離散時間パネル調査の
調査期間, 調査間隔, 样本数の最適化

北村隆一1, 藤井聡2, 山本俊行3

1正会員 Ph.D. 京都大学大学院工学研究科土木システム工学専攻教授 (〒606-8501 京都市左京区吉田本町)
2正会員 博(工) 京都大学大学院工学研究科土木システム工学専攻助手教授 (〒606-8501 京都市左京区吉田本町)
3正会員 博(工) 京都大学大学院工学研究科土木システム工学専攻助手 (〒606-8501 京都市左京区吉田本町)

離散時間パネル調査により行動過程の状態を繰り返し観測し, それら離散時点での観測値に基づく状態間遷移確率などの行動過程のパラメータを推定するとき, 多大な推定誤差が生じる. 常ならびにパネル調査設計に当たっては, この推定誤差による費用と調査費用の双方を考慮されなければならない. 本研究では, 推定誤差費用と調査費用の和を最小化する最適化問題として調査設計問題を定式化し, パネル調査の調査回数, サンプル数, 調査期間を決定している. さらに仮想パラメータ値を用いた数値計算の結果を示し, パネル調査の最適標本数は一般に考えられるより小さく, 調査期間は短なものであることを示している.

Key Words: panel survey design, optimal survey intervals, estimation error cost, Markov processes

1. はじめに

交通行動の動的解析の重要性と, そのための経時的データ収集手法の必要性と, これまでに数々の研究で強調されてきた1521. 日本国内でも, 消耗バイアスの特性と補正1526, 57118. 狀態依存と系列相関1923, 151719. 時間的同質性と経時的行動変化の特性2021. パネル調査設計22. 連続時間軸の導入と準在時間2227. 2829. また交通行動の動的特性全体3021. 31に, 交通行動を観測する方法の開発が進められてきた. これらの研究を通じ, パネルデータの特性, その利用法として, 行動の経時的変化の観測が可能となる, 変化の推定を統計的にとらえることにより定める. 多様な時間ごとに観測される交通情報の集計, 個々の行動を把握できる. 母集団の動的なモニタリングが可能となる, などの点が挙げられている. これに加え, Duncan et al. 32は, 調査設計の視点から, 速度を表現する変数などの経時的測定はパネル調査を通じてのみ可能であること, また, 比較的短い時間間隔で調査が行われるため, 過去の事象回顧 (episodic recall) の精度を向上させることが可能であることなどを, パネルの利点として挙げている.

多くのパネル調査は, 離散観測時点で対象とする変数の値を測定するという形を採っている. 観測時点は, ほぼ一定の間隔で, 外生的変数が設定されることが多い. このような離散時間パネル調査 (discrete-time panel survey) と呼ぶことにしよう. このような離散時間パネル調査で収集されるデータは, 離散調査時点での変数の測定値からなり, 対象とする行動過程の変化に関わる情報提供するもの, その変化がどのようなものである, その変化がどのようなものである, したがって, 離散時間パネルデータを用いて行動の変化を解析しようとする場合, 行動過程の変化の変化の変化を推定することが求められる. しかしながら, 次章で述べるように, 常ならびにパネル調査の変化から行動の変化を推定するとき, 場合によっては大きな誤差が生じることを避け得ない.

交通行動の動的解析に最も適したデータは, 対象とする行動およびそれに影響を及ぼす説明変数の値を, 時間軸上に沿って連続的に記録し, 連続データ (continuous data) である. しかしながら, 対象とする行動を, 数々の(必ずしも特定されていない)説明変数の値を連続的に観測, 記録することは, 不可能ではないとして, 多大の困難を伴う非現実的な作業である. より実用的な方法として, 回答者による事象回顧に基づき行動や説明変数の変化を連続的に行うという試みが考えられる. しかし, 事象回顧に関しては数の膨大となるため, その各々について事象回顧データを収集することは実際的でない場合が多いと言えよう. さらに, 事象回顧そのものについて, 一般的な記憶の不正確さに加え, telescoping effects などの系統的バイアスの問題が指摘されており, 調査方法としての信頼性が疑問視されている.
事象回帰は調査手法として信頼性が低く、直接的な測定に基づき連続データを取得することが非現実的なとき、上述のように、離散調査時点で状態の測定値に基づき、行動変化の特性を分析することが残された可能性として浮かび上がる。例えば Golob et al. は、約1年間でなされた通勤交通手段のパネル観測結果から、手段間の連移確率行列を求めている。しかしながら、すでに述べたように、ここで得られた観測値はパネル調査時点での行動過程の状態を示し続けるものの、行動変化そのものを記録したものではない。Kitamura et al. は、離散時点での状態の観測値に基づき離散的な行動の経時的変化を推定するときに生じる誤差を分析し、離散時点で収集されたパネルデータから行動過程を規定するパラメータを推定することは、多大の誤差を伴うことを示している。

