

## 平成13年度研究報告会要旨

と き：2002年3月19日 午前9時30分～午後5時15分  
20日 午前9時30分～午後5時15分  
ところ：統計数理研究所 講堂

### プログラム

3月19日(火)

あいさつ

所長 清水 良一

#### 【調査実験解析研究系】

国民性調査のための基礎的検討

種多様性を考慮した森林資源管理

二段抽出因子分析モデルについて

Poisson Voronoi Cells in 4 and 5 Dimensions

均等散布を持つ順列の列について

亜熱帯林における種及びサイズ構成の空間パターンの発展

発電プラント・データのクオホート分析

ノンパラメトリック平滑化処理による市区町村別 SMR

疾病地図アニメーションの作製

インターネット調査——第4次実験調査の概要——

学級を集落抽出した調査データの性質について

Web サイト上における統計科学の利用について

マーク依存型更新過程とベイズ的推論による大地震の予測確率と誤差推定

環境データの統計解析

坂元 慶行

吉本 敦

前田 忠彦

種村 正美

丸山 直昌

島谷健一郎

中村 隆

柳原 宏和

大隅 昇

土屋 隆裕

清水 信夫

尾形 良彦

柏木 宣久

#### 【予測制御研究系】

モデル駆動型アプローチによるデータ処理

多変量時系列に対する主成分・因子分析

哺乳類進化の分子系統学的解析

セミマルコフ切替モデルについて

マルチスケールブートストラップとその応用

多変量 AR モデルを使った経済分析

不完全情報下における制御系設計に関する研究

PLRM の応用

複雑な系のシミュレーションと多変量解析

ヌジ人名資料再考

磁気シールドのロバスト最適化

樋口 知之

川崎 能典

長谷川政美

北川源四郎

下平 英寿

佐藤 整尚

宮里 義彦

田邊 國士

伊庭 幸人

上田 澄江

土谷 隆

尤度関数のかたちについて  
 通信のための情報理論と変復調  
 複雑なダイナミックシステムの予測制御

石黒真木夫  
 瀧澤 由美  
 尾崎 統

3月20日(水)

【統計科学情報センター】

強制分類と主成分分析

馬場 康維

Order Statistics of the Triangular Distribution

金藤 浩司

外乱に対する弱非線形系の応答の一般的取扱い

岡崎 卓

信用リスクの計測モデル

山下 智志

Local Conditions Separating Periodic and Non-periodic

Structures and on the High-dimensional Random Packing

Problem of A. Renyi

(客員, Steklov Mathe Inst.) Dolbilin, Nikolai

Some Results for Waiting Time Problems of Pattern and Scar(客員, 華東師範大学) 韓 清

【統計基礎研究系】

待ち時間分布に関する研究

平野 勝臣

有限混合分布のコンポーネント数の検定とチューブ法

栗木 哲

拡散過程のドリフト項のノンパラメトリック最尤推定

西山 陽一

一般指数相反分布とその近似

松縄 規

多変量ラグランジュ分布の多変量逆正規分布への収束

南 美穂子

有限混合モデルの尤度関数の性質とその EM アルゴリズムへの応用

福水 健次

On the Tail of Semistable Laws

志村 隆彰

多重分光画像による土地被覆分類とその応用

(客員, 広島大学) 西井 龍映

大偏差理論の  $k$  標本選択問題への応用

(客員, 明治学院大学) 竹内 啓

【調査実験解析研究系】

ストレス関連語と項目反応

(客員, 大阪樟蔭女子大学短期大学部) 土井 聖陽

【統計計算開発センター】

計算精度の検証

田村 義保

密行列計算専用計算機システムの開発

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 の言語について

中野 純司

【領域統計研究系】

質の高い証拠を得るための要件

柳本 武美

結晶群の出現頻度と確率モデル

伊藤 栄明

文化を計る

村上 征勝

東アジアの日本語観センサス

吉野 諒三

日中国民性比較の基礎研究

鄭 躍軍

スプレッドシート上での統計計算環境の構築

(客員, 立教大学) 山口 和範

シミュレーション試験による母集団薬物動態試験法の評価\* (客員, 北里大学) 竹内 正弘

---

\* 都合により当日は報告されなかったが, 要旨のみ掲載する.

## 国民性調査のための基礎的検討

坂元 慶行

日本人の国民性の統計的研究には、日本人の意識動向の解明、調査法の研究、解析法の研究の3つの目的があり、それぞれに大きな課題が残されている。特に、第一の目的の意識動向の解明に関しては、1953(昭和28)年以来の質問には既に耐用年数を越え、回答の変化幅が縮小し、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、21世紀の意識の動きを捉え得る質問文を開発しなければならない。殊に、来年2003年には第11次全国調査を予定しており、そのための質問の作成は急務である。そこで、今年度は、主として、このための検討を行い、吟味調査(オムニバス調査)を実施した。年度研究発表会では、この調査から、2つの例について検討結果を紹介した。

一つは、「国や社会のためであっても個人の生活は犠牲にすべきでない」という意見が7割も占めている(NHKによる2000年の調査)という林(2001)の指摘を検討するため、「国民性調査」の「公共の利益か個人の権利か(#7.5b\*)」を24年振りに再調査してみた。その結果、個人派が伸びたことは認められたが、「公共」という用語のせい、多数意見が個人派に逆転するほどの劇的な変化は見られなかった。

もう一つは、「めんどろをみる課長(#5.6)に関するもので、「国民性調査」では「時には規則をまげて無理な仕事をさせることもあるが、仕事のこと以外でも人のめんどろをよくみる」課長が、めんどろもみないが無理な仕事をさせない課長を抑え、80~90%の圧倒的な支持を得てきた。しかし、この回答肢の主節と従節を逆転させて、「仕事のこと以外でも人のめんどろをよくみるが、時には規則をまげて無理な仕事をさせることもある」とすると、この課長の人気はかなり落ちることが林らによってすでに知られている。そこで、これら2種類の質問を同時に行い、その格差の動きを追えば、「めんどろはみてほしいが、無理な仕事はいやだ」といっようなドライな人の増加の程度が測定できるのではないかと考え、実査に付した。しかし、この試みは不首尾に終わった。ドライになったか否かより、むしろ、職場における人間関係の低温化とでも称すべき現象の方が顕著であったからである。

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、以上の2例の検討は、図らずも、この特徴を再認識させる結果となった。今後の質問文の開発に当たってはこの点への配慮を欠くことができないように思われる。

## 参 考 文 献

林 知己夫(2001) 柳は緑、花は紅、市場調査, No. 248, 4-19.

## 種多様性を考慮した森林資源管理

吉本 敦

森林資源管理に関連した主な生態学的な問題の1つに、時空間的な林分の伐採パターンによって形成されるパッチ状態から生じる生物多様性がある。生物多様性を形成するいくつかの原因の一つにエッジを共有する林分の林齢格差が指摘されている。このことは、伐採・植林に

より林分の空間的な配置を操作すれば、エッジを囲む林分の状態を人為的に変化させることができ、その結果、森林全体の生物多様性の度合いを調整することが可能になるというを示している。

このような時空間的な問題を解決するには、空間的制約がある森林伐採最適化モデルが必要不可欠となる。本研究では、空間的制約のある森林伐採最適化モデルを使用し、エッジにおける林齢格差の状態を操作することにより森林の状態がどのように変化するかを分析した。

ここで使用するモデルは、空間的制約として隣接林分制約だけを持つものである。最小伐採排除期間を、ある林分が伐採された時にその林分に隣接する林分が伐採されるまでの最小期間と定義すると、最小伐採排除期間を操作することにより共有するエッジにおける林分の林齢格差を調整することができ、その結果、生物多様性の度合いすなわち森林構造の非均一性を制御することができる。

分析の結果、最小伐採排除期間の増加により、年平均伐採量と全期間から得られる総収益の現在価値が減少する結果となった。これは、最小伐採排除期間が長くなればなるほど、伐採の機会が減少するためである。特に、最小伐採排除期間が 3 期から 5 期になったとき、年平均伐採量と現在価値が 17% から 20% まで急激に減少した。また、伐採された全林分面積と伐採林分総数については、最小伐採排除期間が 1, 2 期となった時、徐々に増加した。これは、1 期, 2 期の排除期間でより多くの林分が現在価値を増加させるためにより早い時期に伐採されるからである。林分がより早い時期に伐採されれば、そこから得られる伐採量は自ずと少なくなり、その結果、総現在価値と年平均伐採量は減少するが、伐採される林分数とその総面積は増加するためである。

森林構造への影響については、より長い最小伐採排除期間を採用すればするほど、森林の林齢分布とエッジにおける林齢格差分布に非均一性がかなり生成されることが分かった。時間に対する林齢分布の平均と分散の変化については、最小伐採排除期間を 3 期にすれば、安定な分布が持続されることが分かった。

## 二段抽出因子分析モデルについて

前 田 忠 彦

通常の社会調査では、単純無作為抽出よりも複雑な標本抽出方式——例えば層化二段無作為抽出——を採用することが多い。しかし、そうした枠組の下で得られたデータが標本抽出方式に則した方法で分析されることは、必ずしも多くない。特に多変量解析の文脈では単純無作為抽出を仮定したモデルによる分析で済まされるのが現状と言える。

二段抽出因子分析モデルは、 $p$  個の変量に関するデータが二段抽出の下で得られる場合に、この標本抽出方式に則した因子分析を行うためのモデルである。二段抽出における第 1 次抽出単位を PSU と略記すると、一般に二段抽出では総共分散行列  $\Sigma_T$  が、PSU 間共分散行列  $\Sigma_B$  と PSU 内共分散行列  $\Sigma_W$  の二つの成分に分解される。二段抽出因子分析モデルは、やや省略して言えばこの  $\Sigma_B$  と  $\Sigma_W$  が共に因子分析モデルにしたがうとするモデルと考えてよい。他方単純無作為抽出を仮定した分析とは、 $\Sigma_T$  に直接因子分析モデルを当てはめることである。ここでの問題は、二段抽出モデルで分析すれば得られたであろう本来の興味深い構造を単純無作為抽出モデルでは見落としてしまうケースである。

本研究ではデータに基づく推定の問題としてではなく母集団レベルでの共分散構造のモデル

化という文脈において、次の形で二段抽出因子分析モデルの性質と必要性を考えることにした。すなわち、 $\Sigma_T$  に対する因子分析モデル(単純無作為抽出モデル)では二段抽出モデルでの構造をうまく捉えられないケースとして(1)  $\Sigma_W$  と  $\Sigma_B$  がそれぞれ(同一の)因子分析モデルに従うのに、 $\Sigma_T$  は因子分析モデルの仮定を満たさない(2)  $\Sigma_T$  が因子負荷行列の母数配置が  $\Sigma_W$  と同じモデルにしたがうが、 $\Sigma_W$  の因子負荷行列から一定基準を超えて大きく隔たっている、の 2 通りのケースを想定した。

1 因子 3 変数 ( $p = 3$ ) という極めて限定的な条件ではあるが、数値実験によって、そうしたケースの発生について次のような結果が得られた。

- ・ 各変量の相関比が小さくなると、ケース(1)(2)ともに発生率は低くなる。
- ・ ケース(1)は 3 変数が PSU 内共分散構造において一様に高い因子負荷を持つ傾向が強くなるにつれて、発生率が低まる。

第 1 点は社会調査データでは相関比がそれほど大きくないために、 $\Sigma_T$  に対する分析だけでも大きく判断を誤らないことも多かることを予想させる。第 2 点は、実用場面では、PSU 内構造で明確な因子が同定できる場合に、 $\Sigma_T$  の分析でも同一の因子を見出し得る場合が多くなることを示唆している。

二段抽出因子分析モデルでは、標本に基づく推定でも不適解などの問題に遭遇するケースが多いので、どういう条件下で推定がうまくいかなくなるのか、その場合に上記で考えたような母集団レベルでのモデル構成の困難に原因が見出せるのか、といった観点からの診断法および対処法を類型化することが有効と言える。こうした点が今後の検討課題である。

#### 参 考 文 献

- 前田忠彦(2001) 二段抽出因子分析モデルにおけるデザイン効果の検討, 研究集会「因子分析と共分散構造分析に関する諸問題」講演報告集, 45-63.

### Poisson Voronoi Cells in 4 and 5 Dimensions

種 村 正 美

ポロノイ(Voronoi)領域は空間に散布された粒子の配置パターンの特徴づけや粒子配置のモデルとして非常に有用である。

本研究では、ポアソン配置のポロノイ分割によって生じるポロノイ領域(Poisson Voronoi Cell と呼ぶ)の幾何学的特徴量の統計分布を 4 次元および 5 次元空間の場合に求めた。

この問題は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。2 次元および 3 次元の場合については Tanemura(2001), 種村(2001)に述べた。しかし、われわれの知る限り、4 次元以上ではまだ研究が行われていない。今回のわれわれの研究がおそらく最初である。今回、われわれは標本数を 4 次元および 5 次元に対してともに  $n = 5,000,000$  として計算機実験を行った。

ポロノイ領域の計算のために Tanemura et al.(1983)のアルゴリズムをそれぞれ 4 次元, 5 次

元の場合に拡張したプログラムを開発した。

計算機実験から得られた特徴量のヒストグラムに一般化ガンマ分布

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

を当てはめた。これはパラメータの調節によって広範囲の分布を表現できることが知られている(Hinde and Miles(1980))。われわれは最尤推定法によって4次元ポロノイ多面体体積に対して  $\hat{a} = 1.27657$ ,  $\hat{b} = 5.01598$ ,  $\hat{c} = 6.53950$ , 5次元ポロノイ多面体体積に対して  $\hat{a} = 1.41743$ ,  $\hat{b} = 5.88045$ ,  $\hat{c} = 8.53601$  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極めてよく再現した。これらは、1次元ポロノイ線分長の理論分布 ( $a = 1, b = 2, c = 2$ ), 2次元, 3次元のわれわれの結果(Tanemura(2001))と合わせると、Kiangの予想(Kiang(1966))〔すなわち、 $a = 1, b = 2d, c = 2d$ ;  $d$  は空間の次元〕を否定するものであることが明確になった。その他の幾何学量についても同様の当てはめを行って、新しい知見を得た。

### 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.
- Kiang, T. (1966). Random fragmentation in two and three dimensions, *Zeitschrift für Astrophysik*, **64**, 433–439.
- Tanemura, M. (2001). Statistical distributions of Poisson Voronoi cells in two and three dimensions, Research Memo., No. 796, 1–42, The Institute of Statistical Mathematics, Tokyo.
- 種村正美(2001). Poisson Voronoi Cell の統計分布, *統計数理*, **49**, p. 369.
- Tanemura, M., Ogawa, T. and Ogita, N. (1983). A new algorithm for three-dimensional Voronoi tessellation, *J. Comput. Phys.*, **51**, 191–207.

