

## 平成 6 年度研究報告会要旨

と き : 1995 年 3 月 15 日 午前 10 時 ~ 午後 4 時 30 分  
16 日 午前 10 時 ~ 午後 5 時  
と ころ : 統計数理研究所 講堂

### プログラム

3 月 15 日

あいさつ

所 長 清水 良一

#### 統計基礎研究系

離散確率分布と統計の応用	平野 勝臣
無限次元マルチンゲールの中心極限定理	西山 陽一
パラメトリックなモデルの構築について	松縄 規
マルチンゲールの漸近展開について	吉田 朋広
緩慢変動する truncated moment をもつ確率変数の独立積とその応用	志村 隆彰

#### 統計教育・情報センター

分子系統樹推定法の諸問題	長谷川 政美
原生物の系統と進化	橋本 哲男
指数演算子の分解による Fokker-Planck 方程式の一般化	岡崎 卓
年金アクチュアリーに関するリスクシミュレーション	山下 智志

#### 領域統計研究系

統計的推測における検定の役割再考	柳本 武美
ロジスティックモデルでの条件付きスコア検定	佐藤 俊哉
Jackknife 信頼区間と bootstrap 検定	汪 金芳
結晶群の出現頻度と群の上のランダムウォーク	伊藤 栄明
文献の計量分析	村上 征勝
7ヶ国国際比較調査の総合報告	吉野 諒三
同性間の性愛に関する法律の国際比較分析	釜野 さおり
射影による点集合の近似の限界	(客員, 琉球大学) 前原 潤
工業データにおける因子分析の適用	(客員, 大阪電気通信大学) 猪原 正守

#### 統計データ解析センター

モンテカルロフィルタによる季節調整	北川源四郎
-------------------	-------

語順規則の2パラメータモデル  
 統計科学と並列計算  
 吹雪時における気象要素の伝播に関する統計的解析

上田 澄江  
 田村 義保  
 荒畑恵美子

3月16日

予測制御研究系

神奈川県内製造業者へのアンケートとその解析  
 モンテカルロフィルタの実装について  
 A Likelihood Analysis of Chaos  
 不完全情報下における制御系設計に関する研究  
 最適化の微分幾何学  
 内点法の局所的収束性について  
 動的システムの最適制御  
 男女差の分析  
 内点法に対する新しい探索方向について  
 マルコフ連鎖モンテカルロ法の改良  
 大気ゆらぎの時空間モデル

川崎 能典  
 樋口 知之  
 尾崎 統  
 宮里 義彦  
 田辺 國士  
 水野 眞治  
 伊藤 聡  
 鈴木義一郎  
 土谷 隆  
 伊庭 幸人  
 石黒真木夫

調査実験解析研究系

いくつかの調査に参加して  
 継続調査データの蓄積とコウホート分析  
 国民性調査への共分散構造分析の適用  
 非球粒子系のモンテカルロ法  
 パソコン UNIX による統計学研究環境  
 3母数逆ガウス型分布について  
 グラフィックスによる数量化分析  
 非計量データの解析  
 構造変化を探る等質性分析について  
 統計情報資源の利用環境について  
 南勢町コホート研究における生活習慣調査  
 点過程コホート解析  
 平滑化による解析

坂元 慶行  
 中村 隆  
 前田 忠彦  
 種村 正美  
 丸山 直昌  
 金藤 浩司  
 駒澤 勉  
 馬場 康維  
 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

## 離散確率分布と統計的応用

平野 勝 臣

## 本年度の研究

(1) 離散確率分布の統計的応用の研究を行った。Consecutive- $k$ -out-of- $n$ :F システム (Con/ $k$ / $n$ /F と略す) の信頼性と確率分布との関係を調べ、成分に寿命分布を仮定したとき、Con/ $k$ / $n$ /F システムの寿命分布を調べた。またシステムの故障時刻を観測したとき、母数の推定問題についてまとめた (Aki and Hirano(1995))。

(2) オーダー  $k$  の離散確率分布に関する研究を行い、いくつかの結果をまとめた。この研究は来年度に継続される。

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

Consecutive- $k$ -out-of- $n$ :F システムの母数推定

$n$  個の成分から成るシステムにおいて、連続した少なくとも  $k$  個の成分が故障したとき、システムが機能しなくなるシステムを Con/ $k$ / $n$ /F システムという。各成分の信頼性を  $p$  (故障の確率を  $q = 1 - p$ ) とすると、Con/ $k$ / $n$ /F システムの信頼性は  $p$  の関数で  $h(p)$  とかける。一般的に  $h(p)$  は複雑で、値を求めることに適さない。さらに  $p$  が時刻  $t$  に依存するとき、このようなモデルの解析には適切な式が必要となる。

Con/ $k$ / $n$ /F システムの各成分の lifetimes を  $\xi_1, \xi_2, \dots, \xi_n$  とし、独立同分布に従うとする。この順序統計量を  $\xi_{(1)} \leq \dots \leq \xi_{(n)}$ ,  $\xi_{(i)}$  の cdf を  $G_{(i)}(t)$  とかく。また Con/ $k$ / $n$ /F システムの lifetime を  $T$ , システムの最小  $m$ -cutsequences の数を  $r_{mk}$  とする。このとき

$$P(T \leq t) = \sum_{i=1}^n \{r_{ik}/(n(n-1)\cdots(n-i+1))\} G_{(i)}(t)$$

である。ここに  $r_{mk} = (n-m+1)(m-1)! \cdot N_k(n, m-1) - m! \cdot N_k(n, m)$  で

$$N_k(n, m) = \sum_{i=0}^{\lfloor n/k \rfloor} (-1)^i \binom{n-m+1}{i} \binom{n-ki}{n-m}.$$

この結果を用いて、成分の故障の確率が時間経過とともに変わるモデルの解析を行った。本報告ではその解析事例を示すことが目的であり、Con/ $k$ / $n$ /F システムの寿命分布とその母数推定の議論を示した。

成分が指数分布、ワイブル分布、パレート分布に従うとき、Con/ $k$ / $n$ /F システムの failure times  $T_1, \dots, T_m$  の観測に基づいて、母数の推定について議論した。いずれの場合でも、最尤推定は少し無理があるので、モーメント法で推定した。システムの寿命分布は成分の順序統計量の分布の一次結合であることから順序統計量の結果を利用でき、積率が求まる。数値計算の結果から指数分布の場合では、モーメント法は極めて高い効率を持つことが分かった。

## 参 考 文 献

- Aki, S. and Hirano, K. (1995). Lifetime distribution and estimation problems of consecutive- $k$ -out-of- $n$ -F:systems, Research Memo., No.539, The Institute of Statistical Mathematics, Tokyo.

## 無限次元マルチンゲールの中心極限定理

西山陽一

Aalen (1978) は, 生存解析においてマルチンゲールを用いたアプローチが極めて有力かつ自然であることを示し, 同分野が 1980 年代に爆発的に発展する契機を与えた. それらの成果を豊富な具体例を交えながら系統的に解説した文献として Andersen et al. (1992) が挙げられるが, その中で繰り返し用いられる数学的道具の 1 つが「マルチンゲール中心極限定理」である.

これを踏まえ, 次のような研究を行った.  $\Psi$  は (無限個の要素を含むことを許された) 任意の集合であるとし,  $X = \{X^\psi | \psi \in \Psi\} = \{(X_t^\psi)_{t \in \mathbb{R}_+} | \psi \in \Psi\}$  は  $\Psi$  の要素によって添字付けられたマルチンゲールの族であるとする (特に  $\Psi = \{1, 2, \dots, d\}$  ならば  $X$  は  $d$ -次元マルチンゲールである). このようなものの列  $X^n$  ( $n = 1, 2, \dots$ ) が幾つかの条件のもとで正規確率場に分布収束することを証明した. 特に時間パラメータ  $\mathbb{R}_+$  を離散的な  $\mathbb{N}$  に取り替えることによって, 確率場の従属列に対する中心極限定理が得られるが, これ自身も応用上有益な結果である.

この「無限次元マルチンゲール中心極限定理」の統計的応用として, マーク付き点過程に対するノンパラメトリック推測問題を考察した. 特に Aalen の推定量を一般化したものに対する汎漸近正規性を示した. また時刻とマークの独立性を検定するための統計量を 3 つ提案し, それらが正規確率場の汎関数に分布収束することを示した.

## 参考文献

- Aalen, O. (1978). Nonparametric inference for a family of counting processes, *Ann. Statist.*, **6**, 701-726.  
 Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1992). *Statistical Models Based on Counting Processes*, Springer, New York.  
 Nishiyama, Y. (1995a). A central limit theorem for infinite dimensional local martingales, Research Memo., No.542, The Institute of Statistical Mathematics, Tokyo.  
 西山 陽一 (1995b). マーク付き点過程のノンパラメトリック推測に関するノート, 推定論とその応用の研究, 統計数理研究所共同研究レポート, No.71, 77-82.

## パラメトリックなモデルの構築について

松縄規

これまでに得ていた結果を改善し, 連続, 離散のいずれにも適用できる, 一変量パラメトリックモデルの構築を, いわゆる最尤法を修正して特定化方程式を与えることによって行った. これに基づき代表的な一変量分布を組織的に誘導出来ることを示した.

$x_1, \dots, x_n$  を  $n$  個の i.i.d. 実ランダム数列,  $\theta$  を我々が興味を持つ主要なパラメータ,  $\tau$  を  $\theta$  と関数的に独立な潜在パラメータとする. モデルが密度  $p(x; \theta, \tau)$  を持つとして, これがあるいくつかの正則条件と次の諸条件を満たしているものとする:

$$\sum_{i=1}^n \partial \ln p(x_i; \theta, \tau) / \partial \theta = 0 \quad \text{および} \quad \sum_{i=1}^n \Delta(x_i; \theta, \tau) = 0.$$

但し  $\Delta(x_i; \theta, \tau)$  は観測対象とモデルの間の観測誤差関数を意味し, 分布生成に重要な役割を持つ.

この時パラメトリックモデル分布の特定化方程式が次のように与えられる:

$$\frac{\partial \ln p(x; \theta, \tau)}{\partial \theta} = \xi(\theta, \tau) \Delta(x; \theta, \tau).$$

ここで  $\xi(\theta, \tau)$  は  $\theta, \tau$  のある関数で観測の精度を表す.

また, 上記特定化方程式を満たすパラメトリックモデルの密度は次の様に与えられる:

$$p(x; \theta, \tau) = C^*(x; \tau) \cdot \exp \left[ \int \xi(\theta, \tau) \Delta(x; \theta, \tau) d\theta \right],$$

ここに  $C^*(x; \tau) > 0$  は分布の規格化関数を表す.

以上の結果から, 様々なパラメトリックモデルを与えることが出来る. ここでは離散型のモデルの代表的な分布の構築についてのみ記す:

$$\Delta(x; \theta) = x - \frac{h(\theta) k'(\theta)}{h'(\theta) k(\theta)}, \quad \xi(\theta) = \frac{h(\theta)}{h'(\theta)}$$

と設定すると modified power series 分布を得る:

$$p(x; \theta) = C(x) \cdot \frac{[h(\theta)]^x}{k(\theta)}, \quad (x \in T; h(\theta), k(\theta) > 0)$$

但しここで

$$h(\theta) = \sum_{x \in T} C(x) \cdot [h(\theta)]^x.$$

上の  $h(\theta), k(\theta)$  を適切に選択し  $C(x)$  を定めることにより, 一般化された, 負の二項分布, ポアソン分布, 対数級数分布, など多数の離散分布を誘導できる.

### マルチンゲールの漸近展開について

吉田 朋 広

Wiener 空間上で定義されたマルチンゲールのまわりの確率展開をもつ確率変数  $X_n$  の (局所) 確率密度の漸近展開を示した. これによって, 任意の有界可測関数  $f$  に対して平均値  $E[f(X_n)]$  の漸近展開が得られる. 独立観測の場合に確率変数の分布の滑らかさを保証する条件として, たとえば Cramér 条件が仮定されるが, ここでは Malliavin 共分散の非退化性の条件が本質的な役割を演じる. 応用として, エルゴード的な拡散過程の汎関数に対する中心極限定理の精密化が得られる.

### 参考文献

- Yoshida, N. (1994). Malliavin calculus and asymptotic expansion for martingales I, II, Research Memo., No.504, 517, The Institute of Statistical Mathematics, Tokyo.  
 Yoshida, N. (1994). Asymptotic expansion for local density of martingales, Research Memo., No.520, The Institute of Statistical Mathematics, Tokyo.

