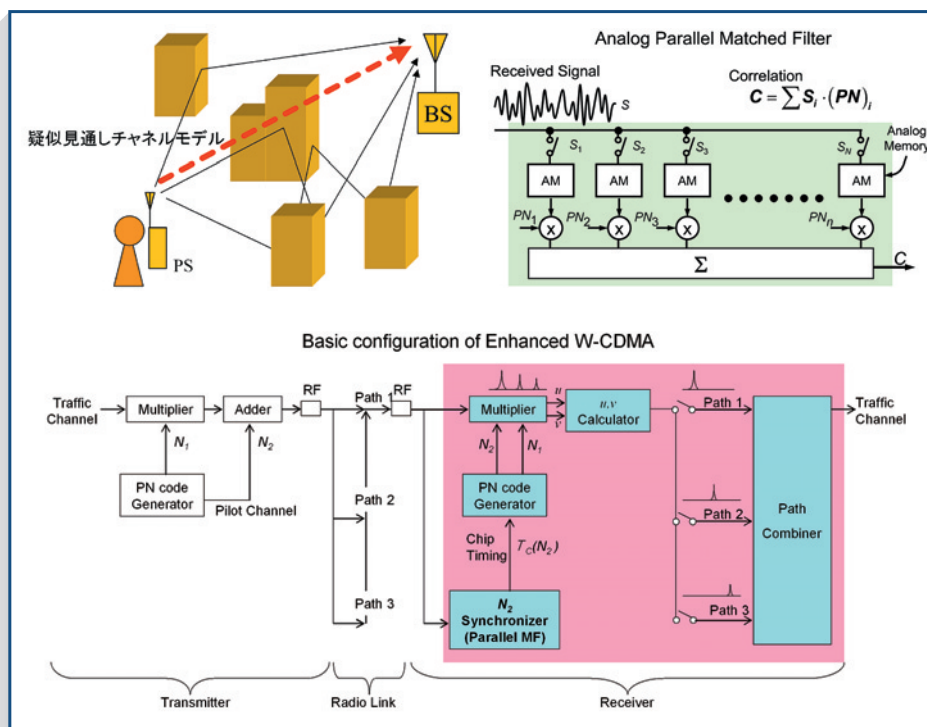




モデル化と高効率データ処理に基づく無線データシステムの研究

大学共同利用機関法人 情報・システム研究機構

統計数理研究所ニュース

C
O
N
T
E
N
T
S

■ 研究紹介	2	■ 共同利用	10
■ シンポジウム報告	4	平成19年度共同利用公募追加採択課題	
国際ワークショップ Data-Mining and Statistical Science 2007 (DMSS2007) 報告		■ 外部資金・研究員等の受入れ	10
■ 研究教育活動	5	共同研究の受入れ／受託研究の受入れ／外来研究員の受入れ／寄附金の受入れ	
新入研究教育職員紹介／2007年9月～11月の公開講座実施状況／「ISM オープンフォーラム第13シリーズ」の実施		■ 人 事	10
■ 統数研トピックス	6	■ 会議開催状況	11
平成19年度文部科学省「子ども見学デー」に参加・出展／イノベーション・ジャパン2007ー大学見本市に出展／大学院生及び学生のため夏期大学院の講座を開講		運営会議の開催／共同利用委員会の開催	
■ 総合研究大学院大学複合科学研究科統計科学専攻関係	7	■ 所外誌掲載論文等	11
■ お知らせ	8	■ 刊行物	12
統計数理セミナー／公開講座／ISM オープンフォーラム／「統計数理」特集論文募集について		Research Memorandum／統計数理研究所研究リポート／統計計算技術報告／Annals of the Institute of Statistical Mathematics	
		■ コラム	14
		■ あとがき	14

モデル化と高効率データ処理に基づく無線データシステムの研究

瀧澤 由美 モデリング研究系



研究目的と背景

この研究は、新領域融合研究プロジェクト：機能と帰納のテーマのひとつとして行っている。(1)有益な科学技術をいかに生み出し、日本のリーダーシップを確保できるか、(2)広く社会の発展(進歩とアメニティ)に寄与し格差を生じさせない、の2点に重点をおき、情報・通信の融合による新しい公衆情報サービスの実現を目的としている。

IT社会を支えるツールとしてのコンピュータ技術は米国がリードし、一方、広帯域通信技術の点では携帯電話は日本がリードしてきた。「組み込みシステム」は日本の活路として可能性がある。組み込みシステムとして、情報と通信を結ぶ専用コンピュータに着目する。

研究の内容と特徴

パーソナル通信サービス(PCS)はユーザ主体の新しいサービスを開拓した。その中で高品質画像、高速大規模データへの需要が急進展する一方、技術的には対応困難となっている。通信コストの過大、アクセスの困難、低い通信品質などを解決するため、(1)無線チャネルの広帯域化と(2)組み込みシステムの高機能化の研究を進めている。特に、ホットスポットに依存しない広域公衆移動無線通信を実現する。これは無線LANとは異なる。さらに、(3)大規模データの圧縮を実現する高効率情報源符号化を残された命題として捉えている。

具体的内容

(1)広帯域無線チャネルの研究

上記の困難な問題解決のために、情報通信理論、数理統計学を含む科学的方法の原点に立ち返り、モデル化とアルゴリズム(帰納)に基礎を置き、かつ高性能、省電力、低コストの通信システム(機能)の実現を図る。

デジタル通信システムにおける最も困難な技術は同期であろう。同期とは、信号の入力時刻を正確に決

定することである。同期の高速化、高信頼化、コヒーレント伝送特性の向上と高速化に焦点を絞り、理論とシミュレーション技術の向上を図り、実現特性を明らかにしている。

(2)通信用高機能組み込みシステムの研究

実際のシステムに求められる多様で複雑な機能の中で、高品質画像を含む高速データ通信機能を最重要課題として組み込みシステムをモデル化(帰納)し研究を行っている。これにより低電力での画像処理を可能とするシステムの構築を目指す。

データの処理量および処理速度から以下の構成を実現する。言語系を備えることによりソフトウェアのフレキシブルな構成を実現する。通信処理を基幹機能として含むOSを構成し、さらに高度な通信処理に対応するソフト/ハードウェアを並列的に設ける。また、メモリおよび制御ソフトウェアの管理により、矛盾するリアルタイム処理と必須の並列処理の実現を図る研究を行っている。

