

統計数理研究所研究活動

1983年度研究発表会要旨

とき：1984年3月21日，午前9時30分～午後4時

ところ：統計数理研究所 講堂

あいさつ

所長 林 知己夫

第1研究部

確率分布の近似

清水 良一

1. 基準化された和の分布 $F_n(x)$ と標準正規分布 $\Phi(x)$ に関して、

$$(1) \quad F_n(u) = \Phi(v)$$

と満足する u, v を互いに近似することを考える。十分に大きな n に対しては、もちろん $u \approx v$ である。 n が余り大きくないときが問題になるが、Cornish-Fisher 展開

$$(2) \quad v_1 = u + \frac{\alpha_3}{6\sqrt{n}}(1-u^2) + \dots,$$

$$(3) \quad u_1 = v - \frac{\alpha_3}{6\sqrt{n}}(1-v^2) + \dots,$$

が一般的な近似法を与えてくれる。しかし、展開(2)は必ずしも十分によいとはいえず、この展開を基礎に適当な変換が必要である。

分布が非対称、したがって $\alpha_3 \neq 0$ のとき、変換

$$v_2 = \frac{\alpha_3}{6\sqrt{n}} + \frac{6\sqrt{n}}{\alpha_3} \left\{ \left(1 + \frac{\alpha_3 u}{2\sqrt{n}} \right)^{1/3} - 1 \right\}$$

は分布の対称化に有効であり、とくに

$$\sup_{u \geq -2\sqrt{n}/\alpha_3} |F_n(u) - \Phi(v_2)| \leq 0.92 \cdot \frac{\alpha_3^2}{n} + R_n$$

である。ただし、 R_n は Edgeworth 展開の剰余項とする。より高次の展開に対する変換も可能である。

2. 同種の変換は展開(3)に関連しても行うことができる。とくに例えば、 U が $N(0, 1)$ に従う変数のとき、

$$G_1 = \left(2 - \sqrt{9 - \frac{1}{2\alpha} + \frac{3U}{2\sqrt{\alpha}}} \right)^4 \cdot \alpha$$

や

$$G_2 = \left(4 - \sqrt{9 + \frac{1}{\alpha} - \frac{3U}{\sqrt{\alpha}}} \right)^2 \cdot \alpha$$

は近似的にガンマ分布に従う。近似の精度は Wilson-Hilferty 近似と同程度である。

3. 一様分布に従う変数の和

$$Z_n = U_1 + \dots + U_{2n} - n/2$$

の分布が $N(0, 1/12)$ で精度よく近似されることはよく知られている。この場合には Cornish-Fisher 展開は有効で $X_n = \sqrt{12} \cdot Z_n / \sqrt{n}$ と基準化すれば

$$Y_n = X_n + \frac{1}{20n}(X_n^3 - 3X_n)$$

の分布は急速に $N(0, 1)$ に近づく。しかし、小さな n に対しては Y_n の分散は 1 よりだいぶ小さくなるから、 Y_n の系列を正規乱数として使うには適当な変換が必要である。

$$V_{12} = 0.98693 \cdot Z_{12} + 0.004493 \cdot Z_{12}^3,$$

$$V_4 = 1.66197 \cdot Z_4 + 0.077655 \cdot Z_4^3$$

は分散 1 で、正規分布への近似の誤差はそれぞれ、0.00007, 0.0007 程度である。

シルベスターの内挿公式の別表現と応用

松 縄 規

標記公式の一般形についての新しい表現を与える。結果は従来のもものと異なり、係数中に対象とする行列の特性多項式又は最小多項式の微係数を含まない閉じた形である。

応用としていくつかの線形結合統計量の精密標本分布などの誘導が可能であることを報告した。

1. 補題 (部分分数展開)

複素数 s の次の形の有理関数を考える：

$$(1) \quad G(s) = \{(s + \lambda_1)^{\nu_1} (s + \lambda_2)^{\nu_2} \cdots (s + \lambda_k)^{\nu_k}\}^{-1},$$

ここで $\lambda_j (j=1, \dots, k)$ は全て異なり、 $\nu_j (j=1, \dots, k)$ は極の重複度を表すものとする。このとき $G(s)$ は次の様に部分分数展開できる：

$$(2) \quad G(s) = \sum_{l=1}^k \sum_{m=1}^{\nu_l} C_{l,m} (s + \lambda_l)^{-m},$$

ここに係数は

$$(3) \quad C_{l,m} = \prod_{r=1}^k \frac{1}{(\lambda_n - \lambda_r)^{\nu_r}} \sum_{q=0}^{\nu_l - m} \sum^* \prod_{j=0}^{\nu_l - m} \frac{1}{q_j! j^{q_j}} \left\{ \sum_{p=1}^k \frac{(-1)^j \nu_p}{(\lambda_p - \lambda_l)^j} \right\}^{q_j},$$

ここで \sum^* は $q_0 + q_1 + \dots + q_{\nu_l - m} = q, 0 \cdot q_0 + 1 \cdot q_1 + \dots + (\nu_l - m) \cdot q_{\nu_l - m} = \nu_l - m, (q_0 = 0 \text{ for } \nu_l - m = 0)$ を満たす非負の整数の組 $\{q_0, \dots, q_{\nu_l - m}\}$ 全体に渡る和を表す。

(注) (3) 式の右辺は通常と違って、 $G(s)$ の (高階) 微係数を陽に含まない、より具体的な表現である。

2. 定 理

A : 複素数を要素とする $N \times N$ 行列, I : $N \times N$ 単位行列, $f(d)$: A の全ての固有値を含む複素 α -平面の単連結領域での解析関数, $R(\alpha)$: A の Lagrange-Sylvester 内挿多項式, $H(\alpha)$: A の最小多項式, $\alpha_1, \dots, \alpha_k$: $H(\alpha) = 0$ の互に異なる根, ν_1, \dots, ν_k : $\alpha_i (i=1, \dots, k)$ の重複度, とする。このとき行列 A の任意の解析関数 $f(A)$ は次の様に表現することが可能である：

$$(4) \quad f(A) = \sum_{l=1}^k \sum_{m=1}^{\nu_l} \sum_{n=0}^{m-1} C_{l,m,n}^* \frac{f^{(n)}(\alpha_l)}{n!} \prod_{r=1}^k \frac{1}{(\alpha_l - \alpha_r)^{\nu_r}} \\ \times (A - \alpha_l I)^{m-1} \prod_{r=1}^k (A - \alpha_r I)^{\nu_r},$$

ここに,

$$(5) \quad C_{l,m,n}^* = \sum_{q=0}^{m-1-n} \sum^* \prod_{j=0}^{m-1-n} \frac{1}{q_j! j^{q_j}} \left\{ \sum_{p=1}^k \frac{m-1}{(\alpha_l - \alpha_p)^j} \right\}^{q_j},$$

\sum^* は $q_0 + q_1 + \dots + q_{\nu_l - n} = q, 0 \cdot q_0 + 1 \cdot q_1 + \dots + (m-1-n) q_{m-1-n} = m-1-n, (q_0 = 0 \text{ for } m-1-n = 0)$ を満たす非負の整数の組 $\{q_0, \dots, q_{m-1-n}\}$ 全体に渡る和を表す。

3. 応用

ディリクレ確率変数やガンマ確率変数の線形結合の精密標本分布の誘導なことを報告した。

“ゆらぎ”のある現象に対する予測関数とモデル推定

川合伸幸

1. 真のモデルと予測 2 乗誤差基準によって選ばれるモデルの違い

回帰モデル $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$ を考察する。モデルに階層性,

$$\Theta_k = \{\boldsymbol{\beta}(k) = (\beta_1(k), \dots, \beta_k(k), 0, \dots, 0)\}, \Theta_1 \subset \Theta_2 \subset \dots \subset \Theta_K$$

を仮定する。そして、真のモデルは m 次のモデル族 Θ_m に含まれるとする。モデル族 Θ_k の下での最尤推定量 $\hat{\mathbf{b}}(k)$

$$\hat{\mathbf{b}}(k) = (\hat{b}_1(k), \dots, \hat{b}_k(k), 0, \dots, 0)'$$

は、 $\hat{\mathbf{b}}(K)$ の一次変換として表現できる。

$$\begin{aligned} \hat{\mathbf{b}}(k) &= (I_K - W_k) \hat{\mathbf{b}}(K) \\ W_k &= S^{-1} P_k (P_k' S^{-1} P_k)^{-1} P_k', \quad S = X'X \\ P_k &= [0 : I_{K-k}]' \quad K \times (K-k) \end{aligned}$$

$\hat{\mathbf{b}}(i)$; $i=1, \dots, K$ の一次結合

$$\hat{\mathbf{b}}_w = \sum_{i=1}^K \phi_i \hat{\mathbf{b}}(i), \quad \sum_{i=1}^K \phi_i = 1$$

を考え、 $\boldsymbol{\phi}$ を予測関数の選択とモデル推定のための指標とする。

予測 2 乗誤差 $E \| X \hat{\mathbf{b}}_w - X \boldsymbol{\beta} \|^2$ を最小にする $\boldsymbol{\phi}^*$ は次のように求まる。

$$\begin{aligned} \phi_1^* &= 1 / (1 + C_1 - C_2) \\ &\vdots \\ \phi_k^* &= 1 / (1 + C_k - C_{k+1}) - 1 / (1 + C_{k-1} - C_k), \quad k=2, \dots, m-1 \\ &\vdots \\ \phi_m^* &= 1 - 1 / (1 + C_{m-1}) \\ \phi_{m+1}^* &= \dots = \phi_K^* = 0 \end{aligned}$$

ここに、 $C_i = \boldsymbol{\beta}' P_i (P_i' S^{-1} P_i)^{-1} P_i \boldsymbol{\beta} / \sigma^2$ 。

よって、モデルが m 次であることと指標が m 次であることが同値、また、真のモデルの次数が m であっても $\phi_m^* = 1$ にはならず、小さな次数の方にも分散するが、 $\sigma^2 \rightarrow 0$ のとき、 $\phi_m^* \rightarrow 1$ となることがわかる。

2. 重みの推定とモデル推定

今、同一の観測行列の上で N 回の独立なくり返しを観測できたとする。くり返しについての $\hat{\mathbf{b}}_w$ の残差平方和の平均を、 $RSS_N(\hat{\mathbf{b}}_w) = \boldsymbol{\phi}' A \boldsymbol{\phi}$ $A = (a_{ij})$, $a_{ij} = a_j (i \leq j)$, $a_j = RSS_N(\hat{\mathbf{b}}(j))$ とする。 $RSS_N(\hat{\mathbf{b}}_w)$ を最小にする $\boldsymbol{\phi}^*$ を求めると、常に、 $\boldsymbol{\phi}^* = (0, \dots, 0, 1)'$ となる。 $N \rightarrow \infty$ としても、 $\boldsymbol{\phi}^*$ は依然として $(0, \dots, 0, 1)'$ である。そのため、このとき、 $\phi_{m+1}^* = \dots = \phi_K^* = 0$ という解が出るよう修正行列を加える。修正を加えた残差平方和は、 $\sum_i \phi_i = 1$ を使って書き下すと、

$$RSS_N(\hat{\mathbf{b}}_w) + 2\sigma^2 \sum_i \phi_i + n\sigma^2$$

となり、 $\sum \phi_i = 1$ の下に、この目的関数を最小にする解は、

$$\begin{aligned} \phi_1^* &= \sigma^2 / (a_1 - a_2) \\ &\vdots \\ \phi_k^* &= \sigma^2 \{1 / (a_k - a_{k+1}) - 1 / (a_{k-1} - a_k)\}, k=2, \dots, K-1 \\ &\vdots \\ \phi_K^* &= 1 - \sigma^2 / (a_{K-1} - a_K) \end{aligned}$$

