

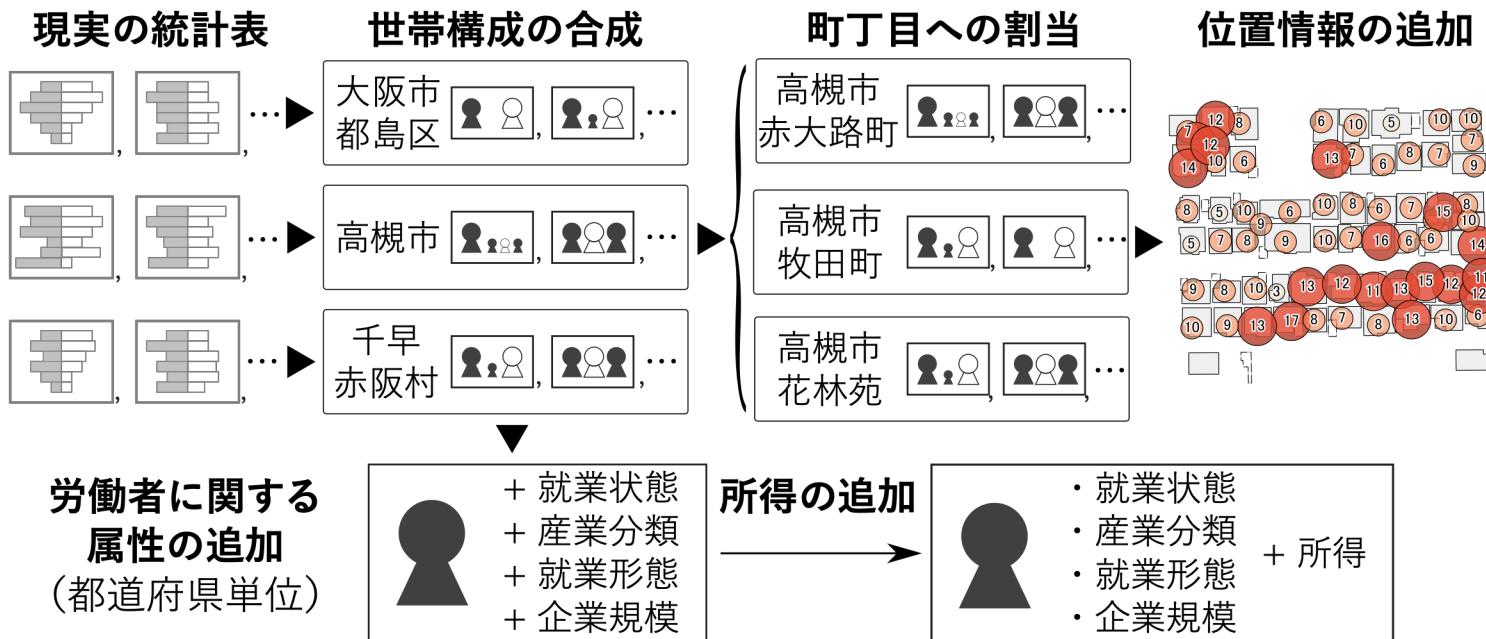
リアルスケール社会 シミュレーションの実現に向けて

村田 忠彦
関西大学 総合情報学部
関西大学ソシオネットワーク戦略研究機構
(全国共同利用・共同研究機関)



本講演の目的

仮想個票の合成には、大阪大学サイバーメディアセンターの計算機を、
地図情報作成には国土地理院基盤地図情報を利用しています。



分析・シミュレーション対象となる国・自治体・コミュニティの実人口の
仮想個票データ(年齢, 性別, 世帯構成, 産業分類, 所得)
を研究者に提供し、開発コストを低減することです。
仮想個票データの利用希望者は、murata@kansai-u.ac.jpまでご連絡ください。



リアルスケール 社会シミュレーションの課題



現実



将来



Challenge 1: モデル化



モデル

Challenge 2: 内部要素の決定

Challenge 4: シナリオ分析

A

B

C

Challenge 3: 計算速度と再現性



社会シミュレーションをより現実的に

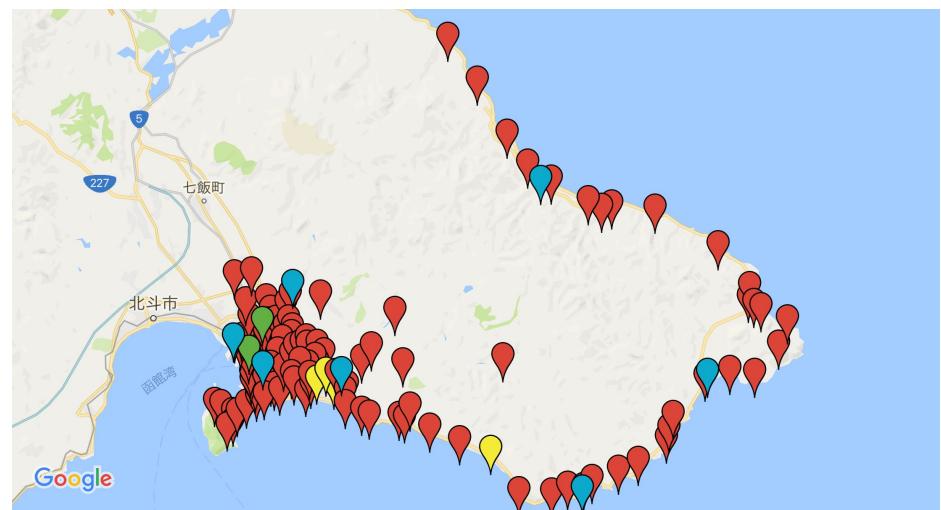
Example



地図データ ©2019 Google、ZENRIN

2016年の参議院議員通常選挙において、共通投票所が導入された。
どれくらい効果的か？
新たな選挙制度の有効性を検証する。

2016年7月の選挙で共通投票所を導入した
のは4つの自治体のみ。
北海道函館市が最も大きな自治体だった
(232,352人)。



地図データ ©2019 Google、ZENRIN



Challenge 1: モデル化

投票区 i の有権者 j の効用関数のモデル化

数理政治学の活用

投票に対する期待効用モデル (Downs1957, Riker1968)

$$R = P \cdot B - C + D$$

R : 投票・棄権

P : 選挙結果への影響の主観的確率

B : 候補者間の効用差

C : 投票にかかるコスト

D : 投票行動に対する満足感

どのように変数の”値”を定義するか？



Challenge 1: モデル化

投票区 i の有権者 j の効用関数のモデル化

$$D_{vij} = \alpha_i \cdot U_{vij} - (1 - \alpha_i) \cdot U_{aij}$$

投票の効用 棄権の効用

If 投票の効用 > 棄権の効用

→ $D_{vij} > 0$, 投票する.