## 緩慢変動する truncated moment をもつ確率変数の独立積とその応用

志村 隆 彰

$\alpha$  次の truncated moment  $\int_0^x t^\alpha \mu(dt)$  が緩慢変動するような非負実数上の分布の族を  $M(\alpha)$  とし、これに関連する分布の Mellin-Stieltjes convolution (確率変数で独立積の分布に相当するもの) とその意味での分解について考える。これは Maller (1981) の正規分布の吸引域 (truncated variance  $\int_0^x t^2 \mu(dt)$  の緩慢変動性で特徴付けられ、以下、 $D_2$  と書く) についての結果を正則変動関数論、特に非減少緩慢変動関数の分解の応用という立場から考察したもので、彼の結果を拡張しその予想に答えたものである。彼は  $D_2$  の分布に従う確率変数の独立積の分布もまた  $D_2$  に属すること、逆に独立積の分布が  $D_2$  に属し、一方の分散が有限ならばもう一方は  $D_2$  に属するということを示し、さらに独立積の分布が  $D_2$  に入っているならば、各確率変数の分布も  $D_2$  に属すると予想している。また彼はこれを線形回帰モデルについて応用している。2つの分布  $\mu, \nu$  の Mellin-Stieltjes convolution を  $\mu \circ \nu$  とかき、 $F(\alpha), S(\alpha), C(\alpha)$  はそれぞれ  $\limsup_{x \rightarrow \infty} \int_x^{2x} t^\alpha \mu(dt) < \infty$ ,  $\lim_{x \rightarrow \infty} \int_0^x t^\alpha \mu(dt) / \int_0^x t^\alpha \mu(dt) = 1$ ,  $\limsup_{x \rightarrow \infty} \int_0^x t^\alpha \mu(dt) / \int_0^x t^\alpha \mu(dt) < \infty$ , を満たす分布族を表すものとする。得られた結果は以下の様なものである。まず、 $M(\alpha)$  に属する  $\mu \circ \nu$  の truncated moment とその factor ( $\mu$  と  $\nu$ ) の truncated moment の積に関する基本的な評価  $\limsup_{x \rightarrow \infty} \int_0^x t^\alpha \mu \circ \nu(dt) / (\int_0^x t^\alpha \mu(dt) \int_0^x t^\alpha \nu(dt)) \leq 1$  を証明した。分解については  $F(\alpha)$  あるいは  $S(\alpha)$  ならば、全ての factor は  $M(\alpha)$  に属すること (特に、 $S(\alpha)$  は Mellin-Stieltjes convolution について閉じたクラスであり、 $\mu \circ \nu$  の truncated moment と  $\mu$  と  $\nu$  の truncated moment の積は漸近的に等しい)、及び、 $\mu \in M(\alpha)$  が発散する  $\alpha$  次モーメントをもち、 $\nu \in C(\alpha)$  ならば  $\mu \circ \nu$  は  $M(\alpha)$  に入ることを示した。 $M(\alpha) \setminus C(\alpha) \neq \emptyset$  であるから、最後の結果は Maller の予想に対する反例を与えることになる。

### 参考文献

- Maller, R.A. (1981). A theorem on products of random variables, with application to regression, *Austral. J. Statist.*, **23**, 177-185.  
 志村 隆彰 (1995). 緩慢変動する truncated moment をもつ分布の独立積とその応用, 統計数理研究所共同研究レポート, No.75, 1-9.

### 統計教育・情報センター

#### 原生生物の系統と進化

橋本 哲男

原生生物界の系統関係を明らかにすることは、真核生物の初期進化の過程を解明するうえで極めて重要である。従来この分野の分子系統学的研究は、主として小亜粒子リボソーム RNA (SrRNA) の塩基配列比較に基づいて行なわれてきた。ところが、SrRNA の塩基組成値は生物種間で大きく偏る傾向が強いため、その偏りの影響により SrRNA の系統樹が誤りを犯しているとの可能性は否定できない (Hasegawa and Hashimoto (1993), Hashimoto et al. (1994))。そこで、各種保存的蛋白質のアミノ酸配列データに基づいて原生生物界の系統関係を再検討することを目的として研究を行なった。

本年度遺伝子解析の対象とした原生生物種は、ミトコンドリアをもたない寄生虫である *Giardia lamblia* (ランブル鞭毛虫) [Diplomonads], *Entamoeba histolytica* (赤痢アメーバ) [Entamoebidae],

*Glugea plecoglossi* (アユ寄生虫) [Microsporidia] の3種, ミトコンドリアを有するが解析上その位置づけが重要となる *Trypanosoma cruzi* (Kinetoplastids), および, チトクロームをもたないミトコンドリアを有するとされる寄生虫 *Blastocystis hominis* の5種である. 対象とした保存的蛋白質は, ペプチド鎖伸長因子 EF-1 $\alpha$  および EF-2, イソロイシル tRNA 合成酵素 (ileRS) などであり, 高度保存部分に対して設計した縮重オリゴヌクレオチドプライマーを用いる PCR 法によりこれらの遺伝子の部分配列を増幅しその配列を決定した. また, 一部については cDNA の単離, 配列決定も行った. 推定アミノ酸配列データをもとに, アミノ酸置換に関するさまざまな確率モデルに基づいて真核生物内部の系統関係を最尤法により検討した結果, 以下のことが明らかとなった.

1. Microsporidia に属するアユ寄生虫 *Glugea* の EF-1 $\alpha$  は, 明らかに真核生物型の特徴を示しているが通常の真核生物のものに比べると極めて特異的である. EF-1 $\alpha$  の解析結果は, *Glugea* の分岐が *Giardia* よりもさらに早いとの可能性の高いことを示唆している. もしこれが真実であるとする, *Glugea* はミトコンドリアの細胞内共生が起こる以前の真核生物の祖先型により近い生物であるということになる.
2. EF-1 $\alpha$ , EF-2 の解析はともに, *Giardia* の分岐のすぐ後に *Trypanosoma* が分岐したとする SrRNA の一般的な解析結果を支持せず, *Trypanosoma* の分岐がミトコンドリアをもたない *Entamoeba* の分岐よりも後であるとの可能性の高いことを示唆している. ileRS や他の保存的蛋白質を含めて総合評価すると, (outgroup, (*Giardia*, (*Entamoeba*, (*Trypanosoma*))) という関係が尤もらしいとの結論が得られる.
3. EF-1 $\alpha$ , EF-2, ileRS のアミノ酸組成値は生物種間でほとんど偏っていない. このことは, 保存的蛋白質に基づく解析が SrRNA に基づく解析に比べ, より頑健な推定を与えることを示唆している.

## 参考文献

- Hasegawa, M. and Hashimoto, T. (1993). Ribosomal RNA trees misleading?, *Nature*, **361**, p. 23.  
 Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K. and Hasegawa, M. (1994). Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*, *Molecular Biology and Evolution*, **11**, 65-71.

## 指数演算子の分解による Fokker-Planck 方程式の一般化

岡崎 卓

### 1. 初めに

外乱によって変動する系変数の確率密度を定める一般化 Fokker-Planck (GFP) 方程式は, 非ガウス外乱に対しても有効であることにその一般性をもつが, 実用に供するには Fokker-Planck 作用素  $i\hat{L}t$  と射影子  $\mathbf{p}$  からなる指数演算子

$$e^{i\hat{L}(1-\mathbf{p})t}$$

の具体的表現が必要である。しかし、通常の摂動展開と繰込み近似に訴えるには限界があり、flatness等、高次の非ガウス性を扱うことができない。

そこで、quantum Monte Carlo simulationをはじめ、物理に関する多くの分野で成果を挙げている「指数演算子の分解」(例えば、Suzuki(1992))を適用し、高次の非ガウス性を反映する可能性を失わずに GFP 方程式の拡散係数を簡約する方法について報告する。

## 2. 指数演算子 $e^{\hat{L}(1-\mathbf{p})t}$ の分解と拡散係数の表現

Fokker-Planck 作用素  $i\hat{L}$  は外乱  $W$  による拡散、系変数  $U$  の移流、およびその相互作用を表す 3 項からなる；  $i\hat{L} = i\hat{L}_W + i\hat{L}_U + \mu(U, W)D_U$  ( $i\hat{L}_W = N(W)D_W + \frac{\sigma^2}{2}D_W^2$ ,  $i\hat{L}_U = M(U)D_U$ ,  $D_W = \frac{\partial}{\partial W}$ ,  $D_U = \frac{\partial}{\partial U}$ )。ここで

$$\begin{aligned} A &= (i\hat{L}_W + i\hat{L}_U)(1 - \mathbf{p}) \\ B &= \mu D_U(1 - \mathbf{p}) \end{aligned}$$

と置けば、 $e^{\hat{L}(1-\mathbf{p})t}$  は 2 次の近似で次のように分解される。

$$e^{\hat{L}(1-\mathbf{p})t} = e^{t(A+B)} = e^{\frac{1}{2}tA} e^{tB} e^{\frac{1}{2}tA} + O(t^3)$$

外乱  $W$  の統計的定常性および射影子  $\mathbf{p}$  に関する各種の性質を用いて  $e^{tA}$  を表現し、さらに  $e^{tB} = e^{\mu D_U t}(1 - \mathbf{p}) + \mathbf{p} + O(\mu^2)$  なる近似を施せば、GFP 方程式で中心的役割を担う拡散係数を

$$\begin{aligned} D(u, v, t) &\equiv \text{tr}\{g(W)\mu\psi_v(1 - \mathbf{p})e^{\hat{L}(1-\mathbf{p})t}\mu\psi_u\} \\ &= \text{tr}\{g(W)[(1 - \mathbf{p})\mu\psi_v]e^{\hat{L}t}(1 - \mathbf{p})\mu\psi_u\} \end{aligned}$$

と表わすことができる ( $\psi_v = \delta(U - v)$ ,  $\text{tr}\{\dots\} = \int dW \int dU$ )。指数関数上の  $(1 - \mathbf{p})$  なる因子が指数関数の外に移動したことは、 $\exp$ (微分作用素) 型の演算を実行する通常の手段を利用して拡散係数を具体的に計算し得る可能性を示すものである。また、時間進行演算子  $e^{\hat{L}t}$  が外乱と系変数との相互作用項を含んでいることから、拡散係数には  $\langle \mu(U, W) \cdot \mu(U, W(t)) \rangle$  以外の高次相関が関与し、従って 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。

## 参考文献

- Suzuki, M. (1992). General nonsymmetric higher-order decomposition of exponential operators and symplectic integrators, *J. Phys. Soc. Japan*, **61**, 3015-3019.

## 年金アクチュアリーに関するリスクシミュレーション

山下 智志

年金 ALM では、資産・負債双方の将来の推移をいかに記述するかが最大の問題になる。しかし、これまで行われてきた年金 ALM に関する研究は、資産サイドの将来の推移を表現することに重点が置かれており、負債サイドの詳細な分析は行われていない。これは、負債サイドの年金数理計算は多くの制度的制約があり、資産サイドのような統計的アプローチが用いられないことに原因がある。このことは、年金 ALM 問題を多期間最適化問題としてモデル化することを困難にしているなど、多くの実務的な問題を発生させている。



そこで本研究では、年金数理計算の本質に回帰し、年金数理計算の外生変数(基礎率等)がALMにたいしていかに影響を及ぼしているかを実験的に解明する。そしてその結果を用い、実用的な多期間最適化モデルに挿入できる簡便的な負債モデルを構築することを目的とする。具体的には以下の分析を行った。

- 1) 年金ALMの負債モデルの外生変数の性質について、ALMに対する影響の視点から分析・整理する。
- 2) 資産モデルを構築する。その際年金ALMの特徴をより正確に反映するよう、従来年金ALMで用いられていた時系列モデルに修正を加える。
- 3) 資産・負債モデルを結合し、モンテカルロシミュレーションを行う。その結果負債モデルの外生変数と掛け金率の関係について把握することができる。また、この結果について解析的分析と比較する。
- 4) 以上の結果を踏まえ、年金ALMモデルの修正・簡略化を行い、資産選択等にも利用可能なモデルを構築する。

これらの分析をふまえ、年金数理計算の性質について統計的に考察を加えた。また、多期間無限期間の資産選択モデルに対応する年金数理計算式の定式化を行った。

### 参考文献

- Daykin, C.D., Pentikainen, T. and Pensonen, M. (1994). *Practical Risk Theory for Actuaries*, Monographs on Statistics and Applied Probability 53, Chapman & Hall, London.
- 矢頭 智夫, 山下 智志(1995). 年金ALMにおける感度分析と掛け金率計算の簡略化, Research Memo., No.545, The Institute of Statistical Mathematics, Tokyo.

## 領域統計研究系

### ロジスティックモデルでの条件付きスコア検定

佐藤 俊哉

疫学研究では、リスク要因への曝露と疾病発生との関係を調べるためにケース1例に対し、性・年齢などをマッチした1~数例のコントロールを取る研究が行われる(一般にはケースもコントロールも複数)。疾病発生リスクにロジスティックモデル、

$$\log \frac{p(x)}{1-p(x)} = \alpha + \beta x$$

を仮定すると、ケース・コントロール研究では曝露効果を示す $\beta$ を推定することができる。「曝露によって疾病発生リスクは変わらない」という帰無仮説は「 $\beta = 1$ 」に相当し、曝露が2値の場合、ロジスティックモデルにもとづく条件付きスコア検定が<sup>s</sup>Mantel-Haenszel 検定である。

曝露が連続量の場合に、条件付き尤度(Breslow and Day(1980), p.205)から帰無仮説「 $\beta = 1$ 」のもとでのスコア検定統計量は、

$$Z_s = \frac{\sum_{k=1}^K \frac{n_k m_k}{N_k} (\bar{x}_{1k} - \bar{x}_{0k})}{\sqrt{\sum_{k=1}^K \frac{n_k m_k}{N_k} s_k^2}}$$

と導くことができる。ただし、 $\bar{x}_{1k}$  はケースの平均曝露量、 $\bar{x}_{0k}$  はコントロールの平均曝露量、

$$s_k^2 = \frac{\sum_{l=1}^{N_k} (x_{lk} - \bar{x}_k)^2}{N_k - 1}$$

はケースとコントロールを込みにした曝露量の不偏分散、 $N_k = n_k + m_k$  である。各層でのケースとコントロールの数  $(n_k, m_k) \Rightarrow \infty$  または層の数  $K \Rightarrow \infty$  のとき、スコア統計量は漸近的に正規分布に従う。

ロジスティックモデルの条件付き尤度は、第  $k$  層を第  $k$  番目の死亡、 $N_k$  を死亡時刻直前のリスク集合に読みかえると、Cox 回帰の部分尤度に一致することが知られている。曝露が 2 値の場合には、Mantel-Haenszel 検定は logrank 検定であり、このことから曝露が連続量の場合、上で与えたスコア検定統計量を用いることができる。

### 参考文献

Breslow, N.E. and Day, N.E. (1980). *Statistical Methods in Cancer Research, Vol. 1 — The Analysis of Case-Control Studies*, Oxford University Press, New York.

