

▼ CONTENTS

- 02 響き合う人とデーター 統数研プロジェクト紹介
第16回「公的調査データの統計開示抑制」
- 06 シンポジウム報告
「Workshop on Statistical Analysis for Forest Resource Management Level 4
—Generalized Linear Model—」 共同開催報告
「Training Workshop Series V —Advanced Statistical Analysis in “R” for Forestry Related Data」
共同開催報告
- 07 研究教育活動
2018年10月-12月の公開講座実施状況／統計数理セミナー実施報告（2018年11月～2019年1月）
平成30年度若手研究者クロストークの開催報告
2018年公開講演会「創薬のフロンティア：データサイエンスの挑戦」
- 09 統数研トピックス
国際外部評価委員会の開催について
「立川市アカデミックアドバイザー」任命式と連絡協議会幹事会の開催
数学・数理科学4研究拠点合同市民講演会「AI社会の基盤は数学!」を開催
統計数理研究所子ども見学デー2018を開催
「大学共同利用機関シンポジウム2018 ～最先端研究大集合～」に参加・出展
実践女子大学より生活心理学専攻の1・2年生が来所
- 11 総合研究大学院大学複合科学研究科統計科学専攻関係
大学院説明会について
- 11 お知らせ
公開講座
- 12 共同利用
H30年度共同利用公募追加採択課題
- 12 外部資金・研究員等の受入れ
受託研究・受託事業等の受入れ／外来研究員の受入れ／寄附金の受入れ
- 14 人事
- 14 刊行物
統計数理／Annals of the Institute of Statistical Mathematics
- 16 コラム

統計数理研究所二ユース

大学共同利用機関法人 情報・システム研究機構



調査票情報を安全に提供 持ち出し審査基準を開発



▲南 和宏モデリング研究系准教授

提供対象の拡大で 調査票情報の活用が進む

国勢調査や国民経済計算（GDP統計）、経済センサス（企業経済活動の実態調査）など、政府は長年にわたりさまざまな公的調査を実施してきた。その位置づけが「行政のための統計」から「社会の情報基盤としての統計」へ大きく転換したのは、2007年の統計法全面改正をきっかけとする。

これにより、国民が合理的な意思決定を行うための共有財産として、公的統計を活用する道が拓かれた。学術研究など公益目的に限り、調査票情報の二次利用が認められるようになったのである。

ただし、利用者である研究者には、データを操作する部屋の入退室管理

や外部ネットワークからのシャットアウトなど、厳しい要件が課せられる。このため、研究室のセキュリティ環境によっては利用申請できないケースが少なかった。また、提供方法も、必要最低限のデータをDVDの形で渡されることから、使い勝手がよくなかった。

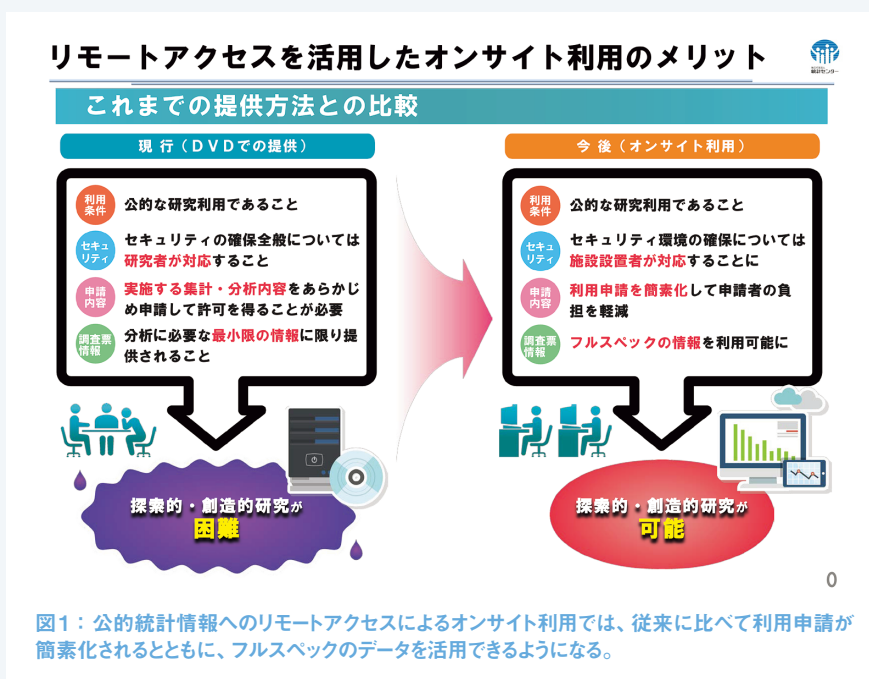
そこで、2010年に情報・システム研究機構とデータの提供元である統計センターが二次利用の普及・啓発に関する連携協定を結び、2012年に東京・立川の統計数理研究所内に「オンサイト施設」を開設。研究者は事前申請のうえで来訪し、調査票情報に直

接アクセスできるようになった。

さらに今年度から、全国9カ所の大学などに設置されたオンサイト施設で、リモートアクセスの試行運用も始まった。今後は、施設を訪れた研究者が、シンクライアントからサーバにアクセスし、フルスペックのデータを用いて分析できるようになる。このことが、探索的・創造的な研究に資すると期待されている（図1、2）。

オンサイト施設利用での 分析結果安全基準を作成

リモートアクセスを活用したオンサイト



する

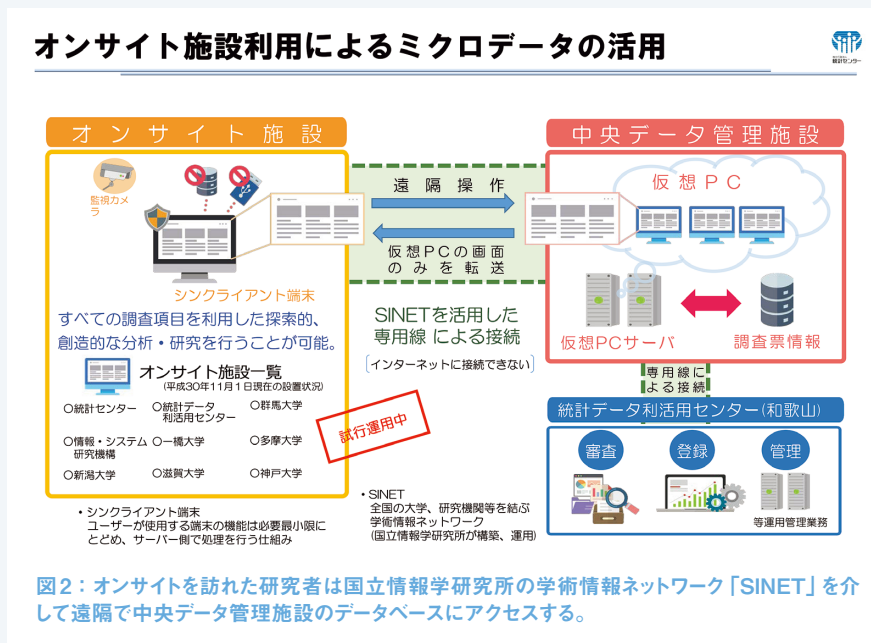
政策立案のために官庁が実施してきた公的調査のデータを学術研究にも活用しようとの動きが顕著になってきた。その大前提となるのが、調査協力者のプライバシー保護だ。統計数理研究所は、データの提供元である統計センターと連携し、研究者が全国の大学に設置されたオンサイト施設で調査票情報にアクセスして分析した結果を持ち出す際の安全性審査基準を策定。審査における経験則ルールを決定するとともに、研究者を支援するツールの開発を手掛けている。

