

膨大なデータを基にした量子統計的アプローチによる行動分析*

The behavior analysis using massive PP data based on quantum statistics*

羽藤英二**, 武智環***

By Eiji HATO and Tamaki Takechi

1. はじめに

ネットワーク上の主体が自らの効用最大化を目指すとき、同一空間上の存在するその他の主体の存在をはずして考えることは難しい。同時に時間軸上に存在する主体自身の行動の影響を無視して意思決定過程の記述を考えることもまた現実的とはいえないだろう。たとえば、週末の行動を先に決定したことで平日の交通行動がなんらかの制約を受けるように、限られた週末の時間の使い方は平日の時間の使い方と互いの意思決定を緩やかに拘束しあうであろう。従前の研究ではこのような時間的空間的社会的相互作用を明示的にシステムとして記述する方法論を確立し得ていない。高齢化社会において求められる相互扶助型の交通システム、季節ごとの休日の変更による観光需要の政策的誘導や、スマートグリッドにおける電気自動車と連動したエネルギーマネジメント問題といったような政策課題を考える上で、こうした問題は無視できない。ネットワーク上のアクティビティ分析をより動学的な枠組みで捉えなおすことの意義は高いといえるだろう。

海外の交通センサスがAll the yearで行われることが多いのは、このような政策課題に配慮した場合、代表的な一日の交通調査が現実的な交通政策を考える上でもはや不十分であることが理解されているからであり、母集団の代表性のみを根拠にネットワーク上の動的な交通行動を静的なものとして取り扱うことのメリットはもはや薄いといわざるを得ない。

従来の交通行動モデルの世界では、交通行動に固有の日々変動の存在を認め、一連の連続日の行動の変動に個人内と個人間で仕分けして記述することの重要性が、Koppelman & Pas (1984) によって指摘されている。Kitamura (1990) のパネル分析の枠組みでは、こうした事実を発展的に捉え、差異から変化へと向かった行動モデルが変化をどのように記述し得るのかについて、差分方程式型の動的モデルの提案や、その検証は見られたもの

*キーワード：量子統計，プローブパーソン

**正員,工博, 東京大学大学院工学系研究科

***日立情報システム

東京都文京区本郷七丁目三番地十一号

E-mail:saito@bin.t.u-tokyo.ac.jp

の、実証的な研究が90年代進められてから以降、わが国の交通政策においてこうした研究の枠組みは殆ど存在しない。こうしたことの理由のひとつとして、動的なデータ収集の困難さをあげることができる。こうした問題に対して著者らはプローブパーソン技術の開発を1998年から行っている。プローブパーソン技術は移動体自身がセンサーなどの情報を用いて半自動的に人の行動データおよび周辺のネットワーク環境情報を収集する技術の総称である。なかでもBCALs(Behavioral Context Addressable Logger in the shell)は複数のセンサー情報を利用して人々の行動を自動記録する端末であり、9割程度の精度で行動履歴の識別が可能である(Hato, 2010)。Hato and Teratani(2006)はBCALsをはじめとするプローブパーソン技術によって収集された膨大な行動記録データを用いてアクティビティパターンの多様性に関する研究を行っている。またGonzalesら(2008)は、100,000人の携帯電話利用者の移動軌跡6ヶ月分を分析し、人の行動軌跡が時空間的規則性を示すこと、移動距離が時間に依存せず、頻繁に訪れる少数の地点に戻る傾向があることを示している。今、我々が関心するのは、膨大な関連しあう個人の行動データの蓄積から、どのような交通行動のよりよい理解が得られるかにある。本研究では携帯電話をベースにして収集可能な膨大な位置記録データを基にした全く新しい行動分析の可能性についての議論を行うことを目的としている。

2. フレームワーク

膨大な行動データと、その系列の中の関係性を踏まえたとき、微視的挙動の単純な積み上げで説明できないマクロ現象をどう記述すべきかという問題をここでは考えたい。このような問題は、たとえば台湾の集集地震の大規模化(1999年)における広域連鎖破壊が発生したような長距離相関の問題や、ルビジウム原子、ナトリウム原子のレーザー冷却によるボーズアインシュタイン凝縮と同じような問題と考えられる

