

平成 9 年度研究報告会要旨

と き：1998 年 3 月 17 日 午前 10 時～午後 4 時 45 分
18 日 午前 10 時～午後 5 時 30 分
ところ：統計数理研究所 講堂

プ ロ グ ラ ム

3 月 17 日 (火)

午前 (10 時～12 時 30 分)

あいさつ

所長 清水 良一

【統計基礎研究系】

連の分布論

平野 勝臣

曲面への縮小推定とその応用

栗木 哲

予測分布を利用した分類について

駒木 文保

観測のバイアスについて

江口 真透

非増加正則変動関数の分解

志村 隆彰

オービタル分解の下での複合変換モデルの特徴付け

紙屋 英彦

標識再捕魚の不完全な報告に関するある問題について (客員, 東京水産大学) 山田作太郎

微小パラメータを持つガンマ確率の評価

松縄 規

【領域統計研究系】

一次元複核白金錯体における分子内振動に伴う

構造の乱れと構造相転移

(客員, 兵庫県立姫路工業大学) 鳥海幸四郎

グラフと確率モデル

伊藤 栄明

午後 (13 時 30 分～16 時 45 分)

「単収/人口」からみた穀物生産動向

柳本 武美

Modelling Higher-order Inter-correlated Data

汪 金芳

大学における統計教育の実態調査

村上 征勝

公理的測定理論について

吉野 諒三

家庭生活と仕事に関する調査結果の分析

釜野さおり

【統計計算開発センター】

半無限計画法

伊藤 聡

物理乱数について

泰地真弘人

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

重回帰分析と情報量規準

石黒真木夫

【統計科学情報センター】

種内の遺伝的多型と種間の変異
 記述的多変量解析法と信頼性：官庁データの分析
 日本人の児童・生徒の体型の変化について
 一般化 Fokker-Planck 方程式の簡約と補正

長谷川政美
 馬場 康維
 金藤 浩司
 岡崎 卓

3月18日(水)

午前(10時~12時15分)

【予測制御研究系】

線形方程式の新解法
 情報量規準 GIC について
 季節調整の適切さについて
 地球電流系構造の自動同定
 情報量規準による重み付き最尤法の最適化
 複雑な系の非線形時系列解析
 不完全情報下における制御系設計に関する研究
 資産価格のトレンドとボラティリティの関係
 近似計算を使った実用的な内点法

田辺 國士
 北川源四郎
 川崎 能典
 樋口 知之
 下平 英寿
 尾崎 統
 宮里 義彦
 佐藤 整尚
 水野 眞治

午後(13時15分~17時30分)

ヌジ人名資料による家系図の復元について
 Non-Empty Cell (NEC) の個数の分布について
 凸2次制約凸2次計画に対する主双対内点法について
 シミュレーションから分布の構造を知る—Finite mixtureを用いたアプローチ—
 あるクラスの半正定値計画 (SDP) 問題に関する考察

上田 澄江
 鈴木義一郎
 土谷 隆
 伊庭 幸人
 (客員, 大阪大学) 小原 敦美

【調査実験解析研究系】

国民性調査に関する基礎的検討(2)
 95年SSM調査の標本設計と標本精度
 階層帰属意識と生活満足感の分析
 MCMC法に基づくランダムネットワークの推論
 パソコンUNIXによる統計学研究環境
 Statistical Distribution Theory in Markov Chains
 健診情報の可視化について
 いくつかの調査に参加して
 電子調査法について
 WINHELPによる統計学情報ナビゲーションシステム
 事象発生時刻の不確定な点過程の推論
 平滑化に関する考察

坂元 慶行
 中村 隆
 前田 忠彦
 種村 正美
 丸山 直昌
 内田 雅之
 駒澤 勉
 土屋 隆裕
 大隅 昇
 高木 廣文
 尾形 良彦
 柏木 宣久

統計基礎研究系

連の分布論

平野 勝 臣

本年度の研究

1. m 次 $\{0,1\}$ -値マルコフ系列において、長さ k の 1 の連が初めて起こるまでに、長さ l の 1 の連が起こる回数の分布についての結果をまとめた (Hirano et al. (1997)).
2. 2次元 $\{0,1\}^2$ -値マルコフ系列において sooner and later waiting time の問題についての結果をまとめた (Aki and Hirano (1997)).

ここでは Hirano et al. (1997) に基づいて述べる。

長さ k の 1 の連が初めて起こるまでに、長さ l の 1 の連が起こる回数の分布

オーダー k の幾何分布を $G_k(p)$ と表わし、 $G_k(p, a)$ を台 $\{a, a+1, \dots\}$ 上のオーダー k の幾何分布とする。また $m \leq k$ とし、確率変数の系列 $X_{-m+1}, X_{-m+2}, \dots, X_0, X_1, X_2, \dots$ は $\{0,1\}$ -値 m 次マルコフチェーンに従うとし、 $p_{x_1, \dots, x_m} = P(X_i=1 | X_{i-m}=x_1, \dots, X_{i-1}=x_m) = 1 - q_{x_1, \dots, x_m}$ for $i \geq 1$ とする。このとき X_1, X_2, \dots において長さ k の 1 の連が初めて起こるまでに、長さ l の 1 の連が起こる回数について次の結果を得た。

はじめに $m \leq l < k$ とする。ただし $p_{1, \dots, 1} = p$ とおいた。

1. overlapping count とした場合は $G_{k-l}(p, k-l+1)$ に従う。
2. non-overlapping count とした場合の pgf は

$$\frac{p^{k-l} t^\nu}{1-t(1-p^l)\{1-(p^l t)^{\nu-1}\}/(1-p^l t) - p^{(\nu-1)l}(1-p^m) t^\nu}$$

で与えられる。ここに $k = \nu l + \mu$, ν と μ は非負整数で $0 \leq \mu < l$ 。

3. 丁度長さ l の場合は $G_1(p^{k-l}/(p^{k-l}+1-p), 0)$ に従う。
4. 長さ l 以上の場合は $G_1(p^{k-l})$ に従う。

つぎに $l < m$ のとき、前のように美しくはないが求めることはできる。例えば overlapping count とした場合の pgf を求める。 $N_m = \{0, 1, 2, \dots, 2^m - 1\}$ とおく。 N_m から N_m への関数 $f_i (i=0, 1)$ を

$$f_i(x) = 2x + i \pmod{2^m}, \quad \text{for } i=0,1$$

とし、また長さ l の 1 の連 (overlapping) を数えるため関数 g_l を

$$g_l(j) = \begin{cases} j-l+1 & \text{if } j \geq l \\ 0 & \text{if } j < l \end{cases}$$

とする。 p や q の添字の 2 進数を 10 進数とみる。 $x \in N_m$ に対して $\phi^{(x)}$ を、初期状態が x であるとき、長さ k の 1 の連がはじめて起こるまでに、長さ l の 1 の連 (overlap) の起こる回数の分布の条件付き pgf とする。最初の 0 が起こることで排反に分ける。各 $x \in N_m$ に対して

$$\begin{aligned} \phi^{(x)} = & q_x \phi^{(f_0(x))} + p_x q_{f_1(x)} \phi^{(f_0 \circ f_1(x))} t^{g_1(1)} \\ & + p_x p_{f_1(x)} q_{f_1^2(x)} \phi^{(f_0 \circ f_1^2(x))} t^{g_1(2)} \\ & + \cdots + p_x p_{f_1(x)} \cdots p_{f_1^{k-2}(x)} q_{f_1^{k-1}(x)} \phi^{(f_0 \circ f_1^{k-1}(x))} t^{g_1(k-1)} \\ & + p_x p_{f_1(x)} \cdots p_{f_1^{k-1}(x)} t^{g_1(k)}, \end{aligned}$$

となる。この 2^m 本の線形方程式を解けばよい。Hirano et al. (1997) に解説的な例を与えてある。

参 考 文 献

- Aki, S. and Hirano, K. (1997). Sooner and later waiting time problems for runs in Markov dependent bivariate trials, Research Memo., No. 641, 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 (Research Memo., No. 603, The Institute of Statistical Mathematics, Tokyo).

曲面への縮小推定とその応用

栗 木 哲

$x \in R^p$ を平均 μ , 分散 I (単位行列) の p 次元正規変量とする。平均ベクトル μ を 2 乗損失で推定する問題において, x 自身はミニマクス推定量ではあるが, 次元 p が 3 以上 ($p \geq 3$) の場合には James-Stein 推定量によってそのリスクが一律に改善されることはよく知られている。James-Stein 推定量の一般化として, Bock (1982) は x を凸多面体 K への直交射影方向に縮小することによりそのリスクを改良するミニマクス推定量を構成した。

本研究の主要な結果は, Bock 推定量の一般化として, x を (多面体とは限らない) 凸集合 K への直交射影方向に縮小するミニマクス推定量を新たに構成したことである。ここでその縮小係数は, x の K への射影の足における曲面 ∂K の埋め込み曲率の関数として決められる。提案する推定量は, その曲率により定義された非整数次元 \hat{p} が 2 より大きい ($\hat{p} > 2$) とき \hat{p} に応じた縮小率で x を縮小する推定量と解釈できる。また, 凸集合 K にハウスドルフ距離の意味で収束する多面体の列 K_1, K_2, \dots が存在するときには, 提案する推定量は各多面体に対応して定義される Bock 推定量の列の収束先と解釈することができる。

本研究の応用として非線形回帰に基づく平滑化を考えることができる。いま事前情報として非線形回帰モデル $y_i = f(x_i; \theta)$ ($\theta \in \Theta$ は未知母数) が想定できるとき, 観測値 $y = (y_1, \dots, y_n) \in R^n$ をモデルのなす多様体

$$F = \{(f(x_1; \theta), \dots, f(x_n; \theta)) \mid \theta \in \Theta\}$$

の直交射影方向に縮小することにより観測値の平滑化が可能となる。ただし一般に F は凸ではないので若干の工夫が必要である。

参 考 文 献

- Bock, M. E. (1982). Employing vague inequality information in the estimation of normal mean vectors

(estimators that shrink to closed convex polyhedra), *Statistical Decision Theory and Related Topics-III*, Vol. 1 (eds. S. S. Gupta and J. O. Berger), 169-193, Academic Press, New York.

Kuriki, S. and Takemura, A. (1997). James-Stein type estimator by shrinkage to closed convex set with smooth boundary, Discussion Paper 97-F-22, Faculty of Economics, University of Tokyo.