### Jackknife 信頼区間と bootstrap 検定

汪 金 芳

本報告は二つの研究対象に関する考察が含まれている。まず、滑らかな母数のジャックナイフ信頼区間の構成について考察を行なう。即ちある確率ベクトル  $X$  の平均  $\eta$  の滑らかな関数  $\theta = h(\eta)$  の信頼区間の構成が問題である。  $X$  の分布に関する情報が特になければ、通常ノン・パラメトリック plug-in 推定量  $\hat{\theta} = h(\bar{X})$  を用いて  $\theta$  を推定する。ここで、 $\bar{X}$  は標本平均を表す。標本相関係数で母集団相関係数を推定する場合が 1 つの例である。これらの推定量はバイアスを持ち、分散の推定も難しい。より高次のバイアスを持つ、かつ漸近的に pivotal な統計量

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

は、ジャックナイフ法を用いて、簡単に構成できる。ここで  $\hat{\theta}_J$  は  $\theta$  のジャックナイフ推定量であり、 $\hat{\sigma}_J^2$  は  $\sqrt{n}\hat{\theta}_J$  の分散のジャックナイフ推定量である。  $T$  に基づく  $\theta$  の信頼区間は、解析的に難しいため、従来  $T$  の極限分布あるいは 2 次の Edgeworth 近似を用いて構成されている。Wang(1994) は  $\theta$  が多次元の場合について  $T$  に関する 3 次までの漸近理論を展開した。本報告は、その応用として、 $\theta \in \mathbf{R}^1$  のよりよい信頼区間の構成に関するものである。

具体的に、 $n$  を標本数とするときの  $T$  に基づく

$$\text{Prob}(\theta \in \hat{I}) = \alpha + O(n^{-3/2})$$

を満たすような信頼区間  $\hat{I}$  を構成したい。ただし、 $\alpha$  を名目上の被覆確率とする。これは  $T$  に関する 3 次までの Edgeworth 展開の逆展開

$$\text{Prob}\left[T \leq t - \frac{1}{\sqrt{n}}\hat{Q}_1(t) + \frac{1}{n}\left(\hat{G}_2(t) + \frac{1}{2}t\hat{a}(t)\right)\right] = \Phi(t) + O(n^{-3/2})$$

を与えることによって解決できる。以上の漸近理論を具体的な場合に適用し、高次漸近理論に基づく推測の良さも確認されている(汪 他(1994a, 1994b))。

次に、2標本問題におけるブートストラップ- $t$ 検定について簡単に報告する。ブートストラップ法は観測されるデータに基づく条件つき推測法である。これはブートストラップ法の統計的検定における難点でもある。それは仮説検定では検定統計量の帰無分布を推定しなければならないからである。一方観測されるデータは、対立仮説の決まる分布から発生される。このようなデータから、如何に異なる分布に関する情報(帰無分布)を抽出するのかが問題となる。Naiveブートストラップ法を修正しなければ、帰無分布の推定はできない。Efron and Tibshirani(1993)の修正案も、満足できる精度までは改善していない。それに対する改良策はmixing法である。詳細は、汪・田栗(1995)を参照されたい。

### 参考文献

- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall, London.  
 Wang, J.F. (1994). Jackknife methods and higher order asymptotic expansions, 京都大学数理解析研究所講究録, **879**, 120-144.  
 汪 金芳, 田栗 正章(1995). 2標本問題におけるブートストラップ- $t$ 検定, 京都大学数理解析研究所講究録(予定).  
 汪 金芳, 大内 俊二, 田栗 正章(1994a). ジャックナイフ統計量に関する高次漸近的性質の検討—相関係数の場合, 京都大学数理解析研究所講究録, **879**, 145-154.  
 汪 金芳, 田栗 正章, 大内 俊二(1994b). 変動係数に対するジャックナイフ- $t$ 推定とその数値的検討, 計算機統計学, **7**(1), 57-72.

### 結晶群の出現頻度と群の上のランダムウォーク

伊 藤 栄 明

3次元空間における周期的な構造の対称性を記述する230個の群があり、それらは空間群と呼ばれている。結晶における原子の配列を解析する際に空間群は用いられる。結晶の対称性は230の空間群のいずれかにより表される。幾何学的対称性を平行移動の操作を考慮に入れず記述する点群と言われている32個の群がある。230個の空間群は32個の点群のいずれかに基づいて構成されている。正六面体を不変に保つ合同変換からなる有限群で位数が最大であるものを考える。それは $O_h$ と呼ばれている群である。正六角柱の高さが一様な柱を不変に保つ合同変換からなる有限群で位数が最大であるものは $D_{6h}$ といわれている。32個の点群は $O_h$ ,  $D_{6h}$ 及びそれらの部分群のいずれかである。

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、空間群の分布を調べるという課題にとりくんできた。ICSDは、無機化合物の完全な構造情報に関するデータベースである。それには、化合物名・分子式・その化合物の基本構成要素である単位格子(Unit Cell)の大きさ・その中での各原子の座標・単位格子の広がり方を表す対称操作パラメータ、更に、各データが掲載されている文献の情報等が収録されている。結晶物質の対称性の統計的分布をもとめる際に種の定義が問題となる。ICSDには約25000件の結晶物質がある順番で登録されている。空間群、化学組成、単位格子パラメータ(単位格子における各軸の長さ $a, b, c$ , 各軸のあいだの角度 $\alpha, \beta, \gamma$ )のそれぞれについて近さを定義し、それらのすべてが、互いに近いときふたつの結晶物質は互いに近いと定義する。ICSDにおける登録順に種を逐次定義して行く。すなわち種 $1, 2, \dots, n$ が定義されているとき、つぎの結晶

物質はすでに定義されている  $n$  種のそれぞれと近くないときにかぎり結晶物質の種  $n+1$  と定義する。ICSD における登録の順番にもとづいてこのようにして種を逐次定義してゆく。このように定義したそれらの種の空間群、点群についての頻度をもとめる。この方法により、 $ANX$  記号ごとに空間群、点群の出現頻度をもとめた(藤原 他(1993))。  $ANX$  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。単位格子パラメータは結晶軸の選び方に依存し、結晶物質の対称性が高くない場合、その選び方は一意的でない。このことから本年度は単位格子パラメータのかわりに単位格子体積をもちいることにより、自然に種を定義してゆけることを示し空間群、点群についての出現頻度をもとめた(藤原 他(1994))。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。

### 参考文献

- 藤原 美也子, 伊藤 栄明, 松本 崧生, 武田 弘(1993). 無機結晶データベース (ICSD) を用いた結晶群の出現頻度, 統計数理研究所共同研究レポート, No.50.
- 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レポート, No.66.
- 伊藤 栄明(1985). 群, グラフを値としてとる確率分布, 統計数理, **33**, 68-70.
- 伊藤 栄明(1986). 幾何学的対称性の統計的分布—結晶における群とその頻度—, 統計数理, **34**, 19-27.
- Itoh, Y. and Matsumoto, T. (1991). Random-generation model for statistical distribution of point groups, *Acta Cryst.*, **A47**, 204-206.

### 文献の計量分析

村上 征勝

日本語の文献の文章を計量分析することによって、文献の真贋の判定を行ったり、著者、著作年、著作順序などの推定を行おうとする研究の歴史は浅く、まだ多くの未解決の問題が残されている。

そこで「源氏物語」を具体的な研究対象として、日本語の文章の計量分析法について研究を続けているが、計量分析を行うためには、計量分析に耐え得るような文章を単語に分別し、品詞情報を付加したデータベースが不可欠である。著者達は中央公論社の「源氏物語大成」(池田亀鑑 編著)を底本としたデータベース(約40万語)を構築したので、これを用いて、各巻での品詞の出現率、初出単語の比率、延べ語数に対する異なり語数の比率などの計量分析に必要となる基礎的な統計量を算出した。また、著者に関する問題を解決するため、「源氏物語」を模倣して書かれたとされる「手枕」、「雲隠六帖」及び「山路の露」を、成立過程に関する問題を解決するため「紫式部日記」及び「紫式部集」を入力し、単語分割し、品詞情報を付加する作業を行った。

なお、「源氏物語」に関しては、54巻に出現する約22万語の自立語に関し、語彙用例索引(約5500頁)を作り、出版した。

## 参考文献

- 村上 征勝(1994).『真贋の科学—計量文献学入門—』,朝倉書店,東京.  
 上田 英代,村上 征勝,今西 祐一郎,樺島 忠夫,上田 裕一 共編(1994).『源氏物語語彙用例  
 総索引 自立語編』,勉誠社,東京.

## 7ヶ国国際比較調査の総合報告

吉野 諒 三

文部省科学研究費補助金 特別推進研究(1)61060002「意識の国際比較方法論の研究—新しい統計的調査法の確立とその展開—」(昭和61年度~平成2年度)と文部省科学研究費補助金 試験研究A(1)045090001「意識の国際比較における連鎖的調査分析方法の実用化に関する研究」(平成4年度~平成6年度)による日,米,英,独,仏,伊,蘭の7ヶ国における大規模な国民性意識調査の計画,立案,遂行,データ収集,分析の各プロセスとその結果を,総合報告6分冊としてまとめ上げた.

特に,第1分冊には,「国際比較可能性」を追求した我々の研究パラダイム「連鎖的調査法」の理念,各国での標本計画,実施,翻訳に伴う問題(バック・トランスレーションの各プロセスの詳細),データ分析についてまとめた.第2分冊は,国際比較調査に興味を持つより多くの方々  
 に情報を提供するために「国民性の国際比較の為のマニュアル」を作成した.これらは,研究  
 リポート76~82としても発刊した.

## 参考文献

- 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—総合報告書—,統計数理研究所研究リポート, No.76.  
 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—国民性の国際比較のためのマニュアル—,統計数理研究所研究リポート, No.77.  
 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—イタリア調査の回答コードブック—,統計数理研究所研究リポート, No.78.  
 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—イタリア調査の自由回答アーター—,統計数理研究所研究リポート, No.79.  
 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—オランダ調査の回答コードブック—,統計数理研究所研究リポート, No.80.  
 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—イタリア調査の自由回答アーター—,統計数理研究所研究リポート, No.81.  
 統計数理研究所(1995).意識の国際比較における連鎖的調査分析方法の実用化に関する研究—共通ファイルコードブック—,統計数理研究所研究リポート, No.82.

