

論文審査の報告

論文題目：超母集団モデルによる寸法指標推測について

氏名：星野伸明

論文の内容：

統計的母集団が多数の副集団からなる場合に、特定の頻度の副集団数を「頻度の頻度」または「寸法指標」と呼ぶ。統計的生態学、計量文献学、統計的開示制限等の分野では、有限母集団の寸法指標を標本から推測する必要性がある。特に統計的開示制限の諸問題は官庁統計や経済統計の分野において今後重要な意味を持っている。本論文では、超母集団モデルを用いた寸法指標の推測の方法論を詳細に検討している。

一章では、寸法指標推測問題において超母集団モデルを用いる事の根拠が論じられている。

有限母集団解析においては、有為抽出と確率抽出の優劣が問題であるが、ネイマン以来有為抽出の恣意性を排除した確率的な標本抽出の方法が主流となり、ネイマン流の標本抽出理論においては主に標本設計の良し悪しが評価の対象となった。しかしながら本来標本設計の良さは母集団の性質に依存するものである。従って、近年になって超母集団によって集団の性質を明示的に記述し、それに依存した標本設計の評価が考慮されるようになった。これは有為抽出への一種の回帰とも考えられる。ある種の問題については、母集団の性質を仮定しない限り、十分に鋭い結果を得られないことがある。寸法指標推測問題においては、標本から得られる情報が少いために、母集団の性質を仮定することは不可欠である。

本論文が想定する応用分野では、寸法指標の凸性（Zipfの法則）が経験的に知られている。従って、このような知見を超母集団モデルとして利用することが自然である。また、寸法指標推測問題においては経験ベイズ法が使用されるが、これはモデルの誤特定について比較的頑健という利点がある。そして一つのモデルを事前に特定するのではなく、複数のモデルを集合的に候補とし、その中からデータ所与でモデルを選択すれば良い。そして選択されたモデルによる推測を採用する。

超母集団を導入することで、有限母集団解析は形式的に無限母集団解析と同等になる。そこではネイマン流とは異なり、 n 個の標本は所与とみなされる。すなわち大きさ N の母集団に関する推測は、 $N - n$ 個の追加抽出個体を予測することにより行われる。これが超母集団モデルにおける予測アプローチである。

二章では、特定の超母集団モデル所与のもとで、寸法指標の推測方法を考察している。まず、超母集団モデルによる母集団寸法指標推定値の精度が検討される。実は寸法指標推測問題においては標準的な予測アプローチを用いることができな

いため、便宜的に新たに大きさ N の集団を同じ超母集団から抽出した場合の寸法指標を推測すると考える。ここで予測尤度アプローチの観点からは、モデルの母数は局外母数と考えられるが、局外母数の除き方は frequentist の枠組みでは一意に定まらない。そこで利便性の観点から、単純に標本からモデルの母数を推定して推定した値を代入することとする。この場合決定理論的に考えれば、寸法指標の点推定値についてはモデルの母数推定値の下での母集団寸法指標の期待値として良いが、区間推定を母数の推定値所与で構成した場合、信頼区間が正確に見えすぎるといった問題点がある。つまりこのように構成された信頼区間は、母数の推定誤差を考慮に入れていない他、モデル選択における不確実性も反映されていない。もちろんこれらの要因を考慮して信頼区間を補正する事も原理的には可能である。しかし正確な区間推定が必要なならば、ブートストラップのような数値的方法を用いる方が現実的であろう。

次に母集団と標本の関係が議論される。モデルのもとでの母集団分布は、母集団サイズ N に依存する。ここで、大きさ n の標本分布が母集団分布の N を n で置き換えて得られる場合、すなわち適当な標本設計についてモデルが閉じる場合、が便利である。このようなモデルでは、標本から得た母数の推定値の下で母集団寸法指標を推測することが正当化される。本論文では、 N が固定されたモデルについては非復元単純無作為抽出、 N がランダムなモデルについてはベルヌーイ抽出を仮定し、このような設計について閉じたモデルのみ取り扱う。個体のラベルに分布が依存しなければ、そのようなモデルが構成される。

