統計数理（1997）
第45巻 第2号 409-442

平成8年度研究報告会要旨

と　き：1997年3月18日　午前10時～午後5時15分
　　　　　　　　19日　午前10時～午後4時45分
ところ：統計数理研究所 講堂

プ ロ グ ラ ム

3月18日（火）

午前（10時～12時30分）

あいさつ　所長　清水　良一

【調査実験解析研究系】

日本人の国民性調査に関する基礎的検討	坂元　慶行
有限母集団からのサンプリングとブートストラップ	中村　　隆
偏対応分析とその応用例について	前田　忠彦
ランダムネットワーク上の空間配置	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のある一般化について	金藤　浩司
集団健康検診と個体評価	駒澤　　勉
官庁統計におけるリサンプリング法の活用	馬場　康維
一次元尺度構成のための質的項目の選択法	土屋　隆裕
自由回答データの分析――いくつかの実験調査でみられた特徴，問題点――	大隅　　昇

午後（13時30分～17時30分）

HALWINの開発	高木　廣文
質的データを用いた森林のパターン分類について	鄭　　躍軍
地震活動の時空間モデル	尾形　良彦
ベイズ的方法によるデータ解析	柏木　宣久

【予測制御研究系】

情報量規準EICについて	北川源四郎
A Visual ‘Test’ Tour of the Extended Nelson-Plosser Data	川崎　能典
モデル選択とその応用	下平　英寿
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予測	佐藤　整尚
線形計画問題のセンターパスと内点法	水野　眞治
離散分布の平均差	鈴木義一郎
半正定値計画法について	土谷　　隆

平均場近似の統計学における応用　伊庭　幸人
非定常時系列の解析における2次元スペクトルの帯域制限　瀧澤　由美

【客員】
ミクロデータを用いた生産関数の推定　(新潟大学)　和合　肇
統計教育としての配列数学とJ言語　(千葉工業大学)　西川　利男

3月19日（水）

午前（10時～12時30分）

【統計データ解析センター】
偽データ混合法　石黒真木夫
ゲームにおける内部モデルの生成　泰地真弘人
ヌジ人名資料概観　上田　澄江
必要病床数の予測　田村　義保
時系列の欠測値処理のアルゴリズムについて　荒畑恵美子

【統計基礎研究系】
成分間に依存性のある consecutive systems の生存時間分布　平野　勝臣
滑らかな凸錐を対立仮説とする尤度比検定について　栗木　哲
K-L情報量による分布の近似の再考　松縄　規
非線型回帰の情報幾何　江口　真透
On Distributions with Regularly Varying Tails　志村　隆彰

午後（13時30分～16時45分）

【領域統計研究系】
SLTEC関連の事故における原因究明から　柳本　武美
相互作用，交互作用，加法モデル　佐藤　俊哉
擬似尤度の構成について　汪　金芳
群，部分群関係と結晶群の出現頻度　伊藤　栄明
高校・大学における統計教育の現状と問題点　村上　征勝
世論調査データにおけるマジック・ナンバー　吉野　諒三
米国の大学教員募集要項の分析　釜野さおり

【統計教育・情報センター】
分子系統学の諸問題　長谷川政美
翻訳関連蛋白質全アミノ酸配列データのアライメント　橋本　哲男
Distribution Theory in a Higher-order Markov Chain　内田　雅之
非ガウス外乱に対する線形系の応答　岡崎　卓
第2次銀行自己資本規制と監査モデル　山下　智志

【客員】
超多項変動を持つデータの分析　(大分大学)　越智　義道

調査実験解析研究系

日本人の国民性調査に関する基礎的検討

坂 元 慶 行

今年度は，(1) EIC による密度関数推定法の再構成，(2) 国民性調査のデータの整備と将来への準備等を行った．(1) では，情報量規準 EIC を導入することによって，各種の密度関数推定法のデータだけに基づく比較が可能になり，ノンパラメトリック検定等への応用も明快になり，性能も向上するという知見を得た．また，(2) では，国民性調査の一部項目の再入力，調査システムによる結果の違いの考察，質問文による回答の揺らぎの検討等を行った．

報告では，(2) に関する成果のうち，特に人間関係に関連した問題を中心に，質問文と回答の揺らぎについて，下記のような例を挙げて述べた．

国民性調査の最大の知見は「(他の領域のさまざまな問題に関する意識の変化は大きかったが) 身近かな人間関係に関する意識は変化が小さかったこと」であるが，その人間関係の有力な維持基盤の一つは日本的経営に特徴的な企業内人間関係にあると考えられる．しかし，近年の経済社会的環境条件の変化は人間関係（観）の変化をも促進しつつあるとの指摘があり，世論調査でも「能力主義を是とする会社員が7割」などという結果が発表されたりもする．しかし，確かに，「給料を決めるのに，功労を重視すべきか現在の能力を重視すべきか」と質問すると約6割が能力派だが，「功労を重視すべきか，能力だけで決めるべきか」とすると多数意見が逆転するという結果が得られる．“能力は重視すべきだが，能力だけでの評価は困る”ということであろうか．いずれにせよ，この種の調査結果の解釈は単純にはいかないようである．

つぎに，総務庁青少年対策本部による日米韓3国での「子供と家族に関する国際比較調査」によれば，子供に望む性格特性として「他人のことを思いやる心」を挙げる比率が，米韓に比べ，日本では62%（1位）と際だって高い（これは，たとえば「責任感」の選択率がどの国でも高いのと対照的である）．この結果は，「思いやり」という語の語感に因る懸念も残るため，この語の代わりに「他人の気持ちを考えられること」として調査してみたが結果に変わりはなかった．のみならず，日本では「規則を守り，人に迷惑をかけない公共心」も2位と高く，これらがいずれも他人を意識した性格特性であり，それが幼少時から望ましい性格特性とされていることは，日本における人間関係（観）がもつ特別の地位を象徴する結果とも見られ，注目すべき項目かと思われる．

有限母集団からのサンプリングとブートストラップ

中 村 隆

有限母集団からのサンプリングで得られたデータにブートストラップ法を適用する際には注意が必要である．たとえば，大きさ N の母集団から非復元単純ランダムサンプリングで大きさ n の標本を得たとき，標本平均の推定分散は $\mathrm{var}(\bar{y})=(1-f)(s^2/n)$ で与えられるが，この標本から復元単純ランダムサンプリングで大きさ n の再標本を得たときのブートストラップ分散は $\mathrm{Var}_*(\bar{y}^*)=((n-1)/n)(s^2/n)$ であり，推定分散と一致しない（ここで，抽出率 $f=n/N$，標本分散 s^2）．この問題はスケーリング問題と呼ばれ，層の数が多い層別サンプリングや多段抽出では問題が深刻になる．

スケーリング問題を克服する方法として，次のような方法が提案されている．BWO

(Without-replacement Bootstrap; Gross (1980)), BWR (With-replacement Bootstrap; McCarthy and Snowden (1985)), BRS (Rescaling Bootstrap; Rao and Wu (1988)), BMM (Mirror-match Bootstrap; Sitter (1992)). いずれも線形推定量の場合に，その推定分散とブートストラップ分散を一致させることを目指している．

本年度は，「日本人の国民性調査」データを対象に，上の諸法およびジャックナイフ法を適用して，調査の精度に関する情報を得るとともに，それぞれの方法の特性・得失などについて比較・検討を行なった．

参考文献

Gross, S. (1980). Median estimation in sample surveys, *Proceedings of the Section on Survey Research Methods*, 181-184, American Statistical Association.

McCarthy, P. J. and Snowden, C. B. (1985). The bootstrap and finite population sampling, *Vital and Health Statistics*, 2-95, Public Health Service Publication 85-1369, US Government Printing Office, Washington, D.C.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data, *J. Amer. Statist. Assoc.*, **83**, 231-241.

Sitter, R. R. (1992). A resampling procedure for complex survey data, *J. Amer. Statist. Assoc.*, **87**, 755-765.

偏対応分析とその応用例について

前 田 忠 彦

第三の項目 C によって層別化された二つの質的変数（項目）A, B 間の二元分割表がある場合に，C の A, B に対する影響を考慮しつつ二項目間の連関を分析する方法について，対応分析（以下 CA）関連手法の適用という文脈から考察した．特に柳井（1994）による「偏対応分析」（Partial Correspondence Analysis; 以下 PCA）について，実データの分析例及び人工データの解析に基づき，その実用的な意味について考察することに主眼を置いた．

PCA は，第三の項目によって層別化された二元分割表がある場合に，当該2項目から，層別変数の影響を取り除いた上で，2項目の各カテゴリに与える重みを，項目間の（偏）相関が最大になるように定めるための偏正準相関分析に相当する手法であり，上記の様な層別化された分割表の分析方法として有用である（前田（1996））．

層別変数をいずれも「性別」とした二種類の実データについて，層別の二元分割表の CA，層を無視した（全体の）二元分割表の CA，及び PCA による分析結果を比較し，偏対応分析の性質について考察した．この結果から，1）PCA は概して男女別に行った CA の結果の中間的な解を与え，かつ全体の二元分割表に対して行った CA の結果と類似度の高いものになる場合が多いが，全体データの CA と異質な解を与える場合もある，2）PCA により与えられる偏連関の値が，全体データの CA の連関よりも非常に小さい場合には，全体データの分析による連関は一種の疑似連関と解釈するのが妥当と考えられる，といった知見が得られた．

上記の他に，小さなサイズの二元分割表の人工データに対する各種分析手法の適用結果から，偏対応分析の解のいくつかの性質について言及した．

参考文献

前田忠彦(1996). 層別化された2元分割表の解析と偏対応分析，統計数理研究所共同研究レポート，No. 86「多変量質的データの構造解析に関する研究」，52-59.
柳井晴夫(1994). 『多変量データ解析法──理論と応用──』，朝倉書店，東京.

ランダムネットワーク上の空間配置

種 村 正 美

空間に配置する点集合が何らかの構造を持つかのように観測されることがしばしばある．そのような配置の中で，ランダムなネットワーク構造の上の配置およびそこからのランダムな変位によって派生した配置に注目する．

われわれはまず，2次元平面でのランダムネットワーク上の配置データをシミュレーションによって作成し，その配置を特徴づけるために，2次モーメント測度である K 関数およびその変形である L 関数を用いた．K 関数の経験分布は2点間の距離が r 以下である対の個数の期待値に比例するものとして定義される．特に，一定の密度の Poisson 配置（完全ランダム配置）の場合は K 関数は漸近的に(すなわち，十分広い領域で観測されたという条件の下で)，半径 r の円の面積 πr^2 で与えられることが知られている．通常は，与えられた配置が Poisson 配置からどの程度はずれているかを診断的に見易いように，K 関数を変形した L 関数が多用される．L 関数は Poisson 配置に対して $L=r$ で与えられる．実際の観測は，有限領域における有限個の点配置のデータであるため，そのための補正が K 関数等の推定に必要である．われわれは，これらの関数の不偏推定のための補正の計算方法を示した．

続いて，上記の人工データに対して，有限性の補正を加えた2次モーメント測度を求めた．その結果，ランダムネットワークの各辺の上で一定密度で生成した Poisson 配置に対して，集中型の K, L 関数が得られた．また，ネットワークの辺から所与の分散の正規分布によって変位させた人工データに対して K, L 関数を求め，分散を大きくすると集中の程度が少なくなることを示した．

さらに，3次元空間の実データを2次モーメント測度で解析した．このデータは，アルミニウムの多結晶中に観測される空孔（ポア）の配置である．解析の結果，K, L 関数は集中型を示した．一方，この空孔はアルミニウムの結晶粒界（すなわち多面体分割構造の境界）付近にでき易い可能性も示唆されている．実際，空孔の観測データを立体視できるグラフを作成したところ，多面体構造の境界付近に空孔が散布していることが目視できることが分かった．そこで，われわれは3次元のランダムなネットワークの上の Poisson 配置をシミュレーションで作成し，2次モーメント測度を求めたところ，アルミニウムの空孔のデータに対する K および L 関数と非常に類似した集中型の結果が得られた．この結果は上記の実験的示唆を強く支持するものである．