従前の行動分析では、同一個人内の行動は互いに関連しあっているはずだが、こうしたデータをどのように取り扱うかについて明快な回答を与えない。これに対して限定合理性の立場にたつSimon, A(1981)は、複雑性は

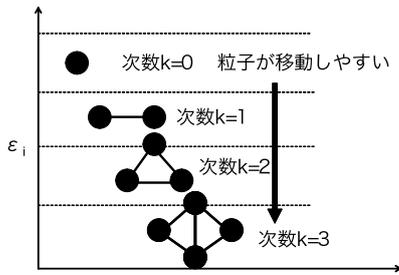


図-1 量子統計による状態推移の考え方

しばしば階層的な形態をとり、階層的システムはそれぞれのシステムの個別的内容から独立した共通の特性を持つこと、階層こそ複雑性の構築に利用される構造的仕組みの中心的なもの1つであることを指摘するとともに、相互作用の緊密度は空間的近接度によること示唆している。同じような問題をボーズアインシュタイン凝縮を例に考えよう。膨大な分子挙動を考えたいとき、分子間相互作用をシミュレーションで解いてもいいが、アボガドロ定数： $6.02 \times 10 \times 1E23$ は多すぎる。そこで、たとえば、ある分子が運動エネルギー $1/2mv^2$ と速度 v を持つ確率を考えよう。

$$\text{Prob}(\epsilon) \propto \exp(-\epsilon/kT) \quad (1)$$

温度が上がるほど ϵ =大(高エネルギー準位)の分子が増える。しかしこれは高温の話(マックスウェルボルツマン統計)で、低温になるとボーズ粒子では、2個以上の粒子が互いにひきつけあい(同じ状態に入る性質を有するのだが)高温では粒子はバラバラの動きを示すが、低温では非常に多くの粒子にひきつけあう力が働き、同じ状態に陥ることで、マクロレベルで確認可能な状態に遷移することを意味する(Chu, S. et al, 2001)。

このような分子の動きを考える際に、低温になると波長が伸びる性質($\lambda = h/(3mkT)^{1/2}$)を用いて、低温の場合はド・ブロイ波長が伸びて原子間距離に近づき、個別識別できなくなる(量子力学的には同種粒子化という)。同種粒子化するという事は、状態が原子の入れ替えに対して対称になるので、場合数が1個減ってひとつの状態に集まりやすくなることを意味する。エネルギー準位の分布関数は、以下の式(2)であらわすことができよう。

$$F(\epsilon) = 1/(\exp((\epsilon - \mu)/kT) - 1) \quad (2)$$

このような関係性を図-1に表す。加算型で制約条件なし(ネットワーク成長過程の仮説)の場合、日々どのような行動パターンをとるかについて、個人の中で最も平均的なパターン(頻度が多いパターン)を優先選択するという仮説を設ける。このとき優先度パラメータ η を

優先されやすい度合いをエネルギー準位と対応づけて考える($\beta \epsilon = -\ln \eta$)。このとき、平衡状態にあるネットワークはエネルギー準位が下がるほどつながりやすいことになるので、図-1のように量子統計における状態推移とアクティビティパターンの生起確率を関連付けて考えることができる。

ここで、式(3)に時刻 t あたりの移動パターンの発生頻度を k で表すことを考える。

$$\frac{\partial k_i}{\partial t} = m \prod (k_i) = \frac{m \eta_i k_i}{\sum_{j=1}^n \eta_j k_j} \quad (3)$$

ここで、発生頻度(次数) k が、時刻 t のべき乗で表されることを仮定すると式(4)のように次数 k は t と t_0 の関数で表すことができる。

$$k_{\eta_i}(t, t_0) = m \left(\frac{t}{t_0} \right)^{\beta(\eta_i)} \quad (4)$$

次に式(3)の分母を集団平均近似すると、

$$\sum_j \eta_j k_j \approx \int d\eta \eta \rho(\eta) \int_0^t dt_0 k_{\eta}(t, t_0) \quad (5)$$

