく，$\rho_k, \rho_{2k}, \cdots\cdots$ だから。この結果は k 個の系統的標本の何れについても同じである。故に

(5) E（系統的標本内の S，S）$= k(n-1)\sigma^2\left\{1 - \frac{2}{n(n-1)}\sum_{u=1}^{n-1}(n-u)\rho_{ku}\right\}$

5　無作為標本に対する平均的分散

有限母集団の平均に対する単純無作為、層化無作為、系統的標本の平均的分散を表わすのに σ_r^2, σ_{st}^2, σ_{sy}^2 なる記号を用いる。

この平均はこれまでの節で詳述した無限母集団からとり出したすべての有限母集団にわたってとられたものである。

無作為標本との比較は、ここでの主目的ではないが興味ある問題であるので一緒に論ずることとする。

1つだけの有限母集団の無作為標本に対する平均値の分散は多くの著者によって

$$(6)\quad \frac{1}{n}\cdot\frac{(kn-n)}{(kn-1)}\cdot\frac{1}{kn}\sum_{i=1}^{kn}(X_i-\bar{X})^2$$

であることが示されている。ここで $\bar{X}$ は有限母集団の平均値である。(3)から

$$(7)\quad \sigma_r^2=\frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{1-\frac{2}{(kn)(kn-1)}\sum_{u=1}^{kn-1}(kn-u)f_u\right\}$$

6. 層化無作為標本の平均的分散

層化無作為標本の平均値を $\bar{X}_{st}$ とすれば定義から $\bar{X}_{st}$ の抽出分散（sampling variance）は

$$(8)\quad E(\bar{X}_{st}-\bar{X})^2$$

である。

まず1つの有限母集団のみに対するこの値の平均を考える。n個の層の平均値を $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n$ とし、また各層から選ばれた要素を $X_{1j}, X_{2j}, \cdots\cdots X_{nj}$ とする。このとき

$$\sum_{i=1}^{n}X_{ij}=n\bar{X}_{st}\quad\text{で}\quad\sum_{i=1}^{n}\bar{X}_i=n\bar{X}$$

だから(8)は

$$(9)\quad \frac{1}{n^2}E\left\{(X_{1j}-\bar{X}_1)+(X_{2j}-\bar{X}_2)+\cdots+(X_{nj}-\bar{X}_n)\right\}^2$$

とかける。

この有限母集団から得られる k^n 個のすべての標本についての平均をとる。Cross product term はすべて消える。何故なら、例えば、X_{ij} は $X_{21}, X_{22}, \cdots, X_{2k}$ と同じ回数だけ現れるから、従って単一の有限母集団に対する分散は

$$(10)\quad \frac{1}{kn^2}\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij}-\bar{X}_i)^2$$

となる。この平方和は勿論単なる層内の平方和である。

よって（4）から

$$(11)\quad \sigma_{st}^2=\frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{1-\frac{2}{k(k-1)}\sum(k-u)f_u\right\}$$

7. 系統的標本の平均的分散

$\bar{X}_{sy}$ を代表的な系統的標本の平均値とするとき、単一の有限母集団の分散は

$$(12)\quad E(\bar{X}_{sy}-\bar{X})^2=\frac{1}{kn}\left\{n\sum(\bar{X}_{sy}-\bar{X})^2\right\}$$

である、ここで和は k 個の系統的標本のすべてにわたる。標本間平方和は、（母集団の全平方和－標本内平方和）に等しいから、(12)は、

$$(13)\quad \frac{1}{kn}\sum_{i=1}^{kn}(X_i-\bar{X})^2-\frac{1}{kn}\quad（系統的標本内SS）$$

に等しい。

すべての有限母集団について平均をとるため、オ1項及びオ2項にそれぞれ(4)と(5)を代入して

$$(14)\quad \sigma_{sy}^2=\frac{(kn-1)}{kn}\sigma^2\left\{1-\frac{2}{(kn)(kn-1)}\sum_{u=1}^{kn-1}(n-u)f_{ku}\right\}$$
$$-\frac{(n-1)}{n}\sigma^2\left\{1-\frac{2}{n(n-1)}\sum_{u=1}^{n-1}(n-u)f_{ku}\right\}$$

を得る。

変形して

$$(15)\quad \sigma_{sy}^2=\frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{1-\frac{2}{kn(k-1)}\sum_{u=1}^{kn-1}(kn-u)\rho_u+\frac{2k}{n(k-1)}\sum_{u=1}^{n-1}(n-u)\rho_{ku}\right\}$$

上述の式と記号はMadow が用いたものと違っていることに注意。Madow は ρ と σ^2 を単一の有限母集団について考え、単一の有限母集団の抽出分散（sample variance）を論じた。

8. 無作為標本と層化無作為標本の相対的な精度

まず幾つかの一般的なことを述べる。

(7), (11), 及び(15)の三種類の標本の相対的な効率は $\sigma_{\bar{x}}^2$, σ_{sy}^2 に現れる ρ の一次函数にのみ関係することがわかった。どの場合についても ρ の係数の和が1に等しいことは容易に証明できる。

無作為標本の場合この一次函数には、log が増加するにつれて直線的に減少し、標本の大きさとは無関係でただ有限集団の要素数 $N=(kn)$ にのみ依存する log $(kn-1)$ までのすべての系列相関係数（Serial correlation coefficient）が含まれる。

層化無作為標本では log $(k-1)$ までの系列相関（serial correlation）が現れる。ここで k は層内の要素数である。(15) に示した様に系統的標本に対する公式は二つの一次函数（linear function）に分解される。

第1のものは、すべての係数が $(kn-1)/(k-1)$ 倍になっている事を除けば無作為標本に出てくるものと同じである。2番目は正の符号をもち、log が k の倍数であるような相関（correlation）が含まれる。従ってこの公式には ρ についての制限は全然不要である。ρ が正で単調減少の場合を考えれば次の Lemma が有用である。

Lemma　$S_i\ (i=1,\cdots,m)$ が正で、単調減少即ち

$S_i\geqq S_{i+1}>0$ で　$(\alpha_1+\alpha_2+\cdots+\alpha_m)=0$　なら

(16) $L=\alpha_1\rho_1+\alpha_2\rho_2+\cdots+\alpha_m\rho_m\geqq 0$　（可能な ρ のすべての組について）なる為の必要かつ十分な条件は

(17) $\alpha_1+\alpha_2+\cdots+\alpha_i\geqq 0\qquad i=1,2,\cdots(m-1)$

なることである。

（証明）何故なら　$\rho_i=\rho_{i+1}+S_i$（ここで仮定から $S_i>0$）とおくと　$\alpha_1,\alpha_2\cdots,\alpha_{m-1}$ の代りに次に $\beta_1,\beta_2\cdots,\beta_{m-1}$ を代入してゆくと

$$(18)\quad L=\alpha_1S_1+(\alpha_1+\alpha_2)S_2+(\alpha_1+\alpha_2+\alpha_3)S_3+\cdots\cdots+(\alpha_1+\alpha_2+\cdots+\alpha_{m-1})S_{m-1}$$

となって、$(\alpha_1+\cdots+\alpha_m)=0$ だから最後の項 ρ_m が消える。

すべての S_i について $\geqq 0$ だから（17）が十分条件であることは明らか。もしも任意の i に対する S_i の係数が負なら、S_i と正にとり他を0とおくことによって L を負とすることができる。これで必要性の条件が得られた。

系（corollary）・S_i が強い意味で単調即ち $S_i>S_{i+1}$ で S_i の少くとも一つが0でなければ（17）は L が0以上であるための十分条件となる。

（証明）なぜなら（18）の S_i はすべて0より大であるから（17）によって S_i の係数はすべて正である，また S_i の少くとも1つの係数は0より大でなければならないが他の S_i は全部0となってもよい。

故に $L>0$

さて ρ_u が単調減少のとき

$$(19)\quad L(k)=\frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)\rho_u$$

が k の単調減少函数であることを示そう。

これは層化無作為標本の分散中に現れる一次函数である。

$$(20)\quad L(k)-L(k+1)=\frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)\rho_u-\frac{2}{(k+1)k}\sum_{u=1}^{k}(k+1-u)\rho_u$$

$$(21)\quad =\frac{2}{k(k^2-1)}\sum_{u=1}^{k}(k+1-2u)\rho_u$$

$L(k)$ と $L(k+1)$ の ρ_u の係数の和は1だから（21）の和は0である。従って Lemma が適用できる。しかし係数は $u\leqq(k+1)/2$ については すべて正で，$u>(k+1)/2$ についてはすべて負だから、（21）

の係数の最初の i 個の和が0よりも大きいことは明らかである。

故に

(22) $\quad \Delta(k)-\Delta(k+1)>0$

さらに，系から，もし p_u が強い意味で単調なら，$\Delta(k)$ は強い意味で単調である。p_u は負だからこの結果によって

(23) $$1-\frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)p_u<1-\frac{2}{(nk)(nk-1)}\sum_{u=1}^{nk-1}(nk-u)p_u$$

を証明できる。

結局任意の大きさの標本について，層化思作為標本の平均的な分散は思作為標本のそれよりも大きくはならない。また，層化標本の思作為標本に対する効率（efficiency）は層の大きさが小さくなるにつれて，即ち標本の大きさが大となるにつれて単調に増大する。勿論これらは予期されなかったものではない。

(22) の方程式はまた，第3節で注意した単調減少な p についての結果即ち層内の平均的分散は，層の大きさが増すにつれて急速に増大するという性質をも説明するものである。何故なら，層内の平方和の自由度が $n(k-1)$ なら，上の式(4) は層内の平均的分散が

(24) $$\sigma^2\left\{1-\frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)p_u\right\}=\sigma^2\left\{1-\Delta(k)\right\}$$

であることを示す。

9. 系統的標本と思作為標本との比較

母集団の形について立入った限定を設けない限り，思作為標本と比較したときの系統的標本の効率について一般的な結果が得られないことは上述の研究から明らかである。Lemma を応用するため，公式 (7)，(11)，(15) に現れる p の一次箱数の最初の i 個の係数の和を求めてみる。簡単な方法でこれらの和が

$$\Sigma_r=\frac{i(2nk-i-r)}{nk(nk-1)},$$

(25) $$\Sigma_{st}=\frac{i(2k-i-1)}{k(k-1)},\qquad 1\leqq i\leqq(k-1)$$

$$1,\qquad i\geqq k$$

$$\Sigma_{sy}=\frac{i(2nk-i-1)}{nk(k-1)}-\frac{rk(2n-r-1)}{n(k-1)}.$$

であることがわかる。ここで r は $(r+1)k\geqq i>rk$ なる整数。Lemme から，$\sigma_{sy}^2\leqq\sigma_{st}^2$ をいうには $\Sigma_{sy}\geqq\Sigma_{st}$ が各 i について成り立つことを示せばよい。いまもし $i<k$ ならば r は0だから $n=1$ のとき以外は

(26) $$\Sigma_{sy}>\Sigma_{st}>\Sigma_r\qquad i=1,2,\cdots\cdots,(k-1)$$

$n=1$ のときはこの三つはすべて等しい。

しかし i が k の整数倍 rk なる場合には

(27) $$\Sigma_r=\frac{r}{n}\left[1+\frac{(n-r)k}{(nk-1)}\right],\quad \Sigma_{st}=1,\quad \Sigma_{sy}=\frac{r}{n}$$

なることがわかるから

(28) $$\Sigma_{st}>\Sigma_r>\Sigma_{sy}$$

である。

前記 Lemma の条件は systematic sample に対しては満足されないから単調減少な p をもつ母集団全体に対する一般的定理は存在しない。(26) と系から，$p_u=0$，$u>(k-1)$ なる class に含まれるような性質の母集団に対しては，系統的標本は層化無作為標本よりはるかに効率がよいことがわかる。一方 (28) によれば，p の最初の j 個が等しく残りが0であるような母集団については，系統的標本の分散は無作為標本のものより大となる。

これらの二つの結果を一諸にすると，最初の j 個の p が等しく残りが0であるような母集団では大きさ j の層をもつ系統的標本は比較し得る無作為標本より精度が悪く，また $(j+1)$ なる大きさの層をもつ標本は対応する層化無作為標本より精度がよい。その様な母集団が実際に出てくる事がないにしても，この結果は系統的標本の平均値の分散の標本の大きさに対するグラフ（graph）が、無作為標

本と同じ様な規則性は表われそうにないことを示している。

10. コレログラムが上に凹な母集団

更に研究すれば系統的標本と無作為標本の相対的精度をきめる決定的な因子は，一次階差（first difference）ではなく，むしろ二次の階差（second difference）であることが示される。

次の結果を証明してみよう。

定　理

$$f_i \geq f_{i+1} \geq 0, \qquad i = 1, 2, \cdots, (kn-1)$$

で

$$\delta_i^2 = f_{i-1} + f_{i+1} - 2f_i \geq 0 \quad i = 2, 3, \cdots, (kn-2)$$

なるようなすべての無限母集団については，任意の大きさの標本に対して

$$\sigma_{sy}^2 \leq \sigma_{st}^2 < \sigma_r^2$$

なり立つ。

さに $\delta_i^2 = 0, \quad i = 2, 3, \cdots, (kn-2)$ でない限り

$$\sigma_{sy}^2 < \sigma_{st}^2$$

但し $i = 2, 3, \cdots, (kn-2)$

この結果は f_u の一次函数を二次の階差で表わし，二次階差に適用できる新しい Lemma を構成することによって証明できる。

もう一つの方法は，より簡単でまたおそらくよりわかりやすい。

f_u は単調減少だから section 8 の結果から $\sigma_{st}^2 \leq \sigma_r^2$。上の (11) において特定の有限母集団の系統的標本の平均値の分散は，

$$\frac{1}{kn}\sum_{i=1}^{kn}(X_i - \bar{X})^2 - \frac{1}{kn}(\text{系統的標本内の全 } S.S.)$$

$$(29) \qquad = \frac{1}{kn}\sum_{i=1}^{kn}(X_i - \bar{X})^2 - \frac{1}{n}(\text{系統的標本内の平均的 } S.S.)$$

と表わすことができる。

層化無作為標本についてもこれに対応する方程式がなり立つ。というのはもし平均値 $\bar{X}_{st}$ をもつ任意の層化無作為標本の要素を X_{ij}，

$X_{2j}, \cdots, X_{nj}$ とすると

$$(30) \quad \Sigma(X_{ij} - \bar{X})^2 = \sum_{i=1}^{n}(X_{ij} - \bar{X}_{st})^2 + n(\bar{X}_{st} - \bar{X})^2$$

こで k^n 個のすべての標本について平均すれば，

$$(31) \quad \frac{1}{k^n}\sum_{L=1}^{k^n}(X_L - \bar{X})^2 = (\text{平均的標本内 } S.S.) + nE(\bar{X}_{st} - \bar{X})^2$$

一番右の項は層化無作為標本の分散の n 倍だからこいでも (29) と同様な結果が得られる。

結局系統的標本の平均的な"組内 (within)"平方和が層化無作為標本の"組内 (within)"平方和より大きいかあるいは等しい場合には

$$\sigma_{sy}^2 \leq \sigma_{st}^2$$

となる。いま (2) から，(kn) の代りに n をおきかえればこれらのそれぞれの平均は

$$(32) \quad \frac{(n-1)}{2}E(X_{ij} - X_{kj})^2$$

に等しい。こで X_{ij} 及び X_{kj} はそれぞれ i 層及び k 層からとった標本の要素で，平均は層の対のすべての可能な組についてとるものとする。

我々は層の対を固定して考え $k - i = u$ とおくことにする。系統的標本で i 層及び k 層の対応する要素は常に (ku) 要素だけ離れているから，

$$(33) \quad E_{sy}(X_{ij} - X_{kj})^2 = 2\sigma^2(1 - \rho_{ku})$$

層化無作為標本では二つの層の要素を一つ宛組合せる仕方の数は k^2 例ある。$(ku-k+1)$ 要素離れている対は1つで $(ku-k+2)$ 要素だけ離れている対は2個等々……である。

対の個数は k まで直線的に増加しそれ以上では直線的に減少して $(ku+k-1)$ だけ離れた最後の対の個数は1個となっている。これから

$$(34) \quad E_{st}(X_{ij} - X_{kj})^2 = 2\sigma^2\left\{1 - \frac{1}{k^2}\sum_{i=-(k-1)}^{(k-1)}(k - |i|)\rho_{ku+i}\right\}$$

故に $\sigma_{sy}^2 < \sigma_{st}^2$ を完全に証明するには

$$(35)\quad \sum_{i=(k-1)}^{(k-1)} (k-|i|)\, f_{ku+i} - k^2 f_{ku} \geqq 0$$

を $u = 1, 2, \cdots, (n-1)$ 即ち層の任意の組に対して示せば十分である。これは

$$(36)\quad \sum_{i=1}^{(k-1)} (k-i)(f_{ku+i} + f_{ku-i} - 2f_{ku}) \geqq 0$$

とかける。しかし $\delta_{ku}^2 = f_{ku-1} + f_{ku+1} - 2f_{ku}$ を第二中央階差 (second central difference) とすると容易に

$$(37)\quad f_{ku+i} + f_{ku-i} - 2f_{ku} = \sum_{j=-(i-1)}^{(i-1)} (i-|j|)\, \delta_{ku+j}^2 \geqq 0$$

を証明できる。何故なら仮定から $\delta_j^2 \geqq 0$, $j = 2, 3, \cdots, (kn-2)$ だから。このことは，任意の固定された層の対に対しては，系統的標本の要素間の分散が層化無作為標本の要素間分散よりも大きいから，等しいことを示している。これから全平均に対する結果も得られる。故に，$\sigma_{sy}^2 < \sigma_{st}^2$ 、まだ更に標本が1つ宛の場合を除いて，$\sigma_j^2 = 0$ でない限り明らかに $\sigma_{sy}^2 < \sigma_{st}^2$ である。

証明の本質的な部分は次の様である。i 番目及び j 番目の層の要素は、系統的及び層化無作為標本の何れにおいても，(ku) 要素丈離れている。層化無作為標本の二つの要素が $(ku+i)$ 要素だけ離れているとき，これらの相関は平均以下である。何故なら，$f_{ku+1} \leqq f_{ku}$ であるから，これはより多くの独立な情報を与えるからである。この要素間の分散は系統的標本の分散より $2\sigma^2(f_{ku} - f_{ku+i})$ だけ大きい。しかしこの様なことの起るのは要素が $(ku-i)$ 丈離れていて分散が系統的標本のものより $2\sigma^2(f_{ku-i} - f_{ku})$ だけ小さくなる場合と同数だから釣合がとれる。f_u が凹だから平均的な損失は釣合いがとれるかあるいは利益の方が多い。

Section 9 で論じた $f_u = f$, $u = 1, 2, \cdots, j$, $f_u = 0$, $u > j$ なる母集団では $\delta_j^2 < 0$, $\delta_{j+1}^2 > 0$, 他の u については $\delta_u = 0$ が得られる。

第二の差の符号が変っていることは層の大きさが j と $j+1$ の系統的標本の異なる性質を説明するものである。

上の定理は層化無作為標本に対する系統的標本の相対的精度が、n の単調函数で表わされることを証明するものでもなく，また，σ_{sy}^2 が n の増加につれて減少するということすら証明するものではない。実際何れの結果も成立たない様な母集団も存在する。これを次の section で述べることにする。実際的な応用を考えている場合には，f_u が上に凹でなければならないという様な条件が満たされないかもしれない、例えばこの条件はコレログラムが直線的 (linear) 即ち $f_u = (l-u)/l$ は場合には満足される。この f_u の形は，Wald [6] が経済のデータに適用し得ると考えた型のコレログラムである。凹になるという性質は Osborn [7] が森林及び土地利用調査に対して提案した $f_u = \theta^{-u}$ なる函数を、Fisher と Mackenzie が二つの観測所で測った週毎の雨量の相関を観測所間の距離を用いて表わした $f_u = \tanh(u^{-2/5})$ なる関係式についても成立つ。実際もし f_u が正にして表わされ、すべての u について連続なら，上に凹な函数はそれ自身自然的なものであることを示す。(naturally)

11. 直線的なコレログラム (linear correlogram)

コレログラムが (i) 直線的 (ii) 指数的 (exponential) な場合に得られる幾つかの結果を述べることは興味ある問題である。というのは、この両方の型は実際に出現する可能性の多い母集団だということがいわれているからである。

直線 (linear) の場合

$$(38)\quad f_u = (l-u)/l, \quad u \leqq l; \qquad f_u = 0 \quad u > l$$

もし $l > (nk-1)$ ならコレログラムは有限母集団の全範囲を通じて直線である。この場合には二次階差はすべて0だから，

$$\sigma_{sy}^2 = \sigma_{st}^2 < \sigma_r^2$$

なることが期待される。

$l < (nk-1)$ ならすべての第二階差は，正である σ_0^2 を除いては

すべて0となる。故に $\sigma_{sy}^2 < \sigma_{st}^2 < \sigma_r^2$ これらの各々の場合における結果は、(7), (11), (15) の基礎公式を初等的な方法で加え上げれば得られる。和をとることは詳しく述べない。

L が $(nk-1)$ なる場合には

$$(39)\quad \sigma_{sy}^2 = \sigma_{st}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\frac{k+1}{3L} \;:\; \sigma_r^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\frac{(nk+1)}{3L}$$

となる。比 σ_r^2/σ_{sy}^2 は $(nk+1)/(k+1)$ だから抽出比が大きい場合の外はこれは大体標本の大きさこれに等しい。このようにすべての無作為標本を上まわる非常に大きい効率 (efficiency) の利得 (gain) が得られる。

L く $(nk-1)$ なら公式はより複雑となる。まず k が L を考える。即ち抽出これに割合が10%以下の場合である。$N=nk$ なら

$$(40)\quad \sigma_r^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{\frac{3N(N-L)+(L^2-1)}{3N(N-1)}\right\}$$

$$(41)\quad \sigma_{st}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{\frac{3k(k-L)+(L^2-1)}{3k(k-1)}\right\} \qquad k > L$$

$$(42)\quad \sigma_{sy}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{\frac{3N(k-L)+(L^2-1)}{3N(k-1)}\right\} \qquad k > L$$

$\sigma_{sy}^2 < \sigma_{st}^2$ なることは明らか、また層化無作為標本に対する系統的標本の効率は標本の大きさが増すにつれて急速に増大することは容易に証明できる。

標本の大きさがもっと大きくなると $k \leq L$ となって (40) はそのまま成立する。一方、σ_{st}^2 はこいでは (39) と同じ公式で与えられる。σ_{sy}^2 に対する公式はもっと複雑である。

L を k で割った時の商の整数部分を g とし、残りを r とすると $L = (gk+r)$ となるから、公式は

$$(43)\quad \sigma_{sy}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{\frac{gk(k^2-1)+3rk(n-g)(k-r)+r(r^2-1)}{3NL(k-1)}\right\}$$

中括弧内の分子の最後の二項は、L が丁度 k で割り切れる時は0となることに注意せよ。また第二項の order は $nk=N$ で今の場合に

は、第一次より大きい影響を与えている。故に σ_{sy}^2 は L が k の倍数となるとき急激に低下する。実際 $L = gk$ なるときは (42') は

$$(43)\quad \sigma_{sy}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\frac{(k+1)}{3N} \qquad L = gk$$

となるから N が十分大きいときこの分散は0になる。σ_{sy} に対する公式 (39) の比較によって、$L = gk$ なる場合には系統的標本と無作為標本の効率の比率は N/L であることがわかる。これは N が十分大きいときには無限に増加する。普通 r が0でない場合には、N が大きいときの相対的効率の主要項は $(k^2-1)/3r(k-r)$ である。これは L と k の間の関係に従って幾分不規則に変動する。

説明のため $L=10$ で、有限母集団が十分大きく、従って、$1/n$ の項を無視し得る様な場合の数値を下に示す。

V_{st}, V_{sy} なる量は係数 σ^2/N を別にして対応する分散を表わす。層化標本の分散は抽出比が大きくなるにつれて急速に減少する。一方系統的標本の分散は $k=2,5,10$ で0に、相対的効率は ∞ となる。こうして中間の場合、即ち $k=3,4,6,7,8,9$ では分散及び相対的効率は抽出比と全然満足な関係を示さない。10%以下の標本では、表に表してない様な場合も含めて、相対的効率が $k=11$ の4から k が大きいときの1まで連続的に減少する。

12. 指数的コレログラム (Exponential Correlogram)

指数函数 (exponential) $f_x = e^{-\lambda x}$ の場合には結果はもっと規則的である。f の一次函数 (linear function) がそれぞれ $(1-x)^{-2}$ の展開式の有限個の項からなっている。もし σ_r^2 の和が

$$(44)\quad f(N,\lambda) = \frac{2}{N(N-1)}\left\{\frac{(N-1)e^{\lambda}-N+e^{-(N-1)\lambda}}{(e^{\lambda}-1)^2}\right\}$$

なら

$$(45)\quad \sigma_r^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{1-f(N,\lambda)\right\}$$

(46) $\sigma_{st}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{1-f(k,\lambda)\right\}$

(47) $\sigma_{sy}^2 = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left\{1-\frac{(N-1)}{(k-1)}f(N.\lambda)+\frac{k(n-1)}{(k-1}f(n,k\lambda)\right\}$

なることがわかる。

標本の大きさが増すにつれて，系統的標本の分散は直線的に減少し同時に層化抽出（*stratified sampling*）に対する相対的効率（*relative efficiency*）も直線的に増大する。

効率の利得（*gain*）の大きさを知るためkの反対が小が大きい場合を考える。この場合相対的効率は実際にはk，n，λの函数であるが，これは殆んど全部が（kλ）なる量のみによってきまる。即ち系統的標本の次々の層の調査対象（*item*）同の相関（*correlation*），$e^{k\lambda}$にのみ依存する。もしも=（kλ）とすると，$\sigma_T^2=\sigma^2/n$，

(48) $\sigma_{st}^2 = \frac{\sigma^2}{n}\left\{1-\frac{2}{t}+\frac{2}{t^2}-\frac{2e^{-t}}{t^2}\right\}$

(49) $\sigma_{sy}^2 = \frac{\sigma^2}{n}\left\{1-\frac{2}{t}+\frac{2}{(e^t-1)}\right\}$

を得る。

Table 1. コレログラムが直線の場合の因数σ^2/Nを除いた分散と系統的（*systematic*）と層化無作為標本の相対的な効率（*relative efficiency*）

k	2	3	4	5	6	7	8	9	10	11	20
% sampled	50	33	25	20	17	14	12	11	10	9	5
V_{st}	10	.27	.50	.80	1.17	1.60	2.10	2.67	3.30	4.00	11.65
V_{sy}	0	.20	.40	0	.80	1.20	1.20	.80	0	1.00	10.00
V_{st}/V_{sy}	∞	1.33	1.25	∞	1.46	1.33	1.75	3.33	∞	4.00	1.16

Table に幾つかのe^{t}に対する相対的効率を与えてある。e^{-t}は次々の層の調査対象（*item*）同の相関である。

この相対的効率にはρが1に近すく（*tend*）するときの2なる最低値がある。こうしてこれはρが0となるときの1にまでゆっくりと減少する。効率の利得（*gain*）はρが0.1をこえる場合には相当大である。

Table 2

ρ	.9	.8	.7	.6	.5	.4	.3	.2	.1
$\sigma_{st}^2/\sigma_{sy}^2$	1.96	1.90	1.84	1.78	1.71	1.64	1.55	1.46	1.33

単一の標本からは系統的標本でも層化無作為標本でも誤差の不偏推定値は得られないということをsection 1で注意しておいた。これは誤差の推定が全然できないということではない。しかし任意の推定値は，抽出される母集団の形についてのはっきりした仮定にもとづいていなければならない。これらの仮定が誤っている場合には推定値が全然誤ることになる。

例えば，コレログラムを指数型（*exponential*）と仮定すると，n，kが大きいとき公式（47）あるいは（49）は1つの系統的標本から誤差を推定する場合の適切な基礎となり得る。系統的標本の次々の調査対象（*item*）同の相関（*correlation*）はe^{-t}の推定値となるから，tの推定値が得られる。またもし1/nの項が無視し得るなら，系統的標本の組内（*within*）の平均平方（*mean square*）はσ^2の不偏推定値であることがわかる。

指数型（*exponential*）という仮定が正しいときは（49）にこれを代入して単一の系統的標本の分散の一致推定量（*consistent estimate*）が得られる。こうして層化及び単純無作為抽出をこえる効率の利得も推定することができる。

By. W. G. Cochran A.M.S. Vol. 17 (1946) より

引用文献

[1] W. G. and L. H. Madow, On the theory of systematic sampling, Annals of Math. Stat., Vol. 15 (1944) pp. 1-24.

[2] H. Fairfield Smith, An empirical law governing soil heterogeneity, Jour. Agr. Sci., Vol. 28 (1938). pp. 1-23

[3] R. J. Hessebm Statistical investigation of a sample survey for obtaining farm facts, Iowa Agr. Exp. Sta., Res. Bull., No, 304, (1942).

[4] P. C. Mahalanobis, On large scale sample surveys, Roy. Soc. Phil. Trans., B.231 (1944), pp. 329-451.

[5] M. H. Hansen and W. N. Hurwitz, On the theory of sampling from finite populations, Annals of Math. Stat., Vol. 14 (1943), pp. 333-362.

[6] H. Wald, A study of the analysis of stationary time series, Uppsala (1938).

[7] J. G. Osborne, Sampling errors of systematic and random surveys of covertype areas, Amer. Stat. Assoc. Jour., Vol. 37 (1942), pp. 256-270.

[8] R. A. Fisher and W. A. Mackenzie, The correlation of weekly rainfall, Quart. Jour. Roy. Met. Soc., Vol. 48 (1922), pp. 234-245.

4. 系統的抽出の理論について

On the theory of Systematic Sampling. II

1. 要約および序

前論文 (1) では系統的抽出の問題を研究し、困難する分散をえた。母集団の型を幾つか想定し、これらについて系統的抽出計画の効率を単純および層化無作為標本の効率を比較した。系統的抽出計画では無作為抽出計画でのものと反対に、標本の大きさが増すと分散が増加することがわかった。この可能性は (2) において正しいことが証明された。

Cochran (3) によって与えられた系統的抽出計画の一つの研究方法は、母集団の要素を確率変数と仮定し、期待分散の一つに問題を転換することによってこの結果をある程度まで解決した。彼はこれらの確率変数のコレログラムが上に凸なら、系統的計画の期待分散は層化計画の分散より小となり、また多くの場合相当小さくなるということを示した。

この論文では前論文の結果を大きさが等しい場合、および等しくない場合の集落の系統的抽出に拡張することにする。二次元における系統的抽出についても若干の説明を行なった。

Sec.2 ではサムプリングの多くの分野でかなり広く応用できる二つの定理を導びいた。サムプリング理論の研究者の間では、これらの定理の成立は当然とされていたものであるが、しかしこれについて言及している文献は今まで存在しなかった。

Sec.3 では、標本調査の設計の際、層間に負の相関を与えるようにすべきだという注意 (I.P. 10) の内容を明らかにした。定理 5 では相関が負となるための十分条件を求めた。Sec. 4 の Lemma および定理 4 によった、定理 3 が実際の場合に適用できることになる。これらの結果の応用として、我々は母集団のコレログラムが上に凸で、各層から 1 要素を無作為にとり出す抽出の形式が最適とな

るように層を定めれば、各層からの独立な無作為抽出よりも更に効率の良い系統的抽出計画を定めることができる。

Sec 5 においては、大部分それまでの節の一般定理の応用である集落の系統的抽出に対する色々な結果を導いた。大体において、この結果は系統的抽出の効率が無作為または層化無作為抽出のものより高くなることが予想される条件を、公式として示している〔1〕および〔3〕と性格を同じくするものである。しかし我々はこれらの公式を母集団のタイプを指定して応用することはしなかった。〔1〕および〔3〕から、このことは有用でもあり、またそのような研究は一般的な論文の例で扱うより重要なタイプの母集団またはデータについて進める方が一層貴重だということがすでにわかっている。

2. 無作為標準量および条件付期待値

標本は殆んどの場合幾つかの段階で抽出される。たとえばあるシティから世帯の標本を抽出するには屢々次のような二段抽出法が用いられる。

a. 各ブロックの位置を示すシティの地図を手に入れ、その時点のものに修正する。

b. この地図を用いてシティのブロックの標本を選ぶ（これが第1段）

c. 第1段で選んだブロックの世帯から世帯の副次標本を選ぶ（これが第2段）

この節では多段抽出の平均値と分散の評価のための一般的方法を与えることにする。この方法は各段階で生ずる分散への寄与を同時に出せるという利点がある。その上ここに示す定理は興味が多段抽出にない場合でも分散の計算においても有用なものである。この定理はサムプリングにおいて広い応用範囲をもつから一般的な用語で提出することにする。

ある操作を行った結果、確率 $P_1 \cdots\cdots P_m$ で m 個の状態 $A_1 \cdots\cdots A_m$

が起りうると仮定できるとき、この操作の結果を確率準変数 A^* という。ここで

$$P\{A^* = A_i\} = P_i \qquad \sum_{i=1}^{m} P_i = 1$$

で $P\{A^* = A_i\}$ は"確率準変数 A^* が A_i なる状態となる確率"という。

操作の1例はブロックの標本を抽出する手続きである。シティに N 個のブロックがあって、その中の n 個づつのブロックの組がいずれも同じ機会で標本となるような方法で選び出すとすると、C_n^N 個の可能な標本がありうる。この場合 $m = C_n^N$ で、C_n^N 個の可能な標本は"ブロックの標本を選んだ結果" A^* の m 個の状態である。さらにブロックの可能な標本がそれぞれ同じ機会で選ばれるときには

$$P\{A^* = A_i\} = \frac{1}{C_n^N} = \frac{1}{m}$$

である。

確率準変数 A^* はまた確率変数の可能な値の一つで示すこともできる。確率変数 x^* のとりうる値が $z_1 \cdots z_m$ でその確率が $P_1 \cdots P_m$ なら A^* が A_i となる状態を"$x^* = z_i$"なる場合と定義できる。

このように確率準標の概念には標本抽出の際にみられる二つのタイプの無作為性が含まれている。

x' を確率変数とする。このとき確率準変数 A^* に関する x' の条件付期待値とはそのとりうる値が $E(x' \mid A_i)$; $i = 1 \cdots m$ で確率が P_i であるような確率変数 $E^*(x' \mid A)$ をいう。すなわち

$$P\{E^*(x' \mid A) = E(x' \mid A_i)\} = P_i = P\{A^* = A_i\}$$

ここで

$$(2.1) \qquad E(x' \mid A_i) = \sum_{j=1}^{N_i} x_{ij} p_j(A_i)$$

で、x_{ij} は A_i が起ったときに x' のとりうる N_i 通りの値のうち、第 j 番目のものであり、

$$p_j(A_i) = P\{x' = x_{ij} \mid A_i\}$$

は A_i が起ったとき $x' = x_{ij}$ となる確率である。注意すべきことは

$$p_{ij} = P\{x' = x_{ij}\}$$

なら

$$p_{ij} = P\{x' = x_{ij},\ A^* = A_i\}$$

となることである。というのは $x' = x_{ij}$ は A_i の生じたことを示すからである。そうすると

(3.2) $$p_i \cdot p_j(A_i) = p_{ij}$$

である

ここで定理1と2を証明なしで述べておく。証明は容易である。

定理1。確率変数 $E^*(x' \mid A)$ の期待値は Ex' である。すなわち

$$E\{E^*(x' \mid A)\} = Ex'$$

$\sigma^*_{x'y'|A}$ なる記号によって、とりうる値が $\sigma_{x'y'|A_i}$, $i = 1, \cdots\cdots$ なる確率変数を表わす。ここで

$$\sigma_{x'y'|A_i} = E\{(x' - E(x' \mid A_i))(y' - E(y' \mid A_i)) \mid A_i\}$$

であり

$$P\{\sigma^*_{x'y'|A} = \sigma_{x'y'|A_i}\} = p_i = P\{A^* = A_i\}$$

である。すなわち

$$\sigma^*_{x'y'|A} = E^*\{(x' - E^*(x' \mid A))(y' - E^*(y' \mid A)) \mid A\}$$

さらに記号 $\sigma E^*(x' \mid A) E^*(y' \mid A)$ は"二つの確率変数 $E^*(x' \mid A)$ と $E^*(y' \mid A)$ の共分散"を表わす。分散についての対応する定義は上の y' を x' でおきかえることによってえられる。

定理2 x', y' が確率変数なら

$$\sigma_{x'y'} = E^* \sigma_{x'y'|A} + \sigma E^*(x' \mid A) E^*(y' \mid A)$$

および

$$\sigma^2_{x'} = E\sigma^2_{x'|A} + \sigma^2 E^*(x' \mid A)$$

である

P_{ij}, p_i, $p_j(A_i)$ は指定されていないから定理1と2は任意の二段抽出計画で成立することに注意。定理1および2の多段抽出計画への拡張はすぐわかるが、実際にはこの定理を何回かにわけて適用する方が簡単なことが多い。定理1と2の応用を示すのはたやすいがしかしこれは本論文の目的には本質的なものでない。序に述べた通りこれらの二つの定理は長い間サムプリングの伝承とよびうるものの一つとなっていたものである。

3 層化抽出と負の相関ならびにその系統的抽出に対する応用

層化された母集団からのサムプリング計画を論ずる際には、仮定として、x' が推定値で $x' = x'_1 + \cdots + x'_L$（ここで x'_j は L 個の層の j 番目のものから生じた x' への寄与である）なら、確率変数 x'_i と x'_j $j \neq i$ が独立となるように抽出するものとする。

〔1, P. 13〕で、もし母集団が層化されており、異なる層からの寄与が負の相関をもつように要素を抽出すれば、この寄与が独立でしかも層内の共分散が同じ場合より推定値の精度は低くなることを注意した。もちろんこの結論は

$$\sigma^2_{x'} = \sum_{i=1}^{L} \sigma_{x_i} x_i$$

だからもし

(3.1) $$C = \sum_{i \neq j} \sigma_{x_i} x_j < 0$$

なら $C = 0$ の場合より $\sigma_{x'}^2$ は小となることからすぐわかる $C < 0$ なる次々はこの標本計画は"負の相関"をもつとよぶ

容易にわかるように、任意の母集団はそれ自身1つの標本があると考えることができる。すなわち現実の母集団を定める力の過り出す可能な母集団からとられた標本である。抽出計画は支配力の知識である種の過去の経験をもととして選ぶことが多いから、特定の母集団の期待値や分散のみならず、同じ力で決定されるすべての可能な母集団上でのそれらの期待値をも考えるのが現実的である。標本計画の期待分散を考えることの有用性は Cochran〔3〕が一つの例で示したところである。彼はそれ自身、確率変数であるところの母集団の要素 $x_1, \cdots\cdots, x_n$ を考え、$Ex_i = \mu$ および $E(x_i - \mu)^2 = \sigma^2$ と仮定した。彼の目的からいうと更に $u > 0$ のとき $E(x_i - \mu)(x_{i+u} - \mu) = \rho_u \sigma^2$ とおくと都合がよい。これによって、彼はコレログラムについての現実的な仮説、すなわち特定の母集団ではこれまで普通とは考えられなかったところの"ρ_u は u の 数である"という

仮説を設けることができた。このようにして彼は無作為および層化無作為計画と比較したときの系統的計画の期待効率についての一般的な結論をえた。

この論文では与えられた母集団に対する期待値や分散だけでなく，母集団の要素自身が確率変数であるという仮定のもとにこれらの期待値および分散の期待値をも考察することにする。我々は母集団の要素を確率変数と考えた場合の期待値を表わすのに E なる記号を用い，また前と同じく特定の有限の母集団についての期待値を E と書くことにする。

$$E\delta\bar{x}' = \frac{1}{N^2}\sum_{i,j=1}^{N} E\delta x_i x_j$$

であるからもし $E\delta C < 0$ ならこの計画は"負の期待相関をもつ"という。

さて我々は標本計画中に負の相関や負の期待相関を導入したり，それを利用したりできる場合の標本計画を論ずることから始めよう。

簡単のためにまず層が二つの場合を考え，x' のとりうる値を $x_1, \cdots\cdots x_n$ また y' のとりうる値を $y_1, \cdots\cdots y_m$ とする。さらに

$$P\{x'=x_i\} = P\{y'=y_i\} = P\{x'=x_i,\ y'=y_i\} = p > 0$$

なるように抽出を行なうものとする。したがって $\sum_{i=1}^{n} p_i = 1$ でもし $i \neq j$ なら $P\{x'=x_i,\ y'=y_j\} = 0$ である。上記の仮定から

$$\delta x'y' = \sum_{i=1}^{n} p_i x_i y_i - \sum_{i,j=1}^{n} p_i p_j x_i x_j \tag{3.2}$$

が導びかれる。

記号 $\varphi_{ij} \gtrless 0$ はすべての i, j について $\varphi_{ij} \geqq 0$ であり少くとも一対の i, j については $\varphi_{ij} > 0$ なることを表わす。我々は $(x_i - x_j)(y_i - y_j) \geqq 0$ なら組 (x) と (y) は同様に順序づけられており，$(x_i - x_j)(y_i - y_j) \leqq 0$ ならこれらの組は順序が逆であるという。ただし (x) は $x_1, \cdots\cdots x_n$ を表わし (y) は $y_1, \cdots\cdots y_m$ を表わす。このとき数値が逆の順序のときは $\delta x'y' < 0$，また同じ順序なら $\delta x'y' > 0$ ということ

と $[x, P, y]$ は直接容易に証明することができる。

これより幾分一般な結果は次の通りである

定　理 3

$n \leqq N$ とし

$$b = \sum_{i=1}^{n}\sum_{j=1}^{N} a_{ij} w_i z_j$$

を実双一次形式，また

$$t = \sum_{i=1}^{n} a_{ii} w_i$$

実一次形式とする。ただし $w_i > 0,\ z_i > 0$ かつ

$$\sum_{i=1}^{n} w_i = \sum_{i=1}^{N} z_i = 1$$

このとき $b > t$ なるための十分条件は

$$a_{ij} \gtrsim a_{ii} \tag{3.3}$$

$n = N$ で $w_i = z_i$ なら $b > t$ となるのは

$$a_{ij} + a_{ji} \gtrsim a_{ii} + a_{jj} \tag{3.4}$$

のときである。

証　明

$$b - t = \sum_{i=1}^{n} a_{ii}(w_i z_i - w_i) + \sum_{i \neq j} a_{ij} w_i z_j$$

で，また

$$1 - z_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} z_j$$

だから

$$b - t = \sum_{i \neq j} (a_{ij} - a_{ii}) w_i z_j$$

が与えられる。よって (3.3) が成りたてば $b > t$，同様にして $n = N$ で $w_i = z_i$ なら (3.4) が成りたてば $b > t$ となる。定理3の自明な拡張は，ここでは必要がないので省略する。

組 (x) と (y) が逆の順序のとき $\delta x'y' < 0$ なることを証明するため $a_{ij} = x_i y_j$ および $z_i = w_i = p_i$ とおきえれば

$$a_{ii} + a_{jj} - a_{ij} - a_{ji} = (x_i - x_j)(y_i - y_j) \tag{3.5}$$

をうる。したがってこの値が逆に順序づけられていれば $\sigma_{x'y'}<0$ であるから，二つの層は負の相関をもつ。

負の期待相関を考えるため

$$(3.6)\quad \mathcal{E}\sigma_{x'y'}=\sum_{i=1}^{n}p_i\sigma_{ii}+\sum_{i\neq j=1}^{n}p_ip_j\rho_{ij}$$

なることに注意する。ここで $\mathcal{E}x_i=\mu,\ \mathcal{E}y_i=\nu$ および $\mathcal{E}(x_i-\mu)(y_j-\nu)=\sigma_{ij}$ と仮定する。したがって σ_{ii} は分散でなく共分散である。

$a_{ij}=\sigma_{ij}$，$z_i=w_i=p_i$ とおくと（3.4）が成立して，$\mathcal{E}\sigma_{x'y'}$ が負となるための十分条件，

$$(3.7)\quad \sigma_{ij}+\sigma_{ji}\geqslant\sigma_{ii}+\sigma_{jj}$$

が与えられる。あるいはまた ρ_{ij} を

$$\sigma_x\sigma_y\rho_{ij}=\sigma_{ij}$$

によって定義すると，$\mathcal{E}\sigma_{x'y'}<0$ なるための十分条件が

$$(3.8)\quad \rho_{ij}+\rho_{ji}\geqslant\rho_{ii}+\rho_{jj}$$

のように与えられる。ただし $\sigma_x^2=\mathcal{E}(x_i-\mu)^2$，$\sigma_y^2=\mathcal{E}(y_i-\nu)^2$ とする。

単一要素の系統的抽出を考えよう。系統的抽出では kn 個の順序づけられた要素 $x_1, x_2, \cdots x_k, x_{1+k}, \cdots x_{2k} \cdots x_{1+(n-1)k}, \cdots x_{nk}$ からなる母集団を仮定し，その算術平均値 $\bar{x}$ を推定したいものとする。推定値としては

$$\bar{x}'=(x_1'+\cdots\cdots+x_n')/n$$

を用いる。ここで x_i' は $x_1, \cdots\cdots x_k$ から無作為に選び，もし $x_j'=x_{j+(i+1)k}$，$i=1, \cdots\cdots n$ とする。よって $\bar{x}'$ は次の形が

$$x_i+(i-1)k, \qquad x_k+(i-1)k$$

からなっていて，

$$P\{x_i'=x_{j+(i-1)k}\}=P\{x_i'=x_{j+(i-1)k}, x_j'=x_{\alpha+(j-1)k}\}=\frac{1}{k}$$

かつ $j\neq\beta$ のとき

$$P\{x_i'=x_{j+(i-1)k}, x_j'=x_{\beta+(j-1)k}\}=0$$

なるような層化母集団よりの推定値と解釈できる。このとき，

$$\sigma_{\bar{x}_i\bar{x}_j}=\left(\frac{1}{k}\right)\sum_{\alpha=1}^{k}x_{j+(i-1)k}\cdot x_{j+(j-1)k}-\bar{x}_i\bar{x}_j$$

ただし $\bar{x}_i=\left(\frac{1}{k}\right)\sum_{\alpha=1}^{k}x_{j+(i-1)k}$

ゆえに逆に順序づけられている二つの層は分散に対して負の寄与を与えることになる。しかしすべての層を負に順序づけることはできないからこれでは有用な結果はえられない。それで（1）でのように C または $\sigma_{\bar{x}}^2$ そのものに立ちかえって考えなければならない。しかしもし Cochran の仮定を認め，$\mathcal{E}\sigma_{x'y'}$ を考えるとし層 j 層について

$$\rho_{\alpha\beta}=\rho_{(j-i)k+\beta-\alpha}$$

なる関係が導かれ（3.8）は

$$(3.9)\quad \rho_{(j-i)k+(\beta-\alpha)}+\rho_{(j-i)k+(\alpha-\beta)}\geqslant 2\rho_{(j-i)k}$$

となる。すなわち相関函数 ρ_α は上方に凹でなければならない。これは Cochran が別の方法で証明したところのものである。$\mathcal{E}C$ を考えることによって，ある種の平均的凹型関係をもつということだけが系統的抽出の分散を層化無作為抽出以下とするのにコレログラムに要求される条件だということが示しうる。

4　層の大きさが等しくない場合に相関が負となるための条件と系統的抽出に対するその応用

大きさに比例した確率による集落の系統的抽出（sec 5 で論ずる）のように，定理3で取扱って簡単な情況は直接あてはまらないことが多い。しかし定理3は次のようにして利用することができる。

x のとりうる値を $x_1, \cdots\cdots x_n$ y のとりうる値を $y_1, \cdots\cdots y_k$ とし

$$P\{y=y_\beta\mid x=x_\alpha\}=p_{\beta|\alpha}$$

とする。ただし $k\geqslant n$。このときもし

$$(4.1)\quad \eta_\alpha=\sum_{\beta=1}^{m}y_\beta p_{\beta|\alpha}$$

と定義すれば

$$\eta_\alpha=E(y\mid x=x_\alpha)$$

である。

y'をその取りうる値が$y_1, \cdots y_m$で，それからの確率が$p_1 \quad p_m$である確率変数だとすると

$$y' = E^*(y_0' \mid x')$$

で

$$\sigma_{x'y_0} = \sigma_{x'y'}$$

ただし $p_\alpha = P\{x' = x_\alpha\}$

確率変数だとy'についても定理3の成立つことは明らかである。

結局我々には組$x_1, \cdots x_n$および$y_1 \quad y_m$が並に順序づけられる，あるいは(3.7)が成立するために満足しなければならない条件付確率$p_{\beta|\alpha}$，およびy^0の値に対する制限が何であるかを定めることだけが必要である

(3.5)のy_iおよびy_iを代入すると，もし

(4.2) $\quad (x_\beta - x_\gamma)\sum_{\delta=1}^{m} y_\delta^0 (p_{\beta|\alpha} - p_{\beta|\alpha}) \leqq 0$

なら $\quad \sigma_{x'y'} - \sigma_{x'y_0} < 0$ である。

$$\sigma^0_{\alpha\delta} = \xi(x_\alpha - M)(y_\delta^0 - r)$$

とする。よって(3.7)に代入すると

(4.3) $\quad \sum_{\beta=1}^{n} (p_{\beta|\alpha} - p_{\beta|\gamma})(\sigma^0_{\gamma\beta} - \sigma_{\gamma\beta}) \leqq 0$

あるいは

(4.4) $\quad \sum_{\beta=1}^{n} (p_{\beta|\alpha} - p_{\beta|\gamma})(\rho^0_{\gamma\beta} - \rho^0_{\delta\beta}) \leqq 0$

なら $\quad \xi\sigma_{x'y_0} < 0$

である。(4.2)および(4.3)を用いるには次のよく知られたLemmaが屡々有用である。

Lemma $\quad \xi_1 \leqq \xi_2 \leqq \cdots\cdots \leqq \xi_n \leqq 0$で並$E_1, \cdots E_n$について

$$\sum_{\beta=1}^{\delta} E_\beta \xi_\beta \leqq 0 \text{ が成立てば } \sum_{\beta=1}^{\delta} \xi_\beta E_\beta \leqq 0, \quad \delta = 1, \cdots, n$$

層間の頁の相関または頁の期待相関を示すのに有用な他の定理を導びくためにこのLemmaを使うことにする。

定理4．一次形式

$$\alpha = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} w_i z_j$$

について

$$\sum_{i=1}^{\delta} w_i \geqslant 0 \quad , \quad \sum_{j=1}^{\delta'} z_j \geqslant 0 \qquad \delta = 1, \cdots n-1 \quad \delta' = 1, \cdots m-1$$

および

(4.5) $\quad \sum_{i=1}^{n} w_i = \sum_{j=1}^{m} z_j = 0$

とする。

$$\delta_{ij} = a_{ij} - a_{i+1,j} - a_{i,j+1} + a_{i+1,j+1}$$

とおく。そうすると$\alpha \leqq 0$なるための十分条件は$\delta_{ij} \leqq 0$である。

証明 $\quad \alpha$のw_nとz_mに(4.5)を代入すると

$$\alpha = \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} \delta'_{ij} w_i z_j$$

なることがわかる。ただし

$$\delta'_{ij} = a_{ij} - a_{im} - a_{nj} + a_{nm}$$

あるいは

$$\xi_j = \sum_{i=1}^{i-1} \delta_{ij} w_i$$

と定義すれば

$$\alpha = \sum_{j=1}^{m-1} \xi_j z_j$$

よってLemmaによって$\alpha \leqq 0$なるための十分条件は $\quad \xi_1 \leqq \xi_2 \leqq \cdots\cdots \leqq \xi_{m-1} \leqq 1$

である。同様に

$$\xi_j - \xi_{j+1} \leqq 0$$

なるための十分条件は

$$\delta'_{ij} - \delta'_{i,j+1} \leqq \delta'_{i+1,j} - \delta'_{i+1,j+1}$$

である。よって証明で残されたものは

$$\delta_{ij} = \delta'_{ij} - \delta'_{i,j+1} - \delta'_{i+1,j} + \delta'_{i+1,j+1}$$

を確めることだけである。

我々は前の頁で層内の抽出が独立でなく，一つの層からの要素の

抽出が他層の抽出を決定してしまうような層化抽出は系統的抽出と同じものだということを証明した。しかしこの場合層内の要素数は等しいことを仮定している。我々はここで層内の要素数が異なる場合についてこの標本抽出値を拡張してみよう。そうしてその過程の中で Lemma および定理4の用法を例示することにしよう。

