統計数理（2004）
第 52 巻 第 1 号 217–254

平成15年度研究報告会要旨

と　き：2004 年 3 月 18 日　午前 9 時 30 分〜午後 5 時 15 分
19 日　午前 9 時 45 分〜午後 5 時 15 分
ところ：統計数理研究所　講堂

プ ロ グ ラ ム

3 月 18 日（木）

あいさつ　　**所長　北川源四郎**

【統計基礎研究系】
待ち時間分布の研究　　**平野　勝臣**
チューブ法の誤差の漸近評価　　**栗木　　哲**
拡散過程の変化点問題　　**西山　陽一**
修正 K-L **情報量と修正** χ^2**-ダイバージェンスの近似同等性**　　**松縄　　規**
混獲されたイルカ・サメの数の解析　　**南　美穂子**
遺伝子発現データの統計解析の問題点と展望　　**江口　真透**
Subexponential Class **の拡張について**　　**志村　隆彰**
極値分布と極値データ解析　　**（客員，神戸大学）高橋　倫也**
Optimal Control Problem Associated with Jump Processes　　**（客員，愛媛大学）石川　保志**

【統計科学情報センター】
統計的評価法について　　**石黒真木夫**
観測精度を考慮した多変量解析　　**馬場　康維**
統計科学における E-learning **の一考察**　　**金藤　浩司**
外乱と応答──非ガウス性の繰込み──　　**岡崎　　卓**
Hyperparameter Estimation in MAP Image Reconstruction
（客員，ヴロツワフ工科大学）Rafal Zdunek

【調査実験解析研究系】
環境データの解析　　**柏木　宣久**
第 11 **次国民性調査の企画について**　　**坂元　慶行**
郵送調査の返送を規定する要因について　　**前田　忠彦**
空間分割の統計とその発見的応用　　**種村　正美**
インターネットにおける抽籤と順列の列　　**丸山　直昌**
環境傾度に沿って変化する樹木の分布パターン　　**島谷健一郎**
コウホート分析に基づく将来推計(2)　　**中村　　隆**
予測分布の解析　　**伏木　忠義**

インターネット環境下での定性調査の課題 大隅 昇
子どもを対象とした自記式調査法の特性について 土屋 隆裕
統計科学関連WWWサイトの現状について 清水 信夫
地震活動変化と地殻の歪み変化と統計モデル――2003年の活動 尾形 良彦

【統計科学情報センター】
債権回収率モデルの深い闇 山下 智志

3月19日（金）

【予測制御研究系】
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
自己修復機能付きデータベースシステム 樋口 知之
生産関数のノンパラメトリック推定 川崎 能典
分子系統学の最近の話題 長谷川政美
遺伝子情報解析 足立 淳
Nearest Neighbor ARX Model with Application to Dynamic Brain Functioning Analysis 尾崎 統
データ同化にもとづくエルニーニョの発生予測 上野 玄太
'Universal' Induction Machine dPLRM 田邉 國士
マルチカノニカル法による状態空間のLandscapeの研究 伊庭 幸人
家系間の世代の同定について 上田 澄江
半正定値計画問題による密度関数推定について 土谷 隆
モンテカルロフィルタを用いた金融時系列分析 佐藤 整尚
グリッド環境に適した遺伝的アルゴリズムについて 染谷 博司
アナログデータ処理とNeuron MOSによるスペクトル拡散通信の研究 瀧澤 由美

【領域統計研究系】
学習指導要領に準拠した項目プール 柳本 武美
Adaptive Parameter Estimation Both for Robustness and Efficiency 藤澤 洋徳
A Pythagorean Relationship in the Conjugate Analysis 大西 俊郎
繰り返し3人一般化ジャンケンゲームの漸近安定性 伊藤 栄明
刑事事件に係る文章の計量分析 村上 征勝
東アジア価値観調査――第2年次報告―― 吉野 諒三
Lattice データのHotspots検出 （客員，岡山大学）栗原 考次
分子シミュレーション専用計算機の開発 （客員，理化学研究所）泰地真弘人

【統計計算開発センター】
覚せい剤乱用者調査について 田村 義保
時系列のスペクトル解析について 荒畑恵美子
統計解析システムJaspにおける並列処理 中野 純司
半無限計画法とその周辺 伊藤 聡

待ち時間分布の研究

平 野 勝 臣

本年度の研究

1. **多値マルコフ系列において，長さ有限の複数のパターンのうちどれかがはじめて起こるまでの待ち時間分布**(Han and Hirano (2003a))，**長さ有限の任意のパターンがはじめて起こるまでに，その部分パターンが起こる回数の分布が幾何分布に従う条件**(Hirano and Aki (2003))，**などを調べた．**
2. n **人でジャンケンをはじめ，以後，勝者でジャンケンを繰り返す．一人の勝者になるまでのジャンケンの回数の厳密分布とジャンケンの性質を調べた**(平野・安芸 (2003))**．**

以下，ここでは Han and Hirano (2003b)**に基づき，**{0,1}-**値マルコフ系列において，長さ** k **のウインドウを系列に沿って動かしたとき，その中に** $k-1$ **個以上の** 1 **がはじめて入るまでの試行数(殆ど一致がはじめて起こるまでの待ち時間)について述べる．引用など詳しいことはこの論文を参照されたい．**

殆ど一致の待ち時間分布

$X_1, X_2, \cdots$ **を** $\{0,1\}$-**値マルコフチェインとし，初期確率を** $\Pr(X_1=1)=p_1, \Pr(X_1=0)=p_0$ **とし，推移確率行列** $\mathbf{P}$, j-step **推移確率行列** $\mathbf{P}^j$, $\mathbf{P}^0$ **をそれぞれ**

$$\mathbf{P}=\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},\ \mathbf{P}^j=\begin{pmatrix} p_{00}^{(j)} & p_{01}^{(j)} \\ p_{10}^{(j)} & p_{11}^{(j)} \end{pmatrix},\ j=0,1,2,\cdots,\ \mathbf{P}^0=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

とする．このとき，このマルコフ系列において，殆ど一致がはじめて起こるまでの待ち時間(試行数)W **の確率生成母関数** $\phi(x)$ **は**

$$\phi(x)=\frac{\{p_0xp_{01}x+p_1x(1-p_{00}x)\}(1-p_{11}x)g(x)}{(p_{10}xp_{01}x)g(x)+(1,(p_{11}x)^{k-3}p_{10}xp_{01}x)(\prod_{j=2}^{[\frac{k-1}{2}]}A_j)\beta(k)h(x)}$$

で与えられる．ここに

$$g(x)=(p_{11}x)^{k-3}p_{10}xp_{01}x\left\{\left(0,\frac{p_{11}x}{p_{10}xp_{01}x}\right)\left(\prod_{j=2}^{[\frac{k-1}{2}]}A_j\right)+(1,1)\right.$$
$$\left.-\frac{1}{2}(1,1)I(k \text{ is an odd })+(1,1)\sum_{i=2}^{[\frac{k-1}{2}]}\left(\prod_{j=i}^{[\frac{k-1}{2}]}A_j\right)\right\}\beta(k),$$

$$A_i=\begin{pmatrix} 1 & (p_{11}x)^{k-i-2}p_{10}xp_{01}x \\ -(p_{11}x)^{i-2}p_{10}xp_{01}x & 1-(p_{11}x)^{k-4}(p_{10}xp_{01}x)^2 \end{pmatrix},$$

$$\beta(k)=\begin{cases} \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \text{if } k \text{ is an odd,} \\ \begin{pmatrix} 1 \\ 1-(p_{11}x)^{\frac{k}{2}-2}p_{10}xp_{01}x \end{pmatrix} & \text{if } k \text{ is an even} \end{cases}$$

である．

p_0 と p_1 をそれぞれ p_{10} と p_{11} に置換えれば，オーバーラップしない数え方で殆ど一致間の分布の確率生成母関数を得ることができる．さらに $X_1, X_2, \cdots, X_n$ において，殆ど一致の起こる回数(non-overlapping)の分布の確率生成母関数も得ることができる．

この分布の応用例を述べると，連続システムのひとつである(k-1)-within-consecutive-k-out-of-n system with dependent components の信頼度は $\Pr(W \leq x)$ を使って計算できる．また，2つの遺伝子系列の一致について調べるとき，文字パターンが完全に一致しなくても同じ機能を持つことが知られている(TATA box, DnaA box, *etc.*)．

参 考 文 献

Han, Q. and Hirano, K. (2003a). Sooner and later waiting time problems for patterns in Markov dependent trials, *Journal of Applied Probability*, **40**, 73–86.

Han, Q. and Hirano, K. (2003b). Waiting time problem for an almost perfect match, *Statistics and Probability Letters*, **65**, 39–49.

平野勝臣，安芸重雄(2003)．ジャンケンの厳密な待ち時間分布と性質，統計数理，**51**(1)，167–172．

Hirano, K. and Aki, S. (2003). Number of occurrences of subpattern until the first appearance of a pattern and geometric distribution, *Statistics and Probability Letters*, **65**, 259–262.

チューブ法の誤差の漸近評価

栗 木 哲

単位球面 S^{m-1} の部分多様体 M で定義された正規確率場 $Z(x) = \langle \xi, x \rangle$, $x \in M$, $\xi \sim N(0, I_m)$, の最大値分布 $P(u) := \Pr(\sup_{x \in M} Z(x) \geq u)$ と，M のまわりの半径 θ の球面チューブの体積 $V(\theta) := \mathrm{Vol}(M_\theta)$ はラプラス変換，逆変換の関係にある：

$$\frac{P(u)}{u^m e^{-\frac{u^2}{2}}} = \frac{1}{2(2\pi)^{\frac{m}{2}}} \int_0^\infty V\Big(\cos^{-1} \frac{1}{\sqrt{\eta+1}}\Big)(\eta+1)^{\frac{m}{2}-1} e^{-\frac{u^2 \eta}{2}} d\eta. \tag{1}$$

チューブ座標を用いた計算によって，チューブの体積の近似公式 $\widehat{V}(\theta)$ が得られる．$\widehat{V}(\theta)$ は $\theta \in [0, \theta_c]$ (θ_c は臨界半径)の範囲で $V(\theta)$ と等しい．チューブ法とは，$V(\theta)$ の代わりにその近似値 $\widehat{V}(\theta)$ を(1)式右辺に代入し，最大値分布の近似 $\widehat{P}(u)$ を求める方法である．

いま，臨界半径をわずかに超えたときのチューブ体積公式の誤差が

$$\widehat{V}(\theta_c + \varepsilon) - V(\theta_c + \varepsilon) \sim c \cdot \varepsilon^\nu \qquad (\varepsilon \downarrow 0)$$

であるとき

$$\widehat{P}(u) - P(u) \sim d \cdot u^{-(2(\nu+1)-m)} e^{-\frac{u^2}{2}(1+\tan^2 \theta_c)} \qquad (u \to \infty),$$
$$d = c \times \frac{\nu!}{(2\pi)^{m/2}} \tan^{-1} \theta_c \, \cos^{2(\nu+1)-m} \theta_c,$$

である．定数 c, ν が分かれば，チューブ法近似の誤差の漸近評価が得られる．

定理．M は一次元で，弧長パラメータを用いて $M = \{x(s)\}$ とパラメトライズされているとする．臨界半径が大域的で，点のペア (p, p') でのみ達成される場合

$$\widehat{P}(u) - P(u) \ \sim \ d \cdot u^{-2} e^{-u^2(1+\tan^2 \theta_c)/2},$$

$$d = \frac{\tan^{-1}\theta_c \cos^2\theta_c}{2\pi}\left\{\frac{(1-\langle\ddot{x}_0, x_1\rangle)(1-\langle x_0, \ddot{x}_1\rangle)}{\langle\ddot{x}_0, x_1\rangle\langle x_0, \ddot{x}_1\rangle - \langle\dot{x}_0, \dot{x}_1\rangle^2}\right\}^{1/2},$$

$p = x_0$, $p' = x_1$, $\langle x_0, x_1\rangle = \cos(2\theta_c)$.

臨界半径が局所的に与えられる場合の誤差評価は今後の課題である．

拡散過程の変化点問題

西 山 陽 一

Θ は $\mathbb{R}$ の有界部分集合であるとする．確率微分方程式

$$dX_t = V_0(X_t, \theta)dt + V(X_t)dW_t$$

を考える．ただし W は標準 Wiener 過程であるとする．仮説検定問題

$$H_0 : \theta \text{ は } t \in [0,T] \text{ において変化しない}$$
$$H_1 : H_0 \text{ でない}$$

を考える．$\widehat{\theta}_t$ は時間区間 $[0,t]$ の上の観測に基づく推定量であるとし，統計量 $S_T = \sup_{u\in[0,1]} u\sqrt{T}|\widehat{\theta}_{uT} - \widehat{\theta}_T|$ を考える．単純に言えば，$\widehat{\theta}_t$ として最尤推定量を採用すればよいのではないかと考えられるが，それは簡単な具体例をもって不適切であることがわかる．代わりに，one-step 推定量を用いる．

まず，対数尤度比は

$$\ell_T(\theta) = \int_0^T \frac{V_0}{V^2}(X_t,\theta)dX_t - \frac{1}{2}\int_0^T \left(\frac{V_0}{V}\right)^2(X_t,\theta)dt$$

によって与えられる．次に，$\psi_T(\theta)$ と $\eta(\theta)$ を以下のように定義する：

$$\begin{aligned}\psi_T(\theta) &:= \partial_\theta \ell_T(\theta)\\ &= \int_0^T \frac{\dot{V}_0}{V^2}(X_t,\theta)dX_t - \int_0^T \frac{V_0\dot{V}_0}{V^2}(X_t,\theta)dt\\ &= \int_0^T \frac{\dot{V}_0}{V}(X_t,\theta)dW_t;\\ \eta(\theta) &:= -E_\theta\left[\left(\frac{\dot{V}_0}{V}\right)^2(X_0,\theta)\right].\end{aligned}$$

ここで $\widehat{\theta}^0_T$ は与えられた初期推定量であるとする．このとき，one-step 推定量は $\widehat{\theta}_T = \widehat{\theta}^0_T - [T\eta(\widehat{\theta}^0_T)]^{-1}\psi_T(\widehat{\theta}^0_T)$ と定義される．ここで，

$$\widehat{\theta}_T - \theta_0 = -\frac{1}{T}\eta(\theta_0)^{-1}\psi_T(\theta_0) + \Delta_T$$

と書けることに注意しよう．ただし，$\Delta_T = \Delta_T^{(1)} + \Delta_T^{(2)} + \Delta_T^{(3)}$ であり，また

$$\Delta_T^{(1)} = (\widehat{\theta}^0_T - \theta_0) - \frac{1}{T}\eta(\theta_0)^{-1}\dot{\psi}_T(\theta_0)(\widehat{\theta}^0_T - \theta_0)$$
$$\Delta_T^{(2)} = \frac{1}{T}\eta(\theta_0)^{-2}\dot{\eta}(\theta_0)\psi_T(\theta_0)(\widehat{\theta}^0_T - \theta_0)$$

$$\Delta_T^{(3)} = -\frac{1}{T}\int_0^1 (1-\alpha)\partial_\theta^2(\eta^{-1}\psi_T)(\theta_0+\alpha(\widehat{\theta}_T^0-\theta_0))d\alpha \times(\widehat{\theta}_T^0-\theta_0)^2$$

