

離散時間パネル調査の 調査期間, 調査間隔, 標本数の最適化

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き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。パネル調査設計に当っては, この推定誤差による費用と調査費用の双方が考慮されねばならない。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, 調査間隔は密なものでありう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としてのパネル調査の必要性は, これまでに数々の研究で強調されてきた^{1)~8)}。日本国内でも, 消耗バイアスの特性と補正^{9), 9)~16)}, 状態依存と系列相関^{5), 14), 15), 17), 18)}, 時間的同質性と経時的行動変化の特性^{19)~25)}, パネル調査設計²⁶⁾, 連続時間軸の導入と滞在時間^{27), 28), 29)}, また交通行動の動的特性全般^{7), 30), 31)}について, 幅広く研究がすすめられてきた。これらの研究を通じ, パネルデータの持つ利点として, 行動の経時変化の観測が可能となる, 変化の推定を統計的により効率的に行える, 行動の動的側面の分析が可能となる, より整合的で正確な予測が可能となる, 個体間の非観測異質性の影響を把握できる, 母集団の動向の監視が可能となる, などの点が挙げられてきた。これらに加え, Duncan et al.³²⁾ は, 調査設計の視点から, 態度を表す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)の精度を向上す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, 外生的に設定されるのが一般である。この種のパネル調査を, 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状態についての情報を提供するものの, 行動の変化そのものについての情報を必ずしも含むものではない。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, 行動過程の状態の変化に基づき行動そのものの変化を推定することが一般である。しかしながら, 次章で見るように, 観測された状態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。

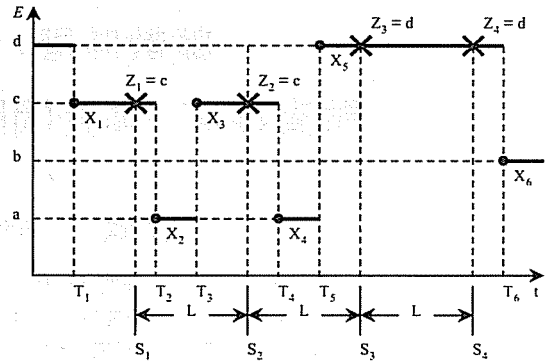
交通行動の動的解析に最も適したデータは, 対象とする行動およびそれに影響を及ぼす説明変数の値を, 時間軸に沿って連続的に記録した, 連続データ(continuous data)である。しかしながら, 対象とする行動と, 数々の(必ずしも特定されていない)説明変数の値を連続的に観測, 記録することは, 不可能ではないとしても, 多大の困難を孕む非現実的な作業である。より実用的な方法として, 回答者による事象回顧に基づき行動や説明変数の変化を連続的に捉えるという試みが考えられる。しかし, 説明変数についてはその数が膨大となるため, その各々について事象回顧データを収集することは実際的でない場合が多いと言えよう³¹⁾。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^{34), 35)}, 調査法としての信頼性が疑問視されている³²⁾。

事象回顧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、離散調査時点での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.³⁶⁾ は、約1年間隔でなされた通勤交通手段のパネル観測結果から、手段間の遷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、ここで用い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はするものの、行動変化そのものを記録したものではない。Kitamura et al.³⁷⁾ は、離散時点での状態の観測値に基づき離散的行動の経時的変化を推定するとき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.³⁷⁾ 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、および総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、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、事象回顧に基づくデータは含ま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有し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に留意されたい。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ろうと、遷移確率、滞在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、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、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に焦点をあてる。消耗などが存在するときどのような調査設計が可能となるかという問題は、今後の課題としたい。

次章では、まず本稿で対象とする問題を例示した後、マルコフ過程を定義し、その特性を規定する諸パラメータと、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、標本数、調査回数、及び総調査期間を決定変数とし、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する。これに続き、総調査費用が固定された場合と



図一 行動過程の例

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.³⁷⁾ 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) 上で定義されており、状態間の遷移は瞬時に行われるものとする。従って遷移と遷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。状態の例としては通勤交通手段、雇用状態、一日の生成トリップ数などが挙げられる。遷移が起こる時刻および遷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 、非負の整数の集合を N_+ 、また $\mathcal{R}_+ = [0, +\infty)$ とし、 $X_n, n \in N_+$ 、により n 番目の遷移後の状態を、 T_n により n 番目の遷移が生じる時刻を表す。ここに $X_n \in E, T_n \in \mathcal{R}_+, 0 = T_0 \leq T_1 \leq T_2 \leq \dots$ である。この確率過程を $(X, T) = \{X_n, T_n; n \in N_+\}$ と表そう。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。