さてN個の要素 $x_1 \cdots\cdots x_N$ からなる母集団が n 個の層に分類されており、その i 番目の層には N_i 個の要素

$$x_{N_1+\cdots+N_{i-1}+1} \cdots\cdots x_{N_1+\cdots+N_i}$$

が含まれるものとする。これらの要素を $x_{i1} \cdots x_{iN_i}$ とする。

これらの n 個の層からそれぞれ要素を一つづつ選ぶ。i 番目の層から選んだ要素を x_i' と書く。母集団の算術平均値 $\bar{x}$ の推定値として

$$\bar{x}' = \sum_{i=1}^{n} \frac{N_i}{N} x_i'$$

をとる。このときよく知られているように抽出が各層から独立かつ無作為に行なわれるとすると

$$\sigma^2_{\bar{x}'} = \sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2 \sigma_i^2$$

である。ただし σ_i^2 は x_i の分散すなわち i 番目の層の分散である。

ここで普通に用いられるものと異なるもう一つの方法を考えてみよう。我々は一般性を失なうことなく $N_i > 1$ と仮定できる。（この方法は $N_i = 1$ なるいづれの層についても同じで、またすべての $N_i = 1$ であるかまたは唯一つの N_i を除き他のすべてが1に等しくなっているような母集団についても同じ結果がえられる。しかし少くとも二つの N_i が1と異なれば結果は違ってくる。）

我々はまず最初の層から無作為に一要素を選ぶ。$x_1' = x_\alpha$ とおく。$N_2 > 1$ と仮定して第2の層から次のようにして一要素を選ぶ。k_2 が整数のとき $N_2 t_2 / N_1 = k_2$ となるような正の整数 t_2 をとり N_2 に乗ずる。第2層の各要素にサイズ t_2 の大きさ（measure）を与え、二組の累積和 $t_2, 2t_2, \cdots N_2 t_2$ および $k_2, 2k_2, \cdots N_1 k_2$ を作る。そうすると層2の各要素に与えられたサイズ t_2 の大きさと層1の各要素に与えられたサイズ k_2 の大きさで層1と2は同じサイズをもつことになる。

以下に与える計算の一例として次のような簡単な場合を考える。$N_1 = 3$ で $N_2 = 4$ と仮定する。このとき t_2 を6にとれば $k_2 = 8$ となる。我々は整数1,2,3の中の一つを等確率で選ぶ。もし整数1、与えられたならば1層の1要素が選ばれたのであるから1から8までの整数の中一数を等確率で抽出する。選ばれた整数が1と6の間の数であれば2層の1要素が選ばれる。同様にもし1層の2要素が選ばれたなら9と16の間の整数から1数を等確率で選ぶ。この数が9……12ならば2層の2要素をとり、13……16ならば3の要素をとる。

2層に対する抽出手続の一般的方式は次の通りである。

β_0 を $(\alpha-1)k_2 + 1 \leq \beta_0 t_2$ なる最小の整数とし、また β_1 は $(\beta_1 - 1)t_2 < \alpha k_2 \leq \beta_1 t_2$ なる整数とする。1,……6の中から無作為に一つの整数を選んで β とする。

このときもし

$$(\alpha-1)k_2 < (\alpha-1)k_2 + \beta \leq \beta_0 t_2$$

ならば2層から β_0 番目の要素を選び

$$\beta_0 t_2 < (\alpha-1)k_2 + \beta \leq (\beta_0+1)t_2$$

ならば (β_0+1) 番目の要素を選ぶ。　そうしてもし

$$(\beta_1 - 1)t_2 < (\alpha-1)k_2 + \beta \leq \alpha k_2$$

ならば2層から β_1 番目の要素を選出する。

このようにすると容易に証明できるように層2の各要素は等確率で抽出されることになる。かえにこの手続を各層に用いれば

$$\sigma^2_{\bar{x}'} = \sum_{i=1}^{n} \left(\frac{N_i}{N}\right)^2 \sigma_i^2 + \sum_{i \neq j} \frac{N_i N_j}{N^2} \sigma_{x_i x_j}$$

がえられる。

このタイプの抽出での $\sigma_{x_i x_j}$ を評価してみよう。いま

$$\sigma_{x_i' x_j'} = E(x_i' - \bar{x}_i)(x_j' - \bar{x}_j)$$

である。ここで $\bar{x}_i$ は i 層の要素の算術平均値である。よって定理

ゆえ

$$\sigma_{x_i x_j} = \frac{1}{N_1}\sum_{\alpha=1}^{N_1} E\{x_i - E(x_i|x_{i\alpha})\}\{x_j - E(x_j|x_{i\alpha})\} + \frac{1}{N_1}\sum_{\alpha=1}^{N_1}\{E(x_i|x_{i\alpha}) - \bar{x}_i\}\{E(x_j|x_{i\alpha}) - \bar{x}_j\}$$

$\sigma_{x_i x_j}$ の第1項は上記の抽出法では0になることが容易にわかる。さらに $\bar{x}_i$ は条件付期待値の算術平均値だから問題はこの条件付期待値が，真の相関または真の期待相関をもつための条件を満たすかどうかを定めることに帰着する。

$E(x_i|x_{i\alpha})$ を $y_{i\alpha}$ で表わす。そうすると組 $y_{i1}\cdots y_{iN}$ と $y_{j1}\cdots y_{jN}$ が逆順かどうかを調べなければならない。いま

$$(y_{i\alpha} - y_{i\beta})(y_{j\alpha} - y_{j\beta}) = \sum_{g=1}^{N_i}\sum_{h=1}^{N_j} x_{ig}x_{jh}E_{ig\alpha\beta}E_{jh\alpha\beta}$$

である。ただし

$$E_{ig\alpha\beta} = P\{x_i = x_{ig}|x_{i\alpha}\} - P\{x_i = x_{ig}|x_{i\beta}\}$$

である。 $\alpha < \beta$ ならこの抽出法にしたがって

$$\sum_{g=1}^{\Delta} E_{ig\alpha\beta} \geq 0 \qquad \Delta = 1 \cdots\cdots N-1$$

で，また一方

$$\sum_{g=1}^{N_i} E_{ig\alpha\beta} = 0$$

である。 定理4で

$$n = N_i,\ m = N_j,\ w_g = E_{ig\alpha\beta},\ Z_h = E_{jh\alpha\beta},\ a_{gh} = x_{ig}x_{jh}$$

とおく。そうすると

$$\delta_{gh} = (x_{ig} - x_{i,g+1})(x_{jh} - x_{j,h+1})$$

であるから，層間に真の相関が存在するためには組 $x_{i1}\cdots\cdots x_{iN_i}$ と $x_{j1}\cdots\cdots x_{jN_j}$ が $\delta_{gh} \leq 0$ で表わされるような型の真の順序のあれば十分である。同様に

$$\sigma_{gh} = \mathcal{E}(x_{ig} - \mu_i)(x_{jh} - \mu_j),\quad \mu_i = \mathcal{E}x_{ig}$$

なら，真の期待相関が生ずるための十分条件は

$$\sigma_{gh} - \sigma_{g,h+1} - \sigma_{g+1,h} + \sigma_{g+1,h+1} \leqq 0$$

である。

もちろんこれらの関係はコレログラムが上に凹なら満足される。したがってもし母集団が上に凹なコレログラムをもつN個の確率変数 $x_1,\cdots\cdots x_N$ からなっているとき，これらの要素がどの層に分類されていても要素の出現の順序が変らない限り最適配分による標本要素の層化無作為抽出より分散の小さい推定値を与えるような標本要素の系統的抽出を計画することができる。もし最適配分のもとで一個以上の要素を層から選出するなら，目数の要素を系統的に抽出すれば十分である。もし最適配分だけでなく最適層化（optimum definitions of strata）をも用いて，各層から一要素だけを選出するものとすれば，この節で述べた方式にしたがう系統的抽出の分散は層化無作為抽出のものよりも大きくはならない。しかし層を考えないで系統的抽出に夢中になってもは小さくならないことに注意しなければならない。分散を減少させるためにはあらかじめしておかねばならない手続きが存在するのである。

この例の方法によって母集団の大きさが標本サイズの倍数でない場合の要素の系統的抽出について答えられるであろう。

5. 大きさに比例した確率を用いる集落の系統的抽出

集落抽出で確率をその大きさに比例させた場合には推定値の分散が屡々大きく減少することが知られている〔5〕しかし確率を大きさに比例させて行なう幾つかの集落の抽出理論はこれまで研究されていない。この節の目的はこの理論に対して何らかの寄与を与えることである。

大きさに比例した確率で集落を抽出するのに最も多く用いられる方法は次に述べるものと同等である。集落を $C_1,\cdots\cdots C_M$ で表わし，これらの円周の集落の n 番目のものに P_n なる確率を与える。次々

の累積和 $P_1, P_1+P_2, P_1+P_2+P_3, \cdots\cdots P_1+\cdots\cdots+P_M$ を作る。これらの中から m 個の集落を選びたいものとすると，$\bar{P}_m=(P_1+\cdots+P_M)/m$ を計算し，次に $P_j \leqq \bar{P}_m$，$j=1,\cdots\cdots M$ と仮定して $1,\cdots\cdots \bar{P}_m$ の中から等確率で1つの整数を選ぶ。この整数を P' とし，m 個の数 P', $P'+\bar{P}_m$, $P'+2\bar{P}_m$, $P'+(m-1)\bar{P}_m$ を計算する。もし任意の整数 $i, i=1,\cdots\cdots m$ に対し

(5.1) $\quad P_1+\cdots\cdots+P_{k_1-1}+1 \leqq P'+(i-1)\bar{P}_m \leqq P_1+\cdots\cdots+P_{k_1}$

ならば集落 C_{k_1} を標本に選ぶ。$P_k > \bar{P}_m$ なる集落はすべて自動的に標本内に含まれ，更にもしそのような集落がたとえば s 個あったとすると，これらの s 個の集落が標本に含まれた後に残る $M-s$ 個の集落に対して $\bar{P}_{m-s}$ を計算し前と同様な手続を繰返えす。

我々の用いる推定値の分散を算出する際，我々はこの推定値を層化抽出の推定値と考えることにする。推定値の分散はこのような解釈なしても容易に求められるが，あとで分散の推定で必要となるのでここで完全な式を少し綴めて示しておく。

集落 $C_1,\cdots\cdots C_{k_1-1}$ は

$$P_1+\cdots+P_{k_1-1} < \bar{P}_m+\cdots\cdots+\bar{P}_{k_1}$$

を満たすものとする。このとき層1は集落 $C_1,\cdots\cdots C_{k_1}$ からなるものとする。容易にわかる通り上の抽出方法をとれば

$$P\{C_{k_1} \text{が層1から選ばれる} \quad \ell < k_1\} = \frac{P_\ell}{\bar{P}_m}$$

$$P\{C_{k_1} \text{が層1から選ばれる}\} = \frac{\bar{P}_m - P_1 - \cdots - P_{k_1-1}}{\bar{P}_m}$$

である。

さらに C_{k_1}, $C_{k_1+k_2}$ について

$$P_1+\cdots+P_{k_1+k_2-1} < 2\bar{P}_m \leqq P_1+\cdots\cdots+P_{k_1+k_2}$$

が成立つとする。このとき層2は集落 $C_{k_1},\cdots\cdots C_{k_1+k_2}$ からなると定める。容易にわかるように上記の抽出方法を用いれば

$$P\{C_{k_1} \text{が層2から選ばれる}\} = \frac{P_1+\quad+P_{k_1}-\bar{P}_m}{\bar{P}_m}$$

$$P\{C_{k_1+\ell} \text{が層2から選ばれる}, 1 \leqq \ell < k_2\} = \frac{P_\ell}{\bar{P}_m}$$

$$P\{C_{k_1+k_2} \text{が層2から選ばれる}\} = \frac{2\bar{P}_m - P_1 - \quad P_{k_1+k_2-1}}{\bar{P}_m}$$

である。$P_\ell \leqq \bar{P}_m$ だから C_{k_1} が層1と層2の両方から選ばれることは不可能である

一般に集落 $C_{k_1+\cdots+k_{i-1}}\cdots C_{k_1+\cdots+k_i}$ について

(5.2) $\quad P_1+\cdots+P_{k_1+\cdots+k_{i-1}} < i\bar{P}_m \leqq P_1+\cdots+P_{k_1+\cdots+k_i}$

なり，第 i 層はこれらの k_i+1 個の集落からなるから，我々は確率 $P_{i\alpha}$, $\alpha=0,\cdots\cdots k_i$ を次のように定義する。

$$P_{i0} = P\{C_{k_1+\cdots+k_i-1} \text{が}i\text{層から選ばれる}\} = \frac{P_1+\cdots+P_{k_1+\cdots+k_i-1}-(i-1)\bar{P}_m}{\bar{P}_m}$$

$$P_{i\alpha} = P\{C_\ell \text{が第}i\text{層から選ばれる}, k_1+\cdots+k_{i-1} < \ell < k_1+\cdots+k_i\}$$

(5.3) $\quad = \frac{P_\ell}{\bar{P}_m} \qquad \alpha = \ell - k_1 - \cdots\cdots - k_{i-1}$

$$P_{ik_i} = P\{C_{k_1+\cdots+k_i} \text{が}i\text{層から選ばれる}\} = \frac{i\bar{P}_m - P_1 - \cdots\cdots - P_{k_1+\cdots+k_i}}{\bar{P}_m}$$

次のは

$$P_{i-1,k_{i-1}} + P_{i0} = \frac{P_{k_1+\cdots+k_{i-1}}}{\bar{P}_m} \tag{5.4}$$

に注意しておく。

いま母集団の要素を x_{kj}, $k=1,\cdots M$, $j=1\cdots N_k$ とし, k 番目の集落の算術平均値を $\bar{x}_k$ とおく。N_k は通例未知だが大きさの尺度 P_k はわかっているから, 抽出確率を N_k に比例させず P_k に比例した確率で抽出を行なう。i 層の集落 $C_{i0}\cdots C_{ik_i}$ と書き

$$C_{i\alpha} = C_{\alpha+k_1+\cdots+k_{i-1}} \tag{5.5}$$

とおく

さらに集落の要素数を $N_{i0}\cdots N_{ik_i}$ また集落平均値を $\bar{x}_{i0}\cdots\bar{x}_{ik_i}$ とする。

ここで

$$N_{i\alpha} = N_{\alpha+k_1+\cdots+k_{i-1}}$$

$$\bar{x}_{i\alpha} = \bar{x}_{\alpha+k_1+\cdots+k_{i-1}} \tag{5.6}$$

だから

$$\bar{x}_{i0} = \bar{x}_{i-1,k_{i-1}}$$

また

$$N_{i0} = N_{i-1,k_{i-1}} \qquad i=1\cdots m$$

である。更に

$$\bar{x}_{i\alpha} = N_{i\alpha}\,\bar{x}_{i\alpha}/P_{i\alpha} = \bar{x}_{\alpha+k+\cdots+k_{i-1}} \tag{5.7}$$

と定義する。

我々は i 番目層の平均値を

$$\bar{\bar{x}}_i = \sum_{\alpha=0}^{k_i} P_{i\alpha}\,\bar{x}_{i\alpha}/\bar{P}_m \tag{5.8}$$

と定義し i 層の分散を

$$\sigma_i^2 = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P_m}(\bar{x}_{i\alpha} - \bar{\bar{x}}_i)^2 \tag{5.9}$$

とする。

このとき母集団の平均値と分散を

$$\bar{x} = \sum_{k=1}^{M} P_k\,\bar{x}_k / P \tag{5.10}$$

および

$$\sigma^2 = \sum_{k=1}^{M} \frac{P_k}{P}(\bar{x}_k - \bar{x})^2 \tag{5.11}$$

と定義すると容易に

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} \bar{\bar{x}}_i \tag{5.12}$$

および

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}\sigma_i^2 + \frac{1}{m}\sum_{i=1}^{m}(\bar{\bar{x}}_i - \bar{x})^2 \tag{5.13}$$

が証明できる。

特性統計の不偏推定値

我々は x の推定値を求めうることを証明しよう。ただし

$$x = \sum_{i=1}^{M}\sum_{j=1}^{N_i} x_{ij}$$

すなわち x は母集団要素の総計である。N は未知だから x の推定値として用いられるのは $\bar{x}$ と N の不偏推定値の比である。よく知られているように, この比は偏りをもつ。我々はここでは比推定値の研究を行なうのでないから, この推定値の分散の近似式を導びくことはしない。しかしそれはここで与えた結果の単純な拡張で与えられることは注意できる。

我々はこの推定値の一般式の次のようになることをみてみよう。もし母集団の i 番目集落が選ばれれば, それからの n_i 個の要素の特性値統計を x_i で表わす。さらに i 番目層からとられたあるいは同じ

ことだが，i 層からとった集落より抽出された副次標本内の要素の統計を x_i とし，これらの要素の統計を x_i'' とかく。このようにしてj番目の集落がi番目に選ばれたなら $n_i = n_j$ で $x_i = x_j$ である．我々は母集団統計 χ の推定値 χ'' を

(5.14) $\qquad \chi'' = K(x_1'' + \cdots + x_m'')$

と定義する。

このときもし $K = P/mn$ で $n_i = nN_i/P_i$ ならば χ'' が χ の不偏推定値であることは容易にわかる。

推定値の分散

χ'' の分散が計算できる。ここで

(5.15) $\qquad \chi'' = \bar{P}_m (\tilde{x}_1'' + \cdots + \tilde{x}_m'')$ および $\tilde{x}_i'' = x_i''/n$

いま定理2によって

(5.16) $\qquad \sigma^2_{\chi''} = E\sigma^{2*}_{\chi''|A} + \sigma^2_{E^*(\chi''|A)}$

である。ここで A^* は前に定義したものがある。我々は $E\sigma^{2*}_{\chi''|A}$ の評価は行なわない。というのはこれには無作為または系統的抽出を用いる副次抽出法あるいは大きさに比例した確率を用いる．副次抽出法に関する新らしい問題は何も含まれていないからである。

(5.15) から

(5.17) $\qquad E^*(\chi''|A) = \bar{P}_m(\tilde{x}_1' + \cdots + \tilde{x}_m')$

が与えられる．いいかえれば $E^*(\chi''|A)$ は標本間の集落を全部調べた場合に与えられる推定値である．(5.16) の右辺項を σ_B^2 と書けば

(5.18) $\qquad \sigma_B^2 = \bar{P}_m^2 \left\{ \sum_{i=1}^{m} \sigma^2_{\tilde{x}_i'} + \sum_{i \neq j} \sigma_{\tilde{x}_i' \tilde{x}_j'} \right\}$

である．

いま，

(5.19) $\qquad \sigma^2_{\tilde{x}_i'} = \sum_{\alpha=0}^{n_i} \frac{P_{i\alpha}}{\bar{P}_m} (\tilde{x}_{i\alpha} - \bar{\tilde{x}}_i)^2 = \sigma_i^2$

$i \neq j$ のときの $\sigma_{\tilde{x}_i' \tilde{x}_j'}$ を計算するため定理1を用いる

(5.20) $\qquad \sigma_{\tilde{x}_i' \tilde{x}_j'} = E(\tilde{x}_i' - \bar{\tilde{x}}_i)(\tilde{x}_j' - \bar{\tilde{x}}_j) = E\{(\tilde{x}_i' - \bar{\tilde{x}}_i) E^*[(\tilde{x}_j' - \bar{\tilde{x}}_j) | \tilde{x}_i']\}$

$E^*[(\tilde{x}_j' - \bar{\tilde{x}}_j)|\tilde{x}_i']$ を計算するためまず

(5.21) $\qquad E^*[(\tilde{x}_j' - \bar{\tilde{x}}_j)|\tilde{x}_i'] = E^*[(\tilde{x}_j' - \bar{\tilde{x}}_j)|C_i]$

に注意する．ここで C_i^* は i 番目の標本の集落としては $C_{i0}, \cdots C_{in_i}$ の何れかが選ばれるという n_i+1 通りの可能な場合のある確率変数である．いまもし $C_{i\alpha}$ が i 層の集落なら

(5.22) $\qquad E[(\tilde{x}_j' - \bar{\tilde{x}}_j)|C_{i\alpha}]$

を計算する．我々はまず i 層から $C_{i\alpha}$ が選ばれることがわかっているとき j 層で標本の集落となり得るのはどの集落であるかをきめることにする．しかも j 層の大きさはどちらも $\bar{P}_m$ だから

$$P_{j0} + \cdots + P_{j\beta_0 - 1} \leq P_{i1} + \cdots + P_{i\alpha-1} < P_{j1} + \cdots + P_{j\beta_0}$$

と

$$P_{j0} + \cdots + P_{j\beta_1 - 1} < P_{i1} + \cdots + P_{i\alpha} \leq P_{j1} + \cdots + P_{j\beta_1}$$

となるような整数 β_0 と β_1 が存在する．ゆえにもし i 層から $C_{i\alpha}$

が選ばれていることがわかっていれば、お層からは $C_{j\beta_0}$, $C_{j\beta_0+1}$, …… $C_{j\beta_1}$ のどれか1つを選ばなければならずまた

$$P\{C_{j\beta} \text{が選ばれる} \mid C_{i\alpha} \text{が選ばれる}\} = P'_{j\beta}/P_{i\alpha}, \quad \beta=\beta_0, \beta_0+1, \cdots, \beta_1 \text{のとき} \quad =0 \quad \text{他の場合}$$

なることがわかる。ただし、

$$P'_{j\beta_0} = P_{j1} + \cdots + P_{j\beta_0} - P_{i1} - \cdots, P_{i,\alpha-1}$$

$$P'_{j\beta} = P_{j\beta}, \quad \beta=\beta_0+1, \cdots, \beta_1-1$$

$$P'_{j\beta_1} = P_{i1} + \cdots + P_{i\alpha} - P_{j1} - \cdots - P_{j\beta_1-1}$$

で

$$\sum_{\beta=\beta_0}^{\beta_1} P'_{j\beta} = P_{i\alpha}$$

である。

このとき

(5.23) $$E\{(\bar{\tilde{x}}'_j - \bar{\tilde{x}}_j \mid C_{i\alpha}) = \bar{\tilde{x}}_{j|\alpha} - \bar{\tilde{x}}_j\}$$

ただし

(5.24) $$\bar{\tilde{x}}_{j|\alpha} = \sum_{\beta=\beta_0}^{\beta_1} \frac{P'_{j\beta}}{P_{i\alpha}} \tilde{x}_{j\beta}$$

よって (5.20) に代入すると

$$\sigma_{\bar{\tilde{x}}'_j \bar{\tilde{x}}_i} = E(\bar{\tilde{x}}'_j - \bar{\tilde{x}}_i)(\bar{\tilde{x}}_{j|i} - \bar{\tilde{x}}_j)$$

たゞし i 層から $C_{i\alpha}$ が選ばれゝば $\bar{\tilde{x}}_{j|i} = \bar{\tilde{x}}_{j|\alpha}$ である。それゆえ

(5.25) $$\sigma_{\bar{\tilde{x}}_i \bar{\tilde{x}}_j} = \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P_m}(\bar{\tilde{x}}_{i\alpha} - \bar{\tilde{x}}_0)(\bar{\tilde{x}}_{j|\alpha} - \bar{\tilde{x}}_j)$$

が えられる。(5.23) を用いれば (5.25) の条件付期待値は消去できることは容易にわかるが しかしそうしても簡略化とか一般

化とかの点では何等うるところがない。

特別な場合には可能な標本を全部書きあらわして $\bar{x}'_i$ の分散共分散を求めることができよう。これを一般化するためには必要な記法を与えさえすればよい。

(5.18) に代入すると

$$\sigma_B^2 = \bar{P}_m^2 \left\{ \sum_{i=1}^{m} \sigma_i^2 + \sum_{i \neq j} \sigma_{\bar{\tilde{x}}_j \bar{\tilde{x}}_j} \right\}$$

がわかる。たゞし $\sigma_{\bar{\tilde{x}}_j \bar{\tilde{x}}_j}$ は (5.25) で与えられたものである

$$\sum_{i=1}^{m} (\bar{\tilde{x}}_j - \bar{x}) = 0$$

なる事実を用いれば

$$\sigma_B^2 = \bar{P}_m^2 \sum_{i=1}^{m} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{\bar{P}_m}(\bar{\tilde{x}}_{i\alpha} - \bar{x})^2 + \bar{P}_m^2 \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P'_{i\alpha}}{P_m}(\bar{\tilde{x}}_{i\alpha} - \bar{\tilde{x}})(\bar{\tilde{x}}_{j|\alpha} - \bar{\tilde{x}})$$

あるいは一部を "層化する前" の記法にもどして

(5.26) $$\sigma_B^2 = \frac{P^2}{m} \sum_{h=1}^{M} \frac{P_h}{P}(\bar{\tilde{x}}_h - \bar{\tilde{x}})^2 + \frac{P^2}{m} \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P}(\bar{\tilde{x}}_{i\alpha} - \bar{\tilde{x}})(\bar{\tilde{x}}_{j|\alpha} - \bar{\tilde{x}})$$

がえられる。

(4.24) の2番目の部分の項を組合わせると (1) でえた公式の結果が容易にえられる。

σ_B^2 を少し違った形で表わせば

(5.27) $$\sigma_B^2 = \frac{P^2}{m}\left\{\sigma^2 - \sigma_{b.s}^2 + \sum_{i \neq j} \sum_{\alpha=0}^{k_i} \frac{P_{i\alpha}}{P}(\bar{\tilde{x}}_{i\alpha} - \bar{\tilde{x}})(\bar{\tilde{x}}_{j|\alpha} - \bar{\tilde{x}})\right\}$$

となる。ここで

$$\sigma_{b.s}^2 = \frac{1}{m} \sum_{i=1}^{m} (\bar{\tilde{x}} - \bar{x})^2$$

これは選ばれた集落を復元し (with replacing)、大きさに比例した確率を用いて抽出する方法と比較したとき 効率と差が出てくる二つの原因を示すものである。(勿論 $P^2\sigma^2/m$ は その前に選ばれたものを必ずもとに戻すようにして m 個の集

差を、大きさに比例した確率で抽出したときの $E^{*}(\bar{x}''|A)$ の分散であることは容易にわかる。

(5.26) と (5.27) をみると、単一要素の抽出の場合と殆んど同じ条件のもとで、大きさに比例した確率で行う系統的抽出は復元 P.P.S 抽出 (sampling with probability proportionate to size) よりも有効になることがわかる。詳細は省略する。これらは Lemma と定理4を適用すればえられる。この条件はまとめると次のようになる。P.P.ρ で系統的に標本を抽出して、二つの組

$$x_{i1}, \cdots x_{ik_i} \quad \text{および} \quad y_{j1} \cdots y_{jk_j}$$

が単調で、一方が単調非増加、一方が単調非減少になっているとすると、i層とj層間の共変動は負となるから層から独立な抽出を行なうよりも有利である。

$$\sigma^{\circ}_{\alpha\beta} = E(\bar{x}_{i\alpha} - E\bar{x}_{i\alpha})(\bar{x}_{j\beta} - E\bar{x}_{j\beta})$$

と定義すると、系統的 P.P.A 抽出の分散が、各層からの独立な P.P.ρ 抽出のものより小となるための concavity の条件は、$\alpha < \beta$ とすると

$$\sigma^{\circ}_{\alpha 1} - \sigma^{\circ}_{\beta 1} \leq \sigma^{\circ}_{\alpha 2} - \sigma_{\alpha 2} \leq \cdots \leq \sigma^{\circ}_{\beta k_\beta} - \sigma^{\circ}_{\alpha k_\beta} \leq 0$$

で与えられる。

b 等しい大きさの集落に対する系統的抽出

いま母集団が要素の集落からなっており、集落の大きさは等しい。すなわち同数の要素が含まれるものとしよう。はっきりさせるため母集団はM個の集落からなり、各集落にはN個の要素が含まれるものとする。ただし $M = cm$, $N = kn$ である。このときi集落のa要素で測定した特性の値を x_{ia} で表わし、i集落の全要素の合計を x_i で表わす。母集団の算術平均は、それから

$$M\bar{x} = \sum_{i=1}^{M} \bar{x}_i$$

ここで

$$N_i \bar{x}_i = x_i$$

a 標本内の集落を完全に調査する場合

まず $\bar{x}$ を $\bar{x}'$ で推定したいものとする。ここで $\bar{x}'$ はM個の集落のうちから大きさ m なる系統的標本をとり、標本内の各集落の全要素を調べてえられる標本算術平均である。

そうすると、

$$m\bar{x}' = \sum_{i=1}^{m} \bar{x}_i \tag{6.1}$$

と書ける。ここで $\bar{x}_i$ は標本に選ばれたi番目集落の平均値である。そうすると (1) から

$$\sigma^2_{\bar{x}'} = \frac{\sigma^2_b}{m}\{1 + (m-1)\bar{\rho}_c\}$$

となる。ここで $M\sigma^2_b = \sum_{i=1}^{M}(\bar{x}_i - \bar{x})^2$、また $\bar{\rho}_c$ は (1.P.b) の $\bar{\rho}_k$ で定義されるが $\bar{x}_i$ のところは $\bar{x}_i$ でおきかえる。いま集落の無作為抽出の理論から

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{N}\{1 - (N-1)\rho\}$$

がえられる。この σ^2 は母集団分散すなわち

$$MN\sigma^2 = \sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij} - \bar{x})^2$$

で、ρ は集落内要素の級内相関係数、すなわち

$$\sigma^2\rho = \sigma^2_b - \sigma^2_w/N-1$$

で与えられる。ただし

$$MN\sigma^2_w = \sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)^2$$

よって

$$(6.2)\quad \sigma^2_{\bar{x}'} = \frac{\sigma^2}{mN}\{1+(N-1)\rho\}\{1+(m-1)\bar{\rho}_c\}$$

となる．(6.2) の3つの因子の中 σ^2/mN は繰返しを許して大きさ mN の標本をとったときの分散で $1+(N-1)\rho$ は集落を用いたことによって生ずる因子．また $1+(m-1)\bar{\rho}_c$ は集落を系統的に抽出したことによる因子である．

6 層化と副次抽出

層化および副次抽出の可能性を考える場合には，可能な計画の数は極めて大きくなる．たとえば，母集団を層化して，抽出単位を大きさに比例した確率でぬき，次に系統的な副次抽出を行ない，再び系統的に副次抽出をして最後に恣作為な副次抽出を行なうというようにしてえた算術平均値の分散は簡単に計算できるであろう．しかしそのような研究はそれを用いなければならない実際問題との関連ややるべきであろう．ここでは実際に生ずるであろう多くの可能性を考察するよりむしろ系統的標本の系統的副次抽出の結果のみを述べることにする．他の多くの計画の分散は定理1および2を用いれば容易に求められる．

いま系統的に選ばれた m 個の集落の標本の各々から系統的に n 個の要素を副次抽出するものとする．そうして $\bar{x}$ の推定値を $\bar{x}'$ とする．ただし i 標本集落から選ばれた α 要素を $x'_{i\alpha}$ とすると

$$\bar{x}' = \left(\frac{1}{mn}\right)\sum_{i=1}^{m}\sum_{\alpha=1}^{n} x'_{i\alpha} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}'_i$$

で，

$$\bar{x}'_i = \left(\frac{1}{n}\right)\sum_{\alpha=1}^{m} x'_{i\alpha}$$

である．

定理2から再び

$$(6.3)\quad \sigma^2_{\bar{x}'} = \frac{\sigma^2}{mN}\{1+(N-1)\rho\}\{1+(m-1)\bar{\rho}_c\} + \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i^2}{mn}\{1+(m-1)\bar{\rho}_c\}$$

が与えられる．ここで σ_i^2 は i 集落内の分散で ρ_i は〔1.P.b〕で定義された．i 集落内の平均的系列相関である．同様に系統的標本を抽出した標本内の m 個の集落を1つの母集団と考えて副次抽出を行なった場合にも，$\bar{x}'$ の分散は簡単に計算できる．これは block の標本を選び，標本 block のすべての世帯に通番号をつけ，このリストから系統的標本を選ぶような場合に生ずるものである．しかしここでの目的には (6.2) の分析が重要であるから，前にもどって (6.2) を簡単に論じてみよう．

(6.2) からひき出せる最も重要な結論は，集落を系統的に抽出することは，それが望ましい場合でさえ集落を使うために分散が増加するという不利益を償いえないだろうということである．系統的な抽出によって同様な相対的利得が与えられるけれども，これらの利得は不等式

$$\{1+(N-1)\rho\}\{1+(n-1)\bar{\rho}_c\} < \frac{MN-mN}{MN-1}$$

を成立させるほど大きくはない．

我々がこれまでに論じなかった問題は次の通りである．母集団の要素を確率変数と考えることによって，異なる集落の要素間の平均的相関についてと同じく，単一の集落の要素間の平均的相関についても，同じ大きさの集落の系統的抽出の方が集落または恒々の要素の恣作為または層化恣作為抽出より分散が小さくなるといいうるための条件が与えられる．この際は直ちに求められるはずである．

7 二次元の系統的抽出

二次元の系統的抽出の問題は city から block の標本を抽出するとか，圃場からプロットの標本を抽出するとかの場合に

生じてくる．

city から block を選ぶ場合 問題を効果的に一次元に還元するため最も多く用いられる方法は，まず city の block またはその一部に，たとえば city の地図の右上隅から始めて第1行の右から左に，次に第2行の左から右というように通番号をつけてゆくやり方である．そうするとこれらの block 番号の系統的標本，したがって block 自体が選ばれることになる．明らかにこの方法はもし隣接 block に高度の相関があるときは最も有効というわけにはゆかないはずである．というのは非現実的な可能性を例にとれば，可能な標本を city のブロックの列になってしまうかもしれないからである．第2の二次元の系統的抽出の方法は，行の系統的標本および列の系統的標本を並び，グリッドの標本をうるものである．この計画は行または列に沿って"肥沃度の傾斜"がある場合には非常に効率が低くなる．

これら両者の方法の効率が悪いことの理由は，系統的標本の分散公式を調べれば判明する．もし上のように番号をつけてゆくとコレログラフが上に凹とは考えられなくなるから，層型との取り違いが非常に大きくなるであろう．グリット計画は，農業の系統的抽出と系統的副次抽出を行なう場合の特別な形のものであるが，これについては (6.3) と $\bar{\rho}_c$ が負のときでさえ，級内相関係数 ρ が，$\sigma^2_{\bar{x}}$ の値を大にする程十分大きくなることを考えてみればよい．

明らかに (6.3) は可能な標本が ρ をできるだけ小さくするように定義されることを示している．正方形の圃場には同じ処理をもつ Knut Vik の方形ブロットを可能な標本であると定義すればそうなるであろうが，不規則な圃場に対しても同様な可能な標本の定義が容易に与えられる．この問題はしかし

将来の研究にまたねばならない[3]．

By W. G. Madow (A. M. S. Volao. 1949. PP333-

3) この論文の引用文献の1つ〔6〕は著者の注意をひいた．その中のデータ，特に第3表は上に述べた見解と一致している．

参考文献

(1) W. G. Madow and L.H Madow "On the theory of Systematic sampling. I" A.M.S. Vol. 15 (1944) PP 1-24

(2) L.H. Madow "Systematic Sampling and its relation to other sampling design" J.A.S.A Vol 41 (1946) PP 204-207

(3) W.G. Cochran "Relative accuracy of systematic and stratified random samples for a certain class of populations" A.M.S. Vol 17. (1946) PP 164-177

(4) G.H Hardy. J.E. Littlewood and G. Polya Inequalities, Camb. Univ Press Lond & Newyork 1934. P43

(5) M. H. Hansen an W.N. Hurwitz "On the theory of sampling from finite populations A.M.S Vol 14 (1943) PP 333-362

(6) P.G. Homeyer & C.A. Black "Sampling replicated field experiments on oats for yield determinations" Socl. Sci. Soc. Proc. Vol 11 (1946) PP341-344

5 二次元のサンプリングについて
Problems in Plane sampling

1 要約

1次元における系統的抽出（systematic）と層化無作為抽出（stratified random sampling）の相対的精度を考察した後、一次元抽出の抽出誤差（linear sampling error）の推定の問題を論じた。

個別的な標本抽出（sampling）の方法を提案し、これらの方法の精度を表わす式を導いた。これらの式を特に理論的にも実際的にも当を得ていると思われる相関函数（Correlation function）に関連づけて、大標本に対して比較した。その結果多くの場合系統的抽出（systematic sampling）は層化抽出（stratified sampling）よりすつと精度のよいことがわかつた。抽出誤差（sampling error）を推定する方法を再び考察し、実例を与えた。この論文には抽出の行なわれる母集団の傾向（trend）の問題に関しても幾つかの注意を与えてある。

2 一次元における系統的標本と層化無作為標本の精度

(accuracy systematic and stratified random samples in one dimension) W.G. Cochran (1) は $x_1, x_2 \cdots x_{nk}$ なる母集団から、単純無作為（random）（r）、層化無作為（stratified random）（St）、系統的（systematic）（これは抽出を行なつて大きさ n の標本を抽出したときの平均値の分散を求める式を与えている。

今、要素 $x_1, x_2 \cdots\cdots x_{nk}$ が

$E(x_i)=\mu$, $E(x_i-\mu)^2=\sigma^2$ $E(x_i-\mu)(x_{i+u}-\mu)=\rho_u\sigma^2$

なる母集団から抽出されると仮定した。

ここに、$u<v$ なるとき常に $\rho_u > \rho_v > 0$

そうして、

(1) $\sigma_r^2=\frac{\sigma^2}{n}\left(1-\frac{1}{nk}\right)\left\{1-\frac{2}{kn(kn-1)}\sum_{u=1}^{kn-1}(kn-u)\rho_u\right\}$

(2) $$\sigma^2_{st} = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\left[1-\frac{2}{k(k-1)}\sum_{u=1}^{k-1}(k-u)\rho_u\right]$$

(3) $$\sigma^2_{sy} = \frac{\sigma^2}{n}\left(1-\frac{1}{k}\right)\cdot\left[1-\frac{2}{kn(k-1)}\sum_{u=1}^{kn-1}(kn-u)\rho_u + \frac{2k}{n(k-1)}\sum_{u=1}^{n-1}(n-u)\rho_{ku}\right]$$

なる式を導いた。

ρ_u の線形函数であるこれらの式を用いて、Cochran は幾つかのタイプのコレログラム（Correlogram）についてそれぞれの抽出方法の相対的な効率（efficiency）を比較した。もし我々が

(a) 各 z_i が平均値 μ_i、分散 σ_i^2 の母集団から抽出された標本で、

(b) μ_i は平均値 μ のまわりに、分散 σ^2 で分布する、

(c) $$E\{(\mu_i-\mu)(\mu_j-\mu)\} = \rho_{ij}\sigma^2$$ かつ

(d) $$\rho_u = \frac{1}{kn-u}\sum_{i=1}^{kn-u}\rho_{i,i+u}$$

という仮定を置くならば、(1)、(2)、(3)、が成立するためには、右辺に、

更に $$\frac{1}{n}\left(1-\frac{1}{k}\right)\cdot\frac{1}{kn}\sum_{i=1}^{kn}\sigma_i^2$$

なる変動を掛けなければならないことは容易にわかる。従って、Cochran の結果は色々な抽出方法の理論的な最大の相対効率を与えるものであることを忘れてはならない。ここで ρ_u は u だけ離れた標本間の平均的な相関である。この結果はおそらく、各点で平均的な要求度があり、標本がこの平均値のまわりで poisson 分布をしている様な昆虫要求度の標本調査で興味あるものである。

この場合加えらるべき変動は、

$$\frac{1}{n}\left(1-\frac{1}{k}\right)\cdot\frac{1}{kn}\sum_{i=1}^{kn}\mu_i \sim \frac{1}{n}\left(1-\frac{1}{k}\right)\mu$$

である。

我々が連続過程（continuous process）を標本抽出[1]するとき、n が大なら (1)、(2)、(3) の積分による等式を得ることができる。 $\sigma_n^2 \sim \frac{\sigma^2}{n}$

(4) $$\sigma^2_{st} \sim \frac{\sigma^2}{n}\left[1-\frac{2}{d^2}\int_0^d (d-u)\rho_u\,\delta u\right]$$

(5) $$\sigma^2_{sy} \sim \frac{\sigma^2}{n}\left[1-\frac{2}{d}\int_0^\infty \rho_u\,\delta u + 2\sum_{u=1}^{\infty}\rho_{du}\right]$$

ここで ρ は標本の u だけ離れた次々の要素間の相関で d は標本間の平均距離である。従って我々は次式を得る。

$$\frac{\sigma^2_{st}-\sigma^2_{sy}}{\sigma_n^2} \sim \frac{2}{d}\left[\int_0^d \frac{u}{d}\rho_u\,\delta u + \int_d^\infty \rho_u\,\delta u - d\sum_{u=1}^{\infty}\rho_{du}\right]$$

これは我々が層化無作為および系統的標本抽出の効率間の差のグラフを利用して研究に屢々用い得るものである。Fig 1 は4種類のコレログラムについてどの様に行なわれるかを示すものである。連続的マルコフ過程（Continuous Markov scheme）の場合には $\rho_u = \rho^u$ だから

$$\sigma^2_{st} \sim \frac{\sigma^2}{n}\left[1+\frac{2}{\log\rho d}+\frac{2}{(\log\rho d)^2}-\frac{2\rho d}{(\log\rho d)^2}\right]$$

$$\sigma^2_{sy} \sim \frac{\sigma^2}{n}\left[1+\frac{2}{\log\rho d}+\frac{2\rho d}{1-\rho d}\right]$$

で、これは Cochran の結果と一致する。

3. 反復と誤差の推定（Replication and the estimation of error）

Yates [2] 系統的標本（systematic sample）の誤差の推定にかゝわる困難を指摘した。しかしこの点について上記の公式を用いて研究することはやり甲斐があると思われる。

単純無作為、層化無作為、系統的抽出で、n が大きく k が一定と考えられるなら、平均値の推定量の分散は σ^2 乗 $(k)/n$ の形となる。ここで $F(k)$ は事実上 n と独立である。従って、もし標本誤差を

---

[1] 実際において我々が連続過程から標本を抽出する場合には、どれか恰も k の大なるときの離散過程（discontinuous process）と みなすときにのみ可能である。

定する方法があれば、我々は標本抽出を行なう系列（Series）を幾つかの部分（即ちブロック）に等分して、各部分の平均値の誤差を求めるとともに、全平均値の誤差のより正確な推定量を得るため、これを結合（Combine）することが可能となる。実際それが非常に大きい場合には、これらの部分を無作為に選び、誤差の推定量を求めることによって、観測値の個数を減少させることが考えられる。層化無作為抽出では、下（k）はれと完全に独立だから、我々は各層（strata）からの誤差の推定量を結合することができる。このことから、各層あたりg個の無作為に選ばれた要素をとり、誤差の推定量をきめるのに、自由度$g-1$の分散の組を結合するという普通に用いられている方法が導びかれる。もし我々が、標本を排反（exclusive）、すなわちどの二要素も一致しないように作れば、標本平均の推定分散を求めるにはこの分散に$1-g/k$をかければよい。

同様にして、任意に選ばれた出発点から始まる十分な長さをもつg個の系統的標本（systematic sample）の組を用いて、系統的標本の平均値の分散を推定することができる。しかしこの標本抽出は、実際に用いるには一層難しいから他の方法を考える必要がある。我々の系統標本は、この系列を分割した部分、即ちブロックの各々の中では一定となるよう選ぶことができるから、この抽出手続には全部でg個の系統標本だけが含まれる。即ち、我々のg個の標本を等間隔で選ぶというYatesの提唱する方法に従うべきである。この場合それらはより大きい系統的標本の副次標本（sub-sample）である。

この後者の方法は簡単でもあり広範囲の計画にも取入れることが出来るという長所を有するが、その使用には非常に慎重な考慮が必要である。離散的な場合を考えれば、我々は

$$(b)\quad \sigma^2\left(1-\frac{2}{k-1}\sum_{u=1}^{\infty}\rho_u+\frac{2k}{k-1}\sum_{u=1}^{\infty}\rho_{ku}\right)$$

を推定したい。しかしg個の等間隔系統的標本（evenly-spaced systematic sample）にもとずく分散の推定量には、ρ_{ku}/gなる形の項だけしか含まれない。一方g個の無作為に出ばれた（randomly chosen）系統的標本にもとずく分散の推定量には明らかに制限があり、多くの場合においてより代表的なものとなる。例えば$k=16$，$g=4$とすると、我々は分散の推定量における観測された相関（observing the correlation）$\rho_1 \cdots\cdots \rho_{15}$の相対的な出現度を比較することができる。

これの6つの例を第1表にあげる。乱数はFisherとYatesの表からとつた。ρ_uとρ_{16-u}との出現度数は等しいから、一諸に示してある。この表は、最初の二つの無作為に選ばれた標本のように殆んど系統的標本に近い無作為標本でも、このコレログラムを系統的に抽出するのを避けられるかを示している。多くの場合、どちらの方法を用いても可成り良い結果の得られることは明らかである。しかし、後の方法を用いた方がより正確である。第1表に示した標本を用いて色々な型のコレログラムに対する比較を第2表で行なった。勿論等間隔の系統的標本の組を分解できるような多くの種類のコレログラムを理論的に仮定することも可能であるが、しかし我々は終局的には、経験から得られたコレログラムの型について決定を下さなければならない。この点については、二次元の標本抽出を考察したあとで、もつと詳しく考えることにする。

第1表　それぞれ1メ単位毎に抽出を行なって4つの系統的標本をとるとき、分散推定値に現れる系列相関係数$\rho_1, \rho_2, \cdots\cdots, \rho_{15}$の出現度

ρ	等間隔に抽出した4つの系統的標本	4つの系統的標本の無作為に選んだ出発点						頻度数
		4, 7, 8, 12	3, 7 9, 12	3, 6 10, 13	4, 6, 7, 14	2, 8 11, 15	2, 6 11, 16	
1, 15		1	1		1			3
2, 14					1		1	2
3, 13		1		2	1	2		6
4, 12	4	2	2	1		1	1	7
5, 11		1	2				2	5
6, 10				1	1	1	1	4
7, 9			1	2	1	2	1	7
8	4	2			2			4

表2表　系統的標本で推定した $\frac{2}{15}\sum_{u=1}^{15} \rho_u$ の値

ρ_u	等間隔抽出の系統的標本	系統的標本の無作為に選んだ出発点						平均値	期待値
		1	2	3	4	5	6		
$1-0.2\,|u|$, $|u|\le 5$	0.17	0.27	0.20	0.17	0.30	0.17	0.13	0.21	0.27
$1-0.1\,|u|$, $|u|\le 10$	0.53	0.62	0.58	0.53	0.60	0.53	0.53	0.57	0.60
2^{-u}	0.04	0.13	0.12	0.06	0.15	0.06	0.07	0.10	0.13
$2^{-u/4}$	0.58	0.66	0.64	0.60	0.66	0.60	0.60	0.63	0.65
Kendallの系列1	−0.14	0.03	0.00	−0.05	0.16	−0.05	−0.05	0.01	0.07

このような抽出誤差の計算は当然系列を分割した各部分即ちブロックの対応する要素間の相関は無視できると仮定している。例えばこの場合には ρ_{16} と上記は無視できる。この場合 $\rho_{16}=1/16$ であるから結局（6）において要求される項は

$$2\left(\frac{1}{15}\sum_{u=1}^{\infty}\rho_u-\frac{16}{15}\sum_{u=1}^{\infty}\rho_{16u}\right)=0.56$$

となつて我々が推定しようとする項 $\frac{2}{15}\sum_{u=1}^{15}\rho_u=0.65$ と少し違つている。

4、二次元での標本抽出法（Methodo of sampling in 2 dimensions）

ある二次元空間から標本を抽出する方法の数は非常に多い。(3) 何故ならどの方向に対しても単純無作為、層化無作為、あるいは系統的抽出を用いることができるからである。この様に我々はこれらの方法の可能な組合せの何れをも考察することができる。例えば一方向に単純無作為で他の方向で系統的なとり方をする抽出は n.S 1 という記号で表わす。

更に我々は一方向が他に対して一直線になるような標本の組立を考え

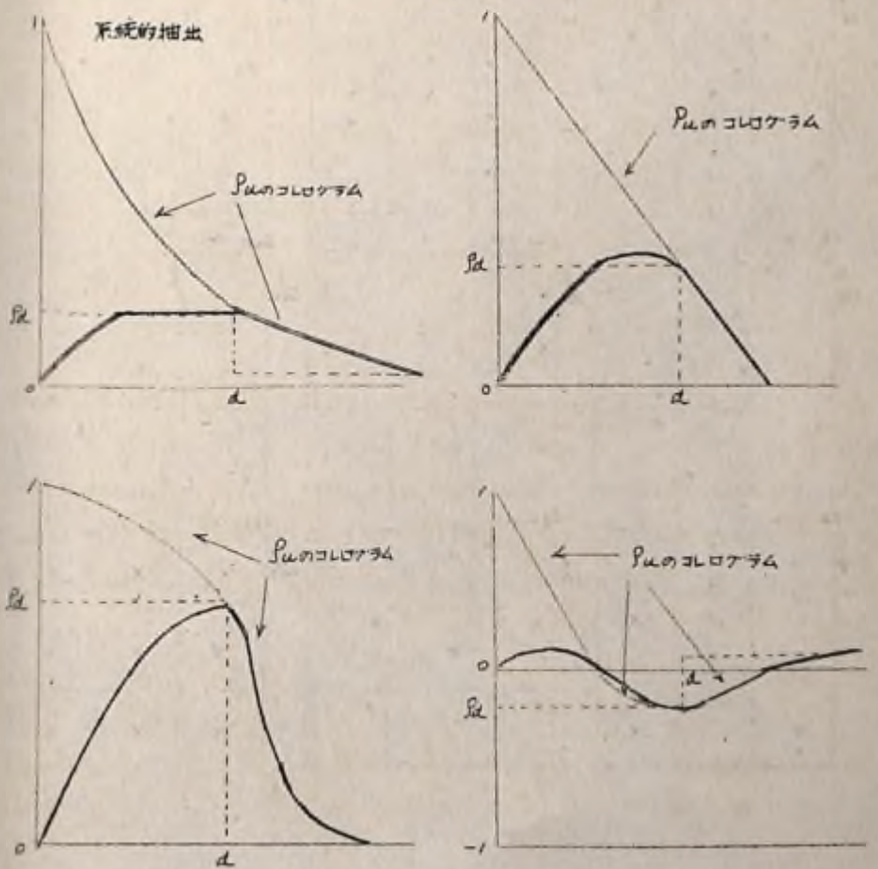
(3) 一般に我々はこの二次元空間が矩形であると考えるが、任意の形の地域に対しても容易に同様な結論を導びくことができる。

ることもできるし、また、独立にきめることもできる。添字 1 は一直線にとった aligned 標本を、また 0 は独立な標本を表わすのに用いる。例えばシステム n.S10 に従つて標本抽出ができるであろう。　標本抽出の幾つかの方法の例を Fig 2 に与える。

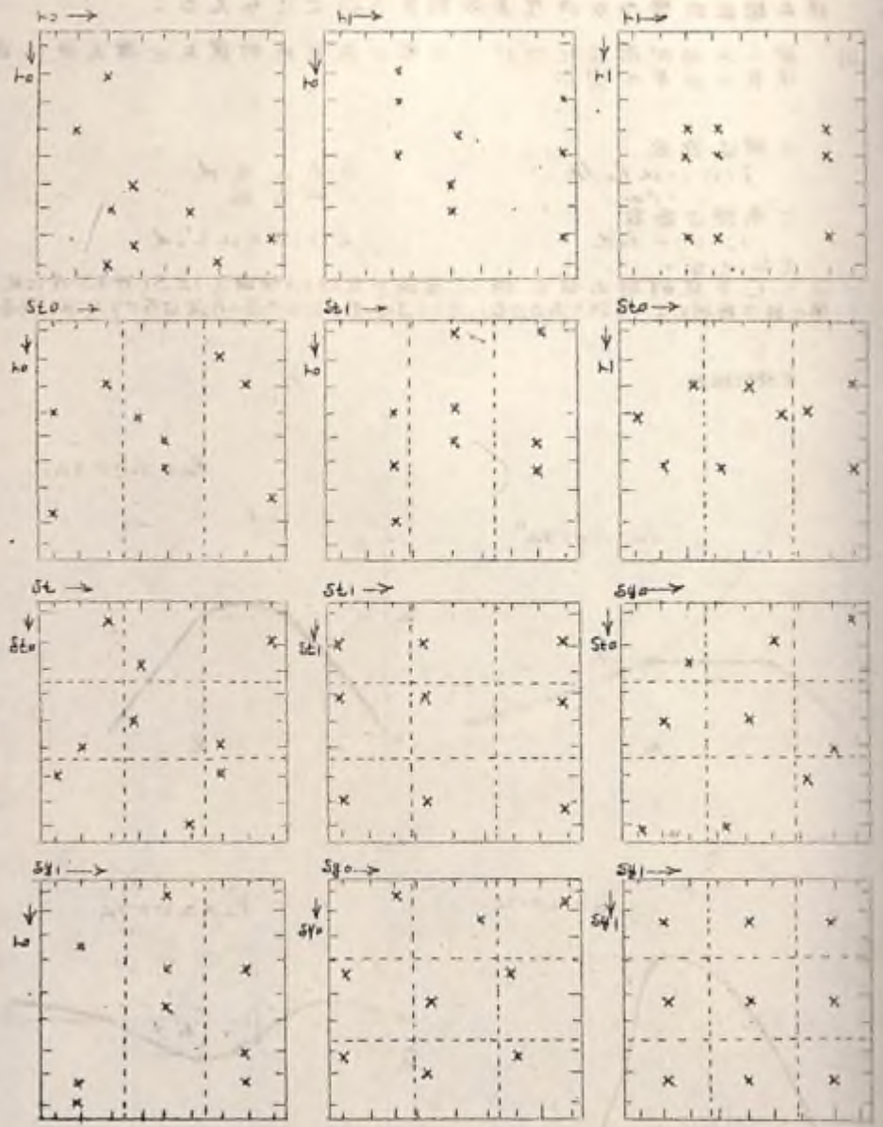
第1図　種々の相関函数について比較した系統的標本と層化無作為標本の効率の図示

太線は函数
$f(u)=u\rho_u/d$　　$0\le u\le d$
$=\rho_u$　　$d\le u$
で点線は函数
$f_2(u)=\rho_i d$　　$(i-1)d<u\le id$
を与える。