Proposed embedded system for advanced mobile communications	Conventional embedded system
1. Language system programming is available.	1. No programming is not applicable.
2. Communication protocol embedded in OS kernel as standard function.	2. on-demand
3. Multi-process and priority control based on memory protections.	3. Restricted
4. Unique OS for security	4. Open OS easy for development, low security

図2 情報・通信端末用組み込みシステムのコンセプト

(3)検証のためのプロトタイプハードウェアの開発(検証)

方式とアルゴリズムの検証のため、広帯域無線方式および高機能組み込みシステムについて、プロトタイプハードウェアの開発を行っている。

今後は、神経と情報の視点から画像の本質的研究に進みたいと考えている。

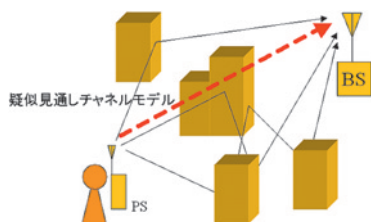


図1 実際のマルチパス無線伝播路と疑似見通チャンネルモデル。



図3 無線通信用組み込みシステムプロトタイプ。図中、右は組み込みコンピュータ、左は音響出力及びディスプレイ。

ゲノム時代の分子系統樹推定と分子進化の解明

足立 淳

予測発見戦略研究センター



地球上の生物が、ある共通の祖先から進化によって種分化して現在の多様性を獲得したのなら、それを「系統樹」として表現できる

はずである。ダーウィン以降の生物学者はそう考え、生物の形態情報から生物群を分類して系統樹を描き始めた。そして、生物の遺伝情報の実態である DNA 分子が発見されると、その情報を基にして「分子系統樹」の推定が行われるようになった。遺伝情報の違いはランダムな突然変異の蓄積に由来するので、分子の配列の変化を確率モデルで表現できる。そこで DNA の塩基配列やタンパクのアミノ酸配列の置換をマルコフモデルとして記述し、最尤法により分子系統樹を推定する方法が主流となった。

分子系統樹を推定するためには、二つの点で適切な遺伝情報を持つ分子(データ)を選択する必要がある。第一に、同じ共通祖先から由来した相同な分子であり、その分子の系統が種の系統と一致しなければならない。多重遺伝子族などゲノム上で重複が起きた遺伝子は、相同な遺伝子の選択が難しいので避けられてきた。第二は、点突然変異に由来する遺伝情報の変化が適度に蓄積している分子を選ぶことである。変異が少なすぎると系統関係を示す情報量が少なく、逆に変異が多すぎると情報は様々なゆらぎやノイズに埋もれてしまう。これまでの分子系統樹の推定では情報量を増やすために比較する遺伝子の数を増やし、また、ゆらぎやノイズに埋もれた情報を引き出すために、置換モデルを現実の進化(点突然変異の蓄積)に近づけるように改良を重ねてきた。

ヒトゲノムの配列データが発表されて以来、様々な種のゲノムが公表され、分子系統学は新たな局面を迎える。これまで、試行錯誤を繰り返しながら積み重ねた経験に基づいて選択してきたデータを、多種のゲノム情報を比較することにより客観的で網羅的な選択を行うことが可能となった。これまでデータが少なかつたために難しかった問題を解決できるかも知れないのである。ただし、ゲノムデータは完全なものではないということに注意しなければならない。ゲノムの多くは効率を重視したショットガン法により配列が決定されてきた。この方法は大量の断片からアセンブルによってゲノムを復元しなければならない。だが、プロジェクトによっては、その終了時に予算の関係で質よりも量を追うことになり、配列の精度が落ちたりアセンブルが不完全な部分を残していることがある。特に重複

によって同じような配列が散在する場合、アセンブルミスを誘発する。多重遺伝子族などの解析では、ゲノム配列の不完全さを考慮したデータの選択が必要である。さらに、ゲノム上の個々の遺伝子は互いに独立に進化してきたので、解析でも独立なモデルを仮定しなければならない。これまでの最尤法は、少数の遺伝子データを対象にしていたので、現在、大規模なデータを扱える新たな方法やアルゴリズムを開発し、ソフトウェアを整備しているところである。

ゲノム情報の急速な蓄積により、さまざまな生物間でゲノム全体の比較ができるようになった意義は大きい。遺伝情報全体を扱えるようになったということは、情報が単に量的に増えたということではなく、質的にも全く新しい基盤の上に立った議論が可能になったということを意味する。これまでの分子系統学は、生命の種分化の歴史の流れを知ることができたが、どのように進化してきたかというメカニズムの解明には力不足であった。だが、遺伝情報全体を比較することにより、ゲノム構造の進化の歴史を探ることが可能となったのである。

その第一歩として、遺伝子ファミリーの進化の解明を行っている。遺伝子の重複は遺伝子ファミリーの進化の引き金になるので、その歴史を遡ることはとても重要であるが、ゲノム上には互いに似た配列が多数存在するために、単純に比較しただけでは重複の歴史的順番を推定することは困難である。ある遺伝子重複の後一方の機能が変化すれば、その機能的制約によって進化速度が変化してしまう。また、遺伝子の機能が途中で失われると機能的制約が解かれて進化速度は急激に速くなる。つまり、遺伝子の進化速度の変化を無視した単純な配列比較では正しい結論が引き出すことができない。遺伝子の重複の順番を解明するためには、個々の遺伝子の機能的変化や機能消失を考慮した分子系統学的手法を新たに開発して解析する必要がある。

現在、相同な遺伝子の系統関係と、ゲノム上での互いの遺伝子の位置関係から、種間におけるゲノム構造の変異の歴史的順番を数値的最適化の手法を適用することによって解明することを試みている。分子系統学とゲノム比較を組み合わせることによって、ある遺伝子の進化の引き金となった突然変異が、ゲノム上でいつどのように起こったかを推定する。個々の突然変異が定着してきた歴史を解明することは、進化のメカニズムを知るための第一歩となるであろう。

