統計数理 (1996)
第44巻 第2号 285-320

平成7年度研究報告会要旨

と　き：1996年3月 18日 午前10時～午後5時15分
19日 午前10時～午後5時

ところ：統計数理研究所講堂

プログラム

3月18日

あいさつ	所長　清水　良一
領域統計研究系	
再審査・再評価再考	柳本　武美
長期予防試験でのサンプルサイズ設計	佐藤　俊哉
Effects of Jackknifing on Asymptotic Pivots	汪　金芳
結晶群の出現頻度の確率モデル	伊藤　栄明
人文科学データの計量分析	村上　征勝
社会調査データに見られる力動的変動	吉野　諒三
マクロ的およびミクロ的国際比較分析	釜野さおり
予測制御研究系	
複雑な系の時系列解析	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
同時転換自己回帰モデルにおける検定について	佐藤　整尚
Centered Newton法再説	田辺　國士
層分割を使った内点法	水野　眞治
動的システムの最適制御	伊藤　聡
J言語による統計解析	鈴木義一郎
内点法の探索方向の計算と最小二乗法	土谷　隆
人間乱数	伊庭　幸人
情報量規準によるranking data解析, 続報	石黒真木夫
非定常信号処理と関連研究	瀧澤　由美
スギ人工林資源活用の支援情報	(客員, 新潟大学)　阿部　信行
統計データ解析センター	
モンテカルロ平滑化について	北川源四郎
予測分布の理論と時系列のスペクトル推定について	駒木　文保
語順のシミュレーション・モデル	上田　澄江

統計数理研究所の計算機環境	田村　義保
欠測値処理のアルゴリズムについて	荒畑恵美子

3月19日

統計教育・情報センター

分子系統学について	長谷川政美
真核生物の起源	橋本　哲男
連続して起こる事象における確率分布	内田　雅之
Fokker-Planck 方程式の高精度化をめざして	岡崎　卓
非集計分析における不確実性の表現法	山下　智志

統計基礎研究系

離散確率分布と統計的応用	平野　勝臣
確率過程のノンパラメトリック推測について	西山　陽一
統計基礎モデルの構築と推定方程式	松縄　規
マルチンゲールの漸近展開とその応用	吉田　朋広
Convolution equivalent distribution に属するための条件	志村　隆彰

調査実験解析研究系

意識, 調査法	坂元　慶行
コウホート分析における識別問題再考	中村　隆
反復測定データの因子分析	前田　忠彦
ランダム充填過程の一性質	種村　正美
パソコン UNIX による統計学研究環境	丸山　直昌
指数逆ガウス型分布について	金藤　浩司
探索的多次元データ分析と応用研究	駒澤　勉
多項分布モデルに基づくクラスタリング	馬場　康維
3相データにおける尺度構成法	土屋　隆裕
テキスト型データの解析法 −自由回答データの分析を中心に−	大隅　昇
HALBAU for Windows の開発	高木　廣文
線形計画法による収穫規整モデルの分析比較	鄭　躍軍
前震の識別に関する統計的予測の研究	尾形　良彦
平滑化によるデータ解析	柏木　宣久
射影行列に関するいくつかの話題　(客員, 大学入試センター)	柳井　晴夫

領域統計研究系

長期予防試験でのサンプルサイズ設計

佐 藤 俊 哉

薬剤の使用や生活習慣の改善により疾病の発生を予防できるかどうかを調べるための予防試験は，観察が長期にわたるため厳密な試験計画を立てることが要求される．試験計画の立案の中で重要な要素の一つに，試験に必要なサンプルサイズの設計がある．ここでは，胃がんの一次予防試験で行ったサンプルサイズ設計について報告する．

Helicobacter pylori (ピロリ菌) は慢性胃炎，胃潰瘍などと関連があり，最近では胃がんと関連があるという疫学研究が複数報告されている．ピロリ菌が胃がん発生の原因であるかどうかを調べるため，薬剤によるピロリ菌の除菌・非除菌を参加者にランダムに割り付け，その後の胃がん発生状況を比較する一次予防試験を計画した．試験参加予定者は，20-39, 40-59 歳のピロリ菌感染者で，参加者の登録は 1996 年 4 月から 1 年間，その後 2004 年 12 月まで 8 年間追跡し，毎年内視鏡検査などを実施して胃がん発生状況を調べる．主要評価項目は胃がん発生までの時間である．胃がん発生率は昭和 63 年全国の性別, 年齢階級別胃がん発生率から推定した．20-39 歳では性別で大きな差はみられなかったので 10 万人年あたり 24.7 人を用いた．40-59 歳では, 男性 10 万人年あたり 225.1 人，女性 10 万人年あたり 88.8 人とした．これらの胃がん発生率はピロリ菌陽性者と陰性者を含んだ値であるので, 年齢階級別ピロリ菌陽性割合 (20-39 歳: 40 %, 40-59 歳: 70 %), 陰性者に対する陽性者の胃がん発生率の比から，今回の試験参加予定者である陽性者のみの胃がん発生率を求めた．胃がん発生率の比 (2 倍，3 倍)，40-59 歳での男女比，20-39 歳と 40-59 歳の参加者の比，検出力，をいくつかの値に設定して片側 5 %水準でログランク検定を実施するときに必要なサンプルサイズを計算した．その結果，検出力 60 %の場合でも 1000 ～ 4400 名の参加者が必要となった．現在，試験の手続き，サンプルサイズ設計をより確かなものにするためにパイロットスタディを実施している．

Effects of Jackknifing on Asymptotic Pivots

汪 金 芳

変動係数や，相関係数などの確率変数のモーメントの滑らかな関数であるパラメーター θ は，適当に新たに導入される確率ベクトルの平均の滑らかな関数として記述できる．したがって，これらの母数に関する推測を統一的な枠組みの中で行うことが可能である．筆者はこれまでにこの滑らかな関数モデルの下で，ジャックナイフ法における高次の漸近理論を展開し，いくつかの結果を得た (Wang (1996), Wang and Taguri (1996))．

ジャックナイフ法の効果をよりよく理解するため，通常の非ジャックナイフ型の studentized 統計量に基づく高次の漸近理論が必要になる．本研究では，θ の推測に関して T および T_J

$$T = \frac{\sqrt{n}(\widehat{\theta} - \theta)}{\hat{\sigma}}$$

$$T_J = \frac{\sqrt{n}(\hat{\theta}_J - \theta)}{\hat{\sigma}_J}$$

に関する高次の漸近理論を展開し，その理論および数値的な比較・検討を行った．より具体的に，例えば，3 次の Edgeworth 展開の逆展開に基づく相関係数に対する信頼区間を考えた場合

に，被覆確率がより名目上の被覆確率に近いという意味で，ジャックナイフ法の効果がよく見られた(詳細は，Wang et al.(1996) を参照されたい)．

参考文献

Wang, J.F. (1996). The Jackknife and the bootstrap–higher order asymptotics and numerical examinations, Ph.D. Thesis, Faculty of Natural Sciences, Chiba University.

Wang, J.F. and Taguri, M. (1996). Jackknifing: higher order accurate confidence intervals, *J. Japan Statist. Soc.*, **26**(1), 69-82.

Wang, J.F., Ohuchi, S. and Taguri, M. (1996). Higher order accurate confidence intervals for smooth functions of means with application to the correlation coefficient, *J. Japanese Soc. Comput. Statist.*, **8**(1), 17-36.

結晶群の出現頻度の確率モデル

伊 藤 栄 明

群，グラフを値としてとる確率分布，確率過程という視点から自然にとりあつかうことができる問題について興味をもってきた．結晶群の出現頻度の問題はその典型的なもののひとつである．

結晶の対称性は230の空間群のいずれかにより表される．幾何学的対称性を平衡移動の操作を考慮にいれず記述する点群と言われている32個の群がある．230個の空間群は32個の点群のいずれかに基づいて構成されている．無機非金属結晶構造データベース ICSD (Inorganic Crystal Structure Database) にもとづいて，無機結晶物の種を定義し，点群の分布を調べるという課題にとりくんできた(藤原 他(1996)). ANX記号ごとにもとめた空間群，点群の出現頻度について群，部分群の関係のうえのランダムウオークからえられる平衡分布より説明することをこころみた．ANX記号はAX, AX_2 等を言い，化合物の分子式を酸化数によりおおまかに示すものである．AXは例えば$NaCl$，AX_2 は例えばCO_2をふくむ．金属結晶データベース CRYSTMET についても同様な方法で種を定義し，点群の出現頻度をもとめた．群のうえのランダムウオークにより説明できる場合は金属結晶，単一の元素からなる結晶等のように最密充填構造が基本にありそれが変形したと考えられる場合と思われる．ANX記号,AB_2X_6, 及びABC_2X_6も群のうえのランダムウオークというモデルがデータに比較的よくあう．ICSD には結晶構造の模型を画面に表示するソフトがあり，これから判断するとこれらの場合も最密充填構造が基本にあるように思われる．

参考文献

伊藤 栄明(1985). 群，グラフを値としてとる確率分布, 統計数理, **33**, 68-70.

Itoh, Y. and Matsumoto, T. (1991). Random-generation model for statistical distribution of point groups, *Acta Cryst. Sect. A*, **47**, 204-206.

藤原 美也子, 伊藤 栄明, 松本 崧生, 武田 弘(1996). 無機結晶データベース (ICSD) を用いた結晶群の出現頻度 III, 統計数理研究所共同研究リポート, No.83.

人文科学データの計量分析

村 上 征 勝

文章データの計量分析と考古学データの計量分析の現状，及び計量分析の際の問題点について幾つかの研究例にもとづき報告した．

文章データの計量分析に関しては，まず『源氏物語』の計量分析のために作成した『源氏物語』54巻（約37万語）及び『紫式部日記』,『手枕』などの『源氏物語』関連文献のデータベース（約4万語）の概要について説明した．次に，これらのデータベースを用いた『源氏物語』の計量分析について,『源氏物語』複数著者説の中でも有名な「宇治十帖」他作家説に関する計量分析を中心に紹介した．更に「法華経」37巻の成立順序の推定や，西鶴作品の計量分析などについても紹介した．

また考古学データの計量分析に関しては，前方後円墳の築造形式の推定に関する研究と古代寺院の伽藍配置のデータから寺院の築造年代を推定する研究を紹介し，これらの研究から得られた考古学データの計量分析の問題点について報告した．

社会調査データに見られる力動的変動

吉 野 諒 三

統計数理研究所が長年の調査研究のなかで収集してきたデータのなかで，今日では貴重なパネル調査データのいくつかをとりあげ，その力動的変動の分析を進めている．これまで，各質問項目毎に，パネルの各時点での回答分布%から線形近似で推測される収束点を計算したデータを考察し，収束点においては，各回答カテゴリーに対応する回答反応の分布比がある簡単な整数比になる傾向があるのではないかという推測をしてきた．本年度は，さらにこの考察を進めた．

理論的には，特に，ある数学的条件や近似のもとでMaster equationと呼ばれるかなり一般的な統計的数式より量子力学で有名なSchroedinger equationが導かれる事を確認し，これに基づいてパネル・データ解析のために，計量心理的尺度モデルを構成した．

これまでの成果の概要は，少なくとも意識のパネル社会調査データに関しては，同一の質問項目に対して，

1. 個々の回答者がパネルの各時点毎に異なる回答をする率（回答変動率）は少なくはない（項目によるが，数十%に及ぶ場合もある）．しかし，
2. 総体としての回答変動率（周辺分布の変動）は，少ない（数%程度である）．
3. 回答カテゴリーが2の場合（例．Yes/No項目）の各質問項目の回答カテゴリーに対する比率は，Schroedinger equationの解に対応するある限られた比率の集合の要素に収束する途上の状態とみなせる傾向が確認された．

ただし，いまだ理論的にも実証データの取扱いの点でも正当化されるべき点も多く，将来の課題として残されている．特に，回答カテゴリーが3以上の場合は，それが2の場合のモデルの単純な拡張では，実証データをうまく説明できず，考察が必要である．

参考文献

林 知己夫, 鈴木 達三, 中村 隆, 林 文, 上笹 恒, 堀 洋道, 竹村 研一, 穐山 貞登, 児玉 好信, 井上 佳朗 (1984). 社会調査による国際比較方法の研究, 統計数理研究所研究リポート, No.59.
Helbing, D. (1995). Quantitative sociodynamics, Kluwer, Boston.
鈴木 達三, 高橋 宏一 (1971). 調査における回答機構の統計的研究, 統計数理研究所研究リポート, No.26.
鈴木 達三, 水野 欽司, 大隅 昇, 中村 隆, 田中さえ子 (1981). 社会調査の実施過程における調査誤差の研究, 統計数理研究所研究リポート, No.52.