なお，今回われわれはランダムネットワークの構造のモデルとして，2次元および3次元の Voronoi セルをそれぞれ用いた．今後は，観測された配置データからネットワーク構造を推定する問題を含めた研究が必要である．

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 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一昨年，昨年に引き続いて行なっている．

昨年度は FreeBSD 1.5.1.1 上で行なっていた開発を今年度は FreeBSD 2.1.5 に移して開発を行ない，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，UNIX を素人が使う場合の便宜に重点に置いた作業を行なった．これによりある程度の完成度を持ったシステムをパッケージ化して提供できるようになった．

FreeBSD に関する趨勢を見ると，昨年度は，S や Netscape など，人気が高い有償のソフトウェアの FreeBSD 2.1 対応版が出た．このような有償ソフトウェアが FreeBSD では使えなかったことが従来 FreeBSD を使うことのデメリットだっただけに，昨今のこのような動きは歓迎されるべき事態である．統数研で従来開発されてきたソフトウェアにも S を基盤としたものが多くあるので，今後それらを FreeBSD 上に移植する作業を行なっていくことが可能である．

指数逆ガウス型分布のある一般化について

金 藤 浩 司

実数上で定義される 2 母数の歪んだ分布，極値分布 (Gumbel 分布) および指数逆ガウス型分布 (Kanefuji and Iwase (1996)) に対して，新たに一つの無名数の母数を加え，その母数の関数として母歪度が表される一般化の方法を提案した．ここで提案した一般化の方法は，Iwase and Kanefuji (1994) で行った 3 母数化の方法に関係している．まず，3 母数ガンマ分布および 3 母数逆ガウス型分布を，以下の様に定義する．

$$\left(1+\lambda\frac{X-\mu}{\sigma}\right) \sim Ga(1, |\lambda|^2)$$
$$\left(1+\lambda\frac{X-\mu}{\sigma}\right) \sim IG(1, |\lambda|^2)$$

ここで，$Ga(m, c^2)$ および $IG(m, c^2)$ は，それぞれ母平均 m，母変動係数 c のガンマ分布および逆ガウス型分布を表し，$\mathrm{sgn}(\lambda)\cdot\{X-(\mu-\sigma/\lambda)\}>0$, $-\infty<\mu<\infty$, $0<\sigma<\infty$, $0<|\lambda|$ である．また，$\mathrm{sgn}(x)$ は，x の符号を表す．

本報告では，$\exp(\lambda(X-\mu)/\sigma)$ の級数展開の近似として上記の左辺を捉え，極値分布および指数逆ガウス型分布の一般化を以下の様に提案した．

$$\exp\left(\lambda\frac{X-\mu}{\sigma}\right) \sim Ga(1, |\lambda|^2)$$
$$\exp\left(\lambda\frac{X-\mu}{\sigma}\right) \sim IG(1, |\lambda|^2)$$

で定められる X の従う分布をそれぞれ，一般化極値分布および一般化指数逆ガウス型分布と定義した．また，$-\infty<X<\infty$, $-\infty<\mu<\infty$, $0<\sigma<\infty$, $0<|\lambda|$ である．

本報告では，無名数の母数を一つ導入することで，正規分布を含み母歪度が一つの母数の関数となる二つの分布を提案した．

参　考　文　献

Iwase, K. and Kanefuji, K. (1994). Estimation for 3-parameter lognormal distribution with unknown shifted origin, *Statist. Hefte*, **35**, 81-90.

Kanefuji, K. and Iwase, K. (1996). Exponential inverse Gaussian distribution, *Comput. Statist.*, **11**, 315-326.

集団健康検診と個体評価

駒　澤　　勉

今後，ヒトの健康に関する問題は高齢化社会に伴い，健康管理対策が医学的にも社会的にもますます重要な時代である．健康を害してからの病院における診断・治療の守備型の個人医学もさらに充実させねばならないことは云うまでもないが，これからの健康問題を考えるとき，特に病気にさせない各種集団健康検診や高齢者のためのライフ・スタイル指導を中心にした攻撃型の集団医学が，積極的に健康の予知・予防を進めるうえで，また保健経済面から見ても重要である．これからの集団健康検診は，今までの判定法に加えて，各々の健康に関する基本的な各種検査項目の結果を暦年齢と生物学的年齢の差について的確に評価すると共に，総合的健康指標化による経年的予測の判断が必要である．この健康情報を得るには多次元的統計データ解析の方法論を利用することは云うまでもないが，病気になってからの個人医学では得られない，健常群から潜在的非健常群（検査内容から見てのいろいろなタイプの異常群）まで含んだ，被検診者の長年にわたる大量な集団検査データ，及びそれら対象についての検診後に実施した発症追跡調査の発症群と非発症群の情報分類のデータを活用することによって，健康の個人評価を実践的に行い，今後の健康管理に関する深度化のあり方について有効な方法を報告した．

官庁統計におけるリサンプリング法の活用

馬　場　康　維

1.　研究の背景

戦後 50 年を経て，我が国の社会•経済のあり様は大きく変化し，それに伴い，統計行政の見直しが必要とされている．その第一の要因は，企業経営の多角化や世帯構造の変化など，統計が把握すべき対象の変化であり，第二にプライバシー意識の顕在化や昼間在宅世帯の減少など統計調査環境の変化，そしてさらに，統計情報に対するニーズの多様化など統計そのものに対するニーズの変化である．

官庁統計における多くの統計が標本調査に基づいているが，半数を超える標本調査では推定値の精度が公表されていないのが現状である．このため，統計調査の見直しとともに標本調査による推定値の精度の公表について検討されている．この研究は，このような背景の下で，統計審議会，調査技術開発・情報処理合同部会により設置された研究会の研究の一環として始めたものである．

2.　リサンプリング法による標準誤差の推定

推定値の精度の評価は，理論的に導出された標準誤差の算出式によるのが一般的である．標本設計が，教科書的・標準的であれば，問題はない．しかし，標準的ではない複雑な標本設計の調査では，調査の設計をする度に標準誤差の算出式を導出する必要があるが，非常に複雑であったり，専門的であったり，時には，導出そのものが困難な場合がある．定期的に人事異動のある官庁統計の現場を考えると，高度な専門知識がないと導出できない理論式の導出を試みることは現実的な対応とはいいがたい．

推定値の精度の理論式による評価に代わるものとして，リサンプリングによる方法がある．この研究では，リサンプリングの方法として知られているジャックナイフ法及びブートストラップ法が，推定値の誤差評価に有効であるかどうか，妥当性，利便性，実用性等の観点から検討した．また，標準化の可能性について検討した．そのために既存のデータを用いて，リサンプリング実験を行った．実験対象としたのは

1)　国民生活基礎調査（厚生省）
2)　労働力調査（総務庁）

である．これらのデータを基にした実験の結果，リサンプリング法が十分実用的であるとの結果を得た．

参　考　文　献

馬場康維，土屋隆裕，中村好宏，山崎伸彦(1996)．ブートストラップ法による標準誤差の推定の試み，第10回日本計算機統計学会大会論文集，68-71.

一次元尺度構成のための質的項目の選択法

土　屋　隆　裕

質的な項目の中から項目選択を行い，一次元尺度を構成する方法を紹介した．数量化III類や最適尺度変換を伴った主成分分析等では，必ずしも一次元尺度構成に最適な尺度変換が行われるとは限らないうえ，項目の選択を図に頼らざるを得ないというあいまいさがある．そこで土屋（1996）では，数量化III類において，尺度得点と各項目の最適尺度変換値との相関係数を a_i という重みとして導入した方法を提案した．この方法は，一次元尺度構成に最適な尺度変換が行われる，a_i の値により項目の選択ができる，といった特徴を持つ．実際に数学試験データの項目分析に適用したところ，相関係数を用いる従来の項目選択の方法に比べ，より信頼性の高い尺度を構成できることが示された．さらに，不安を測定する心理検査の一つであるMASの項目分析についても紹介した．数量化III類では図が繁雑になるため項目の選択が不可能であったが，提案した方法では a_i に基づいて容易に項目選択を行うことができ，その結果は従来の知見とも一致するものであることが示された．

参　考　文　献

土屋隆裕（1996）．質的項目の選択による一次元尺度の構成法，応用心理学研究，No. 21，21-30.

自由回答データの分析
──いくつかの実験調査でみられた特徴，問題点──

大　隅　　昇

調査環境の変化に伴い，自由回答形式の調査法やその取得データの解析法についての方法論が求められている．また調査に限らず，各種情報の電子化に伴ってテキスト型データの蓄積が一般的となり，これについての解析法も必要となってきた．データベースやそのミドルウェア等の普及により，かなりの分析が可能となってはきたが，未だ有効な分析手法は模索の段階にある．こうした事情を背景に，昨年度に続いて自由回答文を用いた実験調査を行った．この取得データに基づき，この種のデータの分析に必要な解析法やコンピュータ・プログラム(SPAD.T, MDSA 他）の利用可能性の検証を行った．実験調査（略称：消費行動に関する調査）は一次調査と二次調査からなる．まず，一次調査では，回答者にとって卑近なジャンルを質問対象分野として設定し(百貨店の利用価値，パソコンの役割，TV の影響，シャンプーの認知度，ビールの認知度)，それぞれについて自由回答の設問を工夫した(関係固定法，結果固定法，原因固定法と言う)．設問作成の要点は，自由回答だけではなく，通常の選択肢型設問の併用，自由回答にリーディングとなる説明文を付与する等の工夫にある．こうして取得した自由回答データについて，従来のアフターコーディング処理方式との比較，統計システム SPAD.T による解析等を試みた．たとえば，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（分かち書き処理），これから導かれる出現頻度等の解析，構成要素と回答（被験者，回答者）との関係の対応分析，多重対応分析，ならびに類型化処理による分類操作等を行った．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有用な知見が得られることを検証した．さらに，通常の選択肢型設問や属性と自由回答との比較分析法の検討を進めた．また，二次調査を実施して，一次調査との対比を行った．二次調査の特徴は，二つの調査を行ったことにある．一つは，一次調査の回答者を対象としたパネル調査を行い，一次調査で得た自由回答の分析情報を用いて作成した新たな選択肢型設問を用いて，回答の出方を調べるというものである．第二は，あらたに抽出した回答者を対象に，前回とほぼ同じ内容の設問による調査を行い，回答の分布の比較を行う，というものである．今回の調査はきわめて小規模ではあるが，自由回答方式調査でのこうした試みは少なく，分析方法やデータ処理方式について，多々工夫を必要とした．とくに，データ処理上の問題として，ローマ字化規則，分かち書き方式，同音異義語の処理，品詞の区分法等についての検討も行った．加えて，SPAD.T の日本語版の開発作業も進め，そのプロトタイプを完成させた．ここでも，自由回答やテキスト型データの分かち書き処理の問題が残る（日本語には元来分かち書きの概念がない）が，これも幾つかの既存応用ソフトの流用で解決される見込みであり，実際に SPAD.T 日本語版を試用した分析では興味ある結果を得ている．

HALWIN の開発

高　木　廣　文

これまで DOS での使用が中心であったが，多くのパソコンで採用されている MS-Windows (3.1 または 95) に対応できるように，HALBAU の全面的な改定を行っている．プログラム言語としては，マイクロソフト社の Visual BASIC を用いている．

HALWIN開発の基本は，使用しやすいマンーマシン・インターフェイスの開発にある．この点に関しては，MS-Windowsに共通する操作方法になり，使用はより容易になったと考えられる．分析方法での大きな変更点は，多変量解析の多くの手法で量的変数と質的変数の混在を可能にした点である．数量化理論のように，もともとカテゴリカルデータの解析のために開発された手法もある．しかし，現実にデータ解析を行う場合，量的変数のみ，質的変数のみというのは少ない．これまで，質的変数を用いる場合，まずダミー変数にデータを変換してから，解析を行っていた．このカテゴリカルデータからダミー変数への変換を，個々の解析で簡単な指定を行うことで，分析可能になるように開発した．量的変数と質的変数の混合が可能となった方法は，重回帰分析，判別分析，正準相関分析，主成分分析，多重ロジスティックモデル，指数ワイブルモデル，条件付き多重ロジスティックモデル，比例ハザードモデルである．これらの方法では，質的変数をダミー変数群に変換するだけではなく，変数選択や各種の検定などを，ダミー変数群をセットとして解析するように変更した．今後は，操作方法および分析方法の説明のためのヘルプ機能の充実を図る予定である．