## 均等散布を持つ順列の列について

丸 山 直 昌

1 から  $n$  までの  $n$  個の数字を任意に並べ替えたものを長さ  $n$  の順列と呼ぶが、長さ  $n$  の順列の  $m$  個の列で、ある種の均等散布性をもつもの考えることが、インターネットを利用した公開抽籤に役に立つ。このような順列の列を  $m$  行  $n$  列の行列  $M$  で表すことにする。ここで考える抽籤とは、 $n$  人の応募者に対して1位から  $n$  位の順序を与えるような抽籤である。応募者が一堂に集まることができない条件下で、抽籤のプロセスを公平かつ公正に行うために、 $M$  をあらかじめ公表しておき、日経株価平均のように多くの人々が容易に知り得る数値で、ある程度のランダム性をもって決まる数値に基づいて行列  $M$  の一つの行を選び、当籤順位を決定する。この抽籤が公平感を持つためには、行列  $M$  の各列に1から  $n$  の数字が「均等」に散布していることが重要である。散布の均等性を計る尺度を導入し、適度の均等散布性を持つ行列  $M$  を群論的な手法を用いて構成する方法を説明した。

## 亜熱帯林における種及びサイズ構成の空間パターンの発展

島 谷 健一郎

樹木群集の構造や動態を調べるため、プロットを設定しその中の(一定サイズを越える)全個体を識別し種を同定し、さらにサイズ(直径または樹高)を測定する毎木調査は各地で進められている。ところで、この作業では個体の位置も同時に測定する事が多いが、それらは野外調査を円滑にするための補足(例えば調査者が再測定時に個体をスムーズに見つけるための目安)で終わることも多く、それが解析データとして用いられることはあまり見られない。その要因のひとつは、空間構造をうまく表現してくれる統計量が少ないからであろうが、空間構造は野外観察によって認識することが難しいため、それを解析しようという意識があまり呼び起こされなかったこともあるだろう。例えば、個体のサイズ構成なら野外において見た目でそのパターンをイメージできるが、特に種数が 100 種にもおよぶ群集では、野外観察で各種の空間パターンを思い浮かべることがほとんど不可能である。

そこで、多種からなる沖縄島北部ヤンバル地域の亜熱帯林で、二次林から極相林まで 4 段階の林分のデータ(久保田康裕(鹿児島大・教育)との共同研究)について、点過程論に基づく空間統計を用いた解析を行ない、見た目ではイメージしにくい多種系の空間パターンを、群落構成種の空間分布を集約した統計量で表現できないものか考察している。群集生態学で広く用いられる種多様度指数(種数, Simpson 指数(相対頻度として個体数を用いたものと胸高断面積のもの 2 種))を、個体間の距離に関する関数に拡張し、空間分布を表現する指標として適用してみた。さらに、樹高に関する林分内での順位で各個体の階層を定め、上層 10%, 20%, 等々の個体だけでこれらの値を計算し、垂直方向の解析に利用した。すると、初期には多様な種が一斉に発育を開始したのに、しだいに競争効果によって限られた種による林冠の独占が進むが、次にはそれら上層木の枯死によりその下で待機していた種が林冠に到達する様子が、グラフで定量的に記述できた。このような空間データの解析は、森林の再生動態のパターンを定量的に表現する上で有効であると期待される。

## 発電プラント・データのコウホート分析

中 村 隆

発電プラントの運営に関する各種の指標(運転保守費や設備利用率など)の時間的変動を分析する際に、暦年・経年・運開年の要因を区別することが必要となる。ここで、暦年要因とは経済循環や障害対策の同時実施といった暦時間によって各プラントに同時に影響を及ぼすもの、経年要因とは機器の故障率の変化など各プラントの運開後の経過時間によって変化するもの、運開年要因とは各プラントが採用した技術水準など運開した年代によって異なるものを指す。

発電プラント指標データの分析モデルは、暦年 = 調査時点(時代)、経年 = 年齢、運開年 = コウホート(同時出生集団)という対応をとれば、コウホートの視点を取り入れ継続的な社会調査データから年齢・時代・コウホート効果を分離するコウホート分析モデルと本質的に変わらず、3 要因を分離できないという識別問題を抱えている。そこで、パラメータの漸進的変化の条件と ABIC を用いるベイズ型コウホートモデルの方法を適用する。

具体的なモデルは以下のとおりである。発電プラントが  $N$  個あるものとし、プラント  $n$



( $n = 1, \dots, N$ ) の運開年を  $k_n (\in \{1, \dots, K\})$  とすれば, 暦年  $j (= 1, \dots, J)$  には経年は  $i_{jn} = j - k_n$  ( $\in \{1, \dots, I\}$ ) となる. 暦年  $j$  における何らかの指標値を  $y_{jn}$ , 発電容量などの共変量を  $x_{jn}$  として,

$$y_{jn} = \beta_0 + \beta^X x_{jn} + \beta_{i_{jn}}^A + \beta_j^P + \beta_{k_n}^C + \varepsilon_{jn}, \quad (j = 1, \dots, J; n = 1, \dots, N)$$

のように分解する. ここで,  $\beta_0, \beta_i^A, \beta_j^P, \beta_k^C$  はそれぞれ総平均, 経年, 暦年, 運開年効果のパラメータであり, ゼロ和制約

$$\sum_{i=1}^I \beta_i^A = \sum_{j=1}^J \beta_j^P = \sum_{k=1}^K \beta_k^C = 0,$$

を満たすようにする. また,  $\{\varepsilon_{jn}\}$  には独立な正規分布を仮定する. モデルの詳細と適用例については小野・中村(2002)を参照.

### 参 考 文 献

小野賢治, 中村 隆(2002) 原子力発電コスト分析における「ベイズ型コウホートモデル」技法の適用, 電力中央研究所報告, P01001.

## ノンパラメトリック平滑化処理による市区町村別 SMR 疾病地図アニメーションの作製

柳 原 宏 和

本研究は, 任意年次時点での各種死因別危険度地図製作およびその動画表示のためのコンピュータシステムの開発を目的とする. この地図を各年次時点ごとに連続に表示することにより, 各種死亡危険度の時空間分布の視覚化が可能となる. 解析に用いたデータは, 厚生省から目的外使用を許可された 1974 年から 1994 年までの 20 年間における全国 3324 市区町村別死因別死亡数, および 1975 年から 1994 年までの期間に行われた 5 回の国勢調査による市区町村別年齢階級別人口数である. 今回は Bladder(膀胱), Brain(脳腫瘍), Breast(乳), Colon(結腸), Kidney(腎臓), Leukemia(白血病), Liver(肝臓), Lung(肺), Pancreas(すい臓), Skin(皮膚), Stomach(胃), Throat(咽頭), Tongue(舌), Uterus(子宮)のそれぞれの 14 種類の癌に関する死亡危険度を取り扱った. 肝臓癌, 肺癌など観察死亡数が比較的多い癌に関しては,  $B$ -spline による標準化死亡比(SMR, Standardized Mortality Rate)の平滑化値に基づく危険度により地図を作成した. 膀胱癌, 脳腫瘍など観察死亡数が少ない癌に関しては, 各市区町村, 年次の観察死亡数がそれぞれ独立であるという仮定のもと Poisson-Gamma Model をあてはめ, 経験ベイズ法を用いて相対的な危険度を推定し,  $B$ -spline によるその相対危険度の平滑化値により地図を製作した. 結果は CG によるアニメーションで危険度が高い順に「赤」「黄」「緑」「水色」および「青」の 5 色のカラーで表現され, 正方形のマークの面積が市区町村の人口数にほぼ比例するように表示される(結果であるアニメーション地図の一例としては大瀧 他(2000), 佐藤・大瀧 2001 等を参照). 本システムを適用することで, 従来は困難であった任意年次時点での死因別危険度の地図を描くことが可能となり, アニメーション化することで経

年変化や地理分布の把握がより容易になった。

### 参 考 文 献

- 大瀧 慈, 川崎裕美, 佐藤健一, 柳原宏和, 山口直人(2000). ノンパラメトリック平滑化処理による市区町村別 SMR 疾病地図アニメーションの作製, 第 68 回日本統計学会講演報告集, 261-262.  
佐藤健一, 大瀧 慈(2001). 統計グラフィブラリ開発と Web への応用, 統計数理, 49, 261-274.

## インターネット調査——第 4 次実験調査の概要——

大 隅 昇

インターネット調査, 特に Web 調査の調査分野への普及・浸透は著しく, 数年の間に調査システムそのものが急速に様変わりしている。迅速, 安価, 簡便, 集計速報性等が Web 調査の長所とされたがこれが当てはまらず, 質の高い調査データの取得には, かなりの投資や信頼できる調査システム構築が必須であることが明らかになってきた。Web 調査は従来型調査とは異なるので, 実験調査に際して様々な工夫が必要である。実験調査の基本方針として, ①なるべく調査現場を模擬できるように実績のある調査機関の協力を仰ぐこと, ②実査の時期を揃えること, ③同一設問, 類似設問による調査票設計とすること, ④同一設問を用いて, 従来型調査と Web 調査との比較実験を行うこと, ⑤調査設問として, 時系列的比較が可能で, しかも従来調査で実績のある設問とインターネット関連設問とを併用・比較すること等を掲げてきた。この方針の下に, 複数の調査機関と協力し, 1997 年に第 1 次調査を開始してから昨年の第 3 次調査まで継続的に実験調査を行ってきた。第 1 次調査では 1 サイトのみの登録パネルを用いて Web 調査を連続して複数回行った。第 2 次調査は 3 社の協力を得て, Web 調査と従来型調査(オムニバス, オンライン)の比較を, また第 3 次調査では 2 社の協力で Web 調査とオムニバス調査を複数回行った(大隅(2001))。

こうした条件で行われる継続比較実験調査は国内では初めての試みであり, これにより従来から類推で議論されてきた諸事象が具体的に見えてきた。Web 調査は調査方式が多数の場合はセルフ・セレクション型であることから, 調査対象とする集団がインターネット利用者を完全に代表しているか保証されないことがある(そもそも, インターネット利用者数の推計も困難である)。また, 登録者集団(リソース)の構成が, すでにインターネット利用者を代表していないことも見えてきた(性差, 年齢区分差, 地域差等, 種々の偏りを含むことが分かってきた)。さらに回答者が, 状況によってはリソースを代表していないことも分かった。また従来調査との比較で, Web 調査特有の種々の事象が次第に明らかになってきた。とくに, リソースの構築方法, 維持管理等, 登録者集団の捕捉手段が, 回答率や回収データの質を左右することが具体的に見えてきた(リソースの鮮度や登録者への対応方法, つまり管理維持が回答率に影響する等)。

こうした調査結果を踏まえて, 第 4 次実験調査を計画し進めることとした。3 調査機関(電通リサーチ, 博報堂・東京サーベイリサーチ・グループ, 日本リサーチセンター)の協力を得て, Web 調査ならびにオムニバス調査, 郵送調査の比較実験を計画し進めてきた。原則として従来の調査の基本方針を踏襲したうえで, ①第 3 次調査以降のインターネット技術の改善要素を考慮する, ②回答者行動を追跡するトラッキング調査を工夫する, ③「調査不能」あるいは無回

答の発生要因を探查する等に重点をおいて進めた。トラッキング調査はインターネット調査の利点の一つで、回答者別の回答所要時間や回答時間帯、完遂率や脱落率、接触率、あるいはそれらの発生理由の探查等が可能である。また、調査不能に関しては、従来型調査に類似する事象とインターネット調査特有のそれとをできるだけ明らかにすること、加えて、欧米の調査研究で指摘されてきた、調査票のデザインが回答行動へ及ぼす影響効果の測定等も試みた。とくに、調査票デザインについては、欧米(米国、スロベニア等)の研究者との共同研究として、設問選択肢設計にラジオボタン、コンボボックス、プルダウンメニューを用いることが回答結果(回答率・無回答率、誤回答等)に及ぼす影響評価等を検証した。実験調査はかなりの規模となり、2001年度末から2002年度前期にかけて、Web調査を延べ8回、オムニバス調査を4回、郵送調査を2回行う。また、調査で得られた成果は、参加機関の合意の下に共有情報としてすべて一般に開示される。

### 参 考 文 献

- 大隅 昇(2001) 調査環境の変化と新しい調査法の抱える問題, 統計数理, 49, 201-213.  
大隅 昇(2002) インターネット調査の適用可能性と限界——実験調査から見えること——, 輿論科学協会創立56周年記念講演, 市場調査, 250, 4-23.  
横原 東(2001) マーケティングにおけるインターネット調査の実状と課題, 統計数理, 49, 215-222.  
吉村 幸(2001) インターネット調査にみられる回答者像, その特性, 統計数理, 49, 223-229.