である．検定統計量は，$S_T = \sup_{u\in[0,1]} u\sqrt{T}|\widehat{\theta}_{uT}-\widehat{\theta}_T|$ である．ここで，

$$\begin{aligned}\widehat{\theta}_{uT}-\widehat{\theta}_T &= (\widehat{\theta}_{uT}-\theta_0)-(\widehat{\theta}_T-\theta_0)\\ &= \frac{\eta(\theta_0)^{-1}}{T}\left\{-\frac{1}{u}\psi_{uT}(\theta_0)+\psi_T(\theta_0)\right\}\\ &\quad +(\Delta_{uT}-\Delta_T)\end{aligned}$$

が成り立つ．このとき，緩い条件のもとで，

$$Y_T^u = \frac{\eta(\theta_0)^{-1}}{\sqrt{T}}(\psi_{uT}(\theta_0)-u\psi_T(\theta_0))$$

によって与えられる確率過程 $u \rightsquigarrow Y_T^u$ の弱収束と，

$$\sup_{u\in[0,1]} u\sqrt{T}|\Delta_{uT}| = o_P(1)$$

であるという主張が証明できる．

ちなみに，$t \rightsquigarrow \psi_t(\theta)$ は連続マルチンゲールであり，よって汎関数中心極限定理が適用できる．一方，後者に関しては，テクニカルな計算が必要となる．

修正 K-L 情報量と修正 χ^2-ダイバージェンスの近似同等性

松 縄 規

修正 χ^2-ダイバージェンス(or W-ダイバージェンス)を修正 K-L 情報量を用いて，定性的及び定量的近似を理論的に考察し応用を図った．そのために，関連諸積分の発散を回避する工夫を行い，χ^2-ダイバージェンスを，適切に設定された近似主領域上で考え，それを上下から，修正 K-L 情報量により精密に評価する不等式を与えた．(R,B) を抽象可測空間とする．R は任意の抽象空間，B は R の部分集合の σ-集合体を表す．X と Y をこの空間上で定義される確率変数とし，P^X および P^Y をそれぞれ，X と Y の確率分布とする．μ をこの空間 (R,B) 上で定義される σ-有限測度とする．A を B に属する可測集合とし，f および g は各々，P^X および P^Y の A 上で定義される，測度 μ に関するラドン・ニコディム導関数を表す．修正 χ^2-ダイバージェンスを

$$W^*(X,Y;A) := \int_A ((f/g)-1)^2\cdot g d\mu$$

で，修正 K-L 情報量を

$$I^*(X,Y;A) := \int_A f\ln(f/g)d\mu$$

で定義する．この時次の不等式が成立することを示した．

$$\begin{aligned}0 &\le I^*(X,Y;A)+(P^Y(A)-P^X(A)) \le W^*(X,Y;A)\\ &\le u\left(\sup_{A^+}(f/g)\right)\cdot I^*(X,Y;A)+(P^Y(A)-P^X(A))\end{aligned}$$

ただし，$A^+ = \{x; f(x) \geq g(x) > 0\}$,

$$u(t) = \frac{5(1+t^{1/3})^3(1+t^{1/3}+t^{2/3})(1+12t^{1/3}+t^{2/3})}{61+436t^{1/3}+686t^{2/3}+436t+61t^{4/3}}, \quad (t>0).$$

また，任意の $\varepsilon > 0$ に対し，可測集合 $A \in \mathrm{B}$ が存在して，$\min\left\{P^Y(A), P^X(A)\right\} \geq e^{-\varepsilon}$ が成立するものとし，$A := \{x \in R; |\ln(f(x)/g(x))| \leq \varepsilon\}$ と設定する．この時

$$W^*(X,Y;A) \geq \max[2 \cdot I^*(X,Y;A) + (e^{-\varepsilon} - 1 - \varepsilon) \cdot e^{-\varepsilon}, 0]$$
$$W^*(X,Y;A) \leq \{u(e^{\varepsilon}) + 1\} \cdot I^*(X,Y;A) + e^{\varepsilon} - 1 + \varepsilon$$

が成り立つことを示した．

また，I^*-近似同等性と a.s.-近似同等性の強さの間には一般に包含関係が成立しないことを例示した．応用として，単位区間のランダム・カバレッジについて，その個数をサンプルサイズと共に増加させて選択する際に，同時分布が修正情報近似の意味で近似独立分性を示すための選択個数の条件を与えた．

混獲されたイルカ・サメの数の解析

南　美穂子

東部太平洋沿岸においてはマグロの群れはイルカと共にいることが多く，マグロ漁もこれを利用した巻網漁が行われている．国際協定による規制で，各漁船の年間累積イルカ混獲数が一定値を超えるとその船は操業ができなくなるため，イルカを逃すよう可能な限りの努力は払われるがそれでもイルカが混獲される場合がある．全米熱帯マグロ類委員会(IATTC)はイルカ混獲規制などを実施する国際機関で，各マグロ漁船に監視員を派遣し，混獲されたイルカやサメの数，気象条件や漁に関する様々なものを計測している．

本研究はこれらのデータを用いて混獲されたイルカの数の解析，および，混獲されたサメの数の解析を行うものであり，IATTC の研究者との共同研究である．イルカに関しては，Lennert-Cody et al. (2004)で 1993 年から 2001 年までのデータをもとに混獲数に影響を与える要因を探った．

サメの場合，漁の約 70%で混獲数は 0 である．イルカはマグロ漁の目印となっているがサメは偶然に捕獲されるものであるから，混獲されたサメの数が 0 である場合の多くは，サメがマグロの群れと一緒にいなかったからであると推測される．そこで，サメがマグロの群れと一緒にいたかどうかにロジステック回帰モデル，一緒にいた場合の混獲数にポアソン回帰モデルを用いた Zero-Inflated ポアソン回帰モデルを考える．共変量としては，緯度，経度，年月日，えさの漁獲量などがあるが，緯度，経度に関しては 2 次元平滑化スプライン，その他は 1 次元平滑化スプラインを用いる．多次元平滑化スプラインはサンプル数の 3 乗のオーダーの計算量が必要という欠点があるが，それを補うために Wood (2003)による thin-plate regression splines を用いる．推定値の計算は EM アルゴリズムを用いると既存のプログラムを用いて簡単に行える．

参　考　文　献

Lennert-Cody, C., Minami, M. and Hall, M.A. (2004). Incidental mortality of dolphins in the Eastern

Pacific Ocean Purse-Seine Fishery: Correlates and their spatial association, *The Journal of Cetacean Research and Management* (to appear).

Wood, S. N. (2003). Thin-plate regression splines, *Journal of Royal Statistical Society. Series B*, **65**(1), 95–114.

遺伝子発現データの統計解析の問題点と展望

江口真透

最近，ゲノム科学の分野から，活発な研究活動に伴い膨大なデータが生産されるようになった．マイクロアレイの方法によって大量の遺伝子発現を定量的に同時計測することや，SNPs (一塩基置換)も多様な遺伝子座において同時タイピングすること，また最近，広がりを見せているプロテオームの波形データも同様である．このような高次元ゲノムデータから，疾病，薬剤感受性の関連遺伝子の発見の問題は，単純に統計的パタン認識の問題に帰着される．しかしながら例題数(サンプル数)n と特徴次元 p (遺伝子数)の不均衡から困難な問題が生じている．典型的には n は数十であるのに対して p は数千から数万のオーダーである．通常の統計解析では多数の見せ掛けの関連遺伝子が見つかってしまい，その中に真の関連遺伝子が埋没されてしまう．具体的にはトレーニングエラーはほとんど 0 にできるが，テストエラーは数十パーセントを下回ることができないということである． 現在，キーワード“ $n << p$ ”問題として様々なアプローチが挑戦されている．本報告ではアダブーストの基本設計を変更して新たに「グループブースト」を提案し，理論的考察と幾つかの実データの解析を示した．グループブーストとは“ 重み付エラーレイトを最小にする判別マシンの選択ステップ ”をとるのではなく,“ 重み付エラーレイトに関して上位 g の判別マシンの選択ステップ ”への変更である．こうして選ばれた g 個からなる判別マシンのグループの平均をとることによって次ステップに更新する．このようにして，最も優秀な判別マシンだけに注目するのではなく，上位 g の判別マシンの平均を取ることが特徴である．グループブーストのロス関数についての理論的結果によって，この方法が支持された．実解析では公開されている遺伝子発現データに対して行った．予想に反してグループブーストの g は比較的大きな数，典型的には 50 から 100 に固定した場合が安定した判別結果を示した．この研究は竹之内高志君(総研大学 D3)との共同研究の一部である．

Subexponential Class の拡張について

志村隆彰

Subexponential distribution は $[0,\infty)$ 上の分布で $\lim_{x\to\infty}\overline{\mu*\mu}(x)/\overline{\mu}(x)=2$ で特徴付けられる ($\overline{\mu}(x)=\mu(x,\infty)$)．その全体を S であらわす．分布族 S は分枝過程，更新過程，乱歩，破産問題等において重要な分布族であり，その研究の発展に伴い，convolution equivalent class をはじめとする S を拡張した分布族や逆に密度をもつなどの制限した分布族が提起，研究されている．

まず，分布族をあらわす記号を示す．すべて $[0,\infty)$ 上の分布とする．L: 任意の $k\in\mathbb{R}$ に対して $\lim_{x\to\infty}\overline{\mu}(x+k)/\overline{\mu}(x)=1$ を満たす分布族．$L(\gamma)(\gamma>0)$: 任意の $k\in\mathbb{R}$ に対して $\lim_{x\to\infty}\overline{\mu}(x+k)/\overline{\mu}(x)=e^{-\gamma k}$ を満たす分布の族．$S(\gamma)(\gamma\geq 0)$: $L(\gamma)$ の分布のうち，

さらに $\lim_{x\to\infty}\overline{\mu*\mu}(x)/\overline{\mu}(x) = \int_0^\infty e^{\gamma t}\mu(dt) < \infty$ となる分布の族(convolution equivalent class)．ただし，$S(0) = S$．OS: $\limsup_{x\to\infty}\overline{\mu*\mu}(x)/\overline{\mu}(x) < \infty$ を満たす分布の族．D : $\limsup_{x\to\infty}\bar{\mu}(x)/\bar{\mu}(2x) < \infty$ を満たす分布の族．M : 平均有限の分布の族．S^* : $M \cap L$ のうち，$\lim_{x\to\infty}\int_0^x \bar{\mu}(x-t)\bar{\mu}(t)dt/\bar{\mu}(x) = 2\int_0^\infty t\mu(dt)$ を満たす分布の族．OS^* : $\limsup_{x\to\infty}\int_0^x \bar{\mu}(x-t)\bar{\mu}(t)dt/\bar{\mu}(x) < \infty$ を満たす分布の族．

定理．これらの分布族の間に次のような関係が成り立つ．包含関係は等しい場合を含まず，真に差があることを意味する．(i) は S の，(ii) は $S(\gamma)(\gamma > 0)$ の系列であり，両者の違いをしめしている．

(i) $D \cap L \cap M \subset S^* \subset OS^* \cap L \subset S \cap M \subset S \subset OS \cap L \subset L$.
(ii) $S(\gamma) \subset OS^* \cap L(\gamma) = OS \cap L(\gamma) \subset L(\gamma)$

この他にも様々な各分布族について，合成積及びその意味で根に関する閉性に関することや γ 変換についても話したが，それらについては参考文献，あるいはこれからまとめる予定の論文をみていただきたい．尚，この研究は渡部俊朗氏(会津大学)との共同研究である．

参 考 文 献

Shimura, T. and Watanabe, T. (2003). *Infinite divisibility and generalized subexponentiality*, Research Memo., No. 885, The Institute of Statistical Mathematics, Tokyo.

志村隆彰，渡部俊朗(2004)．OS に関連するいくつかのクラスについて，統計数理研究所共同研究リポート，No. 170，96–106.

極値分布と極値データ解析

(客員)神戸大学 髙 橋 倫 也

極値理論では，非常に大きな値の T に関する T-return level と呼ばれる極値分布の上側 $1/T$ 確率点の推定が重用である．古典的な方法では，単位領域(または単位期間)ごとの最大データに極値分布の Gumbel 分布または一般極値(GEV)分布を適合させて推定する．これに対して，上位 $r(>1)$ 個までのデータを用いることにより T-return level の推定精度がどの程度改善されるか漸近相対効率を用いて調べた．漸近相対効率は r，T と GEV 分布の形状パラメータに依存する．Gumbel モデル(GEV モデルで形状パラメータが 0)で，r または T を固定した場合，他の変数に関して漸近相対効率は増加関数になることを示した．また，それは r に関してそれ程は増加しないことも示した．同様の事が GEV モデルの場合の漸近相対効率の図でも言える．

一方，形状パラメータが 0 に十分近い GEV 分布からのデータに Gumbel 分布を適合させて T-return level を推定すると，推定値は GEV 分布を適合させるよりも安全側になり，推定精度が良くなる傾向がある事をシミュレーション実験で示した．

参 考 文 献

高橋倫也，渋谷政昭(2004)．上位 r 個の観測値に基づく確率点の推定，統計数理，52(1), 93–116.

Optimal Control Problem Associated with Jump Processes

（客員）愛媛大学 石 川 保 志

$A = A^{\pi,c}$ を次の微積分作用素とする

$$\begin{aligned} Av(x,y) = &-\alpha v - \beta y v_y \\ &+ \left\{ (r + \pi(\hat{b} - r))x v_x + \int (v(x + \pi x(e^z - 1), y) - v(x,y) - \pi x v_x (e^z - 1))\nu(dz) \right\} \\ &+ U(c) - c(v_x - \beta v_y) \end{aligned}$$

さらに

$$\begin{aligned} Nv &= v_x, \\ Mv &= \beta v_y - v_x \end{aligned}$$

とおく．ここで $\beta > 0, U$ は狭義増加，微分可能な凹関数，また $\pi = \pi_\cdot, c = c_\cdot$ はある制御関数である．$\mathcal{S} = \{(x,y); y > 0, y + \beta x > 0\}$ における，これらの作用素に対応した HJB 方程式は

$$\begin{aligned} &\max\{Nv, \sup_{\pi,c}\{Av\}, Mv\} = 0 \quad \text{in} \quad \mathcal{S}. \\ &v = 0 \qquad \text{outside of } (\mathcal{S}). \end{aligned}$$

であたえられる．

A に対応して確率過程 $X = X_\cdot^x, Y = Y_\cdot^y$ を次により構成する．

$$X_t = x - C_t + \int_0^t (r + (\hat{b} - r)\pi_s) X_s ds + L_t + \int_0^t \pi_{s-} X_{s-} \int_{\mathbf{R}\setminus\{0\}} (e^z - 1)\tilde{N}(dsdz), X_0 = x.$$

$$Y_t = y e^{-\beta t} + \beta \int_0^t e^{-\beta(t-s)} dC_s, Y_0 = y.$$

これらから構成される効用関数（value functions）を

$$v(x,y) = \sup_{(\pi_\cdot, c_\cdot, L_\cdot) \in \mathcal{A}} E^{(X,Y)} \left[\int_0^\infty e^{-\alpha s} U(c_s) ds \right]$$

とする．ここで上限は制御過程の集合 $\mathcal{A}$ 内の admissible 制御 $(\pi_\cdot, c_\cdot, L_\cdot)$ について，期待値は (X_t, Y_t) の法則についてとる．

定理．効用関数 $v(x,y)$ は定義され，それは $\bar{\mathcal{S}}$ 上の粘性解になる．

参 考 文 献

Ishikawa, Y. (2004). Optimal control problem associated with jump processes, *Applied Mathematics and Optimization* (to appear).