利用の体制を整えると同時に、より強化しなければならないのが、調査対象者の機密情報保護だ。例えば、高額商品の購買情報、家計の収入や土地などの資産額、雇用契約期間や仕事内容など、プライバシーに関わる情報が、公開された研究成果から漏洩することのないよう、慎重な取り扱いが求められる。

そこで、オンサイト施設で分析した成果物を外部に持ち出す場合には、その安全性・秘匿性のチェックを受ける仕組みを構築することとなった（図3）。統計センターの非常勤職員となり、その安全基準を作成したのが、統計数理研究所の南和宏准教授と、匿名化技術の研究で実績のあるNTTの菊池亮博士だ。

「今回の安全基準を策定するにあたり、まず二つの前提条件を決めました」と南は説明する。一つは、研究者は悪意の情報漏えいを行わない、と仮定すること。つまり、機密情報を含む分析結果が不注意によって論文に掲載され、第三者によって漏えいするケースを想定する。

もう一つは、研究者がデータの持出しを複数回行う場合、中間結果にも最終結果と同じ基準を適用すること。中間結果に対しては、論文に掲載しない前提で基準を緩和することも検討したもの、うっかり掲載してしまうミスを防



止できないと判断し、この結論に達したという。

基準作成にあたっては、欧米諸国の統計開示抑制に関する「Eurostat 基準（SDCハンドブック）」を参考にした。ここでは「安全性基準の作成は『どのような動物を入れるかわからない動物園の檻』を設計するようなもの」と汎用的な基準作成の難しさを説き、審査対象を標準的なデータ形式に限定している。南は「今回の基準でも同様に、対象を度数表や数量表、線形回帰係数および非線形回帰係数、パーセンタイル値などに絞りました」と話す。

同ハンドブックでは、実際の審査は「経験則」による一次審査、「原則ルール」による二次審査の2段階で行う。経験則は、明示的に記述されたルールで、非常に厳格なものだ。これに対し原則ルールは、データの意味や特徴を考慮し、個別の状況に柔軟に対応するもの。日本でもこの2段階審査の枠組みを採用することにした。

菊池博士は「安全性と利便性はトレードオフの関係にあり、完璧な基準をつくることは不可能。運用面でいかにバランスを取るかがポイントです」と話す。



▲菊池 亮博士 (工学) NTTセキュアプラットフォーム研究所

経験則での原則は、①個票データは機密情報、②客体10の原則、③占有性の原則、④グループ開示の原則の四つ。①の原則から、個票データの最小値、最大値の持ち出しは不可となる。②は、個票値の推定ができないように客体数を10以上確保するというもの。③は同様の理由で、いずれかの個票値の割合が全体の50%以上を占めてはいけないというものだ。また④の原則から、ある属性でグループ化された個票の90%以上が、別の属性に関して同

一の区分に属することを禁ずる。

原則ルールでは、研究者が持ち出しを希望する分析結果を「複数の個票値を入力とする関数の出力」と捉え、関数の出力を使って個票値をどの程度特定できるのか、データの特長も加味して判定する。場合によっては経験則ではNGであっても安全であるような場合があり、そのような場合にも柔軟に対応するのが、原則ルールの目的だ。

申請者の負担を軽減するRツールを新たに開発

通常の分析結果は機密情報を含むことから、そのままでは安全性審査を通らない。このため、研究者は事前にデータの秘匿処理を行った分析結果を用意する。一方で、審査担当者は秘匿済みの分析結果を見ても、基準を満たしているか否かの判断ができない。

このため研究者には、審査用に説明資料を添付してもらうのだが、これが研究者にとって作業負担を増大させる要因となる。例えば、占有性ルールの審査で、全体の50%を占める個票値がないかどうかは、元の個票データから第一

位、第二位の客体の値を抽出する必要がある(図4、5)。

また、秘匿化をするにも、集計表のうち占有性ルールを侵害する数値を持つセルだけを隠したのでは、行計や列計から復元できる可能性がある。この場合、秘匿セル変数の取りうる範囲が十分広くなるように、追加でセルを隠さなければならない(図6)。こうした処理を手計算で行うのはほぼ不可能だ。

実際に、オランダの統計局は表データの秘匿処理ツール「^{タウ アルガス}T-Argus」を開発している。南らは、日本にもこれを導入できないかと考えた。だが、調べてみると、このツールはどう考えても、分析者と審査者が同一でなければ使えないものだった。南はオランダへ赴き、直接、担当者に話を聞いた。ところが、意外なことにオランダでは安全性審査にはツールを使わず、審査担当者の経験に基づく審査が行われていた。

使えるものがないのなら、つくるしかない——。そう考え、南と菊池博士は統計ソフト“R”を使い、秘匿処理と説明文書が同時に自動で作成できる新たなツールの開発に乗り出した。

