

平成22年 3月31日現在

研究種目：基盤研究 (B)
 研究期間：2007～2009
 課題番号：19300098
 研究課題名（和文） 秘匿された個票データのリスク評価法の確立と官庁統計データの公開への応用
 研究課題名（英文） Disclosure risk assessment for microdata sets and its application to the official statistics

研究代表者
 佐井 至道 (SAI SHIDO)
 岡山商科大学・経済学部・教授
 研究者番号：30186910

研究成果の概要（和文）：個票データのリスク評価で用いられる種々の超母集団モデルについての性質を明らかにした。2 時点以上で得られた標本から作られた複数の個票データや、層化抽出された標本から作られた個票データをリスク評価するために、寸法指標の多重寸法指標への拡張が提案され、ノンパラメトリック推定法についての理論構築がほぼ完成した。また、表形式データの秘匿措置の一つであるスワッピングに应用されるマルコフ基底を用いた手法についても進展があった。

研究成果の概要（英文）：The properties of several superpopulation models which are used for assessing the risk of a microdata set, were investigated. The size indices were extended to the multi-indices, to carry out the risk assessment for the microdata sets which had been made from the samples on two or more occasions, or a microdata set which had been made from a stratified sample. The nonparametric estimation method for the multi-population size indices was almost established. The swapping method is one of the disclosure control procedures for a tabular data. The theory of swapping method using Markov bases was developed.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	5,400,000	1,620,000	7,020,000
2008年度	4,800,000	1,440,000	6,240,000
2009年度	4,000,000	1,200,000	5,200,000
年度			
年度			
総計	14,200,000	4,260,000	18,460,000

研究分野：数理統計学，標本調査論

科研費の分科・細目：情報学・統計科学

キーワード：官庁統計，個票データ，リスク評価，寸法指標，多重寸法指標，秘匿措置，分割表，統計法

1. 研究開始当初の背景

(1) 欧米の個票データ公開の先進国では、

1990 年頃から個票データや表形式データに対する秘匿方法やリスク評価方法などについての研究が行われていた。最近も研究が継

続されているものの、研究の量は一時よりは少なくなっている。これは、それらの国々における官庁統計データの公開方法がある程度確立されたためである。例えば米国では、1990年と2000年のセンサスの調査結果から1%あるいは5%抽出した個票データに秘匿措置を施したものを、Web上において無料で公開している。

(2) 国内では1996年頃から個票データの秘匿方法とリスク評価方法の研究が始まった。本研究に参加している研究者の多くは、その当時から研究を継続している。1990年代中頃は、諸外国における研究成果の吸収が主な目的であったが、その後は国内独自の研究も行われ、特に、超母集団モデル（確率分割モデルとも呼ばれる）を用いた個票データのリスク評価の研究では、数多くの成果が得られていた。

(3) 内閣府に設置された統計制度改革検討委員会（委員長：吉川洋）と総務省の統計法制度に関する研究会（座長：廣松毅）が、それぞれ2006年に取りまとめた報告の中で、統計データの二次的利用が促進されるべきと結論づけていた。

2. 研究の目的

官庁統計の個票データとしての公開を実現するために必要となる理論の構築を図るために

- (1) 個票データの秘匿と公開のリスク評価
 - (2) 表形式データの秘匿と自動秘匿プログラム作成
 - (3) 秘匿データの有用性の評価
- の3つの研究を柱とする。

また本研究では理論の構築と完成にとどまらず、官庁統計の目的外使用申請を積極的に行い、個々の統計の個票データに対して、利用価値が高く、しかもリスクの小さい個票データセットを明示することも目的とする。

これらの研究を統合することによって、多くの官庁統計を始め、大規模調査によって得られたデータについて、どのような秘匿方法を行えば、安全かつ有用な個票データの公開が可能かという点が明示できると共に、仮に個票データとしての公開が困難と判断された場合にも、安全でしかも現行のものよりも有用な表形式データを示すことができる。