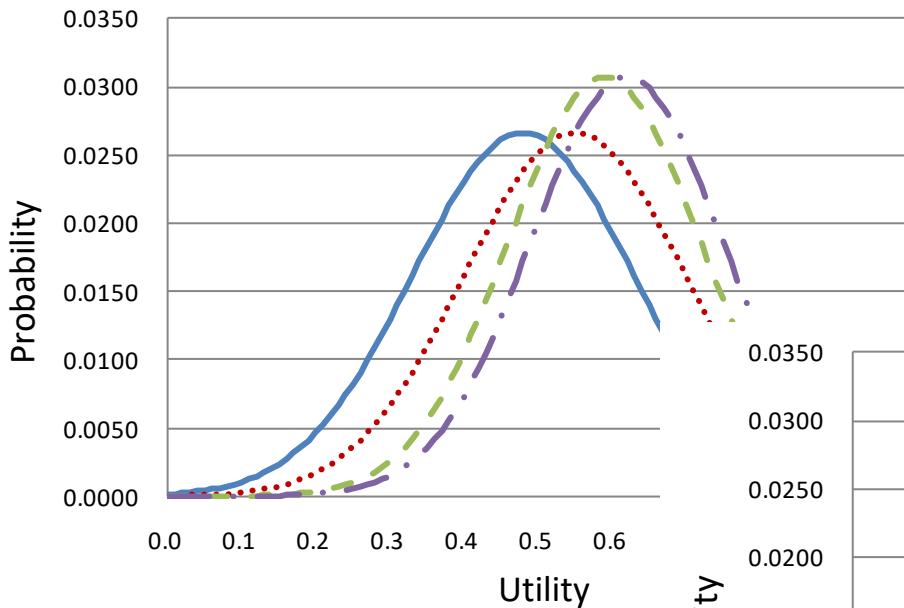
If 投票の効用 ≤ 棄権の効用

→ $D_{vij} \leq 0$, 棄権する.

K. Konishi, T. Murata, R. Natori, "Examination of regional voting parameter in voting simulation," Proc. of 1st International Workshop on Aware Computing, pp. 670-673, 2009.

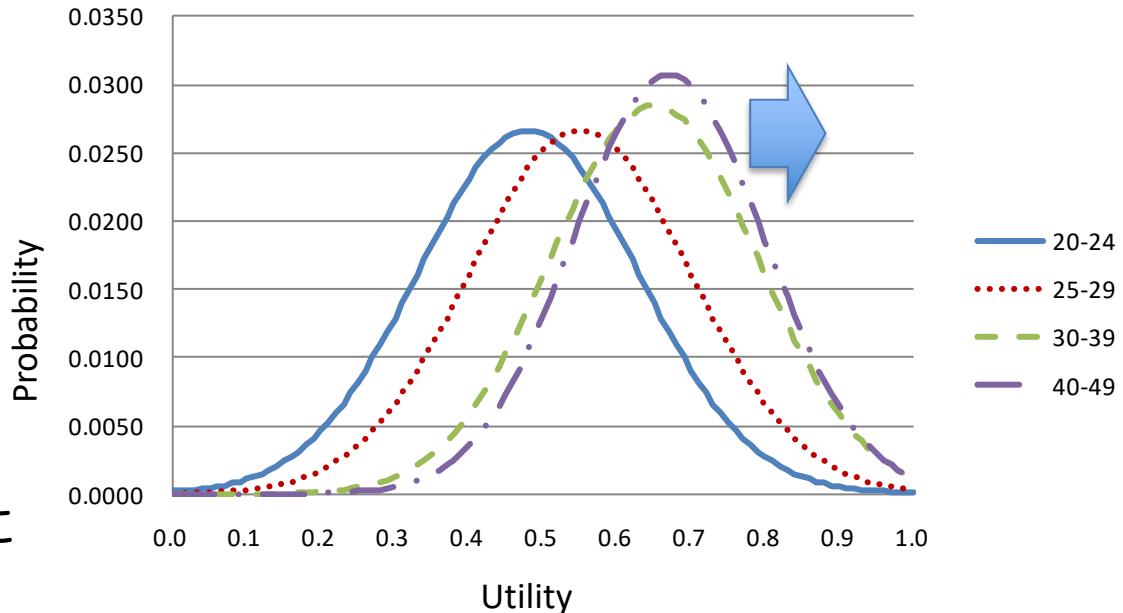


投票の効用 各世代の投票に対する関心



2010年の効用分布
明るい選挙推進協会のデータに
基づいて作成

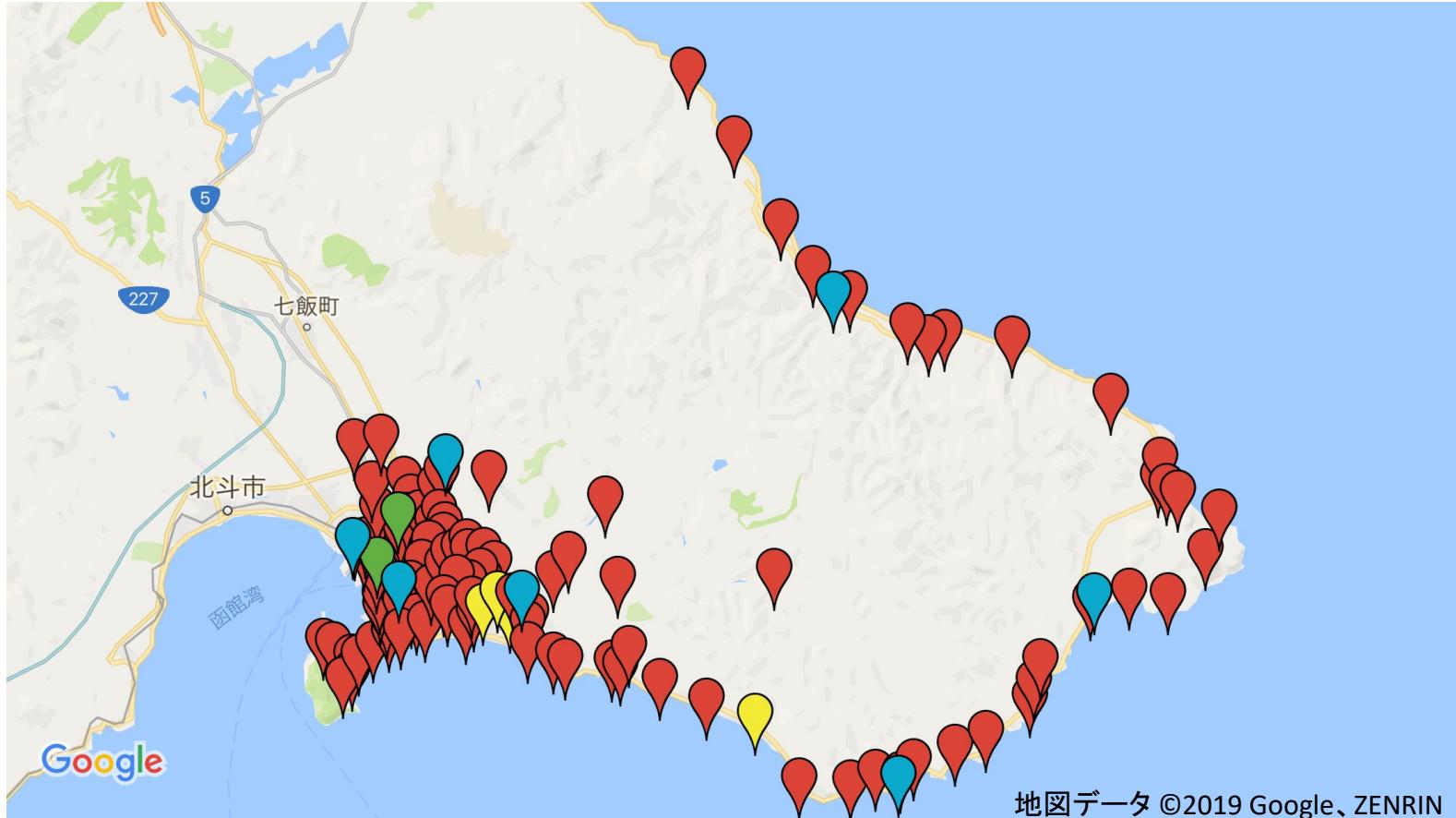
30代, 40代
の関心が高い。



2013年の効用分布
明るい選挙推進協会のデータに
基づいて作成



棄権の効用 投票所までの距離



遠ければ遠いほど、棄権する確率が高くなる。



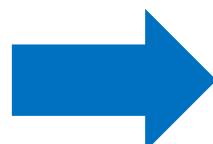
$$D_{vij} = \alpha_i \cdot U_{vij} - (1 - \alpha_i) \cdot U_{aij}$$

投票の効用 棄権の効用

Model 1: 2010年の投票結果から α_i の値を決める.