分析結果はオフラインで提供

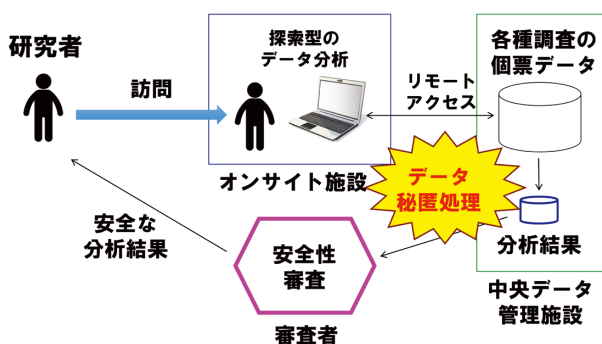


図3：オンサイト施設からのリモートアクセスによる分析結果の持ち出し手順のイメージ。研究者自身が、安全性審査を受けるまえにデータ秘匿処理を行う。

ただし、持出し審査には安全性を証明する追加の説明資料が必要

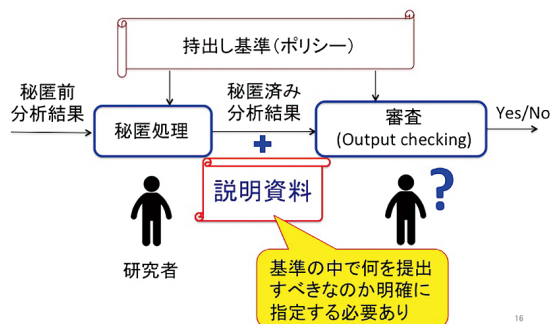


図4：データ秘匿処理を施した分析結果を審査するには、研究者自身が説明資料を揃える必要がある。



▲阿部 穂日統計センター上席統計専門職

分析者と審査者をつなぐという思いで「SDCLink」と名付けた新ツールはすでに施行運用段階に達している。「SDC」は Statistical Disclosure Control（統計開示抑制）の略。近日中にオンサイト施設に導入するとともに、完成度を高めてRのパッケージとして提供することも計画しているという。

運用状況の課題をフィードバックし 利用者の利便性を高める

安全性基準の原則ルールは、裁判における判例のように、運用しながら実

例を積み上げるなかででき上がっていくものだ。調査データ開示の事務を取り仕切る統計センターの阿部穂日上席統計専門職は、審査側の立場から次のように語る。「経験則ではどんなデータも基準を満たさなければ持ち出し不可とするものの、例えば消費実態調査の客体の性別や年齢のクロス集計などのように、公開しても実質的に問題ないものもあります」。

現状では、どの情報を秘匿すべきか、明確な線引きはなされていない。統計調査の所管官庁の判断は、協力者を確保するためにどうしても安全側へ傾く。しかし、そのために研究者の手間や審査時間が増大したのでは、利用促進に歯止めがかかる。阿部上席統計専門職は「統計家の研究成果を実務に反映し、審査時間の短縮を進めるなど、利用者の利便性を確保できるシステムをつくっていきたい」と話す。

このプロジェクトに携わった意義について、南は「個人のプライバシー保護は、ビッグデータ時代の最重要課題です。オンサイト利用における安全性審査プロセスの確立をビッグデータ利活

用の成功例にしたい」と抱負を語る。データサイエンスの最前線を支える取り組みを統数研の知見が牽引している。

（広報室）

占有性ルールの審査プロセス

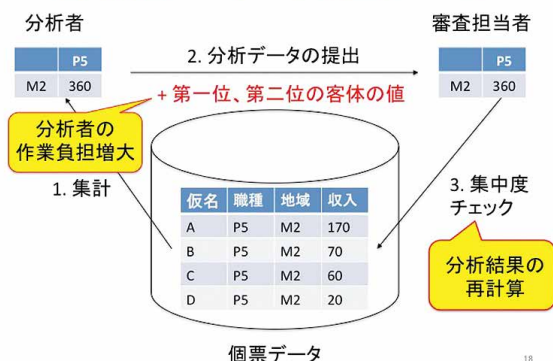


図5：占有性ルールの審査では、分析データに加えて、第一位、第二位の客体の値を抽出して提出しなければならない。

秘匿セルをもつ表データへの追加要件

- 秘匿セル変数の可能範囲（秘匿インターバル）の幅 w の長さがしきい値 t （度数分布表では10）以上であること

	P ₁	P ₂	P ₃	合計
M ₁	x_{11}	24	x_{13}	72
M ₂	x_{21}	38	x_{23}	116
M ₃	40	39	42	121
合計	98	101	110	309

1. 最小値問題

$$a_{23} = \min x_{23}$$

$$\text{拘束条件: } \begin{aligned} x_{11} + x_{13} &= 72 - 24 \\ x_{21} + x_{23} &= 116 - 38 \\ x_{11} + x_{21} &= 98 - 40 \\ x_{13} + x_{23} &= 110 - 42 \\ (x_{11}, x_{13}, x_{21}, x_{23}) &\geq 0 \end{aligned}$$

2. 最大値問題

$$a_{23} = \max x_{23}$$

$$\text{拘束条件: } \begin{aligned} x_{11} + x_{13} &= 72 - 24 \\ x_{21} + x_{23} &= 116 - 38 \\ x_{11} + x_{21} &= 98 - 40 \\ x_{13} + x_{23} &= 110 - 42 \\ (x_{11}, x_{13}, x_{21}, x_{23}) &\geq 0 \end{aligned}$$

$$\text{秘匿インターバル } w = \max x_{23} - \min x_{23} = 68 - 20 = 48 > 10$$

図6：秘匿化にあたっては、当該セル（この例の場合 x_{23} 以外に、 x_{11} , x_{13} , x_{21} のセルを隠す必要がある。

「Workshop on Statistical Analysis for Forest Resource Management Level 4 —Generalized Linear Model—」共同開催報告

カンボジアMOU締結研究機関Institute of Forest and Wildlife Research and Developmentにおいて、平成30年9月4-5日の二日間に渡り、ワークショップ「Workshop on Statistical Analysis for Forest Resource Management Level 4」を開催しました。本ワークショップは、統計解析ソフトRを用いたデータ解析の習得を目的としたものであり、その中でも森林データ解析に特化した内容を取り扱ってきました。これまでの継続した開催効果を踏まえ、今回はLevel4に設定した内容を教授しました。しかし、今回が初めての参加者も混在していたため、参加者全員に習熟度のセルフチェックシートおよび自己学習用の資料を事前配布することにより、参加者全体の習熟度の差異を解消することを試みました。このことにより、参加者自らの疑問点などもクリアになり、初日から積極的な質疑応答が繰り返された印象を受けました。Level4は一般化線形モデルに重点をおき、連続変数による回帰（線形回帰および成長関数に基づく非線形回帰）から、離散型（今回はロジスティック回帰とポアソン回帰）ま



でを網羅することにより、一般化線形モデルの基本的な概念やデータに依存した使い方を実践しました。今回のワークショップは実践的データ解析に主眼を置いて展開しましたが、その背景にある理論的な側面についても講義して欲しいと、引き続きの開催要望を受けました。ワークショップの運営は、Forest and Wildlife Research and Development Forestry AdministrationのSokh Heng氏らが中心に行いました。尚、今回の参加者は32名でした。（吉本 敦）

「Training Workshop Series V —Advanced Statistical Analysis in “R” for Forestry Related Data」共同開催報告