このように系統的抽出は太線下の面積が点線下の面積より大きいかどうかに従つて、層化無作為抽出より有効であるかないかができる。最も効率の高い方法は各グラフに示してある。



第2図　層別の抽出法　この場合 $n_1=n_2=k_1=k_2=3$



5　二次元の標本抽出の精度

（Accuracy of sampling in two dimensions）

我々は，二方向における平均間隔が k_1，k_2 となる様に，要素 x_{ij} $(i=1,2,\cdots\cdots,n,k_1;\ j=1,2\cdots\cdots n_2 k_2)$，から抽出された，$n_1 n_2$ 個の要素からなる標本を考えよう．（これは仮説的無限母集団から抽出された1つの有限母集団を形成する．）これらのパラメータ（parameter）はもし必要なら抽出方法のあとに［　］をつけて示すことにする．例えば $\bar{x}.Sy_0(n_1k_1;n_2k_2)$

$\bar{x}$ は考えている抽出方法で作られた標本平均を表わし，どでこの標本の構成因子を表わす。同様に x_{ij} は

$$E(x_{ij})=\mu \qquad E(x_{ij}-\mu)^2=\sigma^2$$

$$E(x_{ij}-\mu)(x_{i+u\ j+v}-\mu)=\rho_{iju v}\sigma^2$$

なる母集団から抽出されると仮定する。

更に我々は

$$\rho_{uv}=\rho_{-u,-v} \quad を$$

$$\sum_i\sum_j \rho_{iju v}=(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}$$

なる関係で定義するため，ijとuvのすべての可能な値にわたつて，$\rho_{iju v}$ を平均することができる．これらの定義の目的は，有限母集団のパラメータを取扱う場合の困難性を，この母集団自身が無限母集団から得られた1つの標本であると考えることによって，除去することにある。cochronも同様な手法を用いている。

5.1.　単純無作為抽出

$$\sigma^2(\bar{x})=\frac{1}{2}E(x_1-x_2)^2=E(x_1-\mu)^2-E(x_1-\mu)(x_2-\mu)$$

なることは容易に証明できる。ここで x_1 と x_2 は独立な標本である。同様にして

$$E(x_1-\mu)(x_2-\mu)=E(x_1'-\mu')(x_2'-\mu')$$

$$=\frac{\sigma^2}{k_1k_2n_1n_2}\left[1+\frac{1}{k_1k_2n_1n_2}\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\right]$$

ここで二重和（double summation）(4) は $u=v=0$ を除いた $|u|\leq k_1 n_1$, $|v|\leq k_2 n_2$ で与えられる範囲Sの上で存在する。ここで我々は種々な型の無作為標本について $E(\bar{x}_1-\mu)^2$ を評価しなければならない。

$$E(\bar{x}_1-\mu)^2=\frac{\sigma^2}{n_1 n_2}\cdot\left[1+\frac{n_1 n_2-1}{k_1k_2n_1n_2(k_1k_2n_1n_2-1)}\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\right]$$

r_0 r_{2v} に対して

$$=\frac{\sigma^2}{n_1n_2}\left[1+\frac{n_1-1}{k_1k_2n_1(k_1k_2n_1n_2-1)}\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\right.$$

$$\left.+\frac{2(n_2-1)}{k_2n_2(k_2n_2-1)}\sum_{v=1}^{k_2n_2}(k_2n_2-v)\rho_{0v}\right]$$

r_1 r_0 に対して

$$=\frac{\sigma^2}{n_1n_2}\left[1+\frac{(n_1-1)(n_2-1)}{k_1k_2n_1n_2(k_1k_2n_1n_2-1)}\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\right.$$

$$\left.+\frac{2(n_2-1)}{k_2n_2(k_2n_2-1)}\sum_{v=1}^{k_2n_2}(k_2n_2-v)\rho_{0v}+\frac{2(n_1-1)}{k_1n_1(k_1n_1-1)}\sum_{u=1}^{k_1n_1}(k_1n_1-u)\rho_{u0}\right]$$

r_1 r_1 に対して

故に

(7)
$$\begin{cases}\sigma^2(r_0r_0)=\dfrac{1}{n_1n_2}\left(1-\dfrac{1}{k_1k_2}\right)\sigma^2\\ \cdot\left[1-\dfrac{1}{k_1k_2n_1n_2(k_1k_2n_1n_2-1)}\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\right]\end{cases}$$

(8)
$$\begin{cases}\sigma^2(r_1r_0)=\dfrac{1}{n_1n_2}\left(1-\dfrac{1}{k_1k_2}\right)\sigma^2\left[1-\dfrac{k_1k_2n_2-1}{(k_1k_2-1)k_1k_2n_1n_2(k_1k_2n_1n_2-1)}\right.\\ \left.\cdot\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}+\dfrac{2(n_2-1)}{k_2n_2(k_2n_2-1)}\sum_{v=1}^{k_2n_2}(k_2n_2-v)\rho_{0v}\right]\end{cases}$$

(9)
$$\begin{cases}\sigma^2(r_1r_1)=\dfrac{1}{n_1n_2}\left(1-\dfrac{1}{k_1k_2}\right)\sigma^2\left[1-\dfrac{k_1k_2(n_1n_2-1)-1}{(k_1k_2-1)k_1k_2n_1n_2(k_1k_2n_1n_2-1)}\right.\\ \cdot\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}+\dfrac{2k_1(n_2-1)}{(k_1k_2-1)n_2(k_2n_2-1)}\\ \left.\cdot\sum_{v=1}^{k_2n_2}(k_2n_2-v)\rho_{0v}+\dfrac{2k_2(n_1-1)}{(k_1k_2-1)n_1(k_1n_1-1)}\sum_{v=1}^{k_1n_1}(k_1n_1-v)\rho_{v0}\right]\end{cases}$$

なることは容易に証明できる。

5.6 層化無作為抽出（Stratified random sampling）

もし第i層で抽出される要素の平均値 $\bar{x}_i'$ が $\bar{x}_j'$ と独立なら、層化標本抽出の方法に対する分散を推定することができる。何故なら このとき

$$E(\bar{x}-\bar{\bar{x}})^2=E(\bar{\bar{x}}_i'-\bar{\bar{x}}_i)^2/n$$

であるから。ここで $\bar{\bar{x}}$ は標本抽出の行なわれる有限母集団の平均値である。故に

(10)
$$\begin{cases}\sigma^2(\text{sto}\,r_0)=\dfrac{1}{n}\sigma^2\{r_0r_0(1/k_1;n_2k_2)\\ =\dfrac{1}{n_1n_2}\left(1-\dfrac{1}{k_1k_2}\right)\sigma^2\left[1-\dfrac{1}{k_1k_2n_2(k_1k_2n_2-1)}\right.\\ \left.\cdot\sum\sum(k_1-|u|)(k_2n_2-|v|)\rho_{uv}\right]\end{cases}$$

(11)
$$\begin{cases}\sigma^2(\text{st}_1\text{to})=\dfrac{1}{n_1n_2}\left(1-\dfrac{1}{k_1k_2}\right)\sigma^2\left[1-\dfrac{1}{k_1k_2n_2(k_1k_2-1)}\right.\\ \cdot\sum\sum(k_1-|u|)(k_2n_2-|v|)\rho_{uv}\\ \left.+\dfrac{2k_1(n_2-1)}{(k_1k_2-1)n_2(k_2n_2-1)}\sum_{v=1}^{k_2n_2}(k_2n_2-v)\rho_{0v}\right]\end{cases}$$

(12)
$$\begin{cases}\sigma^2(\ldots\text{st}_0)=\dfrac{1}{n_1n_2}\left(1-\dfrac{1}{k_1k_2}\right)\sigma^2\left[1-\dfrac{1}{k_1k_2(k_1k_2-1)}\right.\\ \left.\cdot\sum\sum(k_1-|u|)(k_2-|v|)\rho_{uv}\right]\end{cases}$$

(4) 特に断らない限り、一般に二重和（double summation）は $u=v=0$ を除く係数が正となる適当な範囲上で存在する。

他の抽出方法の分散を推定するため、我々は（8）−（12）式を導びくのに用いた一般公式を使うことにする。

x'_i を標本 x の任意の要素とすると

$$(\bar{x}-\bar{\bar{x}})^2=\frac{1}{n_1n_2}\left[\sum(x'_i-\bar{\bar{x}})^2-\sum(x'_i-\bar{x})^2\right]$$

$$=\frac{1}{n_1n_2}\left\{\sum(x'_i-\bar{\bar{x}})^2-\frac{n_1n_2-1}{n_1n_2}\sum(x'_i-\mu)^2\right.$$

$$\left.+\frac{2}{n_1n_2}\sum\sum(x'_i-\mu)(x'_j-\mu)^\delta\right\}.$$

よって

$$(13)\begin{cases}\sigma^2(\bar{x})=E(\bar{x}-\bar{\bar{x}})^2\\ =\frac{k_1k_2n_1n_2-1}{k_1k_2n_1n_2}\sigma^2\left\{1-\frac{1}{k_1k_2n_1n_2(k_1k_2n_1n_2-1)}\sum\sum(k_1n_1-|u|)\right.\\ \left.\cdot(k_2n_2-|v|)\rho_{uv}\right\}-\frac{n_1n_2-1}{n_1n_2}\sigma^2+\frac{n_1n_2-1}{n_1n_2}E(x'_i-\mu)(x'_j-\mu)\\ =\frac{1}{n_1n_2}\left(1-\frac{1}{k_1k_2}\right)\sigma^2\left\{1-\frac{1}{k_1k_2n_1n_2(k_1k_2-1)}\sum\sum(k_1n_1-|u|)\right.\\ \left.\cdot(k_2n_2-|v|)\rho_{uv}+\frac{k_1k_2(n_1n_2-1)}{k_1k_2-1}\frac{E(x'_i-\mu)(x'_j-\mu)}{\sigma^2}\right]\end{cases}$$

このように $E(x'_i-\mu)(x'_j-\mu)/\sigma^2$ の推定が可能でありさえすれば（13）式はすべての抽出方法に対する誤差を与える。

1例として（12式）を導いてみよう。任意の構成要素 x'_i をとれば、次の構成要素 x'_j は、x'_i と同じ層の中に x'_j が占めることのできない k_1k_2-1 個の場所が存在するということを除いては、x'_i に関して無作為に配置されるであろう。

かくして期待された相関 $E(x'_i-\mu)(x'_j-\mu)/\sigma^2$ は

$$(14)\begin{cases}\frac{1}{k_1^2k_2^2n_1n_2(n_1n_2-1)}\sum\sum k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\\ -\frac{1}{k_1^2k_2^2(n_1n_2-1)}\sum\sum(k_1-|u|)(k_2-|v|)\rho_{uv}\end{cases}$$

で与えられる。

（14）を（13）に代入すれば、St_0 St_0 に対する分散の式（12）が得られる。

同様にして St_1 St_1 に対する式

$$(15)\begin{cases}E(x'_i-\mu)(x'_j-\mu)\\ =\frac{1}{k_1k_2(n_1n_2-1)}\left\{\frac{1}{k_1k_2n_1n_2}\sum\sum(k_1n_1-|u|)(k_2n_2-|v|)\rho_{uv}\right.\\ -\frac{1}{k_1k_2n_1}\sum\sum(k_1n_1-|u|)(k_2-|v|)\rho_{uv}\\ -\frac{1}{k_1k_2n_2}\sum\sum(k_1-|u|)(k_2n_2-|u|)\rho_{uv}+\frac{1}{k_1k_2^2}\sum\sum(k_1-|u|)\\ \cdot(k_2-|v|)\rho_{uv}+\frac{2(k_1k_2n_1-1)}{k_1n_1(k_1n_1-1)}\sum_{u=1}^{k_1n_1}(k_1n_1-u)\rho_{ov}\\ -\frac{2(k_1k_2-1)}{k_1(k_1-1)}\sum_{u=1}^{k_1}(k_1-u)\rho_{uv}+\frac{2(k_1k_2n_2-1)}{k_2n_2(k_2n_2-1)}\sum_{v=1}^{k_2n_2}(k_2n_2-v)\rho_{ov}\\ \left.-\frac{2(k_1k_2-1)}{k_2(k_2-1)}\sum_{v=1}^{k_2}(k_2-v)\rho_{uv}\right]\end{cases}$$

このようにして我々はすべての型の層化無作為抽出について $\sigma^2(\bar{x})$ を計算することができる。

5c. 系統的抽出（Systematic sampling）

層化無作為抽出において用いたと同様なやり方で、系統的抽出の分散を計算するために（13）式を用いることができる。

可能な3つの抽出方法に対する $E(x'_i-\mu)(x'_j-\mu)$ の値を下に掲げる。

Sy_1 Sy_1 については

$$(16)\quad E(x'_i-\mu)(x'_j-\mu)=\frac{1}{n_1n_2(n_1n_2-1)}\sum\sum$$

$$(n_1-|u|)(n_2-|v|)\rho_{k_1u,k_2v}$$

S_{ij} r_0 については．

$$(17)\begin{cases} E(x_i'-\mu)(x_j'-\mu) = \dfrac{1}{k_2^2 n_1 n_2 (n_1 n_2 - 1)} \sum\sum (n_1 - |u|)\cdot \\ \qquad (k_2 n_2 - |v|)\rho_{k_1 u, k_2 v} \\ -\dfrac{2(k_2-1)}{k_2^2 n_2 (n_1 n_2 - 1)(k_2 n_2 - 1)} \sum_{u=1}^{k_2 n_2} (k_2 n_2 - v)\rho_{0v} \end{cases}$$

S_{yo} S_{yo} については

$$(18)\begin{cases} E(x_i'-\mu)(x_j'-\mu) = \dfrac{1}{k_1 k_2 (n_1 n_2 - 1)}\Big\{ \dfrac{1}{k_1 k_2 n_1 n_2} \sum\sum (k_1 n_1 - |u|) \\ \cdot (k_2 n_2 - |v|)\rho_{uv} - \dfrac{1}{k_1 k_2 n} \sum\sum (k_1 n_1 - |u|)(k_2 - |v|)\rho_{uv} \\ -\dfrac{1}{k_1 k_2 n_2} \sum\sum (k_1 - |u|)(k_2 n_2 - |v|)\rho_{uv} + \dfrac{1}{k_1 k_2} \sum\sum (k_1 - |u|) \\ \cdot (k_2 - |v|)\rho_{uv} + \dfrac{k_2}{k_2 n_2} \sum\sum (n_1 - |u|)(k_2 - |v|)\rho_{k_1 u, v} \\ -\dfrac{2k_1}{k_2} \sum_{v=1}^{k_2} (k_2 - v)\rho_{0v} + \dfrac{k_2}{k_1 n_2} \sum\sum (k_1 - |u|)(n_2 - |v|)\rho_{u, k_2 v} \\ -\dfrac{2k_2}{k_1} \sum_{u=1}^{k_1} (k_1 - u)\rho_{u0} \Big\} \end{cases}$$

（18）の導き方は（15）のそれを比較できるものである．

6　標本を直線的にとる場合の影響（Effect of alignment）

直線的抽出（alignment）の効果は，色々の標本の分散を調べるかあるいは直接（13）式を用いて研究することができる．

逐次的および層化無作為抽出での直線的抽出の効果は標本分散を

$$\sum\sum a_{uv}(\rho_{0v} - \rho_{uv}) + \sum\sum b_{uv}(\rho_{v0} - \rho_{uv})$$ ここで $a_{uv} \geq 0$, $b_{uv} \geq 0$

なる量だけ増加させる．

これは単調減少の相関函数および、実際において出現する大多数の函数に対しては正である．この様に直線的抽出は逐次的無作為および層に無作為抽出の分散を増加させる．

系統的標本については問題はもっと複雑であるが、しかし大抵

において、分散は

$$\sum\sum a_{uv}(\rho_{k_1 u, k_2 v} - \bar{\rho}_{k_1 u, k_2 v})$$

なる量だけ増加する．ここで $a_{uv} \geq 0$ で、$\rho_{k_1 u, k_2 v}$ は 0 でない u, v に対しては中心が $\rho_{k_1 u, k_2 v}$ 中心 $\rho_{0 k_2 v}$ なる長さ k_1 の直線上の平均値である．（$v=0$ なるときも同様）

これが正となるか負となるかは、相関函数に関係してくるから、この問題は出現する相関函数の型について研究しなければならない．

7　極限における式（Limiting forms）

連続的過程で，n_1　n_2 が大きいとき、我々は一次元の抽出（linear sampling）の場合と同様にして、$\sum\sum \delta_{d_1 u} \delta_{d_2 u}$ が収斂する限り、抽出分散の積分近似を得ることができる．

即ち

$$(19)\quad \sigma^2(r_0 r_0) = \sigma^2(\mathrm{stor}_0) \sim \sigma^2/n_1 n_2$$

$$(20)\quad \sigma^2(r_1 r_0) \sim \frac{\sigma^2}{n_1 n_2}\left[1 + \frac{2}{d_2}\int_0^\infty \rho_{0u}\,\delta u\right]$$

$$(21)\quad \sigma^2(r_1 r_1) \sim \frac{\sigma^2}{n_1 n_2}\left[1 + \frac{2}{d_2}\int_0^\infty \rho_{0v}\,\delta v + \frac{2}{d_1}\int_0^\infty \rho_{u0}\,\delta u\right]$$

$$(22)\quad \sigma^2(\mathrm{st}_1 r_0) \sim \frac{\sigma^2}{n_1 n_2}\left[1 - \frac{1}{d_1^2 d_2}\int_{-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1 - |u|)\rho_{uv}\,\delta v + \frac{2}{d_2}\int_0^\infty \rho_{0v}\,\delta v\right]$$

$$(23)\quad \sigma^2(\mathrm{st}_0 \mathrm{st}_0) \sim \frac{\sigma^2}{n_1 n_2}\left[1 - \frac{1}{d_1^2 d_2^2}\int_{-d_2}^{d_2}\int_{-d_1}^{d_1}(d_1 - |u|)(d_2 - |v|)\rho_{uv}\,\delta u\,\delta v\right]$$

$$(24)\begin{cases} \sigma^2(\mathrm{st}_1 \mathrm{st}_1) \sim \dfrac{\sigma^2}{n_1 n_2}\Big\{1 - \dfrac{1}{d_1^2 d_2}\int_{-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1 - |u|)\rho_{uv}\,\delta u\,\delta v \\ \dfrac{1}{d_1 d_2}\int_{-d_1}^{d_1}\int_{-\infty}^{\infty}(d_2 - |v|)\rho_{uv}\,\delta u\,\delta v + \dfrac{1}{d_1^2 d_2^2}\int_{-d_2}^{d_2}\int_{-d_1}^{d_1}(d_1 - |u|) \\ \cdot (d_2 - |v|)\rho_{uv}\,\delta u\,\delta v + \dfrac{2}{d_1}\int_0^\infty \rho_{u0}\,\delta u - \dfrac{2}{d_1^2}\int_0^{d_1}(d_1 - u)\rho_{u0}\,\delta v \\ + \dfrac{2}{d_2}\int_0^\infty \rho_{0v}\,\delta v - \dfrac{2}{d_2^2}\int_0^{d_2}(d_2 - v)\rho_{0v}\,\delta v\Big\} \end{cases}$$

$$(25)\quad \begin{cases} \sigma^2(S_{y_1+\tau_0}) \sim \dfrac{\sigma^2}{n_1 n_2}\Big[1-\dfrac{1}{d_1 d_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{u v}\,\delta_u\,\delta_v \\ +\dfrac{1}{d_2}\sum_{u=-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{d_1 u_1 v}\,\delta_v-\dfrac{2}{d_2}\int_0^{\infty}\rho_{0v}\,\delta_v\Big] \end{cases}$$

$$(26)\quad \sigma^2(S_y S_{y_1}) \sim \frac{\sigma^2}{n_1 n_2}\Big[\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty}\rho_{d_1 u, d_2 v}-\frac{1}{d_1 d_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{uv}\,\delta_u\,\delta_v\Big]$$

$$(27)\quad \begin{cases} \sigma^2(S_{y_0} S_{y_0}) \sim \dfrac{\sigma^2}{n_1 n_2}\Big[1-\dfrac{1}{d_1^2 d_2}\int_{-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1-|u|)\,\rho_{uv}\,\delta_u\,\delta_v \\ -\dfrac{1}{d_1 d_2}\int_{-d_2}^{d_1}\int_{-\infty}^{\infty}(d_2-|v|)\,\rho_{uv}\,\delta_u\,\delta_v \\ +\dfrac{1}{d_1^2 d_2^2}\int_{-d_2}^{d_2}\int_{-d_1}^{d_1}(d_1-|u|)(d_2-|v|)\,\rho_{uv}\,\delta_u\,\delta_v \\ +\dfrac{1}{d_2^2}\sum_{u=-\infty}^{\infty}\int_{-d_2}^{d_2}(d_2-|v|)\,\rho_{d_1 u, v}\,\delta_v-\dfrac{1}{d_2^2}\int_{-d_2}^{d_2}(d_2-|v|)\,\rho_{0v}\,\delta_v \\ +\dfrac{1}{d_1^2}\sum_{v=-\infty}^{\infty}\int_{-d_1}^{d_1}(d_1-|u|)\,\rho_{u,d_2 v}\,\delta_u-\dfrac{1}{d_2}\int_{-d_2}^{d_2}(d_1-|u|)\,\rho_{u0}\,\delta_u\Big] \end{cases}$$

3. $\rho_{uv}=\rho_u\rho_v$ なる特別の場合 (particular case where $\rho_{uv}=\rho_u\rho_v$)

$\rho_{uv}=\rho_u\rho_v^{(5)}$ なら、これらの式の大部分は非常に簡単となることに注意せよ。

もし我々が S_{yu} および S_{tv} に対して同様に式で

$$S_{yu}=1-\frac{2}{d_1}\int_0^{\infty}\rho_u\,\delta_u+2\sum_{u=1}^{\infty}\rho_{d_1 u},$$

$$S_{tu}=1-\frac{2}{d_2}\int_0^{d_1}(d_1-u)\,\rho_u\,\delta_u,$$

と書き、同様に

(5) これが正しい自己相関函数 (auto correlation function) であるための必要十分条件は ρ_u および ρ_v がともに自己相関函数となることである。

$$f_1=\frac{2}{d_2}\int_0^{\infty}\rho_v\,\delta_v,\quad f_1'=\frac{2}{d_2^2}\int_0^{d_2}(d_2-v)\rho_v\delta_v,\quad f_1''=2\sum_{v=1}^{\infty}\rho_{d_2 v},$$

$$f_2=\frac{2}{d_1}\int_0^{\infty}\rho_u\delta_u,\quad f_2'=\frac{2}{d_1^2}\int_0^{d_1}(d_1-u)\rho_u\delta_u,\quad f_2''=2\sum_{u=1}^{\infty}\rho_{d_1 u}$$

と書くことにすれば、例えば

$$(28)\quad \sigma^2(\gamma_1\gamma_0) \sim \frac{\sigma^2}{n_1 n_2}(1+f_1)$$

$$(29)\quad \sigma^2(\tau_1\tau_0) \sim \frac{\sigma^2}{n_1 n_2}(1+f_1+f_2)$$

$$(30)\quad \sigma^2(S_{t_0} S_{t_0}) \sim \frac{\sigma^2}{n_1 n_2}(S_{tu}S_{tv}+S_{tu}+S_{tv})$$

$$(31)\quad \sigma^2(S_{t_1} S_{t_1}) \sim \frac{\sigma^2}{n_1 n_2}(S_{tu}S_{tv}+f_1 S_{tu}+f_2 S_{tv})$$

$$(32)\quad \sigma^2(S_{y_1} S_{y_1}) \sim \frac{\sigma^2}{n_1 n_2}(S_{yu}S_{yv}+f_1 S_{yu}+f_2 S_{yv})$$

$$(33)\quad \sigma^2(S_{y_0} S_{y_0}) \sim \frac{\sigma^2}{n_1 n_2}(S_{tu}S_{tv}+f_1' S_{yu}+f_2' S_{yv})$$

これらの式から

$$(34)\quad \begin{cases} \sigma^2(S_{t_1} S_{t_1})-\sigma^2(S_{y_1} S_{y_1}) \sim \dfrac{\sigma^2}{n_1 n_2} \\ \cdot\big[(S_{tu}S_{tv}-S_{yu}S_{yv})+f_1(S_{tu}-S_{yu})+f_2(S_{tu}-S_{yv})\big] \end{cases}$$

$$(35)\quad \begin{cases} \sigma^2(S_{y_1} S_{y_2})-\sigma^2(S_{t_0} S_{t_0}) \sim \dfrac{\sigma^2}{n_1 n_2} \\ \cdot\big[\{(1-S_{t_1 u})(1-S_{yv})-(1-S_{t_2 u})(1-S_{tv})\}+f_1' S_{yu}+f_2'' S_{yv}\big] \end{cases}$$

$$(36)\quad \sigma^2(S_{t_0} S_{t_0})-\sigma^2(S_{y_0} S_{y_0}) \sim \frac{\sigma^2}{n_1 n_2}\big[f_1'(S_{tu}-S_{yu})+f_2'(S_{tu}-S_{yv})\big]$$

が得られる。

(34), (35), (36) 式から、一次元の結果を用いて、二次元の

標本の分散の比較が可能となる。

実際的場合の大部分では、f は正であり $S_{tu} \gg S_{yu}$, $S_{tv} \gg S_{yv}$ だから

(37) $\sigma^2(S_t, S_{t'}) \geqslant \sigma^2(S_y, S_{y'}) \geqslant \sigma^2(S_{t_0}, S_{t_0}) \geqslant \sigma^2(S_{y_0}, S_{y_0})$

となる。$S_{d,u} = \rho_1^{|u|}$ および $S_{d,2v} = \rho_2^{|v|}$ に対する $\sigma^2(S_{t_0}, S_{t_0})/\sigma^2(t_0, t_0)$, $\sigma^2(S_y, S_{y'})/\sigma^2(t_0, t_0)$, $\sigma^2(S_{y_0}, S_{y_0})/\sigma^2(t_0, t_0)$ の値を第3表に与える。与えられた数の標本に対して(d_1, d_2 一定)は、$\rho_1 = \rho_2$ のとき $\sigma^2(S_{t_0}, S_{t_0})$, $\sigma^2(S_y, S_{y'})$ および $\sigma^2(S_{y_0}, S_{y_0})$ が最小となることは容易にわかる。

表示された式は $\rho_1 = \rho_2 = 0$ に対しては 1 をとり、ρ_1 および ρ_2 が 1 に近づくに従って、それぞれ極限値 0, 2/3, 0 および 2 に近ずく。

0.4 以上異なる ρ_1 と ρ_2 に対しては、$S_y, S_{y'}$ によつて組まれた格子(grid)は単純無作為抽出法より効率が低いことは注目すべきことである。

しかし $\rho_{uv} = \rho_u \rho_v'$ なる形の函数は centrally-symmetric function よりも現実において現れることが少ないようである。これは座標軸の選び方と独立である。この理由で、我々は最後の型の函数を次に考察する。

9. Centrally-symmetric な相関函数

Dedebant と Wehrle [3] は $\rho(u,v)$ が相関函数(Correlation function)であるための必要十分条件を

(38) $\rho(u,v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos(wu - \mu v)\, \delta F(w, \mu)$

あるいは逆に

(39) $f(w,\mu) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos(wu - \mu v)\, \rho(u,v)\, \delta u \delta v$

のように与えている。

(6) 都市調査を行なう場合には、二点間の相関が street 内および street 間の相関からきまるから、この函数が現実的となる。

第3表 種々の ρ_1 および ρ_2 の値に対する系統訂抽出と無作為抽出の効率の比較

ρ_2 \ ρ_1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	1.000 1.000 1.000 1.000	1.000 1.222 1.030 1.000	1.000 1.500 1.000 1.000	1.000 1.857 1.000 1.000	1.000 2.333 1.000 1.000	1.000 3.000 1.000 1.000	1.000 4.000 1.000 1.000	1.000 5.667 1.000 1.000	1.000 9.000 1.000 1.000	1.000 19.000 1.000 1.000	1.000 ∞ 1.000 1.000
0.1		0.720 0.739 0.596 1.21	0.689 0.754 0.534 1.25	0.632 0.827 0.493 1.28	0.601 0.956 0.462 1.30	0.575 1.160 0.437 1.31	0.551 1.488 0.416 1.32	0.529 2.055 0.398 1.33	0.508 3.215 0.382 1.33	0.489 6.734 0.367 1.33	0.471 ∞ 0.354 1.33
0.2			0.609 0.706 0.462 1.32	0.565 0.721 0.446 1.36	0.529 0.788 0.380 1.39	0.497 0.914 0.354 1.41	0.469 1.134 0.328 1.43	0.443 1.532 0.307 1.44	0.419 2.362 0.289 1.45	0.396 4.911 0.272 1.46	0.375 ∞ 0.257 1.46
0.3				0.516 0.689 0.365 1.41	0.476 0.797 0.327 1.45	0.441 0.938 0.297 1.49	0.409 0.924 0.271 1.51	0.380 1.209 0.249 1.53	0.354 1.825 0.229 1.54	0.328 3.751 0.212 1.55	0.305 ∞ 0.196 1.55
0.4					0.432 0.680 0.298 1.56	0.398 0.902 0.256 1.54	0.360 0.787 0.229 1.57	0.329 0.983 0.206 1.60	0.300 1.437 0.185 1.62	0.272 2.900 0.167 1.63	0.247 ∞ 0.151 1.64
0.5						0.354 0.695 0.223 1.59	0.317 0.703 0.195 1.63	0.284 0.821 0.171 1.66	0.253 1.139 0.150 1.68	0.223 2.228 0.132 1.70	0.196 ∞ 0.115 1.71
0.6							0.279 0.671 0.167 1.67	0.243 0.712 0.142 1.71	0.210 0.908 0.121 1.74	0.180 1.679 0.102 1.76	0.151 ∞ 0.085 1.78
0.7								0.206 0.664 0.118 1.75	0.172 0.742 0.096 1.79	0.139 1.226 0.077 1.82	0.109 ∞ 0.059 1.84
0.8									0.136 0.667 0.074 1.84	0.102 0.863 0.055 1.87	0.070 ∞ 0.037 1.89
0.9										0.067 0.667 0.035 1.92	0.034 ∞ 0.018 1.95

centrally - symmetric 相関函数について

$u=r\cos\theta$, $v=r\sin\theta$ とおけば $\rho(u,v)=\rho(r)$ で

$$f(w,w)=\frac{1}{(2\pi)^2}\int_0^\infty\int_0^{2\pi}\cos(r\sqrt{w^2+w^2}\cos\theta_1)\rho(r)r\,d\theta\,dr$$

ここで $\theta_1=\theta+\tan^{-1}(w/w)$

$$=\frac{1}{2\pi}\int_0^\infty J_0(rT)\rho(r)dr \qquad \text{ここでて}=\sqrt{w^2+w^2}$$

従って $\rho(u,v)$ が centrally - symmetric なら、逆に $f(w,w)$ もそうなるから、

(40) $$f(T)=\frac{1}{2\pi}\int_0^\infty J_0(rT)\rho(r)r\,\delta r$$

および

(41) $$\rho(r)=2\pi\int_0^\infty J_0(rT)f(T)T\,\delta T$$

が成り立つ。

故に $\rho(r)$ および $f(T)$ の適当な形を求めることができる。

このことから、式

$$\int_0^\infty J_0(yz)e^{-ay}\delta y=1/(a^2+z^2)^{1/2},\quad a>0$$

が有用である。何となれば、我々が式を選ぶ場合、収斂性の問題とともに $\rho(r)$ の確率的性質によって、制約を受けるのであるが、

$$\frac{\delta^n}{\delta a^n}\left(e^{-ay}/y\right)\text{と}\frac{\delta^n}{\delta a^n}(a^2+z^2)^{-1/2}\text{は}2\pi f(T)\text{と}\rho(r)\text{となる可能性をもつ函数である。故に例えば}a=n=0\text{とすればスペクトラル(spectral)および相関函数(correlation function)として、}1/2\pi T\text{および}1/r\text{が得られるがこれは収斂しない。}$$

線型(linear)な場合マルコフ過程(markoff process)$\rho(u)=e^{-au}$ のスペクトラル函数(spectral function)は、一次元の Cauchy 分布、$f(T)=1/\pi(a^2+T^2)$ である。

二次元の Cauchy 分布[7]をスペクトラル函数としてとれば

(7) 同様にして普通の Cauchy 分布は距離 a の点を中心としてすべての方向に伸ばされた直線の一つの上の密度分布(density distribution)と考えられるから、二次元の分布は、距離 a から始まる平面の一つ上の密度分布として考えることができる。

$$f(T)=a/2\pi(a^2+T^2)^{\frac{3}{2}}$$

および

$$\rho(r)=-\frac{\delta}{\delta a}(e^{-ar}/r)=e^{-ar}$$

が得られる。即ち一般化された Cauchy 分布は、一般化されたマルコフ過程のスペクトラル函数であることがわかる。勿論我々は

(42) $$\rho(u,v)=\exp-\left[\frac{u^2}{a^2}-\frac{2muv}{ab}+\frac{v^2}{b^2}\right]^{1/2}$$

で与えられる。[8]"楕円的"マルコフ過程("elliptical" Markoff process)を考えることもできる。

しかし $m=0$ とすれば計算を簡略化することができるから、これからは a_1, a_2 の測定単位を変えて過程 $\rho(r)=e^{-ar}$ を研究することにする。

この過程は農業の圃場研究[9]で経験される相関函数とそう大きくかけ離れている様には思えない。Osborne [4] は $\rho_w=e^{-\lambda w}$ の可能な用法を指摘している。Mahalanobis [5] は 800 区劃の水田に対する相関を計算した。その値を第4表に函数 e^{-r} の値とともに示しておく。

緯度及び北東方向の軸をもつ elliptical process が確かにあてはまりの良いことはわかるが Mahalanobis の値は、それぞれ標準誤差が約 0.035 であることに注意すれば、あてはまりは十分良いことがわかる。

(10) 系統的抽出と層化無作為抽出の相対的効率

(The efficiencies of systematic and stratified random sampling)

戦後の節で概念的に展開した相関函数は (19)-(27) にお い

(8) これらの点から、$\rho(r)=e^{-ar}$ は circular マルコフ過程とよばれるが、一方 $\rho_{uv}=\rho_1^{|u|}\rho_2^{|v|}$ および $\rho_{uv}=\exp\{-|\frac{u}{v}+\frac{v}{z}|\}$ は次数1およびこの退化した(degenerate)マルコフ過程として知られている。

(9) このことは、広い範囲の値にわたって Fairfield-Smith の法則と事実上一致する法則が数値的に得られることによって支持される。

第4表　観測された系列相関とcentrally-symmetric相関函数からもとめた理論値の比較

距離 マイル	行		列		北－東		南－東	
	観測値	計算値	観測値	計算値	観測値	計算値	観測値	計算値
1	0.332	0.368	0.310	0.368	—	—	—	—
$\sqrt{2}$	—	—	—	—	0.264	0.243	0.264	0.243
2	0.129	0.135	0.090	0.135	—	—	—	—
$2\sqrt{2}$	—	—	—	—	0.050	0.059	0.129	0.059
3	0.009	0.050	0.029	0.050	—	—	—	—
$3\sqrt{2}$	—	—	—	—	0.050	0.018	0.070	0.018
4	0.034	0.018	0.041	0.018	—	—	—	—
$4\sqrt{2}$	—	—	—	—	0.020	0.004	0.060	0.004

て用いるべきであるが、しかしこれらの函数は簡単には積分できない。

$$(43)\quad \begin{cases} \dfrac{\sigma^2(S_{to}S_{to})-\sigma^2(S_{yo}S_{yo})}{\sigma^2(h\ h_o)} \sim \dfrac{1}{d_1}\int_{-d_1}^{d_1}\left(1-\dfrac{|w|}{d_1}\right)F(w,d_2)\delta v \\ \qquad +\dfrac{1}{d_2}\left(1-\dfrac{|v|}{d_2}\right)F(v,d_1)\delta u \end{cases}$$

ここで、$F(u,d_2)=\frac{2}{d^2}\left[\int_0^{d_2}\frac{v}{d_2}\rho_{uv}\delta v+\int_{d_2}^{\infty}\rho_{uv}\delta v-d_2\sum_{v=1}^{\infty}\rho_{u,d_2v}\right]$

$$F(v,d_1)=\frac{2}{d_1}\left[\int_0^{d_1}\frac{u}{d_1}\rho_{uv}\delta u+\int_{d_1}^{\infty}\rho_{uv}\delta u-d_1\sum_{u=1}^{\infty}\rho_{d_1u,v}\right]$$

なることに注意すればもう一つの研究方法が可能である。$F(u,d_2)$および$F(v,d_1)$は第2章の$(\sigma_{st}^2-\sigma_{sy}^2)/\sigma_r^2$について得られた式の拡張である。故にもし$F(u,d_2)$および$F(v,d_1)$がともに正の函数なら系統的抽出は層化無作為抽出よりも正確である。これの特別の場合は$\rho_{uv}=\rho_1^u\rho_2^v$なるときに生ずる。しかし$\rho_{uv}=\exp\{-(u^2+v^2)^{1/2}\}$なるとき、$F(u,d_2)$は必ずしも正でない。何故なら$u$が増加するとき$\rho_{uv}$はvについて凹(Convex)となる。これによつて、uが0からd_1まで動くとき、$F(u,d_2)$は$+\infty$からある未知の値Xまで動くことから、(43)の解釈が非常に面倒になる。この値は$d_2\gg d_1$なるとき正で、$d_1\gg d_2$なるときは抽出が二つの方向で比例しないから系統的抽出は層化無作為抽出より有効である。その上$d_1=d_2=d_0$で$d\to 0$ならF(u, d)→∞で、この場合も系統的抽出の効率の高いことは明らかである。この種の系統的抽出即ちSyo Syoは非常に多くの場合無作為抽出より正確な結果を与える。

4. 抽出誤差の推定（Estimation of sampling errors）

公式の(7)－(18)を見れば二次元抽出（Plane sampling）においても一次元抽出の抽出誤差（linear error）の推定に用いられた原理が使えることがわかる。いま各標本が、独立な単位に分割され、その各々がS個の層の何れか1つに位置を占めることができるとすると、g個の反復に対しては誤差の自由度はg(s−1)となる。例えば、hohoおよびSthoの自由度はそれぞれ$g(n_1n_2-1)$, $g(n_2-1)$, $g(n_1n_2-n_2)$, および$g(n_2-1)$となるから、単一の標本は誤差の不偏推定量を含む。しかしStoSto, StSt, StSt, SyoSyoおよびSySyの自由度は$n_1n_2(g-1)$, $n_2(g-1)$, $g-1$および$g-1$だから、適切な誤差の推定量を求めるためには反復が必要である。しかし我々は、標本を十分正確な誤差の推定量を与えるような幾つかの部分に分割するという方法を用いることができる。

更に系統的標本の組を用いることの可能性も考えることができる。これらは抽出誤差を推定するため単純に配置される。それに標本の次々の項の間の相関が小さくない限り、低次のfを無視するとはっきりした偏りが生ずることがわかる。しかしYatesが指摘した様に、この方法によれば我々の抽出誤差の上限を求めることができるのである。

これらの抽出法は例によつて下に示される。

5. 例（Example）

我々は系統的標本の抽出誤差を推定する三つの方法を考えることにする。

(1) 各々無作為に配置された系統的標本の組を用いる方法即ち、抽出対象を小地域（sub-area）即ちブロック（block）の系列に分割し、各ブロックにおいて幾つかの系統的標本を抽出する誤差の分散は各ブロックの系統的標本の分散から計算される。

(2) 無作為に配置された一組の系統的標本を用いる方法即ち、幾つかの系統的標本をとり、この地域を小地域即ちブロックに分割する。誤差分散は各ブロック内で系統的標本が占める割合の分散から計算される。

(3) 1つの系統的標本を用いる方法、即ち1つの系統的標本をとって、それをより広い間隔の幾つかの系統的標本に分解する。例えばもとの間隔の4倍の標本を4つとると、この地域は4つかのsub-areaに分割されるから、誤差の分散は各ブロック内の副次系統的標本の割合の分散から計算される。

これらの三つの方法は、平均値の推定は益々正確に、また抽出分散の推定は次第に偏りが大きくなり、また実際的な適用は順次容易となるから、段々の抽出方法は母集団および結果の使用目的に従って変ってくる。例えば、抽出を次々に行えば、誤差の推定は改善されてくると思われるからはじめには大体の見当だけで良いかも知れない。

α) もし段々が系列を分割したときそれぞれの部分内の観測値の個数が非常に大きい様な連続な一次元母集団（Continuous linear population）について抽出を行なうなら、(1)と(2)の方法はともに項（term）当り分散

$$\sigma^2\left(1-\frac{2}{\alpha}\int_0^\infty \rho_u\,g_u+2\sum_{u=1}^{\infty}\rho_{\alpha u}\right)$$

の正確な推定量を与えるであろう。

しかし(3)は正しい項当りの分散

$$\sigma^2\left(1-\frac{2\beta}{\alpha}\int_0^\infty \rho_u\,g_u+2\sum_{u=1}^{\infty}\rho_{\alpha u/\beta}\right)$$

の代りにσ^2を推定する。

従って(3)による抽出分散の推定量は、実際の分散が小さい場合でも、(1)，(2)の方法による推定量よりも一般に過大である。

β) Kendall は480項からなる人為的な系列、

$u_{n+2}=1.1u_{n+1}-0.5u_n+\varepsilon_{n+2}$ を作った。ここでε_nは-49から49までの矩形分布である。この系列では$\sigma^2=2379.81$および$S^2=2535.11$である。この系列を80項の6つの部分にわけた。これは何れも$n=5$，$\alpha=16$，$\beta=4$であるから誤差に対しては自由度18が利用できる。これのsampling の結果を第5表に示す。

第5表　自己回帰模型に対する系統的標本

抽出誤差の三つの推定方法の比較

方法	自由度18にもとずく項当りの抽出分散の推定値S^2	$E(S^2)$	項当りの真の抽出分散
(1)	3228	2170	2170
(2)	1872	2170	2167
(3)	3709	2577	423

この表の値は連続母集団の大標本に対する結論を保証するものである。

γ) 多数の一様性の実験（uniformity trials）が行われ、システム（System）St, St', Sy, Sy' に対応して標本抽出が行われた。システム St, St' に対応する抽出について、隣当り二つの標本をとって誤差を推定した。一方システム Sy, Sy' に対応する抽出では、(2')および(3)の方法による系列の各部分においてとられた4つの標本の組を比較して誤差を推定した。この抽出の結果を第6表に示す。一方試行の数が小さければ、結果に現れた差の傾向は上に得られた結論と非常によく一致する。

δ) 母集団における trend（Trend in the population）

母集団から標本を抽出するとき我々は屢々 trend の問題に当面する。これは無作為抽出や層化無作為抽出による母集団平均値の推定の場合にはそう大きな影響を与えない。しかし系統的標本の場合これによって非常に大きな影響を受ける。いま一次元の抽出（linear sampling）を考え、x_iが順番の要素となっている標

本をS_iとかけば、標本S_iの組は普通iについて単調であつて、S_1とS_kの差は大となる（大体$x_i - x_k$に等しい）

yates〔1〕はこの難点を克服する方法を提案したS_iを

$$\frac{1}{n-1}\left\{\frac{i}{k}x_i + x_{i+k} + \cdots\cdots + x_{i+(n-2)k} + \frac{k-i}{k}x_{n+(n-1)k}\right.$$

と表せば，trandにもとずく系統的標本間の差は大部分除かれる。この方法を用いると情報を少し失なうことがすぐわかる、特に無作為な連続的母集団（continuous random population）については、分散はσ^2/nでなく$(n-\frac{3}{2})\sigma^2/(n-1)^2$である。二次元の抽出（plane sampling）における対応する補正された標本は

$$S_{ij} = \frac{1}{(n_1-1)(n_2-1)}\left[\frac{ij}{k_1k_2} + \frac{j}{k_2}x_{i+k_1,j} + \cdots + \frac{j(k_1-i)}{k_1k_2}x_{i+(n_1-1)k_1,j}\right.$$

$$+ \frac{i}{k_1}x_{i,j+k_2} + x_{i+k_1,j+k_2} + \cdots + \frac{(k_1-i)}{k_1}x_{i+(n_1-1)k_1,j+k_2}$$

$$\left.+ \frac{i(k_2-j)}{k_1k_2}x_{i,j+(n_2-1)k_2} + \cdots + \frac{(k_1-i)(k_2-j)}{k_1k_2}x_{i+(n_1-1)k_1,j+(n_2-1)k_2}\right]$$

これにも同様な損失がある。

しかしtrandがはっきりした形をとるのは大標本の場合である。この場合においては，端末補正（end adjuoment）による情報の損失は無視できるから上に得られた結論はそのまゝ成立する

第6表 三つの斉一性実験における種々の抽出法の効率の比較

出所	Kalenkar (8)			Wiebe (9)			Uyane Sayes及びKarishna. Iyer (10)		
Cochran〔11〕の分類における番号	72			132			108		
作物の種類	ジャガイモ			小麦			サトウキビ		
プロット数	576			1440			960		
平均値	23.262			587.95			270.89		
項当り平均値	15.555			10.018.0 ★			1794.42		
抽出形式	St, St1	Sy, Sy1	Sy, Sy1	St, St1	Sy, Sy1	Sy, Sy1	St, St1	Sy, Sy1	Sy, Sy1
抽出比	1/6	1/6	1/6	1/9	1/9	1/9	1/8	1/8	1/8
誤差の推定法		(2)	(3)		(2)	(3)		(2)	(3)
余例の数	1	4	4	1	4	4	1	5	5
n_1	3	3	6	4	2	4	4	2	4
k_1	2	2	1	3	6	3	2	4	2
n_2	16	2	4	20	5	10	15	3	6
k_2	6	12	6	6	6	3	8	8	4
q	2	4	1	2	4	1	2	4	1
平均値	23.140	23.435	23.323	586.54	598.65	275.29	275.29	266.72	271.27
項当りの分散推定値	9.763	2.689	4.889	5151.6	5772.7	7038.5	1320.15	799.29	1269.54
分散推定値の自由度	48	12	12	80	12	12	60	15	15

★ もとの1500個のプロットから求めたもの

引　用　文　献

1. W. G Cochran, "The relative accuracy of systematic and stratified random samples for a certain class of populations", annals of Math. Stat., Vol. 17 (1946) P. 164
2. F. Yates, "A review of recent statistical developments in sampling surveys", Roy. stat. Soc. Jour., Vol. 109 (1946) P. 12
3. G. Dedebant and P. Uchrte "Mecanique aleatoir", Portugaliae physics, Vol 1 (1945)
4. J. G. Osborne, "Sampling errors of systematic and random surveys of cover-type areas", Am. Stat. Asso. Jour., Vol. 37 (1942), P. 256.
5. P.C. Mahalanobis, "On large-scale sample surveys," Roy. Soc. Phil. Trans., B. 231 (1944), P. 329
6. M. G. Kendall, "On the analysis of oscillatory time-series," Roy. Stat. Soc Jour., Vol. 108 (1945), P. 93
7. M. G. Kendall, Contributions to the Study of Oscillatory Time-Series, Nat. Inst. Econ. Soc. Res, 1946.
8. R. J. Kalamkar, "Experimental errors and the field-plot technique with potatoes," Jour, Agr. Sci., 1932, P. 373.
9. G. A. Wiebe, "Variation and correlation in Grain-yield among 1500 wheat nursery plots," Jour. Agr. Res., 1935.
10. Wynne Sayer and P. V. Krishna Iyer, "On some of the factors that influence the error of field experiment with special reference to Sugarcane" Ind. Jour. Agr. Sci., 1936. P 917.
11. W. G. Cochran, "Catalogue of uniformity trial data", Roy. Stat. Soc. Suppl. Jour., Vol. 4 (1937), P. 233

12. F. Yates, "Systematic sampling" Roy. Soc Phil. Trans., Vol. 241 (1948), P. 345.

By. M. H. Quenouille

A. M. S. Vol. 20 (1949) I 1)

6、二次元の系統的抽出およびこれと関連する層化および無作為抽出について

Two dimentional systematic sampling and the associated stratified random sampling

緒　言

今日系統的抽出法は理論および応用統計の両面において非常に興味をもたれている。この抽出方式についてはその色々な修正方式とともによく知られている。Madow（1944）、Cochran（1946）およびYates（1948）は、一次元の場合の幾つかの特別なタイプの系統的抽出を論じた。この論文では一次元の場合のCochranの結果を二次元の場合に一般化することを考える。

抽出すべき母集団（univers）がそれぞれれ個の小区劃（cell）からなるmℓ個の行（row）から構成されていて、全部でmn個の層をなしていると仮定する*。この小区劃は場合に応じて正方形でもよくまた短形でもよい。座標を次のように定める。すなわち $y=Const$ なら行、また $x=Const$ なら列を表わす。このようにすると、れ個の小区劃のn個の組である各行はx方向で、ℓ個の小区劃のm組からなる各列はy方向にある。特定の層内において、ℓれ個の小区劃中からv個を無作為に選ぶ。そうして選ばれたこれらの小区劃が含まれる行から出発して、ℓ番目ごとの小区劃の行（y方向）を選ぶ。これらの選ばれた行の各々において無作為に抽出されたものと同じ列の小区劃とこれから℄番目ごとの小区劃（x方向に沿って）をとれば、大きさmnvなる標本ができる。これは二次元の場合の系統的抽出の一種である。このような系統的抽出はまた次のようにしても定義できる。すなわち、特定の層においてℓれ個のなかからv個の小区劃を無作為に選ぶ。次に組内および層に対する配置が最初の層と同じになるようにして他の各層から同様なv個の小区劃の組を選ぶ。我々はまた全部で大きさmnvなる層化標本となるようにそれぞれの層内から大きさvの無作為標本を抽出することもできる。これとは別に全部でmnℓれ個の小区劃の中から層化なしでmnv個の区劃を無作為に選ぶこともできる。これは単純無作為標本（perfectly random sample）である。

我々はこのユニバースが $E(z_i)=\mu$ なる意味で等一であり、また観測値は共通の分散と異なる空間的相関（different space correlation）を有するものとする。ただしzは小区劃の観測値である。xおよびy方向にそれぞれu およびvだけ離れた点の間の空間的相関係数を $\rho(u,v)$ とする。すなわち小区劃 (x,y) を P とすればその収量と区劃 $(x+u,y+v)$ の収量と組合せ、それらの可能なすべての組から $\rho(u,v)$ を計算するのである。

これはユニバースの有限性のためにある変動をもつかもしれない。もしこのユニバースが無限ならそのような影響は存在しない。しかしこのユニバースが有限でも、Cochran（1946）がしたようにそれを無限母集団から得られた標本と考えることができる。このとき無限母集団では、$E(z)=\mu$、$v(z)=\sigma^2$ および

$$E\{(z_{i+u,j+v}-\mu)(z_{ij}-\mu)\}=\rho(u,v)$$

である。色々なuおよびvの値に対するこの $\rho(u,v)$ は三次元の図形となるから、これはCorrelogripeとよぶことができる。この図形は区劃が無限小のときに連続となる。

この論文においては、無限母集団からとられたすべての有限母集団について平均したときの無作為、層化および系統的標本の総平均の分散の期待値をCochran（1946）の用いた記号にしたがってそれぞれ σ_n^2、σ_{st}^2、σ_{sy}^2 で表わす。

Δ_1、Δ_2 S_1、S_2 は

$$\Delta_1\rho(u,v)=\rho(u+1,v)-\rho(u,v)$$

$$\Delta_2\rho(u,v)=\rho(u,v+1)-\rho(u,v)$$

$$S_1\rho(u,v)=\rho(u+\tfrac{1}{2},v)-\rho(u-\tfrac{1}{2},v)$$

$$S_1^2\rho(u,v)=\rho(u+1,v)+\rho(u-1,v)-2\rho(u,v)$$

$$\Delta_1\Delta_2\rho(u,v)\ \Delta_1\rho(u,v+1)-\Delta_1\rho(u,v)$$

なる演算子（operator）で S_2、S_2^2 および S_1^2、S_2^2 の定義も同様である。我々はまた $E_1\rho(u,v)=\rho(u+1,v)$ および

$E_2\rho(u,v)=\rho(u,v+1)$ なる E_1, E_2 をも用いる。

これらの E_i は"期待値"を表わす E, (添字のない) とは異なる。

また $\rho(u,v)=\rho(u,-v)$ と $\rho(u,-v)=\rho(-u,v)$ は別のものであることに注意しなければならない。そうして $\rho(u,v)+\rho(-u,v)$ を $\psi(u,v)$ と書くことにする。これは u および v に関して対称である。$\Delta_1\psi$ および $\Delta_2\psi$ は特に $\Delta_1\psi(u,v)$ および $\Delta_2\psi(u,v)$ を表わす。

§1においては $\sigma_{\bar{x}}^2$, $\sigma_{s\bar{z}}^2$, σ_{sy}^2 を計算し、それらの相対的割合が γ と独立であり、したがって相対的効率 (relative efficiency) も γ と独立であることを示した。§2では層化抽出が無作為抽出より効率が良くなるための幾つかの条件を多数の定理の形で述べ、また§3では系統的抽出が層化抽出よりも有効となるための十分条件を定理4として与えた。したがって、直接二つの比較はしなかったが、これらの結果を用いて系統的抽出が無作為抽出よりも有効となるための十分条件の組を見出すことができる。Cochran の結果の特別な場合として出てくる。しかしこれらは ρ が正また負の何れであっても、また1より大きい γ に対しても成立するという点で Cochran (1946) の結果よりも多少改善されている。

§1

$$\sum_{j=1}^{n}(z_j-\bar{z})^2=\frac{1}{n}\sum_{j=1}^{n}\sum_{i<j=1}^{n}(z_i-z_j)^2$$

が成立する。

Lemma 1　(i,j) 区劃に対応する確率変数として z_{ij} をもつような区劃の N 個からなっている M 個の行を含むプロットにおいて

$$E\left\{\sum_{i=1}^{M}\sum_{j=1}^{N}(z_{ij}-\bar{z})^2\right\}=\frac{2\sigma^2}{MN}\left[\frac{MN(MN-1)}{2}-N\sum_{v=1}^{M}(M-v)\rho(0,v)-M\sum_{u=1}^{n-1}(n-u)\rho(u,0)\right.$$
$$\left.-\sum_{u=1}^{N-1}\sum_{v=1}^{M-1}(N-u)(M-v)\psi(u,v)\right]$$

ただし $i=1,\cdots\cdots,N$; $j=1,\cdots\cdots,M$.