統計的評価法について

石 黒 真 木 夫

平均 μ, 分散 σ^2 の正規分布からとった N ヶの乱数の最大値 m ヶから元の分布の平均 μ と分散 σ^2 を推定せよ, という問題を考える.「大学評価」などの場面で出て来る問題から抽象した問題である.

たとえば $\mu = 50$, $\sigma^2 = 100$, の分布からそれぞれ $N = 100, 1000$ として乱数を取り出して, 上位 5ヶの数値をとり出すという実験を 10 回繰り返した結果を表 1 と 2 に示す. 明らかに表 2 の方に大きい数字が並んでいる. これを見て表 2 の方の集合の方が「良い集合である」と結論してはいけない. 集合のサイズを考慮しなくてはならない. N ヶの集合からとり出した数字の上

表 1. $N = 100$ の場合.

74.3	67.1	64.7	64.3	63.8
76.6	72.2	70.7	70.5	69.5
74.3	73.3	71.4	69.7	68.4
73.6	72.0	71.4	70.6	66.3
74.3	73.3	70.3	69.8	66.2
75.7	73.1	65.3	65.2	64.1
82.8	73.2	72.8	72.2	70.1
74.2	73.6	72.5	71.6	68.2
80.7	78.2	77.8	75.5	74.1
73.7	71.9	71.6	70.5	68.0

表 2. $N = 1000$ の場合.

81.9	81.9	78.9	78.0	77.6
85.3	78.3	76.9	76.5	75.2
87.4	80.2	79.0	77.6	76.4
80.6	76.5	75.7	74.2	74.1
82.5	80.8	78.3	78.2	76.4
82.1	80.9	78.0	75.9	74.1
83.8	76.4	75.1	75.1	74.2
79.1	78.5	78.0	77.7	76.3
81.4	81.2	79.6	78.7	78.6
78.2	78.1	77.9	77.7	76.7

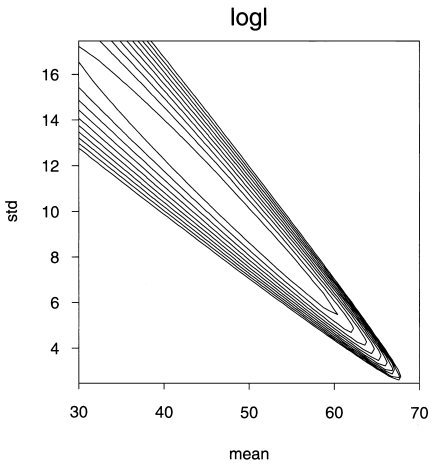


図 1. 表 2 の上から 2 段目のデータに対応する対数尤度関数. $\log f_{1000}\left(85.3, 78.3, 76.9, 76.5, 75.2|\mu, \sigma^2\right)$

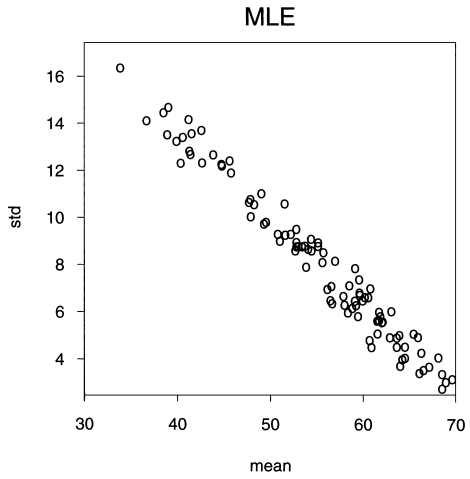


図 2. 100 組のデータそれぞれから求めた μ と σ の最尤推定値.

位 m ヶが $\{x_1, x_2, \ldots, x_m\}$ である場合 μ, σ^2 の尤度関数は, g と G をそれぞれ正規分布の確率密度関数, 分布関数として, 次の式で与えられる.

$$f_N\left(x_1, \ldots, x_m|\mu, \sigma^2\right) = \prod_{i=1}^{m} g\left(x_i|\mu, \sigma^2\right) G\left(x_m|\mu, \sigma^2\right)^{N-m}$$

表 2 の 2 行目のデータに対する μ と σ の対数尤度の等高線を描いてみると図 1 のようになる. 表を見ると結構大きな数字が並んでいるが, 尤度関数は「集合の平均は 50 より小さい. 標準偏差は 10 程度」と言っている.

100 組のデータを作ってみて, それぞれに求めた最尤推定値をプロットしたのが図 2 である. 「上位データ」から全体をおしはかるのは出来なくはないが誤差が大きく難しいということが良く分る. グラフが右下がりなのは上位データからは全体につぶ揃いで大きな値をとる集合なのか, ばらつきの大きい集合からたまたま大きな値が出ているのかを見分けるのは難しいということである.

観測精度を考慮した多変量解析

馬 場 康 維

極めて低濃度の物質の濃度を測定するという場面を考える. 濃度のレベルによって観測結果は次のように分類できる.

A) 濃度が非常に低く, 0 と見なせる
B) 濃度は数値的には求まらないが, 物質の存在は検出できる
C) 濃度が十分高く, 測定値が得られる

A), C)のケースは数値として扱えば良い. しかし, B)の場合が問題である. B)の場合は便法として, 適当な大きさの数値に置き換えて通常のデータとして扱うと言う方法がよく用いられる. しかしこの方法では, 結果にバイアスが生ずる.

ここでは, A), B), C)が混在する場合を想定し

1) 繰り返し観測が行われる場合
2) 繰り返し観測が行われない場合

のそれぞれについての回帰分析の方法を考えた.

外乱と応答──非ガウス性の繰込み──

岡 崎 卓

外乱を受けつつ発展するシステムについて, 外乱の非ガウス性およびシステムの非線形性の繰り込みの観点から応答と外乱の関係を整理し纏めた結果を報告する.

ここに取り上げるシステムは, その変数 $U = (U_1, U_2, \cdots, U_N)$ が外乱 W の作用のもとに

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

に従って発展するものとする. 外乱の種類(白色ガウス, 有色ガウス, あるいは非ガウス)とシ

ステムの構造(線形，あるいは非線形)に応じて応答確率密度 $f(U)$ の解析的表現を求め，外乱の非ガウス性とシステムの構造が $f(U)$ に如何に反映されるかを調べて行く．

1. 白色ガウス外乱

外乱が白色ガウス過程の場合にはシステムの線形性，非線形性を問わず通常の Fokker-Planck (FP)方程式を利用できるから，応答密度 $f(U)$ を外乱の分散とシステムの構造関数 $M(U)$ で容易に表せる．

2. 有色ガウス外乱

システムが線形ならば，FP 方程式により応答密度 $f(U)$ を外乱の分散と 2 点相関関数，およびシステムの構造関数で表現できる． 非線形システムでは結合系 (U,W) に対する FP 方程式を解くのは困難となるが，変数 U のみに縮約した GFP 方程式を用いれば，応答密度 $f(U)$ の表現を前項と同様に得ることができる．

3. 非ガウス外乱──線形システム

外乱が非ガウス過程の場合は，システムが線形であっても もはや FP 方程式を解いて $f(U)$ を求めるのは不可能に近く，GFP 方程式

$$\frac{\partial}{\partial t}f(U,t) = -\ D_U M(U) f(U,t) + D_U D_U \lambda_A(t) f(U,t) + \ D_U D_U \int dV \lambda_B(V,t) f(U-V,t) \qquad \left(D_U = \frac{\partial}{\partial U}\right)$$

に訴えねばならない． 外乱確率密度の特性汎関数を介して 2 種の拡散係数 λ_A と λ_B を外乱の統計量(分散と高次に亘る相関関数群)およびシステムの構造関数 $M_i(U) = -\alpha_{ij}U_j$ (同一添字の反復は $1,2,\cdots,N$ に亘る和を意味する)で表せば，GFP 方程式の解を解析的に表現することができる．すなわち，定常状態における応答確率密度の特性関数 $\hat{f}\ (k)$ は Gauss 外乱に対する応答 $\hat{f}_{Gauss}$ に変調 $e^{\hat{f}_{modL}(k)}$ を施したものとして表される．

$$\hat{f}\ (k) = \hat{f}_{Gauss}\ (k) \cdot e^{\hat{f}_{modL}(k)}$$

$$k_i \alpha_{ij} \frac{\partial}{\partial k_j}\ \hat{f}_{modL}\ (k) + k_i k_j\ \hat{\lambda}_{B\,ij}\ (k) = 0$$

外乱の非ガウス性はシステムの構造関数 $M(U)$ と拡散係数 λ_B を介して変調関数 $\hat{f}_{modL}$ に繰り込まれる．

4. 非ガウス外乱──非線形システム

システムの非線形性が弱い($M(U) = -\alpha U + \gamma M_1(U),\ O(\gamma) << 1$)との前提下に GFP 方程式を具体化すると，拡散項は $\lambda_A(U)f(U,t) + \int dV \lambda_B(U,V) f(U-V,t)$，$\lambda_A(U) \sim \lambda_{AL} e^{\gamma \lambda_{AN}(U)}$, $\lambda_B(U,V) \sim \lambda_{BL}(V) e^{\gamma \lambda_{BN}(U)}$ の形をとる． これに応じて応答確率密度の特性関数 $\hat{f}\ (k)$ は外乱の非ガウス性に因る変調に加えてシステムの非線形性に起因する変調 $\hat{f}_{modN}$ を受けることになる．

$$\hat{f}\ (k) = \hat{f}_{Gauss}\ (k) \cdot e^{\hat{f}_{modL}(k)} \cdot e^{\gamma \hat{f}_{modN}(k)}$$

従って外乱の非ガウス性とシステムの非線形性は，システム構造関数 $M(U)$ と外乱統計量を経由して変調関数 $\hat{f}_{modL}$ と $\hat{f}_{modN}$ にそれぞれ組み込まれることが判る．

第11次国民性調査の企画について

坂元慶行

今年度の活動の中心は,「日本人の国民性 第11次全国調査」等の企画・実施・速報値のとりまとめと,国際会議AIC2003への参加であった.

年度研究報告会では,まず,第11次国民性調査に関連して,調査法の研究のため,電話調査や郵送調査等を含めて5つの調査を行ったことを報告したが,結果公表が可能な状況には至っていなかったので(調査結果の分析は2005年6月刊行予定の『統計数理』等で発表の予定),標題とはややずれるが,後半部では,AIC2003に関連して行った活動について以下のような報告をした.

AIC2003ではCATDAPのweb版のデモンストレーションを行ったが,そこでとりあげた例のうち,「13ヵ国価値観調査」(余暇開発センター,1979年調査)のデータに対して実際にプログラムを起動し,「日本人とアメリカ人とを識別するための質問の探索」を目的として,再度デモンストレーションを行った.この調査データは,幸福感,階層帰属意識,政治的態度,人生観,生活領域ごとの欲求充足度・満足度,家族観,社会への関心度,宗教観,財の保有率と必要度,余暇活動,基本的属性等に関する274項目から成る.そこで,日米以外の国のデータをスキップした上で,国名を目的変数に指定し,プログラムを作動させた結果,たとえば上位10位までの項目はつぎのようになった.「一神教か多神教か」,「子供夫婦と同居するか」,「プロテスタントか否か」,「自分と国家との関係についてどの程度深く考えたことがあるか」,「セントラルヒーティングはあるか」,「宗教的戒律に支配されるべきか」,「仏教徒か否か」,「国の将来を考えているか」,「家庭にどの程度満足か」,「余暇にどの程度満足か」.これらの結果は,日米両国民の識別には,宗教観,国家観,親との同居,満足度(満足度に関する表現の強度)等が有効であることを示唆している.これらは,伝統的な日本人論の,いわば常連とも言える項目であるが,これらが機械的にほとんど瞬時に探索できる意義は小さくない.CATDAPは,目的変数がカテゴリカルでありさえすれば適用可能なデータ・マイナーで,適用範囲は極めて広いが,現在のweb版には若干の手直しを要する点があり,その点を改良した後,一般の利用に供したい.

郵送調査の返送を規定する要因について

前田忠彦

2003年秋に実施した「日本人の国民性 第11次全国調査」に合わせ,類似の調査項目を用いた郵送調査を,東京都及び近郊3県の有権者を対象として行った.この郵送調査は他の調査方法との比較を意図した実験的調査であるが,実施の背景には,郵送調査だけを単独で実施しても,その特徴を把握するには不十分であるとの認識がある.データは2004年3月時点で分析中であるので,ここでは類似の実験的調査(前田・土屋(2001))の一環として2000年度に実施した郵送調査(前田(2002))からの知見をまとめ,今後の分析に備える.

検討したのは,郵送調査ではどのような属性を持つ層からの返送が得られやすいかという点で,これを主に同時期の面接調査における有効票・不能票の属性と比較した.ただし,いわゆる不能票について得られる属性情報は限られており,多くの属性については有効票のみの分布を国勢調査などから知られる母集団分布と比較するに止めざるを得ない.