ベトナムMOU締結研究機関Forest Inventory and Planning Institute, Vietnamにおいて、平成30年12月5-6日の二日間に渡って、統計数理研究所リスク解析戦略研究センターとの共催によりワークショップ「Training Workshop Series V —Advanced Statistical Analysis in “R” for Forestry Related Data」を開催しました。本ワークショップは、統計解析ソフトRを用いたデータ解析の習得を目的としたものであり、その中でも森林データ解析に特化した内容を取り扱い、今回が5回目の開催となります。しかし、今回が初めての参加者も混在していたため、参加者全員に過去のワークショッ



プ内容に関するエクササイズを用意し、参加者全体の習熟度の差異を解消することを試みました。このことにより、参加者自らの疑問点などもクリアになる一方で、エクササイズから派生する最新の論文における手法等に関して初日から積極的な質疑応答が繰り返されました。今回は多変量解析に重点をおき、一般化線形モデル、特に連続変数による回帰（線形回帰および成長関数に基づく非線形回帰）から、離散型（今回はロジスティック回帰とポアソン回帰）までを網羅しました。また、実解析上のニーズとして主成分分析とクラスター分析についても教授しましたが、全てのテーマを完了するに至らなかったため、次回以降の課題となりました。一方で、実践的データ解析手法の背景にある理論的な側面についても講義して欲しいという要望もあり、この点も次回以降の課題となっております。5回のワークショップを終え、受講者もRの利活用に慣れてきた印象を持ちました。近年、Rに関する書籍もベトナム語で出版されるようになってきたのも、R普及の契機になってきたと聞きます。ワークショップの運営は、Forest Inventory and Planning Institute, VietnamのNguyen Hung Dinh氏らが中心に行いました。尚、今回の参加者は30名でした。（吉本 敦）

2018年10月-12月の公開講座実施状況

10月1日(月)に当研究所の小山慎介講師による「イベント時系列解析入門」の講義が行われました。神経スパイク発火、地震、Twitter・YouTubeなどのオンライン上の行動履歴や金融市場における注文履歴など、様々な現象がイベント時系列として表現できることを説明し、イベントが起きた時のデータに基づき、モデリングや統計解析の方法などについて講義をしました。

11月14日(水)は、当研究所の島谷健一郎講師による「なぜ分散の推定は $n-1$ で割るのか? 統計学における推定の再入門」の講義が行われました。 $n-1$ で割る推定は「不偏推定量」という統計学における主要な概念を根拠の第1とするが、これを出発点に、有効性、一致性、漸近正規性等々、推定

量に関する諸概念について、最尤推定とベイズ推定を中心に講義しました。

11月29日(木)~30日(金)は、筑波大学東京キャンパスにおいて、北海道大学の久保拓弥講師と当研究所の野村俊一講師、伊庭幸人講師によるリーディングDAT講座L-B1「統計モデリング入門」の講義が行われました。GLMから階層ベイズモデリングについては久保講師が担当し、状態空間モデルによる時系列解析—KFASによる実践モデリングに関しては野村講師が担当しました。伊庭講師は補足と展望について講義しました。

12月13日(木)~14日(金)は、筑波大学東京キャンパスにおいて、当研究所の伊庭幸人講師、福水健次講師、日野英逸講師、今泉允聡講師、電気通信大学の庄野逸講師、北海道大学の瀧川一学講師によるリーディングDAT講座L-B2「機械学習とデータサイエンスの現代的手法」の講義が行われました。伊庭講師によるイントロダクションから始まり、福水講師によるカーネル法：基礎から最近の発展までを説明し、今泉講師が行列データ・テンソルデータの機械学習について講義しました。翌日は庄野講師によるニューラルネットワーク入門から始まり、瀧川講師による決定木・回帰木に基づくアンサンブル学習について説明し、最後は日野講師によるスパース推定入門の講義をしました。

(情報資源室)

行列・テンソルデータの増加

多くのデータ分析はベクトルを扱ってきた
→ データの多様化で行列・テンソルが増加



「リーディングDAT 講座L-B2」資料より

統計数理セミナー実施報告(2018年11月~2019年1月)

毎週水曜16時から所内研究教育職員および外部の方が1人40分ずつ、1日に2人の講演を行っています。2018年11月~2019年1月のセミナーは下記の通り行われました。

日程	氏名	タイトル
2018年11月 7日	島谷 健一郎	欠損を伴う個体ベースデータからの個体群動態パラメータ推定法
11月 7日	豊田 充	確率ブーリアンネットワークの最適制御と応用
11月14日	松井 知子	話者認識技術の研究動向と課題
11月14日	長幡 英明	多変量時系列に対する分散分析
11月21日	小山 慎介	加法的な計数時系列ネットワークを用いたイベント連鎖のモデリング
11月21日	江口 真透	一般化ミクスチュアモデル
11月28日	前田 忠彦	社会調査の実施に関わるいくつかの話題
11月28日	三分一 史和	時系列データにおける変化点の検出
12月 5日	加藤 昇吾	2変量分布の上側と下側の裾確率を比較するための尺度
12月 5日	山田 寛尚	機械学習による有機・無機分子の物性予測
12月12日	森井 幹雄	Angular resolution booster を用いたX線望遠鏡のイメージ再構成
12月12日	張 俊超	高齢者向け無料交通制度の政策評価およびそのメカニズムの解明
2019年 1月 9日	足立 淳	古代DNA解析による絶滅生物と生物多様性の歴史
1月 9日	栗木 哲	クロネッカー積共分散構造の最尤推定量の存在・非存在 — あるテンソルランク最小化問題

日 程	氏 名	タイトル
1月30日	早水 桃子	細分系統樹の構造定理がもたらす系統データ解析の新展開

セミナーの開催予定はホームページにてご案内しています。 <https://www.ism.ac.jp/>

(メディア開発室)

平成30年度若手研究者クロストークの開催報告

11月20日、TKP 東京駅日本橋カンファレンスセンターにおいて大学共同利用機関法人4機構連携企画の若手研究者クロストークが開催されました。当初このクロストークは9月3日から4日にかけて南房総での開催を予定していましたが、台



風の影響により中止となりました。この中止を残念に思われる多数の声を受け、時期・会場・規模を変更して開催する運びとなりました。今年度は情報・システム研究機構の若手研究者や総合研究大学院大学の大学院生に加え、人間文化研究機構の若手研究者にもご参加いただき、3名の講師による講演とポスターセッションが行われました。統数研の上野玄太教授と遺伝研の小田祥久准教授による講演では、分野を横断した融合研究について当時の苦労話を交えながらの紹介がありました。藤井良一機構長の講演では、若手研究者に期待するメッセージが語られました。また、ポスターセッションでも活発な意見交換が行われました。このクロストークを端緒とする融合研究が今後生まれるとしたら企画委員・幹事としての苦労が報われるのですが、果たしてどうでしょうか。今後が期待されます。(田中未来, Stephen Wu)