## 学級を集落抽出した調査データの性質について

土 屋 隆 裕

子どもを調査対象とした自記式調査においては、学校あるいは学級を集落抽出することがある。実際の調査データに基づいて、どのような抽出方法が適当なのかを検討した。

まず、各対象校から学級を集落抽出する場合に、対象校の抽出確率を、全校児童・生徒数に比例させる方法と、推測した学級数に比例させる方法のそれぞれについて、単純集計値と比推定値を比較した。その結果、小学校では学級数の推測がうまくいくために後者の方法、中学校では学級数の推測がうまくいかないために前者の方法がよいことが分かった。

次に、学校内での学級間の変動と、学校間の変動とを比較した。一般に、学級間の変動に比べ、学校間の変動の方が大きいため、サンプルサイズが同じであるならば、学級を最終抽出単位とする方法に比べ、学校を最終抽出単位とする方法の方が誤差が大きくなる。比較の結果、子どもの年齢が下がるほど、学級間の変動が大きくなる、という傾向が見られた。したがって、低年齢の子どもほど、学校を最終抽出単位とすることによる誤差の拡大が少ないのに対し、中学校では学級を最終抽出単位とする方がよい、という結論が得られた。

## Web サイト上における統計科学の利用について

清水 信 夫

コンピュータ・ネットワークが急速に普及し、多種多様な情報が氾濫している現在、電子的に蓄積された膨大な量の統計データの有効利用や高度な統計分析手法の普及や開示のあり方についての検討は急務の課題と考えられる。特に、インターネットの普及により、様々な分野において蓄積した統計関連情報を WWW (World Wide Web) 環境下で公開する動きが広がっている。統計科学の分野においても、WWW の利用により、研究・教育の発展に供する目的で多くの Web サイトが立ち上げられており、内容も次第に多様化する傾向にある。しかしながら、これらの Web サイトの実態や特徴を総括的にレビューしたり具体的に整理した報告は例が少ない。

Murdoch (2000) においては統計に関連した Web サイトをいくつかの区分により大まかに分類しているものの、急激に多様化する統計科学関連 Web サイトの実態をより詳しく知るには一次元的な分類ではなく複数の方向軸に沿った分類が必要である。

本研究においてはこれらの統計科学関連 Web サイトのうち国内外における主要なサイトに関していくつかの区分を設け、多次元尺度構成法の利用により多次元的な分類を試みた。また、各区分における Web サイトの主な特徴についても考察した(清水・大隅(2001))。

## 参 考 文 献

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

清水信夫, 大隅 昇(2001). Web サイト上における統計科学環境, 第 29 回日本行動計量学会発表論文抄録集, 40-43.

マーク依存型更新過程とベイズ的推論による大地震の  
予測確率と誤差推定

尾 形 良 彦

古地震や歴史地震の発生時刻データはイベント数が少ないため、ポアソン過程を除き最も簡単な点過程モデルとして更新過程が良く使われる。最尤法に際しては、なるべくパラメタ数を節約することが推奨される。例えば発生間隔に対数正規分布を考える場合、分散を示すパラメタを典型値に固定したものを標準モデルとする。そして、それより複雑なモデルの可否を AIC によって判定し、最も優れたモデルを使って将来の発生確率を予測するのである(地震調査委員会 長期評価部会(1998)など)。

しかし特に地震数が少ない場合、また地震間隔がほぼ揃っている場合、AIC を適用すると直感に合わない結果が得られることがある。本報告では、限られたデータ数のもと、最尤法の限界を超えるために尤度関数に立ち返り、ベイズ推論による確率予測の方式について議論した。解析例として南海地震などの確率予測やその誤差の分布を議論した。また地震に伴うスリップ量データがあれば、発生時期予測可能モデル(Time-Predictable model, Shimazaki and Nakata (1980))を検証できる。指数・対数正規・ワイブル・Brownian Passage Time の各分布の拡張型

更新過程を考え、ベイズ推論によって適合性や予測確率などの誤差について議論した。

## 参 考 文 献

- 地震調査委員会長期評価部会(1998) 『(試案)長期的な地震発生確率の評価手法およびその適用例について』,地震調査推進本部。
- Ogata, Y.(2001). Slip-size dependent renewal processes and Bayesian inferences for uncertainties, Research Memo., No. 820, The Institute of Statistical Mathematics, Tokyo.
- Shimazaki, K. and Nakata, T.(1980). Time-predictable recurrence model for large earthquakes, *Geophysical Research Letters*, 7(4), 279-282.

## モデル駆動型アプローチによるデータ処理

樋口 知之

モデル駆動型アプローチという言葉は日本語としては認知されていないが、英語では model-driven approach として定着しつつあり、意味する内容も、データからの情報抽出に明示的にモデルを想起・利用する方法論として理解されている。この概念に基づいた私が行っている共同研究のうち、今年度大幅に進展がみられた二つの成果の一部を紹介した。

一つは、「統計モデルによる大規模電流系の自動同定：DMSP-F シリーズデータの全解析」である。このプロジェクト研究は5年前からスタートしたもので、地球を頻りに周回(100分で1周)する極周り低高度人工衛星によって恒常的に観測・蓄積された超大量の磁場データを取り扱う。特に、太陽風と地球磁場のダイナミックな相互作用の結果、地球磁気圏に形成される大規模な電流系が引き起こす磁場変動に注目する。大規模電流系は極域地方にリング状に存在し、またその構造は経度方向に広がったシート状である。従って、1周で得られるデータを緯度情報を元に単純に4分割したものを(ファイルと呼ぶ)一つ一つに大規模電流系に関連した磁場変動が一つ含まれる。このファイル中の3次元磁場データをうまく座標変換を行うと、ある一成分のみに区分的に直線と近似できる空間系列となる変動磁場が表出してくる。これらの知見を数値的に表すために、主成分分析的なアイデアと、可変節点リニアスプライン関数のあてはめを組み合わせた手法を開発し、DMSP-F7という人工衛星によって得られた5年間の磁場データに適用することでその有効性は検証済みである。今回それを別の人工衛星(DMSP-F12,13,14,15の4つ)データセットに適用した。今回のプロジェクトにより解析されたデータ総量はF7から通算すると40Gb超、判別された磁場変動ファイル数は30万超にもなる。判別された結果は、その他諸々の大規模電流系に関連する異種情報と結合し、太陽と地球磁気圏との相互作用を総合的に理解することが可能なサマリーファイルを作成した。このサマリーファイルは、大規模電流系に関する諸情報を、適切に圧縮かつ体系的に整理した、当該分野にとって初めての成果物である。

もう一つは、「Geomagnetic Jerk の自動検出のための統計モデルの開発」である。Geomagnetic Jerk とは、地球固有磁場の長期的な変動(トレンド成分)の2回微分値に特徴的なものが見られる現象である。この現象の発生メカニズムはまだ未知の部分が多いため、さまざまなアプローチからの研究が現在なされている。我々はトレンド成分に2次のスプライン関数を採用した季節調整モデルを構成し、Jerk 検出と同定を自動的に行う手法を開発した。このモデルを利用し

た解析の結果, Geomagnetic Jerk を特徴付けるパラメータの分布と下部マントル内の地震波速度分布に興味深い類似点があることが分かってきた。

## 多変量時系列に対する主成分・因子分析

川崎能典

主成分分析や因子分析において標準的に想定されるデータセットの代表的な形式は, 異なる被験者に関して観察される値が測定項目ごとに並べられているものと言えるだろう。ここで個体を表す添え字を時間の添え字に読み替えることで, 多変量時系列に対して形式的に主成分分析・因子分析を行うことは可能であるが, このような形式的な適用については古くから問題点が指摘されてきた。本報告の目的は, 多変量時系列解析における主成分分析・因子分析に理論的に妥当性を与える 2 つの切り口(分析を時間領域で行うか周波数領域で行うか)から, これまでの主要な結果を整理することにある。第 1 の方法は, 時系列の離散 Fourier 変換によって漸近的に独立なデータに変換し, 古典的な主成分・因子分析の枠組みに帰着させるものである。時間をインデックスとしたとき近接する観測値にある相関も, 周波数をインデックスとすると漸近的に独立性が確保され, 独立同一分布ケースでの主成分・因子分析の複素変数版を考慮することで多変量時系列に対する分析(スペクトル密度行列の固有値分解)が可能になる。第 2 は観測されない因子過程に直接モデルで表現を与える方法である。接近法としては, 従来の因子分析にラグ構造を入れて共分散構造分析の観点から解析するもの, 構造時系列モデルの枠内で共通因子を与えるもの, 共通因子を多変量 ARMA 過程で記述するものなどがある。これらはすべて明示的に因子に対してモデルを与えるものである。明示的に共通因子をモデル化しない場合は, 作用素のランク条件に議論が集約される。そのような手法としては縮小ランク回帰モデルや誤差修正モデルがある。最後に, 古典的な主成分分析の方法論を時系列データにそのまま適用するやり方であるが, これは時間方向の相関を無視して同一時点での確率変数の相関のみに着目することに等しく, その有効性はノンパラメトリックなトレンド推定やインデックス作成に代表されるデータ圧縮等に限られる。このような立場で一度主成分分析を行った結果に新たにモデルを与える手法には注意が必要である。

## 参 考 文 献

川崎能典(2001). 多変量時系列に対する主成分・因子分析, 統計数理, 49, 109-131.

## セミマルコフ切換モデルについて

北川源四郎

IIP データのような経済データの分析においては, 景気の転換にともなって時折トレンドの傾きに急激な変化が現れる。このような時系列  $y_n$  を想定し,  $y_n = t_n + w_n$  のようにトレンド成分  $t_n$  と観測ノイズ  $w_n$  への分解を考える。ここで, トレンド成分の時間変化は  $t_n = t_{n-1} + \Delta t_n$  と表されるが,  $\Delta t_n$  に関してランダムウォークモデル  $\Delta t_n = \Delta t_{n-1} + v_n$  を仮定すると通常の

トレンドモデルが得られる。

一方、トレンドの傾斜に関する通常の 2 状態マルコフ切替モデルでは、トレンドの上昇期と下降期ではそれぞれ常に同じ傾斜をもつことになり現実的でない。そこで、以下のようなモデルを用いて、 $\Delta t_n$  が時間とともに徐々に変化できるようにするとともに、マルコフチェイン  $S_n$  に依存した切替によって急激な変化にも対応できるようにした。

$$(1) \quad p(\Delta t_n | \Delta t_{n-1}, S_n = i, S_{n-1} = j) \\ = \begin{cases} N(\Delta t_n, \tau^2) & i = 1, j = 1 \text{ または } i = 0, j = 0 \\ U(-c, 0) & i = 0, j = 1 \\ U(0, c) & i = 1, j = 0 \end{cases}$$

さらに、セミマルコフ過程を導入することによって、各状態における滞在時間、すなわち前回の切替の後、始めて切替が発生するまでの時間の分布を直接モデル化することができるようにした。滞在時間分布としては、例えば負の二項分布

$$(2) \quad P(\ell, p) = \binom{k - \ell - 1}{k} p^k (1 - p)^\ell$$

を用いればよい。以上のモデルは一般状態空間モデルで表現でき、したがって、トレンドや傾斜の推定、マルコフチェインの状態推定やパラメータ推定をモンテカルロフィルタで実現することができる。

1973 年 1 月から 2000 年 12 月までの IIP データに対して、提案したモデルをあてはめ、ガウス型、非ガウス型トレンドモデル、2 状態マルコフ切替モデルなどとの比較を行った結果、提案したモデルが AIC 最小となった。得られたモデルを用いると IIP データから景気の転換点を自動的に推定することができる。また、このモデルは長期予測においても望ましい性質を持つことが確認された。

## 参 考 文 献

- Kitagawa, G. and Hakamata, M. (2001) Semi-Markov switching slope model, Research Memo, No. 810, The Institute of Statistical Mathematics, Tokyo.

## マルチスケールブートストラップとその応用

下 平 英 寿

近似的に不偏な検定を簡単なリサンプリングだけから計算するための理論とアルゴリズムを提案した。新たに考案したマルチスケールブートストラップ法は、母数空間の一般の形状の領域を帰無仮説とする検定において、棄却確率を漸近的に 3 次の精度で有意水準に等しくする。これに対して単純なブートストラップ法は 1 次の精度しかない。ダブルブートストラップ法は 3 次の精度であるが、計算量がリサンプル数の 2 乗のオーダーになり多くの応用で現実的でない。マルチスケールブートストラップ法は 3 次の精度でありながら、計算量は単純なブートストラップ法と同様にリサンプル数に比例するオーダーである。マルチスケールブートストラップ法では、リサンプリングにおけるサンプルサイズをいくつか変えて問題をスケール変換し、

単純なブートストラップ法を複数回行う。帰無仮説に対応した領域の境界の曲率とデータからの符号付距離を、マルチスケールブートストラップを使って計算するための漸近展開に基づく方法は Shimodaira (2000, 2002) で述べられている。このためのソフトウェア (Shimodaira and Hasegawa (2001)) は WWW で一般に公開されている。

提案した方法を分子系統樹推定の信頼性評価に応用した。特に哺乳類の進化における論争を呼んだ問題に満足な回答を与えることができた。候補となる各々の系統樹は確率モデルによって表現されており、尤度を最大にする系統樹が選択される。この信頼性を評価するための先行研究として、Shimodaira and Hasegawa (1999) では対数尤度の多重比較を行っている。これは下平 (1993)、Shimodaira (1998) で提案されていた方法を分子系統樹推定に応用したものであり、SH test として多くの系統樹ソフトウェアに組み込まれるなどして定着しつつある。Shimodaira and Hasegawa (1999) の被引用数も急速に増えている (2000=7, 2001=29, 2002=25, 2002 年 4 月現在)。インパクトファクター (IF) に換算すると最初の 2 年間の引用数を足して 36 となり、Nature 誌の平均的な論文 (IF=25) よりも注目度が高かったことを示している。しかしながら、これらの応用を通して多重比較法の保守的な性質が大規模問題で障害になることを実感し、これを解決するためにマルチスケールブートストラップ法が考案された。

### 参 考 文 献

- 下平英寿 (1993) モデルの信頼集合と地図によるモデル探索, 統計数理, 41, 131-147.
- Shimodaira, H. (1998) An application of multiple comparison techniques to model selection, *Ann. Inst. Statist. Math.*, 50, 1-13.
- 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.
- Shimodaira, H. (2001) Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection, *Comm. Statist. Theory Methods*, 30, 1751-1772.
- Shimodaira, H. (2002) Approximately unbiased test of phylogenetic tree selection, *Systematic Biology*, 51, 492-508.
- Shimodaira, H. and Hasegawa M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference, *Molecular Biology and Evolution*, 16, 1114-1116.
- Shimodaira H. and Hasegawa M. (2001) CONSEL: For assessing the confidence of phylogenetic tree selection, *Bioinformatics*, 17, 1246-1247.