予測分布を利用した分類について

駒 木 文 保

与えられたランダムな特徴ベクトルからもとのデータがどのクラスに属しているかを推定する分類の問題で、未知パラメータに事前分布を仮定すると、事後分布を利用して分類したほうが、推定値をそのまま代入した分布を利用するより良いことがいえる。この研究ではモデルに属さない分布を利用する予測分布に基づく手法が、最尤推定量等の推定量を未知パラメータのかわりに置き換えた分布をそのまま使う方法より、事前分布の選択によらず漸近的に優れていることを示した。また、特徴ベクトルのみからなるクラスの間から学習用のデータが与えられた場合には、パラメータ推定のためにEMアルゴリズムがひろく利用されている、しかし、必ずしも尤度の高い値をあたえるパラメータの値が良い推定量になるとは限らず、明らかに望ましくない尤度の高い状態に落ち込んでしまうことがあることが知られている。この研究で、拡張したベイズ型の予測分布を導入し、Mステップのかわりにこの予測分布を利用するPステップを採用することにより明らかに望ましくない尤度の高い状態に落ち込む現象を回避できることを示した。

非増加正則変動関数の分解

志 村 隆 彰

指数が負 ($-\alpha < 0$) の非増加正則変動関数 F を次のような形の非増加関数の和で表現することを考える。

$$F(x) = \sum_{k=0}^{n-1} b_k f(a^k x).$$

ここで、 f は非増加関数、 $0 < a < 1$, $b_k > 0$.

F がこのように表されたとしたとき、 f が正則変動性をもつかということを考える。そのことを判定するために、次の方程式の解の絶対値に注目する。

$$\sum_{k=0}^{n-1} a^{-k} b_k z^k = 0.$$

この式の解の絶対値が全て1未満か、全て1を超えるならば、 f の正則変動性が導かれる。一方、絶対値1の解が存在すれば、正則変動しない非増加関数 g で

$$F(x) \sim \sum_{k=0}^{n-1} b_k g(a^k x)$$

を満たすものが構成できる (ただし、 F によっては、等号が成立するような f が存在しない)。このことは正則変動する裾をもつ分布の分解問題に応用できる。台が非負実数に含まれるふたつの分布 μ と ν の確率変数で独立積の分布にあたるものを μ と ν の Mellin-Stieltjes convolution と呼ぶ ($\mu \circ \nu$ で表す) が、 $\mu \circ \nu$ の裾が指数負の正則変動性を持ち、 ν の台が等比数列をな

す場合の μ の裾の漸近挙動について関数の結果に対応するものが得られる。

オービタル分解の下での複合変換モデルの特徴付け

紙屋 英彦

標本空間 \mathcal{X} に、ある群 G が作用しているとする： $(g, x) \mapsto gx$ 。このとき、 \mathcal{X} 上の確率測度の全体 \mathcal{P} への G の作用が、 $(g, P) \mapsto gP, (gP)(\cdot) := P(g^{-1}\cdot)$ により自然に誘導される。 $\mathcal{P}_0 \subset \mathcal{P}$ がこの作用の下で不変 ($g \in G, P \in \mathcal{P}_0 \Rightarrow gP \in \mathcal{P}_0$) なとき、 \mathcal{P}_0 は不変確率モデル (Eaton (1983)), 或いは複合変換モデル (Barndorff-Nielsen (1988), Murray and Rice (1993)) と呼ばれる。

Θ をパラメタ空間とし、 $\theta \in \Theta$ によりパラメトライズされた密度の族 $\mathcal{D}_0 = \{p(\cdot|\theta) | \theta \in \Theta\}$ を考える。ただし密度は χ を multiplier とする相対不変測度 $\lambda(\lambda(dgx) = \chi(g)\lambda(dx))$ に関するものとする。このとき、 \mathcal{D}_0 に対応する確率測度の族 $\mathcal{P}_0 = \{P_\theta | \theta \in \Theta\}, dP_\theta = p(\cdot|\theta)d\lambda$ が不変確率モデルとなる条件を考える。

これに関して次のことが知られている。(A): $p(x|\theta) = p(gx|g\theta)\chi(g)$ が成り立つとき、密度族 $\mathcal{D}_0 = \{p(\cdot|\theta) | \theta \in \Theta\}$ は χ -不変であると言われる。 \mathcal{D}_0 が χ -不変なとき、それに対応する \mathcal{P}_0 は不変確率モデルとなる。またこのとき、 $gP_\theta = P_{g\theta}$ である (Eaton (1983, 1989))。

ここでは $p(x|\theta)$ が (A) の性質を持つための必要十分条件を、 $p(x|\theta)$ の関数形を指定する形で与える。

\mathcal{X} 及び Θ はオービタル分解可能とする： $x \leftrightarrow (g_x H, Gx), \theta \leftrightarrow (g_\theta K, G\theta)$ 。ただし H, K はそれぞれ、ある $x_0 \in \mathcal{X}, \theta_0 \in \Theta$ における固定部分群 $H = G_{x_0}, K = G_{\theta_0}$ である。また K はコンパクトであると仮定する。

このとき $p(x|\theta)$ が (A) を満たす、即ち密度族 $\mathcal{D}_0 = \{p(\cdot|\theta) | \theta \in \Theta\}$ が χ -不変となるための必要十分条件は、ある関数 $q(\cdot, \cdot, \cdot)$ に対して $p(x|\theta)$ が $p(x|\theta) = \chi(g_\theta)^{-1} q(K(g_\theta^{-1}g_x)H, Gx, G\theta)$ と書けることである。ここで g_x, g_θ はそれぞれ、オービタル分解において x, θ に対応する $g_x H, g_\theta K$ の任意の元である。

証明は、 \mathcal{X}, Θ が等質空間 ($\mathcal{X} = G/H, \Theta = G/K$) の場合の対角作用 $(g, (g'H, g''K)) \mapsto (g(g'H), g(g''K))$ に関する結果 (例えば寺田・原田 (1997), 補題 1.84) を利用する。

報告会では例として、Eaton (1989), Example 3.4 にある多変量正規分布に関連した作用、及び Murray and Rice (1993), Example 3.4.9 にある von Mises-Fisher 分布に関連した作用を取り上げた。

参 考 文 献

- Barndorff-Nielsen, O. E. (1988). Parametric Statistical Models and Likelihood, *Lecture Notes in Statist.*, 50, Springer, New York.
- Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*, Wiley, New York.
- Eaton, M. L. (1989). *Group Invariance Applications in Statistics*, Regional Conference Series in Probability and Statistics, Vol. 1, Institute of Mathematical Statistics and American Statistical Association, Hayward and Alexandria.
- Murray, M. K. and Rice, J. W. (1993). *Differential Geometry and Statistics*, Chapman & Hall, London.
- 寺田至・原田耕一郎 (1997). 『群論 (岩波講座 現代数学の基礎)』, 岩波書店, 東京.

標識再捕魚の不完全な報告に関するある問題について

(客員) 東京水産大学 山田 作太郎

M 尾の魚に標識をつけて放ちその後の捕獲のなかの標識魚尾数から資源尾数 N を推定する方法は、Petersen 法としてよく知られている。これには標識の脱落が無いとか、放流から捕獲までに魚の死亡、逸散は無いとか、標識再捕魚の試験当局への報告は完全であるとかの条件がついている。Paulik (1961) は標識再捕魚の報告が不完全の時の報告率と資源尾数を推定するために捕獲を二つのグループに分け、第一のグループでの標識再捕魚の報告は完全であるが第二のグループからの標識再捕魚の報告率は未知であるとの場面を設定した。グループ i ($i=1, 2$) での捕獲尾数を n_i 、その中の標識魚尾数を X_i 、 X_i のうちの報告数を R_i とする。 n_i も本来確率変数であるがここではその実現値で条件づけて考える。 $R_1 = X_1$ とし、グループ 2 の各捕獲標識魚が報告される確率は ρ とする。 R_i ($i=1, 2$) の観測から ρ と N (もしくは $\tau = M/N =$ 母標識率) についての推測を考える。Paulik は R_1, R_2 は独立とし、

$$R_1 \sim Po(n_1 \tau), \quad R_2 \sim Po(n_2 \tau \rho)$$

なるモデルを仮定し、 $R = R_1 + R_2$ が与えられた時の R_1 の条件付分布

$$P(R_1 = r_1 | R = r) = {}_r C_{r_1} \left(\frac{n_1}{n_1 + n_2 \rho} \right)^{r_1} \left(\frac{n_2 \rho}{n_1 + n_2 \rho} \right)^{r - r_1}$$

から

$$\frac{r_1}{r} = \frac{n_1}{n_1 + n_2 \rho}, \quad \hat{\rho} = \frac{r_2 / n_2}{r_1 / n_1}$$

を得た。Seber (1982) はこの時 X_2 を $r_2 / \hat{\rho}$ で推定し、二つのグループのデータを合わせて Petersen 法から $\hat{N} = n_1 M / x_1$ と推定した。問題はなぜ R で条件づけて ρ を推定するのかと、 N の推定に完全な報告のあったグループ 1 だけのデータを用い、不完全な報告ではあってもどうしてグループ 2 のデータを N の推定に用いないのかである。本報告ではこれらの問題に対して局外母数が存在する時の補助性、充分性の立場から議論した。

参 考 文 献

- Barndorff-Nielsen, O. (1976). Nonformation, *Biometrika*, **63**, 567-571.
 Godambe V. P. (1980). On sufficiency and ancillarity in the presence of a nuisance parameter, *Biometrika*, **67**, 155-162.
 Paulik, G. J. (1961). Detection of incomplete reporting of tags, *Journal of the Fisheries Research Board of Canada*, **18**, 817-829.
 Seber, G. A. F. (1982). *The Estimation of Animal Abundance*, Charles Griffin, London.

微小パラメータを持つガンマ確率の評価

松 縄 規

ガンマ分布は連続型の基本的な確率分布として馴染み深い。その確率分布関数の標準的表現

は次の不完全ガンマ関数比

$$\gamma(p, x) = \frac{1}{\Gamma(p)} \int_0^x y^{p-1} e^{-y} dy, \quad (x > 0, p > 0)$$

で与えられる。この確率の近似計算は種々の方法で行われて来た。しかし解析的に定量的な評価を与えたものは意外と見かけない。特にパラメータ p が $0 < p < 1$ の場合には、筆者の知る限り、これはという不等式による評価は未だ与えられていないように思われる。そこで本報告では上記の場合に $\gamma(p, x)$ を上下からできるだけ精密に近似することを試みた。誘導の核心は分母が主要変数の1次式である分数式を逆階乗展開で表現すること、それを用いて関連する積分を評価することにある。結果として、 $0 < p < 1, x > 0, n \geq 2$ の時次の不等式を得た：

$$\begin{aligned} \gamma(p, x) &< 1 - \frac{x^p e^{-x}}{\Gamma(p)} \left[\frac{1}{x} + \sum_{i=1}^{n-1} (-1)^i \frac{\alpha_i(p)}{x(x+1)(x+2)\cdots(x+i)} - \frac{\Gamma(n)}{x(x+1)(x+2)\cdots(x+n-1)} \right], \\ \gamma(p, x) &> 1 - \frac{x^p e^{-x}}{\Gamma(p)} \left[\frac{1}{x} + \sum_{i=1}^{n-1} (-1)^i \frac{\alpha_i(p)}{x(x+1)(x+2)\cdots(x+i)} + \frac{\Gamma(n)}{x(x+1)(x+2)\cdots(x+n-1)} \right], \end{aligned}$$