質的データを用いた森林のパターン分類について

鄭　躍軍

森林には木材等供与の生産機能の他，水源涵養や国土保全，快適環境形成，保健文化や野生動植物保護などの環境的，文化的な機能を有している．しかし，これらの多元的機能を総合的に発揮させるために，まず森林の機能を量的指標で評価しなければならない．本研究では森林のもつ各種機能を自然的，生物学的な要因に基づいて量的に評価する第一歩として森林の特質を代表する質的データの構造を分析したうえで，森林のパターン分類方法を構築することを目的とした．

これまで林齢や密度などの量的変量を利用して，森林の蓄積量または成長量を予測する研究については多数の文献が見られる．しかし地質条件や地理条件，樹種や林分起源などの自然的，生物学的要因は質的でありながら，森林の各種機能にも反応している．したがって森林の生産的，環境的並びに文化的機能をそれぞれある量的尺度で代表させるより，むしろ量的及び質的要因を利用して間接に森林を類型化して，類型ごとに各種機能を評価する指標を開発した方が現実的であろうと筆者はずっと考えてきた．この動機を駆使して，森林の各種機能はその置かれた自然的，生物学的な要因の下で発生した自然現象であると考えて，これらの機能に反応する要因データから一次元尺度を構築することによって森林を類型化する方法を探索してみた．まず林分面積，地位，林道からの距離，方位，傾斜度，林分起源，樹種，林齢，平均樹高，平均密度などの自然的，生物学的要因項目をアイテムとして選択し，各アイテムを適切なカテゴリーに区切った林分サンプルを独立で等確率に抽出すると想定する．これらのデータに外的基準のない数量化III類をかけて実際のデータの各側面を考察した結果に基づいて，さらに人工林，二次林及び天然林の3つの部分集団に分けて要因項目データの構造を解明することにした．このように部分集団ごとに森林の生物的要因及びその置かれた自然的要因に基づいて森林のパターン分類を行うことにより，森林の特質を記述する要因を縮約できる手法を導出した．起源の異なる森林の内部構造が著しく差があることを明らかにした．とくに人工林の部分集団に対して数量化III類によって一次元尺度がうまく構成されることがわかった．今後の課題としては社会的要因を加えて森林の各種機能を総合的に評価できる指標体系を開発する研究を続けてゆこうと考えている．

参 考 文 献

林知己夫 (1992). 『数量化──理論と方法』, 朝倉書店, 東京.
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III類とデータ処理』, 朝倉書店, 東京.
鄭 躍軍, 南雲秀次郎(1994). GISを利用した森林機能による類型区分, 日本林学会誌, **76**(6), 522-530.

地震活動の時空間モデル

尾 形 良 彦

一定の地域の地震活動は一つ一つの地震の余震活動の重ね合わせで表現されるというEpidemic Type Aftershock Sequence (ETAS) Modelを10年前に提案し，地震発生のデータをあてはめて地震活動計測する点過程モデルとして，その有用性を実証してきた．今回は，このETASモデルを時空間データへの拡張するための時空間モデルを考えた．

余震域の面積（の対数）と本震のマグニチュードの正の相関関係を示す経験法則として「宇津・関の公式」があるが，ここでは更に，余震発生の空間の広がりに関して，次の統計的な仮説が考えられた．それらは，

1. 余震の分布は従来から考えられているような確固とした有界（コンパクト）な領域内に限られているものか (short-range distributed ?).
2. それともそのような境界を設けられるものではなく，場合によっては，かなり遠方にも影響を及ぼすものか (long-range distributed ?).
3. 地震のマグニチュードが領域の広さに影響するものなのか，もしそうなら，どのような関数によるのか.
4. それとも余震の数密度には影響しても，領域の広さには無関係なスケールを持つものであるか.

これらの仮説に基づき時空間モデルの空間についての応答関数に関して3つの関数形を考え，これらのモデルの適合度を調べた．モデルを当てはめるデータとして気象庁の震源カタログの東北沖（プレート境界）と西日本（プレート内）の場合を考え，いくつかのマグニチュード下限を設定してデータ数を加減した．この結果，いずれの場合も共通した一定の仮説を支持する結果が得られた．それは，

1. 空間的な広がりの分布は逆べキ型 (long-range) が優れており，
2. 空間的な広がりのスケールはマグニチュードに関して指数関数としたものが優れているということである.

この結果は地震の活動度が地域的に異なる場合 (non-homogeneous) にも余震の空間分布が本震について非等方 (non-isotropic) な場合でも変わらないことが分かった．特に前者の仮定 (non-homogeneity) を考慮したものが，時間的な定常性（すなわち改良大森減衰p値が1.0以上であること）を保証し，地震活動の時空間モデルとしての実用的なものであることが分かった.

モデル比較はAICで行ったが，尤度計算の為の数値積分の正確さをうまく克服できた点が大きな成果といえる．推定されたパラメタをもとに時空間点過程を生成するシミュレーションア

ルゴリズムも与え，それらが実際の地震活動の時空間パタンに良く似ていることがわかる．しかし，一方で実際の活動パタンとシミュレーションとの微妙な違いも浮き彫りになり，これはモデルの限界でもあろうが，これが地震活動の重要な変化や前兆異常を知る手がかりとしての積極的な意味があるものと期待して研究を続行中である．

以上の結果は論文として *Annals of the Institute of Statistical Mathematics* に掲載予定である．

ベイズ的方法によるデータ解析

柏 木 宣 久

時空間季節変動調整法の開発過程で生じた問題点に対する検討結果について述べた．

予測制御研究系

情報量規準 EIC について

北 川 源 四 郎

情報量規準 AIC, TIC, GIC などは統計モデルの対数尤度を平均対数尤度の推定量とみたときのバイアス

$$b = E_G[\log f(X|\hat{\theta}(X)) - nE_G \log f(Y|\hat{\theta}(X))] \tag{1}$$

を評価し補正したものである．EIC ではこのバイアスをブートストラップを用い

$$b^* = E_{\hat{G}_n}[\log f(X^*|\hat{\theta}(X^*)) - nE_{\hat{G}_n} \log f(Y^*|\hat{\theta}(X^*))] \tag{2}$$

と評価することによって，パラメータを最尤法で推定するという仮定なしに適用範囲の広い情報量規準を定義することができる (Ishiguro et al. (1997))．本報告では Konishi and Kitagawa (1996) の方法を利用して EIC のバイアス補正項のブートストラップ分散およびバイアスを減少する方法およびその理論的根拠を示した．

1. バイアスの減少

EIC のバイアス項を $D = D_1 + D_2 + D_3$ と 3 つに分解し，$D_1 + D_3$ の項だけをブートストラップすることにより推定の分散が減少することが経験的に知られているが，最尤推定量だけでなく統計的汎関数で定義される一般の推定量に関してもこの現象が説明できることがわかった．すなわち，この場合，3 つの項の分散はそれぞれ $\mathrm{Var}\{D_1\} = n^{-1}\nu'\Sigma\nu + O(n^{-2})$, $\mathrm{Var}\{D_2\} = a/n$, $\mathrm{Var}\{D_3\} = n^{-1}\nu'\Sigma\nu + O(n^{-2})$ の形で表され，すべて n^{-1} のオーダーであるが，

$$\mathrm{Var}\{D\} = O(n^{-1}), \quad \mathrm{Var}\{D_1 + D_3\} = O(n^{-2}) \tag{3}$$

となることを示すことができる．メディアン型の推定量の場合について，数値実験の結果も報告した．

2.　2次のバイアス評価

統計的汎関数で表される一般的な推定量に対して，2次のバイアス項を解析的に求めることができる（Konishi and Kitagawa (1996)）．ただし，バイアス項は極めて複雑なので実際の計算では1次のバイアス補正を行った対数尤度をブートストラップする方が実用的であると考えられる．報告では両側指数分布に正規分布をあてはめた場合と簡単なベイズモデルの予測分布の場合について，シミュレーションにより真のバイアス，1次補正，2次補正の結果を比較し，2次補正により実際にバイアス推定の改良が可能であることを示した．

参　考　文　献

Ishiguro, M., Sakamoto, Y. and Kitagawa, G. (1997). Bootstrapping log likelihood and EIC, an extension of AIC, *Ann. Inst. Statist. Math.*, **49**(3), 411-434.

Konishi, S. and Kitagawa, G. (1996). Generalized information criteria in model selection, *Biometrika*, **83**(4), 875-890.

モデル選択とその応用

下　平　英　寿

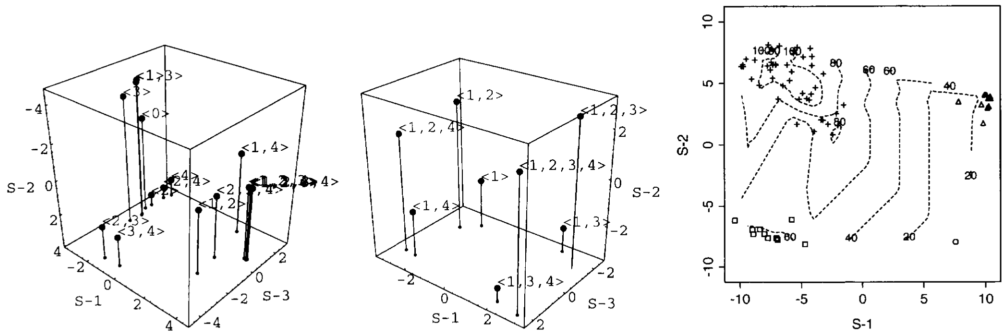
有限次元パラメタを持つ確率モデル $p_\alpha(x|\theta_\alpha)$ がいくつか与えられていて，その添字の集合を $\mathcal{M}$ で表す．候補となる $p_\alpha, \alpha\in\mathcal{M}$ のうちどれが良いのか？　という漠然とした問題を，広義の「モデル選択」と考える．古典的なアプローチは，いわゆる仮説検定で，帰無仮説 p_α を対立仮説 p_β に対して適当な有意水準で検定する．一方で AIC 最小化法などのいわゆるモデル選択のアプローチがあり，互いに入れ子でない複数のモデルに対しても適用が容易である．ところが，仮説検定も同様に広いクラスの問題に適用可能である（多重検定，Cox 検定，LM 検定，不等式制約での検定などの手法）．また，仮説検定における有意水準の恣意性が問題であると指摘されるが，検定の p-値は判断の信頼性を確率論的に表現したものであり，むしろ利点とも言える．このような意味での仮説検定の良さを，AIC 最小化法に採り入れて，モデル選択におけるモデルの信頼集合が構成される（下平（1993）など）．

AIC 最小モデル $\hat{\alpha}$，AIC の期待値を最小にするモデルを α^* とする．各 $\alpha\in\mathcal{M}$ に対して，$\mathrm{AIC}_\alpha=\mathrm{AIC}_{\alpha^*}$ であることを帰無仮説 H_α とした検定を考える．これが棄却できない α を集めてモデル信頼集合 $\mathcal{T}\subset\mathcal{M}$ とする．$\hat{\alpha}$ は α^* の点推定であり，$\mathcal{T}$ は区間推定に相当する（なお $\hat{\alpha}\in\mathcal{T}$ がなりたつ）．ところが，実際に信頼集合を求めて見ると，その解釈が問題になることがある．このために，「モデル地図」を利用する（下平（1993）など）．

モデル間の「遠さ（dissimilarity）」を定義し，多次元尺度法（MDS）を利用して各モデルを点で表した地図を描く．ここでは，AIC の第1項である最大対数尤度の分散を考慮した二種類の遠さの定義を考える．一つからは，良いモデルの塊を見つけ出すための T-map が得られ，もう一つからは，その良いモデルに潜む構造パタンを見つけ出すための σ-map が得られる．AIC の差の分散を $\sigma^{*2}_{\alpha\beta}=\mathrm{var}\{\mathrm{AIC}_\alpha-\mathrm{AIC}_\beta\}$ と書き，その推定を $\hat{\sigma}^2_{\alpha\beta}$ とする．T-map では，$T_{\alpha\beta}=(\mathrm{AIC}_\alpha-\mathrm{AIC}_\beta)/\hat{\sigma}_{\alpha\beta}$ の絶対値を遠さとし，σ-map では $\hat{\sigma}_{\alpha\beta}$ を遠さとする．