## 多変量 AR モデルを使った経済分析

佐藤 整 尚

マクロ計量経済分析においては、構造モデルを用いた方法と時系列モデルを用いた方法があるが、これら 2 つのアプローチは互いに反するのではなく補完的である。つまり、構造モデルであっても適当な変換と近似を行うことによって、時系列モデルに焼き直すことが可能である。大きな違いは、変数間の関係が事前に規定されているかどうかである。現在の経済構造は、大きな変革期を迎えており、多分に流動的であると考えざるをえない。従って、現状においては事前の仮定をなるべく置かず、ここで提案するような多変量時系列モデルにより変数間の関係を捉え、それに基づいて構造モデルの構築にあたるのが望ましいと考えられる。その意味で、



ここで提案する時系列分析は、あくまでも現状を正しく捉えるための道具であると言える。

複数の経済変数が与えられたとき、その変数間の関係を調べるのによく用いられるのが（同時）相関を見る方法である。しかしながら、これだけでは、時間差を伴う関係については捉えきれない。そこで、時間差を伴う相関をとった相互相関を考えるやり方がある。これをもとにモデル化したのが多変量自己回帰モデル（多変量 AR モデル）である。多変量 AR モデルでは、変数の選択さえ行えば、AIC 基準を用いることにより、ほぼ自動的にモデルの当てはめを行うことが可能である。また、このモデルでは、ある変数の今期の値は、自分自身および他の変数の過去の値の線形結合によって表される。このため、過去の値から将来の値を予測することが容易である。また、それぞれにかかる係数（パラメータ）は、モデルにとっては各変数間の関係を表す重要なもので、与えられたデータセットから推定される。変数間の関係が安定的であれば、この係数はほぼ一定の値をとることが期待されるが、もし不安定であれば、時間とともに変化する可能性がある。そこで、ここでは、每期毎期、逐次パラメータを推定していきやり方を採用することにする。これによって、経済構造の変化を抽出することが可能になる。

通常、経済システムは安定的なシステムではない。したがって、そこから得られる予測値は、構造変化等が起きると信頼できないものになってしまう。今回は、その性質を逆に応用して、予測値を、足元の状態を計る尺度として活用しようとする試みを行っている点が特徴である。

## 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

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 Hanba and Miyasato (2001) の研究を行っている。

従来の適応制御理論が漸近安定性に主眼を置くのに対して、制御性能をより定量的に考慮する観点から、適応制御過程を  $H_2/H_\infty$  最適制御問題として定式化することを試みている。これまでにモデル規範形適応制御を含む一般的な形式の適応制御問題について、安定解析に用いるリアプノフ関数の一部と Hamilton-Jacobi-Isaacs 方程式の解を同一視することで、特定の評価関数に対して最適（または準最適）な 3 つの型の適応制御系の構成法を導出した（宮里 (2001)）。またシステムに含まれる未知のパラメータを  $H_\infty$  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  $H_\infty$  制御系を構成する手法を求めて、参考文献 Miyasato (2001a) にあげるような結果を得た。またそれをロボットマニピュレータの軌道追従制御に応用した結果 Miyasato (2002b) も得ることができた。さらにその手法を非線形パラメータモデルの非線形・適応制御に拡張することも考えている。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  $H_\infty$  制御系を設計する手法を求めて、参考文献 Miyasato (2001b) にあげるような成果を得た。

これらとは別に LPV (Linear Parameter Varying) システムのゲインスケジューリング制御において、スケジューリングパラメータが未知な場合に、その推定値を用いて適応的にゲインスケジューリング制御系を構成する手法について考察した。その結果、LMI (Linear Matrix Inequality) に基づく LPV システムの適応型ゲインスケジューリング  $H_\infty$  制御系を構成する手法を導出し、参考文献 Miyasato (2002a) にあげるような成果を得た。さらにそれらと非線形  $H_\infty$

制御手法を組み合わせたスケジューリングパラメータの時間変動成分への対処法や, モデル規範制御も含めたより広いクラスの制御手法に対する LPV システムの適応型ゲインスケジューリング制御についても, いくつかの萌芽的な結果を得ることができた.

### 参 考 文 献

- Hanba, S. and Miyasato, Y. (2001). Output feedback stabilization of bilinear systems using dead-beat observers, *Automatica*, **37**(6), 915–920.
- Miyasato, Y. (2001a). Adaptive nonlinear  $H_\infty$  control for processes with bounded variations of parameters — General forms and general relative degree case —, *Proceedings of IFAC Workshop on Adaptation and Learning in Control and Signal Processing (ALCOSP2001)*, 419–424.
- 宮里義彦 (2001). 適応  $H_\infty$  制御系の構成法, 計測と制御, **40**(10), 692–697.
- Miyasato, Y. (2001b). Design method of adaptive nonlinear  $H_\infty$  control systems via neural network approximators, *Proceedings of the 40th IEEE Conference on Decision and Control*, 3158–3163.
- Miyasato, Y. (2002a). Adaptive gain-scheduled  $H_\infty$  control of linear parameter-varying systems, *The 15th IFAC World Congress* (preprints).
- Miyasato, Y. (2002b). Nonlinear adaptive  $H_\infty$  control for robotic manipulators, *The 15th IFAC World Congress* (preprints).

## 複雑な系のシミュレーションと多変量解析

伊 庭 幸 人

物理系などを対象とした大規模シミュレーションが盛んであるが, 系の振舞いが複雑になるにつれて, 結果の中から系を特徴付ける情報をどのように取り出したらよいか, という問題はますます自明でなくなってくる. そこで, 「シミュレーションで得られたデータの多変量解析」という分野が成立するのではないかと考えられる. 実際, そうした方向の研究がいろいろな分野で目につくようになってきており, 伊庭が編集した統計数理の特集号(統計数理 第 49 卷 第 1 号(2001 年), 特集「地図を描く・風景を眺める — 主成分分析・多次元尺度法とその周辺 —」)でもそのいくつかを取り上げた. 講演では, こうした流れについて説明し, 具体的な例として, 伊庭と福島孝治氏(東大物性研)との共同研究である「スピングラス磁性体のシミュレーションの多変量解析」について報告した.

## ヌジ人名資料再考

上 田 澄 江

古代メソポタミア, イラク北東部の都市ヌジより出土した粘土板に記載された人名から家系図を再構成した.

ヌジ文書とは遺跡から出土した粘土板に記された主に土地に関する契約書(牧野(1991))で, ヌジ人名資料(Gelb et al.(1943))はヌジ文書の名による索引である. そこから人名にかかわ

る親族関係, 内容を記した文献名, その巻数, 行番号などが参照できる。民族の流出入の繰り返しによる表記文字のバリエーションおよび近距離名の検討により, 所属不明の名前を減少することができた。家系図の再構成をした結果, 2人世帯が減少した。世帯数と世帯の大きさの分布, 男女の出現数, 契約当事者間の世代の同一視, 加えて遺跡の面積による人口の換算, などを考慮に入れて古代社会の人口の推定を試みている。

## 参 考 文 献

- Gelb, Ignace F., Purves, Pierre M. and Macrae, Allan A. (1943) *Nuzi Personal Names*, The University of Chicago Press, Chicago, Illinois.  
 牧野久美 (1991) 偽装養取に見られる社会変動——古代メソポタミア・ヌジ遺跡出土文書による——, 史学, 60(1), 91-120.

## 尤度関数のかたちについて

石 黒 真木夫

最尤法は(古典的には現在でも「変分法」などがそうであるような)数式処理の方法として統計的推定量の計算式「最尤方程式」を導き出す方法, あるいは統計量の意味づけを与える方法であったが, 現在では数値計算の方法と考えられていると言ってよい。その数値計算が実際の場面でとんでもない挙動を示すことがあり得る。

たとえば, 平均  $\mu$ , 分散  $s^2$  の正規分布  $f(x|\mu, s^2)$  を 2 つ重ねた「ソンプレロ分布」を次のような形で定義する。

$$\text{sombbrero}(x|\mu, s^2) = 0.2 \times f(x|\mu, s^2) + 0.8 \times f(x|\mu, 100.0)$$

この分布で  $s = 1$ ,  $\mu = 0$  としてシミュレーションデータを 30 ケ発生させた(図 1)。このデータに対するソンプレロモデルの対数尤度には  $s$  が 0 に近づくに従って無限大に近づいていくつも

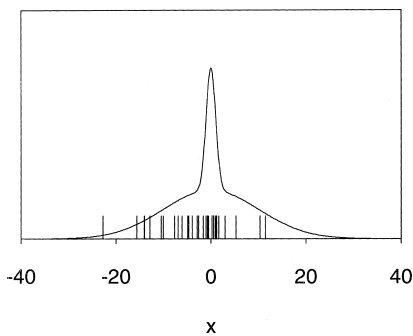


図 1. ソンプレロ分布とシミュレーションデータ。

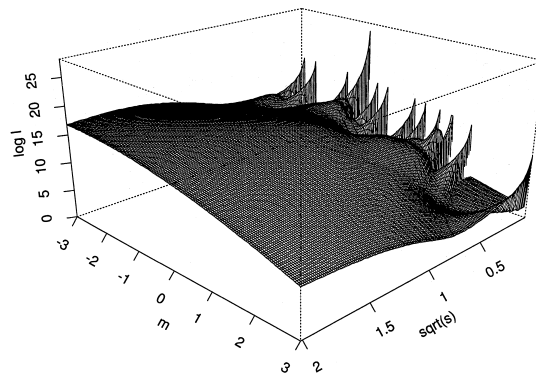


図 2. ソンプレロモデルの対数尤度。

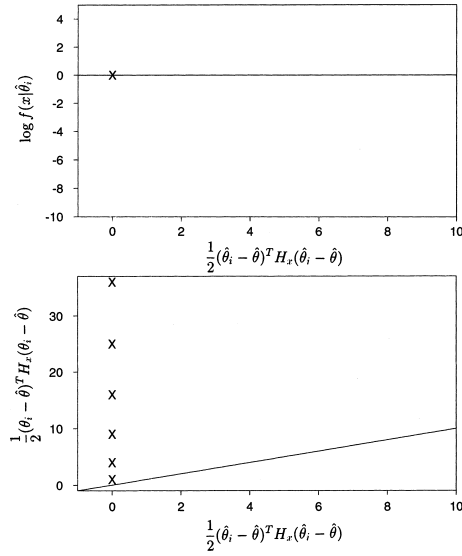


図 3. 「獅子の子プロット」ガウス分布モデルの場合.

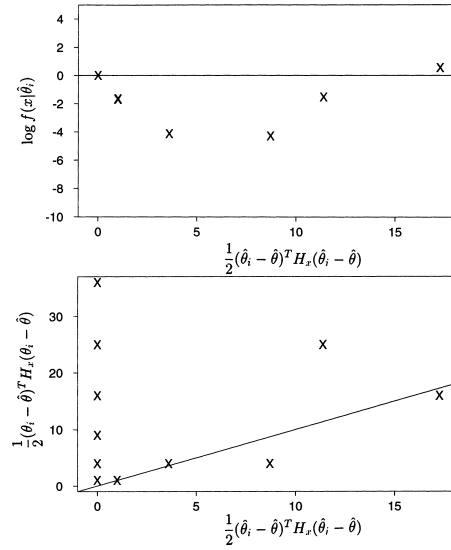


図 4. 「獅子の子」プロット, ソンブレロモデルの場合.

の「山」がある(図 2).

対数尤度関数がこのような形をしている場合, 単純に数値的最尤法を利用してモデルのパラメータを推定するのは危険である. 数値的最適化の結果得られた値を診断する方法を考えた.

獅子の子の谷落とし法.  $\hat{\theta}$  を数値的最尤法で得られたパラメータの推定値とする. 初期値

$$\theta_i = \hat{\theta} + L^{-T} u_i$$

から出発した「最適化計算」が停止した位置を  $\hat{\theta}_i$  とする. ただし  $L$  は  $\hat{\theta}$  における対数尤度関数の Hessian を  $H_x$  として,

$$LL^T = H_x$$

で定義される下 3 角行列である. 対数尤度関数が  $\hat{\theta}$  で最大値をとる素直な形をしていれば  $i = 1, 2, \dots$ , に対して

$$\left( \frac{1}{2}(\hat{\theta}_i - \hat{\theta})^T H_x(\hat{\theta}_i - \hat{\theta}), \log f(x|\hat{\theta}_i) \right) \text{ と } \left( \frac{1}{2}(\hat{\theta}_i - \hat{\theta})^T H_x(\hat{\theta}_i - \hat{\theta}), \frac{1}{2}(\theta_i - \hat{\theta})^T H_x(\theta_i - \hat{\theta}) \right)$$

をプロットすると, 図 3 のような形になるはずである. 図 1 のデータにあてはめた  $\hat{\theta}$  に関してのプロットは図 4 になった. 最適化の結果を解釈するにあたって注意を要する事が明らかである.

## 複雑なダイナミックシステムの予測制御

尾崎 統

今年度は非線形ダイナミックシステムの研究を産業への応用を重点目標におき研究した。火力発電プラントのボイラーの温度の非線形制御法を開発した他、発電プラントの排ガス中の窒素酸化物の濃度をアンモニアによって一定値以下に押さえながら効率的に脱硝制御する非線形制御法などを開発した。我々の新しい手法(NPC法)を取り入れた製品の開発を進める日本ベレーに技術協力すると同時に関連の計算技術などを特許申請した。この他以下の2つの応用研究をおこなった：1)非線形予測制御法の応用として金融資産のポートフォリオ手法の開発をみずほ信託銀行と共同でおこなった。2)生理学研究所とキューバ神経科学センターの研究者と協力して複雑な脳機能ダイナミックスの解明に向けてfMRIデータ、EEGデータなど脳神経科学における大量の計測データの解析手法を研究した。

## 強制分類と主成分分析

馬場 康維

双対尺度法で提案されている強制分類は、判別分析の一種である(Nishisato(1984, 1986))。強制分類の考え方は様々な応用が可能である。分割表や射影に関連した強制分類についてはNishisato & Babaが報告している(Nishisato and Baba(1999))

強制分類の原理は非常に簡単である。 $F$ を $p$ アイテムに対する $n$ 個体のレスポンスデータとする。 $F$ を $p$ 個の列ベクトルに分解して考える。この中の一つのベクトルに定数 $k$ をかけたデータ行列を $G$ とする。 $k$ を十分大きな定数とし $F$ の代わりに $G$ を用いて、双対尺度法を適用すると、重みをかけたアイテムが主となるスケーリングが得られる。したがって、 $k$ を大きくすることにより、アイテムは、このアイテムに関連のあるものとなしものものに分類される。この手続きは変数のグループに定数をかけても適用できる。

強制分類の方法は、主成分分析にも拡張できる(馬場(1997))。 $n$ 個体、 $p$ 変数のデータ行列を $X$ とする。一つの変数ベクトルに定数 $k$ をかけて得られる行列を $G$ とする。 $k=1$ として $G$ から主成分分析を行えば、通常の主成分分析である。 $k$ を小さくすると感度分析に対応し、 $k$ を大きくすれば、このベクトルに主要なものとなる主成分分析が行える。

## 参考文献

- 馬場康維(1997) 強制分類の主成分分析への応用, 統計数理研究所共同研究レポート, No. 100, 41-49.
- Nishisato, S. (1984) Forced classification: A simple application of a quantification technique, *Psychometrika*, **49**, 25-36.
- Nishisato, S. (1986) Generalized forced classification for quantifying categorical data, *Data Analysis and Information* (eds. Diday, E. et al.), 351-362, North-Holland, Amsterdam.
- Nishisato, S. and Baba, Y. (1999) On contingency, projection and forced classification of dual scaling, *Behaviormetrika*, **26**(2), 207-219.