2018年公開講演会「創業のフロンティア:データサイエンスの挑戦」

統計数理研究所では毎年、教育・文化週間(11月1日～7日)関連行事の一つとして公開講演会を開催しています。この講演会では、一般の方を対象に、統計数理に関連したテーマをわかりやすくお話しています。

今年度のテーマは「創業のフロンティア:データサイエンスの挑戦」で、2018年11月1日に統計数理研究所大会議室で開催されました。3件の講演を企画し、91名の聴講がありました。

1. データ化される健康と病気: バイオインフォマティクスの挑戦
井元 清哉(東京大学医科学研究所 教授 ヘルスインテリジェンスセンター センター長)
2. データサイエンスを活用して薬剤分子を設計・合成する
吉田 亮(統計数理研究所 ものづくりデータ科学研究センター センター長)
3. 医薬品評価における医学統計学の役割
伊藤 陽一(統計数理研究所 医療健康データ科学研究センター センター長)

DNAシーケンシングの低コスト化により、個人ゲノムの情報を安価に取得できるようになり、個別化医療の実現に向けた動きがいよいよ活発になってきました。最初の講演では、最新の研究事例にもとづき、大量かつ多様なゲノム情報を有効活用するための統計科学の基盤技術と将来展望が示されました。

二番目は、薬剤分子の設計を目的とする統計科学に関する講演でした。広大な探索空間から薬の要求特性を満たす新規化合物を発掘する人工知能技術が紹介されました。三番目の講演では、医薬品開発において薬の有効性と安全性を定量化するために統計科学がどのように活用されているのかが述べられ、さらに市販後の薬の安全性評価における統計科学の活用事例が紹介されました。

公演後のアンケートでは「最先端の研究や概念の説明を、無料で聴くことができるのは、とても有益だと感じました」「分かりやすくデータサイエンスの研究が聞くことができてよかった」「どのプレゼンテーションもわかりやすく説明いただき、統計学をもっと理解したいと思いました」などのご意見をいただきました。(オーガナイザー・吉田 亮)



国際外部評価委員会の開催について

平成30年10月11日、12日の二日間、国際外部評価委員会が統計数理研究所において開催されました。本委員会のミッションは、本研究所のNOE (Network Of Excellence) 形



国際外部評価委員会

成事業を中心に、基幹的研究組織、国際連携、大学共同利用機関として備えるべき施設（スーパーコンピュータ、Akaike Guest House等）、大学院教育の評価です。本委員会のメンバーは、Sadaoki Furui先生（Toyota Technological Institute at Chicago）【委員長】、Chun-houh Chen先生（Institute of Statistical Science, Academia Sinica）【副委員長】、Arnaud Doucet先生（University of Oxford）、Mark Girolami先生（Imperial College London）、Satoshi Itoh先生（National Institute for Materials Science）、Sadaaki Miyamoto先生（University of Tsukuba）、Peter Jan van Leeuwen先生（University of Reading）です。今後、本委員会から研究所に提出される報告書は、本研究所の活動をより発展させるために活用いたします。（評価室）

「立川市アカデミックアドバイザー」任命式と連絡協議会幹事会の開催

7月19日（木）、朴准教授が初代「立川市アカデミックアドバイザー」に任命され、立川市役所市長室にて任命式が行われました。統数研と立川市は2015年9月に連携・協力に関する協定を締結しており、今回の任命はその連携協力の一環となります。清水市長から朴准教授へ辞令交付が行われた後、今後の統数研と立川市の協力事業が話題となり、樋口所長より統数研で開催する講座などへの立川市職員の参加という人材育成協力について提案が行われました。

10月17日（水）、立川市と統数研の連携協力に関する連絡協議会幹事会が本研究所にて開催されました。立川市総合政策部小林部長と統数研金藤副所長の挨拶の後、今後の連携・協力事業について話し合いが行われました。最後に来年10月に迎える統数研と極地研の立川移転10周年やイ

ベント開催について説明が行われ閉会となりました。（広報室）



立川市アカデミックアドバイザー任命式での樋口所長、朴准教授、立川市清水市長（左より）

数学・数理科学4研究拠点合同市民講演会「AI社会の基盤は数学!」を開催

11月3日（土）、数学・数理科学4研究拠点合同市民講演会「AI社会の基盤は数学!」を開催しました。本講演会は京都大学数理解析研究所、九州大学マス・フォア・インダストリ研究所、明治大学先端数理科学インスティテュート、統計数理研究所の数学・数理科学の4研究機関が主催となり2015年から行われ、今年度は統計数理研究所が幹事として渋谷のシダックスカルチャーホールにおいて開催されました。

伊藤聡 統計数理研究所副所長の司会のもと、樋口知之

同研究所所長より開会の挨拶が行われ、続いて文部科学省研究振興局学術機関課 西井知紀課長よりご挨拶をいただきました。

その後、明治大学先端数理科学インスティテュート 砂田利一副所長による「AIは数学者になれるか?」と題した講演に続き、九州大学マス・フォア・インダストリ研究所 藤澤克樹教授による「AI + グラフ解析 + 数理最適化による新しい産業応用」、京都大学数理解析研究所 小林佑輔准教授による

「効率的なアルゴリズムと数学」、最後に統計数理研究所 島谷健一郎准教授による「自然観察で始まる生物多様性研究のどこに統計数理がある?」と題した講演が行われました。

来年度開催時に幹事となる明治大学先端数理科学インスティテュート 杉原厚吉所長による閉会の挨拶が行われ、来場者からは「意外な分野での応用に驚きと面白さを感じた」「実施中の研究開発に応用できる内容で有益だった」という感想や今後聞いてみたい講演テーマについて多くのご意見をいただき、盛況な講演会となりました。(広報室)



統計数理研究所子ども見学デー2018を開催

10月27日(土) 統計数理研究所子ども見学デー 2018が立川市・立川観光協会主催の立川体験スタンプラリーとの同時開催で実施されました。当日は240名の来場者を迎え、ランダムサンプリングの実験やじゃんけんゲームコーナー、コン



ピュータ見学ツアーなどのプログラムが行われました。ランダムサンプリングの実験「白玉・黒玉数当て」では、少量のサンプルをもとに10万個の白玉黒玉のうち黒玉の個数がどのように推定されるのかを実験し、多くの方が実験に参加されました。「じゃんけんゲーム」ではゲームを進めていくうちに出す手の確率がどのように変化していくのか考えながら挑戦してもらい、これらのプログラムを通じて楽しみながら統計学の一端に触れていただきました。