マクロ的およびミクロ的国際比較分析

釜 野 さ お り

今年度行った国際比較研究の中で用いたデータ解析は，分析レベルに応じて (1) ミクロ的国際比較，(2) マクロ的国際比較，(3) ミクロ-マクロ的国際比較の 3 種類に分類することができる．本報告会では,「7カ国の意識調査データ (統計数理研究所 (1992,1995)) に基づいた研究」と「ジェンダーシステム，同性間の性愛の Naming ならびにゲイ・リベレーション運動の相互関係に関する研究 (Kamano(1995))」の中で行った分析から，各レベル (またはクロスレベル) での国際比較の例を示した．

ミクロ的国際比較分析では，ミクロ変数に基づいて何らかの形で各国の特徴を表現し，それを比較する．例えば，(a) 個人レベルの現象や現象間の構造の比較，(b) 個人レベルの現象を国別に集計した比較，(c) 個人レベルの現象間の因果関係の比較などがある．マクロ的国際比較分析では，国のいろいろな特徴を国レベルで測定し，それらの間の相互関係を分析する．各国の特徴を直接比較するのではなく，世界全体（または特定の地域）においてのマクロ現象の関連パターンを解明する．例えば，(a) マクロ現象間の相関関係に焦点を置くもの，(b) マクロ変数間の因果関係に注目するものなどがある。

マクロ-ミクロ的国際比較分析とは，国のミクロとマクロの現象を同時に分析するもので，(a) ミクロ現象をマクロ変数で説明するもの，(b) マクロ変数によって国を分類し，グループごとにミクロ現象の構造をモデル化してそれらを統計的に比較するものなどがある。マクロ-ミクロ的国際比較分析は，ミクロ的国際比較において単に「国の特徴」と記述され，理論的な説明がされていないものを，国のマクロ的特徴によって説明する方向に進めることができ，社会科学理論的にも意義がある．ミクロ現象とマクロ現象の関連を明確にできるようなマクロレベルのデータを作って行くことが今後の国際比較の課題の一つであると考える．

参考文献

Kamano, S. (1995). Same-sex sexual/intimate relationships: a cross-national analysis of the inter-linkages among naming, the gender system, and gay and lesbian resistance activities, Ph.D Dissertation, Department of Sociology, Stanford University.
統計数理研究所 (1992). 意識の国際比較方法論の研究—5カ国調査共通ファイルコードブック—, 統計数理研究所研究リポート, No.72.
統計数理研究所 (1995). 意識の国際比較における連鎖的調査分析方法の実用化に関する研究—共通ファイルコードブック—, 統計数理研究所研究リポート, No.82.

予測制御研究系

複雑な系の時系列解析

尾崎　統

工学や医学におけるダイナミックな問題，火力発電プラントの制御法，脳神経データからの脳の活動の特徴づけ，心臓拍動の変化の特徴づけなど種々の現実の具体的問題を複雑な系の時系列解析という視点から研究した．

1, 火力発電プラントの制御に限らず，現実の大規模かつ複雑な系の制御は殆どの場合「PID制御」とよばれる古典的方法で行われている．1980年代に入って中村秀雄と赤池弘次はPID制御の下で動いているシステムをPIDコントローラーを込みにして，両者合体したものを一つの複雑なシステムとみなし，それにさらに統計的方法に基づく最適制御をかけることによって，システムの安定性を保ちつつ制御効率を飛躍的にあげることが出来ることを示した．現在この中村-赤池法は日本国内，九州から北海道までの殆どの電力会社において，火力発電プラント制御に採り入れられ，更に国外の火力発電でも関心を集め続々と採り入れられている．中村-赤池法の特徴は定常ガウス確率過程の線形理論に関する推測問題を徹底的に追及する中で生まれたものであるが，火力発電プラントの良く知られた特徴に炉の蒸気温度と蒸気圧力に関係する部分の応答特性が一定せず電力負荷の大きさに対応して大きく変化するという点がある．中村-赤池法では電力負荷要求レベルが大きく変化する時点である種の線形補間をすることによってこの問題をクリアーしているが，彼等の方法で改良の余地があるとすればこの非線形性の取扱いであることは早い段階から知られていた．本研究では赤池のオリジナルTIMSAC法にのっとって，多入力多出力システムのCausality解析と，鍵となる重要変数の選択の方法を利用しながら，システムの非線形性を考慮し，非線形モデルに基づくある種の局所的最適性を追及する非線形制御法を導入した．モデルの非線形性は状態遷移行列の要素（したがって固有値）を時々刻々の電力負荷要求のレベルの関数にすることによって実現されている。実際の火力発電所の試験データ（約40チャンネルの時系列データ）から最終的に8チャンネルのデータを用いて，モデル推定をおこない，推定離散時間モデルによるデジタル計算機実験でこの新手法の有効性を確認した．新しい制御手法を現実の炉の制御に使う前には色々な角度からのチェックが必要で，現在この手法を17次元プラント微分方程式モデルやもっと精密な40次元微分方程式モデルによるシミュレーションに適用することにより制御効率の確認テストを行っている段階にある．

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

モデル規範形適応制御系を構成するためには，対象の次数の上界や相対次数の厳密な値があらかじめ既知でなければならない．さらに外乱や非線形成分が存在するときには、それらの関数形が事前に特定化されねばならない．しかし適切な次数の上界を設定したり，相対次数の厳密な値，外乱や非線形成分の関数形を特定化するのは困難なことが多く，そのような制約をどこまで緩和できるか明らかにすることが適応制御の一つの課題とされている．

これに対してこれまでに，可変構造制御やハイゲインフィードバックの手法を用いてロバスト適応制御系を構成する研究を行い，まず相対次数が1次の系について，次いで相対次数が1次か2次かで未知の場合について，次数や非線形成分，外乱に依存しないロバスト適応制御系の構成法を提案した．

今年度は以上の研究をさらに相対次数が既知で2次以上の一般の場合に拡張して次数や非線形成分に依存しないロバスト適応制御系の構成法を求めるとともに，相対次数が2から3次の範囲で不確定な場合についても同様のロバスト適応制御系の構成法を求めた．

まずBackstepping法とハイゲインフィードバックに通常の線形補償を加えたロバスト適応制御系の一般形式を求めて，次数や非線形性に関するモデル化誤差が存在するときには適応系の大域的有界性と制御誤差の任意に小さな領域への収束性が保証され，さらにモデル化誤差が存在しない理想的な環境の時には制御誤差の零収束も達成されるような非線形ロバスト制御系の構成法を示した．また提案する適応制御系の最小次元構成が3個の適応パラメータで実現できることも示した．

次に超安定ループ中の正実関数の相対次数に-1～+1次の自由度があることに注目して，対象の相対次数が一般にr～r+1次またはr～r+2次の範囲で不確定な場合でも，単一の適応機構で大域的漸近安定性の保証される適応安定化制御系とモデル規範形適応制御系の構成法を求めた．さらにこのような手法が高調波ゲインの符号が未知な場合に拡張できることも，簡単な例で確認した．次年度以降これらの研究を発展させて，超安定ループと正実化に基づく適応制御理論で実現し得る適応制御系のクラスの限界を明らかにしていく予定である．

参考文献

宮里 義彦(1995). 次数に依存しない非線形モデル規範形適応制御系の構成法(一般の相対次数の場合), 計測自動制御学会論文集, **31**(3), 324-333.

Miyasato, Y. (1995). Model reference adaptive control for nonlinear systems with unknown degrees (minimal number of tuning parameters), *Proceedings of 1995 American Control Conference*, **4**, 2505-2509.

Miyasato, Y. (1995). Globally bounded robust adaptive controller with arbitrarily small residual tracking error, *Proceedings of the 34th IEEE Conference on Decision and Control*, **4**, 3947-3952.

宮里 義彦(1996). 大域的有界性と制御誤差のバイアス補償機能を有するロバスト適応制御, 計測自動制御学会論文集, **32**(1), 26-34.

同時転換自己回帰モデルにおける検定について

佐 藤 整 尚

近年，経済データを分析するうえで様々な非線形時系列モデルが適用されている．国友・佐藤(1994), Kunitomo and Sato(1996)で提案された同時転換自己回帰モデル(SSARモデル)もそのひとつで，データの持つ非対称性を表現するモデルと言える．今年度は以下のような1変量SSARモデル,

$$y_t = \begin{cases} Ay_{t-1} + \sigma_1 u_t & \text{if} \quad y_t \geq y_{t-1} \\ By_{t-1} + \sigma_2 u_t & \text{if} \quad y_t < y_{t-1}, \end{cases}$$

ただし,

$$\frac{1-A}{\sigma_1} = \frac{1-B}{\sigma_2} \ , \quad \sigma_1, \sigma_2 > 0 \ , \quad u_t \sim \text{ i.i.d.} N(0,1)$$

を考え，この検定問題を研究した．具体的には帰無仮説に

$$H_0: \ A = B$$

という線形仮説を置き, 現実のデータから SSAR モデルが支持されるかどうかを調べる方法を提案した．まずはじめに尤度比検定を考え，その性質をシミュレーションで調べた．その結果，対立仮説に対する検出力は高かったものの，$\{u_t\}$ の正規性に対しては頑健性がないことが分かった．次に尤度に基づかない検定方法 (PCT テスト) を提案した．この方式によって，尤度比検定の欠点を克服することができた．

参考文献

国友 直人, 佐藤　整尚 (1994). 経済時系列における非線形性と不均衡計量経済モデル,『数理統計学の理論と応用』(竹内 啓・竹村　彰通 編), 東大出版会, 東京.

Kunitomo, N. and Sato, S. (1996). Asymmetry in economic time series and simultaneous switching autoregressive model, *Structural Change and Economic Dynamics*, **7**, 1-34.

近似計算を使う内点法の収束性

水　野　眞　治

内点法は，線形計画問題などの最適化問題を数値的に解くアルゴリズムである．N. Karmarkar (1984) が射影変換法と呼ばれる内点法を提案して以来，さまざまな内点法が開発され，その収束性が理論的に解析されてきた．内点法は，反復解法であり，1 反復毎に線形方程式を解く．内点法の収束性の解析では，今までこの線形方程式の解が正確に計算できるものと仮定されてきた．しかし，大規模な線形計画問題においては，線形方程式の解を必ずしも正確に計算できるとは限らない．また，正確な解を求めるために多くの計算時間を必要とする場合には，近似解を高速に求める算法を使うこともある．

そこで，Mizuno and Jarre (1996) では，線形方程式の解が近似的にしか得られない場合について，内点法の収束性を解析した．その結果，線形方程式の近似解を非常に荒く計算しても内点法が収束することを明らかにした．すなわち，線形方程式を解く場合に，残差のノルムが右辺定数ベクトルのノルムよりもある一定の割合で少ないような近似解が計算できるならば，大域的に収束する内点法を構築できることを示した．さらに，多項式オーダと呼ばれる高速な収束性を理論的に達成するために求めなければならない近似解の精度を明らかにした．すなわち，線形方程式を近似的に解くという条件の下で，多項式オーダを達成する内点法を提案した．

参考文献

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming, *Combinatorica*, 4, 373-395.

Mizuno, S. and Jarre, F. (1996). Global and polynomial-time convergence of an infeasible-interior-point algorithm using inexact computation, Research Memo., No.605, The Institute of Statistical Mathematics, Tokyo.

H_∞ノルム条件の数理計画法的解釈

伊 藤 聡

外乱など不確実さの存在の下での線形制御系の設計仕様として，外乱に対するロバスト性を保証する H_∞ ノルム評価基準が知られているが，ここでは数理計画法的観点からみた H_∞ノルム条件の解釈について述べる．

次のような線形時不変系を考える．

$$(1)\qquad \begin{cases} \dot{x}(t) = Ax(t) + B_1u(t) + B_2v(t), & x(0) = x_0 \\ z(t) = Cx(t) + D_1u(t) + D_2v(t) \end{cases}$$

ここで x は状態，u は制御入力，v は外乱，z は評価出力を表わす．許容な u および v のクラスとして $[0,\infty)$ で定義された L_2 関数の空間を考える．制御入力 u あるいはこれを生成する制御則すなわちコントローラが与えられれば，初期状態 x_0 を 0 としたときの外乱 v から評価出力 z への伝達特性が一意に定まるが，これを時間領域において $z = T_{vz}v + d$ と表わすことにする．例えば状態フィードバック則のもとではこの伝達特性は線形になるが，時間領域における伝達特性は一般にはアフィンとなることに注意されたい．一方，線形作用素 T_{vz} の H_∞ノルムは次のように定義される．

$$\|T_{vz}\|_\infty = \sup_{v\neq 0} \frac{\|T_{vz}v\|_2}{\|v\|_2}$$

いま，ある希求水準 $\gamma > 0$ に対して，閉ループ系を内部安定にし

$$(2)\qquad \|T_{vz}\|_\infty \leq \gamma$$

となるようなコントローラ K を求める問題を考える．ここで外乱から評価出力への伝達特性の線形性を仮定すると，(2) 式は等価に

$$(2)'\qquad \sup_v \int_0^\infty z(t)^T z(t) - \gamma^2 v(t)^T v(t)\, dt \leq 0$$

と書けるが，この左辺をコントローラ K の関数 G とみると，結局 $G(K) \leq 0$ を満たす許容解を求める問題となる．H_∞ノルム評価基準 (2) は線形システム (1) に対する一つの満足化設計仕様に過ぎず，この他にもいろいろな評価基準が存在する．そこで，ある評価基準 F に対して次のような形の満足最適化設計仕様が考えられる．

$$(3)\qquad \min_K F(K) \quad \text{subj. to} \quad G(K) \leq 0$$

(2)$'$式の左辺を $G(K) = \sup_v g(K,v)$ と書いたとき，許容なコントローラ K に対しては汎関数 g は v に関して最大値を持ち，$G(K) = \max_v g(K,v)$ と書けることが次の命題よりわかる．

命題 1. $G(K) \leq 0$ なる許容なコントローラ K に対して g は v に関して凹となる．また $G(K) < 0$ ならば強凹となる．

一方，H_∞ノルムに対して希求水準 γ が適当に定められていれば不等式制約 $G(K) \leq 0$ は制約想定を満たす．

命題 2. 汎関数 g は K に関して凸であるとする．このとき最大値汎関数 G も許容領域上で凸になり，$\|T_{vz}\|_\infty < \gamma$ を満たすコントローラ K が存在することと不等式制約 $G(K) \le 0$ が Slater の制約想定を満たすことは等価である．