●国際ワークショップ Data-Mining and Statistical Science 2007(DMSS2007)報告

さる2007年10月5、6日の二日間にわたり、国際ワークショップ「The International Workshop on Data-Mining and Statistical Science(DMSS2007)」が統計数理研究所にて開催された。本ワークショップは統計数理研究所の主催でとりおこなわれたもので、(社)人工知能学会『データマイニングと統計数理研究会』(略して DMSM。鷲尾教授(阪大)と樋口が共同主査)及び、情報・システム研究機構の新領域融合研究センターの研究プロジェクト『マルチモーダルデータからの不変情報の発見とその方法論の研究』(松井准教授 代表)の支援を受けている。二日間の開催中、約100名の参加の下、三つの招待講演と13の一般講演が行われた。プログラム等の詳細は、<http://daweb.ism.ac.jp/dmss07/>をご覧ください。

近年、社会の情報化が進み、我々は多種多様な情報を、計算機ネットワークを通じて入手し利用できる。一方で、生成・収集・蓄積されるデータの量は急激に増加しており、現在では蓄積された膨大なデータをいかに分析し活用するかが課題となっている。その有力な解決策の一つとして、大量のデータに埋没した有益な知識を発掘(マイニング)するデータマイニング技術が注目されている。この分野の広がりを受け、機械学習の研究も相乗効果的に盛んとなり、また統計科学の有効性と先駆性も昨今あらためて近隣分野から再評価されている。このような背景の下、本ワークショップは、機械学習を中心とする従来のアプローチに加え、統計数理学的アプローチを中核的な研究対象に据え、人工知能分野と統計数理分野の研究者間交流を深めるとともに、両アプローチの融合による新たなデータマイニング技術の創出を目指して企画された。

このワークショップは、科研費特別推進研究「半構造データマイニング」(研究代表者 有村教授(北大))とDMSMの第2回との共催で昨年(2006年)9月末に札幌にて開催した、DMSS06の成功を受けている。今回のDMSS07にも有村教授と鷲尾教授にはプログラム委員に入っただけで、多大なるご協力と助言を賜った。また本ワークショップの開催時期は、仙台にて開催された国際会議ALT07/DS07(Algorithmic Learning Theory/Discovery Science)の直後とし、相互に連動させることで海外からの研究者が帰国途中に立ち寄ることを期待した。具体的に行った協力作業はホームページで相互にリンクを張っただけであるが、実際に数名の海外の研究者が本ワークショップに参加くださった。また、松井准教授が招待講演者の招聘にあたり面倒な事務処理を含め諸々ご尽力くださったおかげで、最前線で活躍されている世界的に著名な研究者を招聘できた。

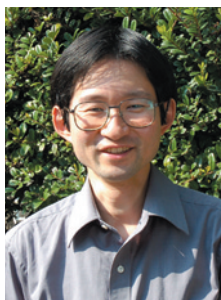
海外からの招待講演者の一人 Vert 教授は、グラフカーネルの理論的研究とそのバイオインフォマティクス分野への実践的应用で著名な方で、若手ながらすでに Centre for Computational Biology のセンター長を務めている、当該分野のフランスの代表者である。またもう一人の Smola 教授は、世界を股にかけた、機械学習理論分野で最も活躍されている研究者の一人である。本ワークショップではグラフに関する研究発表が多くなることを予想していたので、招待講演として東工大の下平准教授と共同研究をしている大学院生 Sheridan 氏にスケールフリーネットワークのベイジアンモデリングにおける活用法について講演してもらった。いずれの講演も講堂いっぱいになるほど聴講者が多く、講演後の議論も熱気あふれるハイレベルなものであった。

初日の夕方に懇親会を研究所内にて小規模に行う予定でいたが、あまりにも参加希望者が多く、秘書の方々等のサポートにより急遽部屋を変更するハプニングがあった。このような想定外のことも、ワークショップが活況を呈していた一つのあらわれである。また、本ワークショップでの発表論文に対しては査読を行う体制としたので(残念ながらご発表いただけなかった投稿論文もあった)、どの一般論文も発表準備が入念になされていた。最後の講演は土曜の夕方遅くなってしまったが、その時にも大勢の方が残られて発表を熱心に聴講されている姿は印象的であった。

本会議のあとに、データマイニングや統計数理に基づく解析を目的とした、理論・手法・技術に関する基礎から応用までを対象とした論文を広く募集し、本研究所の二つの学術雑誌、AISMと「統計数理」で特集号を組む。二誌同時に特集を組むことで、この分野にさらなる刺激を与え、近い将来諸関連研究分野が融合し、一つの領域名を冠した研究分野に発展していく端緒となることを願っている。

(モデリング研究系 樋口知之)

●新入研究教育職員紹介



数理・推論研究系学習推論グループ助教

逸見 昌之

医学研究あるいは健康科学において、新しい薬や治療の効果を評価したり、ある疾患の原因を探索したりすることは重要な問題ですが、そのプロセスには人間を対象とするが故の制約やメカニズムの複雑さから来る不確実な要素が多く、そのためさまざまな統計的手法が発展してきました。私はこれまで、英国のウォーリック大学統計学科において、メタアナリシス（目的を同じくするいくつかの研究から得られる統計的エビデンスを統合して、より強力なエビデンスを得るための統計解析）の方法論の研究に従事してきましたが、メタアナリシスもまたそのような手法の1つで、さらに、近年言われている EBM（根拠にもとづく医療）の中核を成す方法にもなっています。今後もこの研究を続けつつも、統計数理研究所の研究員として高い志しと広い視野を持ってさまざまな問題に取り組み、医学や健康科学などの実質科学に寄与する統計科学の発展に貢献していきたいと思います。



数理・推論研究系統計基礎数理グループ助教

小林 景

私は、東京大学情報理工学系研究科在学中より、主にカーネル法、ネットワークモデルなどの機械学習理論および、モデル選択やバイズ予測理論を研究してきました。その後、学術振興会の特別研究員として統計数理研究所に在籍して以来、当研究所に所属しています。その間、カリフォルニア大学バークレー校に1年半ほど滞在する機会に恵まれ、現地の機械学習や統計学の優秀な若手研究者と議論するという大変刺激的な経験ができました。また、バークレーには大規模ランダム行列の専門家たちが多く集まっていた影響もあり、現在はその機械学習理論への応用に関心をもっています。機械学習理論も大規模ランダム行列理論も、数学、工学、物理を合わせた広大な複合領域的性質をもつ点が面白く感じます。幸い当研究所には様々な専門知識をもつ方々が集っており、ゼミや共同研究を通してその地の利を生かしていこうと考えておりますので、どうぞよろしくお願いいたします。