図一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、 $E = \{a, b, c, d\}$ 、初期状態は d 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, \dots, T_6) の6時点で遷移が生じており、 $(X_1, X_2, \dots, X_6) = (c, a, c, a, d, b)$ 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, \dots, S_4) 、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 $(Z_1, Z_2, Z_3, Z_4) = (c, c, d, d)$ 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、時点 S_1 と時点 S_4 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。図から明らかなように、実際には、 S_1 と S_4 の間で遷移は 4 回生じており、状態は (c, a, c, a, d) と変化している。

この例に示されるように、離散パネルデータから行動過程の挙動を推定しようとするとき、多大な誤差が生じ得る。この誤差は、連続時間軸上の行動過程を離散時点での観測値に基づき推定するときには不可避免的に生じるものである。したがって、仮に消耗によるバイアスや、回答疲労に伴う誤差が全て除去された理想的なパネル調査がなされたとしても、この不可避免的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過程の特殊形で、

$$\begin{aligned} \Pr[X_{n+1}=j, T_{n+1}-T_n \leq t | X_0, \dots, X_n; T_0, \dots, T_n] \\ = \Pr[X_{n+1}=j, T_{n+1}-T_n \leq t | X_n] \end{aligned} \quad (1)$$

をすべての $n \in \mathbb{N}_+, j \in E, t \in \mathbb{R}_+$ 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

$$\Pr[X_{n+1}=j, T_{n+1}-T_n \leq t | X_n=i] = Q(i, j, t) \quad (2)$$

が成立すると仮定する。これは時間的一様性を意味し、将来の行動過程が最後の遷移が起こった時刻、 T_n に依存しないことを意味する。確率群、 $Q = \{Q(i, j, t) : i, j \in E, t \in \mathbb{R}_+\}$ 、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、

$$P(i, j) = \lim_{t \rightarrow \infty} Q(i, j, t) \quad (3)$$

と与えられる。時間的一様なマルコフ過程³⁸⁾ の Q は

$$Q(i, j, t) = P(i, j)(1 - e^{-\lambda_i t}), \quad t \geq 0, \quad \forall i, j \in E \quad (4)$$

と定義される。すなわち、状態 i ($\forall i \in E$) での滞在時間はパラメータ λ_i 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 での滞在時間から独立である。本章の議論では

$$P(i, i) = 0, \quad \forall i \in E \quad (5)$$

を仮定する。

ここまでの議論は (X, T) そのものについてのものだったが、次にパネル調査のように行動過程の状態が特定の時刻に観測される場合、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

$$P_t(i, j) = \Pr[Y_{s+t} = j | Y_s = i], \quad i, j \in E, \quad t, s \geq 0 \quad (6)$$

により、行動過程が時刻 s に状態 i にあったとして、時刻 $s+t$ に状態 j にある条件付確率を表す。 $Q(i, j, t)$ と $P(i, j)$ が式 (3), (4) として与えられるとし、 $I(i, j) = \lim_{t \rightarrow 0} P_t(i, j)$ とすると

$$P_t(i, j) = e^{-\lambda_i t} I(i, j) + \int_0^t \lambda_i e^{-\lambda_i s} \sum_k P(i, k) P_{t-s}(k, j) ds \quad (7)$$

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 ($\exp(-\lambda_i t)$) と、時間軸上のある時点で (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、 $I(i, i) = 1$ で、この積は遷移が一切生じず、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 ($0 < s \leq t$) に状態 i から状態 k ($k \neq i$) への遷移が生じ、さらに時間 $(t-s)$ 後に状態 k から状態 j に (直接 k から j へ、あるいは、 k, j 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。 $P(i, i) = 0, \forall i \in E$ 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ことに留意されたい。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 $P_{t-s}(i, j)$ が含まれており、式 (7) が再帰的な構造を持つことに注目されたい。この $P_t(i, j)$ 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³²⁾。

確率過程 Y の生成作用素 (generator), $A = \frac{d}{dt} P_{t|0}$, を

$$A(i, j) = \begin{cases} -\lambda_i & \text{if } i = j \\ \lambda_i P(i, j) & \text{if } i \neq j \end{cases} \quad (8)$$