汎関数 F, G が凸ならば，非線形計画法あるいは微分不可能最適化手法により，問題 (3) に対して最適コントローラ K が満たすべき必要十分条件を求めることができる．しかしながら，最適コントローラの形状およびこれに対する F, G の凸性をあらかじめ仮定することはできない．コントローラを例えば状態フィードバックに限定 (このとき z の v に対する線形性，G の凸性は成立する) して閉ループゲイン決定問題 (3) を解くことはできるが，一般には問題 (3) においてコントローラ K のかわりに制御入力 u を決定変数とした開ループ制御問題をまず解析し，最適制御 u が満たすべき必要十分条件からこれを生成する最適コントローラの構造を導き出す必要がある．

例えば，Shimizu and Ito (1995) では，目的関数 F として min-max 型の評価基準を持つ設計問題

$$
(4) \quad \begin{cases} \min_u \max_v \int_0^\infty x(t)^T Q x(t) + u(t)^T R_1 u(t) - v(t)^T R_2 v(t)\, dt \\ \text{subj. to} \quad \dot{x}(t) = Ax(t) + B_1 u(t) + B_2 v(t), \quad x(0) = x_0 \\ \qquad \max_v \int_0^\infty z(t)^T z(t) - \gamma^2 v(t)^T v(t)\, dt \le 0 \\ \qquad \text{subj. to} \quad \dot{x}(t) = Ax(t) + B_1 u(t) + B_2 v(t), \quad x(0) = 0 \\ \qquad\qquad z(t) = Cx(t) \end{cases}
$$

について考察し，その最適解が一種の状態フィードバックコントローラで与えられることを示した．このとき F, G の凸性，z の v に対する線形性が成立し，問題 (4) の制約条件は H_∞ ノルム条件 (2) に一致する．

参 考 文 献

Shimizu, K. and Ito, S. (1995). Satisfactory optimal control: min-max control under the H_∞ norm constraint, *Proceedings of the 1995 American Control Conference*; **6**, 4445–4450.

J 言語による統計解析—カテゴリカルデータの散布度—

鈴 木 義 一 郎

順序尺度のカテゴリカルデータが与えられている場合に，i 番目のカテゴリーの観測度数を f_i とする $(i = 0, 1, \ldots, m)$．隣あったカテゴリー間の距離を 1 と考えると

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| \sim \sum_{i \ne j}^{m} |i - j| f_i f_j
$$

といった対応になる．

そこで，次のような第 1，第 2 累積度数を定義する．

$$
g_i = \sum_{i=j}^{m} f_i, \quad h_k = \sum_{j=k}^{m} g_j = \sum_{j=k}^{m} (i - k + 1) f_i
$$

$$v = \frac{2}{n(n-1)} \sum_{k=1}^{m} f_{k-1} h_k, \;\; n = \sum_{j=0}^{m} f_i$$

で与えられるものが，順序尺度データの散布度を表わすことになる．

次に順番はないが，便宜上カテゴリーに番号を付して，i カテゴリーの観測度数を f_i とする $(i = 1, 2, \ldots, m)$. カテゴリー間には距離のようなものがないから

$$\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| \sim \sum_{i \neq j}^{m} \sum_{j=1}^{m} f_i f_j = n^2 - \sum_{i=1}^{m} f_i^2 \;\; (n = \sum_{i=1}^{m} f_i)$$

のように考えると

$$u = \{n - (\sum f_i^2)/n\}/(n-1)$$

といったものが平均差に対応する散布度と解釈できる．

これらの散布度の指標を，『日本人の国民性調査』のいくつかの質問の回答パターンに適用してみた結果を例示した．

人間乱数

伊 庭 幸 人

人間にできるだけランダムに数字を書かせることで生成される数列の性質についての研究を発表した．本研究は田中美栄子氏 (相山女学園大学) との共同研究であり，その一部は平岡千佳，可児美佳子の両氏の同大学における卒業研究に基づくものである．内容及び関連文献については下記を参照されたい．

参 考 文 献

伊庭 幸人, 田中 美栄子, 平岡 千佳, 可児 美佳子 (1996). 人間乱数, 物性研究, **66**(5), 914-924.

Iba, Y. and Tanaka-Yamawaki, M. (1996). Statistical analysis of human random number generators, Methodologies for the Conception, Design, and Application of Intelligent Systems, *Proceedings of IIZUKA '96*, **2**, 467-472, World Scientific, Singapore.

情報量規準による ranking data 解析, 続報

石 黒 真 木 夫

イギリスの作家 Arthur Ransome が書いた 12 冊の本「ツバメ号とアマゾン号」「ツバメの谷」「ヤマネコ号の冒険」「長い冬休み」「オオバンクラブの無法者」「ツバメ号の伝書バト」「海へ出るつもりじゃなかった」「ひみつの海」「六人の探偵たち」「女海賊の島」「スカラブ号の夏休み」「シロクマ号となぞの鳥」のファンのクラブで実施されたアンケート調査の結果がある．質問項目に「あなたの好きな三冊を (好きな順に) あげてください．」と「あなたがいちばんすきでな

い巻はどれですか?」が含まれていた. このデータからファン・クラブにおける本の人気の順を推定するために回答者が n 冊の本に $j_1 > j_2 > \cdots > j_n$ という順序をつける確率が

$$P_n^{\text{top}}(j_1, j_2, \ldots, j_n) = P(j_1) \times \frac{P(j_2)}{1-P(j_1)} \times \frac{P(j_3)}{1-P(j_1)-P(j_2)} \times \cdots \times \frac{P(j_n)}{1-P(j_1)-\cdots-P(j_{n-1})}$$

で与えられ, $j_1, j_2, \ldots, j_T$ が同順位の1位であるとする確率が

$$P_T^{\text{tie}}(j_1, j_2, \ldots, j_T) = \sum_{(j_1, j_2, \ldots, j_T)\text{のすべての並べ替え}} P_T^{\text{top}}(j_1, j_2, \ldots, j_T)$$

であるとする "Luce モデル" をあてはめて

「冬休み」(0.159) > 「ツバメ号」(0.144) > 「伝書バト」(0.137) > 「ツバメの谷」(0.120) 「海へ」(0.120)
> 「スカラブ」 > 「ひみつ」 > 「オオバン」 > 「シロクマ」 > 「ヤマネコ」 > 「六人」 > 「女海賊」

なる結果を得た. かっこ内の数字が回答者がその本を一位にあげる確率の推定値である. この結果の AIC は 1464.8 であった. ファン・クラブが好みの構造を異にする2種類の人達の集まりであるとするモデルを仮定した時の AIC は 1454.8 となり, およそ 80 %が

「伝書バト」(0.167) > 「冬休み」(0.159) > 「海へ」(0.113) > 「ツバメの谷」(0.100) >
「ツバメ号」(0.097) > 「スカラブ号」(0.095) > 「ひみつ」 > 「オオバン」 >
「シロクマ」 > 「六人」 > 「ヤマネコ」 > 「女海賊」

という好みの順を持ち, 20 %が

「ツバメ号」(0.6504) > 「ツバメの谷」(0.1308) > 「海へ」(0.0841) > 「冬休み」(0.0826) >
「伝書バト」 > 「シロクマ」 > 「ひみつ」 > 「ヤマネコ」 > 「スカラブ号」 > 「六人」 「オオバン」
> 「女海賊」

という順で好むという結果を得た. このモデルを利用して rating の結果を解析する方法も紹介した.

非定常信号処理と関連研究

瀧 澤 由 美

非定常信号のスペクトル解析法として開発した瞬時化最大エントロピー法 (Instantaneous Maximum Entropy Method: IMEM) を用いて, 自動車エンジンの振動信号を解析し, スペクトル, エントロピーの時間応答から故障診断を行う研究を行った.

自動車・航空機等のエンジン故障の事前検知は安全性の上で重要で, エンジンオイル・冷却水・排気の温度や圧力等の異常を監視し, 警報を発する故障診断システムが実用化されている. しかし, 温度の上昇等は欠損・摩耗等がかなり進行した後でないと現れない. このため故障の内容や程度により事前に検知することが困難であった. ところが習熟した技能者や操縦者はエ

ンジンの音や振動の微妙な変化を感知し，故障の兆候を知る場合が多くある．このことから振動・音響信号に着目した．しかし，これらの信号は走行状態に対応して時々刻々変化する非定常信号であり，従来の定常を仮定した方法では故障の特徴を検出することが困難であった．

本研究では，非定常信号のスペクトルを追跡的に解析するため，まず最大エントロピー法に基づき評価関数を瞬時化し，次に変化するスペクトルを2次元スペクトルとしてとらえ，非定常性を新たなもう一つの周波数軸に投影してそれに帯域制限を行った．これにより導かれるIMEMを用いて故障状態のエンジンの振動信号を解析した．エンジン内部の回転部分の欠損に対応したスペクトルが観測でき，さらに同時に得られる情報エントロピーの値は故障状態の有効な指標となることを示した．

参考文献

Takizawa, Y., Fukasawa, A. and Yamano, C. (1995). Nonstationary spectral analysis method and its application to turbo charger of automobile engine, *Proceedings of IECON95*.

スギ人工林資源活用の支援情報

（客員）新潟大学 農学部　阿 部 信 行

はじめに

本研究では，GIS(地理情報システム) の特質を生かして，地図情報と属性データとの結合による新たな視点からの森林資源の有効活用に役立つ情報提供例について検討した．

解析方法

解析対象地は，新潟県加茂市管内の民有林約210haである．この民有林を対象に，1/5,000の基本図を用いて，林班，小班，林道，等高線の各位置をデジタイザを用いてデジタル化を行った．等高線に対しては，標高値をキィーボードから入力した．デジタル化された地図情報を基に、GISを利用し，伐期に達したスギ人工林を対象に，利用しやすい度合を基に，13タイプの施業タイプに分けた．林業は，斜面傾斜角の影響を強く受ける．また，伐採木を搬出することを考慮すると，伐採現場から林道までの距離が作業能率に大きな影響を与える．そこで，林道に10 mごとのバッファをかけて，平均集材距離を算出し，それと斜面傾斜角を組み合わせて，(1) 平坦地 (斜面傾斜角6度以下) を対象に，平均集材距離を4区分，(2) 緩傾斜地 (7～20度) を対象に，平均集材距離を5区分，(3) 傾斜地 (21度以上) を対象に，平均集材距離を4区分，計13タイプの作業区分を行った．

解析結果

50年以上のスギ人工林 (62.4ha) を対象に，各タイプ別の面積を求めると，タイプ (1) から順に，3.38ha，15.74ha，43.3haとなる．従って，傾斜地が非常に多く，ハーベスタ等の高性能機械の導入が困難であることが分かる．各タイプにおける平均集材距離の変動を見てみると，タイプ (1) では約50 %が126-400mであることが分かった．高性能林業機械が稼働できる地帯では，この集材距離は作業能率にそれほど大きな影響は与えない．一方，タイプ (3) は傾斜地であり，チェンソー等を利用した従来型の作業システムを取らざるを得ない．搬出距離が長いと，作業能率に大きな影響を与える．最も面積の比率が大きいのは，201-400mの平均集材距離を示すものであった．

まとめ

GISを用いた支援情報では，手計算では不可能な基準化が可能であること，作業に即した分類が容易にできること，分かりやすい地図作成が可能であること等を示すことができた．

統計データ解析センター

モンテカルロ平滑化について

北川源四郎

非線形・非ガウス型の状態空間モデル

$$x_n = F(x_{n-1}, v_n), \quad y_n \sim H(\cdot|x_n)$$

の未知パラメータをθとする．このとき，状態ベクトル x_n にθを付加し、$z_n = (x_n^T, \theta^T)^T$ で新たに状態ベクトルを定義すると自然に z_n に関する状態空間モデルが導かれる．この拡大されたモデルに対して非線形平滑化を行うと状態と未知パラメータの同時推定が実現できる(自己組織型のモデリング)．ただし，そのためには高度に非線形な高次元の状態空間モデルに対して適用可能な非線形フィルタおよび平滑化の方法が必要となる．モンテカルロ法を用いて平滑化を行うモンテカルロ平滑化のアルゴリズムはこのような目的に適している。

ただし，モンテカルロ平滑化のアルゴリズムには粒子の集中化に伴う近似精度の悪化の問題があり，自己組織型のモデリングの実用化のためにはこの問題の解決が必要であるが，その研究の過程ですべての粒子が確率1で事後周辺分布の平均値に収束するような新しいアルゴリズムが得られた．この方法は適当な規則で粒子の重み付き平均で新しい粒子を生成する．特殊な場合には m を粒子数とするとき，粒子の分散と共分散がそれぞれ漸化式

$$\begin{aligned} d_n &= \frac{3}{m} + \frac{2}{3}\left(\frac{2}{3} - \frac{1}{m}\right)^{n-1} - \frac{3}{m}d_{n-1} \\ c_n &= \frac{3}{m} + \frac{2}{m}\left(\frac{2}{3} - \frac{1}{m}\right)^{n-1} - \frac{3}{m}c_{n-1} \end{aligned}$$

で与えられ，したがってすべての粒子が一点に集中することがわかる．

参考文献

Kitagawa, G. (1995). Self-organizing state space model, Research Memo., No.578, The Institute of Statistical Mathematics, Tokyo.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Comutational and Graphical Statistics*, **5**(1), 1-25.

