



Title	出生コーホートを用いた日本全国の位置情報と所得属性付き仮想個票の合成
Author(s)	原田, 拓弥; 村田, 忠彦
Citation	サイバーメディアHPCジャーナル. 2021, 10, p. 45-48
Version Type	VoR
URL	https://doi.org/10.18910/83288
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

出生コーホートを用いた日本全国の位置情報と所得属性付き仮想個票の合成

原田 拓弥¹⁾、村田 忠彦²⁾

¹⁾ 青山学院大学 理工学部 経営システム工学科

²⁾ 関西大学 総合情報学部 総合情報学科

1. はじめに

本研究では、母の生年別、出産時の年齢別、出生数の統計表である出生コーホートを用いて日本全国の仮想個票の合成に取り組む[1]。前年度公募型利用者制度で用いていた従来手法[2]では、ある年度において出生した子と親の年齢差の統計表を用いて仮想個票が合成されていた。したがって、全ての親子の年齢差が、ある1つの年度において調査された親子の年齢差に合わせて最適化されている。そのため、現実社会と異なる傾向をもつ仮想個票が合成される恐れがあった。本研究では、出生コーホートを用いることで親の出生年ごとの出生の傾向を反映した仮想個票を合成する手法を提案する。

国家的また国際的な災害対策や経済政策において、より精度が高く、きめの細かい対応が求められている。これらの分析と将来の可能性を可視化する社会シミュレーションへの関心が高まっている。これまでの多くの社会シミュレーションではモデルを単純化せよという Keep It Simple, Stupid (以下、KISS 原理) [3]に基づいてモデル化されていた。しかし、KISS 原理では現実社会の複雑な現象のモデル化は不可能であると指摘されている[4]。そのため、可能な限り忠実に現実社会を模倣するモデルを用いた社会シミュレーションが期待されている。

このようなモデルを作成するためには、環境のデータと市民のデータが必要となる。環境のデータは地理情報や地域メッシュといった地理情報システムの利用が可能である。このようなモデルにおいて、モデルの粒度を現実社会に近づけるほど、エージェントの意思決定においても、可能な限り現実社会を模倣する必要がある。

現実社会を可能な限り模倣するエージェントの意思決定の実現には様々な課題がある。その課題の 1

つがエージェントが保持する属性の設定である。エージェントの属性の設定に政府や行政が収集している戸籍や納税のデータを用いることができれば、現実社会と整合するエージェントの属性の設定ができる。しかし、これらの市民のデータは個人情報保護やプライバシーの観点から利活用が困難である。このような状況から、政府統計をはじめとする利用可能な統計情報から、仮想的な属性を持つ個人で構成される人工社会を生成し、その人工社会の中でどのような事象が発生するかを観察する社会シミュレーションが行われるようになっている[3, 4]。

統計情報に基づく個票データの合成に関する研究の歴史は古く、合成的再構成法 (Synthetic Reconstruction Method; 以下、SR 法) [5]として知られている。SR 法は、個票データのサンプルをもとに、Iterative Proportional Fitting Procedure (IPFP)を用いて個票データを合成している。なお、本稿では個票データの「復元」ではなく「合成」という用語を用いている。復元の場合、実際の人口構成と同一の個票の復元が期待されるが、合成される個票はあくまでも統計的特徴が類似した仮想的な個票である。Lenormand と Deffuan は[6]、サンプルを用いて合成する SR 法と、サンプルを用いない合成手法とを比較し、後者が個人と世帯をよりよく合成できていることを示した。

これらの海外の研究ではそれぞれの国において利用可能な統計表と特徴に基づいた手法が開発されており、他の地域へ適用する際には留意が必要と指摘されている[7]。日本においても、日本の利活用可能な情報と特徴に基づいた仮想個票が合成されている。日本における合成手法として、国勢調査のサンプルを用いた花岡の手法[8]とサンプルを用いない池田の手法[9]がある。

花岡の手法[8]は探索手法の 1 つである Simulated Annealing (以下 SA 法) を用いて、複数の統計表と整合するように国勢調査のサンプルの組み合わせを探索している。そのため、サンプルを用いない手法と比べると短時間で妥当な解を探索できる。しかし、手法に国勢調査のサンプルを使用しているため、国勢調査のサンプルの利用規約により第三者提供できない。一方で、池田らの手法[9]を基にした手法は公開されている統計情報に適合するように SA 法を用いて最適化している。したがって、合成された仮想個票を第三者提供可能である。