## 同性間の性愛に関する法律の国際比較—その理論的背景

釜野 さおり

行動としての同性間の性愛は様々な社会で存在しているが,その naming (その概念の有無,それに対する社会の関心度や反応など)にはかなりの地理的な違いがある.本研究では,同性間の性愛の naming の国間の相違はそのジェンダーシステムの相違に関連しており,ジェンダー

カテゴリーの厳格さが異なると、ジェンダーシステムが同性間の性愛から受ける脅威の度合いも違うので、脅威への反応としての naming もそれに対応して異なってくる、という考えを追求する。

同性間の性愛は少なくとも次の3点でジェンダーシステムを脅かしている。1) 「同性間の性愛」は、異性との性愛関係をもつという理想的な男、女の社会的定義と規範から逸脱し、その定義をあやふやにする。2) 性愛とは男女の間だけで起きるものなので男と女を空間的に区別すれば性愛はない、という前提に反する。3) 生殖に結び付かない性愛関係、ジェンダー規範からはずれた労働の分担を営むことによってそれらを規制する家族制度を脅かす。

同性間の性愛から受けるこれらの脅威を削減するメカニズムとしての naming は少なくとも次の2通りがある。1) 同性間の性愛の naming を行わないことを通してその存在を否定し、無視し、黙殺してそれを抑圧し、脅威に対応する。2) 否定的な naming を通してそれを罪扱い、悪者扱いし、同性間の性愛を抑制してその脅威を免れる。

ジェンダーシステムとそれに脅威をもたらす同性間の性愛への反応としての naming について、次のような仮説が成り立つ。ジェンダーカテゴリーの厳格な社会は、ジェンダーの定義や役割、家族制度がより厳格であり、性別に区切られた空間も多いので、同性間の性愛からより強い脅威を受ける。脅威が強いと、それに対する反応も強くなり、naming を行わない可能性、また、naming が否定的なものになる可能性が高くなる。逆に、ジェンダーカテゴリーが厳格でない社会では、ジェンダーの定義や役割、また家族制度もそれほど厳格でなく、性別に区切られた空間も少ないので同性間の性愛によって受ける脅威も比較的弱い。脅威への反応も弱くなるので同性間の性愛の naming が行われない可能性が低く、また、naming が否定的なものである可能性も低い。

本研究では、世界全体を母集団とし、法律上での同性間の性愛の naming にしほって、それとジェンダーシステムとの関連を分析している。

## 統計データ解析センター

### モンテカルロフィルタによる季節調整

北川 源 四 郎

状態空間モデルにおいて、ノイズの分布を非ガウス型に拡張するとトレンドや季節成分の急激な変化の検出や異常値の処理を自動的に行なうことができる。さらに、非線形・非ガウス型あるいは一般型状態空間モデルに拡張するといろいろなモデリングを自由に行なうことができる。

このような一般的な形の状態空間モデルの状態推定、尤度計算などを行うためにモンテカルロ・フィルタが開発されている (Kitagawa(1993))。このモンテカルロ・フィルタおよび平滑化のアルゴリズムを利用すると

$$y_n \sim \text{Poisson}(T_n \times S_n)$$

というモデルを用いてポアソン系列の季節調整を行うことができる。平均のレベルを変化させた実験を行ったところ、平均が小さな場合でも良好な結果が得られた。

また、この方法では通常の連続値をとる時系列についても、従来の加法型だけでなく、乗算型のモデル

$$y_n = T_n \times S_n \times I_n$$

や混合型のモデル

$$y_n = T_n \times S_n + I_n$$

を自由にあてはめ相互比較することができるようになった。

更に、モンテカルロ・フィルタではシステムの非線形性が自由に取り扱えることから、状態ベクトルの中に未知のパラメータを取り込むことにより従来の最尤推定法にかえて、ベイズ法により状態とパラメータの同時推定を行うことが可能となった。

### 参考文献

- Kitagawa, G. (1993). Monte Carlo filtering and smoothing method for non-Gaussian nonlinear state space model, Research Memo., No.462, The Institute of Statistical Mathematics, Tokyo.

### 語順規則の2パラメータモデル

上田 澄江

言語を語順規則からみたとき、「側置詞と名詞」の順序が分類に関わる重要な指標であることが従来から指摘されてきたが (Greenberg(1966), Hawkins(1983)), そこでは側置詞をもたない言語は無関係とされた。前論文 (Tsunoda et al.(1995)) では、19項目にわたる語順表のデータ 130 言語 (角田 (1991)) の階層クラスタ分析から側置詞をもたない言語を日本語などの後置詞をもつ言語と統合して考えることによって寧ろ自然な分類づけが可能であることを指摘した。

今年度は日本語と同順序の語順をもつか、或いは対立する語順をもつか、に着目してデータを見直し単純化することによって次に続く重要な指標が「数詞と名詞」の順序であることを導いた。この2つの指標、「側置詞と名詞」および「数詞と名詞」の順序によって4グループ分割をおこなうのが、それぞれのグループの相違を最も有効に表現するものである。これらは、階層クラスタ分析、主成分分析、分割表データにおける変数選択の方法 (坂元 (1985)) を用いて分析した。すなわち、階層クラスタ分析によってほぼ 120 言語を便宜的に7グループに分割して分割表の目的変数とし、主成分分析によって説明変数の候補となる項目を選び出し、そして、それらの関係を示す分割表を作成しあらゆる変数の組み合わせからグループを最も特徴づける分割表を AIC の値の比較によって選択した結果である。

### 参考文献

- Greenberg, J.H. (1966). Some universals of grammar with particular reference to the order of meaningful elements, *Universals of Language* (ed. J.H. Greenberg), 73-113, MIT Press, Cambridge, Massachusetts.
- Hawkins, J.A. (1983). *Word Order Universals*, Academic Press, New York.
- 坂元 慶行 (1985). 【カテゴリーカルデータのモデル分析】, 共立出版, 東京.
- 角田 太作 (1991). 【世界の言語と日本語—言語類型論から見た日本語—】, くろしお出版, 東京.
- Tsunoda, T., Ueda, S. and Itoh, Y. (1995). Adpositions in word order typology, *Linguistics*, **33**(4) (to appear).

## 統計科学と並列計算

田村 義保

統計科学における計算統計学のように大規模な計算が不可欠な研究が多くの自然科学の分野で必要度を増している。一方、コンピュータの速度は、素子製造技術が限界に近づいており、単独のCPUではこれらの研究者を満足させるだけの十分な演算性能を出せないようになってきている。このため、並列計算機、超並列計算機と言われる計算機が製造されるようになり、実際に使用されつつある。統計科学の分野でも、並列計算機を利用した研究が、中心的なテーマの一つになるものと思われる。統計科学の研究者に並列計算機とは、どのようなものであるかを紹介するために本講演を行った。

並列計算機を構造により分類すると次のように分類される。

- SMP(Symmetric Multi Processor)
- SIMD(Single Instruction Multiple Data)
- MIMD(Multiple Instruction Multiple Data)

メモリー形式による分類として、分散メモリー、共有メモリー、分散共有メモリーがある。さらに、ハイパーキューブ、クロスバー等のネットワーク方式により分類される。これらが組合わされて、種々の並列計算機が市販されている。SMPで共有メモリーのものがもっとも使用しやすいものと思われるが、CPU数が増えたとき、線形的に演算能力が増えることが期待できないため、MIMD方式の計算機が現在の主流になっているように思われる。プログラミングモデルもメッセージパッシング方式やSPMD方式等種々のものが考えられている。計算統計学の中で数値積分や乱数を使った繰り返し計算の多いものは、ハードウェアとして分散メモリーのMIMD方式、プログラミングモデルとしてメッセージパッシング方式が最適であると考えている。これらを実現するためにはメッセージパッシングのライブラリである、PVMやMPIがある。しかし、これらで実際にプログラムを書くのはかなりの技量が必要である。HPFのような自動並列化ソフトウェアを用いた方が良いと思われる。

並列計算機による統計データ解析の例は下平 他(1995)を参考にして欲しい。重回帰分析ではデータ数10,000程度で8変数の場合、80%以上の並列処理効率(16CPUで1CPUの14.8倍程度の速度になる)が得られている。また、主成分分析では65%程度の効率となっている。

## 参考文献

- 下平 文彦, 小林 覚, 白川 友紀, 田村 義保(1995). 統計データ解析の並列処理, 計算機統計学(掲載予定).

## 吹雪時における気象要素の伝播に関する統計的解析

荒畑 恵美子

吹雪は、気温や湿度等の気象要素の変動と共に空間を伝播している。2地点A点とB点で気象要素を測って、その伝わり方を調べてみた。風を入れたとき、どう伝わっていくかを調べて



みた。観測変数として温度、湿度、風速、風向を用いた。これらの時系列データをスペクトル解析やランダム変動解析をした。A から B へ温度も湿度も伝わっていることがわかった。又、風が強いほど、吹雪は、生起しやすいと言われていることが、示された。吹雪に他の気象要素がどのようにかかわっているかを見るために、他の気象要素も考慮に入れて、もう少し調べてみる必要があるようである。これは、共同研究 (6-共研-A-52) の一部である。

## 予測制御研究系

### 神奈川県内製造業者へのアンケートとその解析

川崎能典

神奈川県庁商工部産業政策課が行なった「かながわ産業構造モデル分析事業」に参加し、県内製造業 60 社に対するアンケート調査データの解析を行なった。自動車・電機・一般機械(その他)から 20 社づつ、それぞれ各業種の上位大企業と、規模及び製品内容から代表的な中堅・中小企業を 1 対 2 程度の比で選定した。調査自体は典型調査であったが、質問項目は 100 以上に及ぶため、大量データの情報処理という観点から AIC による高次元分割表解析プログラム (CATDAP) を援用した。その際、AIC の差の小さな分割表どうしについては、尤度のバイアス補正の項 (カテゴリ数によって決まる) を場合によっては外して評価を下すこととした。

単純な集計結果も含めて総合的に判断すれば、以下のような事柄がうかがえる：海外調達の増加、価格低下の状況、海外生産へのシフト、研究開発のグローバル化への動き、企業間取引引きの全国展開を背景にした神奈川の立地優位性の変化、系列取引の崩壊、県内機能の開発特化、など。特に CATDAP による解析で顕著に見られた事柄としては、コスト削減項目の選択状況がある。自動車部品メーカーが生産工程の自動化・人件費削減でコストダウンしてゆく方針である以外は、ほとんどが設計変更・海外調達の増加でコスト削減を行なう方針であることが明瞭に示されている。

円高を通じて海外での部品調達やオフショア生産などが促進されたこと、世界的に供給体制が充実してきたこと、情報化社会の本格的到来によって組織のあり方が変わってきたことなどにより、神奈川に存在してきた「ものづくり」を中心とした諸機能の集積及びその相互関係は崩れてきていると言える。神奈川テクノコンプレックスとは、輸入促進が替レートに守られて日本が国際価格から孤立していられた時代のものづくりの形であったかも知れない。

## 参考文献

- Katsura, K. and Sakamoto, Y. (1980). CATDAP, a categorical data analysis program package, *Comput. Sci. Monographs*, No.14, The Institute of Statistical Mathematics, Tokyo.  
Sakamoto, Y. (1991). *Categorical Data Analysis by AIC*, Kluwer, Tokyo.

## モンテカルロフィルタの実装について

樋口知之

パラメータ間に一次元的な特殊な構造をもつようなもの—例えば時系列—は、一般化された状態空間表現の枠組みで問題解決を行うのが、計算手続きの見通しが良く様々な点に於て都合

が良い。実際、状態ベクトルの次元がきわめて低い(例えば4以下)時には、確率分布を数値的に近似表現することで解決されている。ただこの方法による解法は、状態ベクトルの次元に強い制約を課すため、問題対象の自在な *parametrization* という観点からは満足できなかった。この制約を取り除く目的で提案されたものが、モンテカルロフィルタ (Monte Carlo Filter, 以後 MCF と略す) である (Kitagawa (1993))。このアルゴリズムは、自然淘汰と発生のメカニズムを模倣する最適化の一手法の遺伝的アルゴリズム (Genetic Algorithm, 以後 GA (Goldberg (1989))) とほぼ同じ構造をもつことが指摘されている (樋口 (1993, 1994))。GA と対応づけることで MCF の理解は容易になるうえに、アルゴリズムに内在する本質的な問題点の認識も深まる。また、MCF の予測の操作に GA の Crossover や Mutation の操作をとりこんだ手法が考えられる。実際にその試みが季節調整などの具体的な問題に適用されつつある。

### 参考文献

- Goldberg, D.E. (1989). *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley, Massachusetts.
- 樋口 知之 (1993). 遺伝的アルゴリズムとモンテカルロフィルター, Research Memo., No.489, The Institute of Statistical Mathematics, Tokyo.
- 樋口 知之 (1994). モンテカルロフィルタにおけるランダムサンプリング, 情報処理学会研究報告 (情報研報) 94-HPC-54, 94(108), 7-14.
- Kitagawa, G. (1993). Monte Carlo Filtering and smoothing method for Non-Gaussian Nonlinear state space models, Research Memo., No.462, The Institute of Statistical Mathematics, Tokyo.

### A Likelihood Analysis of Chaos

尾崎 統

カオスモデルの尤度法についての否定的俗説を正し、その有効さをローレンツカオスの例を使って数値例で示した。

### 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、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わらかにすることが適応制御の研究の課題の一つとされている。

これに対してこれまで、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次の系について次数や非線形成分、外乱に依存しない(ロバストな)適応制御系の構成法を提案した。次にそれらの手法を相対次数が1次か2次かで未知の場合に拡張し、同様に次数や非線形性にロバストな適応制御系が構成されることを示した。

今年度は以上の研究をさらに相対次数が2次以上の一般の場合に拡張した。近年、非線形制御の分野で広いクラスの対象を正実化する手法として Backstepping 法が着目されているが、そ

こに含まれる一般の相対次数の系に対する正実化法を、これまでに導出した高次出力微分の近似精度を適応的に決定する手法に応用することで、一般の相対次数の場合について次数や非線形成分に依存しないロバスト適応制御系の構成法を求めた。次に可変構造制御のパラメータ表現を含む制御方式も導出し、同様に次数依存性のない適応制御系が構成できることを示した。最後にこの一連の手法の最小次元構成法と、理想的な条件下で制御誤差の零収束を実現する高次元構成法を求めた。最小次元構成においては、制御誤差と非線形成分の整定を行うハイゲインパラメータがそれぞれ1つずつと、高調波利得を推定するパラメータ1つの合計3個の適応パラメータで、任意の次数に対して大域的有界性と制御誤差の任意に指定した(小さな)領域への収束が保証されることを示した。また高次元構成では、ハイゲインフィードバックに基づく非線形補償に加えて、適当な次元と非線形関数を含む線形動的補償器を用いた適応機構も並列に配置することで、モデル化誤差(次数や非線形性に関する)が存在するときには前記の大域的有界性と制御誤差の任意に小さな領域への収束性が保証され、さらにモデル化誤差が存在しない理想的な環境のもとでは制御誤差の零収束も達成されるような、ロバスト適応制御系が構成できることを示した。

### 参考文献

- Miyasato, Y. (1994). Adaptive control for nonlinear systems with unknown degrees, *Proceedings of the 1st Asian Control Conference*, Vol.2, 593-596.
- Miyasato, Y. (1994). Adaptive variable structure control for nonlinear systems with unknown degrees, *Preprints of IFAC Symposium on Robust Control Design*, 69-74.
- 宮里 義彦 (1994). 適応制御の新しい展開—Backstepping, システム/制御/情報, **38**(9), 477-484.
- Miyasato, Y. (1994). Model reference adaptive control for nonlinear systems with unknown degrees, *Proceedings of the 33rd IEEE Conference on Decision and Control*, Vol.3, 3056-3061.
- 宮里 義彦 (1995). 次数に依存しない非線形モデル規範形適応制御系の構成法 (一般の相対次数の場合), 計測自動制御学会論文集, **31**(3), 324-333.