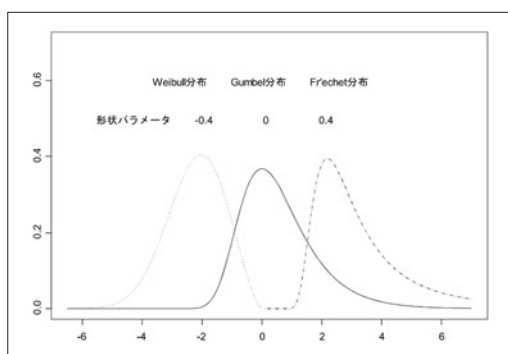
●2007年9月～11月の公開講座実施状況

9月10日(月)から11日(火)の2日間にわたり、今年度5回目の公開講座「極値統計学」を実施しました。

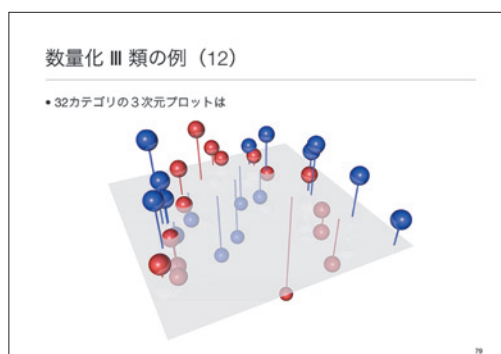
講師は当研究所の志村隆彰助教、渋谷政昭慶應義塾大学名誉教授、高橋倫也神戸大学教授、牧本直樹筑波大学准教授でした。受講生は、定員を大幅に上回り、61名で首都圏以外からも、静岡県から2名、京都府、兵庫、愛知、広島の各県から1名ずつ参加されました。

10月3日(水)から4日(木)の2日間にわたり、今年度6回目の公開講座「質的データの数量化分析法」を実施しました。

講師は当研究所の馬場康維教授、土屋隆裕准教授でした。受講生は、47名で首都圏以外からも、京都府から3名、岐阜県から1名参加されました。



公開講座「極値統計学」



公開講座「質的データの数量化分析法」

●「ISM オープンフォーラム第13シリーズ」の実施

現在、インターネット上などでいろいろなデータが利用できるようになりました。それらの中には音声や画像、DNAなどの系列データも多く含まれます。今回のシリーズでは統計数理研究所准教授松井知子の企画で、「系列データの統計科学的処理－ベイズアプローチ、カーネル法の最先端－」をテーマとし、7月から9月まで3回にわたり、系列データのための統計科学的方法について紹介が行われました。7月19日には松井知子准教授による「音声データのための統計科学的方法」が、8月17日、9月28日にはそれぞれ Arnaud Doucet 教授(University of British Columbia)、Jean-Philippe Vert 教授(Ecole des Mines de Paris)を招いて、「Introduction to particle filtering and applications to sequential data」、「Kernel for strings and applications in bioinformatics」について最先端の方法が説明されました。本シリーズには計54名の参加者があり、それぞれの講演終了後には活発な議論が行われました。統計数理研究所では音声や映像、バイオなどの系列データの統計科学的処理に関する研究を進めており、今後も講演会などを企画して研究紹介の機会を設けていく予定です。

(松井知子)

統計研トピックス

●平成19年度文部科学省「子ども見学デー」に参加・出展

統計数理研究所の展示及び体験コーナーは、「割合を求めてみよう！」というテーマで出展しました。夏休みもいよいよ終わりに近づいていることもあり、8月22日(水)と23日(木)の2日間で親子合わせて1082人もの入場者がありました。世論調査、視聴率調査等で、支持率、視聴率といった割合が発表されています。調査で得られる割合の値が確率的な数字であり、誤差を含んでいるものであることを理解してもらうために、白いBB弾を7万5千個、黒のBB弾を2万5千個入れた水槽の中から、両方合わせて300個のBB弾をランダムに計量カップですくってもらって、黒のBB弾がどれ位あるかを木箱と定規を使い実際に数えてもらいました。理想的には0.25になるはずですが、必ずしもそうはなりません。ゆらぎ・不確かさというものを体験してもらいました。参加者には、ISMメモ用紙と国民性トランプが配付されました。

(総務課 庶務係)



●イノベーション・ジャパン2007－大学見本市に出展

平成19年9月12日(水)～14日(金)の3日間、東京国際フォーラムにおいて開催されました、(独)科学技術振興機構(JST)及び新エネルギー・産業技術総合開発機構(NEDO)主催による「イノベーション・ジャパン2007－大学見本市」に、統計数理研究所としては昨年に引き続き出展し、モデリング研究系樋口知之教授による「エキスパートの聴(みみ)、目を創る：高精度スペクトル判別装置及び判別方法」に関する研究紹介をしました。入場者は3日間で4万4千人もの来場者があり、訪れた研究者からは、手法やその応用可能性についての質問などがあり好評でした。

(総務課 庶務係)



●大学院生及び学生のため夏期大学院の講座を開講

統計数理研究所では、統計数理研究所運営会議の外部委員や統計関連学会関係者からのご提案を受けて、平成18年度より全国の大学院生および学生のための夏の学校を開催しております。今年度は下記のような2つの講義を実施しました。

講義1 極値理論

日時 2007年9月12日(水)、13日(木) 10時から16時
場所 統計数理研究所・講堂
講師 高橋倫也(神戸大学)、渋谷政昭(慶應義塾大学名誉教授)、
牧本直樹(筑波大学)、志村隆彰(統計数理研究所)

講義内容

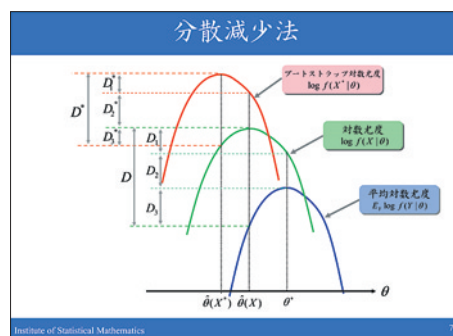
自然災害や経済損失などのリスクを研究するには、極値データの挙動を調べるのが重要です。極値データを用いて統計的推測を行う分野は極値統計学と呼ばれています。今回は、理論面に重点を多くおき極値統計学(極値理論)の理論と応用について講義が行われました。