予測分布の理論と時系列のスペクトル推定について

駒木文保

パラメトリックモデルによる定常ガウス時系列のスペクトルの推定の問題で，一般に行われているスペクトル密度推定の方法が漸近的に改良できること，その補正がある意味で最良のも

のであること，を示した．特に AR モデルの場合を取り上げて改良の方法を具体的に与えた．また，計算統計学支援システムを利用したシミュレーションによる補正の効果の評価を行った．

スペクトル密度のモデル $\{S(\lambda;\theta):\theta\in\Theta\}$ が真の時系列のスペクトルを充分良く近似しているとする．このとき，パラメータ θ の適当な漸近有効推定量 $\hat{\theta}$ を $S(\lambda;\theta)$ に代入した $S(\lambda;\hat{\theta})$ によりスペクトル密度を推定することが通常行われている．ここで，スペクトルの推定量 $\hat{S}(\lambda)$ を Kullback-Leibler ダイバージェンス $D(S,\hat{S})=(1/4\pi)\int\{S/\hat{S}-1-\log(S/\hat{S})\}d\lambda$ を損失関数としてリスク関数 $E_{\theta_0}[D\{S(\lambda;\theta),\hat{S}(\lambda)\}]$ で評価することを考える．モデル $\{S(\lambda;\theta)\}$ をより大きなスペクトル密度の族 $\{S(\lambda;\theta,n)\}$ に埋め込み，スペクトル密度 $S(\lambda;\hat{\theta})$ をモデルに直交する方向にシフトして得られる $S(\lambda;\hat{\theta},\hat{n}_{\mathrm{opt}})$ で推定をすることにより $S(\lambda;\hat{\theta})$ を改良できることを示した．この議論は Amari (1987) で導入されている定常ガウス時系列モデルの微分幾何学的枠組みを利用することにより見通しよく行える．最良のシフトを表すベクトルと改良されるリスクはそれぞれスペクトル密度全体の空間の中でのモデルの θ での m-平均曲率ベクトルと m-平均曲率とみなせる量であることを示した．

この研究は，独立同一分布の場合に予測分布の問題として知られている問題についての対応する理論 (Komaki(1996)) の時系列版とみなせるものである．

参考文献

Amari, S. (1987). Differential geometry of a parametric family of invertible linear systems— Riemannian metric, dual affine connections, and divergence, *Math. Systems Theory*, **20**, 53-82.

Komaki, F. (1996). On asymptotic properties of predictive distributions, *Biometrika*, **83**, 299-313.

語順のシミュレーション・モデル

上 田 澄 江

言語によって語順規則はかなり異なる．しかし，ランダムな並び方をしているわけではない．例えば，所有格，指示詞，数詞，形容詞と名詞の順序をみると，私の家，このペン，3 冊の本，美しい景色，と日本語では一貫して修飾語が先行する．逆にタイ語では一貫して後行する．英語にはこのような一貫性はない．角田の語順の表 (角田 (1991)) の主成分分析はこの 3 言語が 3 極端の語順規則を代表していることを示している. また，130 言語の語順の表の解析から「名詞と側置詞」および「名詞と数詞」の語順が最も分類に関わっているようにみえる (上田・伊藤 (1995))．これらの語順の特徴をシミュレーションによってとらえることを試みた．また消滅モデルおよび変遷モデルの生成を試みた．

1. 語順データを乱数によって生成
   (a) 19 項目の語順をランダムに生成した場合．
   (b) 相関を入れて生成，すなわち項目 (1, 2, 7, 9, 10, 19), 項目 (4, 5, 6, 12), 項目 (14, 16) に高い相関を与えて生成した場合．
2. 言語の消滅モデル
   (a) 言語 A は一定距離 d 内にある言語 (A を含む) の 1 つ Bに変化する場合．

$$\mathrm{dist}(A,B)\le d,\quad d=1,2,3,4,5.$$

3. 1 言語に注目したときの言語の変遷モデル

(a) 言語が別の言語に変化したとき元の言語が消滅する場合.

(b) 言語が変化したとき元の言語が重複して存在する場合. 従って, 同一言語が何度でも選択される.

1. において語順の表との混合データの主成分分析では, ランダム生成モデルでは先の3言語から離れた中心部に集中して分布した. 一方, 相関を考慮したモデルでは3言語の周辺近くに, 且つ内側に分布し中心部は殆ど空白地帯となった.

参考文献

角田 太作 (1991).『世界の言語と日本語―言語類型論からみた日本語―』, くろしお出版, 東京.
上田 澄江, 伊藤 栄明 (1995). 語順規則による言語の分類と2パラメータモデル, 統計数理, **43**(2), 341-365.

統計数理研究所の計算機環境

田 村 義 保

統計数理研究所に初めて計算機が導入されたのは1956年のことである. 富士通のリレー式計算機の第一号機が導入されている。現在の計算速度とは比べようもないくらい遅いものであるが, その当時の統計数理研究所彙報を見ると, 所員の興奮が伝わってくる. また, 現在も統計数理研究所固有の装置としてある物理乱数発生機の1号機が接続されていた. この1号機の乱数源は放射能であったようであるが, 乱数が統計科学に必須のものであると40年前から考えていたことが読みとれる. その後, パラメトロン計算機を経て, LSIを用いた電子計算機と変わって行くが, 常に先進的で, その時点で最高速の計算機環境を構築しようと試みていたように思える.

現在の計算機は統計科学計算機システムと計算統計学支援システムを中核としている. 残念ながら, 最高速のシステムとは言えないが, 大容量のスカラー計算機, ベクトル計算機, 並列計算機とバランスのとれた構成にはなっている. また, イーサネットを用いているために, 高速性は要求できないが, 所内LANも整備しており, 上記2システム, 統計データ解析センター管理のワークステーションのみならず各研究室にあるワークステーション, パーソナルコンピュータが接続されている. データの共有化や計算資源の有効利用が可能になっている. また, SINETを通してインターネットに接続されており, 電子メールやWWWの利用が可能になっている. 統計データ解析センター管理のワークステーションにはLaTeXのようなPDSやSAS, Mathematica, Matlab, S等のソフトウェアがインストールされており, 所員及び共同利用者の便を図っている.

計算機システムの運営をしている立場としては言えることではないかもしれないが, まずまずの環境であると思う. 勿論, 計算機の速度から見ると, ベクトル計算機は遅すぎて話しにならない. また, PDSのインストールも十分ではないと思う. さらに, 利用可能である機能が利用者に十分に伝わっていないようにも思え, 残念でしかたがない.

計算機についての技術革新の速度が速すぎるために, 他の機関と比べて見劣りしない環境を維持していくには, 今のメンバー数では不可能に近いように思える. 不可能を可能にしている技官の方々にこの場を借りて御礼を述べたい. これからも統計数理研究所の計算機環境が向上

していくことを願っている．計算機環境の変遷の詳細については統計数理研究所50年のあゆみを参考にして欲しい．

欠測値処理のアルゴリズムについて

荒 畑 恵 美 子

観測機器の故障や観測システムの物理的制約などにより，時系列の一部分が観測できないことがある．このようなとき，観測できなかったデータのことを欠測値 (missing value) という．欠測値のあるデータでも，自動的に処理出来るような方法が必要である．状態空間モデルとカルマンフィルタを用いることにより，時系列に欠測値が含まれるときにも，パラメータを推定することが出来る．推定されたモデルを用いて，欠測値を推定することが出来る． 観測値が欠測のときは，カルマンフィルタにおいて，フィルタの部分を省略するだけでよい．すなわち，予測のステップはすべてについて，フィルタのステップは観測値のあるところについてだけ実行すればよい．欠測値を含む，一変量時系列に，AR モデル (又は ARMA モデル) をあてはめて，ユールウオーカー法 (Yule-Walker 法)，Parcor 法，状態空間表現法を用いて，モデルのパラメータを推定してみた．推定したモデルの良さは，カルバック-ライブラー情報量で評価した．また，スペクトルがどのようになっているかもみた．Canadian Lynx Data を用いて実験してみた．その結果，状態空間表現を用いたのが一番良いことがわかった．これは，統計数理研究所の北川源四郎教授との研究の一部分である．

統計教育・情報センター

真核生物の起源

橋 本 哲 男

原生生物や古細菌の系統的位置を明らかにすることは，真核生物の起源を探るうえで重要であるが，この問題はこれまで主に小亜粒子リボソーム RNA(SrRNA) の配列比較に基づいて行なわれてきた．我々は，この問題をより多くのデータに基づいて再検討するため，原生生物のペプチド鎖伸長因子などの遺伝子解析を通して配列データを蓄積するとともに，多くの保存的蛋白質による総合評価を行なうことを目指して研究を進めてきた．現時点で利用できる全てのデータセットに対し，さまざまなアミノ酸置換確率モデルに基づく最尤法を適用して詳細な分子系統解析を行なった結果，現在までに以下のことが明らかとなっている (Hashimoto and Hasegawa (1996))．

[1] ミトコンドリアをもたない原生生物のうちのある種のものが，真核生物の進化の非常に早い時期に，他の高等真核生物に至る系統から分岐したという可能性が高いが，これらがミトコンドリアの共生以前の真核生物の祖先型に近い生物であるかどうかは，今後さらに検討が必要である．

[2] 古細菌が単系統群を成すという可能性はあまり高くなく，硫黄細菌 (*Sulfolobus*) は，他の古細菌よりはむしろ真核生物に近いという傾向を示す．すなわち，硫黄細菌が真核生物の起源となったとの可能性が示唆される．

[3] 保存的蛋白質のアミノ酸組成が生物種間で大きく偏ることはほとんどないため，保存的蛋白質に基づく解析は，塩基組成の極端な偏りの影響が問題となる SrRNA に基づく解析に比べ，より頑健な推定を与えるものと考えられる．

参 考 文 献

Hashimoto, T. and Hasegawa, M. (1996). Origin and early evolution of eukaryotes inferred from the amino acid sequences of translation elongation factors 1α/Tu and 2/G, *Advances in Biophysics*, **32**, 73-120.

連続して起こる事象における確率分布

内 田 雅 之

オーダー k の離散確率分布において、興味深い点の１つに連の数に関係する分布，例えば，オーダー k の二項分布や オーダー k の負の二項分布などは連の数え方によってその分布が異なることがあげられる．また，最近では，マルコフ従属を仮定したモデルの下でのオーダー k の離散確率分布論の研究が盛んに行なわれている．本年度の研究の要旨は以下の通りである．

(1) $X_0, X_1, \ldots$ を 0 か 1 のいずれかの値をとる time-homogeneous な first-order マルコフ連鎖とする．まず，確率変数 $X_1, X_2, \ldots, X_n$ の系列において，ちょうど長さ k の 1 の連の起こる回数の確率関数とその漸化式，確率母関数の漸化式を与えた．これは，ちょうど長さ k の 1 の連を数える数え方によるオーダー k の二項分布のマルコフ連鎖バージョンである．さらに、$X_1, X_2, \ldots$ において、長さ k の 1 の連が起こるまでに起こるちょうど長さ l $(= 1, 2, \ldots, k)$ の 1 の連の回数の分布を求めた．

(2) $X_{-1}, X_0, X_1, X_2, \ldots$ を 0 か 1 のいずれかの値をとる time-homogeneous な second-order マルコフ連鎖とする． この時，$X_1, X_2, \ldots$ において，長さ k の 1 の連が起こるまでに起こる長さ l $(= 2, 3, \ldots, k)$ の 1 の連の回数の同時分布を求め，さらにその周辺分布は first-order マルコフ連鎖の場合とパラレルな結果であることがわかった．連の数え方はちょうど長さ l の連を数える数え方，長さ l 以上の連を数える数え方，長さ l の連を重複しないで数える数え方，そして長さ l の連を重複して数える数え方の４つの数え方を用いた．

参 考 文 献

Uchida, M. (1996a). On number of occurrences of success-runs of exact length k, Research Memo., No.597, The Institute of Statistical Mathematics, Tokyo.

Uchida, M. (1996b). Joint distributions of numbers of success-runs until the first consecutive k successes in a second-order two-state Markov chain, Research Memo., No.598, The Institute of Statistical Mathematics, Tokyo.

