出生コーホートを考慮した日本全国の仮想個票の合成

○原田拓弥　村田忠彦 （関西大学）

Synthesizing Real-Scale Populations of Whole Japan considering Birth Cohort

∗T. Harada and T. Murata (Kansai University)

概要– 本研究では親の生年別，出生時の親の年齢別，出生数の統計表である出生コーホートを用いた仮想個票合成手法を提案する．従来手法では親子の年齢差の統計表として，ある年に出生した 0 歳の子とその親の年齢別に集計された統計表から年齢差を計算し，全ての親子に対して適用していた．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．

キーワード: Real-Scale Social Simulation, Synthetic Population, 統計, 出生コーホート

1　はじめに

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 [1]．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）[2] 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 [3]．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可能である．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 [4),5)]．

統計情報に基づく個票データの合成[1] に関する研究の歴史は古く，Synthetic Reconstruction Method (SR 法)[6] 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) [7] を用いて個票データを合成している．その後，数多くの個票データ合成法が提案されているが，基本的に SR 法に基づく，個票データのサンプルを用いたアルゴリズムとなっている．Barthelemy ら [8] は，IPFP の弱点として，個人の統計表と世帯の統計表のどちらかに適合する合成ができたとしても，両方に適合する合成が困難であることを指摘している．この課題を解決するため，Gargiulo ら [10] や Barthelemy ら [8] は，サンプルを用いない合成手法を提案している．Lenormand と Deffuan は [9]，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 [10]．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 [11] とサンプルを用いない池田ら [12] 及びその派生手法 [13)–16)] がある．これらの手法の比較を Table 1 に示す．まず，合成される仮想個票の情報に着目すると，花岡 [11] の手法は合成に費やす計算時間は少なく，サンプルデータが保持する 28 属性を保持した仮想個票を合成できる．一方，池田ら [12] 及びその派生手法 [13)–16)] では，合成に費やす時間は多く，仮想個票が保持する属性数は 5 属性である．花岡 [11] は 2005 年の国勢調査のサンプルデータ[2]を用いて，2015 年の国勢調査結果に整合する集団をサンプルデータの組み合わせにより仮想個票を合成している．もととなる国勢調査のサンプルデータは 28 属性を保持していることに加え，花岡 [11] の手法ではサンプルデータの内容は変更せず，組み合わせ方を最適化していることで，短い処理時間で多くの属性を保持した仮想個票の合成できている．一方，池田ら [12] 及びその派生手法 [13)–16)] では，世帯構成員の属性を変更して最適化している．そのため，合成に時間がかかることに加え，一度に多くの属性を合成することが困難である．

次に，合成した仮想個票の利活用に着目すると，花岡 [11] の手法により合成された仮想個票は一定期間のみ利活用可能であり，第三者提供は不可能である．これは，花岡 [11] の手法が用いる国勢調査のサンプルデータは申請と承認の後に一定期間のみ利活用でき，利用期間終了後はサンプルデータの返却と中間データ[3]を削除

[1] ここで，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．

[2] 国勢調査のサンプルデータ [17] では，個票データから情報の削除や入れ替えを行い，約 1%（約 124 万件）抽出したデータが提供されている．

[3] サンプルデータの個々の情報を判別できるもの．

Table 1: サンプルを用いる手法 [11] と用いない手法 [12)–16)] の比較

	サンプル有 [11]	サンプルなし [12)–16)]
計算時間	少ない	多い
属性数	28 属性	5 属性
合成方法	2005 年の 1%抽出個票から，2015 年の国勢調査小地域集計に適合する組み合わせを Simulated Annealing 法を用いて探索	2015 年の国勢調査を用いて世帯を生成し，年齢差や人口分布に整合するように Simulated Annealing 法を用いて構成員の年齢を変更し探索
第三者提供	不可	可
合成個票の利活用	一定期間のみ	長期間

する必要があるためである．そのため，利用期間終了後には合成した仮想個票を削除する必要がある．また，サンプルデータの利用者の追加・交代は審査と承認を得る必要があり，不特定多数への仮想個票の提供は不可能である．一方，池田ら [12] 及びその派生手法 [13)–16)] では，公開されている情報のみを用いているため，このような制限は存在しない．これらの手法により合成された仮想個票が公開されることで，シミュレーションを実施する研究者が独自に仮想個票を合成する必要がなくなる．

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 [12] をもとにした手法 [16] を改良する．従来手法 [16] は Simulated Annealing（以下，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」[18] と「母の年齢別，出生数」[19] の統計表を用いている．従来手法 [16] 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 [16] 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ーホート [20] が公開されている．出生コーホートとは 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 [16] の年齢差の統計表の代わりに出生コーホートを用いて仮想個票を合成する．出生コー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．出生コー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 [21]．そのため，本研究では，出生コーホートを考慮し，日本全国を一度に合成する手法を提案する．

2 従来手法

2.1 概要

従来手法 [16] は，統計情報を基に作成した仮想の世帯構成を，複数の統計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情報に対する，コンピュータ上で合成した世帯構成のデータ集合（仮想個票，合成データ）の誤差を計算し，SA 法を用いて誤差を最小化する．