と定義すると、 $t > 0$ について

$$P_t = e^{tA} \quad (9)$$

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

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n \quad (10)$$

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 番目の固有値を π_i , i 番目の固有ベクトルを f_i とすると

$$Af_i = \pi_i f_i, \quad i=1, 2, \dots, n \quad (11)$$

が成立する。ここで

$$N = [f_1, \dots, f_n], \quad D = \begin{bmatrix} \pi_1 & & 0 \\ & \ddots & \\ 0 & & \pi_n \end{bmatrix} \quad (12)$$

を定義すると、

$$A^k = ND^kN^{-1}, \quad k=0, 1, \dots \quad (13)$$

が成立し、

$$e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n = Ne^{tD}N^{-1} \quad (14)$$

が得られる。ここに

$$e^{tD} = \begin{bmatrix} e^{\pi_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\pi_n t} \end{bmatrix} \quad (15)$$

である。

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

$$A = \frac{1}{t} \ln \hat{P}_t \quad (16)$$

と推定することが可能である。ここに

$$\ln \hat{P}_t = S \begin{bmatrix} \ln(\vartheta_1 + 1) & & & \\ & \ln(\vartheta_2 + 1) & & \\ & & \ddots & \\ & & & \ln(\vartheta_n + 1) \end{bmatrix} S^{-1} \quad (17)$$

$\vartheta_1, \vartheta_2, \dots, \vartheta_n$ 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 & Spilerman³⁹⁾ は

$$\inf_i [\hat{P}_t(i, i)] > 1/2 \quad (18)$$

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験的理由はない」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を提供するものの、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、遷移確率と滞在時間分布の推定にバイアスが生じる。対応策の一つとして、事象回顧に基づき(擬似)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回顧を行うことは困難である、事象回顧そのものの信頼性に疑問が残る、等の問題が孕まれる。本章の議論は、行動過程がマルコフ過程と見なされるとき、離散時間パネルデータから $P_t(i, j)$ を推定し、それに基づき状態間遷移確率、 $P_t(i, j)$ 、と滞在時間分布のパラメータ、 λ_{ij} 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、滞在時間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、離散時点において観測された状態から遷移確率と滞在時間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鑑みつつ、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絞っている。

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、母集団の均一性が保たれていると仮定している。また、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、行動過程を純粋な確率過程として扱っている。状態間遷移確率や滞在時間分布のパラメータを説明変数の関数として表した例として、Lerman⁴³⁾や山本ら²⁹⁾がある。これらのモデルの枠組みを用い、行動変化の時間遅れなど、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、対象とする事象により大幅に異なるため、数値例を除き具体的なものを設定することは避けている。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間隔)、および総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 $\hat{P}_t(i,i)$ 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用に、現実的に実施可能な調査間隔で式(18)の条件が満たされるかどうか、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、方策として考えられるのが、離散時点で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_i と $P(i,j)$ が推定されることは希で、これまでの研究でも、行動過程の状態に基づき遷移確率を直接推定するという方法が頻りに採られてきた^{5), 7), 19), 25), 45)}。しかしながら、こうして推定される確率は式(6)に示される $P_t(i,j)$ 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孕むものとなる。本章では、調査時に観測された行動過程の状態にのみ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.³⁷⁾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、調査期間中の総遷移頻度は必然的に減少するとともに、滞在時間の打ち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。すなわち、短い調査間隔を用いることにより観測漏れとなる状態間遷移の頻度は減少するものの、 $P_t(i,j)$ の推定に援用可能な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して、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⁴⁾。次に総調査費用が固定されていない場合の標本数および最適調査間隔決定問題を考慮する。後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総費用である。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_i が推定され、推定値 $\hat{\lambda}_i$ を用い平均滞在時間 $1/\hat{\lambda}_i$ 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。

新たな変数を

- τ = パネル調査期間,
- η = 標本数,
- κ = パネル調査期間 τ における調査回数;
- $\kappa = 2, 3, 4, \dots$

$L =$ 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いから、調査間隔はパネル調査期間と調査回数により $L = \tau/(\kappa - 1)$ と定義される。さらに $n = 1, 2, \dots, \kappa - 1$ について

$f_{ni} = n$ 次の調査と $(n + 1)$ 次の調査の双方において状態 i が観測された標本の頻度、

$f_{ni} = n$ 次の調査において状態 i が観測された標本の頻度

とする。これら観測値に基づき、 $P(i, i)$ の推定量として

$$\hat{p} = \frac{\sum_{n=1}^{\kappa-1} f_{ni}}{\sum_{n=1}^{\kappa-1} f_{ni}} \quad (19)$$