郵送調査の標本の性・年齢構成を調べると,回収標本で若年層の重みが小さくなりかつ母集団からの乖離は女性より男性に大きいという,面接法などと共通方向の偏りをもつことを指摘

できるが，偏りは面接法に比べて小さめである．その他の属性では，郵送調査の回収標本は母集団に比べて，学歴が高めであり，未婚者が少なく，一戸建住宅の居住者が多く，単身世帯や非持ち家率も少ない，などの特徴があった．

属性別の回収率という観点からまとめなおすと，男性・若年層の回収率が低く，その他に電話帳に番号を掲載しない層の回収率も低いことが知られる．前記のような属性の母集団分布からの乖離の一部の原因として，郵送調査の未返送についても対象者のプライバシーに対する意識などが関与していることを想像させる結果とも言える．

こうした郵送調査における返送・未返送(有効・不能)の差が，安定して観察されるものかどうか，あるいはこれは他の調査方法と比べたときにどのように特徴づけられるのかを，2003 年度実施の調査の結果と合わせて引き続き検討中である．

参 考 文 献

前田忠彦(2002)．郵送調査法の特性に関する研究 ── 2000 年度 1 都 3 県有権者調査報告 ──，統計数理研究所 研究教育活動報告，No. 14．

前田忠彦，土屋隆裕(2001)．日本人の国民性 2000 年度吟味調査報告 ～電話・郵送・面接調査の比較～，統計数理研究所研究リポート，No. 87．

空間分割の統計とその発見的応用

種 村 正 美

生物細胞や金属粒界といった多面体セルによる空間分割の構造はしばしば自然界において観測される．一方，空間に散布された多数の点の配置を特徴づける一つの方法として，Voronoi セルの統計分布が有用であることが知られており，また一般に Voronoi セルが自然界に観測される空間分割の構造に対する発見的モデルとして有用であることも知られている．本年度は Voronoi セルの統計分布に関する研究と，Voronoi セルの発見的応用の研究とを行った．

空間統計学では，所与の点配置データの予備的解析として，データが Poisson 点配置から外れているか否か，またその外れはどの程度かが問題になる．その際，データから Voronoi セルによる空間分割を構成して，Poisson 配置の Voronoi セルの統計分布と比較することは一つの強力な方法である．そのため，Poisson 配置に対する Voronoi セル(Poisson Voronoi セル)の統計分布を可能な限り正確に求めておくことが重要である．われわれは Poisson Voronoi セルの種々の幾何学量に関する統計分布を次元 2, 3, 4, 5 に対して，計算機実験によって大量の Voronoi セルの独立標本を生成することによって求めた(2, 3 次元については Tanemura (2003)参照; 4, 5 次元については種村(2002)参照)．Voronoi セルの換算体積 v $(= \rho V$: ρ は Poisson 点過程の強度; V は Voronoi セルの体積)の分布に関して，Gilbert (1962)が 2 次積率 $\mu_2' \equiv E[v^2]$ の理論的結果を多重積分を含む形で任意次元 d に対して与えている．われわれは今回，Gilbert が与えた式の数値積分を実行して，われわれの計算機実験結果($\hat{\mu}_2'$)との比較を行った(次の表)．

d	$\hat{\mu}_2'$	μ_2' (Gilbert, 1962)
2	1.28031	1.28018
3	1.17830	1.17905
4	1.12034	1.12046
5	1.08316	1.08336

この数値積分において，われわれは Berntsen et al. (1991) の適応型多重積分アルゴリズムを用いた．Gilbert (1962) は $d = 2, 3$ に対する数値積分の結果として，それぞれ $\mu_2' = 1.280, 1.180$ を与えている．上の表は，われわれが行った計算機実験の結果がすべての次元に対して極めて正確であることを示唆している．Poisson Voronoi セルの換算体積の統計分布に関する上の考察や Kiang (1966) の予想の検討などについては Tanemura (2004) にまとめた．

Voronoi セルの発見的応用の一つとして，生物細胞集団に重力や遠心力などの外力が作用するとき生じる細胞の変形や再配列という実験事実を説明・予測するための頂点ダイナミックス細胞モデルについても報告した (Honda et al. (2003))．

参 考 文 献

Berntsen, J., Espelid, T. O. and Genz, A. (1991). An adaptive algorithm for the approximate calculation of multiple integrals, *ACM Transactions on Mathematical Software*, **17**, 437–451.

Gilbert, E. N. (1962). Random subdivision of space into crystals, *Annals of Mathematical Statistics*, **33**, 958–972.

Honda, H., Tanemura, M. and Nagai, T. (2003). A three-dimensional vertex dynamics cell model of space-filling polyhedra simulating cell behavior in a cell aggregate, *Journal of Theoretical Biology*, **226**, 439–453.

Kiang, T. (1966). Random fragmentation in two and three dimensions, *Zeitschrift für Astrophysik*, **64**, 433–439.

種村正美 (2002). Poisson Voronoi cells in 4 and 5 dimensions, **統計数理**, **50**, 308–309.

Tanemura, M. (2003). Statistical distributions of Poisson Voronoi cells in two and three dimensions, *Forma*, **18**, 221–247.

Tanemura, M. (2004). Statistical distributions of the shape of Poisson Voronoi cells, *Proceedings of the Third Voronoi Conference on Analytic Number Theory and Spatial Tessellations* (in press).

インターネットにおける抽籤と順列の列

丸 山 直 昌

n 人の応募者に対して 1 位から n 位の順序を与えるような抽籤を，応募者が一堂に集まることができない条件で行うことを考える．0 から $n-1$ の n 個の整数を並べた順列が複数個並んだ列を考え，偶然性を持って決まる数 m により，その列の中の m 番目の順列を当籤順位とする，という形でこの問題を定式化することができる．さらにこの列中の各順列を行とする行列 M を考えれば，抽籤は行列 M で表現される．すると，抽籤のプロセスの公正さはこの行列 M をあらかじめ公開することにより確保され，公平さは行列 M の均等散布性という条件に置き換えられる．抽籤に必要とされる意外性は，数値 m が偶然性をもって決まり，かつ想定さ

れる m の変動の幅がある程度あって，事前に抽籤結果についての予感を与えてしまうことがないことが必要である．今回はこれらの条件に加えて抽籤の継続性という問題に注目した．実用例においては，抽籤は繰り返し何回も行われることも多い．その場合毎回同じ抽籤では参加者に飽きが生じる．繰り返し行っても参加者を飽きさせないためには，毎回行列 M を取り替える(シャッフルする)ことが必要である．シャッフルの方法が恣意的であると抽籤参加者に疑念を与えるので，第三者にも事後で確認できるアルゴリズムを採用する必要があるが，同時にシャッフル操作自身が意外性を持っていないと参加者の飽きを防げない．そこで，当たり籤を使って M をシャッフルするアルゴリズムを考案してその性質を調べている．考案したアルゴリズムではシャッフルによって候補となる籤は完全に入れ替わり，重複は生じない．複数回シャッフルを行った場合には，何回か前の候補籤と同じものが候補に現れる可能性はあるので，計算機実験も用いて候補籤の遷移状況を調べている．

環境傾度に沿って変化する樹木の分布パターン

島 谷 健 一 郎

樹木の空間分布パターンは，標高などの環境傾度に沿って変化する場合がある．例えば低地ではランダムに分布するが標高が高いとパッチ分布をなし，かつパッチの密度や大きさも変化する．このような点分布をもたらすモデルとして，Thomas process と inhomogeneous Poisson process の融合が考えられる．即ち，inhomogeneous Poisson process では点密度を傾度に沿って変化させられる．Thomas process では，ランダムに密度 λ で分布する親のまわりに，強度 μ のポアソン分布に従う個数の子供が分散 σ の 2 次元正規分布に従って散布される．これらを組合わせれば，パッチ密度，パッチ内個数，パッチサイズが傾度に沿って変化する空間パターンを創作できる．この点過程の 2 次モーメントはその 2 地点の位置に依存する 4 変数函数であるが，簡略に 1 地点の傾度値と 2 点間の距離の 2 変数で近似できる．これを使えば，傾度に沿って変化する空間パターンを視覚的にグラフで表示でき，実データからのパラメータ推定も簡易に行うことができる．本研究では，これらを北海道知床半島トドマツ個体群に適用し，その空間パターンの標高に沿った連続的変化を撹乱履歴と関係付けて議論した．

コウホート分析に基づく将来推計(2)

中 村 隆

昨年度のスポーツ・レクリエーション参加人口の将来推計に続き，今年度は全国がん罹患数・率の将来推計を行った．このような将来推計を行う際に，人口の年齢・世代構成の変化のみならず，個々人の加齢に伴う変化(年齢効果)，世代特性の違い(コウホート効果)，社会の成員全体に及ぶ時勢による一定方向への変化(時代効果)，を区別することが重要である．これらが区別できれば，将来動向について見通しがつけやすくなる．

がん罹患は，当然ながら年齢効果が大きいが，コウホート分析の結果によると罹患部位により時代効果やコウホート効果のあり方に特徴がみられる．全般にがん罹患の時代効果は急激な変動はないものの，がんそのものの発生を抑える対策や診断技術の発達・検診の普及などの影響を反映し，変化している．またコウホート効果は通過した歴史的な社会・生活環境を反映し

た世代の違いを捉えている．

がん罹患の将来推計は以下のようにして行う．がん罹患調査データの年齢階級数を I，調査時点数を J，コウホート区分数を K とし，$\widehat{\beta}^G$，$\widehat{\beta}^A_i$，$\widehat{\beta}^P_j$，$\widehat{\beta}^C_k$ をそれぞれ，コウホート分析の結果得られた総平均，年齢，時代，コウホート効果のパラメータの推定値とする．がん罹患数の予測値 $\widehat{N}_{ij}$ は，将来の時代効果 $\widehat{\beta}^P_{J+1}$，$\widehat{\beta}^P_{J+2}$，$\ldots$ と新規参入のコウホート効果 $\widehat{\beta}^C_{K+1}$，$\widehat{\beta}^C_{K+2}$，$\ldots$ を適当なシナリオに基づいて設定することにより，

$$\log \widehat{N}_{ij} = \log P_{ij} + \widehat{\beta}^G + \widehat{\beta}^A_i + \widehat{\beta}^P_j + \sum_k c_{k,ij}\, \widehat{\beta}^C_k\,, \qquad \widehat{N}_j = \sum_{i=1}^{I} \widehat{N}_{ij}$$

として求めることができる．ここで，P_{ij} は年齢階級別の将来推計人口であり，ポワソンモデルのオフセットとして用いている．$c_{k,ij}$ はセルとパラメータのコウホート区分の重なりに比例する重みである．

将来推計結果に対しては，時代効果 $\widehat{\beta}^P_j$（$j = J+1, \ldots$）の設定が大きく影響し，その設定の問題が残される．これまで時代効果の推定値に回帰直線をあてはめて延長する方法，2 次階差の自己回帰的平滑化を取り込んだベイズモデルにより延長する方法などが試みられているが，結果的にはトレンドを直線で延長することに変わりはない．そこでわれわれは，近い将来は得られている時代効果のトレンドを反映し，20 年後には何らかの対策を期待して時代効果が水平となるとのシナリオに基づいて，諸検討シナリオの比較の基準となる将来推計を行った．

がん罹患の将来推計についての詳細は，大野 他(2004)を参照．

参 考 文 献

大野ゆう子, 中村 隆, 村田加奈子，津熊秀明，味木和喜子，大島 明(2004)．日本のがん罹患の将来推計――ベイズ型ポワソン・コウホートモデルによる解析に基づく 2020 年までの予測，『がん・統計白書――罹患／死亡／予後――2004』(大島 明，黒石哲生，田島和雄 編)，201–218, 篠原出版新社，東京．

予測分布の解析

伏 木 忠 義

統計的予測問題においてブートストラップを用いることの効果について調べた．Harris の提案したパラメトリック・ブートストラップ予測は，漸近的には最尤推定量の plug-in 分布よりも良い予測を与えるということがわかっている．また，学習理論の分野で研究されている bagging に相当するノンパラメトリック・ブートストラップ予測も同様に最尤推定量の plug-in 分布よりも漸近的に良い予測を与える．このように，理論的には，漸近論を用いてブートストラップ予測の有効性は示されているが，現実的にどの程度の予測性能の改良があるかということは，実際に用いられるモデルなどに依存しており，明らかではない．

今回の発表では，混合正規分布モデルを用いて，ブートストラップ予測の有効性を調べた結果について報告した．複雑で柔軟な密度を扱う場合，混合分布によるモデリングが一般に用いられている．コンポーネント数を適当にとることで混合分布により十分に真のモデルに近いモデルを得ることができる．混合分布モデルはこのように表現力豊かな有用なモデルであるが，尤度が有界ではなく通常の意味での最尤推定量が存在しないなど理論的には取り扱いが難しい．

応用の現場では，EM アルゴリズムなどで尤度の極大値を求めることで混合分布モデルは利用されている．

今回は，EM アルゴリズムで有界な尤度の極大値を求め，その plug-in 分布とブートストラップ予測を行った場合のリスクを比較した．前述のように，混合分布においては通常の漸近理論が成り立たないため，最尤推定量の漸近分布を用いたこれまでの一般論を直接適用することはできないが，尤度の「適当な」極大値をとってきた場合には通常の漸近理論に近い結果が得られると予想され，ブートストラップの効果が期待できる．今回の数値実験の結果から，複雑なモデルにおいては，適度なサンプル数のもとで，ブートストラップを行うことによるリスクの良化が 10–20%みられ，ブートストラップの効果が大きいということや，尤度の極大値がいくつもある場合には，パラメトリック・ブートストラップ予測よりもノンパラメトリック・ブートストラップ予測の方が有効であるということが示唆された．

インターネット環境下での定性調査の課題

大隅　昇

インターネット調査(Web 調査，電子メール調査)の急速な普及に伴い，定性情報のデータ取得環境や取得データの解析方法のあり方にも大きな変化が生じている．従来から行われてきた定性調査(グループ・インタビュー，フォーカス・グループ，談話・発話分析，エスノグラフィー等)は，いずれも電子化によるデータ取得方法に移行する傾向がある．フォーカス・グループはオンライン・フォーカス・グループ(OFG)に，従来の郵送調査や留置自記式による自由回答方式は Web 調査にそれぞれ移行しつつある．

このようにテキスト型データ取得環境は電子的データ取得法(CASIC: Computer-assisted Survey Information Collection，CADAC: Computer-assisted Data Collection)の普及と一般化により調査方式(調査モード)が大きく変容した．また，容易にテキスト型データが電子的に取得できるという利点から，テキスト・マイニングやコンピュータ支援の内容分析(CACA: Computer-assisted Content Analysis)の研究潮流が様変わりした．これに連動して，ここ数年，テキスト・マイニング関連ソフトウェアが無数に登場している．筆者も産学協同研究の一つとしてテキスト型データ解析ソフトウェア WordMiner™ の開発と普及に努めてきた．

とくに Web 調査を用いたときに，「豊富なデータ取得が可能」「書き込み量(率)が高い」「表現内容が豊かで情報量も多い」「本音を書きやすい」「思いがけない意見，出現率の低い稀な意見や回答が得られやすい」等々，科学的根拠のない誤信があるようにみえる．むしろ，自由回答設問を用いる調査は，調査方式が電子的取得に移行という大きな変化があったことで，従来よりも問題が複雑となり科学的な検証がより困難となった．例えば，回答の制御や抑制が可能なことから回答者の正しい回答行動が読み取りにくく回答取得履歴が不透明となり完全な捕捉が困難となったこと，電子調査票を用いることで調査票方式や設問形式の設計，設問の文脈効果の評価が難しいことなど，従来の調査方式では見られなかった事象が生じている．自由回答設問ではテキスト・ボックスやテキスト・フィールドの設計方式の影響があること，マルチメディア機能の利用(静止画・動画などのイメージや音声)も回答に影響することなどが実験的に検証されているが，現実の調査現場ではこれらの情報はほとんど考慮されることなく実査が行われている．