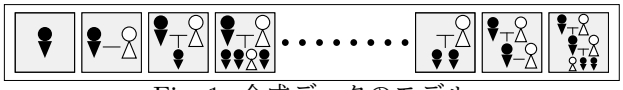


Fig. 1: 合成データのモデル

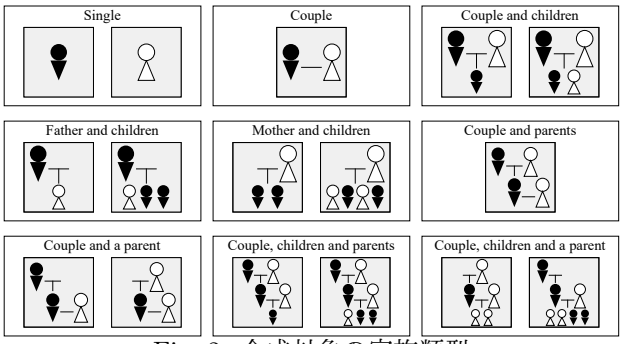


Fig. 2: 合成対象の家族類型

合成データは複数の世帯とその構成員である個人によって構成している．合成データのモデルを Fig. 1 に示す．従来手法 [16] は，統計情報に基づく世帯構成の合成を試みている．対象とする家族類型[4]を世帯員と世帯主の続柄により区分した分類である．従来手法 [16] は Fig. 2 に示す 9 種類の家族類型を合成の対象としている．なお，これらの 9 種類の家族類型の世帯数で日本全体の全世帯数の約 95%を占めている．その他の世帯数の家族類型は，他の親族を含む世帯や兄弟姉妹のみからなる世帯，非親族を含む世帯，他に分類されない世帯である．これらの家族類型は，公開されている複数の統計情報に整合させることが困難である．

世帯の中には世帯を構成する構成員が存在している．それぞれの個人は年齢，性別，所属する家族類型，世帯の役割，親族関係の 5 つの属性を持っている．世帯の役割は，所属する世帯の中での役割を表している．

2.2 世帯構成の合成法

従来手法 [16] は以下の 4 つの要素から成り立っている．

要素 1 初期世帯生成法
要素 2 SA 法により適合させる統計表と調整方法
要素 3 目的関数
要素 4 最適化の手続き

2.2.1 要素 1 初期世帯生成法

従来手法 [16] は対象地域と同じ規模の世帯構成を合成するために，対象地域を集計対象とする統計表を用いて世帯数や人口などの統計情報通りに初期世帯を生

[4]家族類型とは，一般世帯を世帯員と世帯主の続柄により区分した分類である．一般世帯とは，住居と生計を共にしている人の集まりや一戸を構えて住んでいる単身者，間借り，下宿，独身寮に住んでいる単身者の世帯である．

成している．その際に,「家族類型，世帯人員別世帯数」の統計表である国勢調査 人口等基本集計 [22] 表番号 11 を用いている．表番号 11 を考慮することにより，合成した世帯数と国勢調査が相違ない初期世帯を合成できる．また，個人の性別と初期の年齢を設定する際には，人口分布の統計表である国勢調査 人口等基本集計 [22] 表番号 16-1 と表番号 16-2 を用いている．これらの属性は統計表の男女の数や年齢別の人口に相違が発生させずにランダムに設定する．詳細は文献 [16] を参照されたい．

2.2.2 要素 2 適合させる統計表と調整方法

従来手法 [16] は SA 法を用いて合成データと複数の統計表の差を最小化している．従来手法 [16] では，統計表の調査範囲を参考に仮想の統計表を集計している．以下の統計表は合成対象の都道府県下のすべての市区町村を対象に集計している．

P-1 父子の年齢差[5]
P-2 母子の年齢差[6]

人口 20 万人以上の市区は以下の統計表を人口 20 万人以上の市区別に集計している．

BC-1 夫婦の年齢差[7]
BC-2 9 種類の家族類型別，男女別，人口分布[8]

人口 20 万人未満の市町村は以下の統計表を集計している．なお，統計表 SC-1 は国勢調査 人口等基本集計 [22] 表番号 17 を参考に，人口 20 万人未満の市を市部として集計し，町と村を郡部として集計している．また，統計表 SC-2 と統計表 SC-3 は人口 20 万人未満の市町村別に集計している．

SC-1 夫婦の年齢差[7]
SC-2 9 種類の家族類型別，男女別，人口分布[8]
SC-3 男女別の人口分布[9]

統計表 P-1，統計表 P-2 は都道府県単位の統計表のみ公開されている．そのため，仮想の統計値 v^c_{sj} を作成する際においても都道府県下の全ての個人を対象に集計する．統計表 BC-1 と統計表 SC-1 は人口 20 万人以上の市区と市と区を合算した市部，町と村を合算した郡部を集計した統計表が公開されている．人口 20 万人未満の市に居住する個人は市部，人口 20 万人未満の町と村に居住する個人は郡部単位の仮想の統計表を作成する．統計表 SC-2 は 5 歳階級である．そのため，人口 20 万人未満の市町村は 1 歳階級の人口分布である統計表 SC-3 も最適化の対象に加えている．