これらの国内外の研究を踏まえ、本研究では第三者提供可能な日本の仮想個票を合成するために、池田らの手法[9]をもとにした手法[2]を改良する。従来手法[2]は SA 法を用いて適合させる統計表に、親子の年齢差や夫婦年齢差、人口分布などの統計表を用い、統計表と仮想個票の差を最小化している。親子の年齢差として合成対象年度の人口動態統計における「父の年齢別、出生数」と「母の年齢別、出生数」の統計表を用いている。従来手法[2]では、これらの統計表に整合するように父子と母子のすべての関係を最適化していた。しかし、これらはある 1 つの年度において出生した 0 歳の子とその親について、親の年齢別に集計された統計表である。そのため、従来手法[2]では、全ての親子の年齢差が、ある 1 つの年度において調査された親子の年齢差に合わせて最適化されている。

生年別の出生に関する統計表として出生コホートが公開されている。出生コホートとは 1947 年以降、各年の「母の年齢別、出生数」の統計表を「母の生年別、出産時の年齢別、出生数」で整理した統計表である。本研究では従来手法の年齢差の統計表の代わりに出生コホートを用いて仮想個票を合成する。出生コホートは日本全国を対象に集計された統計表のみ公開されている。出生コホートを各都道府県の規模に合わせて調整し、都道府県単位で合成するなど、統計表を過度に調整することは好ましくない結果が得られる[2]。そのため、本研究では、出生コホートを考慮し、日本全国を一度に合成す

る手法を提案する。

2. 従来手法

従来手法[2]は、統計情報を基に作成した仮想個票を複数の統計表に適合させる手法である。個人の年齢や親子の年齢差の統計表に対して、コンピュータ上で再現した世帯構成のデータ集合（仮想個票）の統計値との誤差を計算し、SA 法を用いて誤差を最小化している。仮想個票は複数の世帯とその構成員である個人によって構成される。著者らの手法は、統計表の対象地域と同じ人口規模の仮想個票の合成を試みている。

著者らは、対象地域と同じ人口規模の仮想個票を合成するために、世帯数や人口などの統計表通りに初期世帯を合成している。初期世帯合成後、以下の手続きにより仮想個票と統計表との差を SA 法により最小化する。

- Step 1 仮想個票内の同じ家族類型かつ性別の個人をランダムに 2 人選択する。
- Step 2 選択した 2 人の年齢を入れ替える。
- Step 3 仮想個票と統計表との差を計算する。
- Step 4 メトロポリス法により解を遷移判定する。
- Step 5 探索回数に達するまで Step 1 に戻る。

Step 3 では、次式を用いて仮想個票と統計表との差を計算している。

$$f(A) = \sum_{s=1}^S \sum_{j=1}^{G_s} |c_{sj} - R_{sj}| \quad (1)$$

ここで、 A は仮想個票、 S は最適化に用いる統計表の数、 G_s は統計表 S の項目数、 R_{sj} は統計表 S の項目 j における統計量、 c_{sj} は統計表 S の項目 j における仮想個票から作成した仮想の統計量である。

Step 3 では次の統計表との差を計算している。

- 統計表 1 父子年齢差
- 統計表 2 母子年齢差
- 統計表 3 夫婦年齢差
- 統計表 4 家族類型別、男女別の人口分布
- 統計表 5 男女別人口分布

3. 提案手法

本研究では、従来手法の「統計表 1 父子年齢差」と「統計表 2 母子年齢差」を出生コーホートの以下の統計表に置き換える。

統計表 1' 父の生年別、父子年齢差別、出生数

統計表 2' 母の生年別、母子年齢差別、出生数

出生コーホートとは「親の生年別、出生時の親の年齢別、出生数」が記載された統計表である。出生コーホートから 1932 年生まれ、1952 年生まれ、1972 年生まれ、1982 年生まれの母に関して抽出した図を図 1 (a)に示す。図 1 の横軸は出産時の年齢を縦軸は出生数を示している。また、各折れ線は母の生年を示している。図 1 (a)から母の生年ごとの折れ線の傾向が変化している。特に、出生数のピークの年齢が 26 歳付近から 29 歳付近に変化しており、1932 年生まれや 1952 年生まれに比べ、1972 年生まれの女性は出産時の年齢が高くなっている。

