

平成 14 年度研究報告会要旨

と き : 2003 年 3 月 18 日 午前 9 時 30 分 ~ 午後 5 時 15 分
19 日 午前 9 時 30 分 ~ 午後 5 時 30 分
と ころ : 統計数理研究所 講堂

プ ロ グ ラ ム

3 月 18 日 (火)

あいさつ

所長 北川源四郎

【統計計算開発センター】

救急需要の将来予測
音声認識の課題と統計科学的方法
時変係数 AR モデルについて
統計解析システム Jasp のグラフィック機能の拡張
無限次元線形計画法

田村 義保
松井 知子
荒畑恵美子
中野 純司
伊藤 聡

【予測制御研究系】

測定ノイズロバストなベイジアンネットワークの推定法
イールドカーブの推定法について
分子系統学の問題点
遺伝子コーディング領域の推定
分子系統樹推定における置換モデルのミススペシフィケーションに関する研究
10,000 次元空間カルマンフィルターのパラレル計算法と
その EEG ダイナミック逆問題への応用
不完全情報下における制御系設計に関する研究
CG-Newton Method for Penalized Logistic Regression Machine
逐次モンテカルロの観点からみた Annealed Importance Sampling 法
とその応用
ヌジ人名資料について
PE モデルについて
進化型計算のバイオインフォマティクスへの応用
スペクトル拡散通信におけるマルチユーザ検波方式の研究
無線によるコヒーレント通信の研究

樋口 知之
川崎 能典
長谷川政美
足立 淳
曹 纓
尾崎 統
宮里 義彦
田邊 國士

伊庭 幸人
上田 澄江
佐藤 整尚
染谷 博司
瀧澤 由美
(客員 , 千葉大学) 深沢 敦司

【統計基礎研究系】

待ち時間分布に関する研究 (2)
逐次数値積分を用いた多次元正規分布の確率評価

平野 勝臣
栗木 哲

Functional CLT の応用例について	西山 陽一
不完全 K -型変形ベッセル関数の導入による対流・拡散モデルの拡張	松縄 規
β -ダイバージェンス最小化 ICA 法における β の選択方法	南 美穂子
Classification from DNA Microarray Data	江口 真透
Subexponential Class の拡張と関連する分布族	志村 隆彰
極値分布と極値データ解析	(客員, 神戸商船大学) 高橋 倫也

3 月 19 日 (水)

【領域統計研究系】

教育試験と臨床試験を支える共通要因	柳本 武美
モデルに基づく遺伝子型のクラスタリング	藤澤 洋徳
Model Induced Link Functions in GLM	大西 俊郎
高次元ランダムパッキングの単純化	伊藤 栄明
文化を計る	村上 征勝
東アジア価値観調査 幸福な家庭は皆似ているが, 不幸せな家庭には それぞれの様相がある	吉野 諒三
抽出名簿がない場合の社会調査とそのバイアス問題	鄭 躍軍
スプレッドシート上での統計計算環境の構築	(客員, 立教大学) 山口 和範
大規模臨床疫学研究の基盤整備	(客員, 東京大学) 大橋 靖雄

【調査実験解析研究系】

「国民性調査に関する若干の検討」補遺	坂元 慶行
貿易自由化の森林資源管理への影響 最適確率制御モデルの応用	吉本 敦
社会調査データの解析における潜在変数モデルの活用	前田 忠彦
Voronoi-Delaunay Cell の統計分布とその周辺	種村 正美
公開の抽籤の公平性, 意外性, 継続性についての考察	丸山 直昌
分子生態学における点過程モデル	島谷健一郎
コウホート分析にもとづく予測	中村 隆
インターネット調査 実験調査から見てきたこと	大隅 昇
電話による実験調査	土屋 隆裕
WWW 上の統計科学情報について	清水 信夫
時空間統計モデルによる地震活動研究の新展開	尾形 良彦
環境データの解析	柏木 宣久
確率伝搬法の解析と改善手法	池田 思朗

【統計科学情報センター】

じゃんけんソフトのその後	石黒真木夫
重み付けをした変数による主成分分析	馬場 康維
水環境データの解析手法について	金藤 浩司
外乱に対する多変数弱非線形の応答	岡崎 卓
会計学と統計学	山下 智志
Molecular Evolution and Phylogenetics of Mangroves	(客員, Fudan Univ.) 鐘 揚

救急需要の将来予測

田 村 義 保

平成 9 年 1 月から平成 13 年 12 月までの東京都における救急出動件数データを用いて、救急需要の将来予測を行った。救急需要が上記期間で急増しているが、その要因を調べ、要因の将来予測値から救急出動件数を予測することが研究の目的であった。

東京都の総人口は増加しており、救急出動件数と高い相関がある。しかし、年齢層に分けると、14 歳以下及び 15 歳以上 64 歳以下の人口は減少している(区部の 15 歳以上 64 歳以下を除く)にも関わらず、救急出動件数は増加している。このことから、人口と救急出動件数とは関係ないものと考えられる。単身世帯、特に 65 歳以上の単身世帯が増えていることは救急出動件数の増加に影響を与えていると思われるが、数量化できるほどのデータはない。人口動態統計の死亡数や、在院患者数、外来患者数の変化を見ても、疾病構造が変化しているとは考えることはできない。このために、急病による出動件数の増加は有病者が増えたためではなく、その他の要因に関係しているものと考えられる。65 歳以上を除けば交通事故による出動は増えておらず、減少している期間もある。64 歳以下の交通事故がそれほど増えていないのは、交通事故を起こすあるいは巻き込まれるような機会がほとんど変化していないためであると考えられる。介護保険の導入により、病院や介護施設の役割は変わったであろうが、救急出動に対する影響は見ることができなかった。その他の福祉に関する事業についても救急出動との関係は見られなかった。景気に関連する指数や失業率などは、不景気であることを裏付けるものばかりである。不景気であり、交通量などが増えないことは、上述のように交通事故が増えないことと関係しているが、直接的に救急出動件数との関係を見ることはできない。

このように救急需要の急増をそれだけで説明できるものは存在しない。お金を使うことを極端に避けるという風潮や、安易に救急車を呼ぶという風潮が高まっているためと思われるが、このような質的な要因は数量化することができない。このように要因は不明であるが、増え続けている人口 1 万人あたりの出動要請件数の予測値と厚生労働省の人口予測値を用いて、平成 17 年、22 年、27 年の救急出動件数を予測した。

音声認識の課題と統計科学的方法

松 井 知 子

本研究では、従来の音声・話者認識技術における音響的な課題を解決するための統計手法について考察した。

音声波には音韻、話者、韻律の三つの情報が含まれている。音声認識では主に音韻情報を、話者認識では話者情報を利用して、それぞれ発声内容、発声者を自動判定する。ここ数十年、音声・話者認識の問題は統計的な枠組みで考えるのが常套手段となっている。音声認識は入力音声から抽出した特徴量時系列 Y 、単語列 W について、 $P(W|Y)$ の条件付確率を最大にする単語列 W を見つける問題として定式化する。ベイズの定理により、 $P(W|Y)$ は $P(Y|W)$ と $P(W)$ の積を最大にする単語列 W を見つける問題に置き換えられ、 $P(Y|W)$ は単語もしくは音素単位に学習した HMM の音響モデルから、 $P(W)$ はマルコフモデルで表した言語モデルから計算する。一方、話者認識では単語列 W の代わりに話者 S を考え、 $P(S|Y)$ の条件付確率に基づいて話者判定を行う。 $P(S|Y)$ は同じくベイズの定理により、 $P(Y|S)$ と $P(S)$ の積で表され、 $P(S)$ は一定として無視し、 $P(Y|S)$ は話者単位に学習した HMM の音響モデルから計算

する．なお，特徴量時系列 Y としては音声・話者認識ともに，パワースペクトルの対数の逆フーリエ変換であるケプストラム(10～20次元)が利用されることが多い．また，HMM 計算には EM アルゴリズムが一般的に用いられる．

音声・話者認識の音響的な課題，つまり音響 HMM に関する課題としては，雑音，話者の違い，話し方(発声速度や声の大きさ，体調や気分にも依存)の違い，言い誤り，言い怠けや淀みが挙げられる．従来の HMM は学習データに含まれているような発声の揺らぎには対応できるが，上記の課題に関連した予期せぬ要因による揺らぎに対応することは難しい．それらの揺らぎの問題に対処するためには，音響 HMM の状態空間を拡張してそれらを取り込めるようにしたり，状態パラメータ自体がそれらに適應する自己組織型のモデルが必要になると考える．また，それらのモデルでは探索空間が広がるので，逐次モンテカルロ法などを適用する必要があると考える．

今後は従来の HMM を超える可能性のある新しいモデルの模索を含め，上記のモデルについて検討を続けていく．

時変係数 AR モデルについて

荒畑 恵美子

以前，赤池弘次先生や北川源四郎所長がされた時変係数 AR モデルについての紹介をした．非定常時系列 $Y(n)$ に対し，時変係数自己回帰モデルを考えることにより，時変スペクトルを求めることが出来る．それにより，より詳しい解析を行なうことができる．

参 考 文 献

- 北川源四郎(1986) 時変係数自己回帰モデル プログラム TVCAR の紹介 ，統計数理，34，273-283.
北川源四郎(1993) 『時系列解析プログラミング』，岩波書店，東京．

統計解析システム Jasp のグラフィック機能の拡張

中野 純 司

われわれが作成している統計解析システム Jasp は，Java 言語で書かれており，それを利用して最新の計算機技術を利用できるように設計されている．Jasp では，散布図やヒストグラムのような統計解析のための基本的なグラフィックスは最初から組み込まれている．しかし，より進んだデータビジュアライゼーション機能，例えば散布図行列や平行座標プロット，ダイナミックな 3D グラフィックスによるグラウンドツアーなどは装備されていなかった．近年，これらの機能はビジュアルデータマイニングのためのツールとして必要性が高まっている．韓国で開発されている Davis は，データビジュアライゼーションのために Java 言語で作成されたシステムであり，多くの最新のグラフィックスを実現している．われわれは Jasp の Java プログラムに対する高い拡張性を利用して，Davis のプログラムを変更することなく，Davis のほとんどの機能を Jasp の機能として利用できるようにした．Jasp はクライアント/サーバ構造の分散型統計解析システムであるので，Davis はクライアント側の拡張機能とすることにし，同時に，

分散型統計解析システムにおけるクライアント側の一般的な拡張法についても考察した。

無限次元線形計画法

伊藤 聡

X, Y を有限次元ユークリッド空間のコンパクト集合, $f \in C(X), g \in C(Y), \varphi \in C(X \times Y)$ を与えられた関数とし, 線形計画問題を

$$P: \begin{cases} \inf_{h \in L_1(X)} \int_X f(x) h(x) dx \\ \text{subject to} \int_X \varphi(x, y) h(x) dx \geq g(y) \quad \forall y \in Y \\ h(x) \geq 0 \quad \text{a.e. on } X \end{cases}$$

を切除平面法を用いて解くことを考える。ここで, $M(X)$ を X 上の符号つき正則ボレル測度の空間とし, P に関連した線形計画問題を

$$P': \begin{cases} \min_{\mu \in M(X)} \int_X f(x) d\mu(x) \\ \text{subject to} \int_X \varphi(x, y) d\mu(x) \geq g(y) \quad \forall y \in Y \\ \mu \geq 0 \end{cases}$$

において, 不等式制約条件を規定するインデックス集合 Y をその有限部分集合 $Y_k := \{y_1^k, y_2^k, \dots, y_{n_k}^k\}$ に緩和して得られる半無限計画問題を $SIP(Y_k)$ と書くことにすると, 最も基本的な切除平面法のアルゴリズムは以下ようになる。

- step 1. ε を微小な正数, $Y_1 := \{y_1^1, y_2^1, \dots, y_{n_1}^1\}$ を Y の有限部分集合とし, $k := 1$ とする。
- step 2. 半無限計画法を用いて問題 $SIP(Y_k)$ の最適解 $\mu^k \in M(X)$ を求める。
- step 3. $\delta(\mu^k) := \min_{y \in Y} \{ \int_X \varphi(x, y) d\mu^k(x) - g(y) \}$ を計算し, その最小値を与える解 $y_{n_k+1}^k$ を求める。
- step 4. $\delta(\mu^k) > -\varepsilon$ ならば終了。
- step 5. $Y_{k+1} := Y_k \cup \{y_{n_k+1}^k\} = \{y_1^{k+1}, y_2^{k+1}, \dots, y_{n_{k+1}}^{k+1}\}$, $n_{k+1} := n_k + 1$, $k := k + 1$ として step 2 に戻る。

このアルゴリズムの収束に関して, 適当な条件のもとで, step 2 で生成される点列 μ^k が問題 P' の最適解 μ^* に収束すること, $\delta(\mu^k)$ が $\delta(\mu^*)$ に収束すること, また問題 P と補助問題 $SIP(Y_k)$ の最適目的関数値の差が $\delta(\mu^k)$ を用いて評価できることを示し, これらと同様の結果がそれぞれの双対問題に対しても成り立つことを示した。本研究は台湾・国立成功大学数学科の呉順益教授および呉自力博士との共同研究である。

測定ノイズロバストなベイジアンネットワークの推定法

樋口知之

マイクロアレイ遺伝子発現データから遺伝子の相互関係をモデル化する作業、つまり遺伝子ネットワークのグラフィカルモデル構成において、ベイジアンネットワーク(以後 BN と略す)は有用な表現方法の一つである。各遺伝子発現量を確率変数として取り扱った時、BN の持つ最大の特性である連鎖法則は、超多数の確率変数の同時分布を条件付き確率分布の積へ分解することを可能にする。この分解により、ある一つの子遺伝子が(複数の)親遺伝子にどのように依存しているのかを特徴づける、子遺伝子の条件付分布に主に我々は注目すればいい。一般にこの依存性は非線形の振舞いを示す。また遺伝子発現データが、実験による様々な人工ノイズの異常値を含むことは有名である。これらの 2 つの問題—非線形性と異常値処理—に適切に対応できる、条件付分布の強健で信頼できる推定法が、BN を用いた遺伝子ネットワーク同定には鍵である。

これらの問題に対処するために Imoto らは、データの heteroscedasticity を考慮に入れた、スプライン回帰をベースとしたノンパラメトリック回帰モデルによって条件付分布を表現する方法を提案した。そこでは、heteroscedasticity を考慮しながら非線形な関係を促えることを目指しているが、遺伝子発現データの heteroscedasticity の一般的な特徴を表わすことにはまだ不十分である。というのも heteroscedasticity を、親遺伝子観測値から構成されたベクトルデータの、その平均からの距離の関数で表現しているからである。本研究の主たる動機の一つは、遺伝子発現データから heteroscedasticity も含めた種々様々の異常値を除外できる統計方法論を開発することである。我々はこのために、moving(boxcel)median filter(以後 MMBF)を利用した、モデルを仮定しない統計手法を提案した。MMBF は異常値に対して強健である上、もちろんそれはノンパラメトリックアプローチであるので、非線形な依存性を促える能力を保持する。