証明

(1.1) によって　$\sum_{i=1}^{N}\sum_{j=1}^{M}(z_{ij}-\bar{z})^2=\frac{1}{MN}\sum(z_t-z_{t'})^2$

ここで和は、緒言で述べたタイプの無限個のプロットからとられた標本であるところの有限個のプロット内の z_t と $z_{t'}$ のすべての組合せの上にわたるものである。全部で ${}^{MN}C_2$ 個の組合せがあるが、これらは

(i)　各行の中で ${}^{N}C_2$ 通りの組合せが全部で $M\cdot{}^{N}C_2$ 通り

(ii)　各列の中で ${}^{M}C_2$ 通りの組合せが全部で $N\cdot{}^{M}C_2$ 通り

(iii)　(x,y) と $(x+u,y+v)$, $(u\neq 0, v\neq 0)$ 間で $2\,{}^{M}C_2\,{}^{N}C_2$ 通り

のように分けられる。

いま

$$E\left\{\sum_{ij}(z_{ij}-\bar{z})^2\right\}=\frac{1}{MN}E\left\{\sum(z_t-u+u-z_{t'})^2\right\}$$
$$=\frac{2\sigma^2}{MN}\sum\left\{1-\rho(z_t,z_{t'})\right\}\quad\cdots\cdots(1.2)$$

ここで $\rho(z_t,z_{t'})$ は z_t と $z_{t'}$ 間の相関係数で、和はもとのすべての組合せ (すなわち全部で ${}^{MN}C_2$ 通り) にわたる。まず $\sum\rho(z_t,z_{t'})$ を求めてみよう。これは上に述べた3つの場合 (i)(ii)(iii) に対応する3つの式の和である。(i) の場合には、各行で $(u,0)$ の距離のものは $(N-u)$ 区劃ある。したがってすべての行についてのこのタイプの $\sum\rho(z_t,z_{t'})$ は $N\sum_{u=1}^{N-1}(N-u)\rho(u,0)$ である。同様に (ii) の場合の $\sum\rho(z_t,z_{t'})$ は $N\sum_{v=1}^{M-1}(M-v)\rho(0,v)$ である。(iii) の場合は4点 (x,y), $(x+u,y)$, $(x,y+v)$ および $(x+u,y+v)$ $(u\neq 0, v\neq 0)$ を考えれば、明らかにこれらの4点の中で (iii) の場合の異なる距離のタイプは2つのみ、すなわち (u,v) および $(-u,v)$ だけである。距離 $(-u,-v)$ と (u,v) は同じである。何となれば空間的相関はこの双方の距離に対して同一だからである。同様に距離 $(u-v)$ は $(-u,v)$ と同等である。よってこのような4点対は全部で $(N-u)(M-v)$ 組あるからこのタイプの全体について $\sum\rho(z_t,z_{t'})$ を加え上げれば、$\sum_{u=1}^{N-1}\sum_{v=1}^{M-1}(N-u)(M-v)\psi(u,v)$ を得る。

注意すべき点は

$\rho(u,0)=\rho(-u,0)$ および $\rho(0,v)=\rho(0,-v)$

である。

このときもすべての組合せ上での

$\sum 1$ は ${}_{MN}C_2$ である。

よって(1.2)から Lemma がえられる。

これは次のように書ける。

$$E\left\{\sum_{i=1}^{N}\sum_{j=1}^{M}(z_{ij}-\bar{z})^2\right\}=(MN-1)\sigma^2\{1-\phi(M,N)\}$$

$$=(MN-1)\sigma^2 L(M,N)\cdots\cdots\cdots\cdots(1.3)$$

ただし

$$\phi(M,N)=\frac{2}{M(MN-1)}\sum_{v=1}^{M-1}(M-v)\rho(0,v)+\frac{2}{N(MN-1)}\sum_{u=1}^{N-1}\rho(u,0)$$

$$+\frac{2}{MN(MN-1)}\sum_{u=1}^{N-1}\sum_{v=1}^{M-1}(N-u)(M-v)\psi(u,v)\cdots(1.4)$$

および

$$L(M,N)=1-\phi(M,N)\qquad(1.5)$$

ゆえに(1.3)で $M=ml$ および $N=nk$ とおけば有限母集団の全体について

$$E\left\{\sum(z-\bar{z})^2\right\}=(mnlk-1)\sigma^2\{1-\phi(ml,nk)\}\cdots(1.6)$$

をうる。

同様に各層に対して(1.3) $M=l$ および $N=k$ とおけば

$$E\left\{\sum(z-\bar{z})^2\right\}=(lk-1)\sigma^2\{1-\phi(l,k)\}\cdots\cdots(1.7)$$

したがって mn 個の層に対する

$$E(\text{層内の}S.S.)=mn(lk-1)\sigma^2\{1-\phi(l,k)\}\cdots\cdots\cdots(1.8)$$

$V=1$ なる各系統的標本については、(1.3) および (1.4) で $M=m$, $N=n$ とおき、$\rho(u,v)$ を $\rho(uk,vl)$ でまた ψ を ψ_{sy} でおきかえると

$$E\left\{\sum(z-\bar{z})^2\right\}=\sigma^2\{1-\phi_{sy}(m,n)\}$$

である。

ゆえに $V=1$ なるとき

$$E(\text{系統的標本内}S.S.)=lk(mn-1)\sigma^2\{1-\phi_{sy}(m,n)\}\qquad(1.9)$$

いま大きさ N の有限母集団からとられた大きさ n' の無作為標本については

標本平均値の分散 $=\frac{1}{n'}\cdot\frac{N-n}{N-1}\left\{\frac{1}{N}\sum(z-\bar{z})^2\right\}$

である。ただし和は全有限母集団の上ごとる。

よってこの場合 $N=mnlk$ とおけば $n'=mnv$ であるから、(1.6)を用いて

標本平均値の分散の期待値は

$$\sigma^2_{\bar{x}}=\frac{1}{mnv}\cdot\frac{mn(lk-v)}{(mnlk-1)}\cdot\frac{1}{mnlk}(mnlk-1)\sigma^2\{1-\phi(ml,nk)\}$$

$$=\frac{\sigma^2}{mnv}\left(1-\frac{v}{lk}\right)\{1-\phi(ml,nk)\}\cdots(1.10)$$

となる。

更にそれぞれ lk 個の小区劃からなる層が mn 個あるから

$x_{i1},x_{i2},\cdots\cdots,x_{iv}$ を i 層からとった大きさ v なる標本とする。

ここで x_{ij} は確率変数で、繒言で述べた座標を表わすところの添字をつけないことと混同してはならない。

そうすると

$$\bar{x}=\frac{1}{mnv}\sum_{i=1}^{mn}\sum_{j=1}^{v}x_{ij}=\frac{1}{mn}\sum_{i=1}^{mn}\bar{x}_i,\quad\text{ただし}\ \bar{x}_i\equiv\frac{1}{v}\sum_{j=1}^{v}x_{ij}$$

$$V(\bar{x})=\frac{1}{m^2n^2}\sum_{i=1}^{mn}V(\bar{x}_i)$$

しかるに(1.7)から

$$E\{V(\bar{x}_i)\}=\frac{1}{v}\cdot\frac{(lk-v)}{(lk-1)}\cdot\frac{(lk-1)}{lk}\sigma^2\{1-\phi(l,k)\}$$

よって

$$\sigma^2_{s\bar{x}}=E\{V(\bar{x})\}=\frac{\sigma^2}{mnv}\left(1-\frac{v}{lk}\right)\{1-\phi(l,k)\}\cdots\cdots(1.11)$$

いま系統的標本の分散の期待値 σ^2_{sy} を求めるため $x_{11},x_{12}\cdots\cdots,x_{1v}$ を 1 層からとった標本とする。標本総平均は $\bar{x}_{sy}=\frac{1}{mnv}\sum_{i=1}^{mn}\sum_{j=1}^{v}x_{ij}$

で、$\sigma_{sy}^2 = E(\bar{x}_{sy} - \bar{\bar{x}})^2$ である。ただし $\bar{\bar{x}}$ は相関母集団の総平均である。

ここで $\sum_{j=1}^{mn} x_{ij}/mn$ を ${}_j\bar{x}_{sy}$ と書くことにすると

$$(\bar{x}_{sy} - \bar{\bar{x}}) = \frac{1}{v}\sum_{j=1}^{v}({}_j\bar{x}_{sy} - \bar{\bar{x}})$$

よって

$$(\bar{x}_{sy} - \bar{\bar{x}})^2 \frac{1}{v^2}\left\{\sum_{j=1}^{v}({}_j\bar{x}_{sy} - \bar{\bar{x}})^2 + \sum_{\substack{j,j'=1 \\ (j \neq j')}}^{v}({}_j\bar{x}_{sy} - \bar{\bar{x}})({}_{j'}\bar{x}_{sy} - \bar{\bar{x}})\right\} \cdots\cdots (1.12)$$

$l_k C_v$ 個の系統的標本全体について (1.12) を平均する。

もし x_{ij} が第 i 層からとられた照作為点（観測可能な）なら、${}_j\bar{x}_{sy}$ は対応する系統的標本の平均値である。ゆえに l_k こ2個の系統的標本全体での (1.12) の和において、${}_j\bar{x}_{sy}$ ($j = 1, 2, \cdots, mn$) は ${}_{l_k-1}C_{v-1}$ 回現れる。というのは大きさ v の標本をとるためには残りの $(l_k - 1)$ 個の中から $(v - 1)$ 個だけとり出せばよいからである。同様に ${}_j\bar{x}_{sy}$ および ${}_{j'}\bar{x}_{sy}$ $(j \neq j')$ は同時に ${}_{l_k-2}C_{v-2}$ 個現れる。よって ${}_{l_k}C_v$ 個の標本全部について加え上げると $({}_j\bar{x}_{sy} - \bar{\bar{x}})$, $(j = 1, 2, \cdots\cdots, l_k)$ は ${}_{l_k-1}C_{v-1}$ 個、また $({}_j\bar{x}_{sy} - \bar{\bar{x}})({}_{j'}\bar{x}_{sy} - \bar{\bar{x}})$, $j \neq j'$ は ${}_{l_k-2}C_{v-2}$ 個でてくる。しかるに $\sum_{t=1}^{l_k}({}_t\bar{x}_{sy} - \bar{\bar{x}}) = 0$

よって $\sum_{t=1}^{l_k}({}_j\bar{x}_{sy} - \bar{\bar{x}})({}_t\bar{x}_{sy} - \bar{\bar{x}}) = 0$ かつ

$$\sum_{j=1}^{l_k}\sum_{\substack{t=1 \\ t \neq j}}^{l_k}({}_j\bar{x}_{sy} - \bar{\bar{x}})({}_t\bar{x}_{sy} - \bar{\bar{x}}) = -\sum_{j=1}^{l_k}({}_j\bar{x}_{sy} - \bar{\bar{x}})^2$$

故に和においては $({}_j\bar{x}_{sy} - \bar{\bar{x}})^2$ は $({}_{l_k-1}C_{v-1} - {}_{l_k-2}C_{v-1}) = {}_{l_k-2}C_{v-1}$ 個でてくる。したがって求める (1.12) の平均は

$$\frac{{}_{l_k-2}C_{v-1}}{{}_{l_k}C_v}\frac{1}{v^2}\sum_{j=1}^{l_k}({}_j\bar{x}_{sy} - \bar{\bar{x}}) = \frac{1}{mnv}\frac{(l_k - v)}{l_k(l_k - 1)}\sum_{j=1}^{l_k}mn({}_j\bar{x}_{sy} - \bar{\bar{x}})^2$$

$$= \frac{1}{mnv}\frac{(l_k - v)}{l_k(l_k - 1)}\left\{\sum_{i=1}^{mn}\sum_{j=1}^{l_k}(x_{ij} - \bar{\bar{x}})^2 (i=1 \text{ なるときの系統的標本以外}.)\right\} \cdots (1.13)$$

ゆえに期待値をとると

$$\sigma_{sy}^2 = \frac{\sigma^2}{mnv}\frac{(l_k - v)}{l_k(l_k - 1)}\left\{(mnl_k - 1) - (mnl_k - 1)\phi(ml, nk) - l_k(mn - 1) + l_k(mn - 1)\phi_{sy}(m, n)\right\}$$

(1.6) および (1.9) から

$$= \frac{\sigma^2}{mnv}\left(1 - \frac{v}{l_k}\right)\left\{1 - \frac{(mnl_k - 1)}{(l_k - 1)}\phi(ml, nk) + \frac{l_k(mn - 1)}{(l_k - 1)}\phi_{sy}(m, n)\right\} \cdots (1.14)$$

また (1.10) と (1.11) から

$$\sigma_r^2 = \frac{\sigma^2}{mnv}\left(1 - \frac{v}{l_k}\right)\left\{1 - \phi(ml, nk)\right\} \cdots\cdots (1.15)$$

$$\sigma_{st}^2 = \frac{\sigma^2}{mnv}\left(1 - \frac{v}{l_k}\right)\left\{1 - \phi(l_k)\right\} \cdots\cdots (1.16)$$

ただし $\phi(M, N)$ は (1.4) で与えたものでありまた

$$\phi_{sy}(m, n) = \frac{2}{m(mn - 1)}\sum_{u=1}^{m-1}(m - u)\rho(u, ul) + \frac{2}{n(mn - 1)}\sum_{u=1}^{n-1}\rho(uk, 0) + \frac{2}{mn(mn - 1)}\sum_{u=1}^{n-1}\sum_{v=1}^{m-1}(n - u)(m - v)\psi(uk, vl)$$

である

ゆえに相対的効率

$$\frac{1}{\sigma_r^2} : \frac{1}{\sigma_{st}^2} : \frac{1}{\sigma_{sy}^2} \quad \text{は } v \text{ と独立で層化および correlopifad}$$

の程度によってきまる。もし後者が既知なら前者はその目的によって調整することができる。

さて

ここで層化抽出が純正無作為よりも有効となる、すなわち $\sigma_r^2 \geq \sigma_{st}^2$ となるための十分条件の組を求める。まず基礎な2つの Lemma を証明しておく。

Lemma II　条件

(i) $\Delta_1\psi(u, v) \leq 0$

(ii) $\Delta_2\psi(u, v) \leq 0$

(iii) $2\psi(u+1, v) \geq \psi(u, v+1) + \psi(u+1, v+1)$

が満足されるとき、lおよびkのすべての値に対して

$$L(l+1,k) \gg L(l,k)$$

である。

証　明

(1.4)および(1.5)のMをlで、Nをkでおきかえれば$L(l,k)$がえられる。

そこで $l'=l+1$, $A=lk-1$, $A'=l'k-1$ および

$$\beta_{0v}=2\left(\frac{l-v}{lA}-\frac{l'v}{l'A'}\right),\quad \alpha_{u0}=\frac{k-u}{AA'},\quad \alpha_{uv}=\frac{(k-u)}{k}2\left(\frac{l-v}{lA}-\frac{l'-v}{l'A'}\right) \tag{2.1}$$

ただし $u\neq 0$ および $v\neq 0$, とおくと

$$L(l+1,k)-L(l,k)=\sum_{v=1}^{l}\beta_{0v}P(0,v)+\sum_{u=1}^{k}\sum_{v=0}^{l}\alpha_{uv}\psi(u,v) \cdots (2.2)$$

かえられる。

いま $\psi(u,v)=-\Delta_2\psi(u,v)+\psi(u,v+1)$ であるから

$$\sum_{v=0}^{l}\alpha_{uv}\psi(u,v)=\sum_{v=0}^{l-1}S_{uv}\Delta_2\psi(u,v)+S_{ul}\psi(u,l) \cdots\cdots (2.3)$$

ただし $S_{uv}=\sum_{j=0}^{v}\alpha_{uj}$

また

$$\sum_{v=1}^{l}\beta_{0v}P(0,v)=-\sum_{v=1}^{l-1}S'_{0v}P(0,v)+S'_{0l}P(0,l) \cdots\cdots (2.4)$$

ただし

$$S'_{0v}=\sum_{j=1}^{v}\beta_{0j}$$

よって

$$S'_{0v}=v\left\{y\left(\frac{1}{lA}-\frac{1}{l'A'}\right)+\frac{(k+1)}{l'AA'}\right.$$

ただし $y=l-1-v$

ゆえに $y\gg 0$、すなわち $v\leqslant (l-1)$ のとき

$$S'_{0v}\gg 0 \tag{2.5}$$

および

$$S'_{0l}=-\frac{(k-1)}{AA'} \tag{2.6}$$

いま (2.5) から $v\leqslant(l-1)$ なるとき

$$S_{uv}=\alpha_{u0}+\sum_{j=1}^{v}\alpha_{uj}=\frac{k-u}{AA'}+\frac{k-u}{k}S'_{0v}$$

は正であるから (2.6) を用いて

$$S_{ul}=\frac{(k-u)}{KAA'} \tag{2.7}$$

$$L(l+1,k)-L(l,k)=-\sum_{u=1}^{k}\sum_{v=0}^{l-1}\frac{(k-u)}{AA'}\Delta_2\psi(u,v)$$

$$-\sum_{u=1}^{k}\sum_{v=1}^{l-1}\frac{(k-u)}{k}S'_{0v}\Delta_2\psi(u,v)-\sum_{v=1}^{l-1}S'_{0v}\Delta_2P(0,v)+\sum_{u=1}^{k}\frac{(k-u)}{k2AA'}\psi(u,l)-\frac{(k-1)}{2AA'}\psi(0,l)\cdots \tag{2.8}$$

いま

$$\sum_{u=1}^{k}\frac{(k'-u)}{KAA'}\psi(u,l)-\frac{(k-1)}{2AA'}\psi(0,l)=\sum_{u=1}^{k}\frac{(k-u)}{k2AA'}\left\{\psi(u,l)-\psi(0,l)\right\}$$

$$=\sum_{u=1}^{k}\frac{(k-u)}{k2AA'}\sum_{i=0}^{u-1}\Delta_1\psi(i,l) \cdots\cdots (2.9)$$

$$=\sum_{u=0}^{k}\frac{(l-u)(k-u)}{2kAA'}\Delta_1\psi(u-1,l)+\sum_{u=1}^{k}\frac{(k-u)}{2AA'}\Delta_1\psi(u-1,l) \cdots (2.10)$$

$\Delta_1\psi\leqslant 0$ なるとき (2.10) の最初のところからは0に等しいかまたは0より大である。しかし他の部分は正にならないから $\Delta_2\psi\leqslant 0$ および $\Delta_1\psi\leqslant 0$ のとき (2.8) から

$$L(l+1,k)-L(l,k)=Q+\sum_{u=1}^{k}\frac{(k-u)}{2AA'}\Delta_1\psi(u-1,l)$$

$$-\sum_{u=1}^{k}\sum_{v=0}^{l-1}\frac{(k-u)}{AA'}\Delta_2\psi(u,v) \cdots (2.11)$$

ここで $Q\gg 0$ である。

また

$$\sum_{v=0}^{l-1}\Delta_2\psi(u,v)=\psi(u,l)-\psi(u,0)$$

が成立する。

よって (2.11) から

$$L(l+1,k)-L(l,k)=Q+\sum_{u=1}^{k}\frac{(k-u)}{2AA'}\left\{2\psi(u,0)-\psi(u,l)-\psi(u-1,l)\right\}\cdots$$

$$\cdots\cdots (2.12)$$

いま $\Delta_2\psi\leqslant 0$ なるとき

$$2\psi(u,0)-\psi(u,l)-\psi(u-1,l)\geqq 2\psi(u,l-1)-\psi(u,l)-\psi(u-1,l)$$

ゆえにすべてのvの値に対して

$2\psi(u+1,v)\geqq\psi(u,v+1)+\psi(u+1,v+1)$

$2\psi(u,l-1)-\psi(u-1,l)-\psi(u,l)\geqq 0$、すなわち(2.12)およびこれから(2.11)'は0または正である。

よってLemmaは証明された。

同様にして次のLemmaを証明することができる。

Lemma (III)　条件

(i) $\Delta_1\psi(u,v)\leqq 0$

(ii) $\Delta_2\psi(u,v)\leqq 0$

(iii) $2\psi(u+1,v)\geqq\psi(u+1,v)$

が満たされるとき、lおよびkのすべての値について

$$L(l,k+1)\geqq(l,k)$$

である。

定理 1

(i) $\Delta_1\psi\leqq 0$

(ii) $\Delta_2\psi\leqq 0$

(iii) $2\psi(u,v+1)\geqq\psi(u+1,v)+\psi(u+1,v+1)$

(iv) $2\psi(u+1,v)\geqq\psi(u,v+1)$

が満たされるすべての無限母集団においては、

、任意の大きさの標本について

$$\sigma_{St}^2\leqq\sigma_r^2$$

が成立する。また上の4つの場合のそれぞれについて等号が成立しなければ

$$\sigma_{St}^2<\sigma_r^2$$

である。

証　明

これらの3つの条件はLemma Iのすべての条件を満足している。

よって

$$L(l,k)\leqq L(l+1,k)\leqq\cdots\leqq L(ml,k)$$

すなわち

$$1-\varphi(l,k)\leqq 1-\varphi(ml,k)$$

この場合のkは切論層内の全E個数であるから

$$\sigma_{St}^2\leqq\sigma_r^2$$

で定理がえられる。

いまもし $\Delta_1\psi\geqq\Delta_2\psi$ なら(iii)が満足される。ゆえに更に強い条件の組を次のように与えることができる。

系1　u方向(例えば)に沿った平行なストリップで層化がおこなわれているすべての無限母集団で、もし $\Delta_2\psi\leqq\Delta_1\psi\leqq 0$ なら

$$\sigma_r^2\leqq\sigma_{St}^2$$

で、またこれらの場合の各々において等号が成立しない限り

$$\sigma_r^2>\sigma_{St}^2$$

である。

系2　上記においてk=1すなわちv一方向に沿った一次元の圃場でもし $\Delta_2\psi(u,v)\leqq 0$ または $\Delta_2\rho(0,v)\leqq 0$ なら、

u=0であるから

$$\sigma_r^2\geqq\sigma_{St}^2$$

である。この場合はCochran (1946)がk=1について論じている。これらの定理はすべて l, m, n, k の値と無関係な条件を与えており、したがってどのような層化方式が有効であるかということには考えていない。しかし着者(1949)はパラメーターの関係に依存する他の条件の組を与えた。ここではそれについて論ずることにする。この場合にもある程の層化の任意性(some degree of freedom in stratification)を得た。これは次の定理のように述べられる。

定　理 3　$\sqrt{n}\leqq k\leqq l^2$ および $(n-1)k=(m-1)l$ なる型のすべての層化方法に対して、もし

(i) $\Delta_1\psi\leqq 0$

(ii) $\Delta_2\psi\leqq 0$

(iii) $\Delta_1\Delta_2\psi \geqslant 0$

なら

$$\sigma_{St}^2 \leqslant \sigma_r^2$$

で、これらの3の場合の各々において等号が成立しない限り

$$\sigma_{St}^2 < \sigma_r^2$$

である。

このパラメトリックな関係は勿論 $k=l$, $m=n$ なら成立する。

この定理は次の Lemma N を用いて説明する。

Lemma N 条件

(i) $u \leqslant k$, $v \leqslant l$ のとき

$\sum_{i=0}^{u}\sum_{j=0}^{v} a_{ij} \geqslant 0$ ただし $\sum_{j=0}^{k}\sum_{i=0}^{l} a_{ij} = 0$

(ii) $\Delta_2\psi(k, v) \leqslant 0$

(iii) $\Delta_1\psi(u, l) \leqslant 0$

(iv) $\Delta_1\Delta_2\psi(u, v) \geqslant 0$

が満されるとき

$$\sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij}\psi(j, j) \geqslant 0$$

である。

証明

$$\sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij}\psi(i,j) = -\sum_{j=0}^{l}\sum_{u=0}^{k-1} S_{uj}\Delta_1\psi(u,j) + \sum_{j=0}^{l} S_{kj}\psi(k,j), \text{ ただし } S_{uj} = \sum_{i=0}^{u} a_{ij}$$

$$= \sum_{u=0}^{k-1}\sum_{v=0}^{l-1} a_{ij}\psi(u,v) - \sum_{u=0}^{k-1} T_{uk}\Delta_1\psi(u,l) - \sum_{v=0}^{l-1} T_{kv}\Delta_2\psi(k,v) + T_{kl}\psi(k,l)$$

ここで $T_{uv} = \sum_{j=0}^{v} S_{uj} = \sum_{i=0}^{u}\sum_{j=0}^{v} a_{ij}$

これで Lemma は証明された。

ここで定理3を証明しよう。

$$L(l+1, k+1) - L(l, k) = \sum_{v=1}^{l} a_{0v}\psi(0,v) + \sum_{u=1}^{k} a_{u0}\psi(u,0) + \sum_{u=1}^{k}\sum_{v=1}^{l} a_{uv}\psi(u,v)$$

$$= \sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij}\psi(i,j)$$

$k' = k+1$, $l' = l+1$, $A = lk-1$ および $A' = l'k-1$

$a_{00} = 0$, $a_{0j} = \frac{lj}{lA} - \frac{l(j-1)}{l'A'}$, $a_{i0} = \frac{k-i}{kA} - \frac{k'-i}{k'A'}$

また $i \neq 0$, $j \neq 0$ のとき

$$a_{ij} = \frac{2(k-i)(l-j)}{lkA} - \frac{2(k'-i)(l'-j)}{l'k'A'}$$

とおく。

そうすると

$$T_{uv} = \sum_{i=1}^{u} a_{i0} + \sum_{j=1}^{v} a_{0j} + \sum_{i=1}^{u}\sum_{j=1}^{v} a_{ij}$$

よって

$$2T_{uv} = \frac{u(k+x-1)}{kA} + \frac{v(l+y-1)}{lA} + \frac{uv(k+x-1)(l+y-1)}{lkA}$$

$$- \frac{u(k'+x)}{k'A'} - \frac{v(l'+y)}{l'A'} - \frac{uv(k'+x)(l'+y)}{l'k'A}$$

ただし $x = (k-u)$, $y = (l-v)$ である。

これから $\sqrt{2} \leqslant k \leqslant l^2$ なら $T_{uv} \geqslant 0$ および特に $T_{kl} = 0$ すなわち、$\sqrt{2} \leqslant k \leqslant l^2$ なら

$$\sum_{i=0}^{k}\sum_{j=0}^{l} a_{ij} = 0 \quad \text{および} \quad \sum_{i=0}^{u}\sum_{j=0}^{v} a_{ij} \geqslant 0$$

が証明できる。

よって Lemma N を適用すれば

(i) $\Delta_1\psi(u, l) \leqslant 0$

(ii) $\Delta_2\psi(u, l) \leqslant 0$

(iii) $\Delta_1\Delta_2\psi(u, v) \geqslant 0$

なら

$$L(l+1, k+1) - L(l, k) \geqslant 0$$

ゆえに

(i) $\Delta_1\psi \leqslant 0$

(ii) $\Delta_2\psi \leqslant 0$

(iii) $\Delta_1\Delta_2\psi(u, v) \geqslant 0$

(iv) $(m-1)l = (n-1)k$

(v) $\sqrt{2} \leqslant k \leqslant l^2$

なら

$$L(l, k) \leqslant L(l+1, k+1) \leqslant \cdots\cdots \leqslant L(ml, nk)$$

である。　　証明終り。

§ 3

我々はここで系統的抽出が層化抽出より有効となるための十分条件の組を求めることにする。

定　理　4

(i) $S_1^2 \rho(u,v) \geqslant 0$　　(ii) $S_2^2 \rho(u,v) \geqslant 0$

なるすべての無限母集団において、任意の大きさの標本に対して

$$\sigma_{sy}^2 \leqslant \sigma_{st}^2$$

が成立する。また上の2つの場合の各々において等号が成立しない限り

$$\sigma_{sy}^2 < \sigma_{st}^2$$

である。

証　　明

相対的効率 $\sigma_{sy}^2/\sigma_{st}^2$ は v と独であるから、これを1にとることができる。すなわち mn 個の層の各々から1つの標本をとる。(1.13) から $v=1$ のとき $(\bar{x}_{sy}-\bar{\bar{x}})^2$ の平均値すなわち σ_{sy}^2 の推定値は

$$\frac{1}{mn\,\ell k}\left\{\sum_{i=1}^{mn}\sum_{j=1}^{\ell k}(x_{ij}-\bar{\bar{x}})^2 = \frac{1}{mn}\times(\text{系統的標本内の平均的SS.})\right\} \cdots (3.1)$$

mn 個の層からとられた系統的標本を $x_{1j}, x_{2j}, \cdots\cdots x_{mn,j}$ とし、またその平均値を $\bar{x}_{st}$ とおく。x_{ij} は第 i 層からとられた標本である。

そうすると

$$\bar{x}_{st} = \sum_{i=1}^{mn} x_{ij}/mn .$$

第 i 層の平均値を $\bar{x}_i$、総平均を $\bar{\bar{x}}$ とおくと

$$\bar{x}_i = \frac{1}{\ell k}\sum_{j=1}^{\ell k} x_{ij}, \quad \bar{\bar{x}} = \frac{1}{mn}\sum_{i=1}^{mn}\bar{x}_i, \quad \sigma_{st}^2 = E(\bar{x}_{st}-\bar{\bar{x}})^2$$

このとき

$$\sum_{i=1}^{mn}\sum_{j=1}^{\ell k}(x_{ij}-\bar{\bar{x}})^2 = \sum_{i=1}^{mn}\sum_{j=1}^{\ell k}(x_{ij}-\bar{x}_{st})+mn(\bar{x}_{st}-\bar{\bar{x}})^2$$

$(\ell k)^{mn}$ 個のすべての標本について平均すれば

$$\frac{1}{\ell k}\sum_{i=1}^{mn}\sum_{j=1}^{\ell k}(x_{ij}-\bar{\bar{x}})^2 = (\text{標本内の平均的 SS})+mn(\bar{x}_{st}-\bar{\bar{x}})^2$$

$$\therefore \sigma_{st}^2 = E(\bar{x}_{st}-\bar{\bar{x}})^2 = \frac{1}{mn\,\ell k}E\left\{\sum_{i=1}^{mn}\sum_{j=1}^{\ell k}(x_{ij}-\bar{\bar{x}})^2 - \frac{1}{mn}(\text{標本内平均的 S.S.})\right\} \cdots\cdots\cdots (3.2)$$

標本を $x_1, x_2, \cdots\cdots, x_{mn}$、その平均値を $\bar{\bar{x}}$ とおくと標本内の S.S は

$$\sum_{i=1}^{mn}(x_i-\bar{\bar{x}})^2 = \frac{1}{mn}\sum_{\substack{ij=1\\ j>i}}^{mn}(x_i-x_j)^2 \cdots\cdots\cdots\cdots (3.3)$$

これは系統的または層化標本内の平均的 S. S である。

系統的および層化標本内の平均的 S. S の期待値をそれぞれ Σ_{sy}, Σ_{st} とすると (3.1) および (3.2) から $\Sigma_{sy} \geqslant \Sigma_{st}$ なら

$$\sigma_{st}^2 \geqslant \sigma_{sy}^2$$

である。

(3.3) から、Σ_{sy} および Σ_{st} は層の対 (i,j は層を表わす) のすべての組合せについて $E(x_i-x_j)^2$ なる項を加え上げたものに $\frac{1}{mn}$ を乗じたものである。

ゆえに Σ_{sy}, Σ_{st} は個別に比較することができる。

E_{sy} および E_{st} で系統的および層化標本に対する期待値を表わす。また x_i, x_j はそれぞれ $(i_1-j_1)=u$ および $(i_2-j_2)=v$ なるような (i_1, i_2) および (j_1, j_2) 層に属するものとする。そうすると

$$E_{sy}(x_i-x_j)^2 = E_{sy}(x_{i_1}-\mu+\mu-x_{j_1})^2 = 2\sigma^2\{1-\rho(uk, vl)\} \cdots\cdots (3.4)$$

ここで x_j は x_i がきまればきまり、逆も成立する。

$E_{st}(x_i-x_j)^2$ については、この場合 ℓ^2k^2 通りの対の組合せが可能である。したがって我々はこれらのすべての組合せについて平均値をとる。

$(k-|u|)$ 個の対は距離 $(kl+v)$ だけ離れており、$(\ell-|j|)$ は $(\ell v+j)$ だけ離れていることがわかる。

l^2k^2 通りの組合せについて平均すると

$$E S_{kl}(x_i - x_j)^2 = 2\sigma^2\left\{1 - \frac{1}{k^2l^2}\sum_{i=-(k-1)}^{(k-1)}\sum_{j=-(l-1)}^{(l-1)}(k-|i|)(l-|j|)\rho(uk+i, vl+j)\right\} \cdots (3.5)$$

をうる．

(3.4) と (3.5) によって，$(\sum_{sy} - \sum_{st})$ は

$$\frac{1}{k^2l^2}\sum_{i=-(k-1)}^{(k-1)}\sum_{j=-(l-1)}^{(l-1)}(k-|i|)(l-|j|)\rho(uk+i, vl+j) - \rho(uk, vl)$$

$$= \frac{1}{k^2l^2}(T_1 + T_2 + T_3 + T_4)$$

のような項の和に比例することがわかる．

ここで

$$T_1 = \sum_{i=1}^{(k-1)}\sum_{j=1}^{(l-1)}(k-i)(l-j)\{\rho(uk+i, vl+j) + \rho(uk+i, vl-j) + \rho(uk-i, vl+j) + \rho(uk-i, vl-j)\}$$

$$T_2 = k\sum_{j=1}^{(l-1)}(l-j)\{\rho(uk, vl+j) + \rho(uk, vl-j)\}$$

$$T_3 = l\sum_{i=1}^{k-1}(k-i)\{\rho(uk+i, vl) + \rho(uk-i, vl)\}$$

$$T_4 = -lk(lk-1)\rho(uk, vl)$$

である．

いま，T_1 から $4\rho(uk, vl)$ を，T_2, T_3 からそれぞれ $2\rho(uk, vl)$ を引いて，T_4 にその分を加えると

$$T_1' = \sum_{i=1}^{k-1}\sum_{j=1}^{l-1}(k-i)(l-j)\{\rho(uk+i, vl+j) + \rho(uk+i, vl-j) + \rho(uk-i, vl+j) + \rho(uk-i, vl-j) - 4\rho(uk, vl)\}$$

$$T_2' = k\sum_{j=1}^{l-1}(l-j)\{\rho(uk, vl+j) + \rho(uk, vl-j) - 2\rho(uk, vl)\}$$

$$T_3' = l\sum_{i=1}^{k-1}(k-i)\{\rho(uk+i, vl) + \rho(uk-i, vl) - 2\rho(uk, vl)\}$$

となり，この差引いた分の合計 $lk(lk-1)\rho(uk, vl)$ を T_4 に加えると0になる．

よって式は

$$T_1' + T_2' + T_3' \cdots\cdots\cdots\cdots\cdots\cdots (3.6)$$

となる．

いつものように

$S^2 \equiv \Delta^2 E^{-1} \equiv (E^{\frac{1}{2}} - E^{-\frac{1}{2}})^2$ とすると

$$\sum_{j=-(i-1)}^{(i-1)}(i-|j|)S^2E^j \equiv E^i + E^{-i} - 2$$

が証明できる．

よって

$$\sum_{j=-(i-1)}^{(i-1)}(i-|j|)S_2^2\phi(u+j) = \phi(u+i) + \phi(u-i) - 2\phi(u) \cdots\cdots (3.7)$$

ゆえに T_1' においては

$$\rho(uk+i, vl+j) + \cdots + \rho(uk-i, vl-j) - 4\rho(uk, vl)$$
$$= (E_1^i + E_1^{-i} - 2)\{\rho(uk, vl+j) + \rho(uk, vl-j)\} + 2(E_2^j + E_2^{-j} - 2)\rho(uk, vl)$$

(3.7) を用いて

$$= \sum_{s=-(i-1)}^{(i-1)}(i-|s|)\{S_1^2\rho(uk+s, vl+j) + S_1^2\rho(uk+s, vl-j)\} + 2\sum_{t=-(j-1)}^{(j-1)}(j-|t|)S_2^2\rho(uk, vl+t)$$

同様に T_2' については

$$\rho(uk, vl+j) + \rho(uk, vl-j) - 2\rho(uk, vl) = \sum_{t=-(j-1)}^{(j-1)}(j-|t|)S_2^2\rho(uk, vl+t)$$

また T_3' については

$$\rho(uk+i, vl) + \rho(uk-i, vl) - 2\rho(uk, vl) = \sum_{s=-(i-1)}^{(i-1)}(i-|s|)S_1^2\rho(uk+s, vl)$$

よって

(i) $S_1^2\rho(u, v)$ (ii) $S_2^2\rho(u, v) \geqslant 0$ なら T_1', T_2', T_3' はいずれもすべての u, v について正である．

ゆえにこれらの条件のもとでは $\sum_{sy} \geqslant \sum_{st}$ であるから $\sigma_{sy}^2 \leqslant \sigma_{st}^2$

この証は明らかに上の3つの場合の各々について等号が成立するときにのみ0となる．よって定理は証明された．

この節と前節の定理の証明から $\sigma^2 \geqslant \sigma_{sy}^2$ なための条件を求めることができる．ある場合には，層化した系統的抽出法を採用しなければならないことがあるから，効率の悪さの程度

(degree of inefficiency) を求めなければならない。
これは容易ではないが、$5/1$ の (1.14), (1.15), (1.16) から求めることができる。

ここでは1つの例を考えてみよう。

$$p(u,v) = \sum_{i=1}^{P} A_i e^{-\lambda_i|u| - \mu_i|v|}$$

ただし $\sum_{i=1}^{P} A_i = 1$ で λ_i および μ_i は正である。

とする。　明らかに $p(u,v) = p(-u,v)$ であるから u, v の正の値についてのみ考えればよい。
これは自然的な条件であることが示されている (Ghosh, 1949)

いま　$\psi(u,v) = 2p(u,v)$

よって　$\Delta_1\psi(u,v) = 2\sum_{i=1}^{P} A_i e^{-\lambda_i u - \mu_i v}(e^{-\lambda_i} - 1)$

しかるに　$\lambda_i \leqslant 0$ のとき

$$e^{-\lambda_i} \leqslant 1$$

ゆえに正の相関の場合

$$\Delta_1\psi(u,v) \leqslant 0$$

$\Delta_2\psi(u,v)$ についても同様である。

しかし λ_i および μ_i の値がわからなければ定理1の条件が満足されるかどうかはわからない。ただしのすべての値について $\lambda_i = \mu_i$ となるときにのみ定理1の系が適用できる*。

更に

$$\Delta_1\Delta_2\psi(u,v) = 2A_i e^{-\lambda_i u - \mu_i v}(e^{-\lambda_i} - 1)(e^{-\mu_i} - 1)$$

でこれは正である。

よって定理3が適用できる。すなわち $\sqrt{e} \leqslant k \leqslant l^2$ および $(n-1)k = (m+1)l$ なる層化抽出は、ユニバースが期待平均値 (expected mean) に関して一様であるとしても無作為抽出より

* 著者は、$\Delta_1\psi(u,v) \leqslant 0$, $\Delta_2\psi(u,v) \leqslant 0$, $S_1^2\rho(u,v) \geqslant 0$ および $S_2^2\rho(u,v) \geqslant 0$ なるとき $\sigma_n^2 > \sigma_{sL}^2 \geqslant \sigma_{sy}^2$ が成立し、また上の4つの場合の各々において等号が成り立つときに $\sigma_n^2 = \sigma_{sL}^2$ で、またそのときには $\sigma_{sL}^2 = \sigma_{sy}^2$ なることも証明した。これらの結果は Science and culture の次号に発表する予定である。

有効である。

同様に

$$S_1^2\psi(u,v) = 2\{\rho(u+1,v) + \rho(u-1,v) - 2\rho(u,v)\}$$

$$= 2\sum_{i=1}^{P} A_i e^{-\lambda_i|u| - \mu_i|v|}(e^{\lambda_i/2} - e^{-\lambda_i/2})^2$$

でこれは ρ が正のとき正である。

同様に

$$S_2^2\psi(u,v) \geqslant 0$$

ゆえに定理4を適用すれば、系統的抽出は層化抽出より有効なることがわかる。

この論文の印刷中、この中で論じた結果の幾つかが Quenouille (1949) によってもえられているという注意があつた。著者は1949年にこれらの結果を Science and Culture に発表した。

幾つかの有益な助言を与えられた Prof. B. N. Ghosh にお礼を申上げる。

By A. C. Das (Sankhya vol. 10, 1950 より)

参　考　文　献

1. Cochran, W. G. (1946) Relative accuracy of systematic and stratified random for a certain class of populations. A. M. S. 17, 164-177.
2. Das, A. C. (1949): Two dimensional systematic, stratified and random sampling proc. Ind. Sci Congress. 36th Session Prot III. 6
3. Ghosh, B. N. (1949): On a particular type of natural field proc. Ind. Sci. Congress. 36th Session, part III, 7.

4 Madow, W. G and L. H. (1944) : On the theory of systimatic sampling I. a. m. s. 15. 1-24.

5 Quenouille, M. H (1949) : problem in plane sampling a. m. s. 20, 355-375

6 Yates, F. (1948) : Systematic sampling, phil. trans. Roy. Soc 241 (A), 345-377

7 種々の形式の二重抽出法による推定値の誤差について

On errors of estimates in various type of double sampling procedure

1 緒言

二重抽出法なる用語は二つの抽出調査を含むようなサムプリング方法に適用されるようになつてきた。他の多くの種類のサムプリングと同じく、費用の減少と精度の増加はこの種のサムプリングの主な利益でもある。

Neyman (1938.) は、補助変量の大きい大標本を用いて大きい(主たる)特性の変動を小さくするように母集団をグループわけし、特性に相関がある場合には比較的小数の大きい標本から主特性の良好な推定値を求めうるような抽出方法を与える。

他のよく知られた種類の二重抽出法は、二つの特性についてとられた大きさNの大きい標本を用いて主特性yの他のxに対する回帰を定め、補助特性xのみについて観察された大きさNの大きい標本を用いて主特性yの推定値を求めるものである。この方法は特に、主特性の調査に非常に費用がかかるのに対して、それと相関を有する補助変量は容易に測定できるというような場合に適用できる。Cochran (1943) は Snedecor and King (1942) および C. Bose (1943) によつて与えられた線型回帰の仮定のもとに、この種の推定値の分散に対する抽出公式 (sampling formulae) の例を与えた。補助変量が一つの場合の特別な形の非線型回帰に対する推定値の分散の式は Bose と Groyen (1946) によつて導びかれた。

しかしながら、この種の二重抽出法の推定値の精度は、一つでなく多くの相関ある補助変量を含めることによつて向上させることが可能であろう。B. Ghosh (1947) は、回帰が線型で無作為抽出を行なう場合は推定値は不偏であることを示しまた多くの補助変量にもとずく推定値の分散の近似公式を与えている。

前に述べたように二重抽出法は補助変量が1つでもあつても多数であつても，それ自身が標本の精度を増加させる1つの方法なのである．したがつて更に他の情度の増すことのわかつているサムプリング方法と組合せて用いることが可能であろう．すなわち第1および第2標本の単位の選択には色々な方法を用いることができる．例えば第1標本を無作為に，また第2標本を系統的に選ぶというように，その上費用と精度の考慮から，ある時は予め定められたものを第1の標本としあるいは時に選ばれた個の組にとることも是認されるであろう．

この論文では色々な形式の二重抽出法に対する推定値および推定値の分散の式を求めた．補助変量が1つの場合と2つ以上の場合を別々に取扱つた．殆んどの場合線型回帰を仮定したが，2例だけは非線型回帰を仮定して考察した．第1標本の補助変量の期待値が他の変量の期待値の一定倍になつているように修正された場合を考察した．

第1，第2標本に対する標本単位の最適配分の問題は，*Schumacher and Chapman* (1942) の採用した方針にしたがつて，色々なタイプの中の1つに対する最適個数を等びくことによつて論じておいた．

附録においてまだ知られていないと思われるある型の興味ある結果，たとえば多変量正規母集団での回帰係数の同時分布，および標本変動行列の代表要素の期待値などを等びいた．これとともに，すでに知られている偏回帰係数の分布も直交座標を用いるもう一つの方法で等びいた．

2．多変補助変量の組 — サムプリングにおける色々な場合

2.1 $x_{n\iota}$ ー ランダム，$y_{n\iota}$ ー ランダム，x_N ー ランダム：

補助変量が $x_1, x_2, \cdots\cdots, x_k$ なる組からなつているとき，線型回帰を仮定すると，無作為抽出 n の y の母集団平均値の推定値は

$$Y = \bar{y}_n + \sum_{i=1}^{k} b_{ni} (\bar{X}_{Ni} - \bar{x}_{ni}) \qquad (2.11)$$

で与えられる．ここで $\bar{y}_n$，b_{ni}，$\bar{x}_{ni}$ は第1標本から，また

$\bar{X}_{Ni}$ は第2標本から等びかれる．

多変量正規母集団を仮定し

$$E(y) = \eta \qquad E(x_i) = \xi_i \ (i = 1, 2, \cdots\cdots, k)$$
$$V(y) = \sigma_y^2 \qquad V(x_i) = \sigma_{ii} = \sigma_i^2$$
$$Cov(x_i, x_j) = \sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$
$$Cov(y, x_i) = \sigma_{yi} = \rho_{yi}\sigma_y\sigma_i$$
$$E(b_{ni}) = \beta_{ni} = -\frac{R_{yi}}{R_{yy}} \frac{\sigma_y}{\sigma_i}$$

ただし R_{ij} は行列式 $|\rho_{ij}|$ $(i, j = y, 1, 2, \cdots\cdots, k)$ の ρ_{ij} の余因子．とおけば附録Aの結果を用いて直ちに次のような関係が与えられる．

$$V(b_{ni}) = \sigma_{y\cdot x}^2 E(c_{ii}) = \frac{\sigma_{y\cdot x}^2 \sigma^{ii}}{n-k-2} = \frac{\sigma_y^2 (1 - R_{y.12\cdots k}^2)\sigma^{ii}}{n-k-2}$$

$$Cov(b_{ni}, b_{nj}) = \sigma_{ij\cdot x}^2 E(c_{ij}) = \frac{\sigma_{ij\cdot x}^2 \sigma^{ij}}{n-k-2} = \frac{\sigma_y^2 (1 - R_{y.12\cdots k}^2)\sigma^{ij}}{n-k-2}$$

ここで $\sigma_{y\cdot x}^2$ は x を固定したときの y の残差分散で，$R_{y.12\cdots k}$ はまた $x_1, x_2, \cdots\cdots x_k$ 間の重相関係数

また行列 (σ^{ij}) は (σ_{ij}) の逆行列である．

したがつて第1標本内の y と x，第2標本内の x をすべてランダムに選ぶ場合の二重抽出においては

$$E(Y) = \eta + \sum_{i=1}^{k} \beta_{ni} (\xi_i - \xi_i) = \eta$$

$$V(Y) = E\left\{(\bar{y}_n - \eta) + \sum_{i=1}^{k} b_{ni} (\bar{x}_{Ni} - \xi_i) - \sum_{i=1}^{k} b_{ni} (\bar{x}_{ni} - \xi_i)\right\}^2$$

$$= \frac{\sigma_y^2}{n} + \left\{\sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij}\right\} \beta_{ni}\beta_{nj} + \frac{\sigma_{y\cdot x}^2 \sigma^{ij}}{n-k-2}\Big] \times \left(\frac{1}{n} + \frac{1}{N}\right) - 2\sum_{i=1}^{k} \beta_{ni} \frac{\sigma_{yi}}{n}$$

$$= \frac{\sigma_y^2}{n} + \left(\frac{1}{n} + \frac{1}{N}\right) \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ni}\beta_{nj}\sigma_{ij}$$

$$+ \frac{\sigma_y^2 (1 - R_{y.12\cdots k}^2)}{n-k-2} \cdot \left(\frac{1}{n} + \frac{1}{N}\right) \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{ij}\sigma^{ij} - 2\sum_{i=1}^{k} \beta_{ni} \frac{\sigma_{yi}}{n}$$

が成立する．

いま $\sum_{j=1}^{k} \sigma_{ij}\sigma^{ij} = 1$ $\quad \therefore \sum_{i=1}^{k}\sum_{j=1}^{k} \sigma_{ij}\sigma^{ij} = k$

$$\sum_{j=1}^{k} \beta_{ni}\sigma_{ij} = \sigma_{yi}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{k} \beta_{ni}\beta_{nj} = \sum_{i=1}^{k} \beta_{ni}\sigma_{yi} = \sum_{i=1}^{k} - \frac{R_{yi}}{R_{yy}}\frac{\sigma_y}{\sigma_i} z_{yi}\sigma_i\sigma_y$$

$$= \frac{-(\sum_{i=1}^{k} z_{yi} R_{yi})}{R_{yy}}\sigma_y^2 = \left(\frac{R - R_{yy}}{R_{yy}}\right)\sigma_y^2$$

$$= \sigma_y^2 R^2_{y.12\cdots k}$$

よって

$$V(Y) = \frac{\sigma_y^2}{n} + \frac{k}{n-k-2}\sigma_y^2(1-R^2_{y.12\cdots k})\left(\frac{1}{n}+\frac{1}{N}\right)$$

$$+ \sigma_y^2 R^2_{y.12\cdots k}\left(\frac{1}{N} - \frac{1}{n}\right)$$

$$= \sigma_y^2(1-R^2_{y.12\cdots k})\left(\frac{1}{n} + \frac{k}{n-k-2}\left(\frac{1}{n}+\frac{1}{N}\right)\right) + \frac{\sigma_y^2 R_{y.12\cdots k}}{N} \quad (2.13)$$