本研究は Kitamura et al. の結果を受け、パネル観測値に基づくパラメータの推定誤差を考慮しつつ、数値計算に基づきパネル調査間隔、調査回数、サンプル数、および各調査費用の最適化を行い、離散時間パネル調査設計の一助とすることを目的としている。分析の前提として、対象とする行動は通勤交通手段選択などの確率的離散選択行動であるとする。パネル調査は一定間隔で実施されるものとして、得られるデータは各離散時点での行動過程の状態のみを含み、事象回帰に基づくデータは含まれないものと仮定する。パネル設計の最適化に当たっては、行動過程はマルコフ過程として記述されると仮定し、パラメータ推定に際しての誤差と調査費用を勘案しつつ、調査設計を最適化することを試みる。

本研究で対象としている問題は、回答誤差や消耗などのパネル調査の問題を一切含まない理想的なパネルデータが得られたとしても、回避できないものであることご留意されたい。すなわち、離散時点における観測値を用いるとき、それら観測値がいかに正確であるか、仮説の選定、観測時間分布推定のバイアスは生じる。したがって、本研究では、解析を始めとするパネル調査の諸問題は視認の外に置き、問題を純化した上で、どのように効率的なパネル調査を設計するかという点を基に考えて行く。消耗などの存在するときにどのような調査設計が可能となるかという問題は、今後の課題となる。

次章では、まず本稿で対象とする問題を例示した後、マルコフ過程を内包し、その特性を規定するパラメタと、パネル調査から得られる行動過程の状態の観測値との関係を示す。3 章ではパネル調査設計を、基本数、調査回数、及び総調査期間を決定変数とする、仮定誤差と調査費用の和を最小化する最適化問題として定式化する。これに続き、総調査費用が固定された場合と

固定されない場合について、パネル調査最適設計の数値例を第 4 章に示す。5 章に結論を挙げる。

2. マルコフ過程

本章では Kitamura et al. に沿って、離散交通行動の経時的過程をマルコフ過程として定式化する。対象とする交通行動は離散状態 (discrete states) の集合からなる状態空間 (state space) 上で演算されており、状態間の連移は瞬時に行われるものとする。従って連移と連移の間で行動過程はただ一つの状態にある。これをその状態での滞在 (sojourn) と呼ぶ。状態の列挙としては交通手段、用途、一日の生成トリップ数などが挙げられる。連移が起こる時刻および連移後の状態は確率的に決定され、そのメカニズムは時間によって変化しないものとする (時間的一様性; time homogeneity)。

状態空間を \( E \) に非負の整数の集合を \( N \)、また \( \mathbb{R} = \{0, \pm \infty\} \) とし、\( X_n, n \in N \) により \( n \)番目の連移後の状態を \( T_n \) により \( n \)番目の連移が生じる時刻を表す、ここで \( X_n \in E, T_n \in \mathbb{R}, 0 = T_0 \leq T_1 \leq T_2 \leq \cdots \) である。この確率過程を \( (X, T) = (X_n, T_n, n \in N) \) と表す。本稿ではこの確率過程により行動過程が記述されるものとし、さらに \( (X, T) \) はマルコフ過程であると仮定する。

図-1 に行動過程の例を示す。横軸が時間、縦軸が行動過程の状態を示し、\( E = \{a, b, c, d\} \), 初期状態が \( \langle a \rangle \) である。各々の状態での滞在は、水平な太線により示されている。この図では \( (T_1, T_2, \cdots, T_d) \) の \( 6 \)時点での連移が生じており、\( X_n = (a, a, c, a, d, b) \) の離散状態が示されている。さて、この行動過程が、間隔 \( L \) の離散時点 \( (S_1, S_2, \cdots, S_d) \) で観測されたとしよう。この場合、データとして記録される行動過程の状態は \( Z_1, Z_2, Z_3 = (c, c, c, d, d) \) である。したがって、離散時間パネル調査の観測結果からは、時刻 \( S_1 \) と時刻 \( S_2 \) の間で、状態 c
から状態 $d$ への遷移がただ 1 回のみ生じたという推定結果が得られる。図から明らかのように、実際には、$S_i$ と $S_d$ の間で遷移は 4 回生じており、状態は $(c, a, c, e, a, d)$ と変化している。

この例に示されるように、離散パラメータデータから行動過程の推定を行うとき、多大な誤差が生じる。この誤差は、連続時間軸上の行動過程を離散時間での観測値に基づき推定するときに不可避的に生じるものである。したがって、仮に消去によるバイアスや、回答の不正確な影響を示したとしても、この不適切な誤差は依然存在することに留意されたい。

さて、上述のように本研究では $(X, T)$ はマルコフ過程であると仮定する。マルコフ過程はマルコフ再生成過程の特殊形で、

$$
Pr[X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_0, \ldots, X_n, T_0, \ldots, T_n] = Pr[X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n]
$$

をすべての $n \in N_0, j \in E, t \in \mathbb{R}$, について満たし、現在の状態 $X_n$ が与えられたとき条件付過去独立である。さらにすべての $i, j \in E, t \in \mathbb{R}$, について

$$
Pr[X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i] = Q(i, j, t)
$$

が成立すると仮定する。これは時間の一様性を意味し、将来の行動過程が過去の観測が起こったと仮定し、$T_n$ に依存しないことを意味する。確率群 $Q = \{Q(i, j, t) \mid i, j \in E, t \in \mathbb{R}\}$ は状態空間 $E$ 上のセミ・マルコフ核 (semi-Markov kernel) と呼ばれる。状態 $i$ から状態 $j$ への遷移確率、すなわち、現在の状態が $i$ のとき、遷移後の状態が $j$ である確率は、

$$
P(i, j) = \lim_{t \to \infty} Q(i, j, t)
$$

と与えられる。時間の一様なマルコフ過程の状態変遷確率

$$
Q(i,j,t) = P(i,j)(1-e^{-\lambda t}), \quad t \geq 0, \quad \forall i, j \in E
$$