ここで、

$$\alpha_i(p) := \sum_{j=1}^i (-1)^{i-j} d_{i-j, i-1} (j-p)(j-1-p)\cdots(1-p), \quad (0 < p < 1; i=1, 2, \dots),$$

と与えられる。ただし $d_{i-j, i-1}$ は第1種スターリング数の絶対値を表す。

上記の不等式の応用として、次の関係

$$N(x) := \int_0^x \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv = \frac{1}{2} \gamma\left(\frac{1}{2}, \frac{x^2}{2}\right)$$

を利用し、標準正規分布の裾の確率を上下から精密に評価できることを示した。

領域統計研究系

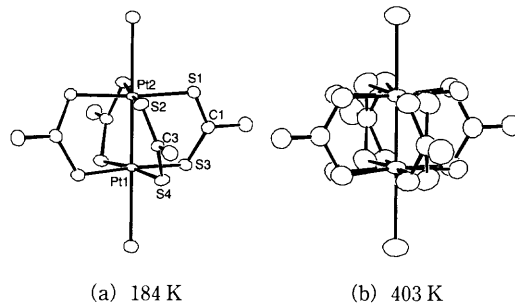
一次元複核白金錯体における分子内振動に伴う構造の乱れと構造相転移

(客員) 兵庫県立姫路工業大学 鳥海幸四郎

結晶は、原子・イオン・分子が引力と斥力で密に集合し、3次元的な周期構造を形成したものである。このとき、原子・イオン・分子の集合の様子(すなわち結晶構造)は、結晶を構成する原子・イオン・分子の相対的な半径比、球形からのずれ、相互作用の異方性、などによって変化する。例えば、熱振動の小さな低温では、構成粒子の形状や相互作用の異方性の影響が顕著に現れ、低対称な結晶構造となる。一方、熱振動が大きな高温では、異方的な影響が熱振動によって隠され、高対称な結晶構造となる。実際、下記の一次元白金錯体の相転移においては、熱振動に伴う分子構造の乱れに伴い、高温相では結晶構造の対称性の上昇が見られる。

伊藤らは、ランダムウォークモデルを無機結晶に対して適用し、結晶の属する点群の分布がほぼ統計的に予想される分布と一致することを示した。しかし、無機結晶を組成で分類してそれぞれについてランダムウォークモデルを適用した場合、統計的分布から大きくずれる集団があることが分かった(Fujiwara et al. (1996))。この統計的分布からのずれは、何らかの構造科学的要因に帰着できると考えられ、どのような要因が統計的分布を乱すかは大変に興味深い問題である。この点に着目し、さらに検討を進めていく予定である。

分子性結晶の一つとしてハロゲン架橋一次元金属錯体の結晶があり、金属錯体とハロゲンが交互に積層して一次元鎖状構造をとっており、興味深い固体物性を示す。特に、ヨウ素架橋複

Fig. 1. Molecular structures of $\text{Pt}_2(\text{dta})_4\text{I}$.

核白金錯体では、室温付近で電気伝導性が半導体から金属へ転移することが報告された (Kitagawa et al. (1997)). X線結晶解析法により結晶構造の温度変化を測定したところ、 90°C 付近で一次相転移が存在し、相転移温度以下では単斜晶系で空間群 $C2/c$ (No. 15) であるが、相転移点以上では空間群 $A2/m$ (No. 12) に変化することが分かった。 $C2/c$ と $A2/m$ の関係を *International Tables for Crystallography*, Vol. A (1992) で調べると、両者はともに $2/m$ の点群に属するが、 $C2/c$ の supergroup の一つとして $C2/m$ ($A2/m$ と同じ) があり $C2/m$ の subgroup として $C2/c$ があることが分かる。このような対称性の変化がどのような構造変化と対応しているかを結晶構造から調べると、2回回転軸上に位置している複核白金錯体ユニット、 $\text{Pt}_2(\text{dta})_4$ の対称性が2から $2/m$ へ変化していることが分かる。さらに、 $\text{Pt}_2(\text{dta})_4$ ユニットの分子構造を詳細に検討すると、低温では図1(a)のようにラセン状の分子構造をとっているが、相転移温度以上では熱的にツイスト振動をしており、結晶学的には時間平均されて図1(b)のように対称性が $2/m$ へ上昇している。

しかし、物性変化を説明するにはこの対称性の変化だけでは説明できない。この点については、熱振動に伴う Pt-Pt 間距離の変化 → Pt 原子の酸化状態の変化 → 複核白金錯体ユニット間の電子移動、という電子格子相互作用の関与した価数揺動によって説明される。

参 考 文 献

- Fujiwara, M., Itoh, Y., Matsumoto, T. and Takeda, H. (1996). Statistical distribution of symmetry groups for inorganic crystal structure database (ICSD) III, ISM Cooperative Research Report, No. 83.
- Kitagawa, H., Onodera, N., Ahn, J. -S., Mitani, T., Toriumi, K. and Yamashita, M. (1997). The metallic state in a MMX-chain complex, $\text{Pt}_2(\text{dta})_4\text{I}$, *Synthetic Metals*, **86**, 1931-1932.

Modelling Higher-order Inter-correlated Data

汪 金 芳

Higher-order inter-correlated data とは、2次元の場合を例に、次のような構造をもつデータのことをいう

$$(1) \quad y_{st}, \quad s=1, \dots, S; t=1, \dots, T.$$

ここで、 s, t はそれぞれ空間、時間を表すインデックスである。もちろん、一般に、時・空間は仮想的なものでよく、データを生み出す連続体(continuum)の数は2以上でよい。本報告では、2次元の場合に限って、状態空間モデルの適用について述べる。工夫がいるのは状態遷移方程式なので、観測方程式については省略する。

最も単純なランダムウォーク・プラスノイズモデルを、コンポーネント毎に次のように決める。

$$(2) \quad \mu_{1t} = \mu_{1,t-1} + \omega_{1t}, \quad t=1, \dots, T$$

$$(3) \quad \mu_{st} = \mu_{s-1,t} + \omega_{st}, \quad s=2, \dots, S; t=1, \dots, T$$

ここで、 μ_{st} は y_{st} の平均で、システムノイズ ω_{1t}, ω_{st} は互いに無相関である。式(2)と(3)から、固定した $s \geq 2$ に対して、 μ_{st} はやはりランダムウォークになることがわかる。

ここで、**相関次元** (correlation continuum) の概念を導入し、アレイ・データ(1)を1次元的に整理し直す。 $S < T$ の場合に次のようになる

$$\begin{aligned} \mathbf{y}_1 &= (y_{1t} | t=1, \dots, l+1)' \\ \mathbf{y}_r &= \{(y_{rt} | t=1, \dots, l+r)', (y_{s,t+r} | s=1, \dots, r-1)\}' \\ \boldsymbol{\theta}_1 &= (\theta_{1t} | t=1, \dots, l+1)' \\ \boldsymbol{\theta}_r &= \{(\theta_{rt} | t=1, \dots, l+r)', (\theta_{s,t+r} | s=1, \dots, r-1)\}' \end{aligned}$$

ここで、 $\theta_{st}, \boldsymbol{\theta}_r$ は構造パラメーターで、 $l = T - S, r = 2, \dots, R (= S)$ である。そうすると、システム方程式は、次のように書ける

$$(4) \quad \boldsymbol{\theta}_r = G_r \boldsymbol{\theta}_{r-1} + \boldsymbol{\omega}_r, \quad r=1, \dots, R.$$

システム方程式(4)は通常の状態方程式と本質的に異なる点がある。相関次元や、状態ベクトル $\boldsymbol{\theta}_r$ は、(1)の相関構造をモデリングするため、人工的に作られたもので、一般には物理的な意味を持ってない。また、システム行列 G_r や、状態ベクトル $\boldsymbol{\theta}_r$ の次元は r に依存してしまう。方程式(4)と観測方程式に基づいて、Kalman filter の拡張や、平滑化のアルゴリズムを容易に構成することができる。

参 考 文 献

- Wang, J. (1998). Modelling higher-order inter-correlated data, 「確率過程の漸近分布論の研究」, 統計数理研究所共同研究レポート, No. 11, 1-6.

大学における統計教育の実態調査

村上 征勝

日本の大学に統計学科が1つも存在しないことからわかるように、大学での統計教育環境は非常に厳しい。その一方、企業は大学での統計教育に大きな期待をいただいている。このような状況の中で、大学での統計教育はいかにあるべきかを検討する資料を得るため、「大学における統計教育の実態調査」を実施した。

調査は1995年の夏から秋にかけて、全国の四年制大学550校から平成7年度の講義要項、時

間割などの資料を収集するという型で行われ、国公立大学144校(含む放送大学)の中の121校(84.0%)、私立大学406校の中の343校(84.5%)の協力を得た。収集した講義要項の中から統計学及び統計学関連科目を抜き出し、9項目(1. 大学名, 2. 学部名, 3. 学科名, 4. 講義名, 5. 一般教育科目・基礎科目・専門科目の別, 6. 単位数, 7. 対象学年, 8. 必修・選択の別, 9. 使用教科書)を含むデータベースを構築した。

このデータベースに基づき、一般教育としての統計学は私立大学では91.0%にあたる312校で、国公立大学では95.0%にあたる115校で開講されており、統計学関連科目の総数は国公立大学1543科目、私立大学2518科目で、大学一校当りの平均開講数は国公立大学で12.8科目、私立大学で7.3科目である等に関し報告した。

公理的測定理論について

吉野 諒三

この研究の源流は、18世紀のBernoulliの「富の主観的効用価値の研究」や19世紀のFechnerの精神物理学に求めることができる。数学的にはKleinのErlanger Programが哲学的基盤にあり、また20世紀初頭の論理実証主義などの影響も受けてきた。直接には、今世紀初頭のHelmholtzとHölderのextensive measurement theoryが先駆的研究である。その後、数理心理学者の数十年にわたる努力の結晶として公理的測定論の基盤が確立し、広い分野に影響を与えている。

この流れの中で、Stevens(1957)は広い分野の科学的測定や法則を吟味して、尺度を名義尺度、順序尺度、間隔尺度、比例尺度に分類した。Luce(1959)は、この分類を理論的に正当化し、一方で「理論構築の原理」から「科学的法則の可能な形」を特定し、多くの議論を巻き起こした。1971年発刊の*Foundation of Measurement Theory*において、当時までの成果がまとめられ、将来への発展の基礎が確立された。Yoshino(1989)は、Luceの研究を受けて、力学系理論の「構造安定性」という概念を用いて、「可能でかつ安定な法則」の形を特定した。とくに、精神物理学でよく知られているStevensのべき乗法則のべきの値域が限られているという実験心理学的知見に理論的正当化を与えた。

今日では、これまでのerror-freeであった公理的測定論の発展の可能性として、統計的誤差を加味した理論が模索されている。

参 考 文 献

- Krantz, D., Suppes, P., Luce, R. D, and Tversky, A. (1971). *Foundation of Measurement Theory*, Academic Press, New York.
- Luce, R. D. (1959). On the possible psychophysical laws, *The Psychological Review*, **66**(3), 81-95.
- Stevens, S. S. (1957). On the psychophysical laws, *The Psychological Review*, **66**(2), 153-181.
- Yoshino, R. (1989). On the possible and stable psychophysical laws, *J. Math. Psych.*, **33**(1), 68-90.