講義2 情報量規準と統計的モデリング

日時 2007年9月14日(金)、15日(土) 10時から16時
場所 統計数理研究所・講堂
講師 小西貞則(九州大学)、北川源四郎(統計数理研究所)

講義内容

さまざまな現象を記録したデータから有益な情報やパターンを抽出・処理し、現象の解明と新たな知識発見に重要な役割を果たすのが統計的モデリングです。本講義では、モデリングの基本的な考え方、最尤法、情報量規準などの関連手法およびモデリングの実例について説明しました。さらに、現象解明のための線形手法が非線形手法へとどのように拡張されていったかを、モデルの想定、推定そしてモデルの評価という一連のプロセスを通して、最近の成果を踏まえた講義が行われました。



講義2より

極値理論には全国の大学院生等20名、情報量規準には全国の大学院生等25名もの受講者がありました。両講座とも受講した大学院生等は12名でした。来年度以降も実施して行く予定です。

(田村義保)

総合研究大学院大学複合科学研究科統計科学専攻関係

総合研究大学院大学複合科学研究科統計科学専攻入学者選抜試験結果について

【5年一貫制博士課程】

試験年月日	受験者数	合格者数
H19.8.27(月)～ 8.28(火)	平成19年10月入学 1名	0名
	平成20年4月入学(第1回) 0名	0名

【博士後期課程】

試験年月日	受験者数	合格者数
H19.8.28(火)～ 8.29(水)	平成19年10月入学 3名	1名
	平成20年4月入学(第1回) 6名	3名

学生研究発表会

平成19年9月20日(木)に講堂において、統計科学専攻学生による研究発表会が行われました。当日は学生の他、教員等多数の参加者がありました。

専攻修了式

平成19年9月20日(木)に講堂において、統計科学専攻修了式が行われ、石垣 司が本専攻を修了しました。

平成19年度学位記授与式

9月28日(金)に総合研究大学院大学学位記授与式が葉山キャンパスにて挙行され、石垣 司(課程博士)及び福井義成(論文博士)の2名が学位記を授与されました。



【課程博士】

氏 名	論文題目
石垣 司	Diagnostic System for Time Series Data Measured by New Sensing Approach with Statistical Signal Processing

【論文博士】

氏 名	論文題目
福井義成	数値シミュレーションの基礎と応用の研究

平成19年度入学式(10月期)

10月11日(木)に総合研究大学院大学入学式が葉山キャンパスにて挙行され、本専攻の入学者1名を含む33名の新生が迎えられました。

(総務課 研究協力係)

お知らせ

●統計数理セミナー

(平成19年12月～平成20年3月)

毎週水曜日、午後1時30分から所内研究教育職員及び外部の方による「統計数理セミナー」を開催します。多くの方々にご参加いただき活発な討論が展開されることを期待しています。

12/5 地震クラス特性の確率的再構築および点過程の二階残差分析
庄 建倉

12/12 外れ値の割合が大きい場合にもバイアスが小さいロバスト推定
藤澤 洋徳

12/19 メタアナリシスにおける統計的諸問題について
逸見 昌之

1/9 ロガーデータに基づく鳥類行動パターン分析ー深度、加速度、GPSー
島谷健一郎

1/16 コウホート分析における2つの識別問題について
中村 隆

1/23 カイ2乗確率場に対するチューブ法
栗木 哲

1/30 金融時系列における非対称性
佐藤 整尚

2/6 大規模コホート研究の実施：日本ナースヘルス研究
藤田 利治

2/13 標本調査の実際ー各国の事情ー
松本 涉

2/20 パターンの待ち時間分布に関する研究
平野 勝臣

2/27 線形回帰問題のベイズ縮小予測
小林 景

3/5 データ・解析・グラフ表現
馬場 康維

開場：13時

場所：統計数理研究所研修室(新館2階)

2月20日及び3月5日は統計数理研究所講堂(本館2階)にて行います。

時間：13時30分～14時30分

(事前予約不要、入場自由)

(教育情報室)

●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。

(N)共分散構造分析：回帰分析から因果分析へ

日時：2月4日(月)～5日(火)

10時～16時(10時間)

講師：椿広計(筑波大学/統計数理研究所)

申込受付：1月7日(月)10時～1月18日(金)17時

受講料(税込)：5,000円(学生2,000円)

定員：40名(先着順)

講義レベル：初級

回帰分析に関する基礎知識を前提として、パス解析、検証的因子分析、共分散構造分析などを計算機の支援の下に行う方法を習得する。講義としては、なぜ因果分析的な考え方が必要なのかを変数誤差と効果の分解という2つの観点で述べる。また、統計モデルにおける構造モデルと測定モデルの役割についても触れた上で、それらを統合した共分散構造モデリングの意義を確認する。実際の例題は、AMOSやEQSのようなパス図に基づくモデリングが可能な商用ツールの出力を用いるが、演習は何らかのフリーソフトを用いて行う予定である。

(P)計数データに対する回帰モデルとその拡張

日時：2月15日(金)

10時～17時(6時間)

講師：南美穂子(統計数理研究所)

申込受付：1月15日(火)10時～1月25日(金)17時

受講料(税込)：3,000円(学生1,200円)

定員：30名(先着順)

講義レベル：中級

.....
計数データに対する回帰モデルとして代表的なものに誤差にポアソン分布を仮定するポアソン回帰モデルがあります。ポアソン分布はいくつかの仮定の下に自然に導かれるものですが、実際には、様々な原因によりその仮定が満たされず、データがポアソンモデルから乖離していることがよくあります。本講座では、まずポアソン回帰モデルおよび一般化線型モデルについて紹介し、overdispersion の問題、負の2項回帰モデル、zero-inflated モデル、zero-truncated モデル、検定とモデル選択、生物学・医学・社会学・経済学などにおける応用例などを扱います。さらに時間が許す範囲で経時測定データや個体差をモデルに取り入れるためのGEE (一般化推定方程式) やGLMM (一般化線型混合モデル) による方法と応用例を紹介します。

受講対象者のレベル：学部4年生以上で統計学の基礎知識がある方。一般化線型モデルについては講義でも簡単な説明をしますが、事前に知識があった方が内容を理解しやすいと思われます。

詳細は、以下の web サイトをご覧ください。

<http://www.ism.ac.jp/> (教育情報室)