と定義される。すなわち、状態 $i (i \in E)$ での潜在時間はパラメータ $\lambda$ の負の指数分布を持ち、状態間の遷移確率 $P(i, j)$ は状態 $i$ での潜在時間から独立である。本章の議論では

$$
P(i, i) = 0, \quad \forall i \in E
$$

を仮定する。

ここまでの議論は $(X, T)$ のものについてのものであったが、次にパネル調査の行動過程の状態が特徴の有無、観測される場合、それが $(X, T)$ を規定するパラメータによりどう表現されるかを考える。連続時間軸上で定義される時刻 $t$ における行動過程の状態を $Y_t$ で表し

$$
P(i, j) = Pr[Y_t = j | Y_0 = i], \quad i, j \in E, \quad t, s \geq 0
$$

により、行動過程が時刻 $s$ に状態 $i$ にあった場合、時刻 $s + t$ に状態 $j$ に変化する確率 $Q(i, j, t)$ と $P(i, j)$ が式 (3), (4) として与えられるとし、$I(i, j) = \lim_{t \to 0} P(i, j)$ とすると

$$
P(i, j) = e^{-\lambda t}I(i, j) + \int_0^t e^{-\lambda t} \sum_k P(k, i)P_s(k, j) ds
$$

が成立する。式 (7) は Chapman-Kolmogorov の方程式と呼ばれ、右辺第 1 項は、時刻 0 から時刻 $t$ まで遷移が起こらない確率（$exp(-\lambda t)$）と、時刻軸上のある時点で（この場合は時刻 $t$）、状態 $i$ から状態 $j$ へと変化する確率（$I(i, j)$）の積である。$i = j$ の場合は、$I(i, j)$ は 1 で、この積は遷移が一時生じず、時刻 $t$ まで状態が $i$ のままである確率である。右辺第 2 項は、時刻 0 の状態 $i$ から、1 回以上の遷移が生じ、時刻 $t$ に状態 $j$ に遷移している確率である。被積分関数は、時刻 $s (0 \leq s \leq t)$ に状態 $i$ から状態 $k (\neq i)$ への遷移が生じ、さらに時刻 $t - s$ 後に状態 $k$ から状態 $j$ に（直接 $k$ から $j$ へ、あるいは、$k$ 以外の状態を 1 度以上経て、$k = j$ の場合には、遷移を含む場合も遷移が全く生じない場合も含まれる）移っている確率密度である。$P(i, j) = 0, \forall i, j \in E$ が仮定されているため、ある状態からそれ自身への遷移は生じないように留意されるべき。この関数を $s$ について積分することにより、第 3 の状態を少なくとも 1 度経て、状態 $i$ から状態 $j$ に移動する確率が得られる。ここで被積分関数には $P_s(i, j)$ が含まれており、式 (7) が再帰的な構造を持つことに着目されたい。この $P(i, j)$ の値は以下に示す関係を用いて算定することが可能である。$P_{i0}$

確率過程 $Y$ の生成作用素（generator）、$A = \frac{d}{dt}P_{i0}$ を

$$
A(i,j) = \begin{cases} -\lambda, & \text{if } i = j \\ \lambda, & \text{if } i \neq j \end{cases}
$$

と定義すると、$t > 0$ について
\[ P_t = e^{tA} \]  \hspace{1cm} (9)

が得られる。ここに \( P_t \) は \( P(i,j) \) を要素として持つ \( n \times n \) の行列、

\[ e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n \]  \hspace{1cm} (10)

である。式 (10) の \( e^{tA} \) は以下の手順で算定される。行列 \( A \) の \( i \) 行 \( j \) 列の固有値を \( \pi_i \), \( i \) 列の固有ベクトルを \( f_i \) とすると

\[ Af_i = \pi_i f_i, \quad i = 1, 2, ..., n \]  \hspace{1cm} (11)

が成立する。ここで

\[ N = [f_1, ..., f_n] \quad D = \begin{bmatrix} \pi_1 & 0 \\ 0 & \cdots & \pi_n \end{bmatrix} \]  \hspace{1cm} (12)

を定義すると、

\[ A^k = ND^kN^{-1}, \quad k = 0, 1, ... \]  \hspace{1cm} (13)

が成立し、

\[ e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n = Ne^{tD}N^{-1} \]  \hspace{1cm} (14)

が得られる。ここに

\[ e^{tD} = \begin{bmatrix} e^{\pi_1 t} & 0 \\ 0 & \cdots & e^{\pi_n t} \end{bmatrix} \]  \hspace{1cm} (15)

である。

逆に観測データから \( \hat{P}(i,j) \) が \( \hat{P}(i,j) \) と推定される場合、生成作用素 \( A \) を

\[ A = \frac{1}{t} \ln \hat{P}_t \]  \hspace{1cm} (16)