### Correct Definition of the Lagrangean Multiplier Test Statistic

田 辺 國 士

最適化問題を微分幾何学的に取り扱うことには多くの利点がある。非線形問題のためのアルゴリズムの設計や解析においては様々な近似式が用いられるが、不用意にこの近似を行うと正しくない結果に導かれる。統計学の理論分野においても最適化問題が頻繁に現れ、近似を慎重に行わなかったため誤った結果を導いてしまったケースがある。これを避けるためには微分幾何学的アプローチが有用である。ここでは関数制約仮説の検定に登場する Lagrangean Multiplier Test を取り上げ、このアプローチによるより適切な結果を与える。

未知パラメータ  $\theta$  に関する検定問題: 帰無仮説:  $a(\theta) = 0$ , 対立仮説  $a(\theta) \neq 0$  を考えよう。ただし,  $a(\theta) \equiv (a_1(\theta), a_2(\theta), \dots, a_m(\theta))^t$  とする。この関数制約仮説の検定問題は統計学において大変重要な問題であり, 対数尤度比統計量, Wald 統計量, Lagrange 乗数統計量などの検定統計量がよく知られている。教科書などにおいて, Lagrange 乗数統計量はしばしば

$$LM \equiv S(\bar{\theta})^t I(\bar{\theta})^{-1} S(\bar{\theta})$$

と定義されている。ただし,  $n$  はサンプルサイズ,  $S(\theta) \equiv \nabla_{\theta} \log L(\theta; x)$  はスコア,  $L(\theta; x) \equiv \prod_{i=1}^n f(x_i; \theta)$  は尤度関数,  $I(\theta) \equiv -E(\nabla_{\theta}^2 \log L(\theta; x))$  は Fisher の情報行列,  $\bar{\theta}$  は帰無仮説の下

( $a(\bar{\theta}) = 0$ ) での最尤推定値とする。この  $LM$  は、仮説の下での最尤推定値のみを用いて定義されるといふ意味で便利な統計量であるため、経済学などの分野においてモデルの検定に広く用いられている。

仮説が線形の場合には上式は正しい表現であるが、非線形の場合には、これは適切ではない。帰無仮説が定義する多様体の曲率に関する項 (すなわち、 $a(\theta)$  の高次の項) がこの表現には欠けているというメタ・計算的な観察からも、これが不自然であることは容易に理解できる。Lagrange 統計量は正しくは、

$$LM \equiv S(\bar{\theta})^t N(\bar{\theta}) S(\bar{\theta}) = \bar{\lambda}^t A(\bar{\theta}) N(\bar{\theta}) A(\bar{\theta})^t \bar{\lambda}$$

と定義すべきである。ただし、 $\bar{\lambda}$  は  $\bar{\theta}$  に対応する Lagrangean Multiplier、 $A(\theta) \equiv \nabla_{\theta} a(\theta)$  は  $a(\theta)$  の Jacobi 行列で、条件

$$S(\bar{\theta}) + A(\bar{\theta})^t \bar{\lambda} = 0$$

を満足する。また、 $M(\theta)$ 、 $N(\theta)$  の定義は

$$M(\theta) \equiv I(\theta) - \sum_i \bar{\lambda}_i \nabla_{\theta}^2 a_i(\theta)$$

あるいは

$$\begin{aligned} M(\theta) &\equiv -\nabla_{\theta}^2 \log L(\theta : x) - \sum_i \bar{\lambda}_i \nabla_{\theta}^2 a_i(\theta) \\ N(\theta) &\equiv M^{-1}(\theta) \left( M(\theta) - \sum_i \bar{\lambda}_i \nabla_{\theta}^2 a_i(\theta) \right) M^{-1}(\theta) \\ &= M^{-1}(\theta) - M^{-1}(\theta) \left( \sum_i \bar{\lambda}_i \nabla_{\theta}^2 a_i(\theta) \right) M^{-1}(\theta) \end{aligned}$$

とする。 $a(\theta)$  が線形である場合には、上記の  $N(\theta)$  の定義式における  $\sum_i \bar{\lambda}_i \nabla_{\theta}^2 a_i(\theta)$  の項が 0 となるので、最初に掲げた式に帰着する。

最初の式の不適切性は、Lagrange 乗数統計量を提案した Aitchison and Silvey(1958) に発するもので、多くの解説論文、辞典、教科書において無批判に引用されている (Kmenta(1971), Buse(1980), Engle(1984), 竹内(1989))。Lagrange 乗数統計量は広く用いられているようであるが、非線形仮説の検定問題の場合には注意して用いる必要がある。

### 参考文献

- Aitchison, J. and Silvey, S.D. (1958). Maximum likelihood estimation of parameters subjects to constraints, *Ann. Math. Statist.*, **29**, 813-828.
- Buse, A. (1980). Likelihood ratio, Wald, Lagrange multiplier tests: an expository note, *Amer. Statist.*, **36**(3), 153-157.
- Engle, R.F. (1984). Wald, likelihood ratio, Lagrange multiplier tests in econometrics, *Handbook of Econometrics* (eds. Z. Griliches and M.D. Intriligator), 776-827, North-Holland, Amsterdam.
- Kmenta, J. (1971). *Elements of Econometrics*, Macmillan, New York.
- 竹内 啓 編 (1989). 仮説検定, 『統計学辞典』, p.103, 東洋経済新報社, 東京.

## 内点法の局所的収束性について

水野 眞 治

線形相補性問題を解く内点法に関する研究を行なった。Mizuno et al.(1994)では、線形相補性問題を幾何的に定義することにより、さまざまなタイプの線形相補性問題を統一的に解析できることを示し、その問題を解く内点法のアルゴリズムを提案した。そして、そのアルゴリズムがグローバルに解を求め、初期点として実行可能内点を使えば反復回数が多項式オーダーで抑えられ、強相補性解を持つ場合には局所的に二次収束することを明らかにした。Mizuno(1994)では、強相補性解を持たない線形相補性問題の場合にも局所的に二次収束する新しいアルゴリズムを提案した。

## 参 考 文 献

- Mizuno, S.(1994). A superlinearly convergent infeasible-interior-point algorithm for geometrical LCPs without a strictly complementarity condition, *Math. Oper. Res.*(to appear).  
 Mizuno, S., Jarre, F. and Stoer, J. (1994). A unified approach to infeasible-interior-point algorithms via geometrical linear complementarity problems, *Appl. Math. Optim.*(to appear).

満足条件としての  $H_\infty$  ノルム評価基準

伊 藤 聡

外乱など不確かさの存在の下での線形制御系の設計仕様として、外乱に対するロバスト性を保証する  $H_\infty$  ノルム評価基準が知られているが、ここではこの  $H_\infty$  ノルム評価基準を満足条件とする線形システムの満足最適化について述べる。

次のような線形時不変系を考える。

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

ここで  $x(t) \in R^n$  は状態,  $u(t) \in R^r$  は制御入力,  $v(t) \in R^q$  は外乱,  $z(t) \in R^l$  は評価出力である。許容な  $u$  および  $v$  のクラスとしてそれぞれ  $U = L_2([0, \infty), R^r)$ ,  $V = L_2([0, \infty), R^q)$  を考える。制御入力  $u$  あるいはこれを生成する制御則が与えられたとき、外乱  $v$  から評価出力  $z$  への伝達特性が一意に定まるが、これを  $T_{vz}$  で表わすことにする。伝達作用素  $T_{vz}$  は一般にはアフィンであるが、特に線形であると仮定すると、その  $H_\infty$  ノルムは時間領域において次のように定義される。

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

いま、ある希求水準  $\gamma > 0$  に対して

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

となるような制御入力  $u$  あるいはこれを生成するコントローラ  $K$  を求める問題を考える。(2)式は等価に

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

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

$$\begin{array}{ll} \min_u F(u) & \text{あるいは} \\ \text{subj. to } G(u) \leq 0 & \min_K F(K) \\ & \text{subj. to } G(K) \leq 0 \end{array}$$

Shimizu and Ito (1995) では、目的関数  $F$  として min-max 型の評価基準を持つ凸計画問題

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

について考察し、その満足最適解が一種の状態フィードバック則で与えられることを示した。

研究報告会当日は、この他に最急降下法に基づく非線形系のフィードバック制御 (Shimizu et al. (1995)) および最適制御問題に対する内点法の近似的な実装 (Ito et al. (1995), 伊藤 (1994)) について報告した。

## 参考文献

- 伊藤 聡 (1994). ヒルベルト空間における内点法とその最適制御問題への応用, 計測自動制御学会第20回システムシンポジウム・第20回知能システムシンポジウム・第5回ニューラルネットワークシンポジウム・第14回離散事象システム研究会合同シンポジウム講演論文集, 335-342.
- Ito, S., Kelley, C.T. and Sachs, E.W. (1995). Inexact primal-dual interior point iteration for linear programs in function spaces, *Computational Optimization and Applications* (to appear).
- Shimizu, K. and Ito, S. (1995). Satisfactory optimal control: min-max control under the  $H_\infty$  norm constraint, *Proceedings of the 1995 American Control Conference* (to appear).
- Shimizu, K., Nukumi, H. and Ito, S. (1995). Direct steepest descent control of nonlinear dynamical systems, *Proceedings of the 3rd IFAC Symposium on Nonlinear Control Systems Design* (to appear).

## 男女差の分析

鈴木 義一郎

『日本人の国民性調査』では、「もう一度生まれかわるとしたら、あなたは男と女のどちらに生まれてきたいと思いますか」と尋ねている。

そこで、1958年から93年までの  $2 \times 2$  分割表のデータを対角型と独立型に分割して、対角型のほうの係数  $a$  の値を算出して

$$0.243 \quad 0.319 \quad 0.471 \quad 0.487 \quad 0.508 \quad 0.542 \quad 0.589 \quad 0.653$$

という結果が得られる。35年間で、 $a$  の値が25%から65%まで増大した。

また、人口問題研究所の『出産力調査』の結果で、初婚・再婚別の組合せの比率を眺めてみると、初婚同士のカップルが94%もある。ところが、人口動態統計による比率をみると、初婚同士の組合せは80%をやや上まわる程度に過ぎない。つまり、常に“初婚同士でない回答者で初婚である”と回答するタイプが、1割程度はいることが分かる。

出産力調査の初婚同士以外の比率を人口動態統計のほうの比率で割算してみると、これらの値が3回の調査についてはほぼ安定して一定である。比較的正直な層は夫再婚・妻初婚のカテゴリーで、ほぼ半数が正直に回答していることが分かる。最もウソ回答の多いのが再婚同士のカテゴリーで、約4分の3が正しく答えていない。

そこで、人口動態統計のほうのデータを三角型と対称型に分割して、三角型のほうの係数  $a$  の値を算出してみると

0.40 0.36 0.29 0.19 0.14 0.13 0.15

といった結果が得られる。75年以降、 $a$  の値はかなり小さくなってきており、“初婚の夫が再婚の妻を嫌う”傾向はほとんど無くなってきている。

88年と93年の『日本人の国民性調査』で、「子供を一人だけもつとしたら、男の子の方がよいですか、女の子の方がよいですか、それとも、どちらでもよいですか?」という質問を行った。93年の結果では半数の女子が女の子を希望し、5年前より10%もの伸びをみせた。男子の回答でも女の子を希望する比率が増えたが、依然として男を希望する回答のほうが多い。

## 内点法に対する新しい探索方向について

土 谷 隆

### 線形計画問題

$$(1) \quad \begin{aligned} & \text{minimize}_x \quad c^T x, \\ & \text{subject to} \quad Ax = b, \quad x \geq 0 \end{aligned}$$

### およびその双対問題

$$(2) \quad \begin{aligned} & \text{maximize}_{(y,s)} \quad b^T y, \\ & \text{subject to} \quad A^T y + s = c, \quad s \geq 0, \end{aligned}$$

を考える。ここで、 $A \in \mathbf{R}^{m \times n}$ ,  $c, x, s \in \mathbf{R}^n$  and  $b, y \in \mathbf{R}^m$  とし、次の仮定を置く:

- (1) と (2) は主双対内点可能解を持つ。
- Rank  $A = m$ .