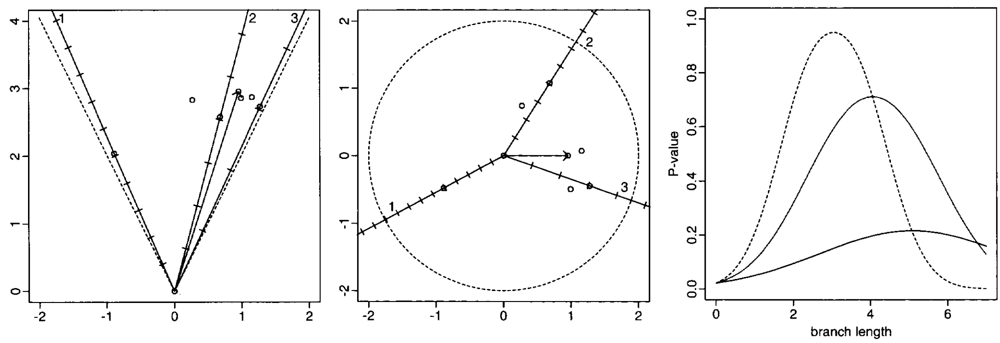
サンプルサイズを n, $\hat{\theta}_\alpha$ を MLE として，$\mathbb{R}^n$ におけるモデル α の位置を

$$s_\alpha=(2\log p_\alpha(x_t|\hat{\theta}_\alpha):\ t=1,\dots,n)$$



(a) (b) (c)

図1. (a) HALD：$\mathscr{M}$ のσ-map, (b) BABY：$\mathcal{D}$ のσ-map, (c) BOSTON：良いモデルの σ-map.



(a) (b) (c)

図2. (a) LMR side view, (b) top view, (c) 同時 p-値.

とおくと，これを，$1_n=(1,\ldots,1)$ の直交補空間に射影してから，主成分分析などで低次元に射影したものが，σ-map になる．ただし，回帰モデルや時系列などでは，$p(x_t|\theta)$ を条件付密度 $p(x_t|z_t,\theta)$ で置き換える ($z_t=\{x_{t-1},\ldots\}$ など)．$D(\cdot,\cdot)$ を Kullback Leibler 情報量とすると，適当な条件下で，$\hat{\sigma}^2_{\alpha\beta}\approx 8nD(p_\alpha(\hat{\theta}_\alpha), p_\beta(\hat{\theta}_\beta))$ である．

HALD のセメントデータ(説明変数4個，$|\mathscr{M}|=2^4=16$, $n=13$) では $\hat{\alpha}=\langle 1,2,4\rangle$ である．図1(a) では，クラスタ $\langle 1\rangle$, $\langle 3\rangle$, $\langle 1,3\rangle$ から，説明変数1と3は，回帰に関して独立な情報をあまり持っていないことがわかる．説明変数2と4も同様．また，説明変数を3つ以上持つ4個のモデルは，互いに近い．このようなパタンは，一種の多重共線性の結果と言える．他のデータセット BABY (説明変数4個，$n=15$) や，BOSTON (説明変数13個，$\mathscr{M}$ は説明変数を3個だけ使う286個のモデル，$n=506$) では，特徴的なパタンが観察される．図1(c) では，$\langle 13\rangle$ を含む66個のモデルが，$\langle 1\rangle$ と $\langle 11\rangle$ を含むかどうかによって $2^2=4$ 個のクラスタに分かれる．

ところで，系統樹のトポロジ推定は，モデル選択とみなせる．種 (のグループ) が4つの場合，unrooted tree のトポロジは，T_1, T_2, T_3 と，star トポロジ T_0 の4通り．T_0 の周りで一種の LM 法を用い，すべてのトポロジを含むモデルを構成する．これは Davidson and MacKinnon (1987) の "artificial regression" になる (図2(a) と 2(b))．必要なのは，サンプル毎のスコアベクタ：

$$\left(\frac{1}{\sqrt{n}}\frac{\partial \log p_\beta(x_t|\theta_\beta)}{\partial\theta_\beta}\bigg|_{\theta_\beta=\theta_\beta^*(\hat{\theta}_\alpha)}: t=1,\dots,n\right),$$

ただし，$\alpha=T_0, \beta\in\{T_1, T_2, T_3\}$. これから，各種の検定やモデル信頼集合，bootstrap 確率などが計算できる．一つの検定として，各枝長に関する「同時 p-値」を図2 (c) に与える．詳しくは Shimodaira (1997) を参照．

σ-map は $\log p$ に関係した大域的な情報を使っている．一方 artificial regression は $\partial \log p$ に関係した局所的な情報を利用している．じつは，これらの二つを統合することは，それほど難しくない．

参 考 文 献

Davidson, R. and MacKinnon, J. G. (1987). Implicit alternatives and the local power of test statistics, *Econometrica*, **55**, 1305–1329.

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. (1997). Testing multiple nonnested hypotheses using score statistics and its application to phylogenetic inference, Research Memo., No. 648, The Institute of Statistical Mathematics, Tokyo.

この他，論文やプレプリントについては，ホームページ http://www.ism.ac.jp/~shimo/ を参照するか，shimo@ism.ac.jp まで EMAIL を下さい．

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．さらにその後，backstepping の手法を用いて，一般の既知の相対次数の場合に，同様のロバスト適応制御系を構成する手法を導出した．

今年度は以上の研究で得られた知見をもとにして，相対次数が不確定な場合のロバスト適応制御系の構成法について研究を行い，さらにその一部を高調波利得の符号が未知な場合にも適用可能なように拡張した．また backstepping 法に代わる適応系の安定化手法についても考察を進め，簡単な例で有用性を確認した．

まず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．次にこれらの手法に適応的な内部モデルの生成機能を加えて，適応サーボ系を構成する手法も求めた．さらに Nussbaum ゲインを導入して，相対次数が3次の範囲で不確定

な場合に，高調波利得の符号にも依存しない非線形適応安定化制御系の構成法も求めた．今後，同手法のモデル規範形適応制御系や適応サーボ系への応用，あるいは一般的な次数や非線形成分の不確定性に対処するロバスト適応制御系への応用を進めていく予定である．

これらと平行して，backsteppingに代わる適応系の安定化手法についても考察を進め，ハイゲインオブザーバを用いた構造の簡単な適応制御系が構成できることを示した．その適応制御器のロバスト性や次数・相対次数の不確定性に対する特性についても部分的な結果を得た．今後，この新しい適応系についてもより一般的な解析を行う予定である．

以上の研究をもとにして，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することが出来ると考えている．

参考文献

Hanba, S. and Miyasato, Y. (1996). Model reference adaptive control of bilinear systems using Volterra series expansion, *Proceedings of the 35th IEEE Conference on Decision and Control*, **4**, 4673-4678.

半場 滋，宮里義彦，川邊武俊，磯邉 修(1997)．セミアクティブサスペンションシステムに対するオブザーバの構成，計測自動制御学会論文集，**33**(1)，66-68.

宮里義彦 (1996)． Backstepping法によるロバスト適応制御，計測と制御，**35**(6)，430-436.

Miyasato, Y. (1996). Adaptive stabilizing controller for nonlinear systems with unknown degrees and uncertain relative degrees, The International Symposium on the Mathematical Theory of Networks and Systems 1996 (MTNS96), St. Louis, USA.

Miyasato, Y. (1997). Adaptive servo controller for systems with unknown degrees and uncertain relative degrees, *Proceedings of 1997 American Control Conference*, **3**, 1662-1666.

Miyasato, Y. (1997). Model reference adaptive control for systems with uncertain relative degrees, *Proceedings of 1997 European Control Conference*, **2**.

Miyasato, Y. and Hanba, S. (1997). Adaptive control for nonlinear systems with unknown degrees and uncertain relative degrees, Preprints of the 11th IFAC Symposium on System Identification, **2**, 943-948.

同時転換自己回帰モデルにおける予測

佐 藤 整 尚

同時転換自己回帰モデルは経済データ等に応用可能な非線形時系列モデルである．今回はこのモデルの予測について考察をした．まず始めに，1期先予測を条件つき期待値で与えてやると，ARモデルに比べてそれほど良くはならなかった．しかしながら，上がったか下がったかの情報で条件をつけた期待値で予測をすると，ARモデルに比べてかなり良くなることが分かった．

線形計画問題のセンターパスと内点法

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．内点法には，大きく分けてパス追跡法とポテンシャル減少法があるが，パス追跡法は，センターパスを追跡することにより，最適解に収束する点列を生成する方法である．

線形計画問題のセンターパスは，問題のデータと初期点に依存して定まる．センターパスは，初期点の近くから出発し，最適解に到達するパスであり，パス追跡法では，センターパスを数値的に近似する．したがって，パス追跡法のパフォーマンスは，このパスの性質に大きく依存する．

Megiddo et al. (1996) では，このセンターパスがほとんど直線的な部分と曲がっている部分から成るという性質を利用した内点法を提案し，必要とされる計算量の上限値を求めた．また，Mizuno et al. (1996) では，このセンターパスの曲がった部分の最大数に関する性質を求めた．

Stoer et al. (1997) では，センターパスを高次の曲線により近似するアルゴリズムを提案した．そして，線形計画問題，線形相補性問題などに適用した場合に，そのアルゴリズムが局所的に高次の超一次収束することを明らかにした．

参 考 文 献

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

Megiddo, N., Mizuno, S. and Tsuchiya, T. (1996). A modified layered-step interior-point algorithm for linear programming, IBM Research Report 10028, San Jose, California.

Mizuno, S., Megiddo, N. and Tsuchiya, T. (1996). A linear programming instance with many crossover events, *J. Complexity*, **12**, 474-478.

Stoer, J., Wechs, M. and Mizuno, S. (1997). High order infeasible-interior-point methods for solving sufficient linear complementarity problems, Research Memo., No. 634, The Institute of Statistical Mathematics, Tokyo.

半正定値計画法について

土 谷 隆

$X, S, C, A_i\,(i=1,\ldots,m)$ を $n\times n$ 対称行列とする．次の最適化問題を半正定値計画問題といい，制御や組合せ的最適化問題への応用が期待されている（小島 (1996)，小原 (1993)）．

$$(1)\qquad \begin{cases} \min \sum_i b_i y_i\,, \\ \text{s.t.} \sum_i A_i y_i - C = S, \quad S \geq 0\,. \end{cases}$$

ここで，$S\geq 0$ は，S が半正定値対称行列であることを示す(以下同様に定義する)．また，$\bullet$ は $n\times n$ 行列を n^2 次元ベクトル空間と見た時の内積である．この問題を解くことは，次の方程式を解くことと等価であることが知られている．

$$(2)\qquad \begin{array}{l} XS=0\,, \quad A_i \bullet X = b_i\,, \quad i=1,\ldots,m\,, \\ \sum_i A_i y_i - C = S\,, \quad X\geq 0\,, \quad S\geq 0\,. \end{array}$$

(この方程式の解の (S, y) 成分が(1)の最適解である) 方程式(2)を解くため，次のような，$\nu>0$ をパラメータとする方程式を考える．

$$
(3)\qquad \begin{aligned} &XS=\nu I, \quad A_i\bullet X=b_i, \quad i=1,\dots,m, \\ &\sum_i A_i y_i - C = S, \quad X\geq 0, \quad S\geq 0. \end{aligned}
$$

この解を(X_ν, S_ν, y_ν)と記す．この解集合を中心曲線といい，$\nu\to 0$で(X_ν, S_ν, y_ν)は(2)の解に収束する．この中心曲線を追跡して(2)を解くためのホモトピー法においては，νを固定した時の(3)の解を求めるためにどのような方向を用いるかによって，いくつかの方法が提案されている．ここでは，Monteiro and Tsuchiya (1996) に基づいて，(3)と等価な方程式

$$
\begin{aligned} &X^{1/2}SX^{1/2}=\nu I, \quad A_i\bullet X=b_i, \quad i=1,\dots,m, \\ &\sum_i A_i y_i - C = S, \quad X\geq 0, \quad S\geq 0 \end{aligned}
$$

に対するNewton方向が線形方程式

$$
\begin{aligned} &USX^{1/2}+X^{1/2}SU+X^{1/2}\Delta SX^{1/2}=-X^{1/2}SX^{1/2}+\nu I, \\ &UX^{1/2}+X^{1/2}U=\Delta X, \\ &A_i\bullet\Delta X=0, \quad i=1,\dots,m, \quad \sum_i A_i\Delta y_i=\Delta S, \quad X\geq 0, \quad S\geq 0 \end{aligned}
$$

の解として与えられることを示し，この方向を用いて(2)を解くためのホモトピー法を提案した．そして，各反復においてXSの固有値の和が収束率が次元にしか依らない一次収束をすることを証明した．また，この新しい方法が，座標系のスケーリングと組み合わせることで，これまで提案されてきた方法とも関連づけられることも示した．

参 考 文 献

小島政和(1996)．半正定値計画問題と内点法，応用数理，**6**(4)，270-279．

Monteiro, R. D. C. and Tsuchiya, T. (1996). Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming, Research Memo., No. 627, The Institute of Statistical Mathematics, Tokyo.