● ISM オープンフォーラム

本研究所では、原則毎月最終金曜日の夜1～2時間程度、専門的業務に従事されている社会人及び研究者の方を対象に、ISM オープンフォーラムを開催しております。このフォーラムでは、3回を一つのシリーズとして構成し、本研究所の教育研究職員を主とした国内の第一線級の講師が、先端的統計科学の応用成果を具体的に解説します。

ー第14シリーズー

「数理ファイナンスと統計科学」

総合コーディネータ：

佐藤整尚(統計数理研究所・准教授)

金利モデルなどを題材にしながら、数理ファイナンスと統計科学の融合について考えてみることにします。

.....
12月14日(金)18:00～19:30

「数理ファイナンスの視点から」

講師：高橋明彦(東京大学)

「統計科学の視点から」

講師：佐藤整尚(統計数理研究所)

.....
(樋口知之)

● 「統計数理」 特集論文募集について

統計数理研究所の和文誌「統計数理」の第56巻第2号(2008年12月発行)に、「データマイニングと統計数理」と題する特集を企画しています。この特集への論文を以下の要領で公募致します。執筆要項については<http://www.ism.ac.jp/editsec/toukei.html>をご覧ください。

特集のオーガナイザー

樋口知之(統計数理研究所)

特集の主旨

近年、社会の情報化が進み、我々は多種多様な情報を計算機ネットワークを通じて入手し利用できるようになりました。一方で、生成・収集・蓄積されるデータの量は急激に増加しており、現在では蓄積された膨大なデータをいかに分析し活用するかが課題となっています。その有力な解決策の1つとして、大量のデータに埋没した有益な知識を発掘(マイニング)するデータマイニング技術が注目されています。この分野の広がりを受け、機械学習の研究も相乗効果的に盛んとなり、また統計科学の有効性と先駆性も昨今あらためて近隣分野から再評価されています。

本特集号では、データマイニングや統計数理に基づく解析を目的とした、理論・手法・技術に関する基礎から応用までを対象とした論文を広く募集いたします。

多くの方のご投稿を期待しております。

対象分野

統計的モデリング／ベイズモデリング／機械学習／確率計算／統計数理／データベース／半構造データマイニング／テキストマイニング／データ可視化インターフェイス／データマイニング固有技術／メタウェア／その他データマイニングと統計数理解析に資する研究

投稿先

〒106-8569 東京都港区南麻布4-6-7

情報・システム研究機構

統計数理研究所 メディア情報室

締め切り

2007年12月31日

投稿の際には、特集「データマイニングと統計数理」に対する投稿である旨と、論文の種別を明記してください。

問合せ先

「統計数理」 特集「データマイニングと統計数理」

編集委員 樋口知之

e-mail: higuchi@ism.ac.jp

なお、特集以外の原稿も随時受け付けております。いずれの原稿も「統計数理」投稿規定(上記 URL からご覧になれます)に従ってご執筆下さい。

(メディア情報室)

共同利用

●平成19年度共同利用公募追加採択課題

共同利用登録（1件）

分野	研究課題名	利用登録者(所属)
g5	極値統計学の理論と応用	渋谷政昭(慶應義塾大学理工学部)

(総務課 研究協力係)

外部資金・研究員等の受入れ

●共同研究の受入れ

受入年月日	委託者の名称	研究題目	研究期間	研究経費 (千円)	研究代表者
H19.8.1	味の素(株)ライフサイエンス研究所所長 尾道一哉	血中アミノ酸プロファイルから生体状態を統計的に推定する技術に関する研究	H19.8.1 ～ H20.3.31	420	データ科学研究系 馬場康維 教授
H19.10.1	(株)博報堂 (株)博報堂 DY メディア パートナーズ	タグチメソッドを主とした統計的手法によるマーケティングへの応用研究	H19.10.1 ～ H20.3.31	1,000	データ科学研究系 河村敏彦 助手

(総務課 研究協力係)

●受託研究の受入れ

受入年月日	委託者の名称	研究題目	研究期間	研究経費 (千円)	受入担当研究 教育職員
H19.4.1	文部科学省研究振興局 振興局長 徳永 保	次世代生命体シミュレーションソフトウェアの研究開発	H19.4.1 ～ H20.3.31	9,961	モデリング研究系 樋口知之 教授

(総務課 研究協力係)

●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
Ilia Negri	ベルガモ大学管理情報工学科・ 准教授	拡散過程の統計的推測	H19.12.1～H19.12.28	西山陽一 助教
石垣 司	独立行政法人 科学技術振興 機構・博士研究員	データ同化論	H19.10.1～H20.3.31	樋口知之 教授

(総務課 研究協力係)

●寄附金の受入れ

受入決定 年月日	寄附者の名称	寄附金額 (千円)	担当教員	寄附目的
H19.9.5	財団法人 流通経済研究所 理事長 宮下正房	200	データ科学研究系 中村 隆 教授	統計数理研究所データ科学研究系における 研究助成として

(会計課 総務係)

人 事

平成19年10月1日転入者(事務)

異動内容	氏 名	新職名等	旧職名等
採 用	中村 透	管理部総務課専門職員(研究協力担当)	東京大学地震研究所係長

平成19年10月1日所内異動(事務)

異動内容	氏 名	新職名等	旧職名等
配置換	須藤文雄	管理部総務課専門職員(庶務担当)	管理部総務課庶務係長

平成19年10月1日兼務(事務)

異動内容	氏 名	兼務先	本 務
兼 務	須藤文雄	管理部総務課庶務係長	管理部総務課専門職員
兼 務	中村 透	管理部総務課研究協力係長	管理部総務課専門職員

外国人研究員(客員)

氏 名	現 職	国 籍	所 属	職名	研究課題	期 間	受入教員
ジャンフィリップ ヴェール Jean-Philippe Vert	Director, Center for Computational Biology (Ecole des Mines de Paris)	フランス 共和国	モデリング研究系 知的情報モデリング グループ	客員 教授	大規模マルチモーダルデー タのためのカーネルマシン に関する研究	H19.9.28 ～ H19.10.25	松井知子 准教授
レオニード フェイボソビッチ Leonid Faybusovich	ノートルダム大学 数学科教授	アメリカ 合衆国	数理・推論研究系 計算数理グループ	客員 教授	種々の凸錐に対する自己整 合障壁関数の計算法とその 応用	H19.10.1 ～ H19.11.30	土谷 隆 教授
陳 素雲 チェン スーユン Su-Yun Chen	Institute of Statistical Science, Academia Sinica 主任研究員	中国 (台湾)	予測発見戦略研究セ ンター遺伝子多様性 解析グループ	客員 教授	統計的カーネル法とバイオ インフォマティクス	H19.10.16 ～ H19.11.15	江口真透 教授