となる。またモデルの次数の推定としては、

$$(a_{k-1} - a_k) / (a_k - a_{k+1}) \rightarrow \max_k$$

となる k を選ぶのが適当である。この値を m^* とすると、

$$\begin{aligned} \phi_1^* &= \sigma^2 / (a_1 - a_2) \\ &\vdots \\ \phi_i^* &= \sigma^2 \{1 / (a_i - a_{i+1}) - 1 / (a_{i-1} - a_i)\}, i=2, \dots, m^*-1 \\ &\vdots \\ \phi_{m^*}^* &= 1 - \sigma^2 / (a_{m^*-1} - a_{m^*}) \\ \phi_{m^*+1} &= \dots = \phi_K^* = 0 \end{aligned}$$

という重みをもつ予測関数が、得られる。 σ^2 の推定は、一致性のある推定量、 $RSS_N(\hat{b}(K)) / (n - K)$ 、で代用する。

2 母数指数分布の尺度母数の検定推定量

平 野 勝 臣

1. 問題および結論

2 母数指数分布 $(1/\theta) \exp \{-(x-\eta)/\theta\}$, $x \geq \eta$, $\theta > 0$, $-\infty < \eta < \infty$, (EP(η , θ) と書く) からの大きさ n の標本の Type II 打ち切りの標本を $X_1 < \dots < X_r$ とする。 $T_r(\eta) = \sum_{i=1}^r (X_i - \eta) + (n-r)(X_r - \eta)$, $S_r = T_r(X_1)$ とおく。 η に対する近似値 $\eta_0 (\leq \eta)$ がある。 θ の検定推定量 $\hat{\theta}_{PT}$ を、 $\eta = \eta_0$ が採択のとき $a(r)T_r(\eta_0)$, $\eta = \eta_0$ が棄却のとき $b(r)S_r$ とする。 $\hat{\theta}_{PT}$ の危険 $R(\hat{\theta}_{PT}) = E[(\theta/\hat{\theta}_{PT}) + (\hat{\theta}_{PT}/\theta) - 2]$ を求める。ここに $a(r)$, $b(r)$ は対応する損失関数で、夫々のモデルで最も良い値とする。 $R(\hat{\theta}_{PT})$ は検定の有意水準 α に依存するので、各 α に対してこの値を求めグラフにした。グラフはここでは略すが、当日 OHP で示した。問題は最適な α を指定したい。この問題の解決のための方法としてミニマックス・リグレット規準の適用が考えられるが難しい。この問題に対する AIC を導き、適用すると $\alpha = e^{-1} = 0.3678$ を得る。この値はグラフから見てミニマックス的な最適性を持つ。この問題に対して、少なくとも 0.01 や 0.05 を用いることは "統計的な考え方" の教えるところではない。以上の内容に関しては *Annals, A*, Vol. 36 No. 1 に発表されている。グラフはそれを参照されたい。

2. 説明および補足

上の例は AIC の導出に仮定される正則条件を満たしていない。しかし一般に AIC はこの正則条件を満たさない場合でも適用できることが少なくない。また上の例の様に小標本が前提の場合、精密理論に基づく AIC を必要とする。その考え方に従い、この場合の規準を精密理論で導びける。

正規分布 $N(\mu, 1)$ と $N(0, \sigma^2)$ の比較、EP($\eta, 1$) と EP($0, \theta$) の比較は正則条件を満たさない。よく知られている様に AIC の導出には尤度比統計量の偏りを修正し、不偏とするための項を必要としている。従ってこの比較のために AIC 統計量の差の期待値を評価する。精密理論で考え、評価のために必要な性質を示す。 X がガンマ分布 $G(s, \theta) : (x/\theta)^{s-1} e^{-x/\theta} / \{\theta \Gamma(s)\}$ に従うとき、1) $-1/(s-1) \leq E[\log X / (s\theta)] \leq 0$, 2) $E[\log X / (s\theta)] = \log \theta + \Psi(s)$ を得る。ここに Ψ は Psi 関数である。そこで

$$E[AIC \{N(\mu, 1)\} - AIC \{N(0, \sigma^2)\}] = 1/(3n) - 2/(15n^3) + \dots \text{under } N(0, 1),$$

$E\{AIC\{EP(\eta, 1)\} - AIC\{EP(0, \theta)\}\} = -1 + 1/(6r) - 1/(60r^3) + \dots$ under $EP(0, 1)$,
(但し n は標本数, r は打ち切り個数) を得る.

測度空間の写像: 条件付け (conditioning)

久保木 久 孝

1. 例 (序に代えて)

写像 $T: R \times R \rightarrow R$ を, $t = T(x_1, x_2) = x_1 + x_2$ で定義する. $f(x_1, x_2)$ は, $R \times R$ 上の Lebesgue 測度 $m \times m$ に関する確率密度関数 (p.d.f.) であるとする. T の従う p.d.f. — R 上のどんな測度に関する p.d.f. であるのかをはっきりさせる必要があるが — を求めよう.

R 上の測度 λ を次で定義する: $\lambda = (m \times m)T^{-1}$. 更に, 次のような測度も同時に考える: $Q(\cdot) = \int_{T^{-1} \cdot} f(x_1, x_2) m(dx_1) m(dx_2)$. 明らかに, $Q \ll \lambda$. しかし, λ は σ -有限でない: 事実, $\lambda(\text{点}) = 0, \lambda(\text{区間}) = \infty$. したがって, Radon-Nikodym の定理 (R-D の定理) が使えず, 次の式をみたく (関数) の存在は保証できない:

$$\int_{T^{-1}B} f(x_1, x_2) m(dx_1) m(dx_2) = \int_B \langle \text{関数} \rangle \lambda(dt), B \in \mathcal{B}.$$

しかし, われわれは教科書などで, 以下のような手続きを学んでいる——

- (1) 適当なダミー変数 v を用意し, 変換 $(x_1, x_2) \rightarrow (t, v)$ を考える.
- (2) この変換の Jacobian を計算し, $m(dt)m(dv)$ に関する積分に変更する.
- (3) v に関する積分を実行すると, T の Lebesgue 測度 m に関する p.d.f. を得る.

このことは, 次の事実が裏付けとなっている. λ は σ -有限ではないが, $\lambda \ll m$, よって, $Q \ll m$; R-D の定理により, 次をみたく (関数) が m -一意にきまる:

$$\int_{T^{-1}B} f(x_1, x_2) m(dx_1) m(dx_2) = \int_B \langle \text{関数} \rangle m(dt), B \in \mathcal{B}.$$

この (関数) が, T の Lebesgue 測度 m に関する p.d.f. に他ならない.

2. 写 像 T^*

上の考察は, 以下で述べる一般的な定義へと発展する. T を測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ から可測空間 $(\mathcal{Y}, \mathcal{G})$ への写像とする. μ は σ -有限であるとし, \mathcal{Y} 上には測度 $\nu_0 = \mu T^{-1}$ を誘導する. ν_0 は一般に σ -有限ではないが, 測度 ν — σ -有限で, $\nu_0 \ll \nu$ — は必ず存在する. $f(x)$ は \mathcal{X} 上の関数で, μ -積分確定であるとする. このとき, 次の関係をみたく \mathcal{Y} 上の関数 $g(t)$ が, ν -一意にきまる:

$$\int_{T^{-1}A} f(x) \mu(dx) = \int_A g(t) \nu(dt), A \in \mathcal{G}.$$

この g を, T^*f と書く. 特に, μ が確率測度, ν が T で誘導される確率測度の場合には, $T^*f = E\{f | T\}$. この写像 T^* は, 条件つき期待値と同様な諸性質をもっているが, T^* 固有の性質もある. これに関する議論, および統計への応用については著者の論文(1984, *Ann. Inst. Statist. Math.*, p. 181-197, 36) で詳しく取り扱った.

多変量解析における標本分布の計算法

安 芸 重 雄

多変量解析に使われる統計量の分布は, 直交群上で, その上の Haar 測度による積分を含んだ形に書けることが多い. 正値対称行列上に値をとる確率行列の固有値のように直接書けるもの他に, 行列変数の超幾何関数で表わされる分布がそれに当る. James (1961, *Ann. Math.*) は zonal 多項式を用

いてその積分を展開したが、実際には高次の zonal 多項式を求めることが困難であるため、2変量の場合を除いて、標本分布を計算することには殆ど成功していない。

しかし、直交群上の積分を展開せずに、Cayley 変換を用いて Euclid 空間上の積分に置き換えると、標本分布の計算に都合の良い場合がある。

例 Wishart 行列の最大固有根の分布関数

A は $m \times m$ の確率行列で、自由度 n , covariance matrix Σ の Wishart 分布に従い、 $l_1, l_2, \dots, l_m (l_1 > l_2 > \dots > l_m > 0)$ を A の固有根とする。 l_1, \dots, l_m の同時密度は m 次直交群上の積分を含んだ形で書けるが、Cayley 変換を使うと $m(m-1)/2$ 次元 Euclid 空間上の積分で表わすことができる。その式を l_2, \dots, l_m について積分すると l_1 の分布が導出できるが、これに関し次の結果を得た。(詳細は Aki (1983, Research Memo. No. 256) を参照)

有界な連続関数 f, g が存在し、

(1) $m=2$ のとき、 n が奇数であれば、

$$P(l_1 \leq x) = \int_{\pi/2}^{\pi/2} f(n, \Sigma, x, \theta) d\theta$$

と書ける。

(2) $m=3$ のとき、 n が偶数であれば、

$$P(l_1 \leq x) = \int_{\pi/2}^{\pi/2} \int_{\pi/2}^{\pi/2} \int_{\pi/2}^{\pi/2} g(n, \Sigma, x, \theta_1, \theta_2, \theta_3) d\theta_1 d\theta_2 d\theta_3$$

と書ける。 f, g の具体的な形については上記文献を参照。

また、この結果を使って、 l_1 の分布関数を求めるプログラムを作成した。

意思決定問題における情報システムの比較

村上 征 勝

意思決定問題において、複数個の情報システムが存在する時、それらのシステムを比較する有効な方法はいまだ提案されていない。しかし、次の様な場合には partial order がつく。

いま、 $B_X(W), B_Y(W)$ をシステム S_X, S_Y からの情報 X, Y を用いた時の、事前確率 W に対するベイズ期待危険度とする。もしこの時

$$\textcircled{1} \quad \begin{cases} B_X(W) \leq B_Y(W) \text{ for all } W \\ B_X(W) < B_Y(W) \text{ for some } w \end{cases}$$

なら、ベイズ期待危険度で比較する限り S_X の方が良い。問題は、 $\textcircled{1}$ が成立する為の条件である。状態数が 2 の場合に、次のことがいえる。

情報 X, Y および損失 $L(\theta, a)$ が表 1, 2, 3 で与えられた時

i) 任意の $k (k=2, \dots, n-1)$ に関し

$$\frac{\nabla_{n-k+2} \Delta_{k+1}}{\nabla_{n-k+1} \Delta_k} > \frac{q_{11} q_{2m'}}{q_{21} q_{1m'}}$$

で、かつ

ii) 任意の $i (i=1, \dots, m-1)$ に関し

$$0 < \frac{\sum_{k=1}^j q_{1k} - \sum_{k=1}^i p_{1k}}{\sum_{k=1}^j q_{2k} - \sum_{k=1}^i p_{2k}} > \frac{p_{2,i+1}}{p_{1,i+1}}$$

となる $j (j=1, \dots, m')$ が存在するならば $\textcircled{1}$ が成立する (証明略)。

例にあげた情報は i) ii) の条件を満たす。したがって $\textcircled{1}$ が成立 (図参照)。

表1. 情報 X

	x_1	x_2	...	x_m
θ_1	p_{11}	p_{12}	...	p_{1m}
θ_2	p_{21}	p_{22}	...	p_{2m}

表2. 情報 Y

	y_1	y_2	...	y_{m^1}
θ_1	q_{11}	q_{12}	...	q_{1m^1}
θ_2	q_{21}	q_{22}	...	q_{2m^1}