小原敦美(1993)．システム制御理論における情報幾何，数理科学，No. 366，26-30．

平均場近似の統計学における応用

伊 庭 幸 人

平均場近似の概念は統計物理では古くから用いられてきたが，統計学の周辺で話題になったのは比較的最近である．よく知られた適用例としては，ボルツマン・マシンの学習（Peterson and Hartman (1989)），マルコフ場による画像処理（Geiger and Girosi (1991)）などがある．画像処理については，ほぼ同じ内容の近似が統計学者によってそれ以前に議論されている(Kay and Titterington (1986))．

平均場近似は組合わせ的最適化における Hopfield-Tank の方法と近い関係にあるが，Hopfield-Tank 法では local optima を避けるための方便としてシグモイド関数を導入するのに対し，平均場近似という場合には(事後)分布の周辺分布に対する近似としての側面が意味を持つ点が異なっている．この関係は，simulated annealing 法では最適化のための手段として確率分布からのサンプリングを利用するのに対し，マルコフ連鎖モンテカルロ法の統計学への応用では分布からのサンプリング自体が重視されるのに似ている．

統計学一般においても，平均場近似の概念は有効なはずであるが，意外に使われていないようである．本講演では，ベイズ的なモデル選択において，モデル空間の上の事後分布の周辺分布を近似するために平均場近似を適用することを提案した．この方法は，乱数を使わないアルゴリズムを用いて，“ある変数が選択される事後確率”が 0 と 1 の間の実数値として求まる点に特徴がある．講演では，重回帰分析における変数選択を論じ，ボストンの住宅に関するデータ (Belsley et al. (1980)) への適用結果を示した．詳しくは伊庭 (1996) を参照されたい．

参 考 文 献

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics*, Wiley, New York.

Geiger, D. and Girosi, F. (1991). Parallel and deterministic algorithms for MRF's: surface reconstruction and integration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 401-412.

伊庭幸人 (1996). 重回帰分析の変数選択におけるベイズ的方法，──平均場近似に基づくアプローチ──，第 19 回情報理論とその応用シンポジウム (SITA96) 予稿集，533-536. (http://www.ism.ac.jp/~iba/jdoc.html に原稿あり．)

Kay, J. W. and Titterington, D. M. (1986). Image labelling and statistical analysis of incomplete data, Proc. 2nd. Int. Conf. Image processing and Applications, Conf. Publ., No. 265, London Inst. Elec. Engrs., 44-48.

Peterson, C. and Hartman, E. (1989). Explorations of the mean field theory learning algorithm, *Neural Networks*, **2**, 475-494.

ミクロデータを用いた生産関数の推定

(客員) 新潟大学　経済学部　和　合　　肇

本研究の目的は，工業統計表の個票データを用いて生産曲面を推定することであり，企業別にではなく事業所別のデータを用いることによって，産業別の技術的な投入―産出関係を調べることである．このような分析は，個票データを用いることによって可能になる．

本年度は，通商産業省で収集している工業統計表の個票（ミクロ）データを用いて，産業別の技術的な投入―産出関係を調べる生産フロンティア関係を推定する場合に，従来の典型的な計量経済分析で行われていたように，コブ・ダグラス型やトランスログ型のように生産関数を特定化して，モデルのパラメータを推定するパラメトリック・アプローチではなく，データから適切な制約を与えて生産関数を導き出すベイズ的なノン・パラメトリック方法を提案し，グラフィカルに生産曲面を推定する方法を研究した．

各年度における工業統計表の膨大なデータを，今後の分析に便利なようにデータベース化し，事業所毎の多数の項目の中から，産業別の生産，労働，資本，原材料などの要素に集計した．各種の予備的計量分析を迅速に行うために分析プログラムを作成し，計算結果を加工し，グラフィカルに表示するための分析環境を整備した．本年度は試験的に，1992 年の工業統計表の精密工業のデータをとり上げて分析した．資本―労働の投入要素を 30×30 の 2 次元のメッシュに区切り，各セルでの最高の生産量を上げている事業所をそのセルの代表として考えた．次に，このデータに関数が滑らかであるとの制約を付けた，ベイズ型ノンパラメトリック密度推定の方法を応用し，生産フロンティア関数を求めて，グラフ表示する方法を開発した．これにより，従来単一の数値で分析されていた代替弾力性が，投入要素と生産量の規模毎に計算され，多面的な分析が可能になった．さらに各事業所毎の非効率性を，フロンティア関数面からの下方へ

の乖離として定義し，その分布をグラフ的に表示することによって，生産規模によって非効率性を表す分布が異なることが分かるようになる．しかしながらいくつかの分析方法上の問題があり，今後これらの問題点を修正した方法を用いて上記の分析を継続する予定である．

統計教育としての配列数学とJ言語

(客員) 千葉工業大学　西　川　利　男

大学にせよ，高校にせよ統計教育の第一歩は記述統計であり，表タイプのデータのタテ，ヨコの和，平均，標準偏差などの計算から始まる．それと併行してこれらのデータをグラフ化し，視覚的にとらえることも大切である．当然，これらの計算はコンピュータを用いて行われるが，最近ではLotus 1-2-3, EXCELをはじめとするスプレッドシート・ソフトウェアがよく使われる．

一方,統計解析の各種手法には線形代数と行列の計算が不可欠であることは言うまでもない．しかしながら，統計教育としての立場からは，線形代数において定義された概念である行列も当然含めるものの，もっと融通性をもった配列をベースにした配列数学なるとらえ方が必要であると思う．

配列の具体的イメージは，実用的にはタテ，ヨコのセルとして表わされるLotus 1-2-3などのワークシートであるが，数学の概念として考えたときはやはり充分とはいいがたい．

配列処理に強力なプログラミング言語としては，古くからAPLがあるが専用文字を用いる等の制約により，BASIC, FORTRANのようには一般化していない．J言語はAPLと同じく配列処理言語であるが，通常のASCII文字で行え，かつ最近のWindows, Macintosh, UNIXなどほとんどすべてのコンピュータ環境の上で実行が可能である．

J言語では配列(ランクにより規定される)なるデータ構造と関数形処理とを基本とする．すなわち，ランク -0 を単独の値として，ランク -1 をリスト(一方向の値の集合)，ランク -2 をテーブル(リストを要素とする上位のリスト集合)，ランク -3 をブック(テーブルを要素とするさらに上位のリスト集合)，……なる階層的データに対して，動詞と呼ばれる関数が作用する．つまり，例えば標準偏差をとるという関数は，ランク -1 のリストにも，ランク -2 のテーブルにもランクを指定することによりそのままの形で作用するのである．線形代数の演算である行列の積，逆行列などに対しても一行でコーディングできる．また，グラフィックスについては最近のGUIにより機能的に行われる．これらの典型的な例について紹介する．

統計データ解析センター

偽データ混合法

石 黒 真 木 夫

データに多項式回帰モデルをあてはめる場合，多項式の次数が低すぎればデータの特徴を捉えることができず，高すぎれば結果は不安定になり，ちょうどよい次数を選べばよい結果が得られる．次数選択でなく罰付き最小2乗法も同じように働き，罰が重すぎるとデータの特徴を捉えることができず，軽すぎれば結果は不安定になり，ちょうどよい重さの罰を与えればよい結果が得られる．

次数選択の問題にAIC最小化法を適用する方法はよく知られている．罰付き最小2乗法の「罰」の重みを決めるのにABIC最小化法やEIC最小化法などを用いることができる．罰付き最小2乗法の式を書き直してみると，この方法が推定したいことがらに関する別の側面からのデータを利用することによって推定精度をあげる方法になっていることが分かる．この解釈から重みの決定に利用できる第三の方法が導かれる．偽データ混合法である．

データ

$$y \sim N(X\theta, \sigma^2 I):\ n \text{ 次元実データ}$$
$$y_i \sim N(0, \sigma_q^2/v_i^2 I):\ n_i \text{ 次元偽データ } (i=1,2,\dots,M)$$

にモデル

$$MODEL(k)=\begin{cases} y \sim N(X\theta, \sigma^2 I) & (\text{実データ}) \\ y_i \sim N(L_i\theta, \sigma_p^2/v_i^2 I) & (i=1,2,\dots,k) \\ y_i \sim N(0, \sigma_q^2/v_i^2 I) & (i=k+1,\dots,M) \end{cases}$$

をあてはめる問題を考える．この問題において，MODEL (k) の対数尤度を最大化して求められるAIC (k) から k に依存しない M だけに依存する量をひいてPAICを定義する．

数値実験の結果，情報量規準の意味で良い k というものがあり，それを「PAIC最小化法」で推定できることが示された．

ゲームにおける内部モデルの生成

泰 地 真 弘 人

ジレンマゲームにおける協調の発生の研究は，これまでも盛んにおこなわれてきた．しかし，これまでの多くの研究では戦略の優劣と進化を扱うのみで，対戦中につまり相手の戦略を基に自分の戦略をたてるということを考慮にいれていなかった．こうしたダイナミックな戦略の変化は特に社会現象の研究などでは重要になる．

本研究では，互いに相手の戦略を学習してモデル化し，そのモデルに基づいて自分の戦略を構成する系のシミュレーションをおこなった．モデル化の方法としてはリカレントニューラルネットワークを用いた．ゲームとしては囚人のジレンマゲームをとりあげた．学習戦略同士の対戦では，最終的には自明なNash均衡解である裏切りあい状態に到達した．均衡に至る過程では複雑なふるまいがみられた．こうした複雑なふるまいは，相手のモデルを作成するときのモデル空間の構造の複雑さに起因することがわかった．学習させる相手のふるまいに対し，たくさんのlocal minimaとなるモデルがあるような乱れたlandscapeが現れるのである．このため系はlocal minimaをホップしながらそのたびに違ったふるまいをする．こうした「モデル空間の複雑さ」がふるまいの複雑さを産みだしていることがわかった．しかし全体的な傾向としては，相手の裏切りが増えるにつれ，ますます自分も裏切るようになる．このため徐々に裏切りの頻度が増え，最終的には安定な均衡である常に互いに裏切る状態に到達してしまう．さらに，相手が「自分のことを学習戦略だと思っている」ような場合についてもシミュレーションをおこない，同様の結果を得た．学習戦略というのは基本的に「自分自身をその内部で定義できない」ような戦略である．このため相手も自分と同じである場合，相手を完全に理解することはできず，このことが複雑なふるまいを経て自明な均衡に至るという過程がみられることがわかった．

ヌジ人名資料概観

上 田 澄 江

ヌジ文書とは遺跡から出土した粘土版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))はヌジ文書の人名による索引である．そこから契約内容は窺いしれないが，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．親族関係をもとに家系図の作成，契約書に登場する人数の推定，また派生的に得られるヌジの特徴，例えばヌジの人名の長さの分布，ポピュラーなヌジ名，契約を頻繁に繰り返す有力者の存在などの情報を抽出することができる(伊藤 他(1996))．父系社会を反映して人名が95%程の男性名で占められていることなどを考慮して，契約に関わる人数から，ヌジ文書が単に小都市ヌジに限定されない広範囲にわたる土地契約であったことが想像される．

参 考 文 献

伊藤栄明，石黒真木夫，上田澄江，牧野久美(1996)． 系図データからの古代社会人口の推定，シンポジウム「人文科学における数量的分析」，79-86.