## Order Statistics of the Triangular Distribution

金 藤 浩 司

$X$  を以下の確率密度関数  $f(x)$  に従う確率変数とする .

$$f(x) = \frac{\sqrt{6} - |x|}{6}, \quad -\sqrt{6} < x < \sqrt{6}.$$

その時, 上記に従う順序統計量  $X_{i:n}$  の期待値  $E[X_{i:n}] = \mu_{i:n}$  を以下のように解析的に表現した .

$$\mu_{i:n} = \frac{\sqrt{6}n!}{(i-1)!(n-i)!} \left\{ \left(\frac{1}{2}\right)^{i-1} (A1 - A2) + \left(\frac{1}{2}\right)^{n-i} (B1 - B2) \right\},$$

ここで,  $A1, A2, B1, B2$  は, 以下である .

$$\begin{aligned} A1 &= \frac{2^{i+\frac{1}{2}}}{1+2i} {}_2F_1\left(i + \frac{1}{2}, n + \frac{3}{2}; i + \frac{3}{2}; -1\right), \\ A2 &= \frac{2^{i-1}}{i} {}_2F_1(i, n+1; i+1; -1), \\ B1 &= \frac{2^{-i}}{1-i+n} {}_2F_1(1, 1-i; 2-i+n; -1), \\ B2 &= \frac{2^{1-i}}{3-2i+2n} {}_2F_1\left(1, 1-i; \frac{5}{2} - i + n; -1\right), \end{aligned}$$

ここで,  ${}_2F_1(a, b; c; z)$  はガウスの超幾何関数である .

## 参 考 文 献

Iwase, K. and Kanefuji, K. (2002). Estimation of the fifth percentile using a subset of order statistics, *Measurement and Multivariate Analysis* (eds. S. Nishisato, Y. Baba, H. Bozdogan and K. Kanefuji), 217-226, Springer, Tokyo.

## 外乱に対する弱非線形系の応答の一般的取り扱い

岡 崎 卓

外乱を受けて発展する系を対象として, 系の非線形性が強くないとの前提のもとに, 外乱に対する系の応答確率密度を外乱の統計量で可及一般的に表現する試みについて報告する. その要点は, 任意の外乱を許容する一般化 Fokker-Planck (GFP) 方程式とその拡散係数の具体的表現, および解の解析的表現にある.

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

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

に従って発展する系変数  $U$  の確率密度を定めるために開発された GFP 方程式は、非線形系にあつては複雑な構造をもち実用に供し難い。そこで、系の非線形性が弱く、正規軌道  $U(t)$  ( $\frac{d}{dt}U = M(U)$ ,  $U(0) = U$  の Jacobian  $J = \frac{\partial M(U)}{\partial U}$ ) が  $J(U + \varepsilon V, t) = J(U, t) + O(\varepsilon^2)$  ( $\varepsilon \ll 1$ ) となる場合を取り上げ、GFP 方程式の簡略化を図る。

## 2. 拡散係数の表現

非 Gauss 外乱に対する 1 変数非線形系の応答を GFP 方程式によって求めるには、先ず同方程式中の拡散係数を外乱の統計と系の非線形構造に応じて定めねばならない。外乱の特性汎関数の cumulant 表示

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

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

と系の弱非線形性  $M(U) = -\alpha U + \gamma M_1(U)$ ,  $|\gamma| \ll 1$  により、non-local および local な拡散係数を外乱の cumulant と系の構造関数  $M_1(U)$  で次の形に表わすことができる。

$$\lambda_A(U) = e^{\gamma \lambda_{A1}(U)} \lambda_{A0}, \quad \lambda_B(U, V) = e^{\gamma \lambda_{B1}(U)} \lambda_{B0}(V)$$

## 3. 非 Gauss 外乱に対する 1 変数弱非線形系の応答

1 変数系に対して系応答の特性関数  $\hat{f}(k)$  を定める上記 GFP 方程式を書き下せば

$$\left( \alpha i \frac{d}{dk} - \gamma M_1 \left( i \frac{d}{dk} \right) \right) \hat{f}(k) + ik \left[ e^{\gamma \lambda_{A1}(i \frac{d}{dk})} \lambda_{A0} \hat{f}(k) + e^{\gamma \lambda_{B1}(i \frac{d}{dk})} \lambda_{B0}(k) \hat{f}(k) \right] = 0$$

となり、非線形度を表わす  $\gamma$  による摂動解から、特性関数  $\hat{f}$  に対する次の表現が導かれる。

$$\hat{f}(k) = \hat{f}_{Gauss}(k) \cdot \hat{f}_{mod}(k)$$

即ち非 Gauss 外乱を受ける 1 変数非線形系の応答密度  $f(U)$  は、Gauss 外乱に対する線形部分の応答に外乱の非ガウス性と系の非線形性から定まる変調を施したものと得られることが判る。

## 4. Duffing spring の応答

固いバネに対応する Duffing spring 系

$$M(U) = -\alpha U + \gamma M_1(U), \quad M_1(U) = -U^3$$

を例に取って計算を進めれば、応答特性関数

$$\hat{f}(k) = \exp \left[ A_U ik + \frac{1}{2} \sigma_U^2 (ik)^2 + \frac{1}{6} S_U (ik)^3 + \frac{1}{24} F_U (ik)^4 + \dots \right]$$

の特性量(平均変位  $A_U$ , 分散  $\sigma_U$ , skewness  $S_U$ , flatness  $F_U$ )を、外乱の特性量( skewness および flatness )と系非線形度  $\gamma$  で具体的解析的に表現することができ、任意の外乱に対する系応答の一般的表現が完了する。

## 信用リスクの計測モデル

山下 智志

企業倒産などの理由で融資した資金が回収不可能となる現象をデフォルトといい、そのリスクを計算するモデルを信用リスクモデルと呼ぶ。近年信用リスクモデルの開発を進めてきたが、今年度は CDR 運営協議会や全国信用保証協会の協力を得ることにより、これまで日本に存在しなかった 100 万件レベルの信用情報データを入手し、その分析を行った。これはデータ量が従来の研究の 10~100 倍であるという点に特徴があり、それを生かしたより複雑な分析を試みることができた。例えば業種や規模によるセグメントの効果に関する考察や、潜在変数を含んだ信用リスクモデルの構築である。潜在変数モデルというのは、これまでのモデルのように数十の財務変数を取捨選択しモデルの説明変数とするのではなく、「安定性」「成長性」「収益性」などの直接観測不可能な合成変数(潜在変数)を一旦作成し、それによって信用リスクを説明させるモデルである。このモデルの利点は、実務的に行われていた財務データの縮約を、統計的裏付けをもって行うという点にある。今回の研究では主成分 logit や LISREL-logit, 潜在変数の測定モデルを導入することによって、信用リスクの測定精度の向上を達成した。

## 待ち時間分布に関する研究

平野 勝臣

### 本年度の研究

1. パターンのあるクラスに対して、パターンがはじめて起こるまでに指定された長さ以上のサブパターンの起こる回数は幾何分布に従う(Hirano and Aki(2001))。
2. ある依存性を持つ  $\{1, 2, \dots, \mu\}$ -値系列において、長さ  $m$  のパターンが起こるまでの待ち時間分布はその逆パターンの待ち時間分布に等しいことを示した(Aki and Hirano(2002))。

### パターンが起こるまでにサブパターンの起こる回数の分布と幾何分布

連やパターンに関する統計量の分布は、確率変数の系列が  $\{0, 1\}$ -値で独立、同一の分布に従っていても普通は複雑である。しかしながら、最近、連についてのいくつかの統計量は幾何分布のような簡単な分布に従うことが報告されている(Hirano et al.(1997))。

$X_1, X_2, \dots$  を  $\{1, 2, \dots, \mu\}$ -値 i.i.d. 確率変数列とし、 $j = 1, 2, \dots, \mu$  に対し  $P(X_i = j) = p_j$ ,  $p_j > 0$  とする。また  $(a_1, a_2, \dots, a_k)$  を  $\{1, 2, \dots, \mu\}$ -パターン、整数  $\ell$  ( $1 \leq \ell < k$ ) に対しパターンの左からのサブパターン  $(a_1, a_2, \dots, a_\ell)$  を考える。

もし  $X_{i-\ell+1} = a_1, X_{i-\ell+2} = a_2, \dots, X_i = a_\ell$  で  $X_{i-\ell-h+1} = a_1, \dots, X_i = a_{\ell+h}$  であるような  $(k-\ell)$  以下の正整数  $h$  が存在しないならば、長さ  $\ell$  以上のサブパターンが  $i$  番目の試行で起こったという。また、パターン  $(a_1, a_2, \dots, a_k)$  に対し、次(i)か(ii)のいずれかであれば、サブパターン  $(a_1, \dots, a_\ell)$ , ( $\ell < k$ ) は条件 A を満たすという:(i)  $a_1 = a_{i+1}, a_2 = a_{i+2}, \dots, a_\ell = a_{i+\ell}$  であるような整数  $i$  が存在しない(ii)もしそのような  $i$  が存在すれば、すべての  $j = 1, 2, \dots, (k-\ell-i)$  に対し  $a_{\ell+i+j} = a_{\ell+j}$  を満たす。



定理 . サブパターン  $(a_1, a_2, \dots, a_\ell)$  は条件 A を満たしているとする . そのとき , パターン  $(a_1, a_2, \dots, a_k)$  がはじめて起こるまでに , 長さ  $\ell$  以上のサブパターンの起こる回数は  $G_1(p_{a_{\ell+1}} \cdots p_{a_k})$  に従う . ただし , サブパターンの数え方は隣あったサブパターンの長さ  $(\ell - 1)$  以下の重複を許す数え方で , 確率生成母関数  $\phi(t) = pt/(1 - (1 - p)t)$  をもつ幾何分布を  $G_1(p)$  とかく .

この結果は  $m$ -次マルコフ系列 ( $\ell \geq m$ ) に拡張できる .  
 なお , 本報告は Hirano and Aki (2001) に基づいている .

### 参 考 文 献

- Aki, S. and Hirano, K. (2002). On waiting time for reversed partterns in random sequences, *Ann. Inst. Statist. Math.*, **54**, 713–718.
- Hirano, K. and Aki, S. (2001). Number of occurrences of subpattern until the first appearance of a pattern and geometric distribution, Research Memo., No. 805, The Institute of Statistical Mathematics, Tokyo.
- Hirano, K., Aki, S. and Uchida, M. (1997). Distributions of numbers of success-runs until the first consecutive  $k$  successes in higher order Markov dependent trials, *Advances in Combinatorial Methods and Applications to Probability and Statistics* (ed. N. Balakrishnan), 401–410, Birkhäuser, Boston.

## 有限混合分布のコンポーネント数の検定とチューブ法

栗 木 哲

### 2 項有限混合モデル

$$\alpha \text{Bin}(k, p) + (1 - \alpha) \text{Bin}(k, 1/2)$$

の観測値として i.i.d. 系列  $x_1, \dots, x_n$  が得られているとする . ここで  $0 \leq \alpha \leq 1$  は未知の混合確率 ,  $p$  は未知の 2 項比率 ,  $k$  は既知の定数である .  $\alpha = 0$  または  $p = 1/2$  のとき , またそのときに限り , モデルは 1 コンポーネントモデル  $\text{Bin}(k, 1/2)$  となる . 多因子性疾患の遺伝マッピングにおいてこのモデリングがしばしば用いられる (Ott (1999), 10 章) .

コンポーネント数に関する仮説の尤度比検定を考える .

$$H_0 : 1 \text{ コンポーネント} \quad \text{vs.} \quad H_1 : 2 \text{ コンポーネント}$$

このモデルは局所錐型モデル (Dacunha-Castelle and Gassiat (1997)) の一例であり , 尤度比検定統計量の漸近帰無分布は一般には  $\chi^2$  分布にはならない .

2 項有限混合モデルは離散分布であるので , 高い次元の多項分布に埋め込むことができる . 母数空間が有限次元接錐で近似できる場合の尤度比検定統計量の漸近分布を論じた Chernoff (1954) の結果より , 尤度比検定  $H_0$  vs.  $H_1$  の漸近帰無分布は , 平均 0 , 分散 1 の正規確率過程

$Z(\phi) = \sum_{i=1}^k g_i(\phi) Y_i$ ,  $\phi \in [-1, 0]$ , ただし

$$g_i(\phi) = \frac{1}{\sqrt{(1+\phi^2)^k - 1}} \binom{k}{i}^{\frac{1}{2}} \phi^i, \quad Y_i \sim N(0, 1) \text{ i.i.d.},$$

を用いて

$$T_k = \max \left\{ 0, \max_{-1 \leq \phi \leq 0} Z(\phi) \right\}^2$$

と表すことができる.