Model 2: 2013年の投票結果から α_i の値を決める.

- 投票の効用: 世代ごとの投票に対する関心
- 棄権の効用: 投票所までの距離
- 投票区 i の係数: 2010年と2013年の結果



上記の2つのモデルを使って
2016年の選挙結果を推定する.



2010年と2013年の 選挙の同定結果を用いた推定投票率

	実投票率	推定投票率
2010	44.585	44.579
2013	37.413	37.400

各年度の投票結果に合うように同定が
できている。



Challenge 1: モデル化における課題

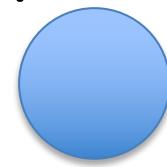
- ・ 社会科学における数理的アプローチをエージェントの設計に活用する
- ・ どんな粒度でモデルを設計するか？
- ・ 数理的アプローチにおける変数の値をどうするか？
- ・ 地域ごとの特性をどのように反映させるか？
- ・ 各分野の社会科学者と社会シミュレーション研究者のコラボレーションが必要



リアルスケール 社会シミュレーションの課題



現実



将来



Challenge 1: モデル化



モデル

Challenge 2: 内部要素の決定

Challenge 4: シナリオ分析

A

B

C

Challenge 3: 計算速度と再現性



Challenge 2: 内部要素の決定

- 実社会のデータをどのように決定するか？

人的データ:居住者(市民)と組織(会社)の情報

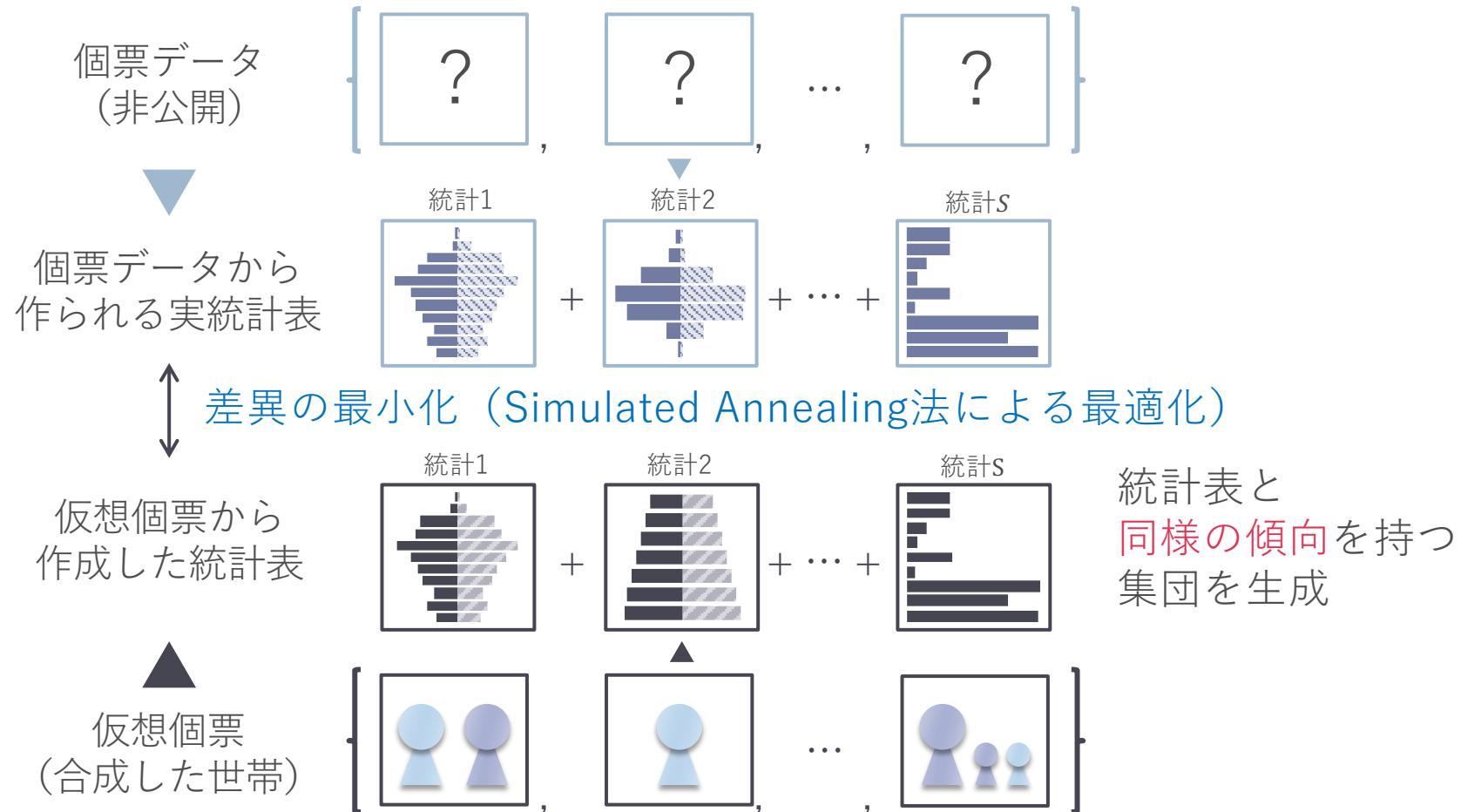
- 個人情報は、個人情報保護の観点から利用が制限されている。
 - 2018年5月に欧州でGDPRが施行。
 - 公表されている統計情報から、個人の情報を合成する。
- 組織に関する情報
 - 政府統計(中小企業実態基本調査、個人企業経済統計など)
 - 行政情報(事業承継補助金事業など)
 - 民間調査会社(帝国データバンク、東京商工リサーチなど)
 - 政府データベース(ハローワークなど)
 - 海外データベース(D&B Hoovers, OpenCorporatesなど)

環境データ:物理条件、制度条件

- 国土地理院基盤地図情報
- 憲法、法律(1,967)、政令(2,157)、府令・省令(3,759)