統計表 P-1，P-2，BC-1，SC-1，SC-3 は全ての人口を対象に集計されており，9 種類の家族類型を対象とする本研究及び従来手法 [16] では統計表を調整する必要がある．統計表 P-1，P-2 は合成データ内の父子/母子の数に合わせて調整している．統計表 BC-1，SC-1 は合成データ内における対象地域の夫婦の数に合わせて調整している．統計表 SC-3 は合成データ内における対象地域の人口に合わせて調整している．詳細は文献 [15),16)] を参照されたい．

2.2.3 要素 3 目的関数

仮想個票合成手法では実統計表と合成データから作成する仮想の統計表との差を次式のように定式化し，目的関数を最小化している．

$$f(A) = \sum_{s}^{S} f_s(A) \tag{1}$$

ここで，式（1）の A は合成データ，S は統計表の数である．$f_s(A)$ は実統計表の各項目と仮想の統計表の各項目の差を定式化した関数である．$f_s(A)$ として，花岡 [11] が用いている式（2）の 2 乗誤差と，従来手法 [16] が用いている式（3）の絶対誤差がある．

$$f_s(A) = \sum_{j}^{G_s} (v_{sj} - r_{sj})^2 \tag{2}$$

$$f_s(A) = \sum_{j}^{G_s} |v_{sj} - rsj| \tag{3}$$

ここで，G_s は統計表 s の項目数，v_{sj} は仮想の統計表 s の項目 j の値，r_{sj} は統計表 s の項目 j の値である．絶対誤差の目的関数である式（3）は実統計表との誤差が直感的に理解しやすい．合成データの利活用先を考えると，同じ誤差の合成データにおいて，特定の項目に多くの誤差をもつ合成データより誤差が平均的に発生した合成データの方が好ましい．そのため，本研究では 2 乗誤差の式（2）を用いて最適化する．なお，本研究では，統計表との誤差を示す際に式（2）の 2 乗誤差と式（3）の絶対誤差の両方を示す．

2.2.4 要素 4 最適化の手続き

従来手法 [16] では以下の手続きにより SA 法を用いた最適化を行う．

Step 1 合成データを初期生成する（要素 1）．
Step 2 統計表 P-1，P-2，BC-1，SC-1，SC-3 を初期解に合わせて調整する（要素 2）．
Step 3 初期解の目的関数値を計算する（要素 3）．
Step 4 探索回数が規定数に達するか，$f(A) = 0$ になるまで探索する．
Step 5-1 合成データ内の個人をランダムに 1 人選択する（要素 4）．
Step 5-2 後述する条件に該当する個人をランダムに 1 人選択する（要素 4）．

---

[5] 人口動態職業・産業別統計 出生 [18] 表番号 2，15 歳～49 歳（5 歳階級）と 50 歳以上の項目の都道府県毎の総数を抽出．

[6] 人口動態統計 出生 [19] 表番号 5-2，15 歳～49 歳（5 歳階級）と 50 歳以上の項目の都道府県毎の総数を抽出．

[7] 国勢調査 人口等基本集計 [22] 表番号 17，−70 歳～+70 歳の 1 歳階級の年齢差毎の夫婦の数を人口 20 万人以上の市区と市部・郡部のみ抽出し，年齢差を計算．なお，市部の統計表から人口 20 万人以上の市区の値を取り除いている．

[8] 国勢調査 人口等基本集計 [22] 表番号 16-1 もしくは表番号 16-2 より年齢不詳の項目を除いて調整し，家族類型，男女別の人口を抽出．人口 20 万人以上の市区は 1 歳階級，人口 20 万人未満の市町村は 5 歳階級である．

[9] 国勢調査 人口等基本集計 [22] 表番号 3-2，1 歳階級の人口分布を人口 20 万人未満の市町村のみ抽出し，同じ市区町村の統計表 (4)～統計表 (21) の年齢別の人口及び全市区町村の年齢別の人口の総和が都道府県の統計表 (4)～(21) が整合するように調整．

Step 5-3 Step 5-1 と Step 5-2 で選択した個人の年齢を交換する（要素 4）.

Step 6 目的関数値を再計算し解をメトロポリス法により遷移判定する（要素 3）.

Step 7 探索回数を更新して SA の温度を冷却する.

Step 8 Step 4 の処理に戻る.

Step 5-2 では，Step 3-1 で選択した個人によって選択する個人の条件を変更する．Step 5-1 で選択された個人が人口 20 万人以上の市区に属している場合，Step 5-2 では，選択されている個人と同じ市区かつ家族類型かつ性別の個人をランダムに選択する．一方，Step 5-1 で選択された個人が人口 20 万人未満の市町村に属している場合，以下の候補からランダムに選択する．

- Step 5-1 で選択された個人と同じ市町村かつ家族類型かつ性別の個人.
- 人口 20 万人未満の市町村の個人のうち，Step 5-1 で選択された個人と同じ家族類型かつ性別かつ年代の個人.

ここで，同じ年代とは，0 歳〜4 歳，5 歳〜9 歳 など，5 歳階級の年齢が等しいという意味である.