Fokker-Planck 方程式の高精度化をめざして

岡崎 卓

1. 初めに

外乱を受けつつ発展する系変数の確率密度を定めるには，Fokker-Planck(FP) 方程式を用いるのが通例である．しかしこの方程式は外乱がもつ情報の一部しか考慮しない．これを改良した一般化 Fokker-Planck(GFP) 方程式は厳密に外乱を考慮するが，実際の適用にあたって，その拡散項を具体的に表現する必要がある．摂動展開と繰込み近似によって拡散項の簡約を試みる方法は，その煩雑さの故に外乱の2次および3次相関を考慮するに留まり，flatness 等，高次の非ガウス性を扱うことができない．

そこで，外乱の非ガウス性をより忠実に反映する精密な FP 方程式の建設をめざし,「指数演算子の分解」を応用して GFP 方程式の拡散項を具体的に表現する方法について報告する．

2.GFP 方程式と拡散係数

系変数 u の確率密度 $f(u,t)$ を定める GFP 方程式の拡散項 I は，拡散係数 D と f の積分として与えられる．

$$
\begin{aligned}
I &= \int dv\, D(u,v,t,s)\frac{\partial}{\partial v} f(v,s) \\
D &= \int dU \int dW\, g(W)[(1-\mathbf{p})\mu(U,W)\psi_v]G(t-s)(1-\mathbf{p})\mu(u,W)\psi_u
\end{aligned}
$$

ここに $g(W)$ 、$\mathbf{p}$ および μ はそれぞれ，外乱 W の確率密度，射影子，および系と外乱の相互作用関数であり，また拡散係数 D の核 G は FP 作用素 $i\hat{L}$ を肩にもつ指数演算子

$$G(t) = e^{ti\hat{L}(1-\mathbf{p})}$$

である (他の記号については末尾参照)．外乱の発展，系の固有な発展，外乱の作用による系の発展を表す3種の微分作用素からなる FP 作用素

$$
\begin{aligned}
&i\hat{L} = i\hat{L}_W + i\hat{L}_U + \mu D_U \\
&\left(i\hat{L}_W = N(W)D_W + \frac{\sigma^2}{2}D_W^2, \quad i\hat{L}_U = M(U)D_U \right)
\end{aligned}
$$

の構造と，射影子が $\mathbf{p}X(U,W) = \int dW\, g(W)X(U,W) = \langle X(U,W) \rangle$ と定義され，従って $i\hat{L}_W\,(1-\mathbf{p})$ や $\mu D_U(1-\mathbf{p})$ が非可換である事実により，拡散係数 D を正確且つ具体的に表現する手段は現時点では存在しない．

3. 指数演算子の分解

拡散項の具体的表現を得るため，作用素 A と B の指数演算子を

$$e^{t(A+B)} = e^{\frac{t}{2}A}e^{tB}e^{\frac{t}{2}A} + O(t^3)$$

と近似する指数演算子分解法を適用する．この分解により

$$e^{ti\hat{L}(1-\mathbf{p})} = e^{\frac{t}{2}(i\hat{L}_W + i\hat{L}_U)(1-\mathbf{p})} e^{t\mu D_U(1-\mathbf{p})} e^{\frac{t}{2}(i\hat{L}_W + i\hat{L}_U)(1-\mathbf{p})}$$

を得るが，更に右辺中央の演算子にも重ねて分解を施す．

$$e^{t\mu D_U(1-\mathbf{p})} = e^{\frac{t}{2}\mu D_U} e^{-t\mu D_U \mathbf{p}} e^{\frac{t}{2}\mu D_U}(1-\mathbf{p}) + \mathbf{p}$$

4. 拡散項の表現

上記の指数演算子分解により，核 G と拡散項の積分 I を具体的に書き下すことができる．簡単のため，外乱が定常で且つ additive （$\mu = \mu(W)$ ）な場合を考えると，拡散項は外乱に関する3個の特別な統計量と確率密度 f との convolution として表現される．

$$\begin{aligned}
I &= I_A + I_B \\
I_A &= e^{-si\,L_U} D_U \int dV\, \Phi_A(V,s)\, f(U-V, t-s) \\
I_B &= e^{-si\,L_U}\, D_U D_U \\
&\quad \cdot \int dV_1 \int dV_2\, \Phi_B^{(1)}(V,s)\, \Phi_B^{(2)}(V,s)\, f(U-V_1-V_2, t-s)
\end{aligned}$$

ここに統計量 Φ_A、$\Phi_B^{(1)}$、$\Phi_B^{(2)}$ は

$$\begin{aligned}
\Phi_A(V,s) &= \Big\langle (\mu[0] - <\mu>)\delta\Big(V - sJ(s)\mu[s/2]\Big)\,(\mu[s] - <\mu>)\Big\rangle \\
\Phi_B^{(1)}(V,s) &= \int_{-\infty}^{\infty} d\tau\, \theta(-\tau) \Big\langle (\mu[0] - <\mu>)K(s)\delta\Big(V - \frac{s}{2}K(s) + (s+\tau) <K(s)>\Big)\Big\rangle \\
\Phi_B^{(2)}(V,s) &= \int_{-\infty}^{\infty} d\tau\, \theta(-\tau) \Big\langle \delta\Big(V - \frac{s}{2}K(s) + \tau <K(s)>\Big)(\mu[s] - <\mu>)\Big\rangle
\end{aligned}$$

なる構造をもっており，通常の相関関数と異なって無限の次数の自己相関関数を含む．これらの統計量を介して外乱の情報を組み込む GFP 方程式は，それだけ外乱の非ガウスを忠実に反映するものと期待される．

[記号註]
$\psi_u = \delta(U-u)$ 、$D_U = \frac{\partial}{\partial U}$、$D_W = \frac{\partial}{\partial W}$、$\theta(t) = 1\ (t > 0)$, $\theta(t) = 0\ (t < 0)$ 、
$D_U J(-t) = e^{ti\hat{L}_U} U e^{-ti\hat{L}_U}$、$\mu[t] = \mu(W(t))$ 、$K(s) = J(s/2)\mu[s/2]$.

非集計分析における不確実性の表現法

山　下　智　志

所要時間の不確実性の影響

旅行者にとって所要時間の不確実性は不効用であり，経路選択，機関選択，出発時刻選択に影響を及ぼしているという指摘は Jucker をはじめ多くの研究で指摘されている．しかし需要予測におけるこの問題の表現方法は，いまだ確立されているとはいいがたい．本研究では，不確実性下の意志決定問題について，経済学 (企業財務論) との比較を行いながら，旅行者の不確実性下の機関選択・出発時刻選択問題について考察する．

所要時間に関する効用関数

実際に非集計行動モデルを適用する場合，所要時間の不確実性をどのように効用関数に反映させるかが問題になる．不確実性に対する消費者行動の計量化についての最も基礎となる概念

は Neurmann = Morgenstern により提案された期待効用の原則である．これを単純に所要時間の不確実性に適用すると以下の式になる．

$$(1)\qquad \begin{aligned} V_j &= u_j(\tilde{t}) + \sum kZ_j. \\ &= \int f_j(t)u(t)dt + \sum kZ_j. \end{aligned}$$

V: 非集計モデルの確定効用項，$f_j(t)$: 交通手段 j の所要時間分布, Z: 個人属性等その他の要因

しかし，Jucker の研究では旅行者の期待効用は Mean=Variance 型の期待効用関数と同様に，平均と分散の線形結合で表現されていた．

$$(2)\qquad \begin{aligned} V_j &= \alpha m_j + \beta\sigma_j^2 + \sum kZ_j. \\ m_j &: \text{交通手段 } j \text{ の所要時間の平均}, \quad \sigma_j^2\text{: 分散} \end{aligned}$$

これが成立する用件として NM 効用関数（$u_j(\tilde{t})$）が HARA 族型の効用関数でなければならない．しかし現実的に旅行者の効用関数をこのように仮定するには根拠が曖昧である．

時刻に対する NM 効用関数

現実の旅行者の結果効用は到着時刻に依存していると仮定しよう．このとき式 (1) は出発時刻を期待効用最大化原則に基づき，以下のように展開できる．

$$(3)\qquad \begin{aligned} V_j &= max_{T_S} E[u_j(T_S, \tilde{t})] \\ &= \int f(T_l|T_S, j)u(T_s, t)dT_l + \sum kZ_j. \end{aligned}$$

T_lは到着時刻 $(= T_S + t)$ であり，旅行者の NM 効用関数は出発時刻の効用関数と到着時刻の効用関数に分解される．

$$(4)\qquad \begin{aligned} u(T_s, \tilde{t}) &= u_s(T_s) + u_l(\tilde{T}_l) \\ &= u_s(T_s) + u_l(T_s + \tilde{t}) \end{aligned}$$

その結果，需要予測に必要なデータは，出発時刻ベースの NM 効用関数 $u_s(T_s)$，到着時刻ベースの NM 効用関数 $u_l(\tilde{T}_l)$，時刻 T_s で出発し交通手段 j を選択したときの到着時刻分布 $f(T_l|T_S, j)$ であり，これらのパラメータを非集計 logit モデルで推定する．

モデルの比較

式 (2)，式 (3) は$\sigma^2 = 0$，$f(T_l|T_S, j)$ の分散が 0 、$u_s(T_s)$ が線形 (出発地における時間価値は時刻に対して不変) の条件下で，同一のモデルとなる．実データへの適合度による比較では，式 (3) のモデルは実データへの適合度も高く，他のモデルより現実の行動を反映していることが確認されてた．

統計基礎研究系

離散確率分布と統計的応用

平　野　勝　臣

本年度の研究

(1) これまでの離散確率分布の研究のうち，k 連続成功が初めて起こるまでに起こる成功連の数と失敗数の同時分布についての結果をまとめた (Aki and Hirano (1995)).

(2) システムの信頼性に関する研究のうち，Consecutive-k-out-of-n:F システムの故障時刻の観測に基づいた母数の推定問題についてまとめた (Aki and Hirano (1996)).

以上について報告した．ここでは Aki and Hirano (1995) の結果について述べる．

長さ k の成功連が起こるまでの成功連の数と失敗数の同時分布

値 $j(j=1,2,\ldots,m)$ を確率 $p_j(\sum_{j=1}^m p_j=1)$ でとる iid の確率変数の系列 $X_1, X_2, \ldots$ において，τ を k 連続 1 が初めて起こるまでの waiting time，$\eta_j(j=1,2,\ldots,m)$ を τ までに j の起こる数とする．このとき $(\tau,\eta_1,\ldots,\eta_m)$ の pgf $\phi_2(r,t_1,\ldots,t_m)$ を陽の形で与えた．これから $E(\tau)$, $\mathrm{Var}(\tau)$, $E(\eta_j)$, $\mathrm{Var}(\eta_j)$, $\mathrm{Cov}(\tau,\eta_j)$, $\mathrm{Cov}(\eta_i,\eta_j)$ が得られ，$\phi_2(1,t_1,1,\ldots,1)$ から η_1 の周辺分布が $G_{k-1}(p,k)$ であることが導かれる．但し，$G_k(p,a)$ は台 $\{a,a+1,\ldots\}$ 上のオーダー k の幾何分布を表わす．

$\xi_l(l=1,2,\ldots,k)$ を τ までに長さ l の overlapping で数えた 1 の連の数とする．このとき $(\tau,\xi_1,\ldots,\xi_k,\eta_2,\ldots,\eta_m)$ の pgf $\phi_3(r,t_1,\ldots,t_k,s_2,\ldots,s_m)$ を陽の形で与えた．$\phi_3(1,\ldots,1,t_l,1,\ldots,1)$ から ξ_l の周辺分布が $G_{k-l}(p_1,k-l+1)$ であることが導かれる．

$\nu_l(l=1,2,\ldots,k)$ を τ までに長さ l 以上の 1 の連の数とする．このとき $(\tau,\nu_1,\ldots,\nu_k,\eta_2,\ldots,\eta_m)$ の pgf $\phi_4(r,t_1,\ldots,t_k,s_2,\ldots,s_m)$ を陽の形で与えた．$\phi_4(1,\ldots,1,t_l,1,\ldots,1)$ から ν_l の周辺分布が $G_1(p_1^{k-l},1)$ であることがわかる．

$\mu_l(l=1,2,\ldots,k)$ を τ までに長さ l の 1 の連の数とする．このとき $(\tau,\mu_1,\ldots,\mu_k,\eta_2,\ldots,\eta_m)$ の pgf $\phi_5(r,t_1,\ldots,t_k,s_2,\ldots,s_m)$ を陽の形で与えた．$\phi_5(1,\ldots,1,t_l,1,\ldots,1)$ から μ_l の周辺分布が $G_m(p_1^l,m+1)$, $k=l(m+1)$ であることがわかる．

確率変数の系列 $X_0, X_1, \ldots$ が $\{0,1\}$ 値のマルコフ系列に従う場合，η を τ までに 0 の起こる数とする．このとき $(\tau,\eta,\xi_1,\ldots,\xi_k|X_0=0)$ の pgf ϕ_6^0, $(\tau,\eta,\xi_1,\ldots,\xi_k|X_0=1)$ の pgf ϕ_6^1, $(\tau,\eta,\nu_1,\ldots,\nu_k|X_0=0)$ の pgf ϕ_7^0, $(\tau,\eta,\nu_1,\ldots,\nu_k|X_0=1)$ の pgf ϕ_7^1, $(\tau,\eta,\mu_1,\ldots,\mu_k|X_0=0)$ の pgf ϕ_8^0, $(\tau,\eta,\mu_1,\ldots,\mu_k|X_0=1)$ の pgf ϕ_8^1 を陽の形で求めた．また確率変数の系列 $X_1, X_2, \ldots$ がオーダー kの 2 値系列に従う場合，$(\tau,\eta,\xi_1,\ldots,\xi_k)$ の pgf ϕ_9, $(\tau,\eta,\nu_1,\ldots,\nu_k)$ の pgf ϕ_{10}, $(\tau,\eta,\mu_1,\ldots,\mu_k)$ の pgf ϕ_{11} を陽の形で求めた．

参 考 文 献

Aki, S. and Hirano, K. (1995). Joint distributions of numbers of success-runs and failures until the first consecutive k successes, *Ann. Inst. Statist. Math.*, **47**, 225-235.

Aki, S. and Hirano, K. (1996). Lifetime distribution and estimation problems of consecutive-k-out-of-n-F:systems, *Ann. Inst. Statist. Math.*, **48**, 185-199.