年齢差の統計表に関して従来手法により合成された、日本全国の仮想個票を用いて作成した「母の生年別、母子年齢差別、出生数の統計表を図 1 (b)に示す。ここで、出生数とは、仮想個票に含まれる同じ世帯の母と子の年齢の組み合わせから、子を出生した時の母の年齢ごとに集計して求めた値である。横軸は出生時の母の年齢を、縦軸は出生数を示している。図 1 (b)の生年別の各線は、出生数が異なるものの概ね同じ形状をしている。これは、2010 年に出生した子と母の年齢差の統計に対して、全ての母子年齢を最適化した結果である。

図 1 (a)と図 1 (b)を比較すると、分布の形状と y 軸の値が異なる。図 1 (a)は出生した子を集計しているが、図 1 (b)は合成時の出生数を集計している。図 1

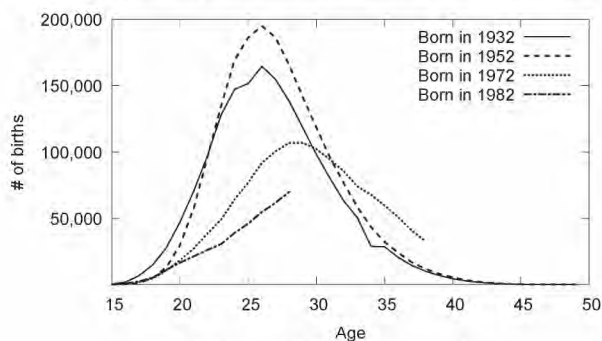
(a)は子の出生時の親の年齢別の統計表であり、図 1 (b)の集計は、世帯分離や親（または子）の死別後の世帯を対象としているため、図 1 (b)は出生数が少なく集計されている。また、母の年齢が 5 歳階級の統計表を用いていたため、15 歳差から 19 歳差、20 歳差から 24 歳差などの各区間の境界では値の変化量が大きい。加えて、図 1 (b)のピークは約 31 歳差であるが、図 1 (a)では生年毎に異なる。

図 1 (a)と図 1 (b)で示したように、出生数と仮想個票から集計される出生数は異なる。本研究では、出生コーホートを、国勢調査から推定した「生年別、父の数」と「生年別、母の数」を用いて調整する。

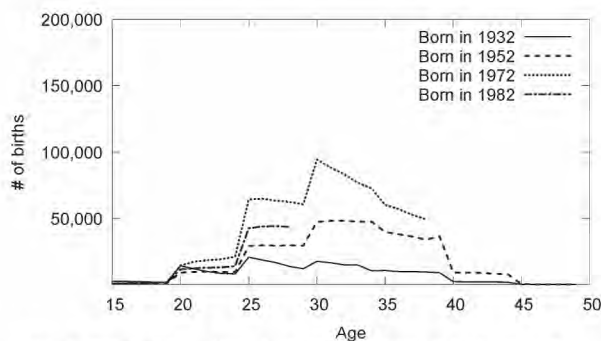
「生年別、父の数」と「生年別、母の数」の推定には国勢調査 人口等基本集計 表 15 を用いる。国勢調査 人口等基本集計表 15 は「世帯主との続柄（12 区分）、世帯の家族類型（16 区分）、年齢（5 歳階級）、男女別一般世帯人員」について集計されている。この統計表の家族類型別の年齢と世帯主との続柄を用いて、生年別の父と母の数を推定し、合成データ内の父/母の総数と合うように調整する。「男女別、年齢別、父/母の数」を用いて出生コーホートの「生年別、出生時の親の年齢別出生数」を調整する。具体的には、「男女別、年齢別、父/母の数」の年齢から生年を算出し、生年毎に父/母の数と合うように出生時の親の年齢別出生数を調整する。

4. 実験結果

本研究では、国勢調査をもとに日本全国の世帯構成を合成する。合成対象は平成 27 年国勢調査から 50,962,785 世帯、その人口は 115,552,530 人である。探索回数は 1 人あたり 1 回、10 回、100 回、500 回



(a) 母に関する出生コーホート



(b) 仮想個票から作成した母子に関する出生コーホート

図 1：出生コーホートと仮想個票から作成したコーホートの比較

表 1：統計表との誤差

探索回数	誤差の総和	1人あたりの誤差
1回 / 人	130,463,564.6	1.1290
10回 / 人	77,217,391.3	0.6682
100回 / 人	27,555,307.3	0.2385
500回 / 人	23,614,503.8	0.2044