「コンピュータ見学ツアー」は3回開催し、43名が参加されました。今年10月より稼動を始めた新しい統計科学スーパーコンピューターや1950年代からの計算機が170台以上展示してある分散コンピュータ博物館の説明に興味深げに耳を傾ける様子が伺えました。(広報室)

「大学共同利用機関シンポジウム2018 ～最先端研究大集合～」に参加・出展

11月14日(日)に名古屋市科学館において、「大学共同利用機関シンポジウム2018 ～最先端研究大集合～」が開催されました。

各研究所の展示ブースがイベントホールに設けられ、本研究所は研究所紹介のポスターやパネル展示を行い、入場者は人材育成事業や日本人の国民性調査の説明に熱心に聞き入っていました。また昨年に引き続き『第2回統計検定統計数理1日マイスター』と題した問題の配布と認定証の発行を行いました。「難しそう」と言いながらその場で挑戦する来場者もあり、本研究所の活動や統計に関心を持ってもらう機会となりました。

イベントホールの一部では研究者によるミニ講演『研究者トーク』が行われました。情報・システム研究機構の発表では樋口所長が司会を務め、4研究所の研究者と講演概要の紹介を行いました。本研究所からは島谷准教授が「観察で



きなかった種数の推定と生物多様性」と題した講演を行い、講演会場には立ち見が出るほど多くの来場者が集まりました。

(広報室)

実践女子大学より生活心理学専攻の1・2年生が来所

昨年に引き続き、10月30日(火)と11月6日(火)に、実践女子大学生生活科学部 生活文化学科生活心理学専攻の1年生(13名)および2年生(15名)が、引率の作田由依子専任講師と共に当研究所を訪れました。学生たちの来訪目的は、



統計学と心理学との関係を学ぶことであり、「統計学を学ぶことで得た知識を、生活の中でどのように活用できるか」、「世論調査において、調査内容が同じなのに実施主体が異なると結果が異なるのは何故か」などの関心を持っているとのことでした。

当日は、はじめに田村義保名誉教授の解説のもと、計算機展示室およびスーパーコンピュータの見学が行なわれました。またその後には、稲垣佑典 特任助教と芝井清久 特任助教によって、統計学と心理学の関係についての解説、国際比較調査の結果解説が行なわれました。

新たに導入されたスーパーコンピュータに驚いたり、国によって大きく異なる国民性の話を興味深そうに聞いたり、今回の訪問は学生たちにとって有意義であったようです。

(社会データ構造化センター 特任助教 稲垣佑典)

総合研究大学院大学複合科学研究科統計科学専攻関係

大学院説明会について

平成30年11月2日(金)に、会議室1(D222)において、平成30年度第2回大学院説明会を開催し、28名の参加者がありました。これは昨年度同時期の参加者の2倍です。「統計」の需要が高まっていることを実感する数字でした。また社会人と学生の割合がそれぞれ65%と35%で、これまでよりも学生が多いという特徴もありました。

大学院説明会の前半は全体説明を行い、募集要項、カリキュラム、修了後の進路、学生の研究テーマなどを紹介しました。後半は、希望者に対して教員との面談を行いました。面談希望者の割合も年々増加しており、今回は28名中21名

が教員と面談しました。(研究推進課・大学院係)



●公開講座

一般社会人・学生を対象に、下記の公開講座を開催します。

① Rによる時系列解析入門(川崎 能典)

日時:6月18日(火)10時~16時(5時間)
講師:川崎能典(統計数理研究所)
申込受付:4月15日(月)10時~5月7日(火)10時
受講料:5,000円
会場:統計数理研究所
定員:100名(応募者多数の場合は抽選)
講義レベル:初級

時系列の図示/定常性/自己共分散関数/自己回帰

(AR)モデル/自己回帰移動平均(ARMA)モデル/時系列の変換/時系列の予測/多変量自己回帰モデル:これらをフリーソフトウェアRで実行し、出力を理解できることを目標に解説を加える。講義内で行うデモに関してはソースコードを配布する。(※時間の制約上、実習と銘打つほどの時間は確保できないので予めご了承の上お申し込みください。)

受講者に期待する予備知識やレベル:

相関、回帰、推定・検定等の基本的知識に基づき、重回帰分析の実行や結果の解釈が一通りできるレベルにあること。

参考書:

●北川源四郎「時系列解析入門」岩波書店、2005年。
その他:本講座(初級)では状態空間モデリングは扱わない。

B スパース推定

日時: 7月17日(水) 10時～16時(5時間)
 講師: 川野秀一(電気通信大学)
 申込受付: 5月27日(月) 10時～6月3日(月) 10時
 受講料: 5,000円
 会場: 統計数理研究所
 定員: 100名(応募者多数の場合は抽選)
 講義レベル: 中級

スパース推定とは、データ発生構造の疎性に着目した統計的推定法である。近年ビッグデータ(特に、超高次元データ)解析の一手法として注目を浴びはじめ、今後その重要性がますます高まるものと期待されている。本講座では、正則化法に基づいたスパース推定の入門的内容について概説する。具体的には、正則化法からはじめ、lasso法を軸としたスパース推定法、スパース推定の推定値を得るための計算アルゴリズム、構築したスパースモデルの評価方法について解説する。

受講者に期待する予備知識やレベル:

大学初級程度の微分積分や線形代数と、学部程度の統計学の基礎知識は前提とします。

参考書:

- 川野秀一, 松井秀俊, 廣瀬慧(2018) スパース推定法による統計モデリング. 共立出版.

C 多変量解析法

日時: 8月20日(火)～23日(金) 10時～16時(20時間)
 講師: 馬場康雄・清水信夫(統計数理研究所) 今泉忠(多摩大学)
 申込受付: 7月1日(月) 10時～7月8日(月) 10時
 受講料: 28,000円
 定員: 90名(応募者多数の場合は抽選)

会場: エッサム神田ホール
 講義レベル: 初級

多くの現象は一つの変数で観測されるものではなく多数の変数の観測によって把握される。すなわち多次元のデータによって現象が表現される。得られた多次元のデータを用いて、数量の推測をする、判別をする、あるいは尺度を作る等の手法の総称が多変量解析法である。重回帰分析、判別分析、主成分分析、因子分析、数量化など、多変量解析の古典的・標準的な手法の解説をする。また、クラスター分析、共分散構造分析などの解説をする。

受講者に期待する予備知識やレベル:

平均、分散、標準偏差等、統計学の基礎的な概念を知っていることを前提とする。手法の数学的な説明よりは、用い方に重点をおいた解説を行うが、理解をたやすくする点から、大学初級程度の線形代数と微分積分の知識があることが望ましい。