筆者等も，インターネット調査の実験調査の課題の一つとして，定性調査とくに自由回答設問の利用方法やデータ解析手法についての研究を行ってきた．例えば，調査方式を変えたとき，

つまり郵送調査(自記式)，オムニバス調査(面接，自記式)，インターネット調査の比較実験を通じて検証を進め，調査機関(サイト)間比較，調査方式の類似差異，設問形式や設問内容の違い，書き込み量の比較など，多面的かつ計量的に研究を進めた．例えば，総単語数や異なり単語数(率)の分布傾向，回答あたりの平均語長等の観察などから客観的な分析を進めた．さらに，テキスト型データのマイニング・ソフトウェア WordMiner™ を解析エンジンとして，調査現場に適した機能のカスタマイズを進め，これを自由回答設問の設計評価や分析への適用を試みた．

参 考 文 献

松田浩幸，大隅 昇(2002)．インターネット調査における調査票設問設計の評価 ── 設問形式が回答に及ぼす影響を測る ──，ISM シンポジウム「インターネット調査の現状を検証する ── 調査法としての評価方法と標準化をどう考えるか ──」，予稿集，33–54．

大隅 昇，Lebart, L.(2002)．テキスト型データの多次元データ解析 ── Web 調査自由回答データの解析事例 ──,『多変量解析実例ハンドブック』(柳井晴夫 他 編)，757–783，朝倉書店，東京．

子どもを対象とした自記式調査法の特性について

土 屋 隆 裕

学校あるいは学級を集落抽出し，子どもを調査対象とした自記式調査データの特性について示した．

まず，学校間変動と学級間変動とを比較し，中学生では学級間に比べ学校間の変動が大きいこと，一方，小学生では学齢が下がるほど学校間と学級間の変動の大きさに違いが見られなくなることを示した．このことから，誤差を抑えるためには，中学生では学級を二段集落抽出し，小学生では学校を集落抽出するのがよい，との示唆を得た．

次に，学級内の等質性について比較し，小規模な(人数が少ない)学級や学年あたりの学級数が少ない学校，人口規模の小さな地域の学校ほど学級内の等質性が高いことを示した．さらに，学年が低くなるほど，教室で記入するのか自宅で記入するのかといった調査の実施場所が，回答に大きく影響していることを示した．

統計科学関連 WWW サイトの現状について

清 水 信 夫

コンピュータ・ネットワークが急速に普及し，多種多様な情報が氾濫している現在，電子的に蓄積された膨大な量の統計データの有効利用や高度な統計分析手法を開示し普及させていく方法についての検討が課題となっている．インターネットの普及が拡大するにつれて，様々な分野において蓄積した統計関連情報を WWW (World Wide Web) 環境下で公開する動きが広がっており，内容も次第に多様化する傾向にある．これらの Web サイトについていくつかの区分により大まかに分類した報告例はあるものの(Murdoch (2000)など)，統計科学関連の Web サイト全般の中からいかに有益なものを広範に選ぶか，さらにそれらのサイトからいかに有用な情報を取り出し表示するかについては研究の余地がある．

本研究においては，2002 年度および 2003 年度それぞれの場合において，各種サーチエンジンを利用して統計科学関連 Web サイトを広範に収集した上で有用なサイトを絞り込んだリストを作成し，その過程で順序化が可能と考え得るいくつかの指標を設けた．その上で，各指標における各区分を数値化して，多次元尺度構成法の利用による分類および自己組織化マップによる分類を試み，それぞれの場合において布置図に影響を与えている要素について考察した（清水（2003））．

参 考 文 献

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

清水信夫(2003)．統計科学関連 WWW サイトの現状，2003 年度統計関連学会連合大会(日本統計学会第 71 回大会)講演報告集，99–100．

地震活動変化と地殻の歪み変化と統計モデル——2003 年の活動

尾 形 良 彦

はじめに

2003 年 7 月 28 日零時に宮城県北部地域に M5.5 の直下型地震が起き，その余震が続発した．本震余震型であると仮定して気象庁は余震確率を公表したが，同日朝 7 時過ぎ M6.2 の地震が同地域に発生し，それまでの本震余震は前震と呼ばれる事になった．同様なケースとして双子地震と呼べるものが少なからずある．最近のケースでは，後発の地震は若干小さくなるが，1997 年鹿児島県西北部の地震（3 月の M6.5 と 5 月の M6.2）や 1994 年三陸はるか沖の本震（M7.5）と最大余震（M7.2）などがある．最初の余震系列の経過情報から，その後に近隣で本震に近い規模またはそれ以上の大きな余震や地震の発生の確率が高まるか否かを判断するのは地震予知の観点から重要である．

統計モデルによる余震系列の静穏化の検出

実際，前者の地震の余震系列に改良大森型のポアソン過程または ETAS 点過程を当てはめ，AIC による適合度を比較してみると，余震活動が静穏化している場合が多い．同様の解析によると今回の宮城県北部地域の前震も静穏化している可能性が大きい．

このほか 2003 年 5 月宮城県沖の余震活動は，例えば M2.5 以上の余震系列については順調に改良大森法則に従って推移している．しかし，M1.5 以上（但し本震後 20 日以降のあてはめ）では，宮城県北部の地震（M6.2）によると思われるコサイスミックな静穏化が見られ，M0 以上ではプレサイスミックな静穏化も見られる．

相対的静穏化現象のメカニズム

余震域近傍での将来のトリガー地震断層内のプレスリップを仮定するとき，これが余震活動にとって stress-shadow になり静穏化すると考えるが，通常そのクーロン破壊関数値は大きくない．しかし，そのような応力変化でも余震活動の数割の低下が起きうる．同様なメカニズムが 2003 年 11 月の十勝沖地震（M8.0）をめぐっても見られる．

債権回収率モデルの深い闇

山 下 智 志

2006年に発行される銀行の自己資本に関する国際協定である新BIS規制では，銀行の貸し出しリスク(信用リスク)の算定にはPD（デフォルト率)の推計だけではなく，LGD（デフォルト時損失率)の算定も各銀行の独自モデルによって行うことを推奨されるようになった．現在，提案されているLGDには，デフォルト債券市場のデータから推計するMarket LGD，実際に回収した実績から推計するWorkout LGD，債券市場などのリスクプレミアムから計算するImplied LGD，過去の実績損失と平均PDよりマクロ的にLGDを推計するHistorical Implied LGDがある．しかし，日本ではデータベースの不備などによりMarket LGDやWorkout LGDに関するモデルを構築することが不可能な状況である．そこで本研究では市場の取引データがあれば計算可能な，比較的データベースの必要要件の少ないImplied LGDについてモデル化した．

Implied LGDモデルの基本は，債券市場のリスクプレミアムからその企業のリスク量を算出するReduced Formアプローチをベースにしている．しかしリスクプレミアムをPD × LGDの形で把握するReduced Formアプローチではは PDとLGDのどちらか一方しか推定できない．本研究ではこの欠点を，株式を永久債券と見なすことによって解決する．つまり，PDとLGDで未知数が2つあるために，Reduced Formの式も2本立てて，連立させて解くことにより，同時推計を可能にした．実際の市場では債券と株式ではデフォルトしたときの回収のシステムが違うので(債券は順位によって回収されるが，株式は一部の例外を除けば回収が不可能である)，高リスクの企業についてはその差異が市場に反映されている．

市場データを用いて時系列分析とクロスセクショナルなLDGの分布を求めた．時系列分析ではLGDはPDの動きと逆相関の関係があった．これは，株価の方が企業のリスクに対してより敏感に反応し，債券市場は動きが鈍いことによる結果である．また，クロスセクショナル分析では全体的にPDとLGDは弱い正の相関があった．しかし，一般に言われているほど顕著ではない．また同一格付の中ではPDとLGDには逆相関の関係が見られた．これは格付がPD × LGDでリスクを評価している結果であると思われる．

不完全情報下における制御系設計に関する研究

宮 里 義 彦

制御のためのモデルの設定と同定から制御手法までを総括的に含む統合化制御系設計理論の構築を考えている．その一環として，モデリングと制御の接点を扱う適応制御の基礎理論の研究や，実用化のための様々な制約を取り除いた適応制御系の設計法，及び関連する非線形制御の研究を行っている．

漸近安定性に主眼を置く従来の適応制御理論に対して，制御性能をより定量的に考慮するために，適応制御過程をH_2/ H_∞最適制御問題として定式化する研究を行ってきた．モデル規範形適応制御を含む一般的な形式の適応制御問題について，安定解析に用いるリアプノフ関数の一部とHamilton-Jacobi-Isaacs方程式の解を同一視することで，特定の評価関数に対して最適(または準最適)な適応制御系の構成法を求めた先の研究(宮里(2002))を発展させて，システムに含まれる未知のパラメータをH_∞制御問題における未知外乱と見なすことにより，パラメータの任意の変動に対して安定な非線形適応H_∞制御系を構成する手法を開発した(宮里

(2003a)). その手法をニューラルネットを制御器に含む非線形適応 H_∞ 制御や，さらにより一般的な非線形パラメータモデルの非線形・適応制御に拡張することも考えている．

これらとは別に LPV (Linear Parameter Varying)システムのゲインスケジューリング制御において，スケジューリングパラメータが未知な場合に，その推定値を用いて適応的にゲインスケジューリング制御系を構成する手法について考察した．これまでの基礎的な結果をもとにして，非線形パラメトリックモデルを導入した LPV システムの適応型ゲインスケジューリング制御についても考察し，Miyasato (2003b)にあげるような結果を得た．またむだ時間要素が含まれる場合の安定解析についても考察した．

さらに反復学習制御を，ハイブリッド適応機構を用いて実現する方法を考案し，ロボットマニピュレータの軌道追従制御に適用して，追従誤差の安定解析と数値計算による検証を行った (Miyasato (2003c))．ハイブリッド適応機構としてより収束特性の優れた方式や，複数の適応プロセスの導入による制御性能の向上の可能性についても，いくつかの結果が得られた．

参 考 文 献

宮里義彦(2002)．最適性に基づく適応制御系の再設計，計測自動制御学会論文集，**38**(9)，765–774．

宮里義彦 (2003a)．パラメータを外乱と見なした非線形適応 H_∞ 制御系の構成法，計測自動制御学会論文集，**39**(10)，914–923．

Miyasato, Y. (2003b). Adaptive gain-scheduled H_∞ control of linear parameter-varying systems with nonlinear components, *Proceedings of 2003 American Control Conference*, 208–213.

Miyasato, Y. (2003c). Iterative learning control of robotic manipulators by hybrid adaptation schemes, *Proceedings of the 42nd IEEE Conference on Decision and Control*, 4428–4433.

自己修復機能付きデータベースシステム

樋 口 知 之

マイクロアレイ遺伝子発現データから遺伝子の相互関係をモデル化する作業，つまり遺伝子ネットワークのグラフィカルモデル構成において，ベイジアンネットワーク(以後 BN と略す)は有用な表現方法の一つである．各遺伝子発現量を確率変数として取り扱った時，BN の持つ最大の特性である連鎖法則は，超多数の確率変数の同時分布を条件付き確率分布の積へ分解することを可能にする． この分解により主に我々は，ある一つの子遺伝子が(複数の)親遺伝子にどのように依存しているのかを特徴づける，子遺伝子の条件付分布に注目すればいい．

我々が提案した方法 (Imoto et al. (2004))は遺伝子ネットワークのモデル構成において，マイクロアレイデータからの遺伝子発現量の情報と，データベース化された生物学的知識の情報を結合させる一般的枠組みである．その手法の利点の一つは，マイクロアレイからの情報と生物学的知識間のバランスをどうとるかを情報量規準が定めることができることである．生物学的知識をベイジアンネットワークに付加することにより，マイクロアレイデータからさまざまなノイズによる影響を排除しつつ精密に遺伝子ネットワークを推定でき，結果としてより多くの情報を抽出することに成功した．

本研究では，この方法論をさらに発展させることを狙った．一般に生物学的知識は，データ

ベースの形に具現化されることが多い．データベースに登録された情報にも信頼度の属性を与え，つまりそれを確率変数として取り扱うことで，データベースの登録上の過誤やその情報の不確実さをモデル化する．この設定のもとで，今までのモデルをさらに階層化したベイズモデルを構成し，マイクロアレイデータからの情報抽出，既存の生物学的知識の有効活用，データベースの信頼性の検証などを統一的に可能にする枠組みを考案し，現在人工データへの適用による性能評価中である．本研究の延長上には，データベースへの登録作業の過誤などが自動的に特定できる，自己修復機能をそなえたデータベースシステムの設計も視野においている．

参 考 文 献

Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S. and Miyano, S. (2004). Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks, *Journal of Bioinformatics and Computational Biology*, **2**(2) (to appear).

生産関数のノンパラメトリック推定

川 崎 能 典

東証一部上場企業を製造業，非製造業に大別した上で，1965年から2001年までの各年で生産関数としてコブ・ダグラス型やトランスログ型といったパラメトリックモデルが妥当といえるかどうかの検定を行った．回帰関数の定式化に関する検定としてはRamseyのRESET検定がよく知られているが，対立仮説の下で想定されるクラス(べき関数)が真の構造を含まないときには，検出力が非常に低いことが知られている．そこで本研究ではHong and White (1995)で提案されたノンパラメトリック検定を適用した．分析結果の示すところでは，製造業では1979年から84年以降，非製造業においては1990年代に入って以降，コブ・ダグラスやトランスログ型生産関数は適切な関数形とは言えない．次のステップとして，Bスプライン基底関数への回帰に基づく一般化加法モデルを考え，恐らくは非線形な回帰関数の形状をノンパラメトリック回帰によって推定した．回帰関数の形状に対して，基底関数の個数や平滑化パラメータの選択は本質的であるが，ここでは一般化情報量規準GICを利用した．推定された生産関数の局面形状は，検定結果と整合的であった．部分残差に基づき，1980年時点での資本・労働に関する非効率性を観察した上で，2001年までに東証一部から削除された企業群を調べてみると，どちらかの生産要素に非効率性を抱えていた企業ほど，吸収合併や倒産が顕著に見られた．なお，本研究は小西葉子氏(日本学術振興会特別研究員)，西山慶彦氏(京都大学経済研究所)，安道知寛氏(九州大学大学院数理学府)との共同研究(15–共研–1019)に基づく研究成果である．

参 考 文 献

Hong, Y. and White, H. (1995). Consistent specification testing via nonparametric series regression, *Econometrica*, **63**, 1133–1159.