と推定することが可能である。ここに

\[ \theta_1, \theta_2, ..., \theta_n \] は固有値、\( S \) は \( \hat{P}_t - I \) の固有ベクトルを列とする行列である。

しかししながら、遷移確率行列の対数は必ずしも一意的に求まるとは限らない。\( A \) が式 (16) により同定できる十分条件として Singer & Spilerman は

\[ \inf \{ \hat{P}(i,i) \} > 1/2 \]  \hspace{1cm} (18)

を挙げると同時に、「この条件が、マルコフ過程モデルが適用される環境において成立していると信じる先駆的の理由はない」と指摘している。また Singer ら 39) 40) 41) によって示された推定値の統計的特性は状態数が 2 の場合にのみ当てはまり、その適用は、状態数が 2 を超えるとき極めて困難なものとなる 42)。

この章を終えるに先立ち、本研究が対象とする問題を整理し、これまでに数々の研究が対象としてきたパネル調査の他の諸々の問題との関連に言及したい。図-1 に示したように、離散時間パネルデータに基づき連続時間軸上の行動過程を完全に規定することは難しい。これは、離散時間パネルデータは、行動過程の離散時点での状態についての情報提供するものので、状態の遷移そのものについての情報は保有していないからである。必然的に、観測されない状態間遷移が存在し、遷移確率と時定常分布の推定にバイアスが生じる。対応策の一つとして、事象観回に基づき（類似）連続データを作成することが考えられるが、これには、多数の変数について事象観回を行うことは困難である。事象観回そのもの信頼性に疑問が残る。等の問題が生まる。本章の議論は、行動過程がマルコフ過程とみなしできるとき、離散時間パネルデータから \( P(i,j) \) を推定し、それに基づき状態間遷移確率、\( P(i,j) \) と時定常分布のパラメータ、\( \lambda_i \) をバイアス無く推定することが可能なことを示している。

ここで強調したいのは、パネル調査が完璧なかたちで設計、実施され、回答誤差や消耗などの問題が皆無であったとしても、遷移確率、時定常分布推定のバイアスは存在するという点である。同様に、行動に時間遅れなどが無く、対象とする状態が均衡状態にあるとしても、時定常点において観測された状態から遷移確率と時定常分布を推定するとき、バイアスは避け得ない。したがって、本研究では、消耗を始めとするパネル調査の諸
問題は、視野の外に置き、離散時点における状態の観測値からいかに行動過程を規定するパラメータを推定するか、さらに、パラメータ推定値の特性を証明する、どのように効率的なパネル調査を設計するか、という点に議論の対象を切り込むでいる。

さらに、問題の単純化のため、行動過程のパラメータは、個体間でも経時的にも変化せず、母集団の均一性が仮定されていると仮定している。行動過程をバイアス無く測定するという点に本研究の主眼があるため、説明変数の存在は無視し得るものとして、行動過程を純粋な確率過程として扱っている。状態空間遷移確率や滞在時間分布パラメータを説明変数の関数として表した例として、Lerman 4) や山本ら 28) がある。これらのモデルの枠組みを用い、行動変化の時間延遲、行動の動的特性を解析することが可能となる。

時間のスケールについては、対象とする事象の影響を大幅に異なるため、数値例を除き具体的なものを設定することに避けている。例えば、世帯構成の変化を対象とする場合、年単位が適切であろうが、通勤交通手段の変遷を対象とするなら、より短い単位——例えば月単位——がより適切となるだろう。

3. パネル調査設計の最適化問題としての定式化

本章では、2 章で整理した知見を基に、一定間隔で繰り返される離散時間パネル調査の標本数、調査期間長、調査回数（あるいは、調査期間長が与えられたとして調査期間）、および調査予算を決定する調査設計問題を、最適化問題として定式化する。行動過程のパラメータ推定の戦略として、式 (18) が満たされるような調査間隔を設定し、式 (16) を適用するという方法が考えられる。この場合、データが所与として、最も精度の高い推定が可能となる。しかし、一般に t の増加に伴い P(t,i,j) は減少するため、式 (18) を満たす t の値は極めて小さなものとなる可能性がある。特に状態の数が多い場合、次章の数値例に示されるように、現実的に実験可能な調査間隔で式 (18) の条件が満たされそうかどうか、疑問視される。したがって、本研究では式 (16) の適用は不可能であるという前提に立つこととする。

式 (16) の適用が不可能だとすれば、効率として考えられるのが、離散時点に観測される行動過程の状態に基づき、状態間の遷移確率を直接推定するという、単純、直感的ではあるが、信頼性の低い方法である。実際のところ、離散時間パネルデータの解析で、式 (16) に示される関係から生成作用素 A を推定し、それに基づき λ と P(i,j) が推定されることを示すことができる。このような研究で、行動過程の状態に基づき選択確率を直接推定するような方法が確立されてきた 5), 7), 19), 25), 48)。しかしながら、こうして推定可能ければ式 (6) に示される P(i,j) に対応するもので、式 (3) の状態間遷移確率 P(i,j) の推定量としては大きな誤差を伴うものである 37)。結果として、状態間遷移確率のみならず、滞在時間分布および遷移の頻度の推定はバイアスを伴うものとなる。本章では、調査時に観測された行動過程の状態により基づいて推定がなされるとの仮定の下、推定誤差を考慮したパネル調査設計問題の定式化を図る。