家庭生活と仕事に関する調査結果の分析

釜野 さおり

スウェーデンと日本における家族関係と家族政策の関係の研究の一環として日本で実施した

家庭生活と仕事に関する調査について報告した。96年にスウェーデンで行なわれた全国レベルの家族関係についての調査(ストックホルム大学ヨーラン・アルネ教授代表)を日本の実状にあわせてアレンジした調査票を作成し、家事・就労・介護の実態、役割分担のプロセスと夫妻の対立、家族に関する価値・規範意識、就労に関する意識、夫妻関係の主観的評価などについての質問をした。ほとんどの家庭が共働きの核家族であるスウェーデンとの比較を可能とするために、連合大阪に所属している共働き核家族の女性を対象とし、その夫にも回答してもらう方式をとった。600カップルに調査票を配布し、492カップルの有効回答が集まった。

家庭生活における性別役割についての質問の回答をいくつか紹介すると、夫妻間の役割についての項目の中で夫にも妻にも一番多く支持されているのは、「父親はできるだけ多く子供の世話をした方が子どものためによい」という意見で、妻の9割近くと夫の8割近くが「賛成」または「やや賛成」を選んでいる。次に多く支持されたのは「男性は女性と同じくらい実際に家事や育児をするべきだ」という意見で、妻の8割、夫の6割弱が「賛成」または「やや賛成」と回答している。妻と夫の違いが大きく現れたのは「子どもが小さいうちは女性は家にいるべきだ」に対する意見で、妻は「反対/やや反対」が4割で「賛成/やや賛成」が2割強であるのに対し、夫は逆に「賛成/やや賛成」が4割、「反対/やや反対」が2割となっている。「女性が仕事に就くのは反対ではないが、あくまでも男性が家庭の主な稼ぎ手であるべき」についても、妻の6割以上が反対しているのに対し、夫では反対が4割以下である。

家庭生活における性別役割についてのもの、就労生活における性別役割についてのもの、結婚・離婚観に関するもの計13の項目の数量化III類を行なうと、男女平等的な考えを示す意見、それを否定する意見、中間回答がそれぞれきれいに固まり、一次元のスケールをなした。それぞれの回答カテゴリーに与えられた得点に応じて個人の得点を求め、男女別の平均値を求めると、女性の平均値は男女平等的な群に、男性の平均値はそれを否定する群に位置することがわかった。

統計計算開発センター

半無限計画法

伊藤 聡

次のような半無限計画問題を考える。

$$(1) \quad \min_x f(x) \quad \text{subject to} \quad g(x, \omega) \leq 0 \quad \forall \omega \in \Omega$$

ただし、 Ω をコンパクトな Hausdorff 空間とし、 $x \in R^n, g: R^n \times \Omega \rightarrow R$ とする。また、 f, g の微分可能性を次のように仮定する。

仮定 1. f は微分可能であり、 g は任意の $x \in R^n$ に対して ω に関して連続で、 $R^n \times \Omega$ 上で x に関して連続微分可能である。

さらに問題(1)の最適解において Guignard の意味での制約想定が成立することを仮定するが、この制約想定は実際に検証することが困難であり実用的ではない。そこで、ここでは Guignard の制約想定 の十分条件として次の Slater (あるいは Robinson) の制約想定を挙げておく。

仮定 2. g は任意の $\omega \in \Omega$ に対して x に関して凸であり、

$$g(x, \omega) < 0 \quad \forall \omega \in \Omega$$

を満たす点 $x \in R^n$ が存在する。

以上の仮定のもとで、問題(1)に対する最適性条件として次の命題が得られる (Ito et al. (1998)).

命題. (問題(1) に対する Karush-Kuhn-Tucker 条件) 仮定 1, 2 が成り立つとする. $x^* \in R^n$ が問題(1)の最適解ならば,

$$\begin{aligned} \nabla f(x^*) + \int_{\Omega} \nabla_x g(x^*, \omega) d\Lambda^* &= 0 \\ \int_{\Omega} g(x^*, \omega) d\Lambda^* &= 0 \end{aligned}$$

を満たし、高々 n 個の点からなる台を持つ (非負の) 正則な Borel 測度 Λ^* が存在する。

Ito et al. (1997, 1998) および Liu et al. (1998) では、この最適 Lagrange 乗数の測度としての有限性と双対理論に基づいた数値解法を提案している。

参 考 文 献

- Ito, S., Liu, Y. and Teo, K. L. (1997). A dual parametrization method for convex semi-infinite programming, The 16th International Symposium on Mathematical Programming, Lausanne, Switzerland (口頭発表).
- Ito, S., Liu, Y. and Teo, K. L. (1998). A dual parametrization method for convex semi-infinite programming, *Optimization: Techniques and Applications* (eds. L. Caccett, K. L. Teo, P. F. Siew, Y. H. Leung, L. S. Jennings, and V. Rehbock), Volume 1, 550-557, Curtin University of Technology, Perth.
- Liu, Y., Teo, K. L. and Ito, S. (1998). A dual parametrization approach to linear-quadratic semi-infinite programming problems, *Optimization Methods and Software*, **10** (to appear).

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

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ているときにも、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。数値例として、多変量2次元ARモデルのシミュレーションを行ない、そのデータを用いて数値実験を行なった。両方の成分ともに欠測値がないとき、両方の成分ともに欠測値があるとき、一方の成分のみに欠測値があるとき等について数値実験を行なった。一方の成分のみに、欠測値があるときのものは、両方の成分に欠測値がないときのものにかなり近い良い結果が得られた。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。

重回帰分析と情報量規準

石 黒 真木夫

重回帰分析における最小2乗法に p 次正則化項を導入することが汎化誤差を減少させるために有効であること、正則化のパラメータの適切な設定が情報量規準 EIC を用いることによって可能であることを示した。この正則化項は $p=2$ においてガウス型、 $p=1$ においてラプラス型に一致する。説明変数の候補の中に無意味な変数が含まれる場合、すべての変数が意味を持つ場合、共線性がある場合における挙動を調べて良好な結果を得ている。

統計科学情報センター

記述的多変量解析法と信頼性：官庁データの分析

馬 場 康 維

1. 数量化と強制分類

データ行列 F が与えられているものとする。ここで F は分割表型データ行列とする。これに、ある要因に対応する行列 A を加えた行列 K

$$K=[F, A, A, \dots, A]$$

を考えてみる。 A をいくら増やしても、 K のランクは同じである。しかし、 K による数量化をする場合、 A の個数が増えるにつれて K に占める A の影響が大きくなる。したがって、個体に与えられる数量から個体を分類する場合について考えると、 A の個数を増やすことによって、 A による強制的な分類を行うことができる (Nishisato (1986))。 K のランクを考えると、これは、

$$K=[F, kA]$$

とおき、 k を大きくしていくことと同等である。この方法を“強制分類法” (forced classification method) という。

2. 主成分分析における強制分類

記述的多変量解析法の代表的な方法である主成分分析法をとり上げる。主成分分析法では説明変数の直交変換により得られる“主成分”によって記述される少次元の空間に個体や変数をプロットすることによって個体間の関係や変数間の関係が解釈される。強制分類の考え方を適用することにより、この解釈の安定性について吟味できる。

主成分分析の場合はデータ行列

$$F=[f_{ij}] \quad (i=1, \dots, p; j=1, \dots, q)$$

は、 i が個体、 j が変数を表すことになる。

一つあるいは複数の変数の重みを変化させることによって強制分類が主成分分析に適用できる。

一般にデータ行列

$$K=[F, kA]$$

を考える。 A は重みを変える変数に対応するベクトルである。 k を大きくしていけば強制分類になり、 k を小さくしていけば、感度分析に呼応する。したがって、一般には

$$k \geq 0$$

と定義される。

参 考 文 献

- 馬場康維 (1996). 等価分割原理の主成分分析への応用, 第10回日本計算機統計学会シンポジウム論文集, 72-73.
- 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リポート, No. 100 (柳井晴夫編), 41-49.
- Nishisato, S. (1986). Generalized forced classification for quantifying categorical data, *Data Analysis and Information* (eds. E. Diday et al.), 351-362, North-Holland, Amsterdam.

日本人の児童・生徒の体型の変化について —文部省学校保健統計調査報告書より—

金 藤 浩 司

日本人の児童・生徒の身長と座高の経年変化を文部省が毎年調査し発行している「学校保健統計調査報告書」に基づき解析した。取り扱ったデータは、1943年生まれから1979年生まれまでの各学年の身長と座高の平均値である。解析においては、各平均値を出生年度毎にまとめ12年間分の経年的データとして取り扱った。また、本研究で用いた解析手法は、ヒトの体格成長を非線形の成長模型を用いて表現した非線形回帰である。

はじめに、身長と座高の成長を非線形の成長模型にあてはめ男女別に解析した。次に、身長と座高から計算される座高比についても議論した。推定した成長模型中の成長母数の推定値や体格成長の思春期のスパート時の年齢等の成長の特性値は、男女とも1960年付近で傾向が変化することが捉えられた。その変化を要約すると、1960年付近までは、各推定値は、出生年度とともに比較的単調に増加や減少傾向を示している。しかし、1960年付近以降はそれまでと異なり、各推定値の傾きが零に近づく傾向がみられる。また、座高比曲線を用いて日本人の平均体型を捉えると、近年、一定の曲線に近づきつつあることが分かった。

本研究では、成長模型としてPreece and Baines (1978)の提案した4つの成長模型の中でモデル1と呼ばれる成長模型を用いた。しかし、この模型を用いた理由は、解析対象が乳幼児期を含まず、かつデータ変動が少ない集団の平均値であることによる。もし、対象とするデータが個体である場合や測定区間が乳幼児期を含む場合では、Shohoji and Sasaki (1984)やJolicoeur et al. (1988)の成長模型を用いた解析が望まれる。

参 考 文 献

- Jolicoeur, P., Pontier, J., Pernin, M.-O. and Sampé, M. (1988). A lifetime asymptotic growth curve for

human height, *Biometrics*, **44**, 995-1003.

Preece, M. A. and Baines, M. J. (1978). A new family of mathematical models describing the human growth curve, *Annals of Human Biology*, **5**, 1-24.

Shohoji, T. and Sasaki, H. (1984). The growth process of the stature of Japanese growth from early childhood, *Acta Medica Auxologica*, **16**, 101-111.