が用いられたとする。前述のように、この推定量は $P_L(i, i)$ に対応するもので $P(i, i)$ の推定量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し^[4]

$$p = P_L(i, i)$$

$$p_i = \Pr[Y_s = i], s \geq 0$$

を定義すると、式 (19) の推定量 \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 $\sqrt{\frac{p(1-p)}{(\kappa-1)\eta p_i}}$ 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 に滞在する真の確率は、 $e^{-\lambda_i L}$ と表される。この関係に着目し、 $\hat{p} = e^{-\lambda_i L}$ と考え、 λ_i と状態 i での平均滞在時間が各々

$$\hat{\lambda}_i = -\ln \hat{p} / L \quad (20a)$$

$$1/\hat{\lambda}_i = -L / \ln \hat{p} \quad (20b)$$

と推定されたとしよう。これらは勿論偏った推定であるが、パネルデータのみから $P_L(i, i)$ についての情報を得ることが不可能なため、これら推定量が用いられたと想定しよう。このときの推定誤差を

$$\xi = 1/\lambda_i - (-L / \ln \hat{p}) \quad (21)$$

と定義し、推定誤差に伴う損失が費用関数、

$$f(\xi) = \theta |\xi|^\gamma, \quad \theta, \gamma > 0 \quad (22)$$

により表されるものとする。ここに θ と γ 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

$$C_e = \theta \int |1/\lambda_i + L / \ln x|^\gamma \phi(x) dx \quad (23)$$

と表すことができる。ここに $\phi(x)$ は \hat{p} の確率密度関数で

$$\phi(x) = \frac{\sqrt{(\kappa-1)\eta p_i}}{\sqrt{2\pi p(1-p)}} \exp\left\{-\frac{(x-p)^2 (\kappa-1)\eta p_i}{2p(1-p)}\right\}, \quad -\infty < x < \infty \quad (24)$$

と表される^[5]。

次に、一回毎の調査の費用が標本数 η の線形関数であるとし、調査期間 τ における総調査費用を^[6]

$$C_s = \alpha_0 + (\alpha + \beta\eta)\kappa \quad (25)$$

と表す。 α_0 はパネル調査全体に関わる固定費用、 α は調査票の改訂などに伴う各調査毎の固定費用、 β は標本一個当りの費用で、各々定数とする。パネル調査の総費用を、式 (23) と (25) に示される推定誤差に伴う損失と調査費用を合わせたものとして、 $C = C_e + C_s$ と表し、調査期間 τ が与えられたとして、標本数 η および調査回数 κ 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 $\alpha_0 = 0$ と仮定し、 $L = \tau/(\kappa - 1)$ を用い、目的関数は

$$\min_{w.r.t. \kappa, \eta} C = \theta \int |1/\lambda_i + (\tau/\kappa - 1)/\ln x|^\gamma \phi(x) dx + (\alpha + \beta\eta)\kappa \quad (26)$$

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 β である。

調査費用 $C_s = \alpha_0 + (\alpha + \beta\eta)\kappa$ が固定されている場合、 $\eta = \frac{1}{\beta} \left(\frac{C_s}{\kappa} - \alpha \right)$ が成立し、最適化問題は

$$\min_{w.r.t. \kappa} C = \theta \int |1/\lambda_i + (\tau/\kappa - 1)/\ln x|^\gamma \phi(x) dx \quad (27)$$

と一元化される。次章では、これらに加え調査期間長も考慮する。

表一 解析に用いられたパラメータ値

θ	500 (万円)
γ	1.0
調査当たりの固定費用 (α)	30 (万円)
標本当たりの費用 (β)	0.1 (万円)