Karmarkar 法の登場以来、最適化の分野では、制約領域の内部に点列を生成し最適解を求める内点法と呼ばれる解法の研究が盛んに進められている。これらは、大まかには、

- 主内点法: 主問題の空間で点列を生成し、それより双対問題に関する情報を副次的に生成する。
- 主双対内点法: 主問題、双対問題を対等に扱い、各反復で主問題、双対問題の可能解の組を同時に更新する。

- 双対内点法: 双対問題の空間で点列を生成し, それより主問題に関する情報を副次的に生成する.

の3つに分けることができる. これらの内点法は次のような性質を共通に持っている:

- $P = DA^T(AD^2A^T)^{-1}AD$  ( $D$  は対角行列) なる形の射影行列が探索方向に現れる.
- 探索方向は, 各方法において, “アフィンスケーリング方向” と “中心化方向” と呼ばれる2つの探索方向の線形結合でかける.
- 主問題, 双対問題の内部で定義され, 最適解に向かう曲線, 中心曲線が重要な役割を果たす.

本研究では, 3つの方法の類似点に着目し, 主双対問題の内点可能解の上で定義されこれら3つの方法の探索方向を含むような1パラメータベクトル場の族を考えた. それは一つのパラメータ  $\alpha \in [-1, 1]$  を含み, 主双対内点可能解  $(x, y, s)$  において, 次の2つの方程式の解として与えられる.

$\alpha$ -アフィンスケーリング方向  $(\Delta x_A, \Delta y_A, \Delta s_A)$ :

$$(3) \quad D^{-1}\Delta x_A + D\Delta s_A = (XS)^{(\alpha^2+1)/2}e,$$

$$(4) \quad A\Delta x_A = 0,$$

$$(5) \quad \Delta s_A = -A^T\Delta y_A$$

$\alpha$ -中心化方向  $(\Delta x_C, \Delta y_C, \Delta s_C)$ :

$$(6) \quad D^{-1}\Delta x_C + D\Delta s_C = (XS)^{(\alpha^2-1)/2}e,$$

$$(7) \quad A\Delta x_C = 0,$$

$$(8) \quad \Delta s_C = -A^T\Delta y_C.$$

ここで,

$$(9) \quad D = \left[ n \frac{XS}{x^T s} \right]^{\alpha/2} XS^{-1}, \quad X = \text{diag}(x), \quad S = \text{diag}(s)$$

である. 上に定義した方向は次の性質を持つ.

- $\alpha = 1$  ( $\alpha = -1$ ) については,  $(\Delta x_A, \Delta y_A, \Delta s_A)$  の  $x$  成分,  $((y, s)$  成分)  $(\Delta x_C, \Delta y_C, \Delta s_C)$  の  $x$  成分 ( $(y, s)$  成分) は, 主(双対)アフィンスケーリング方向, 主(双対)中心化方向になっている.
- $\alpha = 0$  については,  $(\Delta x_A, \Delta y_A, \Delta s_A)$ ,  $(\Delta x_C, \Delta y_C, \Delta s_C)$  は, 主双対アフィンスケーリング方向, 中心化方向となっている.

つまり, これらの探索方向は, これまで存在する内点法の探索方向の一般化になっているわけである. 本研究では (Tsuchiya (1994)) に基づいて, 任意の  $\alpha \in [-1, 1]$  について, これらの探索方向を用いて  $O(\sqrt{n}L)$  回の反復の predictor-corrector 法が構成できることを示した.



## 参考文献

Tsuchiya, T. (1994). A new family of polynomial-time interior point algorithms for linear programming, Research Memo., No.510, The Institute of Statistical Mathematics, Tokyo.

## マルコフ連鎖モンテカルロ法の改良

伊庭 幸人

マルコフ連鎖モンテカルロ法 (動的モンテカルロ法, ボルツマンマシン) の緩和時間を短縮する方策について論じた. 緩和時間の短縮のための方法は, (1) アンサンブルを拡張する方法 (2) ダイナミクスを改良する方法に大別される. 講演では, (1) に属する1つのアルゴリズムと (2) に関する展望を論じた. 以下では前者について述べる.

本年度は, 新しい種類の拡張アンサンブル法をいくつか研究したが, 講演で話したのは, このうち, 「システムサイズを一様分布にするようなアンサンブル」(「マルチグランドカノニカルアンサンブル」) を利用する方法についてである. この方法は, いわば「組み立て式」のシミュレーションであって, その狙いは, 小さい系から成長させてゆくことによって, 準安定状態へのトラップを防ぐことにある. ただし, 詳細釣合の条件を満たすためには成長する一方では駄目なので, 成長と縮小を乱歩として行なわせることになる.

具体例としてスピングラスのSK模型

$$(1) \quad E(\{S_i\}) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} S_i S_j, \quad S_i \in \{\pm 1\}$$

$$(2) \quad P(\{S_i\}) = \frac{1}{Z} \exp(-E(\{S_i\})/T)$$

( $N$ は系の大きさ(変数の数),  $Z$ は規格化定数,  $J_{ij}$ は分散が  $1/\sqrt{N}$  のガウス分布から抽出して固定する) を考える.

システムサイズを可変にするということで, まず考えられるのは, (1) 式において  $S_i$  が値 0 もとれるようにすることである. 値 0 をとった変数は「無い」ということになり,  $n = \sum_i |S_i|$  が大きくなる. これを, 式で書くと,

$$(3) \quad E(\{S_i\}) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} S_i S_j, \quad S_i \in \{+1, -1, 0\}$$

となる. しかし, これでは, 特定の大きさ  $n$  のところに分布の山が出来てしまい, 目的をうまく果たすことができない. ここで, マルチカノニカル法の考えを応用する.  $n$  に関する罰金項  $f(n)$  を

$$(4) \quad E(\{S_i\}) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N J_{ij} S_i S_j + f(n), \quad S_i \in \{+1, -1, 0\}$$

のように加えて, マルコフ連鎖モンテカルロ法のシミュレーションをしながら,  $n$  の (周辺) 分布が一様になるように修正 (学習) してゆけば良い. いったん,  $f(n)$  がうまく定められたら, それを固定してシミュレーションを行えば, 狙いどおりの成長/縮小が生じる. ここで, サンプルを  $n = N$  の時にだけとることにすれば, もとの分布 (1), (2) からの正しいサンプリングに

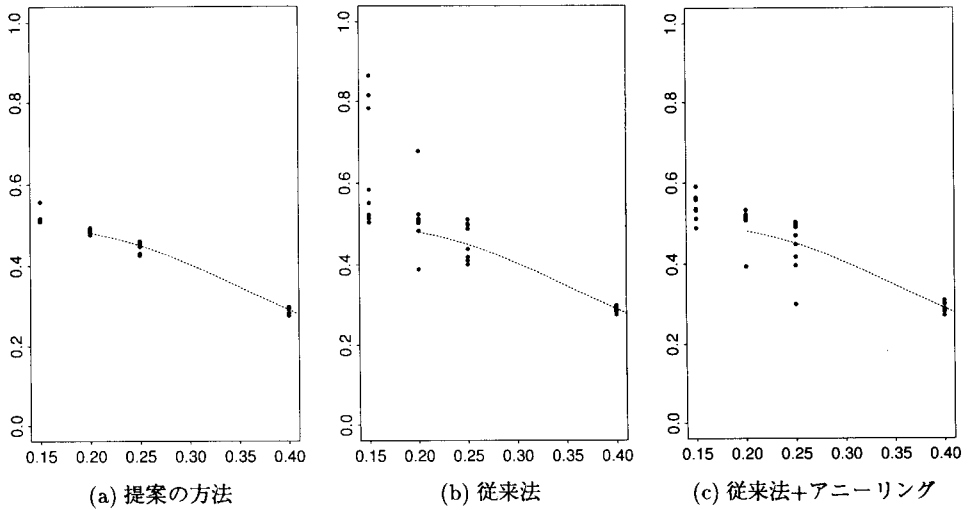


図1. 縦軸は統計量  $\frac{1}{N^2} \sum_{ij} \langle S_i S_j \rangle^2$  (実レプリカを用いずに直接測定したもの). 横軸は温度  $T$ . 各温度につき8回ずつ異なる初期値・乱数で計算した結果を黒丸で示した. 点線 (a, b, c 共通) は“交換法”による値をしめす. サンプルサイズは  $N=50$ , サンプル数は1. (a) は1千MCS捨てて, 20万MCSを  $f(n)$  の学習に費やしたのちの20万MCSで測定 (学習の際には1万MCSごとに  $f(n)$  を更新). (b) は20万MCS捨てた後の20万MCSで測定. (c) は最初の20万MCSの一部を利用して  $T=1.0$  から一定速度 ( $5 \times 10^{-6}/\text{MCS}$ ) で目標温度までアニールし, その後の20万MCSで測定. (c) は講演後に追加.

なっている. もちろんサンプル間の相関は残るが, 相関の消える時間 (緩和時間) は短くなっていることが期待される. そのかわり, 得られるサンプル数は  $1/N$  になる (途中の  $n$  での結果が役に立たないのが欠点である).

結果の例を図1にしめす. 図1に示した例/温度範囲では, 拡張アンサンブル法のひとつである交換法に近い値が得られている. また, 従来法 (熱浴法・Gibbs Sampler) との比較でも優れていることがわかる.

## 大気ゆらぎの時空間モデル

石黒真木夫

雲の密度のように, かたちを変えながら一定速度である空間のなかをながれて行く現象の線形モデルをつくって, 電波天文学の観測データへの応用を示した.

モデルは単純な線形 Gauss 型の状態遷移方程式になり, 時空間現象の simulation を手軽に行なうのに便利であり, 線形の観測が仮定できる場合には Kalman filter を利用することによって現象解析に利用できる.

図1の上段に示したのは, ある2地点におけるある観測値の差の時系列記録である. 下段にモデルを用いた推定値を示す. かなり良い推定とってよいのではないか.

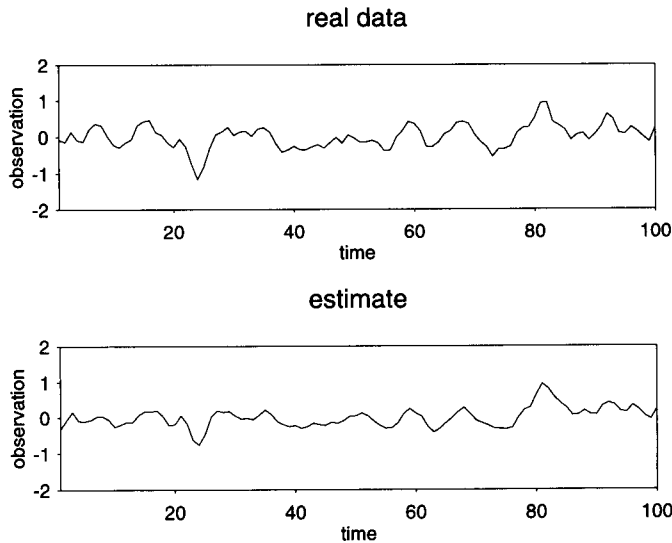


図1. 観測値時系列の推定(下段)と真値(上段).

このモデルによれば、ユークリッド空間以外の空間、たとえば球面上を変形しながら流れる現象も同じように扱うことができる。

#### 参考文献

石黒 真木夫(1995). 大気ゆらぎの時空間モデル II, Research Memo., No.548, 統計数理研究所, 東京.

#### 調査実験解析研究系

##### いくつかの調査に参加して

坂元 慶行

今年度も、「社会階層と社会移動 1995 年調査」をはじめとして、いくつかの調査に関わったが、研究活動の中心は、1993 年 10 月に実施した「第 9 次 日本人の国民性調査」の結果の分析と発表であった。

この調査は、統計数理研究所が 1953 年(昭和 28 年)から 5 年おきに 40 年にわたって行なっているもので、その特徴は、同じ質問文で調査を繰り返すことによって、日本人のものの見方や考え方が時代によってどう変わるかを明らかにするところにある。この調査から、現在は、近代化が終わり、伝統回帰的現象も終わり、それらを超えた新たな次元の意識状況にあることが指摘されていたが、1970 年代以降の意識変化を検討したところ、社会や家庭への満足度が高まり、穏やかな意見が増え、女性人気が高まり、家庭志向が強まっていることがわかった。より端的には、私的心理空間が肥大化し公的空間が縮小する私生活偏重主義とでも表現できよう。そして、価値観の多様化とはいっても、この点は多くの国民に共通した傾向ではあるまいか。人は、つまるところ、自らの価値の実現をめざして行動するものであるとすれば、この価値観の動きは、最近の政治意識や行動などとも無縁ではなく、その含意は小さくない。

## 継続調査データの蓄積とコウホート分析

中 村 隆

継続調査が回を重ねてデータが蓄積してくると、その分析のためにより複雑なモデルが必要となり、また結果的に複雑なモデルが選ばれるようにもなる。しかしこのとき、調査誤差を適切に考慮しないと複雑なモデルほど有利になりすぎるといった問題が生じる。

コウホート分析は、継続調査データを分析する方法の1つであり、人間集団に関する何らかの数量特性を年齢層×調査時点別に集計したコウホート表データから、年齢・時代・世代(コウホート)効果を分離する方法である。本年度はこのコウホート分析について、「日本人の国民性調査」データの蓄積に対応するために、これまでのベイズ型ロジット・コウホートモデルに年齢×時代の交互作用効果をもつモデルと過大分散パラメータを考慮するモデルを加えて ABIC によりモデル選択を行なう方法を研究した。

継続調査の調査期間が長期にわたると、ライフサイクルが変化するので、年齢効果を全期間に共通するものとして想定するのは困難になる。そこで、年齢効果を年齢×時代の交互作用効果に拡張したモデルを提案した。このモデルが適切に機能している1つの例として国民性調査の“家族が大切”の分析結果を得ることができ、晩婚化の影響を捉えることができた。

また、拡張疑似尤度に基づく過大分散パラメータ・モデルを導入し、複雑なモデルが有利になりすぎないように現実の層別多段抽出法による標本誤差の影響評価を試みた。通常の尤度の代わりに拡張疑似尤度を用いる点について議論の余地は十分残されているが、近似的な方法として、また簡便な方法として有用と考えられる。国民性調査データのコウホート分析を通じていえることは、過大分散パラメータの有無が結論を左右することは多くの場合ないということである。しかし、国民性調査の“くらし方(社会につくす)”のような例も存在し、他の社会調査データの分析法についても過大分散を考慮することが有用になる場面のあることが示唆される。

一方、過大分散パラメータの推定値が1よりかなり小さい場合が散見され、誤差の部分拾い過ぎているという印象も免れ得ない。この点への対処は将来の課題の1つである。

## 参 考 文 献

- 中村 隆(1995). 交互作用効果モデルと過大分散モデルを用いたコウホート分析—「日本人の国民性調査」データへの適用—, 統計数理, 43(1), 99-119.