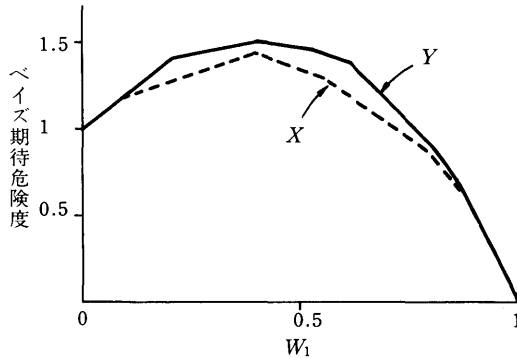
表3. 損失 $L(\theta, a)$

	a_1	a_2	...	a_n
θ_1	Δ_1	$\Delta_1 + \Delta_2$...	$\Delta_1 + \Delta_2 + \dots + \Delta_n$
θ_2	$\nabla_1 + \nabla_2 + \dots + \nabla_n$	$\nabla_1 + \nabla_2 + \dots + \nabla_{n-1}$...	∇_1

(例)

	a_1	a_2	a_3	
θ_1	0	1	3	
θ_2	5	2	1	

	x_1	x_2	x_3		y_1	y_2	y_3
θ_1	0.5	0.3	0.2		0.6	0.25	0.15
θ_2	0.1	0.4	0.5		0.3	0.35	0.35



余震時系列の統計解析

尾形良彦

修正大森法則と尤度法 (Ogata, 1983)

両対数紙を使用する従来の推定法に替えて、発震時をそのまま使い、点過程の尤度法による推定法を提案した。これで推定誤差の算出、異なる余震列間の比較、二次余震の有無等さまざまな統計的議

論がモデルの選択の問題に帰着できる。

余震活動から定常活動への遷移 (1984)

地震サイクルの各時階における地震活動を定量化するために、まず余震活動から定常活動への遷移をとらえてみた。大地震前の静穏期が、単に前の地震の余震活動の終息の結果であり、予兆現象ではないとする考え (Lomnitz ら, 1983) があるが、これは余震活動から定常活動への遷移が確立されていないためであろう。解析の例として、1965年アリューシャン地震 ($M_w=8.7$) の余震をとりあげる。

宇津 (1961) の改良大森公式に基づき、余震系列を危険度関数 $\lambda(t)$

$$(1) \quad \lambda(t) = K_1(t+c_1)^{-P_1}$$

$$(2) \quad \lambda(t) = K_1(t+c_1)^{-P_1} + H(t-t_2)K_2(t-t_2+c_1)^{-P_1}$$

の点過程として取り扱う。ここで t は主震後の経過時間、(1) 式は単一の余震系列の場合、(2) 式は二次余震が $t=t_2$ から始まる場合を示す。危険度関数 $\lambda(t)$ の積分

$$A(t) = \int_0^t \lambda(s) ds$$

を τ とし、時間 t の代りに τ を用いると、発生頻度が τ によらず定常な時系列へと変換できる。すなわち (1) または (2) 式を用いて、 $\lambda(\tau)=1$ となるように時間軸を圧縮・伸張したと考えれば良い。主震発生後 $T_0=3$ 時間から $T_1=1,000$ 日までのデータを用い、Ogata (1983) の方法で (1) 式のパラメータを推定して求め、 τ に対する地震の累積頻度をみると (1) 式が適用できれば、 T_0 から T_1 までほぼ直線となるはずであるが、主震発生後約 50 日後に明瞭な "こぶ" が認められる。これは最大余震 (1965年3月30日, $M_w=7.6$) に伴う二次余震が発生したためである。最大余震の発生時を t_2 として、 T_0 から T_1 までの期間のデータに対し主震発生後約 2,200 日まで、ほぼ直線のグラフが得られる。AIC から二次余震を考慮した方がよいことがわかる。2,200 日以降の活動は改良大森公式から外挿される傾向と著しく異なり、定常活動を表していると解釈した。こうして余震活動から定常活動への遷移を、直截的かつ客観的に検出することができる。定常活動の確立は、前兆としての静穏期をとらえるために不可欠と思われる。なお、これは東大震研の島崎邦彦氏との共同研究である。

第2研究部

大気環境データの統計解析 (II)

樋 口 伊 佐 夫

大気汚染の発生、拡散、輸送、消滅のプロセスに関しては、分子拡散や乱流拡散など、実験室規模や、理想化された条件の下での物理現象の知見を用いて、濃度予測や監視システムの設計計算が行われている。しかし、われわれの生活環境の規模とスコープにおいては、果たしてそのような拡散方程式で記述される現象が支配的であるかどうか疑わしい。それで、現象を実際の場の観測データにもとづいて調べなおす必要がある。まず、風向風速から推定した移動時間を勘定に入れた二地点の濃度のクロス相関を利用して、拡散方程式の移送の項に相当するものを検証することを考えた。それによると、 NO_x などでは風による輸送効果は殆どないことがわかる。これと相関関数の形などを考えあわせると、風によって運ばれるというイメージは正しくなく、風によって吹き上げられると想像した方がよいといえる。なお、拡散項に相当する効果はさらに小さいと思われる。

汚れた空気は、風で吹き上げられても、逆転層の下に閉じ込められると考えられるので、逆転層の高さの逆数を指標として解析に導入すれば、冬期の NO_x の高濃度などで風速だけでは説明のつかない部分が一気に解消するのではないかと考えた。逆転層に関しては、茨城県館野で毎日2回あげられるラジオゾンデの高層気象データ (気圧と温度) から計算で求めるしかない。(観測データからその高さを算出する手続きは簡単ではないが、われわれの統計解析のためには、きわめて簡素な計算で推定した値で充分なことがわかるので、それを用いる。地上温度は館野ではなく、考察地点のを用いる

べきことが結果的にもわかる)。館野に最も近いという意味で足立を選びNO濃度との関係を見たが、期待に反し、高濃度発生機構に結び付く結果は得られなかった。高層気象で問題になる所は高すぎて、地上の汚染の現象には関係がないらしい。それで一時はこの情報の利用を諦めねばならないかと思った。ところが、いろいろの地点、いろいろの種類の汚染物質について調べて比較してみると、例えば、冬期のNO_xでは館野から遠いほど相関が小さいという特徴が顕著であるというようなことが出てきた。そこで、今後逆転層の発生・移動・消滅に関してすこしの推測を行い、時間的・空間的にすこしだけ精密な計算をすれば、地上と空中との濃度の関係などについて相当の知見が得られるのではないかと思われる。

なわばりのパターンと生成機構

種 村 正 美

平面において一つの点の存在が他の点の位置に影響を及ぼす場合、即ち点間に相互作用が働く場合を考えよう。例えばなわばり性動物の個体群を考える。この個体群がつくる配置データが与えられたとき、(1) 個体間相互作用、(2) パターン生成過程、の両者について推論することが本稿での問題である。

(1) については点間距離 r の関数として相互作用ポテンシャル $\phi(r)$ を推定する尤度法 (Ogata・Tanemura, 1981, 1982) により、種々のタイプのモデル族の中から最適のモデルが得られる。しかし、この方法は点配置がギブス分布に従うこと、即ちデータは全ポテンシャルがなるべく小さくなるよう各個体の位置が調節されて生じた配置の一つであると仮定されている。これに関して、Hasegawa・Tanemura (1978, 1980) は動物のなわばりパターン形成過程を二つ考察している。一方は、多数の個体が同時に生息地に入り位置を調節しながらなわばりを形成する場合(調節モデル)、他の一つは個体が逐次到着し先着順になわばりを確保する場合(逐次充填モデル)である。パターン生成に関するこの二つのモデルは、都市の形成や分子の表面吸着等の問題でも一般性をもつと考えられる。

さて、ある相互作用 $\phi(r)$ に対する調節モデル(これはギブス分布を採用)と逐次充填モデルとを比較したいのであるが、われわれは経験分布によるモデルのあてはめで両者を比べた(種村, 1983)。

いま、空間配置を表現する統計量 T のデータの経験分布を $\hat{F}(t)$ とし、モデルのそれを $F(t; \theta)$ とする (θ はモデルの母数)。 $F(t; \theta)$ は実際には実験によって求める。両者のくい違いの尺度として

$$d(\theta) = \int |F(t; \theta) - \hat{F}(t)| dt$$

を選ぶ。そして $d(\theta)$ を小さくするモデルが良いモデルと定める。

T としてここでは最近接距離 r_1 および r_2 をとり上げる。 r_1 はランダム点とこれに最近接の個体との距離で、 r_2 は各個体から最も近い個体までの距離である。このうち r_1 の分布関数 $p(t)$ の計算は、ポロノイ多角形分割を用いれば精密に行なえる(Lotwick・Silverman, 1982)。そのため、 r_2 の分布 $q(t)$ に比べ、 $p(t)$ はバラツキが少なく感度が良い。この方法でいくつかのデータを解析した。

集中解析法の経済分析への適用について

田 口 時 夫

計量的経済分析にとって重要な概念の一つに弾力性係数がある。一般に \tilde{y} の x に関する弾力性係数はその微分可能性を前提として、

$$\eta(x) = \frac{\alpha \log \tilde{y}}{\alpha \log x} = \frac{x}{\tilde{y}}$$

によって規定される。この場合、 y の x に関する集中回帰曲線の局所的解析が必要となる。このとき次に定義するベクトルは重要な役割を果たす。

定義1 「集中回帰曲線上の任意の点 $L_1(x)$ における単位切線ベクトル及び単位従法線ベクトルをそれぞれ $\tau(x) = (\tau_1(x), \tau_2(x), \tau_3(x))$, $\beta(x) = (\beta_1(x), \beta_2(x), \beta_3(x))$ とする。この時、単位弾力性ベクトル $e(x)$ を次のように定義する。

$$(1) \quad e(x) = \frac{(\tau_1(x)\beta_1(x), \tau_2(x)\beta_2(x), \tau_3(x)\beta_3(x))}{\|(\tau_1(x)\beta_1(x), \tau_2(x)\beta_2(x), \tau_3(x)\beta_3(x))\|}$$

又、第 (a) 図の座標原点を含み、均等線に直交する平面を弾力性平面とする。」

この時、容易に次の結果が得られる。

定理1 「一般に

$$(2) \quad e(x) = \frac{(\eta(x)-1, \eta(x), 1)}{\sqrt{2\{\eta^2(x)-\eta(x)+1\}}}$$

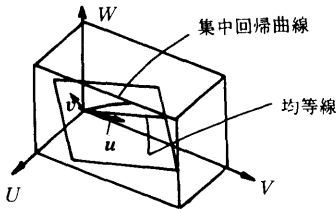
が成立する。又 $e(x)$ を原点を始点とする座標ベクトルと考えると、弾力性平面上の単位円上に終点をもつ。」

今、弾力性平面上に適当な座標軸を設け、 uv 座標系とした時、 $e(x)$ はこの座標系に関して

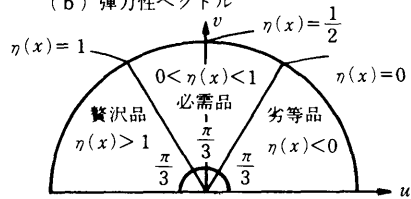
$$(3) \quad u = \frac{2\eta-1}{2\sqrt{\eta^2-\eta+1}}, \quad v = \frac{\sqrt{3}}{2\sqrt{\eta^2-\eta+1}}$$