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.

必要病床数の予測

田 村 義 保

高齢化社会を迎え，医療を含めた福祉が尚一層重要視されていることは言うまでもないことである．病院や福祉施設の適切な施設数，設置場所を定めた医療計画が必要である．性，年齢階級，疾病別に平均在院日数を推定し，必要病床数の予測を行った．

患者個人ごとの在院日数データを解析することに先立ち，年齢階級別，疾病別の在院割合の変化データを用いて解析した．データの形式はいわゆるサバイバル曲線の形になっており，在院1週目のものが100%おり，2週目，3週目と退院するものが現れるごとにその割合が減って行くことになる．ハザードを扱っても良いが，この解析ではサバイバル曲線そのものの解析を行った．

退院という事象は勿論患者個人の病気の種類・程度・回復力と治療効果によって決るのであろうが，統計が本来用いる，個々は決定論に従っていても，集団的な振る舞いはランダム(確率的)であるという考えで解析している．確率過程としては点過程を考える．従って，サバイバル曲線にあてはめるべき分布は，指数分布やワイブル分布ということになる．指数分布を考えることは，退院確率が在院日数によらないという仮定をしていることにあたる．一見不合理に見えるかもしれないが，多種多様な人の集まりを相手にしているのだから，全体としては，このようになっている可能性もある．ワイブル分布を考えることは退院確率が在院日数によるということを仮定している．ある意味では，合理的な仮定であるとも思えるが，サバイバル曲線であるから，ワイブルと短絡思考しているようにも思える．ここでは，さらに，二つの指数分布の和というのも考えてみた．患者集団が平均在院日数の長いグループと短いグループの2

つのグループからなっていると仮定しているのである．勿論，3グループ以上に拡張することも計算的には容易である．

詳細な解析結果は論文として発表したい．疾病，年齢によってワイブルが良い場合も2つの指数分布の和が良い場合もあるということのみ報告した．

時系列の欠測値処理のアルゴリズムについて

荒 畑 恵 美 子

時系列の一部分が観測できないことがある．観測できなかったデータのことを欠測値という．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．観測値が欠測のときには，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．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．例として，多変量2次元ARモデルのシミュレーションを行いそのデータを用いて，欠測値の処理や予測を行った．欠測値がないデータにかなり良くあっていた．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．

統計基礎研究系

成分間に依存性のある consecutive systems の生存時間分布

平 野 勝 臣

システムの n 個の成分のそれぞれの生存時間 $\xi_1, \xi_2, \dots, \xi_n$ は確率空間 $(\Omega, \mathcal{F}, P)$ で定義された確率変数とする．T をシステムの生存時間とする．T は $\xi_1, \xi_2, \dots, \xi_n$ の関数である．T に次の二つの仮定をおく．(1) 任意の $\omega \in \Omega$ に対して $T(\omega) = \xi_{j(\omega)}(\omega)$ を満たす $j = j(\omega)$ が存在する．(2) $j(\omega)$ は $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega)$ の順位の関数である．成分を $\{1, 2, \dots, n\}$ とかき，$\{1, 2, \dots, n\}$ の置換群を $\mathcal{S}_n$ とかく．いくつかの成分の故障がシステム故障を引き起こすとき，その成分の集合を cutset という．cutset の置換を cutsequence という．cutsequence $(a_1, a_2, \dots, a_k)$ はその subsequence $(a_1, a_2, \dots, a_{k-1})$ が cutsequence でないとき minimal であるという．すべての $\pi = (\pi(1), \dots, \pi(n)) \in \mathcal{S}_n$ に対して，$(\pi(1), \dots, \pi(j))$ が minimal cutsequence となる正整数 j が一意的に存在する．この j を π の minimal cutnumber といい，$m(\pi)$ で表す．

定理． 成分 $(\xi_1, \dots, \xi_n)$ の生存時間分布が同時密度 $g(s_1, \dots, s_n)$ を持つならば，システムの生存時間分布の密度 $f(s)$ は次で与えられる．

$$f(s) = \sum_{\pi \in \mathcal{S}_n} \left\{ \int_{s_{\pi(m(\pi))}}^{\infty} ds_{\pi(m(\pi)+1)} \cdots \int_{s_{\pi(n-2)}}^{\infty} ds_{\pi(n-1)} \int_{s_{\pi(n-1)}}^{\infty} ds_{\pi(n)} \right.$$
$$\left. \times \int_0^{s_{\pi(m(\pi))}} ds_{\pi(m(\pi)-1)} \cdots \int_0^{s_{\pi(3)}} ds_{\pi(2)} \int_0^{s_{\pi(2)}} ds_{\pi(1)} g(s_1, \dots, s_n) \right\} \Bigg|_{s_{\pi(m(\pi))} = s}$$

この定理を用い consecutive-k-out-of-n: F システムに二つの種類の依存性を仮定したモデルの生存時間を与える．はじめに Kamps (1995) の sequential order statistics に基づく依存性がある場合，n 個の成分の生存時間の同時密度を求め，システムの生存時間の密度を得る．例え

ば，$i=1,2,3, F_i(x)=1-e^{-\mu_i x}$ として，consecutive-2-out-of-3 : F システムの密度は次で与えられる．

$$\begin{aligned}f(s)=&(2\mu_2\mu_1(3e^{\mu_3 s+2\mu_2 s}\mu_3^2-6e^{\mu_3 s+2\mu_2 s}\mu_3\mu_2-6e^{\mu_3 s+2\mu_2 s}\mu_3\mu_1\\&+12e^{\mu_3 s+2\mu_2 s}\mu_2\mu_1-3e^{\mu_3 s+3\mu_1 s}\mu_3^2+4e^{\mu_3 s+3\mu_1 s}\mu_3\mu_2+9e^{\mu_3 s+3\mu_1 s}\mu_3\mu_1\\&-12e^{\mu_3 s+3\mu_1 s}\mu_2\mu_1+2e^{2\mu_2 s+3\mu_1 s}\mu_3\mu_2-3e^{2\mu_2 s+3\mu_1 s}\mu_3\mu_1))\\&/(e^{\mu_3 s+2\mu_2 s+3\mu_1 s}(2\mu_3^2\mu_2-3\mu_3^2\mu_1-4\mu_3\mu_2^2+9\mu_3\mu_1^2+12\mu_2^2\mu_1-18\mu_2\mu_1^2))\,.\end{aligned}$$

次にもう一つの依存性を成分間にいれた場合，即ち各成分の瞬間故障率はその左成分が機能しているとき μ, 故障したとき λ としよう．このとき $(\xi_1,\ldots,\xi_n)$ の同時密度を求め，システムの生存時間分布の密度を導くことができる．例えば consecutive-2-out-of-3 : F システムの生存時間分布の密度は

$$f(s)=\frac{\mu(2(\lambda-\mu)e^{(\mu+\lambda)s}+3\mu e^{\lambda s}-\lambda e^{3\mu s}-\lambda e^{2\mu s}-\mu e^{2\mu s})}{(\lambda-2\mu)\exp((3\mu+\lambda)s)}$$

で与えられる．

なお，本報告は Aki and Hirano (1997) に基づいている．

参　考　文　献

Aki, S. and Hirano, K. (1997). Lifetime distributions of consecutive-k-out-of-n : F systems, *Nonlinear Analysis, Theory, Methods & Applications*, **30**(1), 555-562.

Kamps, U. (1995). A concept of generalized order statistics, *J. Statist. Plann. Inference*, **48**, 1-23.

滑らかな凸錐を対立仮説とする尤度比検定の分布について

栗　木　　哲

Z を p 次元正規分布モデル $N_p(\mu, I_p)$ からの観測値とし，K を R^p の閉凸錐とする．検定 H_0: $\mu=0$ vs. H_1: $\mu\in K$ は「傾向のある対立仮説問題」として応用統計の各分野でしばしば現れる問題である．尤度比検定統計量は $\bar{\chi}^2=\|Z_K\|^2$ (Z_K は Z の K への直交射影) で与えられ，その H_0 の下での分布は $\bar{\chi}^2$ 分布と呼ばれる χ^2 分布の混合分布である．ところで，閉凸錐 K が有限個の線形不等式で定義される錐 (polyhedral cone) の場合の $\bar{\chi}^2$ 分布の混合確率 (重み) の幾何学的意味あいは明確であり，重みを K の各面 (face) で定義される内角・外角で表現する公式が知られている．しかし non-polyhedral cone の場合は二，三の具体例についてその重みが求められているのにすぎなかった．

本研究では，積分幾何学における Steiner の公式(Schneider (1993))を拡張することにより，一般の閉凸錐 K についてその重みが境界 ∂K の主曲率を用いて陽に表されることを示した．これにより，Kuriki (1993) が導いた $p\times p$ 非負定値行列のなす閉凸錐の重みの幾何学的意味付けを与えることができた．なお，本研究で用いた手法は，Weyl (1939)のいわゆる tube 公式と密接に関連している．

参　考　文　献

Kuriki, S. (1993). One-sided test for the equality of two covariance matrices, *Ann. Statist.*, **21**, 1379-1384.

Schneider, R. (1993). *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge University Press, Cambridge.
Takemura, A. and Kuriki, S. (1997). Weights of $\bar{\chi}^2$ distribution for smooth or piecewise smooth cone alternatives, *Ann. Statist.* (to appear)
Weyl, H. (1939). *On the volume of tubes*, *Amer. J. Math.*, **61**, 461-472.

K-L 情報量による分布の近似の再考

松縄　規

K-L 情報量による分布の近似について，近似されるもの，近似するものの方向性を明確にして考察した．このことに関し，K-L 情報量は Boltzmann や Kullback が与えた解釈に倣って，基礎確率モデルに対する相対ネゲントロピーと捉える必要性を指摘した．これによって統計の理論が関連諸科学の理論と整合することも注意した．例として，K-L 情報量に基づく 1) 中心極限定理，2) Shrödinger 方程式の誘導を行った．

1) について，従来の関連する主な仕事として，Linnik (1959) が truncation 及び独立な正規確率変数で smoothing をした確率変数列の標準化和についての中心極限定理を Shannon および Fisher の情報量を用いて行ったが，これは厳密には通常の中心極限定理とは異なるもので，歴史的な意義の方が強いことが注意されるべきであることを指摘した．また，Barron (1986) による K-L 情報量による強い意味での中心極限定理の証明について触れ，それが Shannon 及び Fisher の情報量に関する Stam 等の仕事に基づく定性的な結果であるが，Linnik の場合と同様に，情報量の定義が不自然であることを注意した．即ち，彼らは基礎確率モデルを正規分布に取ってしまったため，K-L 情報量が本来持っている変化の方向性に従っておらず，熱力学第2法則等の物理法則に整合しない欠点を有することになる．この近似の方向性を認識した上で，改善の方策を示した．

2) について，科学の基礎方程式の一つである Shrödinger 方程式はその誘導の仕方が不明確に思われる．Shrödinger (1926) の第1論文で与えられた方程式に対して，K-L 情報量を用いて行うことを試みた．これにより，Shrödinger が誘導の過程で行った，波動関数の対数を取ったことの意味がある程度説明可能となった．

参考文献

Barron, A. R. (1986). Entropy and the central limit theorem, *Ann. Prob.*, **14**, 336-342.
Linnik, Yu. V. (1959). An information-theoretic proof of the central limit theorem with Lindeberg condition, *Theory Probab. Appl.*, **4**, 288-299.
Schrödinger, E. (1926). Quantisierung als Eigenwertproblem, *Ann. Physik*, **81**, 109-139.