Challenge 2: 人的データの同定

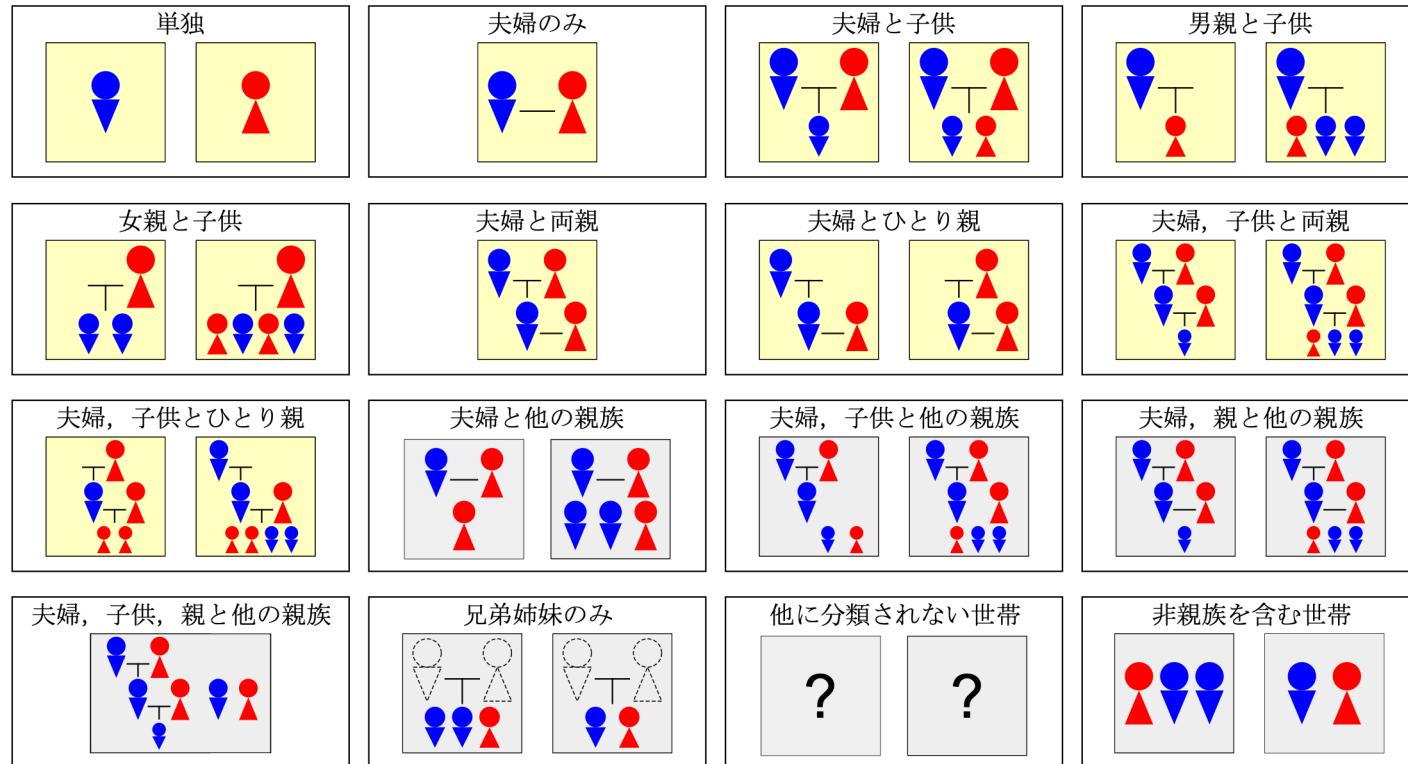


T. Murata, T. Harada, D. Masui: Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method Without Samples, SICE JCMSI, 10-6, pp.513-519 (2017).



Challenge 2: 人的データの同定

- 16分類中9種類を対象
 - 残りの7種類: 世帯人員間の関係を示した統計表なし



日本的人口の95%が9つの家族類型でカバーできる。

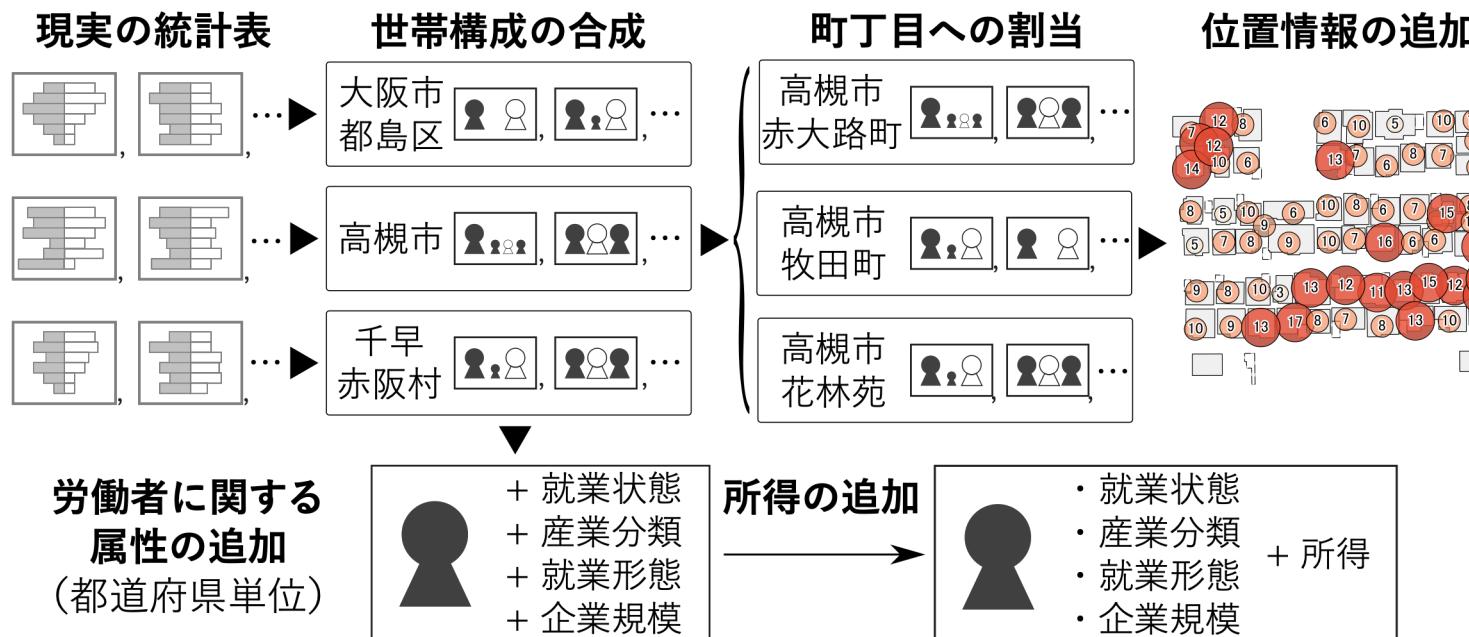


Challenge 2: 地図データへのマッピング

- 合成の流れ(阪大VCC利用)

- 世帯構成の合成, 町丁目への割当
所得の割当

地図情報は国土地理院基盤
地図情報を利用.



T. Harada, T. Murata: Projecting Households of Synthetic Population on Buildings Using Fundamental Geospatial Data, SICE JCMSI, 10-6, pp. 505-512 (2017).



Challenge 2: 合成手法の改善

従来手法

人口20万以上	A市	都道府県の(1)	都道府県の(2)	A市の(3)	A市の(4)	…	A市の(21)
	B市	都道府県の(1)	都道府県の(2)	B市の(3)	B市の(4)	…	B市の(21)
人口20万未満	C市	都道府県の(1)	都道府県の(2)	C市の(3)	C市の(4)	…	C市の(21)
	D市	都道府県の(1)	都道府県の(2)	市部の(3)	D市の(4)	…	D市の(21)
人口20万未満	E市	都道府県の(1)	都道府県の(2)	市部の(3)	E市の(4)	…	E市の(21)
	F町	都道府県の(1)	都道府県の(2)	郡部の(3)	F町の(4)	…	F町の(21)
人口20万未満	G町	都道府県の(1)	都道府県の(2)	郡部の(3)	G町の(4)	…	G町の(21)

同一の統計表を
市区町村に合わせて調整

推計した1歳階級の
統計表

提案手法

人口20万以上	A市	A市の(3)	A市の(4)	…	A市の(21)	
	B市	B市の(3)	B市の(4)	…	B市の(21)	
人口20万未満	C市	C市の(3)	C市の(4)	…	C市の(21)	
	D市	D市の(4)	…	D市の(21)	D市の(22)	D市の(23)
人口20万未満	E市	E市の(4)	…	E市の(21)	E市の(22)	E市の(23)
	F町	F町の(4)	…	F町の(21)	F町の(22)	F町の(23)
人口20万未満	G町	G町の(4)	…	G町の(21)	G町の(22)	G町の(23)