3. 研究の方法

本研究は「2. 研究の目的」に挙げた3つの研究を統合する形で構成される。それぞれの研究については、研究者の専門分野やそれまでの研究の過程などを考慮して、数名ずつ

からなる小グループを配置し、他の研究者がサポートする。

(1) 個票データの秘匿と公開のリスク評価については、佐井、竹村、星野、渋谷、大和、山口（小林）が中心に研究する。これらの研究については成果が蓄積されてきたものの、残されている課題も多く、それらの残された問題の解決を図りたい。標本寸法指標からの母集団寸法指標の推定では、新たなモデルの構築やモデルの相互関係の解明、精度の高い推定方法の確立などを旨とする。また、ノンパラメトリックな推定方法についても、実データへの適用が可能なレベルにまで引き上げる。更に、官庁統計で用いられるサンプリング方法に対応した推定方法についても検討を行う。なお、条件が整えば官庁統計の目的外使用を行い、提案する方法の有用性について検証を行う。

(2) 表形式データの秘匿と自動秘匿プログラム作成については瀧、稲葉、田村が中心に研究する。データの公開方法としては調査された内容を個票データのまま公開する方法と、数次元の分割表、すなわち表形式データに集計してから公開する方法があり、国内の官庁統計では後者の方式がとられている。秘匿方法は少数頻度のセルをxマークで置き換える方法が用いられ、しかも人間の目に頼る部分が多い。それらの改善を図るために、新たな秘匿方法について検討を行うとともに、秘匿の自動化も試みる。

(3) 秘匿データの有用性の評価については和合、大森、稲葉が中心に研究する。個票データの秘匿は秘匿後のデータのリスクと残されている情報量とを両方考慮しなければならないが、これまでの国内外の研究はリスク評価に大きく偏っていた。そこで個票データを分析する有用性を示すとともに、可能であれば個票データに段階的に秘匿を行いながら、各種統計量などの変化についても検討を行う。

以上のように、本研究では研究代表者及び研究分担者、連携研究者が3つの研究を分担して行うが、これらの研究は相互に深く関連している。そのため、本研究参加者全員の共通認識の構築と相互の意見交換のために研究会を開催する。また幅広い意見や知識の収集のために、本研究参加者以外の研究者、特に官庁統計に携わる研究者や実務者にも参加を呼びかけて研究集会を開催する。

4. 研究成果

- (1) 個票データの個人を識別できる変数の

値の組み合わせ一つ一つをセルと呼ぶが、各セルに含まれる個体数の頻度分布を寸法指標と呼ぶ。リスク評価方法としては、標本寸法指標 s_1, s_2, s_3, \dots から母集団寸法指標 S_1, S_2, S_3, \dots を推定することが一般的で、その推定には確率分割モデル（超母集団モデルとも呼ばれる）が用いられることが多い。本研究では、各モデルの性質やモデルの相互関係について研究が深められた。特に実用性の高いピットマンモデルについては、多くの性質が明らかになり、リスク評価に用いられている。

(2) 継続調査のように 2 時点以上で得られた標本から作られた複数の個票データを同時にリスク評価するために、また層化抽出された標本から作られた個票データをリスク評価するために、寸法指標の多重寸法指標への拡張が提案された。多重母集団寸法指標（例えば二次元の場合には $S_{(1,2)}$ のように表す）のノンパラメトリック推定法の理論構築がほぼ完成し、実データへの応用も行われた。ただし、母集団寸法指標の推定と比較してやや不安定であることも指摘され、更なる改善が必要である。また、同じ目的でピットマンモデルを 2 次元に拡張したモデルについても、実データへの適用の可能性が報告された。

(3) 表形式データの秘匿方法では、特にマルコフ基底を用いたスワッピングの理論について研究に大きな前進が見られた。自動秘匿プログラムとしてはオランダ統計局の -ARGUS が知られているが、日本の官庁統計への応用は難しいとの指摘がある。独自のシステムは、まだ完成には至っていない。