遺伝子情報解析

足立　淳

ゲノムや cDNA の塩基配列から，未知のタンパクをコードする遺伝子領域を推定する方法は大きく二つに分けられる．一つは既知の遺伝子と比較し，相同性が高い領域を見つける方法．もう一つは，既知の遺伝子の特徴を抽出した統計的なモデルを使い，未知の領域を評価する方法である．後者の代表的なモデルとして coding potential がある．これまで coding potential として hexa-nucleotide モデルという連続する 6 塩基の使用頻度情報が使われてきた．タンパクの構成要素である 20 種類のアミノ酸は，それぞれコドンと呼ばれる 3 塩基の並びから翻訳されるので，6 塩基の使用頻度は隣り合うコドンのペアの使用頻度を表している．この情報は，同じアミノ酸に翻訳される複数のコドンの中でどのコドンがどれだけ使われるかという「冗長なコドンの使用頻度」と，どんなアミノ酸が隣り合わせになりやすいかという「アミノ酸ペアの使用頻度」が組み合わさったものである．

しかし，アミノ酸のペアの組み合わせは 400 通りしかないのに対して，コドンのペアの組み合わせは 3721 通りもあり，とても冗長である．また，冗長なコドンの頻度は，ゲノム上の塩基配列の偏りの影響を強く受けるのに対し，アミノ酸のペアの頻度はその偏りの影響をそれほど受けない．むしろアミノ酸のペアの頻度は，その遺伝子がコードしているタンパクの種類に大きな影響を受ける．よって，この二つの頻度情報は独立に扱った方がよいことが期待される．そこで新しい coding potential として，塩基組成の偏りを補正した冗長なコドンの使用頻度モデルと，タンパクの種類を考慮したアミノ酸のペアの使用頻度モデルを独立に扱うことを提案した．

最後に，これまで分泌系等の小さなタンパクはノイズに埋もれやすく発見することが難しかった．新しいモデルを用いると，より小さなタンパクを発見することができるなど，その有効性を示した．

Nearest Neighbor ARX Model with Application to Dynamic Brain Functioning Analysis

尾崎　統

今年度も非線形時系列モデリングとその応用の研究を以下の 3 つの応用領域と関連して行った．

1）ガスコンバインド火力発電プラントの排煙脱硝制御
2）リターンとリスクの予測と資産最適配分
3）fMRI と EEG データの時空間因果モデリングとダイナミック脳機能解析

火力発電プラントの排煙脱硝制御に関しては日本ベーレーで NPC（Nonlinear Predictive Controller）という名前の製品化が平成 15 年度夏に完了し販売が始まった．現在の制御法は Clarke 流の制約付き最適化の形になっているがその改良版として Akaike 流状態空間にもとづく制御法の非線形版も準備中．

2）に関してはいくつかの金融業界の企業が実際の資産運用への利用に関心を示しており今年度は実践への利用に備えて最適化に繋げるリターンとヴォラティリティの予測モデルの改良に努めた．

3)に関しては NN-ARX モデルを中心にその fMRI の Source Localization 問題と Connectivity 解析への利用の道を開いた．NN-ARX モデルをダイナミックモデルに持つ状態空間アプローチの EEG ダイナミック逆問題解法の研究も昨年度に続き行った．システムノイズ分散の時空間的変化の適応的推定により逆問題解の解像度を飛躍的に向上させることが出来ることを確認した．

データ同化にもとづくエルニーニョの発生予測

上 野 玄 太

本研究の目的は，非線型の力学方程式や非ガウスの擾乱要素を含む数値モデルに対して，高次元のデータを同化するための新しい計算技法を開発することである．本年度は，Zebiak and Cane (1987)による大気・海洋結合モデルに対し，アンサンブルカルマンフィルタ(EnKF, Evensen (1994))を用いて TOPEX/Poseidon 衛星による海面高度データの同化作業をスタートした．

Zebiak and Cane (1987)のモデルは，太平洋赤道域に着目し，簡単ながらも海洋と大気の相互作用を取り入れ，ほぼ 4 年ごとのエルニーニョらしき海面水温の準周期的変動を再現している．このモデルはいくつもの非線型のプロセスを内包しているため，標準的なカルマンフィルタ・平滑化のアルゴリズムはもはや使えない．そこで，EnKF を用いて，多数個の実現値を用いてカルマンゲインを近似することで対処することとした．このときの状態ベクトルの次元は 72810 となる．同化に使用する観測データは TOPEX/Poseidon 衛星による海面高度偏差である．Zebiak and Cane モデルの海盆範囲とデータ取得位置の関係から，各タイムステップで得られるデータ点数は最大で 8388 点となる．

EnKF による同化にあたって，観測ノイズの分散共分散行列を次のように推定した．まず，各観測グリッド上での海面高度データの時系列に対してカルマンフィルタ・平滑化のアルゴリズムを用いて 1 階トレンドモデルをあてはめ，残差を求める．同様に他の観測グリッド上でも残差を求め，この残差のグリッド間の分散共分散行列を計算し，観測ノイズの推定値とする．一方でシステムノイズは，海洋と大気を結ぶ式にガウスノイズとして導入した．ガウスノイズ発生のための分散共分散行列は，グリッド間距離に応じたノイズの相関を仮定した上で定めた．

参 考 文 献

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, **99**, 10143–10162.

Zebiak, S. E. and Cane, M. A. (1987). A model El Nino-Southern Oscillation, *Monthly Weather Review*, **115**, 2262–2278.

マルチカノニカル法による状態空間の Landscape の研究

伊 庭 幸 人

統計科学や統計物理では複雑な確率分布を扱うことが要求される．その際，高次元の状態空

間における確率分布(ベイズ統計では事後分布，統計物理ではギブス分布)がどのような形状をしているかを知ることは基本的な問題であり，さまざまなアプローチがなされている．

本講演では，対象となる分布族が指数型分布族(あるいはギブス分布)の形になる場合に，多変量の状態密度をマルチカノニカルモンテカルロ法で計算することで，高次元での状況の一端を知るという方法について論じた．ここで，状態密度 $D(E_1', E_2', \dots)$ とは，離散変数の場合，分布の十分統計量(あるいは十分統計量と興味のある統計量を含む組)を $E_1(\boldsymbol{x}), E_2(\boldsymbol{x}), \dots$ としたとき，「すべての i について $E_i(\boldsymbol{x}) = E_i'$」を満たす $\boldsymbol{x}$ の個数として定義される量であり，分布族全体の情報を含んでいる($\boldsymbol{x}$ が連続変数の場合は $E_i' < E_i(\boldsymbol{x}) < E_i' + dE$ 等とする)．

この方法はすでに格子タンパク模型や 2 次元格子上のイジング模型などに適用された例があるが，著者は誤り訂正符号のプロトタイプである 3 体のソウラス符号(3 体相互作用をするランダムネットワーク上のイジング模型に等価)について実験を行い，その結果と解釈について報告した．報告した内容は，高橋久尚氏との共同研究である．

家系間の世代の同定について

上田　澄江

ヌジ人名資料(Gelb et al. (1943))は B.C.15 世紀頃の古代都市ヌジの遺跡から出土した粘土版文書の人名による索引であり，人名にかかわる親族関係, 内容を記した文献名，その巻数, 行番号などが参照できる．今回，この資料の情報を下に構成した家系の各メンバーの契約文書を，2 家系間で比較することにより，ある程度世代を特定することができた．大富豪であったと思われるテヒプティラの家系のメンバーと直接・間接的に同文書に現出する家系は全家系の実に 88%余りに及ぶ．勿論，世代と世代の境界が明確に線引きされるわけではないが，中央集権的な社会構造と急速にその社会形態が崩壊していったさまが契約数の減少により読み取られることは興味深い．

参考文献

Gelb, I. F., Purves, P. M. and Macrae, A. A. (1943). *Nuzi Personal Names*, The University of Chicago Press, Chicago, Illinois.

半正定値計画問題による密度関数推定について

土谷　隆

半正定値計画問題は半正定値対称行列とアフィン空間の交わり上で線形関数を最適化する凸計画問題であり, 制御や信号処理, 組合せ最適化等に多くの応用を持つ．本講演では, 半正定値計画法を用いた(1 変数)確率密度推定法について述べる．この方法では, 密度関数は正規分布, 指数分布一様分布などの基底関数とそのサポート上での非負多項式の積として表現される．非負多項式は, 半正定値対称行列を用いて表現され, サポート上で積分値が 1 であるという制約は, それらの行列の線形等式制約として書ける．そして基底関数が固定されていれば最尤推定は半正定値計画問題となり, 近年発展した内点法によって, 効率良く厳密に解くことができる．

基底関数は，通常1つか2つのパラメータを含むのみなので，多項式部分の最適化が厳密にできるのであれば，基底関数の最適化は比較的容易であり，結局，本モデルにおいては厳密な尤度関数の最大化が可能となる．さらに，密度関数にしばしば要請される条件である単峰性や対称性なども半正定値計画法によって記述でき，同じ枠組みで取り扱うことができる．本発表では，有限混合分布において混合される各分布を上述のアプローチで推定することにより，より柔軟に密度関数を表現する手法について述べた．

参 考 文 献

Fushiki, T. , Horiuchi, S and Tsuchiya, T. (2003). A new computational approach to density estimation with semidefinite programming, Research Memo., No. 898, The Institute of Statistical Mathematics, Tokyo.

モンテカルロフィルタを用いた金融時系列分析

佐 藤 整 尚

近年，金融データに対する時系列解析の応用が広く行われるようになった．特に，時系列のトレンドおよびボラティリティの推定とその予測には時系列解析の手法がよく用いられている．また，複数の時系列間の関係を探る上でも時系列解析的手法を用いることが多い．これらに共通する目的として，観測値の背後に隠れている性質や関係を導き出すということがあげられる．そのような目的では，近年，状態空間表現によるアプローチが盛んに研究されている．たとえば，トレンドの推定には移動平均や多項式回帰といった手法から状態空間表現によってモデルを作りカルマンフィルタ等で推定を行うことが一般的になっている．また，ボラティリティの推定についてもStochastic Volatility Modelという状態空間表現を用いたモデルを使って推定することが行われている．しかしながら，金融工学の発達により，より複雑なモデルが提案されるようになり，線形ガウス型のモデルのみならず，非線形非ガウス型のモデルを推定する必要が出てきた．この場合，カルマンフィルタでは近似的な推定しかできないので，様々な形の非線形フィルタが提案されている．ここで用いているモンテカルロフィルタもその1つで，平易な計算アルゴリズムにより様々な形のモデルを推定することが可能である．本報告では，これまで行ってきた金融時系列に対するモンテカルロフィルタの応用例を紹介し，さらに，並列計算の可能性についても考察を行った．ここで紹介したのは，Stochastic Volatility Modelの推定，金利モデルの推定，および，回帰モデルにおける時変係数の推定である．Stochastic Volatility Modelにおいてはいろいろな分布系を仮定して計算することができ，さらに，ボラティリティの変動がトレンドの変化と連動するような複雑なモデルの推定も可能であることを示した．金利モデルの推定では，観測方程式が解析的に求められないような複雑なモデルであっても，モンテカルロフィルタを用いれば推定ができることを示した．回帰モデルの時変係数の推定では係数のプロセスに非正規分布を仮定することにより，係数の急な変化を検出することが可能であることを示した．これにより，時系列間の関係についても様々なモデリングが可能になったといえる．以上の例を用いて，モンテカルロフィルタの優位性について議論した．また，モンテカルロフィルタでは並列化が難しいとされているが，研究所に導入された共有メモリー型計算機では，工夫を重ねることにより，openMPを使った並列化が有効であることを示した．また，この際，同機に搭載されている物理乱数発生装置が極めて有効であることにも言及した．

グリッド環境に適した遺伝的アルゴリズムについて

染 谷 博 司

近年，遺伝的アルゴリズム（GA）は生命情報科学など解評価に多大な計算量が要求される分野にも応用され始めており，最適化手法としての有効性が注目されている．しかし，GA に必要とされる計算資源は十分とはいえず，次世代のより大きな計算資源の利用が可能な GA が望まれている．一方，高速ネットワークを利用しインターネット上の膨大な計算資源を共有利用した超並列計算を可能とするグリッド技術が注目されている．年度研究報告会では，GA の計算資源としての計算グリッド環境に着目し，計算グリッド環境への GA の適用について述べた．まず，GA が考慮すべき計算グリッドの特徴を述べ，グリッド環境に適した GA およびその設計について考察し，グリッド環境に適した GA とは，(1)世代交代が局所化されている，(2)通信量が小さく通信頻度が少ない，(3)高性能である，といった特徴を有した GA であることを示した．また，遠隔地間を接続するグリッド環境での GA の実装例および最適化問題への応用例を示し，その有効性を確認した（染谷（2003a，2003b，2004））．

参 考 文 献

染谷博司（2003a）．グリッド環境に適した遺伝的アルゴリズムに関する考察とその実現，電気学会 電子・情報・システム部門大会 2003 講演論文集，435–439（OS7–3）．

染谷博司（2003b）．グリッド環境に適した遺伝的アルゴリズムの設計とその性能評価，情報処理学会・電子情報通信学会 第 2 回 情報科学技術フォーラム（FIT2003）講演論文集第 1 分冊，75–77（A–036）．

染谷博司（2004）．進化型計算による適応的探索およびグリッド環境への応用，研究集会「最適化：モデリングとアルゴリズム 17」，180–190．

アナログデータ処理によるスペクトル拡散通信方式と Neuron MOS によるハードウエア実現の研究

瀧 澤 由 美

本研究は，限定された帯域幅の通信路において高速データ伝送を可能とする通信方式とこれを実現するための LSI に関する理論・技術の創出を目的とする．具体的には，まず高速データ伝送のためのスペクトル拡散変調方式の研究を行った．次に，ハードウエア実現のため Neuron MOS によるアナログ LSI 構成について研究を行った．Neuron MOS は本方式に適合し，またアナログ信号処理による低電力小型チップの実現が期待される．

本研究ではシステムを方式と LSI の結合として把え，その中から新技術・新理論の創出を試みている．

(1)スペクトル拡散通信方式の研究

雑音を有するチャネルの通信容量（通信速度の限界値）C（bit/s）は Shannon によって与えられる．

$$C = W \log(1 + S/N)$$

ここで，W は帯域幅（Hz），S, N は信号およびノイズパワー（W）である．しかし Shannon は変調および符号化のための具体方式を示唆してはいない．本研究では，系の特性を支配する同

期と復調特性を改善し，W-CDMA を Shannon 限界に近づけ，高速データ伝送を実現する．研究の結果，(i)同期のためのパイロット信号の捕捉と追跡の機能を空間並列と時間制御による単一マッチドフィルタにより実現し，(ii)相関信号のピーク検出における WTA 回路の特性改善を試み，同期と復調特性の改善を図った．

(2)Neuron MOS による低消費電力 LSI 実現の研究

神経ニューロン型の応答を示す素子として Neuron MOS が東京大学柴田研究室によって開発された．本研究では系の特性を支配するテーマとして同期・復調のためのマッチドフィルタとピーク検出のための伝搬位相等価のためのデータ処理回路に着目し，系の応答速度の向上と消費電力の顕著な軽減を試みた．研究の結果，従来と比較すると，同等のデザインルール，速度で特に消費電力は従来の数分の一以下となった．2003 年度末に第 1 次 LSI 試作を完了した．2004 年度には第 1 次試作の評価と，方式と回路の改良による第 2 次試作を行う．

本研究のこれまでの成果に対して，電気通信普及財団賞・テレコムシステム技術賞を授与された(2004 年 3 月)．

参 考 文 献

Arai, Y., Igarashi, K., Fukasawa, A. and Takizawa, Y. (2004). Multi-user detection to enhance the capacity of W-CDMA based on the conjugate gradient method, *Journal of Circuits, Systems, and Computers*, **13**(2), 1–11.