3 提案手法

本研究では，従来手法の要素 2 適合させる統計表の「P-1 父子年齢差」と「P-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.

J-1 父の生年別，父子年齢差別，出生数

J-2 母の生年別，母子年齢差別，出生数

加えて，従来手法 [15),16)] の最適化の手続きのうち，Step 2 を以下の様に置き換える.

Step 2 統計表 J-1，J-2 を初期解に合わせて調整する（要素 2）.

3.1 出生コーホートと従来手法 [15)] の比較

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．母に関する出生コーホートから 1933 年生まれ，1953 年生まれ，1973 年生まれ，1983 年生まれの母を抽出した図を Fig. 3(a) に示す．Fig. 3(a) 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．Fig. 3(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，1933 年生まれや 1953 年生まれに比べ，1973 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 [16)] と同様の統計表を用いた手法 [15)] 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[10]」の統計表を Fig. 3(b) に示す．横軸は出生時の母の年齢を，縦軸は出生数を示している．Fig. 3(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，全ての母子関係を最適化した結果である.

Fig. 3(a) と Fig. 3(b) を比較すると，分布の形状と y 軸の値が異なる．Fig. 3(a) は出生した子を集計しているが，Fig. 3(b) は合成時の出生数を集計している．Fig. 3(a) は子の出生時の親の年齢別の統計表であり，Fig. 3(b) の集計には，世帯分離や親（または子）の死別後の世帯が親子関係に含まれていないため，Fig. 3(b) の方が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，Fig. 3(b) のピークは約 31 歳差であるが，Fig. 3(a) では生年毎に異なる．Fig. 3(c) から Fig. 3(d) については 3.3 節で，Fig. 3(e) から Fig. 3(g) については次章で言及する.

3.2 母子に関する出生コーホートの補完と父子に関する出生コーホートの作成

従来手法で用いていた年齢差の統計表はある年に生まれた 0 歳の子について，その親の年齢別に集計されている．そのため，ある年の親子の年齢差の統計表を全ての年代に対して適用すると，現実社会と異なる傾向を持つ集団が合成される恐れがある．そのため，本研究では,「生年別，出生時の親の年齢別，出生数」の統計表を用いて最適化する.

母子に関する出生コーホートは平成 22 年度人口動態統計 特殊報告 [20)] に公開されている．しかし，特殊報告であるためか 2011 年以降は出生コーホートが公開されていない．また，父子に関する出生コーホートも公開されていない.

そこで，2015 年の国勢調査を用いて合成する本研究では，母子に関する出生コーホートの補完と父子に関する出生コーホートを作成する．母子に関する出生コーホートでは，2010 年から 2015 年に調査された人口動態統計 [23)] の中巻 表 7 を転写する．人口動態統計 [23)] の中巻 表 7 は「出生数，性・母の年齢（各歳）・出生順位・嫡出子—嫡出でない子別」について集計されている．転写方法は，例えば，1994 年生まれの女性は 2011 年では 17 歳であるため，2011 年の人口動態統計の 17 歳の項目を出生コーホートの 1994 年生まれの列の 17 歳のセルに転写する．同様に，それぞれの調査年ごとに生年を算出し，転写することで 2010 年に公開された 2009 年時点の出生コーホートを補完できる．なお，2009 年以前の人口動態統計の中巻 表 7 の補完時の方法と同様に生年を算出し出生コーホートと比較したところ，同一の値が記載されていた.

父子に関する出生コーホートは人口動態統計 [23)] の中巻 表 8[11]．を用いて作成する．人口動態統計 [23)] の中巻 表 8 は「嫡出出生数，父の年齢（各歳）・母の年齢（各歳）・出生順位別」について集計されている．この統計表と同様の統計表の集計が開始された 1952 年調査から 2015 年調査の統計表を用いて父子に関する出生コーホートを作成する作成方法は母子に関する出生コーホートの保管方法と同様に，1952 年調査から 2015

[10] ここで，出生数は具体的には，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.

[11] 表番号は年毎に異なる．1952 年から 1965 年は「出生 表 11」，1966 年と 1967 年は「出生 表 9」，1968 年は「出生 表 10」，1969 年から 1978 年は「出生 表 7」，1979 年以降は「中巻 表 7」である．なお，1980 年以降の人口動態統計は Web 上（e-Stat）で入手できる．1979 年以前の人口動態統計は国立国会図書館で閲覧できる.