式(5)のように分母をあらわすことが可能となる。次にこの式(5)を式(3)の分母に代入する。

$$\frac{\partial k_{\eta}}{\partial t} = \frac{\eta k_{\eta}}{ct} \quad (6)$$

このとき、時刻 t においてパターンが生起頻度が k 以上となる確率分布は次の式(7)で表されるので、これを展開すると、式(8)を導出できる。

$$P[k_{\eta}(t) > k] = P\left[t_0 < t \left(\frac{m}{k}\right)^{c/\eta}\right] \approx t \left(\frac{m}{k}\right)^{c/\eta} \quad (7)$$

$$\begin{aligned} p(k) &= -\int \frac{\partial P(k_{\eta}(t) > k)}{\partial k} \rho(\eta) d\eta \\ &= \int \left(\frac{m}{k}\right)^{c/\eta} \frac{ct}{\eta k} \rho(\eta) d\eta \propto k^{-\frac{c}{\eta}-1} \end{aligned} \quad (8)$$

式(8)からノードの次数(頻度)は超長期観測によって導かれる様々なトリップの発生原単位として解釈することが可能となる。

ここで重要なのは全数的な相互作用を行動モデルに当てはめ、協調ゲーム的なパラメータ推論を行うのではなく、ミクロとマクロの関係としていったん集約情報(エネルギー準位)に分解した上で扱うというSimonが定義するところのシステム準分解可能性を援用したアプローチだという点にある。時間的空間的ネットワーク上

表-1 調査データの概要

調査名称	調査期間	対象人数	レコード数
MPP2003	2003/01/29- 2003/02/28	84	8269
MPP2004	2004/01/26- 2004/02/29	311	32318
MPP2005	2005/02/01- 2005/02/28	374	46214
TPP	2005/11/20- 2006/02/20	43	9741
長期TPP	2006/03/01- 継続中	4	10109

※ レコード数はユニークな観測トリップ数を表す。

表-2 パターン数の分布

調査名称	総パターン数	ユニークな パターン数
MPP2003	1029	18.93
MPP2004	3584	18.15
MPP2005	5033	21.21
TPP2006	1270	21.90
超長期TPP	1197	20.93

の膨大な位置データを用いた行動分析では、状態の差異から変化を直接的に取り扱うのではなく、量子統計的アプローチに基づいて、マクロ統計的な行動パターンの多様性と安定性解析を行うことに主眼を置くことを考えた。

3. ケーススタディ

3. 1 データ概要

膨大な行動記録のデータを前提に、量子統計的アプローチに基づいて解析可能なトリップ特性について、実際のデータを用いて分析を行うことを試みる。用いたデータについて次に説明する(表-1)。著者らは長期観測データを用いた個人の行動分析を目的として、複数の行動記録データアーカイブの構築を目指してきた。松山プローブパーソン調査データは2003年から2006年まで愛媛県松山市で収集された行動記録データである(以下、MPP2003, MPP2004, MPP2005のデータとして定義する)。これらのデータの一部はパネルデータとして収集されている。一方東京プローブパーソン調査は2005年度の冬に3ヶ月間にわたって行われたもので(以下TPPとする)、TPPのうち4名については超長期PPとして約4年分のデータ蓄積がなされている。

3. 2 量子統計分析

式(7)に基づいてトリップ発生頻度の量子統計的分析を試みる。分析結果を表-2に示す。どのデータでも一ヶ月あたりの行動パターン数は平均で20前後の値を示している。3日のうち2日は同じパターンを示していることがわかる。

次に重複率 (Repetition rate) を定義し、分析を行う。重複率という指標は、「1人1日ごとにすべて異なるアクティビティパターンをしたと仮定し、出現するパターンのうち、どの程度重複しているアクティビティパターンが存在しているのか」を示す。重複率の算出方法は、出現するであろうアクティビティパターンの種類 (= 調査日数 × 人数) (N) における、いずれかのアクティビティパターンと重複しているアクティビティパターンの種類 (平均アクティビティパターン × 人数) (n) の割合として、以下の式 (9) と定義する。