Kitamura et al. 39) の結果にも示されるように、状態間遷移確率や滞在時間分布の推定精度は、パネル調査間隔を縮小するにつれ向上する。しかしながら、総調査回数が一定とすれば、調査間隔の短縮に伴いパネル調査全体の期間は短縮し、調査期間中の状態遷移頻度は必然的に減少するとともに、滞在時間の打切り観測データの割合が増加する。短い調査間隔そのものはより精度の高い推定を可能とするものの、得られるデータが提供する行動過程についての情報は限られたものとなる。するかに、短い調査間隔を用いることにより観測トレンドを示す状態間遷移の頻度は減少するものの、P(i,j) の推定に用いる情報量は減少すると考えられる。

このトレードオフを勘案しつつ、本章ではまずパネル調査期間長および総調査費用が与えられたとき、標本数および調査回数（あるいは調査間隔）を決定する最適化問題を定式化する 48)。次に総調査費用を固定化していない場合の標本数および最適化調査間隔決定問題を考え、著者の解析で目的関数として用いられているのは、推定誤差による損失と調査費用を加算した一般化最適費用である。4 章では、これら 2 ケースに加え、調査費用は固定されているが調査期間が可変である場合について、数値例を示す。

問題の定式化に当たり、対象とする行動過程は前章に示した離散状態空間を持つマルコフ過程であるとする。また、調査時点間での状態の変化に基づき、状態 i に滞在する確率を推定し、それに基づき滞在時間分布のパラメータ λ が推定され、推定値 λ が用い平均滞在時間 1/λ が推定されるものとする。この推定に伴う誤差に応じて誤差損失が発生すると考える。

新たな変数

\[ \tau = \text{パネル調査期間} \]
\[ \eta = \text{標本数} \]
\[ \kappa = \text{パネル調査期間} \tau \text{における調査回数} \]
\[ \kappa = 2, 3, 4, \ldots \]
\( L = \) 調査間隔
と定義する。パネル調査期間の開始時点と終了時点に調査がなされなければならなから、調査間隔はパネル調査期間と調査回数により \( L = \tau(K - 1) \) と定義される。
さらに \( n = 1, 2, \ldots, K - 1 \) について
\[
f_{mi} = n \text{ 次の調査と } (n + 1) \text{ 次の調査の双方において状態 } i \text{ が観測された標本の頻度,}
\]
\[
f_{ni} = n \text{ 次の調査において状態 } i \text{ が観測された標本の頻度,}
\]
とする。これら観測値に基づき、\( P_{(i,i)} \) の推定値として
\[
\hat{p} = \frac{\sum_{i=1}^{n} f_{ni}}{\sum_{n=1}^{n} f_{ni}} \quad (19)
\]
が用いられたとする。前述のように、この推定値は\( P_{(i,i)} \) に対応するもので、\( P_{(i,i)} \) の推定値としての一貫性を持つものではない。行動過程が定常であると仮定する。
を定義すると、式 (19) の推定値 \( \hat{p} \) の期待値は \( p \) である。パネル調査の結果がプールされる場合、大標本を仮定すると \( \hat{p} \) の標準偏差は \( \sqrt{\frac{p(1-p)}{K-1} n p_i} \) である。本稿の解析では大標本を前提とし、\( \hat{p} \) の分布は上記の期待値と標準偏差を持つ正規分布であると仮定する。
マルコフ過程が、\( n \) 次の調査から \((n + 1)\) 次の調査にかけて連続して状態 \( i \) に滞在する確率は、\( e^{\lambda_t} \) と表される。この関係に着目し、\( \lambda_t \) と状態 \( i \) での平均滞在時間が\( \beta \) で
\[
\lambda_t = -\ln \hat{p}/L \quad (20a)
\]
\[
1/\lambda_t = -L/\ln \hat{p} \quad (20b)
\]
と推定されたとする。これらは仮定した推定であるが、パネルデータのみならず \( P_{(i,i)} \) についての情報を得ることが不可能なため、これら推定値が用いられたと仮定しよう。このときの推定誤差を
\[
\xi = 1/\lambda_t - (1/\ln \hat{p}) \quad (21)
\]
と定義し、推定誤差に伴う損失が費用関数、
\[
\theta, \gamma > 0 \quad (22)
\]
により表されるものとする。ここに \( \theta \) と \( \gamma \) は定数である。これらに基づき、推定誤差による期待費用を
\[
C_e = \theta \int [1/L + (1/L) \ln x] \varphi(x) \, dx \quad (23)
\]
と表すことができる。ここに \( \varphi(x) \) は \( \hat{p} \) の確率密度関数で
\[
\varphi(x) = \frac{\sqrt{p(1-p)}}{\sqrt{2 \pi} (1-p)} \exp \left\{ \frac{- (x-p)^2 (1-p) \ln p}{2 (1-p)} \right\} \quad (24)
\]
と表される。
次に、一回毎の調査の費用が標本数 \( n \) の線形関数であるとし、調査期間 \( \tau \) における総調査費用を \( C_\tau \)
\[
C_\tau = \alpha + (\alpha + \beta n) \kappa \quad (25)
\]
と仮定する。\( \alpha \) はパネル調査全体に関わる固定費用、\( \beta \) は調査票の改訂などに伴う各調査毎の固定費用、\( \beta \) は標本一覧に従う費用で、各率定数とする。パネル調査の総費用を、式 (23) と (25) に示す推定誤差に関係する損失と調査費用を合わせたものとして、\( C = C_e + C_\tau \) と仮定し、調査期間 \( \tau \) が与えられたとして、標本数 \( n \) および調査回数 \( \kappa \) を決定変数とし、この総費用を最小化することを考える。一般性を失うことなく \( \alpha = 0 \) と仮定し、\( L = \tau/(K - 1) \) を用い、目的関数は
\[
\begin{align*}
\min_{\kappa \in \mathbb{R}^+} C &= \theta \int [1/L + (1/K - 1)/(\ln x)] \varphi(x) \, dx + (\alpha + \beta n) \kappa \\
& \quad \text{And } \kappa \text{ が決定される。この問題を規定するパラメータは、行動過程に関するパラメータに加え、} \tau, \theta, \gamma, \alpha \text{ および } \beta \text{ である。}
\end{align*}
\]
と表される。この問題を規定するパラメータは、行動過程に関するパラメータに加え、\( \tau, \theta, \gamma, \alpha \) および \( \beta \) である。
調査費用 \( C_\tau = \alpha + (\alpha + \beta n) \kappa \) が固定されている場合、\( \eta = \frac{1}{\beta} \left( \frac{C_\tau - \alpha}{\kappa} \right) \) が成立し、最適化問題は
\[
\begin{align*}
\min_{\kappa \in \mathbb{R}^+} C &= \theta \int [1/L + (1/K - 1)/(\ln x)] \varphi(x) \, dx \\
& \quad \text{And } \kappa \text{ が決定される。この問題を規定するパラメータは、行動過程に関するパラメータに加え、} \tau, \theta, \gamma, \alpha \text{ および } \beta \text{ である。}
\end{align*}
\]
と一次元化される。次章では、これらに加え調査期間長も考慮する。
4. 数値解析結果