4. 数値解析結果

これまでの解析から明らかなように、 $P_i(i, j)$ 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 $P_i(i, j)$ の値を以下に示す。式 (18) に示される条件は、状態 2 で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よう。

$$P_{0.1} = \begin{bmatrix} 0.9074 & 0.0176 & 0.0750 \\ 0.1049 & 0.8201 & 0.0750 \\ 0.0377 & 0.0092 & 0.9531 \end{bmatrix}$$

$$P_{0.5} = \begin{bmatrix} 0.6501 & 0.0549 & 0.2941 \\ 0.3181 & 0.3878 & 0.2941 \\ 0.1499 & 0.0339 & 0.8162 \end{bmatrix}$$

$$P_{1.0} = \begin{bmatrix} 0.4853 & 0.0670 & 0.4477 \\ 0.3745 & 0.1778 & 0.4477 \\ 0.2307 & 0.0491 & 0.7202 \end{bmatrix}$$

$$P_{1.5} = \begin{bmatrix} 0.4043 & 0.0678 & 0.5278 \\ 0.3675 & 0.1047 & 0.5278 \\ 0.2738 & 0.0561 & 0.6701 \end{bmatrix}$$

$$P_{2.0} = \begin{bmatrix} 0.3639 & 0.0664 & 0.5697 \\ 0.3516 & 0.0787 & 0.5697 \\ 0.2965 & 0.0595 & 0.6440 \end{bmatrix}$$

$$P_{5.0} = \begin{bmatrix} 0.3225 & 0.0630 & 0.6145 \\ 0.3225 & 0.0631 & 0.6145 \\ 0.3212 & 0.0629 & 0.6160 \end{bmatrix}$$

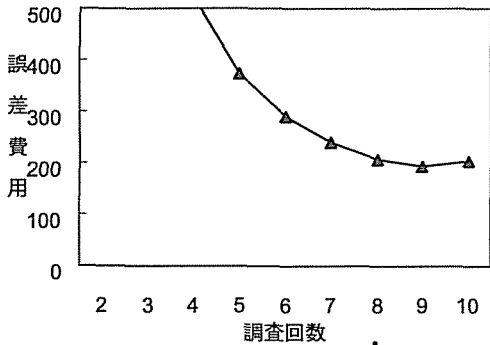
この数値解析では状態 3 の平均滞在時間、 $1/\lambda_3$ をパネル観測結果に基づき推定することを考える。ここで

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⁸⁾。

これまでの調査設計では、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が一般で、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、考察が積み重ねられてきた訳ではない。上述の設定法も、暫定的な性格のものである。商品の売上高の予測のように、推定値の適用が費用要素を含む場合、 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可能であろう。同様のことが土木施設設計の建設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。

(1) 総調査費用が固定された場合

ここでは総調査費用、 C_s が 350 (万円) と固定された場合を考える。まず、調査期間も 3 (年) に固定されたとし、調査回数のみが決定変数となった場合を対象としよう。この場合、調査回数が増加し、調査間隔が狭まるにつれ、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。逆に調査回数が少ない場合は、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で最適調査回数が決定される。

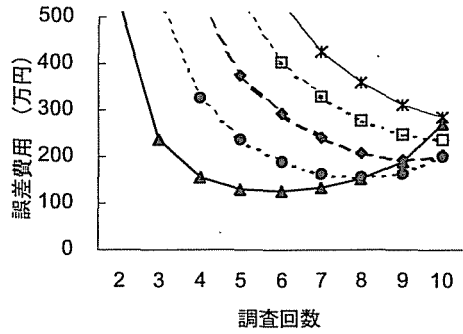


図一2 総調査費用、調査期間が固定された場合の誤差費用 ($C_s = 350, \tau = 3$)

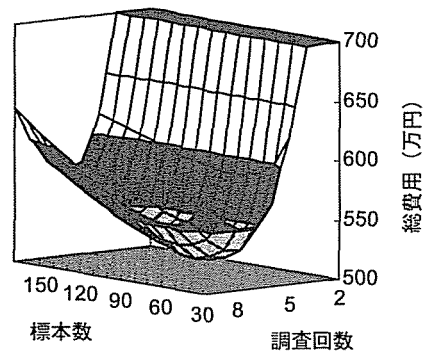
図一2 に示される計算結果から分かるように、式 (27) に示される誤差費用は大幅に変動し、調査回数が 9 回 のとき最小値を採る。この場合標本数は 89、また調査間 隔は 4.5 ヶ月で、状態 3 での平均滞在時間の 1/5 弱 となっている。この数値計算で用いたパラメータ値の下 では、調査間隔を狭め系統的誤差を減少させることにより 総誤差費用が最小化されるという結果となった。この理 由の一つは、ここでの数値例は行動過程の定常性を仮 定し、 $(\tau - 1)$ 回にわたって観測された状態間遷移の観 測値をプールすることにより \hat{p} の推定値を求めている という点にある。このため調査回数の増加に伴いパネル 調査の標本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。ここで用いたパ ラメータ値の下では、調査回数が 3 回の場合にプールさ れた総観測値数は 1733 と最大で、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。図一3 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 ヶ月、標本数は 283 である。