一般化 Fokker-Planck 方程式の簡約と補正

岡崎 卓

外乱を受けて発展する系の確率密度を定める指数分解型一般化 Fokker-Planck (*e*-GFP) 方程式の実用性を高め精度の向上を図るため、簡約と補正を施す方法を述べる。

1. *e*-GFP 方程式の簡約

外乱 W を受けつつ、確率微分方程式

$$(1.1) \quad \frac{d}{dt}U = M(U) + \mu(W)$$

に従って変動する変数 U の確率密度 $f(u, t)$ を定める *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、作用素表示による当初の形式に較べ、通常の数値計算法で解を求め得る程に扱い易い形を獲得したが、拡散項に空間積分のみならず確率密度の過去の状態 $f(U, s)$, $s < t$ に関する履歴積分が存在するため、解析的取扱いを困難とし、有色ガウス外乱に対してさえ解が簡単に求まらない。

そこで系変数の発展方程式に微小量 ε を導入し、*e*-GFP 方程式の形式解を求めて拡散項の大きさを評価することにより、履歴積分を除去する。

2. 拡散項の補正

上に得た簡約版 *e*-GFP 方程式を有色ガウス外乱を受ける線形系に適用しても、既知の厳密な確率密度を与えない、これは指数分解近似により元々の GFP 方程式から何らかの情報が失われたためと考えられる。この近似は指数演算子

$$e^{t(A+\varepsilon B)} = e^{tA} + \varepsilon \int_0^t ds_1 e^{s_1 A} B e^{(t-s_1)A} \\ + \varepsilon^2 \int_0^t ds_1 \int_0^{t-s_1} ds_2 e^{s_1 A} B e^{s_2 A} e^{(t-s_1-s_2)A} + \dots$$

における ε の全ての次数の項を採用するが、しかし全ての項を取らず、無限級数の部分列 Y_{exp} を取ることを意味する：

$$e^{t(A+\varepsilon B)} = Y_{exp} + Y_{excess}.$$

従って *e*-GFP 方程式に精密さを付与するにはこの失われた情報 Y_{excess} を回復し、拡散項に補正を加えなければならない。そこで、ガウス外乱に対して正しい解を与えるように Y_{excess} を決めて固定し、その結果を一般の非ガウス外乱にも当てはめることとする。この補正によって *e*-GFP 方程式は有色ガウス外乱に対して厳密となり、且つその補正量は外乱の skewness が小さい限り無視できる微小量であるので、*e*-GFP 方程式の非ガウス外乱に対する有効性は損なわ

れない。

3. e-GFP 方程式の表現

以上の手続きによって拡散項の簡約と補正を行った結果、e-GFP 方程式は次の簡潔な形を採る。

$$\begin{aligned} \frac{\partial}{\partial t} f(U, t) = & -D_{U_k} M_k(U) f(U, t) + D_{U_k} D_{U_l} \lambda_{Akl}(t) f(U, t) \\ & + D_{U_k} D_{U_l} \int dV \lambda_{Bkl}(V, t) f(U - V, t) \end{aligned}$$

ここに λ_A は local な拡散係数、 λ_B は non-local な拡散係数であり、非確率過程 $(d/dt)U = M(U)$ の Jacobian $R(U, t) = (\partial U(t)/\partial U)^{(-1)}$ を用いて表される ($\mu[t] = \mu(W(t))$)。

$$\begin{aligned} \lambda_{Akl}(t) &= \int_0^t ds R^*(-s) \langle \mu_l[0] \mu_k[s] \rangle \\ \lambda_{Bkl}(V, t) &= \int_0^t ds R^*(-s) (\Phi_{Akl} - \Phi_{0kl}) \\ \Phi_{Akl} &= \int_0^t ds R^*(-s) \langle \mu_l[0] \delta(V - sR^*(s)) \mu_k[s/2] \rangle \\ \Phi_{0kl} &= \langle \mu_l[0] \mu_k[s] \rangle \langle \delta(V - sR^*(s)) \mu[s/2] \rangle > \frac{1}{\langle \mu^2 \rangle} \left(\frac{V}{sR^*(s)} \right)^2 \end{aligned}$$

通常の FP 方程式 (これは白色ガウス外乱に対してのみ正しい) は、 $R(t) = 1$ と置き、且つ $f(U - V, t)$ を含む convolution 項を省いたものである。これに対して e-GFP 方程式は次の特色を有する

1. 拡散係数には Jacobian $R(t)$ を介して、系の構造が考慮されている。
2. 外乱の非ガウス性を反映する non-local な拡散項が存在する。

予測制御研究系

情報量規準 GIC について

北川 源四郎

情報量規準 AIC は最尤法で推定されたモデルの対数尤度を平均対数尤度の推定量と考えたときのバイアスを評価し、補正することによって導かれた。一般化情報量規準 GIC はこの方法を統計的汎関数で定義される一般的な推定量 $\theta = T(G)$ の場合に拡張したものである。この場合、バイアスは

$$(1) \quad b_1(G) = \text{tr} \left\{ T^{(1)}(X_a; G) \frac{\partial \log f(X_a | T(G))}{\partial \theta} \right\},$$

で与えられる (Konishi and Kitagawa (1996))。ただし、 $T^{(1)}(X_a; G)$ は $\theta = T(G)$ のコンパクト微分である。この方法を精密化すると、バイアスの1次補正量のバイアス、その分散、バイアスの2次補正量などを評価することができる (Konishi and Kitagawa (1998))。

モデルと真の分布がともに正規分布であるもっとも簡単な場合には、バイアスの1次補正量のバイアスはデータ数を n とするとき $-3/n$ となる。これは、真のキュミュラントが分かっている

ない場合には、バイアスは上記の分だけ過小評価されることを意味する。この単純な場合には GIC と TIC の理論値は AIC と一致する。したがって、平均と分散を未知パラメータとする正規分布モデルの場合には、TIC と GIC のバイアス補正值は $2-3/n$ となる。 $n=25$ の場合にこの値は 1.88 となりシミュレーションによる評価値をよく説明している。バイアスの 2 次補正を行う場合には、この 1 次補正量のバイアスの補正も同時に行う必要がある。

よく知られているように、バイアス項は 3 つの部分 b_1 , b_2 , b_3 に分解できる。このうち、第 2 の部分 b_2 は平均が 0 であるので実際には無視することができる。各項の分散を評価すると一般には 3 項とも $1/n$ のオーダーの分散を持つが、 b_1+b_3 の分散は $1/n^2$ のオーダーであることが示される。これから、情報量規準 EIC においても b_1+b_3 のブートストラップを行うことにより推定分散を著しく減少できることを説明できる。

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.
- Konishi, S. and Kitagawa, G. (1998). Second order bias correction for generalized Information Criterion, Research Memo., No. 663, The Institute of Statistical Mathematics, Tokyo.

季節調整の適切さについて

川崎能典

四半期あるいは月次等の経済時系列に共通して、ゼロ周波数、季節周波数およびその高調波で鋭いピークを持つスペクトルがしばしば観察される。季節調整を施すとは、時間領域で考えれば一年というある一定期間内での周期的なパターンを取り除くことであるが、周波数領域で見るとこれは、季節周波数でのピークを取り除く操作に該当する。直観的には、ピークが残っていても逆に溝が生じて季節調整は適切でないと考えてしまいそうであるが、これは実は季節調整の善し悪しを測る基準にはなり得ない。

一般に、加法型に帰着される季節調整モデルにおいて、季節成分などの観測不能要素（オリジナルな意味でのパラメータ）を最小二乗推定すると、季節調整済み系列の標本スペクトルは、季節周波数で多かれ少なかれ溝を示す。これは、最小二乗季節調整フィルタが、観測値に対するモデルのスペクトルと構成要素に対するモデルのスペクトルの比の二乗で表されることから容易に理解される。従って、seasonal dip と呼ばれる現象は、過剰調整を示すものではなくある種の「最適な」季節調整の特徴付けになっている。

このように、原理的には必ず存在する季節性の溝が、実際には見えたり見えなかったりするのだが、この現象は季節調整モデルにおける不規則変動項（観測ノイズ）の役割に着目することで解明できる。一般に、観測ノイズの分散が大きければ大きいほど、季節調整済み系列における seasonal dip はノイズレベルに埋もれて見えなくなっていく。また、ここでいう観測ノイズの分散は、その絶対スケールではなく、むしろ季節成分やトレンド成分を駆動するイノベーションの分散に比して相対的にどの程度の大きさを持っているかが本質的である。

明示的にモデルに基づかないセンサス局法に関しても同様のメカニズムが働いていることが、数値実験を通して類推できる。(ARIMA 等による近似モデルの妥当性を認めてしまえば、

これまでの議論はそのまま成立する。) 特にセンサス局法の持つ癖として、観測ノイズが本来ないような系列に対しても無理矢理観測ノイズをひねり出す傾向がある。このような場合は、seasonal dip は生じないかもしれないが、時間領域で見ると季節調整は明らかに失敗している。詳しくは以下の文献を参照。

参 考 文 献

川崎能典, 佐藤整尚 (1997). 季節調整の「最適性」について, 統計数理, 45(2), 245-263.

地球電流系構造の自動同定

樋 口 知 之

極域地方に特有の自然現象であるオーロラは、太陽風と地球固有（内部）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ひきおこす物理プロセスの一部分である。電流が流れることによって生じる磁場変動を宇宙空間中で観測すれば、大規模電流系の構造を直接的に推定することができるため、人工衛星の磁場データを解析した研究はこの20年来かなり精力的になされてきた。最近全く新しいタイプの電流系の存在が4例発見され、その解釈のため新しい相互作用のモデルが仮説として提案されている。この4例しか見つかっていない新しいタイプの電流系は、実はさまざまな偶然の物理条件が重なり合った結果の希な現象なのか、それとも今まで見過ごされてきた実はごく普通に存在するものなのかを明らかにすることは、地球物理の観点から興味深い重要なテーマである。そのため、低高度（周回時間が約100分程度）の極周回人工衛星の今までに蓄積された超大量の磁場データを丹念に調べあげ、新しいタイプの電流系を発見するプロジェクトに取り組み始めた。

この大規模電流系は、同じタイプの電流系であっても、その空間的位置や電流値は、時々刻々と変化する太陽風の影響や日照条件（極域地方の太陽となす角度）の違いでかなり変動するため、電流系の自動的同定は非常に困難である。当然、従来知られているタイプの電流系も同様の時間・空間変動性を示す。従って、電流系の空間的位置や電流値の変動性を容認しつつ、タイプ別に分類する柔軟な手法の開発が必須である。電流系は経度方向にシート状の構造をもつので、もともと3次元の磁場データをうまく座標変換すると、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、大規模電流系に関連する3次元磁場データを適切かつ柔軟に表現できるモデル——具体的には、可変節点リニアスプライン——を組み合わせた新しい手法を開発した。超大量のデータセット（最低でも、10ギガバイトをこえる量）を取り扱うため、その手続き（アルゴリズム）を完全自動化する工夫を行った。

情報量規準による重み付き最尤法の最適化

下 平 英 寿

確率変数 x を観測した時の y の条件付き密度関数を $q(y/x)$ とする。未知の $q(y/x)$ のモデルとして、パラメタ $\theta \in \Theta$ をもつ密度関数 $p(y/x, \theta)$ を考える。ここでは、モデル $p(y/x, \theta)$ は真の