これまでの解析から明らかのように，$P(i,j)$ を解析的に求めることは一般に困難であり，本章の解析も式 (8) ～(15) を用いた数値計算に依るものである．解析の対象となる行動過程は，以下のパラメータを持つマルコフ過程であるとする。

$$
P = \begin{bmatrix} 0 & 0.2 & 0.8 \\ 0.6 & 0 & 0.4 \\ 0.8 & 0.2 & 0 \end{bmatrix} \quad (28a)$$

$$\lambda = (\lambda_1, \lambda_2, \lambda_3) = (1.0, 2.0, 0.5) \quad (28b)$$

ここに $P$ は $P(i,j)$ を要素として持つ行列である.

このマルコフ過程について，いくつかの $t$ 値での $P_t(i,j)$ の値を以下に示す．式 (18) に示される条件は，状態 2 の平均滞在時間に等しい $t = 0.5$ で既に満たされていない[7]．すなわち，式 (16) の援用を可能とするためには，最も短い平均滞在時間よりさらに短い調査間隔を用いることが必要となる．式 (16) を用いこれらのパラメータを推定するためには，極めて短い調査間隔が必要となることが理解されるよう．

$$
P_{0.5} = \begin{bmatrix} 0.09074 & 0.0176 & 0.0750 \\ 0.0249 & 0.8021 & 0.0750 \\ 0.0377 & 0.0092 & 0.9531 \end{bmatrix}$$

$$
P_{1.0} = \begin{bmatrix} 0.6501 & 0.0549 & 0.2941 \\ 0.3181 & 0.3878 & 0.2941 \\ 0.1499 & 0.0339 & 0.8162 \end{bmatrix}$$

$$
P_{1.5} = \begin{bmatrix} 0.4853 & 0.0670 & 0.4477 \\ 0.3745 & 0.1778 & 0.4477 \\ 0.2307 & 0.0491 & 0.7202 \end{bmatrix}$$

$$
P_{2.0} = \begin{bmatrix} 0.4043 & 0.0678 & 0.5278 \\ 0.3675 & 0.1047 & 0.5278 \\ 0.2738 & 0.0561 & 0.6701 \end{bmatrix}$$

$$
P_{2.5} = \begin{bmatrix} 0.3639 & 0.0664 & 0.5697 \\ 0.3516 & 0.0787 & 0.5697 \\ 0.2965 & 0.0595 & 0.6440 \end{bmatrix}$$

$$
P_{3.0} = \begin{bmatrix} 0.3225 & 0.0630 & 0.6145 \\ 0.3225 & 0.0631 & 0.6145 \\ 0.3212 & 0.0629 & 0.6160 \end{bmatrix}$$