## 国民性調査への共分散構造分析の適用

前 田 忠 彦

「日本人の国民性調査」第9次全国調査(統計数理研究所(1994))のM型調査票(新しい意識動向を把握するために用意された項目を中心とする調査票)から、「満足感」に関する項目を中心に14項目をとりあげ、共分散構造モデルに基づく分析を行った。共分散構造分析は、潜在変数(構成概念)の間に構成した因果モデルの妥当性を検証するための多変量解析モデルである(Bollen(1989))。分析の目的は、日本人の満足感に関する二つの側面すなわち「個人生活への満足感」と「社会への満足感」を、それ以外の主観的な要因から説明するための因果モデルを構成し、その妥当性を検証すること、特に「個人生活への満足感」と「社会への満足感」の間の関係を検討することであった。

「 $\eta_1$  : 個人生活への満足感」と「 $\eta_2$  : 社会への満足感」を内生的潜在変数, 他の3つの構成概念 (「 $\xi_1$  : 健康状態」「 $\xi_2$  : 生活水準」「 $\xi_3$  : 日本の物質的豊かさに対する評価) を外生的潜在変数とする因果モデルを構成し, 候補となる複数のモデルから情報量規準に基づくモデル選択を行うという手順で分析を行った。すなわち,  $\eta_1$  と  $\eta_2$  の間に, 1)  $\eta_1 \rightarrow \eta_2$ , 2)  $\eta_1 \leftarrow \eta_2$ , 3)  $\eta_1 \rightleftharpoons \eta_2$  という, 大きく分けて3種類の影響関係を設定した多数のモデルを構成し, それらの中から最適モデルを選択した。

まず男女をあわせた全サンプル ( $N = 1381$ ) に基づく分析では,  $\eta_1 \rightarrow \eta_2$  という方向の影響を仮定したモデルが選択された。選択されたモデルと双方向の因果を仮定したモデルのパラメータ推定値から得られた主たる知見は以下の通りである。

1. 「個人生活への満足感」は「社会への満足感」を高めるが, 逆方向の影響は仮にあるとしても相対的に弱い効果しかもたない。
2. 「個人生活への満足感」に対する影響力が最も強い構成概念は, 自身の「生活水準」に対する評価である。
3. 「生活水準」に対する評価は, 「社会への満足感」への直接効果をほとんどもたない。

更に男女別に分析を行ったところ, 男性と女性では選択されるモデルが異なり, 満足感をめぐる意識構造には男女差がある可能性が示唆された。

#### 参 考 文 献

- Bollen, K.A. (1989). *Structural Equations with Latent Variables*, Wiley, New York.  
 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.

#### 非球粒子系のモンテカルロ法

種 村 正 美

われわれが現実に目にする自然界の多くの空間配置パターンはさまざまな形をもつ対象 (粒子と総称する) から成っている。従来の Spatial Statistics では, このような粒子を円, または球として記述してきた。また, 粒子を等方的な (球対称の) 相互作用をする点としても記述してきた。これらは, もちろん記述を容易にするための一種の近似であった。

しかしながら, より現実的で柔軟な統計モデルを構成するためには, 等方的相互作用をもつ点や円・球といった球対称性をもつ粒子ばかりでなく, 球対称性をもたない粒子 (これを非球粒子と呼ぶことにする) も扱う必要がある。

この目的のために, 今年度, 非球粒子から成る空間配置パターンを生成するためのモンテカルロ法の計算プログラムの開発をして計算機シミュレーションを行った。ここでいうモンテカルロ法とはマルコフ連鎖を生成する方法であり, 所与の確率密度に対する空間配置を平衡配置として次々に実現できる方法である。典型的な確率密度としてギブス分布があり, Spatial Statistics でも多用されている手法である。

非球粒子としては, 無限の可能性がある中で, その第一歩として互いに重なり合えない長方形および楕円をそれぞれ考察した。計算プログラムの作成には, 前年度に行なった長方形・楕

円のランダム充填問題の計算プログラムに用いた工夫が用いられており、計算の効率を上げている。

長方形や楕円の形を特徴づけるために、長方形のサイズを  $2a \times 2b$  と定め、楕円の長軸を  $2a$ 、短軸を  $2b$  と定めて  $k = a/b$  をアスペクト比として導入する。 $k$  が一つのパラメータとなりうる点が非球粒子系の特徴である。発表では、合同な長方形・楕円の粒子系のモンテカルロシミュレーションからつくった平衡配置のサンプルを示した。また、空間構造の特徴づけのために、2次モーメント量である動径分布関数を、いくつかのアスペクト比といくつかの密度に対して求めた結果を示し、現実に観察された配置データに対するデータ解析への適用可能性について論じた。

### パソコン UNIX による統計学研究環境

丸山直昌

IBM-PC 互換パソコンに UNIX を実装し、文書作成ソフト、数式処理ソフト、統計計算ソフトをコンピュータの素人でも使えるように実装して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最近手掛けている。この構想の有用性と実現可能性について述べた。

数年前はいわゆるワークステーション (Engineering WorkStation) の独壇場であった UNIX も、ここ 2, 3 年のコンピュータのめざましい高性能化、小型化によって、低価格のパソコンで使える状況になりつつある。特に IBM-PC パソコンは、ハードウェアの実際上の業界標準の役割を果たしており、その上で動く無償の UNIX が最近いくつか発表されている。このような動きは、今まで職場では UNIX、自宅に帰ると MSDOS というような不便な使い分けを強いられていた人々に取っては大きな恩恵である。最近は大容量の内蔵ディスクを持った IBM-PC 互換のノートパソコンもあり、これに UNIX を実装して必要なソフトウェアを入れれば、持ち運びもできる UNIX 計算機環境が手に入る時代になった。

筆者は現在東芝のノートパソコン DynaBook 上に FreeBSD.1.1.5.1 を搭載して開発を行っており、Ethernet adaptor のドライバソフトにいくらかてこずったものの、現在順調に仕事は進んでいる。また、同機種のパソコンを持つ研究者に出来上がったソフトウェアを丸ごと配布するプランも考えている。その場合、基本となる UNIX が無償であることは重要である。また、配布する統計ソフトウェアは統数研で開発されたものを多く採り入れれば、統数研の研究成果普及にも役立つことになる。

### 3 母数逆ガウス型分布について

金藤浩司

Iwase and Kanefuji(1994) に於て単位を持つものに対する対数変換を行わない形で、原点未知の 3 母数対数正規分布の新しい母数の導入を提案した。これは新しく無名数の母数を導入することによって、統一的な 3 母数化の方法を提案したことになっている。また、3 母数対数正規分布では、この定義によって各母数の最尤推定量が満たす 3 つの最尤方程式と 3 つのモーメント推定量の方程式が完全に一致した。

そこで、対数正規分布で行った 3 母数化の方法を信頼性に於て用いられる逆ガウス型分布、ガンマ分布およびワイブル分布にも適用した。これらの 3 つの分布に於ても、最尤推定量が満た

す方程式とモーメント推定量の方程式が完全に対応した。原点未知の分布に従う確率変数  $X$  を、以下のように定義する。

対数正規分布	逆ガウス型分布
$(1 + \lambda \frac{X - \mu}{\sigma}) \sim LN(1, \lambda^2)$ $X \sim LN_{\lambda}(\mu, \sigma^2)$	$(1 + \lambda \frac{X - \mu}{\sigma}) \sim IG(1, \lambda^2)$ $X \sim IG_{\lambda}(\mu, \sigma^2)$
ワイブル分布	ガンマ分布
$(1 + \lambda \frac{X - \mu}{\sigma}) \sim Weibull(1, \lambda^2)$ $X \sim Weibull_{\lambda}(\mu, \sigma^2)$	$(1 + \lambda \frac{X - \mu}{\sigma}) \sim Ga(1, \lambda^2)$ $X \sim Ga_{\lambda}(\mu, \sigma^2)$

ここで、 $\text{sgn}(\lambda) \cdot \{X - (\mu - \frac{\sigma}{\lambda})\} > 0$ ,  $-\infty < \mu < \infty$ ,  $0 < \sigma < \infty$ ,  $0 < |\lambda|$  である。また、ワイブル分布を除き2母数分布の表現は岩瀬・平野(1990)に従う。

ワイブル分布の確率素分は、以下のように現時点では表現している。

$$Weibull(\mu, c^2) : \frac{1}{c} \left(\frac{x}{\mu}\right)^{\frac{1}{c}-1} \exp\left[-\left(\frac{x}{\mu}\right)^{\frac{1}{c}}\right] \frac{dx}{\mu}$$

また、2母数分布と3母数分布との関係を以下に示す。

$$LN_{\lambda}(\mu, \sigma^2) : \begin{cases} = LN(\mu, \lambda^2) & : \lambda = \frac{\sigma}{\mu}, \quad 0 < \mu < \infty, \\ \rightarrow N(\mu, \sigma^2) & : \lambda \rightarrow 0. \end{cases}$$

$$IG_{\lambda}(\mu, \sigma^2) : \begin{cases} = IG(\mu, \lambda^2) & : \lambda = \frac{\sigma}{\mu}, \quad 0 < \mu < \infty, \\ \rightarrow N(\mu, \sigma^2) & : \lambda \rightarrow 0. \end{cases}$$

$$Ga_{\lambda}(\mu, \sigma^2) : \begin{cases} = Ga(\mu, \lambda^2) & : \lambda = \frac{\sigma}{\mu}, \quad 0 < \mu < \infty, \\ \rightarrow N(\mu, \sigma^2) & : \lambda \rightarrow 0. \end{cases}$$

$$Weibull_{\lambda}(\mu, \sigma^2) : \begin{cases} = Weibull(\mu, \lambda^2) & : \lambda = \frac{\sigma}{\mu}, \quad 0 < \mu < \infty, \\ \rightarrow Gumbel(\mu, \sigma^2) & : \lambda \rightarrow 0. \end{cases}$$

本報告は、統計数理研究所平成6年度共同研究6A-32の一部であり広島大学 岩瀬見盛および統計数理研究所 金藤浩司の担当部分である。

#### 参考文献

- 岩瀬 見盛, 平野 勝臣(1990). べき逆ガウス型分布とその応用, 応用統計学, **19**, 163-176.  
Iwase, K. and Kanefuji, K. (1994). Estimation for 3-parameter lognormal distribution with unknown shifted origin, *Statistical Papers*, **35**, 81-90.

