

幹線旅客純流動調査のサンプル拡大方法に関する考察*

Aggregation Method of the Interregional Travel Survey*

奥村 誠**

By Makoto OKUMURA**

1. 全国幹線旅客純流動調査の特徴

全国幹線旅客純流動調査（以下純流動調査）は、都道府県をまたぐような都市間交通の実態を把握するためのアンケートベースの調査であり、1990年から5年ごとに2005年まで4回にわたり、秋季の平日の調査が行われ、2005年からは休日（日曜日）の調査が実施された。今年の秋には第5回目の調査が実施される予定である。純流動調査では、対象日に航空、新幹線及び特急列車、高速バス、幹線航路などの交通機関を利用した旅客に別々に実態調査を行うが、質問項目として当該交通機関の利用だけではなく、当日のトリップの真の出発地、目的地、トリップ目的、利用交通機関、利用経路、および個人属性を調査している。その結果を統合、集計することにより、一つのトリップの中で複数の交通機関や便を乗り継ぐ行動を把握できることが、他の事業者ごとの統計調査には見られない大きな特徴となっている。

純流動調査の実態調査では、全数調査を行う航空を除いて、特定の列車や車両、調査区間などを抽出したうえで調査が実施される。また航空調査においても全員から回答が得られるわけではない。回答サンプルから日データ、もしくは年間データを集計するために、「それぞれのサンプルと同等の旅行者が当該日、あるいは年間に何人いるか？」を表す拡大係数を乗じたうえで集計を行う必要がある。これまで、純流動調査データに基づいて、特定の距離帯やODにおける交通機関の分担率、駅や空港などの選択率などの分析に個人属性や移動目的を考慮することは試みられてきたが、発生段階のモデルの構築には大きな限界があった。それは純流動調査における拡大係数が、別途事業者から報告されている日断面交通量、あるいは年間断面交通量のみに基づいて定められており、個人属性や移動目的を考慮していないことが大きな限界であったためと考えられる。

*キーワード：交通行動調査、調査論、

**正員，博(工)，東北大学東北アジア研究センター
(仙台市青葉区川内4-1、TEL:022-795-7571、
E-mail:mokmr@m.tohoku.ac.jp)

今後都市間交通の需要は量的に減少する中で、特定のニーズを持ったユーザーに特徴的な行動を見出し、選択的なサービスの提供によって需要を喚起するという政策の重要性が大きくなっていく。このような分析ニーズに対応するためには、純流動調査における拡大係数を、トリップ目的や個人属性やに対して集計してもひずみを生まないように設定しておく必要がある。

本論文は、この拡大係数の設定問題について論じることを目的とするが、具体的なデータを用いた手法の開発の段階には至っていないことをご容赦願いたい。研究発表会では何らかの分析結果を示す予定である。

2. 第4回純流動調査の拡大方法

(1) データ拡大の全体的手順

純流動調査の実態調査は幹線交通機関事業者の協力を得て交通機関ごとに行われている。拡大処理の手順は大きく2段階に分かれる。まず第1段階として、各サンプルがとられた断面ごとの回収数と、別途事業者から報告された平日1日、休日1日、もしくは年間の利用者数との比率を用いて拡大処理を行う。第二段階として、同種交通機関の中での乗り継ぎ、異種交通機関間の乗継サンプルについて、第一段階の回収率から計算される分散の逆比を用いて加重平均化する。なお一方の交通機関においてのみ観測されたトリップパターンについては、ダミーのトリップが付与される。

なお、純流動調査では秋季の連休のない週について、平日調査を水曜日に、休日調査を日曜日に行っている。そのため休日調査のトリップには観光等の「行き」のトリップに比べて「帰り」のトリップが多く含まれてしまう。そのため第一段階の拡大処理の際には、平均的な休日の分布に近づけるため、逆方向のトリップを持つ仮想的なサンプルを付加したうえで2で割って平均化する。同様に第4回調査では平日のトリップについても同様の手順により、方向別の偏りをなくしたデータとしている。

(2) 鉄道におけるデータ拡大の手順

航空、高速バス、幹線旅客船は基本的にポイント対ポ

イントのサービスであるため、調査サンプルは当該区間利用者からの単純な標本とみなしてよいのに対して、幹線鉄道においては一つの列車の途中での乗降客が存在し、異なる区間の利用者が混在している中での標本となり、特別な取り扱いが必要となる。