となる。従って $e(x)$ は $\eta(x)$ の値に応じて下図 (b) のような存在領域を示す。特に η が一定のとき、これはエンゲル係数と考えられ、(b) 図のような商品の分類領域を与えることが出来る。 $e(x)$ を利用すると、更に有用な結果が得られることが予想される。

(a) 弾力性平面



(b) 弾力性ベクトル



空間補間と物理モデル

岸 野 洋 久

大気汚染濃度の空間分布およびその時間変動を解析するとき、問題となるのは、いくつかの常設測定点における観測値から他の場所における濃度を推定することと、さらにその濃度分布を予測することである。前者は補間であるが、後者においては時間変動のモデル——例えば島国日本では海陸風により風が日変動する為、これの移流の情報は欠かせない——が必要となる。そこで観測値からの情報に物理的な先験情報および空間的平滑性を加味すると、1つの量に複数の事前情報を持つことになるが、これらを最小二乗の意味で良く再現する1つの事前分布に集約することを考える。

時刻 t 地点 (i, j) における濃度と排出量を $x_{ij}(t)$, $\lambda_{ij}(t)$ とし、それらを並べて x_t, λ_t とおくと物理モデルは

$$x_{t+1} = A_t x_t + B_t \lambda_t$$

と定式化される。そこで次のように濃度分布の推定・予測値は逐次更新されてゆく。

[1] 時刻 t における濃度分布の推定量を $\bar{x}_{t|t}$ とおくと、このときの x_{t+1} に関する情報は

$$\mathbf{x}_{t+1} = A_t \bar{\mathbf{x}}_{t|t} + B_t \boldsymbol{\lambda}_t + \boldsymbol{\varepsilon}_{t+1}^{(1)} \quad \text{Var } \boldsymbol{\varepsilon}_{t+1}^{(1)} = \sigma^2 I$$

で与えられる。予測値は $\bar{\mathbf{x}}_{t+1|t} = A_t \bar{\mathbf{x}}_{t|t} + B_t \boldsymbol{\lambda}_t$ である。

[2] [1] における情報に加えて時刻 $t+1$ において観測値 \mathbf{y}_{t+1} が入って来るとこの時点での \mathbf{x}_{t+1} に関する情報は

$$\begin{cases} \mathbf{x}_{t+1} = \bar{\mathbf{x}}_{t+1|t} + \boldsymbol{\varepsilon}_{t+1}^{(1)} & \boldsymbol{\varepsilon}_{t+1}^{(1)} \sim N(\mathbf{0}, \sigma^2 I) \\ D\mathbf{x}_{t+1} = \alpha \boldsymbol{\varepsilon}_{t+1}^{(2)} & \boldsymbol{\varepsilon}_{t+1}^{(2)} \sim N(\mathbf{0}, \sigma^2 I) \\ \mathbf{y}_{t+1} = F\mathbf{x}_{t+1} + \beta \boldsymbol{\varepsilon}_{t+1}^{(3)} & \boldsymbol{\varepsilon}_{t+1}^{(3)} \sim N(\mathbf{0}, \sigma^2 I) \end{cases}$$

で表される。このうち第2式は

$$x_{i+1,j}(t+1) + x_{i,j+1}(t+1) + x_{i-1,j}(t+1) + x_{i,j-1}(t+1) - 4x_{i,j}(t+1) = \alpha \boldsymbol{\varepsilon}_{t+1}^{(2)}(t+1)$$

をベクトル表現したもので、空間的平滑性の事前情報を表わす。

上式において、 \mathbf{x}_{t+1} に対して2つの分布が対応していることに注意を要する。そこでいま、これをまとめて

$$\mathbf{x}_{t+1} = (I + \alpha^{-2} D' D)^{-1} \bar{\mathbf{x}}_{t+1|t} + (I + \alpha^{-2} D' D)^{-\frac{1}{2}} \boldsymbol{\eta}_{t+1} \quad \boldsymbol{\eta}_{t+1} \sim N(\mathbf{0}, \sigma^2 I)$$

という事前分布を考えると、 \mathbf{x}_{t+1} と \mathbf{y}_{t+1} の同時分布は

$$C \exp \left[-\frac{1}{2\sigma^2} \left\{ \|\mathbf{x}_{t+1} - \bar{\mathbf{x}}_{t+1|t}\|^2 + \frac{1}{\alpha^2} \|D\mathbf{x}_{t+1}\|^2 + \frac{1}{\beta^2} \|\mathbf{y}_{t+1} - F\mathbf{x}_{t+1}\|^2 \right\} \right]$$

となって、上記事前分布は2つの事前情報を最小二乗的に見て自然な形に取り込んだものであることがわかる。

DNA 配列データから系統樹を推定するための最尤法

長谷川 政 美

DNA, RNA, たんぱく質など情報高分子の配列データは、生物の系統進化を研究する上で客観的な手掛りを与えるものである。配列データから系統樹を推定する方法としては各種の行列法があり、簡便なためにひろく使われている。しかし、これらの方法では配列間の pair-wise な difference のみを用いるために、せっかくの配列データに含まれる情報の多くを有効に生かしていない。この点で一歩進んだものとして、最大節約法がある¹⁾。塩基置換数が最小になるような系統樹が最も真実に近いとするものである。この方法では進化の過程で起こる塩基置換は確率的なもので、統計的なゆらぎを伴うものであるということが考慮されていない。従って、得られた系統樹が別の系統樹にくらべてどの程度有意なものであるかを評価することができないという欠点がある。ここで簡単のために塩基の状態にはプリンRとピリミジンYの2通りしかないものとする。図1のように4種の配列から成る系統樹を考え、各枝での塩基置換の確率を p, q とし、枝 $\overline{15}$ と $\overline{26}$ での置換確率 p は他の枝での確率 q よりも高いものとする。この場合 $p^2 > q(1-q)$ ならば、最大節約法はデータが十分多い(配列が長い)と、必ず3と4とが近縁という間違っ系統樹を与えることになる^{2,3)}。

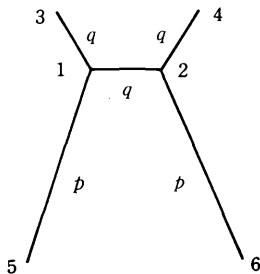


図1.

表1. シミュレーションによる適中率の比較 ($k=5$)

α	最尤法	節約法
0.1	6/6	3/6
0.2	5/6	0/6
0.3	6/6	0/6
0.4	5/6	0/6
0.5	5/6	0/6
0.6	1/5	0/5
0.7	0/5	0/5

これに対して Felsenstein⁴⁾ は dt 時間で i が j に遷移する確率 $P_{ij}(dt)$ が $P_{ij}(dt) = (1 - udt)\delta_{ij} + udt \pi_j$ で与えられる定常マルコフ過程を考え、尤度を最大にするように系統樹のトポロジーと各枝の長さ (ut) を選ぶ最尤法を提案した。ここでは時間当りの塩基置換率が系統ごとに大きく異なる場合でも、この最尤法はかなりの程度まで使えることをシミュレーションで示す。図1のような系統関係のもとでシミュレーションを行なう。長さ1,000の配列を考え、node 1で RYRYRY... の alternate copolymer であるとし、各塩基座位が

$$p = (1 - e^{-a})/2, \quad q = (1 - e^{-a})/2$$

なる確率で独立に進化して、現存配列 3, 4, 5, 6 ができてくるとする。 $k=5$ の場合いろいろな a の値について真の系統樹を再現する割合 (適中率) を表1に示す。

References

- 1) 長谷川政美 (1982). 化学と工業, 35, p 102.
- 2) Felsenstein, J. (1978). Syst. Zool., 27, p 401.
- 3) 長谷川政美 (1984). 数理科学, No. 248, p 40.
- 4) Felsenstein, J. (1981). J. Mol. Evol., 17, p 368.

本研究は特別研究員・矢野隆昭氏 (昭和大学) との共同研究である。

比例ハザードモデルにおける最大全尤度法の利用

柳 本 武 美

1. 問 題

比例ハザードモデルは生存時間モデルの中で実用的なモデルとして導入され、広く用いられている。モデルは (離散型の場合)、

$$\frac{\lambda(t; z)}{1 - \lambda(t; z)} = \frac{\lambda_0(t)}{1 - \lambda_0(t)} \exp \beta z,$$

ただし $\lambda(t; z)$ は共変量 z をもつ個体が時刻 t に死亡する確率を $p(t; z)$ としたとき、 $\lambda(t; z) = p(t; z) / \sum_{s \geq t} p(t; z)$ で、 $\lambda_0(t)$, β は母数である。

比例ハザードモデルは多重 2×2 分割表 (ケース・コントロール研究) 順序カテゴリカルな目的変数に対する回帰分析、点過程における比例強度モデル解析的に同じ問題をもっている。

2. 全尤度と部分尤度

得られたデータから (一般化) 全尤度と部分尤度を求めることができる。全尤度の β に対する尤度方程式を $\sum_t LF_t(\beta) = 0$ 部分尤度の尤度方程式を $\sum LP_t(\beta) = 0$ とする。以下で β をスカラー量とすると

$$\begin{aligned} LP(0) &= LF(0) \\ LP'_t(0) &= n_t - 1/n_t LF'_t(0) \\ &\quad (n_t \text{ は } t \text{ での観察個体数}) \\ LP(\infty) &= LF(\infty) \\ LP(-\infty) &= LF(-\infty) \\ LP(\beta), LF(\beta) &\text{ は狭義単調減少} \end{aligned}$$

が成り立つ。

更に弱い仮定の下で $\beta(LP(\beta) - LF(\beta)) < 0$ for $\beta \neq 0$ が成立する。このことは最大全尤度推定量 $\hat{\beta}_F$ の絶対値は最大部分尤度推定量 $\hat{\beta}_P$ より大きいことを示している。表は $\sum LF(n_t / (n_t - 1)\beta) = 0$ が

表

	共 変 量	死 亡 数		
		t_1	t_2	t_3
群 1	0	2K	2K	4K
群 2	1	2K	3K	2K
群 3	2	3K	2K	K

結果

K	$\hat{\beta}_P$	(CPU Time)	$\hat{\beta}_A$	(CPU Time)
1	.59317	(.012)	.59429	(.007)
2	.61178	(.037)	.61236	(.007)
3	.61798	(.079)	.61838	(.007)
4	.62108	(.136)	.62138	(.007)
5	.62294	(.211)	.62318	(.007)
6	.62418	(.302)	.62439	(.007)
7	.62507	(.409)	.62524	(.007)
8	.62574	(.533)	.62589	(.007)
9	.62625	(.670)	.62639	(.007)
10	.62667	(.825)	.62679	(.007)
$\infty(\hat{\beta}_F)$.63039			

$\hat{\beta}_F$ の近似値を与えることを示している。

3. 全尤度の長所

- モデルの中のすべての母数を推定するのでモデルの比較ができる。 $\lambda_0(t)$ は Nuisance であるという常識は実際的ではない。
- 表にもみられるように $\hat{\beta}_F$ の計算は計算量が遙かに大きい。

生存時間分布の推定におけるベイズモデルの利用

鎌 倉 稔 成

生存時間分布の統計的研究では、2つの集団の生存時間分布関数の差を知りたいということがよくある。通常、差のあるパラメータによってモデル化し、分布を決定する他のパラメータは攪乱母数として扱う。しかしながら、分布の形状も知りたいし、差のパラメータの推定もしたいという欲ばった要求が現実には多い。たとえば、Cox (1972) の比例ハザードモデルでは、ハザードとして、

$$\lambda(t) = \lambda_0(t) \exp(z\beta)$$

を考えているが、この場合、回帰係数 β が構造母数 (structure parameter) として扱われ、 λ_0 は攪乱母数 (nuisance parameter) の扱いを受けることが多い。しかし、実際のデータ解析に際しては、回帰係数 β の値だけでなく分布を決定するのに必要な $\lambda_0(t)$ の推定もなんらかの方法でおこなう (たとえば、Kaplan-Meier のようなノンパラメトリックな方法) 必要性がでてくる。