正則変動する裾をもつ分布の積の意味での分解

志村隆彰

$\boldsymbol{R}^1$ 上の分布 μ は任意の $k>0$ に対して $\lim_{x\to\infty}\mu(kx,\infty)/\mu(x,\infty)=k^{-\alpha}$ が成り立つとき指数 $\alpha\geq 0$ の regularly varying tail をもつといわれる（$\mu\in\boldsymbol{D}(\alpha)$ と表す）．ふたつの非負値独立確率

変数 X, Y の分布がともに regularly varying tail をもてば，積 XY の分布は $\boldsymbol{D}(\alpha)$ に属するが，問題としたいのはその逆，すなわち，積 XY の分布が $\boldsymbol{D}(\alpha)$ に属する分布であるとき，因子 X, Y の分布はどのような性質のものかということである．X の分布を μ, Y の分布を ν とするとき，XY の分布を $\mu\circ\nu$ と書く．$\mu\in\boldsymbol{D}(\alpha)$ で，ある $\varepsilon>0$ に対して，$\int_0^\infty t^{\alpha+\varepsilon}\nu(dt)<\infty$ ならば，$\mu\circ\nu\in\boldsymbol{D}(\alpha)$ であり，

$$\lim_{x\to\infty}\mu\circ\nu(x,\infty)/\mu(x,\infty)=\int_0^\infty t^\alpha\nu(dt)$$

が成り立つことが知られている（志村(1993)）．この逆，すなわち，$\mu\circ\nu\in\boldsymbol{D}(\alpha)$ とある $\varepsilon>0$ に対して $\int_0^\infty t^{\alpha+\varepsilon}\nu(dt)<\infty$ を仮定したとき，$\mu\in\boldsymbol{D}(\alpha)$ がいえるかについて以下の結果を得た．$\alpha=0$（緩慢変動する tail をもつ場合）に対しては肯定的な答が得られた．しかしながら，$\alpha>0$ の場合は必ずしもそのようなことは成り立たない．ν は2点を台とする離散確率分布 $(\nu(\{t_i\})=\nu_i\ (i=1,2))$ であるとしよう．このとき，$\mu\circ\nu\in\boldsymbol{D}(\alpha)$ かつ $t_1^\alpha\nu_1\neq t_2^\alpha\nu_2$ ならば，$\mu\in\boldsymbol{D}(\alpha)$ となる．逆に，$t_1^\alpha\nu_1=t_2^\alpha\nu_2$ ならば，$\mu\notin\boldsymbol{D}(\alpha)$ で，$\mu\circ\nu\in\boldsymbol{D}(\alpha)$ となるようなものが存在することが示された．証明には緩慢変動する truncated moment をもつ分布に対する同種の命題の証明に使われた単調緩慢変動関数の分解の理論に加え（志村（1995)，Shimura (1997)），指数が0でない正則変動関数に対する新しいタイプの分解についての結果が手法として用いられる（志村（1997)）．

参 考 文 献

志村隆彰（1993）． Mellin-Stieltjes convolution of distributions characterized by regular variation，統計数理研究所共同研究リポート，No. 51，6-12.

志村隆彰(1995)． 緩慢変動する truncated moment をもつ確率変数の独立積とその応用，統計数理研究所共同研究リポート，No. 75，1-9.

志村隆彰（1997）． Decomposition of distributions with regularly varying tails in the sense of Mellin-Stieltjes convolution，統計数理研究所共同研究リポート，No. 102，54-61.

Shimura, T. (1997). The product of independent random variables with slowly varying truncated moments, *J. Austral. Math. Soc. Ser. A*, **62**, 186-197.

領域統計研究系

相互作用，交互作用，加法モデル

佐 藤 俊 哉

要因Bと要因Cの相互作用とは，両方に曝露されることによってそれぞれ単独のリスクをあわせた場合よりも疾病発生リスクが増加すること，と定義できる．しかし，「それぞれ単独のリスクをあわせた場合」は，ロジスティックモデルなどの乗法モデルの場合は，

要因Bのリスク×要因Cのリスク

となるし，加法モデルの場合は，

要因Bのリスク＋要因Cのリスク

となるから，どちらの統計モデルを用いるかによって「それぞれ単独のリスクをあわせた場合よりも疾病発生リスクが増加する」という判断がことなってしまう．この問題は統計的交互作用の問題であり，交互作用の定義は用いる統計モデルに依存して変わるので，因果的な相互作

用の定義とは区別しなければならない．

要因BとCに同時に曝露した場合，要因Cにのみ曝露した場合，要因Bにのみ曝露した場合，どちらにも曝露していない場合，それぞれの疾病発生割合を$(P_{11}, P_{10}, P_{01}, P_{00})$とする．Rothman (1986) はcomponent causeモデルを用いて，相互作用がない場合，

$$P_{11}-P_{00}=(P_{10}-P_{00})+(P_{01}-P_{00})$$

とリスク差が加法モデルにしたがうことをしめした．本報告では，簡単な因果モデルを用いて，相互作用がない場合でも明らかに，

$$P_{11}-P_{00}\leq(P_{10}-P_{00})+(P_{01}-P_{00})$$

であることを明らかにした．したがって，データから超加法性が示せれば，相互作用があると結論できるが，相互作用がある場合でも加法モデルのあてはまりがよい場合がありうることになる．

参考文献

Rothman, K. J. (1986). *Modern Epidemiology*, Little, Brown, and Co., Boston.

擬似尤度の構成について

汪　金芳

推定関数の考え方は，ガウスによる最小自乗法とフィッシャーの最尤法を融合させた新しい統計的推測法である．特に擬似スコアと呼ばれる推定関数は，一般化線形模型の枠組みを本質的に拡張し，計量生物学の分野ではすでに標準的なデータ解析法として確立されている．しかし，推定関数に基づく推論では，多くの問題点や困難も含み，基礎的な理論を整備していかなければならないのが現状である．

本研究では，推定関数による推測理論における多くの問題点の中から，最も基本的と思われる，推定関数の不可積分問題にテーマを絞って研究を進めてきた．推定関数の不可積分問題とは，推定関数から構成されるベクトル場が勾配場にならない，という問題である．すなわち，ベクトル場としての推定関数に対応するスカラー•ポテンシャルが存在しないことを意味する．フィッシャーリアンの尤度推測の世界では，尤度がこのポテンシャルの役割を果たしている．ポテンシャルが存在しないことから，今までの多くの成熟した推測理論の取り入れが困難となる．たとえば，適合度統計量の欠けていることや，母数に関する事前情報の取り入れの難しさなど，様々な困難に直面する．また，推定方程式に重根が存在するとき，“正しい”ものを見分けることも問題である．

本研究では，不可積分問題をベクトル解析の観点からの考察を行ってきた．3次元のユークリッド空間の中でのあらゆるベクトル場は，適当なポテンシャル場とダイバージェンス•フリーなベクトル場によって分解される．この事実はヘルムホルツの分解定理として知られている．統計的推測の場合には，一般の$p\geq 2$次元まで考える必要があるため，本研究では，まずヘルムホルツの分解定理を任意の次元まで拡張した(Wang (1997))．ヘルムホルツ型のポテンシャルは調和関数の不定性を持ち，また陽に表現することが一般に困難である．Wang (1997) では，推定関数を帰無仮説の近傍で線形近似を行い，二次のヘルムホルツ型のポテンシャルを提案し

ている．また Wang (1997) は，擬似尤度比統計量を構成し，投票者の遷移確率の推定問題と測定誤差が存在する場合のロジスチック回帰問題へ適用し，二次のヘルムホルツ型のポテンシャルの妥当性を確認した．

参 考 文 献

Wang, J. (1997). Nonconservative estimating functions and approximate quasilikelihoods, Research Memo., No. 629, The Institute of Statistical Mathematics, Tokyo.

群，部分群関係と結晶群の出現頻度

伊 藤 栄 明

無機結晶構造データベース ICSD (Inorganic Crystal Structure Database) を使って，無機結晶物の種を定義し，幾何学的対称性の分布を調べるという課題にとりくんできた．ICSD における登録の順番にもとづいて種を逐次定義してゆく方法を考案し，それらの種の空間群，点群についての出現頻度を ANX 記号ごとにもとめた (藤原 他 (1994))．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．群，部分群の関係にもとづき，群のうえのランダムウォークを考えることにより，えられた点群の分布の説明をこころみた．最密充填構造が基本になっていると考えられる場合，情報量規準をもちいるとモデルがデータをよく説明していると思われるが，その他の場合は，必ずしもそうではない．まだ分類の方法が不十分であるためであると思われる．ANX 記号だけでなくさらに構造に立ち入った分類をする必要がある．モデルのシミュレーション法および情報量規準をもちいたデータの解析法についての計算機プログラムを整理し，共同研究リポートにまとめた．

参 考 文 献

藤原美也子，伊藤栄明，松本崧生，武田 弘 (1994). 無機結晶データベース (ICSD) を用いた結晶群の出現頻度II，統計数理研究所共同研究リポート，No. 66.

伊藤栄明，杉本真貴子 (1997). 結晶群のうえのランダムウォークと結晶群の出現頻度，統計数理研究所共同研究リポート，No. 98.

高校・大学における統計教育の現状と問題点

村 上 征 勝

調査データ等に基づき，高校と大学のそれぞれにおける統計教育の現状と問題点について報告した．

高校の統計教育に大きな影響を与えているのは，文部省が定める高等学校指導要領と大学入試の二つであると考えられる．そこで学習指導要領に関しては，昭和26年の第一回の改訂から平成元年の第六回の改訂まで，確率・統計に関する項目の扱いがどのように変わってきたか，

その変遷をまとめ，特に最も新しい平成元年の改訂で，確率は必修科目の中に含まれているが，確率分布，統計処理が選択科目の中に含まれてしまったことの問題点を指摘した．そしてこのことが，高校の統計教育に影響を与える二番目の要因である大学入試にも影響し，大学入試において統計に関する問題の出題率の低さとなって現れ，それがまた高校の授業に大きく影響していることを，高校における数学科目の履修調査のデータでもって示した．

次に大学での統計教育に関しては

1. 大学で統計教育の基盤となる組織（学科・学部）が存在しないという問題点
2. 一般教育における統計学の開講数の少なさと，大学設置基準の改訂による教養部改組の悪影響
3. 企業側の統計教育に対する期待の高さと少ない専門教員数の問題
4. 情報処理教育重視の流れの中における統計学の重要性に対する認識の低さ

等の問題点について報告した．

世論調査データにおけるマジック・ナンバー

吉　野　諒　三

統計数理研究所が収集してきたパネル調査データの分析を進め，各質問項目毎に回答分布％の変化から線形近似で推測される「一時収束点」を計算し，回答反応の分布比がある簡単な整数比（1：1，1：2，1：3，2：3 やそれらの鏡映）になる場合が多い傾向について考察を進めてきた．理論的には，ある数学的条件や近似のもとで Master 方程式と呼ばれるかなり一般的な統計的数式より Schrödinger 方程式が導かれる事を確認し，これに基づいて「波動尺度」と称する計量心理的尺度モデルを構成した．考察を進める過程で，非線形力学的現象に広く見られる「均衡点付近での一種の共鳴現象」の相互関連を推察した．

これまでの成果の概要は，意識のパネル社会調査データに関しては

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）が，
2. 総体としての回答変動率（周辺分布の波動）は，少ない（数％程度）．
3. 回答カテゴリーが２つの場合（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，Schrödinger 方程式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ある．特に，解のエネルギーと安定性を論ずる意味がありそうである事等などである．

本年度は，さらに，回答カテゴリーが３以上の場合の尺度モデルとして，階層二進構造を考察し，現実のデータと照合検討を進めた．また，一般の社会調査データ分析への応用可能性を追求するために，国政選挙の「投票率」の時系的な推移を考察した．

参　考　文　献

Yoshino, R. (1997). A social quantum theory for the analysis of public opinion survey data, Research Memo., No. 636, The Institute of Statistical Mathematics, Tokyo.

米国の大学教員募集要項の分析──社会学の場合

釜野さおり

本研究では,「社会学」という分野が,その中でどの専門領域に正当性を与え,学問としてどう構築されているのかを,アメリカ社会学会が発行する就職情報誌に掲載された教員募集要項の分析を通して探究する.