残念ながら、microarray 遺伝子発現データにこの方法を適用するために実際的な問題が残る。MMBF によって安定した統計量を得るためには、各固定された領域に十分なサンプル数を要求する。従って、1 セットの発現データから作る多数の仮想親遺伝子サンプルの生成によりこの問題を緩和した。仮想親遺伝子観測値ベクトルは、各親遺伝子が子遺伝子として取り扱われた条件付分布推定の際に得られた、MMBF 推定値の観測値からの残差をブートストラップすることで作成される。親遺伝子として取り扱われる遺伝子変数の観測値は当然統計的誤差を含んでいるという現実的な予想を、この処理により、仮想的ではあるが実現している。我々は、以前の数値実験の中で検討された人工データの分析を通じて提案された方法の強健さを実証し、実際のデータセットに適用することで方法の有効性を示した。

イールドカーブの推定法について

川崎能典

McCulloch(1971)以降、利付債のクロスセクションデータを適当な基底関数の組に回帰させて金利の期間構造を推定するという方法は、学術・実務の両面において広く定着してきた。しかし、推定されたイールドカーブ(特にフォワード・レート)の安定性や、基底関数の個数の決定などについては今なお問題が残されており、これらに関しては主に基底関数の選択や節点の置き方などを中心に、多くの研究・提案が行われてきている。本研究では、推定曲線の不安定さは、基底関数や節点の位置の選択以上に推定問題の不適切性に起因すると考え、より一般的

な解決策を目指して正則化法によるモデリングを行った。すなわち、従来の単純な最小二乗法ベースの手法に代わって、罰則付き最尤法の利用を提案した。また、割引関数とフォワード・レートと同時に安定的に推定することを可能にするために、新たにガウス基底関数 (Gaussian Radial Basis Function) に基づく統計モデルを導入した。ガウス基底関数の広がりや節点に関して共通に制御するパラメータを導入することで、割引関数のような単調性を持つ関数の推定に関する安定度が向上する。モデル構築に当たっては、正則化 (平滑化) パラメータ、基底関数の個数、基底関数の広がりやを表すパラメータの選択が本質的となるが、本研究では一般化情報量規準 GIC の枠組みに基づいて選択規準を与えた。最後に日本国債のクロスセクションデータを用いて、提案した手法を実際に適用すると、フォワード・レートが安定的に推定されることがブートストラップ法で確認された。

参 考 文 献

- 川崎能典, 安道知寛 (2002) 正則化非線形回帰モデルによるイールドカーブの推定, 統計数理, 50(2), 149-164.
- McCulloch, J. H. (1971) Measuring the term structure of interest rates, *Journal of Business*, 44(1), 19-31.

遺伝子コーディング領域の推定

足 立 淳

ゲノムや cDNA の塩基配列から、未知のタンパクをコードする遺伝子領域を推定する方法は大きく二つに分けられる。一つは既知の遺伝子と比較し、相同性が高い領域を見つける方法。もう一つは、既知の遺伝子の特徴を抽出した統計的なモデルを使い、未知の領域を評価する方法である。後者の代表的なモデルとして coding potential がある。これまで coding potential として hexa-nucleotide モデルという連続する 6 塩基の使用頻度情報が使われてきた。タンパクの構成要素である 20 種類のアミノ酸は、それぞれコドンと呼ばれる 3 塩基の並びから翻訳されるので、6 塩基の使用頻度は隣り合わせるコドンのペアの使用頻度を表している。この情報は、同じアミノ酸に翻訳される複数のコドンの中でどのコドンがどれだけ使われるかという「冗長なコドンの使用頻度」と、どんなアミノ酸が隣り合わせになりやすいかという「アミノ酸ペアの使用頻度」が組み合わさったものである。

しかし、アミノ酸のペアの組み合わせは 400 通りしかないのに対して、コドンのペアの組み合わせは 3721 通りもあり、とても冗長である。また、冗長なコドンの頻度は、ゲノム上の塩基配列の偏りの影響を強く受けるのに対し、アミノ酸のペアの頻度はその偏りの影響をそれほど受けない。むしろアミノ酸のペアの頻度は、その遺伝子がコードしているタンパクの種類に大きな影響を受ける。よって、この二つの頻度情報は独立に扱った方がよいことが期待される。そこで新しい coding potential として、塩基組成の偏りを補正した冗長なコドンの使用頻度モデルと、タンパクの種類を考慮したアミノ酸のペアの使用頻度モデルを独立に扱うことを提案し、その有効性を示した。

分子系統樹推定における置換モデルの ミススペシフィケーションに関する研究

曹 纓

生物の系統関係を明らかにすることは、生物学の基本であり、そのための方法として分子系統学が近年注目されてきた。そこでは確率モデルに基づいた統計解析がもっとも有力な手法であった。それは進化における DNA の塩基置換は、確率的な過程と見なすべきものだからである。確率モデルに基づく統計解析と言えば塩基置換あるいはアミノ酸置換のモデルの仮定が必要となるが、そのモデルは現実的なものでなければ、偏った系統樹推定が行なわれる可能性がある。データ解析をする際には、DNA の塩基や蛋白質のアミノ酸が置換する確率法則をモデル化しなければならないが、このような素過程に関する私たちの知識は限られているために、現実的なモデル化ができないことが多いのである。現存の分子系統樹推定法には、種のサンプリング依存性、遺伝子サンプリング依存性、置換モデル依存性などの問題点がある。今年度の研究報告会では脊椎動物の分子系統学的研究で見られる、置換モデルのミススペシフィケーションに関わるいくつかの問題点を、事例を通じて紹介した。

不完全情報下における制御系設計に関する研究

宮 里 義 彦

制御のためのモデルの設定と同定から制御手法までを総括的に含む統合化制御系設計理論の構築を考えている。その一環として、モデリングと制御の接点を扱う適応制御の基礎理論の研究や、実用化のための様々な制約を取り除いた適応制御系の設計法、及び関連する非線形制御の研究を行っている。

従来の適応制御理論が漸近安定性に主眼を置くのに対して、制御性能をより定量的に考慮する観点から、適応制御過程を H_2/H_∞ 最適制御問題として定式化することを試みている。これまでにモデル規範形適応制御を含む一般的な形式の適応制御問題について、安定解析に用いるリアプノフ関数の一部と Hamilton-Jacobi-Isaacs 方程式の解を同一視することで、特定の評価関数に対して最適（または準最適）な 3 つの型の適応制御系の構成法を導出した(宮里(2002a))。またシステムに含まれる未知のパラメータを H_∞ 制御問題における未知外乱と見なすことにより、パラメータの任意の変動に対して安定な非線形適応 H_∞ 制御系を構成する手法を求めて(Miyasato(2002a))、それをロボットマニピュレータの軌道追従制御に応用した結果も得ることができた(Miyasato(2002b))。その手法はニューラルネットを制御器に含む非線形適応 H_∞ 制御にも適用されたが(宮里(2002b))、さらにより一般的な非線形パラメータモデルの非線形・適応制御に拡張することも考えている。

これらとは別に LPV (Linear Parameter Varying) システムのゲインスケジューリング制御において、スケジューリングパラメータが未知な場合に、その推定値を用いて適応的にゲインスケジューリング制御系を構成する手法について考察した。その結果、LMI (Linear Matrix Inequality) に基づく LPV システムの適応型ゲインスケジューリング H_∞ 制御系を構成する手法を導出し、参考文献(Miyasato(2002c))にあげるような成果を得た。またそれらと非線形 H_∞ 制御手法を組み合わせたスケジューリングパラメータの時間変動成分への対処法や、非線形パラメトリックモデルを導入した LPV システムの適応型ゲインスケジューリング制御について

も考察し、参考文献(Miyasato(2002d, 2003))にあげるような結果を得た。

さらに反復学習制御を、ハイブリッド適応機構を用いて実現する方法を考案し、ロボットマニピュレータの軌道追従制御に適用して、追従誤差の安定解析と数値計算による検証を行った。ハイブリッド適応機構としてより収束特性の優れた方式や、複数の適応プロセスの導入による制御性能の向上の可能性についても、いくつかの萌芽的な結果が得られた。

参 考 文 献

- 宮里義彦(2002a). 最適性に基づく適応制御系の再設計, 計測自動制御学会論文集, 38(9), 765-774.
- 宮里義彦(2002b). ニューラルネットを用いた非線形適応制御, 計測と制御, 41(12), 881-887.
- Miyasato, Y.(2002a). Adaptive nonlinear H_{∞} control for processes with bounded variations of parameters—General forms and general relative degree case—, *Proceedings of IFAC Workshop on Adaptation and Learning in Control and Signal Processing(ALCOPS2001)*, 419-424.
- Miyasato, Y.(2002b). Nonlinear adaptive H_{∞} control for robotic manipulators, the 15th IFAC World Congress(preprints)
- Miyasato, Y.(2002c). Adaptive gain-scheduled H_{∞} control of linear parameter-varying systems, the 15th IFAC World Congress(preprints)
- Miyasato, Y.(2002d). Adaptive gain-scheduled H_{∞} control of linear parameter-varying systems by adding nonlinear feedback schemes, *Proceedings of the 41st IEEE Conference on Decision and Control*, 3173-3178.
- Miyasato, Y.(2003). Adaptive gain-scheduled H_{∞} control of linear parameter-varying systems with nonlinear components, 2003 American Control Conference(to be presented).

ヌジ人名資料について

上 田 澄 江

ヌジ人名資料(Gelb et al.(1943))とは、B.C.15世紀頃の古代都市ヌジの遺跡から出土した粘土版文書の人名による索引であり、人名にかかわる親族関係、内容を記した文献名、その巻数、行番号などが参照できる。今回、文献 No. の情報に注目することにより資料のより正確な情報を得ることができた。前回に試みた親族情報の抽出による家系図の構成に加え、より正確な家系図を再構成することができるかと期待される。また、文献 No. の記述から複数の家系図との関連をたどることで同世代の人物が特定できる可能性がある。男女の出現数を考慮しながら、家系図からの古代都市の人口の試算、および出生率と世帯数の推定からの人口の試算を試みている。

参 考 文 献

- Gelb, Ignace F., Purves, Pierre M. and Macrae, Allan A.(1943). *Nuzi Personal Names*, The University of Chicago Press, Chicago, Illinois.

PEモデルについて

佐藤 整 尚

これまでの経済分析では、経済理論に基づいた計量モデルを構築し、それらを同時方程式モデルとしてとらえて、構造式を推定するという手法が広く用いられていた。PEモデルでは特定の構造を仮定しないで、データから経済メカニズムや循環を抽出しようという考え方をとった。具体的には、注目する経済変数をいくつか選び、定常化して、多変量自己回帰モデルを当てはめるという手法を用いる。こうして得られたモデル式から相対寄与率や予測値を計算することができ、その内容を分析することで、データのみでは捕らえきれないような構造変化やショックの度合いを見ることが可能となる。もともと、このモデルはマネタリーベースを中心とした金融政策変数を動かすことで、実体経済にどのような影響を及ぼしうるかを評価するために構築されたモデルで、PEはPolicy Evaluationの略である。用いた変数は、マネタリーベース、長期金利、卸売物価、鉱工業生産指数、機械受注、為替相場の6変量であった。われわれは、このモデルにアメリカの変数を加えた11変数モデルを構築し、日米の経済情勢の変化を抽出するモデルとして捉え、特に予測値の変化に注目した。これにより、2001年から2002年の景気後退局面と2003年にかけての景気回復局面を捉えることに成功した。また、局所的な変化として、2001年9月11日のテロの影響や、エンロン問題の波及効果を評価することが可能となった。今後の課題としては、現在、データの公表時期の関係で、2ヶ月遅れの解析しかできない点がある。このため、独自調査によるデータの取得を目指すことを考えており、調査の企画を行なっているところである。また、解析を通じて感じたことは、経済統計はかなりの頻度で改訂やリニューアルが行なわれる点であり、統計の断絶をどのように扱うかも重要なテーマになってくる。今後はさらに解析を進めるとともに、広く社会に対して公開していきたいと考えている。さらに、マクロ経済分野のみならず、もうすこしミクロな領域、たとえば、企業の経営データなどに対しても適用できないかを研究していくつもりである。現在のような、経済情勢が内外とも不安定な状況では、従来型の硬い構造モデルでは限界があり、今回使ったような柔軟な時系列モデルが有効であり、また、利用範囲も広いと考えられる。

進化型計算のバイオインフォマティクスへの応用

染谷 博 司

年度研究報告会では、進化型計算による最適化の話題について紹介し、バイオインフォマティクス分野への応用例を示した。近年、高次元、非線形、多目的、多峰性、といった特徴を有する困難な最適化問題への強力な接近法として進化型計算が注目されている。しかし、進化型計算は設計の自由度が高く、適切に設計されれば高い性能を示すものの、設計に失敗するとランダムサーチ以下の性能しか示すことができないこともある。既に提案済みである適応的な探索を行う進化型計算、Genetic Algorithm with Search Area Adaptation (GSA) の設計指針を紹介した(染谷・山村(1999, 2002))。また、VLSIレイアウト設計問題、関数最適化問題への応用例を示した。進化型計算の交叉による探索にはサンプリング・バイアスと呼ばれる探索の偏りがあり、最適解が探索空間のコナーに位置する問題を不得手とすることが知られている。サンプリング・バイアスを無にする手法、Toroidal Search Space Conversion (TSC) を紹介した(染谷・山村(2001))。

進化型計算のバイオインフォマティクス分野への応用例として、分子計算への応用を示し

た．近年，DNA などの分子に解候補をコード化し最適化計算を行う分子計算が注目されている．年度研究報告会では，進化型計算の分子計算への応用例を紹介した．また，進化型計算のグリッド環境への応用例を示した．グリッドは，インターネット等に接続されている計算機を高速ネットワークでつなぎ並列計算機として利用可能にする技術である．バイオインフォマティクス研究のために構築されているグリッド環境である OBIGrid 上に進化型計算を実装し，その性質を示した(OBIGrid(2002)，染谷(2003))．

参 考 文 献

- OBIGrid(2002). Open Bioinformatics Grid, <http://www.obigrid.org/>
- 染谷博司(2003). グリッドでの最適化について，統計数理研究所研究会「並列計算・乱数・グリッド・HPC」講演資料．
- 染谷博司，山村雅幸(1999). 探索領域を適応的に調整する遺伝的アルゴリズムによるフロアプラン設計問題の一解法，電気学会論文誌 C, 119-C(3), 393-403.
- 染谷博司，山村雅幸(2001). 最適解の位置にロバストな実数値 GA を実現する Toroidal Search Space Conversion の提案，人工知能学会誌, 16(3), 333-343.
- 染谷博司，山村雅幸(2002). 探索オペレータの機能分担を考慮した進化型計算による関数最適化，電気学会論文誌 C, 122-C(3), 363-373.

スペクトル拡散通信におけるマルチユーザ検波方式の研究

瀧澤 由美

DS-CDMA では拡散符号の相互相関が多重接続干渉を引き起こし，システムの容量を制限する原因となっている．本研究ではこの干渉の影響を軽減する新しい検波方式として共役勾配法に基づくマルチユーザ検波方式を提案した．

近年，符号分割多重接続(Code-Division Multiple Access:CDMA)方式が注目を集めている．CDMA は大容量，高品質パーソナル無線通信方式として提案され，2001 年秋から W-CDMA の商用サービスが開始された(Iijima et al. (2003), Arai et al. (2003))．直接拡散符号分割多重接続(Direct Sequence CDMA:DS-CDMA)方式は各ユーザに拡散符号を割り当てて通信を行うもので，時間，周波数ともに重なり合った状態で多重接続を行う．しかしこの方式は，各局に割り当てられた符号間の相互相関値がゼロにはならないため，他のチャネルの信号が多重接続干渉(Multiple Access Interference: MAI)として作用しシステムの加入者容量を制限する．

この多重接続干渉の解法は計算の複雑さから減算型の SIC(Successive Interference Cancellation)や PIC(Parallel Interference Cancellation)といった干渉キャンセラが注目されている．本研究では共役勾配法を利用し，この干渉を信号モデルから直接解法する新しいマルチユーザ検波方式を提案した．DS-CDMA システムにおける受信信号モデルの干渉除去問題の解法は大規模な連立一次方程式に帰着できる．共役勾配法は連立一次方程式の解法として知られており，本研究ではこれを応用して検波器を構成し，コンピュータシミュレーションによりその有効性を示した．

参 考 文 献

- Arai, Yoshikazu, Igarashi, Kiyomi, Fukasawa, Atsushi and Takizawa, Yumi (2003). Multi-user detection to enhance the capacity of W-CDMA based on the conjugate gradient method, *Journal of Circuits, Systems, and Computers (JCSC)*, Manuscript # JCSC-MWN-004.
- EIA/TIA/IS-665 (1995). W-CDMA (Wideband Code Division Multiple Access) air interface compatibility standard for 1.85-1.99 GHz PCS applications, EIA/TIA/95.09.1, Sep. 1995.
- Fukasawa, A., Takizawa, Y., Kawabe, M., Kato, T. and Sato, T. (1995). Configuration and its performance of wideband CDMA system, *IEEE MT & ED France Proceedings (Microwave Theory and Electronic Devices)*, Nov. 1995.
- Fukasawa, A., Takizawa, Y., Kawabe, M., Sato, S., Sugimoto, D. and Sato, T. (1995). Configuration and its performance of an interference cancellation system using pilot signal for Radio Channel estimation, *IEICE, J77-B-11* (11), 628-640.
- Fukasawa, A., Sato, T., Takizawa, Y., Kato, T., Kawabe, M. and Fisher, R. E. (1996). Wideband CDMA System for Personal Radio Communications, *IEEE Communications Magazine, Topics in Personal Communications*, **34**(10), 116-123.
- Fukasawa, A., Ohkawa, S. and Takizawa, Y. (1998). The latest technology of wideband-spread CDMA for mobile radio communications, *Journal on C.C.C.C.*, **XLIX**(9-10), 3-8.
- Fukasawa, A., Iijima, Y., Igarashi, K., Inoue, S. and Takizawa, Y. (2002). Estimation and compensation of phase rotation in radio propagation using pilot channel for radio data transmission, *Asian Simulation Conference, International Conference on System Simulation and Scientific Computing, Shang-hai*, 963-967.
- Igarashi, K., Arai, Y., Sakai, T., Fukasawa, A. and Takizawa, Y. (2002). Multi-user detection for wideband CDMA based on the conjugate gradient method, *WSEAS ICOSMO*, Skiathos, Greece, Sep. 25-28, 447-350.
- Iijima, Y., Inoue, S., Kashima, T., Fukasawa, A. and Takizawa, Y. (2003). Coherent SS demodulation with estimation and compensation of phase rotation using a pilot channel, *Journal of Circuits, Systems, and Computers (JCSC)*, Manuscript # JCSC-MWN-007.
- Hestenes, M.R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards*, **49**(6), 409-436.

無線によるコヒーレント通信の研究

(客員)千葉大学 深 沢 敦 司

本研究はスペクトル拡散変調符号分割多重アクセス通信方式の性能向上に関わる。この成果は正論文2件、プロシーディングス2件として学会発表した。

スペクトル拡散変復調のための Pilot channel を用いた無線伝搬路における伝搬路推定と位相等化の構成とその特性を記述し、明らかにした。

無線によるコヒーレント伝送法の構成を以下に示す。

- (i) Pilot 信号の連続送信を Pilot channel で行う。
- (ii) Pilot channel の構成を、Reverse Link では O-QPSK, Forward Link では QPSK とする。
- (iii) O-QPSK, QPSK のそれぞれの I channel に Pilot 信号を加算する。
- (iv) Pilot channel を用いた伝搬路推定と位相等化によりコヒーレント復調を行う(図1, 2)。

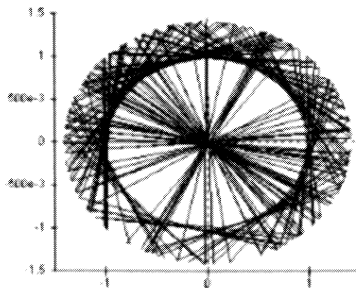


図 1. 位相等化なしの軌跡 .

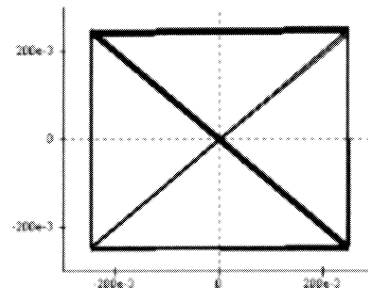


図 2. 位相等化後の軌跡 .

QPSK, O-QPSK 変調それぞれについて位相回転と線形性を評価した。複素平面上のベクトル(星座)として I, Q channel からなる入力信号の位相と振幅を観測した。図 1 では時間と共に星座が回転する。図 2 では星座が本来の位置に留まり等化の効果が明らかである。本提案の無線システムのデータ変復調特性および、スペクトル拡散変復調特性を BER vs. E_b/N_0 を用いて表した。

本提案の方式は以下の点で従来の方法より優れている。

- (i) 周波数における高速ドップラーシフトをともなうマルチパス伝搬での高速な位相回転の推定が, Pilot 信号の連続伝送(Pilot channel)によって実現する。
- (ii) 速い位相回転の等化が, 容易な論理演算により行われる。
- (iii) コヒーレント復調により, 高信頼の高速データ伝送が実現できる。これにより, 本提案システムの高信頼で簡易な無線 LAN への適用が考えられる。

参 考 文 献

- Arai, Yoshikazu, Igarashi, Kiyomi, Fukasawa, Atsushi and Takizawa, Yumi(2003). Multi-user detection to enhance the capacity of W-CDMA based on the conjugate gradient method, *Journal of Circuits, Systems, and Computers(JCSC)*, Manuscript # JCSC-MWN-004.
- Fukasawa, A., Iijima, Y., Igarashi, K., Inoue, S. and Takizawa, Y.(2002). Estimation and compensation of phase rotation in radio propagation using pilot channel for radio data transmission, *Asian Simulation Conference, International Conference on System Simulation and Scientific Computing, Shang-hai*, 963-967.
- Igarashi, K., Arai, Y., Sakai, T., Fukasawa, A. and Takizawa, Y.(2002). Multi-user detection for wide-band CDMA based on the conjugate gradient method, *WSEAS ICOSMO*, Skiathos, Greece, Sep.25-28, 447-350.
- Iijima, Y., Inoue, S., Kashima, T., Fukasawa, A. and Takizawa, Y.(2003). Coherent SS demodulation with estimation and compensation of phase rotation using a pilot channel, *Journal of Circuits, Systems, and Computers(JCSC)*, Manuscript # JCSC-MWN-007.

待ち時間分布に関する研究(2)

平野 勝 臣

本年度の研究

1. 多値 exchangeable 系列において,長さ有限の任意のパターンに対し,その逆順パターンの待ち時間分布はもとのパターンが起こるまでの待ち時間分布に等しいこと,などがわかった(Aki and Hirano(2002)).
2. 確率生成母関数が有力な道具である解析事例として k -match problems(Hirano and Aki(2003))に適用された.

以下,ここでは 1 について述べる.本報告は Aki and Hirano(2002)に基づいている.

逆順パターンの待ち時間分布

X_1, X_2, \dots を $\{1, 2, \dots, \mu\}$ -値 iid 系列とし, $P(X_i = 1) = p_i, i = 1, 2, \dots, \mu, \sum_{i=1}^{\mu} p_i = 1$ とする.例えば,長さ 4 のパターン 3123, 1233, 3213 の待ち時間分布が等しいかどうかを知りたいとする.直感的にはわからないだろう.そこで 3123 と 1233 の待ち時間分布の確率生成母関数を求めると,それぞれ $\phi^{3123}(t) = p_1 p_2 p_3^2 t^4 / (1 - t + p_1 p_2 p_3 t^3 - p_1 p_2 p_3 (1 - p_3) t^4)$, $\phi^{1233}(t) = p_1 p_2 p_3^2 t^4 / (1 - t + p_1 p_2 p_3 t^4)$ となり,異なることがわかる.しかし 3123 の逆順パターン 3213 の待ち時間分布を調べると, $\phi^{3213}(t) = \phi^{3123}(t)$ となり,等しいことがわかる.

逆順パターンの待ち時間分布について,一般的なことが Aki and Hirano(2002)で示されている.すなわち,多値 exchangeable 系列(この系列をやや広くできる)において,長さ有限の任意のパターンに対して,その逆順パターンの待ち時間分布はもとのパターンの待ち時間分布に等しい.証明は,確率生成母関数を利用するのではなく,両者の typical sequence の数を調べることによって示される.また r 番目のパターンが起こる(overlapping count)までの待ち時間分布は,その系列において r 番目の逆順パターンが起こる(overlapping count)までの分布に等しい.大きさ n の観測で,あるパターンが起こる回数(overlapping count)の分布は,その逆順パターンの起こる回数(overlapping count)の分布に等しい.

この結果, iid または exchangeable 系列において,あるパターンがはじめて起こるまでの試行数の観測値をとる.最後のパターンが起こってから終わりまでの試行数の観測値もこのデータと考えてよい.すなわち 2 倍の観測値を取ることができる.

参 考 文 献

- Aki, S. and Hirano, K. (2002). On waiting time for reversed patterns in random sequences, *Ann. Inst. Statist. Math.*, **54**, 713–718.
- Hirano, K. and Aki, S. (2003). On k -match problems, *J. Statist. Plann. Inference*, **109**, 67–79.

逐次数値積分を用いた多次元正規分布の確率評価

栗 木 哲

$z = (z_1, \dots, z_n) \in R^n$ を各成分が独立に標準正規分布 $N(0, 1)$ に従う確率ベクトル, $A \subset R^n$

を凸多面体とすると、確率 $P(A) = \Pr(z \in A)$ の数値計算について考える。この確率の計算は、多重比較、多重検定の有意水準あるいは検出力の評価において本質的である。しかしながら次元 n が小さい場合を除くと、素朴な数値積分やモンテカルロ法は役に立たない。

R^n の (閉) 単体錐の全体を S とおく。その要素 $Q \in S$ は半平面 $H_i = \{x \in R^n \mid \langle a_i, x \rangle \leq b_i\}$ の交わり $Q = \bigcap_{i=1}^n H_i$ と表すことができる (a_i は一次独立な単位法ベクトル)。 $\langle a_i, a_j \rangle = 0$, $\forall |i-j| \geq 2$, のとき単体錐 Q はオーソスキームと呼ばれる。オーソスキーム錐の全体を S_0 とおく。

Miwa et al. (2003) は (i) $P(Q)$, $Q \in S_0$ を次元 n に依存しない手間で計算する数値積分アルゴリズムが存在すること (ii) 任意の単体錐 $Q \in S$ は $Q = \sum_i \epsilon_i Q_i$, $Q_i \in S_0$, $\epsilon_i = \pm 1$, a.s. のようにオーソスキーム錐に有向分割 (符号付き分割) できることを示した。

これより任意の単体錐 Q に対して確率 $P(Q)$ を $P(Q) = \sum_i \epsilon_i P(Q_i)$ の形で数値計算できる。さらにいくつかの方法によって (iii) 凸多面体 A の補集合 $A^c = R^n \setminus A$ は単体錐に有向分割できることも示される (i)-(iii) を組み合わせることにより任意の凸多面体 A に対して $P(A)$ の数値計算が可能となる。

参 考 文 献

Miwa, T., Hayter, T. and Kuriki, S. (2003). The evaluation of general non-centred orthant probabilities, *J. Roy. Statist. Soc. Ser. B*, **65**(1), 223-234.

Functional CLT の応用例について

西 山 陽 一

Functional CLT を応用する際の注意点を、次のような非線型ノンパラメトリック時系列モデルの適合度検定の問題を通じて解説した。

次の式で与えられる実数値時系列 $\{X_i\}_{i \in \mathbb{Z}}$ を考える：

$$X_i = h(X_{i-1}) + \varepsilon_i,$$

ただし h は実数値関数であり、 ε_i は実数値確率変数であって

$$P(\varepsilon_i \leq 0 \mid \mathcal{F}_{i-1}) = \frac{1}{2} \quad \text{almost surely}$$

を満たすものとする。ここに $\mathcal{F}_i = \sigma\{X_j : j \leq i\}$ である。 $K : \mathbb{R} \rightarrow [0, \infty)$ はコンパクトな台をもつ核関数であるとし、 $\{b_n\}_{n \in \mathbb{N}}$ は $b_n \downarrow 0$ であるような正数の列であるとする。我々は、確率過程 $V^n = (V^n(x) \mid x \in \mathbb{R})$ を

$$V^n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i^n(x) Z_i$$

によって導入する。ここで

$$Y_i^n(x) = \frac{1}{b_n} \int_{-\infty}^x K\left(\frac{X_{i-1} - u}{b_n}\right) du,$$

$$Z_i = \text{sign}(X_i - h(X_{i-1})),$$

$$\text{sign}(x) = \begin{cases} -1 & \text{for } x \leq 0, \\ 1 & \text{for } x > 0 \end{cases}$$

とする．このとき，緩い仮定のもとで，確率過程 V^n が標準ブラウン運動を時間変更した確率過程に $\ell^\infty(\mathbb{R})$ 空間の中で分布収束することが証明できる．また，その系として， $S^n = \sup_{x \in \mathbb{R}} |V^n(x)|$ によって定義された統計量が $[0, 1]$ 上の標準ブラウン運動の最大点に分布収束することが得られる．つまり，漸近的に distribution-free であるような検定統計量 S^n が得られた訳である．時系列 $\{X_i\}_{i \in \mathbb{Z}}$ がマルコフ的であることは必要でなく，ノイズ ε_i は過去全体に依存してよいことに注意されたい．

証明は，マルチンゲール確率場に対する functional CLT を適用することによってできる．ポイントは，分割エントロピー条件をチェックするために，ヘルダーの不等式が使える形まで式変形する，ということである．

不完全 K -型変形ベッセル関数の導入による対流・拡散モデルの拡張

松 縄 規

不完全 K -型変形ベッセル関数を新たに導入することにより，一般化した対流・拡散方程式の初期-境界値問題に基づいて一般指数相反分布族を構築した．その際，これまで一般的な状況下では未解決の課題であった，一般指数相反分布の累積分布関数型を不完全 K -型変形ベッセル関数比 $Q^*(x)$ として与えた．

拡張した対流・拡散方程式として

$$w^{1-\alpha} \frac{\partial U(\eta, w)}{\partial w} = \alpha \eta^{2(1-\alpha)} \left[\kappa \beta^2 \frac{\partial^2 U(\eta, w)}{\partial \eta^2} - \nu \frac{\partial U(\eta, w)}{\partial \eta} \right],$$

$$U(\eta, w) > 0 \quad (\eta, w, t, \kappa, \nu, p > 0, \alpha, \beta \in \mathbb{R} - \{0\})$$

を考察した．ここに $U(\eta, w)$ は位置 η ，一般化時刻 w における非定常温度場の温度分布で κ は拡散係数， ν は対流速度を表す．温度分布は次の初期-境界条件 $U(\eta, +0) = 0, U(+0, t) = \{2 \cdot K_{-\rho/\alpha}(\eta\nu/(2\kappa))\}^{-1} \eta^{\rho/|\alpha|} \nu^{-\rho/|\alpha|} \cdot 1(t), \lim_{\eta \rightarrow \infty} U(\eta, w) = 0$ を満たすものとし， $\eta^{2(1-\alpha)} w^\alpha := t$ の関係にあるものとする．

上記方程式は逆ガウス分布に対する対流・拡散方程式の関係の一般化であり，次の一般化冪逆ガウス分布を生成する． $a = 2\kappa/(\eta^{2(1-\alpha)} \nu^2), b = \eta^{2|\alpha|}/(2\kappa)$ として，その累積分布関数型は

$$Q^*(x) = K_{\rho/\alpha}(\sqrt{b/a}; x^{|\alpha|}/(2a)) / K_{\rho/\alpha}(\sqrt{b/a}), \quad (x, a, b > 0, \alpha \in \mathbb{R} - \{0\})$$

で与えられる．ここで分母は次数 ρ/α の第 2 種変形ベッセル関数を，分子は

$$K_{\rho/\alpha}(z; x) = \frac{1}{2} \left(\frac{z}{2}\right)^{\rho/\alpha} \int_0^x \exp\left\{-\left(t + \frac{z^2}{4t}\right)\right\} \cdot t^{-\rho/\alpha-1} dt, \quad (z > 0, x > 0)$$

を表す．後者を不完全 K -型変形ベッセル関数と呼ぶことにする．これはまた次のように表現できる：

$$K_{\rho/\alpha}(\sqrt{b/a}; t^\alpha/(2a)) = \int_0^t x^{\rho-1} e^{-\frac{1}{2}(a^{-1}x^\alpha + bx^{-\alpha})} dx / \{2 \cdot |\alpha|^{-1} (\sqrt{ab})^{\rho/\alpha}\}$$

$$= 2 \cdot (\sqrt{b/a})^{\rho/\alpha} \times \begin{cases} \int_{\frac{1}{2}(t^\alpha/a-b/t^\alpha)}^{\infty} \frac{(\sqrt{z^2+b/a}+z)^{\rho/\alpha} e^{-\sqrt{z^2+b/a}}}{\sqrt{z^2+b/a}} dz, & (t > 0, \alpha < 0), \\ \int_{-\infty}^{\frac{1}{2}(t^\alpha/a-b/t^\alpha)} \frac{(\sqrt{z^2+b/a}+z)^{\rho/\alpha} e^{-\sqrt{z^2+b/a}}}{\sqrt{z^2+b/a}} dz, & (t > 0, \alpha > 0). \end{cases}$$

β -ダイバージェンス最小化 ICA 法における β の選択方法

南 美穂子

独立成分分析 (Independent Component Analysis, ICA) は、いくつかの独立な発生源からの信号が線形に混合されて観測されるとき、その観測された信号からもとの互いに独立な信号を復元する手法である。線形変換によって復元できるので、問題は復元行列の推定に帰着される。Minami and Eguchi (2002) は頑健な推定方法として、 β -ダイバージェンスの最小化による方法を提案した。推定量は β が大きいほど頑健で外れ値の影響を受けにくくなるが、反面、 β が大きすぎると推定効率が悪くなる。そこで、本研究ではデータに対して適切な β の選択方法を提案した。

β を評価するための尺度には、独立な信号が復元された場合に良い評価を与えることはもちろんであるが、外れ値に対しあまり大きなペナルティを与えないことを望む。これを満たすものとして、 β の値をある程度大きく取った ($0.3 \sim 0.5$) β -ダイバージェンスを用いることにする。また、汎化能力を測るための方法として K-fold クロスバリデーション法を用いる。復元行列のスケールは復元信号の独立性には影響を及ぼさないが β -ダイバージェンスで評価する場合はスケールによって結果が違ってくる。適切な評価をするために復元行列のスケールを評価に使う β -ダイバージェンスに合わせて調整することにする。

評価のための β -ダイバージェンスの値もまた確率変数であるから、ばらつきの尺度としてクロスバリデーションの際の標準誤差を計算し、それぞれの β に対する評価の値と標準誤差ひとつ分を足し引きしたのも同じ図にプロットする。判断方法としては、もし評価値が途中で急激に下がっているようであれば、下がった直後の値を選択する。あるいは、'one-standard-error' ルールを採用する。

例として、この方法を 2 次元一様分布に従う確率変数に外れ値を加えたデータとスパイクノイズのあるスピーチデータに適用した。いずれの場合も評価値が途中で急激に下がっており、下がった直後の値は確かにもとの独立信号を復元し、かつ、大き過ぎない適切な β であった。また、外れ値やノイズがないデータに適用すると、評価値はいずれの β に対してもあまりかわらず、 $\beta=0$ 、つまり通常の K-L ダイバージェンスによる方法を選択することになり、適切な選択であった。

参 考 文 献

- Minami, M. and Eguchi, S. (2002). Robust blind source separation by β -divergence, *Neural Computation*, **14**, 1859–1886.

Classification from DNA Microarray Data

江口 真透

個人の遺伝子発現データと疾病の関連について統計的判別に関する研究を発表した。遺伝子発現データは近年のゲノム工学の発展により、より大量なデータが獲得されるようになってきている。1人に対して数千～数万の遺伝子の発現量が計測される。今回の発表では White Head 研究所 (MIT) の Gulab 博士の研究グループが公開しているデータに基づいて統計的考察を行った。扱われたデータは Affimetrix 社が開発した高密度オリゴヌクレオチドによる遺伝子発現の方法から白血ガン 72 人の検体データに基づき、7109 次元の発現量ベクトルからなる。この特徴ベクトルから白血ガンの中でも 2 つの疾病 AML が ALL を判別できるかどうかという問題に帰着される。

この問題に適した方法論の候補としてロジステック型判別法 (江口 (2002), Eguchi and Copas (2002)) や AdaBoost を含む U -Boost (Eguchi (2002a, 2002b), Murata et al. (2002), Takenouchi and Eguchi (2002)) の適応が考えられた。しかしながらパターン識別の問題というよりも、その前処理について大きな問題点があることを指摘した。例えば、線形ロジステックモデルを考えるとパラメータ次元は 7109 で標本サイズが 72 になる。過剰な特徴空間の情報を例題一つに対して百の答えを用意しなくてはならないことになる。簡明な弱学習機の構成から AdaBoost をかけたらトレーニングエラーを 0 にする 3 個の遺伝子の組が数百組も見つかってしまう。これでは、たとえ良い検証法を設計しても何も望ましい結果はでないだろう。この問題に対して KS 検定によって p 値が 0.01% 以下の遺伝子 100 個を特定し、その下で判別問題を考察した。この遺伝子発現の測定方法が cDNA, GenTips, Aligent など発展しつつあること、原データの加工の問題など未だ絞りきれない問題点があることを指摘し今後の方向について報告した。

参 考 文 献

- 江口真透 (2002). 統計的識別の方法について ロジスティック判別からアダブーストまで, 応用統計学会 第 24 回 シンポジウム, 特別講演.
- Eguchi, S. (2002a). U -Boosting method and information geometry (invited talk) SRCCS Internatinal Statistical Workshop, Seoul National University.
- Eguchi, S. (2002b). Information geometry of Bregman divergences, Conference on Information Geometry and Its Applications, Pescara, Italy.
- Eguchi, S. and Copas, J. (2002). A class of logistic-type discriminant functions, *Biometrika*, **89**, 1–22.
- Murata, N., Takenouchi, T., Kanamori, T. and Eguchi, S. (2002). Information geometry of U -Boost and Bregman divergence, Research Memo., No. 860, The Institute of Statistical Mathematics, Tokyo.
- Nishii, R. and Eguchi, S. (2002). Image segmentation by structural Markov random fields based on Jeffreys divergence, Research Memo., No. 849, The Institute of Statistical Mathematics, Tokyo.
- Takenouchi, T. and Eguchi, S. (2002). Robustifying AdaBoost by adding the naive error rate, Research Memo., No. 859, The Institute of Statistical Mathematics, Tokyo.

Subexponential class の拡張と関連する分布族

志村 隆 彰

Subexponential distribution は $[0, \infty)$ 上の分布で $\lim_{x \rightarrow \infty} \overline{\mu * \mu}(x) / \overline{\mu}(x) = 2$ で特徴付けられる(ここで, $\overline{\mu}(x) = \mu(x, \infty)$). 片側安定分布, パレート分布など正則変動する裾をもつ分布やワイブル分布, 対数正規分布などは, subexponential であり, このような分布全体からなる分布族を S とかく. この分布族 S に関連する様々な分布族についていくつかの結果が得られた. S は任意の $k \in \mathbb{R}$ に対して $\lim_{x \rightarrow \infty} \overline{\mu}(x+k) / \overline{\mu}(x) = 1$ を満たす分布族 L の部分族である. また, $L(\gamma)$ ($\gamma > 0$) を任意の $k \in \mathbb{R}$ に対して $\lim_{x \rightarrow \infty} \overline{\mu}(x+k) / \overline{\mu}(x) = e^{-\gamma k}$ を満たす裾をもつ分布の族とし, 逆正規分布のようにさらに $\lim_{x \rightarrow \infty} \overline{\mu * \mu}(x) / \overline{\mu}(x) = \int_0^\infty e^{\gamma t} \mu(dt)$ となる分布は convolution equivalent と呼ばれる(全体を $S(\gamma)$ ($\gamma \geq 0$) であらわす. 以上は既知). これらの拡張として, 半安定分布のように $\limsup_{x \rightarrow \infty} \overline{\mu * \mu}(x) / \overline{\mu}(x) < \infty$ を満たす分布の族を OS , 任意の $k \in \mathbb{R}$ に対して $\limsup_{x \rightarrow \infty} \overline{\mu}(x+k) / \overline{\mu}(x) < \infty$ を満たす分布の族を OL で表す. これらの各分布族について, 裾の重さや合成積及びその意味で根に関する閉性といった諸性質や判定条件, 相互の関係, 具体例の構成などを行った. 具体的には, 以下のようなことがらである. ハザード関数が非増加という条件の下で, $\mu \in S(\gamma)$ となるための必要十分条件が $g(t) = -\log \overline{\mu}(t)$ としたとき, $\lim_{x \rightarrow \infty} \int_0^x \exp\{tg'(x) - g(t)\} g'(t) dt = \int_0^\infty e^{\gamma t} \mu(dt)$ で与えられること(Pitman の S に対する条件 (1980) の拡張)を示した. L かつ OS だが S でない分布, $L(\gamma)$ かつ OS だが $S(\gamma)$ でない分布等の具体例の構成による存在証明を与えた. 最初にあげた具体的な分布は S であると同時に無限分解可能分布でもあり, その裾と対応する Lévy 測度の裾の比が遠方で 1 に収束するのは, その無限分解可能分布自身が S であることが必要十分であることが知られている(Embrechts et al. (1979)). このことの一般化を試みた結果, 比が遠方で定数に収束するのは, $L(\gamma)$ の条件のもとでは, $S(\gamma)$ に限られることなどがわかった. 尚, この研究は渡部俊朗氏(会津大学)との共同研究である.

参 考 文 献

- Embrechts, P., Goldie, C. M. and Veraverbeke, N. (1979). Subexponentiality and infinite divisibility, *Z. Wahrsch. Verw. Gebiete*, **49**, 335–347.
- Pitman, E. J. G. (1980) Subexponential distribution functions, *J. Austral. Math. Soc. Ser. A*, **29**, 337–347.
- 志村隆彰, 渡部俊朗 (2003). Infinite divisibility and generalized subexponentiality, 統計数理研究所共同研究レポート, No. 157, 118–135.

極値分布と極値データ解析

(客員)神戸商船大学 高橋 倫也

いくつかの単位領域(または単位期間)での観測データから, 与えられた領域(または期間)における最大データを予測する問題を考える. 古典的な極値データ解析では, 各単位領域での最大データのみを用いて与えられた領域における最大データの予測を行う.

ここでは, 上位 r (> 1)個までのデータを用いて, 与えられた領域における最大データの予測法について研究した. まず, 上位 r 個のデータに適合させる(漸近)同時分布を示し, その性

質について調べた。これ等の性質を用いて、上位何個までのデータが利用可能か(r の決定)について研究した。これらの結果を、上位 3 個まで測定した孔食深さデータに適用した。

また、T-return level と呼ばれる上側確率点の推定精度が、上位 r 個を用いる事により極値データのみの場合よりどれだけ改善されるかを詳しく調べた。

教育試験と臨床試験を支える共通要因

柳 本 武 美

本年度の研究の一つとして、医歯学共用試験 CBT の実施に関して、試行試験の設計と解析について指導と支援を行った。この研究に関して、共著で研究報告を一編まとめると共に、各種研究集会・会合において口頭発表を行った。

医・歯学生に対して臨床実習前に、診療に必要な医歯学の知識の修得を確認するために、全国的規模での試験が企画されている。このプロジェクトに統計的側面で協力した。データ収集の企画とその解析が求められているが、特に前者について参画した。全国の医・歯学生を対象にして試験の実施の都合で CBT が採用された。CBT には二つの側面 1) 計算機を活用する機能(情報科学的側面)と 2) 能力と特性を推定する機能(統計学的側面)がある。後者は、無作為化項目試験と呼ぶことができる。その中ではランダムな出題と問題データベースの構築がキーとなることを明確に指摘した。

臨床試験との共通要因としては、1) 無作為化の利用、2) 最終消費者の視点、3) 公開と秘匿の必要性、4) 設計の重要性が挙げられる。無作為化はいわゆる「傾向と対策」からの決別に必要であり、出題者の個人的趣向の駆逐である。試験は受験者のために行うのではなく、最終消費者である実習生が診療する患者の利益を優先する。また、コアカリキュラムの内容と問題作成の方針と手続きは公開すべきだが、問題そのものは公開すべきでない。工夫された設計では、試験問題の評価ができる。なお、試験とは試験問題の試験でもあることに注意する。だから、勉強をしない受験者と規準を満たさない問題が落第となる。臨床試験と共通するこれらの要因は、正に質の高い証拠を得るための科学的技法でもある。そして、この推論を支えるためには単なる思いつきと努力では無理で、基盤の整備が必要になる。基盤の創成を通じて望ましい推論が行われる。このことは、耕作農業への転換には農地と灌漑設備の整備が必要であることと同様である。そうした意味では、今日の教育試験の改善は、教育評価における文明の黎明になぞらえることができる。

モデルに基づく遺伝子型のクラスタリング

藤 澤 洋 徳

疾病の治療に遺伝子型を利用する試みが行われ続けている。それを実現する一つの方向として遺伝子多型を利用した疾患関連遺伝子の探索が行われている。その研究の基礎となるのが遺伝子多型に基づいた個人の遺伝子型の同定である。本研究では、インベダー法を利用した一塩基多型(SNP)の遺伝子型を同定する方法を扱った。

インベダー法は適当な蛍光量を各個人に与える。我々は、その蛍光量を基にして、各個人を高々 3 グループにクラスタリングして、各個人の遺伝子型を判定する。現在までに使われ

ていたクラスタリングの手法は、あまり柔軟性がなく、しばしば誤った遺伝子型判定を与えていた。

問題を克服するために、正規混合分布を利用することでクラスタリングに柔軟性をもたせることを検討した。最初は、単純なグループ数が一定の場合を検討し、柔軟な機能の発揮を十分に確認した。ただし、実際には、グループ数が一定ではなく、正規混合分布を適切に機能させるためには、事前にグループ数を同定する必要がある。このようなグループ数の同定は困難な問題として広く知られている。

本研究では、実験に伴う事前情報を積極的に取り入れた罰則付き尤度法を採択し、誤ったグループ数が出にくい方法を提案した。この方法は、適当な場合には、存在しないグループの同定をも可能にした。そのような現象は理論的にも確認できた。加えて、提案した方法を多くのデータに適用し、非常に安定的に機能することを実証した。この方法はシステムの一部として組み込まれる予定である。

Model Induced Link Functions in GLM

大西 俊郎

1. 問題設定

一般化線形モデル (McCullagh and Nelder (1989)) におけるリンク関数の選択について、モデルビルディングの観点から議論する。確率変数ベクトル $x = (x_1, \dots, x_n)^T$ の各成分が指数型分布族

$$p(x_i; \mu_i) = \exp\{c(\mu_i) x_i - M(\mu_i)\} a(x_i)$$

に従っているとし、平均ベクトル $\mu = (\mu_1, \dots, \mu_n)^T$ が回帰式

$$f(\mu) = \alpha \mathbf{1}_n + Z^T \beta$$

によって共変量行列 $Z = (z_1, \dots, z_n)$ と関係づけられているとする。適当な関数 $g(\mu)$ を用いて新しいパラメータ

$$\gamma = \sum_{i=1}^n g(\mu_i)$$

を導入する。スロープ β が μ_i の z_i に対する依存性を表しているのに対し、 γ は (μ_1, \dots, μ_n) の全体的な傾向を示す。モデルを (β, γ) でパラメトライズし、同時密度関数 $\prod p(x_i; \mu_i)$ を $p(x; \beta, \gamma)$ と表示する。次の 2 つのケースを考えよう：

- (i) β が興味あるパラメータであり、 γ には関心がない
- (ii) γ が興味あるパラメータであり、 β には関心がない

本発表の目的は、上記の 2 つのケースに適した関数の組 $(f(\mu), g(\mu))$ を選択することである。

2. スコア関数の強不偏性

密度関数 $p(x; \beta, \gamma)$ に対応するスコア関数を $l_\beta(x; \beta, \gamma)$ および $l_\gamma(x; \beta, \gamma)$ と書く。リンク関数の選択に際して指針となる要請を定義として与える。それは関心のないパラメータのミススペシフィケーションに対するある種の頑健性である。

定義 . 共変量行列 Z を任意に固定する .

(i) スコア関数 $l_\beta(x; \beta, \gamma)$ が強不偏であるとは , 次の等式が成り立つことである .

$$(2.1) \quad E[l_\beta(x; \beta, \gamma^*) | p(x; \beta, \gamma)] = \mathbf{0} \quad \text{for any } (\beta, \gamma, \gamma^*).$$

(ii) スコア関数 $l_\gamma(x; \beta, \gamma)$ が強不偏であるとは , 次の等式が成り立つことである .

$$(2.2) \quad E[l_\gamma(x; \beta^*, \gamma) | p(x; \beta, \gamma)] = 0 \quad \text{for any } (\beta, \beta^*, \gamma).$$

上記の要請(2.1)および(2.2)に対する必要条件は ,

$$E[l_{\beta\gamma}(x; \beta, \gamma) | p(x; \beta, \gamma)] = \mathbf{0} \quad \text{for any } (\beta, \gamma)$$

であり , これについて次の定理が得られる .

定理 1 . 任意の共変量行列 Z に対して (β, γ) の Fisher 情報行列がブロック対角になるための必要十分条件は , 次の等式が成り立つことである .

$$\frac{c'(\mu)}{f'(\mu)g'(\mu)} = \text{Const.}$$

3 . 対数リンク , 恒等リンクおよび正準リンクの特徴づけ

スコア関数 $l_\beta(x; \beta, \gamma)$ の強不偏性からは対数リンクと恒等リンクが特徴づけられる . 特に後者は前者の特殊ケースである . 一方 , スコア関数 $l_\gamma(x; \beta, \gamma)$ の強不偏性によって特徴づけられるのは正準リンクである .

定理 2 . 任意の共変量行列 Z に対してスコア関数 $l_\beta(x; \beta, \gamma)$ が強不偏であるのは

$$f(\mu) = \frac{a_1}{a_2} \log \left(1 + \frac{a_2}{a_1} \frac{\mu - a_0}{a_1} \right),$$

$$g(\mu) = \frac{a_2 b_1}{a_1} M(\mu) + \left(a_1 - \frac{a_0 a_2}{a_1} \right) b_1 c(\mu) + b_0$$

のときに限る . ただし , a_0, a_1, a_2, b_0 および b_1 は定数であり , $a_1 b_1 \neq 0$ である .

定理 3 . 任意の共変量行列 Z に対してスコア関数 $l_\gamma(x; \beta, \gamma)$ が強不偏であるのは

$$f(\mu) = \frac{c(\mu)}{a_1} + a_0,$$

$$g(\mu) = a_1 b_1 \mu + b_0$$

のときに限る . ただし , a_0, a_1, b_0 および b_1 は定数であり , $a_1 b_1 \neq 0$ である .

4 . 多層問題への拡張

第 1 節の単層問題を拡張して , 層が K 個あり , かつ , 層によらず共通なスロープ β をもつモデルを考える .

$$p(x_{ki}; \mu_{ki}) = \exp\{c(\mu_{ki}) x_{ki} - M(\mu_{ki})\} a(x_{ki}) \quad (1 \leq k \leq K, 1 \leq i \leq n_k),$$

$$f(\boldsymbol{\mu}_k) = \alpha_k \mathbf{1}_{n_k} + Z_k^T \boldsymbol{\beta} \quad \text{where } \boldsymbol{\mu}_k = (\mu_{k1}, \dots, \mu_{kn_k})^T \text{ and } Z_k = (z_{k1}, \dots, z_{kn_k}).$$

パラメータベクトル $\gamma = (\gamma_1, \dots, \gamma_K)^T$ を

$$\gamma_k = \sum_{i=1}^{n_k} g(\mu_{ki})$$

のようにを導入すれば、定理 1, 2 および 3 の系としてそれらの多層問題版の命題が得られる。

5. 強不偏性が含意するもの

Fisher 情報行列がブロック対角になる、すなわち、Fisher 直交は局所的な性質である。一方、Kullback-Leibler 情報量に関する Pythagoras 関係が成り立つことは大域的な直交性である。本発表で定義した強不偏性は両者の中間の概念である。実際、恒等リンクと正準リンクでは Pythagoras 関係が成り立ち、対数リンクでは変形 Pythagoras 関係が成立する。

参 考 文 献

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

高次元ランダムパッキングの単純化

伊 藤 栄 明

向きをそろえて辺の長さ 1 の cube を辺の長さ x の cube にランダムにつめてゆくという問題は以前から議論されている。1 次元ランダムパッキングについては解析的な接近が可能であるが 2 次元以上の場合については解析的な研究は非常に困難であることが知られている。 d 次元の場合の充填率 β_d については Palásti の予想と呼ばれている予想が以前から議論されている。 x を大きくしてゆくときの極限 β_d が (i) 存在し (ii) β を $d=1$ のときの充填率の極限としたときに $\beta_d = \beta^d$ となるというものである。極限の存在 (i) は最近 Penrose (2001) により示されたが (ii) については計算機実験の結果より成り立っていないと考えられている。高次元のランダムパッキングについての解析的に議論できる単純なモデルを見つけることを試みてきた。向きをそろえて辺の長さ 2 の cube を辺のながさ 4 の cube にランダムにつめてゆくというもっとも単純な離散的ランダムパッキング (伊藤・上田 (1983)) を考えると充填率は α を約 0.44 にとると $d^{-\alpha}$ によく fit することに最近気づいた。これが Palásti の予想にかわる予想となりうるものではないかと考えている。スピンを持つ粒子のランダムパッキングという問題も 2 次元のランダムパッキングの最も簡単な場合であると考えられる (Itoh and Shepp (1999))。2 進探索木の連続モデルとして確率逐次 2 分割 (Sibuya and Itoh (1987)) について考えてきたがこの問題に関連した 1 次元片側ランダムパッキングを考えたい (Itoh and Mahmoud (2003))。この問題も興味深い高次元の問題を提起する。

参 考 文 献

Itoh, Y. and Mahmoud, H. (2003). One-sided variations on interval trees, *J. Appl. Probab.*, **40**, 654–670.
Itoh, Y. and Shepp, L. (1999). Parking cars with spin but no length, *J. Statist. Phys.*, **97**, 209–231.

- 伊藤栄明, 上田澄江(1983). 多次元立方体の離散的ランダムパッキングによる充填率について, 統計研究彙報, **31**, 65-69.
- Penrose, M. D.(2001). Random parking, sequential adsorption and the jamming limit, *Comm. Math., Phys.*, **218**, 153-176.
- Sibuya, M. and Itoh, Y.(1987). Random sequential bisection and its associated binary tree, *Ann. Inst. Statist. Math.*, **39**, 69-84.

文化を計る

村上 征 勝

文化を計量的な観点から研究すべく、「文化を計る」「美を計る」「古代を計る」「心を計る」の4つの課題の下に、幾種類かの文化現象の分析を試みたので、その概要を報告した。特に「文化を計る」の課題の下に行った、平安時代の『源氏物語』と『うつほ物語』の文体比較で、以下のような分析結果が得られていることをやや詳細に報告した。

『源氏物語』と『うつほ物語』は成立時期がほぼ同じと考えられるが、しかし用いられている言葉はかなり異なっていることが数量化Ⅲ類の分析で判明した。たとえば形容詞についてみると、『うつほ物語』は「よし」「おなじ」「なし」「おほし」など基本的、形式的な語が多く用いられているのに対し、『源氏物語』では多様な情意的形容詞が用いられている。同様に副詞の場合にも、『うつほ物語』では「かく」「いかで」「さらに」などの基本的、形式的な語が多く用いられているのに対し、『源氏物語』では「なかなか」「おのずから」「げに」といった情意性豊かな語が用いられている。このように、『うつほ物語』は文章が単純単調であり『源氏物語』の文章は複雑微妙な陰影に富むことが、言葉の出現頻度に基づく計量分析で裏付けられた。

東アジア価値観国際比較調査

幸福な家庭は皆似ているが、不幸せな家庭にはそれぞれの様相がある
(トルストイのアンナ・カレニーナの冒頭より)

吉野 諒 三

平成14年度より17年度までの「東アジア価値観国際比較調査」のプロジェクトの一環として、本年度は、日本、中国(北京、上海、香港)の人々の意識や価値観に関する面接調査を遂行した。現在の中国国内の事情により、日本で通常行われるような標本抽出のための基礎資料(住民票など)が一般公開されていないために、中国調査のためにはエリア・サンプリング法を工夫して遂行した。

中国調査データ解析については、昨年度の北京・上海調査の試行が既にあり、本年度はまず、主としてそれらのデータと過去の日欧米7か国国際比較調査データとの統合データについて遂行した。特に、「信頼感」のあり方について国際比較の焦点を当てた。F. フクヤマらによると、中国やイタリア、フランスは歴史上、中央集権の強い時代が長く、これが血族の結束を強めた。逆に血族の枠の外に対しては不信を示す傾向を発達させることになったとされる。しかし、GSS(米国の一般社会調査)の採用している信頼感の3項目スケールで比較すると、イタリアとフランスは低信頼、英米は高信頼、日本や中国は中程度となる。これは、中国はこれまでそう

見なされている程には低信頼ではないということか、中国人の意見表明の本音と建前の差が大きいということなのか、あるいは当該のスケールの限界なのか、今後の検討が必要である。

さらに、いくつかの満足感に関する項目等のデータを含めて比較すると、イタリアと中国の類似性と比較して、フランスは「家庭、家族 (Family)」のあり方について、それらから分離してくるのが明瞭になった。このように「信頼感」の観点からは、今まで同様と見なされてきた国々も、多面的なデータ解析の中で、それぞれの様相が浮かび上がってくることを確認した。

参 考 文 献

- 山岡和枝, 吉野諒三, 鄭 躍軍 (2002) 健康感と信頼感, 行動計量学会第 30 回大会発表論文集, 350-351.
Yoshino, R. (2002). A time to trust—A study on peoples' sense of trust from a viewpoint of cross-national and longitudinal study on national character—, *Behaviormetrika*, 29(2), 231-260.
鄭 躍軍, 吉野諒三 (2002). 標本抽出名簿がない場合の個人標本抽出, 行動計量学会第 30 回大会発表論文集, 346-349.

抽出名簿がない場合の社会調査とそのバイアス問題

鄭 躍 軍

日本のように調査対象と食い違いの殆どない抽出名簿(住民基本台帳や選挙人名簿など)が利用できるということは、世界では例のないことである。しかし、近年日本でも住民基本台帳や選挙人名簿の閲覧制限が厳しくなっている自治体が増えつつある。このような社会調査環境の変化に備えて、国際比較研究において比較可能性ならびに信頼性の高いデータを収集するために、いわゆる日本型標本抽出法を改めることが求められており、抽出名簿のない場合の確率標本法を考えておく必要がある。本研究では国勢調査などの統計データに基づき、調査地点が抽出された後、家や世帯が書かれている世帯配置図を基に、抽出された調査地点別に世帯リストニングを行い、調査時点での最も正確な世帯リストを作成し、そして無作為に個人を抽出するような確率標本法の構築について研究してきた。また、中国の北京、上海、杭州、昆明の4都市で遂行した意識調査の実際を通し、抽出名簿のない場合に適応した確率標本法を実践的に考案し、標本誤差および非標本誤差にかかわる諸問題も統計的に解析した。

抽出名簿のない場合の確率標本の抽出手順としては(1)国勢調査などの統計データを利用して、人口や世帯の規模に比例してPPSで調査地点の抽出を行う;(2)抽出された調査地点から予め作成した世帯配置図を基に等間隔で調査対象世帯を抽出する;(3)抽出された世帯を直接訪問する際、世帯住所の通し番号(最下位1桁)と成人家族構成員の生年月日順位(降順)を基に発生した乱数により、調査対象個人を抽出する、という3段階抽出方法を提案した。

なお、エリア・サンプリングを基に考案した3段階抽出法を実際の調査データに基づき検討した結果、標本誤差と非標本誤差の影響によって性別、年齢と学歴などの属性については母集団と標本の間に顕著な差があり、無回答誤差による非標本誤差も大きいことが明らかになった。今後の課題としては、バイアスを最小にするための調査地点の適切な規模の設定や標本誤差及び非標本誤差の見分け及び定式化などが残っている。

「国民性調査に関する若干の検討」補遺

坂元 慶行

日本人の国民性の統計的研究には、日本人の意識動向の解明、調査法の研究、解析法の研究の3つの目的があるが、まずもって重要なのは、第1の目的である意識動向の解明のための質問文作りである。特に、2003年秋には第11次全国調査が予定されており、そのための質問文の作成は急務である。そこで、今年度は、12月に吟味調査を行い、主として、過去の国民性調査で一度は調査されたことのある質問を再調査し、その利用可能性についての検討を行った。この結果については、2003年1月の統計数理セミナーで「国民性調査に関する若干の検討」と題して報告した。年度研究報告会では、この吟味調査から「公益と個人の権利」をとりあげ、上の統計数理セミナーの報告を補足した。

1978(昭和53)年の国民性調査には、「公益と個人の権利」についての現状評価と対策についての質問が2問あり、現状に関しては「個人の権利が軽視されている」という評価が多いが、対策としては「公共の利益を優先せよ」が多いという結果が得られている。また、2問のクロス表を見ても、現状評価がどうであれ、「公共の利益を優先せよ」が多数意見であった。しかし、ほぼ四半世紀後の2002年12月の吟味調査では、どちらの質問でも「個人の権利」派が増え、2問のクロス表でも、現状評価と対策が一貫する回答が多数派になった。しかし、これは論理的な一貫性が高まったことを意味するというより、他のいろいろな調査や質問で共通に見られる、従来のタテマエが崩れホンネが露わになるという現象の一例に過ぎないのではないか、という指摘を行った。ここでの質問の場合、「個人の権利が軽視されている」はいわばホンネであり、「公共の利益を優先せよ」は従来のタテマエであると見られるからである。

筆者は、戦後の意識動向の基調の一つは私生活優先という価値観の顕在化であると考えているが、一般的にホンネとはこの私生活優先という考え方にほかならない。とはいえ、私生活の実体は、家庭や職場等、さまざまな生活領域で、いわば身近な人間関係にくるまれた“私”からひとりぼっちの“私”へ変わりつつある。この点の推移が今後の質問で追跡していかなければならないポイントの一つではないかと思われる。

貿易自由化の森林資源管理への影響——最適確率制御モデルの応用——

吉本 敦

森林環境の変化はローカルな土地利用の変化に起因するものであり、土地所有者レベルでの森林から農地へ、あるいは農地から放牧地、休耕地へといった利用の変化が地域的に集積され、その結果が恒久的な森林環境の変化に繋がっていく。そこでは、土地所有者ないし経営者は不確実な社会、経済条件の下で、自身の土地利用に関し最適選択を行っているものと考えられ、選択に際しての個々の所有者の持つ情報と判断が、土地利用の変化に大きな影響を与えることとなる。本研究では、こうした土地所有者の意思決定と行動様式をシミュレーションできるような土地利用変動プロトタイプモデルを開発し、貿易自由化に伴う市場環境の不確実性の高まりが伐採地の拡大や土地の転用等を引き起こす可能性について分析した。なお、本研究で構築したモデルは森林土地利用の転換を考慮できる土地利用変動モデルであり、金融分野のオプション理論を適用したものである。

分析では、経営を持続できる価格を最低許容価格とし、価格ダイナミックスの変化に対する最低許容価格の変化に注目した。分析の結果、市場価格のボラティリティの増加に対しては、最

低許容価格は減少し、その付近に市場価格が近づくと、収穫時期の遅延が生じることが分かった。次に、市場価格の傾向的な減少を仮定した場合、最低許容価格は逆に上昇し、その付近に市場価格が近づくと収穫時期が短縮され、管理放棄への土地利用の転用が収穫時期の延長を伴わず即座に行われることが分かった。

一般に、自由化などの新たな政策が導入されると、まず短期的に市場価格のボラティリティが増加することが予想される。仮にボラティリティが増加すれば、経営持続の最低許容価格は減少する。その結果、減少した最低許容価格と市場価格との乖離が広がり、その利ざやを狙った供給サイドの競争の激化が予測される。一般的に供給サイドの競争が激化すれば、市場価格は最低許容価格に近づくように減少し、傾向的な価格の減少が生じる結果となる。そして、この減少は逆に最低許容価格を上昇させる結果になり、供給サイドの経営継続が困難になる確率が増すものと考えられた。

社会調査データの解析における潜在変数モデルの活用

前 田 忠 彦

潜在変数の形式的な定義にはいくつかの流派があるが、潜在変数(モデル)を日常的な語彙で表現すると「データとして値が観測不可能な変数(を含むモデル)」としておけば十分であろう。本研究では、潜在変数を導入したモデルによる分析が、社会調査データの解析にあたりどのような有利さを持ち、調査法研究にどのような点で貢献するのかという点について考察した。

潜在変数モデルには、導入する潜在変数 f が離散型か連続型か、潜在変数の指標 観測変数 X が離散型か連続型か、の二つの観点により分類した場合に、いくつかのモデル群が考えられる。ここでは f, X ともに連続型変数と想定されるモデルのうち構造方程式モデル(SEMwLV: Structural Equation Models, with Latant Variables)および、その特殊ケースとも言える因子分析モデル(FA: Factor Analysis)を主たる考察対象とする。

SEMwLV は潜在変数を伴わないモデル(パス解析型 SEM)が潜在変数を伴うモデルに発展したものであることができる。他方、因子分析 FA については、探索的な方法(EFA: Exploratory FA)から確認的(仮説)検証的な方法(CFA: Confirmatory FA)が展開した。問題は、1)パス解析型 SEM から SEMwLV への発展において、また 2)EFA から CFA への展開によって、それぞれの分析法の文脈で何が改善され、また調査データ解析の文脈からは、その改善によって調査方法論研究に対してどのような貢献がなされたか、という点である。

1)の点について、パス解析への潜在変数の導入によって多くの変数間の関係をより縮約的に記述出来る可能性が高まった点を指摘することができ、またこの有利さを支える一つの側面として、いわゆるパス係数の解釈がよりシンプルになる点を挙げられる。異論・反論もあり得ようが、それらを含め、ここではこの論点の詳細には立ち入らない。

他方、2)の点すなわち、EFA から CFA への展開においては、端的に言って「観測変数の測定状態に対するきめ細かいモデリング」が可能になった点が最大の進歩と考えられる。例えば計量心理学における「(古典的)テスト理論」の再解釈と発展により、観測指標 X の分散成分の分解を、測定状況に合わせたより自然な設定で考察することができる点などはその帰結である。調査データ解析の文脈では、そのことにより調査項目選択の妥当性や測定方法による影響の評価を行うための方法論が提供された点を指摘できる。

この第 2)の点に関する事例を、社会調査から得られた多特性・多方法行列と呼ばれる特殊な構造を持つ相関行列に対する CFA による分析モデルの適用に求め、CFA におけるモデル選択

アプローチの有効性を示した。更に分析結果を利用し、測定モデルに即した“信頼性”や“項目妥当性”の定義による計算例の検討を行い、調査で採用された測定方法が観測指標の分散にどの程度寄与しているかの考察を行うことができることも示した。

結局、CFA で EFA よりも柔軟な測定モデルの表現を獲得したことにより（上記の例では）測定目的の特性そのものによる変動と、測定方法の効果の分離を考察することができた。複数の測定条件で測定がなされていれば、項目測定法以外の要因であってもよく、それらの要因の貢献を知ることは、調査項目の設計や調査モードの効果の検討などにおいて威力を発揮すると考えられる。

Voronoi-Delaunay Cell の統計分布とその周辺

種 村 正 美

ポロノイ・セル(Voronoi cell)とその双対であるデローネイ・セル(Delaunay cell)は空間に散布された粒子の配置パターンの特徴づけや粒子配置の確率幾何学モデルとして非常に有用である。

本年度は、ポアソン点配置のポロノイ分割によって生じるポロノイ・セル(Poisson Voronoi cell と呼ぶ)の幾何学的特徴量の統計分布に関して、これまで行ってきた 5 次元までのユークリッド空間における計算機実験の結果に対する考察と、ポアソン点配置のデローネイ分割から生じるデローネイ・セル(Poisson Delaunay cell)の幾何学的特徴量の統計分布に関する理論および計算機実験の結果に対する考察を行った。

一般に、 d -次元ユークリッド空間における Poisson Voronoi cell は凸多面体(convex polytope)であるが、凸多面体は k -次元 ($0 \leq k \leq d-1$) の要素(k -face)をもつ。0-face は「頂点」で、1-face は「辺」、 \dots 、 $(d-1)$ -face は「面」または facet である。いま、 k -face の個数を f_k で表すとき、 $(f_0, f_1, \dots, f_{d-1})$ を f -vector とよぶ(たとえば Grünbaum(1967)参照)。 f -vector の要素の間には Euler の定理

$$\sum_{k=0}^{d-1} (-1)^k f_k = 1 - (-1)^d$$

が成り立つことが知られている(Grünbaum(1967))。また、 d -次元 Voronoi cell に対して関係式

$$d f_0 = 2 f_1$$

が成り立つ。これは、 d -次元 Voronoi cell の一つの「頂点」(0-face)から d 本の「辺」(1-face)が出ていていることを表し、Voronoi 分割の「頂点」が一つの Delaunay cell($(d+1)$ -面体)に対応することを反映している。これらの二つの関係から、2 次元および 3 次元では f -vector の一つの要素を与えると他の要素は一意的に定まる。なぜなら、2 次元では $f_1 = f_0$ となり、3 次元では $f_0 = 2f_2 - 4$ 、 $f_1 = 3f_2 - 6$ となるからである。それに対して、4 次元以上ではそうはならないことが分かる。すなわち、4 次元では $f_1 = 2f_0$ 、 $f_2 = f_0 + f_3$ となるので二つの要素 (f_0, f_3) を与えて初めて f -vector が定まり、5 次元では例えば $f_1 = (5/2)f_3$ 、 $f_2 = f_4 + f_3 - (3/2)f_0 - 2$ となるから、三つの要素 (f_0, f_2, f_4) を与えて f -vector が定まる等々である。

そこで、上の議論から Poisson Voronoi cell の f -vector の統計分布に関して、2 次元では「辺数」 f_1 の分布、そして 3 次元では「面数」 f_2 の分布が知られれば他の要素の分布は導かれることが分かる。それでは 4 次元以上ではどうなるかが興味のある点である。われわれは 4 次元お

よび 5 次元の Poisson Voronoi cell に対して計算機実験で生成したそれぞれ 5,000,000 個の独立標本(種村(2002a))から, 4 次元では (f_0, f_3) , 5 次元では (f_0, f_2, f_4) の同時分布を求めた. その結果, それぞれの場合に変数間に強い相関のある特徴的な分布をもつことが判明した(種村(2002b)).

また, 2 次元の Poisson Voronoi cell に対して, われわれの得た「辺数」分布から, N -角形の出現する確率 p_N の推定値 \hat{p}_N が与えられる. これに関して, 最近 Hayen and Quine(2000a, 2000b)が三角形の出現確率 p_3 の理論的導出に成功し, 複雑な被積分関数をもつ多重積分を実行した結果, $p_3 \approx 0.011240$ を求めた. 一方, われわれの推定値は $\hat{p}_3 = 0.01125$ であった. さらに, Hayen and Quine(2000a)は三角形の面積および周囲長のモーメントを理論的に導いたが, その数値計算結果は, われわれの計算機実験の結果と良い一致を示している. これらの事実は, 提出された理論計算とわれわれの実験結果の正しさを裏付けていると同時に, $N > 3$ に対しては理論計算が困難で, 信頼できる計算機実験が必要であることをもものがたっている.

Poisson Voronoi cell および Poisson Delaunay cell の幾何学量の統計分布に関する理論的導出の試みは他にもなされていて, Muche(1996)は 3 次元の Poisson Delaunay cell の統計分布に関して, 多重積分を含む一般的公式を提出している. そこでは一つの Poisson Delaunay cell の外接球が空(empty)であるという性質が利用されている. われわれは Muche(1996)が提出した一般公式を用いて, 3 次元の Poisson Delaunay cell の二面角の確率分布を導くことに成功し, 計算機実験から得られた二面角の分布がこれに良い一致を示すこと, いくつかのモーメントの理論結果と実験結果の間的一致も見出した.

参 考 文 献

- Grünbaum, B.(1967). *Convex Polytopes*, John Wiley & Sons, London.
- Hayen, A. and Quine, M.(2000a). The proportion of triangles in a Poisson-Voronoi tessellation of the plane, *Adv. in Appl. Probab. (SGSA)*, **32**, 67-74.
- Hayen, A. and Quine, M.(2000b). Calculating the proportion of triangles in a Poisson-Voronoi tessellation of the plane, *J. Statist. Comput. Simulation*, **67**, 351-358.
- Muche, L.(1996). Distributional properties of the three-dimensional Poisson Delaunay cell, *J. Statist. Phys.*, **84**, 147-167.
- 種村正美(2002a). Poisson Voronoi cells in 4 and 5 dimensions, *統計数理*, **50**, 308-309.
- 種村正美(2002b). 4 次元, 5 次元における random Voronoi cell の f-vector の統計分布, *形の科学会誌*, **17**, 50-51.

公開の抽籤の公平性, 意外性, 継続性についての考察

丸 山 直 昌

n 人の応募者に対して 1 位から n 位の順序を与えるような抽籤を, 応募者が一堂に集まることができない条件で行うことを考える. 各行が 1 から n の n 個の自然数を並べた順列であるような行列 M を使って, 偶然性を持って決まる数 m により, 行列 M の第 m 行目を当籤順位とする, という形でこの問題を定式化することを考える. すると, 抽籤のプロセスの公正さはこの行列 M をあらかじめ公開することにより確保され, 公平さは行列 M の均等散布性という条件に置き換えられる. しかし, 抽籤にはさらに意外性が必要である. 例えば, あらか

じめ決めておいた日の日経株価平均のように多くの人々が容易に知り得る数値で m を決めるとすると、公平さ、公正さという意味では問題ないが、一日の株価平均の変動幅がある程度の範囲内に収まるであろうという予感から、前日のうちに抽籤結果についての予感を与えてしまうようでは、抽籤としてはつまらない。蓋を開けるまで結果がわからないことが望ましく、それが意外性という概念である。また、実用例においては、抽籤は繰り返し何回も行われることも多い。その場合毎回同じ抽籤では参加者に飽きが生じる。繰り返し行っても参加者を飽きさせないためには、毎回行列 M を取り替える方法が考えられる。適切な取り替えの方式があれば、その方式は抽籤の方法として継続性があると言えるであろう。この意外性と継続性という点を行列の言葉でどのように表現できるか、考察した。

分子生態学における点過程モデル

島谷 健一郎

天然林モニタリングデータから森林動態を解明する研究は古来から世界各地で行われているが、1990年代後半から、そこに遺伝子データを加味する事で、交配・繁殖も踏まえた考察が始められた。ところで遺伝子の空間分布及びその時間発展については集団遺伝学で古くから研究が進められており、森林モニタリングへの応用ではそれを援用する形で用いている。しかし集団遺伝学では遺伝子に焦点を当て、遺伝子の器としての生物個体の空間分布は考慮しない。遺伝子拡散モデルでは全平面に遺伝子〔個体〕は存在しうると仮定して拡散方程式を適用するし、その離散化・個体ベース化では格子点上に個体を配置してその間で遺伝子を動かす。ところが現実の森林樹木では、各種はそれぞれパッチを形成して不均一に分布する場合が見られるが、そのような状況でも従来の集団遺伝学的手法を適用していいものであろうか。ここでは、マーク付点過程モデルを導入し、個体分布とその上の遺伝子分布を同時に扱い、両者が及ぼしあう影響について考察した。初等的な試みでも、個体分布の違いにより、同一の交配様式でもその上の遺伝子分布パターンは異なってくる例が普通に見られることから、従来の集団遺伝モデルの個体ベース森林モニタリングへの適用は適切ではなく、マーク付点過程による検証が必要と思われる。

コウホート分析にもとづく予測

中村 隆

スポーツやレクリエーションの参加人口を予測しようとするとき、人口の年齢・世代構成の変化のみならず、個々人の加齢による変化(年齢効果)、世代特性の違い(コウホート効果)をもたらす世代交代による変化、社会全体の時勢による一定方向への変化(時代効果)を区別することが重要である。これらが区別できれば、スポーツ参加の将来動向について見通しがつけやすくなる。

年齢効果が大きいスポーツ種目への参加は、個人は変化するが社会全体では安定している。ただし、人口構成の変化がある程度の変動をもたらす。一方、コウホート効果が大きい種目への参加は、個人は変わりにくいが社会全体については緩やかに変わってゆき、人口構成の変化がそれを加減速する。時代効果が大きい種目への参加は、個人の変化の総体として社会全体の

変化も大きく流動的なものとなる。以上の各効果の大きさは、継続調査データを分析するベイズ型コウホートモデルによって明らかにすることができる。

継続調査データの年齢階級数を I 、調査時点数を J 、コウホート区分数を K とし、 $\hat{\beta}^G$ 、 $\hat{\beta}_i^A$ 、 $\hat{\beta}_j^P$ 、 $\hat{\beta}_k^C$ をそれぞれ、コウホート分析の結果得られた総平均、年齢、時代、コウホート効果のパラメータの推定値とする。コウホート分析結果にもとづけば、たとえば、ゴルフの参加割合の予測値 $\hat{\pi}_{ij}$ は、将来の時代効果 $\hat{\beta}_{J+1}^P$ 、 $\hat{\beta}_{J+2}^P$ 、... と新規参入のコウホート効果 $\hat{\beta}_{K+1}^C$ 、 $\hat{\beta}_{K+2}^C$ 、... を適当なシナリオに基づいて設定することにより、

$$\hat{\pi}_{ij} = [1 + \exp(-\hat{\eta}_{ij})]^{-1}, \quad \hat{\eta}_{ij} = \hat{\beta}^G + \hat{\beta}_i^A + \hat{\beta}_j^P + \hat{\beta}_k^C$$

として求めることができる。さらに、年齢階級別の将来推計人口 P_{ij} を利用して、次のようにゴルフの将来参加人口 \hat{N}_{J+1} 、 \hat{N}_{J+2} 、... を予測する。

$$\hat{N}_j = \sum_{i=1}^I \hat{N}_{ij}, \quad \hat{N}_{ij} = \hat{\pi}_{ij} P_{ij}.$$

将来の各効果に関する現状維持の設定を含む複数のシナリオによる結果を比較すれば、人口構造の変化やコウホート効果をもたらす変化の方向について考察することができる。ただし、予測結果に対しては時代効果 $\hat{\beta}_j^P$ ($j = J+1, \dots$) の設定の影響が大きく、その設定問題が依然残される。

コウホート分析にもとづく予測法の詳細と適用例については、山本 他(2002)を参照。

参 考 文 献

山本達三、菊池秀夫、中村 隆(2002)。スポーツ参加人口の推定と予測—年齢・時代・コウホート効果・人口変動を考慮して—、スポーツ産業学研究、12(2)、33-46。

インターネット調査 実験調査から見てきたこと

大 隅 昇

インターネット調査(とくに Web 調査)が急速に普及している。我々は、1997 年以降、計 4 回の実験調査を行い、現行のインターネット調査の特性の検証に努めてきた。例えば、顕著な特徴として、回答率(調査依頼数に対する有効回答数の割合)がかなり低いことがあるが、これは登録者集団の特質(構築法、出处、鮮度等)に依拠することが分かってきた。また、インターネット調査における回答者は、従来型調査に比べ、かなり異なる特徴的な回答を示すこと、しかも回答者は、現状では明らかにインターネット・ユーザを代表するとは言えず、インターネット・ユーザの中のある一部の限定された集団である可能性が高いこと等も分かってきた。

従来型調査では、調査情報を表わす諸事項や指標、例えば、調査目的や項目、調査対象(母集団、標本数、抽出方法)、調査時期(期間)、回収結果(有効回収数・率、調査不能数・率やその内訳)等を調査概要として報告することが一般的な事とされ、こうした情報から当該調査の質的内容を客観的に評価し、それを前提に取得データに基づく調査結果の分析・解釈が行われてきた。

しかし、現状のインターネット調査では、回収率の定義も容易ではなく、前述の調査情報に相当する事項や指標として何を報告すべきかの共通した考え方も定まっておらず、調査の精度

や信頼性を議論する土壌も十分に育まれているとはいえない。今までの実験調査結果が示唆することは、実証研究により調査法としての特質を根気よく解読する作業を続けることである。

このようなことで、第1次実験調査(1997年)では、あるパネルを用いて計12回の継続調査を行いWeb調査の実態把握や諸特性の整理に努めた。この知見に基づき第2次、3次の実験調査(1999年~2001年)では、基本方針として、①従来型調査法との比較検証を行う、②複数の調査機関(複数のWebサイト)でほぼ同時期に複数回実施する、③原則として調査票は同じ形式・設問内容とする、④調査票へのアクセス方式はなるべく各機関の現状方式にならうなどを設定条件として進めた。第4次実験調査(2002年、2003年)ではこれに加えて、調査法としての質的側面の評価を行うため、①トラッキング(回答者行動を時間軸に沿って電子的に追跡し関連情報を収集する操作)によりWeb調査の技術的な問題や事象をより詳細に把握する、②登録者集団(リソース、パネル)構築方法の類似・差異と調査結果との関連を検証する、特に公募・自己参加型の構築方法と、従来型調査の調査対象者抽出方法(住民基本台帳やマスターサンプルの利用等)に基づくリソース構築方法との比較等を組み入れた。こうして第4次実験調査では、Web調査を「3つの異なるWebサイトで」「4つの異なるリソースまたはパネルを対象に」「同じ設問をほぼ同じ調査票で」「ほぼ同時期に」「2種の調査を各1回」実施し、また従来型調査を「可能な限りWeb調査に同等の設問で、ほぼ同時期に」実施した。この他Web調査では、調査票レイアウトの実験として、ページ分割の有無、設問形式の差異の回答への影響評価の実験も加えた。

この第4次実験調査から、インターネット調査の質を評価するうえで、登録者集団の構築法の性質・差異が、調査結果にどう影響し、何が現れるかが具体的に見えてきたこと、よって登録者集団の構築に関わる諸要素の重要度が見えてきたことなどがある。

とくに、インターネット調査の特徴の一つに、計画標本のうちの調査依頼発信数と、実際に回収されるアクセス延べ数とが一致しないということがある。ここに注目し、またトラッキング情報を活用することで、調査の質を測る目安とすべき指標が分かってきた。例えば、「調査不能(無回答)を、インターネット調査特有の区分(「未着数」「無接触数」「アクセスのみ数(接触数)」「回答送受信異常」等)で測ること、回答所要時間、回答所要アクセス数を知ること、さらに有効回答回収数(率)と調査不能の関係を定量的に求めること、完答数(率)、の測定等、様々な指標を示すことで調査結果の客観的な評価が可能となることが見えてきた。

なお本研究の一部は、研究課題「調査環境の変化に対応した新たな調査法の研究」として平成14年度研究所リーダーシップ支援経費の助成を得て行った。

電話による実験調査

土屋 隆裕

二種類の電話調査法による実験調査の結果を示した。一つの方法は、選挙人名簿から抽出したサンプルのうち電話番号判明者を対象とするTD(Telephone Directory)法であり、もう一つの方法は電話番号をランダムに発生させるRDD(Random Digit Dialing)法である。RDD法の回収サンプルのうち、電話番号非掲載者の結果は、明らかに他とは異なった。さらに、RDD法のサンプルを掲載者に限定しても、TD法の結果とは違いが見られた。この違いを説明するため、いずれの方法によっても得られる共通クラスと、各方法によってのみ得られる二つの独自クラスという三つの潜在クラスを導入した。混合多項分布モデルによる分析では、三つのクラスを適切に分類できないことを示した。一方、混合正規分布モデルを用いると、掲載者に限定

した RDD 法のサンプルにはやや大きな独自クラスが含まれていることが示唆された。

WWW 上の統計科学情報について

清水 信 夫

コンピュータ・ネットワークが急速に普及し、多種多様な情報が氾濫している現在、電子的に蓄積された膨大な量の統計データの有効利用や高度な統計分析手法を開示し普及させていく方法についての検討は急務と考えられる。最近ではインターネットの普及により、様々な分野において蓄積した統計関連情報を WWW (World Wide Web) 環境下で公開する動きが広がっている。統計科学の分野においても、WWW の利用により、研究・教育の発展に供する目的で多くの Web サイトが立ち上げられており、内容も次第に多様化する傾向にある。これらの Web サイトについていくつかの区分により大まかに分類した報告例はあるもの(Murdoch (2000) など)、統計科学関連の Web サイト全般の中からいかに有益なものを広範に選ぶか、さらにこれらのサイトからいかに有用な情報を取り出し表示するかについて研究が求められている。

本研究においては各種サーチエンジンを利用して統計科学関連 Web サイトを広範に収集した上で有用なサイトを絞り込み、その過程で順序化が可能と考え得るいくつかの指標を設けた。その上で各指標における各区分を数値化し、多次元尺度構成法の利用により多次元的な分類を試みた。また、各指標において各区分に属する Web サイトを表示するシステムを WWW 上で設計し表示した(清水・大隅 (2002))。

参 考 文 献

Murdoch, D. J. (2000). On the edge: Statistics & computing, *Chance*, 13(1), 49-51.

清水信夫, 大隅昇 (2002). WWW 上の統計関連情報について, 2002 年度統計関連学会連合大会 (日本統計学会第 70 回大会) 講演報告集, 140-141.

時空間統計モデルによる地震活動研究の新展開

尾 形 良 彦

1. 余震の統計モデルと地震活動変化の意味

大きな地震の破壊に伴う急激で局所的なストレス変化によって主断層近傍のフラクタル的断層群の破壊やすべりが誘発されることによって、余震の連鎖性・集中性を生じていると考えられている。このように余震のメカニズムは本質的にミクロで複雑であるため、その予測には、粒子系の統計力学のごとく、統計的なアプローチをとらざるを得ない。

本報告で言う「地震活動の変化」の意味は、実際の地震活動がモデルによる予測から有意に逸脱していることを意味する。地震活動の活発化や静穏化はこの意味でモデルを基準にしたもので、その相対的な比率を推定し、これが地震の誘発や抑制をする大域的なストレスの増減の反映であるとする。ミクロで複雑な現象の統計モデルによって相対的に地震活動を解析することによって広域のストレス変化が見易くなる。これからの地震活動の研究はこのような展開

なしには難しいであろう。

2. 時空間 ETAS モデルと地震活動変化の検出

時空間 ETAS モデル(Ogata(1998, 2002))は改良大森公式や余震域のスケーリング則などに依拠したもので、パラメタが地震活動の地域的特性に適應するようになっている。すなわちモデルのパラメタは位置の関数と考え、各地の地震活動の違いを定量化する。推定は 2 次元空間上の曲面としての滑らかさを制約条件として入れてベイズ法で最適な解として得られる。また各地の実際の地震活動と時空間モデルで計算された理論的活動度の相対比率の推定は時空間関数としての滑らかさを制約条件として考えてベイズ法で最適な解として得られる。その解である相対比率が有意に 1 より小さいか大きいかによって、相対的静穏化や活発化した地域や期間を可視化するのである。

3. 中部・近畿地方における 1990-2002 の活動

1995-2001 年の M2.5 以上のデータに時空間 ETAS モデルをあてはめ、実際の活動の相対的な活発化や静穏化を相対比率として時空間(経度, 緯度, 年)上の画像を得た。最初の 5 年間(1995-1999)の相対比率は大体の地域で 1.0 前後になっているが、特徴的なのは 1995 年兵庫県南部地震震源域に隣接する丹波地域と、常時活動の活発な和歌山地方の更なる相対的な活発化である。これらは、兵庫県南部地震の断層モデル(橋本(1996))に対する、当該地域の地震の最頻メカニズム(馬場 他(2002))によるクーロン破壊応力(CFS)変化と調和的である。特に注目すべきことに、両地域での活発化は兵庫県南部地震に先行しており、この断層やその深部のどこかで前駆すべりがあった可能性を考えたいところである。また断層より北部での S 波反射面の正断層のすべり仮説(飯尾(1996))による CFS 変化についても、前述の 2 つの地域の先駆的な活発化と調和的である。

2001 年を通した相対比率はこの年に例えば琵琶湖東岸の活動の低下、岐阜県中部や和歌山市海岸部周辺の活発化を示している。前者 2 地域はほぼ東西圧縮の地震メカニズムが卓越しており、そこで検出された地震活動の減増は、Ozawa et al.(2002)によって GPS 観測値の変化から求められた浜名湖付近直下のすべりを根拠に計算された当該地域の CFS 変化の増減に調和的である。

参 考 文 献

- 馬場俊孝, Cammins, P., 堀高峰, Townend, J., 宇平幸一(2002). 西南日本における震源メカニズムデータの統合: 応力場インバージョンに向けて, 地球惑星科学関連学会 2002 年合同大会予稿集, S041-P013.
- 橋本 学(1996). 兵庫県南部地震に伴う応力変化: 断層モデルによるクーロン破壊関数の変化の計算と地震活動の変化との比較, 地震(ii) 48, 521-530.
- 飯尾能久(1996). 兵庫県南部地震の発生過程 断層の固着とディタッチメントのすべり, 地震(ii) 49, 103-112.
- Ogata, Y.(1998). Space-time point-process models for earthquake occurrences, *Ann. Inst. Statist. Math.*, 50, 379-402.
- Ogata, Y.(2002). Modelling of heterogeneous space-time seismic activity and its residual analysis, Research Memo., No. 808, The Institute of Statistical Mathematics, Tokyo.
- Ozawa, S., Murakami, M., Kaidzu, M., Tada, T., Sagiya, T., Hatanaka, Y., Yarai, H. and Nishimura, T. (2002) Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan,

Science, **298**, 1009–1012.

確率伝搬法の解析と改善手法

池 田 思 朗

確率伝搬法とは、ベイジアンネットによって表現される確率分布から各ノードの周辺分布を求めるための計算手法である。グラフが木の構造ならば確率伝搬法によって効率良く正しい周辺分布を得られるが、グラフがループを持つ場合には繰り返し演算となり、収束性が保証されないだけでなく、仮りに収束したとしても得られる結果は正しい周辺分布ではないことが分っている。

しかしながらノードの数が 100 を超える大規模なグラフにおいては他に有効な計算手法がなく、近似解しか得られないにもかかわらず様々な分野で広く用いられている。例えば統計物理におけるペーテ近似、クラスタ変分法は確率伝搬法と同様の計算手法であり、1950 年代から広く研究がなされている。一方 1990 年代に登場したターボ符号といった誤り訂正符号においても、復号に際し確率伝搬法が用いられている。特にターボ符号は次世代携帯電話の誤り訂正技術として用いられることが決定している。

確率伝搬法の問題点は前にも挙げたように収束性と近似精度の評価である。実験的な検証については様々な研究が行なわれていて、実用上十分な結果が得られている。しかし理論的解析は満足いく結果が得られていない。これは既存の理論では捉えられない新しい解析手法が必要なことを示している。我々は情報幾何を用いてこの問題を扱うことを考えている。現在はターボ符号に関する初歩的な結果を得ており(池田 他(2002))、これを一般的な確率伝搬法に拡張し収束性と近似精度を向上させる手法を提案すべく、現在研究を行なっている。

参 考 文 献

池田思朗, 田中利幸, 甘利俊一(2002). ターボ復号の情報幾何, 電子情報通信学会論文誌 D-II, J85-D-II(5), 758-765.

じゃんけんソフトのその後

石 黒 真 木 夫

今千年紀の最初の年である西暦 2000 年に、21 世紀の科学技術分野での夢の実現に向けての営みを紹介する「夢テク」というイベントが開催された。統計数理研究所はエキジビションの一つとして、統計的思考というものを分かりやすい形で見せるじゃんけんソフトを出展したが、ここで得られたデータを解析してみた。

じゃんけんソフトは会場において、不特定多数のプレーヤーによって操作される状態であった。このゲームを一人のプレーヤーがどの位継続しておこなったかを見るために、ゲーム終了時の得点の最多得点の分布を調べてみると図 1 のようであった。ゲーム終了である 30 点までゲームが継続された場合が多い事が分かる。又、ゲームを途中で終了する行動に関しても納得のいく結果になっている。

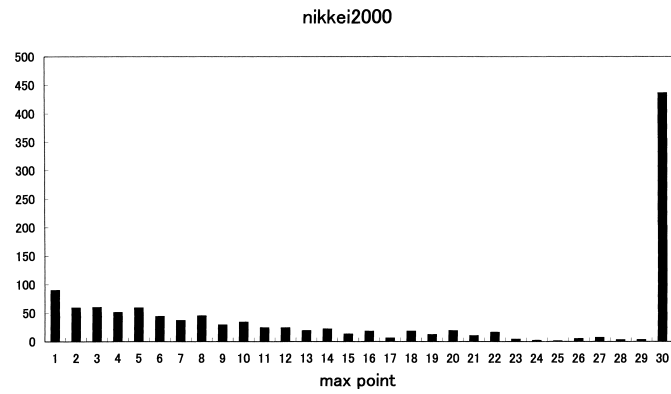


図 1. ゲーム終了時の得点の最多得点の分布 .

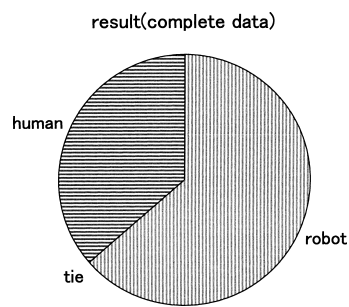


図 2. 勝敗の結果(縦縞が「ソフト」の勝率, 横縞が人間の勝率) .

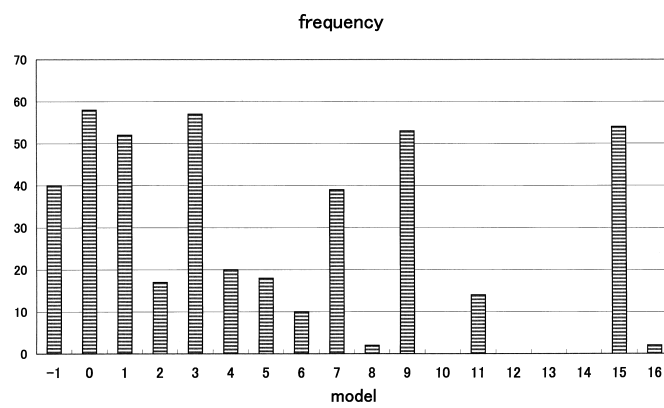


図 3. モデル出現頻度 .

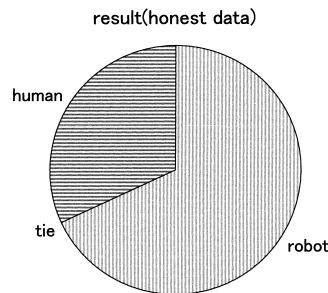


図 4. 「後出し」を除いた勝敗結果 .

このゲーム終了まで行われた結果についてソフト側と人間側の勝敗を調べてみたのが図 2 である . ソフトがかなり強い事が分かる . 統計的思考の有効性をアピールするという所期の目的は達成されたと考えていいだろう .

ソフトは人間の行動のモデルを有しており , プレイにあたって人間がどのように行動をしていたかをモデルあてはめによって判定する機能を備えている . この結果判定されたモデルの出現頻度は図 3 のようであった . 詳細は参考文献 (石黒・佐藤(2000)) にゆずるが , 「モデル 0」は人間のプレイヤーが「後出し」をしているとするモデルである . 統計的思考の一面を見せるとともにゲームにおいて計算機が後だしをしていない事を証明することを一つの目的として計算機の次の手をあらかじめ予告する裏画面が用意されていたので , 人間のプレイヤーが後だしをする事が可能だったのである . 図 3 によればかなりの数のプレイヤーが後出しをおこなっていたと判定される .

この後出しの場合を除いた場合についての勝ち負けの分布を求めたのが図 4 である .

参 考 文 献

石黒真木夫, 佐藤整尚(2000). 2 次元 3 値時系列としてのジャンケンデータの研究, Research Memo., No.759, The Institute of Statistical Mathematics, Tokyo.

重み付けをした変数による主成分分析

馬 場 康 維

主成分分析は多変量データの記述の方法の一つとしてよく知られている方法である . 一般に用いられる方法は ,

- 1) 中心化されたデータから主成分を求める方法 ,
- 2) 標準化されたデータから主成分を求める方法 ,

である . これらを , 変数ベクトルに重みを付けるという観点から見直してみると , 前者は変数に重みをつけずに分散共分散行列から主成分を求める方法 , 後者は , 変数にそれぞれの変数の標準偏差の逆数の重みをつけた変数の分散共分散行列から主成分を求める方法と解釈できる .

上記のいずれの場合も、重みは固定されている。ここで、一変数(あるいは複数)の変数に注目しこの変数に重みをつけ、主成分を求める場合を考える。重みを小さくすれば、感度分析に対応し、大きくしていけば強制分類に対応する。したがって、重みを変化させることによって、データに内在する様々な構造を浮かび上がらせることができる。

重みを変化させるという方法は、異なる時点で観測されたデータ行列を滑らかにつなぎ、主成分の変化を追跡することにも応用できる。ここでは、実例を交えながら、2 時点間の主成分の対応関係を示す方法について述べた。

参 考 文 献

- Baba, Y. and Nakamura, T. (2002). Time dependent principal component analysis, *Measurement and Multivariate Analysis* (eds. S. Nishisato, Y. Baba, H. Bozdogan, and K. Kanefuji), 155–162, Springer-Verlag, Tokyo.
- Baba, Y. and Baba, E. (2002). Principal component analysis based on weighted variables, *Proceedings of the 4th ARS Conference of the IASC*, 110–112.

外乱に対する多変数弱非線形系の応答

岡 崎 卓

外乱を受けて発展する多変数系を取り上げ、外乱の特性と系の応答確率密度との関係を出る限り解析的に表現する試みについて報告する。その要点は、系の非線形性が弱いとの前提に立って一般化 Fokker-Planck (GFP) 方程式を簡略化して解を求め、応答確率密度の一般的な表現を導くことにある。

1. 弱非線形系に対する一般化 Fokker-Planck (GFP) 方程式の表現
非ガウス外乱 W を受けつつ運動方程式

$$\frac{d}{dt}U = M(U) + \mu(W(t))$$

に従って発展する系変数 $U = (U_1, U_2, \dots, U_N)$ の確率密度 $f(U, t)$ は GFP 方程式の解として定まる。系の非線形性に起因する困難を克服するため、非線形性が弱い場合に限定すれば、 λ_A 、および λ_B をそれぞれ local な、および non-local な拡散係数として GFP 方程式は次の形を採る。

$$\begin{aligned} \frac{\partial}{\partial t}f(U, t) = & -D_U M(U)f(U, t) + D_U D_U \lambda_A(U, t)f(U, t) \\ & + D_U D_U \int dV \lambda_B(U, V, t)f(U - V, t) \end{aligned}$$

2. 拡散係数の表現

非 Gauss 外乱に対する系の応答を GFP 方程式によって求めるには、先ず拡散係数を外乱の統計と系の非線形構造に応じて定めねばならない。外乱の特性汎関数の cumulant 表示

$$\hat{G}[\phi] = e^{\frac{1}{2}h(1,2)i^2\phi(1\sim 2) + \frac{1}{6}h(1,2,3)i^3\phi(1\sim 3) + \frac{1}{24}h(1,2,3,4)i^4\phi(1\sim 4) + \dots}$$

$$\text{(ここに } h(1,2)\phi(1\sim 2) = \int \int h(t, s)\phi(t)\phi(s)dt ds \text{ 等と略記)}$$

と系の弱非線形性 $M(U) = -\alpha U + \gamma M_N(U)$ 、 $|\gamma| \ll 1$ により、2 種の拡散係数を外乱の cumulant

群と系の構造関数 $M_N(U)$ の関数として次の形に表すことができる．

$$\lambda_A(U) = e^{\gamma \lambda_{AN}(U)} \lambda_{AL}, \quad \lambda_B(U, V) = e^{\gamma \lambda_{BN}(U)} \lambda_{BL}(V)$$

3. GFP 方程式の解と系の応答

上記 GFP 方程式に Fourier 変換を施せば，定常状態における系応答の特性関数 $\hat{f}(k)$ を定める式が次の形に得られる．

$$\left(\alpha \mathbf{i} \frac{d}{dk} - \gamma M_N \left(\mathbf{i} \frac{d}{dk} \right) \right) \hat{f}(k) + \mathbf{i} k \left[e^{\gamma \lambda_{A1}(\mathbf{i} \frac{d}{dk})} \lambda_{A0} \hat{f}(k) + e^{\gamma \lambda_{B1}(\mathbf{i} \frac{d}{dk})} \lambda_{B0}(k) \hat{f}(k) \right] = 0$$

非線形度を表すパラメータ γ による摂動解を求めると， \hat{f} に対する次の表現が導かれる．

$$\begin{aligned} \hat{f}(k) &= \hat{f}_{Gauss}(k) \cdot \hat{f}_{mod}(k) \\ \hat{f}_{mod}(k) &= e^{\hat{f}_{modL}(k)} e^{\hat{f}_{modN}(k)} \end{aligned}$$

即ち非 Gauss 外乱を受ける 1 変数非線形系の応答密度 $f(U)$ は，Gauss 外乱に対する線形部分の応答 \hat{f}_{Gauss} に変調 \hat{f}_{mod} を施したものと表される．その変調は，外乱の非ガウス性に因る部分 \hat{f}_{modL} と系の非線形性に因る部分 \hat{f}_{modN} からなり，それぞれ次の Lagrange 型偏微分方程式の解として定まる（以下の式で，同一添字の反復は 1, 2, ..., N に亘る和を意味する）．

$$\begin{aligned} k_i \alpha_{ij} \frac{\partial}{\partial k_j} \hat{f}_{modL}(k) + k_i k_j \lambda_{BL\ ij}(k) &= 0 \\ k_i \alpha_{ij} \frac{\partial}{\partial k_j} \hat{f}_1(k) + \hat{f}_0^{-1}(k) \mathcal{M}(k) \hat{f}_0(k) &= 0 \\ \left(\begin{array}{l} \mathcal{M}(k) = \mathbf{i} k_i M_{Ni} \left(\mathbf{i} \frac{\partial}{\partial k} \right) \\ + k_i k_j \frac{(e^{\gamma \lambda_{AN}(\mathbf{i} \frac{\partial}{\partial k})} - 1)}{\gamma} \lambda_{AL\ ij} \\ + k_i k_j \frac{(e^{\gamma \lambda_{BN}(\mathbf{i} \frac{\partial}{\partial k})} - 1)}{\gamma} \lambda_{BL\ ij}(k) \end{array} \right) \end{aligned}$$

会計学と統計学

山下 智 志

従来，企業会計は過去の経営状態を金額ベースで数値化し，損益計算書や貸借対照表などの会計諸表を作成していた．しかし，資産価値の変動が企業の経営状態に大きく影響を与えるようになり，また企業倒産などにより貸倒れや売掛金の未回収などの損失が発生している．そのため企業会計に，将来の不確定要素を数値化して，企業がもつ様々なリスクを決算書に反映させる試みが始まっている．

最近提案された貸倒引当金の算出方法である DCF 法（ディスカウント・キャッシュ・フロー）は，このような考えに基づいたもので，貸出のデフォルトリスクを確率的に把握し，期待損失を求める方法である．しかし，将来のデフォルト確率を長期的に推計することは，現状では

データベースの不備などの理由により困難な状況にある。発表では DCF 法に関わる問題点と、解決策のバリエーションについて解説を行った。そのなかで、Cox 比例ハザードモデルが DCF 法と相性がいいが、正確な推定ができるようになるには、あと数年間企業財務データを蓄積しなければならないことを示した。

Molecular Evolution and Phylogenetics of Mangroves

(客員) Fudan University Yang Zhong

Mangroves form the dominant intertidal ecosystems throughout the tropical regions of the world and enter the subtropics. Because of the adaptive characteristics of the habit and morphology associated with the ecology of a marine environment, mangroves and their close terrestrial relatives have been extensively studied in terms of morphology, palynology, anatomy, ecology, and cytology. In this study, phylogenetic analyses of Rhizophoraceae, a major mangrove family, were conducted based on the sequences of the cpDNA *matK* and *rbcL* genes and nrDNA ITS regions using the maximum likelihood method. The phylogenetic trees constructed by both independent and combined analyses strongly supported that the mangrove genera formed a monophyletic group. The relative-rate tests also showed evolutionary rate heterogeneity in the Rhizophoraceae for the plastid *matK* and *rbcL* genes: the mangrove genus *Bruguiera* had relatively slow substitution rates than the terrestrial genus *Carallia* at both synonymous and nonsynonymous sites in the *matK* sequences, and the synonymous and nonsynonymous substitution matrices were correlated. However, the *rbcL* nonsynonymous sites exhibited a high degree of rate heterogeneity among mangroves and related terrestrial groups, and uncoupling of rates with the synonymous sites. Selection is likely an important influence on the rate variation, suggesting further investigation for better understanding of various forces contributing to the rate heterogeneity and molecular adaptation in mangroves.