(2.13)式 は各 x の分散 σ_{ii} と独立であることに注意せよ

(2.2) x_n ー ランダム， y_n ー ランダム， x_N ー システマティック：

ここで更に一般な場合，すなわち母集団の x と y はランダムに選ぶが，第2標本はシステマティックな方法で選ぶことにしても，推定値 Y については (2.11) と同じ関係が成立する。

$$E(Y) = \eta$$

$$E\left\{(\bar{x}_{Ni} - \bar{x}_i)(\bar{x}_{Nj} - \bar{x}_j)\right\} = \frac{1}{N^2} E\left\{\sum_{p=1}^{N}\sum_{q=1}^{N}(x_{Nip} - \bar{x}_i)(x_{Njq} - \bar{x}_j)\right\}$$

$$= \frac{\sum_{p=1}^{N}\sum_{q=1}^{N} z^{(xx)}_{ipjq}\sigma_{ij}}{N^2}$$

ここで $z^{(xx)}_{ipjq}$ は $(p = 1, \cdots N : q = 1, \cdots, N)$ は母集団における x_{Nip} と x_{Njp} の相関を表わす。

$$V(Y) = E\left\{(\bar{y}_n - \eta) + \sum_{i=1}^{k} b_{ni}(\bar{x}_{Ni} - \bar{x}_i) - \sum_{i=1}^{k} b_{ni}(\bar{x}_i - \bar{x}_i)\right\}^2$$

$$= \frac{\sigma_y^2}{n} + \sum_{i=1}^{k}\sum_{j=1}^{k}\left[\beta_{ni}\beta_{nj} + \frac{\sigma_{yx}^2\sigma^{ij}}{n-k-2}\right]\left[\frac{\sum_{p=1}^{N}\sum_{q=1}^{N} z^{(xx)}_{ipjq}\sigma_i\sigma_j}{N^2} + \frac{\sigma_{ij}}{n}\right]$$

$$- 2\sum_{i=1}^{k}\beta_{ni}\frac{\sigma_{yi}}{n}$$

$$= \frac{\sigma_y^2}{n} + \sum_{i=1}^{k}\sum_{j=1}^{k}\left[\beta_{ni}\beta_{nj} + \frac{\sigma_y^2(1-R^2_{y.12\cdots k})\sigma^{ij}}{n-k-2}\right] \times$$

$$\left\{\frac{\sum_{p=1}^{N}\sum_{q=1}^{N} z^{(xx)}_{ipjq}\sigma_i\sigma_j}{N^2} + \frac{k\sigma_y^2(1-R^2_{y.12\cdots k})}{n(n-k-2)} - \frac{\sigma_y^2 R^2_{y.12\cdots k}}{n}\right.$$

$$= \frac{\sigma_y^2(1-R^2_{y.12\cdots k})}{n}\frac{n-2}{n-k-2} + \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\sum_{p=1}^{N}\sum_{q=1}^{N} z^{(xx)}_{ipjq}\sigma_i\sigma_j}{N^2}$$

$$\times\left[\beta_{ni}\beta_{nj} + \frac{\sigma_y^2(1-R^2_{y.12\cdots k})}{n-k-2}\sigma^{ij}\right] \quad \cdots\cdots\cdots\cdots (2.22)$$

2.1で考察したのは，これの

$$z^{(xx)}_{ipjq} = z_{ij} \qquad p = q \text{ のとき}$$

$$= 0 \qquad p \neq q \text{ のとき}$$

であるような特別な場合になっている。

(2.3) x_n ー 固定， y_n ー ランダム， x_N ー ランダム

次に我々は，第1標本の x とは固定するが，第2標本の x と第1標本の y はランダムに選ぶという第3の場合を考える。 y の母集団平均値の推定値はこの場合も (2.11) と同じである。

$$E(y_{np}) = \alpha_n + \sum_{i=1}^{k}\beta_{ni}x_{nip} \qquad (p = 1, 2, \cdots, n)$$

と仮定すると

$$E(\bar{y}_n) = \alpha_n + \sum_{i=1}^{k}\beta_{ni} \quad (\equiv \eta' \text{とおく})$$

よって

$$E(Y) = \eta' + \sum_{i=1}^{k}\beta_{ni}(\bar{x}_{ni} - \bar{x}_{ni}) = \eta \quad \cdots\cdots\cdots (2.31)$$

$$V(Y) = E\left\{(\bar{y}_n - \eta) - \sum_{i=1}^{k}(\bar{x}_{ni} - \bar{z}_i)(b_{ni} - \beta_{ni}) + \sum_{i=1}^{k} b_{ni}(x_{Ni} - \bar{z}_i)\right\}^2$$

$$= A + \frac{B}{2} + \frac{C}{N} \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2.32)$$

ただし

$$A = \sigma_{ix}^2 \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{x}_{ni} - \bar{z}_i)(\bar{x}_{nj} - \bar{z}_j) C_{ij} \right\}$$

$$B = \sigma_{ix}^2$$

$$C = \sum_{i=1}^{k} \sum_{j=1}^{k} \sigma_{ij} \left\{ C_{ij} \sigma_{yx}^2 + \beta_{ni} \beta_{nj} \right\}$$

系　総費用 T が

$$T = \alpha + \beta n + \gamma N \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (2.33)$$

ただし α, β, γ はデーターから推定されるパラメーターである．なる形で表わされるとすれば，与えられた費用 T に対して分散を最小とするような n および N の最適値は

$$\frac{\partial V}{\partial n} + \lambda \frac{\partial T}{\partial n} = 0$$

$$\frac{\partial V}{\partial N} + \lambda \frac{\partial T}{\partial N} = 0$$

すなわち

$$\frac{B n}{\sqrt{B\beta}} = \frac{\gamma N}{\sqrt{C\gamma}} = \frac{T - \alpha}{\sqrt{B\beta} + \sqrt{C\gamma}}$$

なる方程式を満足しなければならないから

$$\left.\begin{aligned} n &= \frac{T-\alpha}{\beta} \cdot \frac{\sqrt{B\beta}}{\sqrt{B\beta} + \sqrt{C\gamma}} \\ N &= \frac{T-\alpha}{\gamma} \cdot \frac{\sqrt{C\gamma}}{\sqrt{B\beta} + \sqrt{C\gamma}} \end{aligned}\right\} \qquad \cdots\cdots\cdots\cdots (2.34)$$

である．

(2.4) x_n ー固定，y_n ーランダム，x_N ー システマティック

ここでは第 1 標本の x は固定するが，y はランダムで，第 2 標本の x には相関があるという第4の場合を考える．ここでも (2.11) は成立して

$$E(Y) = \eta + \sum_{i=1}^{k} \beta_{ni}(\bar{z}_i - \bar{x}_{ni}) = \eta \qquad \cdots\cdots\cdots\cdots (2.41)$$

$$V(Y) = E\left\{(\bar{y}_n - \eta) - \sum_{i=1}^{k}(\bar{x}_{ni} - \bar{z}_i)(b_{ni} - \beta_{ni}) - \sum_{i=1}^{k} b_{ni}(\bar{x}_{Ni} - \bar{z}_i)\right\}^2$$

$$= \frac{\sigma_{y.x}^2}{n} + \sum_{i=1}^{k} \sum_{j=1}^{k} E\left\{ \frac{\sum_{p=1}^{N} \sum_{q=1}^{N} (x_{Nip} - \bar{z}_i)(x_{Njq} - \bar{z}_j)}{N^2} \right\} E(b_{ni} b_{nj})$$

$$+ \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{x}_{ni} - \bar{z}_i)(\bar{x}_{nj} - \bar{z}_j) C_{ij} \sigma_{y.x}^2$$

$$= \frac{\sigma_{y.x}^2}{n} + \frac{1}{N^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left\{ \beta_{ni} \beta_{nj} + C_{ij} \sigma_{y.x}^2 \right\} \left\{ \sum_{p=1}^{N} \sum_{q=1}^{N} L_{ip,jq}^{(xx)} \sigma_i \sigma_j \right\}$$

$$+ \sum_{i=1}^{k} \sum_{j=1}^{k} (\bar{x}_{ni} - \bar{z}_i)(\bar{x}_{nj} - \bar{z}_j) C_{ij} \sigma_{y.x}^2 \qquad \cdots\cdots (2.42)$$

(2.42) から V(y) は第 1 標本の x が広範囲にわたり，しかも平均値 $\bar{x}_{ni}$ が母集団平均 $\bar{z}_i$ ($i = 1, 2, \cdots\cdots, k$) に近い程小となることがわかる．

明らかに (2.32) の結果は (2.42) から

$$L_{ip,jq}^{(xx)} = L_{ij} \qquad p = q \text{ のとき}$$

$$= 0 \qquad p \neq q \text{ のとき}$$

3．補助変量が 1 つのとき ー カムプリングにおける色々な場合

(3.1) x_n ー固定，y_n ーシステマティック，x_N ー ランダム：

我々はここで補助変量が 1 つしかないという特別な場合の二重抽出法を考察する．x は第 1 標本では固定するが y には相関があり，また第 2 標本の x はランダムに選ぶものとする．

$$E(x_{Ni}) = \bar{z}_i \quad (i = 1, 2, \cdots\cdots, N)$$

$$E(y_{ni}) = \alpha_n + \beta_n x_{ni} \quad (i = 1, 2, \cdots\cdots, n)$$

$$V(y_{ni} \mid x_{ni}) = \sigma_{y.x}^2$$

$$Cov(y_{ni}, y_{nj} \mid x_{ni}, x_{nj}) = L_{ij}^{(yy)} \sigma_{y.x}^2$$

このとき

$$Y = \bar{y}_n + b_n(\bar{x}_N - \bar{x}_n) \qquad \cdots\cdots\cdots\cdots\cdots\cdots (3.11)$$

$$E(b_n) = E\left\{ \frac{\sum_{i=1}^{n} (x_{ni} - \bar{x}_n)(y_{ni} - \bar{y}_n)}{\sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2} \right\}$$

$$= \beta_n$$

$$E(Y) = (\alpha_n + \beta_n \bar{x}_n) + \beta_n(\xi - \bar{x}_n)$$

$$= \alpha_n + \beta_n \xi \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (3.12)$$

が成立する。

さらに

$$V(\beta_n) = \frac{\sigma_{Y\cdot x}^2 \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ni} - \bar{x}_n)(x_{nj} - \bar{x}_n) \rho_{ij}^{(YY)}}{\left[\sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2 \right]}$$

$$Cov(b_n, \bar{y}_n) = \frac{1}{n} E\left\{ \frac{\sum_{i=1}^{n} y_{ni}(x_{ni} - \bar{x}_n) \sum_{i=1}^{n} (y_{ni} - \alpha_n - \beta_n x_{ni})}{\sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2} \right\}$$

$$= \frac{\sigma_{Y\cdot x}^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)} (x_{ni} - \bar{x}_n)}{n \sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2}$$

$$V(\bar{y}_n) = \frac{\sigma_{Y\cdot x}^2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)}$$

である。

よって

$$V(Y) = E\left\{ (\bar{y}_n - \alpha_n - \beta_n \bar{x}_n) + b_n(\bar{x}_N - \xi) + (\beta_n - \beta_n)(\xi - \bar{x}_n) \right\}^2$$

$$= \sigma_{Y\cdot x}^2 \left\{ \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)}}{n^2} + \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)} (x_{ni} - \bar{x}_n)(x_{nj} - \bar{x}_n)}{\left[\sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2 \right]^2} \right.$$

$$\left. \times \left\{ \frac{\sigma_x^2}{N} + (\xi - \bar{x}_n)^2 \right\} + \frac{2(\xi - \bar{x}_n) \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)} (x_{ni} - \bar{x}_n)}{n \sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2} \right\} + \frac{\sigma_{Y\cdot x}^2 \beta_n^2}{N}$$

(3.13)の特別な場合、すなわち

$$\rho_{ij}^{(YY)} = 1 \qquad (i = 1, 2, \cdots\cdots, n)$$

$$\rho_{ij}^{(YY)} = 0 \qquad (i \neq j)$$

なる場合が $k = 1$ のときの(2.32)である。$V(Y)$の推定値を求めるには相関係数 $\rho_{ij}^{(YY)}$ について $|i-j| = u$ なるとき $\rho_{ij} = \rho_u$ というようなある仮定を設けなければならない。

(3.2) x_n 一固定、y_n ーシステマティック、x_nーシステマティック

ここでは第1標本の x は固定するか、第2標本の x と第1標本の y には相関があるものとする。

このときも(3.11)と(3.12)は成立するから

$$Cov(x_{ni}, x_{nj}) = \rho_{ij}^{(xx)} \sigma_x^2 \qquad (i, j = 1, 2, \cdots\cdots, N)$$

と仮定すれば

$$V(Y) = \sigma_{Y\cdot x}^2 \left[\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)} + \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)} (x_{ni} - \bar{x}_n)(x_{nj} - \bar{x}_n)}{n^2} \right.$$

$$\left. \times \left\{ \frac{\sigma_x^2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{ij}^{(xx)} + (\xi - \bar{x}_n)^2 \right\} + \frac{2(\xi - \bar{x}_n)}{\sum_{i=1}^{n} (x_{ni} - \bar{x}_n)^2} \right.$$

$$\left. \times \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}^{(YY)} (x_{ni} - \bar{x}_n)}{n} \right] + \frac{\sigma_x^2 \beta_n^2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{ij}^{(xx)} \cdots (3.21)$$

が成立する。

$$\rho_{ij}^{(xx)} = 1 \qquad (i = 1, 2, \cdots\cdots, N)$$

$$\rho_{ij}^{(xx)} = 0 \qquad (i \neq j)$$

ならこの式は(3.13)に帰着する。

(3.3) x_nーシステマティック、y_nーシステマティック、x_Nーシステマティック

我々はここで更に一般な場合、すなわち第1標本内の x と y および第2標本内の x がすべてそれら対の間で相関を有するという場

合を取扱う，この場合のYの期待値および分散は近似公式でさえ非常に複雑である．ここでも（3.11）は成立する．N個の標本のXとYについて多変量正規分布を仮定し，

$$\left.\begin{aligned} &E(X_{ni}) = E(X_{Ni}) = \xi \\ &E(Y_{ni}) = \eta \\ &Cov(X_{ni}, X_{nj}) = \zeta_{ij}^{(XX)}\sigma_x^2 \\ &Cov(Y_{ni}, Y_{nj}) = \zeta_{ij}^{(YY)}\sigma_y^2 \\ &Cov(X_{ni}, Y_{nj}) = \zeta_{ij}^{(XY)}\sigma_x\sigma_y \end{aligned}\right\} (i,j = 1, 2, \cdots\cdots, n)$$

$$E(X_{ni} - \xi)(X_{nj} - \xi)(Y_{nk} - \eta) = 0 \quad (i, j, k = 1, 2 \cdots\cdots, n)$$

とおけば

$$E(X_{ni}, X_{nj} Y_{nk}) = \left\{\xi^2 + \zeta_{ij}^{(XX)}\sigma_x^2\right\}\eta + \xi\sigma_x\sigma_y\left\{\zeta_{ik}^{(XY)} + \zeta_{jk}^{(XY)}\right\}$$

であるから

$$E\left\{\sum_{i=1}^{n} X_{ni} Y_{ni} - \frac{\sum_{i=1}^{n} X_{ni} \sum_{i=1}^{n} Y_{ni}}{n}\right\}$$

$$E(b_n) \simeq \frac{}{E\left\{\sum_{i=1}^{n} X_{ni}^2 - \frac{\left(\sum_{i=1}^{n} X_{ni}\right)^2}{n}\right\}}$$

$$\simeq \frac{\sigma_y}{\sigma_x} \cdot \frac{\left\{n(n-1)\zeta - \sum_{i\neq j=1}^{n} \zeta_{ij}^{(XY)}\right\}}{\left\{n(n-1) - \sum_{i\neq j=1}^{n} \zeta_{ij}^{(XX)}\right\}} \equiv \beta_n \text{（とおく）}$$

ここで $\zeta = \zeta_{ij}^{(XY)}$ $(i = 1, 2, \cdots\cdots, n)$ が成立する。

さらに

$$E\left\{\sum_{i=1}^{n} X_{ni} \sum_{i=1}^{n} X_{ni} Y_{ni}\right\} = \sum_{i=1}^{n} \sum_{j=1}^{n} E(X_{ni} Y_{ni} X_{nj})$$

$$= n^2\xi^2\eta + \sum_{i=1}^{n}\sum_{j=1}^{n} \zeta_{ij}^{(XX)}\sigma_x^2\eta + n^2\xi\zeta\sigma_x\sigma_y + \sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(XY)}\xi\sigma_x\sigma_y$$

$$E\left\{\left(\sum_{i=1}^{n} X_{ni}\right)^2 \sum_{i=1}^{n} Y_{ni}\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} E(X_{ni} X_{nj} Y_{nk})$$

$$= n^2\xi^2\eta + n\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(XX)}\sigma_x^2\eta + 2n\sum_{i=1}^{n}\sum_{j=1}^{n}\zeta_{ij}^{(XY)}\xi\sigma_x\sigma_y$$

よって

$$E\left(b_n \bar{X}_n\right) \simeq \frac{E\left[\sum_{i=1}^{n} X_{ni} Y_{ni} \sum_{i=1}^{n} X_{ni} - \frac{\left(\sum_{i=1}^{n} X_{ni}\right)^2 \left(\sum_{i=1}^{n} Y_{ni}\right)}{n}\right]}{nE\left\{\sum_{i=1}^{n} X_{ni}^2 - \frac{\left(\sum_{i=1}^{n} X_{ni}\right)^2}{n}\right\}}$$

$$\simeq \xi\frac{\sigma_y}{\sigma_x} \frac{\left\{n(n-1) - \sum_{i\neq j=1}^{n} \zeta_{ij}^{(XY)}\right\}}{\left\{n(n-1) - \sum_{i\neq j=1}^{n} \zeta_{ij}^{(XX)}\right\}}$$

$$\simeq E(\bar{X}_n) E(b_n)$$

$$\therefore E(Y) \simeq \eta + \beta_n(\xi - \xi) = \eta \quad \cdots\cdots\cdots\cdots\cdots (3.31)$$

また

$$E(b_n^2) \simeq \frac{E\left\{\sum_{i=1}^{n}(X_{ni} - \xi)(Y_{ni} - \eta)\right\}^2}{E\left\{\sum_{i=1}^{n}(X_{ni} - \xi)^2\right\}^2}$$

$$\simeq \frac{E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} X'_{ni} X'_{nj} Y'_{ni} Y'_{nj}\right\}}{E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} X'^2_{ni} X'^2_{nj}\right\}}$$

ここで X'，Y'_n は対応する変量のその平均値からの偏差を表わす。

（これ）のときの $E(X'_{ni}, X'_{nj}, Y'_{ni}, Y'_{nj})$ は4変数の積率母函数（この分布は多変量正規分布であると仮定されている）から直ちに求められて

$$E(x'_{ni}\,x'_{nj}\,y'_{ni}\,y'_{nj}) = \sigma_x^2\sigma_y^2\{\zeta_{ij}^{(xx)}\zeta_{ij}^{(yy)}+\zeta_{ij}^{(xy)}\zeta_{ji}^{(xy)}+\zeta^2\} \quad (i\neq j)$$

となる．

同様に（$i=j$）のときは

$$E(x'^2_{ni}\,x'^2_{nj}\,y'_{ni}\,y'_{nj}) = E(x'^2_{ni}\,y'^2_{ni}) = \sigma_x^2\sigma_y^2(1+2\zeta^2)$$

$$E(x'^2_{ni}\,x'^2_{nj}) = \sigma_x^4\{1+2\zeta_{ij}^{(xx)2}\}$$

$$E(x'^4_{ni}) = 3\sigma_x^4$$

となる．ゆえに

$$E(\beta_n^2) \simeq \frac{\sigma_y^2}{\sigma_x^2}\,\frac{\{n+n(n+1)\zeta^2+\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xx)}\zeta_{ij}^{(yy)}+\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xy)}\zeta_{ji}^{(xy)}\}}{\{n(n+2)+2\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xx)2}\}}$$

$$\simeq \beta_n^2 \quad (\text{とおく})$$

$$\therefore V(Y) \simeq E\{(\bar{y}_n-\eta)+\beta_n(\bar{X}_N-\xi_0)-\beta_n(\bar{x}_n-\xi)\}^2$$

$$\simeq E(\bar{y}_n-\eta)^2+\{E(\bar{X}_N-\xi)^2+E(\bar{x}_n-\xi)^2+E(\bar{x}_n-\xi_0)^2E(\beta_n^2)-2E(\beta_n)E(\bar{x}_n-\xi)(\bar{y}_n-\eta)$$

$$\simeq \frac{\sigma_y^2}{n^2}\{n+\sum_{i\neq j=1}^{n}\zeta_{ij}^{(yy)}\}-\frac{2\beta_n\sigma_x\sigma_y}{n^2}\{n\zeta+\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xy)}\}$$

$$+\beta_n^2\sigma_x^2\{\frac{1}{N}+\frac{1}{n}+\frac{1}{N^2}\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xx)}+\frac{1}{n^2}\sum_{i\neq j=1}^{n}\zeta_{ij}^{(xx)}\} \cdots (3.42)$$

(3.4) x_n ― ランダム，y_n ― ランダム，x_N ―ランダム ただし $E(\bar{x}_n)\neq E(\bar{x}_N)$ とする．

ここでは少し違った問題を考える．第1の標本のxとyがともにランダムであるばかりでなく、第2標本のxもランダムである．しかし

$$E(\bar{x}_n) = \lambda E(\bar{x}_N) = \xi$$

なる関係は成立するとする．

$$E(\bar{y}_n) = \eta$$

$$V(x_n) = \sigma_{xn}^2,\ V(x_N) = \sigma_{xN}^2$$

$$V(y_n) = \sigma_{yn}^2,\ \mathrm{cov}(x_n y_n) = \zeta_{x_n y_n}\sigma_{xn}\sigma_{yn}$$

とする．

この場合の実例は作物収穫の未乾燥重量をx，またyを対応する乾燥重量と考えてえられる．xおよびyは特定の抽出単位によって書成される第1標本では両方とも測定されるが，第2標本のxは他の抽出単位の組について測定される．

この場合yの母集団平均ηの不偏測定値は

$$Y = \bar{y}_n + b_n(\lambda\bar{x}_N - \bar{x}_n)$$

$$E(Y) = \eta+\beta(\xi-\xi) = \eta \quad \cdots\cdots\cdots\cdots\cdots\cdots (3.41)$$

で与えられる．

$$V(Y) = E\left\{(\bar{y}_n-\eta)+b_n\lambda\left(\bar{x}_N-\frac{\xi}{\lambda}\right)-b_n(\bar{x}_n-\xi)\right\}^2$$

$$= \frac{\sigma_{yn}^2}{n}+\left(\frac{\lambda^2\sigma_{xN}^2}{N}+\frac{\sigma_{xn}^2}{n}\right)\frac{\sigma_{yn}^2}{\sigma_{xn}^2}\left\{1-\zeta_{xnyn}^2\frac{\gamma_n}{n-3}+\zeta_{xnyn}^2\right\}-2\zeta_{xnyn}^2\frac{\sigma_{yn}^2}{n}$$

$$= \frac{\sigma_{yn}^2}{n}(1-2\zeta_{xnyn}^2)+\frac{\sigma_{yn}^2}{\sigma_{xn}^2}\left\{\frac{\lambda^2\sigma_{xN}^2}{N}+\frac{\sigma_{xn}^2}{n}\right\}\left\{1+\frac{(n-4)\zeta_{xnyn}^2}{n-3}\right\} \cdots (3.42)$$

ここで $\sigma_{xn}^2 = \mu^2\sigma_{xN}^2$ なら（3.42）は

$$V(Y) = \frac{\sigma_{yn}^2}{n}(1-2\zeta_{xnyn}^2)+\sigma_{yn}^2\left\{\frac{\lambda^2}{\mu^2}\frac{1}{N}+\frac{1}{n}\right\}\left\{1+\frac{(n-4)\zeta_{xnyn}^2}{n-3}\right\} \cdots (3.43)$$

となる．

さらにもし第1標本と第2標本のxとの相関係数が等しい，すなわち $\lambda=\mu$ のときは，（3.43）は

$$V(Y) = \frac{\sigma_{yn}^2}{n}\,\frac{n-2}{n-3}(1-\zeta_{xnyn}^2)+\frac{\sigma_{yn}^2}{N(n-3)}\{1+\zeta_{xnyn}^2(n-4)\} \cdots (3.44)$$

となる．

(3.44)を b_n が1となるときの（2.13）と比較すると興味がある．

非線型の場合

(3.5) x_n ― 固定，y_n ― ランダム，x_N ― ランダム

これまで我々はxとyの間に線型関係の成立が仮定されるような場合を取扱ってきた．この関係を非線型とし，二次の放物線に限定

すれば、

$$y = \alpha + \beta(x - \bar{x}_n) + \gamma(x^2 - \bar{x}_n^2)$$

と書ける．ここで $\bar{x}_n$ と $\bar{x}_n^2$ はn個の標本のxとx²の平均値，すなわち，

$$\bar{x}_n = \frac{\sum_{i=1}^{n} x_{ni}}{n} \quad \text{および} \quad \bar{x}_n^2 = \frac{\sum_{i=1}^{n} x_{ni}^2}{n}$$

である．このとき α，β，γ を推定するための正規方程式は

$$\bar{y}_n = a_n$$

$$\Sigma(x - \bar{x}_n)y = b_n \Sigma(x - \bar{x}_n)^2 + c_n \Sigma(x - \bar{x}_n)(x^2 - \bar{x}_n^2)$$

$$\Sigma(x^2 - \bar{x}_n^2)y = b_n \Sigma(x - \bar{x}_n)(x^2 - \bar{x}_n^2) + c_n \Sigma(x^2 - \bar{x}_n^2)^2$$

ここで a_n b_n および c_n はそれぞれα，β，γ の推定値である．

上式から

$$b_n = C_{11}\Sigma(x - \bar{x}_n)y + C_{12}\Sigma(x^2 - \bar{x}_n^2)y$$

$$c_n = C_{21}\Sigma(x - \bar{x}_n)y + C_{22}\Sigma(x^2 - \bar{x}_n^2)y$$

と書ける．ここで C_{11}，C_{12}，C_{22} はすぐ求められる．

n個の標本のxを固定し，n個の標本のyとn2個の標本のxをランダムと仮定すると

$$V(a_n) = \frac{\sigma_{Y\cdot x}^2}{n}, \quad Cov(a_n, b_n) = Cov(a_n, c_n) = 0$$

$$V(b_n) = C_{11}\sigma_{Y\cdot x}^2 \quad Cov(b_n, c_n) = C_{12}\sigma_{Y\cdot x}^2$$

$$V(c_n) = C_{22}\sigma_{Y\cdot x}^2$$

この場合yの母集団平均値は

$$Y = \bar{y}_n + b_n(\bar{x}_N - \bar{x}_n) + c_n(\bar{x}_N^2 - \bar{x}_n^2) \quad \cdots\cdots (3.51)$$

である．ここで $\bar{x}_N$ と $\bar{x}_N^2$ は $\bar{x}_n$ $\bar{x}_n^2$ と同様な意味をもつ．

$$E(\bar{x}_N) = \mu_1', \quad E(\bar{x}_N^2) = \mu_2'$$

とおけば

$$E(Y) = \alpha + \beta(\mu_1' - \bar{x}_n) + \gamma(\mu_2' - \bar{x}_n^2) = \eta \quad \cdots\cdots (3.52)$$

および

$$\begin{aligned} V(Y) = E\Big\{&(\bar{y}_n - \alpha) + (b_n - \beta)(\bar{x}_N - \mu_1') - (b_n - \beta)(\bar{x}_n - \mu_1') \\ &+ \beta(\bar{x}_N - \mu_1') + (c_n - \gamma)(\bar{x}_N^2 - \mu_2') - (c_n - \gamma)(\bar{x}_n^2 - \mu_2') + \gamma(\bar{x}_N^2 - \mu_2')\Big\}^2 \\ = &\frac{\sigma_{Y\cdot x}^2}{n} + C_{11}\sigma_{Y\cdot x}^2\left\{\frac{\mu_2' - \mu_1'^2}{N} + (\bar{x}_n - \mu_1')^2\right\} + \beta\frac{\mu_2' - \mu_1'^2}{N} \\ &+ C_{22}\sigma_{Y\cdot x}^2\left\{\frac{\mu_3' - \mu_1'\mu_2'}{N} + (\bar{x}_n^2 - \mu_2')^2\right\} + \gamma^2\frac{\mu_4' - \mu_2'^2}{N} \\ &+ 2C_{12}\sigma_{Y\cdot x}^2\left\{\frac{\mu_3' - \mu_1'\mu_2'}{N} + (x_n - \mu_1')(\bar{x}_n^2 - \mu_2')\right\} \\ &+ 2\beta\gamma\cdot\frac{\mu_3' - \mu_1'\mu_2'}{N} \quad \cdots\cdots (3.53) \end{aligned}$$

ここで

$$\mu_r' = E\left\{\frac{\sum_{i=1}^{n} x_{ni}^r}{n}\right\}$$

(3.5α) (3.5) の上に述べた結果はyとx間の非線型関係にP次の放物線すなわち

$$y = \alpha + \beta'(x - \bar{x}_n) + \beta^2(x^2 - \bar{x}_n^2) + \cdots + \beta^P(x^P - \bar{x}_n^P)$$

ここで $\bar{x}_n^k = \sum_{i}^{n} \frac{x_{ni}^k}{n}$ (βの添字は巾ではない．)

を仮定した場合にも応用できる．

係数 α，β'，β^2，$\cdots$ β^P は P+1 個の正規方程式を解いて得た a_n，b_n' b_n^2 b_n^P で推定できる．

$$b_n^k = C_{k1}\Sigma(x - \bar{x}_n)y + C_{k2}\Sigma(x^2 - \bar{x}_n^2)y + \cdots + C_{kP}\Sigma(x^P - \bar{x}_n^P)y$$

$$(k = 1, 2, \cdots\cdots, P)$$

と書ける．

ここで $V(a_n) = \frac{\sigma_{Y\cdot x}^2}{n}$，$Cov(b_n^i, b_n^j) = C_{ij}\sigma_{Y\cdot x}^2$ $(i, j = 1, 2, \cdots, P)$

大きさNのxの他の独立な標本をとったときのyの母集団平均の不偏推定値は $Y = \bar{y}_n + b_n'(\bar{x}_N - \bar{x}_n) + b_n^2(\bar{x}_N^2 - \bar{x}_n^2) + \cdots + b_n^P(\bar{x}_N^P - \bar{x}_n^P) \cdots (3.51a)$

$$E(Y) = \eta$$

$$\begin{aligned} V(Y) = &\frac{\sigma_{Y\cdot x}^2}{n} + \sigma_{Y\cdot x}^2\sum_{i=1}^{P}\sum_{j=1}^{P} C_{ij}\left\{\frac{\mu_{i+j}' - \mu_i'\mu_j'}{N} + (\bar{x}_n^i - \bar{x}_n^i)(\bar{x}_n^j - \mu_j')\right\} \\ &+ \sum_{i=1}^{P}\sum_{j=1}^{P} b_n^i b_n^j \frac{\mu_{i+j}' - \mu_i'\mu_j'}{N} \quad \cdots (3.53a) \end{aligned}$$

となる。

(3.6) 非線型回帰の場合：x_n ランダム、y_n ランダム、z_n ランダム

ここでわれわれの標本の大きさも固定されることなくランダムにとられるものとすれば前と同じく

$$Y = \bar{y}_n + b_n(\bar{x}_N - \bar{x}_n) + c_n(\bar{x}_N^2 - \bar{x}_n^2) \quad \cdots\cdots (3.61)$$

である。

$$E(\bar{x}_n) = E(\bar{x}_N) = \mu_1'$$
$$E(\bar{x}_n^2) = E(\bar{x}_N^2) = \mu_2'$$
$$E(y_n) = \eta$$

とおき、大標本近似（b_n が $\bar{x}_n$ と独立、かつ c_n が $\bar{x}_n^2$ と独立）を仮定すると、Y の期待値は近似的に

$$E(Y) \simeq \eta + E(b_n)E(\bar{x}_N - \bar{x}_n) + E(c_n)E(\bar{x}_N^2 - \bar{x}_n^2) \simeq \eta \quad \cdots (3.62)$$

で与えられる。

また $V(b_n) = \sigma_{yx}^2 E(C_{11}) = \sigma_{bb}$ （とおく）

$V(c_n) = \sigma_{yx}^2 E(C_{22}) = \sigma_{cc}$ （とおく）

$Cov(b_n, c_n) \cdot \sigma_{yx}^2 E(C_{12}) = \sigma_{bc}$ （とおく）

ここで

$$C_{11} = \begin{vmatrix} \Sigma(x-\bar{x}_n)^2 & \Sigma(x^2-\bar{x}_n^2)(x-\bar{x}_n) \\ \Sigma(x^2-\bar{x}_n^2)(x-\bar{x}_n) & \Sigma(x^2-\bar{x}_n^2)^2 \end{vmatrix}$$

で、C_{22} および C_{12} も同様である。

ここで

$$E\left\{\Sigma(x-\bar{x}_n)^2\right\} = E\left\{\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right\} = (n-1)(\mu_2' - \mu_1'^2)$$

$$E\left\{\Sigma(x^2-\bar{x}_n^2)^2\right\} = E\left\{\Sigma x^4 - \frac{(\Sigma x^2)^2}{n}\right\} = (n-1)(\mu_4' - \mu_2'^2)$$

$$E\left\{\Sigma(x-\bar{x}_n)(x^2-\bar{x}_n^2)\right\} = E\left\{\Sigma x^3 - \frac{(\Sigma x)(\Sigma x^2)}{n}\right\} = (n-1)(\mu_3' - \mu_1'\mu_2')$$

$$E(C_{11}) \simeq \frac{\text{分子の期待値}}{\text{分母の期待値}}$$

$$\simeq \frac{1}{(n-1)} \frac{(\mu_4' - \mu_2'^2)}{\mu_2(\mu_4' - \mu_2'^2) - (\mu_3' - \mu_1'\mu_2')^2}$$

$$E(C_{22}) \simeq \frac{1}{(n-1)} \frac{\mu_2}{\mu_2(\mu_4' - \mu_2'^2) - (\mu_3' - \mu_1'\mu_2')^2}$$

$$E(C_{12}) \simeq \frac{1}{(n-1)} \frac{(\mu_3' - \mu_1'\mu_2')}{\mu_2(\mu_4' - \mu_2'^2) - (\mu_3' - \mu_1'\mu_2')^2}$$

ここで

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$V(Y) \simeq E\left\{(\bar{y}_n - \eta) + b_n(\bar{x}_N - \mu_1') - \beta_n(\bar{x}_n - \mu_1') + c_n(\bar{x}_N^2 - \mu_2') - \gamma_n(\bar{x}_n^2 - \mu_2')\right\}^2$$

$$\simeq \frac{\sigma_y^2}{n} + \left(\frac{1}{N} + \frac{1}{n}\right)\left\{\mu_2(\sigma_{bb} + \beta^2) + (\mu_4' - \mu_2'^2)(\sigma_{cc} + \gamma^2) + 2(\mu_3' - \mu_1'\mu_2')(\sigma_{bc} + \beta\gamma)\right\} \quad \cdots (3.63)$$

研究中、親切な援助と助言を与えられたC. R. Rao 博士にお礼を申上げる。さらに原稿の作成と有益な示唆を与えられたMatthai氏に謝意を表する次第である。

附　録　A

$\langle\langle S_{ij} \rangle\rangle$ を母集団の標本変動行列とし、$\langle\langle C_{ij} \rangle\rangle$ を対応するその逆行列とする。このとき次の結果をうる。

$$\left.\begin{aligned} E(C_{ij}) &= \frac{\sigma^{ij}}{n-k-2} \quad (i \neq j) \\ E(C_{ii}) &= \frac{\sigma^{ii}}{n-k-2} \end{aligned}\right\} \quad (A.1)$$

（証明）　行列 $\langle\langle C_{ij} \rangle\rangle$ の要素 C_{ij} から出発して

$$E(C_{kk}) = E(S^{kk}) = E\left\{ \begin{vmatrix} S_{11} & S_{12} & \cdots & S_{1,k-1} \\ S_{21} & S_{22} & \cdots & S_{2,k-1} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k-1,1} & S_{k-1,2} & \cdots & S_{k-1,k-1} \end{vmatrix} \div \Delta \right\}$$

ここで

$$\Delta = \begin{vmatrix} S_{11} & S_{12} & \cdots & S_{1k} \\ S_{21} & S_{22} & \cdots & S_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k1} & S_{k2} & \cdots & S_{kk} \end{vmatrix}$$

よって

$$E(C_{kk}) = \frac{1}{n} E\left\{ \frac{(t_{11} t_{12} \cdots t_{k-1,k-1})^2}{(t_{11} t_{22} \cdots t_{kk})^2} \right\}$$

$$= \frac{1}{n} E\left(\frac{1}{t^2_{kk}} \right)$$

ここでも直角座標(rectangular coodinate)を表わす
(Mahalanobis, Bose and Roy (1937))
他のtを積分して了ったときの t_{kk} の分布は

$$\text{Const}\ e^{-\frac{n}{2}\left\{ T^{kk} t^2_{kk} \right\}} (t^2_{kk})^{\frac{n-k-1}{2}} dt_{kk}$$

で、この積数は

$$\text{Const} \int_{-\infty}^{\infty} e^{-\frac{n}{2}\left\{ T^{kk} t^2_{kk} \right\}} (t^2_{kk})^{\frac{n-k-2}{2}} d(t^2_{kk}) = 1$$

で与えられる。

$$\therefore E\left(\frac{1}{t^2_{kk}} \right) = \text{Const} \int_{-\infty}^{\infty} e^{-\frac{n}{2}\left\{ T^{kk} t^2_{kk} \right\}} (t^2_{kk})^{\frac{n-k-2}{2}} d(t^2_{kk})$$

$$= \frac{\left(\frac{n}{2} T^{kk} \right)^{\frac{n-k}{2}} \Gamma\left(\frac{n-k-2}{2} \right)}{\Gamma\left(\frac{n-k}{2} \right) \left(\frac{n}{2} T^{kk} \right) \left(\frac{n}{2} T^{kk} \right)^{\frac{n-k-2}{2}}}$$

$$= \frac{n}{n-k-2} T^{kk}$$

$$= \frac{n}{n-k-2} \sigma^{kk}$$

すなわち

$$E(C_{kk}) = \frac{\sigma^{kk}}{n-k-2}$$

で一般に

$$E(C_{ij}) = \frac{\sigma^{ij}}{n-k-2}$$

更に

$$E(C_{k,k-1}) = E\left\{ \begin{vmatrix} S_{11} & \cdots & S_{1,k-2} & S_{1k} \\ S_{21} & \cdots & S_{2,k-2} & S_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ S_{k-1,1} & \cdots & S_{k-1,k-2} & S_{k-1,k} \end{vmatrix} \div \Delta \right\}$$

$$= -\frac{1}{n} \left\{ \frac{t_{k-1,k} (t_{11} t_{22} \cdots t_{k-2,k-2}) (t_{11} t_{22} \cdots t_{k-1,k-1})}{(t_{11} t_{22} \cdots t_{kk})^2} \right\}$$

$$= \frac{1}{n} E\left\{ \frac{t_{k-1,k}}{t_{k-1,k-1} t_{kk}^2} \right\}$$

$$A_{11} = \frac{n}{2} T^{kk} \qquad A_{12} = A_{21} = \frac{n}{2} T^{k-1,k}$$

$$A_{22} = \frac{n}{2} T^{k-1,k-1} \qquad t_{kk} = x_1,\ t_{k-1,k-1} = x_2,\ t_{k-1,k} = x_3$$

とおく。

このとき t_{kk}, $t_{k-1,k}$ および $t_{k-1,k-1}$ の同時分布は

$$C \iiint \exp -\frac{n}{2} \left\{ T^{kk} t^2_{kk} + (T^{k-1,k-1} t^2_{k-1,k-1} + 2 T^{k-1,k} t_{k-1,k-1} t_{k-1,k} + T^{kk} t^2_{k-1,k}) \right\} \times$$

$$t_{kk}^{n-k-1} t_{k-1,k-1}^{n-2} dt_{kk}\, dt_{k-1,k-1}\, dt_{k-1,k}$$

で与えられる。ここで常数 C は

$$C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp -\frac{n}{2} \{\cdots\} t_{kk}^{n-k-1} t_{k-1,k-1}^{n-k} dt_{kk}\, dt_{k-1,k-1}\, dt_{k-1,k} = 1$$

あるいは t_{kk} で積分して

$$C_1 \frac{\Gamma\left(\frac{n-k}{2}\right)}{A_{11}^{\frac{n-k}{2}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp-\left\{A_{22}x_2^2+2A_{21}x_2x_3+A_{11}x_3^2\right\} x_2^{n-k}\,dx_2dx_3=1$$

すなわち

$$C_1 \frac{\Gamma\left(\frac{n-k}{2}\right)}{A_{11}^{\frac{n-k}{2}}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp-\left[A_{11}\left(x_3+\frac{A_{21}}{A_{11}}\right)^2+A_{33}x_2^2\right] x_2^{n-k}\,dx_2dx_3=1 \quad \cdots\cdots (A\cdot II)$$

で与えられる。

ここで $A_{33}=\dfrac{A_{11}A_{22}-A_{21}^2}{A_{11}}$

$$y_2=x_3+\frac{A_{21}}{A_{11}}x_2$$

$$y_3=x_2$$

とおけば

$$\frac{\partial(y_2-y_3)}{\partial(x_2,x_3)}=1$$

よって (A. II) は

$$C_2 \frac{\Gamma\left(\frac{n-k}{2}\right)}{A_{11}^{\frac{n-k}{2}}} \frac{\Gamma\left(\frac{1}{2}\right)}{A_{11}^{\frac{1}{2}}} \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{A_{33}^{\frac{n-k+1}{2}}}=1$$

すなわち

$$C_2=\frac{\begin{vmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{vmatrix}^{\frac{n-k+1}{2}}}{\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-k+1}{2}\right)}$$

となる。

いま $E\left\{-\dfrac{t_{k-1,k}}{t_{k-1,k-1}t^2_{kk}}\right\}=E\left\{-\dfrac{x^3}{x_2x_1^2}\right\}$

$$=C_1\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{A_{11}^{\frac{n-k-2}{2}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-A_{11}y_2^2-A_{33}y_3^2}\, y_2^{n-k-1}\left(\frac{A_{21}}{A_{11}}y_2-y_3\right)dy_2dy_3$$

$$=C_2\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{A_{11}^{\frac{n-k-2}{2}}}\left\{\frac{A_{21}}{A_{11}}\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-k+1}{2}\right)}{A_{11}^{\frac{1}{2}}A_{33}^{\frac{n-k+1}{2}}}+0\right\}$$

第2項は奇函数で範囲は $-\infty$ から ∞ までであるから

$$=C_2\frac{\Gamma\left(\frac{n-k-2}{2}\right)\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{n-k+1}{2}\right)A_{21}}{(A_{11}A_{33})^{\frac{n-k+1}{2}}}$$

$$=\frac{\Gamma\left(\frac{n-k-2}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)}A_{21}$$

$$=\frac{n}{n-k-2}\,\sigma^{k,k-1}$$

$$=\frac{n}{n-k-2}\,\sigma^{k,k-1}$$

$$\therefore\quad E(C_{k,k-1})=\frac{\sigma^{k,k-1}}{n-k-2}$$

よって一般に

$$E(C_{ij})=\frac{\sigma^{ij}}{n-k-2}\qquad \begin{pmatrix} i=1,2\cdots\cdots,k\\ j=1,2\cdots\cdots,k\end{pmatrix}$$

附　録　B

x が多変量正規分布に従うとき、x_i, $i=1,2\cdots\cdots,k$ のうちに対する重相関係数 $(\rho_{y\cdot i})$ の同時分布は

$$\frac{\Gamma\left(\frac{n}{2}\right)|\sigma^{ij}|^{\frac{n-1}{2}}(\sigma^{yy})^{k/2}\cdot\prod_{i=1}^{k}d\cdot b_{yi}}{\Gamma\left(\frac{n-k}{2}\right)\pi^{k/2}\left|\sigma^{ij}+(b_{yi}+(b_{yi}-\beta_{yi})(b_{yi}-\beta_{yi})\sigma^{yy}\right|^{\frac{n}{2}}}$$

で与えられる。

証明、各 x_i $(i=1,2,\cdots\cdots,k)$ が一定の場合 b_{yi} の多重分布は

$$\frac{|S_{ij}|^{1/2}}{(\sigma_{y\cdot x}\sqrt{2\pi})^k}e^{-\frac{1}{2\sigma_{y\cdot x}^2}\sum_{i=1}^{k}\sum_{j=1}^{k}S_{ij}(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})}\prod_{i=1}^{k}db_{yi}$$

$$\left\{ \text{ただし} \quad \sigma^2_{Y\cdot X} = \frac{1}{\sigma^{YY}} \right\}$$

で与えられる。

一方 Z の分布は Wishart

$$\frac{|\sigma^{ij}|^{\frac{n-1}{2}}}{2^{\frac{(n-1)k}{2}} \pi^{\frac{k(k-1)}{4}} \prod_{i=1}^{k} \Gamma\left(\frac{n-i}{2}\right)} |S_{ij}|^{\frac{n-k-2}{2}} e^{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\sigma^{ij}S_{ij}} \prod_{i=1}^{k}\prod_{j=i}^{k} dS_{ij}$$

に従う。よって Z が変動するときの b_{yi} の分布は

$$\text{Const} \prod_{i=1}^{k} db_{yi} \int\cdots\int |S_{ij}|^{\frac{n-k-1}{2}} e^{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left\{\sigma^{ij} + \frac{(b_{yi}-B_{yi})(b_{yj}-B_{yj})}{\sigma^2_{Y\cdot X}}\right\}S_{ij}} \prod_{i=1}^{k}\prod_{j=i}^{k} dS_{ij}$$

$$= C \frac{\prod_{i=1}^{k} db_{yi}}{|\sigma^{ij}|^{\frac{k}{2}}}$$

で与えられる。

ここで

$$\sigma^{ij} = \sigma^{ij} + \frac{(b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})}{\sigma^2_{Y\cdot X}}$$

$$= \sigma^{ij} + (b_{yi}-\beta_{yi})(b_{yj}-\beta_{yj})\sigma^{YY}$$

また

$$C = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)} \frac{|\sigma^{ij}|^{\frac{n-1}{2}} (\sigma^{yy})^{\frac{k}{2}}}{\pi^{\frac{k}{2}}}$$

附　　録　　C

$x_1, x_2, \cdots\cdots, x_k$ が多変量正規分布に従うときの $b_{k,k-1,12\cdots k-2}$（すなわち x_{k-1} の x_k に対する回帰係数）の分布は

$$\frac{1}{a\beta\left(\frac{n-k+1}{2}, \frac{1}{2}\right)} \left\{1 + \frac{(b_{k,k-1,12\cdots k-2} - \beta_{k,k-1,12\cdots k-2})^2}{a^2}\right\}^{-\frac{n-k+2}{2}} db_{k,k-1,12\cdots k-2} \quad \cdots\cdots (C.1)$$

ここで

$$a = \frac{\begin{vmatrix} \sigma^{kk} & \sigma^{k,k-1} \\ \sigma^{k-1,k} & \sigma^{k-1,k-1} \end{vmatrix}^2}{\sigma^{kk}}$$

証明

$$b_{k,k-1,12\cdots k-2} = \frac{t_{k-1,k}}{t_{k-1,k-1}}$$

$t_{k-1,k-1}$ および $t_{k-1,k}$ の同時分布は

$$\text{Const} \cdot |A|^{\frac{n}{2}} \left\{-\frac{n}{2}\left(T^{k-1,k-1} t^2_{k-1,k-1} + 2T^{k-1,k} t_{k-1,k-1} t_{k-1,k} + T^{kk} t^2_{k-1,k}\right)\right\} t^{n-k}_{k-1,k-1} \, dt_{k-1,k-1} \times dt_{k-1,k} \cdots\cdots (C.11)$$

$$u = \frac{t_{k-1,k}}{t_{k-1,k-1}}, \quad v = t_{k-1,k}\, t_{k-1,k-1}$$

とおくと変換の Jacobian は

$$\frac{\partial(u,v)}{\partial(t_{k-1,k-1}\, t_{k-1,k})} = 2\frac{t_{k-1,k}}{t_{k-1,k-1}} = 2u$$

となる。

(C.11) から u と v の同時分布は

$$\text{Const} \cdot e^{\left\{-\frac{n}{2}\left(T^{kk}u + 2T^{k-1,k} + \frac{T^{k-1,k-1}}{u}\right)\right\}} \times v^{\frac{n-k+2}{2}-1} u^{-\frac{n-k+2}{2}} du\, dv \cdots \quad \cdots (C.12)$$

なることがわかる。

v について 0 から ∞ まで積分し、$T^{ij} = \sigma^{ij}$ および

$$E(b_{k,k-1,12\cdots k-2}) = \beta_{k,k-1,12\cdots k-2} = \frac{\sigma^{k,k-1}}{\sigma^{kk}}$$

なることに注意すると、u の分布は

$$\text{Const} \cdot \left\{1 + \frac{(b_{k,k-1,12\cdots k-2} - \beta_{k,k-1,12\cdots k-2})^2}{a^2}\right\}^{-\frac{n-k+2}{2}} db_{k,k-1,12\cdots k-2}$$

となる。ここで

$$a = \frac{\begin{vmatrix} \sigma^{kk} & \sigma^{k,k-1} \\ \sigma^{k-1,k} & \sigma^{k-1,k-1} \end{vmatrix}^{1/2}}{\sigma^{kk}}$$

で、積分は $-\infty$ と ∞ の範囲で積分すると

$$a\, B\left(\frac{n-k+1}{2}, \frac{1}{2}\right)$$

となることがわかる。

by K.C. Seal (Sankhya Vol. 11. 1951)

文　　献

Bose, C. (1943) Note on sapling error in the method of double sampling. Sankhyā 6. 329-30

Bose C. and Gayen. A. K (1946) Note on the expected discrepancy in the estimation (by duble sampling of a variate in terms of a concomitant variate when there exists a nonlinear regression between the two variates. Sankhyā 8, 73-74

Cochran. W. G. (1939) The use of analisis of variance in enumeration by sampling. J. A. S. A 34, 492-96

Ghosh. B. (1947). Double sampling with many auxiliary variales Cal. stat Assoc. Bull. 1. 91-93

Mahalanbis. P. C. Bose. R. C and Roy. S. N. (1937): Normaligation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions Sankhyā 3. 8-40

Neyman, J (1938): Contribution to the theory of sampling human populations J. A. S. A 33, 101-116

Schumacher. F. X. and Chapman. R. A. (1942): Sampling methods in forestry and range management (186-189)

Snedecor. G. W and King. A. J. (1942) Recent development in sampling for agricultral statistics. J. A. S. A. 37. 99-116

8　プロットの大きさにともなう収量の分散の変動について

On the variation of yield variance with plot size

研究した問題は、プロットの大きさおよび形状にともなう収量分散の変動状態の把握から、収量密度の空間的共分散函数（Spatial coveriance function）を求めることである。数種のプロットについて得られた結果を多少に示してある。またプロットの幾何学的形状に対する収量分散の依存関係についても、非常に小さいプロットおよび非常に大きいプロットの結果から多々で得られた。共分散が距離の大きいところで市法則（Power law）に従う場合に特別な注意を払った。

1　緒　　言

よく知られているように収量分散の観測された変動をプロットの大きさおよび形状について説明するためには、プロット内の任意の二点の収量密度の間に相関が存在するという可能性を考慮することが必要である。（我々は問題を定常的（stationary）な場合だけに限定する。この場合には収量密度の期待値はその地域内で一定である。）更にこの空間的相関は普々二点間の距離が離すにつれて、比較的ゆるやかに即ち指数的というよりは距離の巾函数として減少しなければならないように思われる。

同様な性質は灯き糸の直径、洪水位（flood height）（Feller, 1951）および母集団成本からの距離などの観測で示される。この論文の計算はこれらの場合にもあてはまるものであるが、具体性をもたせるため我々は引続きプロットおよび収量の問題で話をすることにする。（たとえ時として二次元の場合だけに制限されないことを示すために、"プロット"の代りに"領域（region）"なる用語を用いるとしても）。

最も普通に行なわれる計算は、一定の大きさと形状をもつプロッ

トおよび与えられた空間的共分散函数（Spatial covariance function）$\rho(S)$に対して収量の分散を評価するものである．しかし一般にはおそらくプロットの数学的形態の函数としての収量分散の知識からcovariance functionを決定する逆の計算の方が成立ちそうに思われる．この様な方法によれば実験結果を用いて少くとも$\rho(S)$の部分的推定を得ることができるであろう．我々はこの問題を§3で考えることにする．

同様の一つはMellin変換を用いて容易に解くことができる．従って距離の巾函数に従って減少する共分散

$$\rho(S) \sim c|S|^{-\nu} \quad (S\text{が大のとき}) \quad \cdots\cdots (1)$$

は苦労なしに得られる．

観測される多くの共分散函数が(1)の形であることを示す証拠が数多く存在する．(Fairfield Smith, 1938)．収量反動に対するより単純な線型模型は(1)式を予想するものでなく，むしろ一層急速に減少する

$$\rho(S) \sim c|S|^{-\nu} e^{-\lambda|S|} \quad (S\text{が大のとき}) \quad \cdots\cdots (2)$$