一般に、こうした攪乱母数を含むような問題では、計算の複雑さを避けたり、推定量の偏りをなくするために条件付尤度・周辺尤度・部分尤度を用いて攪乱母数を除いた形で推定をおこなう方法がとられている。この場合、攪乱母数は推定しないことを前提としているが、比例ハザードモデルでは β を推定したのち、その β を用いて λ_0 を推定しなおすということがよくおこなわれている。比例ハザードモデルの場合には、攪乱母数として取り扱われている λ が時間軸上で滑らかに変動していると考えら

れ、Akaike (1980) のベイズモデルの利用が考えられる。

ここでは比例ハザードモデルに限らずに、一般に構造母数に先験情報がなく、時間軸上に整列した攪乱母数に滑らかさの情報を持ったモデルを考えて、ABIC 最小の基準によるパラメータ推定の問題を議論する。ABIC の計算の困難さはパラメータの分布による尤度の積分にあるが、初期値 β_0 での対数尤度の 2 次近似を利用して ABIC を計算し、超母数とそれに対応する推定量を求め、初期値を更新するといったイテラティブな方法を用いている。

第3研究部

変化する標本の大きさをもつ経験的ベイズ決定

藤 本 照

比較のため、まずベイズ推論のモデルの構造に触れておく。観測値 x を与えた母数 θ の尤度関数を $f(x|\theta)$ とし、 θ の先験確率分布の密度関数または確率関数を $w(\theta)$ とするとき、ベイズの定理から、事後確率分布は、

$$w(\theta|x_1) \propto w(\theta)f(x_1|\theta).$$

観測値 x_1 を得たあと、これに独立な観測によって、観測値 x_2 を得たとき、 x_1, x_2 を与えた θ の事後確率分布は、

$$w(\theta|x_1, x_2) \propto w(\theta|x_1)f(x_2|\theta).$$

同様に大きさ n の標本の観測結果 (x_1, x_2, \dots, x_n) に対して、

$$w(\theta|x_1, x_2, \dots, x_n) \propto w(\theta) \prod_{k=1}^n f(x_k|\theta).$$

つまりベイズ推論は、固定された母数 θ の値に対して、観測が繰返されるわけで、 $w(\theta)$ は実験には先験的に想定される確率分布であることになる。それ故、どのような考え方で $w(\theta)$ を扱うとしても、ベイズ推論は使用する確率に対して主観主義派であることになる。あるいは、確率モデルの中の $f(x|\theta)$ に含まれる x と θ の間の関係を逆転して、 $w(\theta|x)$ をつくるための便宜のために $w(\theta)$ を使用するともいえる。

これに対して経験的ベイズは、使用する確率に対して客観主義派である。そこで、そのモデル構成を示すと、顕在変数 X と潜在変数 θ の対の系列

$$(\theta_1, X_1), (\theta_2, X_2), \dots, (\theta_n, X_n), (\theta_{n+1}, X_{n+1})$$

を考え、 $\theta_i (i=1, 2, \dots)$ は $w(\theta)$ をもち、 $\theta_i = \theta$ のとき、 X_i は $f(x|\theta)$ をもつとする。ただし $w(\theta)$ は未知である。このとき問題は、 X_{n+1} の観測結果から θ_{n+1} に関する決定を為すのに、 (X_1, X_2, \dots, X_n) の結果を利用して、ベイズ危険度へ収束するような、決定法則をつくることである。

ところで、この問題は、1970年代はじめ迄は、各 (θ_i, X_i) における決定は、同一構造で、且それが繰返されてゆくものであった。例えば、

$$X_i = (X_{i1}, X_{i2}, \dots, X_{im_i})$$

と考えるとき、添字 $m_i = m$, すなわち同一の標本の大きさ—というようなものであったが、この制限を除いた場合の問題が、T.E. O'Bryan などによって考えられるようになった。

本研究は今迄行ってきたものを、この傾向へむけ拡張しようとするものである。

多次元ランダムパッキングの充填率

伊 藤 栄 明

m 次元立方体を向きをそろえてランダムに充填して行くという問題について議論する。1次元の場合は解析的研究が可能である。充填率は

$$1 \leq x \text{ において} \quad f(x+1) = \frac{2}{x} \int_0^x f(y) dy + 1,$$

$0 \leq x < 1$ においては $f(x) = 0$, なる式より導かれる (Renyi (1958)). 隙間の長さの最小値あるいは最大値の確率分布の満す方程式として

$$f(x+1) = \frac{1}{x} \int_0^x f(x-y)f(y) dy$$

が導かれる (Itoh (1980)). $0 \leq x < 1$ における $f(x)$ の値を与え, 解の挙動をしらべるのは興味ある解析学の問題と思われる. 多次元の場合, 辺の長さが十分長い m 次元立方体への m 次元単位立方体のランダムパッキングの充填率 C_m の期待値は 1 次元の場合の充填率の期待値の m 乗になるという Palasti の予想がよく知られている. これは近年計算機実験により否定され Blaisdell と Solomon (1982) は実験式

$$C_m^{1/m} - C_1 = (m-1)(C_2^{1/2} - C_1)$$

を $m=3, 4$ について適用している. 一辺の長さ $2n$ の m 次元の立方体があるとする. その中に一辺が長さ n の m 次元立方体をランダムにつめて行くという離散的な場合を考える. $n=2$ とし, 充填率の期待値を d_m とする. この場合も Blaisdell と Solomon の式

$$d_m^{1/m} - d_1 = (m-1)(d_2^{1/2} - d_1)$$

が $m=3, 4$ についてよく適合することがわかる. このようなもっとも単純な場合について m を大きくして行くかどうかという問題を考える. 計算機実験による $m=11$ まで求めた. この結果より m を大きくして行くと $d_m^{1/m}$ が 1 に近づいて行くことが予想される. $d_m^{1/m} - d_1$ についての実験式を求めるのも興味ある問題と思われる. 計算機実験によらずに数値計算により充填率の期待値を求めるという方法も考えられる. すべての場合をつくして行くこの方法は計算時間及び誤差の問題により 3 次元までしか求めることができない. これに関連して 1 次元の場合にパッキングにより生成される tree のかぞえ上げの問題がある. これより

$$f(x+m) = \sum_{y=0}^x f(x-y)f(y)$$

なる方程式が導かれるがこれは組合せ数学においてよく知られた式である.

多次元ランダムパッキングの問題は符号化理論と関連があり解析的な結果を得るのが今後の課題である.

参 考 文 献

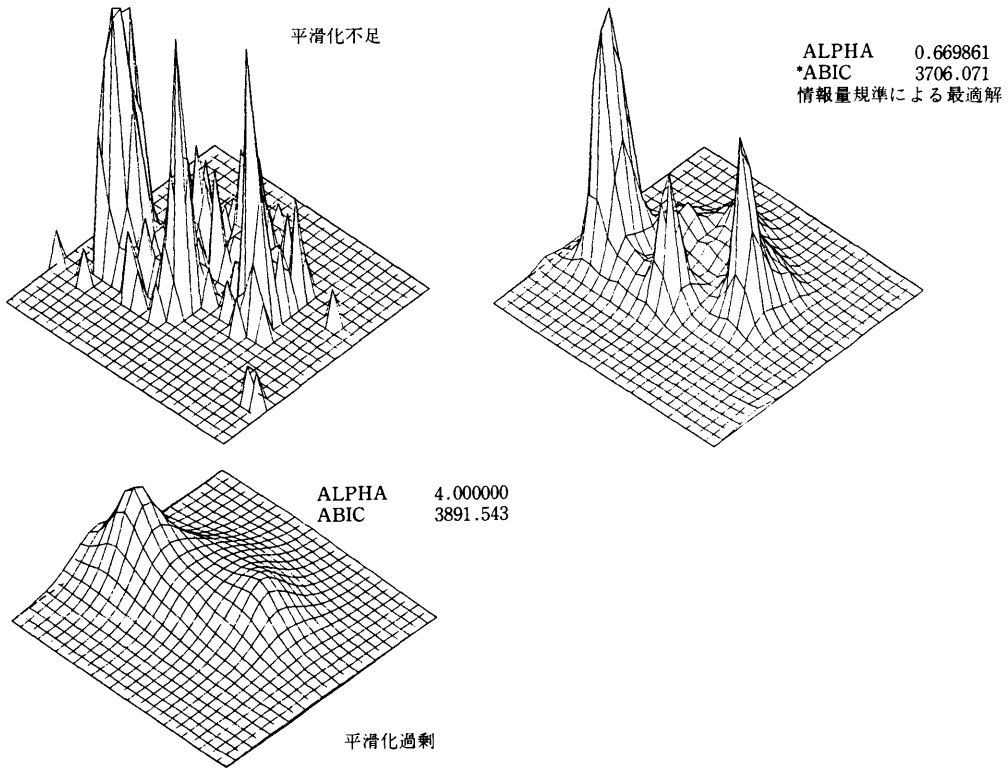
- Itoh, Y. (1980). On the minimum of gaps generated by one-dimensional random packing, *J. Appl. Prob.*, **17**, 134-144.
 伊藤栄明, 上田澄江 (1983). 多次元立方体の離散的ランダムパッキングによる充填率について, 統計数理研究所彙報第 31 巻第 1 号, 65-69.

2 次元の確率密度関数のノンパラメトリックな推定

田 辺 國 士

一次元および二次元の確率密度関数をノンパラメトリックに推定するため, Leonard の logistic density transform と変則先験分布を用いるベイズモデルを構成し, 推定関数の滑らかさを制御するハイパーパラメータの値をデータに基づいて適切に決めるための (ABIC に対応する) 情報量規準を導びいた. また singular なこのモデルにおける情報量を計算するための Monte Carlo 法による方法を開発した. 下図はこの方法による地震データの適用例である.

