

機関番号：13301

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21730173

研究課題名（和文） 条件付き複合ポアソン分布族による個票開示リスク評価

研究課題名（英文） Microdata disclosure risk assessment using the family of conditional compound Poisson distributions

研究代表者

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

研究成果の概要（和文）：

本研究は実証と理論の両面で進めた。まず実証面では労働力調査個票データを匿名化し、匿名化手法のリスクや有用性を実証研究するための基盤を構築した。また実証の妥当性を確保し、日本の匿名化実務について指針を得るため、外国の先進的事例を調査してまとめた。理論面では、リスク評価の対象として頻出する疎な分割表の挙動を説明するため、極限条件付き複合ポアソン分布族の性質を評価した。特に、ベル多項式に依存した新しい漸近論を提案した。

研究成果の概要（英文）：

This project has progressed both on an empirical basis and on a theory. One empirical contribution is an anonymized data set of the Japanese Labor Force Survey, which will be used to test the risk and utility of anonymization methods. Another contribution is a survey on the advanced practices and researches of anonymization. This survey has been especially conducted to provide a sound view on the current Japanese anonymization practices and researches. On a theory, the limiting conditional compound Poisson distribution has been investigated to explain the behavior of a sparse contingency table, which often arises in assessing disclosure risk. In particular, a new asymptotics based on a Bell polynomial has been proposed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	2,600,000	780,000	3,380,000
2010年度	800,000	240,000	1,040,000
年度			
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：統計学

科研費の分科・細目：経済学・経済統計学

キーワード：個票開示リスク、母集団一意、プライバシー、寸法指標

1. 研究開始当初の背景

日本でも匿名化処理を施した個票データの研究用提供を目前に控え、匿名化の正確な理解は喫緊の課題であった。特に、調査客体のセンシティブな情報が暴露される危険性（個票開示リスク）の評価は重要である。

研究代表者は分割表に多変数離散分布を入れる事でリスク評価の改善を続けており、匿名化の実務的要請と整合的な分布族（条件付き複合ポアソン分布）を新しく着想するに至った。特に、疎な分割表のモデルとして極限分布の族（極限条件付き複合ポアソン分布）を構成可能で、この族は理論的に個票開示リスクの上限を与える。

2. 研究の目的

本研究課題では、条件付き複合ポアソン分布族と極限条件付き複合ポアソン分布族の解析を、理論及び実証の両面で発展させる。理論的には、族のメンバーの性質を族の性質として一般化する。実証的には、労働力調査の個票データを用い、個票開示リスクの理論的な上限を検証する。

3. 研究の方法

実証研究のためのデータセットを、まず作成する。労働力調査の基礎調査個票を、平成14年1月から平成19年12月まで目的外使用する。リスク評価に用いるキー変数を選択し、程度を替えて再符号化する。そして各段階で結果を保存しておく。

理論研究については、過去に個別の分布で示した性質をベル多項式を用いて表現し、こ

れを族の性質とする。またこのように考えると、極限条件付き複合ポアソン分布族と既存の確率分割族との関係が明確になる。故に既存の族との関係をサーベイして整理する。

4. 研究成果

実証研究用データセットは、予定通り完成させた。このデータセットについてはピットマンモデルによるリスク評価を行い、最尤推定量の経験分布を得た。リスクの上限概念について明確な証拠は得られなかったが、個票開示リスクの時系列構造評価という新しい問題系に取り組む事が出来た。

母集団のリスク構造を推定するとして、一時点の標本のみ使用する場合よりも、異時点の標本を利用すればより正確な推定が可能なのである。特に労働力調査の場合、ローテーションサンプリングに起因するリスクの自己相関が存在すると予想される。

また実務との関係でも、個票開示リスクの時系列変動は無視できない。例えば、より新しい時期の匿名データを作成する場合の審査は、リスクの時系列変動が大きいなら、匿名化の程度を変えなければならないかもしれない。これは変数のカテゴリー化を伴うということであり、データの使い勝手を悪化させる。つまり匿名データの審査では、時系列変動も考慮に入れた上で意思決定されるべきである。

ところが既存研究では、単純化のため時系列構造を無視し、各時点独立と仮定してリスクが評価される。一方、研究代表者の多変数離散分布アプローチでは、少数の母数にリスク構造を縮約する。故に時系列要因をモデルに入れても操作可能性を失わない。本研究課

題ではこのように考えて、重要な問題に見通しをつける事が出来た。

リスクの時系列構造評価の精緻化は今後の課題としたが、労働力調査データセットの探索的解析の結果、リスクの一次の自己相関やトレンドは確認された。また最尤推定量の理論分布と経験分布を比較し、データ期間内の最低リスクと最高リスクは母集団のリスク構造が違うかもしれないという結果を得た。

ここまで述べた内容はデータセットのリスクに関わる評価だが、有用性評価も現実問題としては重要である。このような要請に応えるため、本研究課題では有用性評価の方向性も検討した。

匿名化の先進的研究、事例は残念ながら海外に存在する。従ってそれらの文献をサーベイし、日本の実務に役立てるという視点で整理してまとめた。結果として、データセットに「嘘」の要素をランダムに混在させるという手法が、制度の改善に不可欠ということになる。

理論的な結果について述べる。極限条件付き複合ポアソン分布族はそもそも条件付き複合ポアソン分布族の極限分布として得られた族だが、より広い分布族の極限として得られる事が明らかになった。極限条件付き複合ポアソン分布族は小数法則によって特徴付けされるので、確率分割族としてはそれが本質ということになる。そしてベル多項式による漸近的性質表現は予想通りに進展した。結果として、個票開示リスク評価に限らず、より一般性を持つ疎な分割表解析の理論枠が整備された。

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 2 件)

- (1) N. Hoshino (2011) Random partitioning over a sparse contingency table, *Annals of the Institute of Statistical Mathematics*, To Appear, 査読有.
- (2) 星野伸明 (2010)「公的統計マイクロデータ提供制度の課題」, *日本統計学会誌*, Vol. 40, 23-45, 査読有.

[学会発表] (計 4 件)

- (1) 星野伸明, 模造個票データの必要性について, 統計関連学会連合大会, 2010/9/6, 早稲田大学早稲田キャンパス(東京都).
- (2) 星野伸明, 個人情報保護の数理的方法論, 第三回横幹連合コンファレンス, 2009/12/3, 東北大学片平さくらホール(宮城県).
- (3) 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, 2009/9/7, 同志社大学京田辺キャンパス(京都府).
- (4) N. Hoshino, On the time series variation of microdata disclosure risk, The 57th session of the International Statistical Institute, 2009/8/22, Durban (South Africa).

[その他]

ホームページアドレス :

<http://stat.w3.kanazawa-u.ac.jp/owner/papers.html>

6. 研究組織

(1) 研究代表者

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

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし