統計数理 (2001)
第 49 巻 第 2 号 345–373
［統計数理研究所研究活動］

平成 12 年度研究報告会要旨

と　き：2001 年 3 月 13 日　午前 9 時 30 分〜午後 5 時 15 分
14 日　午前 9 時 30 分〜午後 4 時 45 分
ところ：統計数理研究所 講堂

プ ロ グ ラ ム

3 月 13 日（火）
午前（9 時 30 分〜12 時）

あいさつ　所長　清水　良一

【予測制御研究系】
インフレ率長期予測誤差の分析　北川源四郎
季調済系列の滑らかさについて　川崎　能典
哺乳類における収斂進化　長谷川政美
発見科学との接点　樋口　知之
スケール変換したブートストラップによる多様体の検定　下平　英寿
Innovation Approach to Dynamic Brain Imaging　尾崎　統
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
Ｒを使った並列計算　佐藤　整尚
A New Algorithm for Statistical Discrimination　田邉　國士
モンテカルロ法の研究と学際性　伊庭　幸人

午後（13 時〜17 時 15 分）
ｎ人のジャンケン　上田　澄江
食料の自給問題について　鈴木義一郎
大規模 2 次錐計画による磁気シールド設計最適化　土谷　隆
在庫変動と価格変動について　石黒真木夫
脳システムの行動学と信号処理　瀧澤　由美
ノンパラメトリック統計モデルと局所モーメント法　（客員，岐阜大学）寒河江雅彦

【統計科学情報センター】
重み付けをした変数による多変量解析　馬場　康維
順序統計量に基づいた 5 パーセント点の推定　金藤　浩司
外乱に対する弱非線形システムの応答　岡崎　卓
McFadden のノーベル賞とその応用例　山下　智志

【領域統計研究系】
臨床試験の方法から見る質の高い証拠　柳本　武美
Bayesian Generalized Linear Models and Extensions　汪　金芳

語順規則の Ising モデル 伊藤 栄明
文章の統計分析 村上 征勝
ハワイ日系・非日系人調査 吉野 諒三
仮想評価法 (CVM) のバイアス問題について 鄭 躍軍
可逆から不可逆へ （客員，東京大学）伊藤 伸泰

3 月 14 日（水）
午前（9 時 30 分〜12 時）

【統計基礎研究系】
On k-match Problems 平野 勝臣
多重線形形式の最大値の裾確率とその応用 栗木 哲
ジャンプをもつマルチンゲールに対するブラケティング中心極限定理 西山 陽一
一般相反分布とその周辺 松縄 規
頑健な独立成分解析法 南 美穂子
Local Regression Analysis in a Near-parametric Model 江口 真透
識別不能性を持つモデルの尤度比と錐型の特異点 福水 健次
Asymptotic Expansion for Stochastic Differential Equation with Jumps
（客員，東京大学）吉田 朋広

【調査実験解析研究系】
第 10 次国民性調査の成果と課題 補遺 坂元 慶行
日本型森林セクターモデル：持続的森林資源管理の可能性 吉本 敦

午後（13 時〜16 時 45 分）
社会調査データ解析をめぐる 2 つの課題 前田 忠彦
Poisson Voronoi Cell の統計分布 種村 正美
インターネットにおいて行なう公開の抽選についての考察 丸山 直昌
遺伝子型を含んだ樹木分布地図データの応用例 島谷健一郎
コウホート分析から見た調査の継続性 中村 隆
マラリア原虫の系統的位置の解析 橋本 哲男
定性情報のマイニング 大隅 昇
——テキスト型データ解析システム：WordMiner について——
電話調査法について 土屋 隆裕
Trimmed k-means 基準を用いた主要点について 清水 信夫
環境データの解析 柏木 宣久
Echelon 解析とその応用 （客員，岡山大学）栗原 考次

【統計計算開発センター】
統計解析とパソコンクラスター 田村 義保
密行列計算向け CPU の開発 泰地真弘人
TIMSAC72 の計算法について 荒畑恵美子
統計解析システム Jasp のグラフィカルユーザインタフェース 中野 純司

予測制御研究系

インフレ率長期予測誤差の分析

北　川　源四郎

我が国の消費者物価指数の前年同期比伸び率がどの程度の精度で予測可能かを時系列解析の立場から検討した．具体的には，国内総生産，有効求人倍率，鉱工業生産指数，TOPIX 終値，長期貸出金利などを用いた多変量 AR モデルや状態空間モデルによる 1 年（4 期）先予測誤差の大きさの評価を行った．

多変量 AR 型モデルに関しては，通常の多変量 AR モデルのほか，4 期先の物価指数の伸び率を直接表現する回帰モデルも考慮した．パラメータの推定には最小二乗法と 4 期先の予測誤差分散最小の二つの方法を，変数選択には MULMAR の方法と最適部分回帰モデルを選択する方法の二つを利用した．この結果，アウトサンプル予測精度が上位のモデルはすべて有効求人倍率，TOPIX 終値とその他の系列の差分からからなっており，0.45 程度の予測精度 (RMSE) が達成できることがわかった．

状態空間モデルに関しては MTCAR（個別トレンド共通 AR）モデルを新たに開発した．このモデルは各系列を個別のトレンドと時間遅れと係数を除いて共通の AR 成分で表現するものである．このモデルが想定する共通の循環変動は，フィリップス曲線やオーカンの法則といった，マクロ経済学における経験則のモデリングと対応していることから，経済理論に整合的な予測モデルの枠組みとして近年注目を浴びているものであるが，現時点では予測精度に関してはランダムウォークモデルと比較して有意な改善は得られていない．

参　考　文　献

北川源四郎, 川崎能典 (2001)． 時系列モデルによるインフレ率予測誤差の分析, 日本銀行 Working Paper 01–13.

季調済系列の滑らかさについて

川　崎　能　典

季節調整モデルの特徴付けに関連して，次に挙げる 3 つの問題を考察した．(Q1) モデルベースの季節調整法は X-11 等に比べて滑らかな季節調整済系列を生成するという指摘が，経験的分析を通じてしばしば行われる．これをどう説明したらよいだろうか．(Q2) DECOMP においては，「各成分を駆動するイノベーションの分散が観測ノイズを超えない」という制約の下でパラメータ推定が行われる．この制約はどのようにして正当化されるか．(Q3) ホワイトノイズではなく相関を持つノイズで駆動される季節成分モデルは形式的には既に提案されている．そうしたモデルを考えることの，実際的なメリットは何であろうか．また，このようなモデリングによって季節成分の変動は増すのだろうか，減るのだろうか．これらの問いに対する答えは，以下のように与えられる．(A1) 状態空間モデルにおいては，モデルの識別性の観点から，現在の季節成分と 1 年前の季節成分がほぼ等しいという差分型のモデルよりも，季節成分（正確には一周期に含まれる季節数から 1 引いただけの個数）を足し込んだものがほぼゼロという

和分型のモデルが採用される．季節調整フィルタのゲインを調べると，和分型の季節調整フィルタは低周波成分をより通しやすく，特にシステムノイズが大きいときにこの傾向は顕著に見られることがわかる．(A2) システムノイズを観測ノイズ以下に押さえ込むことで，和分型季節調整モデルが持っている low pass 傾向を抑制することができる．従って，季節成分に高周波ノイズが過度に振り分けられるのを防ぐ制約として有効である．(A3) ここでの相関を持つノイズは，AR(1) 過程から生成されるものとしよう．このようなモデルは，和分型の季節調整モデルを差分型の季節調整モデルに近づける効果を持つ．従って，システムノイズが大きいときに，季節調整フィルタの low pass バイアスを修正する役割を果たす．有色ノイズで駆動される季節調整モデルは，モデルの形式上一定パターンの季節成分を累積してゆく形になっていることから，より柔軟な季節性の表現に役立つと考えられがちである．実際はむしろ逆に，季節成分に高周波ノイズが混入することを防ぎ，より固定的な季節成分の抽出を行うものであることがフィルタ特性を吟味することでわかる．

スケール変換したブートストラップによる領域の検定

下 平 英 寿

未知母数がパラメタ空間のある領域に含まれるとする帰無仮説を検定する問題において，3 次の漸近精度をもつ不偏検定を構成した．すなわち，未知母数が領域の境界上にあるとき棄却確率を $O(n^{-3/2})$ の誤差で有意水準に等しくすることができる．ただし，ここではデータは多変量正規分布に従うと仮定しておく．

領域の境界の曲率，及び，境界からデータまでの符号付距離を使うと p–値の漸近的な表現が得られる．ところが曲率や符号付距離を実際の応用で得るのは難しいことが多い．これをスケール変換したブートストラップから効率よく推定するのが本方法のオリジナルな点であり，計算機プログラムへの実装は非常に容易である．領域は陽に与える必要がなく，データが領域に入っているかどうかを判定することさえできれば本方法を適用できる．これにより，モデル選択の信頼性を検定する問題や，classification の信頼性評価への応用のように，複雑な形状をもつ領域に適用できる．また，検定を反転することにより実数パラメタの信頼区間を 3 次の精度で計算することにも応用できる．

本方法の詳細は Shimodaira (2000) を参照．漸近理論は Efron (1985)，Efron and Tibshirani (1998) に基づいており，これは尤度比検定の Bartlett 補正と密接なかかわりがある．Efron et al. (1996) では 2 次の漸近精度をもつアルゴリズムが与えられていて，それが分子系統樹推定へ応用されている．これより漸近精度が良く計算も容易な方法を Shimodaira (2000) で与えている．

参 考 文 献

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems, *Biometrika*, **72**, 45–58.

Efron, B. and Tibshirani, R. (1998). The problem of regions, *Ann. Statist.*, **26**, 1687–1718.

Efron, B., Halloran, E. and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees, *Proc. Nat. Acad. Sci. U.S.A.,* **93**, 13429–13434.

Shimodaira, H. (2000). Another calculation of the p-value for the problem of regions using the scaled bootstrap resamplings, Tech. Report, No. 2000-35, Stanford University, California.

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．

この数年間は外乱や非線形成分，次数や相対次数に依存しない適応制御系の構成法の研究に関わり，外乱・非線形成分・次数，そして部分的に相対次数について従来の制約を緩和した適応制御手法を導出することができた．今年度は特に相対次数に 3 次の幅の不確定性がある場合や，対象の高調波利得の符号に不確定性がある場合でも対処できる適応制御系（モデル規範形適応制御系，適応安定化制御系）の構成手法についてこれまでの研究をまとめて，Miyasato (2000c, 2000e) の結果を得た．

これらと平行して，ハイゲインオブザーバを用いた構造の簡単な適応制御系の構成手法を求め，そのロバスト特性について解析を行い，研究成果をまとめて宮里 (2000) の結果を得た．

また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，数年前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．今年度は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，それらについて Miyasato (2000a, b, d, 2001) の結果を得た．さらにその手法を非線形パラメトリックモデルの非線形・適応制御に拡張することも考えている．その最初のステップとして，ニューラルネットを制御器の中に含んだ非線形適応 H_∞ 制御と，ニューラルネットの安定な学習則に関して萌芽的な結果を得ることができた．

参 考 文 献

宮里義彦 (2000). 構造の簡単な適応制御系の設計法とそのロバスト性, 計測自動制御学会論文集, **36**(5), 424–430.

Miyasato, Y. (2000a). General forms of adaptive nonlinear H^∞ control for processes with bounded variations of parameters, *Preprints of IFAC Symposium on System Identification* (*SYSID2000*).

Miyasato, Y. (2000b). Adaptive nonlinear H^∞ control for processes with bounded variations of parameters, *Proceedings of 2000 American Control Conference,* **1**, 297–301.

Miyasato, Y. (2000c). A model reference adaptive controller for systems with uncertain relative degrees r, $r+1$ or $r+2$ and unknown signs of high-frequency gains, *Automatica,* **36**(6), 889–896.

Miyasato, Y. (2000d). Adaptive nonlinear H_∞ control for processes with bounded variations of parameters —— General relative degree case ——, *Proceedings of the 39th IEEE Conference on Decision and Control,* 1453–1458.

Miyasato, Y. (2000e). A design method of universal adaptive stabilizer, *IEEE Trans. Automat. Control,* **45**(12), 2368–2373.

宮里義彦 (2001). 適応制御理論の回顧と展望, 計測と制御, **40**(1), 56–62.

Miyasato, Y. (2001). Adaptive nonlinear H_∞ control for processes with bounded variations of parameters —— General forms and general relative degree case ——, *Preprints of IFAC Workshop*

on Adaptation and Learning in Control and Signal Processing (*ALCOSP2001*), 419–424.

R を使った並列計算

佐 藤 整 尚

コンピュータの高性能化と低価格化によって，並列コンピュータでの統計計算が注目されている．しかしながら，ソフトウエアの並列化については，MPI (Message Passing Interface) 等を使って個別に対応しているのが一般的で，Fortran や C 言語を利用している．本報告では統計アプリケーションを直接，MPI 化することを提案した．ここで取り上げたソフトウエアは R と呼ばれる統計用のプログラム言語で，S 言語のフリー版である (http://www.ci.tuwien.ac.at/R/)．今回は MPI 部分の共有ライブラリ化によって，ソフトウエアのソースに手を加えない方法をとった．ここで提案した環境の中心は共有ライブラリとそれへアクセスするための R で書かれたドライバールーティンである．これにより，R(S) 言語のみの知識で並列計算が可能になる．ただし，MPI に関する知識は多少必要である．そうして，すべてのノードで同時に R が動き，MPI によって，通信することによって並列計算を実行するものである．また，この環境を使ったモンテカルロフィルタの計算例を示し，これをメモリ共有システムと Linux クラスターシステムで動かした時の実行速度の比較を行った．その結果，Linux クラスターは通信量が少ないときに有効で，通信量が多く，ノード数が多い場合はメモリ共有型のほうが有利であることがわかった．今回提案した方法は，R にとどまらず，他の統計ソフトにも応用可能なもので，このような拡張によって，並列計算がより手軽に行えることが望まれる．

モンテカルロ法の研究と学際性

伊 庭 幸 人

多変量の確率分布からのサンプリングや場合の数（エントロピー）の数値計算は，統計物理，統計学の諸領域，ニューラルネットや符号理論など，広い範囲の科学・工学の基盤となるものである．この分野では，異なる領域で類似したアルゴリズムが独立に開発されたり，ある領域で開発された手法が他の領域の突破口を与えることがしばしば起きている．マルコフ連鎖モンテカルロ法は後者の好例である．

講演では，モンテカルロ・フィルタをはじめとするポピュレーション型のモンテカルロ法が前者（「並行進化」）の例であることを示した著者のサーベイ論文 (Iba (2001)) を紹介した．また，マルコフ連鎖モンテカルロ法やポピュレーションモンテカルロ法の新たな適用領域として，魔方陣の数えあげ (Pinn and Wieczerkowski (1998)) をはじめとした組み合わせ論の諸問題が興味深いことを述べた．

参 考 文 献

Iba, Y. (2001). Population Monte Carlo algorithms, *Transactions of the Japanese Society for Artificial Intelligence*, **16**, 279–286.

Pinn, K. and Wieczerkowski, C. (1998). Number of magic squares from parallel tempering Monte Carlo, *Internat. J. Modern Phys. C*, **9**, 541–546.

在庫変動と価格変動について

石　黒　真木夫

（多次元）経済データを

$$y_t = T_t + S_t + X_t + R_t$$

の形にトレンド，季節変動，循環変動および誤差に分解して経済現象を理解する方法を研究してきた．

循環変動のモデル

$$X_t = \sum_{m=1}^{M} A_m X_{t-m} + \epsilon_t$$

が得られると，複数の時系列の間のダイナミクスを調べられる．

トレンドの間の関係にモデルを導入することによって，トレンドの間の関係が明示的に得られるとともに，モデルの過剰な自由度をおさえて循環変動モデルの精度が向上する．Rahman (2001) による 2 次元トレンド $T_t = (T_{1t}, T_{2t})$ のモデルの一つ

$$\begin{aligned} \Delta^2 T_{1t} &= v_t \\ T_{2t} &= \beta_{-1} + \beta_0 T_{1t} + \beta_1 \Delta T_{1t} + \beta_2 \Delta^2 T_{1t} \end{aligned}$$

が現実の在庫と価格のデータ解析に有効である事が確認された．

この方法をさまざまな業種のデータに適用することによって，循環変動がその業種の構造によって引き起こされるものなのか，外部「雑音」によって引き起こされるものなのか等，業種の個性を細かく調べることができる．

参　考　文　献

Rahman, Md. Moshiur (2001). Modeling trend of multiple nonstationary time series, PhD. Dissertation, Department of Statistical Science, The Graduate University for Advanced Studies, Tokyo.

ノンパラメトリック統計モデルと局所モーメント法

（客員）岐阜大学　寒河江　雅　彦

ノンパラメトリック統計モデルは局所的な標本情報に基づいた複数の局所統計モデルを統合したものと見ることができる．ノンパラメトリックな確率密度関数の中で最もよく知られてい

るカーネル型確率密度関数は全ての標本点をカーネル関数でモデル化した線形混合モデルである．ビン型標本情報（度数）を利用した推定法としてはビン型推定量と呼ばれ，ヒストグラム型密度関数がよく用いられている．ヒストグラム等のビン型推定量の特徴としては計算上の簡便さとモデルの簡単さがよく知られている．しかし，ビン型推定法はカーネル推定法に比べ，理論的な効率の悪さから多くの研究者の興味の対象外であった．しかし，近年，大規模データ処理のニーズと関心の高まりの中でビン型推定法の計算上のメリットも再評価されるべきと考える．また，理論面での進展も見られ，高次の効率をもつビン型推定量も近年提案 (Jones et al. (1998)，Minnotte (1998)，Sagae and Scott (1997)) され，新しい展開が期待される．ビン毎に集約されたデータに基づく推定法はこれまで度数データのみを利用したモデリングであった．Sagae and Scott はビン型情報を高次の局所モーメント情報まで広げたデータ集約化に基づいた局所モーメント法と Histogram を拡張した Polynomial Histogram を提案した．そして，高次の Polynomial Histogram は高次の漸近効率をもち，カーネル推定と同等な効率をもつことが示された．また，Sagae and Kogure (2000) では与えられた局所標本集約情報（局所モーメント情報を含む）からノンパラメトリック確率密度関数の形を同定する Local Maximam Entropy Density Estimator を提案し，理論的にも高次の漸近効率をもつことが示された．このモデルは予め確率密度関数の構造を仮定した従来の統計モデルとは異なり，与えられた様々な標本モーメント情報によって密度関数の構造が決まる可変的な構造を許す統計モデルである．最近，ノンパラメトリック密度推定法の分野で関心の高い局所尤度法と局所モーメント法の関係についてもこの論文の中で論じている．本研究報告では，同分野の研究の展望と著者の最近の当該分野の研究成果の報告を行った．

参 考 文 献

Jones, M. C., Samiuddin, M., Al-Harbey, A. and Maatouk, T. A. H. (1998). The edge frequency polygon, *Biometrika*, **85**, 235–239.

Minnotte, M. C. (1998). Achieving higher–order convergence rate for density estimation with binned data, *J. Amer. Statist. Assoc.*, **93**, 663–672.

Sagae, M. and Scott, D. W. (1997). Bin–interval method of locally adaptive nonparametric density estimation, Tech. Report, Department of Statistics, Rice University.

Sagae, M. and Kogure, A. (2000). A local maximum entropy density estimation, Cooperative Research Report in ISM, No. 134.

統計科学情報センター

重み付けをした変数による多変量解析

馬 場 康 維

データ行列の中の一つあるいは複数の特性ベクトルの重みを変えることにより様々な分析が行える．ここでは，下記の二つの場合を扱った．

1. 強制分類法の主成分分析への適用

一つの特性ベクトルの重みを変化させて主成分分析を行う．重みを大きくすると，この変数の影響の大きな主成分が得られ，小さくすると影響の少ない主成分が得られる．この操作を，

各特性ベクトルに適用して，変数空間の構造を把握できる．

2. 異なる時点間のデータ行列を各時点の行列に重みをつけることにより滑らかに変化させる．そうすることにより，異なる時点間の主成分，因子付加量などの関係が把握できる．

順序統計量に基づいた 5 パーセント点の推定

金 藤 浩 司

水環境における水質基準には，大きく分けると急性毒性と慢性毒性に関する 2 つの基準がある．2 つの基準の差は，毒性物質が生物に及ぼす影響の時間的な速さである．ここでは EPA（アメリカ合衆国：Environmental Protection Agency）で定められている急性毒性値の推定法の改良について研究した．その主な点は，提案されていた急性毒性値の推定方法において必要な測定時点数よりも少ない測定時点数で，急性毒性値の推定が十分であることを提案したことである．同時に，シュミレーションによりその有効性を検証した．また，従来の研究では，急性毒性値が従う確率分布として不自然な確率分布が考えられていた．そこで，急性毒性のデータが得られる基となる確率分布に関しての考察を行い，対数ロジスティック分布が急性毒性値の従う確率分布として適切であることを示した．

参 考 文 献

Erickson, R. J. and Stephan, C. E. (1988). Calculation of the final acute value for water quality criteria for aquatic organisms, U.S. Environmental Protection Agency, EPA/600/3-88/018.

Iwase, K. and Kanefuji, K. (2000). Estimation of the fifth percentile using a subset of order statistics, *Proceedings of the International Conference on Measurement and Multivariate Analysis*, Vol. 2, 218–220.

外乱に対する弱非線形システムの応答

岡 崎 卓

外乱を受けて発展する系の確率密度と外乱の非ガウス性との関係を明らかにするため，系の非線形性が強くないとの前提のもとに，外乱の統計的特性に基づき系応答確率密度を定める方法について報告する．

1. 弱非線形系に対する一般化 Fokker-Planck (GFP) 方程式の表現

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 の確率密度を定めるために開発された GFP 方程式は，外乱が非ガウス過程であっても成立する． しかし，その拡散項に存在する演算子 Super Translator (ST) は線形系では単なる変位演算子となるものの，非線形系にあっては複雑な構造をもち正確に表

現し難い．そこで系の非線形性が弱く，正規軌道 $\boldsymbol{U}(t)$ ($\frac{d}{dt}\boldsymbol{U} = M(\boldsymbol{U})$, $\boldsymbol{U}(0) = U$) の Jacobian ($J = \frac{\partial \boldsymbol{U}(t)}{\partial U}$) が

$$J(U + \epsilon V, t) = J(U, t) + O(\epsilon^2) \quad (\epsilon \ll 1)$$

なる場合を考えると，ST 演算子を具体的に書き下し GFP 方程式を実用に供し得る形に表現することができる．

2. 有色 Gauss 外乱に対する 1 変数弱非線形系の応答

系に加わる外乱が Gaussian ならば，外乱の相関を $\langle \mu\, \mu(s) \rangle = \frac{\sigma^2}{\beta} e^{-\beta s}$ として上記 GFP 方程式を解き，直ちに系の定常応答確率密度を得る．

$$f(U) = \text{const.}(\beta - M'(U))\, e^{-\frac{2\beta}{\sigma^2}(\frac{1}{2}M(U)^2 - \beta \int_0^U M(V)dV)}$$

3. 非 Gauss 外乱に対する 1 変数弱非線形系の応答

系が弱非線形であって，その構造関数 M が非線形度を表す γ により

$$M(U) = -\alpha U + \gamma M_1(U)$$

と表されるものとすれば，GFP 方程式の γ に関する摂動解から，定常応答確率密度の特性関数に対する次の表現が導かれる．

$$\left(\begin{array}{c} \hat{f}(k) = \hat{f}_{Gauss}(k) \cdot \hat{f}_{mod}(k) \\ \hat{f}_{Gauss}(k) = e^{-\frac{\lambda_{A0}}{2\alpha}k^2}, \quad \hat{f}_{mod}(k) = e^{\int^k d\omega\, \omega \hat{\lambda}_{B0}(\omega)} \cdot e^{\gamma \hat{f}_1(k)} \end{array} \right)$$

即ち，非 Gauss 外乱を受ける 1 変数弱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 $\hat{f}_{Gauss}$ (Gauss 分布) に，外乱の非ガウス性と 系の非線形性に基づく変調 $\hat{f}_{mod}$ を施したものとして得られることが判る（外乱の非ガウス性は $\hat{\lambda}_{B0}(\omega)$ を含む項に，非線形性の効果は $\gamma \hat{f}_1(k)$ の項に集約されている).

領域統計研究系

臨床試験の方法から見る質の高い証拠

柳 本 武 美

1. 臨床試験の方法

今日広く受け入れられている臨床試験は，通常の研究方法に比べて，形式化された手続きが多い．その結果しばしば「硬直した発展性に乏しい方法」とみなされることがある．一方で，既存のデータから本質的な情報を引き出すと宣伝される，新しい手法が山のように提案されている．これらの新しい手法に比べて，臨床試験では

1. 厳密な比較により評価する
2. 誤った仮説を否定する
3. 仮説と解析の方法を事前に設定する

ことが強調されている．これに比べて新しい方法では，専ら後智恵により自らが支持する仮説を肯定しようとしている．

2. 質の高いデータ

近年提案されている新しい方法と臨床試験の方法との間の大きな違いは，本質的な情報がありのままの自然の中に容易に見いだされるのか，人間の努力によって質の高い情報を含むデータを作り込むのかの違いに求められる．後者の立場に立てば，質の高い証拠を得るためには，目的に則して人間の努力を通してデータを作り出すことになる．言い換えれば，質の高いデータは自然界から採集するのではなく手間をかけて栽培することである．改めて強調するまでもなく，信頼のおける推論結果を得るためには，質の高いデータが必要である．臨床試験で開発されてきた厳密な手続きが，他の重要な問題に関連した科学的な推論の際にも必要となる．

このテーマに関連して，本年度は 8 月を除く各月に研究会を開催した．また総研大の共同研究グループでの会合でその一部を報告した．

Bayesian Generalized Linear Models and Extensions

汪　金芳

この報告では次のようなベイズ的一般化線形モデル (BGLM)

$$p(y_j|\theta) \propto \exp\{y_j\theta_j - b_j(\theta_j)\} \tag{1}$$

$$g(\mu_j) = x_j^T\beta \quad [\mu_j = EY_j = b_j'(\theta_j)] \tag{2}$$

$$\mu_j \propto \pi(\mu_j|\lambda) \tag{3}$$

についての考察を行った．BGLM を構成する (1) と (2) は通常の一般化線形モデルである．(3) 式は構造パラメータ β の事前分布に関する仮定であり，λ はハイパー・パラメータである．重みつき最小 2 乗法を一般化し，パラメータの推定は次の反復アルゴリズム

$$\beta_1 = (X^TWX)^{-1}X^TWU \tag{4}$$

を用いればよい．ただし，(4) 式における X はデザイン行列で，W は適当な重み行列である．また，最適ハイパー・パラメータ λ の選択は次の交差確認法によって得られる基準

$$\begin{aligned}\mathrm{ACV}(\lambda) = &\sum_{j=1}^{n}\log f(y_j|\hat\theta_\lambda) + \sum_{j=1}^{n} s^t(y_j|\hat\theta_\lambda) \\ &\times\{\dot s_n(\hat\theta_\lambda|\lambda) - [\dot s(y_j|\hat\theta_\lambda) + \dot s_j(\hat\theta_\lambda|\lambda)]\}^{-1}\{s(y_j|\hat\theta_\lambda) + s_j(\hat\theta_\lambda|\lambda)\}\end{aligned} \tag{5}$$

に従って行えばよい．

BGLM のセミ・パラメトリックの場合への拡張や，尿路感染症データへの適用については，汪 (2000) を参照されたい．

参 考 文 献

汪　金芳 (2000)．ベイズ的一般化線形模型におけるハイパー・パラメータの選択について，シンポジウム「統計科学における予測の可能性と限界に関する研究」報告集，千葉大学，2000.9.18–19.

語順規則の Ising モデル

伊 藤 栄 明

語順規則において，側置詞（前置詞・後置詞）が重要であることは従来より指摘されてきた．世界の 130 の言語の 19 項目についての角田による語順のデータを数値化し，階層クラスタ分析をもちいると 130 の言語は前置詞をもたない言語（後置詞をもつ言語と無側置詞言語）と前置詞をもつ言語に 2 分割された．さらに数詞と名詞の順序を考慮すると階層クラスタ分析の結果が自然に理解できることがわかった．

19 項目の語順のうち 7 項目が側置詞と関連が強い．日本語の語順を + とすると後置詞言語は側置詞およびそれと関連の強い 7 項目の全部が + になる傾向があり，英語等の前置詞言語は全部 − になる傾向がある．世界の言語は前置詞型構造と後置詞型構造のあいだをランダムに変動すると考え，有限格子の Ising モデルに類似の確率モデルを提案しデータと比較した．

参 考 文 献

Tsunoda, T., Ueda, S. and Itoh, Y. (1995). Adpositions in word order typology, *Linguistics*, **33** (4), 741–761.

上田澄江， 伊藤栄明 (1995). 語順規則による言語の分類と 2 パラメーターモデル，統計数理，**43**, 341–365.

Ueda, S. and Itoh, Y. (2001). Classification of natural languages by word ordering rules (投稿中).

ハワイ日系・非日系人調査

吉 野 諒 三

本研究の目的は，文化の伝搬変容の統計科学的解明のために，意識の国際比較調査により，1. 海外の日系人・非日系人と日本人の比較，2. 日系一世，二世，三世以上の世代間の比較，3. 他の諸国との比較，4. 継続調査データに基づく時系列比較，さらに，5. 我々が収集してきた国際比較調査データの一般公開を図り，世界の相互理解を通じた平和的発展の一助に供することである．

平成 12 年度は，前年度のハワイ・ホノルル日系・非日系人の調査結果を集計・分析した成果を，研究代表者（吉野）のもとで報告書としてまとめる作業を中心として，計画を遂行した．特に，8 月にはハワイにおいて，分析結果の日米合同討議（日本側共同研究者を派遣，米側の共同研究者を招聘）を行い，報告書の焦点についての検討を行った．

今回の調査では，我々が 1971 年以降に幾度か遂行してきたハワイ調査とは社会事情がかなり変化し，調査環境の著しい悪化が見られた．これは，標本抽出に用いた選挙人名簿には，過去 2 回の国政選挙時の投票者の名前が掲載されるようになっているが，近年，ハワイの経済状況が悪く，米国本土への移住者が増えていて，結果として「引越し」が増加し，適正な抽出が困難であったこと，また調査当時に現地の日系人の起こした殺人事件の影響などのため，安全への不安による「回答拒否」が増加したこと等が考えられる．

今後，調査結果の詳細な分析を推進すべきであるが，世界全体が「伝統的社会システム」から「高度情報社会」へ移行する過渡期にあるためか，日系人社会においても，様々な側面で伝統的システムの崩壊に伴う「信頼感の低下」の影響が生じているようである．

これらの結果は，研究代表者（吉野）の所属機関を中心に，「ハワイ日系・非日系人調査報告書」を統計数理研究所リポートとしてまとめ，発刊の準備作業を推進し，近く完了する予定である（文化の伝搬変容の統計科学的研究 ── ハワイ日系人・非日系人国際比較調査 ──，No. 86, として 2001 年 3 月刊行済み）．

仮想評価法 (CVM) のバイアス問題について

鄭　躍軍

仮想評価法 (CVM: Contingent Valuation Method) は環境資源の利用価値と非利用価値をともに評価できる数少ない表明選好法の一つとしてもっとも注目を集めている．しかし，CVM における非標本抽出バイアスが評価結果に大きな影響を及ぼすことが指摘されている．したがって，戦略的バイアス，仮想市場の設定バイアス，付値方式バイアスと支払手段バイアスの発生メカニズムを統計学的観点から総合的に解明することが CVM の信頼性を高める一つの鍵となっている．そこで，本研究では CVM の評価バイアスならびにその解決策を理論的・実証的に研究することを続けてきた．

本研究では，まず仮想市場の設定バイアスと支払手段バイアス問題を取りあげ，CVM の信頼性に与える影響について理論的に分析した上で，発生メカニズムの各種シナリオを作り出した．次いで，バイアス問題の実証的な考察を行うために，東京湾防波堤内側埋立地の「海上森林公園」の価値を評価するための仮想評価調査を遂行した．ここで，とりわけ実際の調査データに基づき，仮想市場の設定バイアスと支払手段バイアスのメカニズムを総合的に考察した．本調査に当たっては，「中央防波堤内側埋立地」の家庭ゴミに関わりのある東京都 23 区のうち，まず 8 区を抽出して，各区の住民基本台帳より約 800 人に 1 人の割合でサンプル個人を無作為に抽出する作業を行った．このサンプリングによって抽出した 2,649 サンプル個人を被調査対象とした．「中央防波堤内側埋立地」において，「海上森林公園」をつくる事業への賛否を質問した．さらに，上の質問で「賛成」を表明した場合，海上森林公園造成の初期段階，植栽基盤整備のための資金に対する，支払意志額をダブルバウンド二項選択方式により質問した．支払手段については，「海上森林造成基金への募金」と「税金による海上森林造成」のニつのシナリオで調査票を作成し，サンプル個人全体に郵送調査法で調査を行った．

調査データを統計的に解析し，仮想市場の設定バイアスと支払手段各種バイアス問題を検証した結果，ダブルバウンド二項選択方式による支払意志額 (WTP) の推定値は，シングルバウンド二項選択方式より大きいことと，「基金」と「税金」の支払手段による WTP には顕著な差があることを明らかにした．関連するバイアス問題の解決策を探ることは今後の重要な課題の一つである．

可逆から不可逆へ ── 非平衡緩和法 ──

（客員）東京大学　伊藤伸泰

物質がどのような状態をとり，その状態がどのような性質を持っているかを理論的に記述し予測する学問は統計力学と呼ばれる．数多くの原子分子からなる物質を扱うため，統計力学は個々の要素の運動を記述する力学理論と統計的な記述とを合わせた理論体系となっている．19

世紀末以来のこれまでの研究から，物質の状態・性質は，その物質を構成する要素がとりうる状態の全体の上の平衡分布（ボルツマンあるいはギブス分布とも呼ばれる）により精密に記述されることが明らかにされている．

平衡分布では，各状態はそのエネルギーの値でしか区別されない．エネルギーが同じならば同じ重みで考慮される．平衡分布で記述される状態（平衡状態）は時間変化せず，外からの作用が変化しない限り不変である．外からの作用により変化した後も各状態はエネルギーのみで区別され，各状態の履歴にはよらない．このため状態の変化は一般には不可逆となる．

可逆な運動方程式に基いて運動が履歴を持ち続ける力学理論と，この不可逆性とは相容れない．現実には，要素が増えてくると運動方程式の可逆性の根拠となる運動の履歴自身が不安定となり，だんだんと不可逆となってしまうのである．

構成要素間が相互作用をしている問題の重要性は，ますます高まっている．しかしこのような場合の平衡分布を手計算で解析することは困難で，計算機による解析が強力である．計算機による統計力学研究は，1950 年代の黎明期より今日まで 4 年で 10 倍の成長を遂げ続けており，21 世紀の今日，無くてはならないものとなっている．計算機による研究戦略は，長時間平均が平衡分布となる動力学あるいはモンテカルロシミュレーションにより平衡分布を実現し，解析するというものである．とはいえ最大級の計算機でも歯が立たない問題も未だ珍しくはない．

このため異なる研究戦略の確立が待たれており，その有力候補の 1 つが「非平衡緩和法」と呼ばれる方法である．これまではシミュレーションが完全に不可逆な平衡状態に到達するのを待ってから解析を行ってきたが，非平衡緩和法は平衡状態の兆から必要な解析を目指す方法である．これまでの研究から，自由エネルギーの 1 階微分の量を使って相図を解析する方法は確立されたといえよう．平成 12 年度は，自由エネルギーの 2 階微分の量（ゆらぎ）を使って臨界現象を解析する手法の研究を進めた．その結果，種々の臨界指数を直接評価する方法に目途がついた．

強磁性を例として説明しよう．（ゆらぎをもたない）完全秩序状態を初期状態としてシミュレーションを行うと，秩序変数である磁化とエネルギーのゆらぎがシミュレーション時刻とともに成長する．相転移点以外では（1 自由度当りの）ゆらぎは一定値に収束するが，相転移点では時刻とともに巾的に成長を続ける．この成長の巾から動的なものもふくめて臨界指数が評価出来ることが明らかとなったのである．

最後に非平衡緩和法の持つ長所をまとめて紹介しよう：

- 平衡状態を完全に実現する必要がないため，計算時間が短い．
- 相転移点も含め，系の大きさが有限であることによる効果をなくし，無限に大きい系の振舞いを直接観察する事が容易である．
- 解析手続きが簡明なため結果の信頼性が高く，検証も容易である．

統計基礎研究系

On k-match Problems

平 野 勝 臣

本年度の研究

1. $\{1, 2, \ldots, \mu\}$–値系列において，長さ m のあるパターンが起こるまでの待ち時間分布の

性質をまとめた (Aki and Hirano (2000)).

2. k-match problems と呼ばれる問題に確率生成母関数の方法を用いていくつかの結果をまとめた (Hirano and Aki (2001)).

On k-match problems

壷の中に，相異なる m 色のボールが入っている．1 個ずつ復元抽出で取り出し，色の番号 $1,\ldots,m$ を記録して行く．このようにしてできる $X_1, X_2, \ldots$ は i.i.d. $\{\ 1,2,\ldots,m\ \}$–値確率変数列である（1 回の試行でひとつの色が記録される確率は $1/m$ である）．k を正の整数とし，固定する．試行 i で "k-match が起こる" とは，X_i とその直前の k 回の試行結果のうちのどれかが一致することをいう (Arnold (1972))．試行 i で "k-bimatches が起こる" とは，この i 番目の試行結果がその直近の過去 k 回の試行結果のうちのどれかと一致し，且つ，この k 回の試行結果の中に 1 組の同一の試行結果があることをいう．

ところで，確率母関数は離散分布の厳密分布を導出する有力な方法である．そこで，この方法を用いて，1st k-match，2nd k-match，1st k-bimatches の待ち時間の厳密分布を導出した．

確率生成母関数を用いて解く方法は次の通りである．$\phi(t)$ を 1st k-match が起こるまでの待ち時間の分布の確率母関数とする．$\phi_i(t)$ を今まで i 回の試行を行っていて，これまでに match は起こっていないという条件の下で，これからはじめて match が起こるまでの待ち時間の条件付分布の確率母関数とする．そこで $\phi(t)$, $\phi_i(t), i = 1,2,\ldots, \min\{m,k\}$ についての方程式系を作り，これを解いて $\phi(t)$ を得る．また $\phi(t)$ を $t=0$ で展開することによって確率を求めることができる．

同様の方法によって 2nd k-match, 1st k-bimatches の待ち時間分布についても調べることができる．さらに，この方法を用いることによって系列を高次マルコフ系列に拡張したときもこれらの問題を調べることができることを注意したい．

なお，本報告は Hirano and Aki (2001) に基づいている．応用や他の問題との関係，および k-match problems に関する最近の研究など，詳細についてはこれを参照されたい．

参 考 文 献

Aki, S. and Hirano, K. (2000). On waiting time for reversed patterns in random sequences, Research Memo., No. 752, The Institute of Statistical Mathematics, Tokyo.

Arnold, B. C. (1972). The waiting time until first duplication, *J. Appl. Probab.,* **9**, 841–846.

Hirano, K. and Aki, S. (2001). On k-match problems, *J. Statist. Plann. Inference* (to appear).

多重線形形式の最大値の裾確率とその応用

栗 木 哲

z を各成分が独立に $N(0,1)$ に従う $p = q^k$ 次元の確率ベクトルとする．h を q 次元係数ベクトルとし，z の k 次形式 $(\underbrace{h \otimes \cdots \otimes h}_{k})'z$ を考える．ただし $\otimes$ はクロネッカー積である．またより一般に z の次元を $p = \prod_{i=1}^{k} q_i$，h_i の次元を q_i として，k 重線形形式 $(h_1 \otimes \cdots \otimes h_k)'z$ を考える．Kuriki and Takemura (2001) は tube 法（積分幾何に基づく正規確率場の分布論）を用い

ることにより，制約 $\|h\| = 1$，$\|h_i\| = 1$ の下での z の k 次形式，k 重線形形式の最大値，およびそれらの $\|z\|$ による基準化

$$T_k = \max_{\|h_i\|=1} (h_1 \otimes \cdots \otimes h_k)' z\,, \quad \tilde{T}_k = \max_{\|h\|=1} (h \otimes \cdots \otimes h)' z\,,$$

$$U_k = T_k / \|z\|\,, \quad \tilde{U}_k = \tilde{T}_k / \|z\|$$

の上側確率評価式を与えた．

ここではこれらの確率変数の意味あいとその応用について述べる．まず T_2^2 は Wishart 行列 $W(I_{q_2}, q_1)$ の最大固有値である．U_2^2 は T_2^2 と同じ Wishart 行列の最大固有値を，その trace で割ったものである． Johnson and Graybill (1972) は，繰り返しのない 2 元配置における交互作用の検定を構成したが，その検定統計量の帰無分布は U_2^2 の分布に帰着する．Johnson-Graybill の方法は繰り返しのない k 元配置の場合に拡張することができるが，その場合の検定統計量の帰無分布は U_k^2 の分布に等しい．$\tilde{T}_2$ の分布は $q \times q$ 対称正規分布行列の最大固有値の分布である．$\tilde{T}_k$，$k = 3, 4$，の分布は，多変量正規性に関する Malkovich and Afifi (1973) の検定統計量の極限分布である．

参 考 文 献

Johnson, D. E. and Graybill, F. A. (1972). An analysis of a two-way model with interaction and no replication, *J. Amer. Statist. Assoc.*, **67**, 862–868.

Kuriki, S. and Takemura, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications, *Ann. Statist.*, **29**, 328–371.

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality, *J. Amer. Statist. Assoc.*, **68**, 176–179.

ジャンプをもつマルチンゲールに対するブラケティング中心極限定理

西 山 陽 一

μ^n はマーク空間 E をもつマーク付き点過程であるとし，ν^n はそのカンペンセイターであるとする．Ψ を任意の集合とし，これによって添え字付けられた予測可能関数の族 $\{W^{n,\psi} : \psi \in \Psi\}$ が与えられたとする．確率過程 $t \rightsquigarrow X_t^{n,\psi}$ を

$$X_t^{n,\psi}(\omega) = \int_{[0,t]\times E} W^{n,\psi}(\omega, s, x)(\mu^n(\omega; dsdx) - \nu^n(\omega; dsdx))$$

によって定義する．これをパラメータ t および ψ をもつ確率場 $(t, \psi) \rightsquigarrow X_t^{n,\psi}$ とみなし，$n \to \infty$ としたときの漸近挙動を考える．まず，2 次共変量の収束および Lindeberg 条件の成立を，通常のマルチンゲール中心極限定理のときと同様に仮定する．つぎに，“quadratic modulus” という量を新たに導入し，これが確率有界であることを仮定する．さらに，パラメータ集合 Ψ に対しエントロピー条件を仮定する．このとき，確率場の列 $(t, \psi) \rightsquigarrow X_t^{n,\psi}$，$n = 1, 2, \ldots$ が $\ell^\infty([0, T] \times \Psi)$ の中において正規確率場に分布収束することが証明される．この定理は，Donsker の定理および Ossiander の定理の一般化にあたる．

参 考 文 献

Nishiyama, Y. (2000a). Entropy methods for martingales, *CWI Tract*, **128**.

Nishiyama, Y. (2000b). Weak convergence of some classes of martingales with jumps, *Ann. Probab.*, **28**, 685–712.

一般相反分布とその周辺

松 縄 規

正の範囲に分布する連続型分布の密度関数の主要部分に, 互いに牽制して定義域全体での積分の発散を防ぐ, 相反性を持つ分布族は統計学や関連分野において興味深い. そこで, 次の密度関数を持つ一般相反分布を考え, その分布関数の誘導を中心に考察した: $x>0, a>0, b>0, \gamma \in R-\{0\}$, $f(\cdot)>0$ に対し

$$
\begin{aligned}
p_{a,b,\rho,\gamma}(x) &= x^{\rho-1} f\left(\frac{1}{2}(x^{\gamma}/a + b/x^{\gamma})\right) \Big/ \int_0^{\infty} x^{\rho-1} f\left(\frac{1}{2}(x^{\gamma}/a + b/x^{\gamma})\right) \cdot dx \\
&= \begin{cases} \dfrac{|\gamma| x^{\rho-1} f\left(\frac{1}{2}(x^{\gamma}/a + b/x^{\gamma})\right)}{a^{\rho/\gamma} \displaystyle\int_0^{\infty} w^{\rho/\gamma-1} f\left(\frac{1}{2}(w + (b/a)/w)\right) \cdot dw} & (\rho \neq \gamma), \\ \dfrac{|\gamma| x^{\gamma-1} f\left(\frac{1}{2}(x^{\gamma}/a + b/x^{\gamma})\right)}{2a \displaystyle\int_{-\infty}^{0} f\left(\sqrt{z^2 + b/a}\right) \cdot dz} & (\rho = \gamma). \end{cases}
\end{aligned}
$$

この密度関数に対する, 累積分布関数が次のように求まった:

$$
F_{a,b,\rho,\gamma}(x) = \begin{cases} J_{\rho,\gamma}(x)/J_{\rho,\gamma}(\infty) & (\rho \neq \gamma), \\ J_{\gamma,\gamma}(t) \Big/ \left\{ 2|\gamma|^{-1} a \displaystyle\int_{-\infty}^{0} f\left(\sqrt{z^2 + b/a}\right) \cdot dz \right\} & (\rho = \gamma). \end{cases}
$$

ここに

$$
J_{\rho,\gamma}(t) = \begin{cases} \left[\begin{array}{l} -\displaystyle\int_{\frac{1}{2}(t^{\gamma}/a + b/t^{\gamma})}^{\infty} \left(y - \sqrt{y^2 - b/a}\right)^{\rho/\gamma-1} f(y) \cdot dy \\ +\displaystyle\int_{-\infty}^{\frac{1}{2}(t^{\gamma}/a - b/t^{\gamma})(<0)} \left(\sqrt{z^2 + b/a} + z\right)^{\rho/\gamma-1} f\left(\sqrt{z^2 + b/a}\right) \cdot dz \end{array} \right] & \left(0 < t^{\gamma} \leq \sqrt{ab}\right); \\ \left[\begin{array}{l} -\displaystyle\int_{\sqrt{b/a}}^{\infty} \left(y - \sqrt{y^2 - b/a}\right)^{\rho/\gamma-1} f(y) \cdot dy \\ +\displaystyle\int_{\sqrt{b/a}}^{\frac{1}{2}(t^{\gamma}/a + b/t^{\gamma})} \left(y + \sqrt{y^2 - b/a}\right)^{\rho/\gamma-1} f(y) \cdot dy \\ +\displaystyle\int_{-\infty}^{0} \left(\sqrt{z^2 + b/a} + z\right)^{\rho/\gamma-1} f\left(\sqrt{z^2 + b/a}\right) \cdot dz \\ +\displaystyle\int_{-\frac{1}{2}(t^{\gamma}/a - b/t^{\gamma})}^{0} \left(\sqrt{z^2 + b/a} - z\right)^{\rho/\gamma-1} f\left(\sqrt{z^2 + b/a}\right) \cdot dz \end{array} \right] & \left(t^{\gamma} > \sqrt{ab}\right). \end{cases}
$$

ここで得た結果から，逆ガウス分布，逆数逆ガウス分布等の分布関数を組織的に誘導できることを示し，関連事項について触れた．

頑健な独立成分解析法

南　美穂子

独立成分解析は，複数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に，その観測された信号 $\boldsymbol{x}$ からもとの互いに独立な信号を復元すること，具体的には，$W\boldsymbol{x}$ の各成分が独立になるような正則行列 W を求めることを目的とする．ただし，独立成分の順序とスケールは不定である．また，識別可能性の要請から正規分布に従うものは高々一つでなければならない．原信号はその分布も未知であり，特定の分布を仮定せずに解析を行う．本報告では，独立成分解析のなかでも，時間的混合のない Blind source separation と呼ばれる問題を取り扱う．

各成分が独立であるとは，結合密度が周辺密度の積で表せるということであるから，推定の基準としては，結合密度と周辺密度の積との何らかの距離（隔たり）を用いる．ただし，密度は未知であるから適当な密度関数を用いて推定量を求め，その性質を議論する際は，密度を未知とするセミパラメトリックなモデルの下で行う．

今までに提案されている多くの推定量は，隔たりを測る尺度として Kullback-Leibler ダイバージェンスを用いて導出したものとして捉えることができ，推定関数は同じ形で表現できる．これらは独立成分の期待値が 0 であるという仮定の下で一致性を持つことが示されるが，どのような密度関数を用いても推定関数は有界にならず，外れ値によって推定値が大きく左右されるという欠点を持つ．また，実際には観測信号の期待値は 0 ではなく，算術平均を引くという前処理をしているにもかかわらず，理論的な解析の際にはそれが考慮されていない．

本報告では，β ダイバージェンスを推定の基準とし，シフトパラメータをモデルに含めた，外れ値の影響が少ない頑健な推定量を提案する．β ダイバージェンスは分布間の隔たりを測る尺度で，非負の値をとり，0 となるのは 2 つの分布が等しいときのみである．β は 0 以上の値をとり，$\beta = 0$ の場合は，Kullback-Leibler ダイバージェンスとなる．この推定量は一致性をもち，$\beta > 0$ の場合は，用いる密度関数がいくつかの条件を満たせば推定関数は有界となり頑健性を持つ．通常用いられる密度関数はこの条件を満たしている．漸近安定となるための必要十分条件と計算アルゴリズムも示した．4 種類のデータセットに対してシミュレーションを行ったが，いずれの場合も，従来の方法では外れ値や異質のデータに推定値が大きく影響されるが，本研究で提案した推定量は，β を大きくするにつれ，外れ値や異質のデータに影響されなくなった．

Local Regression Analysis in a Near-parametric Model

江 口 真 透

回帰関数の推定において局所尤度法について発表した．ノンパラメトリックとパラメトリックの推定の理論の統一的理解を得るために次の仮定 α が成された： 真の回帰関数のパラメトリックモデル関数への射影の長さが標本サイズ n の $(-1+\alpha)$ のべき乗である．

この仮定において α は -1 から ∞ まで値をとり，その両端においてちょうどノンパラメトリックとパラメトリックな仮定を表している．主結果として，仮定 α の下で局所尤度法の最適

バンド幅は α の符号によって，漸近的に ∞ と 0 へ分岐されることが示された．

識別不能性を持つモデルの尤度比と錐型の特異点

福 水 健 次

混合モデルやニューラルネットワーク，ARMA など複雑なパラメトリゼーションを持つ統計モデルでは，あるひとつの分布を表すパラメータ集合が連続集合になることがあり得る．このような現象は，パラメータの識別不能性と呼ばれる．混合モデルやニューラルネットの例では，設定したモデルのサイズ（コンポーネント数や中間素子数）より小さいサイズで実現可能な関数を定めるパラメータは識別不能となる．サンプルを発生している真のパラメータが識別不能だという仮定のもとで最尤推定量の挙動を考えると，当然，漸近正規性などは成立しなくなり，モデルサイズの検定やモデル選択をはじめとする多くの統計的手法に特別の考察が必要となる．

本研究では，識別不能性を持つモデルの定式化として局所錐型モデルを用い，真のパラメータが識別不能な場合に，最尤推定の尤度比の漸近特性を考察した．識別可能なモデルにおいては，ある正則条件のもとで，尤度比はパラメータ数の自由度を持つカイ 2 乗分布に法則収束することはよく知られている．一方，識別不能性を持ついくつかの例においては，サンプル数を無限大にしたとき尤度比が発散する現象が知られている．そこで特に，局所錐型モデルにおいて，サンプル数に対する尤度比のオーダーについて研究を行った．

局所錐型モデルでは，真の密度関数に相当する点は錐型の特異点となっており，その近傍の様子は接錐によってよく記述される．本研究の結果，接錐を構成する L2 ノルム 1 の関数族が 0 に確率収束する関数列を含むことが，最尤推定の尤度比が発散するための十分条件となることを示した．この結果は，具体的な統計モデルに関してチェックしやすい十分条件を与えており，従来個別に考察されてきた尤度比の発散現象が一般的に理解できるようになった．実際この十分条件を用いると，3 層パーセプトロンや正規混合モデルに関して，真の分布がモデルよりも小さいサイズで実現できる場合に尤度比が発散することが簡単に示せる．

さらに，3 層パーセプトロンに関して，サンプル数に対する尤度比のオーダーを詳細に調べた．その結果，真の関数がモデルよりも 2 個以上少ない中間素子数で実現可能であれば，ノイズモデルに関するある種の条件のもと，サンプル数 n に対する尤度比のオーダーは $\log n$ 以上であることが示された．今後，尤度比の正しいオーダーやその漸近分布などを研究していきたいと考えている．

調査実験解析研究系

第 10 次国民性調査の成果と課題 補遺

坂 元 慶 行

今年度は，拡張情報量規準 EIC の性能についての検討と発表を除けば，研究活動の中心は，昨年度に引き続き，1998（平成 10）年の「第 10 次 日本人の国民性調査」の結果の分析と発表

であった．

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 つの目的があり，それぞれについて課題があるが，まず重要なのは，日本人の意識動向，つまり，21 世紀の意識の動きを捉えるための質問文の開発である．1953（昭和 28）年以来の質問には回答の変化が縮小し時代の動きを測る機能を失ってしまった質問も少なくなく，新しい質問文の開発が急務であるからである．年度研究発表会ではこの面での課題の若干の整理を試みた．

筆者は，戦後日本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るが，この特徴を最も典型的に示す調査結果が “ 一番大切なのは家族 ” という意見の 1970 年代以降の激増である．この調査結果は，前回 1993（平成 5）年調査の結果発表以来総じて好感されてきたが，いくつかの問題点も指摘された．その第 1 は，この質問が自由回答法を採用していることに起因する “ 家族 ” の選択率の確かさであり，第 2 は，“家族” の激増の意味である．

第 1 の問題に対しては，1998（平成 10）年調査の結果発表に際して過去 10 回の調査のコーディングの見直しを行ったほか，（自由回答法ではなく）多項選択法（プリコード）の採用による追試や，選択肢の順番を逆にした追試によって検討した結果，“ 家族 ” の選択率は底固いことが分かった．つぎに，第 2 の問題については，現在，家族のさまざまな問題が取り上げられ，また，意識調査でも家族関係等に変化が見られるところから，“ 家族 ” の増加は両義的で，その選択率は固いが内実は揺れていると考えられ，21 世紀の「国民性調査」に当たっては組織的な研究を要する課題の一つであると考えられる．

日本型森林セクターモデル：持続的森林資源管理の可能性

吉 本 敦

昭和 30 年代に林野庁の指導の基で行われた拡大造林により，現在我が国の森林資源の大部分は伐採可能な林齢に到達してきており，木材自給率の増加が期待された．しかしながら，当初は木材供給不足を補うために輸入された，いわゆる外材が市場価格に弾性的に反応し，今では我が国市場の 80%を占めるに至る結果となった．すなわち，世界最大の木材輸入国である我が国では利用可能な森林資源が増加しているにも拘わらず，急激な円高の影響等により国産材の価格競争力が低下し，木材消費の自給率が 20%を切ろうとしている状態に陥った訳である．このような経営環境の悪化は森林経営の放棄に結びつき，我が国の森林資源は本来あるべき姿とはほど遠く，不適切な管理状態になってきていると言っても過言ではない．このように，豊富な森林資源を保有しながらそれを利用できない日本と，環境問題を抱えつつ日本へ木材を輸出する産出国との間の資源利用の不均衡を国際的視点から是正するには，貿易構造と環境問題の関係を定量的に分析し，その背景にある各国森林資源の利用・保全のあり方をシミュレーション分析により検討する必要がある．

本研究は部分均衡モデルにより日本を中心とした木材貿易モデルを構築し，市場構造の変化に伴い，持続可能な森林資源管理が可能か否かについて分析を行った．木材貿易モデルの対象として，まず日本を 8 つの地域（東北・北陸・関東・中部・近畿・中国・四国・九州）に分け，外材に対しては，輸出国を考慮せず，全てまとめて 1 つの地域（外国）とした．また，製品については，国産製材，国内挽き米製材，輸入製材の 3 つである．シミュレーションでは，まず 1) 現状の需給傾向が続く場合，2) 海外からの供給が減少する場合，最後に 3) 国内の生産性が上昇した場合の 3 ケースを考慮した．分析の結果，現状の需給傾向が続く限り，現在成熟して

いる森林に対しそれほど伐採の機会がなく，将来それらの多くは老齢林として存続するだけであることが分かった．また，供給の絶対量の変化はあるものの，市場シェアについては全体的な変化は期待されないことが分かった．次に海外からの供給が減少するに従い，国産材の供給はシェア的に増加するものの，価格の高騰により，供給量そのものはそれほど増加しない結果となった．最後に，国内での生産性の増加を仮定すれば，国産材のシェアの増加は期待できることが分かった．人工林の多くは管理がされてはじめて森林の持つ公益的機能が発揮されると言われている．従って，適切な管理へ繋げていくためには，生産性の向上あるいは政策的な手段が必要不可欠になると考えられる．

社会調査データ解析をめぐる 2 つの課題

前 田 忠 彦

表題の 2 つの課題とは，社会調査データの解析，特に構造方程式モデル (SEM)，あるいは因子分析 (FA) の文脈で

1. 複数データセットの効果的解析方法（SEM，FA における多母集団モデルの活用）
2. 標本抽出デザインを考慮したデータ解析，あるいは階層構造をなす調査データの解析

を指す．特に 1 については既にいくつかの応用研究を行ったが，2 の話題も含めていずれも継続検討中という意味で課題と表現したものである．

このうち 1 の課題すなわち「複数データセットの解析」に関し，本年度は「日本人の国民性調査」から満足感関連項目の分析を行った (Maeda (2000))．これは前田 (1995) で同調査の第 9 次全国調査（1993 年）データについて検討したモデルを，第 10 次全国調査（1998 年）のデータにも適用し，モデルの当てはまりを第 9 次，第 10 次間で比較したものである．分析モデルは，2 種類の満足感（“生活満足感”と “社会への満足感”）の相互関係およびそれらの規定因を SEM で表現したもので，第 9 次データに基づき最適と判断したモデルが，第 10 次データでもより明確な形で妥当と判断され，相関構造の安定性が示唆された．しかし第 10 次においては，モデルで使用した「日本への評価」や「社会への満足感」に関する項目の平均値が第 9 次から大幅に低下しており，期待値構造を含むモデル化を行うことにより，両データの差異を記述可能である点を指摘できる．

2 番目の課題に関連し，本年度は科学研究費（奨励研究（A））の補助を受けて，分析に適した社会調査データの取得を目的として，首都圏の有権者を対象とする郵送調査を実施した（日本人の国民性 2000 年度吟味郵送調査）．層化二段無作為抽出を用いて，地点当たりの計画標本サイズを通常の面接調査などよりもやや大きめに設計した．回収率は約 60%となり，2001 年 3 月現在で集計中である．

参 考 文 献

前田忠彦 (1995). 日本人の満足感の構造とその規定因に関する因果モデル ——共分散構造分析の「日本人の国民性調査」への適用——, 統計数理, **43**(1), 141–160.

Maeda, T. (2000). Analyses of satisfaction related items in the Japanese National Character Survey by structural equation modeling, *Proceedings of the International Conference on Measurement and Multivariate Analysis*, Vol. 2, 152–155.

Poisson Voronoi Cell の統計分布

種 村 正 美

空間に散布された多数の粒子の配置パターンを特徴づけたり，粒子同士の局所的な配置のモデルを作るために，ボロノイ (Voronoi) 領域を利用することはしばしば有用である．

特に，ポアソン配置（完全ランダム配置ともいう）に対するボロノイ分割は空間統計学において配置の標準モデルとして重要である．

本研究では，ポアソン配置のボロノイ分割によって生じるボロノイ領域（Poisson Voronoi Cell と呼ぶ）の幾何学的特徴量の統計分布を求めた．これらの統計分布が知られれば別の配置モデルに対する同種の分布と比較するなど，利用価値が高い．

この問題はすでに多くの研究がなされているが，平均値など一部の情報が理論的に得られている (Meijering (1953)) ものの統計分布そのものは理論的に求められていない．そこで，計算機実験によって Poisson Voronoi Cell の独立な標本を大量に生成し，それらから幾何学量の統計分布を推定する方法がとられる．これまでの研究ではわれわれの知る限り，2 次元では標本数が $n = 2,000,000$ が最大 (Hinde and Miles (1980))，3 次元では $n = 358,000$ が最大であった (Kumar et al. (1992))（Tanemura（1988）では $n = 100,000$（3 次元）で行った）．今回，われわれは標本数をそれぞれ $n = 10,000,000$（2 次元）および $n = 5,000,000$（3 次元）として計算機実験を行った．

計算方法の詳細については省略する（ボロノイ領域の計算には Tanemura et al.（1983）のアルゴリズムを用いた）．

計算機実験から得られた特徴量のヒストグラムに理論分布

$$f(x) = ab^{c/a}x^{c-1}\exp(-bx^a)/\Gamma(c/a) \quad (a, b, c > 0)$$

を当てはめた．これは一般化ガンマ分布と呼ばれ，広範囲の分布を表現できる (Hinde and Miles (1980))．最尤推定法によって 2 次元ボロノイ多角形面積に対して $\hat{a} = 1.07946$, $\hat{b} = 3.01623$, $\hat{c} = 3.31134$, 3 次元ボロノイ多面体体積に対して $\hat{a} = 1.16794$, $\hat{b} = 4.03059$, $\hat{c} = 4.79786$ を得た．推定された分布はそれぞれ観測されたヒストグラムをよく再現した．上記の結果は，1 次元ボロノイ線分長の理論分布 ($a = 1$, $b = 2$, $c = 2$) と合わせると，空間次元との明確な関係が予想される．他にも，2 次元では辺数分布，周囲長分布など，また 3 次元では，面数分布，表面積分布などについて同様のあてはめを行い，興味深い知見を得た．

参 考 文 献

Hinde, A. L. and Miles, R. E. (1980). Monte Carlo estimates of the distributions of the random polygons of the Voronoi tessellation with respect to a Poisson process, *J. Statist. Comput. Simulation*, **10**, 205–223.

Kumar, S., Kurtz, S. K., Banavar, J. R. and Sharma, M. G. (1992). Properties of a three-dimensional Poisson-Voronoi tesselation: A Monte Carlo study, *J. Statist. Phys.*, **67**, 523–551.

Meijering, J. L. (1953). Interface area, edge length, and number of vertices in crystal aggregates with random nucleation, *Philips Res. Rep.*, **8**, 270–290.

Tanemura, M. (1988). Random packing and random tessellation in relation to the dimension of space, *J. Microsc.*, **151**, 247–255.

Tanemura, M., Ogawa, T. and Ogita, N. (1983). A new algorithm for three-dimensional Voronoi tessellation, *J. Comput. Phys.*, **51**, 191–207.

遺伝子型を含んだ樹木分布地図データの応用例

島 谷 健一郎

ブナ科の樹木のように種子散布の大半が樹冠からの単なる重力落下による場合，実生は当然母樹のまわりに集中する．その分布地図データから，種子（実生）散布を適当な分散の二次元正規分布あるいは樹冠くらいの大きさの円内の一様分布と仮定し，各母樹の種子生産力と合わせて inhomogeneous Poisson process としてモデル化できる．Likelihood も容易に書き下せるので最尤法により種子散布や種子生産を推定できる．

近年，個体分布地図に加えて，それらの遺伝子型も調べ，遺伝子から各実生の母樹を絞り込んだり父系（花粉の移動）も推定したりする試みが盛んに行われている．本研究では，遺伝子型をマークとする marked point process により遺伝子型付実生分布地図データをモデル化し，最尤法により種子及び花粉の移動並びに各母樹の種子生産を推定することを試みた．並行して遺伝子を含まない単なる分布地図だけから推定した数値との比較も行った．データは Kawano and Kitamura (1997), のブナの実生データを用いた．

遺伝子を含めるか否かで，種子生産の最尤推定値の大小が逆転する母樹の対などもみつかり，遺伝子を加える事の意義は大きいように思える．但し，モデルの検定など，細部は 2001 年 3 月現在ではまだ研究途上にある．

参 考 文 献

Kawano, S. and Kitamura, K. (1997). Demographic genetics of the Japanese Beech, *Fagus crenata*, in the Ogawa Forest Preserve, Ibaraki, Central Honshu, Japan. III. Population dynamics and genetic substructuring within a metapopulation, *Plant Species Biology*, **12**, 157–177.

コウホート分析から見た調査の継続性

中 村 隆

調査結果は調査方式（調査票の設計やサンプリング・実査の方法）の影響を受ける．したがって，調査間の結果の比較可能性を確保するためには調査方式を同一にすることが望ましい．しかし，長期にわたる継続調査の場合，調査環境の変化によって，総合的に判断して比較可能性を維持するためにも，調査方式の変更を余儀なくされることがある．このときその影響の程度を把握しておくことが必要である．

調査の継続性があるかどうかは，まずは各調査回の標本全体についての集計結果を時系列比較して判断することになる．ここではさらに，コウホート（同時出生集団，世代）の視点を取り入れて継続調査データ全体を有機的に結びつけながら分析する方法 —— 年齢×調査時点別の集計データから年齢・時代・コウホート（世代）効果を分離するコウホート分析法 —— を適用して検討する．

一般に，調査方式の変更の影響（調査方式効果）は，もしあれば理想的には，調査回単位で調査対象者全体に影響が及ぶという意味で，時代効果にのみ含まれると考えられる．したがって，影響が時代効果に止まりそれが分離できれば，継続性は保たれたともいえる．調査方式効

果が時代効果から分離できるか否か，年齢・時代・コウホート効果のあり方がどう違ってくるのかを分析することにより，調査方式効果に関して一段深い知見を得ることができる．

具体的な分析モデルとしては，第 4 変数を導入したベイズ型コウホートモデル（中村（2000））を適用する．第 i 年齢階層，第 j 調査年のあるカテゴリの反応確率を π_{ij} とするとき，そのロジット変換を

$$\log[\pi_{ij}/(1-\pi_{ij})] = \beta_0 + \beta^X x_j + \beta_i^A + \beta_j^P + \beta_k^C$$

のように分解する．ここで，β_0，β_i^A，β_j^P，β_k^C はそれぞれ総平均，年齢，時代，コウホート効果のパラメータである．x_j が第 4 変数（調査方式）であり，調査方式変更前なら $x_j = 0$，変更後なら $x_j = 1$ とする．β^X が対応する調査方式効果のパラメータである．適用例などの詳細は中村 (2000) を参照.

参 考 文 献

中村 隆 (2000). コウホート分析における第 4 変数の導入に関する研究，平成 10〜11 年度科学研究費補助金基盤研究 (C)(2) 研究成果報告書 (課題番号 10680324).

定性情報のマイニング
── テキスト型データ解析システム：WordMiner ──

大 隅 昇

市場調査や社会調査の分野では，ここ数年の傾向として，定性情報の活用方法，とくに自由回答・自由記述等のテキスト型データの取得環境の構築やそれらデータの解析手法への関心が高い．インターネットの普及により，電子的手段によるテキスト型データの取得が容易となったことが，かつての定性調査とは様相の異なる種々のアプローチの可能性を拡げつつある．その一つが，テキスト・マイニング手法とそれに関連した応用ソフトウェアの登場という形で現れている（例えば，VextSearch，Survey Analyzer，Symfo WARE Mining Server）．しかし，多くのソフトウェアは高価でありしかも PC 上で簡便に使えるものは少ない．

もちろん社会調査分野におけるテキスト型データの利用に限って考えると，従来は定性調査の一環として，様々な方法が日常的に行われてきた．グループ・インタビューや自由回答設問で取得のデータのアフターコーディング処理等がその例であるが，これは多くの場合主観的な操作となり易く，より客観的で具体性のある解析手法の研究や関連したコンピュータ処理システムの開発はそれほど進んでいるとはいえなかった．

しかし，テキスト型データの電子的取得の環境が急速に整い，調査における自由回答取得はもとより様々な場面でテキスト型データの取得が容易となってきた．例えば，Web 調査における電子調査票による自由回答取得，企業のコールセンターにおける顧客意見の電子的取得等が日常化している．しかも，このような取得データは一般に膨大な量にのぼり，従来型の分析手法だけでは対応が十分とはいえない．とくに，この種の定性情報の適切な計量化や客観的評価は容易ではなく，しかも調査法としての反復可能性や再現性の確保が難しい．

このようなことから，筆書は，調査における自由回答取得やデータ解析のあり方として，従来からある科学的調査法の援用が得られる選択肢型設問方式と，自由回答設問とを併用する調

査法が妥当と考え，この主張を多くの実験調査を通じて実証的に検証することを進めてきた．同時に，データ科学の観点から，適切な自由回答取得方法の検討に始まり，取得データの多次元データ解析手法の方法論構築，それを具現化したソフトウェアの開発まで，一貫した研究を進めてきた．

WordMiner は，この主張を具現化したテキスト型データ解析ソフトウェアである．既に 10 年程前から，仏国の研究者等との共同研究で SPAD.T (Systèm Portable pour l'Analyse des Données Textuelles) を開発してきたが，その後，これに日本語対応のための分かち書き処理等を付加した InfoMiner with WinAiBASE を開発し，一部の研究者や企業にモニター用として提供してきた．これを，新たな設計指針に沿って強化改善を図った解析ソフトが WordMiner である．これは PC 上（Windows 対応）で動作し，とくに特別なコンピュータ環境を必要とせず操作も簡単である．

WordMiner は，日本語テキストの解析に適した基本機能（分かち書き，キーワード抽出，不要単語削除，単語置換編集，辞書化等）に加えて，独自開発の多次元データ解析機能（対応分析，クラスター化等）が含まれる．しかし，単なる多次元解析手法だけでなく，①自由回答パターンと単語，回答者クラスター化情報と単語，それぞれの関連性評価，②選択肢型設問・属性情報と単語の関連性の分析，③単語群の有意性テストから典型的な回答例を知ること，④回答者クラスターを意味づける単語群の生成と典型的な回答例を検証すること等，自由回答と選択肢型設問・属性情報あるいはクラスター化で得た類型との関連等を探索的に行う，テキスト型データ解析専用統計システムである．

電話調査法について

土 屋 隆 裕

首都圏の 1 都 3 県の有権者を対象に電話による世論調査を実施し，調査法の違いが結果に与える影響について検討した．具体的には，事前告知のある TD 法，事前告知のない TD 法，RDD 法という 3 つの調査方法の比較を行った．TD 法は，各市区町村の選挙人名簿から層別二段系統抽出法により選び出した調査対象者の電話番号を電話帳で調べ，電話番号の判明した対象者に電話をかけて調査を行う方法である．TD 法の対象者は 2 群に分け，一方には事前にハガキによって調査への協力依頼を行った．RDD 法は，1 都 3 県で使われている可能性のある電話番号をランダムに発生させ，その電話番号に電話をして世帯につながれば，世帯内でランダムに有権者を 1 人選んでもらって調査対象者とする，という方法である．事前告知を行わないと回収率が低下すること，RDD 法に比べ TD 法は保守的な意見が多くなること，といった傾向が見られた．

Trimmed k-means 基準を用いた主要点について

清 水 信 夫

主要点 (Principal Points，Flury (1990)) は，最適分割された確率密度関数の各領域の代表点のうち，それらの点からの 2 乗距離の期待値が最小となるような点として定義される．対称な 1 変量分布における主要点の対称性は必ずしも成り立たないことが知られているが (Flury

(1990))，正規分布やロジスティック分布など今日の統計学で広く利用されている多くの対称な 1 変量確率分布族において主要点は対称となる（清水 他（1999））．また，2 変量正規分布における主要点の配置についても数値計算により求められている（清水 他（1998））．

一方，Trimmed k-means 法 (Cuesta-Albertos et al. (1997)) は，全データのうちの一部分をトリミングした上で行われる k-means クラスタリングである．この手法では，特異データの混在により潜在的なクラスター構造が識別できない場合においても，元のクラスターを十分に識別し得る利点がある．

本研究においては主要点に Trimmed k-means 基準を適用し(以下 Trimmed k-Principal Points)，1 変量混合正規分布における Trimmed 2-Principal Points の対称性について検討を行った．

参 考 文 献

Flury, B. A. (1990). Principal points, *Biometrika*, **77**(1), 33–41.

Cuesta-Albertos, J. A., Gordaliza, A. and Matrán, C. (1997). Trimmed k-means: An attempt to robustify quantizers, *Ann. Statist.*, **25**(2), 553–576.

清水信夫，水田正弘，佐藤義治 (1998). Principal Points の性質について，応用統計学，**27**(1)，1–16.

清水信夫，水田正弘，佐藤義治 (1999). Principal Points の対称性に関する定理について，計算機統計学，**12**(1)，45–53.

Echelon 解析とその応用

（客員）岡山大学 栗 原 考 次

Echelon 解析は，矩形上に分けられた地図上の 1 変量データに対して，空間的な位置を表面上のデータ高低に基づき分割し，空間データの位相的な構造を系統的かつ客観的に見つける新しい解析法 (Myers et al. (1997)) である．本年度は，1) 人口データ及びリモートセンシングデータの echelon 解析，2) 大容量データに対する echelon 解析，3) 都道府県データに対する echelon 解析，に関する研究を行った．1) については，東京都心における緑地と人口密度の空間構造とそれらの関連について調べた．緑地データは，ランドサット TM データのバンド 3 と 4 に基づく正規化植生指数，人口データは，国勢調査の第 3 次地域区画における人口数を利用した．ランドサットデータは位置に関する幾何補正を行った後，ピクセルのグループ化を行い人口データの矩形領域に適合させた．Echelon 解析の結果，(1) 山の手地区，(2) 山の手地区から 7 km 未満（環状 8 号線内）の住宅地区，(3) 郊外地区，の 3 地区でその構造が大きく異なり，山の手地区から 7 km 未満の住宅地区において，人口数と緑地に強い相関が示された．さらに，人口の過疎及び緑地のピークはともに皇居及び代々木公園を中心とした山の手地域内と西部の多摩川流域にあることや東京都心における人口の過疎及び緑地の全体的な位相構造が示された．2) については，リモートセンシングデータのような大容量データの解析を行うため，echelon tree を limb と bough に階層的に分解し，これら値に基づく 4 つのプロフィール divergence，scope，bunching，stacking を定義した．また，具体的なリモートセンシングデータに対して 4 つのプロフィールに基づく構造分析を行った．3) については，各都道府県の隣接情報を与えることによって，echelon デンドログラムを構築すると共に，都道府県別人口及び日本酒の消費量データに基づきこれらの位相的な構造の把握を行った．

参 考 文 献

Kurihara, K., Myers, W. M. and Patil, G. P. (2000). Echelon analysis of the relationship between population and land cover pattern based on remote sensing data, *Community Ecology*, **1**(1), 103–122.

Myers, W. M., Patil, G. P. and Joly, K. (1997). Echelon approach to areas of concern in synoptic regional monitoring, *Environmental and Ecological Statistics*, **4**, 131–152.

統計計算開発センター

統計解析とパソコンクラスター

田 村 義 保

「データマイニング」,「超大量データ」という用語を，日々，耳にするようになっている．「知識獲得・発見」という用語も良く耳にする．自然科学のみならず，生命科学，社会科学の研究者が統計科学の必要性を，未だかつてないくらい強く感じ取っているものと思われる．さらに，データを扱うための科学が統計科学であり，その利用分野が広がりつつあることを否定する者はいないと思う．しかし，データの大量化，複雑化，品質の低下（悪い意味ではなく，非常に困難な条件で測定しているので，S/N 比が悪いデータしかとれないという意味である）等のために，高度な統計モデルを考えることが必要になり，複雑な計算が必要になっている．データの大量化と計算の複雑化のために，必要とする計算資源が増大している．

パーソナルコンピュータのクロックが 1 GHz 以上になり高速化している．スパコンの生みの親クレイ氏が夢見た机の上に乗る Cray が実現したと言っても過言ではない．しかしながら，パーソナルコンピュータでは処理できないような問題にチャレンジすることが統計科学者に課されている．高速ではあるが高価な並列計算機を用いなければならない問題も多い．しかしながら，比較的，安価で高速計算環境を実現する手段として，パソコンクラスターを紹介したい．複数のパソコンをネットワークで接続するだけである．ミリネットのような高速ネットを使うと経費が跳ね上がるが，100Base/TX を使うと，比較的，安価で並列計算環境を実現することができる．プログラムもフリーである Mpich を使用すればよく，FORTRAN であれば，

```
call MPI_INIT(ierror)
call MPI_COMM_SIZE(MPI_COMM_WORLD, nprocs, ierror)
call MPI_COMM_RANK(MPI_COMM_WORLD, myid, ierror)
call MPI_BCAST(n, 1, MPI_INTEGER,0,MPI_COMM_WORLD, ierror)
```

のようなサブルーチンコールをするだけである．問題にもよるが，用いる計算機の台数を 10 倍にした場合，計算速度は 5 倍から 10 倍程度になることが期待できる．安価なスパコンとしてパソコンクラスターの構築を勧めたい．

密行列計算向け CPU の開発

泰地　真弘人

密な行列であらわされる連立一次方程式の解や，対角化，行列式の値などの計算には，行列

の大きさ N の 3 乗の計算時間を要するため，非常に計算量が多い．しかしこうした問題に対しては，多数の演算器に同一のデータを供給し同一の演算を行なう変形 SIMD 方式の CPU で高速化できる．各演算器は共通のデータ入力ポートと演算器に固有のローカルメモリを持ち，これらのデータを用いて計算を行なう．こうした CPU は入出力の問題が少ないため，通常のマイクロプロセッサに比べてはるかに高い集積度・性能を達成でき，結果として通常の並列計算機に比べて 10 倍以上コストパフォーマンスを上げることができる．本 CPU では例えば LU 分解，QR 分解，行列積などの演算を高速に実行することが可能である．

今年度は実際の CPU の試作を行なった．倍精度浮動小数点積和演算器 8 個を 1CPU に搭載し，最悪条件下では 222 MHz，標準条件下では 333 MHz で動作するように設計した．従って，最悪条件下では 3.5 Gflops，標準条件下では 5.3 Gflops の性能を持つことになる．これは計画値 (1.6 Gflops) の倍以上であり，倍精度演算を行なう CPU としては世界最高速となる．現在 CPU の評価用のボードを製作中である．本 CPU を用いたシステムの原価は，1 Tflops あたり約 4000 万円程度になる．1 Tflops のシステムを用いると，10 万元の連立一次方程式を約 30 分で解くことができる．

TIMSAC72 の計算法について

荒畑 恵美子

TIMSAC72 の計算法についての説明を ASTEC-X で作った．それを用いて，TIMSAC72 の計算法について説明をした．

参 考 文 献

赤池弘次，中川東一郎 (1972). 『ダイナミックシステムの統計的解析と制御』, サイエンス社，東京.

統計解析システム Jasp のグラフィカルユーザインタフェース

中 野 純 司

パーソナルコンピュータの高性能化と低価格化，およびインターネットの普及などにより，計算機はあらゆる場面でごく普通に用いられるようになっている．従って，それらを用いて統計解析を行いたいという初心者ユーザも増えている．そのようなユーザにとって，使いやすいユーザインタフェース (UI) は統計解析システムが備えるべきもっとも重要な機能のひとつである．

現在のソフトウエアにおいては，そのわかりやすさのために，グラフィカルユーザインタフェース (GUI) が標準として使われている．ただ，統計解析システムにおいては，既存の機能を定型的に利用するだけですむことは少なく，それらを複雑に組み合わせたり，新しい手法を開発したりする必要があることが多い．そのために，統計解析言語，またはキャラクタユーザインタフェース (CUI) も同時に重要である．さらに技術的な理由もあり，著名な統計解析シ

ステムでは CUI が最初に設計され，その上に GUI をかぶせたものが多い．そのため，GUI が CUI の制約を受けるような設計も見受けられる．

われわれが開発している統計解析システム Jasp においては，はじめの段階から GUI と CUI を同等に考えている．そして，それらが独立に利用できること，しかし交互に等価的にも利用できること，さらに GUI はその利点を十分に生かすこと，を目的として設計されている．まず，GUI ではデータやモデルなどをアイコンで表し，それらを直観的に直接操作できるようにした．この機能を実現するためにオブジェクト指向の考え方を利用した．また，それらのアイコンは解析の履歴を表すように，木構造をもつように整理されて表示される．これにより，試行錯誤的な解析がやりやすくなることが期待できる．そして，これらの GUI に必要とされる情報を与えるために，プログラムの中で特殊な形式のコメントを付加するようにし，言語構造が GUI のために複雑にならないようにしている．