年調査の各表から生年を算出し，転写している．人口動態統計 [23] の中巻 表 8 の父の年齢は 17 歳から記載されているため，1935 年以降生まれの男性を対象とし，父子に関する出生コーホートを作成した．

3.3 出生コーホートの調整

Fig. 3(a) と Fig. 3(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，父–子の数と母–子の数をもとに調整する方法（以下，親–子の数による調整法）と生年別の父と母の数をもとに調整する方法（以下，父母の数による調整法）の 2 つに取り組む．

3.3.1 父–子の数と母–子の数をもとに調整する手法

父–子の数と母–子の数をもとに調整する手法では，父子に関する出生コーホートを合成データ内の父–子の数を用いて調整し，母子に関する出生コーホートも同様に母–子の数を用いて調整する．具体的には，父子/母子それぞれの出生コーホートを父–子の数/母–子の数と整合するように調整する．

3.3.2 生年別父の数と生年別母の数をもとに調整する手法

生年別父の数と生年別母の数をもとに調整する手法では,「生年別，父の数」と「生年別，母の数」を推定し，推定した統計表を用いて出生コーホートを調整する.「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 [22] 表 15 を用いる．国勢調査 人口等基本集計 [22] 表 15 は「世帯主との続き柄 (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する.「夫婦と子供」,「男親と子供」,「女親と子供」,「夫婦と両親」,「夫婦とひとり親」世帯では，世帯主との続柄として「世帯主」,「子の配偶者」,「世帯主の父母」,「世帯主の配偶者の父母」の項目を男女別かつ年齢別に加算する．なお,「男親と子供」は男性のみ「女親と子供」は女性のみ加算する.「夫婦，子供と両親」,「夫婦，子供とひとり親」世帯では「世帯主」,「子の配偶者」,「世帯主の父母」,「世帯主の配偶者の父母」,「祖父母」の項目を男女別かつ年齢別に加算する．これら「7 種類の家族類型別，男女別，年齢別，父/母の数」から「男女別，年齢別，父/母の数」を計算し，合成データ内の父/母の総数と合うように調整する．

その後に,「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．

3.3.3 調整後の統計表の比較

親–子の数による調整法で調整した統計表を Fig. 3(c) に，父母の数による調整法で調整した統計表を Fig. 3(d) に示す．横軸は年齢を縦軸は出生数を示している．親–子の数により調整した Fig. 3(c) は Fig. 3(d) に比べ 1933 年生と 1953 年生の出生数が高めに，1973 年生と 1983 年生の出生数が低めにでている．これは，生年別，年齢別の出生数の 2 次元の統計表を単一の値である親–子の数に整合するように調整したため，Fig. 3(a) に示したもとの統計表の親の生年別の傾向が反映されたからである．

4 実験結果

本研究では，平成 27 年度の国勢調査をもとに日本全国の世帯構成を合成する．合成対象世帯は 50,962,785 世帯，その人口は 115,552,530 人である．本研究が用いた計算機の CPU は，AMD Ryzen Threadripper 1950X（3.4 GHz, 16 コア）で，メモリは DDR4-2400 16 GB×8，OS は Microsoft Windows 10 Pro 64 bit である．

親–子の数による調整法について，式（3）の絶対誤差で評価した統計表と合成データの誤差の総和を Table 2 に，式（2）の 2 乗誤差で評価した誤差の総和を Table 3 に示す．また，父母の数による調整法について，式（3）の絶対誤差で評価した統計表と合成データの誤差の総和を Table 4 に，式（2）の 2 乗誤差で評価した誤差の総和を Table 5 に示す．本研究が用いた出生コーホートは従来手法 [16] で用いていた年齢差の統計表と大きく異なる．最適化後に双方の手法で評価した場合，他の手法での評価結果は大きく悪化する．そのため，本研究では提案手法を用いて探索回数を変化させた比較をする．探索回数は 1 人あたり 1 回，10 回，100 回，500 回とした．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．なお，試行回数は 8 回であり，Table 2 から Table 5 は平均値を示している．また，Table 2 から Table 5 の統計表 BC-1，BC-2，SC-1，SC-2，SC-3 は各統計表毎に市区町村別の統計表との誤差の総和を示した．

Table 2 と Table 4 を，また，Table 3 と Table 5 を比較すると，父母の数による調整法の Table 4 と Table 5 が親–子の数による調整法と比べ統計表との誤差を削減できている．これは，生年別に合成データに合わせた父の数を推定したことで，単位の値で調整した親–子の数による調整法と比べ統計表との誤差の削減が容易だったと考えられる．しかし，詳細に比較するためにはより多くの探索回数を設定した実験が必要である．

親–子の数による調整法で合成した仮想個票から作成した母子に関する出生コーホートを Fig. 3(e) に，父母の数による調整法で合成した仮想個票から作成した母子に関する出生コーホートを Fig. 3(f) に示す．Fig. 3(e) と Fig. 3(f) の横軸は年齢を縦軸は出生数を示している．Fig. 3(e) と Fig. 3(f) 共に探索回数 1 人あたり 500 回の結果より作成した．従来手法 [16] から作成した Fig. 3(b) の母子に関する出生コーホート（探索回数を 1 人あたり 10 万回に設定）と比較すると，Fig. 3(e) と Fig. 3(f) のどちらも成年毎の傾向の違いが反映された仮想個票を合成できている．

一方で，Fig. 3(e) は Fig. 3(f) を比較すると，前者は 1953 年生の出生数が高めに，1973 年生と 1983 年生は低めにでている．しかしながら，どちらも適合させる統計表（Fig. 3(c) と Fig. 3(d)）と比較すると 1983 年生の出生数全体的に高めにでている．また，1933 年生まれと 1953 年生の 43 歳以上も統計表と比べると高めにでている．これは，Fig. 3(e) と Fig. 3(f) のどちらも