Employment Bulletin の1994年5月から1995年8月号に掲載された,1995年秋採用のアメリカ国内の社会学/社会学関連の学部におけるアカデミック職計569のポジションを分析に使った.それぞれのポジションを,(a) ランク(1年,tenure-track, tenured),(b) 機関のタイプ(コミュニティカレッジ,4年制大学,社会学のマスタープログラムのある機関,社会学のドクタープログラムのある機関),(c) 領域指定の形(リストされた領域の数,主副の分類など),(d) 指定された専門領域についてコーディングし,募集ポジション全体の組織的な特徴ならびに各専門領域の需要を記述した.

全体の40%が4年制大学のポジション,同じく40%がPh.D.プログラムのある機関のポジションであった.ランク別に見ると,64%がtenure-trackのポジションで10%がtenuredのポジションであった.最も頻繁に指定された領域は「犯罪学」と「人種-民族」(20%以上のポジションで指定)であった.「組織」と「医療社会学」を指定した40%以上は,ドクタープログラムのある機関のポジションであった.また,「ジェンダー」,「家族」,「都市社会学」の領域を指定したポジションの中でtenuredのものは5%以下だが,「不平等」,「医療社会学」,「組織」の領域を指定したtenuredのポジションは10%以上あった.

このデータは,統計的研究の材料として,いろいろな「可能性」に富んでいる.今後は,クラスター分析や数量化などの手法を適用して,専門領域がどう構造しているのか,さらに,1974年からの時系列データを使って,領域の需要,領域の組み合わせ,組織的特徴と専門領域の関連などの時間的変化をたどる分析を行なう計画である.

統計教育・情報センター

翻訳関連蛋白種全アミノ酸配列データのアライメント

橋本哲男

1996年8月,古細菌であるメタン産生菌 *Methanococcus jannaschii* のゲノムDNAの全塩基配列が発表された.インフルエンザ菌,マイコプラズマ菌等の真正細菌および酵母(真核生物)については,既にゲノムプロジェクトが完了していたために,これにより,三大生物界(真正細菌,古細菌,真核生物)のいずれに属する生物についても,それらの生命活動に必要最小限の遺伝子のセットが明らかになったものとみなせる.このような状況下において,進化的な視点からゲノム全体をさまざまな生物種間で比較することは,分子進化学のみならず,生物学のあらゆる分野の研究にとっても重要な課題となるが,その基礎となるのは複数配列アライメントである.そこで我々は,生命維持にとって必須と考えられる「ハウスキーピング遺伝子」に注目し,これら全てのアミノ酸配列アライメントを構築することを目的として研究を開始した.

本年度は翻訳関連蛋白種を中心に転写関連蛋白種も含めて解析を行なった.まず,*M. jannaschii* の全対象蛋白種について,各アミノ酸配列データをキーとする相同性解析(fasta)(Pearson and Lipman (1988))を実施した.その結果をもとに,相同種であると考えられる全てのデー

タをコンパイルし，それらのアライメントを構築した．第一段階のアライメントには，Hidden Markov Model (HMM) に基づく複数配列アライメント法（Hughey and Krogh (1996)）を利用した．その結果，さらに修正が必要と考えられた場合には，ペアワイズの相似度行列やペアワイズの最尤法アライメント（Thorne and Churchill (1995)）の結果に基づいて，マニュアルで修正を加えた．

三大生物界に共通に存在するような蛋白種の場合，各生物界内部でのアライメントは容易であったが，全生物界を通してのアライメントはほとんどの場合困難であり，高度に保存された部分以外はアライメント不可能という例が続出した．一部のアミノアシル tRNA 合成酵素を除く多くの転写・翻訳関連種では，*M. jannaschii* を含む古細菌の配列は，アライメントパターンおよび無根系統樹の解析結果からみる限り，真正細菌よりは真核生物のものに近縁であるように思われた．この点は，三大生物界の分岐以前に重複したと考えられる遺伝子群であるイソロイシン/ヴァリン tRNA 合成酵素，および，ペプチド鎖伸長因子 EF-1α・Tu/2・G の複合系統樹解析によっても支持された．

参考文献

Hughey, R. and Krogh, A. (1996). Hidden Markov models for sequence analysis: extension and analysis of the basic method, *Computer Applications in the Biosciences*, **12**, 95-107.

Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison, *Proc. Nat. Acad. Sci. U.S.A.*, **85**, 2444-2448.

Thorne, J. L. and Churchill, G. A. (1995). Estimation and reliability of molecular sequence alignments, *Biometrics*, **51**, 100-110.

Distribution Theory in a Higher-order Markov Chain

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．

まず，確率変数 $X_{-m+1}, X_{-m+2}, \ldots, X_0, X_1, X_2, \ldots$ を 0 か 1 のいずれかの値をとる time-homogeneous な m-th-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 $l(=m, m+1, \ldots, k)$ の1の連の回数の同時分布と長さ $l(=1, 2, \ldots, m-1)$ の1の連の回数の同時分布を求めた．このことによって higher-order のマルコフ連鎖の下でのオーダー k の幾何分布も求めることができる．さらにオーダー k の幾何分布の characterization として長さ k の1の連が起こるまでに起こる長さ $l(=m, m+1, \ldots, k)$ の1の連の回数の分布はある種のオーダー k の幾何分布になっていることもわかった．連の数え方はちょうど長さ l の連の数える数え方，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

Hirano, K., Aki, S. and Uchida, M. (1996). Distributions of numbers of success-runs until the first

consecutive k successes in higher-order Markov dependent trials, Research Memo., No. 603, The Institute of Statistical Mathematics, Tokyo.
Uchida, M. (1996a). 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.
Uchida, M. (1996b). Joint distributions of numbers of success-runs until the first consecutive k successes in a higher-order two-state Markov chain, Research Memo., No. 619, The Institute of Statistical Mathematics, Tokyo.

非ガウス外乱に対する線形系の応答

岡 崎 卓

1. 外乱を受ける系の確率密度と指数分解型一般化 Fokker-Planck 方程式

外乱 W を受けつつ，確率微分方程式

$$\frac{d}{dt}U=M(U)+\mu(W) \tag{1}$$

に従って発展する変数 U の確率密度 $f(u,t)$ を定める指数分解型一般化 Fokker-Planck e-GFP 方程式は，射影子法と指数演算子の分解近似によって導かれ，

$$\begin{aligned}\frac{\partial}{\partial t}f(U,t)=&-D_UM(U)f(U,t)\\&+D_U\int_0^t dse^{-siL_U}D_U\Big[\int dV\Phi_A(V,s)f(U-V,t-s)\\&+D_U\int dV_1\int dV_2\Phi_B(V_1,s)\Phi_C(V_2,s)f(U-V_1-V_2,t-s)\Big]\end{aligned}$$

の形をもつ ($e^{-siL_U}=e^{-sD_UM(U)}$, $D_U=\partial/\partial U$). ここに拡散項の核をなす3種の関数（「拡散核」と呼ぶ）は，作用する外乱に応じて決まる統計量であり，系の構造に依存する関数 $A(\cdot)$ および $B(\cdot)$ を介して次のように表される；

$$\begin{aligned}&\Phi_A(V,s)=\langle\mu(0)\delta(V-A(s)\mu(s/2))\mu(s)\rangle\\&\Phi_B(V,s)=\langle\mu(0)\mu(s/2)\delta(V-B(s)\mu(s/2))\rangle\cdot 2B(s)\\&\Phi_C(V,s)=\langle\delta(V-B(s)\mu(s/2))\mu(s)\rangle\\&\mu(s)=\mu(W(t))\end{aligned}$$

通常の Fokker-Planck 方程式では，外乱の情報が単なる拡散係数として考慮されるに過ぎず，外乱の個性を反映しない．これに対し，e-GFP 方程式の拡散核 Φ_A, Φ_B, Φ_C は，V を引数とするデルタ関数の存在により，時刻 $s/2$ の外乱 $W(s/2)$ に関しては確率密度に類似の構造をもち，それだけ外乱の統計的性質を詳細に捉えている．従って e-GFP 方程式は，外乱の非ガウス性を忠実に反映するものと期待される．以下では，この方程式を線形系 $M(U)=-\alpha U$ に適用し，外乱の特性が系確率密度を左右するメカニズムについて考察する．

2. 外乱と拡散核

有色ガウス外乱に拡散核 $\Phi_A\sim\Phi_C$ は $W(0)$ と $W(t)$ の結合確率分布 $P(W_0, W|t)$，あるいは相似変換 $\boldsymbol{W}(t)\equiv e^{ti\hat{L}_W}We^{-ti\hat{L}_W}$, $i\hat{L}_W=-\beta W(\partial/\partial W)$ により計算することができ，3種の拡散核はいずれも空間座標 V について対称性をもつ．

一方，非ガウス外乱の典型例として，船舶に及ぼす波力のモデルとしてもよく使われるRayleigh過程

$$\frac{d}{dt}W=-\beta W+\frac{\gamma}{W}+\sigma\nu(t), \qquad \gamma=\frac{\sigma^2}{2}$$

を採り，結合確率密度 $P(W_0, W|t)$ によって拡散核を計算すると，非対称で且つ一方に長い裾をもつことが判る．

3. 系確率密度の非ガウス性が発現する機構

非ガウス外乱を受ける線形系の確率密度 $f(U,t)$ が非ガウス性を呈することは当然であるが，如何なる機構によってその非ガウス性が現れるかについては，これまで不明であった．*e*-GFP方程式によれば，その形は有色ガウス外乱に対してもRayleigh外乱に対しても同型であるが，拡散核は外乱によって異なり，Rayleigh外乱に対する拡散核は大きな歪みをもつ．従ってこの歪みが *e*-GFP方程式拡散項のconvolution積分を通じて，確率密度 f のガウス型からの乖離をもたらすものと結論される．

超多項変動を持つデータの分析

（客員）大分大学　越　智　義　道

多値離散反応における共変量効果の評価に関するモデルの適合性の研究を行った．このような反応を多項反応とみなして，多項分布にもとづいて多項ロジスティックモデル等の適合を行うと，飽和モデルの下でも十分な適合が得られない場合がある．この場合に適合度診断を行うと，いくつかの外れ値や影響度の高い点があるためにモデルの適合がうまくいかない場合と特定の影響の高い点が存在するわけではないが全般的に大きな残差が生じる場合がある（越智 他(1994)）．催奇形性試験等で得られるデータは多くの場合後者のケースであり，このようなデータのもつ変動を超多項変動と呼んでいる．

超多項変動を示すデータに対して，ここではディリクレ-多項分布の最尤法にもとづく分析法とこの分布の平均と分散構造に着目した一般化推定方程式による分析法について検討した（椺木 他 (1996)）．ディリクレ-多項分布では多項分布の平均，分散構造と類似した形態の表現が可能であり，また超多項変動は実験単位ごとの対象者の相関パラメータを用いて表現することもできる．ここでは平均構造については多項ロジスティック回帰モデルや累積ロジスティック回帰モデルなどのモデル化を用いた共変量効果の評価を想定し，超過変動についても相関パラメータに関するロジスティック回帰による共変量効果を想定したモデルを考えた．一般化推定方程式では分散構造の推定に関して，Zhu et al. (1994) の提案したピアソン残差にもとづく推定方程式の構成法を拡張して用いた．

本報告では，これらの方法を紹介し，Chen et al. (1991) の催奇形性試験に関する文献データへの適用例を紹介するとともに，適合されたモデルの診断についても考察を加えた．

参　考　文　献

Chen, J. J., Kodell, R. L., Howe, R. B. and Gaylor, D. W. (1991). Analysis of trinomial responses from reproductive and developmental toxicity experiments, *Biometrics*, **47**, 1049-1058.

越智義道，岸野清広，上杉博紀，小畑経史 (1994). 多項反応における回帰分析と診断法，計算機統計学，

7, 111-125.

椛木浩世，越智義道(1996)． 一般化推定方程式による超過変動多項反応データの解析，大分大学工学部研究報告，第34号，49-54.

Zhu, Y., Krewski, D. and Ross, W. H. (1994). Dose-response models for correlated multinomial data from developmental toxicity studies, *Applied Statistics*, **43**, 583-598.