確率過程のノンパラメトリック推測について

西 山 陽 一

LeCam が創始した漸近有効推測の理論は，1980 年代に未知母数が無限次元である場合まで拡張された．その概略は次の通りである：(1) 統計実験が局所漸近正規性をもつことをチェックする；(2) 未知母数の微分可能性をチェックする；(3) すると，正則推定量の漸近有効性の限界点が計算できる．詳細は Bickel et al. (1993) や van der Vaart (1988) を参照．

この理論によると,「よい」推定量を構成し，それがステップ (3) で算出された「漸近有効性の限界点」を達成することを示せばよい．ところが，未知母数が ℓ^∞-空間の値をとる場合には，それは容易ではない．原因は，ℓ^∞-空間の中心極限定理が，独立同一分布列や強定常列の場合などの極めて限定された状況においてしか証明されていないからである．詳細は Wellner (1992) を参照．

そこで，ℓ^∞-空間値のセミマルチンゲールに対する中心極限定理を 4 タイプ用意し，その応用を議論した．詳細は拙著 Nishiyama (1996). を参照．

参 考 文 献

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore, Maryland.

Nishiyama, Y. (1996). Some central limit theorems in ℓ^∞-valued semimartingales and their applications, Research Memo., No.599, The Institute of Statistical Mathmatics, Tokyo.

van der Vaart, A.W. (1988). Statistical estimation in large parameter spaces, CWI Tract 44, Centrum voor Wiskunde en Informatica, Amsterdam.

Wellner, J.A. (1992). Empirical processes in action: a review, *Internat. Statist. Rev.* **60**, 247-269.

統計基礎モデルの構築と推定方程式

松 縄 規

パラメトリックな基礎モデルを構築する際に，推定方程式が果たす役割を考察した．

(1) まず，一変量の場合に，観測値の一般化算術平均を導入した．これと最尤方程式を組み合わせればパラメトリックな統計基礎方程式が導かれることを，理論展開の端緒とした．主要なパラメーター θ の推定が以下の量を含むいくつかの典型的な平均でなされる場合を考えた：

$$\hat{\theta} = \left(\frac{1}{n}\sum_{i=n}^{n} x_i^\alpha\right)^{\frac{1}{\alpha}} \qquad (\alpha \neq 0).$$

(2) 次に，上記の結果を一般化して，観測対象量とその近似関数の間で定義される観測誤差関数$\Delta(x;\theta,\tau)$ に関連した推定関数

$$\sum_{i=1}^{n}\Delta(x;\theta,\tau) = 0, \qquad \Delta(x;\theta,\tau) := \Phi(x;\theta,\tau) - \Psi(x;\theta,\tau)$$

を取り上げた．但し，上で τ は潜在パラメーターを，x は観測値を表す．この方程式を取り込んだパラメトリックな統計基礎モデルを構築することを考えた．それらの考え方は，離散型のモデルの構築にも適用可能であることも示した．

(3) 更に，以上の結果が多変量推定関数およびパラメトリックな多変量基礎方程式の構築へ拡張できることを示した.

(4) 以上の事柄を踏まえて，統計モデル構築の基盤として，観測誤差が，ある意味で，統計的ポテンシャルと見なされるべきものであることを指摘し，関連するスカラポテンシャルとベクトルポテンシャルを産み出す条件および各々のポテンシャルの表現を与えた．それらの条件は，ある種のベクトルの外積および内積を用いて与えられることを示した．また，これらの条件には，それぞれ，一般化指数分布族，一般化混合分布族が対応する事を示し，それらの分布型を構成できることも示した.

(5) 一般化の場合として，上記の二つのポテンシャルに関連する，ベクトルが互いに直交する場を構成し，その際のポテンシャルの表現をポアソン方程式と関連させて考察した.

マルチンゲールの漸近展開とその応用

吉 田 朋 広

確率変数列 $\{X_n\}$: $X_n = M_{n,T_n} + r_n N_n$ を考える．ここで、M_{n,T_n} は(一般にジャンプのある)局所マルチンゲール M_n の時刻 T_n での値で，r_n は0に収束する正数列，N_n は任意の確率変数である. $\xi_n = r_n^{-1}\left[\frac{1}{3}[M_n]_{T_n} + \frac{2}{3} < M_n >_{T_n} -1\right]$ とおく．$n \to \infty$ のとき，$(M_{n,T_n}, \xi_n, N_n) \to^d (Z, \xi, \eta)$ と仮定する．Y_n は X_n か M_{n,T_n} を表すものとするとき，適当な正則条件のもとで,

$$\sup_{x \in \mathbf{R}} |P(X_n \le x) - P_n(x)| \le C r_n^{-a_1} P(\sigma_{Y_n} < s_n)^{a_2} + o(r_n)$$

を得る．ここで，C, a_1, a_2 はある正定数，s_n は $\limsup_{n\in\mathbf{N}} E[s_n^{-p}] < \infty$ $(p > 1)$ を満足する確率変数で，σ_{Y_n} は Y_n の Malliavin 共分散を表し，

$$\begin{aligned} P_n(x) &= \int_{-\infty}^{x} \Big[\phi(z) + \frac{1}{2} r_n \partial_z^2 (E[\xi | Z = z]\phi(z)) \\ &\quad - r_n \partial_z (E[\eta | Z = z]\phi(z))\Big] dz. \end{aligned}$$

Convolution equivalent となるための条件

志 村 隆 彰

$[0, \infty)$ 上の分布 F は任意の $u \ge 0$ に対して $\lim_{t\to\infty} \overline{F}(t-u)/\overline{F}(t) = e^{\alpha u}$ が成立つとき (ここで，$\overline{F}(t)$ は分布の tail $F(t, \infty)$ である) 指数 $\alpha > 0$ の exponential tail をもつといわれる ($F \in \mathbf{L}_\alpha$ と表す)．緩慢変動の言葉を使えば，F が $\mathbf{L}_\alpha$ に属するとはある緩慢変動関数 l によって $\overline{F}(t) = e^{-\alpha u} l(e^t)$ と書けるものということになる．$\alpha = 0$ のときは long-tailed distribution と呼ばれる．Convolution equivalent な分布は $\mathbf{L}_\alpha$ の部分族として $\lim_{t\to\infty} \overline{F * F}(t)/\overline{F}(t) = 2d$ (ここで，$d = \int_0^\infty e^{\alpha u} F(du)$ である。$F \in \mathbf{S}_\alpha$ と表す) と定義される．問題は $\mathbf{L}_\alpha$ の中で，$\mathbf{S}_\alpha$ はどのように特徴付けられるか．すなわち，tail を $\overline{F}(t) = e^{-\alpha u} l(e^t)$ としたときの緩慢変動関数 l はいかなるものかということである．このためのアプローチとして0に収束する非増加緩慢変動関数 l に対して，$(l(x))^2/l(x^2)$ の $x \to \infty$ のときの挙動について考察し，以下の様な結果を得た．l の ε-関数が単調であるならば，これは0に収束する．しかし，一般には上極限は

正でありうる(下極限は常に0).さらにε-関数の単調性を仮定したときの0に収束する速さについて必ずしもl自体の0にいく速さとの相関があるわけではなく,ある意味で任意に遅くなりうることがわかった.このことを分布の問題に応用することで 非増加緩慢変動関数lで$\overline{F}_\alpha(t) = e^{-\alpha u}l(e^t)$により 分布$F_\alpha$のtailを定義したとき,これが$\alpha$が 0 のときも正のときも(正の時は$\int_0^\infty e^{\alpha u}F_\alpha(du) < \infty$を満たす)$\mathbf{L}_\alpha \setminus \mathbf{S}_\alpha$に入る分布となるものが存在することが示された.これは 新しいタイプの$\mathbf{L}_\alpha \setminus \mathbf{S}_\alpha$に入る分布の例である.また,$\alpha > 0$の場合に$\int_0^\infty e^{\alpha u}F_\alpha(du) < \infty$と$\overline{F}(t) = e^{-\alpha u}t^{-1}l_0(e^t)$($l_0$の$\varepsilon$-関数は単調)という形を仮定しても,$\mathbf{S}_\alpha$になるとは限らないこともいえた.これはClineの$\mathbf{S}_\alpha$に属するための必要条件を満たすもので ある意味で$\mathbf{L}\alpha$と$\mathbf{S}_\alpha$との境界にある例であると言える.

参考文献

Cline D. B. H. (1987). Convolution tails, product tails and domain of attraction, *Probab. Theory Related Fields*, **472**, 529-557.

志村 隆彰 (1995). Exponential tail をもつが convolution equivalent でない分布について, 統計数理研究所共同研究リポート, No.75, 59-61.

志村 隆彰 (1996). Convolution equivalent となるための条件について, 統計数理研究所共同研究リポート, No.88, 62-67.

調査実験解析研究系

意識,調査法

坂 元 慶 行

今年度は,(1)1993年10月に実施した「第9次 日本人の国民性調査」の分析結果の発表(坂元 他(1995)),(2)この第9次調査の精度の考察,(3)将来の国民性調査のための準備,等を行った.

以上のうち,(2)では,第9次調査を中心に,調査方式と調査結果の関係について考察した.国民性調査では,1953年以来,全国数十の拠点大学の学生調査員によって実査を行ってきたが,第8次調査(1988年)では回収率が急落した.しかし,回収率の回復を狙って安易に専門調査機関に委託すれば,調査方式の違いから,国民性調査の結果数値に処理不可能な時系列上の断層を招くのは必至であった.そこで,第9次調査では,調査員の調達・管理は委託するが,それ以外は可能な限り従来の調査方式を継続することによって,断層を一応回避することができた.

しかしながら,問題はこれで終りではない.たとえば(3)では,将来に備えて新しい質問作りを試みたが,その中で,一番大切なものを自由に挙げてもらう質問に対し,「家族」という答えは,1973年国民性調査以降,大幅増加を続けており,いずれ飽和状態に達する可能性もある.そこで,専門調査機関による1995年1月の吟味調査で「二番目に大切なもの」も聞いてみた.その結果,「二番目に大切なもの」は,「一番大切なもの」だけでは不可能な価値観の別の面を描写できる可能性をもつように思われたが,結論は留保せざるを得なかった.この調査だけでなく,1年前の1994年3月の別の専門調査機関の調査でも,「一番大切なもの」として「生命・健康」を選んだ比率は,国民性調査に比べて約15ポイント多く,これが他の回答肢の選択率に影響を及ぼしている可能性もあるため,詳細な分析は難しいと思われたからである.

われわれは頻繁に独自の調査を実施できるわけではなく,世の多くの世論調査も専門調査機関によって実施されており,今後もこの傾向は強まるものと思われるところから,それらの結

果を最大限活用するのが得策であることは言うまでもない．また，統計調査環境の悪化という背景もあって，最近，電話調査をはじめとする種々の調査方式も導入されている．したがって，各種の調査方式による調査結果の違い，その読み方といったことは今後もっと研究されるべき，古くて新しい課題であると思われる．

参考文献

坂元 慶行 他(1995). 特集 日本人の国民性調査, 統計数理, **43**, 1-176.

コウホート分析における識別問題再考

中村 隆

コウホート分析は，継続調査データを分析する方法の1つであり，人間集団に関する何らかの数量特性を年齢層×調査時点別に集計したコウホート表データから，年齢・時代・世代(コウホート)効果を分離する方法である．コウホート分析については，3効果のパラメータを一義に分解できないという識別問題が知られているが，ベイズ型コウホートモデルでは，各効果のパラメータの1次階差の2乗和を小さくするという付加条件—パラメータの漸進的変化の条件—によって識別問題を克服している．本年度は，この付加条件によってどのように識別問題が克服されるのかについて再考察を行なった．

調査間隔と年齢区分幅の等しいコウホート表を考え，調査時点数を J，年齢区分数を I とする．第 j 調査時点，第 i 年齢区分のセルの期待値を q_{ij} とすると，モデルは次のようにかける．

$$q_{ij} = \mu + \mu_i^A + \mu_j^P + \mu_k^C, \quad i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, K.$$

ここで，μ，μ_i^A，μ_j^P，μ_k^C はそれぞれ総平均効果，年齢，時代，コウホート効果である($k = j-i+I$，$K = I + J - 1$)．このモデルの何らかの意味の不定解は，ある解と任意の実数 α によって

$$\breve{\mu}_i^A = \mu_i^A - \left(i - \frac{I+1}{2}\right)\alpha, \quad \breve{\mu}_j^P = \mu_j^P + \left(j - \frac{J+1}{2}\right)\alpha, \quad \breve{\mu}_k^C = \mu_k^C - \left(k - \frac{K+1}{2}\right)\alpha,$$

のように表わすことができる．

各効果の重み付き1次階差の2乗和 S は

$$S = \frac{1}{\sigma_A^2}\sum_{i=1}^{I-1}\left\{(\mu_i^A - \mu_{i+1}^A) + \alpha\right\}^2 + \frac{1}{\sigma_P^2}\sum_{j=1}^{J-1}\left\{(\mu_j^P - \mu_{j+1}^P) - \alpha\right\}^2 + \frac{1}{\sigma_C^2}\sum_{k=1}^{K-1}\left\{(\mu_k^C - \mu_{k+1}^C) + \alpha\right\}^2$$

であり，α についての2次関数となる(2次以上の階差では α の関数にならないことに注意)．ここで，たとえば，σ_A^2 に小さな値を与えたとすると，S を小さくするためには $\{\mu_i^A\}$ の変動をあまり大きくできず，逆に σ_A^2 に大きな値を与えたとすると，$\{\mu_i^A\}$ の変動が大きくなっても S はそれほど大きくならない．

上のような機構により，ベイズ型モデルにおける超パラメータにあたる σ_A^2，σ_P^2，σ_C^2 を(ABICによって)適当に決めれば S が小さくなるような解，すなわち，パラメータの変動が小さく抑えられた節約原理に基づく解が得られ，コウホート分析における識別問題が克服できる．

参考文献

中村 隆(1995). 交互作用効果モデルと過大分散モデルを用いたコウホート分析—「日本人の国民性調査」データへの適用—, 統計数理, **43**, 99-119.