チューブ法 (例えば Takemura and Kuriki (2002)) の適用により次の公式を得る.  $x \geq 0$  に対して

$$\begin{aligned} P(T_k > x) &= P \left( \max_{-1 \leq \phi \leq 0} Z(\phi) > \sqrt{x} \right) \\ &= \frac{\ell_k}{2\pi} P(\chi^2(2) > x) + \frac{1}{2} P(\chi^2(1) > x), \end{aligned}$$

ただし

$$\ell_k = \int_{-1}^0 \frac{\sqrt{k \{ (1+\phi^2)^k - 1 - k\phi^2 \}} (1+\phi^2)^{k-2}}{(1+\phi^2)^k - 1} d\phi.$$

## 参 考 文 献

- Chernoff, H. (1954). On the distribution of the likelihood ratio, *Ann. Math. Statist.*, **25**, 573–578.
- Dacunha-Castelle, D. and Gassiat, É. (1997). Testing in locally conic models, and application to mixture models, *ESAIM Probab. Statist.*, **1**, 285–317.
- Ott, J. (1999). *Analysis of Human Genetic Linkage*, The John Hopkins University Press, Baltimore, Maryland.
- Takemura, A. and Kuriki, S. (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains, *Ann. Appl. Probab.*, **12**, 768–796.

## 拡散過程のドリフト項のノンパラメトリック最尤推定

西 山 陽 一

確率微分方程式

$$dX_t = \theta(X_t)dt + n^{-1/2}dW_t, \quad X_0 = x_0 \in \mathbb{R}$$

における未知関数  $\theta(\cdot)$  を推定する問題を考える. ただし母数空間  $\Theta$  は  $C(\mathbb{R})$  の部分集合であるとする. 常微分方程式

$$dx_t = \theta(x_t)dt \quad x_0 \in \mathbb{R}$$

の解を  $x^\theta = (x_t^\theta)_{t \in [0,1]}$  とし, 擬距離

$$d_{\theta_0}(\theta, \vartheta) = \sqrt{\int_0^1 |\theta(x_t^{\theta_0}) - \vartheta(x_t^{\theta_0})|^2 dt}$$

を導入する. 尤度比が

$$\log \frac{dP_\theta^n}{dP_{\theta_0}^n} = n\{\Gamma_{\theta_0}^n(\theta) - \Gamma_{\theta_0}^n(\theta_0)\}$$

と書けることに注目しよう. ただし

$$\Gamma_{\theta_0}^n(\theta) = \int_0^1 \theta(X_t^{n,\theta_0}) dX_t^{n,\theta_0} - \frac{1}{2} \int_0^1 |\theta(X_t^{n,\theta_0})|^2 dt$$

である.  $\Theta$  を “近似” する sieve を  $\Theta_n$  とし, sieved MLE を  $\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta_n} \Gamma_{\theta_0}^n(\theta)$  と定義する. 目標は, 上で定義した擬距離に関する  $\hat{\theta}_n$  の  $\theta_0$  への収束率を導出することである.

我々は van der Vaart and Wellner (1996) の Theorem 3.2.5 によるアプローチを用いる. ポイントは,

$$\gamma_{\theta_0}^n(\theta) = \int_0^1 \theta(X_t^{n,\theta_0}) \theta_0(X_t^{n,\theta_0}) dt - \frac{1}{2} \int_0^1 |\theta(X_t^{n,\theta_0})|^2 dt$$

とおくと  $\Gamma_{\theta_0}^n(\theta) - \gamma_{\theta_0}^n(\theta)$  が連続マルチンゲールの終点値になり, Nishiyama (1999) による最大不等式を用いて

$$E_{\theta_0}^{n*} \sup_{d_{\theta_0}(\theta, \theta_0) < \delta} |(\Gamma_{\theta_0}^n(\theta) - \gamma_{\theta_0}^n(\theta)) - (\Gamma_{\theta_0}^n(\theta_0) - \gamma_{\theta_0}^n(\theta_0))| \leq \frac{\phi_n(\delta)}{\sqrt{n}}$$

の形の評価を行うことができることである. 結論は, もしも

$$\int_0^\delta \sqrt{\log(1 + N(\Theta_n, \|\cdot\|_\infty; \varepsilon))} d\varepsilon \leq \phi_n(\delta)$$

が満たされるならば,  $n^{-1/2} \phi_n(r_n^{-1}) \leq r_n^{-2}$  を満たす  $r_n$  に対し

$$\lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{\theta_0 \in \Theta} P_{\theta_0}^{n*}(r_n d_{\theta_0}(\hat{\theta}_n, \theta_0) > L) = 0$$

が成り立つ, ということである.

## 参 考 文 献

- Nishiyama, Y. (1999). A maximal inequality for continuous martingales and  $M$ -estimation in a Gaussian white noise model, *Ann. Statist.*, **27**, 675–696.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York.

## 一般指数相反分布とその近似

松 縄 規

正の範囲に分布する連続型分布の密度関数でその主要部分に相反性を持つ一般指数相反分布族について (1) その分布関数の生成基礎偏微分方程式を提案し, それを基に (2) 一般化冪逆正規分布の密度関数の導出とその近似を行った:  $U(\eta, w)$  を分布関数とし, 初期値-境界値条件  $U(\eta, +0) = 0, U(+0, t) = a_0(w)$  および  $\lim_{\eta \rightarrow \infty} U(\eta, w) = 0$  が満たされているものとする. 一般指数相反分布族の基礎方程式が次の偏微分方程式で与えられる.

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

$$a_0(w) = \frac{\nu^{\frac{\rho}{\alpha} + \frac{1}{2}} K_{\pm \frac{1}{2}} \left( \frac{\eta \nu}{2\kappa \alpha^2} \right)}{\eta^{\{2(\alpha-1)\beta+1\} \left( \frac{\rho}{\alpha} + 1 \right)} K_{\frac{\rho}{\alpha}} \left( \frac{\eta \nu}{2\kappa \alpha^2} \right)} w^{\rho + \frac{\alpha}{2}}.$$

ここに,  $\eta > 0, w > 0, \kappa > 0, \nu > 0, \beta > 0, \alpha \& \rho \in R - \{0\}$  であり,  $K_\nu(\cdot)$  は次数  $\nu$  の第2種変形ベッセル関数を表す. 上の方程式を解いて, 微分することにより次の一般化冪逆正規分布の密度関数を得ることを示した:

$$f_{\text{GPIG}}(w; \eta, \kappa, \nu, \alpha, \beta, \rho) = \frac{|\alpha| \nu^{\frac{\rho}{\alpha}} e^{-\frac{\eta \nu}{2\kappa \alpha^2} w^{\rho-1}} w^{\rho-1}}{2 K_{\frac{\rho}{\alpha}} \left( \frac{\eta \nu}{2\kappa \alpha^2} \right) \cdot \eta^{\frac{\rho}{\alpha} \{1+2(\alpha-1)\beta\}}} \exp \left\{ - \frac{\left( \eta - \nu \frac{w^\alpha}{\eta^{2(\alpha-1)\beta}} \right)^2}{4\kappa \alpha^2 \frac{w^\alpha}{\eta^{2(\alpha-1)\beta}}} \right\}$$

この分布族の理論的評価をするために関数  $K_\nu(x)$  ( $x > 0, \nu > 0$ ) を不等式でできるだけ精密に評価することを行った.  $x$  の値が (A) 小の場合 (B) 大の場合に分け, 更に各々に対し (a)  $0 < \nu < 1/2$  および (b)  $\nu > p - 1/2 \geq 1/2$  の場合を考察した. 但し  $p$  は有限項級数近似の際の項数を表す. その結果の一部分として (A) かつ (a) の場合次の不等式を得た:

$$\frac{\gamma\left(\nu + \frac{1}{2}, 2x\right)}{2^{\frac{1}{2}-\nu}} + \frac{x^{\frac{1}{2}-\nu} \left(\frac{1}{2} + \nu\right) \Gamma(\nu) \Gamma(2\nu, 2x)}{2^{\frac{1}{2}-\nu} \sqrt{\pi}} < \frac{K_\nu(x)}{K_{\frac{1}{2}}(x)} < \gamma\left(\nu + \frac{1}{2}, 2x\right) + \frac{x^{\frac{1}{2}-\nu} \Gamma(\nu) \Gamma(2\nu, 2x)}{2^{\frac{1}{2}-\nu} \sqrt{\pi}},$$

ここに,  $\Gamma(\cdot), \gamma(\cdot, \cdot)$  は各々ガンマ関数, および下側不完全ガンマ関数比を表す. これらの近似に関する種々のグラフ化および数値計算も行った.

## 多変量ラグランジュ分布の多変量逆正規分布への収束

南 美穂子

多変量ラグランジュ分布は, 窓口が一つのカウンターで複数のタイプの客がいる場合に, サービスを開始してから初めて客が誰もいなくなるまでの間にサービスを受けた客の, タイプ別の数の分布として特徴付けられる分布である.  $k$  タイプの客がいるとしよう. 初めにいる客の  $k$  次元分布を初期分布, タイプ  $i$  の客へのサービス中に来る客の  $k$  次元分布を到着分布, 初めにタイプ  $i$  ( $i = 1, \dots, k$ ) の客が一人だけいた場合の多変量ラグランジュ分布を基本ラグランジュ分布と呼ぶことにする. 到着分布と基本ラグランジュ分布の確率母関数はラグランジュ変換式を満たす.

一般の多変量ラグランジュ分布は初期分布と基本ラグランジュ分布の複合分布である。これより、到着分布が一定、つまり基本ラグランジュ分布が一定で、初期分布が「大きくなる」、例えば、定数ベクトル  $nl$  の値を確率 1 で取り、 $n$  が無限大に大きくなると、中心極限定理により、基準化した多変量ラグランジュ分布は多変量正規分布に収束する。

Minami (2002) は 2 組の  $k$  個の  $k$  次元分布の間の逆関係をキュムラント母関数を用いて定義した。分布が離散型の場合、ラグランジュ変換式は多変量逆関係を定義する式と同値であることが示せる。ただし、多変量逆関係は、到着分布ではなく、サービス前後の客の減少数の分布と基本ラグランジュ分布の間の関係である。

すべての到着分布のキュムラント母関数がある関数の定数倍になっている多変量ラグランジュ分布の部分族を考える。このとき、基本ラグランジュ分布のキュムラント母関数も同様な構造を持つ。多変量逆正規分布 (Minami (2002)) はこの部分族と同様な形で多変量正規分布と逆関係を持つ分布として定義されている。多変量ラグランジュ分布に従う確率変数を適当に線形変換したとき、いくつかの条件の下で多変量逆正規分布に収束することは、逆関係を持つ分布の 3 次以上のキュムラントが 0 に収束すること、つまり、多変量正規分布に収束することによって示せる。

## 参 考 文 献

- Minami, M. (2002). A multivariate extension of inverse Gaussian distribution derived from inverse relationship, Research Memo., No. 830, The Institute of Statistical Mathematics, Tokyo.

## 有限混合モデルの尤度関数の性質とその EM アルゴリズムへの応用

福 水 健 次

有限混合モデルのように、パラメータを持つ非線形関数(密度関数)の線形和でかかれるような関数に対して(マイナスの)対数尤度関数のようにサンプルに関する和を取った形の目的関数を最小化する問題を扱う。このような特殊な形を持った目的関数の臨界点に関して、以下のような理論的な結果を得た。まず、混合モデルのコンポーネント数が  $H$  のモデルに対する目的関数の臨界点を考え、この臨界点を、コンポーネントを複製して線形係数を任意の比率で分割することによって、コンポーネント数  $H+1$  個のモデルのパラメータを作る。すると、このパラメータはコンポーネント数  $H+1$  個に対する目的関数の臨界点にもなっている。線形係数の分割の比率は任意なので、臨界直線が常に存在していることがわかる。さらに、この臨界直線上の点が、鞍点あるいは極小点であるかを判定するための十分条件を求めた。

以上の理論的結果を、平均と分散共分散行列をパラメータを持つ正規混合モデルに当てはめると、コンポーネント数  $H$  に対する尤度関数の極大点を、任意のコンポーネントの複製によって  $H+1$  個のモデルに埋め込んだ点は、常に鞍点となることがわかった。このことは、ひとつのコンポーネントを 2 つに分割する際に、平均と分散共分散行列を適当な方向に微小に変動させると、必ず尤度が増加することを意味している。

この事実を EM アルゴリズムによる正規混合モデルの最適化に適用し、EM アルゴリズムの問題点として知られている初期値依存性を回避するアルゴリズムを提案した。あるコンポーネント数において EM アルゴリズムが収束した際に、尤度が最も上昇するようにコンポーネント

を分割し、それを初期値としてコンポーネント数が 1 個増加したモデルで EM を行う。コンポーネント 1 個から出発して、このプロセスを適当なコンポーネント数まで繰り返していく。実際にデータを用いて実験を行ったところ、ランダムに初期値を配置する方法に比べ、EM アルゴリズムの局所極大を回避するのに有効に働くことが確認できた。

## On the Tail of Semistable Laws

志 村 隆 彰

$\lim_{x \rightarrow \infty} \mu * \mu(x, \infty) / \mu(x, \infty) = 2$  を満たす  $(0, \infty)$  上の分布は subexponential 分布と呼ばれる ( $S$  と書く)。片側無限分解可能分布においては、分布自身が  $S$  に属すること、正規化された Lévy 測度が  $S$  に属すること、無限分解可能分布とその Lévy 測度の裾が漸近的に等しいことの 3 つが同値であることが知られている。例えば、片側安定分布は  $S$  に属するので、その裾の挙動は分かっていると言える。しかし、不連続な Lévy 測度をもつ半安定分布の場合、 $S$  の理論を使うことは出来ない。そこで、 $S$  の概念を拡張して、

$$\limsup_{x \rightarrow \infty} \mu * \mu(x, \infty) / \mu(x, \infty) < \infty$$

を満たす分布族を考え、O-subexponential 分布族と呼ぶ ( $OS$  と書く)。 $OS$  の研究の基本方針は  $S$  との比較、対応する結果の拡張であるが、単に形式的なものではなく両者には基本的な違いが見られる。例えば、 $S$  が合成積について閉じていないのに対し、 $OS$  は閉じている。一方で、合成積根についてはその逆になることがいえる。考察の結果、 $OS$  と無限分解可能分布との関係、関連する様々な分布族も含めた合成積についての性質、そうした分布族同士の関係性、具体例の構築などの諸結果を得た。表題の問題については特に一般の半安定分布の裾とその Lévy 測度の比の下極限と上極限を求めることが出来た。尚、この研究は渡部俊朗氏(会津大学)との共同研究である。