第4回調査では、調査票記入時間が十分に取れない短距離利用者ほど回収率が小さくなる影響を考慮するため、乗車時間が短いほどサンプルの回答率が低いと仮定してその逆数の重みをつける(乗車時間補正)。次に、乗車している乗客全員に調査票を配布できる調査開始駅に対して、中間駅では調査員が目視で途中乗車者を確認して追加配布するために配布率が落ちるといった影響を考慮して重みを割増す(中間駅補正)。

以上のサンプルごとの補正を行った後に、断面交通量との比較に基づいて拡大係数を付与する。調査区間の各駅間断面について、上記の補正後のサンプルを集計して通過人数を計算する。事業者より提供された調査日の駅間通過人数との比を求め、サンプルの拡大係数を設定する。列車の運行区間の中に複数の駅間通過人数の調査断面を含む場合には、各断面におけるサンプル通過人数と駅間通過人数との比によって別々に拡大係数値を求め、分散の逆数を重みとして平均してトリップデータの拡大係数とする。このような拡大係数の推定計算を長距離利用者から順に行い、区間利用者については残余のトリップ数と通過サンプル数の比に基づいて設定する。

さらに、複数の列車を乗り継いでいるサンプルについては、双方の調査対象列車における乗り継ぎ行動の出現率を勘案したサンプルの誤差率を求め、重み付き平均によって乗り継ぎ客数を算出して、再度両側のサンプルの拡大係数値を修正するという手順をとる。

最後に以上の手順で作成した平日、休日のデータに逆方向のデータを加えて方向を平均化し、各々を247倍、118倍した量を加算して年間の拡大値とする。

以上のように、第4回調査のサンプルの拡大方法は、調査対象列車で得られたサンプルは同一の確率でサンプリングの対象となっているという考え方に立ち、それらを集計した断面通過人数を、事業者から報告された断面交通量との比率で機械的に調整している。

(3) 現行のデータ拡大方法の問題点

調査で特定の属性やトリップパターンの旅行者に欠損が多くみられるような場合には、その拡大集計結果の属性分布はひずんでしまう恐れが大きい。たとえば時間的に余裕がある観光目的の旅客に比べて仕事目的の旅客の回答率が低い場合には、上記のような拡大係数から集計した旅行目的別トリップ数はひずみを持ち、旅行目的を説明変数に用いた交通機関選択モデルも正しく推計することができないという問題が生じる。

3. サンプル拡大問題へのアプローチ

(1) 数理計画問題

まず各サンプルにつける拡大係数を変数とみなす。この時、a) 各断面交通量について拡大集計したトリップ数が観測断面交通量にほぼ等しい、b) 属性やトリップ目的別に出発ゾーンについて拡大集計したトリップ数は、各ゾーンにおける対応する属性を持つ居住人口に比例している、c) サンプルの拡大係数は調査地点ごとの抽出率の逆数にほぼ等しい、というような特徴をできるだけ満足するような拡大係数を求めるという問題を考えることができる。a) ~c)の条件のうちの1つか2つを制約条件とし、残りの条件を2乗誤差で評価したものを目的関数と置くことにより、数理計画問題を考えることができる。この時、各サンプルごとに自由に拡大係数を設定できるとすると、自由に操作できる変数の数が数十万次元の莫大な数となるため実用的とは言えない。

サンプル属性グループごとの回答率と、調査区間ごとの抽出率を操作変数と考え、その積としてサンプルの回収率が定まり、その逆数が拡大係数となると考えれば、未知変数の次元は数百の規模に減らすことができるが、操作変数の積の形を含む非線形計画問題となる。より実用的には、2種類の操作変数のうちの一方を固定し、解空間が線形であらわされる計画問題を互い違いに解いて、収束解を求めるということになる。

(2) 反復比率調整法(iterative proportion fitting)

異なる次元に着目して集計した場合の集計値の観測地との比率に基づいて、個々の要素に与える重みを順次修正する計算を繰り返す反復比率調整法も、この種の問題には多く適用されている。