Table 2: 親–子の数による調整法の統計表との誤差の総和（式（3）の絶対誤差，8 回平均）

統計表	1 人あたりの探索回数			
	1 回	10 回	100 回	500 回
J-1	133,861,197.3	90,809,048.8	44,690,982.8	37,715,772.0
J-2	42,563,827.5	28,838,658.8	17,849,272.0	16,205,882.5
BC-1	51,380,372.5	34,051,917.3	21,149,154.5	19,666,474.0
BC-2	0.0	0.0	0.0	0.0
SC-1	19,303,656.5	12,248,075.8	1,689,403.3	720,251.5
SC-2	18,300,957.8	13,642,888.5	3,379,846.8	713,155.8
SC-3	2,312,383.0	2,027,508.5	623,306.3	410,008.3
合計	267,722,394.5	181,618,097.5	89,381,965.5	75,431,544.0

Table 3: 親–子の数による調整法の統計表との誤差の総和（式（2）の 2 乗誤差，8 回平均）

統計表	1 人あたりの探索回数			
	1 回	10 回	100 回	500 回
J-1	70,162,524,598,028.5	8,162,216,921,609.0	642,677,153,384.5	506,421,673,038.8
J-2	35,765,508,288,667.7	3,810,598,031,001.8	258,463,877,287.3	192,535,727,371.5
BC-1	34,101,885,390,045.5	4,218,667,451,648.3	380,134,095,830.5	313,744,575,317.5
BC-2	0.0	0.0	0.0	0.0
SC-1	56,637,843,470.3	20,601,833,639.0	192,579,581.5	22,840,087.8
SC-2	238,452,618,561.5	112,320,117,193.3	3,883,801,018.5	117,054,909.5
SC-3	40,457,283.5	29,488,126.8	2,799,666.8	1,475,352.5
合計	140,325,049,196,057.0	16,324,433,843,218.0	1,285,354,306,769.0	1,012,843,346,077.5

探索回数が少なく，初期解で発生していた統計表との誤差を削減し切れていないことが原因である．初期解から作成した母子に関する出生コーホートを Fig. 3(g) に示す．Fig. 3(g) から 1933 年生と 1953 年生の 49 歳の項目が突出している．これは，49 歳以上の年齢差を 49 歳の項目として集計したためである．本研究では式（2）の 2 乗誤差の目的関数で最適化したことにより，最適化の過程で突出した 1933 年生と 1953 年生の 49 歳の項目が削減された．しかし，探索回数が不足したことから，Fig. 3(e) と Fig. 3(f) では統計表との差が十分に削減できていない．そのため，より多くの探索を実施する必要がある．

5 おわりに

本研究では，従来手法 [16] 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．出生コーホートの出生数と仮想個票から作成する出生コーホートの出生数は集計方法が異なるため，本研究では，親-－子の数により出生コーホートを調整する方法と親の年齢別，父/母の数を推定し，出生コーホートを調整する方法を比較した．

実験結果から後者が前者より統計表との誤差を削減できていた．しかし，詳細に比較するためにはより多くの探索回数を設定し実験する必要がある．

加えて，本研究では日本全国を一度に合成するため，合成に非常に時間がかかる．本研究が実施した 1 人あたり 500 回の探索回数で計算に約 27 時間かかった．著者らの経験から誤差を十分に削減するためには 1 人あたり 10 万回の探索回数が必要である．提案手法を用いて 1 人あたり 10 万回の実験を実施すると約 8ヶ月かかる．そのため，高速に合成するための手法や探索アルゴリズムも検討する必要がある．

謝辞

本研究の一部は，科学技術融合振興財団，JSPS 科研費 17K03669 の助成を受けたものです．また，本研究成果の一部は，大阪大学サイバーメディアセンターの大規模可視化対応 PC クラスタ（VCC）を利用して得られたものです．

参考文献

1) J. M. Epstein and R. Axtell: *Growing Artificial Societies: Social Science from the Bottom Up*, The MIT Press, 1st edition (1996)

2) R. Axelrod: *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, (1997)

3) 寺野：エージェントベースモデリング：KISS 原理を超えて，人工知能学会誌，**18**-6，710/715 (2003)

4) 市川：医療分野におけるリスクマネジメント　地理情報分析と社会シミュレーション技術を用いた検討，計測と制御，**57**-6，407/412 (2018)

5) Y. Goto: Stylized Fact Analysis of Cash-for-Work Programs in the Disaster Reconstruction Process, *Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics*, 1144/1149 (2018)

6) A. G. Wilson and C. E. Pownall: A New Representation of the Urban System for Modelling and for the Study of Micro-Level Interdependence, *Area*, **8**-4, 246/254 (1976)