またこれに関連して行列 $P - pp^t$ (ただし $P = \text{diag}(p_1, p_2, \dots, p_n)$, $p = (p_1, p_2, \dots, p_n)^t$ の exact Cholesky 分解公式,



$$P - pp^t = LDL^t$$

$$L \equiv (l_{ij}): l_{ij} = -p_i/q_j (i > j), 1 (i = j), 0 (i < j),$$

$$D = \text{diag}(d_1, \dots, d_n)$$

$$d_i = p_i q_i / q_{i-1}, \quad q_i = 1 - \sum_{k=1}^i p_k,$$

および Moore Penrose 一般逆行列の公式を得た。これを用いると、ベイズモデルの smoothness prior の設定において構造の対称性を損うことのないモデル化が可能となり、随伴する数値計算においてモデルの悪条件性を克服することができる。

その他、本年度は以下の成果を得た。

1. 一般の特異モデルと変則先験分布モデルにおける情報量規準の導出。
2. 偏微分方程式モデルを統計的モデルとして利用するための理論的枠組の構成とそれを有限要素法を用いて運用する技術の開発。
3. X線分光学における逆問題の処理法。
4. CT 像による局所脳血流量計測法。

非正規外乱に対するシステム応答の確率分布

岡 崎 卓

1. 非正規外乱とシステム応答

外部から加えられる攪乱に対してシステムがどのような挙動を示すかという問題は近年益々注目を集め、電圧変動によって振動数遷移を起す発振器、風波により過動揺の危険を伴う船舶といった機械系に留まらず環境因子の変動によって繁殖型を変える生物等、生態系においても、外乱の及ぼす影響が研究されている。その中心的課題は外乱の統計的性質からシステムの応答変数の確率分布を導びくことにある。

従来解析では外乱に関する知識の不足を補い、また解析の便を図るため、専ら正規白色過程に従う外乱のみを取扱ってきた。しかしこの仮定の妥当性を検証し、現実の外乱を忠実に模して解析の精度を向上させるにはより一般な外乱を対象とすることが必要である。以下では外乱のモデルとして白色正規過程で駆動される非線型過程を考え、システム応答の確率分布が満たすべき「一般化 Fokker-Planck 方程式」とこれを具体的に運用するための「移流項近似」について概略を報告する。

2. システムおよび外乱のモデル

外乱 W を受けて変化するシステム応答 $U = U(t)$ (t : 時間) は非線型方程式

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

に従い、 $W(t)$ は $\langle n(t)n(s) \rangle = \sigma^2 \delta(t-s)$ なる白色正規過程 $n(t)$ の非線型変換

$$\frac{dW}{dt} = N(W) + n(t)$$

で与えられるものとする。 $N(W) = -\beta W + \gamma W^{-1}$ は Rayleigh 過程、 $N(W) = -\beta W - \gamma W^3$ は硬いバネの乱振動に対応する。

3. 一般化 Fokker-Planck 方程式

応答 U の確率密度関数 $f(u)$ を定める方程式は $pX(U, W) = \int dW g(W) X(U, W)$ (g は外乱 W の定常確率密度) なる射影子 p を用いて次のように表わされる、

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial u} M(u) f(u, t) + \frac{\partial}{\partial u} \int_0^t ds \int_0^\infty dv D(u, v, t-s) \frac{\partial}{\partial v} f(v, s),$$

$$D(u, v, t) = \int dU dW g(W) W \delta(U-u) e^{i\hat{L}(1-p)t} W \delta(U-v),$$

$$i\hat{L} = i\hat{L}_1 + i\hat{L}_2, \quad i\hat{L}_1 = M(U) \frac{\partial}{\partial U}, \quad i\hat{L}_2 = N(W) \frac{\partial}{\partial W} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial W^2}.$$

4. 移流項近似

上記方程式は外乱の正規非正規を問わず厳密に成立つが、拡散係数 D に含まれる時間進行作用素 $e^{i\hat{L}(1-p)t}$ のためこの儘では計算し得ない。そこで $D \rightarrow \varepsilon D$ と書替え、 $f = f_0 + \varepsilon f_1 + \dots$ と展開すると f_0 は移流方程式 $\frac{\partial f_0}{\partial t} = -\frac{\partial}{\partial u} M(u) f_0$ の解となる。この f_0 を用いて拡散項を $O(\varepsilon)$ の精度で評価すれば、

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial u} (M(u) + \hat{M}(u)) f(u, t) + \frac{\partial}{\partial u} \hat{D}(u, t) \frac{\partial}{\partial u} f(u, t)$$

なる具体的計算可能な方程式が得られる。ここに応答と外乱の相互作用を繰込んだ係数 \hat{M} 、 \hat{D} は応答 $U(t)$ が時間と共に歪む度合を意味する $p(u, t)$ と $W(t)$ の時間積分として表現される、

$$\begin{aligned} \hat{M}(u) &= - \int_0^t ds \frac{\partial p(us)}{\partial u} \langle W \hat{W}(s) \rangle, \quad \hat{D}(u) = \int_0^t ds p(u, s) \langle W \hat{W}(s) \rangle, \\ p(u, t) &= \frac{\partial \hat{U}(t)}{\partial U} \Big|_{\hat{U}(t)=u}, \\ \hat{U}(t) &= e^{M(u) \frac{\partial}{\partial U} t} U, \quad \hat{W}(t) = e^{iLst} W, \\ \langle X(W) \rangle &= \int dW g(W) X(W). \end{aligned}$$

分割表に対する潜在構造モデル

鈴木 義一郎

$r \times c$ -分割表の形で与えられるデータに対して, $r \times c$ 行列

$$P = \{p_{ij} : i=1, \dots, r, j=1, \dots, c\}$$

の“確率モデル”を対応させる。特にすべての i, j に対して

$$p_{ij} = p_i \cdot p_{\cdot j} \equiv \left(\sum_k p_{ik} \right) \left(\sum_n p_{nj} \right)$$

が成り立てば, いわゆる独立モデルになる。

所与のデータの, 独立モデルからの“乖離”を調べるのが多くの研究者の関心事であった。しかし, どのような“方向”へずれているかを議論した結果となると, その数はかなり少なくなる。最近, オッズ比の行列に線型構造を仮定したグッドマンのモデルが脚光を浴びているが, 適用面で相応の制約を受けることは否めない。

いま, $P^{(0)}, P^{(1)}$ という2つのモデルを固定して

$$\{P^{(\alpha)} = \alpha P^{(1)} + (1-\alpha) P^{(0)}, 0 \leq \alpha \leq 1\}$$

というモデルの集合を考えてみる。 $P^{(0)}$ を独立モデルにとり, $P^{(1)}$ としていろいろな特殊モデルを考えてやれば, 非常に広範な現象に適合するモデルを対応させることが可能となる。所与のデータの潜在構造を勘案すれば, 然るべき $P^{(1)}$ を想定することができる。

混合モデル $P^{(\alpha)}$ の要素と α を関係づける, 典型例を1つ示してみる。 $P^{(0)}$ を独立モデルとし, $P^{(1)}$ に関して列の周辺分布が $P^{(0)}$ と同じで, 各行には少なくとも $p_{ij}^{(1)} > 0$ となる要素があると仮定する。

そこで各 i に対し, $p_{ij}^{(1)} > 0$ となる j の集合を J_i , その要素の数を c_i とする。 $P^{(\alpha)}$ の列分布は $P^{(0)}$ のそれと同じであり, $j \in J_i$ に対しても

$$p_{ij}^{(\alpha)} = (1-\alpha) p_{ij}^{(0)} + \alpha p_{ij}^{(1)}$$

という関係があるから,

$$\frac{1}{c_i} \sum_{j \in J_i} \frac{p_{ij}^{(\alpha)}}{p_{ij}^{(0)}} = (1-\alpha) + \alpha \frac{p_{ij}^{(1)}}{p_{ij}^{(0)}}$$

これを i について加えて

$$\alpha = \frac{\sum_{i=1}^r \frac{1}{c_i} \sum_{j \in J_i} \frac{p_{ij}^{(\alpha)} p_{ij}^{(0)} - p_{ij}^{(0)2}}{p_{ij}^{(0)2}}}{\sum_{i=1}^r \frac{1}{c_i} \sum_{j \in J_i} \frac{p_{ij}^{(1)} p_{ij}^{(0)} - p_{ij}^{(0)2}}{p_{ij}^{(0)2}}}$$

右辺は混合モデルに対応するデータから推定することができる。

特に $r=c$, $P^{(1)}$ の非対角要素がすべて0の場合を考えてみる。例えば, r 通りある結果の1つを予想することを「行」に, 実際実現した結果を「列」に分類すると, $P^{(1)}$ は完璧な予想, $P^{(0)}$ はデタラメな予想に対応する。従って, 混合モデル $P^{(\alpha)}$ はこの両極端の予想の中間に位し, α はその予想の“的中率”を示す尺度とみなすことができる。

この種の混合モデルをいろいろ考えてやることにより, 分割表に対する「潜在構造分析」が可能となる。

判別分析における誤判別率の推定

小 西 貞 則

Bootstrap 法 (Efron (1979; *Ann. Statist.*, 7)) はデータが正規分布に従うという仮定なしに、統計学の種々の分野の問題に適用でき統計的分析を可能にするということでその応用が試みられている。ここでは判別分析における誤判別率の推定等への応用を、正規性の仮定のもとで展開された理論との対応を踏まえて考察する。

いま母集団 Π_1, Π_2 は各々母集団分布として p 変量分布 $F(\mathbf{x}), G(\mathbf{x})$ をもつとする。判別方式として線形判別関数 $h(\mathbf{x})$ を用いると、新たな観測値 \mathbf{x} は $h(\mathbf{x}) > c$ または $h(\mathbf{x}) \leq c$ に応じて各々 Π_1, Π_2 へ判別される。ここでは先験確率、損失は等しいものと仮定し、したがって $c=0$ とする。このとき実際の誤判別率 (予測誤差) は

$$e_F = P_F\{\mathbf{x} : h(\mathbf{x}) \leq 0 \mid \mathbf{x} \in \Pi_1, \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S\}$$

$$e_G = P_G\{\mathbf{x} : h(\mathbf{x}) > 0 \mid \mathbf{x} \in \Pi_2, \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, S\}$$

で与えられる。ここに $\bar{\mathbf{x}}_i$ は母集団 Π_i からの標本に基づく標本平均ベクトル、 S は共通の標本分散共分散行列とする。

問題は $e_F(e_G)$ の推定方法であるがこれを母集団分布として p 変量正規分布を仮定した場合および母集団分布は未知の場合に分けて考察する。

(i) 正規母集団を仮定する場合

母集団分布 $F(\mathbf{x}), G(\mathbf{x})$ は各々 p 変量正規分布 $N_p(\boldsymbol{\mu}_i, \Sigma)$ ($i=1, 2$) であるとする。線形判別関数 $h(\mathbf{x})$ に基づいて判別を行うとき、 $e_F(e_G)$ は新たな観測値が $\mathbf{x} \in \Pi_1, (\Pi_2)$ であるのに対して

$$e_F \left. \vphantom{e_F} \right\} e_G = \Phi \left[\frac{(-1)^i \left\{ \boldsymbol{\mu}_i - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right\}' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\left\{ \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 \right\}' S^{-1} \Sigma S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \right\}^{1/2}} \right]$$

で与えられる。ここに $\Phi(y)$ は標準正規分布関数とする。 $e_F(e_G)$ の推定法は種々提案されているが、中でも変数選択の問題を考える時、McLachlan (1976; *Biometrics*, 32) の標本数に関する漸近展開式が有効と思われる。

(ii) 母集団分布が未知の場合

この時 $e_F(e_G)$ は直接推定できない。 $e_F(e_G)$ の簡便な推定量として所謂見かけ上の誤判別率 $A_F(A_G)$ が用いられるが、一般に $e_F(e_G)$ を過少に評価する傾向にある。そこで Bootstrap 法によって e_F を A_F で推定したときの偏り w_F を推定し e_F の推定量として、 $\hat{e}_F = A_F + \hat{w}_F$ または $\hat{e}_F = \hat{A}_F + \hat{w}_F$ を用いるものである。

〈Monte Carlo 実験〉 正規理論のもとで導出された実際の誤判別率の推定法、Bootstrap 法にもとづく推定法の比較を行うため、母集団分布として既知の混合多変量正規分布を仮定し Monte Carlo 実験を行った。その結果 Bootstrap 法に基づく実際の誤判別率の推定値 $\hat{e}_F(\hat{e}_G)$ は種々の混合比率に対して、混合多変量正規分布の仮定のもとで導出した値に近く、かなり安定した推定法であるといえる。

第 4 研究部

生活環境としての森林の有効利用に関する統計的研究

石 田 正 次

一般住民の生活環境としての自然を要望する声は高く、自然保護の問題に関しては神経過敏である。しかし、我々のこれまでの研究によれば、自然、特に森林に対する日本人の関心と知識はドイツ人やフランス人などと比較して相当に低い。その結果、観念的には自然に対する人為の介入を拒否しなが

ら、現実を求めるものは芝生や花壇に代表される人工的自然ということになる。生活圏に隣接した森林は必ずしも住民に喜ばれず、生態的に健全な森林も放置された醜い自然としか受けとられない。このような状況下では、いかに素晴らしい自然があっても、これを愛し親しんでその恩恵を享受できないばかりか、国民全体の協力による自然環境の保全を期待することもできない。本研究の目的はどうすれば自然に対する日本人の関心やモラルを向上させ得るか、その具体的な施策を実証的な立場から追求することにある。