この問題では、まず適当な初期値、たとえば各調査区間における乗客数とサンプル回収数の比率を拡大係数の初期値と考える。つぎにa)現在の拡大係数値に基づいてサンプルを断面ごとに集計し、当該断面の観測交通量との比率を求め、その比率の逆数を乗じて拡大係数値を更新する。このような修正操作をすべての交通量観測断面に対して行う。ついでb)各出発地における属性グループごとの人口比を基準として、現在の拡大係数値に基づく発生サンプル集計量を求めてその比率に基づく習性計算を繰り返す。以上のa)とb)の修正計算を互い違いに行い、変化量が十分小さくなるまで繰り返す。

この方法は考え方がシンプルで分かりやすい点、計算の操作も単純であることが優れているが、条件の着目順序の違いによって異なる結果に収束する危険性があること、計算の途中でゼロと区別できない値が出現するとそれ以降の計算の精度が保証できない点、本来連続変数で

ある年齢などの属性変数をどの値で区切るのか、もしくは4区分されたトリップ目的のいくつかを統合して分析すると異なる結果が得られるという可変集計単位問題がある、などの問題点が指摘されている。

(3) MCMCアプローチ

杉木らは、土地利用マイクロシミュレーションを行う際の初期の人口・世帯分布マイクロデータを生成する問題を取り上げ、従来のIPF法では多次元で連続変数を含むような属性の分布を生成できないことから、乱数を用いたモンテカルロ手法によって現在のサンプル集合から取り除くべきサンプルを決め、さらに主成分分析に基づく独立スケール上の乱数によるモンテカルロ手法で新しいサンプル候補の属性の組み合わせを発生して、そのサンプルの置き換えが周辺分布の適合性の向上につながるかを評価してサンプルの更新計算を多数回繰り返すという手順を提案し、その適用性を確認している。

我々の拡大係数設定問題においては、新しいサンプル候補を乱数で作出す代わりに、既存の調査サンプルの一つを取り出してその拡大係数を増加させるという手順をとることになる。

単純に、各サンプルごとの拡大係数を変数と考えれば10万オーダーの次元の変数を扱うことになるため、(1)と同じように積の形で整理するほうが実際的かもしれない。計算手順は以下のようにまとめられる。

- a) 各調査地点の調査時の利用客数とサンプル回収数に応じて、拡大係数の初期値を設定する。
- b) 乱数を発生させ、サンプルを一つ選ぶ。当該サンプルを取り去ることにより、各断面の集計値および発生ゾーンにおける集計値の周辺分布へのあてはまりが改善できるかどうかを評価する。改善できる程度に合わせて1-0の数値を設定する。
- c) 乱数を発生させ、先の設定値と比較して、当該サンプルの拡大係数を減少させるか、もしくは前回の更新で減少させたサンプルの拡大係数をもう一度減少させるかを確率的に決める。
- d) 乱数を発生させ、サンプルを一つ選ぶ。当該サンプルを付け加えることにより、各断面の集計値お

よび発生ゾーンにおける集計値の周辺分布へのあてはまりが改善できるかどうかを評価する。改善できる程度に合わせて1-0の数値を設定する。

- e) 乱数を発生させ、先の設定値と比較して、当該サンプルの拡大係数を増加させるか、もしくは前回の更新で増加させたサンプルの拡大係数をもう一度増加させるかを確率的に決める。
- f) 以上のような拡大係数の調整計算を多数回(100万回オーダー)繰り返し、適合度が大きく変化するまで繰り返す。

以上の方法は乱数を使用するため、計算結果に一意性がなく、計算を志向するたびに異なる結果が得られるという問題があるが、満足させるべき条件が多くなっても同様の方法で計算できるため、拡張性に優れている。

4. おわりに

本論文は、この拡大係数の設定問題について、新しい計算方法のアイデアを提示するのにとどまった。具体的なデータを用いた計算手法を実装し、研究発表会では何らかの分析結果を示したい。

参考文献

- 1) 奥村泰宏・新倉淳史：「全国旅客幹線純流動調査について」第15回鉄道技術・政策連合シンポジウム(J-RAIL2008), pp.419-422, 2008.
- 2) 国土交通省政策統括官：「平成18年度 全国幹線旅客純流動調査報告書」,2006.
- 3) 杉木直・宮本和明・Varameth Vichiensan：「土地利用マイクロシミュレーションにおける初期マイクロ世帯データの推定方法」,土木計画学研究・講演集, No.39,2009.
- 4) 星野崇宏：「調査観察データの統計科学—因果推論・選択バイアス・データ融合」, 岩波書店, 2009.