Table 4: 父母の数による調整法の統計表との誤差の総和（式（3）の絶対誤差，8 回平均）

統計表	1 人あたりの探索回数			
	1 回	10 回	100 回	500 回
J-1	133,375,336.8	88,288,701.0	30,294,105.0	24,118,029.8
J-2	42,234,391.8	27,423,723.8	9,937,878.3	8,807,294.5
BC-1	51,218,669.0	32,908,052.0	14,834,217.5	13,413,265.3
BC-2	0.0	0.0	0.0	0.0
SC-1	19,304,764.8	12,208,331.0	1,672,497.5	575,754.5
SC-2	18,306,070.0	13,655,885.0	3,217,353.5	925,477.8
SC-3	2,311,441.3	2,092,709.3	632,158.3	396,237.8
合計	266,750,673.5	176,577,402.0	60,588,210.0	48,236,059.5

Table 5: 父母の数による調整法の統計表との誤差の総和（式（2）の 2 乗誤差，8 回平均）

統計表	1 人あたりの探索回数			
	1 回	10 回	100 回	500 回
J-1	69,954,618,651,960.5	7,782,935,127,191.3	267,424,734,369.0	188,720,300,575.8
J-2	35,710,217,982,157.0	3,651,255,227,046.3	76,334,493,871.3	50,871,579,208.5
BC-1	33,949,106,573,102.0	3,998,650,251,895.8	187,420,835,349.5	137,676,923,614.5
BC-2	0.0	0.0	0.0	0.0
SC-1	56,667,594,060.8	20,606,350,726.3	180,863,733.3	15,319,685.0
SC-2	238,586,082,796.8	112,392,301,481.0	3,485,795,485.8	155,037,013.5
SC-3	40,419,844.0	30,996,042.0	2,745,929.3	1,441,054.3
合計	139,909,237,303,921.0	15,565,870,254,382.5	534,849,468,738.0	377,440,601,151.5

7) W. E. Deming and F. F. Stephan: On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known, *The Annals of Mathematical Statistics*, **11**-4, 427/444 (1940)

8) J. Barthelemy and P. L. Toint: Synthetic Population Generation Without a Sample, *Transportation Science*, **47**-2, 266/279 (2013)

9) M. Lenormand and G. Deffuant: Generating a Synthetic Population of Individuals in Households: Sample-Free Vs Sample-Based Methods, *Journal of Artificial Societies and Social Simulation*, **16**-4, 12 (2013)

10) F. Gargiulo, S. Ternes, S. Huet, and G. Deffuant: An Iterative Approach for Generating Statistically Realistic Populations of Households, *PLOS ONE*, **5**-1, 1/9 (2010)

11) 花岡：全国版の小地域マイクロデータの構築と災害分析への活用，地域安全学会論文集，**29**，247/255 (2016)

12) 池田, 喜多, 薄田：地域人口動態シミュレーションのためのエージェント推計手法，第 43 回システム工学部研究会，11/14 (2010)

13) 福田, 喜多：エージェントベースの人口推計モデルにおける属性決定手法の評価，システム制御情報学会論文誌，**27**-7，279/289 (2014)

14) 枡井, 村田：統計データからの市民の属性復元のための進化計算と SA による 2 段階最適化，システム制御情報学会論文誌，**30**-6，216/227 (2017)

15) T. Murata, T. Harada, and D. Masui: Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method without Samples, *SICE JCMSI*, **10**-6, 513/519 (2017)

16) 原田, 村田：世帯合成法における世帯構成員の年齢と役割を考慮した初期世帯と近傍解生成法の改良，計測自動制御学会システム・情報部門 学術講演会 2018，1/6 (2018)

17) 統計センター：公的統計のミクロデータの利用，`http://www.nstac.go.jp/services/archives.html`

18) 総務省統計局：人口動態職業・産業別統計 出生，`http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&listID=000001108272` (2013)

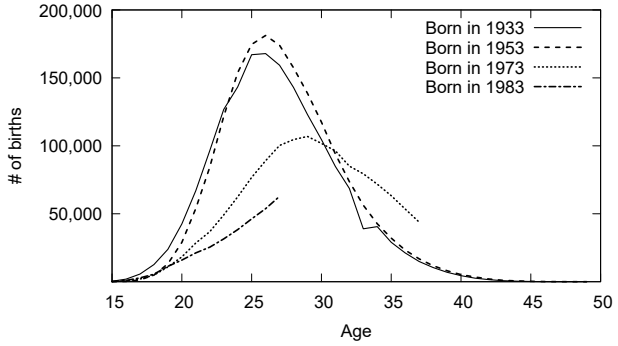
19) 総務省統計局：人口動態統計 出生 年次 2010 年度，`http://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&bunya_l=02&tstat=000001028897&cycle=7&year=20100&month=0&tclass1=000001053058&tclass2=000001053061&tclass3=000001053074&tclass4=000001053084` (2011)

20) 総務省統計局：平成 22 年度　人口動態統計特殊報告　出生に関する統計 `https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450013&tstat=000001040871&cycle=8&tclass1=000001040872` (2015)