(総務課 人事係)

会議開催状況

●運営会議の開催

平成19年度第2回統計数理研究所運営会議が、10月30日(火)に開催され、統計数理研究所の現況及び情報・システム研究機構関連事項についての報告、規則等の制定、研究教育職員の再任、新機軸創発センターの設置、平成20年度共同利用公募案内、平成21年度概算要求について、それぞれ審議が行われ了承されました。

(総務課 庶務係)

●共同利用委員会の開催

平成19年度第2回統計数理研究所共同利用委員会が10月19日(金)に開催され、平成20年度共同利用公募案内の内容及びその審査方法について審議が行われました。

(総務課 研究協力係)

所外誌掲載論文等

本研究所の教員、研究員、総研大(統計科学専攻)大学院生によって発表された論文等を前号に引き続き紹介します。

Ueno, G., Higuchi, T., Ohtani, S. and Newell, P. T., Particle precipitation characteristics in the dayside four-sheet field-aligned current structure, *Journal of Geophysical Research*, Vol.112, A06242, doi:10.1029/2006JA012036, 2007

Nakano, S., Ueno, G. and Higuchi, T., Merging particle filter for sequential data assimilation, *Nonlinear Processes in Geophysics*, Vol.14, 395-408, 2007

北川源四郎, 予測と発見を目指す統計科学, 日本統計学会誌, 第37巻 J-1, 25-36, 2007

Henmi, M., Copas, J. B. and Eguchi, S., Confidence intervals and P-values for meta-analysis with publication bias, *Biometrics*, Vol.63, Issue 2, 475-482, #DOI 10.1111/j.1541-0420.2006.00705.x, 2007.6

江口 真透, 心にのこる1冊:「*Differential Geometrical Methods in Statistics* (統計学の微分幾何の方法)」

Amari, Shun-Ichi (甘利 俊一)著, 科学, 77巻6号, 642-643, 岩波書店, 2007.6

Tanaka, U., Ogata, Y. and Stoyan, D.: Parameter estimation and model selection for Neyman-Scott point processes,

Biometrical Journal, (Early View), doi:10.1002/bimj.200610339, Published Online: 2007.7

尾形 良彦：千島列島沖の地震活動の時空間パターンとストレス変化, 地震予知連絡会会報, Vol.78, 60-66, 2007.8

尾形 良彦：能登半島周辺の地震活動と地殻変動について, 地震予知連絡会会報, Vol.78, 399-404, 2007.8

Kanamori, T., Takenouchi, T., Eguchi, S. and Murata, N., Robust loss functions for boosting, *Neural Computation*, Vol.19, No.8, 2183-2244, #DOI 10.1162/neco.2007.19.8.2183, 2007.8

Fujisawa, H., Isomura, M., Eguchi, S., Ushijima, M., Miyata, S., Miki, Y. and Matsuura, M., Identifying haplotype block structure by using ancestor-derived model, *Journal of Human Genetics*, Vol.52, No.9, 738-746, 2007.9

江口 真透, リスク科学のための人工知能的方法と統計的思考, 人工知能学会誌, 22巻5号, 656-662, 2007.9

Someya, H., Theoretical analysis on an inversion phenomenon of convergence velocity in a real-coded GA, *Proceedings of IEEE Congress on Evolutionary Computation 2007 (IEEE CEC 2007)*, 4531-4537, 2007.9

津本 周作, 鷲尾 隆, 宮本 定明, 椿 広計, 北川源四郎, 特集「定量的リスク科学」にあたって, 人工知能学会誌, 第22巻, 第5号, 630-632, 2007.9

Ogata, Y.: Seismicity and geodetic anomalies in a wide area preceding the Niigata-Ken-Chuetsu earthquake of 23 October 2004, central Japan, *Journal of Geophysical Research*, Vol.112, B10301, doi:10.1029/2006JB004697, 2007.10

吉野 諒三：統計数理研究所の「国民性の調査」, 心理学ワールド, 39号, p.38, 日本心理学会, 2007.10

吉野 諒三：海外の標本抽出面接調査の方法, いんぷおるむ (第53回), 新情報, Vol.95, 7-12, 新情報センター, 2007.10

Minami, M.: Multivariate inverse gaussian distribution as a limit of multivariate waiting time distributions, *Journal of Statistical Planning and Inference*, 137(11), 3626-3633, #DOI 10.1016/j.jspi.2007.03.038, 2007.11

(教育情報室)

刊行物

● Research Memorandum (2007.7~10)

No.1039 : Yamashita, T. and Ozaki, T., Empirical analysis for estimating timevarying volatility with mixture system noise model.

No.1040 : Yamashita, T., Dynamic modeling and estimation for analyzing investment style of Japanese equity mutual funds considering individual stock selection ability.

No.1041 : Tsuchiya, Takahiro, Ono, S. and Hirai, Y., Response times in the item count technique.

No.1042 : Takeuchi, K., Kumon, M. and Takemura, A., A new formulation of asset trading games in continuous time with essential forcing of variation exponent.

No.1043 : 栗木 哲, 竹村 彰通, チューブの体積と正規確率場の最大値の分布.

No.1044 : 栗木 哲, QTL 解析における検定の多重性調整—ロードスコアの最大値分布の近似.

No.1045 : Ikeda, S. and Sakaguchi, Y., Motor planning as an optimization of command representation.

No.1046 : Shimizu, S., Hoyer, P. and Hyvärinen, A.: Learning linear non-gaussian acyclic models for latent factors.

No.1047 : Inoue, K. and Aki, S., Bivariate Fibonacci polynomials of order k with statistical applications.