(4) 2002 年 1 月～2007 年 12 月の労働力調査の調査票情報の目的外使用申請を行い、段階的に秘匿措置を施しながら、個票データ、あるいは複数の個票データのリスク評価を行った。また種々のパラメータの秘匿による変化についても分析が行われた。

(5) 2007 年 5 月に統計法が全部改正され、2009 年 4 月から全面施行されている。改正の柱は「公的統計の体系的整備」、「統計データの利用促進と秘密の保護」、「統計委員会の設置」の 3 つであり、「統計データの利用促進と秘密の保護」には、「従来の目的外使用を継承した調査票情報の二次利用に関する規定」に加えて、「②調査票情報について適切な秘匿措置を施し、必要に応じてサブサンプリングを行った匿名データの提供に関する規定」、「③利用者の求めに応じて、調査票情報を基に集計結果を計算するオーダーメイド集計に関する規定」が含まれており、大きな前進と言える。本研究グループに参加し

ている研究者の 10 年以上に渡る研究がその一助になったと思われる。ただし、匿名データとして提供されているものが 4 統計、オーダーメイド集計が可能なものが 1 統計にとどまっており、今後も利用拡大のために研究を継続しなければならない。

(6) 本研究の成果については国内外の学会、シンポジウム、研究集会などで随時報告した。特に、統計関連学会連合大会においては、主なものだけで 34 件の報告を行っている。また、2007 年 11 月 16 日～17 日、2008 年 10 月 23 日～24 日、2009 年 10 月 22 日～23 日には、それぞれ統計数理研究所において研究集会「官庁統計データの公開における諸問題の研究」を開催した。毎年度、50 名前後の参加者があり、この分野に関係する研究者や官庁統計の実務者と意見交換を行った。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕(計 29 件)

佐井至道, 多重母集団寸法指標のノンパラメトリック最尤推定—2 時点の個票データへの適用—, 統計数理, 査読有, 57 巻, 2009, 425-442.

稲葉由之, 企業グループにおける電子商取引, 統計数理, 査読有, 57 巻, 2009, 357-369.

小林良行, ヨーロッパにおけるデータエディティング及び補定に関する調査報告—EDIMBUS プロジェクトを中心に, 総務省統計研修所・統計研究彙報, 査読無, 66 号, 2009, 101-129.

Hoshino, N., The quasi-multinomial distribution as a tool for disclosure risk assessment, Journal of Official Statistics, 査読有, Vol.25, 2009, 269-291.

竹村彰通, 計算代数統計の最近の話題について, 日本統計学会誌, 査読有, 39 巻, 2009, 137-159.

Sei, T., Aoki, S. and Takemura, A., Perturbation method for determining the group of invariance of hierarchical models, Advances in Applied Mathematics, 査読有, Vol.43, 2009, 375-389.

Hemmecke, R., Takemura, A. and Yoshida, R., Computing holes in semi-groups and its applications to transportation problems, Contributions to Discrete Mathematics, 査読有, Vol.4, 2009, 81-91.

Takahashi, M., Omori, Y. and Watanabe, T., Estimating stochastic volatility models

using daily returns and realized volatility simultaneously, Computational Statistics and Data Analysis, 査読有, Vol.53, 2009, 2404-2426.

Nakajima, J. and Omori, Y., Leverage, heavy-tails and correlated jumps in stochastic volatility models, Computational Statistics and Data Analysis, 査読有, Vol.53, 2009, 2335-2353.

Omori, Y. and Johnson, R. A., Efficient semiparametric Bayesian estimation of multivariate discrete proportional hazards model with random effects, Communications in Statistics-Theory and Methods, 査読有, Vol.38, 2009, 29-41.

佐井至道, 多重寸法指標のノンパラメトリック推定, 岡山商大論叢, 査読無, 43 巻, 2008, 1-18.