反復測定データの因子分析

前田 忠彦

同一特性を測定する複数個 (K個) の項目についての複数回 (J回) の測定データが得られている場合に，確認的因子分析の適用によって，$K \times J$ 個の変数間の共分散行列を以下の４つの因子に由来する成分に分解する方法について検討した: a)「共通因子」(測定目的の特性)，b)「項目特殊因子」，c)「時点特殊因子」，d)「独自因子」(測定誤差)，に由来する成分．

この問題について，Raffalovich and Bohrnstedt(1987) と Marsh and Grayson (1994) によって Model 1 〜 Model 4 までの分析方法が提案されているが，本研究では更に Model 5 による分析方法を提案し，5 つのモデルの優劣を比較した．なお，Model 5 は意味の明瞭性の点では他の 4 つのモデルに比べて明らかに優れている．

提案した Model 5 のモデル特定の方法は概略以下の通りである．$x_{k\cdot j}$ を第 j 時点で測定した第 k 番目の項目得点（平均 0 に調節済み）とし，そのベクトル表記を $\boldsymbol{x}$ として，

$$\begin{aligned} x_{k\cdot j} &= \lambda_{Ck\cdot j}\xi_C + \lambda_{Ik\cdot j}\xi_{Ik} + \lambda_{Tk\cdot j}\xi_{Tj} + \epsilon_{k\cdot j} \\ (\boldsymbol{x} &= \boldsymbol{\lambda}_C\xi_C + \boldsymbol{\Lambda}_I\boldsymbol{\xi}_I + \boldsymbol{\Lambda}_T\boldsymbol{\xi}_T + \boldsymbol{\epsilon}) \end{aligned}$$

と確認的因子分析モデルで表現する．ただし，ξ_C は「共通因子」，ξ_{Ik} は第 k 番目の項目に関する「項目特殊因子」，ξ_{Tj} は第 j 番目の時点に関する「時点特殊因子」，$\epsilon_{k\cdot j}$ は項目 $x_{k\cdot j}$ に関する「独自因子」(測定誤差) を表わす確率変数である．外生変数 $\boldsymbol{z} = [\xi_C\, \boldsymbol{\xi}_I{}'\, \boldsymbol{\xi}_T{}'\, \boldsymbol{\epsilon}']'$ の共分散行列 $\boldsymbol{\Phi}_z$ に適切な仮定をおくことにより，$x_{k\cdot j}$ の共分散行列$\boldsymbol{\Sigma}$ は

$$\boldsymbol{\Sigma} = \boldsymbol{\lambda}_C\boldsymbol{\lambda}_C{}' + \boldsymbol{\Lambda}_I\boldsymbol{\Lambda}_I{}' + \boldsymbol{\Lambda}_T\boldsymbol{\Lambda}_T{}' + \boldsymbol{\Psi}$$

と構造化される．$\boldsymbol{\Psi}$ (対角行列) は $\boldsymbol{\epsilon}$ の共分散行列である．上式の右辺の各項が前述の a)，b)，c)，d) の各成分に相当する．

先行研究に示された計 3 個のデータを再分析し，5 つのモデルの適合度を比較した結果，Model 5 は Model 3，Model 4 と少なくとも同等の適合を示し，分散成分の分解も従来のモデルとほぼ同様の結果が得られた．このことから，Model 5 が十分に実用的でもあることが示された．

参考文献

Marsh, H. W. and Grayson, D. (1994). Longitudinal confirmatory factor analysis: common, time-specific, item-specific, and residual-error components of variance, *Structural Equation Modeling*, 1, 116-145.

Raffalovich, L. E. and Bohrnstedt, G. W. (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data, *Sociological Methods and Research*, **15**, 385-405.

ランダム充填過程の一性質

種 村 正 美

われわれは以前,「先住者ほど大きな場所を確保できる?」という設問をランダム逐次充填の問題として取り組み,地理的に均質な土地にランダムに順次確保された場所の面積が,先住者ほど大きい傾向になるとの計算機シミュレーションの結果を示した(種村(1992)).ここでは,この問題を一次元空間において理論的に解析した.

一次元のランダム逐次充填過程は「ランダム駐車問題」として知られている.いま,長さ L の直線上の有界区間に幅1の線分を次々にランダムに詰めていくものとし,k番目の線分は,すでに詰められた $k-1$ 個の線分と重ならない部分区間に一様ランダムに配置される.この手続きは,許された部分区間が無くなるまで続ける.そして,最終的に得られた状態を「完全充填」という.完全充填の密度 c_L が $L \to \infty$ の極限で理論的に計算されていて(Rényi (1958)),$c_\infty = 0.74759\cdots$ であることが知られている.

さて,一次元のランダム逐次充填の手続きによって実現された線分の配置について,各々の線分に元の区間をすき間なくかつ重なり無く割り振ることにする.その一つの方法は,Voronoi線分に分割することである.これは,充填された線分の両側に隣接する線分との隙間を二等分するだけでよい.その理論計算をするためには,Rényiの方法と異なり,充填された線分配置の確率密度そのものの計算が必要である.。

この確率密度の計算のために,充填される区間は長さ L の円周とすると,最初の線分の中点の位置 x_1 は原点に固定できる.すると,第二の線分の中点の位置 x_2 の確率密度は $g_2(x_2) = 1/(L-2)$,第三の線分の位置 x_3 の確率密度は x_2 に依存して定まるという具合に,順次計算することが出来る.$L(>1)$ が大きくなるにしたがって,関数形は急激に複雑になっていくが,われわれは任意の L に対して確率密度を求めるための一般的手続きを与えた.

そして,$L \leq 7$ の各々の場合に,具体的にVoronoi線分長の期待値および分散を計算した.その結果,Voronoi線分長は相対定着順位が先頭のものから平均長を上回ったまま緩やかに減少し,後位になるほど線分長が平均長を急激に下回っていくという,以前の二次元での"経験則"を,一次元の場合に理論的に示すことが出来た.また,$L \leq 7$ のそれぞれにおいて,充填密度が計算でき,比較的小さなこれらの L に対して,c_∞ に極めて近いことも示された.

参考文献

Rényi, A.(1958). On a one-dimensional problem concerning random space filling, *Publ. Math. Inst. Hung. Acd. Sci.*, **3**, 109-127.

種村 正美(1992). 先住者ほど大きな場所を確保できる?— ランダムパッキングの一側面 —, 形の科学会報 **7**(2), 41-43.

パソコンUNIXによる統計学研究環境

丸 山 直 昌

IBM-PC互換パソコンにUNIXを実装し,統計学研究者の日常的な計算機需要を満たすシステムを作る研究を昨年に引き続いて行なっている.

今年度はノートパソコン上にFreeBSD 1.1.5.1を搭載して開発を行ない,インストール手順の簡略化,日本語ソフトウェア関連の設定の簡略化,プリンタドライバの日本語対応などを中

心にシステムの整備を行なった．これらはUNIXの専門家に取ってそれ程意欲をかきたてる性質のものではないにも拘らず，UNIXの素人が使う場合には非常に手を焼く種類の作業であり，「素人にも使えるシステムを供給する」という，この研究の趣旨からは重要である．

ここ半年ほどの新製品ラッシュにより，IBM-PC互換パソコンの性能は，いわゆるUNIXワークステーションの性能に完全に追い付いた感がある．FreeBSDもVersion 2.1が出て，安定性を増しつつある．ここで重要な点は，単にUNIXの専門家だけではなく，素人でも使えるようにソフトウェアを整備しておくことであり，今年度の成果は重要なステップである．来年度はこの成果をVersion 2.1に移した上で，各種統計ソフトウェア，特に統数研で開発されたものを搭載する作業を行なう予定である．

指数逆ガウス型分布について

金 藤 浩 司

実数上 $(-\infty,\infty)$ で定義された分布は多くあるが，その中で歪度がゼロでない分布は余り多く知られていない．最小値のGumbel分布は、μおよびσをそれぞれ位置母数および尺度母数とすれば

$$\exp\left(\frac{X-\mu}{\sigma}\right) \sim G_a(1,1^2),\ \sigma>0,\ -\infty<\mu<\infty,\ -\infty<X<\infty$$

で定められるXの従う分布として定義される．この分布は歪度がゼロでない数少ない分布である． ここで、μ,σはXと同じ単位を持ち，$\frac{X-\mu}{\sigma}$は無名数になるので指数変換が可能である．また，$G_a(m,c^2)$は母平均m, 母変動係数c, 歪度$2c$のGamma分布である．

Gumbel分布の定義に於いて，Gamma分布のかわりに逆ガウス型分布を考え，

$$\exp\left(\frac{X-\mu}{\sigma}\right) \sim IG(1,1^2),\ \sigma>0,\ -\infty<\mu<\infty,\ -\infty<X<\infty$$

で定められるXの従う分布を考察した．ここで、$IG(m,c^2)$は母平均m, 母変動係数c, 歪度は$3c$の逆ガウス型分布である．これ以降，この分布を指数逆ガウス型分布と呼び、$X\sim EIG(\mu,\sigma^2)$と表記する．

$X\sim EIG(\mu,\sigma^2)$とすると，Xの確率素分$p_X(x)dx$と積率母関数$E\left[e^{\theta X}\right]$が以下の様に示される．

$$\begin{aligned} p_X(x)dx &= \frac{1}{\sqrt{2\pi}}\left[\exp\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(2\sinh\left(\frac{x-\mu}{2\sigma}\right)\right)^2\right\}\frac{dx}{\sigma}, \\ & \quad -\infty<x<\infty, -\infty<\mu<\infty, \sigma>0, \\ E\left[e^{\theta X}\right] &= \exp(1+\mu\theta)\sqrt{\frac{2}{\pi}}K_{\sigma\theta-\frac{1}{2}}(1), \qquad -\infty<\theta<\infty, \end{aligned}$$

ここで，$K_\nu(\cdot)$は変形された第2種ベッセル関数である．

また，以下の様なモーメントを持つ．

$$
\begin{aligned}
E[X] &= \mu + \sigma e^2 E_i(-2),\\
&= \mu - \sigma \cdot 0.36113,\\
E[X^r] &= \frac{e}{\sqrt{2\pi}} \sum_{k=0}^{r} \begin{pmatrix} r \\ r-k \end{pmatrix} \mu^{r-k}(-\sigma)^k\\
&\quad \times \int_0^\infty x^{-\frac{1}{2}} \log^k x \exp\left(-\frac{x}{2} - \frac{1}{2x}\right) dx,\\
\frac{E[(X-E[X])^3]}{(\mathrm{Var}[X])^{\frac{3}{2}}} &= 0.16174\cdots,
\end{aligned}
$$

ここで，$E_i(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt$, $x < 0$ で $\begin{pmatrix} a \\ b \end{pmatrix} = \frac{a!}{(a-b)!b!}$ である．以上のことより，指数逆ガウス型分布の歪度はゼロでないことがわかる．

最後に，$X \sim EIG(\mu, \sigma^2)$ の時，未知母数μ, σの最尤推定について提示し，本報告の枠組みで正規分布及び sinh-normal 分布の解説を加えた．

探索的多次元データ分析と応用研究

駒澤　勉

カテゴリカル・データの順序構造を探索的に多次元データ分析する手法として，数量化 III 類の解の 3 次元散布図で形状を探る方法について，以前から提唱してきた．本年度は記述的かつ探索的なこの手法の研究を明解な可視化表現法によって進展させた．

まず，人工データの代表的な Guttman, L. の完全尺度構成データ (Guttman(1950)) の三角データ行列とセミ対角データ行列，岩坪 (1987)，駒澤 (1982, 1992) の円環構造を持つ正方データ行列を円環 (ドーナツ型) の形状の平面に可視化する円環モデルを提示し，これら人工モデルのデータ行列が円環の表面上に部分集合で全て系統的に可視化表現できることを示した．

次に，これら人工モデルのデータを数量化 III 類を適用し，各要素の 3 次元散布図の形状が円環の表面に表現した各人工データの形状に類似していることを可視化し，数量化 III 類による順序構造を簡便に探る手法としての有用性を提案した．

終わりに，応用として (財) 日本労働文化協会・動脈硬化疫学研究所が集団健康検診を実施した約 20 万例の「QOL に関する健康生活診断」の不定愁訴，ストレス蓄積度，生活活性度 (運動)，食生活の四つの基本調査項目のデータに適用した実践的結果を報告した．

参考文献

Guttman, L. (1950). The principal components of scale analysis, *Measurement and Prediction* (ed. S.A. Stouffer), 312-361, Wiley, New York.

岩坪　秀一 (1987).『数量化法の基礎』, 朝倉書店, 東京.

駒澤　勉 (1982). 数量化理論 III 類とデータ処理, 数理科学, **225**, 24-31.

駒澤　勉 (1992). 順序尺度の探索的数量化の方法, 数理科学, **359**, 82-90.