## 多重分光画像による土地被覆分類とその応用

(客員) 広島大学 西 井 龍 映

多くの衛星や航空機搭載の地球観測センサは地表面の可視光や赤外域の複数の波長帯(バンド)での反射量や放射量を計測している。そのような多重分光画像の主要な解析目的の一つに土地被覆の分類や分類結果の時空間的な定量評価がある。例えば森林減少、植生やサンゴ礁の時系列的変化、流氷の位置の把握は、あるカテゴリで覆われている地表面を特定することが基本である。このような空間データの判別に対してはニューラルネットや機械学習からのアプローチもあるが、空間情報をモデル化しやすい統計的アプローチが重要視されている。本報告では、ある手法によって与えられた判別結果を利用して、判別結果を改良する問題を考察する。

まず与えられた判別結果が必ずしも良いものであるとは限らないことを前提にする。そこで高性能であることが確認されているマルコフ確率場に基づく判別方式を基準にし、与えられた判別結果は参考資料とすることを提案する。テストデータに、カテゴリが与えられたとき特徴ベクトルは独立であり、カテゴリは局所依存するマルコフ確率場に従うと仮定する。ここで、

確率場の母数を減らすために、分布間のマハラノビスを用いた 1 母数で特徴付けされる分布を仮定する。同時に影響を及ぼす近傍半径も未知母数である。

判別の手順は、2 つの未知母数を固定して、ICM( iterative conditional modes )で判別を行い、事後確率を最大にする母数組で最適な判別結果を得る。そこで事後確率は固定した母数に対し、ICM による結果と与えられた判別結果が一致すれば事後確率を大きく、そうでなければそのまま変更したものである。この結果は IEEE の判別のベンチマークデータに適用し、判別率を 89.7% から 94.5% に改良した。

## 大偏差理論の $k$ 標本選択問題への応用

( 客員 ) 明治学院大学 竹 内 啓

### 1. 多次元分布の大偏差近似

積率母関数を持つ連続分布に従う独立確率変数の和の分布に対する大偏差近似 Large deviation approximation についてはよく知られている。全く同じ考え方により多次元分布についてもその密度関数を近似することができることは知られている。

すなわち  $X_1, X_2, \dots, X_n$  を独立に  $p$  次元同一連続分布に従う確率変数、その密度関数を  $f(x)$ 、積率母関数を  $M(\theta) = E(\exp \theta' X_1)$ 、特性関数を  $M(it)$  とし、 $\theta$  を母数とする分布族  $f_\theta(x) = M(\theta)^{-1} e^{\theta' x} f(x)$  を定義する。 $\bar{X}_n = \sum \bar{X}_i / n$  とし、その分布の密度関数を  $f_{n,\theta}(x)$  とすれば

$$f_{n,\theta}(x) = \frac{1}{(2\pi)^n} \int \cdots \int M(\theta) M\left(\theta + \frac{it}{n}\right)^n e^{-itx} dt.$$

また  $f_{n,0}(x) = M(\theta)^n e^{-\theta' x} f_{n,\theta}(x)$  となる。

ここで  $K(\theta) = \log M(\theta)$  とし、 $\partial K(\hat{\theta}) / \partial \theta = x$  により  $\hat{\theta}$  で定義すれば、 $f_{n,\hat{\theta}}(x)$  は  $n^{-1}$  のべきで漸近級数展開され、それから

$$f_{n,0}(x) = c_p M(\hat{\theta})^n e^{-n\hat{\theta}' x} \left(1 + \frac{\square}{n} + \frac{\square}{n^2} + \cdots\right)$$

という形の展開が得られる。また片側確率については、 $\hat{\theta} > 0$  ならば

$$\int \cdots \int_{y \geq x} f_{n,0}(y) dy = c_p M(\hat{\theta})^n e^{-n\hat{\theta}' x} \frac{1}{\prod \hat{\theta}_j} \left(1 + \frac{\square}{n} + \frac{\square}{n^2} + \cdots\right)$$

という形の展開が得られる。これは多次元正規分布の一般 Mill's 比の展開に対応するが、この収束は遅いので

$$c_p M(\hat{\theta})^n e^{-n\hat{\theta}' x} \prod \frac{1 - \Phi(\hat{\theta}_j)}{\phi(\hat{\theta}_j)} \left(1 + \frac{\square}{n} + \frac{\square}{n^2} + \cdots\right)$$

という形の展開を考える。

### 2. ベルヌーイ系列からの選択

上記の理論を次のような問題に適用することができる。

$X_1, X_2, \dots, X_k$  が互いに独立に 2 項分布  $B(n, p_i)$   $i = 1, \dots, k$  に従うとき、

$$P\left\{X_i = \max_{1 \leq j \leq k} X_j\right\} = P\{X_1 \geq X_2, \dots, X_1 \geq X_k\}$$

$$= P\{X_1 - X_2 \geq 0, \dots, X_1 - X_k \geq 0\}$$

として前節の議論を格子点分布について適用する．これを  $k$  個のベルヌーイ系列からの選択問題に応用することができる．

## ストレス関連語と項目反応

(客員)大阪樟蔭女子大学 土 井 聖 陽

小中学生の日常的な疲労感や高校生のストレスなどが問題になっている．本研究では男子大学生をサンプルとして，産業衛生分野の疲労自覚症状しらべ，心理学分野の YG 性格検査，看護学分野のバーンアウトスケールという 3 種類の質問紙とストレスと疲労に関する自由回答を実施した．

3 種類の質問紙は特徴的な結果を示さず，さらに 3 質問紙毎の最高点と最低点の自由回答を比較しても違いは全て不明確であった．次に自由回答はテキストファイル化され，分かち書きを施されて WordMiner によって対応分析やクラスター化などがなされ，項目反応との関連が検討された．自由回答から得られた単語は，疲労に関する「だるい，ねむい，疲れる」など，フラストレーションに関する「イライラ，めんどうくさい，うざい」など，攻撃性に関する「むかつく，バカ，なぐりたい」などとそれら以外の概ね 4 種類に分類された．自由回答で得られた攻撃性の関連語は，YG 性格検査の攻撃性尺度の項目とは表現が異なっていたし，バーンアウトスケールの項目と同じような表現の単語は特に現れなかった．疲労に関しては，「だるい，ねむい，疲れる」などと同じ単語を含む項目があったが，両者の関係は独立であった．この両者の関係を調べると，サンプルは，項目と自由回答の両方に反応する群，片方にのみ反応する群，両方に反応しない群の 4 種類に分類が可能であった．

## 計算精度の検証

田 村 義 保

パーソナルコンピュータの計算速度は 10 年前のスーパーコンピュータを凌駕するようになっている．そのために，以前は汎用計算機やスパコンで行われていたようなデータ解析を机上の高々 30 万円のパソコンで実行することが可能になってきた．さらに，市販の統計ソフトウェアパッケージも充実してきたために，ソフトウェアを FORTRAN や C 言語で自作するのではなく，市販のものを使うようになっている．

計算機が出現した時から利用者はおそらく，計算機の出す結果は正しいと思い続けていたのであろうが，最近利用層が広がったために，ますますそのような傾向にあるものと思われる．このような傾向に警鐘をならすために，計算機は必ずしも正しい答えを与えてはくれないことを例示したい．データとしては，米国 NIST の StRD のアーカイブデータを用いる．URL は次のとおりである(<http://www.itl.nist.gov/div898/strd/>)が，この中から分散分析のサンプルデータ SmLs07 を選ぶ．いくつかの市販ソフトウェアでの検証結果は次のとおりで StRD の列が正解であるが，今回示した結果はほぼ満足できる．有名なソフトウェアでも目を覆いたくなるようなものもある．



表 1. SmLs07 の分散分析の結果比較 .

| ソフト名 |      | NAG              | Splus V4.5 | JMP V4.0.5J | Statistica V5.5 | StRD                  |
|------|------|------------------|------------|-------------|-----------------|-----------------------|
| 処置   | 平方和  | 1.6801563955E+10 | 1.680158   | 1.6832929   | 1.679180        | 1.680000000000000E+00 |
|      | 平均平方 | 2.1001954944E-01 | 0.2100197  | 0.210412    | 0.209898        | 2.100000000000000E-01 |
| 残差   | 平方和  | 1.8000978373E+00 | 1.800125   | 1.7969654   | 1.800027        | 1.800000000000000E+00 |
|      | 平均平方 | 1.0000543541E-02 | 0.0100007  | 0.009983    | 0.010000        | 1.000000000000000E-02 |
| F値   |      | 21.0008134646    | 21.00051   | 21.0767     | 20.98944        | 2.100000000000000E+01 |
| p値   |      | 2.5790717703E-22 | 0          | <.0001      | 0.000000        | 記載無し                  |

## TIMSAC72 の計算法について

荒 畑 恵美子

TIMSAC72 を使用する人にプログラムの内容がわかるようにマニュアルを作った . TIMSAC72 の計算法についての説明をした .

### 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 の言語について

中 野 純 司

われわれが開発している統計解析システム Jasp は, 統計解析のために新たに設計されたプログラム言語を備えている . Jasp 言語では, まず, 個々の比較的単純な手法を実現するために, 関数を中心としたプログラムを書くことが推奨される . Jasp 言語は文法の簡単な, 型のないスクリプト言語であり, 統計解析でよく用いられる, 行列計算, グラフィックス, 確率分布に関する計算, 乱数などの関数が組み込まれている . さらに Java のクラスファイルを直接利用することができる . これらを用いることにより, 新しい統計手法の一部を構成するような関数を, 容易に実現することができる . そして, 作成した一連の関数を, 後の再利用のために整理しておきたい場合には, それらを少数のクラスオブジェクトとしてまとめるのがよい . Jasp クラスは継承関係を持つことができるが, これを利用すると, 統計手法の整理や, 一般的な手法を自分の必要に応じて変更することが, 簡単に行える .

Jasp 言語では, Java 言語における Javadoc と同様な形式のコメントを用いることで, そのプログラムをグラフィカルユーザインタフェース( GUI )から利用しやすくするための情報を付加できる . GUI はそれらのコメントをもとに, 入力の際に変数の説明を表示し, 入力が適当かどうかのチェックを行える . このようなコメントのないメソッドは重要でないと思われ, GUI 上には表示されない .

Jasp 言語は, Java 言語で作成されたスクリプト言語である Pnuts 言語をもとにして実装された . 統計に必要な Java ライブラリを付加するとともに, プリプロセッサを作成することにより, その機能を実現している .

## 質の高い証拠を得るための要件

柳 本 武 美

通常の科学研究では、注目される新しい知見を提出することに主眼が置かれ、その知見を導いた証拠の質の高さを問うことはない。しかしながら一方では可能な限り質の高い証拠を求める必要がある分野も少なくない。そこで質の高い証拠を得るための要件を統計学の立場から研究した。その結果は各種研究会で報告したのを初め、本年度の統計学会チュートリアルセミナーでも概要を解説した。また統計学会誌にも投稿した。

この発表では、データを見てから統計解析の方法を決める危険性について考える。通常はデータを得た後で、異常と思われるデータを除去する、そしてデータから最大限の情報を抽出するための解析法を適用する。ところが臨床試験とか実験物理学の一部では、このような post-hoc 解析が誤りをもたらした例が多く知られている。そこで統計解析法について思考するのは、データを見る以前が望ましいとの理解が進んでいる。そして上に挙げた分野では、研究デザインの段階で解析法を決める。

質の高い証拠を必要とする研究の最前線では、予め解析法を決めておくことの必要性は浸透しつつあるが、多くの研究者には現在でも理解が困難である。この理由の 1 つとして科学的知見の認識論に関連する部分がある。例えばミルの 5 つのカノンは黎明期の我が国の統計学者に大きな影響を与えた。しかしこのような帰納に関する楽観論は、今日では殆ど受け入れられていない。素朴で直観的な推論から、形式的な推論への転換が求められている。そのための 1 つの要件が予め定めた統計解析法である。

### 参 考 文 献

- 柳本武美(2001) 因果推論の方法としての臨床試験, 日本統計学会チュートリアル・セミナー(第9回) テキスト.  
 柳本武美(2002) 実証主義の転換から導かれる統計的推論の視点, 日本統計学会誌(和文誌)(投稿中).

## 結晶群の出現頻度と確率モデル

伊 藤 栄 明

結晶構造の予測の問題は本来自然科学の問題であるが、数理としても興味ある問題を提起し統計学、確率論、計算幾何学、離散幾何学、群論等の接点にある問題である。結晶の対称性は群をもちいてあらわすことができる。化学式についての AX 記号をもちいて結晶物質を分類すると、出現頻度の高い結晶群のあいだには群部分群関係があることが多いことが観察される。群部分群関係のうえでランダムウォークを考えることにより金属結晶における結晶群の出現頻度がよく説明できる。無機結晶構造のデータベース ICSD には 6 万件ほどのデータが収録されている。どのような結晶構造が自然界にあらわれやすいかという問題について ICSD をもちいて統計的研究を行ってきた(Fujiwara et al.(1998), Itoh and Matsumoto(1991))。単位格子中にふくまれる formula unit の個数( $Z$ )の分布と結晶の対称性のあいだには強い関連がある。点群のうえでの分布だけでなく、空間群のうえでの分布をしらべると、たとえば ANX 記号で

$ABX_3$  である物質では  $Z$  は 1 から 32 までの値をとるが、 $Z$  をあたえると少数の空間群のうえに分布していることがわかり、出現しやすい構造はかぎられてくる。ANX 記号および  $Z$  があたえられた場合に結晶構造がどのようなようになるかという問題は統計学に多くの興味深い課題を提起する。

### 参 考 文 献

- Fujiwara, M., Itoh, Y., Matsumoto, T. and Takeda, H. (1998). Statistical distribution of symmetry groups for Inorganic Crystal Structure Database (ICSD), *Comput. Sci. Monographs*, No. 29, 1-92, The Institute of Statistical Mathematics, Tokyo.
- Itoh Y. and Matsumoto T. (1991). Random-generation model for statistical distribution of point groups, *Acta Crystallographica*, **A47**, 204-206.