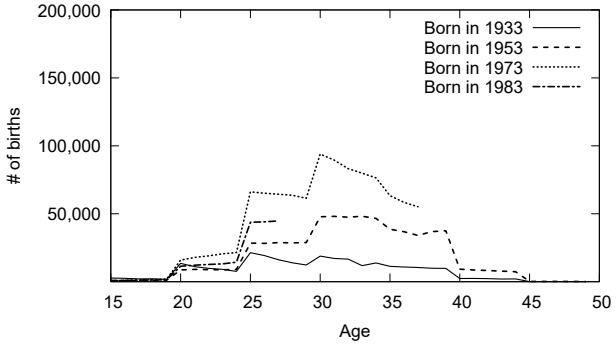
21) 原田, 村田：市区町村の統計表を考慮した都道府県単位の仮想個票の合成，計測自動制御学会第 15 回社会システム部会研究会，30/35 (2018)

22) 総務省統計局：平成 22 年度国勢調査 人口等基本集計　全国結果，`http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001034991` (2011)

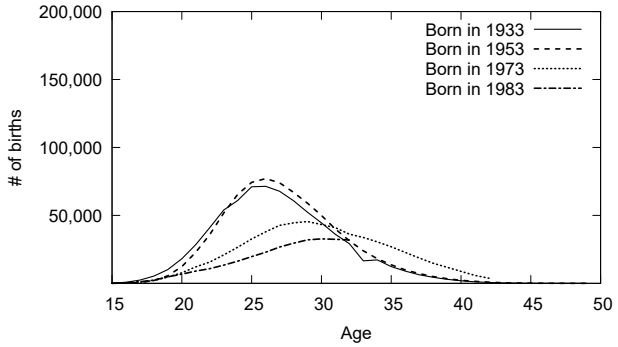
23) 総務省統計局：人口動態統計 出生 年次 2015 年，`https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&bunya_l=02&tstat=000001028897&cycle=7&year=20150&month=0&tclass1=000001053058&tclass2=000001053061&tclass3=000001053064` (2016)



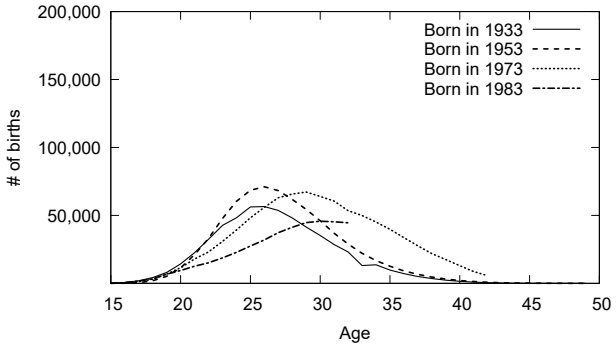
(a) 母に関する出生コーホート [20)]



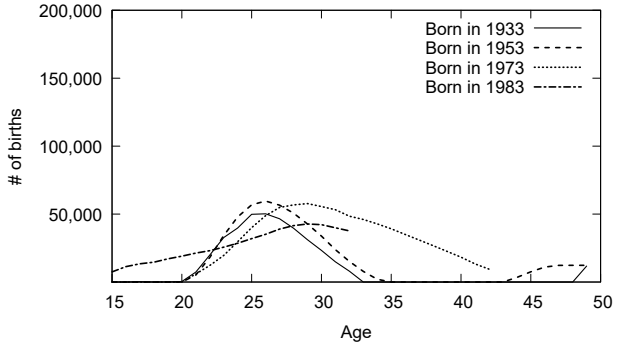
(b) 従来手法 [15)] から作成した母子関係のコーホート



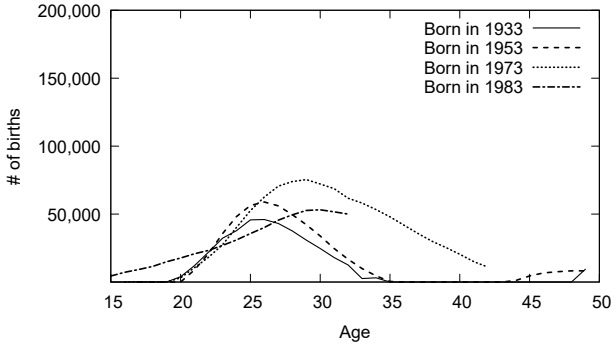
(c) 親–子の数による調整法により調整した母子関係のコーホート



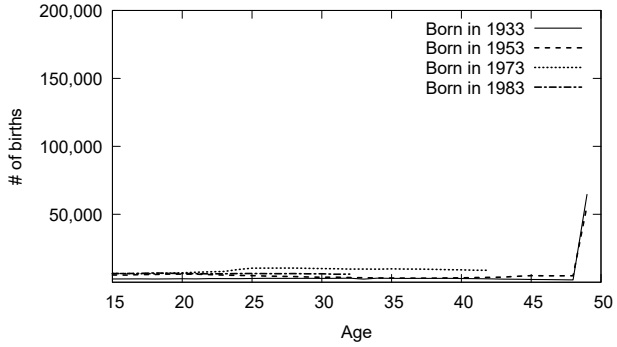
(d) 父母の数による調整法により調整した母子関係のコーホート



(e) 親–子の数による調整法による仮想個票から作成した母子関係のコーホート



(f) 父母の数による調整法による仮想個票から作成した母子関係のコーホート



(g) 初期解生成時の母子関係のコーホート

Fig. 3: 出生コーホート [20)] と仮想個票から作成した母子に関するコーホートの比較