とした。総探索回数はそれぞれ、115,552,530 回、1,155,525,300 回、11,555,253,000 回、57,776,265,000 回である。SA 法の設定として、初期温度を 10.0、収束温度を 0.1 と設定し、冷却関数には指数冷却を用いた。

統計表との誤差を表 1 に示す。表 1 から探索回数を増加するほど統計表との誤差を削減できているが、その効果は限定的である。これは、出生コーホートを用いることで親の出生年毎との出生の傾向を考慮できるものの、提案手法における出生コーホートの調整が不十分であったことが原因である。

5. おわりに

本研究では、従来手法で使用していた年齢差の統計表を出生コーホートに置き換える手法を提案した。従来手法で用いていた年齢差の統計表はある年に出生した 0 歳の子とその親の年齢について集計された統計表である。従来手法では年齢差の統計表を全ての親子の組み合わせに対して適用していた。そのため、親の出生年ごとの出生の傾向が異なる仮想個票が合成されていた。

本研究では、出生コーホートを用いることで親の出生年ごとの出生の傾向を考慮した合成手法を提案した。出生コーホートは親の生年別、出生時の親の年齢別、出生数について集計された統計表である。本研究では出生数を親子の組み合わせ数として用いるため、出生コーホートを親の年齢別、父/母の数を推定し、調整して用いた。出生コーホートを用いて合成するためには、出生コーホートの調整方法を十分に検討する必要がある。

著者らは図 2 に例示する日本全国全ての市区町村において仮想個票の合成に成功しており、公的目的

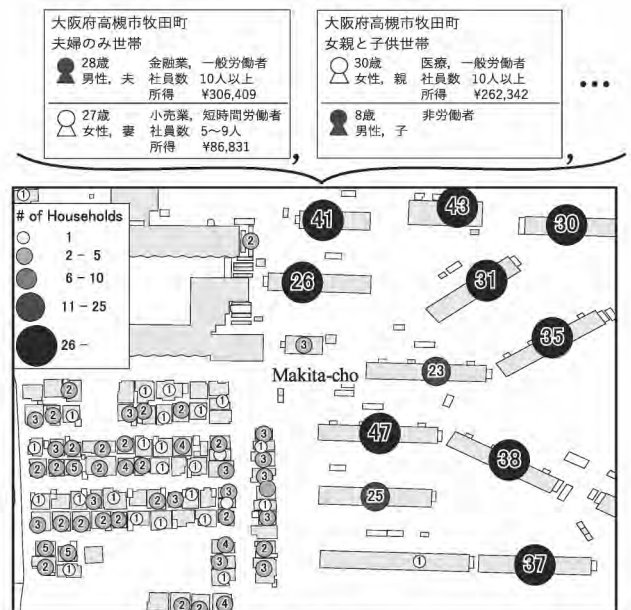


図 2：仮想個票の一例
(この地図は国土地理院の基盤地図情報を用いた。)

に限り無償での提供を開始している。利用を希望する研究者の方がおられれば文献[12]の Web フォームから申請いただきたい。

参考文献

- (1) 原田拓弥, 村田忠彦, 計測自動制御学会 システム情報・部門 学術講演会 2019, 554–559 (2019).
- (2) 原田拓弥, 村田忠彦, 計測自動制御学会 システム・情報部門 学術講演会 2018, 6 pages, (2018).
- (3) R. Axelrod, Princeton University Press, (1997).
- (4) 寺野隆雄, 人工知能学会誌, **18**, 6, 710–715, (2003).
- (5) 市川学, 計測と制御, **57**, 6, 710–715, (2013).
- (6) Y. Goto, Proc. of 2018 IEEE Int. Conf. on Systems, Man, and Cybernetics, 1144–1149 (2018).
- (7) A. G. Wilson and C. E. Pownall, Area, **8**, 4, 246–254, (1976).
- (8) M. Lenormand and G. Deffuant, Journal of Artificial Societies and Social Simulation, **16**, 4, 1–9, (2013).
- (9) F. Gargiulo, et. al., PLoS One, **5**, 1, 266–279, (2010).
- (10) 花岡和星, 地域安全学会論文集, **29**, 247–255, (2016).
- (11) 池田心, 他, 第 43 回システム工学部研究会, 11–14 (2010).
- (12) 合成人口データ提供, <http://www.res.kutc.kansai-u.ac.jp/~murata/rsss-distribution/>