教科書: 資料を配布します。

備考: この講座は社会調査協会の社会調査士認定科目申請中です。

- 一般講座の他、系統的な講座編成により現代的な統計科学の姿を示す「リーディングDAT講座」、医学・健康科学分野のデータサイエンスを基礎から最先端まで詳説する医療健康データ科学研究センターの「系統的教育コース・公開講座」も開講予定です。

詳細は、以下のwebサイトをご覧ください。

<https://www.ism.ac.jp/lectures/kouza.html>

(情報資源室)

共同利用

H30年度共同利用公募追加採択課題

【一般研究1】2件

分野	研究課題名	研究代表者(所属)
a5	津波リスク評価に係る多数アンサンブル津波伝播シミュレーション	北野 利一(名古屋工業大学・教授)
c4	データ同化手法による核融合プラズマの統合輸送シミュレーション	村上 定義(京都大学・教授)

(研究推進課・共同利用係)

外部資金・研究員等の受入れ

受託研究・受託事業等の受入れ

委託者の名称	研究題目	研究期間	研究経費(円)	受入担当研究教育職員
学校法人日本医科大学 理事長 坂本 篤裕	非小細胞肺がんの転移を予測し術後補助化学療法の効果を予測するバイオマーカーの開発のための薬事承認申請のための臨床性能試験実施を目的とした生物統計解析	H30.4.1～ H31.3.31	200,100	医療健康データ科学研究センター 長島 健悟 特任准教授

委託者の名称	研究題目	研究期間	研究経費(円)	受入担当研究教育職員
国立大学法人京都大学 学長 山極 壽一 代理人 医学・病院構内共通事務部長 廣瀬 幸司	京都大学大学院における臨床統計家 育成のための教育カリキュラムの標準 化のための研究開発	H30.4.1～ H31.3.31	5,000,000	樋口 知之 所長 数理・推論研究系 逸見 昌之 准教授
国立大学法人京都大学 学長 山極 壽一 代理人 医学・病院構内共通事務部長 廣瀬 幸司	患者特性に応じた薬物療法・精神療法 の個別化医療とその臨床試験プロトコ ルの開発研究	H30.4.1～ H31.3.31	130,000	データ科学研究系 野間 久史 准教授
国立研究開発法人科学技術振興 機構 分任研究契約担当者 契約部長 岩田 一彦	深層学習の高速化にむけた適応ネット ワークの数学的発見と学習法開発	H30.10.1～ H31.3.31	9,100,000	統計思考院 今泉 允聡 助教
国立研究開発法人新エネルギー・産 業技術総合開発機構 理事長 石塚 博昭	深層確率コンピューティング技術の研 究開発	H30.9.5～ H31.3.31	12,000,000	樋口 知之 所長 数理・推論研究系 福水 健次 教授

(研究推進課・研究推進係)

外来研究員の受入れ

氏 名	職 名	研究題目	研究期間	受入担当研究教育職員
Mitja Alexander Stadje	Ulm University, Institute of Insurance Science, Professor	バックワード確率微分方程式のパラメータ推定問題とマーケットインパクト	H30.10.13～ H30.10.20	荻原 哲平 助教
Gloster Arnaud Edouard	Université d' Evry Val d' Essonne, Professor	拡散過程のパラメータ推定と局所漸近混合正規性について	H30.10.18～ H30.10.26	荻原 哲平 助教
廖 怡雯	National Central University, Earthquake-Disaster&Risk Evaluation and Management (E-DREaM) Center, Research Assistant	ETASモデルとクーロン応力伝達に基づく推定した断層面を組み込んだ余震予測	H30.10.22～ H30.11.21	庄 建倉 准教授
徐 泓	名古屋大学大学院経済学研究科・博士課程後期生	データマイニング方法で企業のデフォルトリスクの分析	H30.11.1～ H31.3.31	山下 智志 センター長
Arthur Pewsey	University of Extremadura, Associate Professor	周期的な確率密度関数を持つ多変量コンピュータに関する研究	H30.11.22～ H30.12.3	加藤 昇吾 准教授
平井 貴裕	住友化学株式会社・研究員	データ科学を用いた材料設計	H30.10.4～ H31.3.31	吉田 亮 センター長
Xiaoling Dou	早稲田大学理工学術院・講師	Bスプラインコンピュータの研究	H30.11.8～ H31.3.31	栗木 哲 教授
伊高 静	早稲田大学人間科学学術院・招聘研究員(平成30年12月より)	森林資源管理における病虫害拡散予測および制御に関わる順応的最適化システムの構築	H30.12.1～ H31.3.31	島谷 健一郎 准教授
Rui Li	Sun Yat-sen University (Zhongshan University), Student	自動材料設計アルゴリズムの開発	H31.2.10～ H31.3.24	吉田 亮 教授
龔 萱	Chinese Academy of Science, Institute of Geology and Geophysics, PhD student	時空間ETASモデルに基づくビルマの地震活動シミュレーション	H31.1.3～ H31.2.2	庄 建倉 准教授
司 政亜	China Earthquake Administration, Institute of Geology and Geophysics, PhD student	地震検出率の日変化の統計特徴	H31.1.3～ H31.2.2	庄 建倉 准教授
王 林海	China Earthquake Administration, Institute of Geology and Geophysics, PhD student	地震活動と重力場変動の相関性のモデリング	H31.1.15～ H31.2.14	庄 建倉 准教授
渡邊 隼史	金沢大学理工学域電子情報学類・助教(平成30年12月より)	アパートローンリスク計量、データ構造化	H31.1.1～ H31.3.31	山下 智志 センター長
Jin Sun	University of Technology Sydney, PhD student	金融対策のための近似的動的計画法アルゴリズムの研究	H31.1.4～ H31.1.30	松井 知子 教授
Ye Chen	University of Technology Sydney, Post-doctoral research fellow	時系列データのためのベイズ計算統計モデルの研究	H31.2.23～ H31.3.3	松井 知子 教授

(研究推進課・共同利用係)

寄附金の受入れ

受入決定年月日	寄附者	寄附金額(円)	担当教員	寄附目的
H30.12.12	新菱冷熱工業株式会社	1,000,000	日野 英逸	「データ解析に関する研究」に対する研究助成

(財務課・予算・決算第二係)

人 事

外国人研究員(客員)

氏 名	現 職	所 属	職 名	研究課題	期 間	受入教員
フォア フレデリック キン ヒン Phoa Frederick Kin Hing	准教授	Institute of Statistical Science, Academia Sinica, Taiwan	客員准教授	Web of Scienceデー タの解析	H30.9.20～ H30.10.16	中野 純司 教授

(総務課・人事係)

外国人客員紹介



●Frederick Kin Hing Phoa 客員教授

I would sincerely thank Professor Nakano for their invitation to my one-month visit the Institute of Statistical Mathematics (ISM). This is by far my longest visit in terms of time to a single university/research institute in my nine-year career. The main purposes for my visit are to complete several projects in the Web of Science and write several research articles with Professor Nakano and his Ph.D. student Ms. Livia Lin-Hsuan Chang, plus additional visits and research collaborations with Japanese researchers outside ISM. During this period, I am grateful to experience on the different working environment, culture and attitude in Japan when compared to Taiwan and USA, which I have yet felt in my past short-time visits. In addition to the discussions with Professor Nakano, I am glad to continue the collaborations with Dr. Keisuke Honda and his team on other projects in the Web of Science. I only regret that the visiting period is still too short to accomplish bigger tasks, and to connect and explore potential research collaboration with other colleagues in ISM. In summary, I can firmly say that this visit is successful and highly productive. I am looking forward to similar visiting opportunity of ISM in the coming future.