分布 $q(y|x)$ を必ずしも含まない、つまり misspecification の状況を想定している。さて、過去の x の密度関数 $q(x)$ と、未来の密度関数 $q_f(x)$ が異なるという設定を考える。つまり、サンプルサイズ n のデータ $(x^{(n)}, y^{(n)}) = ((x_t, y_t) : t=1, \dots, n)$ は i.i.d. で $q(y|x)q(x)$ に従うが、未来の観測での (x, y) は $q(y|x)q_f(x)$ に従うとする。ここでは簡単のため、 $q(x)$ と $q_f(x)$ を既知とする。普通の最尤法では、「重み付き対数尤度」(weighted log-likelihood)

$$(1) \quad L_w(\theta|x^{(n)}, y^{(n)}) := \sum_{i=1}^n w(x_i) \log p(y_i|x_i, \theta)$$

において $w(x) \equiv 1$ とおいた対数尤度の最大化によって最尤推定 $\hat{\theta}$ を与える。これは、未来に $x \sim q(x)$ を観測して $p(y|x, \hat{\theta})$ によって y を予測するなら良い推定量であるが、 $x \sim q_f(x)$ に対しては必ずしも良い推定ではない。そこで、仮に $q(y|x)q_f(x)$ に従うデータがあった時の対数尤度を、“importance sampling”を用いて近似的に実現する。つまり、(1)での重み関数を $w_f(x) := q_f(x)/q(x)$ とした対数尤度を $L_f(\cdot)$ とし、この最大化によって重み付き最尤推定 $\hat{\theta}_f$ を与える。さらに、任意の重み $w(x)$ に関して、対数尤度(1)を最大化するパラメタ値を $\hat{\theta}_w$ と書く。

真の分布 $q(y|x)q_f(x)$ から予測分布 $p(y|x, \hat{\theta}_w)q_f(x)$ への隔たりを K-L 情報量で測ると、その期待値の推定量として情報量規準

$$(2) \quad IC_{\hat{\theta}_w} := -2L_f(\hat{\theta}_w) + 2\text{tr}(J_w H_w^{-1})$$

が与えられる (図1)。ただし、 $\theta_w^* = \text{plim}_{n \rightarrow \infty} \hat{\theta}_w$ として

$$J_w = E_q \left\{ w_f(x) w(x) \frac{\partial \log p(y|x, \theta)}{\partial \theta_w^*} \frac{\partial \log p(y|x, \theta)}{\partial \theta_w^{*'}}} \right\}, \quad H_w := E_q \left\{ -w(x) \frac{\partial^2 \log p(y|x, \theta)}{\partial \theta_w^* \partial \theta_w^{*'}}} \right\}.$$

特に正規回帰では $\hat{\theta}_w$ は WLS になり $\text{tr}(J_w H_w^{-1})$ の推定の計算もハット行列の対角項を使って容易に出来る。一般に、 $w(x)$ が $w_f(x)$ に近い程 $-L_f(\hat{\theta}_w)$ は小さくなり、それとは逆に $w(x) \equiv 1$ に近い程 $\text{tr}(J_w H_w^{-1})$ は小さくなる。そこで、 $\lambda \in [0, 1]$ をパラメタとし、 $w(x) = w_f(x)^\lambda$ を重みとして(2)を最小にする $\hat{\lambda}$ を求める。図2は(2)の各項を λ に関してプロットしたもの。

misspecification の時、 $n \rightarrow \infty$ で $\hat{\lambda} \rightarrow 1$ となるので、 $\lambda = 1$ に相当する $\hat{\theta}_f$ が漸近的には良い。実際、この時の $\hat{\lambda}$ の変動による効果も情報量規準に考慮すると、この項は(2)の第2項よりさらに

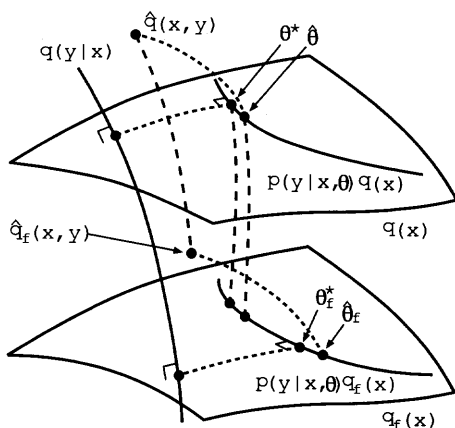


図1. 情報幾何的解釈: (x, y) の密度関数の空間。

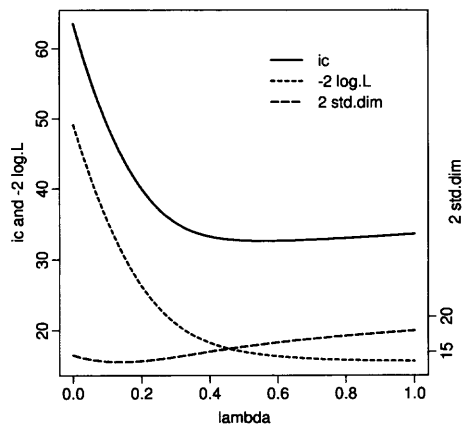


図2. λ の最適化の数値例。

n^{-1} のオーダーで小さく、 λ の最適化は漸近的にはあまり意味が無い。しかし、misspecification の程度に対して n があまり大きくない時は、 $\lambda=0$ に相当する $\hat{\theta}$ のほうが $\hat{\theta}_f$ より (2) の期待値を小さくする。これらどちらの状況でも、 $w(x)=w_f(x)^{\hat{\lambda}}$ を使った $\hat{\theta}_w$ は (その導出はややアドホックなもの)、比較的良好な振舞をする。 $\hat{\lambda}$ の計算が困難な場合は、 $\hat{\theta}$ と $\hat{\theta}_f$ から (2) を小さくするものを選ぶのも、実用的である。

参 考 文 献

- Shimodaira, H. (1997). Predictive inference under misspecification and its model selection, Research Memo., No. 642, The Institute of Statistical Mathematics, Tokyo.
- 下平英寿 (1997). 重み付き最尤法による予測推論とモデル選択, 研究集会「推定論とその応用の研究」資料 51-59.
- 下平英寿 (1998). 情報量規準による重み付き最尤法の最適化, 「非線形統計モデルにおける統計理論とその応用」研究論文集.
- 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.

複雑な系の非線形時系列解析

尾 崎 統

統計数理研究所の赤池弘次らによって導入され現実の多くの複雑なシステムの予測と制御に役立ってきた線形モデルにもとづく TIMSAC 法を非線形なシステムに発展させ、より多くの複雑なシステムの予測と制御に役立てることをめざして研究を行ってきた。非線形複雑系の例としては、本年度は脳神経機能、火力発電プラント予測制御、経済、金融、官庁統計などを扱った。研究成果は以下の論文に収めた。

参 考 文 献

- Jimenez, J. C. and Ozaki, T. (1997). Identification of continuous-discrete state space models with delays, Research Memo., No. 658, The Institute of Statistical Mathematics, Tokyo.
- Jimenez, J. C., Shoji, I. and Ozaki, T. (1997). Simulations of stochastic differential equations through the local linearization method. A comparative study, Research Memo., No. 651, The Institute of Statistical Mathematics, Tokyo.
- 尾崎統 (1997). 動的 X 11 モデルと非線形季節調整 I—モデルと計算法—, 統計数理, 45, 265-285.
- 尾崎統 (1997). 動的 X 11 モデルと非線形季節調整 II—解析例と考察—, 統計数理, 45, 287-300.
- Ozaki, T., Jimenez, J. C., Biscay, R., Valdes, P. (1997). Nonlinear time series models and neural dynamical systems, Research Memo., No. 657, The Institute of Statistical Mathematics, Tokyo.
- Shoji, I. and Ozaki, T. (1997). Comparative study of estimation of continuous time stochastic processes, *J. Time Ser. Anal.*, 18, 485-506.
- Shoji, I. and Ozaki, T. (1998). Statistical method of estimation and simulation for systems of stochastic differential equations, *Biometrika*, 85, 240-243.
- Toyoda, Y., Oda, K. and Ozaki, T. (1997). The nonlinear system identification method for advanced control of the fossil power plants, *Proc. of 11th IFAC Symposium on System Identification*, 8-11, July 1997, Kitakyusyu, Fukuoka, Japan.
- Valdes, P., Jimenez, J. C., Riera, J., Biscay, R. and Ozaki, T. (1997). Nonlinear EEG analysis based on a neural mass model, Research Memo., No. 652, The Institute of Statistical Mathematics,

Tokyo.

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試みている。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の研究を行っている。

この数年間は外乱や非線形成分、次数や相対次数に依存しない適応制御系の構成法の研究に関わってきた。その中でこれまでに、可変構造制御系、ハイゲインフィードバックやバックステッピング法などにより、外乱・非線形成分・次数、そして部分的に相対次数についても従来の制約を緩和した適応制御手法を導出することができた。今年度はさらに相対次数に3次の幅の不確実性がある場合や、対象の高調波利得の符号に不確実性がある場合でも対処できる適応制御系の構成法、具体的には適応安定化制御系、モデル規範形適応制御系、適応サーボ系の構成手法について研究を行い、参考文献にあげるような結果を得た。

これらと平行して、そこで得られた制御系が複雑な構成となる問題にも着目して、バックステッピング法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の構成手法も求めた。実際の応用も考慮して、新たに導出した適応制御系において、対象に寄生要素の影響がある場合でも準大域的な有界性が保証されること、寄生要素の種類によっては制御手法に改良を加えることで寄生要素の影響を任意に小さくできることも示した。

また従来の適応制御理論が漸近安定性に主眼を置くのに対して、制御性能をもより定量的に考慮する観点から、適応制御の過程を最適制御問題や M_∞ 制御問題としても定式化することを試みた。簡単な事例において、いくつかのモデル規範形適応制御系やロバスト適応制御系が、制御入力に適切な補助入力を加えることで、特定の評価規範に関して前出の制御問題に対する最適解（または準最適解）となっていることを示すことができた。

参 考 文 献

- Hanba, S. and Miyasato, Y. (1997). Observers, adaptive observers and output feedback stabilizing controllers for uniformly observable bilinear systems, *Preprints of 11th IFAC Symposium on System Identification*, **2**, 919-924.
- Miyasato, Y. (1996). Model reference adaptive control for systems with unknown relative degrees, *Proceedings of European Control Conference 1997*, **2**.
- 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). A design method of universal adaptive stabilizer, *Proceedings of the 36th IEEE Conference on Decision and Control*, **2**, 1896-1901.
- Miyasato, Y. and Hanba, S. (1997). Adaptive control for nonlinear systems with unknown degrees and uncertain relative degrees, *System Identification 1997 (SYSID'97)* (eds. Y. Sawaragi and S. Setsuo), 903-908, Elsevier, England.
- 宮里義彦 (1998). 不確実な相対次数に対するモデル規範形適応制御系の構成法 (任意の3次の幅の不確実性に対する構成), 計測自動制御学会論文集, **34**(2), 87-95.
- Miyasato, Y. (1998). A design method of universal model reference adaptive controller, *Proceedings of 1998 American Control Conference*, **1**, 609-613.