稲葉由之, SOHO事業者とSOHO事業希望者の実態, 日本テレワーク学会誌, 査読有, 6 巻, 2008, 17-31.

稲葉由之, 専門職員に対する統計教育の設計, 統計, 査読無, 59 巻, 2008, 13-18.

稲葉由之, 携帯電話普及期における電話所帯と通信料の変化—全国消費実態調査における耐久消費財票と家計簿の結合による分析—, 総務省統計研修所リサーチペーパー, 査読無, 11 号, 2008, 1-143.

稲葉由之, 高岡信行, 岡亜由子, 佐藤朋彦, 家計資産格差の推移と要因分解, 総務省統計研修所リサーチペーパー, 査読無, 14 号, 2008, 1-107.

大和元, 自然数の確率分割 (Gibbs分割から見た), 国際数理科学協会会報, 査読無, 59 巻, 2008, 1-9.

瀧敦弘, 野崎祐子, 高齢者就業の現状と問題点, 地域経済研究, 査読無, 19 号, 2008, 77-85.

Kakumu, K. and Wago, H., Small-sample Properties of Panel Spatial Autoregressive Models: Comparison of the Bayesian and Maximum Likelihood Methods, Spatial Economic Analysis, 査読有, Vol.3, 2008, 305-319.

Omori, Y. and Watanabe, T., Block sampler and posterior mode estimation for asymmetric stochastic volatility models, Computational Statistics and Data Analysis, 査読有, Vol.52, 2008, 2892-2910.

山口幸三, 政府統計の個票利用と統計法改正, 経済研究, 査読有, 59 巻, 2008, 139-152.

山口幸三, 政府統計マイクロデータ利用の現状と今後—試行的提供の利用者募集を終えて—, 統計, 査読無, 59 巻, 2008, 16-20.

稲葉由之, 有業者方式と労働力方式に基づく2種類の失業者の比較, 日本経済研究, 査読有, 56 号, 2007, 52-69.

稲葉由之, 失業者の求職開始時期, 統計, 査読無, 58 巻, 2007, 35-40.

Yamato, H., Toda, K. and Nomachi, T., Jackknifing a convex combination of one-sample U-statistics, Communications in Statistics - Theory and Methods, 査読有, Vol.36, 2007, 2223-2234.

Takemura, A. and Hara, H., Conditions for swappability of records in a microdata set when some marginals are fixed, Computational Statistics, 査読有, Vol.22, 2007, 173-185.

Sheena, Y. and Takemura, A., An asymptotic expansion of Wishart distribution when the population eigenvalues are infinitely dispersed, Statistical Methodology, 査読有, Vol.4, 2007, 158-184.

Omori, Y., et al., Stochastic volatility model with leverage: fast and efficient likelihood inference, Journal of Econometrics, 査読有, Vol.140, 2007, 425-449.

Omori, Y., Efficient Gibbs sampler for Bayesian analysis of a sample selection model, Statistics and Probability Letters, 査読有, Vol.77, 2007, 1300-1311.

大森裕浩, 多変量因子確率的ボラティリティ変動モデル, 経済研究, 査読有, 58 巻, 2007, 335-351.

[学会発表] (計43件)

佐井至道, 労働力調査の個票データのリスク評価法に関する種々の改善, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府)。

渋谷政昭, 複合分布モデルとベル多項式, 統計関連学会連合大会, 2009年9月8日, 同志社大学(京都府)。

渋谷政昭, Bell多項式と確率分布論, 応用統計学会, 2009年8月1日, 統計数理研究所(東京都)。

稲葉由之, 荒木万寿夫, 世帯員データを用いた世帯類型の作成, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府)。

大和元, Residual fractions of size-biased permutation of discrete prior associated with Gibbs partition, 統計関連学会連合大会, 2009年9月8日, 同志社大学(京都府)。

小林良行, 一橋大学を通じた匿名データの利用手続きについて—利用相談から利用実績の報告まで—, 統計関連学会連合大会, 2009年9月8日, 同志社大学(京都府)。

清智也, 青木敏, 竹村彰通, 多元分割表の階

層モデルに対する不変群の導出, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府).