なる型のものを要求するという意味においてこれらは一層興味をひくものである．我々はこの問題を§4において更に追求するであろう．

一次元の大きい区間での"収量"の分散はその共分散函数が(1)の形なら漸近的に区間の長さの巾に比例することが容易に証明できる．しかし二次元あるいはそれ以上の次元で類似する結果を得ることはそう容易ではない．この問題も同様に§3および§4で取扱うことにする．

§3 収量の分散に対する公式

我々は平面上の点を表わすのにベクトル座標xを用いる．そうして面積A，収量$y(\Omega)$のプロットもしくは領域Ωを考える（"平面"とか"面積"などの用語を用いるが，この論文の大部分の公式はn次元空間（$n=1, 2, 3 \cdots\cdots$）にあてはまる）．

最終的には距離Sだけ離れた二点になるように収縮する二つの領域Ω_1とΩ_2を考えよう．共分散函数を

$$\rho(S) = \lim_{A_1, A_2 \to 0} \frac{\mathrm{cov}[y(\Omega_1), y(\Omega_2)]}{A_1 \quad A_2} \quad \cdots\cdots (3)$$

で定義し，この極限が存在する場合のみを考えることにする．

$y(\Omega)$の母集団分散を$V(\Omega)$とすると，(3)の定義から

$$V(\Omega) = \int_{\Omega}\int_{\Omega} \rho(x_1 - x_2)\, dx_1\, dx_2 \quad \cdots\cdots (4)$$

が得られる．$\rho(S)$が原点で連続なら(4)から一つの直接的な結論を導びくことができる．

$$V(\Omega) \sim \rho(0) A^2 \quad (\Omega\text{が小なるとき}) \quad \cdots\cdots (5)$$

即ち小さい領域（即ちそれらの次元のすべてにおいて小さい領域）ではVは面積の二乗に比例する．これはすべての状況についてVが面積に比例するとき，空間的相関（spatial correlation）が0の場合と対比さるべきものである．

一次元の場合でΩが長さの線分からなっているときには(4)は

$$V(a) = 2\int_0^a (a-x)\rho(x)\, dx \quad \cdots\cdots (6)$$

のように簡単になる．

Ωかつ座標軸に平行なa，bなる辺を有する平面上の矩である場合に同様にして

$$V(a,b) = 4\int_0^b \int_0^a (a-x)(b-y)\rho(x,y)\, dx\, dy \quad \cdots\cdots (7)$$

となる．(6)の積分は一般的な$\rho(S)$について計算できるが(7)およびその他の二次元の公式は，普通あまり一般的でない形の函数$\rho(S)$に対してしか計算できない．

ここでスペクトル密度函数（Spectral density function）

$$F(w) = \int e^{2\pi i w \cdot x} \rho(x)\, dx \quad \cdots\cdots (8)$$

および固有特性函数（axial characteristic function）

$$G(w) = \int_{\Omega} e^{2\pi i w \cdot x}\, dx \quad \cdots\cdots (9)$$

を導入すれば

(4)式は

$$V(\Omega) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} |G(w)|^2 F(w)\, dw \quad \cdots\cdots (10)$$

と書ける．この公式は往々(4)式よりも便利である．というのは

積分の速度が簡単で種々の波数（wave number）wに対して与えられる重み（weights）が全くはっきりしており、steepest descents法のような近似法を直接用いることができるからである。簡単な形状のプロットに対する函数$G(w)$は直ちに計算できる。即ち巾a, bをもつ矩形に対しては

$$G(w) = \frac{4 \sin(a w_1) \sin(b w_2)}{w_1 \quad w_2} \qquad (11)$$

で、半径a、中心が原点にある円形に対しては

$$G(w) = \frac{2\pi a}{w} J_1(aw) \qquad (12)$$

である。但しJ_1はorder 1 の Bessel 函数である。
またw_1, w_2はwの成分で、wはその絶対値である。

3 逆の問題（The inverse problem）

実用上からは、（4）の関係を逆にして covariance function を多分色々なΩについて観測できると思われる収量分散$V(\Omega)$で表わせれば利用である。この逆関係は、線型および矩形プロットに対応する（5）および（6）の特別な場合には非常に簡単に求められる。これらはそれぞれ、

$$\rho(x) = \frac{1}{2} \frac{\partial^2 V(x)}{\partial x^2} \qquad (13)$$

$$\rho(x,y) = \frac{1}{4} \frac{\partial^4 V(x,y)}{\partial x^2 \partial y^2} \qquad (14)$$

である。しかしこれらと同じように取扱いやすい場合が少しある。

ここでは主として一般的な型の一つに問題を限定することにする。即ち共分散は等方性（isotropic）であると仮定する。従って$\rho(s)$はsのみの函数であるから

$$\rho(s) = \rho(s) \qquad (15)$$

と書ける。相似的にのみ（即ちそのすべての大きさを同じ比で変化させることによって類化することのできる一定の形をしたプロットについてのみを取扱うものと仮定する。与えられたプロットの形状

についてはプロットの大きさをその最も大きい寸法即ちxで規定すると便利である。従って我々は

$$V(\Omega) = V(x) \qquad (16)$$

と書くことにする（このように円形プロットではxは直径に等しくまた矩形プロットでは対角線に等しい）。この場合プロットのそれぞれの形状に対応して、プロット内で無作為に選んだ二点間の距離の分布を表わす次の様な函数$K(s)$が存在する。
即ち

$$V(1) = \int_0^1 K(s) \rho(s) ds \qquad (17)$$

である。こゝでもしすべての一次元の寸法を$x:1$に増加させれば、この積分の類似小要素は、$x^2:1$の割合で増加するであろうから全く一般的に

$$V(x) = x^{2n} \int_0^1 K(s) \rho(xs) ds \qquad (18)$$

Vおよびρの間の関係（18）はMellin 変換を適用できる形であるから

$$\bar{V}(\nu) = \int_0^\infty V(s) s^{\nu-1} ds \qquad (19)$$

$$\bar{\rho}(\nu) = \int_0^\infty \rho(s) s^{\nu-1} ds \qquad (20)$$

$$\bar{K}(\nu) = \int_0^1 K(s) s^{\nu-1} ds \qquad (21)$$

これを用いると（18）は

$$\bar{V}(\nu - 2n) = \bar{K}(1-\nu) \bar{\rho}(\nu) \qquad (22)$$

となる。

我々は§4において変換（19）および（20）が存在し、かつ（22）が成立つようなνの共通的な範囲を考えることにする。普らくの適当な範囲が存在すると仮定して、（22）で与えられる$\bar{\rho}$の式の逆変換をとることによって$\rho(s)$の解を直接求めることができる。実際（18）に対する希望する反転関係（inverse relation）が

$$\rho(x) = x^{-2n} \int_0^1 L(s) V(xs) ds \qquad (23)$$

なら

$$\bar{\rho}(\nu) = \bar{L}(1+2n-\nu) \bar{V}(\nu - 2n) \qquad (24)$$

である。ここで $\bar{L}(\nu)$ は $\bar{K}(\nu)$ と類似をもつものである。(22)と(24)を比較すれば

$$\bar{L}(\nu) = \frac{1}{\bar{K}(\nu - n\pi)} \qquad \cdots\cdots\cdots\cdots\cdots\cdots (25)$$

であることがわかる。

計算を正確に実行し得るためには(21)の積分を求めることができかつ(25)で与えられる $\bar{L}(\nu)$ の反転(invert)できねばならない。これが可能となる小数の場合の一つはどちらかというと非現実的な円形プロットの場合である(x=直径、$n=2$)。円については直接法から

$$\bar{K}(s) = \frac{1}{2}\left[s\cos^{-1}(s) - s^2\sqrt{(1-s^2)} \right] \cdots\cdots (26)$$

を得る。但し

$$\bar{K}(\nu) = \frac{\Gamma(\frac{1}{2})\Gamma\left\{\frac{1}{2}(\nu+2)\right\}}{4(1+\nu)\Gamma\left\{\frac{1}{2}(\nu+5)\right\}} \cdots\cdots\cdots\cdots (27)$$

$$\bar{L}(\nu) = \frac{4(\nu-3)\Gamma\left\{\frac{1}{2}(\nu+1)\right\}}{\Gamma(\frac{1}{2})\Gamma\left\{\frac{1}{2}(\nu-2)\right\}} \cdots\cdots\cdots\cdots (28)$$

しかし $\bar{L}(\nu)$ を反転しようとするとき双々は発散積分を得る。これは $P(x)$ は $V(x)$ だけでなくその微係数 $V^{(j)}(x)$ でも表さるべきであるという事実を示すものとして受け取らねばならない。(13)および(14)の結果をみればこのことは驚くには当らない。もし(23)に対して更に一般な関係

$$P(x) = x^{-2n}\int_0^1 \sum_{j=0}^{\infty} L_j(s)\left[(xs)^j V^{(j)}(xs)\right] ds \cdots (29)$$

を代入すれば

$$\left[s^{\nu+j-1} V^{(j)}(s) \right]_0^{\infty} = o(j-0,1,2,\cdots\cdots) \cdots\cdots (30)$$

なる限り

$$\bar{L}(\nu) = \sum_j (\nu-1)(\nu-2)\cdots(\nu-j)\bar{L}_j(\nu) \cdots\cdots (31)$$

なることがわかる。

いま(28)の $\bar{L}(\nu)$ は

$$\pi^{-1}B\left(\frac{1}{2}, \frac{\nu-3}{2}\right)\left[(\nu-1)-(\nu-1)(\nu-2)+(\nu-1)(\nu-2)(\nu-3)\right] \cdots (32)$$

と書ける。これは(29)および Beta 函数の積分表示を考えれば

$$P(x) = \frac{2x^{-n}}{\pi}\int_0^1 \left[(xs)V'(s) - (xs)^2 V''(xs) + (xs)^3 V'''(xs)\right]\frac{s^{n-3}ds}{\sqrt{(1-s^2)}} \cdots (33)$$

なる関係に対応するものである。

実用的に最も興味のあるのは矩形の場合であろう。しかし $K(s)$ はこの形に対して容易に計算できるのに反して $\bar{K}(\nu)$ はそうでない。

4. 巾法則に従う共分散(Power-law covariances)

まず(22)の関係にたちかえることにする。もし函数 $f(s)$ が正数 s の小さい値および大きい値に対してそれぞれ $O(s^{-\alpha})$ および $O(s^{-\beta})$ ならその Mellin 変換 $\bar{f}(\nu)$ が

$$-\alpha < Re(\nu) < \beta \qquad \cdots\cdots\cdots\cdots\cdots\cdots (34)$$

に対して確かに存在する。そうして変換 $\bar{f}(\nu)$ は $-\alpha$ および β において単純な極(pole)を有する。

いま原点で $P(s)$ が $O(1)$(この点で連続と仮定すれば)で、かつ ν の大きい値に対して $O(s^{-\lambda})$ と仮定すれば $\bar{P}(\nu)$ は

$0 < Re(\nu) < \lambda$ なる範囲で存在する。

小さい s に対しては $K(s)$ は半径 s の n 次元の球の表面積に比例するから、それは $O(s^{n-1})$ である。

一方 $s > 1$ に対しては 0 となる。結局 $\bar{K}(\nu)$ は $(1-n) < Re(\nu) < \infty$ に対して存在する。

故に(22)式の右辺の値はこつとも

$$0 < Re(\nu) < \min(n, \lambda) \qquad \cdots\cdots\cdots\cdots (35)$$

なる範囲で存在する。(22)式は $\bar{K}$ および $\bar{P}$ で $\bar{V}$ を定義しているから、これらの ν の値についてもこの関係が成立つ。よって $\bar{V}(\nu)$ は

$$-2n < Re(\nu) < \min(-n, \lambda - 2n) \cdots\cdots\cdots\cdots (36)$$

で存在し、またこれらの ν の孤立な値では単純な極(Simple pole)をもつ。

故にもし $P(s)$ が s^{-n} あるいは

$$V(x) = O(x^{2n-\lambda}) = O(A^{2-\lambda/n}) \cdots\cdots\cdots (37)$$

より急速に減少しまた $P(s)$ が $s^{-\lambda}$ ($\lambda < n$) と同程度に減少する

なら、小プロットの場合

$$V(x) = 0(x^{2\pi}) = 0(A^2) \quad \cdots\cdots (38)$$

大プロットでは

$$V(x) = 0(x^{\pi}) = 0(A) \quad \cdots\cdots (39)$$

を得る。言いかえれば、Vが小プロットのプロット面積に比例するかあるいはプロット面積よりも急速に増加するかの二つの場合のけじめをつけるのが巾法則（power law）〔$P(S) \sim CS^{-\pi}$〕である。

(22)の関係は反転に用いることができる。もしV(x)が大面積に対してA^{μ}と同様に増加することがわかれば、大きい距離に対してP(S)が$S^{\pi(\mu-2)}$と同程度に減少することが結論できる。

注意しておかねばならないことはこれらの条件はプロットの形が一定の場合にのみ成立するということである。何となれば一般にVは面積の函数であるとともに形の函数でもあるのだから。しかし、P. Fairfield Smith (1938) によって与えられた数つかの結果は、この形に対する依存関係は小さくて、Vは殆んどプロット面積だけから決定し得ることを示しているように思われる。このことは多分プロットの形が特に極端な場合、例えばあまり細長くない様な場合にのみ正しいであろう。Fairfield Smith の結果はまた(37)式に関連する性質について非常に厳密のゆく証明を与えているという意味で興味をひくものである。彼は面積Aのプロット内の単位面積当り収量の分散が面積$A^{-0.74P}$で非常によく近似できることを見出した。指数を$-\frac{3}{4}$でおきかえると共分散函数

$$P(S) \sim \text{Const}\, S^{-\frac{3}{2}} \quad \cdots\cdots (40)$$

に対応する

$$V(A) \sim \text{Const}\ A^{2-\frac{3}{4}} = \text{Const}\, A^{\frac{5}{4}} \quad \cdots\cdots (41)$$

を得る。（勿論これらの等式は無限に小さいAおよびSについては成立しない）。

観測においてよく見られるこの結果は、三つの興味ある可能性の根拠を与えるものである。即ち

(a) 共分散は$S^{-\lambda}$に従って減少するという性質をもつている。

(b) 収量の分散をプロット面積以上の速さで減少させる理論が小である。

(c) 観測された指数λは簡単な有理数である。

これまで上記の型の共分散を導びくようなある地域上の収量の変動を表わす簡単な線型の模型が提出されたことは一度もない。（Whittle, 1954; Heine, 1955 を見よ）。例えば、(x,y)のまわりのすべての点の収量に対して均等に収量密度（yield density）$z(x,y)$を関係づける模型

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 - \chi^2\right] z(x,y) = \varepsilon(x,y) \quad \cdots (42)$$

は共分散函数

$$P(S) = \text{Const}\ .\ SK_1(\chi S)$$

$$\sim \text{Const}.\, S^{\frac{1}{2}}\, e^{-\chi S} \quad (\chi S\text{が大きいとき}) \quad \cdots (43)$$

を導びく。（ここでK_1は修正された Bessel 函数を表わす）

紡糸の直径の変動に対する一次元の問題で、D.R. Cox（未発表と思われる研究についての私信で）は大きいSに対して事実上巾法則（power law）を予想する模型を提案している。

多次元の模型で任意の範囲のSに対して巾法則の共分散を予想する著者の知っているのは騒音（turbulence）のだけである（Batchelor, 1953, P. 122 をみよ）。これらの模型では次元の議論から指数λは簡単な有理数値でなければならないことが示される。

著者の見解では農業研究で観測される巾法則に従う減少を十分説明するこの模正は、Coxの紡ぎ糸の模型および騒音の模型に共通する二つの性質を備えていなければならないように思われる。

即ち

(a) 線型でないこと。

(b) 変量（収量密度（yield density）は空間的座標（spatial co-ordinates）の函数であると同時に、座標の函数としても考えなければならないこと。

である。（Box の場合では "時間" は折り糸の受けた Smoothing の数で不連続である。）

このような模型としては、土壌の傾向（Soil trent）における肥沃度の勾配（fertility gradients）は時の経過とともに拡散過程（diffusion process）のために平滑化（Smoothing）されてしまうことをあげ得るであろう。この過程は一定の値以上の勾配のみが減少するという範囲内では非線型である。

指数 λ の値については当然検討しなければならない。Fairfield Smith の推定した指数 $b=0.749$ が有理数 $\frac{3}{4}$（$\lambda=\frac{3}{2}$ に対応する）に非常に近いということは偶然であろうか？ 確かにこの一例だけから結論を出すのは早計にすぎる。しかし Fairfield Smith は、色々な種類の作物の斉一性試験（uniformity trial）のデーターから推定した b の値をあげている。推定された39個の b の値のヒストグラムでは主な頂点が $b=0.41-0.50$ の間にあり、また次の頂点および上の打切り点（cut-off point）は $b=0.71-0.80$ にある。このことは少くとも $b=\frac{1}{2}$ および $\frac{3}{4}$ なる値（$\lambda=1, \frac{3}{2}$ に対応する）はある方法で区別されることを示している。

更に多数のデーターあるいは理論からの推測が与えられない限りおそらくこれ以上のことは言いえないであろう。

by P. Whittle

参　考　文　献

Batchelor, G. K (1953) The theory of homogeneous turbulance cambridge Univ. press.

Fairfield, Smith. H (1938) J. Agric. Sci, 28, 1-23

Feller (1951) The asymptotic distribution of the range of independent random variables Ann, Math. Stat. 22, 427-32

Moine, V. (1955) Models for two-dimensional Stationary stochastic processes Biometrika, 42. 170-8.

Whittle, P. (1954), on stationary process inthe plane. Biometrika 41, 434-49

9 系統的抽出研究の展望
A Review of the Literature of Sistematic Sampling

要　約

系統的抽出に関する文献は、実際面の必要性とそれらの充足（satisfaction）に関する統計的手法の供給との相互作用を示している。系統的抽出の問題は一般に認められているよりはるかに屢々生じ、また統計的手法の多くも、満足という点から程遠いということから、この事情は一層の進歩に対する強い刺激となっている。

緒　言

この展望の目的は系統的抽出に関する主要論文の見通しを与えることである。この分野におけるサムプリングの大部分の進歩における現代的な性格に照らして、この展望は大体年代順にまとめてある。これによって発表当時の実際的問題の解決を助けるために与えられた手法の、要求に対する相互関係を調べることが可能になる。この点よりすれば、四子分析の方法が心理学者の周辺から発展してきたのと全く同様に、系統的抽出法が林業および土地利用に関する問題から発展してきたことは注目に価する。Stephan の論文（1948）はそれ自身が展望の形になっているから、最初に取り上げるのに都合がよい。

この論文の第2部におけるこつの文章は重要だから次に引用する。

1. "現代のサムプリングの方法は、個体を無作為かあるいはある種の系統的な手続によって選び出すプロセスの上に成立っている"

2. 系統的抽出は単純な規則で場合をある便宜的な順序で数え、標本としては n 番目ごとの「場合」、または測定された間隔でのなにか同様なパターンを用いて「場合」を選出する。

ある種の物理的および工学的科学においては、系統的抽出の形は長い間用いられてきた。他の応用は林業、農業、畜産、および気象学においてなされてきた。しかし経済および社会学の分野の研究の場合 — 例えば Bowley, Caradog Jones, Hilton および Kaier の — には、標本抽出に用いるリストが前もってある順序になっていて、その大部分はランダムかあるいは少くとも全体のある division 内にあるということが根本的な特徴となっている。一方他の分野における対象は、確かにランダムでない順序に配列されていることがわかっている。したがって同様な標本抽出の手続によって現れる二種類の系統的抽出が存在することになるが、この2つの解決は全く異なるものである。経済の分野において Bowley や他の人々の用いた方法は屢々単一無作為をいわれているもので、むしろ層別無作為抽出といった方がよい。この用語の一般的なことは、この展望の最後において簡単にふりかえることにするものの一つである。

Hasel（1938）は林業における数個のサムプリングの文献をまとめているが、その中で標本内に抽出誤差を推定するのに必要な情報が含まれるためのよくいわれている条件を述べている。しかし生育する林木の全例"毎木調査（Cruising"）によって調査され、サムプリングは同時に行なわれる。この標本は普通ある形式の系統的選出でえられるから、抽出誤差の正しい推定のための条件を欠いている。この困難を解決するため Hasel は補助標本を無作為に選ぶことを提案した。しかし彼はこの方法を実行するには費用がかかるといっている。その困難にもかかわらず、彼は管理面および図化の際に非常に都合がよいということから、系統的標本を支持しているが、系統的標本の誤差に対する方法を発展させることが実際上極めて重要であると指摘している。

次の二つの論文は土地利用の分野に関する研究である。Greenleaf（1942）の論文は、土地利用型の分布の標本データーが満足すべきものであるかどうかを推定する問題に transect survey の方法を適用した。"彼は全 transect の intercept が大きい或 transect line の間隔が狭くなるために、平均平方誤差の平方根が小さくなること

を確かめた．これは形式的には抽出比の増加による標本サイズの増大と同じことである．外業調査データーでチェックされたのと同じ色々な間隔のtraverse lineによる推定値の相関から，色々な値数が計算された．これらは"traverseサムプリング"の正確さを予想するためのtraverse精度図表の最初のもの"となった．

Osborne（1942）の論文はその用語の一層厳密な解釈の上に立って，系統的抽出の問題と取り組んだ最初の試みと思われる．色々な種類の林地によって被覆される面積を推定するための系統的ラインサムプリングの背景に対して，彼の論文は無作為および系統的に選ばれた標本からえられた標本推定値の正確度を報告し，後者の抽出誤差を推定する方法を与えている．彼は系統的にとられたデーターを無作為標本の公式で評価すれば，抽出誤差の推定値には必ず偏りが生ずると結論している．これとともに地域をブロックに細分するという解決方法もとれなかった．というのは，このようにしても唯一な層の組を作ることができなかったからである．実際，変量は位置の連続函数と考えられる変化を受けているのである．したがってサムムリングの問題は回帰のあてはめに関するものとなり，観測された値の間に存在するかもしれない内部的な相関を考慮しなければならないことになる．

よってこの初期の段階では，系列相関が問題の中心的な部分を占めていて，空間または時間において順序づけられた系列の一般的分析につながっているということが明らかにされたのである．系統的抽出法の近代的な解釈を過去の厳密ならざる解釈と区別する問題はこの見方に存在するのである

理論の発展

1944年に系統的抽出法を取扱った最初の論文（Madow and Madow, 1944）が発表された．それ以来筆者 William and Silliam Madow の名前は引続く理論の発展の各段階を通じてーきわ目立っていた．この論文において引用されたそれ以前の文献か

Hansen と Hurwitz（1943）による有限母集団からのサムプリングの研究だけであったことは興味がある．

この論文は単一要素の層化および無作為母集団からの系統的抽出を論じている．後の論文（Madow, 1949）では大きさが等しい場合，等しくない場合のclusterに対する理論の拡張を取扱っている．この最初の論文では"サムプリング理論の大きい欠陥は，完全な無作為標本をとるべきかをきめる何等の統計的方法のないことである"と考えられている．系統的抽出の手続は，多くの可能な並び方を特に除外した無作為抽出の手続として定義されている．もう一つの面からは，系統的抽出計画は集落抽出の一形式であるが，母集団における要素の順序の知識が，級内相関を求めるのに用いられるところが異なっているとしている．このためにHansen and Hurwitzの研究が引用されている．このようにして，母集団分散およびある種の系列相関の知識から，系統的標本の抽出誤差を評価できる．3つの基本的な結果が述べられている．

イ　系列相関が正の和なら系統的抽出は抽出分数が大きいという意味で無作為抽出より悪い．

ロ　系列相関が負の和なら系統的抽出は無作為抽出よりも良い．

ハ　系列相関の和が近似的に0なら，二つの抽出手続の間には大きい違いがない．

層別母集団からの系統的標本の数は小さいから，標本平均値が正規分布すると仮定することは難かしい．結局我々は，母集団の要素が，確率変数の単一の観測値であるという考えに従うことになる．この点を理論の展開に完全に組入れたCochran（1946）の論文まで，このような考え方は採用しなかったことに注意せよ．

一つだけの単位の無作為化の系統的抽出の場合，系統的標本の平均値（$\bar{x}$）は母集団平均値の不偏推定値であり，その分散は

$$\mathrm{Var}(\bar{x}) = \frac{\sigma^2}{n}\left(1+(n-1)\bar{\rho}_k\right) \equiv \frac{\sigma^2}{n}\left\{1-2\sum \rho_{km}\right\}$$

であることが示された．ここで ρ_k はそれぞれが要素からなるクラス

の数で、k_mは系列相関のラグ（ずれ）である。またσ^2は母集団分散で、ρ_kは級内相関係数である。筆者は単一の標本から得られた偏りのある分散および系列相関係数の推定値は、2つ以上の標本を用いることのみによつて補正できることを注意している。このような方法は前述のようにHasel（1938）によつても提案されているが、費用がかさむと思われる。（ρ_{mk}の）推定値γ_{km}の偏りは無視できる程度であることを示して、σ^2の推定を解決する方法が数つか与えられているが、これらはかなり複雑である。

層化された場合の系統的抽出については、二つの基本的な形式が考えられている。

(i) 抽出比（または間隔）がすべての層で一定のとき。

(ii) 可変抽出比（または間隔）を用いたとき　標本平均値は層平均値の重みづき平均であるが、分散についての考察から次のような実際的問題が生ずる。

1　異なる層の対応するitemが正の相関をもつときは、抽出の効率以外の考慮が重要でない限り、抽出間隔を一定とすることは得策でない（Madow. W. G. 1949 をも参照）

2　対応するitemが負の相関をもつているときは、抽出間隔を一定とすると抽出分散が小さくなる。

3　可変抽出比を用いた場合には、層間で独立な抽出を行なうと層の対の間の共分散が消えて、平均値の分散の式はよく知られた層化抽出のものとなる。

系統的抽出の分散の無作為抽出のそれに対する比の式が与えられ、ために比して大きく、かつ、

$$\Sigma \rho_{km} < -(n-1)/2 \quad (kn-1) \leq -\frac{1}{2}k$$

なるときには、前者の分散が後者より小さくなることが示された。

次にこの論文はデーターの周期性の効果を考察している。もしデータの周期と系統的標本の層平均値間の相関は± 1であるから、無作為標本の方がよい。しかし、周期が$2k$なら相隣る層は

は位置の差が奇数となるような層については、層間の相関が-1となり、また位置の差が偶数となる層の間では$+1$となるから、系統的標本の分散はおそらく小となるであろう。興味ある岩石は、母集団の型（form）が線型のとき、系統的標本は末知の傾向（trend）の影響を取除くという点で無作為標本よりはるかに有効であるが、層化無作為標本よりは劣つているということである。

この論文の簡潔な説明が筆者の一人（Madow, L. H., 1946）によつて与えられている。この研究方法は集落抽出においてはさらにはつきりしている（Hansen and Hurwitz を見よ）が、その最も単純なのが、集落内で全数調査を行なわないで、唯一つの集落を抽出するものである。系統的抽出計画を採用するために必要な修正は明瞭に述べられている。この論文に与えられた3つの例は、色々な標本サイズに対する系統的標本の分散が標本サイズの増加にともなつてて、単調に減少しないという意味において consistent でないことを示している。したがつて系統的抽出計画の効率は不定だということになる。

理論の次の発展を論ずる前に、3つの論文を検討しておくのが有用であろう。これらはいづれも重要な実際的問題を論じている。

最初の論文（Wadley 1945）は泥炭の母集団の推定一単位面積あたりの数えあげまたは有機体の問題に関するものである。もし予備に空間的な差のあることがわかつていれば、系統的標本は無作為標本より代表性がある。系統的抽出理論の発展の実際面からの必要性と理論的発展の相互作用は次の引用文中に示されている。

"われわれは系統的標本の抽出変動を標準的な方法では正しく推定できない。…… そのような標本の分散について光を与える方法は何かなるものでも有用である。Madow and Madow はこの問題を論じたが、それらの結果はまだ現場での利用に適当な形式では与えられていない。"

次のものは（Fleming and Simmons, 1946）屑肉で系統的抽出を行なう層化母集団からの興味あるサムプリングの例である。

統計的な考察がなされていないのは、その時（1946）は理論におけるそのような発展がまだ十分利用できる程になつていなかつたことによるものである。この種類に属する第3の論文（Bayley and Hammersley, 1946）は、1946年の1月にRoyal Statistical Society, Research Sectionでもたれた時系列における自己相関に関するシンポジウムの討論で筆者が与えた寄与から生れたものである。ここで展望を行なつている幾つかの論文の間につながりをつけるには、Bayley and Hammersleyの論文の第11番目の式は、Madow and Madowの第4の式と形式的に同じだということが示せる。これはまた非常に長い形の式にも適用するのである——これについてはYule and Kendall の Introduction to the Theory of statistics, 14th Edition Page 404の公式17.11および Yule（1945）のJ.R.S.S. 108, 208における時系列についての論文を参照されたい。Bayley and Hammersleyが研究した問題は、対空兵器のテストの連続的記録をとる器械の系統的2段抽出に関するものであつた。

系統的抽出におけるコレログラム

Cochranの論文（1946）は、コレログラムの型およびその系統的抽出におよぼす影響についての最も重要な進歩である——系統的標本の分散がある程度まで系列相関の型に依存することが認識されるであろう。この論文はまた母集団の要素を確率変数と考えることによつて導びかれる期待分散について研究するという考えを展開していることでも重要である。

筆者は相隣る要素のグループ内分散が、グループの大きさとともに増加する、すなわち系列相関が存在するような母集団模型を考察した。この系列相関が正なら、コレログラムの一般的な形は単調減少函数（上に凸）である。平均的にいうと、層化無作為標本の効率は単純無作為標本のそれ以下となることはなく、またその相対的効率は標本サイズの単調増加函数であるが、系統的標本の場合にはこのような一般的結果はえられない。Madowの論文の一部をなすこの結果の確認は、コレログラムが上に凹なら、平均的にいつて系統的標本は任意の標本サイズの層化標本より正確だという点で拡張されている。ついでにいえばコレログラムに対するこのような制限は、実際にありえない程厳格なものでないといえる。Wold（1938）は経済データに線型のコレログラムを用いたしOsborneも林業および土地利用調査で同じ型の函数を用いている。またFisher（1922）は2つの気象観測所における適雨量間の相関の研究に幾分似通つた型のものを用いている。

Cochranは系統的および無作為抽出計画の相対的精度をきめる決定的因子は、系列相関の第1階差でなくむしろ第2階差の方であると述べている。このことはMadow and Madow（1944）の与えた幾つかの不等式の解釈をはつきりさせるものであるが、それでも系統的標本の層化無作為標本に対する相対的効率が、標本サイズの単調函数であるとは推論できない。期待分散についての彼の研究方針にしたがつて、Cochranは無作為、層化無作為および系統的標本の平均値の分散の式を与え、またそれらの相対的効率が式中の系列相関の線型函数とどのように密接な関係にあるかを示した。この論文は

(i) 増加型
(ii) 線型

のコレログラムをもつ母集団についての各種の標本抽出の幾つかの結果で終つている。

こゝらに理論の発展の次の主要段階について述べる前に、系統的抽出法の使用を報告している一群の主要論文に眼を通しておくのが便利である。

最初のもの（Wordskov and Crump, 1948）は畜産の分野での論文である。筆者は

(1) 無作為な日（random day）
(2) 連続した日（consecutive day）

(3) 区画ごとの日 (interval day)

による産卵 — サムプリング方式から得た結果を比較し、この三つの方法間の正確さの差は僅かであると結論した。異なる日について違の卵の産卵数を手エックしなければならないとき（このような事情は大規模な養殖場ではよくある。）は系統的標本（区画ごとの日）の方が僅かに精度がよかった。

このグループに属する次の論文（Finney, 1948）は一般の刊行物および risk of event において屢々言及されている多くの点を強調しているので重要である。立木の材積査定の場合、全数調査に要する時間と労力が大きいため実際的価値の大きい抽出方法を考えることが必要となる。この論文は Hasel (1938) と大体同じように、林業問題に適用した場合のサムプリングの一般的概観と、あとの論文（Finney, 1949）の序論として書かれている。これらの論文の次のもの（Shaul and Myburgh, 1948）は、アフリカの人口の標本センサスを行なうための計画を論じたものである。与えられた抽出誤差に対して必要とされる標本サイズの計算公式を引用しているが、単純無作為抽出を仮定したときのものであることを注意している。そうして層化標本の方が精度がよいと附言している。しかし実際には、それはリストから系統的標本をとり、その結果を無作為標本として取扱うことにより精度を増加させるように計画したのであった。これは明らかに Bowley その他の用いた研究法の現代版であって、我々がこの展望の中で考えてきたような厳格な意味での系統的標本でないことは明らかである。この点については Gray and Corlett (1950) の論文が、これらの方法を幾つかの点で改善していることを注意しておく。

英国学派の寄与

ここにおいてこの国（英国）の統計家による主な寄与を概観することができる — 初期の文献はほとんど北アメリカで発表されている。言語のためこれ以前は研究上の連絡が阻げられていたが、しか

しこの中の一つの結果は今日実用上最も有用な（特に二次元の系統的抽出の分野における）発展の幾つかを含むと考えられる一群の論文である。1948年9月前後の時点で書かれた論文を考える場合は、どれが先かということで順序をつけることは困難である。というのはそれが色々な定期刊行物での出版事情によることが多いからである。この困難を避ける一つの方法は、関連する論文を一まとめにして考えることであろう。そこでまず Yates の研究（1946, 1948 および 1949）に関するものをとりあげることにする。

主要論文は"系統的抽出法"と題して Phil. Roy. Soc. (A) (1948) に発表された。その結果の幾つかおよびもっと一般には論文（Yates, 1946）の一節で簡単に説明されておりまた Sampling Methods for Census and Survey (1949) なる書物に要約と展望が与えられている。最後の書物は僅かの間にこの分野の標準的な教科書となったものである。1948年の Yates の論文は多くの理由のために系統的抽出論の進歩において画期的重要性をもつものであった。これはその名人どが直接的意味で誰れに実際的関心を示し、また計量的な値の系列に対する系統的標本の有用性を与えたということによるものである。しかしこの論文の取扱っているのは計量的データだけではない。というのはその一節で属性 — 二つの値をとる函数の標本として表わされている — の系統的抽出を取扱っているからである。筆者は医学を含めて多くの分野における実験面の応用を引用しているが、ついでにいうと、最近の論文は系統的抽出の問題の所在は認められているが、その解決は尽んどなされていないという感じを受けるであろう。その適例は Carruthers (1949) の論文である。そこで筆者は明らかに気温の読み取りに系統的抽出法を用いることを推奨しているが、しかし毎日一定の時間に単一の観測値をとることから生ずる困難を解決する手段は何もとっていないのである。

Yates は一次元の系統的抽出を取扱い、抽出誤差を推定する一つの方法を展開した。疑作為に配置された系統的標本に固有な偏向

(trend)の誤差を除去するために、彼は前の論文(Yates, 1946)の中で見直して系統的抽出の結果のみからは抽出誤差の完全に信頼できる推定値は作れないことを示し、また補足の証明を与えるために、1/2 および 1/4 間隔で追加標本をとる方法を述べた。系統的抽出の実行は次のような母集団模型について研究された。

(1) 二つの値をとる函数、すなわち属性
(2) 正規分布する資料
(3) 一項の自己回帰函数

Cochran (1946) にならって Yates も期待分散の概念を用いた。彼はまた何もわかっていないかまたは周期性のある資料について系統的標本を用いることの危険性について彼の考え方を強調している。

区分された層に対して二つの函数をサムプリングする場合、属性を有する区分の割合が小さいかまたは1に近い場合、あるいは区分の大部分が抽出間隔に比べて小さいようなときには、系統的抽出と無作為抽出形式のどれを選ぶかということはそう重要でなくなる。しかしもし区間の大部分が抽出間隔より大なら、系統的抽出は大体において他の型の無作為抽出よりも正確と思われる。この中間の場合は更に研究しなければならないが、しかし同様な結果が成立しそうに思われる。抽出対象が正規分布にしたがうときは、垂直的空中写真からのサムプリング――この分野はまだ十分発展していない――において二次元の傾向の成立することが示される。

研究された第3の場合は単一項の自己回帰函数(単純マルコフチェーン)

$$y_{t+1} = b y_t + a_{t+1}$$

のサムプリングである。この種の函数の性質は、連続変量の型とは本質的に異なることが指摘されている。そのコレログラムは指数型で、この場合も Cochran (1946) の論文のときに引用した文献とつながりがある。もし対象が実際に自己回帰であることがわかっていれば、系統的標本の分散を推定することは容易である。

しかしこのようなことは殆んどなく、もし真の自己回帰からのへだたりが多項の変動(long term variation)によるものであれば、その影響は僅かである。しかしもしこのへだたりが短作為変動が加えられたためのものであれば非常に複雑な過少推定が生ずる。この難点を避けるには、部分的(Partial)系統的標本によって補助情報を求めることが提案される。非常に興味ある抽出研究について幾つかの結果が与えられ、またこの分野の抽出理論の実際例が提出される。

英国学派の第2の主な寄与は故A.E. Jonesの論文である。この研究は、1948年5月の不幸かつ悲惨な Dr. Jones の事故により M.G. Kendal (1948)によってうけつがれた。この論文で考えられたのは、連続パラメーターに依存する1次元の有一な母集団からとられた確率変数の平均値を系統的標本で推定することであった。この問題の核心は標本の構成要素を次々の間の間の相関度によって単純な基準で配分する最良の方法を求めることである。この自己回帰模型は Yates' (1948) の研究したものと同様であった。Jones はもしこの相関が非常に小なら、n個の標本点を

$$T/(n+1),\quad 2T/(n+1),\quad \cdots\cdots$$

の距離で配分することを提案した。ここでTは抽出すべきストリップの長さである。一方相関が0.25より大なら最良の配分は距離が

$$T/2n,\quad 3T/2n,\quad 5T/2n$$

のときのものである。どちらの場合でも、最適な分布は点が等距離すなわち系統的配置のときにえられる。

続きの論文で Kendall (1948) はもっと集約なサムプリングと抽出された strip の拡張との間の本質的な違いを指摘し、後者で端の区間の長さにあまり神経質になることは不必要な手数をかけることだと述べている。この注意は Yates (1949) の書物の1つの節にもとり入れられている。

一方これまでにみてきたそれ以前の論文の多くは、抽出率の増加に対して形式的に同一であるところの集約な系統的抽出の結果については、解説されているが、これと相補う抽出された"strip"の拡張は全然ふれられていない。コレログラムが指数型でないようなも

つと一般な場合について、Kendallは同じ距離でとった観測値を同一の重みで考えてえられた結論は誤まりであると述べている。しかしこのために必要とされる一般的方程式から直接的な解がえられるとは思えない。さらにKendallは対象に著しい周期性がある場合、系統的抽出を使用してはならないという強い基準をあげて始めとしている。

Finney (1949)の論文は前年に発表されたもつと一般な論文の続編をなすものである。これは森林調査において精度が高いとされている系統的抽出法を、色々な出所の多数のデータをもとにして吟味検討した研究である。この論文もまた統計的手法の相互作用が関連する問題に導入される仕方のしるしを与えるものである。

最近の進歩

W. G. Madowによるこの分野の研究の次ぎの部分が1949年の終りに発表された。この論文において、彼は前の論文の結果を集落の大きさが等しい場合と等しくない場合の系統的抽出に拡張した。これはまた二次元の系統的抽出についての幾つかの説明が含まれているが、しかしこの問題に関する最も重要な研究はQuenouille (1949)の論文である。集落の大きさが等しい場合の系統的抽出には提案されている方法が二つすなわち

(i) 抽出された集落内の全要素を調べる

(ii) 抽出された集落内で層化および副次抽出を行なう、

あるか、この特別な設計は、人は母集団よりむしろ自然的母集団のサムプリングに適しているように思われる。最も重要な結論はたとえそれが非統計的な理由によるものであっても、集落を系統的に抽出する方法は集落を用いることによって生ずる分散の増加を償えるほど有効なものでないということである。

二次元の系統的抽出の問題については、二つの点が引用されている。(i) city内のブロックに"適当型"に番号をつけてゆくやり方は、ブロック間に相関がある場合には有効な方法でない。

ある系統的標本はcityのブロックの列となることもある。

コレログラムは上に凹となり

(2) グリット上の行および列からなる系統的標本は行または列に沿って肥沃度の傾斜があるときには有効とならないであろう。

しかし上に注意したように、この展望のそして結びともなる論文は、二次元の系統的抽出の問題を取扱う段階を定められるものである。

Quenouille (1949)はその"平面上のサムプリングについて"と題する論文において系統的抽出の問題を実際と非常に近いところまで近づけた。これはそれまで殆んど直観的に取扱われてきたものである。Cochranの論文(1946)の中のある式について、更ね合わせた変動をつけ加える必要のあることを注意してから、彼は無作為抽出の分散に対する比で表わした系統的および層化無作為計画の抽出分散の差をコレログラムとともに図上でどのように用いれば これらの抽出方式の相対的精密を調べることができるかを示した。幾つかの有用な図表を用いて、彼は直線の一方または両方に平行または独立な標本 — の場合は非常に多くの組合せがある — についてだけでなく無作為、層化無作為、系統的抽出方式の二次元サムプリングの問題を説明した。二つの無作為抽出方式の場合には平行に単位をとると母集分散が大きくなる。予想されるように系統的抽出の場合の分散の増加、減少はレログラムの形によってきまる。全領域について函数の式が与えられ、また各座標軸に沿った内部相関の異なる値に対して、4つの典型的抽出方式の相対的精密の表が与えられている。色々な場合を通じての結論は、二次元の独立な系統的抽出は無作為抽出より正確な結果を与えるということである。線型方式の抽出誤差の推定に提案された方法は平面方式の場合にも採用できる

(i) 互いに独立に配置された系統的標本の組、誤差は各ブロック内の系統的標本の分散から計算される。

(ii) 無作為に配置した系統的標本の一組、誤差分散は系統的標本

の一部から計算される。

(iii) 1つの系統的標本を用いて、これを分割して更に間隔の広い系統的標本を幾つかこしらえる。抽出分散は各副次標本の分散から計算する。

これらの3つの方法は、平均値の推定の場合はこの順序に精度がよくなるが、その反面抽出分散の推定にはこの順に偏りが大きくなるから、次第に実用性が低くなる。したがつて方法の選択はある程度まで問題の性質に支配される。

結　論

過去10年の間、系統的抽出を実用に供しようとする直観的試みに対して、多くの理論的裏づけがなされてきた。引続き理論の一層の進歩を促すためには、実際的問題が非常に重要だということが多くの面において強く示されている。しかし、我々は実際には互いに僅かしか類似していないにもかかわらず同一の名称でよばれている2種類の抽出方式があつて、用語の上で根本的な困難が生じていることをみてきた。ある人々は一方を選めて、ある段階で無作為抽出をとり入れない抽出方式はすべて系統的抽出とよぶべきだというかもしれない。これは疑いもなく連続パラメータをもつ母集団からの集落抽出の場合をも含めるものであろう（Kendall J.R.S.S., 109, 230, para. 11）。しかしこれは我々がこの論文でみてきた抽出方式とは明らかに異なるものである。抽出方式の原理的な分類ができるまでにはなお時日を要するものと思われる。

by William R. Buckland

（J. R. S. S. B. Vol 13, 1951 より）

系統的抽出に関する重要文献

Bayley, G. V., and Hammersley, J.M. (1946) "The effective number of 'independent' observations in an autocorrelated time series," J.R.S.S. Suppl. 8, 184

Carruthers, N. (1949), "Accuracy of mean of 'n' temperature observations" Met. Mag., 78, 65*

Cochran, W.G. (1946), "Relative accuracy of systematic and stratified random samples for a certain class of population", Ann Math. stat., 17, 164

Deming, W.E., and Simmons, W.R (1946), "On the design of a sample for dealer inventories", J.A.S.A. 41, 16*

Finney, D.J. (1948), "Volume estimation of standing timber by survey" Forestry, 21, 179*

——— (1949), "Random and systematic sampling in timber surveys" Forestry, 22, 64*

Fisher, R.A. and Mackenzie, W.A. (1922). "The correlation of weekly rainfall" Q.J. Met. Soc., 48, 234.

Gray, P., and Corlett, F. (1950), "Sampling for the Social Survey" J.R.S.S. (A), 113, 150.

Hansen, M., and Hurwitz, W.N. (1943), "On the theory of sampling from finite populations." Ann. Math Stat., 14, 333.

Hasel, A.A. (1938). "Sampling error in timber surveys", J. Agric. Res., 57, 713.*

Jones, A.E. (1948), "Systematic sampling of continuous parameter populations", Biom. 35, 285

Kendall, M.G. (1948), "A continuation of Dr. Jones' paper" Biom, 35, 291

Madow, L. H. (1946), "Systematic sampling and its relation to other sampling designs", J.A.S.A. 41, 204.

Madow, W. G. (1949), "Systmatic sampling: II. A.M.S. 20. 333
——— and Madow, L. H. (1944) "On the theory of systematic sampling: I. A. M. S. 15. 1.
Nordskog, A. W. and Crump, S. L. (1948). "Systematic and random sampling for estimating egg production in poultry", Biometrics. 4. 273*
Osborne, J. G. (1942), "sampling errors systematic and random surveys of cover-type areas", J. A. S. A. 37. 256*
Proudfoot, M. J (1942), "Sampling with transverse transects lines", J. A. S: A. 37. 265*
Quenouille, M. H. (1949), "Problem in plane sampling", A. M. S. 20, 355
Raul, J. R. H. and Myburgh, C. A. L. (1948) "Sampling survey of African population Southern Rhodesia", Population studies, 2. 339*
Stephan, F. F. (1948). History of the uses of modern sampling Procedures", J. A. S. A., 43. 12
Wadley, F. M. (1945), "An application of the poisson series to some problem of enumeration" J. A. S. A. 40, 93*
Wald, H. (1938), A study in Analysis of Stationary Time Series. Uppsala: Almqvist & Wicksells.
Yates, F. (1946), "A review of recent statistical developments in sampling and sampling surveys", J. R. S. S., 109. 12
——— (1948). "Systewatic Sampling, phil. Trans. (A), 241, 347
——— (1949). Sampling Methods for Census and Surveys. London: Griffin & Co.

*印は主として応用を目的とした論文である。

系統的抽出に関する重要文献

Bayley. G. V. and Hammersley. J. M. (1946) "the effective number of 'independent' observations in an autocorreated time series" J. R. S. S Suppl. 8. 184
Carruthers. N (1949) "accuracy of mean of "n" temperature observations" Met. mag. 78. 65*
Cochran. W. G (1946) "Relative accuracy of Systematic and stratified random Samples for a certain Class of population" Ann Math Stat. 17. 164
Deming W. E. and Simmons. W. R. (1946) "On the design of a sample for dealer inventories" J. A. S. A. 41. 16*
Finney. D. J. (1948) "volume estimation of standing timber by survey" Forestry. 21. 179*
——— (1949), Random and systematic Sampling in timber Surveys" Forestry. 22. 64*
Fisher. R. A. and Mackenzie. W. A. (1922) "The correlation of weekly rainfall" J. R. met. Soc 48. 234
Gray P. and Corlett. J. (1950) "Sampling for The Social Survey" J. R. S. S. 113. 150
Hansen M. and Hurwitz. W. N. (1943) "On the theory of sampling from finite populations" Ann Math Stat. 14. 333
Hasel. A. A. (1938) "Sampling error in timber Surveys" J Agric Res 57. 713*
Jones. A. E (1948) Systematic sampling of Continuous

parmeter populations" Biom. 35. 283
Kendall. M. G (1948) "A continuation of Ler Jones' paper" Biom 35. 291
Madow W. G (1949) "Systematic Sampling and its relation to other Sampling designs" J. A. S A 41. 204
Madow W. G (1949) "Systematic sampling : II. A. M. S. 20. 333
— and Madow. L. H. (1944) "on the theory of systematic sampling : I. A. M. S. 15, 1
Nordskog. A. W and Crump. S. L. (1948) "Systematic and random sampling for estimating egg production in poultry" Biometrics. 4. 223*
Osborne. J. G (1942) "Sampling errors of systematic and random surveys of cover-type areas" J. A. S. A. 37. 256*
Proudfoot. M. G. (1942) "Sampling with transverse traves lines" J. A. S. A. 37. 265*
Quenouille. M. H. (1949) "problem in plane sampling" A. M. S. 20. 355
Shrul. J. R. H and Myburgh. C. A. L (1948) "Sample survey of African poplation of Southern Rhodesia", poplation Studies. 2. 339*
Stephan F F (1948). History of the uses of modern Sampling procedures" J. A. S. A. 43. 12
Wadley F. M (1945) "An application of the poisson series to some problem of enumeration" J. A. S. A. 40. 93*
Wold. H. (1938). a study in analysis of

Stationary Time Series. Uppsala: almqvist & wicksells.
Yates F. (1946) "a review of recent statistical developments in sampling and Sampling surveys" J. R. S S. 109, 12
— (1948) "Systematic Sampling. phil Trans (A). 241 347
— (1949) Sampling Methods for Censes and Survey Sandom: Griffin & Co

*印は主として応用を目的とした論文である.