宮里義彦 (1998). 次数と相対次数に依存しない非線形適応制御系の構成法 (相対次数が3次の範囲で不確定な場合), システム制御情報, 11(6) (掲載予定).

資産価格のトレンドとボラティリティの関係

佐藤 整 尚

従来は, 収益率のみに注目し, そのボラティリティを推定することが多いが, 本研究では, 原系列に戻って, トレンド周りのノイズ系列のボラティリティに注目した. そして, このボラティリティとトレンドの関係を考えることによって, データへの当てはまりが良くなった.

近似計算を使った実用的な内点法

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. 内点法は, 反復解法であり, 1反復毎に線形方程式を解き, 探索方向を計算し, 点を更新する. 内点法の収束性の解析では, ほとんどの場合に, この線形方程式の解が正確に計算できるものと仮定される. しかし, 大規模な線形計画問題では, 正確な解を求めるために多くの計算時間を必要とするため, 近似計算を使うことがよくある.

Mizuno and Jarre (1996)では, 線形方程式の解を近似的に計算する内点法を提案し, その収束性を解析した. しかし, この方法は, センターパスの小さな近傍内で点列を生成するために, 実用的に高速な解法ではなかった. そこで, 近似計算を使うが, 大きな近傍内に点列を生成する内点法を提案し, その収束性を解析した. そして, 点列のサイズが大きくなるに従い, 近似精度を上げるという工夫を凝らすことにより, 大域的な収束性を持つ実用的な内点法を開発した.

参 考 文 献

- Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming, *Combinatorica*, 4, 373-395.
- Mizuno, S. and Jarre, F. (1996). Global and polynomial-time convergence of an infeasible-interior-point algorithm using inexact computation, Research Memo., No. 605, The Institute of Statistical Mathematics, Tokyo.

シミュレーションから分布の構造を知る — Finite mixture を用いたアプローチ —

伊庭 幸 人

高次元・多変量の分布の場合, マルコフ連鎖モンテカルロ法などの数値的な方法で分布から

のサンプリングができて、分布の構造の全体像はよくわからないということがしばしば起こる。統計物理には“準安定状態” (metastable state) という概念があり、ギブス分布の全体が近似的にいくつかの準安定状態に分解するというような直観的な捉え方をするが、その定義は有限系では必ずしもはっきりしない。本研究の目的は、local optimum の一般化としての (有限温度での) 準安定状態を有限混合分布 (finite mixture model) を通じて定義し、構成することである。統計物理でいう“準安定状態”はダイナミカルな性質、サンプリング法に依存する概念であって、ギブス分布の性質ではないという考え方もあるが、本研究では分布の性質であるという立場をとる。

ここでは、確率変数が成分が2値 (± 1) のベクトルであるような場合、具体的にはイジング・スピングラスモデル (特に SK モデル) と連想記憶モデル (Hopfield モデル) を解析の対象としてとりあげ、それらのモンテカルロシミュレーションデータ (交換モンテカルロ法 (MCMCMC) を使用) を下記の有限混合分布モデルによって解析した。

$$(1) \quad \log L(d|x) = \sum_i \log \left(\sum_c w_c \frac{\exp(\sum_j x_c^j d_i^j)}{Z_c} \right)$$

$$(2) \quad Z_c = \prod_j \{ \exp(+x_c^j) + \exp(-x_c^j) \}$$

$\log L(d|x)$ がモデルの対数尤度、未知パラメータは $\{x_c^j\}$ と $\{w_c\}$ である。 d_i^j は i 番目のデータ (snapshot) の j 成分を示す。このモデルは、どのデータも各準安定状態 c に対応する“テンプレート” $x_c = \{x_c^j\}$ のいずれかの近くに分布していることを表わす。 w_c は準安定状態 c 全体の重みであって $\sum_c w_c = 1$ を満たすものとする。 d_i^j は ± 1 のいずれかの値をとるが、テンプレートの成分 x_c^j は連続値をとる。これは、Gaussian mixture の場合の分散に相当する部分 (“テンプレートの強さ”) も x_c^j に含まれているからである。物理的には x_c^j は準安定状態 c において変数 (スピン) j に働く内部場であると解釈できる。このモデルのひとつの利点は、対称性を組み入れるのが比較的容易なことである。たとえば磁場なしのスピングラスモデルからのサンプルを解析する場合、up-down 対称性を考慮した解析が可能である。モデルのあてはめは EM アルゴリズムによる繰り返し計算によって行った。

計算機実験の結果、一応もつともな結果が得られたが、特定の c について、テンプレートの成分 x_c^j の絶対値が小さくなる現象がみられた。このひとつの原因は、本来どのテンプレート (準安定状態) にも属さないはずのデータが特定のテンプレート (で定義されるクラスター) に集まるためではないかと思われる。現在この点に関する改良を試みている。

本研究は、物性研究所の福島孝治氏との共同研究であり、その結果は物理学会 秋の分科会 (神戸大学) で発表された。

あるクラスの半正定値計画 (SDP) 問題に関する考察

(客員) 大阪大学 小原 敦 美

近年、半正定値計画 (Semidefinite Programming, 以下 SDP) 問題と呼ばれる凸計画問題が盛んに研究されている。すでによく知られているように、この問題は信号処理、通信などを含むシステム制御工学に線形行列不等式アプローチとして広い応用をもち、最近では組み合わせ最適化などへの適用も活発である。またこの問題は線形計画 (Linear Programming) 問題の素直な拡張となっており、Karmarkar 法を端緒に LP で研究された多項式時間解法である内点法

のメカニズムが適用でき、求解のために必要な反復回数などが解析されている。

一方、甘利・長岡（1993）らによる情報幾何は統計理論を中心として成果を上げているが、そこに現れる双対性の一側面は Legendre 変換を通して現れるということからわかるようにもともと凸解析と関係が深い。特に凸計画問題の内点法に現れる手法やアルゴリズムは情報幾何の観点から興味深い解釈ができる。

Ohara et al. (1996), Ohara (1997) は SDP 問題について情報幾何の観点から考察を行い、考えている SDP 問題の可能領域を正定対称行列集合の部分多様体と見て、それが主・双対両接続の意味で自己平行（以後二重自己平行と呼ぶ）と呼ばれる性質を持つとき、その SDP 問題が Newton 反復無しで解けることを示した。この特別な場合として接空間が Jordan 代数と呼ばれるものを用いて代数的に特徴づけられる部分多様体が現れる。また、二重自己平行でなくとも、双対接続に関する第二基本形式と呼ばれる曲がり方を表す量が小さければ反復回数も少なくなるアルゴリズムを構成できることを示した。

平成9年度の研究報告会では、まず正定対称行列集合に含まれる二重自己平行な部分多様体のいくつかを具体的に構成し、自明ではない SDP 問題が Newton 反復無しで解けるクラスに属することを示した。次にこのクラスに属する SDP 問題に対し最適解の陽な表示を与えた。最後に Vanderbei and Yang (1995) が全く異なる方法によって与えていた Newton 反復なしで解ける SDP 問題のクラスが、二重自己平行性で特徴づけられるクラスに含まれることを示した。

これらの研究を通して得られた結果が、内点法の数理的な側面と隣接した諸分野との関連の一端を明らかにする一方、大規模な SDP 問題を解く際のひとつの実用的な手法への糸口となれば幸いである。

参 考 文 献

- 甘利俊一，長岡浩司（1993）．『情報幾何の方法』，岩波書店，東京。
- Ohara, A. (1997). Information geometric analysis of an interior point method for semidefinite programming, 数理解析研究所講究録, No. 1004, 71-89.
- Ohara, A., Suda, N. and Amari, S. (1996). Dualistic differential geometry of positive definite matrices and its applications to related problems, *Linear Algebra Appl.*, **247**, 31-53.
- Vanderbei, R. J. and Yang, B. (1995). The simplest semidefinite programs are trivial, *Math. Oper. Res.*, **20**(3), 590-596.

調査実験解析研究系

国民性調査に関する基礎的検討（2）

坂 元 慶 行

今年度は、主として、国民性調査の過去のデータの整理や将来の調査のための準備、特に、新しい質問文の試作を行なった。国民性調査の一部の質問項目には、調査開始以来 40 数年を経、いわば意識動向の測定手段としての機能を失ったものがある。そこで、可能なら次々回からの 21 世紀の調査で使える質問文を、それが難しければ次回の全国調査で使える質問文を開発・選定することをめざした。21 世紀の調査では、従来の質問にとらわれず、新機軸を出すのが望ましいが、実際には適当な手がかりがないと考察を進めにくい。そこで、これまでの国民性調査から得られた知見を手がかりにすることにし、人間関係観の変化が小さかったこと、「一

番大切なものは家族」が激増したこと、女性に生まれ変わりたい女性が増加したこと、等の知見に基いて、家族や職場での人間関係観を中心に新質問を試作したが、決定的な成果を得るには至らなかった。

このため、報告では、次回の全国調査のための検討から得られたいくつかの知見について述べた。たとえば、「生まれかわるなら女に」と言う女性が増えたことについてこれまで様々な解釈がなされているが、調査データからはその説明に成功しているとは言い難い。われわれ自身も、既に1960～70年代に「男と女で、楽しみはどちらが多いか」や「苦労はどちらが多いか」との関係を調べ、今年度は「どちらが自由か」等との関連も調べてみたが、決定的と言えるほどの関連は見い出せなかった。要するに、現在のところ、「どういう訳か、女に生まれたがる女性が増えた」とでも言う以外にない。ただ、この意見の増加に対しては時代効果が支配的で、このことから、世代や年代の違いを超えて女性が一斉に同じ理由で意見を変えて来たと思いがちであるが、理由は単一ではなく、それぞれの世代や年代がそれぞれの理由で意見を変えてきたとの見方も検討に値するのではないかと思われる。

95年SSM調査の標本設計と標本精度

中村 隆

「社会階層と社会移動全国調査(SSM調査, Social Stratification and Social Mobility Survey)」は、1955年(昭和30年)から10年ごとに実施されている調査であり、1995年に第5回目が実施された。標本設計が、1953年から5年ごとに実施している「日本人の国民性調査」の一番近い時点のそれとほぼ同じであり、標本抽出のための資料の提供や調査地点の抽出をはじめとして、統計数理研究所とは協力関係が深い。

1995年のSSM調査では、職歴を中心とする本調査A、意識を中心とする本調査B、職業評価を中心とする職業威信調査(P調査)の3種類の調査票が用いられた。本調査AとBは全国336調査地点で男女各12人について1人おきに、P調査は本調査の調査地点から1つおきに選んだ約半数の169地点で男女各5人に割り当てられた。計画標本全体の大きさは9,739になる。

95年SSM調査の標本設計は、一口に層別多段無作為抽出法と呼べるような方法によっている。市区町村をその地域性と人口規模に基づいて層別し、各層に標本の大きさを比例割当てしている。投票区の抽出までは有権者数に基づく確率比例抽出であり、調査対象者は各投票区の有権者名簿上での系統抽出である。

標本誤差の推定は、層別確率比例復元2段無作為抽出法を仮定して行なった。たとえば、P調査の56職業についての5段階評価(+DK)について計算したところ、有効な計331カテゴリーの回答比率の標準誤差は、単純無作為抽出を仮定する場合のほぼ1.1～1.2倍の範囲にあることをみた。

層別多段抽出法で得られた調査データについて、標本抽出法を考慮して分析がなされるのはきわめてまれである。そのような分析法が少ないことと、あっても手順がかなり面倒だからである。このような局面を打開すべく、本年度は一般Wald検定に基づく方法、有限母集団版ブートストラップ法の利用などについて研究を行なった。

参 考 文 献

中村 隆(1998 a). 調査データの標本誤差評価とブートストラップ法——層別確率比例復元2段無作為抽出

法の場合一, *ESTRELA*, 3月号, 11-19.