EIA/TIA/IS-665 (1995). W-CDMA (Wideband Code Division Multiple Access)Air Interface Compatibility Standard for 1.85–1.99 GHz PCS Applications, EIA/TIA/95.09.1.

Fukasawa, A., Iijima, Y., Igarashi, K., Inoue, S. and Takizawa, Y. (2002). Estimation and compensation of phase rotation in radio propagation using pilot channel for radio data transmission, International Conference on System Simulation and Scientific Computing, Shanghai, 963–967.

Fukasawa, A., Sato, T., Takizawa, Y., Kato, T., Kawabe M. and Fisher, R. E. (1996). Wideband CDMA system for personal radio communications, *IEEE Communications Magazine, Topics in Personal Communications*, **34**(10), 116–123.

Iijima, Y., Inoue, S., Kashima, T., Fukasawa, A. and Takizawa, Y. (2004). Coherent SS demodulation with estimation and compensation of phase rotation using a pilot channel, *Journal of Circuits, Systems, and Computers*, **13**(2), 12–24.

Nakano, H., Fukasawa, A. and Takizawa, Y. (2003). Multi-user detection by sequential interference cancellation for W-CDMA, WSEAS International Conference on Simulation, Modelling and Optimization, Manuscript #463–133.

Adaptive Parameter Estimation Both for Robustness and Efficiency

藤 澤 洋 徳

データを解析するとき，しばしば外れ値の影響によって，解析結果にブレが生じる．頻繁に使われる標本平均などは外れ値に多大な影響を受けてしまう．中央値など様々な外れ値に強いパラメータ推定法が提案されているが，分布が対称的でない場合などには，どれも十分に機能するとは言えない．本発表では，そのような場合を含めて外れ値の割合が多い場合にも十分に機能し，かつ，推定効率が保たれるパラメータ推定法を提案した．

外れ値とは尤度を小さくするデータである，という考えに基づいて，Windham (1995)は，

分布が対称的でない場合にも利用できるロバストなパラメータ推定法を提案した．同時にパラメータ推定値を得るための簡単なアルゴリズムも提案されたため，非常に使いやすい推定法であった．Basu et al.（1998）は同じ流れで別のパラメータ推定法を提案した．Jones et al.（2001）は二つの推定法を幾つかの意味で比較している．

それらのロバスト法を吟味することで以下の性質が分かった．前者のロバスト法から導かれるダイバージェンスは，外れ値の構造を本質的に無視しやすい構造になっていた．その結果として，前者のロバスト法は外れ値の割合が多くても機能しやすいと考えられた．その構造は，後者のロバスト法を含めて他の良く知られたロバスト法には見受けられなかった．この構造については過去の文献では全く考慮されていない．そしてこの構造は，これまで外れ値の割合が少ないときにしかできなかった議論を，外れ値の割合が多い場合にも可能にしている．また，その構造を利用すると，上述のロバスト法をチューニングすることが可能となり，その結果として，推定効率をも復元することが可能になった．その復元のからくりは，概パラメトリックモデルに着目することで考察が可能になった．

参 考 文 献

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence, *Biometrika*, **85**, 549–559.

Jones, M. C., Hjort, N. L., Harris, I. R. and Basu, A. (2001). A comparison of related density-based minimum divergence estimators, *Biometrika*, **88**, 865–73.

Windham, M. P. (1995). Robustifying model fitting, *Journal of Royal Statistical Society. Series B,* **43**, 599–609.

A Pythagorean Relationship in the Conjugate Analysis

大 西 俊 郎

1．問題設定

曲指数型分布族に属する標本分布

$$p(\boldsymbol{x};\,\boldsymbol{\mu}) = \exp\bigl\{-d(\boldsymbol{x},\,\boldsymbol{\mu})\bigr\}\,a(\boldsymbol{x}) \tag{1.1}$$

における Bayes 推定を議論する．ここで，$\boldsymbol{x}$ および $\boldsymbol{\mu}$ は p 次元であり，$d(\boldsymbol{a},\,\boldsymbol{t})$ は

$$d(\boldsymbol{a},\,\boldsymbol{t}) = \sum_{j=1}^{p+1} h_j(\boldsymbol{a})\bigl\{f_j(\boldsymbol{t}) - f_j(\boldsymbol{a})\bigr\}$$

で与えられる．関数 $f_1(\boldsymbol{x}),\ldots,f_{p+1}(\boldsymbol{x})$ および $h_1(\boldsymbol{x}),\ldots,h_{p+1}(\boldsymbol{x})$ は適当な条件を満たすものとし，損失関数として $d(\hat{\boldsymbol{\mu}},\,\boldsymbol{\mu})$ を採用する．仮定する事前分布は

$$\pi(\boldsymbol{\mu};\,\boldsymbol{m},\,\delta) = \exp\bigl\{-\delta\,d(\boldsymbol{m},\,\boldsymbol{\mu}) + k(\delta)\bigr\}\,b(\boldsymbol{\mu}) \tag{1.2}$$

である．ただし，規格化定数が δ のみに依存するように適当な $b(\boldsymbol{\mu})$ を選べるものとする．本発表の目的は次の 2 点である．

（i） 曲指数型分布族 (1.1) の共役解析を議論する．

（ii） ピタゴラス関係の観点から共役解析の意味を明らかにする．

2．共役解析

曲指数型分布族 (1.1) **の共役解析について** 2 **つの命題が得られる．命題** 2.1 **において，** $\hat{\boldsymbol{\mu}}_{smap}$ **は** Yanagimoto and Ohnishi **(**2004**)によって提案された標準化事後モードである．**

$$\hat{\boldsymbol{\mu}}_{smap} = \underset{\boldsymbol{\mu}}{\mathrm{Argmin}}\{d(\boldsymbol{x},\, \boldsymbol{\mu}) + \delta\, d(\boldsymbol{m},\, \boldsymbol{\mu})\}.$$

命題 2.2 **の修正ピタゴラス関係は，標準化事後モードが任意の推定量を優越する様子を明らかにしている．**

命題 2.1．事前分布 (1.2) **は標本分布** (1.1) **に対して共役であり，事後分布は適当な** ρ_{min} **を用いて次のように表現される．**

$$\pi(\boldsymbol{\mu}|\boldsymbol{x};\, \boldsymbol{m},\, \delta) = \pi(\boldsymbol{\mu};\, \hat{\boldsymbol{\mu}}_{smap},\, \rho_{min}).$$

命題 2.2．(i**) 任意の推定量** $\hat{\mu}$ **に対して，次の修正ピタゴラス関係が成立する．**

$$\mathrm{E}_{\mathrm{post}}\big[\, d(\hat{\boldsymbol{\mu}},\, \boldsymbol{\mu})\,\big] - \mathrm{E}_{\mathrm{post}}\big[\, d(\hat{\boldsymbol{\mu}}_{smap},\, \boldsymbol{\mu})\,\big] = \frac{1}{\rho_{min}}\,\mathrm{KL}\big(\,\pi(\boldsymbol{\mu},\, \hat{\boldsymbol{\mu}}_{smap},\, \rho_{min}),\, \pi(\boldsymbol{\mu},\, \hat{\boldsymbol{\mu}},\, \rho_{min})\,\big).$$

ここで，$\mathrm{E}_{\mathrm{post}}$ **は事後平均，**KL **は** Kullback-Leibler **分離度を意味する．**

(ii**) 標準化事後モード** $\hat{\mu}_{smap}$ **は最適であり，その事後リスクは** $k'(\rho_{min})$ **である．**

3．ピタゴラス関係

必ずしも共役でない事前分布 $\pi(\mu)$ **を仮定した場合の推定問題を考察することにより，共役解析の意味を明らかにする．仮定した事前分布に対応する事後分布を** $\pi(\mu|x)$ **と書く．**

命題 3.1．評価関数 $\mathrm{KL}\big(\,\pi(\boldsymbol{\mu}|\boldsymbol{x}),\, \pi(\boldsymbol{\mu};\, \boldsymbol{\theta},\, \rho)\,\big)$ **を最小化する** $(\theta,\, \rho)$ **の値を** $(\boldsymbol{\theta}_{min},\, \rho_{min})$ **とする．損失関数** $d(\hat{\mu},\, \mu)$ **に対する** Bayes **推定量** $\hat{\mu}_B$ **は** $\boldsymbol{\theta}_{min}$ **であり，その事後リスクは** $k'(\rho_{min})$ **で与えられる．**

この命題から，事後分布から共役事前分布の平面に垂線を下ろすことで Bayes **推定量が得られることが分かる．次の命題** 3.2 **におけるピタゴラス関係は，その垂線が共役事前分布の平面と局所的に直交しているだけでなく大域的にも直交していることを意味している(図** 1 **参照)．これは命題** 2.2 **(**i**)の一般化である．**

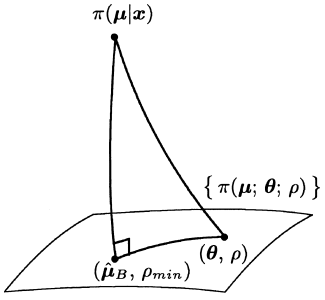


図 1**.**

命題 3.2. 任意の $(\theta,\, \rho)$ **に対して次のピタゴラス関係が成立する.**

$$\begin{aligned}&\mathrm{KL}\big(\,\pi(\boldsymbol{\mu}|\boldsymbol{x}),\, \pi(\boldsymbol{\mu};\, \boldsymbol{\theta},\, \rho)\,\big) - \mathrm{KL}\big(\,\pi(\boldsymbol{\mu}|\boldsymbol{x}),\, \pi(\boldsymbol{\mu};\, \hat{\boldsymbol{\mu}}_B,\, \rho_{min})\,\big)\\ &\qquad = \mathrm{KL}\big(\,\pi(\boldsymbol{\mu};\, \hat{\boldsymbol{\mu}}_B,\, \rho_{min}),\, \pi(\boldsymbol{\mu};\, \boldsymbol{\theta},\, \rho)\,\big).\end{aligned}$$

4. 最小情報性

命題 3.2 **のピタゴラス関係を用いて共役事前分布の最小情報性を示す. 共役事前分布** $\pi(\boldsymbol{\mu};\, \boldsymbol{m},\, \delta)$ **の超パラメータ** $\boldsymbol{m}$ and δ **を任意に固定し, データ** $\boldsymbol{x}$ **が与えられた下で事前分布の集合** $\mathcal{P}_{\boldsymbol{x}}$ **を考える.**

$$\mathcal{P}_{\boldsymbol{x}} = \{\, \pi(\boldsymbol{\mu}) \mid \text{Bayes 推定値とその事後リスクが } \pi(\boldsymbol{\mu};\, \boldsymbol{m},\, \delta) \text{ のものと同一} \,\}.$$

この集合の中で次の汎関数を最小化する事前分布を見つけたい.

$$G[\,\pi(\boldsymbol{\mu})\,] = \mathrm{KL}\big(\,\pi(\boldsymbol{\mu}|\boldsymbol{x}),\, \pi(\boldsymbol{\mu};\, \boldsymbol{x},\, 1)\,\big)$$

標本分布 $p(\boldsymbol{x};\, \boldsymbol{\mu})$ **と事前分布** $\pi(\boldsymbol{\mu};\, \boldsymbol{x},\, 1)$ **は同一視できるので, 汎関数は** $G[\,\pi(\boldsymbol{\mu})\,]$ **は事前分布** $\pi(\boldsymbol{\mu})$ **に含まれる情報量と解釈できる. 次の命題は共役事前分布の最小情報性を意味している.**

命題 4.1. 事前分布の集合 $\mathcal{P}_{\boldsymbol{x}}$ **の中で** $G[\,\pi(\boldsymbol{\mu})\,]$ **を最小化するのは, 共役事前分布** $\pi(\boldsymbol{\mu};\, \boldsymbol{m},\, \delta)$ **である.**

参 考 文 献

Yanagimoto, T. and Ohnishi, T. (2004). Standardized posterior mode for the flexible use of a conjugate prior, *Journal of Statistical Planning and Inference* (to appear).

繰り返し 3 人一般化ジャンケンゲームの漸近安定性

伊 藤 栄 明

二人の player **が一般化ジャンケンゲームを繰り返し行ってゆくとする. 一般化ジャンケンゲームにおいて 各** player **は** $1, 2, ..., m$ **の戦略をとるものとし,** player A **が戦略** x**,** player B **が戦略** y **をとったとき, 戦略** x **をとる** player A **が戦略** y **をとる** player B **に勝つ確率は** $\frac{1}{2} + a_{xy}$ **であるものとする.** $a_{xy} = -a_{yx}$ **とし,** $\frac{1}{2} \le a_{xy} \le \frac{1}{2}$ **であるとする. すなわち 確率** $\frac{1}{2} + a_{xy}$**, で** $x \succ y$ **であり確率** $\frac{1}{2} + a_{yx}$ **で** $x \prec y$ **であるとする.**

非協力 3 **人一般化ジャンケンゲームを繰り返してゆくとする.** x**,** y**,** z **を** player A, B **および** C **のそれぞれの戦略としたとき,**

i) $x \succ y$ and $x \succ z$ **である確率が** $(\frac{1}{2} + a_{xy})(\frac{1}{2} + a_{xz})$,
ii) $y \succ x$ and $y \succ z$ **である確率が** $(\frac{1}{2} + a_{yx})(\frac{1}{2} + a_{yz})$,
iii) $z \succ x$ and $z \succ y$ **である確率が** $(\frac{1}{2} + a_{zx})(\frac{1}{2} + a_{zy})$,
iv) $x \succ y \succ z \succ x$ **である確率が** $(\frac{1}{2} + a_{xy})(\frac{1}{2} + a_{yz})(\frac{1}{2} + a_{zx})$
$x \succ z \succ y \succ x$ **である確率が** $(\frac{1}{2} + a_{xz})(\frac{1}{2} + a_{zy})(\frac{1}{2} + a_{yx})$.

であるものとする. 各 player **は次の混合戦略として前回のゲームで成功をおさめた戦略をとる確率をふやしてゆくものとする.** 2 **人ゲームの場合は各** player **のとる混合戦略の時間的変化**

は 2 体の相互作用による Lotka-Volterra 方程式によりモデル化することができる．この場合各 player の混合戦略は振動する．3 人ゲームの場合は三体の相互作用による Lotka-Volterra 方程式によりモデル化することができ，各 player の混合戦略は Nash 平衡へ近づいてゆく．

参 考 文 献

Itoh, Y. (1975). An H-theorem for a system of competing species, *Proceedings of the Japan Academy*, **51**, 374–379.