さらに頻度のモデルと寸法指標のモデルの関係が考察される。各副集団の頻度が互いに独立に同一分布に従うようなモデル（以下、基本モデルと呼ぶ）では、寸法指標は無項の多項分布に従う。この場合、寸法指標の周辺分布は二項分布である。故に副集団の総数を無限大に近づけるような極限で、周辺の寸法指標の期待値が定数に収束すれば、寸法指標は互いに独立なポアソン分布に従う。このような極限操作を小数法則と呼ぶ。

三章では、非負整数上の離散分布を用いた寸法指標モデリングが考察されている。重要な離散分布のクラスである混合ポアソン分布と複合ポアソン分布の性質が整理され、本論文で提案されているモデリングの方法論と既存のモデリング手法の差異が検討されている。

もし副集団の総数が総頻度に比して多ければ、各副集団頻度の周辺分布をポアソン分布で近似するのは自然である。ここで経験的に over dispersion が観測される為、頻度分布を混合ポアソン分布で記述する。混合ポアソン分布はベルヌーイ抽出について閉じている他、モデリングに都合が良い性質を多く持つ。

一般に各副集団の頻度が互いに独立に同一の非負整数上の分布に従うようなモデルで、平均的に寸法指標が対数凸ならば、頻度の分布は複合ポアソン分布である。なお複合ポアソンが混合ポアソンとなる必要十分条件は、ポアソンに混合された分布が無限分解可能な事である。一章で議論したように、寸法指標の凸性を

記述出来るモデルが必要とされるため、ここでは複合ポアソン分布の基本モデルについて更に考察を加えている。そのようなモデルでは、小数法則が常に適用可能である。そして極限で寸法指標が従うポアソン分布の期待値は、複合ポアソン分布の複合された分布で規定される。また複合ポアソン分布の基本モデルでは、小数法則は総頻度の分布を変えずに副集団の総数を増やす操作となる。そして総頻度の分布を変えずに小数法則が適用出来るのは、複合ポアソンモデルに限られる。

以上の結果を前提とすれば、寸法指標のモデリングは正の整数上の分布の選択に帰着されることとなる。適当な正の整数上の分布で寸法指標の凸性を記述できるとして、この分布を複合して得られる複合ポアソン分布の基本モデルに小数法則を適用すれば、望ましいモデルが現れる。

そこで、次に正の整数上の分布の選択が問題となる。ここでは正の整数上の分布族として標準ベキ級数分布を仮定し、グラフィカルなモデル選択が検討されている。標本寸法指標について対数凸性を要求すれば、代表的な標準ベキ級数分布の中では対数級数分布と拡張負の二項分布のみが条件を満たす。これらを複合分布とする複合ポアソン分布は、それぞれ負の二項分布とポアソン・パスカル分布であり、これらの分布のモデルが重要となる。

更に総頻度を固定したモデルを考察する。ここまでのモデルでは総頻度が確率変数だったので、それを所与とした条件付分布を使えば良い。そのようなモデルは、非復元単純無作為抽出について閉じる。先に述べたように、複合ポアソンの基本モデルでは小数法則を適用した後も総頻度の分布は変化しない。従って複合ポアソンの基本モデルに小数法則を適用して得られるモデルの総頻度所与の条件付分布は、基本モデルの総頻度に関する条件付分布に小数法則を適用した結果と等しい。結局、条件付けと小数法則を組み合わせれば、複合ポアソン分布によるモデリングは4形態に整理される。