$$R(\%) = \frac{n}{N} \times 100 = \frac{N-d}{N} \times 100 \quad (9)$$

ここで n は、出現するであろうアクティビティパターンから実際に出現したアクティビティパターン d を差引いたものと同値である。MPPP では 40%前後であった重複率に対して、TPP では 60%台になるなど東京において重複率が高い。性別では、MPPP2003, MPPP2004, MPPP2005 では女性の平均パターン数が多い。これは女性が多様な行動をとっていることを示している。MPPP2003 では、男性において 1457 人・日中で 891 種類出現している(重複率は 38.8%)。これに対して女性は 1147 人・日中 752 種類が出現しており重複率は 34.4%となっている。男性も女性も期間中の6割以上異なった生活パターンをとっていることがわかる。また、MPPP2005 の女性の場合、重複率 25.7%と低い値を示しており、行動パターンの多様性が伺える。一方、TPP では、女性の平均出現パターン数が少なく、重複率も高いため、似た行動を繰り返す傾向を示しているなど傾向に違いが見られた。

次に職業別の比較を行う。MPPP に関してはどの年の調査でも会社員の重複率が高く、定型のアクティビティパターンを示しているといえよう。一方 TPP では、主婦の出現パターン数が少なく、重複率も高い。学生、パート・アルバイトなどは自由な時間が多いと考えられ、比較的アクティビティパターンも多様化していると考えられる。また、自営業も重複率が低く、行動が多様化している傾向にある。

次に縦軸に次数 k を横軸にランクをとり対数をとったグラフを図-3に示す。観測頻度の多い MPP2004 と MPP2005 では、最大次数 $k > 1000$ を示しており、その他のデータ群に比べて大きな値を示していることがわかる。ランク1位のパターンはHWH型でありこれに頻度が集中していることがわかる。一方他のMPP2003やTPPにおいてもHWHにパターンが集中しているのは同様である。特に僅か4名の超長期データとTPP, MPP2003がほぼ同じ傾向を示して

図-2 アクティビティパターンの重複率

属性		調査名											
		平均出現パターン数/重複率/(サンプル数)											
		MPPP2003		MPPP2004		MPPP2005		TPP					
性別	男性	18.35	38.8%	(47)	17.56	41.5%	(199)	20.50	31.7%	(229)	22.58	64.7%	(22)
	女性	19.67	34.4%	(37)	19.18	36.1%	(112)	22.30	25.7%	(148)	21.19	66.9%	(21)
年代	10代			(-)			(-)			(-)	23.28	63.6%	(5)
	20代	13.47	55.1%	(12)	18.68	37.7%	(49)	21.93	26.9%	(56)	22.87	64.3%	(11)
	30代	19.85	33.8%	(35)	18.79	37.4%	(121)	21.84	27.2%	(153)	18.77	70.7%	(13)
	40代	20.66	31.1%	(29)	17.53	41.6%	(96)	20.39	32.0%	(115)	23.55	63.2%	(13)
	50代	17.00	43.3%	(7)	16.58	44.7%	(26)	19.61	34.6%	(30)	23.53	63.2%	(1)
	60代以上	15.48	48.4%	(1)	19.29	35.7%	(10)	20.09	33.0%	(16)			(-)
	不明			(-)	16.48	45.1%	(9)	24.34	18.9%	(7)			(-)
職業	主婦	20.55	31.5%	(21)	19.75	34.2%	(47)	25.02	16.6%	(34)	20.27	68.3%	(3)
	自営業	21.71	27.6%	(7)	20.57	31.4%	(19)	25.13	16.3%	(20)			(-)
	学生	15.48	48.4%	(1)	20.23	32.6%	(5)	23.69	21.0%	(9)	23.39	63.5%	(11)
	会社員	18.07	39.8%	(49)	16.94	43.5%	(209)	19.96	33.5%	(238)	21.30	66.7%	(26)
	パート・アルバイト			(-)	22.80	24.0%	(5)	21.99	26.7%	(48)			(-)
	その他	17.58	41.4%	(6)	21.86	27.1%	(26)	22.27	25.8%	(28)	23.29	63.6%	(3)
主な交通機関	自動車	19.10	36.3%	(73)	18.13	39.6%	(283)	21.23	29.2%	(356)	22.76	64.4%	(3)
	電車	13.55	54.8%	(1)			(-)	11.79	60.7%	(1)	20.96	67.3%	(24)
	バス			(-)			(-)	7.50	75.0%	(1)	7.07	89.0%	(1)
	バイク・原付	19.11	36.3%	(4)	17.20	42.7%	(15)	19.29	35.7%	(4)			(-)
	自転車	14.76	50.8%	(4)	19.90	33.7%	(9)	20.60	31.3%	(9)	24.81	61.2%	(4)
	徒歩	24.19	19.4%	(1)	18.86	37.1%	(4)	26.07	13.1%	(6)	24.01	62.5%	(11)
その他	22.26	25.8%	(1)			(-)			(-)			(-)	
居住地	都心部	18.95	36.8%	(70)	18.68	37.7%	(144)	21.67	27.8%	(310)	22.05	65.6%	(20)
	郊外部	18.80	37.3%	(14)	17.62	41.3%	(163)	19.05	36.5%	(67)	21.77	66.0%	(23)
総計		18.93	36.9%	(84)	18.15	39.5%	(311)	21.21	29.3%	(377)	21.90	65.8%	(43)