青木敏, 竹村彰通, Affinely full-dimensional factorial design の性質, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府).

原尚幸, 竹村彰通, ロジスティック回帰モデルの正確検定とマルコフ基底, 統計関連学会連合大会, 2009年9月8日, 同志社大学(京都府).

星野伸明, 労働力調査個票開示リスクの時間安定性について, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府).

黒瀬雄大, 大森裕浩, 時変分位点のベイズ解析, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府).

國濱剛, 大森裕浩, Max-stable process のベイズ分析と高頻度株価収益率データへの応用, 統計関連学会連合大会, 2009年9月7日, 同志社大学(京都府).

広瀬要輔, 大森裕浩, Estimating Inefficiency in Online Auctions, 統計関連学会連合大会, 2009年9月8日, 同志社大学(京都府).

大西裕子, 大森裕浩, Bayesian Estimation of Entry Games with Multiple Player and Multiple Equilibria, 統計関連学会連合大会, 2009年9月8日, 同志社大学(京都府).

佐井至道, 個票データに基づく母集団多重寸法指標の推定, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

Sai, S., Nonparametric estimation for population size indexes, CASTA2008, 2008年12月10日, Kyodai Kaikan(京都府).

大和元, 渋谷政昭, Gibbs分割における新観測値までの個数の分布, 統計関連学会連合大会, 2008年9月9日, 慶應義塾大学(神奈川県).

Sibuya, M., An extension of Ewens-Pitman sampling formula to the partition of a multi-index, CASTA2008, 2008年12月10日, Kyodai Kaikan(京都府).

稲葉由之, 全国消費実態調査における家計簿と耐久消費財票の結合による分析, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

原尚幸, 竹村彰通, Markov bases for two-way subtable sum problems, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

青木敏, 竹村彰通, 3水準の一部実施計画に対するMCMC法による要因効果の検定, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

青木敏, 竹村彰通, 一部実施計画における要因の推定可能性と指示関数で表現される交

絡の関係, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

原尚幸, 竹村彰通, A localization approach to improve iterative proportional scaling in Gaussian graphical models, 統計関連学会連合大会, 2008年9月9日, 慶應義塾大学(神奈川県).

Takemura, A., Hierarchical models for contingency tables from the viewpoint of abstract simplicial complex, CASTA2008, 2008年12月10日, Kyodai Kaikan(京都府).

星野伸明, 複合ポアソンから生成される確率分割族について, 統計関連学会連合大会, 2008年9月9日, 慶應義塾大学(神奈川県).

Hoshino, N., On a class of random partitioning distributions, CASTA2008, 2008年12月10日, Kyodai Kaikan(京都府).

宮脇幸治, 大森裕浩, 日引聡, マイクロデータを用いた水道需要の分析, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

Wago, H., Kakamu, K. and Polasek, W., Spatial agglomeration and spill-over analysis for Japanese prefectures during 1991-2000, IFCS 2009, 2009年3月15日, University of Technology (Germany).

石黒真木夫, 中野純司, 田村義保, 統計的メタウェア試作, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

山口幸三, 労働力フローデータの作成と分析, 統計関連学会連合大会, 2008年9月8日, 慶應義塾大学(神奈川県).

山口幸三, 世帯動態統計作成の試み, 日本人口学会, 2008年6月8日, 日本女子大学(東京都).

佐井至道, 多重寸法指標のノンパラメトリック推定, 統計関連学会連合大会, 2007年9月8日, 神戸大学(兵庫県).

渋谷政昭, 佐井至道, 多重指標の確率分割, 応用統計学会年会, 2007年5月24日, 東京工業大学(東京都).

渋谷政昭, 大和元, 確率分割, ベル多項式, スターリング数, 統計関連学会連合大会, 2007年9月8日, 神戸大学(兵庫県).