調査範囲
通りに適用

5歳階級の
統計表

1歳階級の
統計表

- 最適化に用いる統計表の詳細

- 父子・母子・夫婦年齢差 : (1)~(3)
- 家族類型別, 男女別, 人口分布 : (4)~(21)
- 男女別, 人口分布(全人口を対象) : (22), (23)

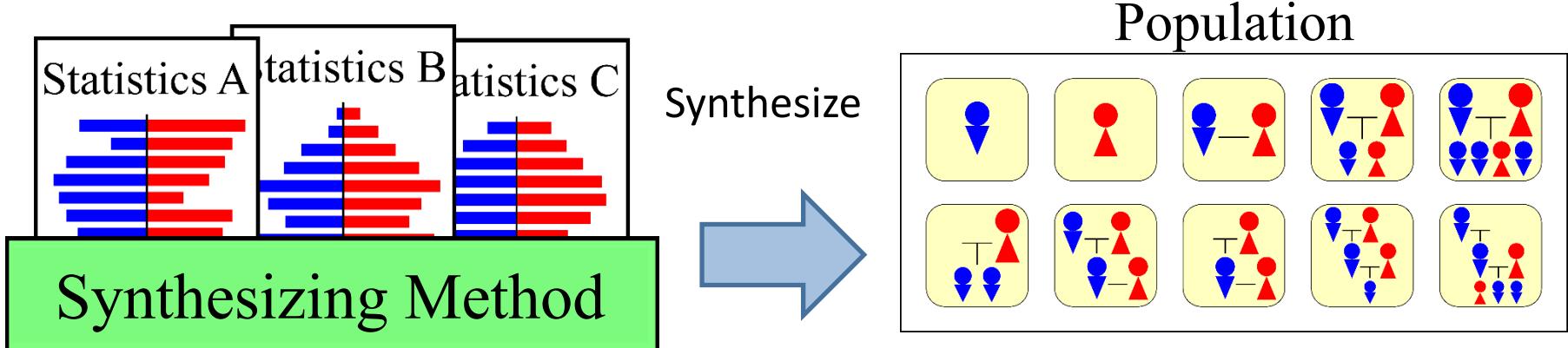


Challenge 2: 合成手法の公開 世界の合成人口の公開データ

組織	合成対象地域	使用統計
RTI (Research Triangle Institute) International	All states, USA Pop: 300 million	2010 US Decennial Census 2007-2011 American Community Survey
FRED (A Framework of Reconstructing Epidemiological Dynamics), Pittsburgh Supercomputing Center, Carnegie Mellon University	All states, USA Pop: 300 million Telangana, India Pop: 35 million	2010 US Census 2005-2009 American Community Survey 5-year sample
CDRC: Consumer Data Research Center	England & Wales, UK Pop: 53 + 3 million	2011 UK Census
関西大学	Japan Pop: 120 million	2000 National Census 2005 National Census 2010 National Census 2015 National Census



Challenge 2: 合成手法の課題



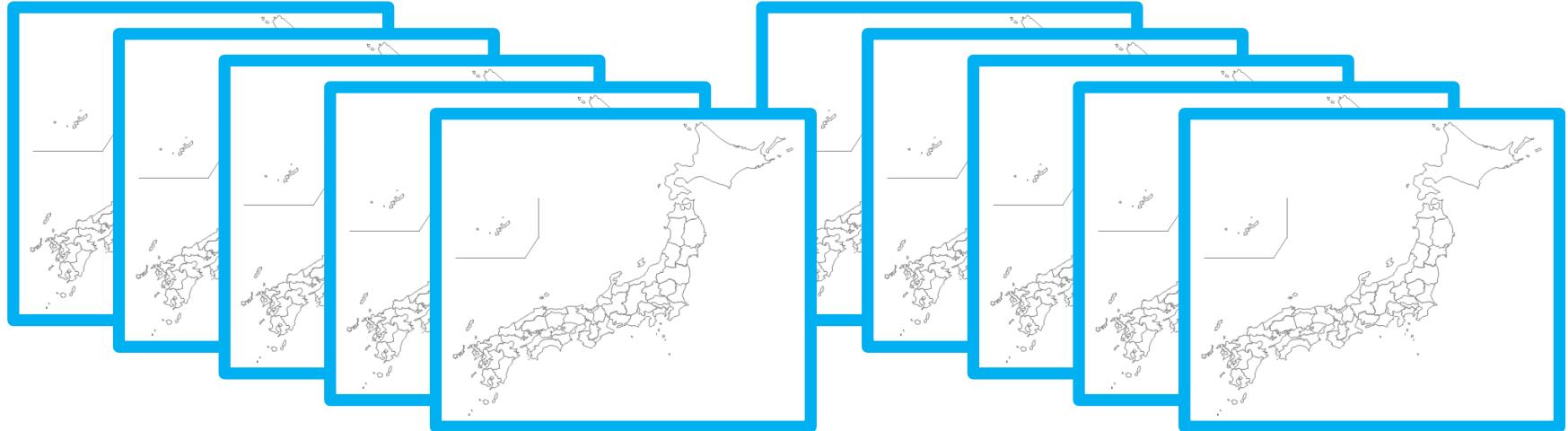
統計から人口を合成する。

FAQ: 合成された人口はどれほど実人口と一致しているのか？

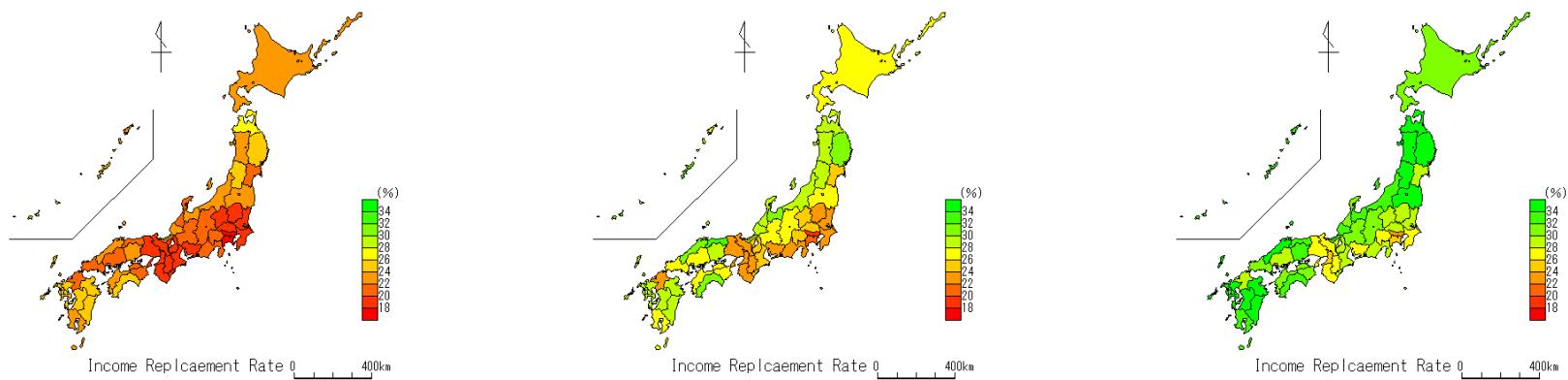
Answer: **一致していない**. 合成人口は**同じ統計的特徴**を持っているだけ. したがって, 複数の合成人口に対して, シミュレーションを行うべき.