は線形費用関数を仮定し($\gamma = 1$)，1 単位の推定誤差が 500 単位の費用に対応するものとする．解析に用いられたパラメータ値を表一に示す．調査費用に関するパラメータは，当研究室がこれまでに実施した諸調査の費用から概算したもので，営利団体が行う調査の費用に対応するものではない．推定誤差費用を示すθの値は，公共交通の需要予測を想定し，予測誤差が過剰投資による損失，あるいは投資不足による運賃収入損失につながると考え，投資不足の場合を想定して設定した[7]．

これからの調査結果では，母数の推定値の信頼性を特定の水準に保つといった観点から標本数を決定するというのは一般で，推定誤差と調査費用との兼ね合いから最適調査形態を決定するという考え方は取られてこなかった．したがって誤差費用のパラメータθの決定法について，考慮が積み重ねられてきた訳ではない．上述の設定法も，暫定的な性格のもとである．商品の売上高の予測のように，推定値の適用が費用要因を含む場合，θのみならず，式 (22) の費用関数そのものを合理的に決定することがあるであろう．同様のことが土木施設設計に関する需要予測についても言えよう．交通行動調査に関しても，その規模，費用水準の決定に当たり，適切な推定誤差の費用関数を適用することが必要とされているといえる．

表一 推計に用いられたパラメータ値

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>500 (万円)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>1.0</td>
</tr>
<tr>
<td>準当たるの固定費用 (a)</td>
<td>30 (万円)</td>
</tr>
<tr>
<td>標当たるの費用 (β)</td>
<td>0.1 (万円)</td>
</tr>
</tbody>
</table>

（1）総調査費用が固定された場合

ここでは総調査費用Cに対する，500万円が固定された場合を考える．まず，調査期間も 3 (年) に固定されているとし，調査回数のみが決定変数となった場合を対象とする．この場合，調査回数が増加し，調査間隔が狭まるにつれて，式 (19) を用いることによる系統的誤差は削減されると同時に，標本数の減少に伴い標本誤差は増加する．逆に調査回数が減少する場合は，系統的誤差が増大，標本誤差は減少する．これら両者の兼ね合いを最適調査回数が決定される．

この数値解析では状態 3 の平均滞在時間，$1/\lambda_3$ をパネル観測結果に基づき推定することを考える．ここで
図-2 総調査費用、調査期間が固定された場合の誤差費用 (Cₖ = 350, τ = 3)

図-2 に示される計算結果から分かるように、式 (27) に示される誤差費用は大幅に変動し、調査回数が 9 回のとき最小値を探る。この場合標準数は 89、また調査回間隔は 4.5 ケ月で、状態 3 での平均滞在時間を 1/5 弱となっている。この数値計算で用いたパラメータの値の下では、調査間隔を狭め系統的誤差を減少させることにより総誤差費用が最小化されるという結果となった。この理由の一つは、ここでの数値例は行動過程の定常性を仮定し、(τ - 1) 回にわたって観測された状態間移動の観測値をプールすることにより値の推定値を求めるということ点である。このため調査回数の増加に伴いパラメータ調査の標準数は著しく減少するにもかかわらず、プールされた観測値の数はさほど減少しないことに、ここで用いたパラメータの下では、調査回数が 3 回の場合にプールされた総観測値数は 1733 で、調査回数が 9 回の場合総観測値数は 711 である。

次に、調査期間が固定されていない場合を考える。この数値計算では、調査期間を 1 年から 5 年まで 1 年刻みに変化させ、誤差費用を算定した。図-3 に示されるように調査期間を、この数値計算例における最小値の 1 年と設定したときに最小誤差費用が得られている。最適調査回数は 6 回で、調査間隔は 2.4 ケ月、標準数は 283 である。

図-3 総調査費用が固定、調査期間が可変の場合の誤差費用 (Cₖ = 350)

図-4 調査期間が 3 年の場合の総費用の最適化

与えないことを示唆している。しかしながら、この結果が、最適種目行列 P 等のパラメータの推定についても当てはまるかどうかを結論づけるには、更なる解析を待たなければならない。

(2) 総調査費用が固定されない場合

表-1 に示されるパラメータ値を用い、パネル調査期間が 3 年と固定された場合、調査回数および標本数の最適化を考える。ここでは標本数を 10 から 200 まで 10 割みで変化させ数値計算を行い、最適値を近似的に求めている。問題の性質から、少なくとも解析的近似では目的関数が凸関数と考えられ、この近似に問題はないと考えられる。結果として得られた最適値は、図-4 に示されるように、調査回数 7 回、標本数 60 である。調査期間は 6 ケ月、調査費用は 252（万円）で、最適点での誤差費用は 268（万円）である。
5. 結論

調査期間および調査回数（調査間隔）を決定変数として含むパネル調査の設計は、断面調査の設計に比べはあるかに複雑な問題である。本稿では、繰り返し観測される行動過程の状態に基づき行動過程のパラメータを推定した場合の誤差に着目し、パネル調査設計を、推定誤差費用と調査費用の和を最小化する最適化問題として定式化し、さらに仮想パネル数を用いた数値計算の結果を示した。数値解析の結果は、パネル調査の最適標本数は一般には考えられるより小さく、全体の調査期間は短く、調査間隔は密なものであることを示している。