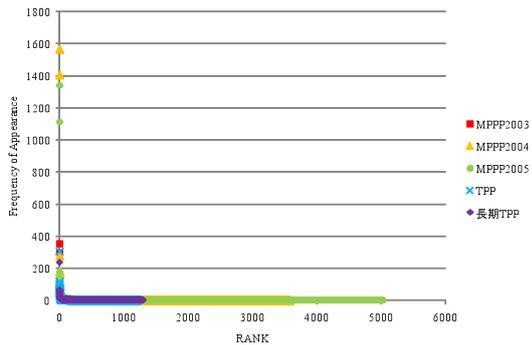


図-3 各サンプルの出現頻度分布

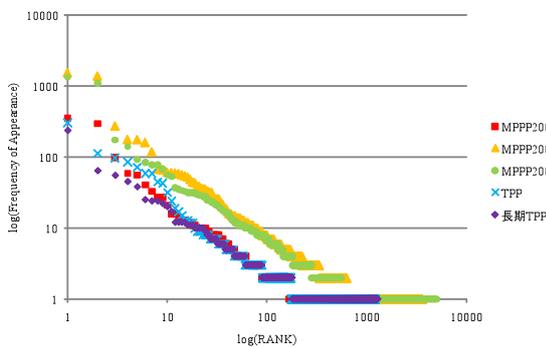


図-4 各サンプルの出現頻度分布(対数表示)

いることは興味深い。

東京と松山、異なる被験者数であってもアクティビティパターンの集中傾向については同一の傾向が示されるといえよう。また、すべてのサンプルについて、ラン

クと出現頻度の傾きはほぼ同じであり、式(8)で定義した個人の行動の多様性分布をエネルギー準位に都市間で大きな違いは見られないことが明らかとなった。

4. まとめ

本研究では量子統計的アプローチに基づいた行動分析の枠組みを示し、ある特定のアクティビティパターンにそのマクロ的分布が収斂している傾向を示した。属性や居住地によってあるばらつきはあるものの、このようなスケールフリー性はどのようなサンプルにも観測することができた。今後は、同一個人の超長期観測における個人の意思決定の相関分析をおこなっていく。

参考文献：

- 1) Pas, E. I., Subramanian Sunder: Intrapersonal variability in daily urban travel behavior: Some additional evidence, Transportation 22, pp.135-150, 1995.
- 2) Kitamura, R., Panel Analysis in Transportation Planning: An Overview, 1990.
- 3) Hato, E., Development of behavioral context addressable loggers in the shell for travel-activity analysis, Transportation Research C, Vol.18(1), Pages 55-67, 2010.
- 4) Hato, E. Evaluation of trip-activity pattern variability using probe person data, Transportation Research Board Annual Meeting, CDROM, 2006.
- 5) Gonzalez, M.C., and Hidalgo, C. A. and Barabasi, A. L. Understanding individual human mobility patterns, Nature 453, pp.779-782, 2008.
- 6) Simon, H.A. The Science of the Artificial, The Massachusetts Institute of Technology, 1981.