多項分布モデルに基づくクラスタリング

馬 場 康 維

多項分布の漸近分布は多次元の正規分布である．したがって，何等かの方法によって得られた頻度表が，多項分布からの標本であると見なせる場合には，頻度の線形結合で表わされる“点”の信頼領域を正規近似によって求めることができる．

頻度の線形結合で表現できる複数の“点”の信頼領域が信頼係数 α の下で重なっているとき，それらの点は互に区別がつかない，即ち同一のクラスターに属するとみなせるものとする．α の大きさによって信頼領域の大きさが変る．したがって，信頼領域の重なり具合から，α をクラスターの距離とした階層的クラスターの構成ができることになる．ここでは，二つの例について述べた．

1) ランキングデータに基づく判定者とアイテムの同時クラスタリング
2) 数量化スコアの信頼性に基づくアイテム及び個体のクラスタリング

参 考 文 献

村田 磨理子, 馬場 康維 (1993). 順位グラフによる分類, 計算機統計学, **6**(1, 2), 25-36.

Baba, Y. (1996). Classification based on confidential region of quantification scores. *IFCS-96 Abstracts*, **1**, p.33.

3相データにおける尺度構成法

土 屋 隆 裕

質的な3相データを基にした4つの尺度構成法を報告した．具体的なデータとしては，$P=33$ 枚の樹木画に対する印象を $N=33$ 人の評定者が $I=20$ 個のSD項目を用いて7段階評定したものを用いた．すなわち，$P \times I \times N$ という3相データを基に樹木画に対する印象を測る尺度を構成することが目的である．最も簡便な第1の方法は，$P \times I$ 行列を縦に N 個並べた $NP \times I$ 行列に対して，項目分割のための数量化法 (土屋 (1995)) を行うことである．その結果，20個のSD項目が3つの次元に分類され,「明るさ」,「ていねいさ」,「大きさ」を測る尺度が構成された．しかしこの方法は，データが3相であるという構造を生かしていない．そこで第2の方法として，同一の絵に対する尺度得点はどの評定者も等しい，という制約を課して尺度構成を試みた．その結果，項目は方法1で得られた3つの次元には分類されなかった．このことから，評定者間で違いがないという制約は，尺度構成を行うためには厳し過ぎるものであり，絵から受ける印象は評定者ごとに異なるということが示唆された．評定者間の個人差を表現するために，第3の方法として各次元の尺度得点のランクを R に制約しながら尺度構成を行った．ランク制約を設けないときの項目の分類が得られるまで R を大きくする，という原則に基づいた結果，データ例では $R=2$ とするのが良いことが示された．平面上に絵画と評定者の位置をプロットすることで評定者間の違いが明確に表現され，特に「ていねいさ」を測る次元2では，評定者に大きく2つのグループがあることが見出された．そこで第4の方法では，項目の相を分割すると同時に，評定者の相も D 個のグループに分割することで，評定者間の差違を表現することを試みた．評定者が全て異なるとしたときの項目の分類が得られるまで D を大きくする，という方針に従った結果，$D=18$ という値が得られた．

参考文献

土屋 隆裕 (1995). 項目分割のための数量化法, 日本行動計量学会第23回大会発表論文抄録集, 240-241.

テキスト型データの解析法 —自由回答データの分析を中心に—

大隅 昇

質問紙調査における回答取得方法の一つに自由回答方式がある. 近年の急速な調査環境の悪化や新しい調査形態の登場で, 調査実施環境や調査方法自体の転換を余儀なくされているという事情を反映し, 郵送調査, 電話調査, 電子メール調査, CAPI/CATI, POSデータ・移動体通信利用調査など様々の試みがなされている. こうした多様化に伴い, 自由回答によるデータ取得や解析の機会が増えてきた. また質問紙調査とは異なる分野で, 自由記述文・テキスト型データ (textual data) を扱うデータ解析手法が求められている. とくに, 情報のコンピュータ化や文書電子化によるテキスト型データの蓄積が進み, これのデータ解析手法が話題となってきた. またデータベースの普及に伴い従来は諦めていた大量文字型データのコンピュータ処理が可能になってきたが, 現状は有効なデータ解析手法が不足し模索の段階にあった. 筆者の抱える例として, 野外調査記述データの分析 (エゾ鹿による農作物被害実態調査), POSデータ売上分析, 歯科矯正治療の患者意見調査, マーケティング消費者行動調査, ハイパーテキスト型データ処理解析 (サーチエンジンの研究) 等々きわめて多彩である. とくに, 自由回答文については, (1) 考えたことがないことには答えにくい, (2) 予想しなかった回答や知見がえられるかという疑問, (3) 無記入が多くなる傾向, (4) 他の設問の選択肢の影響を受ける, (5) 適切なデータ解析法がない, また集計の手間がかかる, (6) 回答に均一性を欠く (と言われている) 等の理由からあまり前向きに利用されてこなかった. そこで, 幾つかの実験調査データ (女性の自立意識調査, 消費者行動調査他) を事例として, 従来のアフターコーディング処理方式と幾つかのテキスト型データの解析方法・分析プログラム (GU-Talk, SPAD.T) とを比較し, それらの利用可能性やデータ処理・分析上の問題点を検討した. まず, アフターコーディング処理方式とGU-Talkによる自動的な回答文分解による抽出単語との対応分析による比較を行った. また, ローマ字化した自由回答文を「構成要素」(単語や文節) に分解して, これから導かれる出現頻度等の解析を行う方法の検討, 共同開発になるプログラムSPAD.Tによる解析等を行った. たとえば, 構成要素と回答 (たとえば被験者, 回答者) との関係を, ある回答に構成要素が出現する回数 (頻度) を成分とする行列 (一種のクロス表) とし, これに基づいた対応分析, 多重対応分析等を行う. これにより, 回答間, 構成要素間それぞれの間の類似性や, 回答と構成要素の間の関連性の理解に役立つ知見が得られることを検証した. また, ローマ字化規則, 分け書き規則, 同音異義語の扱い方, 品詞・助詞等の扱い方などの検討, 従来方式とのクロス表の分析によるデータ表の構造の影響評価, 従来型の設問や属性と構成要素とのクロス表に基づく自由回答型と多項分類型の設問間の比較, その問題点等を実証的に検討した. さらに分け書きの自動化という点で期待されるGU-Talkの機能評価や問題点を併せて検討した.

HALBAU for Windows の開発

高 木 廣 文

HALBAU (はるぼう) は，柳井・高木 (1986) による BACIC サブプログラムをもとに開発されたものである．HALBAU 開発の目的は，質問紙調査等実施時において，統計学がそれほど得意でなくても，誰でも簡単に統計的な解析を行える環境を提供することである．ユーザーからの要望などを加味して，これまで大幅なバージョンアップを行ってきたが (高木 他 (1989))，高木・柳井 他（1995））, 急速なパソコン環境の変化に対処するため，MS-Windows 対応版の作成を行っている．

これまでは開発言語に，NEC-N88BASIC を用いていたが，DOS/V 系のパソコンでの使用も可能にするために，Visual BASIC を用いることにした．HALBAU はその目的から，解析のためにプログラムやコマンドをキー入力するなどの操作を，必要としないように設計されている．この考えは Windows 版でも継承されているが，DOS 版とは基本的な設計が異なるので，現在は各分析に共通するフレームを作成している．具体的には，ファイルの指定画面，分析方法の指定画面，変数の指定画面，結果の出力画面，グラフィック出力画面などの開発を行っている．

各目的に応じた指定画面を作成する場合，とくに使用時の使い易さ，誤操作時のメッセージの表示などを考慮している．どの程度まで，マン-マシン・インターフェイスをよくするかは開発者に依存しているので，今後ある程度開発した段階でユーザーの意見を参考にする必要があるだろう．また，今後のパソコンの進化を踏まえて，開発・改訂のしやすいプログラムを当初から考えておくべきであろう．

参 考 文 献

高木 廣文, 柳井 晴夫 編著 (1995)『HALBAU による多変量解析の実践』，現代数学社, 京都.
高木 廣文, 佐伯 圭一郎, 中井 里史 (1989).『HALBAU によるデータ解析入門』，現代数学社, 京都.
柳井 晴夫, 高木 廣文 編著 (1986).『多変量解析ハンドブック』，現代数学社, 京都.

前震の識別に関する統計的予測の研究

尾 形 良 彦

ある所で地震活動が始まる．それは段違いに大きな地震の前震かもしれないし，ほぼ同規模の地震が続く群発型地震かもしれない．単なる本震・余震型の場合も多い．これらのいずれの型であるかは，勿論，その地震活動が終息してからでないと分からないが，現時点での活動中のデータに基づいてどの程度の統計的な識別予測ができるのであろうか．気象庁震源カタログに載っている M4 以上の地震に関して，ある客観的な規準で「群れ (cluster)」を同定すると千組以上の群れが得られるが，その大多数が数個の地震からなる．「本震」「前震」「余震」「群発型地震」などをある規準を設けて定義すると，群れ全体の中で複数の前震を持つ群れの数は 6% 強，群発型地震群はその数倍，残りは本震-余震型の群れで全体の 70 ~ 80% を占める．

この研究の目標は，ある地域で連発の地震が起きた時，その発生時刻・位置・マグニチュード列に関する情報を使うことによって，群れの型 (特に前震型) を出来るだけ有効に予測するような統計モデル (条件付き確率) を見いだすことである．各地震の発生時刻を時間軸の原点に据えて，後に続く地震の時点を重ね合わせることによって推定される Palm 型条件付き強度関数

を調べてみた．同様の強度関数を震央間の距離やマグニチュードの増減列の重ね合わせによって調べた．それぞれの型の群れのPalm型強度関数の割合 (寄与率) について，次のような有意な特性がみられた．(1) 前震型の群れの時間的寄与率は数日以内に集中し，群内の地震発生の時間差が半日位をピークとしてその前後は小さくなる，(2) 空間的には10km程度以内にまとまっており，群内の震央位置の相互距離が小さくなる (群の集中度が高くなる) にしたがって増大する，(3) 最初の数個でマグニチュード列が増加すれば寄与率は大きくなり，減少すれば小さくなる．

以上の統計的特徴を考慮したLogisticモデルを構成し，推定に使ったデータとは異なる新しいデータに対して確率予測を行うと，おおよそ一% ～十数%の変動が見られた．この確率予測の有効性はエントロピーの比較や分割表のAIC比較によって示すことができた．

さらに，最初の地震が起きた時点でも，それが前震である確率を予測することができる．ただしこのためには単発の地震もデータに含めなければならない．気象庁地震カタログ (M4以上) によると相対頻度は4%弱であるが，単発の地震が前震である地域性の確率 (震央位置の関数) がベイズ型平滑法によって推定され，この条件付確率は1.4 ～ 10%の変動がある．推定に使ったデータとは異なる新しいデータに対して予測の有効性が同様の意味で示された．

連発の地震に関しては以上の2つの条件つき確率を組み合わせれば更に有効な確率予測になっていることが確かめられた．

参考文献

Ogata, Y., Utsu, T. and Katsura, K. (1995). Statistical features of foreshocks in comparison with other earthquake clusters, *Geophysical Journal International*, **121**, 233-254.

Ogata, Y., Utsu, T. and Katsura, K. (1996). Statistical discrimination of foreshocks from other earthquake clusters (under review).

射影行列に関するいくつかの性質

(客員) 大学入試センター 研究開発部　柳　井　晴　夫

射影行列に関する性質とその多変量解析への適用に関して最近筆者が関心を持っている5つのテーマについて解説した．ここでは，そのうちの二つの話題をとりあげて解説する．

1) 制約付き一般化正準相関分析 (Yanai(1996))

$X_1, X_2, \ldots, X_m$ という m 種の多変量データ行列が与えられているとき，

$$C_j{}'a_j = 0, \quad (j = 1, \cdots, m)$$

という制約条件で

$$\sum_{i \neq j}^{m}(X_i a_i, X_j a_j) \Big/ \sum_{j=1}^{m} \| X_j a_j \|^2$$

を最大にする．その解は，Yanai・Takane(1992) の定理2.1を使うことにより

$$\sum_{j=1}^{m}(P(X_j) - P(X_j(X_j{}'X_j)^{-1}C_j))$$

の固有値，固有ベクトルを解く問題に帰着される。ただし，$P(X)$ は X の列ベクトルで生成される部分空間への直交射影行列である．このとき，$X_1, X_2, \ldots, X_m$ をアイテムカテゴリデータとすると，上記の結果から制約付き対応分析 (数量化 3 類) の解が導かれる。

2) 斜交射影行列の積に関する性質 (Takane・Yanai(1996))

P_1, P_2 が直交射影行列の場合，$(P_1P_2) = (P_2P_1)$ と $(P_1P_2)^2 = P_1P_2$ は等価となる．しかし，P_1, P_2 が直交射影行列でない場合，これらは等価にはならない．そこで，P_1, P_2 をそれぞれ W_j にそった $V_j(j = 1, 2)$ の上への射影行列としたとき (ただし，$V_j + W_j = E^n, j = 1, 2$)，

2a) $(P_1P_2)^2 = P_1P_2$ を満たす P_1P_2 はどの空間に沿ったどの空間への射影行列となるか，

2b) $P_1P_2 = P_2P_1$ が成立するための必要十分条件は何か，

といった命題を検証する。

参考文献

Takane, Y. and Yanai, H. (1996). On product of two oblique projectors (in preparation).

Yanai, H. and Takane, Y. (1992). Canonical correlation analysis with linear constraints, *Linear Algebra Appl.*, **176**, 75-89.

Yanai, H. (1996). Generalized canonical correlation analysis with linear constraints, *Proceedings of IFCS-96*, **2**, 177-180.