以上のモデリングの方法論は、特に副集団の総数が未知の場合、伝統的なモデリングとは異なる。伝統的なモデリングでは正の頻度を持つ副集団の記述を重視し、頻度0の副集団は極めて便宜的な扱いをされることが多い。具体的には、有限個の副集団の頻度が互いに独立に同一の正の整数上の分布に従うと仮定することが多い。期待値が有限な正の整数上の分布は複合ポアソンモデルではないので、このようなモデリングは今までの議論とは別に考察する必要がある。しかしこの場合、モデルは標本抽出について閉じない。つまりモデルの母数を推定する際、標本で頻度が0の副集団数という不可知量が必要となる。従って母集団寸法指標を推定する際、このような伝統的モデリングは不十分である。副集団の総数が未知または不確実の場合、前述の小数法則による極限分布を用いるべきである。この極限では分布が副集団の総数に依存せず、各副集団の正の頻度が複合ポアソンの複合された分布に従っていると解釈出来る。

四章ではここまでの一般的議論に基づき、個別の超母集団モデルに関する文脈

が整理される。具体的には各モデルの導出、他モデルとの関係、寸法指標の階乗モメント、母数推定に関する結果が示される。取り上げられているモデルは、ディリクレ=多項、ガンマ=ポアソン(負の二項)、対数正規=ポアソン、一般化逆ガウシアン=ポアソン、逆ガウシアン=ポアソン、対数級数、Ewens、Pitman、条件付逆ガウシアン=ポアソン、拡張負の二項、極限条件付逆ガウシアン=ポアソンである。またポアソン・パスカルモデルについても考察され、結果として一般化された条件付逆ガウシアン=ポアソンモデルと一般化された極限条件付逆ガウシアン=ポアソンモデルが現れる。

最後の五章では、超母集団モデルによる寸法指標推測の応用が整理されている。主な応用分野は統計的生態学、計量言語学、統計的開示制限だが、前二者については標準的なリファレンスとも言うべき文献が存在する。従って統計的生態学と計量言語学における応用については、簡単なサーベイのみ与えられている。統計的開示制限は新しい応用分野であり、統計的開示制限の応用例が詳しく説明されている。また最後に、実データにモデルを当てはめた数値例が示されている。三種類のデータについて Pitman モデルのあてはまりが良好であったため、Pitman モデルによる寸法指標推定がより詳しく数値的に検討されている。標本抽出率を変えながら性能を数値的に評価したところ、Pitman モデルにはバイアスが存在するもののバイアスは標本抽出率の増加につれて縮小することが観察されている。

講評：一章における超母集団モデルの正当化においては、ネイマン流の標本抽出設計に関する評価などにおいてやや一面的な議論がなされている印象があった。二章ではパラメータ推定の精度まで含めた予測の精度が議論されてはいるが、その後三章以降の具体的な検討においてはパラメータ推定の精度の考察は不十分であると思われた。3章における混合ポアソン分布と複合ポアソン分布の理論、四章の超母集団モデル各論においては、著者のオリジナルな研究成果を含めて充実した理論的内容となっており、統計的開示制限問題の理論の発展への貢献が大きい。五章の応用例の数値的検討も興味深いものがある。

論文審査の結論：星野氏は、経済学研究科在学中より統計的開示制限問題にとりくみ、特に寸法指標推測問題において、いくつかの新たな離散分布族を導出するなどのオリジナルな貢献をしてきた。本論文においては、有限母集団の推測理論の中で寸法指標推測問題を位置付け、星野氏自身の貢献を含めて寸法指標推測問題の全貌を明らかにしている。統計的開示制限問題は官庁統計や経済統計の分野において今後重要な意味を持つ問題であり、本論文の方法論は有用である。また統計的生態学、計量文献学等への応用についても整理している。このような意味で、本論文は本研究科が要求する論文博士の基準を十分に満たしていると考えら

れる。したがって、この審査委員会は、本論文を博士（経済学）の学位を授与するにふさわしいと全員一致で判断した。

2003年12月

審査委員：竹村彰通（主査）

国友直人

矢島美寛

久保川達也

大森裕浩