(教育情報室)

● 統計数理研究所研究レポート

No.96 : 土屋 隆裕, 前田 忠彦, 中村 隆, 坂元 慶行, エリアサンプリングによるポスティング調査—郵送調査との比較実験—(2007.8)

(メディア情報室)

● 統計計算技術報告

RSC-037 : 田中さえ子, 蛭田 智則, 桂 康一(編), 平成18年度スーパーコンピュータ利用成果報告書 (2007.9)

RSC-038 : 田中さえ子, 松野 秀夫, ISMLIB のアクセス集計—平成18年度版—(2007.9)

(計算資源室)

● **Annals of the Institute of Statistical Mathematics**
Volume 59, Number 3 (September 2007)

Nunzio Cappuccio and Diego Lubian

Asymptotic null distributions of stationarity and nonstationarity tests under local-to-finite variance errors403

Felix Abramovich, Claudia Angelini and Daniela De Canditiis

Pointwise optimality of Bayesian wavelet estimators425

D. Fourdrinier and S. Pergamenschikov

Improved model selection method for a regression function with dependent noise435

R. Kluszczyński, M. N. M. van Lieshout and T. Schreiber

Image segmentation by polygonal Markov fields465

Fabrizio Durante, José Juan Quesada-Molina and Carlo Sempi

A generalization of the Archimedean class of bivariate copulas487

T. Royen

Integral representations and approximations for multivariate gamma distributions499

Arturo A. Z. Zavala, Heleno Bolfarine and Mário de Castro

Consistent estimation and testing in heteroscedastic polynomial errors-in-variables models515

Solomon W. Harrar and Arjun K. Gupta

Asymptotic expansion for the null distribution of the F -statistic in one-way ANOVA under non-normality531

P. Vellaisamy and V. Vijay

Some collapsibility results for n -dimensional contingency tables557

Kiyoshi Inoue and Sigeo Aki

Joint distributions of numbers of runs of specified lengths in a sequence of Markov dependent multistate trials577

Sigeo Aki and Katuomi Hirano

On the waiting time for the first success run597

Cibele Queiroz da-Silva

Asymptotics for a population size estimator of a partially uncachable population603

(メディア情報室)



●統計数理研究所の大学院教育

吉田 亮（モデリング研究室）

総合研究大学院の学生として統計数理研究所の門をくぐった2001年の春が統計学者としての私の原点である。数日前にようやく東京に引っ越してきたばかりで、荷物の整理もままならない中での慌ただしい第一歩だった。山手線の雑多な風景と広尾駅を降りて目に映る景色のコントラストが今でも鮮明に残っている。当時広尾駅近くの地下にあった薄暗い喫茶店に立ち寄って、コーヒーで気持ちを落ち着かせていた。今思えばコーヒー中毒者としての原点でもあった。

あれから6年と8ヶ月、喫茶店業界におよそ百万円つぎこんだ私は、統計数理研究所の研究者になり、研究室のパソコンの画面には膨大なゲノムデータが映し出されている。2001年の春は、経済学の研究者になる予定だった。学会で講演していると、たまにふと思うことがある。なぜここで遺伝子の転写について語っているのだろうか？なぜ生命科学に惹かれたのだろうか？

統計数理研究所という研究機関は実に稀有な存在である。所内にいる多くの研究者は、データ科学とは別に各々の学問領域を持っている。生命科学、数学、工学、物理学、社会科学、実に多種多様である。6人の同級生の専門分野は、数学、機械学習、脳科学、経済学、鉄道工学、商学だった。この「アカデミックカオス」の中で、人々は共通言語の統計科学を駆使しながら、日々知的創造の営みを積み重ねている。2001年春の学問的アイデンティティが実に希薄だった学生はカオスの渦中に飛び込んだ。

研究者にとって最たる喜びは、生涯を賭して挑戦する研究テーマに出会うことである。多くの学生、若手研究者が、ここ広尾で、自分の「研究」を見つけ出していく様を目の当たりにしてきた。私はコーヒー代につぎこんだ約百万円を対価にそれを得た。今は皆、それぞれの分野で元気一杯に研究に動んでいる。研究テーマ選定の指針は、

所内の研究者が織り成す世界のアカデミックマップだった。

私にとってゲノムデータ研究のスタートは実に遅く、博士後期過程3年生の頃である。指導教官だった樋口知之教授の部屋で、当時東京大学医科学研究所の助手だった井元晴哉先生を交えて、遺伝子発現データの解析結果について議論していたことを覚えている。当時ゲノムの「ゲ」もよく分かっていない自分であったが、なぜか妙にエキサイティングな時間だった。そのときもデータ科学という共通言語で話していた。

学位取得後は、東京大学医科学研究所へと研究の場を移し、約2年間のキャリアを経て、今年度から統計数理研究所に戻ってきた。この期間中、私が学生時代に受けた教育が何だったのかということを感じることになる。それが何かについて一言で説明できる言葉はない。研究分野の特性上、生物や医学の研究者と議論することが多いが、話が専門的になり過ぎると結構分からないことがある。前日あまり寝ていないときなど、かなりつらい。ところが、議論の途中に実験データが登場すると大抵様相が一変する。急に見通しがよくなって、相手が何をしたいのか、どんなデータ解析をすればゴールに辿り着けるのか、アイデアが次から次へと溢れてくる。眠気は吹き飛び、普段以上に元気になる。統計数理研究所の某先生のようにある。これが教育だと思った。

つい先日、総合研究大学院の学生研究発表会に企画委員のひとりとして参加する機会があった。そこにはまさに「アカデミックカオス」が存在した。線虫の行動解析、企業の信用リスクの格付け、薬剤血中濃度分布のモデリング、音声認識と流れる研究群は、まさに混沌であるが、実に心地良い空気が漂っていた。そこはデータ科学という秩序によって支配されていた。データ科学とは実に稀有な学問領域である。

あとがき 散策手帖 秋の訪れ

町田市郊外の里山散策に出かけました。稲刈りの真最中で、農家の人が忙しく「はさがけ」をしておりました。今年もお米は豊作のようです。谷戸田の一角には、コスモスが咲いており、秋晴れのもと爽やかな一日を過ごせました。この地区は、東京都の歴史的景観保存地区に指定され、徒歩のみでしか行けないのですが、旧鎌倉街道と古代東海道からもほど近く、小田急線唐木田駅から図師町を経由して小野路一本杉公園まで足を延ばすと手ごろなハイキングコースになります。

(文と写真 須藤文雄)



統計数理研究所ニュース No.98

発行／大学共同利用機関法人 情報・システム研究機構

統計数理研究所 広報委員会

問い合わせ先／統計科学技術センター TEL (03) 5421-8722, 8723

発行日：平成19年11月25日

〒106-8569 東京都港区南麻布4-6-7

ホームページ <http://www.ism.ac.jp/>