ここまでの結果で明らかなことは、パネル調査の標本 数の拡大を図るより、より短い調査間隔を用い、状態間 遷移の観測漏れを減少させることにより、誤差費用の削 減が達成されるという点である。この理由は、既に述べた ように、観測値をプールするため \hat{p} の推定に用いられ る観測値の総数は、調査回数の増加に関わらずさほど 減少しないという点に求められる。これに加え、図一3 の 結果は、滞在時間の分布関数が特定できる場合、パネ ル調査期間の短縮に伴う打ち切り観測値の増加は必ず しも滞在時間分布関数のパラメータの推定に悪影響を



図一3 総調査費用が固定、調査期間が可変の場合の誤差費用 ($C_s = 350$)



図一4 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、 遷移確率行列 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、更なる解析を待たね ばならない。

(2) 総調査費用が固定されない場合

表一1 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に、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 刻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、少なくとも最適解の近傍で は目的関数は凹関数と考えられ、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、図一4 に 示されるように、調査回数 7 回、標本数 60 である。調査 間隔は 6 ヶ月、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。

最適調査費用は 4. (1) で用いられた値 (350 万円) を下回るものとなっており, 調査回数も図-2 に示されるものを下回る. 図-2 の結果では, 標本数を約 48% 増加させ, 調査間隔を 25% 短縮することに, 過剰な調査費用が割り振られていたことが分かる. 4. (1) の結果と同様, 標本数は極めて少ないものとなっている. 密に調査を行った場合, 式 (18) が成立し, 式 (16) を用いて行動過程の生成作用素の推定が可能となる蓋然性が高くなることを附記する.

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, さらに仮想パラメータ値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考えられるより小さく, 全体の調査期間は短く, また, 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, 行動過程が時間的一様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最適標本数として得られたのは, 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, 母集団が均一と仮定されていることにもよる. 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, これまで余り研究の対象と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, ここでの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, 本研究では事象回顧の精度やパネル消費といった, 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, 高度の過去従属性が存在する場合, 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] 事象回顧に基づき自動車保有行動とその説明変数に関する疑似連続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 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³⁹⁾. この調査は, 前もって郵送された調査媒体を併用した, 高度に個別化 (customize) された computer-aided telephone interviews (CATI) によるもので, 質問の数はのべ 1400 程度に及ぶものである.
- [2] 式 (8) から (15) は, Kitamura et al.³⁷⁾, 藤井ら³⁸⁾にも示されているが, 本論文の自己完結性のためここに繰り返す.
- [3] このトレードオフは, 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. 各々の変数が異なった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る. 本稿ではパネル調査設計問題の研究の初段階として, 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てか不明な場合, 調査設計の最適化は不可能であることに注意されたい.
- [4] すなわち $\Pr\{Y_t = i\} = \Pr\{Y_s = i\}$, $\forall t, s \in \mathcal{R}_+$.
- [5] \hat{p} は確率であるため, $[0, 1]$ の範囲しか取り得ない. したがって, 式 (24) に示すように x の定義域は全実数ではないが, 本稿に示す数値計算ではその分散は非常に小さく, $[0, 1]$ 以外の範囲となる確率は極めて微小であった. 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 (truncate) して得られる確率密度関数を, 式 (24) の代わりに用いた.
- [6] 本研究では, パネル消費を含め, 無回答による費用は存在しないものと仮定する.
- [7] $P_t(2, 2)$ が初めて 0.5 を下回る時点は $t = 0.36$ である.
- [8] ここでの事例では 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 の間での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.,