## 文化を計る

村 上 征 勝

文化現象の計量分析に関する研究の中で、文学作品の計量分析と絵画の計量分析について報告した。

文学作品に関しては、川端康成の作品における文体の変化の有無を計量的に調べた。一般に川端康成の文章は、『みづうみ』(昭和 29 年)という作品から変化したといわれているが、読点のつけ方の情報の計量分析からは、『山の音』(昭和 22 年)から変わったと考えるのが妥当である、つまり戦前と戦後で読点のつけ方が異なるという結果を得た。

絵画に関しては、まず代表的な浮世絵師 9 人の描いた美人の顔の形状の分析を試みた。その結果、絵師ごとに描く顔にそれぞれ特徴があること、また、絵師の活躍期を前・中・後の 3 期に分けてみると、描かれている美人の顔が前期の丸顔から次第に後期の面長な顔へと変化していることなどを報告した。

次に、同じ絵師が描いた顔でも男性の顔と女性の顔では描き方に違いがあるのかどうか、同様に男性の顔と歌舞伎役者の顔では違いがあるのかどうか等の分析結果についても報告した。

## 東アジアの日本語観センサス

吉 野 諒 三

国立国語研究所は、1996～98 年にかけて、世界 28 ヶ国における「日本語」のイメージや好悪、学習状況に関する各国での全国レベルの標本抽出調査を遂行している。我々は、そのデータ分析に共同研究者として参加し、多様な解析作業を試行した。この中で、あまりに多国にわたるデータのため、28 ヶ国全体の明瞭で安定した回答パターンは発見し難いのがわかり、特に通常東アジア・東南アジアと呼ばれる、日本を含む 10 ヶ国に焦点を絞って、分析した。

しかしながら、韓国で程度の差こそあれ、全体の約 50% が日本語学習経験者であるのを例外として、台湾、シンガポール、フィリピン、オーストラリアがそれぞれ約 29.3%、11.3%、

16.0%, 11.0%で, 他の国々は1%以下であり, 日本語に関する印象や好悪を尋ねたデータの妥当性に疑義を感じさせた。このためか, 日本という国に対する好悪や信頼感, 日本人に対する好悪や信頼感, 日本語に対する好悪は, 相関が高く, 多くの国々の人々には, それらのイメージが鋭く弁別されているとはいえない(但し, 個別には, 国によって例えばブラジルは, 日本や日本人に対するイメージは全体の国々の中でもかなり高いが, 詳細に見ると, 日本人に対するイメージは日本という国に対するイメージよりも少し下がっているようである。)

総じて, 分析を進めていく中で, 東南アジア各国の多様性, 各国内の政治・経済・文化の複雑性が浮き上がってきた。一つの見逃せない点は, 華僑・華人の存在と役割が各国の事情や民衆の意識を多様で複雑にしていることである。また, アメリカによる占領により英語を公用語としている国々もある。以上を勘案すると, 今後, 日本語学習者の増大があり得るとするならば, 各国内の経済の向上や政治的安定が達せられた上で, 各国の教育の向上があり, その条件の下で現地での文盲率の低下や公用語の学習レベルの向上がなされ, さらに外国語学習としては, 中国語や英語が進み, そして次に日本語学習へ向く可能性が考えられる。

### 参 考 文 献

吉野諒三, 鄭 躍軍, 朴 承根(2002) 東アジア諸国の日本語観センサス, 行動計量学, 29(2)(印刷中)。

## 日中国民性比較の基礎研究

鄭 躍 軍

日本と中国は重要な隣国として知られており, 両国の国民性を理解し合うことは相互交流の促進に役立つ情報提供の一助となることが期待できる。しかし, 中国では日本のように整備されている住民基本台帳や選挙人名簿がないのが現状である。これまで海外の研究・調査機関により中国でさまざまな意識調査が行われてきたが, 調査状況の実態を把握しないまま現地の関係機関や関係者に委託し, 調査結果が手渡されることを待つ, というような「人任せの調査」がよく見られる。したがって, 中国でどのような標本調査法で意識調査を行うべきかを実証的に研究することが重要である。そこで, 本研究では日本・中国の国民性比較のための基礎研究として, 中国での標本調査システムを実践的に構築し, 厳格なランダム・サンプリング理論に則った意識調査を遂行し, 収集された調査データを探索的に解析し, 国民性の異同を明らかにすることを目的とした。

この問題意識のもとで, 中国本土での標本調査システムを実践的に構築した上で, 中国の国情を代表する北京市と上海市においてランダム・サンプリング理論に則って抽出した市民を対象とした標本調査を個別面接調査法により実施した。標本抽出計画では, 層別3段無作為抽出法を採用して両市からそれぞれ抽出した調査対象個人に2001年11月と2002年2月に本調査をそれぞれ実施した。一方で, 統計的データ分析法をデータ利用上の観点と, 調査企画等調査法上の観点とから検討することにより, 調査方法と調査内容がデータ分析を前提としたものとなり, より一層偏りなく比較可能性の高いデータの収集ができた。

これまでのデータ解析結果により, 日本人と中国人の間で伝統的観念に対する態度がきわめて近い一方で, 現実の諸問題に対する態度は異なることが明らかになった。また, 北京市における標本調査システムの概要と質問文の単純集計表を取りまとめ, 研究レポートとして発行し

た。今後収集されたデータを探索的に解析することによって、日中の国民性の特徴を統計的に解明する方法論を確立していこうと考えている。

## スプレッドシート上での統計計算環境の構築

(客員)立教大学 山口 和 範

近年のパソコンの普及と高機能化、さらには、ネットワーク環境の普及により、統計分析や統計学習の環境が、大学や職場、さらには、家庭へと広がりを見せつつある。そこで、現在、インターネットを利用した統計教育ツールや分析ツールの開発が進められている。旧来、統計分析は自作のプログラムが統計分析専用で作られた統計ソフトウェアで行われてきたが、現在では、教育の場のみならず、実務上の現場でさえ、スプレッドシートソフト上(特に、Microsoft Excel)で行われることが多くなっている。これは、Excel が、個人または企業で購入するほぼすべてのパソコンにインストールされていることに起因していると考えられるが、必ずしも満足な統計分析環境を与えているとは限らない。特に、Excel の統計解析機能が非常に限られているだけでなく、欠測値の処理や数値計算上の精度の問題など種々の問題点の指摘もある。

そこで本研究では、Excel 上での統計計算環境の改良を目的とし、統計数理研究所で公開されている TMSAC の Excel 上での実装を行った。具体的には、TMSAC72 の DLL を読み込む、Excel 上のインターフェースの開発を行った。ここで開発したインターフェースは、VBA マクロによる GUI、ワークシート関数、また、CGI や VBScript などを利用した web 形式の GUI である。ワークシート関数としての利用を可能にしているため、データの変更に伴う結果の自動での再計算にも対応でき、教育リソースとしての利用などで有用であると考えられる。また、通常の GUI に加え、web 形式の GUI の開発を行ったのは、インターネット上での教材資源等の有機的融合を目指してのものである。なお、現状での開発は TMSAC72 の一部機能にとどまっているので、すべての機能の実装が今後の課題である。

## シミュレーション試験による母集団薬物動態試験法の評価\*

(客員)北里大学 竹内 正 弘

### 1. 序論

母集団薬物動態試験を実施し、平均的な薬物動態とそのばらつきを推定することは、医薬品の有効性・安全性を把握する上で重要である。解析するにあたり、非線形混合効果モデルが応用される。ランダム効果が非線形になっているため、母集団パラメータである固定効果を推定する際には、テラー展開により、周辺分布を線形化する方法が一般的に用いられている。本研究では、2つの近似法に焦点をあて、シミュレーション試験を通して、母集団薬物動態試験をデザインする際の統計学的条件を考察する。

---

\* 都合により当日は報告されなかったが、要旨のみ掲載する。

## 2. モデルの定義

各個人で観察される薬物動態データ,  $y_i$  ( $n_i \times 1$  ベクトル,  $1 \leq i \leq N$ ) は次の非線形混合効果モデル,

$$(2.1) \quad y_i = f(X_i, \beta_i) + \varepsilon_i \quad \beta_i = \beta + b_i$$

で定義されるとする. ここで,  $f(X_i, \beta_i)$  は薬物動態の非線形関数,  $\beta_i$  は  $r \times 1$  ベクトルの各個人の回帰パラメータ,  $X_i$  は  $n_i \times t$  の既知デザインマトリックス,  $\beta$  は  $r \times 1$  ベクトルの固定効果パラメータ,  $b_i$  は  $r \times 1$  ベクトルの個人間変動パラメータ,  $\varepsilon_i$  は  $n_i \times 1$  ベクトルの個人内誤差パラメータとする.  $b_i$  は平均 0, 共分散  $\Psi$  の正規分布 ( $N(0, \Psi_{r \times r})$ ),  $\varepsilon_i$  は平均 0, 共分散  $\sigma^2 A_i(\gamma)$  の正規分布 ( $N(0, \sigma^2 A_i(\gamma)_{n_i \times n_i})$ ,  $\gamma$  は未知のパラメータ) に従うとし,  $b_i$  と  $\varepsilon_i$  は互いに独立であると仮定する. 母集団薬物動態試験において最も興味あるパラメータは, 非線形関数で表される薬物動態の固定効果,  $\beta$  である.

ランダム効果の期待値, 0 周りで 1 次のテーラー展開 (Sheiner and Beal(1980, 1985), Vonesh and Carter(1992)) をすると,  $y_i$  の分布は以下のように表される.

$$(2.2) \quad y_i = f(X_i, \beta, b_i = 0) - \left[ \frac{\partial f}{\partial b_i^T} \Big|_{b_i=0} \right] b_i + \varepsilon_i$$

個人間変動パラメータ,  $b_i$  は  $y_i$  の期待値には影響せず, 共分散に影響を与えており (1) 個人間のばらつきが無視できる場合には, 尤度関数への近似として妥当である (Solomon and Cox(1992)) こと (2) 周辺分布の共分散パラメータ,  $\gamma, \sigma^2$  を推定する場合, 期待値が正しく定義できている場合に限り妥当である (Breslow and Clayton(1993)) ことの 2 点が注意点である.

ランダム効果の推定値,  $\hat{b}_i$  の周りでテーラー展開 (Lindstrom and Bates(1990)) をすると, 観察データ,  $y_i$  の分布は以下のように表される.

$$(2.3) \quad y_i = f(X_i, \beta, b_i = \hat{b}_i) - \left[ \frac{\partial f}{\partial b_i^T} \Big|_{b_i=\hat{b}_i} \right] (b_i - \hat{b}_i) + \varepsilon_i$$

$b_i$  の推定値,  $\hat{b}_i$  が期待値, 共分散の関数となっているため,  $\beta$  の推定値  $\hat{\beta}$  の漸近性に影響を及ぼしている. この尤度関数に対して Laplace 近似法を応用 (Wolfinger(1993), Vonesh(1996)) すると,  $\hat{\beta}$  の漸近性については

$$(2.4) \quad \hat{\beta} - \beta = O_p \left[ \max \left\{ \frac{1}{\sqrt{N}}, \frac{1}{\min(n_i)} \right\} \right]$$

が証明されている (Vonesh(1996)). この条件によれば,  $\hat{\beta}$  の一致性を確立するためには, 被験者数  $N$  と各個人の最小観察データ数  $n_i$  が同じ速さで増加しなければならない.

## 3. シミュレーション試験

反復投与における個人毎の血中濃度の薬物動態データをシミュレーションし, 全体の被験者数  $N$ , 個人毎の採集データ数  $n_i$  の 2 条件が, 上記で定義された 2 つのテーラー展開 (期待値周りと推定値周り) で推定される固定効果パラメータへの影響を考察した (Mimani(2002)).

## 4. 考察

母集団薬物動態試験では, 各個人から採取するデータ数を少なくし, 被験者数を増やすことに特徴がある. 各個人間変動のばらつきを無視できる場合には, 0 周りでテーラー展開する方

法が適切である。全体の被験者数  $N$  を増やすことで  $\hat{\beta}$  の一致性が求められる。母集団薬物動態試験では、個人間のばらつきを仮定し、ばらつきを調整して母集団の薬物動態パラメータを推定することが目的である。個人毎の  $b_i$  の推定値  $\hat{b}_i$  の周りでテラー展開する方法が適切であるが、 $\hat{\beta}$  の一致性を得るためには、式(2.4)の条件が伴い、母集団薬物動態試験の特徴である、 $n_i$  を少なくし、 $N$  を大きくすることに相反している。

このシミュレーション試験を通して、個人間のばらつきが無視できない場合には、期待値 0 の周りでテラー展開法では、固定効果パラメータの推定値にバイアスが含まれるが、被験者数が増加すれば固定効果パラメータの推定値の精度はよくなる。推定値  $\hat{b}_i$  の周りでテラー展開法では、固定効果パラメータの推定値に含まれるバイアスは小さくなるものの、固定効果パラメータの推定値の精度は被験者数と各個人毎のデータ数の割合に依存する。

将来、これらの条件を満足できない場合には、データへの goodness of fit を考慮して、どちらのテラー展開法が妥当であるかを研究していく計画である。

### 参 考 文 献

- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *J. Amer. Statist. Assoc.*, **88**, 9–25.
- Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data, *Biometrics*, **38**, 963–974.
- Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measure data, *Biometrics*, **46**, 673–687.
- Mimani, F. (2002). Statistical consideration in population pharmacokinetics (PPK) study — Via simulation study —, Master Thesis at Kitasato University Graduate School.
- Sheiner, L. B. and Beal, S. L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data, *Journal of Pharmacokinetics and Biopharmaceutics*, **8**, 553–571.
- Sheiner, L. B., and Beal, S. L. (1985). Pharmacokinetic parameter estimates from several least squares procedures: Superiority of extended least squares, *Journal of Pharmacokinetics and Biopharmaceutics*, **13**, 185–201.
- Solomon, P. J. and Cox, D. R. (1992). Nonlinear component of variance models, *Biometrika*, **79**, 1–11.
- Vonesh, E. F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models, *Biometrika*, **83**, 447–452.
- Vonesh, E. F. and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures, *Biometrics*, **48**, 1–17.
- Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models, *Biometrika*, **80**, 791–795.