中村 隆 (1998 b). 1995年SSM調査の標本設計と標本精度—標本抽出法を考慮した分析に向けて—, 『社会階層・移動の基礎分析と国際比較』(石田浩 編), 1995年SSM調査シリーズ1, 77-100, 1995年SSM調査研究会.

階層帰属意識と生活満足感の分析

前 田 忠 彦

複数の社会調査でしばしば取り上げられる項目である「階層帰属意識」と「生活満足感」に着目し, この二項目に対する説明要因と二項目間の相互関係を検討するための分析を行った。主な使用データは「1995年社会階層と社会移動調査」(以下1995年SSM調査; 1995年SSM調査研究会(1996))によるもので, 1993年に行われた「日本人の国民性調査第9回全国調査」(統計数理研究所(1994))のデータによる分析結果と随時比較を行いながら考察を進めた。本研究の目的は, 第一に階層帰属意識と生活満足感相互の連関, 及び両者に対する説明要因について検討し, 両者の調査項目としての性質を論ずること, 第二に階層帰属意識と生活満足感の間の関係について, 両者のうちどちらが先行する要因であるかについての考察を因果モデルの分析を通じて行うこと, の2点であった。

第一の目的のために, まず両者の連関が線形連関と呼べる性質のものであることをクロス集計とその対応分析により確認した。続いて, 両者を目的変数とし, それ以外の調査項目を説明変数と考えて連関指標を算出し, 両者に対する主要な説明要因を探索した。更に, 説明変数として共通の潜在変数を導入し, 上記二変数を目的変数とした多変量回帰分析を行い, 両者の規定因を比較した。第二の目的のためには, 両者を内生的変数とする複数の因果モデルを構成し(その中には両者が相互に影響を与えあうとする双方向因果を含むモデルも含まれていた), モデル選択を行うことを通じて一種の因果推論を行った。

以上の分析から本研究で得られた主要な結果は以下の通りである。

1. 階層帰属意識の最も主要な規定因は経済的地位である。学歴, 職業威信などの社会的な地位は, 経済的地位を介して階層帰属意識に影響を与えるが, 直接影響を与えるという側面は弱い。
2. 生活満足感は各種の主観変数(意識項目)と幅広く連関を示すが, 中でも主要な規定因は, 時間的余裕と経済的地位であるように思われる。
3. 生活満足感と階層帰属意識の間の関係については, 前者が後者に対して先行し, 他の変数の媒介要因として働く可能性が示唆された。

主要な結論は1995年SSM調査男性票のデータにのみ基づいて導かれたものである。女性との比較, 時代を通じての比較, 他の調査との比較などの課題が残されている。

参 考 文 献

- 1995年SSM調査研究会 編(1996). 1995年SSM調査コード・ブック, 1995年SSM調査研究会.
統計数理研究所(1994). 国民性の調査第9回全国調査—1993年全国調査—, 統計数理研究所研究リポート, No. 75.

MCMC法に基づくランダムネットワークの推論

種村正美

与えられた点配置データがランダムなネットワーク構造から派生した場合、元のネットワーク構造をデータから推論する問題を考える。このような問題は現実には、材料科学、医学、地球科学などで現れる重要な統計的問題であり、画像解析や不均質点過程における事象の空間構造の探索など幅広い問題と密接な関連がある。

われわれはまず、観測データの背景となるランダムネットワークをVoronoi分割ネットワーク（以下、Voronoiネットワーク）で表す。これは、比較的少数個のパラメータでネットワーク構造を表現できるモデルであり、必要なら重み付きVoronoi分割などを考慮することによって柔軟性に富んだモデル化が可能であるためである。

今、観測データを x とし、Voronoiネットワークの母点を y で表すことにする。そして、 x の各々はVoronoiネットワークのある辺上の点から確率過程によって発生したものとし、具体的には観測点 x と辺上の点 z との誤差が分散 σ^2 の正規分布に従うと仮定する。以上の設定によって、 y の事前分布 $p(y)$ および z の条件つき事前分布 $p(z|y)$ 、そして x の尤度関数 $p(x|z, y)$ を用いると、事後分布 $p(y, z|x, \sigma)$ が構成される。 y, z の事前分布としてPoisson分布を仮定すると、 z に関する積分が実行できて周辺事後分布 $p(y|x, \sigma)$ が求められる。

以上のBayes型枠組みに基づいて、MCMC (Markov Chain Monte Carlo) 法によって、観測データ x に対応して (y, σ) の系列を計算機実験で実現し、Voronoiネットワークの構造と正規分布のパラメータ σ を推定することが可能になる。われわれは、2次元の人工データに対して、このBayes型アプローチが実際に作動することを確かめた。今後は、実データの解析を2次元、3次元の場合を含めて行うことが課題である。

なお、本研究はE. Vedel Jensen (Aarhus Univ.) およびJ. Møller (Aalborg Univ.) との共同研究である。

パソコンUNIXによる統計学研究環境

丸山直昌

IBM-PC互換パソコンにUNIX (FreeBSD. 2.1.5 および2.2.5) 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行なっている。

FreeBSDの世の中への普及に伴い、システムとしてはかなり整備されたものが出回るようになり、利用環境としては整ってきた。SやNetscapeなど、人気が高い有償のソフトウェアもFreeBSD上で利用可能になっている。しかし使い込んでみると、統計数理研究所で従来開発されてきたソフトウェアをFreeBSD上で利用する場合の問題点も見つかった。今回数学関数ライブラリに見つけた問題点を取り上げ、その詳細と改修方法について述べた。

Statistical Distribution Theory in Markov Chains

内 田 雅 之

オーダー k の離散確率分布において、信頼性の分野と密接に関係があるものの1つにオーダー k の二項分布があげられる。さらに、オーダー k の二項分布の分布論として興味深い点は連の数え方によってその分布が異なることである。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が、高次のマルコフ従属性を仮定したモデルの下での研究は少ないと思われる。本年度の研究の要旨は以下の通りである。

$X_{-m+1}, X_{-m+2}, \dots, X_0, X_1, X_2, \dots, X_n$ 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、確率変数列 X_1, X_2, \dots, X_n において、長さ k の1の連の起こる回数の分布の確率母関数の漸化式を与え、さらに n についての母関数を考えることにより、オーダー k の二項分布のマルコフ連鎖バージョンの厳密分布を二重母関数の形で導出することができた。さらに、連の数え方はちょうど長さ l の連を数える数え方、長さ l 以上の連を数える数え方、長さ l の連を重複しないで数える数え方、そして長さ l の連を重複して数える数え方の4つの数え方を用いたが、それらのすべての分布が統一した形式で表現できることもわかった。

参 考 文 献

- Uchida, M. (1998). Statistical distribution theory of runs in higher-order Markov dependent trials, Ph. D. Dissertation, Osaka University.

健診情報の可視化について

駒 澤 勉

研究生活の大半を費やした「ヒトの健康に関する予防」の為の研究成果及び実践応用について報告した。

基礎研究は1963年に東京慈恵会医科大学第四内科の動脈硬化研究班への統計数理の立場での支援研究に始まり、以後、東邦大学医学部の臨床生理機能学研究室、ME学研究室、第二検査部、(財)日本労働文化協会・健診事業部や(財)動脈硬化予防研究会など多くの健康科学領域との共同研究に展開してきた。その成果は高度情報化社会にふさわしい健康の予知・予防の為の集団健診システムを構築し、実践の場で役立っている。

詳細は平成10年度の統計数理、第46巻1号、特集「ヘルスサイエンスと統計科学」の研究詳解「健診データ処理システムの構築」、平成9年度の公開講演会要旨「循環器系機能特性の計測モデル」に記載されているので、ご参照下さい。

事象発生時刻の不確定な点過程の推論

尾形 良彦

1. 定常更新過程の厳密な対数尤度

活断層のトレンチ調査に基づく古地震や歴史地震の発生時刻データに更新過程モデルを当てはめて地震発生強度や長期区間の確率予測がなされている。通常はイベント間隔の確率分布に基づく(近似)尤度を最大化すること(最尤法)によってパラメタを推定するが、これは、特に地震数が少ない場合には、情報を十分に使っていない。厳密尤度を定義するためには観測開始点から最初のイベントまでの時間および最後のイベントから現在までの時間に関する分布を考慮する必要があった。

2. 不確実性を持つ発生時刻を含むデータに対する尤度関数

トレンチ調査に基づく地震発生の推定時刻は堆積地層の年代推定に基づくので、多かれ少なかれ幅をもった時間区間(または不確実性を示す何らかの分布)として与えられる場合が多い。通常はこの区間の中央値などを発生時刻データと見做して更新過程などの統計モデルを当てはめるが、本報告では区間などのデータの不確実性に忠実な推定方式を提案した。

すなわち時間区間や不確実性を示す分布をそのような発生時刻の事前分布とし更新過程の尤度と合わせて事後分布を考えた。不確実な発生時刻に関するこの事後分布の積分(ベイズモデルの積分尤度)を最大化することによって更新過程のパラメタを推定値とし、最大積分尤度を使ってABICを定義し、異なる分布を持つ更新過程同士のモデル比較をすることができる。かくして推定された更新過程のモデルのもとで発生時刻の不確実性(区間)の精度を周辺事後分布として絞り込むことが出来た。

キーワード：ABIC, 更新過程, 厳密尤度, 積分尤度, 事後分布, 古地震。

平滑化に関する考察

柏木 宣久

無作為標本抽出と不偏推定は標本調査の基礎的要件とされてきた。母集団平均を調査の知見目的とする限り、こうした要件を確保する努力は依然として重要である。一方で、不可避的な事情により、こうした要件を十分に確保できない場合もある。そうした場合、しばしば、事前情報を用いて推定精度を向上させる、あるいはデータから情報を抽出するといった視点が有用となる。本報告では、そうした視点から構成された平滑化に基づく推定法と、その標本調査データへの適用可能性について述べた。