

「特集 個票開示問題の統計理論」について

竹村 彰通[†] (オーガナイザー)

計算機やネットワークの急速な発展にともない、データの収集提供のあり方が急激に変化しつつある。膨大なゲノム情報がインターネットを通じて入手可能になりつつある現状などが一つの典型的な例である。政府統計情報についてもホームページ等での提供が進んでいるものの、我が国における統計情報の収集・集計・公表は統計情報の連続性や正確性の確保を目的とした伝統的な手続きに基づいておこなわれており、そのため環境の変化への適応が遅れている面がある。

特に我が国では集計表以前の個票レベルの統計データが提供されていない点が問題である。これに対して諸外国の統計当局は、すでに以前よりさまざまな形で個票データを提供しており、統計の利用者の便宜をはかり統計の存在意義をアピールしている。進んだ事例としては、統計当局と特定の利用者の間に暗号化した通信路を確保することによって、概念的にはオフラインに類する形でのデータの提供なども一部実用化されている。

個票データを提供する際に最も重要な問題は、データに含まれる個体が識別されないという安全性の確保である。統計調査の回答者は個人情報公開されないという約束のもとに回答しているから、間接的にせよ個体が識別されることは深刻な問題となる。個体の識別を防ぎつつ、有用な個票データをどのように提供するか、という問題が個票開示問題である。

本特集では、個票開示問題に関する統計数理的な研究を扱っている。筆者を含めて本特集への寄稿者は、ここ数年この問題に関心を持って研究を進めて来た。そのような中で、国際的なレベルでの理論的な貢献もいくつかなされており、本特集に示されているようにかなりの研究成果が得られてきている。一方で我が国でこの問題に関心を持つ研究者は必ずしも多いとは言えないし、また一定の貢献を生み出すためには対象をある程度しぼって研究をおこなう必要があることから、本特集で扱われている個票開示問題に関する研究成果は必ずしも問題の全貌をバランスよく反映しているわけではない。特に、数理統計の研究者による統計数理的な研究以外の側面として、統計実務家による個票データ提供の実務面の研究やオンライン提供にともなうアルゴリズムに関する研究が遅れている。諸外国では、政府統計当局にも研究者がおり、実務的な観点からの個票開示問題の研究もさかんである。本特集を契機として、統計数理的な研究に加えて、実務面やアルゴリズムの側面にも関心を持っていただける研究者が研究に参入されることを期待したい。

なお、有用なデータの提供と個人情報の保護の問題は、政府統計情報に限らず民間の研究機関によって収集される情報、さらには例えば遺伝情報の収集と提供等においても、今後の重要な課題になると予想される。個票開示問題をそのような広い観点から理解していただき関心を持って本特集を読んでいただければさいわいである。また個票開示問題にあらわれる統計モデルは、集団遺伝学、統計的生態学、計量言語学などにも共通にあらわれるものが多く、そのような観点からも興味を持っていただけるものと思う。

[†] 東京大学大学院 情報理工学系研究科：〒133-0033 東京都文京区本郷 7-3-1