統計からの合成人口



各合成人口に基づいて異なる結果を得る。

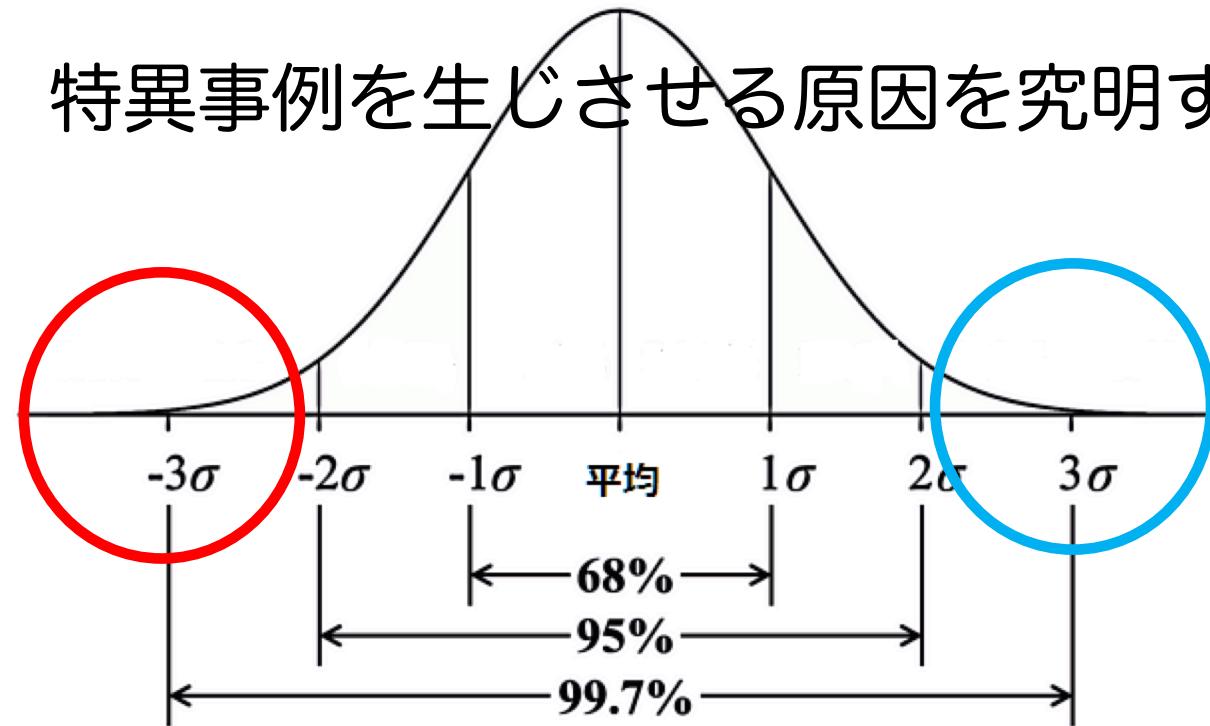


何が共通で、何が特異か？何がその原因か？



シナリオ分析における 共通事例と特異事例

特異事例を生じさせる原因を究明する



← →

悪事例 共通事例 良事例



Challenge 2: 内部要素の決定 における課題

- ・ 合成人口は、使用した統計と同じ統計的特徴をもつ
- ・ 複数の合成人口を対象にして、分析・シミュレーションを行うべき。
- ・ 合成データではあるが、分析・シミュレーションに使用する合成人口に、職業・所得などの個人情報が入るため、位置情報が加わった時に、センシティブな情報になる。



リアルスケール 社会シミュレーションの課題



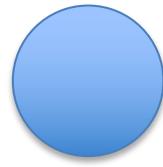
現実



将来



Challenge 1: モデル化



モデル

Challenge 4: シナリオ分析

A

B

C

Challenge 2: 内部要素の決定

Challenge 3: 計算速度と再現性

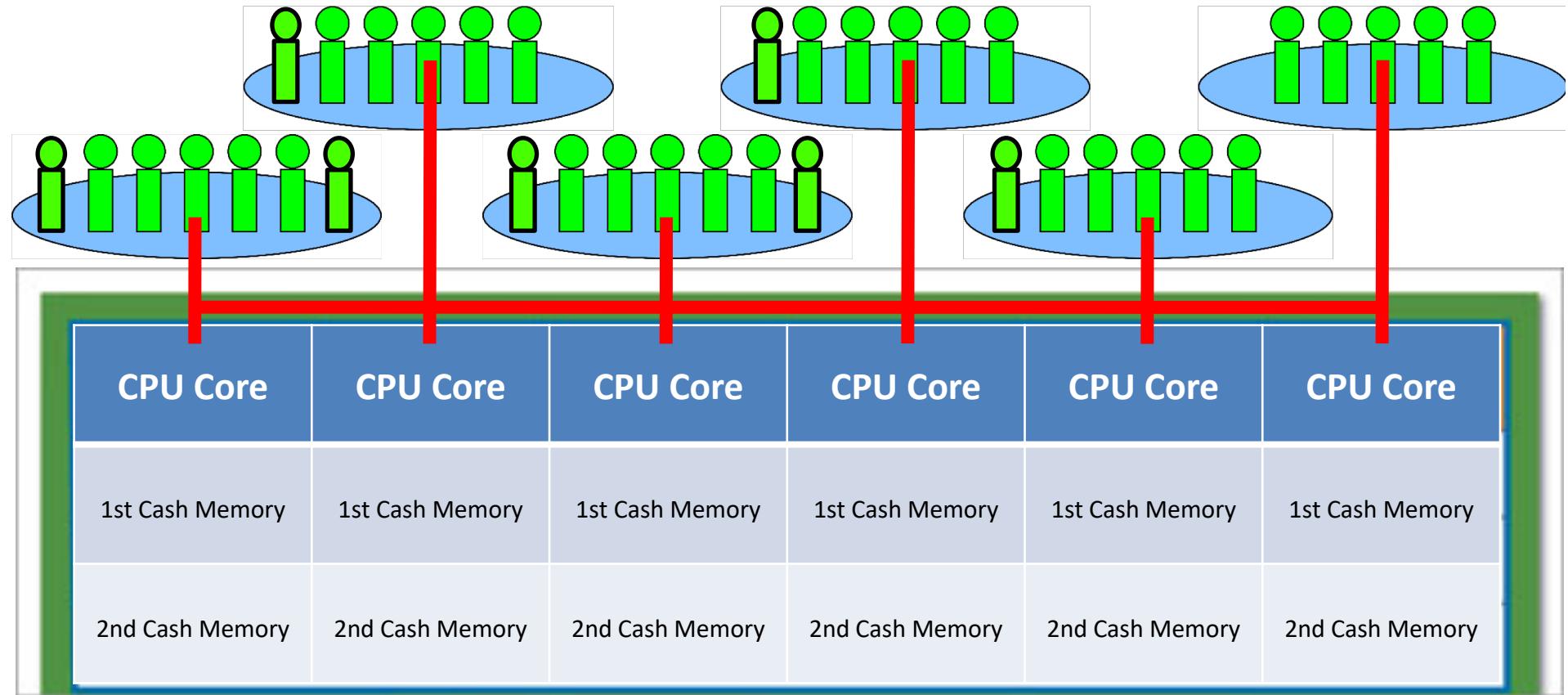


Challenge 3: 計算速度と再現性

- どうやって高速化するか？
- 並列計算と分散計算の技術を用いる。
- 乱数の使い方に注意すべき。
- 全く同じシミュレーションを再現することで、より詳細のログをとることができるようにする。