本年度は樹木、植生、森林動物のガイド・ブックを試作し、一般住民を対象とした現地指導を実施した。この会合に集った人々について意見調査を行ったがその結果は大略次のようなものであった。

1. 野鳥、山菜、昆虫、写真などある特定のものに趣味をもつ人々が主に集った。
2. 樹木や森林そのものに関心をもつ人々は極く少数である。
3. 自分の興味をもつ対象以外にはほとんど関心を示さない。
4. 自然に対する知識は非常に局所的であって、自然全体を見、考えようとする傾向は少ない。
5. テレビや本で得た知識が自然の中のどれに対応するかがわからない。
6. 他人の行動には厳しいが自分の行動にはルーズである。
7. 子供は自然を相手に遊ぶことができない。
8. 道路以外のところは歩かない。
9. 目の上の様子よりは下の様子に注意が向く。
10. 近い風景よりは遠い景色を喜ぶ。

標本分布の高次漸近展開の収束性

仁 木 直 人

標本分布の近似を得るための有力な方法として、標本数に関する漸近展開を導出することがある。この方法に従う場合、小標本数でも実用上十分な精度を確保するためには、展開式の高次の項を求める必要がある。最近、『数式処理システム』すなわち『計算機を用いて、数式の演算を記号のままで行なうシステム』の統計解析への活用により、手ではとても計算が困難と思われる高次の漸近展開式を求めることが可能になりつつあるが、その際とくに注意しなければならないことがある。それは、『より高い次数まで求めた漸近展開式の方が、必ずしも、分布全体に関してより良い近似を与える、とは限らない』という点である。

漸近展開を k 項までで切った(n の $-k/2$ 乗のオーダーまで使った)式の値から正しい分布関数値を引いた誤差についてプロットしてみると、次数 k を上げるとともにグラフの形が一層大きく振動するようになり、近似が却って悪くなる現象が見られることがある。但し、もう少し注意深くグラフを見てみると、平均の付近ではやはり次数の高い展開式の方が近似精度が高く、また、良い近似を与える範囲も広いことが分かるであろう。このような現象は、もともと特定の多項式系では近似しにくいものを、無理やり近似を行おうとするときにしばしば見られるものである。

このような場合、統計量を何らかの関数で変換し、変換された統計量の分布の漸近展開式を求めると、収束性が著しく改善され、また、式の簡略化にも役立つ場合がある。このような変換は、高次の項を考慮した場合、歪度の漸近的性質(0に近づく速さ)を改善するように選択するのが最も得策で、2階の微分方程式を解くことにより定めることができる(『高次のオーダーまで考慮した正規化変換』と呼ぶことができよう)。

◇歪度を改善する変換の例 Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis, *Biometrika*, 68, 3.

◇変換した統計量の漸近展開例 Niki, N. and Konishi, S. (1984). Higher order asymptotic

expansions for the distribution of the sample correlation coefficient, *Commun. Statist. -Simula. Computa.*, B 13, 2.

調査特性の推測

柏 木 宣 久

内閣及び政党の支持率は複数の会社によって独立に、そして継続的に調査されている。ほぼ同時期に実施される各社の調査の結果は、標本調査論の教える所によると、前後2~3%内に収まると期待される。しかし現実には10%以上の差も珍しくない。

従来の誤差評価は過少評価であるとの指摘 [柏木, 岸野 (1983), 日本統計学会] はある。その一方で、もし各社間の差が系統的であるならば、その差は偶然誤差というより各社の調査の特性によって生じたとするのが妥当となる。

今、 $y_i(t)$, $b_i(t)$, $\varepsilon_i(t)$ を t 時点 ($1 \leq t \leq T$) での i 番目 ($1 \leq i \leq m$) の会社の調査に於ける、支持率、調査特性、誤差とする。また $q(t)$ を t 時点での真の支持率とする。更に、 $b_i(t)$, $\varepsilon_i(t)$ は確率変数で、

$$E[b_i(t)] = b_i, \sum_{i=1}^m b_i = 0, E[\varepsilon_i(t)] = 0$$

であり、各時点間で独立に、しかし時点内ではある共分散構造を持つとする。

この時、モデル

$$(1) \quad y_i(t) = q(t) + b_i(t) + \varepsilon_i(t), \quad i \in Q(t), \quad t = 1, 2, \dots, T$$

を考える。但し $Q(t)$ は t 時点で実際に調査を行った会社の番号を要素とする集合である。

$b_i(t)$, $\varepsilon_i(t)$ の共分散がこれ以上のパラメータを使用せずに記述されるならば、 $q(t)$, b_i の最尤推定値は直観的には、

$$\hat{q}(t) = \text{Mean} [y_i(t)], \quad \hat{b}_i = \text{Mean} [y_i(t) - \hat{q}(t)]$$

によって求まる。

さて問題点が2つある。ひとつは柔軟性のある誤差評価をする為には $b_i(t)$, $\varepsilon_i(t)$ の共分散が新たなパラメータを必要とする点。この場合、問題は識別不能となる。もうひとつは $\hat{q}(t)$ の推定値が不自然である事。というのは、各時点で全社が調査を行っているわけではない為、時点毎の会社の構成によって $\hat{q}(t)$ が異様に変動してしまうのである。

こうした点を改善する為、 $q(t)$ は特異な事態が生じた場合を除いて時点間で滑らかに変動するであろうという直観的な事前情報を用いたモデルを提案する。

簡単に説明すると、データに関するモデルとしては (1) を、また $q(t)$ の事前分布として2階差のモデル

$$q(t) = 2q(t-1) - q(t-2) + \delta(t)$$

を利用する。但し $\delta(t)$ は正規分布に従う確率変数である。

疾病の発症と予後に関する統計解析

駒 澤 勉

高齢化社会に伴い今日ほど人の健康管理対策が医学的・社会的に重要な時代はない。動脈硬化および高血圧を基盤として発症する脳血管損傷、虚血性心疾患の予知・診断の諸検査項目は現在のところ血圧、心電図、眼底、血清脂質、尿糖・蛋白などの測定があり、病院のみならず集団健康診断で、それらの検査の結果から予後の予測がなされている。しかし、総合的に判断する統一的規準量はない。また、日本人の正常な検査値といわれている範囲の値からも推測しがたい発症者が出ている。正常な血圧値の範囲からも脳梗塞や心筋梗塞が少なからず発症している。

これら発症の現象を解析するため、昭和57年動脈硬化性循環器疾患の集団検診を受けた約3万例の

被検者に対して発症追跡調査を実施し、またそれら被検者の検査データを収集した。それに加えて過去約20年間の延べ約150万例の被検者の検査データを磁気テープに収録して来ている。それらデータも活用して疾病の発症と予後に関する現象解析を行っている。

現象分析の方針：

(1) 事前処理として経年の検査データの質の均一化をはかると共に、可能な限り欠陥のないデータを磁気テープに収録した。

(2) 循環器疾患に最も重要な検査項目である血圧値について日本人の動態を経年データで調べる。

(3) 年齢による差異がこの種の問題では重要な意味を持っているので年齢に関係する検査項目で層別した上、他の検査項目別にみた疾患発症率との関連の分析をし、どんな要因を持つ被検者が疾病発症の危険が大きいかを調べる。

(4) 被検者の個々がどんな検査パターンを示す諸検査項目のもつ意味と関連を明らかにするために、数量化III類のパターン分類により基本的データ構造を調べる。

(5) 諸検査項目を総合して疾病の発症と予後を予測する。まず、多次元的要因パターンからの発症率を解析するため、数量化II類の方法を用いて健常群と発症群、健常群と脳血管疾患群、健常群と心臓血管疾患群の判別・予測分析を行う。この各解析では2群に属する確率の差が極端に相異なる分析を扱うことになるが、たとえ相関比が非常に小さいことが予想されるとも、相関比とミニマックスの中率の関係式から分析が可能であることを示す。

現在のところ基礎分析の段階であるが、年間を通じて血圧値の計測時期は秋期が安定していることや、加齢に従っての血圧上昇率は女性の方が大きく50歳前後で年代別の血圧値の平均値が逆転する事も明白となった。

ベクトル変換を用いた数量化法

馬 場 康 維

統計的データ解析をする際、多次元のデータを視覚化することは解析を進める上で非常に有用である。このような見地から我々は数量化I類の対象とする問題（質的データから目的の変量の値を予測する問題）の視覚化に適する一つの方法を提案した（馬場・脇本，1983）。我々の方法の特徴は数量化I類と同様の予測値を求めることができることに加えて個々の標本のアイテム・カテゴリーに対する反応パターンを表現できるところにある。

目的変量 y を適当な変換により角度に変換する。

$$\phi = G(y) \quad (0 \leq \phi \leq \pi)$$

一方 j アイテム k カテゴリーには角度 θ_{jk} を対応させ、 j アイテムに対する反応をベクトル

$$X_j = (\cos \eta_j, \sin \eta_j)$$

によって表現する。ただし

$$\eta_j = \sum_{k=1}^{K_j} \theta_{jk} \delta(jk)$$

である。ここで K_j は j アイテムのカテゴリー数を表わし $\delta(jk)$ は個体の j アイテムへの反応パターンを表わす変数である。各アイテムに対する反応を表わすベクトル X_j を連結することにより個体の反応パターンをグラフ上に表現できる。

ϕ に対する予測式を

$$\hat{\phi} = \arctan \left(\frac{\sum_{j=1}^R w_j \sin \eta_j}{\sum_{j=1}^R w_j \cos \eta_j} \right)$$

と定義する。ここで R はアイテム数 w_j は重みである。 $\hat{\phi}$ は $w_j X_j$ の合成ベクトルの方向にほかならない。 y の予測値は G の逆変換

$$\hat{y} = G^{-1}(\hat{\phi})$$

によって求める。

問題は w_j, θ_{jk} を求めることである。 $Y = (R \cos \phi, R \sin \phi)$, $\hat{Y} = \sum_j w_j X_j$ とする。以下 N 個の個体の反応パターンと目的変数の実測値が得られているものとし、各変数あるいはベクトルに添字 α を付けたものは個体 α に対応する測定値等を表わすものとする。 w_j, θ_{jk} を決めるため次の2つの基準を考えた。

[A] $\epsilon_\alpha = Y_\alpha - \hat{Y}_\alpha$ とし $Q = \sum_{\alpha=1}^N |\epsilon_\alpha|^2$ を最小にする w_j, θ_{jk} を求める。

[B] $\epsilon_\alpha = \phi_\alpha - \hat{\phi}_\alpha$ とし条件 $\sum_j w_j = R$ のもとで $Q = \sum_{\alpha} \epsilon_\alpha^2$ を最小にする w_j, θ_{jk} を求める。

$w_1 = w_2 = \dots = w_R = 1$ という条件下で線型近似をすると数量化 I 類と同等の解が得られることは昨年度報告した (馬場・脇本, 1983)。本年度は上記の2つの基準で求めた非線型の解と数量化 I 類による解とを比較検討しその一部を統計学会で発表した (馬場・脇本, 1984)。

なおこの研究は岡山大学脇本和昌氏との共同研究である。

参 考 文 献

- [1] 馬場・脇本：統計研彙報第30巻第2号 (1983)
- [2] 馬場・脇本：日本統計学会 (1984)

第 5 研究部

変動する季節成分を持つ時系列の季節調整

赤 池 弘 次

現在世界的に使用されている季節調整法はアメリカセンサス局による X-11 である。X-11 については、失業者統計の時系列の異常な変動 (1975 年における急上昇) の結果、1977 年当時の調整値が異常な動きを示すことが認められ (アメリカの場合)、その修正が重要な問題となった。われわれの開発したベイズ模型による季節調整法 BAYSEA と X-11 との比較研究の結果、X-11 は過調整の傾向があること、これがトレンドの変動を生み出すこと、更に一般に X-11 のトレンド成分は平滑さに欠けるものであることが分った。BAYSEA は X-11 に比し遙かに自然なトレンドの推定値を与えること

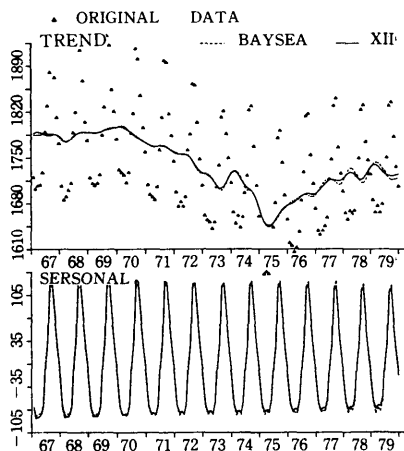


図 1.