刊行物

統計数理 第66巻 第2号

特集「サービス科学の今」

「特集 サービス科学の今」について

丸山 宏、中野 純司189

アジャイルな社会に向けて [総合報告]

丸山 宏193

ビッグデータを活用する確率モデリング技術—社会実装の取り組みと課題— [総合報告]

本村 陽一213

位置情報軌跡の統計的プライバシー保護 [総合報告]

南 和宏225

大規模集計POSデータの高次元スパースモデリング [研究詳解]

李 銀星、照井 伸彦235

統計モデルによる消費者理解の可能性 [研究ノート]

佐藤 忠彦249

地域健康政策へのベイジアンネットワークの応用 [研究ノート]

鳥海 航、生方 裕一、久野 譜也、岡田 幸彦267

集約的シンボリックデータのカイ2乗統計量を用いた非類似度とその不動産情報データへの適用 [原著論文] 清水 信夫、中野 純司、山本 由和	279
B-スプライン及び Adaptive Group LASSO に基づく正則化非線形ロジットモデルによるデフォルト確率の推定 [原著論文] 高部 勲、山下 智志	295
トータルパワー寄与率を用いた海洋生態システムにおける因果性推測 [研究ノート] ソルヴァン加藤 比呂子、Subbey Sam	319
P ³ : Pythonによる並列計算機用粒子フィルタライブラリ [統計ソフトウェア] 中野 慎也、有吉 雄哉、樋口 知之	339
	(メディア開発室)

Annals of the Institute of Statistical Mathematics

Volume 71, Number 1 (February 2019)

Mijeong Kim and Yanyuan Ma Semiparametric efficient estimators in heteroscedastic error models	1
Denis Belomestny, Fabienne Comte and Valentine Genon-Catalot Sobolev-Hermite versus Sobolev nonparametric density estimation on $\mathbb{R}$	29
Omer Ozturk Two-stage cluster samples with ranked set sampling designs	63
Kun-Lin Kuo and Yuchung J. Wang Pseudo-Gibbs sampler for discrete conditional distributions	93
Benedikt Bauer, Felix Heimrich, Michael Kohler and Adam Krzyżak On estimation of surrogate models for multivariate computer experiments	107
Jiajuan Liang, Kai Wang Ng and Guoliang Tian A class of uniform tests for goodness-of-fit of the multivariate L_p -norm spherical distributions and the l_p -norm symmetric distributions	137
Robert E. Gaunt, Satish Iyengar, Adri B. Olde Daalhuis and Burcin Simsek An asymptotic expansion for the normalizing constant of the Conway-Maxwell-Poisson distribution	163
Ursula U. Müller, Hanxiang Peng and Anton Schick Inference about the slope in linear regression: An empirical likelihood approach	181
Dennis Dobler Bootstrapping the Kaplan-Meier estimator on the whole line	213
	(メディア開発室)



悪魔の証明

田中 未来

数理・推論研究系

悪魔の証明とは、もともとは所有権帰属の証明の困難性を比喩的に表現した言葉だったようだが、今では消極的事実の証明の困難性を表す。例えば、“地球上にツチノコはいる”という命題を証明するためには実際にツチノコを見つければよいのに対し、“地球上にツチノコはいない”という命題を証明することは難しい。どんなに探しても見つからないという“証明”に対しては探し方が悪いせいだという反論の余地がある。

私はツチノコの研究をしているわけではなく（それで食っていけるならやってみたいがそれはともかく）、学生時代から数理最適化の研究をしている。数理最適化とはある性質（制約条件）を満たすもの（解）の中である指標（目的関数）が最小あるいは最大となるもの（最適解）を求めるための数理工学的方法論を指す。例えば乗換案内などは出発地と目的地を結ぶ経路の中から移動時間や運賃が最小となるものを教えてくれるシステムであり、数理最適化の身近な例といえる。最尤推定など統計数理に関する応用も枚挙に暇がない。

数理最適化はある意味で悪魔の証明との戦いといえる。“ある解が最適解である”という命題は本質的には“制約条件を満たす解にはその解より目的関数をよくするものは存在しない”という消極的事実である。一般にこれを示すことは悪魔の証明に他ならない。制約条件を満たす解が有限個であれば、すべての解を列挙し、おのおのについて目的関数の値を計算すればよいような気もする。しかし、この方法は問題が大きくなると現実的な計算時間では不可能になる（このよう

な現象はしばしば組合せ爆発とも呼ばれる）。制約条件を満たす解が無数にある場合はそもそもすべての解を列挙することすら不可能である。ところが面白いことに、問題がよい性質をもつ場合やなんらかの仮定のもとでは、最適性の証明が可能となる場合がある。例えば凸性や制約想定などのもとでは、最適性条件を満たす Lagrange 乗数を見つければ、それがいま着目している解が最適解であることの証拠となっている。

よい性質や仮定のもとで悪魔の証明が可能となる例は実社会にも存在する。そのひとつを以下に紹介しよう。あるとき私は友人から次のような相談を受けた。“結婚を考えている交際相手と知り合ったコミュニティが怪しい宗教と関係があるのではないかと交際相手の母親に疑われ、その結果として自分自身も怪しい宗教を信仰していると疑われてしまった。このコミュニティが宗教と無関係であることをもってもらしく説明することはできないか？”悪魔の証明の依頼である。これを聞いた私は少し考えて次のように答えた。“同時に2つの宗教を信仰することはないという仮定は受け入れやすいだろう。この仮定のもとでは、そのコミュニティ内でなんらかの宗教を信仰している人を見つけることができれば、その人が無関係であることの証拠となっている。”友人が実際にこの証明を説得に使ったかどうかを私は知らない（し、使えないような気もする）が、2人の仲睦まじい姿はなにかを“証明”しているようにも思われる。