本稿の数値解析では、各調査から得られた観測値をプールしパラメータの推定が行われることを想定していることもあり、一定の調査費用下で調査回数を増やしたりした場合の標本数減少に伴う推定誤差費用の増加は小さいものとなっている。しかし、観測値をプールし推定することが可能となるのは、行動過程が時間的同一である場合に限られるため、行動過程が経時的に変化する場合の最適標本数は、本稿で得られたものよりも大きなものとなる。また、実際に行われている調査で用いられる標本数よりはるかに小さい標本数が本稿の数値解析で得られたのは、想定されたパラメータ推定が単純なものであることに加え、母集団が均一で仮定されていることもによる。したがってここでの最適標本数は母集団内の各々のセグメントに対応するものと考えるのが妥当であろう。

本稿では、これまでに余り研究の対象はされなかったパネル調査の設計に関して、推定誤差費用と調査費用の和の最小化という観点から、一つの方法論と数値計算結果を示した。当然のことながら、ここで得た結果の一般化のためには、この分野での更なる研究結果の積み上げが必要である。特に、本稿では事象観点の精度やパネル消耗といった、回答者に関わる要素を一切考慮していない。また、単一の母数の推定のみを考慮して最適化問題が定式化されている。異なった特性を持つ複数のパラメータの推定が求められる場合、あるいは異なった推定量が用いられる場合の調査設計は今後の課題である。さらに、マルコフ過程を前提としない場合、高度の過去依存性が存在する場合、行動過程が時間的一様でない場合、母集団が均一でない場合、等についての解析、また推定誤差関数についての考察が課題として残る。

【注】
1. 事象観点に基づく自動車保有行動との説明変数に関する疑問連鎖データがパネル調査により収集された例として、カリフォルニア州で1993から1996年にかけて実施された電気自動車の需要推定を目的とした調査がある。この調査は、電気自動車の需要推定を目的とした調査である。
2. このトーティオフは、対象とする変数を規定する確率過程の特性により最適調査間隔が決定されることを意味する。各々の変数が異なる確率過程を持つと考えられるから、ある変数について最適な調査間隔は、他の変数については必ずしも最適ではないと考えられる。本稿ではパネル調査設計問題の初期段階として、単一の変数についての調査間隔の最適化を図っている。複数変数を考慮した調査設計の最適化は、次の段階で検討したい。なお、対象とする行動についての知識が皆無で、どの変数に焦点を当てるか不明な場合、調査設計の最適化は不可能であることに注意したい。
3. すなわち、Pr[Y=1] = Pr[Y=1] は、i, s ∈ R。...
4. pは確率であるため、[0,1]の範囲しか取らない。したがって、式 (24) に示すように x の定義域は全数値区間ではないが、本稿に示す数値計算ではその範囲は非常に小さく、[0,1] 以外の範囲の確率は極めて微小であろう。ただし、厳密な数値計算を保証するために、0 以下となる確率と 1 以上となる確率をそれぞれ用いて、正規分布を補正 (truncation) して得られる確率密度関数を、式 (24) の代わりに用いた。
5. 本研究では、パネル消耗を含め、無回答による費用は存在しないものと仮定する。
6. P(2,2) が初めに 0.5を下回る時点は 1 = 0.36 である。
7. この事例では 3 つの状態 を想定しているが、この場合のマルコフ仮定は 3 つの鉄道路線が存在するODの間で各鉄道路線の交通需要の動的変化を表現するものと解釈することができる。

参考文献
1) Hensher, D.A.: Longitudinal surveys in transport: An assessment, Ampt, E.S., Richardson, A.J. and Brög, W. eds.,
10) 杉原順幸, 羽藤英男, 山根景也: 選好意識パネルデータに潜在する消耗バイアスの修正, 土木計画学研究・論文集, No. 11, pp. 311-318, 1993.
25) 延 峻雄, 杉原順幸, 藤原章正, 岡村 章, SP パネル調査の実施方法に関する基礎的分析, 科研パネル研究会(文部省科学研究費基盤研究(B)01830519 交通計画におけるパネル調査の方法論およびパネルデータ解析手法に関する研究)にて発表, 京都大学, 1997年 4月.
26) 藤井 隆, 杉原順幸, 藤原章正, 岡村 章, SP パネル調査の実施方法に関する基礎的分析, 科研パネル研究会(文部省科学研究費基盤研究(B)01830519 交通計画におけるパネル調査の方法論およびパネルデータ解析手法に関する研究)にて発表, 京都大学, 1997年 4月.
31) 小林 充, 永井 篤, 坂本 哲, 今野 康之: 交通実験が交通手段選択に与える影響—バーケアドバサルイドの交通実験に関するパネル分析, 土木計画学研究・講演集, No.
OPTIMIZING THE DURATION, SURVEY INTERVAL, AND SAMPLE SIZE
FOR A DISCRETE PANEL STUDY

Ryuichi KITAMURA, Satoshi FUJII and Toshiyuki YAMAMOTO

The problem of determining the interval between waves of panel surveys, the sample size, and the study duration is formulated in this study as an optimization problem in which the sum of the cost of estimation error and the cost of the surveys is minimized. Numerical analyses using hypothetical parameter values indicate that the optimal sample size of a panel study can be much smaller, and the optimal inter-wave interval and the optimal study duration much shorter, than those values that have typically been used in previous panel studies.