Itoh, Y. (1981). Non-associatev algebra and Lotka-Volterra equation with ternary interaction, *Nonlinear Analysis*, **5**, 53–56.

Itoh, Y. and Cohen, J. E. (1994). Competitive ternary interactions and relative entropy of solutions, *Journal of Physics A*, **27**, 6383–6393.

刑事事件に係る文章の計量分析

村 上 征 勝

文化現象の統計分析の研究の一つとして文章の計量分析の研究を行っているが，今回は刑事事件の解明に役立った文章の計量分析について紹介した．

2001 年 9 月 28 日に東京都台東区でひき逃げ事件が発生した．事件発生から 10 日ほど経ってから，ひき逃げを目撃したという匿名の情報提供の手紙と，ひき逃げ事件の犯人と名乗る人物から告白状・遺書が警察に届いた．目撃者の手紙と犯人の告白状・遺書とは，ひき逃げをした経緯やその車が逃走する状況がほぼ一致しており，この二通の手紙を読む限りでは単純な交通事故のようにも見える．

しかし，その後死亡した被害者に対して死亡時に四千万円が支払われる生命保険がかけられており，保険をかけていたのは被害者の兄である事が判明した．犯人は被害者の兄である可能性が高まった．

しかし，証拠がない上に，二通の手紙はワープロで作成されており，筆跡鑑定ができない．そのため，警察より目撃者の手紙と犯人の告白状・遺書は被害者の兄が偽装して書いたことを文章の統計分析を使って調べられないかという依頼があった．

そこでこの二通の文書と，被害者の兄が以前に書いた二通の文書，さらに学生や被害者の兄と同年齢の人の文章等を用いて，文章のクセについて分析を行った．分析では

1. どのような助詞がどの程度用いられているかの頻度情報
2. どの助詞の後にどの助詞が出現するかに関する頻度情報
3. どの文字の後に読点がつけられているかに関する頻度情報

の三種類の情報を用いて分析した．

その結果，目撃者の手紙，犯人の告白状・遺書の二通と被害者の兄が書いた二通の合計四通の文章はいずれの情報を用いた分析でも一つのグループを作ったことから，この四通は同一人物の手で書かれた可能性が非常に高いという結果を得た．この分析結果の報告の一ヵ月後，被害者の兄が目撃者の手紙，及び犯人の告白状・遺書を自分が書いたことを自白し，事件が解決した．文章の計量分析が社会に貢献したということを人々に知らしめたと同時に，これまでの研究方法の妥当性が裏付けられることになった．

東アジア価値観調査 ── 第 2 年次報告 ──

吉 野 諒 三

平成 14 年度より 4ヵ年計画で，東アジアの人々の価値観の国際比較調査を遂行している．本研究の重点は，a)文化の伝播変容の解明のために，東アジア諸国の人々の意識構造について統計科学的「標本抽出法」に則った面接調査を遂行し，b)21 世紀における国際交流の中で，東アジア諸国民の「信頼感」のあり方について焦点を当て世界の政治・経済の平和的発展の一助となる基礎情報を与える分析を推進させることである．過去の国際比較調査研究と同様に,「国際可能性の追求」(異なる国々の言語や調査方法の下で収集されたデータの比較可能性)が，研究の中心テーマである．

本年度は，第 2 年次の計画のとおり，特に，1)調査票の台湾語と韓国語の暫定版を，現地研究者の協力により翻訳・再翻訳を経て作成し，2)台湾と韓国の調査予定地域を視察し，調査環境を確認，現地調査研究者と標本抽出の実践的検討を行い，同時に調査票の検討を経て，3)9 月上旬，台湾と韓国の調査票の最終版を確定し，10 月〜11 月に本調査を実施した．

調査対象は，台湾及び韓国に居住し，各地域の国籍をもつ成人(20 人以上)の男女．調査法は，台湾は電話所有者の人口データを用いて，120 地点抽出，各地点ではランダムウォーク法で 1 軒置きに訪問，各世帯から誕生日法で 1 名抽出，合計 15 名ずつ抽出．韓国は 125 地点をランダム抽出し，各地点では性・年齢層別(10 才刻み)に約 8 名ずつをクォータ法で抽出した．調査項目 は，各地域の人々の一般的意識構造，特に対人関係，集団内や集団間，社会制度やリーダーに関する「信頼感」を主とした項目.(昨年度実施の日本調査票と基本的に同一だが，一部表現を各地域に合わせている)

12 月〜翌 3 月は，回収調査票を詳細に検討して，データ・クリーニング，単純集計表作成，データ分析調査結果の解釈について検討し，第一次報告書をまとめた．これまでに，日本，中国，台湾，韓国という,「儒教文化」の影響が強いといわれる国々や地域の人々の意識や価値観のデータが収集された．しかし，第 1 次分析でも，各国・各地域の共通の側面と固有の側面とが多様に現れ，伝統的な価値観と，現在の社会の急速な変化とが複雑に絡まった様相が浮かび上がってきた．結果は統計数理研究所研究リポート No.91 として発刊を予定している．

Lattice データの Hotspots 検出

(客員)岡山大学 栗 原 考 次

地域における病気の発生率を示した疫学データや環境汚染物質の分布を表した環境データのように空間的構造をもつデータに対して，これらの関連を調査したり，有意に高い値を示す地域(hotspots：ホットスポット)の検出をおこなうことは，環境の状況を把握するとともに，将来の環境や健康への影響を早期に発見するためにも重要である．空間スキャン統計量は，データが得られた地点を中心に円状に領域をスキャンし，有意に高い比率を示す領域を見つける．しかし，円状に領域をスキャンするため円状のホットスポットの検出には優れているが，線状や他の形状をしたホットスポットの検出には適しない．Echelon 解析は，空間的な位置を表面上の高低に基づき分割し，空間データの位相的な構造を系統的かつ客観的に見つけることができる．本研究では，lattice 領域内で観測されるデータに対して，各種の形状をしたホットスポットの検出を行うために，echelon 解析により各種の空間データの空間的な階層構造を求め，その構造に基づき領域をスキャンする方式を提唱した．具体的な例として，ノースカロライナ州

の 100 郡において 1974 年 7 月から 1984 年 6 月の期間に観測された乳幼児突然死症候群(SIDS)データを用いて，提唱した方法を利用することにより，円状のスキャンでは検出できなかった形状のホットスポットの検出が可能であることを示した．さらに，両親の社会経済状態と精神的健康状態を順序尺度で分類した 6 × 4 の順序カテゴリカルデータに対して，有意に独立性から逸脱している隣接したホットスポット(セル群)を検出した．

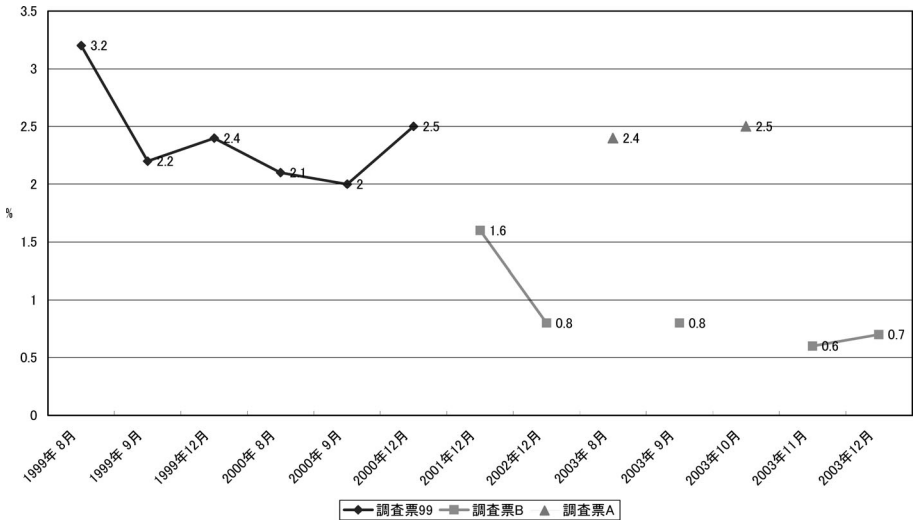
参 考 文 献

栗原考次(2003)．階層的空間構造を利用したホットスポット検出，計算機統計学，**15**(2), 171–183.

覚せい剤乱用者調査について

田 村 義 保

覚せい剤乱用者数を推定するためのオムニバス調査を平成 15 年 8 月から 12 月まで 5 回実施した．標本数は 2,000 (20 歳以上の日本人)で回収率は 70%程度であった．8 月，10 月に用いた調査票を調査票 A，9 月，11 月，12 月に用いた調査票を調査票 B と呼ぶことにする．2 種類の調査票を用いたのは，平成 14 年度の調査で覚せい剤乱用者を知っていると答える者の数が急激に減少したので，実際に減ったのか調査票の影響かを見るためである．調査票 A は，平成 11 年度，12 年度とほぼ同等の質問文，調査票 B は平成 13 年度，14 年度とほぼ同等の質問文を用いている．「覚せい剤乱用者を知っている」と回答した者の割合は図のようになっている．明らかに調査票により回答比率は異なっており，平成 14 年度の調査で，覚せい剤乱用者を知っている者の数が減少したように見えたが，調査票の影響であり，調査票 A を用いた結果は平成 11 年度，12 年度と同程度であることが分かった．本調査は社会安全研究財団の委託調査研究として，薬物使用の状況を広く一般に訴え，薬物問題に対する認識を深めるための情報を集



図．覚せい剤使用者を知っていると回答した者の比率．

めるため，平成 10 年度から 14 年度までの 5 年計画で行う予定であった．すでに述べたように平成 14 年度の調査結果について再考するために，1 年延長して平成 15 年度にも調査を行った．今後も調査を行うのであれば，調査の質問文としては調査票 A のような答える量が少ない方がよいと考えている．なお，6 年間の報告書及び調査結果をまとめた CD-ROM を作成したので，希望者には配布可能である．

時系列のスペクトル解析について

荒畑恵美子

不規則変動を示す現象の解析をする時，次の様にしてみる．定常時系列のときには，時系列を周波数成分にスペクトル分解し，各々の周波数帯にどのように成分が分布しているかを見る．それには，Blackman-Tukey 法，自己回帰モデルのあてはめ(1 次元の場合，多次元の場合)等がある．これらを用いてスペクトルの推定をする．又，非定常な時系列のときには，時間と共に係数が変化する自己回帰モデルをあてはめ，時系列スペクトルを推定する．

参考文献

赤池弘次，中川東一郎(l972)．『ダイナミックシステムの統計的解析と制御』，サイエンス社，東京．
北川源四郎(1986)．時変係数自己回帰モデル，統計数理，34(2), 273–283.
北川源四郎(1993)．『時系列解析プログラミング』，岩波書店，東京．

統計解析システム Jasp における並列処理

中野純司

計算機環境の進化につれて，統計解析の状況も変化している．特に，POS システムに代表されるようなデータの自動収集機構が多く利用されるようになり，扱うべきデータの量は巨大になり，生成過程は複雑になってきた．そのようなデータを解析するために，計算機の莫大な計算能力に依存する新しい統計手法が多く提案されている．そのため，数年前のスーパーコンピュータに匹敵するような強力なパーソナルコンピュータをもってしても，最近の統計解析には十分とは言えない．すなわち，最新のプロセッサよりも強力な計算能力が要求されているのである．それを実現するための技術として，並列処理がある．並列処理は古くから研究されており，現在ではハードウエア，ソフトウエアとも実用的な段階になっている．ただ，統計解析においては，一部の専門家を除いて，並列処理はこれまであまり利用されて来なかった．その原因は，並列処理の可能な計算機が高価であったことと，ソフトウエアが使いやすくなかったことがある．近年，ネットワークで結ばれた複数の計算機を同時に使用することが可能になり，ハードウエアにはほぼ問題がなくなった．ただ，利用できるソフトウエアは，Fortran や C のようなシステム言語を使わなければならないものが多く，高度な統計解析システムの利用になれた統計学者には使いやすいとは言えない．そこで，われわれは統計解析システムの利用者が容易に並列処理を行えるような環境を作成することにした．具体的には，われわれが作成している統計解析システム Jasp において，Jasp 言語だけで並列処理が行えるような環境を設計し，

実現した．使いやすさを第一に考え，数個の関数だけで並列処理が実現できるようにした．この方法では，並列処理の細かい制御はできないが，統計計算に多く見られるような粒度の大きい並列処理には十分な能力をもっている．Jasp は Java 言語で書かれているため，Java 言語の分散処理技術である Jini と JavaSpaces を利用して，これを実現した．

半無限計画法とその周辺

伊 藤 聡

2 次の半無限計画問題

$$\text{(P)} \qquad \begin{aligned} &\min_{x\in R^n} \ \frac{1}{2}x^TQx + b^Tx \\ &\text{subject to} \quad a(y)^Tx \le c(y) \quad \forall y \in Y \end{aligned}$$

を考える．ただし，Y を R^s のコンパクト部分集合とし，$0 \le Q \in R^{n\times n}$, $b \in R^n$, $a \in C(Y, R^n)$, $c \in C(Y)$ とする．問題 (P) に対する双対形式の一つとして，以下のように Haar 測度で表現されたものがある．

$$\text{(D)} \qquad \begin{aligned} &\min_{\substack{x\in R^n \\ \lambda_i\in R,\ y_i\in R^s,\ i=1,2,\ldots,p}} \ \frac{1}{2}x^TQx + \sum_{i=1}^{p} c(y_i)\,\lambda_i \\ &\text{subject to} \quad Qx + b + \sum_{i=1}^{p} a(y_i)\,\lambda_i = 0 \\ &\qquad\qquad \lambda_i \ge 0, \quad y_i \in Y, \quad i = 1, 2, \ldots, p \end{aligned}$$

ただし $p \le n$ である．関数 a および c が通常強い非線形性を持つため，次元 n (そして s)が非常に小さい場合を除き，問題(D)を直接解いてその(大域的)最適解を求めることは容易ではない．そのため一般には，まず離散化法・切除平面法などにより十分な精度の近似解を得てから，より高精度の解を求めて局所収束性のすぐれた解法に移行する，いわゆる二段階法が有効である．

本研究では上述の第一段階における近似解法としての切除平面法について考察した．切除平面法では基本的に，問題(P)においてインデックス集合 Y をその有限部分集合 Y_k に緩和して得られる 2 次計画問題を解き，その解 x_k に対して元の制約を最も侵害する Y の点を Y_k に加えるという手続きを繰り返すため，反復が進むにつれて計算コストが増大する．本研究では，各反復において，制約の最大侵害点の計算を近似的に行ない，また緩和問題を解いた後で $|Y_k| \le n$ となるように Y_k を調整することにより，計算コストをおさえるアルゴリズムを提案し，自然な条件のもとで解点列 $\{x^k\}$ が問題(P)の最適解に収束することを示した．