- New Survey Methods in Transport*, VNU Science Press, Utrecht, pp. 77-97, 1985.
- 2) Kitamura, R.: A panel analysis of household car ownership, 土木学会論文集, No. 383/IV-7, pp. 13-27, 1987.
 - 3) Kitamura, R.: Panel analysis in transportation planning: An overview, *Transportation Research A*, Vol. 24, No. 6, 401-415, 1990.
 - 4) Goodwin, P.B., Kitamura, R. and Meurs, H.: Some principles of dynamic analysis of travel demand, *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Jones, P. ed., Gower Publishing, pp. 56-72, 1990.
 - 5) 杉恵頼寧, 羽藤英二, 藤原章正: パネルデータを用いた交通機関選好意識のダイナミック分析, 土木計画学研究・論文集, No. 10, pp. 31-38, 1992.
 - 6) 内田敬, 飯田恭敬: 交通行動パネル調査の方法論的検討, 土木計画学研究・論文集, No. 11, pp. 319-326, 1993.
 - 7) Lidasan, H. S., 田村亨, 石田東生, 黒川洸: 発展途上国における交通行動のパネル分析, 土木学会論文集, No. 470/IV-20, pp. 135-144, 1993.
 - 8) Raimond, T. and Hensher, D.A.: A review of empirical studies and applications, *Panels for Transportation Planning: Methods and Applications*, Golob, T.F., Kitamura, R. and Long, L. eds., Kluwer Academic Publishers, Boston, pp. 15-72, 1997.
 - 9) 鈴木 聡, 毛利雄一, 中野 敦, 原田 昇: パネルデータに基づく交通手段選択行動の分析, 土木計画学研究・講演集, No. 13, pp. 537-542, 1990.
 - 10) 杉恵頼寧, 藤原章正, 山根啓典: 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.
 - 11) 西井和夫, 近藤勝直, 古屋秀樹, 鈴木隆: パネルアトリションを考慮した買物場所選択モデル: 甲府買物パネルデータを用いて, 土木計画学研究・論文集, No. 12, pp. 389-396, 1995.
 - 12) 藤原章正, 杉恵頼寧, 張 峻屹: Mass Point 手法による交通機関選好モデルの消耗及び回答バイアスの修正, 土木計画学研究・論文集, No. 13, pp. 587-594, 1996.
 - 13) 佐々木邦明, 森川高行, 杉山幸司: パネルサンプルの初期摩擦を考慮した動的な買物目的地選択モデル, 土木計画学研究・論文集, No. 13, pp. 595-602, 1996.
 - 14) 杉恵頼寧, 藤原章正, 小笹俊成: 選好意識パネルデータを用いた交通機関選択モデルの予測精度, 土木学会論文集, No. 576/IV-37, pp. 11-22, 1997.
 - 15) 佐々木邦明: 行動論的背景を持たせた生存時間モデルによるパネルアトリションバイアスを修正した選択モデル, 土木計画学研究・講演集, No. 20 (2), pp. 181-184, 1997.
 - 16) 西井和夫, 近藤勝直, 古屋秀樹, 栃木秀典: 多時点パネルのアトリションバイアスとその修正法に関する研究: 甲府買物パネルデータを用いて, 土木計画学研究・論文集, No. 14, pp. 653-662, 1997.
 - 17) 河上省吾, 三島康生: 通勤・通学交通手段選択行動における動的特性の分析, 土木学会論文集, No. 470/IV-20, pp. 57-66, 1993.
 - 18) 西井和夫, 北村隆一, 近藤勝直, 弦間重彦: 観測されていない異質性を考慮した繰り返しデータに関するパラメータ推定法, 土木学会論文集, No.506/IV-26, pp.25-34, 1995.
 - 19) 西井和夫, 岩本哲也, 弦間重彦, 岡田好裕: パネルデータを用いた休日買物交通パターンの経年変化に関する基礎分析, 土木計画学研究・講演集, No. 15 (1)-1, pp. 163-168, 1992.
 - 20) 飯田恭敬, 内田 敬, 中原正顕, 廣松幹雄: 交通情報提供下の経済選択行動のパネル調査, 土木計画学研究・講演集, No. 16(1)-1, pp. 7-12, 1993.
 - 21) 毛利雄一, 中野 敦, 原田 昇: モノレール開業に伴う事前・事後調査の活用に関する研究—調査方法と交通需要予測手法の改善—, 土木計画学研究・論文集, No. 12, pp. 633-642, 1995.
 - 22) 倉内文孝, 飯田恭敬, 塚口博司, 宇野伸宏: 駐車場案内システム導入によるドライバーの駐車行動変化の実証分析, 都市計画論文集, No. 31, pp. 457-462, 1996.
 - 23) 藤井 聡, 北村隆一, 拓植章英, 大藤武彦: 阪神・淡路大震災が交通行動に及ぼした影響に関するパネル分析, 土木計画学研究・論文集, No. 14, pp. 327-332, 1997.
 - 24) 西井和夫, 近藤勝直, 太田敦夫: 甲府買物パネルデータを用いた休日行動の時間利用特性の実証的分析, 土木計画学研究・論文集, No. 15, pp. 499-508, 1998.
 - 25) 西井和夫, 酒井弘, 西野至, 浅野智弥: 京都市地下鉄東西線開通に伴う交通行動変化に関するパネル分析, 土木計画学研究・講演集, No. 22(2), pp. 687-690, 1999.
 - 26) 張 峻屹, 杉恵頼寧, 藤原章正, 奥村 誠: 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)(1)08305019 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.
 - 27) 山本俊行, 松田忠士, 北村隆一: 保有予定期間との比較に基づく世帯における自動車保有期間の分析, 土木計画学研究・論文集, No.14, pp. 799-808, 1997.
 - 28) 藤井 聡, 米田和也, 北村隆一, 山本俊行: パネルデータを用いた連続時間軸上の個人の離散選択行動の動的モデル化: 均衡状態を仮定しない動的な需要変動解析に向けて, 土木計画学研究・論文集, No. 15, pp. 489-498, 1998.
 - 29) 山本俊行, 木村誠司, 北村隆一: 取替更新行動間の相互影響を考慮した世帯の自動車取替更新行動モデルの構築, 土木計画学研究・論文集, No. 15, pp. 593 - 599, 1998.
 - 30) 飯田恭敬, 内田敬, 宇野伸宏: 交通情報の効果を考慮した経路選択行動の動的解析, 土木学会論文集, No. 470/IV-20, pp. 77-86, 1993.
 - 31) 小林 充, 永井 護, 本多 均, 洞 康之: 交通実験が交通手段選択に与える影響—パークアンドバスラ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.