人口合成の高速化のための並列計算



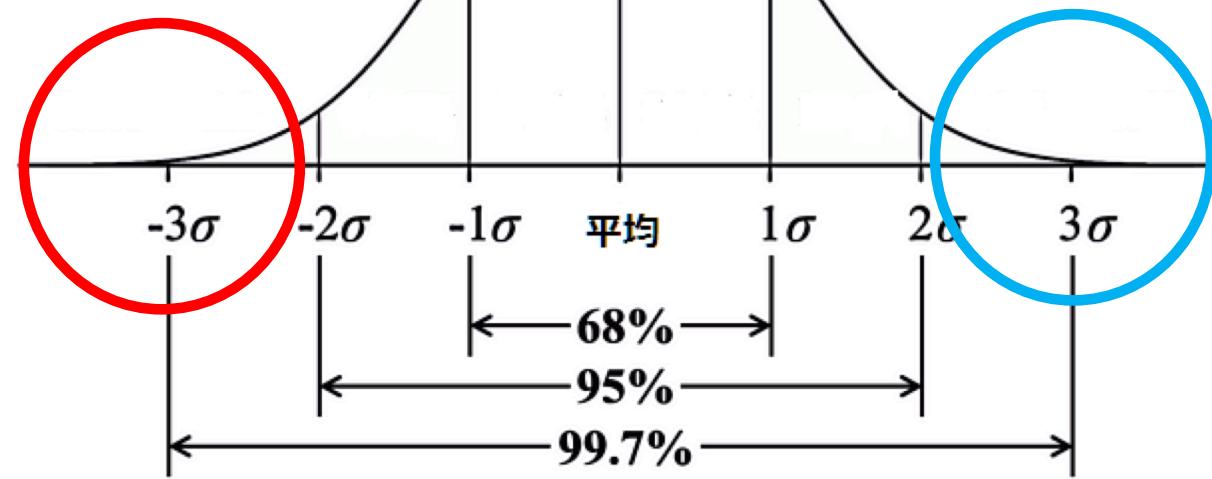
並列計算の際の統計の整合性

T. Harada, T. Murata, "Reconstructing prefecture-level large-scale household compositions using parallel computing," Trans. on Society for Instruments and Control Engineers, Vol. 54, No. 4, pp. 421-429, April, 2018 (in Japanese).



シナリオ分析における再現性の課題

特異事例を生じさせる原因を究明する



← →

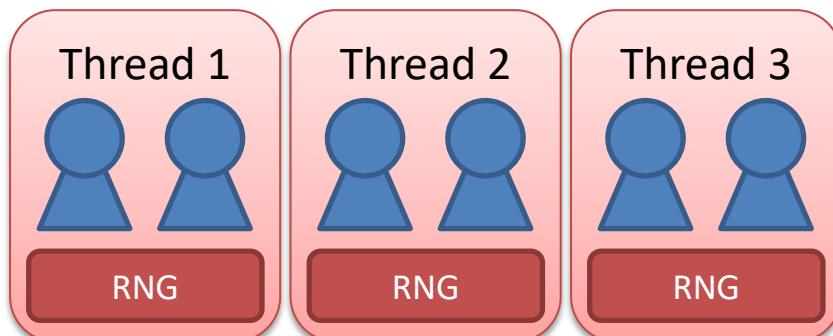
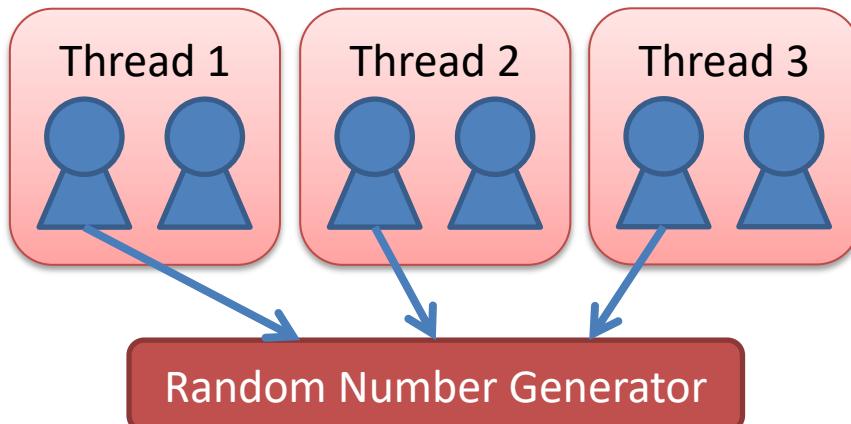
悪事例 共通事例 良事例