## グラフィックスによる数量化分析

駒澤 勉

円環モデルの3次元グラフィックスにより、ガットマンの完全一次元尺度構造の人工データ、岩坪の円環構造の人工データも円環(ドーナツ)の表面上に表現できることを示した。また、種々の円環モデルの部分集合の人工データに数量化III類を適用すると、それら得られた数量化の結果を3次元散布図に描いた形状がドーナツ型の表面に表現し、各人工データの形状と相似的に類似していることを可視化表示した。さらに、円環モデルは順序構造をもつアイテム・カテゴリ形式データに必ず表現できることも提言した。

終りに、1993年度に実施した「第9次日本人の国民性調査」から、従来の満足度、公平度と水準評価度に加えて、第9次に採用された関連項目に数量化III類を適用して、その結果を3次元カラーグラフィックスで表現することによって、各項目の順序尺度構造および系列的類似性を人工データ同様に、空間配置の形状で示した。なお、分析のうち、「満足度」と「公平度」の6項目について次の結果を示した。

日本人のより満足タイプは、まず一般的な集団社会の場は「公平である」を根底に抱き、次に一般的な場の「社会」に満足を抱き、個人が直接的に関係する集団的な場の「仕事や職場」、個人的な場の「生活全体」、「家庭」にと満足を順次、外から内へ向かって満足感の膨らみが集約される。このタイプが日本人に多く、これに対して、何事にも極度に不満足タイプは内から外に向かって、直接的な個人生活の場から一般的社会生活の場への順序に不満を膨らませ、その延長線で社会に対して不公平感を抱いている。しかし、日本人にはこの不満タイプは回答パターンの頻度から非常に少ないことなどの分析の一部を報告した。

## 非計量データの解析

馬場 康維

質的データ、順位データに関する最近の研究の中から、順序構造を持つカテゴリカルデータから1次元の構造を抽出する方法、順位データに基づくクラスター分析の方法について述べる。

## 1. 質的データの1次元構造の抽出

質的データの解析法の一つとして数量化III類が知られている。順序のあるカテゴリーに数量化III類を適用したとき、特性が多次元空間内の曲線の近傍に分布することが良く知られている(駒澤(1982), 岩坪(1987))。個体と特性に(本来持っているはずの)順序を付けるためには、多次元空間内の曲線に沿った座標系で、個体や特性を並べる必要がある。

多次元空間内のデータの分布に曲線を当てはめる手法の一つにPrincipal Curvesを用いる方法がある(Hastie and Stuetzle(1989))。数量化によって得られた、特性の多次元の布置にPrincipal Curveを当てはめることによって、特性に存在する順序構造を見出すことができる。

## 2. 順位データによる同時クラスタリング

順位法は感覚的な評価ではポピュラーな方法である。ここでは、ランキングデータをもとに、評価されるもの(アイテム)と評価する側(判定者)を分類する方法について述べる。

ランキングによって得られるアイテムの評価を、平均順位と一致度によって表現する方法として「順位グラフ」が知られている。順位グラフによってアイテムは2次元の平面に布置される。



判定者が一つのアイテムに与えた順位を、アイテムに与えられる順位の分布からのランダムサンプルとみなすことによって、アイテムを表わす点の信頼領域が求められる。

ある信頼係数 $\alpha$ の下で、信頼領域が重複しているアイテムを同一クラスターに属するものとみなすことにすると、 $\alpha$ を変化させることにより階層的なアイテムのクラスターが得られる。次に、 $\alpha$ の下で定まるクラスターにつけた順位のパターンを判定者の分類に用いることにより、アイテムと判定者の同時分類が可能になる。

### 参考文献

- Baba, Y. (1994). New approaches based on ranking in sensory evaluation, *New Approaches in Classification and Data Analysis* (eds. E. Diday, Y. Lechvallier, M. Schder, P. Bertrand and B. Birtschy), 525-532, Springer, Heidelberg.
- Hastie, T. and Stuetzle, W. (1989). Principal curves, *J. Amer. Statist. Assoc.*, **84**, 502-516.
- 岩坪 秀一 (1987). 『数量化法の基礎』, 朝倉書店, 東京.
- 駒澤 勉 (1982). 『数量化理論とデータ処理』 (林 知己夫 監修), 朝倉書店, 東京.
- 水田 正弘, 馬場 康維 (1993). Principal Curves と数量化 III 類を用いた質的データの1次元構造の抽出, *統計数理*, **41**(1), 1-11.
- 村田 磨理子, 馬場 康維 (1993). 順位グラフによる分類, *計算機統計学*, **6**(1・2), 25-36.

### 構造変化を探る等質性分析について

土屋 隆 裕

「日本人の国民性」調査など、同一の質問項目群に対して複数のデータセットがあるとき、その構造の変化・違いを知るために、各データセットに対して数量化 III 類を適用し、結果の散布図を見比べるという方法が考えられる。ここでは、各カテゴリが  $R$  次元上に布置  $L$  を持ち、それを各データセットごとに異なる  $r (< R)$  次元部分空間  $S(F_i)$  上に射影することによって、各データセットのカテゴリに与えられる数値が得られると仮定する。この仮定の下では、 $F_i$ 、あるいは  $r = R - 1$  のときには、 $F_i' g_i = 0$  を満たす法線ベクトル  $g_i$  を図等で比較することにより、データセット間の構造の類似性が表現される。すなわち、 $F_i$  や  $g_i$  どうしが似ていれば、それらのデータセット間の構造は似ている、ということになる。また、 $L$  の値を調べることににより、データセットの構造間に具体的にどのような差異・変化が見られるのかが示される。

以上の方法を、 $R = 3$ ,  $r = 2$  として、「日本人の国民性」調査データに適用した。その結果、「義理人情」に関する項目群について、1970年代前半までと後半からでは、構造が異なることが見出された。具体的には、 $L$  の三次元散布図を用いることで、前者では義理人情的でないカテゴリ間の関連が強いのにに対し、後者では義理人情的なカテゴリ間の関連が強くなっていることが示された。また、「日本の水準」についての項目群では、1980年代に構造の変化があることが認められた。「伝統対近代」に関する項目群では、構造変化が、同じ方向に、かつ徐々に起こっていることが、 $g_i$  の変化により表わされた。さらに、「伝統対近代」の項目群について、年齢層別にデータセットを分割した結果、若い年齢層のデータセットの構造が、調査年を経るごとに、高年齢層に伝わる様子を図として表現することができた。

## 統計情報資源の利用環境について

大 隅 昇

多くの研究機関と同様に、統計数理研究所も長年にわたって収集・蓄積されてきた統計情報資源(共同研究情報, 図書登録情報, 各種出版刊行物など)が多々あり, これらの有効的利用法の検討が必要とされる。一昨年度から「統計メタ情報ナビゲーション・システム」を所内のサブネットワークとして試験的に利用して, 各種統計情報資源のデータベース化や関連ソフトウェアの開発を進めてきた。本年度はこれらの成果を基盤として, 「統計情報資源統合化ハイパーネットワーク・システム」(特別設備費), 「ハイパーデータライブラリ構築のための基礎的研究」(特定研究費)の両プロジェクトに関連する基礎的研究と応用システムの開発を進めた。まず従来の方針に従い, 統計情報資源の利用度や重要度に応じて, 共同研究情報データベース, 図書情報データベース, 刊行出版物・研究成果物データベース(「刊行物データベース」と略称)を主な対象情報資源とし, これらの共有化ならびに一元的な利用環境の構築を実現するシステム開発を進めた。共同研究情報データベースについては既に業務に組み入れられて定常的に利用されており, また過去10年分のデータの充実化が進み, 本年度で全情報の統合的な利用が可能となる。図書情報データベースについても, 複数のPCから構成されるサブネットワークを構築し, この環境下での図書情報管理システムの運用を開始した。さらに, PC用データベースの強化を進めると共に, これをワークステーション上に展開し, 一般利用者を対象とする図書情報データベース・システムの開発も進めた。刊行物データベースについては, Annals, 統計数理, リサーチ・メモランダム等を共通化されたデータスキーマの下に統一的にデータベース化した。以上の統計情報資源データベースは, いずれも共通のリレーショナル・データベース・エンジン(オラクル)により一元的に管理され, またいわゆるクライアント・サーバ・ネットワーク環境を利用して異機種クライアントから共通のインターフェースにより参照できるような利用環境を実現している。一方, 当面はこうしたデータベース化を要しない情報資源(研究所の歴史沿革, 年報紹介, 統計数理ニュース, その他の諸活動情報等)については, テキスト・ファイル, イメージ(静止画, 動画)による電子化を進め, これも別のメディア形態として利用可能とすることを検討した。このように, 保存形態やメディアの種類が多様であるという特徴を持つ多量の情報資源を一元的に共有し, またそれより一般的な利用環境, とくに広域ネットワークの利用を配慮した情報公開を具体的に実現するには, ハイパーテキスト的な機能を備えたユーザ・インターフェースを提供するWWW(World Wide Web)やインターフェース用ブラウザ(Mosaic, Netscapeなど)を活用することが効果的である。これを用いた実験作業として, 多数クライアントへの統計情報資源サービスを目的とするハイパーデータライブラリとその運用システムであるハイパーネットワーク・システムのプロトタイプを試作を行った。

## 南勢町コホート研究における生活習慣調査

高 木 廣 文

循環器疾患による罹患, 死亡のリスク要因を査定し, 予防するため, また早期にスクリーニングを行うために, 三重県度会郡南勢町住民を対象としたコホート研究を実施している。調査は三重大学医学部内科学第1講座, 国立健康・栄養研究所および南勢町立病院との共同で行われている。調査期間は, 平成5年4月1日から平成18年3月31日までの13年間である。調査

対象者数は、リスク因子の検出が可能なように統計学的に検討した。その結果、相対リスクが3程度以上のリスク因子を検出できるように、平成5年4月1日現在で、40歳以上65歳以下の4910名とした。調査開始から初めの3年間で、対象者のベースラインの情報を収集する予定である。主な調査項目は、コレステロール、グリコヘモグロビンなどの血液学的検査値、糖尿病、高血圧などの既往歴、身長、体重などの身体状況、および質問紙による生活習慣調査である。用いた生活習慣調査票は1980年頃から、日常生活の各種習慣を測定するための用具として開発を進めてきたものであり、現在では136項目22尺度を測定できるように構成されている(Takagi et al.(1991))。22尺度には塩分、糖分などの食生活に関する尺度、外向性などの人格特性に関する尺度、疾病頻度などの健康に関する尺度、娯楽などの社会生活に関する尺度などから構成されている(高木 他(1993))。さらに、喫煙習慣、飲酒習慣などについても情報を得られるように、作成されている。

今後はコホートでの特定の疾患への罹患・死亡などの発生状況と各生活習慣尺度得点、各種検査値などとの関係を、比例ハザードモデルなどを用いて検討する予定である。

### 参考文献

- Takagi, H., Saiki, K., Hinohara, S., Yanai, H., Douba, N. and Mizuguchi, M. (1991). Construction of the life habits inventory, *Methods and Applications in Mental Health Surveys: The Today Health Index* (eds. S. Suzuki and R.E. Roberts), 103-121, 297-306, University of Tokyo Press, Tokyo.
- 高木 廣文, 金子 俊, 佐伯 圭一郎, 西山 悦子, 平野 真澄, 道場 信孝, 日野原 重明(1993). 質問紙を用いた食塩摂取量推定について, 民族衛生, 59(3), 113-122.

### 点過程コーホート解析

尾形良彦

疫学的現象などの事象発生データのデータが(発生時刻, 患者の年齢)の点配置の形で与えられているとき, その発生率の変化を時刻, 年齢およびコーホートの3成分のリスクに分解する点過程モデルを提案し, 実例のデータ解析を報告した。

時刻  $t$  に, 年齢  $a$  のものの或る事象が発生する intensity を  $\alpha(t, a)$  としたとき上記の3成分についての log-linear 分解

$$\log \alpha(t, a) = \mu + f_T(t) + f_A(a) + f_C(t - a)$$

を考察する。但し,  $\mu$  は定数,  $f_T, f_A$  および  $f_C$  はそれぞれ, 時刻, 年齢およびコーホートの成分関数である。

一方, この種のデータは長年に渡る事が多く, 欠測などのため事象の検出率が一律でなく, データの不均質性も考慮しなければならない。検出率関数を  $\eta(t, a)$  とすると, 実際に検出された事象の intensity は

$$\lambda(t, a) = \alpha(t, a) \cdot \eta(t, a)$$

で与えられる。予め検出率関数  $\eta(t, a)$  を推定しておき, リスク分解を実行するにあたって,  $\lambda(t, a)$  に関する尤度関数を書く事ができる。しかし, 意味のある3成分のリスク分解の解を, 最尤法や AIC 最小化法で求める事は subset-regression の最適解を求める場合と同じ理由で困

難である事が分かる。結局、中村隆(中村(1982), Nakamura(1986))がカテゴリカルデータについてコーホート分解したときのように、それぞれの成分に関して平滑化事前分布を仮定して ABIC 最小化の解を求めるのが実用的である。

解析例としては Green et al. (1981) が追跡調査をしたデンマークの一地方における糖尿病の患者のうち、Keiding et al. (1989) が解析を行ったのと同じ 710 (男 410, 女 300) の例である。発病に関するコーホート解析を上記のモデルによって行った結果、男女例ともコーホート成分の変動は見られなく、増加傾向にあった時代(時刻)成分も最近には頭打ちとなり下降傾向にある。年齢成分は男と女とでは異なっている事が分かる。

最後に、(時刻, 年齢)の空間での intensity を考え、これを直接的に平滑化した結果と、上記コーホートモデルで推定された3成分を合成した intensity との比較を試みた。等方性(isotropic)をみたく平滑 prior と非等方性(anisotropic)のものとは後者の方が適合性が良くコーホートモデルによる intensity と類似している事が分かった。

### 参考文献

- Green, A., Hauge, M., Holm, N.V. and Rasch, L.L. (1981). Epidemiological studies of diabetes mellitus in Denmark: II. A prevalence study based on insulin prescriptions, *Diabetologia*, **20**, 468-470.
- Keiding, N., Holst, C. and Green, A. (1989). Retrospective estimation of diabetes incidence from information in a prevalent population and historical mortality, *Amer. J. Epidemiol.*, **130**, 588-600.
- 中村 隆(1982). ベイズ型コウホート・モデル — 標準コウホート表への適用, *統計数理*, **29**, 77-97.
- Nakamura, T. (1986). Bayesian cohort model for general cohort table analyses, *Ann. Inst. Statist. Math.*, **38**, 353-370.