稲葉由之, 就業異動の履歴に基づく区分に関する考察, 統計関連学会連合大会, 2007年9月8日, 神戸大学(兵庫県).

紙屋英彦, 竹村彰通, Hierarchical orbital decompositions and extended decomposable distributions, 統計関連学会連合大会, 2007年9月7日, 神戸大学(兵庫県).

野村俊一, 竹村彰通, Generalized zig-zag products of regular digraphs and bounds on their spectral expansions, 統計関連学会連合大会, 2007年9月7日, 神戸大学(兵庫県).

原尚幸, 青木敏, 竹村彰通, 標本数2のファ

イバーの構造と分解可能モデルのマルコフ基底，統計関連学会連合大会，2007年9月8日，神戸大学（兵庫県）。

青木敏，竹村彰通，一部実施要因計画に対するMCMC法による要因効果の検定，統計関連学会連合大会，2007年9月8日，神戸大学（兵庫県）。

星野伸明，疑似多項分布による個票開示リスク評価（続），統計関連学会連合大会，2007年9月8日，神戸大学（兵庫県）。

高橋慎，大森裕浩，渡部敏明，Estimating stochastic volatility models using daily returns and realized volatility simultaneously，統計関連学会連合大会，2007年9月7日，神戸大学（兵庫県）。

各務和彦，和合肇，Small sample properties of panel spatial autoregressive models，統計関連学会連合大会，2007年9月8日，神戸大学（兵庫県）。

仲西章，山口幸三，政府統計マイクロデータの試行的提供の現状と今後の課題，統計関連学会連合大会，2007年9月8日，神戸大学（兵庫県）。

〔図書〕（計2件）

佐井至道，統計データ科学事典（32章の2項目），朝倉書店，2007，710-713。

安田聖，山口幸三，横内宏至，統計資料シリーズ No.64 政府統計マイクロデータの試行的提供，一橋大学経済研究所附属社会科学統計情報研究センター，2009，317。

6. 研究組織

(1)研究代表者

佐井 至道 (SAI SHIDO)

岡山商科大学・経済学部・教授

研究者番号：30186910

(2)研究分担者

渋谷 政昭 (SIBUYA MASAOKI)

慶應義塾大学・理工学部名誉教授

研究者番号：20146723

瀧 敦弘 (TAKI ATSUHIRO)

広島大学・大学院社会科学部研究科・教授

研究者番号：40216809

竹村 彰通 (TAKEMURA AKIMICHI)

東京大学・情報理工学系研究科・教授

研究者番号：10171670

(H19→H20：連携研究者)

和合 肇 (WAGO HAJIME)

京都産業大学・経済学部・教授

研究者番号：00091934

(H19→H20：連携研究者)

稲葉 由之 (INABA YOSHIYUKI)

慶應義塾大学・経済学部・教授

研究者番号：80312437

(H19→H20：連携研究者)

大森 裕浩 (OMORI YASUHIRO)

東京大学・大学院経済学研究科・教授

研究者番号：60251188

(H19→H20：連携研究者)

大和 元 (YAMATO HAJIME)

鹿児島大学・理学部・名誉教授

研究者番号：90041227

(H19→H20：連携研究者)

田村 義保 (TAMURA YOSHIYASU)

統計数理研究所・データ科学研究科・教授

研究者番号：60150033

(H19→H20：連携研究者)

星野 伸明 (HOSHINO NOBUAKI)

金沢大学・経済学経営学系・准教授

研究者番号：00313627

(H19→H20：連携研究者)

山口 幸三 (YAMAGUCHI KOZO)

一橋大学・経済研究所・准教授

研究者番号：10436751

(H19→H20：連携研究者，H20まで)

(3)連携研究者

小林 良行 (KOBAYASHI YOSHIYUKI)

一橋大学・経済研究所・准教授

研究者番号：80553643

(H21のみ)