再現性における課題

エージェントのグループに乱数生成器を割り当てることで、異種計算機上の再現性を確保する。

再現性のないシミュレーション



再現性のあるシミュレーション

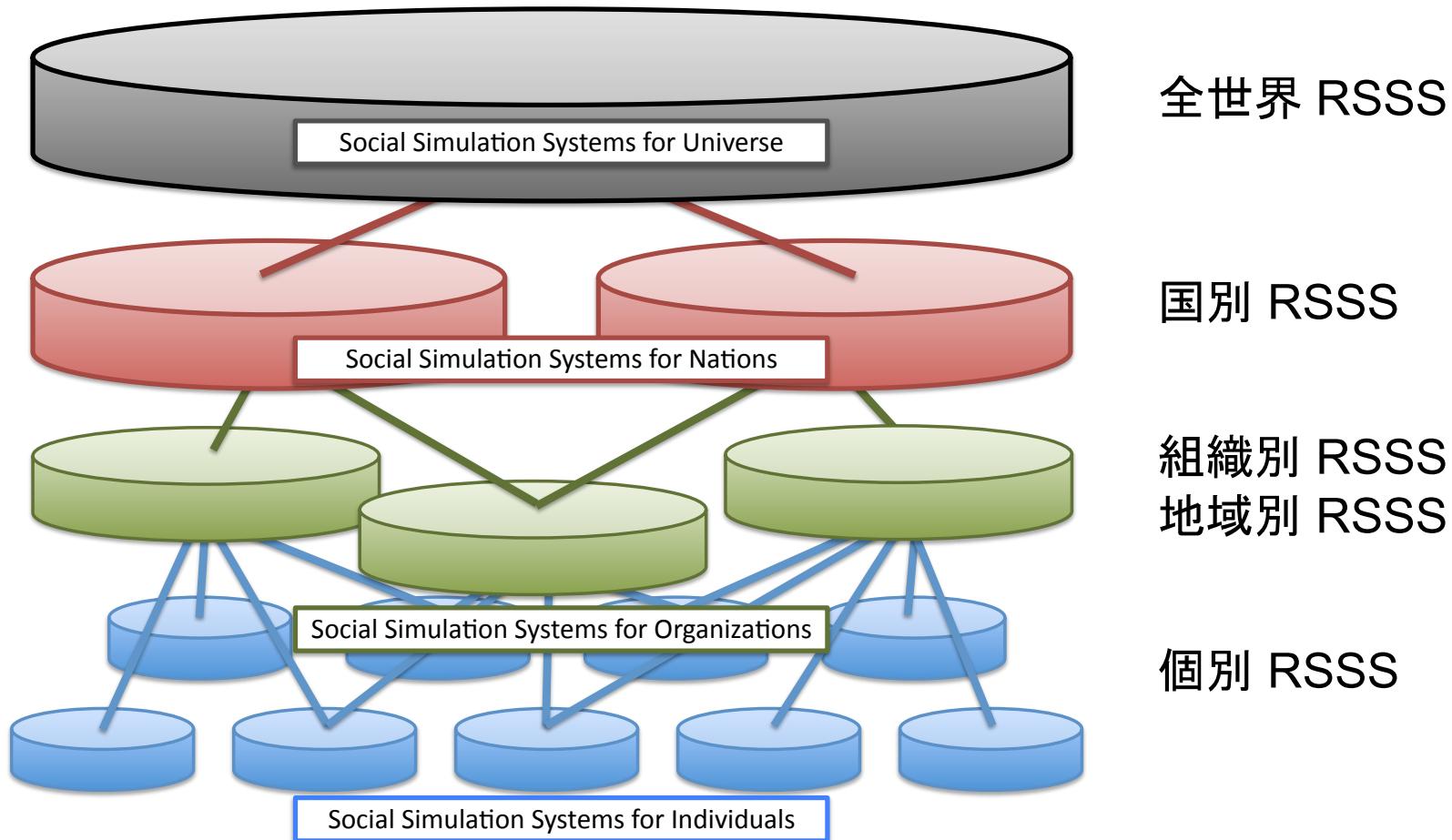


各エージェントグループがどのスレッドでも同じ乱数生成器を用いる。

原田, 村田, 社会シミュレーションのための異種並列計算環境における再現性の確保, システム制御情報学会論文誌 31(2), 37–48, 2018.



多層型リアルスケール 社会シミュレーション(RSSS)



Challenge 3: 高速化と再現性の課題

- ・ 合成人口の高速化では、統計の分割が課題
- ・ シミュレーションの高速化では、エージェントの移動が課題
- ・ 再現性は、エージェント群ごとに乱数生成器をもたせることで対応
- ・ 多層化したRSSSの接合の解決が必要



リアルスケール 社会シミュレーションの課題



現実

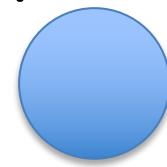


Challenge 1: モデル化



モデル

Challenge 2: 内部要素の決定



将来

Challenge 4: シナリオ分析

A

B

C



Challenge 3: 計算速度と再現性



Challenge 4: シナリオ分析

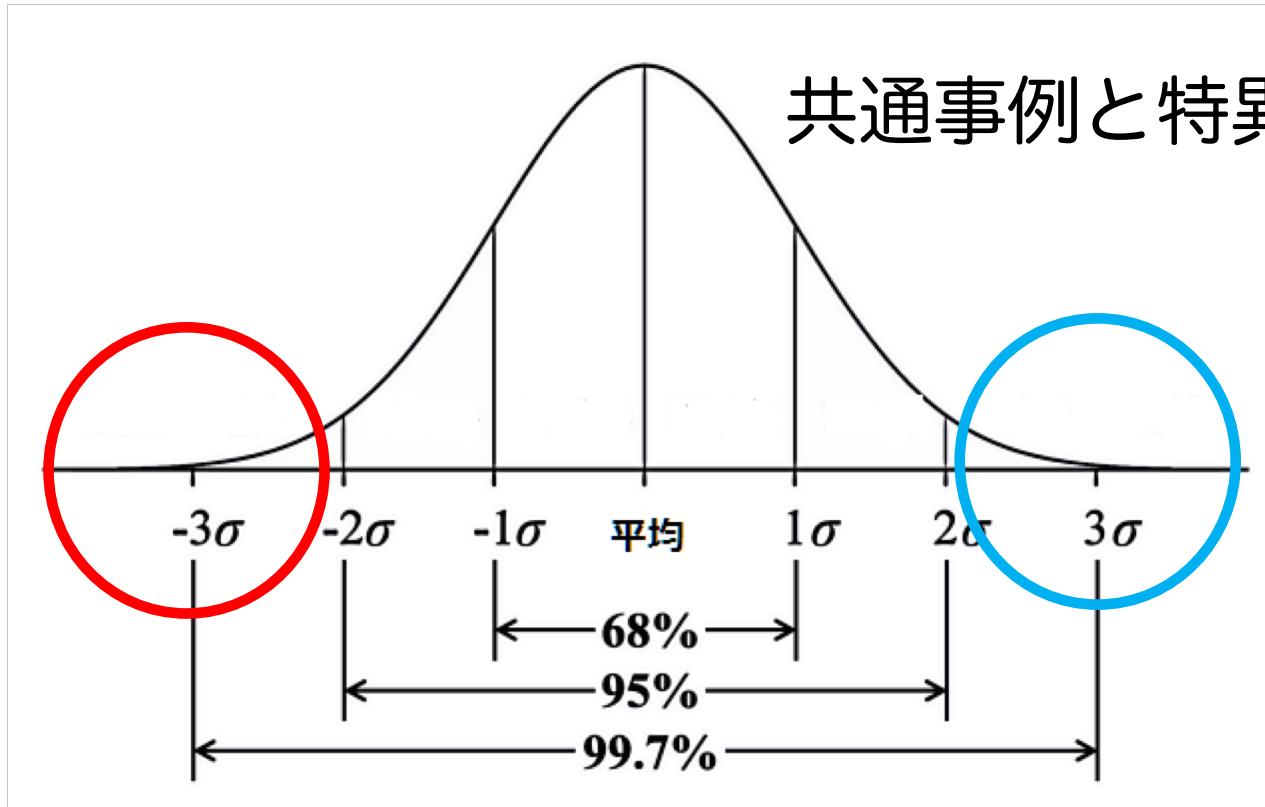
2016年の選挙の投票率の推定(10回平均)

2016年の実投票率	36.337%
2010 model による推定投票率	33.370%
2013 model による推定投票率	35.276%

2010年と2013年には、共通投票所を用いていなかったため、
共通投票所を用いたことで、投票率があがったと推定される。



Challenge 4: シナリオ分析 データ分析における結果の検出



← →

悪事例 共通事例 良事例



Challenge 4: シナリオ分析 データ分析する際のインターフェース

一橋大学経済研究所 社会科学統計情報研究センター
秘密分散・秘密計算技術を用いた
公的統計ミクロデータ分析のセキュリティ強化に関する研究



図出典：NTTセキュアプラットフォーム研究所



Challenge 4: シナリオ分析における課題

- ・ 共通事例・特異事例の検出
- ・ データ提供時のセキュリティの確保



リアルスケール 社会シミュレーションの課題



現実



将来



Challenge 1: モデル化



モデル

Challenge 2: 内部要素の決定

Challenge 3: 計算速度と再現性

Challenge 4: シナリオ分析

A

B

C



Challenge 1: モデル化における課題

- ・ どんな粒度でモデルを設計するか？
- ・ 数理的アプローチにおける変数の値をどうするか？
- ・ 地域ごとの特性をどのように反映させるか？
- ・ 各分野の社会科学者と社会シミュレーション研究者とのコラボレーションが必要

Challenge 2: 内部要素の決定における課題

- ・ 合成データではあるが、分析・シミュレーションに使用する合成人口に、職業・所得などの個人情報が入るため、位置情報が加わった時に、センシティブな情報になる。

Challenge 3: 高速化と再現性の課題 計算機

- ・ 合成人口の高速化では、統計の分割が課題
- ・ シミュレーションの高速化では、エージェントの移動が課題
- ・ 再現性は、エージェント群ごとに乱数生成器をもたせることで対応
- ・ 多層化したRSSSの接合の解決が必要

Challenge 4: シナリオ分析における課題 計算機

- ・ 共通事例・特異事例の検出
- ・ データ提供時のセキュリティの確保





Thank you!