10 系統的抽出の理論 III

中心化された系統的抽出とランダムスタートの系統的抽出との比較

On the theory of systematic sampling III Comparison of centered and random start systematic sampling

1 要約

えられた主な結果は次の通りである。

母集団のコレログラムが単調減少なら、中心化された系統的抽出（centered systematic sampling）はランダムスタートの系統的抽出より効率が高い、これとともに、母集団なら中心化された系統的抽出はランダムスタートの系統的抽出より効率が高いことがおかつたが、しかしこの場合層化無作為抽出の効率はこのいづれよりも高い場合のあることが容易に例証できる。このように、Cochran [1] の証明した、ランダムスタート系統的抽出の効率が層化無作為抽出より高くなるような場合（例えばコレログラムが上に凹で減少する）には中心化された系統的抽出はランダムスタートの系統的抽出より効率が高くなる。

2 緒言 考察した抽出のタイプ

この論文では中心化された系統的抽出方法の理論を考える。よく知られている通り、このような標本抽出の方法は長い間実際的重要性を認められてきたものである。中心化された系統的抽出の理論は端末補正を施したランダムスタート系統的抽出に対しても成立する（Yates [5]）、というのは端末補正の方法によつてランダムスタート系統的抽出は実際上中心化された系統的抽出に帰着するからである。

論証に用いる方法は Cochran [1] および筆者 [3] [4] の以前の論文のものにしたがつたので、証明と記法を簡潔な形で与えてある。

母集団の要素は $X_1, X_2, \cdots, X_N$ で $N = nk$ である。目的は大きさ n の標本から母集団の算術平均値 $\bar{X}$ を推定することである。

ランダムスタート系統的抽出の推定値 $\bar{x}_{sy}$ は、$X_1, \cdots X_k$ の中から等確率抽出法で1要素を選びそのあとの k 番目ごとの要素を標本に含めることにしてえた n 個の要素の算術平均値である。これらの異なる k 個の標本の算術平均値を $\bar{x}_1, \cdots \bar{x}_k$ と書く。ここで $\bar{x}_i$ はその中の要素が X_i となっている標本の平均値である。$\bar{x}_{sy}$ の分散を母集団の要素で表わしてこれを σ^2_{sy} と書く。k が奇数なら中心化された系統的抽出の推定値 $\bar{x}_c$ は $\bar{x}_{(k+1)/2}$ で、k が偶数なら我々は任意に $\bar{x}_c = \bar{x}_{k/2}$ と定義する。（実際には k が偶数なら、人によつて $k/2$ または $(k+2)/2$ どちらかを確率的に選んだり、あるいは要素 $X_{k/2}$, $X_{k+2/2}, \cdots X_{nk/2}$ を選び出す上記の方法の代りに別な形の標本の抽出法をとるかもしれない。たとえば単調な母集団では $X_{k/2}$, $X_{(2k+2)/2}$, $X_{3k/2}$ $X_{(4k+2)/2}$ をとる方が望ましい。我々のいまの目的についていえば、最良の抽出の型を決定することはあまり重要でない）$\bar{x}$ のまわりの $\bar{x}_c$ の平均平方を σ^2_c と書く。

層化無作為抽出の場合には我々は層を構成するため $X_{i+(j-1)k}$, $X_{2+(j-1)k}, \cdots X_{jk}$ を考える。ただし $j=1, \cdots, n$。よつてそれぞれ k 個の要素からなる n 個の層があることになる。我々は n 個の層のおのおのの中の要素が等確率抽出法で選ばれるものとする。この標本平均を $\bar{x}_{st}$ と書き、また母集団要素で表わしたその $\bar{x}_{st}$ の分散を σ^2_{st} と書く。

我々は、母集団の要素が常数と考えられる場合の期待値の記号を E で表わし、また母集団要素が確率変数と考えられる場合のものを $\mathcal{E}$ と書くことにする。

以下においてはコレログラムという用語を多少広義に用いるので、ここで用語について述べておく。$X_1, \cdots X_N$ が順序のついた確率変数列で、添字がちょうど i だけ異なる二つの確率変数の相関係数を ρ_i とするとき（たとえば $\rho_2 = \sigma_{x_1 x_3} / \sigma_{x_1} \sigma_{x_3}$）、この相関からの

みできまるものとすれば、函数 $f(S) = \rho_S$, $S = 1, \cdots, N-1$ は厳々この列のコレログラムとよばれる。通例確率変数は等平均、等分散と仮定される。しかし我々が以下でこの用語を使うときは、$S = |i-k|$ のみに依存する仮定した積 $x_i x_k$ の期待値のみを指すことにする。よって確率変数が等しい平均値をもつなら、我々の論述は普通のコレログラムを指すことになるが、それ以外では、我々の述べる条件は確率変数の平均値の等しいことを仮定しない。

3 単調な母集団

Hotelling と Solomons [2] は量 $z_1, z_2, \cdots z_q$ に対して次の不等式の成立つことを証明した。

$$\frac{q\ (\text{メディアン} - \text{算術平均})^2}{\sum_{i=1}^{q} (z_i - \bar{z})^2} \leq 1 \qquad (3.1)$$

ただしどの項もすべて有限で、分母は0にならないものとする。この場合 q が奇数ならメディアンは普通に定義されるものであるが、q が偶数で、z' と z'' が二つの中央値であれば、メディアンは $z' \leq$ メディアン $\leq z''$ なる任意の値にとることができる。（細かい証明は q が奇数の場合しか与えられていないが q が偶数のときもすぐ証明できる）

母集団が単調ならそれは $\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k$ か $\bar{x}_1 \geq \bar{x}_2 \geq \cdots \bar{x}_k$ のどちらかである。よって k が奇数なら $\bar{x}_c$ は $\bar{x}_1, \cdots \bar{x}_k$ のメディアンで、k が偶数なら $\bar{x}_c$ は $x_{k/2} \leq \bar{x}_c \leq \bar{x}_{(k+2)/2}$ または $\bar{x}_{k/2} \geq \bar{x}_c \geq \bar{x}_{(k+2)/2}$ である。それで (3.1) は

$$\frac{(\bar{x}_c - \bar{x})^2}{\sigma_{St}^2} \leq 1 \qquad (3.2)$$

となる。（$\bar{x}_c - \bar{x}$ は母集団の要素が確率変数でないときの $\bar{x}$ のまわりの $\bar{x}_c$ の平均平方誤差だから次の定理が証明されたことになる。

定理1 母集団が単調なら、中心化された系統的抽出はランダムスタートの系統的抽出より良好である。

仮説母集団が単調で標本が十分大きければ、層化無作為抽出の方が中心化された系統的抽出より効率が高くなる。なぜなら後者の推定値のもつ偏りは標本の大きさが増しても十分な速さで0に収縮しないからである。

しかし実際には母集団が単調なときでさえ中心化された系統的抽出の方が層化無作為抽出より効率が高くなることが多い。なぜこのようなことが起るかを調べるため、σ_c^2 の平均的分散と平均的共分散の項を定義してみよう。

$$\bar{\bar{x}}_j = \frac{1}{k} \sum_{i=1}^{k} x_{i+(j-1)k} \qquad j = 1, 2, \cdots n$$

とおく。そうすると

$$\sigma_c^2 = S + C$$

である。ここで

$$S = \frac{1}{n^2} \sum_{j=1}^{n} (x_{a+(j-1)k} - \bar{\bar{x}}_j)^2$$

$$C = \frac{1}{n^2} \sum_{\substack{j,m=1 \\ j \neq m}}^{n} (x_{a+(j-1)k} - \bar{\bar{x}}_j)(x_{a+(m-1)k} - \bar{\bar{x}}_m)$$

で、k が奇数なら $a = (k+1)/2$ 、k が偶数なら $a = k/2$ である。我々は S、C をそれぞれ σ_c^2 の平均的分散項、平均的共分散項とよぶことにする。

Hotelling と Solomons の結果から

$$(x_{a+(j-1)k} - \bar{\bar{x}}_j)^2 \leq \frac{1}{k} \sum_{i=1}^{k} (x_{i+(j-1)k} - \bar{\bar{x}}_j)^2 \qquad j = 1, \cdots, n$$

が成立つ。よって $S \leq \sigma_{Sc}^2$ このように $C < \sigma_{St}^2 - S$ なら $\sigma_c^2 < \sigma_{St}^2$ である。実際の場合平均的共分散の項 C は C についての上の条件を満足する位小さい値であることが多い。

4 コレログラムが単調減少であるような母集団（実際には4ノ以下で単調減少と仮定しているようなものをいうのである。）

この節では次の記法が必要である。特に断らない限り、i, j は 1 から k までのすべての整数値をとるものとし、α, β は 1 から k までのすべての整数値をとるものとする。または 1 から $n-1$ まで、s は 1 から $k-1$ まで、それに E は 1 から $\frac{1}{2}(k-1)$ までのすべての整数値をとるものとする。

(証明の中で k は奇数と仮定する。k が偶数のときは基本的な結果は変わらないが一層の複雑さと記法が導入される。) 我々はここで母集団の要素が確率変数と仮定し、また

$$E\, x_{\alpha+(j-1)k}\, x_{\beta+(m-1)k} = \mu_{(m-j)k+s}$$

とおく。ここで $s=\beta-\alpha$, $j \leq m$ である。よって $-(k-1) \leq s \leq (k-1)$

定理2　上述の条件のもとで

$$(4.2)\qquad E\sigma_{sy}^2 - E\sigma_c^2 = \frac{4}{n^2k^2}\sum_E E \sum_{j=0}^{n-1} (\mu_{jk+E} - \mu_{(j+1)k-E})$$

が成立する。

もし $\mu_1 \geq \mu_2 \geq \cdots\cdots \geq \mu_{nk-1}$ で不等号が1つでも成立てば中心化された系統的抽出はランダムスタートの系統的抽出より効率がよい。しかし $\mu_1 \leq \mu_2 \leq \cdots\cdots \leq \mu_{nk-1}$ で不等号が1つでも成立てば逆が真である。

定理2の証明の前にその意味を少し考えてみよう。実際、母集団の要素がそれらの距離と無関係に同一の積の期待値 (4.1) をもつ、すなわち、$\mu_1 = \mu_2 = \cdots\cdots = \mu_{nk-1}$ なら (4.2) から $E\sigma_{sy}^2 = E\sigma_c^2$ が得られる。もし母集団の全要素の期待値を同じと仮定すれば、上記の命題は期待積 (expected product) よりはむしろ系列共分散 (serial coveriance) についていえるものである。たとえば母集団の要素の期待値が同じで無相関なら $E\sigma_{sy}^2 = E\sigma_c^2$ となる。

上述の、中心化された系統的抽出の効率がランダムスタートの系統的抽出より低くなるような条件は、現実には殆んど充たされない筈である。しかし実際には、コレログラムが不規則なためランダムスタートの系統的抽出の方が、中心化された系統的抽出よりも効率の高くなることがよくある。

証明 (4.2) の証明は面倒だが、困難ではない。まず次の二つの Lemma

Lemma 1　$f(i-k)$ を整数 i および k の某の函数とすると

$$(4.3)\qquad \sum_{i\,k} f(i-k) = k f(0) + \sum_s (k-s)\{f(s)+f(-s)\}$$

である。同じ様にして

$$(4.4)\qquad \sum_{i\,k} f(|i-k|) = k f(0) + 2\sum_s (k-s) f(s)$$

証明は省略する。

Lemma 2　$s = |i-k|$ とおく。そうすると $i \neq k$ なら

$$(4.5)\qquad E\,\bar{x}_i\,\bar{x}_k = \frac{\mu_s}{n} + \frac{1}{n^2}\sum_j (n-j)\{\mu_{jk+s} + \mu_{jk-s}\}$$

で、また

$$(4.6)\qquad E\,\bar{x}_c^2 = \frac{\mu_0}{n} + \frac{2}{n^2}\sum_j (n-j)\mu_{jk}$$

が成立する

証明　ここで $x_{i+(j-1)k}$ を x_{ji} と書く。

$$\bar{x}_i = (1/n)\sum_j x_{ji}$$ だから

$$E\,\bar{x}_i\bar{x}_k = \frac{1}{n^2}\sum E x_{ji}\,x_{jk} + \frac{1}{n^2}\sum_{j<m} E x_{ji}\,x_{mk} + \frac{1}{n^2}\sum_{j>m} E x_{ji} x_{mk}$$

$$= \frac{1}{n}\left\{\mu_s + \sum_j \frac{n-j}{n}\{\mu_{jk+s} + \mu_{jk-s}\}\right\}$$

よって (4.5) は証明された。(4.6) はその特別な場合である。

定理2の証明にもどって

$$\Delta = E\sigma_{sy}^2 - E\sigma_c^2$$

とおくと $\Delta = \frac{1}{k}\sum_i E\bar{x}_i^2 - E\bar{x}_c^2 + 2E\bar{x}_c\bar{x} - 2E\bar{x}^2$

かえられる．Lemma 2によって $E\bar{x}_i^2$ はiと独立だから

$$\Delta = 2E\bar{x}_i\bar{x} - 2E\bar{x}^2$$

をうる．ここで（4.4）と（4.5）からi=Cにとり，hについて平均すると

$$E\bar{x}_i\bar{x} = \frac{1}{nk}\left\{\mu_0 + 2\sum_\gamma \frac{n-\gamma}{n}\mu_{\gamma k}\right\} + \frac{2}{nk}\left\{\sum_E \mu_E + \sum_\gamma \frac{n-\gamma}{n}\left[\mu_{\gamma k+E} + \mu_{\gamma k-E}\right]\right\}$$

さらに

$$E\bar{x}^2 = \frac{1}{k^2}\sum_{i,h} E\bar{x}_i\bar{x}_h$$

で，Lemma 2により $E\bar{x}_i\bar{x}_h$ は $|i-h|$ にのみ関係するからLemma 1によって

$$E\bar{x}^2 = \frac{1}{nk}\left\{\mu_0 + 2\sum_\gamma \frac{n-\gamma}{n}\mu_{\gamma k}\right\} + \frac{2}{nk}\left\{\sum_\delta \frac{k-\delta}{k}\left[\mu_\delta + \sum_\gamma \frac{n-\gamma}{n}(\mu_{\gamma k+\delta} + \mu_{\gamma k-\delta})\right]\right\}$$

である．そうすると

$$\Delta = \frac{4}{nk}\sum_E \frac{E}{k}\left[\mu_E - \mu_{k-E}\right] + \frac{4}{nk}\sum_\gamma \frac{n-\gamma}{n}\sum_E \frac{E}{k}\left\{\mu_{\gamma k+E} + \mu_{\gamma k-E} - \mu_{(\gamma+1)k-E} - \mu_{(\gamma-1)k+E}\right\}$$

である．いま

$$\sum_\gamma (n-\gamma)\left\{\mu_{\gamma k+E} - \mu_{(\gamma-1)k+E}\right\} = -n\mu_E + \sum_{\gamma=0}^{n-1}\mu_{\gamma k+E}$$

でまた

$$\sum_\gamma (n-\gamma)\left\{\mu_{\gamma k-E} - \mu_{(\gamma+1)k-E}\right\} = n\mu_{k-E} - \sum_{\gamma=0}^{n-1}\mu_{(\gamma+1)k-E}$$

である．

よって

$$\sum_\gamma \frac{n-\gamma}{n}\sum_E \frac{E}{k}\left[\mu_{\gamma k+E} + \mu_{\gamma k-E} - \mu_{(\gamma+1)k-E} - \mu_{(\gamma-1)k+E}\right]$$

$$= -\sum_E \frac{E}{k}(\mu_E - \mu_{k-E}) + \frac{1}{n}\sum_E \frac{E}{k}\sum_{\gamma=0}^{n-1}(\mu_{\gamma k+E} - \mu_{(\gamma+1)k-E})$$

となる．これで（4.2）は証明された．$1 \leqq E \leqq (k-1)/2$ だから，$\mu_1 \geqq \mu_2 \geqq \cdots \geqq \mu_{nk-1}$ なら $\mu_{\gamma k+E} \geqq \mu_{(\gamma+1)k-E}$ である．よってコレログラムが単調減少ならΔ≧0でありまたコレログラムが単調減少かつ一定でなければΔ>0である．

5 注

この論文の結果は二次元のサムプリングおよび高次のサムプリングにも容易に拡張できるこれらの問題は次の論文で論ずることにする．

興味あることは，もしも $E\bar{x}_{i+(\gamma-1)k}^2$ がγおよびiに依存でないと仮定すれば，$E\bar{x}_{i+(\gamma-1)k}^2$ について更に仮定を設けないと上記の結果が成立しないことである．

by W. G. Madow A. M. S. 1953 (PP. 105—)

参考文献

1 W. G. Cochran "Rerative Accuracy of systematic and Stratified random Sampling for a certain population" A. M. S. vol 17 (1946) pp. 164—177

2 H. Hotelling and L. M. Solomons "Limits of Measure of sheminess" A. M. S vol 3 (1932) pp 141—142

3 W. G. Madow and L. H. Madow "on the theory of systematic Sampling I" A. M. S. vol 15 (1944) pp 1—24

4 W. G. Madow "on the theory of systematic sampling, II" A. M. S vol 20 (1949) pp. 333—354

5 F. Yates "Systematic sampling" Philos. Trans Roy Soc. London Ser. A. vol 241 (1948) pp. 345—377

II 平面における確率過程について
On stationary process in the plane

空間における定常過程（stationary process）の抽出理論は定常時系列（stationary time series）と全く同じようには論じられない。というのは、時系列（time series）の変量は過去の値のみによって影響されるのに反し、空間的過程はすべての方向の値によって定まるのだから。この点については節2-4において詳論した。多くで展開した推定と検定の理論は§8で小麦とオレンジのデータの一様性について適用した。最後の節は幾つかの特殊な二次元の過程の研究にあてた。

緒　言

圃場実験、森林および収量調査あるいは居住地域（populated area）の標本調査等における地形的相関（topographic correlation）の攪乱効果については長く知られているところであって、二次元の確率過程（two dimentional stochasstic process）に関するこのデータについても認められるものである。物理学者も高次元の過程（例えば乱雲（turbulence）や分子系）を取扱うが、実際には物理学者の方が主要な研究者であつた。

我々の提起した過程が定常（stationary）と考えられるのは少くとも第一次近似としての意味においてである。しかしこの近似は屢々十分満足であるから定常的な過程は研究する価値がある。

二次元の理論の大部分は時系列の研究で用いられるもの、形式的な拡張にすぎないから、出来るだけ簡単にこれらの見通しをつけておこう。しかし上の要約で注意した様に興味のある一つの新らしい側面が現れるので、これを詳細に考察することにする。

一般に二次元で特定の模型（model）を研究する場合には技術的な面の数学的取扱いが難かしいということだけのために、一次元の場合より困難が大きいものである。ある場合には初等函数の範囲を全く逸脱してしまう様にも思える。

我々はこの障害を指摘するであろうが、しかし幸いなことに多くの目的についてはこれらの困難を避けることができる。

多くの応用では全く非決定論的な過程（purely non-deterministic process）を考えるだけで十分であるから、我々はこの種の過程、特に線型自己回帰過程（linear autoregressive process）のみに問題を限定する。

2　ライントランセクト

まずライントランセクト（line transect）（即ち地域上に設定した直線で、観測値はこの直線に沿って等距離にとる）。トランセクトの観測値は時系列の場合と全く同じく、一次元の過程から生成（generate）されていると考えることができる。しかしこの二つの場合には重要な違いが存在する。すなわち時系列ではどの瞬間についても過去と将来の間に自然的な区別があり、またある瞬間における観測値は過去の値のみによってきまる。すなわちこの従属性は唯一方向（すなわちあと向き）のみに進む。しかしトランセクトの場合にはこのような型というような二つの方向の区別はなく、両方の側に対する依存性が存在する。我々は圃場に対するより一般的な二次元の例を考えることができる。圃場の任意の点で施した小量の肥料は、終局的にはすべての方向の土壌の肥沃度に影響を与える。（勿論、例外もありうる：たとえば圃場の傾斜がひどくて、施肥した点より下側の地域のみが影響される場合のように。）

いま観測値と"誤差"の変量をそれぞれ z_t と ε_t（$t=\cdots,-2,-1,0,1,2\cdots$）と書けば、最も簡単な現実的な時系列の変量はおそらく一次の自己回帰（autoregression）

$$z_t = a z_{t-1} + \varepsilon_t \qquad \cdots\cdots (1)$$

であろう。

しかしトランセクトでは模型（1）は退化する場合のもので、一

つの方向についてのみ縦属性が見られる。非退化（non-degenerate）なトランセクト模型の最も簡単なものは、

$$z_t = a z_{t-1} + b z_{t+1} + \varepsilon_t \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (2)$$

であろう。ここで a, b があまり大きくてはならないということは直観的に明らかである。我々は（1）の一方向の自己相関（unilateral autoregression）と区別して（2）を二方向の自己相関（bilateral autoregression）と呼ぶ。一方向型 scheme に対するこのような仮定を極めるとどの様な結果が得られるかはパラメーター（parameter）の推定を考える場合に明瞭になってくる。（1）のパラメーター a は残差平方和（residual sum of square）$\sum_t (z_t - a z_{t-1})^2$ を最小にすれば consistent な推定値が得られるにもかかわらず、（2）の a, b を求めるにはただ、

$$U = \sum_t (z_t - a z_{t-1} - b z_{t+1})^2 \quad \cdots\cdots\cdots\cdots \quad (3)$$

を最小にしただけでは無意味な結果がえられる。このことは z_{t+1} の値と自身 z_t に関係している場合には z_t の条件付（conditional）の平均値と z_{t+1} を考えることは間違っているということから説明できる。形式的には、ε_t から z_t への変換（transformation）の Jacobian が、（1）のような一方向の関係（unilateral relation）と違って（2）の関係では1でないということができる。しかし最小二乗推定の正しい方程式は ∂U を最小にすれば得られるということが示される。（これは多少でもっと一般的な二次元の場合について証明する）ここで k は

$$\log k = -\frac{1}{2\pi} \int_0^{2\pi} \log (a e^{iw} + b e^{-iw})(a e^{-iw} + b e^{iw}) du \quad \cdots\cdots$$

で与えられるパラメーターのある種の函数である。

3 二方向（bilateral）scheme の形式的な性質

二次元の過程を論ずる準備としてトランセクト模型の性質を形式的に調べてみよう。

一般的な二方向の線型自己相関（bilateral linear autoregression）

$$L(T) z_t = \varepsilon_t \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (5)$$

を考えよう。ここで $L(T) = \sum a_j T^j$ で T は変換の作用素（translation operator）

$$T z_t = z_{t+1}$$

である。

（5）式の解は

$$z_t = \frac{\varepsilon_t}{L(T)} = \sum b_j \varepsilon_{t+j} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (7)$$

で与えられる。ここで b_j は $\{L(e^{iw})\}^{-1}$ の Fourier 展開の e^{ijw} の係数である（Bartlett, 1946, P 60 を見よ）。Fourier 変換を用いれば、scheme（5）のスペクトル函数（spectral function）は簡単にして

$$F(w) = \frac{\sigma^2(\varepsilon)}{L(e^{iw}) L(e^{-iw})} \quad \cdots\cdots\cdots\cdots \quad (8)$$

であることがわかる（Doob, 1944, Daniell, 1946）自己共変動（autocovariance）

$$\phi(j) = \mathrm{cov}(z_t, z_{t+j}) \quad (j = 0, \pm 1, \pm 2, \cdots) \quad \cdots\cdots \quad (9)$$

は $F(w)$ の Fourier 展開の係数として得られる。したがって、簡単のために、$e^{iw} = Z$, $\sigma^2(\varepsilon) = v$ と書くことにすれば

$$F(w) = \frac{v}{L(Z) L(Z^{-1})} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (10)$$

多くの本質的な点が含まれているから、展開式（7）（8）が成立するための条件と考えることも可能である。（5）が"過去のみに対する従属"を表わす一方向の（unilateral）scheme を表わすとすると $|Z| = 1$ での $\{L(Z)\}^{-1}$ の Laurent 展開にはこの箇所に正べきは含まれてはならないから、時系列自己相関（time series）の安定性（stability）に対する普遍の条件が導かれる。すなわち $L(Z) = 0$ のすべての根は単位円の内部にある（Wold, 1938）しかし、もし二方向（bilateral）scheme を承認する場合に必要なことに、係数が何も制限されていなければ $\{L(Z)\}^{-1}$ の Laurent 展開が収斂することである。この場合の定常性の条件は非常に弱くなっていて、必要なのは、$L(Z) = 0$ の根が何れも単位円上にあ

つてはならないということである。

しかしこれは特定の $L(z)$ が定常であることを保証するための条件にすぎない。

$L(z)$ は (5) なる関係式では従属変量である ε_t であることが望ましい。

(例えば関係式 (1) を $a\varepsilon_t = \varepsilon_{t+1} - \varepsilon_t$ と変形したときの ε_{t+1} ではなく) これには $L(z)$ に乙のある整数乗をかければよい (あるいは同じことであるが ε-系列 (sequence) を転移 (translation) によって再定義すれば ε_t は z_t と対応がつく) というのは、z が単位円周上を正の方向に一まわりすれば $L(z)$ は複素平面上で原点のまわりを k 回まわる (ここで k は整数でなければならない) したがって ε-系列を k 段階遅って転移 (translation) すれば正規化 (normalization) できる。ゆえに $L(T)$ を正規化作用素 (normalized operator) $T^{-k}L(T)$ 即ち $L^*(T)$ でおきかえることができる。原点は $\log L^*(z)$ の branch point でないから (k の0でないときの $L(z)$ の場合と同じく) $\log L(z)$ は z の Laurent 級数に展開でき、また作用素 $L^*(T)$ は

$$L^*(T) = e^{\sum c_j T^j} \qquad \cdots\cdots (11)$$

と表わされる。

これらは、この正規化が行なわれているものと仮定する。したがって (11) は常に可能である。

4 推定の非確定性

二方向の (bilateral) scheme を承認すると、Wold (1938) が移動平均 (moving average) について示したの同様な困難が自己相関を求める際に生ずる。これから、オーダーpの自己相関から生成される一連の自己共変動が与えられたものとすれば自己相関をきめることができる。すなわち (5) の多項式 $L(T)$ をきめることができる。ここで (10) を満足する $L(z)$ を求めることが必要である。ここで $F(\omega)$ は与えられた自己共変動で決定される。このような $L(z)$ のとり方は、2^p 通りある。何故なら特定の $L(z)$ の根を $\alpha_1, \alpha_2, \cdots \alpha_p$ とすれば $F(\omega)$ の 2^p 通りの可能な表現

$$F(\omega) = \frac{\text{Const}}{|(z^{\pm 1} - \alpha_1)(z^{\pm 1} - \alpha_2)\cdots(z^{\pm 1} - \alpha_p)|^2} \qquad \cdots (12)$$

に対応して 2^p 通りの可能な有限自己相関 (finite autoregression) が存在する。

これらの二つの scheme は一方向であるが、たた違うのは (時間(time)) 軸が正の方向のみに向かっているという点である。

同様な論法はあるデータの組に p 次の自己回帰を最小自乗法であてはめる場合にも適用できる。というのはこの場合にもまた自己共変動から計算を行うからである。これは 2^p 個の あてはめの自己相関の一つから導けるのと同じようにして他からも容易に求めることができる。

例として、もう一度二方向の scheme (z) を考える。この scheme から3の最後の節のようにして正規化されれば、方程式

$$a - z + bz^2 = 0 \qquad \cdots\cdots (13)$$

の一根は単位円の内部にあり他は外部にある。この根を α, β^{-1} と書くことにする。

$\{|\alpha| < 1, |\beta| < 1\}$。係数 A と B を

$$(z - \alpha)(z - \beta) = z^2 + Az + B \qquad \cdots\cdots (14)$$

で定義すれば自己相関

$$\varepsilon_t + A\varepsilon_{t-1} + B\varepsilon_{t-2} = \varepsilon_t \qquad \cdots\cdots (15)$$

を作ることができる。これは一方向であるけれども (z) と同じ自己相関を生ずる。

積分 (4) を計算すれば a, b は

$$\{1 + \sqrt{(1 - 4ab)}\}^{-2} \sum (\varepsilon_t - a\varepsilon_{t-1} - b\varepsilon_{t-2})^2$$

$$\simeq \{1 + \sqrt{(1 - 4ab)}\}^{-2} \{(1 + a^2 + b^2)C_0 - 2(a + b)C_1 + 2abC_2\} \cdots$$

$$\cdots (16)$$

を最小にすれば推定されることがわかる。

ここで C_s は $\log s$ で観測された自己共変動であるパラメーターAおよびBを変換すれば（16）式は

$$(1+A^2+B^2)C_0+2(A+AB)C_1+2BC_2 \approx \text{Const.}\ \Sigma(\xi_t+A\xi_{t-1}+B\xi_{t-2})^2 \cdots (17)$$

に比例することがわかる．これは *Scheme*（2）と（15）が同等であることを他の面から表わすものである．この同一性の実際的な意義は取扱いにくい（16）式を最小にする代りにより簡単な（17）式を最小にすればよいということにある．したがってaおよびcの推定値は（14）の関係を用いて，A：Bのそれから計算することができる．

二方向型の *scheme* を前景的に一方向型に帰着させうるような場合には，そのような *scheme* を導入する必要はないように思えるかもしれない．実際このやり方は，我々が二方向の機型にどれ位多くの面倒かつ不確実なパラメーターがあるかということを考える場合には確かに望ましくないことがわかる．しかし二次元の場合には，いま考えた一次元の場合と違って，一方向の *scheme* に帰着させるには非常に多くの面倒な問題が生ずることがわかるであろう．このようにすべての方向への従属性を正確に導入することは不可避的である．またそれ以上の点が期待している．すなわち実際に広く対応する多方向の *scheme* は，たとえそれが推定値等の形式的な作業のような場合でも同等な一方向の機型を用いて行えば最も簡単である．

5　二次元の過程の一般性

平面上のいかなる点においても，ある値となる様な数量の連続的な過程を考えることができる．それは普通，数量が距離の平面上の格子でのみ観測される様なものである．我々は主として数学的により簡単でまたより実際的興味の多い離散的な場合を取扱うことにする．

観測値および誤差数量をそれぞれ ξ_{st} および ϵ_{st} $(s,t=\cdots-2,-1,0,1,2\cdots)$ で表わす．連続過程を論ずる場合，この記号は $\xi(x,y)$ および $\epsilon(x,y)$（x,y はすべて実数とする）に変える．

我々がこれから主として取扱う特別な機型は二次元の線型自己回帰

である．これを我々は

$$L(T_s, T_t)\,\xi_{st} = \epsilon_{st} \cdots (18)$$

と書くことにする．ここで T_s，T_t は

$$T_s\,\xi_{st} = \xi_{s+1,t} \qquad T_t\,\xi_{st} = \xi_{s,t+1} \cdots (19)$$

および

$$L(T_s, T_t) = \sum_j \sum_k a_{jk}\, T_s^j\, T_t^k \cdots (20)$$

で定義される．転移作用素（translation operator）である．方程式（9）－（10）に対応して

$$\xi_{st} = \frac{\epsilon_{st}}{L(T_s\ T_t)} = \sum_j \sum_k b_{jk}\,\epsilon_{s+j,\,t+k} \cdots (21)$$

$$F(w_1,w_2) = \frac{2\pi}{L(z_1,z_2)\ L(z_1^{-1}, z_2^{-1})} \cdots (22)$$

$$= \sum_j \sum_k \phi(j,k)\, z_1^j\, z_2^k \cdots (23)$$

をうる．

ここで b_{jk} は $\{L(z_1,z_2)\}^{-1}$，$(z_1=e^{iw_1},\ z_2=e^{iw_2})$ の Fourier 展開の $z_1^j\ z_2^k$ の係数，ひは ϵ の分散，$\phi(j,k)$ は ξ_{st} と $\xi_{s+j,t+k}$ 間の共変動（covariance），また $F(w_1,w_2)$ は（23）によって定義される我々に対応するスペクトル函数である．

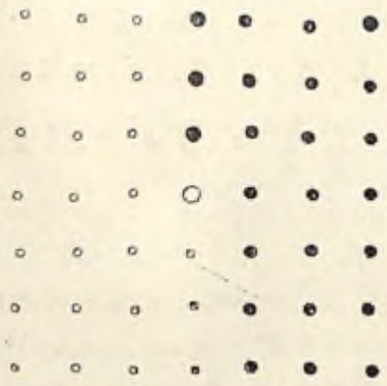
前と同じく，（21）と（22）が成立つための必要かつ十分な条件は，任意の z_1，z_2 に対し $L(z_1,z_2)$ が0でなく，同時に $|z_1|=1$，$|z_2|=1$ を満足することである（少くとも自己回帰（18）が有限であるときに）．

云いかえると，もしくは z_2 の何れかを単位円周に沿って動かすとき，複素平面の原点が $L(z_1,z_2)$ の円内にはいらない様に作用素（operator）L が正規化されていると仮定する．（もし z_1 が単位円の外側をまわり，z_2 が特定の値をとるとき，L が原点とその円内に含まなければ任意の他の z_2 の値に対してもその様な事は起らない．何となればもしその様なことが起れば仮定に反してしが原点を通って動くような z_2 のある中間的な値がある筈である．

6 二次元過程の一方向模型による表現

今日よく知られているタイプの論法（Wiener による。例えば Wiener, 1949. 978 ）によって、与えられた自己相関の組（その場合 z_{st} は z_{su}（$u<t$）および z_{vw}（$v>s$, w は無条件）上の自己回帰、として表わされる）を生成する過程は唯一通りに求めうることを証明しよう。すなわち Fig 1 の格子において、白い点の値は黒点の値によって表わしうるのである。与えられたコレログラム（autocorrelagram）は完全にどんなものでもよいというわけにはゆかないが、必要条件は非常にゆるやかである。

Fig 1



その条件は自己相関が純非一決定的過程（non-deterministic process）によって生成されるということである。これはとりわけ $\log F(w_1, w_2)$ が Fourier 展開でき、したがって $F(w_1, w_2)$ が

$$F(w_1, w_2) = \exp\left[\sum_j \sum_k \alpha_{jk} z_1^j z_2^k\right] \quad \cdots\cdots (24)$$

のように表わされることを意味する。

ここで函数

$$P(z_1, z_2) = \exp\left[-\left(\frac{\alpha_{00}}{2} + \sum_{k=1}^{\infty} \alpha_{0k} z_2^k + \sum_{j=1}^{\infty} \sum_{k=-\infty}^{\infty} \alpha_{jk} z_1^j z_2^k\right)\right] \quad (25)$$

を定義する。

$$F(w_1, w_2) = \frac{1}{P(z_1, z_2) P(z_1^{-1}, z_2^{-1})} \quad \cdots\cdots (26)$$

だから自己回帰

$$P(T_s, T_t)\, z_{st} = \epsilon'_{st} \qquad (\sigma^2(\epsilon') = 1) \quad \cdots\cdots (27)$$

は要求されたスペクトル函数 $F(w_1, w_2)$ をもつ。その上 $P(T_s, T_t)$ は（25）式での z_1 と z_2 との同じ冪指数の T_s, T_t を含む。ゆえにこの自己回帰は要求された形のものである。（27）が自己回帰として正しく書かれうるためには、次の条件を満たす。即ち $P(e^{iw_1}, e^{iw_2})$ は Fourier 展開が可能ということである。

（27）式はトランセクトの場合の一方向表示に対応する。しかしそれは便利なものではないから、推定のような問題については一般にもとの模型について研究する方がよい。1つの点で二次元の場合には一次元の場合と同じことが成立しない。すなわち有限自己回帰（finite autoregression）の一方向模型による表現はまた有限自己回帰であるということである。

このことは次の簡単な例で確かめられる。有限自己回帰

$$(1+B^2) z_{st} = B(z_{s+1,t} + z_{s,t+1} + \varepsilon_{s,t-1}) + \epsilon_{st} \quad \cdots (28)$$

は無限な一方向模型による表現

$$\varepsilon_{st} = 2B\varepsilon_{s,t+1} - B^2 z_{s,t+2} - B^2 \varepsilon_{s+1,t+1} + B(1-B^2) \sum_0^{\infty} B^j \varepsilon_{s+1,t-j} + \epsilon'_{st} \quad \cdots\cdots (29)$$

実に一方向模型による表現の実際的利用さは、それがパラメーターの変換（change）を単純化するということである。（同様なパラメーターの変換は普通積分（∠）の計算から、あるいはその二次元での同等性から示されるものではあるが）二次元模型の大部分ではそれが明らかなものであっても、非常に複雑で、それをやる価値はあまりない。例えば模型

$$\varepsilon_{st} = \alpha(\varepsilon_{s+1,t} + \varepsilon_{s-1,t} + \varepsilon_{s,t+1} + \varepsilon_{s,t-1}) + \epsilon_{st} \quad \cdots\cdots (30)$$

の一方向模型での表現を計算すれば、$\log\{1-\alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1})\}$

のFourier 展開の係数を計算しなければならないか．これらは逆向積分で表わすより他に方法がないのである．

7 描出理論

この節においては，離散的課程の一般な推定式，適合度の検定およびパラメーターの推定値の近似的な分散および共変動を導く描出理論を述べる．理論の展開の大要は，多方向性に特別な注意を払わなければならないという事実を別として，前に出した時系列解析（Whittle，1953）と似通っている．考察した特別の過程は絶対一次定的でその $[F(\omega_1, \omega_2)]^{-1}$ は任意の実数 ω_1, ω_2 に対して0とならないようなものである．したがってこの過程は自己回帰として表現可能である．このように最尤基準で得られた推定式は，変量が正規分布をしない場合には最小二乗推定方程式と考えることができる．しかしこのような場合に我々の結果が成立する限界はより広んだ研究の課題である．

mn 個の観測値 ξ_{st}（$s=1,2,\cdots m$；$t=1,2,\cdots,n$）の組があるものとする．そうするとラグ（lag）j,k の経験的共変動は

$$C_{jk}=\frac{1}{mn}\sum_{s=1}^{m-j}\sum_{t=1}^{n-k}\xi_{st}\,\xi_{s+j,t+k} \qquad \cdots\cdots\cdots (31)$$

である．我々は端末効果（end-effect）は一様に無視することにするから，（31）では $(m-j)(n-k)$ で割る代りに mn で割った．更に次の量

$$f(\omega_1,\omega_2)=\frac{1}{mn}\left\{\left[\sum\sum\xi_{st}\cos(s\omega_1+t\omega_2)\right]^2+\left[\sum\sum\xi_{st}\sin(s\omega_1+t\omega_2)\right]^2\right\}$$

$$=\sum_{-m}^{m}\sum_{-n}^{n}C_{jk}\cos(j\omega_1+k\omega_2)$$

$$=\sum_{-m}^{m}\sum_{-n}^{n}C_{jk}\,\alpha^{|j|+|k|}\cos(j\omega_1+k\omega_2) \qquad \cdots\cdots (32)$$

を定義する．これは（23）と比較すると経験的スペクトル函数であることがわかる．またこれは実際 Schuster のペリオドグラムこの節の終りで我々は次の基本的な結果を証明することにする．

すなわち，もし変量 ξ_{st} が，平均値0で正規に分布し，スペクトル函数 $F(\omega_1,\omega_2)$ をもつと考えられる型の定常過程から生成されたものであれば，これらの同時的尤度（joint likelihood）は端末効果を無視すれば

$$P(\xi_{st})=\frac{1}{(2\pi V)^{\frac{1}{2}mn}}\exp\left[-\frac{mn}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi}\frac{f}{F}\,d\omega_1\,\omega_2\right] \qquad \cdots\cdots (33)$$

で与えられる。ここで

$$V=\exp\left[\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\log F\,d\omega_1\,d\omega_2\right] \qquad \cdots\cdots\cdots (34)$$

(33)の対数をとればFのパラメーターの最尤推定値は

$$\mathcal{L}=\frac{1}{4\pi^2}\iint\left[\log F+\frac{f}{F}\right]d\omega_1\,d\omega_2 \qquad \cdots\cdots (35)$$

を最小にしてえられることがわかる。変量が正規分布と考えられない場合でも，（35）式を最小とすれば最小二乗推定値がえられる．（Whittle，1953，P.132）

実際は（35）式は推定方程式の最も便利な形を与えるものではない．これはペリオドグラムよりむしろ自己共分散による $\mathcal{L}$ の式

$$\mathcal{L}=\frac{1}{4\pi^2}\iint\log F\,d\omega_1\,d\omega_2+\sum\sum C_{jk}\,C_{jk} \qquad \cdots\cdots (36)$$

によってうることができる．

ここで係数 C は

$$\sum\sum C_{jk}\,Z_1^j\,Z_2^k=\frac{1}{F(\omega_1,\omega_2)} \qquad \cdots\cdots\cdots (37)$$

で与えられる．

いま（18）の形の確率的差方程式（stochastic difference equation）によって生成される過程の特別な場合を考えることにすれば

$$F=\frac{\cdots\cdots\cdots}{L(Z_1,Z_2)\,L(Z_1^{-1},Z_2^{-1})} \qquad \cdots\cdots\cdots (38)$$

（36）にこの関係式を代入すれば

$$\mathcal{L}=\log v+\log k+v/mnv \qquad \cdots\cdots\cdots (39)$$

ここで

$$U = \sum_{1}^{m}\sum_{1}^{n} \varepsilon_{St}^{2} = \sum_{1}^{m}\sum_{1}^{n} \left[L(T_s, T_b)\, \varepsilon_{0St} \right]^2 = \sum_{j}\sum_{k} C_{jk} C_{jk} \cdots (40)$$

および $$\log h = \frac{1}{4\pi^2} \iint \log \left\{ L(Z_1 Z_2)\, L(Z_1^{-1}, Z_2^{-1}) \right\} dw_1 dw_2 \cdots (41)$$

無関係なパラメーター (irrelevant parameter) θ に関する (39) の最小値は $\log (hU/mn)$ であるから L のパラメーターに関して最小とすべき量は hU である。

これは今のようなタイプの研究においては最も有用な結果である。最小二乗推定値は普通の通り、'残差平方和'したパラメーターの函数を最小とすることによって得られる。

(41) から $-\frac{1}{2}\log h$ は $(e^{iw}; e^{iw})$ の二重 fourier 展開の絶対項 (absolute term) と考えられることがわかる。屢々問題となるのはその評価である。

F は未知母数 $\theta_1, \theta_2 \cdots \theta_n$ (θ はこれらの中に含まれている。普通はその値は未知である) このとき $\theta_1, \theta_2 \cdots, \theta_n$ の最小二乗推定値の近似的共分散行列が

$$\frac{2}{mn}\left[\frac{1}{4\pi^2} \iint \frac{\partial \log F}{\partial \theta_j} \frac{\partial \log F}{\partial \theta_k} dw_1 dw_2 \right]^{-1} \cdots\cdots (42)$$

であるという よく知られた論法がえられる。

(Whittle, 1953, p.35; ここでの場合と同様な応用について)

ここでまた P 個のパラメーターにあてはめを行なつたときの hU の最小値を $(hU)_n$ と仮定し、q 個のパラメーターを追加してあてはめたときこの量が $(hU)_{n+q}$ になるものとする。そうすると、最初の ν パラメーターの仮説が正しいとすると

$$\psi^2 = (mn - n - q) \log \frac{(hU)_n}{(hU)_{p+q}} \cdots\cdots (43)$$

は近似的に自由度 q の χ^2 分布をするということを用いて、q 個のパラメーターを補足したことによつてえられた改善を検定できる。この論法も前に用いたものである (Whittle 1952, 1953)

(33) の証明

考えている過程は一方向型による表現 (27) をもつものとする。mn 個の残差 ε'_{St} ($S = 1, 2 \cdots\cdots, m$; $t = 1, 2 \cdots\cdots, n$) の同時的頻数函数 (joint frequency function) は

$$f(\varepsilon') = \frac{1}{(2\pi)^{\frac{1}{2}(mn)}} \exp\left\{ -\frac{1}{2} \sum_{1}^{m}\sum_{1}^{n} (\varepsilon'_{St})^2 \right\} \cdots\cdots (44)$$

である。(27) が同等であるような一次変換を行なつて、次のような ε_{St} の頻数函数の式をうる。

$$P(\varepsilon) \approx \frac{A^{mn}}{2\pi^{\frac{1}{2}(mn)}} \exp\left\{ -\frac{1}{2} \sum_{1}^{m}\sum_{1}^{n} \left[P(T_s, T_b) \varepsilon_{0St} \right]^2 \right\}$$

$$\approx \frac{A^{mn}}{(2\pi)^{\frac{1}{2}(mn)}} \exp\left\{ -\frac{mn}{2} \sum\sum C_{jk} C_{jk} \right\}$$

$$\approx \frac{A^{mn}}{(2\pi)^{\frac{1}{2}mn}} \exp\left[-\frac{mn}{8\pi^2} \iint \frac{f}{F} dw_1 dw_2 \right] \cdots (45)$$

ここで A は (27) の ε_{St} の係数である。(24) (25) から

$$A = e^{-\frac{1}{2}d_{00}} = \exp\left[-\frac{1}{8\pi^2} \iint \log F\, dw_1 dw_2 \right] \cdots\cdots (46)$$

これを (45) に代入して (33) が得られる。

(45) の近似的な等式の性質はもつと正確に説明することができる。(45) の右辺を P' と書けば、得られた関係は $\log h$ と $\log h'$ が近似的に等しいということである。(論未効果の無視によつて exponent の項を無視することになる)

この種の近似は全く十分なものである。その理由は尤度の対数に対する相対的なオーダー n^{-1}, π^{-1} の項を加えても、mn の大きな値では最尤推定値あるいは尤度比の有意点には大きい影響がない。

8 数値例

研究さるべき二組のデータはいずれも我々が論じてきた特別な模型のそのままの例 (picture book example) になつていないが、しかしそれらの理由によつて更に価値あるものであろう。

第1のデータはMercerとHall（1911）が実行した小麦（grainの収量）の一様性試験に関するものである。
これは20×25の矩形に配列された500個の11×10.82㎡ものプロットを含むもので、プロット合計が観測値となっている。相関面（correlation field）の第1象限と第4象限がTable 1に与えられている。
第2と第3象限は

$$\phi(-j,\ -k) = \phi(j, k)$$

なる関係で埋めることができる。ここで気のつくことは、南北軸(S)に沿った相関は東北軸(E)に沿った相関よりもかなり大きいことで、少くともプロットが正方形でないということが多分その説明となるものである。
相関は一般に北西方面より北東方向で高く、はっきりした方向の影響があることを示している。もう一つの注目すべき事実は原点から離れるにしたがって相関が単調に減少しないで最低にまで下ってから再び増加することである。距離が小さい場合にはこのような効果を生み出している近くの相関間の競合を考えることもできるが、しかしこの場合もっと真実らしく思える説明は解析された圃場で屢々見られる"肥沃度の波"が存在するということである。（例えばNeyman, 1952, P 75）
Table 2に考察したさまざまなschemeを、あてはめの係数および対応するk_0, k_Uの値とともに要約した。
k_Uはすべて0.7に近いから、我々は大ざっぱに次のようにいうことができる。すなわち、新しいパラメーターの導入によってk_UがΔだけ減少したなら、（43）式によって

$$3.841 < 500 \log\left(\frac{0.7+\Delta}{0.7}\right), \text{ すなわち } \Delta > 0.0054$$

ならΔは5%水準で有意である。
二つあるいは三つのパラメーターを導入したとき、対応する限界は0.0084と0.0110である。このようにして、仮説1と3は適合度

に顕著な差がない。しかし1と4とは非常に明瞭な差がある。
1と5のような仮説はこの方法では正確に比較することができない。なぜなら何れの仮説も他のものへ特別な場合ではないからである。しかし読者は1のあてはまりは5のそれよりはるかに優れていると言いたいように思うであろう。したがって、この結論はそれぞれを順次もっと一般な仮説Hと比較すれば証明される。
この結果で意外に思える点は簡単な一方向のschemeが対称な二次（second order）scheme 5, 6よりずっとよくデータに適合していることである。この節の論題の大きな部分は、空間的過程は一般に多方向的であるということであったが、我々の最初の例では一方向型が相応である。
このことの理由は、Table 1の相関を再検討すれば明らかとなる。これらの相関はラグが1つ導入されると急激に小さくなる（4軸上で0.52も軸上で0.29に）しかしこれはラグが大きくなるにつれて非常にゆるやかにしか減少しない（二単位のラグに対応する数値は0.41と0.15である）。しかし§9で示すように、5または6のようなschemeのコレログラムは、原点から右の方へなだらかに下降していて事実上原点ではその導函数は0となっている（Cf. Fig 2）。したがって5も6も観測値に適合しないことは当然である。

Table 1　小麦のデータに対する自己相関

t	S = 0	S = 1	S = 2	S = 3	S = 4
−3	0.1880	0.1602	0.1509	0.1276	0.1352
−2	0.1510	0.0234	0.0020	−0.0137	−0.1039
−1	0.2923	0.1853	0.1349	0.0788	0.0878
0	1.0000	0.5252	0.4055	0.3639	0.3561
1	0.2923	0.2354	0.1099	0.1205	0.1399
2	0.1510	0.1285	0.0999	0.0749	0.0859
3	0.1880	0.1935	0.2083	0.2415	0.2284

観測されたコレログラムのこの性質については少くとも二つの説明が可能である。

Table 2　あてはめた模型

模型NO	L (T_S , T_t)	k_1	U	k_1 U
1	$1-0.488\,T_S-0.202\,T_t$	1	0.6848	0.6848
2	$1-0.483\,T_S-0.179\,T_t$	1	0.6940	0.6940
3	$1-0.492\,T_S-0.211\,T_t+0.019\,T_S\,T_t$	1	0.6845	0.6845
4	$1-0.402\,T_S-0.168\,T_t-0.172\,T_S^2-0.092\,T_t^2$	1	0.6564	0.6564
5	$1-0.159\,(T_S+T_S^{-1}+T_t+T_t^{-1})$	1.1240	0.6508	0.7314
6	$1-0.213\,(T_S+T_S^{-1})-0.102\,(T_t+T_t^{-1})$	1.1332	0.6217	0.7045
7	$1-0.488\,T_S+0.030\,T_S^{-1}-0.202\,T_t-0.034\,T_t^{-1}$	0.9843	0.6816	0.6709

速度となる scheme は例えば土地の傾斜、主風の方向等のために一方向的であるかもしれない。更に我々は観測値が成育の点観測値（point observation）ではなく、ある地域上で加え上げ（integrate）られ成育の観測値だということを忘れてはならない。このような加え上げはのラグの自己共変動を他のものよりも大きくさせるであろうから、6のような scheme のコレログラムでこのような歪曲が行なわれれば、観測されたコレログラムとは違ったものとなる。しかしそのような初期を考えることは当面の問題と離れすぎるから、我々は観測値の表面上の値でとり扱うこととしよう。

scheme 1 は距離の増加に伴なう相関の低下、増大を説明しないから、全く不適当である。一次（first order）の項と同方向の二次（second order）の項を入れればかなりあてはまりがよくなる。scheme 4 を見よ。

1～4のような scheme では $k=1$ だから最小二乗回帰法で直接にあてはめが可能である。

しかし5～7のような多方向的 scheme ではあてはめはもつと面倒である。

Uは（40）によれば、直接計算できる（41）から λ をもとめる場合には新らしい問題が生ずる。ここでの計算には他にもつとよい方法があるかもしれなかつたがありふれた級数展開の方法と数値積分を用いた。このようにして scheme

$$\xi_{St} = \alpha\xi_{S+1,t} + \beta\xi_{S-1,t} + \gamma\xi_{S,t+1} + \delta\xi_{S,t-1} + \epsilon_{St} \cdots\cdots (47)$$

については

$\log(k) = \log\{1-\alpha z_1-\beta z_1+\beta z_1^{-1}-\gamma z_2-\delta z_2^{-1}\}$ の absolute term の -2 倍

$$= \sum_{j=1}^{\infty}\sum_{k=0}^{j}\frac{(2j)!}{j![k!(j-k)!]}(\alpha\beta)^k(\gamma\delta)^{j-k} \cdots\cdots (48)$$

$\alpha\beta = \gamma\delta$ なるときこれは

$$\log(k) = \sum_{j=1}^{\infty}\frac{1}{j}\binom{2j}{j}(\alpha\beta)^j \cdots\cdots (49)$$

となる。

展開式（48）（49）は $\alpha\beta$ $\gamma\delta$ が非常に小さい場合には有用である。しかしこれらの値がその最大値 1/16 に近すくとき、収斂が緩漫となる。

Table 3 には $\theta=\sqrt{(\alpha\beta)}$ の幾つかの値に対する（49）の $\log(k)$ の値を示した。これらの値は数値積分で求めたが、変化は非常にゆるやかで十分公正な補間が可能である。

Table 3　$\log(k)$ の修正係数の例

θ	0.00	0.05	0.10	0.15	0.20	0.22	0.25
$\log(k)$	0.0000	0.0076	0.0420	0.1010	0.2028	0.2656	0.4406

Table 2 の模型は単にパラメーターの trial value を代入し、対応する k_1U を計算してあてはめた。そうして改善された値は、放物体（hyperboloid）から近似によつてその最小値の位置をきめるというようにしてこの操作を繰返した。この作業は退屈であるが、

しかし、データそのものが提出された努力を考えればそう不当なものではない。

我々のオヌの例は Batchelor と Reed (1924) が1000本のオレンジについて行った一様性試験 (uniformity trial) から得られたもので樹は20×50の矩形格子に配置されている。圃場の相関 $\rho(S_t)$ は Table 4 に、またS軸に沿った相関は Fig 2 にプロットされている。二つの系列の相関はもとの圃場の高度の対称性を反映して、非常に似通った形態を示している。興味のある点は隣接した S_t もの値を通って動くとき大部分の相関 $\rho(S_t)$ の値はなめらかに変化しているのに、原点では非常にはっきりした不連続性が存在することである。

このように相関面 (correlation surface) は原点で大きい突起をもった広いドーム (dome) のまわりを滑らかに取りまいたような形をしている。これは何々の樹の収量 Y_{St} が

$$Y_{St} = \xi_{St} + \eta_{St} \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (50)$$

のように表わされることを示している。ここで ξ_{St} は滑らかなコレログラムをもつ過程にしたがう。一方異なる η_{St} は互いに(あるいは ξ_{St} と)無相関であるから η のコレログラムはただ原点の"突起"のみからなっている。(50)の自然的な解釈が心に浮んでくる: すなわち ξ は樹のまわりの土地の肥沃度を表わし、多分(30)のような単純な対称的な過程 (symmetric process) にしたがい、また η は個々の樹に固有の結実性を表わすということである。コレログラムを見たとき、読者は土壌と木の変異性は全変動に対してそれぞれ5%および95%の寄与を与えているというかも知れない。しかし前の例と同じく樹はまわりの土壌の肥料を"integrate"するかも知れないということを忘れてはならない。

(30)のように単純な自己回帰をあてはめるのと違って(50)+(30)のような混合した scheme をあてはめるのは非常に面倒である。

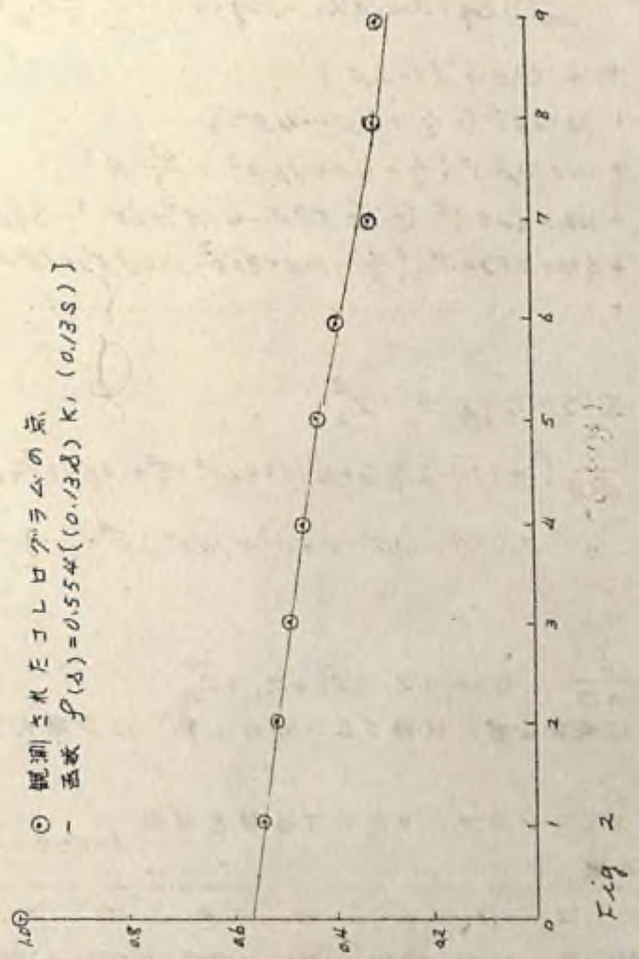


Fig 2

$\mathrm{Var}(\eta) = A \quad \mathrm{Var}(\varepsilon) = B$ とすると

$$F(W_1, W_2) = A + \frac{B}{\{1+\delta(Z_1+Z_1^{-1}+Z_2+Z_2^{-1})\}^2} \quad \cdots\cdots \quad (51)$$