- 18(2), pp. 485-488, 1995.
- 32) Duncan, G.J., Juster, F.T. and Morgan, J.N.: The role of panel studies in research on economic behavior, *Transportation Research A*, Vol. 21 No. 4, pp. 249-263, 1987.
- 33) Bunch, D. S., Brownstone, D. and Golob, T. F.: A Dynamic Forecasting System for Vehicle Markets with Clean-Duel Vehicles, *World Transport Research, Proceedings of the Seventh World Conference on Transportation Research, Vol 1, Travel Behavior*, Hensher, D., King, J. and Oum, T. eds. Elsevier Science, Oxford, pp. 189-203, 1996.
- 34) Kalton, G., Kasprzyk, D. and McMillen, D.B.: Nonsampling errors in panel surveys, *Panel Surveys*, Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. eds., John Wiley & Sons, New York, pp. 249-270, 1989.
- 35) Silberstein, A.R. and Jacobs, C.A.: Symptoms of repeated interview effects in the consumer expenditure interview survey, *Panel Surveys*, Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. eds., John Wiley & Sons, New York, pp. 289-303, 1989.
- 36) Golob, T.F., Kitamura, R. and Supernak, J.: A panel-based evaluation of the San Diego I-15 carpool lanes project, *Panels for Transportation Planning: Methods and Applications*, Golob, T.F., Kitamura, R. and Long, L. eds., Kluwer Academic Publishers, Boston, pp. 97-128, 1997.
- 37) Kitamura, R., Fujii, S. and Yamamoto, T.: The effectiveness of panels in detecting changes, Paper presented at the Fourth International Conference on Survey Methods in Transport, Oxfordshire, 1996.
- 38) Çinlar, E.: *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, 1975.
- 39) Singer, B. and Spilerman, S.: Social mobility models for heterogeneous populations, *Sociological Methodology 1973-74*, Costner, H. L. ed., Jossey-Bass, San Francisco, pp. 356-401, 1974.
- 40) Singer, B. and Cohen, J.E.: Estimating malaria incidence and recovery rates from panel data, *Mathematical Biosciences*, Vol. 49, pp. 273-305, 1980.
- 41) Singer, B.: Estimation of nonstationary Markov chains from panel data, *Sociological Methodology 1981*, Leinhardt, S., ed., pp. 319-337, 1981.
- 42) Carette, P.: Compatibility of multi-wave panel data and the continuous-time homogeneous Markov chain, *Applied Stochastic Models and Data Analysis*, Vol. 14, pp. 219-228, 1998
- 43) Lerman, S.R.: The use of disaggregate choice models in semi-Markov process models of trip chaining behavior, *Transportation Science*, Vol. 13, No. 4, pp. 273-291, 1979.
- 44) von Eye, A., ed.: *Statistical Methods in Longitudinal Research, Vol. II, Time Series and Categorical Longitudinal Data*, Academic Press, New York, 1990.

(2000.1.31 受付)

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.