をうる。

推定の目的に必要な量は $\log \nabla$ と F^{-1} を展開したときの係数である。(36を見よ)

我々は

$$\log V = \frac{1}{4\pi^2}\iint \log F\, dw_1\, dw_2 = \log(A+B) + \sum_{j=1}^{\infty} \frac{1}{j}\binom{2j}{j}\alpha^{2j} t$$

$$+ 4(\alpha\beta)^2 [1-2\beta]$$

$$+ 36(\alpha\beta)^4 [-\tfrac{1}{2} + 4\beta - 4\beta^2]$$

$$+ 400(\alpha\beta)^6 [\tfrac{1}{3} - 6\beta + 16\beta^2 - \tfrac{32}{3}\beta^3]$$

$$+ 4900(\alpha B)^8 [-\tfrac{1}{4} + 8\beta - 40\beta^2 + 60\ {}^3 - 32\beta^4]$$

$$+ 63504(2B)^{10} [\tfrac{1}{5} - 10\beta + 80\beta^2 - 224\beta^3 + 256\beta^4 - \tfrac{1024}{10}\beta 5]$$

$$+ \cdots\cdots \qquad (52)$$

および

$$F^{-1} = \sum\sum C_{jk} z_1^j z_2^k$$

$$= \frac{1}{A+B}\{1 + (2\beta-2)S + (1-5\beta+4\beta^2)S^2 + (4\beta - 12\beta^2 + 8\beta^3)S^3$$

$$+ (-\beta + 13\beta^2 - 28\beta^3 + 16\beta^4)S^4 + \cdots\} \cdots (53)$$

をうる。

ここで

$$\beta = \frac{A}{A+B},\quad S = \alpha(z_1 + z_1^{-1} + z_2 + z_2^{-1})$$

この scheme は純粋な自己回帰でないから、F^{-1} は多項式でなく、

Table 4　オレンジのデータに対する自己相関　$s=0\sim9,\ t=-4\sim4$

sの値

t	0	1	2	3	4	5	6	7	8	9
-4	0.3912	0.3902	0.3771	0.3581	0.3110	0.2923	0.2549	0.2546	0.2440	0.2131
-3	.4609	.3956	.3930	.3696	.3484	.3353	.3137	0.2834	.2894	.2870
-2	.4667	.4459	.3982	.3470	.3243	.3225	.3250	.2801	.2721	.2952
-1	.5462	.4469	.4336	.3854	.3880	.3761	.3495	.2914	.2567	.2606
0	1.0000	.5403	.5052	.4840	.4585	.4233	.3960	.3246	.3210	.3150
1	0.5462	.4458	.4280	.3751	.3914	.3622	.3716	.2948	.2958	.2925
2	.4667	.3858	.3799	.3459	.3603	.3475	.3401	.2905	.2766	.2230
3	.4609	.4190	.3716	.3764	.3509	.3336	.3369	.2765	.2433	.2741
4	.3912	.3543	.3882	.3392	.3537	.3446	.3305	.2893	.2685	.2321

また係数 C_{jk} も有限でない。

したがって今の場合、我々は（36）の和 $\sum\sum C_{jk} C_{jk}$ の十分な近似をうるには大体ラグ10までの相関を含まなければならぬことがわかる。しかし相関は僅か $t=4$ までしかえられない。これ以上続けて求めることは非常に困難である。

一般に scheme が純粋な自己回帰なものでなければ観測されたコレログラムが非常に急速に小さくなる場合に限って最小二乗法を採用することができる。今の問題のような場合にはもっとそぼくなあてはめの方法、例えば観測された方程式と理論的自己相関係数を用いなければならないようになる。

一方減少の率がゆるやかだということは、この離散的な模型が連続な模型で近似しうるという利点を有する。これを単純化するのに最も多く用いられる方法である。このように scheme（30）においてαが 0.25 に接近した値なら、これは自己共分散函数（autocovariance function）

$$\rho(r) = \text{const}\ r K_1(Kr) \cdots\cdots \qquad (54)$$

をもつ連続な scheme で近似しうることが§9において示されている。

ここで r は考えられた二点間の距離であって K_1 は修正された Bessel 函数で

$$K = \sqrt{\left(\frac{1}{\alpha}-4\right)} \cdots\cdots \qquad (55)$$

である。コレログラム（Fig 2）の薄い点は常数が r_1 と r_8 で一致するように調整したときの函数（55）の値を示す。$\alpha=0.2489$ のとき $K=0.13$ であることがわかる。この一致は注目に価する。しかしこれは相当割引して考えなければならない。というのは殆んどどの単調減少函数も端点（end-point）さえ一致するようにすれば観測値に相当よくあてはまるものだからである。

しかし例えば同じような型の指数曲線をあてはめるなら、指数曲線は中間であまり下りすぎて全然あてはまりが悪い。

9. 特別な過程

データの組にあてはめを行うとき模型の性質についてそう厳密に知っている必要はない（前節の例を参照）。しかしこれらの性質を知ることは優れた模型をうるための最も近道である。この場合"性質"の内容は、普通コレログラムと同義である。なぜなら与えられたデータの性質を表わす場合はコレログラムを用いることが最も多いからである。

ここで我々は二、三の特殊な模型とそのコレログラムをどちらかというと系統的な形で論じてみよう。というのは厳密な性質の結果も幾つか得ることができるからと思う。

興味ある問題で最も簡単なものは解

$$\varepsilon_{st} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \binom{j+k}{j} \alpha^j \beta^k \epsilon_{s+j,t+k} \quad \cdots\cdots (57)$$

である。

ε_{st} を独立変量とするこの過程が定常であるためには

$$|\alpha| + |\beta| < 1$$

でなければならない（§5 をみよ）

自己共振動は

$$(1-\alpha z_1 - \beta z_2)^{-1} \quad (1-\alpha z_1^{-1} - \beta z_2^{-1})^{-1}$$

から生成され、相関は

$$\rho(s,0) = A^s \qquad \rho(0,t) = B^t$$

$$\rho(s,t) = \left\{ \sum_{j=0}^{s} \binom{s+t-j}{t} \alpha^{s-j} \beta^t A^j + \sum_{k=0}^{t} \binom{s+t-k}{s} \alpha^s \beta^{t-k} B^k \right\} \quad \cdots (58)$$

となる。

$$\left.\begin{aligned} \text{ここで} \quad A &= \frac{1+\alpha^2-\beta^2-\Delta}{2\alpha}, \qquad B = \frac{1+\beta^2-\alpha^2-\Delta}{2\beta} \\ \Delta &= \sqrt{\{(1+\alpha+\beta)(1+\alpha+B)(1-\alpha+\beta)(1-\alpha-B)\}} \end{aligned}\right\} \cdots (59)$$

である

（58）式は $\alpha, \beta > 0$ について成立つ。

Scheme（56）の連続の場合への類推は一次確率微分方程式（first-order stochastic differential equation）となり形式的には

$$\left(\alpha \frac{\partial}{\partial x} + \beta \frac{\partial}{\partial y} + \gamma\right) \varepsilon(x,y) = \epsilon(x,y) \quad \cdots\cdots (60)$$

のように書かれる。

もとの軸に対して再び $\tan^{-1}(\beta/\alpha)$ だけ回転させた新しい座標軸 $X=0$, $Y=0$ に対して平面を回転させると、この関係は

$$\left(\sqrt{(\alpha^2+\beta^2)}\frac{\partial}{\partial X} + \gamma\right) \varepsilon'(X,Y) = \epsilon'(X,Y) \quad \cdots (61)$$

のように書けることがわかる。

ここで $\varepsilon'(X,Y) = \varepsilon(x,y)$ $\epsilon'(X,Y) = \epsilon(x,y)$ である。

すなわち scheme (60) は $\tan^{-1}(\beta/\alpha)$ の方向に沿って走る Markov 過程の系列を考えることができる。これは互いに独立である。このことは一次の模型（first order scheme）の退化（degenerate）する性質を示している。

最も簡単な二次の（second order）scheme（すなわち最も簡単な非一退化 scheme）は系統的自己回帰

$$\varepsilon_{st} = \alpha(\varepsilon_{s+1,t} + \varepsilon_{s-1,t} + \varepsilon_{s,t+1} + \varepsilon_{s,t-1}) + \epsilon_{st} \quad \cdots (62)$$

である。

$$\left[\Delta_s^2 + \Delta_t^2 + \left(\Delta - \frac{1}{\alpha}\right)\right] \varepsilon_{st} = \epsilon_{st} \quad \cdots\cdots (63)$$

のように書かれる連続的関係に対する類比の stochastic Laplace equation は（ここで Δ は central difference operator）

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 - K\right] \varepsilon(x,y) = \epsilon(x,y) \quad \cdots\cdots (64)$$

となることは明らかである。

この連続的な関係は（関係式（50），（60）などの場合と同じく）本質的に技巧的な性質がないため古くから取扱われている。scheme (62) の正確な結果は得られている。（Von der Pol

& Bremmer, 1950; Stöhr, 1950) がこれらは簡単ではない。公式 (Titchmarsh, 1948, P 201) から

$$\frac{1}{4\pi^2}\iint \frac{e^{i(xw_1+yw_2)}\,dw_1\,dw_2}{(w_1^2+w_2^2+k^2)^{\mu+1}} = \left(\frac{r}{2k}\right)^{\mu}\frac{K_\mu(kr)}{\Gamma(\mu+1)} \quad \cdots (65)$$

$$(r=\sqrt{(x^2+y^2)}),$$

であるから scheme (64) は

$$\xi(x,y)=\int_{-\infty}^{+\infty}\epsilon(x+X,\ y+Y)K_1(kr)\,dx\,dy \quad \cdots (66)$$

および

$$\phi(x,y)=\frac{r}{2\pi}K_1(kr) \quad \cdots (67)$$

となる。

ここで K_ν は次元の修正されたオーダー ν の Bessel 函数である。$\lim_{r\to 0} rK_1(r)=1$ だから

(67) に対応する相関係数は

$$\rho(r)=krK_1(kr) \quad \cdots (68)$$

である。

scheme (64) は一般的 (general) second-order 確率差分方程式 (stochastic difference equation) の特別の場合である。通常の楕円的 (elliptic) 放物線的 (parabolic) 双曲線的形式 (hyperbolic form) への分類にしたがえば、(64) は原点を中心とする円形式 (circular form) と考えることができる。残りの second order scheme を研究することは理論的にも実際的にも興味がある。

相関函数 (68) は一次元の場合の指数式 $e^{-\alpha|x|}$ と同じように、二次元での "elementary" 相関と考えられるから興味がある。どちらの相関曲線も単調減少であるが、(68) は原点で平でありまたその減少の率が指数式よりも遅いという点で異っている。

指数式はそれ自身一次元での当然の選択であることが証明されており、また観測された曲線も同様な単調減少を示すという大きな理

由から、二次元のコレログラムを指数式の和で表わすという多くの試みがなされてきた。

しかし前節で明らかなように、二次元の場合 指数式は必ずしもよくないし また最後の節の例によって、K_1 函数は指数型よりも観測値によくあてはまることが示されている。

指数的相関をもつような二次元の過程も作れるか、それは非常に人工的なものである。例えば Matern (1947) は相関函数 $e^{-\alpha\sqrt{(x^2+y^2)}}$ に対応するスペクトル函数は $(w_1^2+w_2^2-\alpha^2)^{-3/2}$ であることを示した。

このようなスペクトル函数をもつ最も簡単な過程は形式的に

$$\left[\left(\frac{\partial}{\partial x}\right)^2+\left(\frac{\partial}{\partial y}\right)^2-\alpha^2\right]^{\frac{3}{4}}\xi(x,y)=\epsilon(x,y) \quad \cdots (69)$$

と書くことができるが このような関係を導びく自然的なメカニズムを具体的に示すことは難しい。

by P Whittle

(Biometrika, 1954, vol 41, P 434 - 449 より)

参考文献

Bartlett, M. S (1946). Stochastic processes. Mimeographed N. Carolina Lecture notes.

Batchelor, S. Le and Reed, K. S (1924) Relation of the variability of yields of fruit trees to the accuracy of field trials. J Agric Res 12. 245-83

Daniell P J (1946) contribution to discussion on stochastic processes J. R Statist Soc B. 8 88-90

Doob J L (1944) The elementary Gaussian processes Am Math Statist. 15. 229-82.

Matern. B. (1947). Metoder att uppskatta noggrannheten vid linje-och provytetaxering Medd. Skogsforskninst. 36 No 1.

Mercer. W. B and Hall. A. D (1911) the experimental error of field trials. J Agric. Sci. 4 107-32

Neyman J (1952) Lectures and Conferences on Mathematical Statistics and probability Graduate School U S Dep of Agriculture.

Quenouille. M. H. (1949) problems in plane Sampling. Ann. Math Statist 20 355-75

Stöhr. A. (1950) Über einige lineare partielle differenzgleichungen mit konstanten Koeffizienten Math. Nachr 3, 295-315

Titchmarch. E. C (1948) introduction to the Theory of Fourier integrals oxford University press.

Van Der pol B. and Bremmer. H (1950) Operational Calculus. Cambridg University press

Whittle. P (1952) Tests of fit in time series Biometrika, 39, 309-18.

Whittle. P (1953) The analysis of multiple stationary time series J. R. Statist. Soc B 15, 125-39.

Wiener, N. (1949) The Extrapolation interpolation and Smoothing of Stationary time-series. New York Wiley.

Williams. R. M. (1954) The choice of sampling interval for systematic samples from population with stationary correlation. (Unpublished paper)

Wold, H. (1938). A Study in the Analysis of Stationary Time Series. Upsala.

綱泉君 注

Whittle の上記論文 (p 59-60) および Patankar の論文 (1954 Biom. pp 450-462) ではともに Mercer & Hall (1911) の小麦のデータを理論の説明に用いている。一見したところでは Whittle の table の (i) $s=0$ の列と (ii) $t=0$ の行の系列相関は Patankar の相関の (i) "行に沿って、もとの全データについての結果" (table 5.2, P462 の最後の列) を (ii) の列) と対応すべきように思われる。しかし著者の確かめたところ、計算された量は異なる定義のものである。すなわち

$i = 1, 2, \cdots\cdots 25$ 東西方向の列を示す。

$j = 1, 2 \cdots\cdots 20$ 南北方向の行を示す。

とすれば Whittle の r_{30} に対する共変動は

$$\sum_{i=1}^{25}\sum_{j=1}^{20-S} X_{ij}X_{i,j+S} - \sum_{i=1}^{25}\sum_{j=1}^{20-S} X_{ij}\times \sum_{i=1}^{25}\sum_{j=1}^{20-S} X_{i,j+S}/\{25(20-S)\}$$

で、分散についても同じである。しかし Patankar の用いた対応する共変動は

$$\sum_{i=1}^{25}\left\{\sum_{j=1}^{20-S} X_{ij}X_{i,j+S} - \sum_{j=1}^{20-S} X_i X_j \times \sum_{i=1}^{20-S} X_{i,j+S}/(20-S)\right\}$$

で 分散でも同じである。

このように Whittle の系列相関は全変動 (total variation) と共変動から計算されているが、Patankar の相関は列 (または行) 変動にもとづいている。予期されるように Whittle の相関は行および列両の変動だから Patankar のものより大きい。

12 系統的標本の平均値の分散
The Variance of the Mean of Systematic Samples

緒言

系統的標本抽出の方法はこれまで色々な分野、特に林業および生態学者によつて用いられてきた。これは、現場での便利さと調査資料を地図作成に用いる場合に無作為抽出法より優れていることによるものである。これらの用法は Hasel（1938）や Osborn（1942）などの研究によつて正当化されたものである。彼等は無作為、層化無作為および系統的調査の抽出誤差を、これらの三方法で分析するためにとられた詳細な森林調査のデータについて研究した。これらの分析が、系統的抽出の効率が一般に最も高いということが示された。Osborne の論文では実験結果と妥当な一致を示す平均平方系列相関（mean square serial correlation）を用いる一つの方法が与えられてはいるが、大体において彼等は系統的標本の誤差を推定する問題を未解決のまゝ残している。

ごく最近 Finney（1948）も森林調査のデータを用いて、系統的標本（妥当な精度を得るために群の重ね合せ法（overlapping method）によって計算した）と無作為標本、および層当り一つまたは二つの要素をもつ層化無作為標本の分散を比較した。これらの場合において Finney は系統的標本の分散は層当り1個の観測値をもつ同じ大きさの層化無作為標本の分散と少ししか違わないが、層の数を半分にして層当り2個の観測値をとつた層化標本で得られる分散よりは非常に小さくなることを示した。これは補足的な情報なしで分散の不偏推定値を与える最も有効な方法であると思われる。色々な模型（model）に対する系統的標本の抽出分散の更に進んだ研究が有用であるように思われる。

Yates（1948）は多数の特別な場合について抽出分散を研究し、補助情報がある場合に系統的標本の平均値の分散を与える式あるいはある種の仮定のもとで誤差の限界を与える式を示した。Jowett（1952）は、観測値が定常確率過程（stationary process）から選びかれたという仮定のもとに系統的標本の分散を決定する方法を与えた。これは同じ分野における Cochran（1946）および Quenouille（1949）の研究から導びかれたものである。

この論文においては、Jowett の方法に類似したやり方を展開するか、それよりは幾分弱い仮定によるものである。我々はまた観測データのそれより小さい間隔および等しい間隔でとつた標本の分散式を導くであろう。

（これは Jowett の方法には含まれていない）。これは分散が単一の標本から算かれる場合も含まれる。これは屢々言われている原則、すなわち系統的標本それ自身からは誤差の推定値は得られないことと矛盾するものではない。というのは母集団について行なつた仮定は補助情報と同等であるからである。心にも数多くの仮定を設けることができたが、特定のものを選ぶについてはそれが多くの種類のデータにあてはまりそうだという証明がなければならないし、また後で考察する場合において実験と一致する結果を導びくことが必要である。

この研究は、直線上にとつた点で、地上の種々の地被物（various types of cover）の割合（%）を定め観測を行なうとき、この点の間隔をどのようにきめるかということについて行なつた North Canterbury Catchment Board（New Zealand）の担当者との討議から生れたものである。

この理論の主要部は、観測値の組が後出の離散的および連続的変量の何れについても成立する模型にあてはまりさえすれば、時間又は面積上で一様にとられた任意の観測値の系列に対して適用することができる。

便宜のためこれからはいつも観測値をとる直線の全体を意味するトランセクトと、その上の点についてを考えることにする。実験的研究については§4に詳述する。

2. 分散の公式

トランセクトは kn 個の点からなっているとする。標本は k 番毎の点をとることによって抽出されるから、可能な標本の数は k となる。i 点での変量を X_i $(i = 1, 2, \cdots, kn)$ とする。母集団から抽出された系統的標本の平均値の抽出分散 $\sigma_s^2(n, k)$（標本平均値の母集団平均値 $\bar{X}$ からの偏差の平均平方として定義される）の公式は、連続する三つの論文（Madow, 1949, 1953）の最初のものにおいて Madow & Madow (1944) が与えた。

これは平方和の各個の分割と、

$$\sum_{i=1}^{kn}(X_i-\bar{X})^2 = \sum_{i=1}^{kn}\sum_{j>i}(X_i-X_j)^2/kn$$

なる関係から都合のよい形で導くことができる。

すなわち、これから

$$\sigma_s^2(n,k) = \frac{1}{(kn)^2}\sum_{i=1}^{kn}\sum_{j>i}(X_i-X_j)^2 - \frac{1}{kn^2}\sum_{S=1}^{n-1}\sum_{i=1}^{k(n-S)}(X_i-X_{i+kS})^2$$

$$= \left(\frac{f(0)}{n} + 2\bar{f}_n\right) - \left(\frac{f(0)}{kn} + 2\bar{f}_{kn}\right) \cdots\cdots (1)$$

ここで

$$f(0) = \sum_{i=1}^{kn}(X_i-\alpha)^2/kn$$

$$\bar{f}_n = \frac{1}{n}\sum_{S=1}^{n-1}\frac{1}{kn}\sum_{i=1}^{k(n-S)}(X_i-\alpha)(X_{i+kS}-\alpha)$$

$$\bar{f}_{kn} = \frac{1}{kn}\sum_{\gamma=1}^{kn-1}\frac{1}{kn}\sum_{i=1}^{kn-\gamma}(X_i-\alpha)(X_{i+\gamma}-\alpha) \quad \cdots\cdots (2)$$

ここで α は任意の常数である。

k が無限大に tend するものとし、また $S = i/kn$、これに対応する X_i の値を $X(S)$、および $\phi(t)$ を

$$\phi(t)\int_0^{1-t}(X(S)-\alpha)(X(S+t)-\alpha)\,dS \cdots\cdots (3)$$

と定義する。

極限において

$$\bar{f}_n \to \frac{1}{n}\sum_{S=1}^{n-1}\phi\left(\frac{S}{n}\right) = \bar{\phi}_n$$

$$\bar{f} \to \int_0^1 \phi(t)\,dt = \bar{\phi}$$

$$f(0) \to \phi(0)$$

であるから

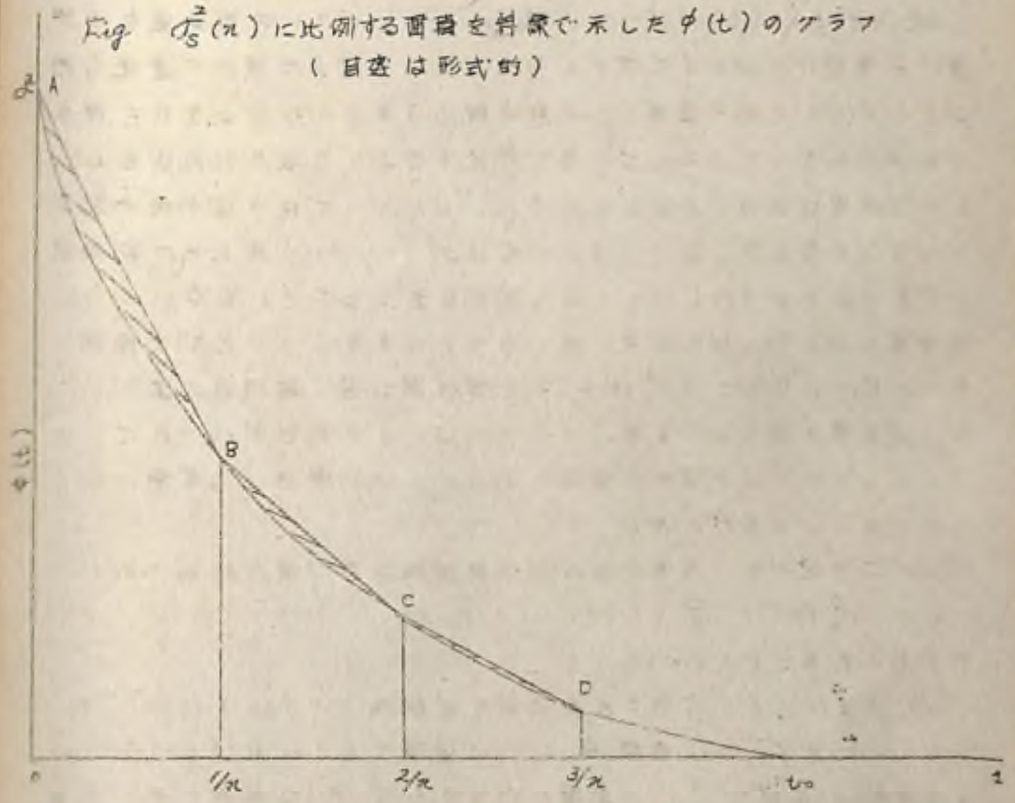
$$\sigma_s^2(n,k) \to \frac{\phi(0)}{n} + 2\bar{\phi}_n - 2\bar{\phi} = \sigma_s^2(n) \cdots\cdots (4)^*$$

$\sigma_s^2(n)$ のもう一つの形が Euler-Maclaurin の定理から得られる。この定理によれば、もし $\phi(t)$ が $(2q+1)$ 次までの連続な導函数をもてば

$$\sum_{S=0}^{n}\frac{1}{n}\phi\left(\frac{S}{n}\right) = \int_0^1\phi(t)\,dt + \frac{1}{2n}(\phi(0)+\phi(1)) - \sum_{\nu=1}^{q}\frac{B_{2\nu}}{(2\nu)!}\frac{1}{n^{2\nu}}(\phi^{(2\nu-1)}(0) - \phi^{(2\nu-1)}(1)) + (-1)^{q+1}\frac{1}{n^{2q+2}}\int_0^n P_{2q+1}(x)\,\phi^{(2q+1)}\left(\frac{x}{n}\right)dx$$

が得られる。

Fig $\sigma_s^2(n)$ に比例する面積を斜線で示した $\phi(t)$ のグラフ
（目盛は形式的）



* Jowett (1952) はこの公式を少し違った形で与えている。Quenouille (1949) の与えたのは n が k と同じ位大きいときの式である。

ここで $\phi^{(i)}(t)$ は i 次の導函数、$B_{2\nu}$ は Bernoulli の数で表わしまた $P_{2s+1}(x) = \sum_{\ell=1}^{\infty} \frac{\sin(2\ell\pi x)}{2^{2s}(\ell\pi)^{2s+1}}$ である。

この級数が収斂すると仮定すると

$$\sigma_S^2(n) = \frac{\phi^{(1)}(1) - \phi^{(1)}(0)}{6n^2} - \frac{\phi^{(3)}(1) - \phi^{(3)}(0)}{360n^4} + \frac{\phi^{(5)}(1) - \phi^{(5)}(0)}{15120\, n^6} + \cdots \quad (5)$$

が得られる。

$\phi(t)$ が必要な次数まで微分可能であるという仮定は、一般には時別なトランセクト内の観測値によって満足されない。しかし Cochran (1946) に従って、特定のトランセクトの観測値を多数の母集団から抽出した標本と考えるなら、個々の値が不連続な場合でも $\phi(t)$ の期待値等（この期待値は母集団から抽出された標本の全体にわたってとられているが希望するような微分可能性をもつという仮定は正当であることが多い。したがって我々は今後分散等の期待値を考えることにする。公式は $\sigma_S^2(n)$, $\phi(t)$ 等をその期待値でおきかえても変らないし、また混乱を生ずるおそれもないから、期待値に対しても同じ記号を用いることにする。この区別を強調したいと思うところでは $\sigma_S^2(n)$, $\phi(t)$ 等の標本値、期待値のように二つの値を考えることにする。この区別は（5）式が用いられているようなところでは本質的に重要であるが　他の場合には実際にあまり重大でないと思われる。

何様な方法から、大きさ n の無作為標本の平均値の抽出分散が

$$\sigma_r^2(n) = \frac{1}{n}(\phi(0) - 2\bar{\phi}) \quad \cdots\cdots\cdots\cdots \quad (6)$$

で与えられることがわかる。

(4) および (6) に含まれる函数を単調減少の $\phi(t)$ について Fig 1 に示す。$\bar{\phi}$ は曲線 $\phi(t)$ 下の面積で与えられ $\phi(0)/2n + \bar{\phi}n$ は多角形 O A B C D … の面積であるから $\sigma_S^2(n)$ は斜線で示した面積の二倍に等しい。[*]

[*] このグラフによる方法は Jowett (1952) が与えたものと同じである。しかし之が定常時系列からとられているという Jowett の仮定は含んでいないことに注意せよ。

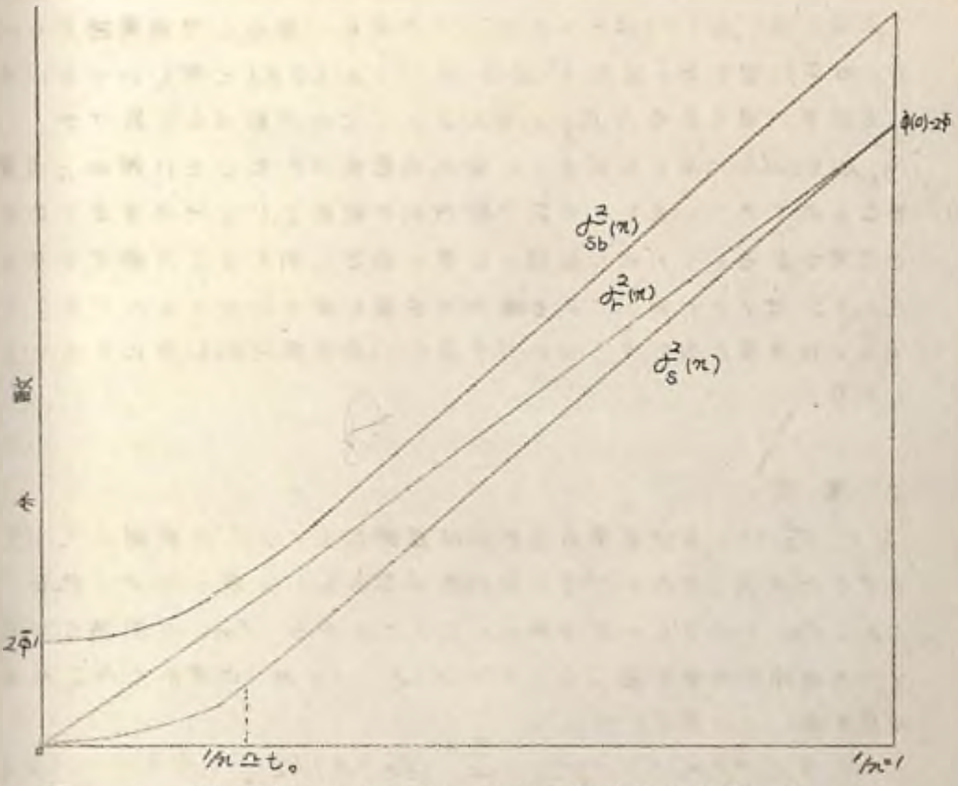


Fig 2　単調減少の $\phi(t)$ に対する相対的な分散の値

(5) から、もし展開が正しければ、まず $\sigma_S^2(n)$ は観測値間で距離 $(1/n)$ の二乗で変化する。一方 $\sigma_r^2(n)$ は $1/n$ とともに一次的に (linearly) 変化する従って十分大きい n に対しては $\sigma_r^2(n)$ は $\sigma_S^2(n)$ よりも大きいから、系統的標本は無作為標本より有効である。もし $\phi(t)$ が 0 かまたはある点 t_0 を越えると一次的に変化するなら、抽出点 $(1/n)$ の間隔が t_0。以上になるとき、$\sigma_S^2(n)$ は $1/n$ とともに一次的に増加する (Fig 2 を見よ)

分散 $\sigma_S^2(n)$ は、同一トランセクトから抽出された全標本の平均値の効率性を表わす尺度である。$\int_0^t \mu(t)\,dt$ に対する標本平均値の偏差の平均平方を $\sigma_{S6}^2(n)$ と表わすと

$$\sigma_{S6}^2(n) = \sigma_S^2(n) + 2\bar{\phi}' \quad \cdots\cdots\cdots\cdots \quad (7)$$

である。但し $\mu(t)$ はトランセクトの点を 抽出した母集団点もにおける平均値でありまた $\bar{\phi}'$ は α が $\int_0^{t'}\mu(t)dt$ に等しいとおいたときの ϕ の値である（Fig 2 をみよ）。この分散はもし我々が $\int_0^{t'}\mu(t)dt$ に生じた疎わしい変化の意味づけをしたい場合には重要なものであり、またZの共分散行列が依然としてそのままであると仮定できる。これには証明が必要である。例えばこの論文で考えたトランセクトにおいてある植物の被覆を徐々に他のものでおきかえると被覆度が変化するばかりでなく、共分散行列も変化するかもしれない。

3 推定

3.1 $\sigma_S^2(n)$ を決定するためには理論的 $(0,1)$ の範囲のすべての点での $\phi(t)$ がわかっていなければならない。我々は n 点を $(0,\ 1/n)$ の中から恣作為にとり、これから $1/n$ の間隔ごとにとった n 点での観測値 X_r $(r=1,2,\cdots,n)$ の標本からこれを推定する。この標本に対して

$$\psi(u/n)=\frac{1}{n-u}\sum_{r=1}^{n-u}(X_r-\alpha)(X_{r+u}-\alpha)\cdots\cdots(8)$$

を定義する。$\psi(u/n)$ は $\phi(u/n)$ の不偏推定値である。

もし $\phi(t)$ に対して特別な形（例えば指数函数）を仮定できればむしろ我々は、$\psi(u/n)$ が非常に多数の点にもとずくことが保証される十分小さい u の値に対しては、その式をこれらの観測された $\psi(u/n)$ にあてはめ、因又は級数展開の式から $\sigma_S^2(n)$ を求めるのがよい。

これが不可能な場合にも、小さいもの値に対する $\phi(t)$ をきめることはむづかしくないから、結局級数式に必要な $t=0$ での係数も容易に決定できる。しかし t が1に近くなると、含まれる項の数が少なくなりすぎて十分な推定値を得られなくなる。グラフによる方法にも同様な欠点があるが、それに加えてもし可能ならすべての点で $\phi(t)$ を計算するという手数を省略することが望ましい。

もし $\phi(t)$ がある値 t_0 で直線なら、グラフによる方法では

t_0。以上のところまで $\phi(t)$ を計算しても $\sigma_S^2(n)$ に影響はない。実際においては、重要と考えられる最大の抽出区間に等しい長さ上でそれが直線と見做せるようになる点まで $\phi(t)$ をつなげばおそらく十分であろう。

展開が正しいと仮定して、t_0 までの部分のみを考える場合に含まれる誤差は

$$\sum_{\nu=1}^{\infty}\frac{B_{2\nu}}{(2\nu)!\,n^{2\nu}}\left[\phi^{(2\nu-1)}(1)-\phi^{(2\nu-1)}(t_0)\right]$$

で与えられる。

もし級数の形を用いるなら、幾つかの両端の仮定が必要である。$\phi(t)$ が $t=t_0$ から先で線型と仮定すれば、$\phi'(1)=\phi(t_0)/(1-t_0)$ で、高次の微係数は0である。これと別な仮定は、領域 $(0,\ 1-t_0)$ で $Z(S)$ が平均値 μ_1 をもち、$(t_0,\ 1)$ なる領域で $Z(S)$ と独立とするものである。但し平均値は μ_2。このとき、$t>t_0$ に対して

$$\phi(t)=(1-t)(\mu_1-\alpha)(\mu_2-\alpha)$$

となる。観測値を観察してこの仮定の正しいことを確められれば $\phi(t)$ はトランセクトの最初と終りの平均値から直接決定することができる。この仮定が正確には成立たなくても、$\phi(t)$ はこのずれにそう影響されない。例えば μ_1 および μ_2 が常数でなくて、それぞれ異なる

$$\mu_1=\alpha_1+\beta_1 t,\quad \mu_2=\alpha_2+\beta_2 t$$

なる傾向に従うとすれば、

$$\phi(t)=(1-t)(\alpha_1-\alpha)(\alpha_2+\beta_2-\alpha)+\frac{(1-t)^2}{2}\left\{\beta_1(\alpha_2+\beta_2-\alpha)-\beta_2(\alpha_1-\alpha)\right\}-\frac{(1-t)^3}{6}\beta_1\beta_2$$

これから

$$\phi^{(2)}(1)=-(\alpha_1-\alpha)(\alpha_2+\beta_2-\alpha)$$
$$\phi^{(3)}(1)=\beta_1\beta_2$$

となる。平均値が領域 $(0,\ 1-t_0)$ および $(t_0,\ 1)$ で常数であるという仮定によつて

$$\phi^{(1)}(1) = -(\lambda_1-\alpha)(\lambda_2+\beta_2-\alpha) - \frac{(1-t_0)}{2}\left[\beta(\lambda_2+\beta_2-\alpha)+\beta_2(\lambda_1-\alpha)\right] + \frac{(1-t_0)^2}{4}\beta_1\beta_2$$

となるから、平均値が妥当な標本にもとずくものでありまたも。をできるだけ大きくとるなら、この誤差は傾向が著しい（即ちβ_1およびβ_2が大きい）場合を除いてそう大きくならないであろう。

3.2　計画のデータをおめるため、パイロット調査（m点の）を行なう場合には、mより大きいかまたは小さいnの値について$\sigma_S^2(n)$を計算したい。このとき 標本平均値の分散は$m=n$になる標本それ自身から計算さるべきである。

nが事実上mより小さいとき例えばその$1/2$もしくはそれ以下であれば、級数式がゆるやかに収斂するからグラフによる方法が最もよい。nがmと同じオーダーかそれよりも大きいときには、$\phi(t)$の値と決定する場合の抽出変動が測定すべき差と更わなくなつてくるから、級数の形を用いなければならない。

$t=0$で必要な微係数はGregory - Newtonの公式によつて決定できる。即ち

$$\phi^{(1)}(0) = m\left(\Delta\psi(0) - \frac{1}{2}\psi(0)\cdots\cdots\right)$$

$$\phi^{(3)}(0) = m^3\left(\Delta^3\psi(0) - \frac{3}{2}\Delta^4\psi(0)\cdots\right)$$

とおけはよい。ここで$\Delta\psi(0) = \psi\left(\frac{1}{m}\right) - \psi(0)$

$\Delta^2\psi(0) = \psi\left(\frac{2}{m}\right) - 2\psi\left(\frac{1}{m}\right) + \psi(0)$　である。

$\Delta^i\psi(0)$の抽出分散は、iとともに急速に増大するから、$\phi^{(1)}(0)$, $\phi^{(3)}(0)$等を推定するには、少数の項のみを用いる方がよい。また同じ理由で、$\sigma_S^2(n)$を実際に展開するときも項数は少ない方がよい。このことは級数展開が有用なのは、nが十分大きくて、収斂が急速な場合に限ることを示している。というのは他の場合には、著しい偏りが生ずるからである。実際においては多くの場合、分散の正確な推定値が要求されることはなく、かなり大きい誤差でも許容できるものである。

3.3　観測値に周期的な傾向があるときには、周期が$\frac{1}{m}$または その倍数に近いと非常な偏りを生ずる原因となる。というのは、グラフによる方法において計算された$\psi(u/m)$の値を通る滑らかな曲線を引くことは妥当性を失うからである。
級数式も確実な結果を与える程急速には収斂しない。もしデータの性質から周期の存在することが予期できれば、この困難は$1/m$が周期の十分小さい部分となるようにすれば避けられる。

Finneyは森林調査において生ずる予期されない周期変動の一例を示した。そうしてこの場合には簡単な説明は何も与えられないように思えた。この種の影響を避ける唯一の方法は、観測値の間隔をできるだけ小さくとり、少なくとも周期効果が見出されないことのないようにすることである。

3.4　$\chi(S)$が0または1の値をとる特別な場合には$\alpha=0$とおくと方程式（8）は非常に簡単な形となる。u個所あとに他の1が続いているような1の個数を$\sqrt{u}$とかくと

$$\psi\left(\frac{u}{m}\right) = \frac{\sqrt{u}}{m}$$

となる。

この場合には

$$0 \leqq \Phi(t) \leqq (1-t)$$

であるから$t=1$における勾配（gradient）の限界を

$$-1 \leqq \phi^{(1)}(1) \leqq 0$$

のようにきめることができる。

もしトランセクトの隣前の点の間に相関がないと仮定すれば$\phi^{(1)}(1) = -p_1 p_2$である。但しp_1とp_2はトランセクトの始めと終りでの1の割合である。

もしmが大きければ第一階差（first difference）は略んと無視できるから

$$\sigma_S^2(n) \doteqdot \frac{m\,n}{12\,n^2} \qquad (9')$$

が成立する。但しnはm個の観測値における1との run (run)

の数である。

この近似式はかが非常に大きいときにのみ正しい。その理由は隣接点間の相関が高いから、ランの数はかが少し変化してもそれにつれてあまり大きく変らないからである。例えば観測値を交互にぬかして系列内のランの数を求めることができる。もしこれがrとあまり違わなければこの式を分散推定の簡便式として用いることができる。

4 実験結果

実験データは遷生の被展度の記録で、これは生きている遷生（living vegetation）、枯死した遷生（dead vegetation）裸地の三つに分類されている。この遷生は、New Zealand 南島（South Island）の南alps の分水嶺以東の山岳地帯にある土着のカヤ草原の一部分からなつている。外業はWaimakariri 川流域において行なわれた。ここの土壌の大部分は、海抜高 7000 フィート以上に直する噴出な硬砂岩質の山岳から導びかれたものである。一世紀にわたつて、自然の開墾によつて多数のブロックにわけられた。このカヤの草地に綿羊が放牧されてきた。

草地解析（pasture analysis）の点分析 point method が用いられ、長さ 10 チエーンおよびそれ以上のラインtransectが恒久的に設定された。解析器は2インチ間隔で鉄のピンを植えた長さ20インチの金属製の枠である。毎年夏の同じ日に解析器をtransect上に注意深く置いて 各々のピンに触れる植物の数を記録した。これによつてtransect上の2インチごとの観測値の組が得られた。

色々異なつた条件の4つのtransectを選んだ。これらは第4、7、8、11transectで、その簡単な説明を APPendix にあげてある（省略）

我々は生活している（living）のと生活していない場合、枯死（dead）しているのと枯死していない場合および裸地または非観

第1表 u区間あとに同じ地被の点が続いている点の数 v_u

u	トランセクト4			トランセクト7			トランセクト11			参考トランセクト8
	L	D	B	L	D	B	L	D	B	L
0	1570	201	2226	4713	348	1227	2955	127	676	3335
1	1362	58	1992	4315	84	933	2787	33	514	2750
2	1252	38	1895	4194	55	848	2718	24	456	2498
3	1169	29	1815	4116	43	802	2668	14	424	2340
4	1100	25	1745	4061	36	750	2635	16	401	2214
5	1046	27	1681	4015	31	717	2601	12	381	2151
6	974	20	1623	3991	33	694	2574	11	365	2077
7	919	16	1579	3974	27	668	2556	8	350	2042
8	872	19	1530	3936	21	647	2534	6	338	2025
9	829	13	1488	3924	24	630	2518	8	320	1999
10	797	11	1545	3909	18	626	2506	4	314	1974
11	773	9	1426	3888	18	615	2499	3	306	1962
12	762	11	1411	3869	18	595	2478	4	294	1968
13	743	15	1398	3855	17	586	2464	7	281	1994
14	733	11	1382	3856	21	581	2451	6	271	1981
15	724	8	1367	3847	17	581	2444	6	268	1979
16	724	9	1374	3832	18	561	2434	8	260	1966
17	730	4	1373	3812	18	550	2427	7	259	1936
18	728	12	1378	3809	17	557	2418	6	251	1927
19	719	12	1363	3788	14	544	2414	3	250	1915
20	718	11	1363	3784	13	530	2408	4	245	1936
21	709	14	1348	3786	21	535	2404	1	243	1951
22	704	13	1336	3788	24	522	2401	5	253	1941
23	693	11	1330	3779	17	514	2404	5	249	1945
24	694	11	1325	3778	21	522	2389	8	236	1942
25	696	10	1322	3767	25	513	2380	3	232	1944
26	700	10	1326	3755	22	501	2387	5	236	1958
27	703	14	1335	3745	16	492	2395	4	236	1945
28	—	—	—	3745	19	498	2398	4	240	1931
29	—	—	—	3746	14	483	2396	4	244	1918
30	—	—	—	3745	19	475	2398	9	237	1910
35	—	—	—	—	—	—	—	—	—	1875
40	—	—	—	—	—	—	—	—	—	1826
面積(m)	3997			6283			3758			6321
被覆率(%)	39.28	5.03	55.69	75.01	5.46	19.53	78.63	3.38	17.99	52.76

地の三つの場合を別々に考えることにする。生態学的にはこの最初のものが最も重要である。最後は点Sでの地面の状態を生活している個生があれば $\chi(S)=1$ なければ $\chi(S)=0$ として記述することにする。その他の場合も同様である。

第2表 $\sigma_S^2(n)$ 一覧表

トランセクト4　　$n=3997$

$$V_L = \frac{10^4}{n^2}\left(46.75 - 0.197\,\frac{n^2}{n^2} \cdots\right)$$
$$V_D = \frac{10^4}{n^2}\left(40.30 - 0.311\,\frac{n^2}{n^2} \cdots\right)$$
$$V_B = \frac{10^4}{n^2}\left(57.03 - 0.381\,\frac{n^2}{n^2} \cdots\right)$$

トランセクト7　　$n=6283$

$$V_L = \frac{10^4}{n^2}\left(102.32 - 0.650\,\frac{n^2}{n^2} \cdots\right)$$
$$V_D = \frac{10^4}{n^2}\left(74.17 - 0.592\,\frac{n^2}{n^2} \cdots\right)$$
$$V_B = \frac{10^4}{n^2}\left(75.85 - 0.473\,\frac{n^2}{n^2} \cdots\right)$$

トランセクト11　　$n=3758$

$$V_L = \frac{10^4}{n^2}\left(40.59 - 0.222\,\frac{n^2}{n^2} \cdots\right)$$
$$V_D = \frac{10^4}{n^2}\left(27.53 - 0.239\,\frac{n^2}{n^2} \cdots\right)$$
$$V_B = \frac{10^4}{n^2}\left(39.99 - 0.217\,\frac{n^2}{n^2} \cdots\right)$$

トランセクト8　　$n=6321$

$$V_L = \frac{10^4}{n^2}\left(138.48 - 0.664\,\frac{n^2}{n^2} \cdots\right)$$

上にあげた場合に対する生活しているもの、枯死、裸地（L.D.B）の V_u の値を第1表に与えてある。

第1表をみると、約50インチ（即ち $u>25$）以上の間隔への場合の枯死点（dead point）の観測値に対する V_u の値（無作為効果を別にすると）相関がない（即ち $(n-u)p^2$. 但しpは枯死の比率）場合に予期されるものと大体等しい。生活および裸地点の場合はこの間隔では多少相関がある。しかしこれは大きい抽出間隔の場合を除いて、直線と考え得るのに十分な程ゆつくり変化する。

このことは、ここに表示していない $\phi(t)$ の値で与えられる分散に対する寄与を無視することと同等である。

級数展開の係数を求めるためも=0における三次までの微分のみを用いた。これは抽出誤差が大きくなるのを避けるためである。

$t=1$ では $\phi^{(1)}(1) = -p^2$ ととり、他のすべての導函数は0ととつた。

第2表に生活、枯死、および裸地（V_L, V_D, V_B）の割合（%）の分散を級数式で与える。トランセクト内の点の総数は n だから、n 点を抽出すると間隔は $2n/n$ インチとなる。

直接計算して求めた分散の値と観測を無作為に行なう同様な標本抽出で期待される分散 V_L', V_D', V_B' を比較のために第3表に示す。間隔が狭いとき V_L と V_B は V_L' と V_B' よりはるかに小さい。枯死した植生の割合は比較的小さく、生活している地被は多少無作為に点在するということのために V_D と V_D' 間の差は小さい。一方生活している植生では、特にトランセクト4でそうであるが、小さい裸地または岩石地に点々と群生する傾向があるから、大体10～20インチの狭い間隔では $\phi(u)$ の値が大きくなる。トランセクト8の差が小さいのは地被かむしろより一様であることを示している。

第3表には10インチ $n=n/5$ の間隔に対する近似公式（9）から計算した分散の値をもあげてある。この一致は満足なものである。この例は§3.2で注意した問題の説明を与えるものである。近似公式は、$\phi^{(1)}(0)$ の推定に $\Delta\psi(0)$ のみを用い $\sigma_S^2(n)$ の級数でそれからあとのすべての項を無視する場合と殆んど同等である。10インチ間隔の場合、この方法で数値積分から分散の近似値を求めると、第2表で与えたと同様に $\Delta^2\psi(0)$ および $\Delta^3\psi(0)$ を用いる級数移よりよい結果が得られる。

理論の正しさおよび間接には模型の十分さをチエックするため、それぞれトランセクト上の最初の20点中から無作為に選んだ1点より出発し、20番目ごとの点をとつた10個の標本と40番目ごとの点をとつた20個の標本を作る。これらの標本平均値の分散は（1）で与えられる $\sigma_S^2(n, n)$ の実験的な推定値を与える。

第3表　相関のあるデータから抽出した系統的標本と無作為標本の分散の比較

間隔 (2m/3インチ)	10インチ	20インチ	40インチ	60インチ	80インチ
			トランセクト4		
V_L	0.604(0.543*)	1.746	6.088(5.852**)	11.156	16.223(9.223**)
V'_L	2.996	5.992	11.985	17.977	23.970
V'_D	0.370(0.373*)	0.902	2.091(3.233**)	3.330	4.568 (2.538**)
V_D	0.571	1.143	2.286	3.429	4.572
V_B	0.586(0.610*)	1.959	6.219(2.059**)	11.118	16.017(1.959**)
V'_B	3.090	6.180	12.360	18.540	24.720
			トランセクト7		
V_L	0.400(0.420*)	1.141	2.942(6.200**)	4.991	7.020(9.120**)
V'_L	1.496	2.992	5.983	8.975	11.967
V_D	0.267(0.273*)	0.644	1.462(1.883**)	2.282	3.100(1.918**)
V'_D	0.409	0.819	1.637	2.456	3.274
V_B	0.341(0.310)	0.842	2.331(3.229**)	3.480	5.300(2.065**)
V'_B	1.257	2.514	5.029	7.543	10.058
			トランセクト11		
V_L	0.435(0.510*)	1.606	4.090(4.994**)	7.819	11.549(15.366**)
V'_L	2.235	4.470	8.940	13.410	17.881
V_D	0.243(0.277*)	0.633	1.503(1.769**)	2.484	3.386(4.796**)
V'_D	0.434	0.869	1.738	2.606	3.475
V_B	0.437(0.463*)	1.389	3.463(3.426**)	6.400	9.337(13.535**)
V'_B	1.917	3.833	7.667	11.500	15.334
			トランセクト8		
V_L	0.661(0.610*)	1.791	4.957(3.776**)	8.474	12.440(6.944**)
V'_L	1.969	3.939	7.878	11.817	15.755

* $\sigma_S^2(n) = mr/12n^2$ から計算した近似推定値

** 抽出実験から求めた推定値

値し $k=20$ または $k=40$ で $kn = m$ である．しかし $\sigma_S^2(n.k)$ はまた $\sigma_S^2(n) - \sigma_S^2(kn)$ の推定値とも考えられるから，$\sigma_S^2(n.k)$ の実験的推定値に，(9)で $n=m$ とおいて与えられる $\sigma_S^2(kn)$ を加えれば $\sigma_S^2(n)$ の推定値を求めることができる．これらは第3表のグラフ法（間隔40および80インチ）で求められた対応する分散の次に括弧書きで示してある．これらの比較は厳正でないがしかし実験的分散のグラフ法で得られた分散に対する比の平均値は1.0946となるから十分と思われる．

5　要約

有限母集団から抽出された系統的標本の平均値の分散を求める Madow & Madow (1944) の公式を連続的母集団に適用し，ある仮定のもとに単一標本から系統的標本の平均値の分散を求める方法を展開した．そうしてこれを生態学研究における植生の被覆度の推定に適用した．

R. M Williams

参考文献

Cochran, W. G (1946) Relative accuracy of systematic and stratified random samples for certain Class of populations. A.M.S. 17, 164-77

Finney D. J (1948) Random and systematic sampling in timber surveys. Forestry. 22. 64-99

——— (1950) An example of periodic variation in forest sampling. Forestry. 23, 96-111

Hasel. A. A (1938) Sampling error in timber surveys J. agric. Res. 57, 713-36.

Jowett. G. H (1952) The accuracy of systematic sampling from conveyer belts. Appl statist 1. 50-9

Madow. W. G (1949) On the theory of systematic

sampling II A.M.S 20, 333-54

Madow. W.G (1953) On the theory of systematic Sampling III A.M.S. 24 101-6

Madow. W.G & Madow L.H (1944) On the theory of systematic sampling I A.M.S. 15, 1-24

Osborne J.G (1942) Sampling errors of systematic and random surveys of cover type areas. J.A.S.A 33 256-64

Quenouille. M.H (1949) Problems in plane sampling A.M.S 20, 355-75

Yates. F (1948) Systematic sampling Phi Trans A 241, 345-77