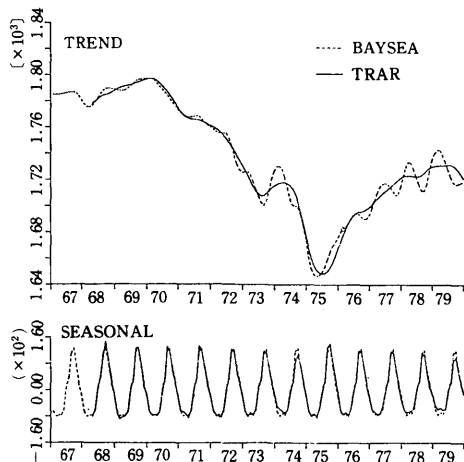


図 2.

が多くの例について確認された。この結果にもかかわらず、いくつかの例では BAYSEA の与えるトレンドもわれわれの通常予期するものとは異った動きを示した (図 1)。

この現象の原因は季節成分の比較的速やかな変化にあり、それが BAYSEA のモデルによっては良く表現されないためであると考えられる。そこで一般に経済時系列の動きがその時に定まった平均的な水準のまわりで安定な変動をくりかえすように見えることに注目して

$$x(n) - t(n) = \sum_{m=1}^M a_m \Delta^d x(n-m) + w(n)$$

というモデルを考え、これを適用してトレンドの動きを追って見た。ここに $t(n)$ はトレンドを示し、 Δ^d は d 階の階差、 $w(n)$ はホワイトノイズである。 $M=13$, $d=2$ とし、 $t(n)$ に対して BAYSEA の場合と類似の smoothness prior を想定した結果、極めて自然な動きを示すトレンドの推定値が得られた (図 2)。 図中 TRAR とあるのがこのモデルで得られた結果である。

電波望遠鏡データにもとづく像の回復

石 黒 真木 夫

I. 電波

光に比べて波長の長い電磁波である。ここで考えているのは波長 1 cm 前後のものである (光の波長は 5×10^{-5} cm 程度)。

II. 電波望遠鏡

原理的には光学望遠鏡と同じように、金属板を鏡とした反射望遠鏡が作れる。ただし望遠鏡の分解能 (分離できる最小の角度) は口径と波長の比できまるので口径 1 m の光学望遠鏡と同じ分解能を得ようとすれば口径 20 km の電波望遠鏡を作ればよいことになる。

III. 電波望遠鏡データ

光学的な反射望遠鏡で焦点面に像が出来る原理をよく見るとわかるようにある面上における電磁波の振動の振幅と位相の情報があればもとの光源 (電波源) がどのようなものであったのか計算できるはずである。

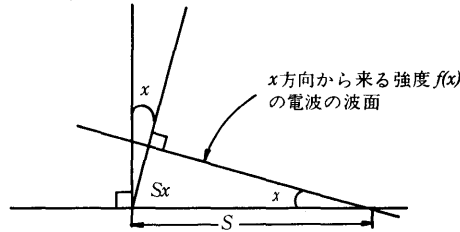
簡単のために (黄道に沿った電波源分布のような) 一次元の単色電波源の強度分布を $f(x)$ とする (x は角度)。このとき (口径 10 m 程度の) 小さなアンテナとこれから距離 S だけ離れたもう一つのアンテナに入ってくる $f(x)$ からの信号の複素相関をとると (x と S の単位が適当にとってあれば)、図 1 を参照して

$$(1) \quad \int \sqrt{f(x)} e^{i(\omega t + Sx)} \times \sqrt{f(x)} e^{-i\omega t} dx = \int f(x) e^{iSx} dx$$

に観測誤差が乗ったものが得られることがわかる。

つまり $f(x)$ のフーリエ変換の S 成分が観測される。 $0 \leq s \leq S$ なるすべての S について観測できれば口径 S の反射望遠鏡で見たのと同様な $f(x)$ の像が回復できることになる。しかし S は 20 km にとれても $0 \leq s \leq S$ なるすべての s について観測することは事実上できない。データとして得られるのは n ケの複素数

図 1



$$(2) \quad y_j = \int f(x) e^{is_j x} dx + \varepsilon_j \quad (j=1, \dots, n)$$

である。

IV. 電波望遠鏡データにもとづく像の回復

ベイイズモデル

$$\begin{cases} f(x) = \sum_{m=1}^M e^{a_m} S_m(x): \text{ 正值スプライン} \\ \varepsilon_j \text{ の実部と虚部} \sim N(0, \sigma^2) \\ a_m - 2a_{m-1} + a_{m-2} \sim O(v^2) \quad (m=1, 2, \dots, M) \end{cases}$$

を利用して $f(x)$ が推定できる。 v^2 は ABIC 最小化によってきめる。

非定常過程の解析

濱 田 義 保

従来、時系列解析の研究は、定常時系列データの解析が中心であった。すなわち、統計的性質が時間変化しない時系列データが与えられた時、次に述べるような解析が行われてきた。

- データの時間変化を説明するためのモデル族を構成する。
- AIC 等の情報量規準を用いて、最適モデルを選択する。
- 選ばれたモデルを、予測、シミュレーション、制御系の設計に用いる。
- パワースペクトルを計算し、周波数領域の情報を求める。

しかしながら、時系列データの中には、その統計的性質が時間変化するものも少なくない。例えば、何らかの経済指標のデータ、地震波、大気の流れの記録、工学的構造物の振動の記録がある。経済データのように、非定常性の原因が政策の変化に起因する場合は、Box・Tiao (1975) により intervention 解析が提案されている。ここでは、このような場合は、取扱わない。地震波、大気の流れの記録は、無限遠の過去から未来までの変化を見ることができれば、定常かつエルゴード的であるかもしれないが、我々が得ることのできるデータの時間スケールでは、非定常性を持っていると言わざるを得ない。また、種々の振動データは、その力学的性質が時間変化するために、非定常性を示すものと思われる。

非定常時系列データが与えられた時、周波数領域における性質の変化を表すために、非定常スペクトルが多く研究者によって導入された (Page (1952), Priestley (1965), Mark (1970))。非定常スペクトルを、ただ一本のデータから、求めるには、統計的性質の時間変化がゆるやかな場合に限られる。また、得られる結果は、あまり満足できるものではない。

ここでは、時間領域でのモデリングを行うことにより、統計的性質がゆるやかに変化する過程を扱うことにする。このような研究は、Ozaki・Tong (1975), Kitagawa・Akaike (1978) によって、既になされている。筆者は、短いスパンごとに局所的に、AR 過程を考え、各スパンの AR 係数の間に、統計的性質の時間変化が、ゆるやかであるという束縛を加えたモデルを考えた。モデル選択には、情報量規準 ABIC を用いる。詳しい報告は、近く発表する予定である。

参 考 文 献

- Box・Tiao (1975). *J. Amer. Statist. Ass.*, **70**, p 70.
 Kitagawa・Akaike (1978). *Ann. Inst. Statist. Math.*, **30**, p 351.
 Mark (1970). *J. Sound. Vib.*, **11**, p 19.
 Ozaki・Tong (1975). *Proc. 8th Hawaii Inc. Conf. on Sys. Sci.*
 Page (1952). *J. Appl. Phys.*, **23**, p 103.
 Priestley (1965). *J.R. Statist. Soc.*, **B 27**, p 204.

非ガウス確率過程のシミュレーションモデル

尾 崎 統

確率微分方程式

$$(1) \quad dx = f(x)dt + g(x)dW(t)$$

のシミュレーションモデルとして、伊藤の確率積分を応用した次のモデル

$$(2) \quad x_{t+\Delta t} = x_t + f(x_t)\Delta t + g(x_t)\Delta W_{t+\Delta t}$$

が世の中にひろく用いられている。ここに $\Delta W_{t+\Delta t}$ は平均ゼロ、分散 Δt の離散時間白色ノイズである。このモデルが推奨される唯一の理由は他の諸々のアドホックな方法と違って $\Delta t \rightarrow 0$ の時、区間 $[t, T]$ 上で、 x_t がもとの確率過程 $x(t)$ に収束することが保証されているからに他ならない。このモデル(2)を非線型時系列モデル、あるいは連続空間上のマルコフチェインとみなす時、次の二つの問題がある。第一に、 $f(x)$ が非線型な場合は、 Δt をどんなに小さく固定しても、このマルコフチェインは確実に発散し、従ってもとの $x(t)$ が定常なマルコフ過程であるにもかかわらず非定常なマルコフチェインを定義する。これは解析的方法が困難な対象をモデルの数値実験で確かめようとする場合に、はなはだ困ることである。第二の問題はモデル(2)の一致性が保証されていることは、保証されていないよりはましだとは言え、有限の Δt に関してどれくらい良い近似かに関して何の手がかりもない点である。

以上の二つの問題点に関して以下の接近法が考えられる。第一の問題に関しては、まず定常な確率過程に対して変数変換を行いその付随力学系表現を求める。次にその決定論的力学系の部分に対して差分近似を行う。その際、その力学系が区間 $[t, t+\Delta t)$ 上で線型とみなすことによって得られる局所的解析解を各小区間上で用いれば、発散しない安定な差分モデルが得られる。第二の問題に関して、一つの考えは、その確率過程が、区間 $[t, t+\Delta t)$ 上で近似的にガウス過程とみなし、各小区間上でその近似ガウス過程に合わせるモデルを考えることである。その為には各小区間、 $[t, t+\Delta t)$ 上で定義される自己共分散関数 $\Gamma(x_t, \tau)$ に $\tau=0, \Delta t, 2\Delta t, \dots$ で一致する自己共分散関数を持つ局所的にガウスな離散時間モデルを用いることである。これを可能にするホワイトノイズの分散表現(x_t -従属：しかし Δt のオーダーで近似すると x_t -独立になる。)が求まる。以上の二つの考えが具体化される為には、確率過程の変数変換によりダイナミクスに関する全ての情報を力学系表現の中に取り入れることができると言う点が基本的である。

非線形最適化のアルゴリズム

北 川 源 四 郎

統計的モデル構成において、しばしば非線形関数の最小化あるいは最大化が必要となる。準ニュートン法は各点 x における関数値 $f(x)$ と微分 $g(x)$ からヘシアン行列の逆行列を逐次推定しながら関数の最適化を行なうもので、ニュートン法の持ついくつかの欠点を補っており、現在のところ最も発達した方法と考えられる。ただし、この方法は二次曲面の最小化を想定して $H_{n+1}y_i = \rho s_i$ を満たす

ように行列 H_n を更新するもので、一般の非線形関数に対する性質は十分には解明されていない。

ここでは $g_n = g(x_n)$, $S_n = x_{n+1} - x_n$, $y_n = g_{n+1} - g_n$ とし、ヘシアン の逆行列を H と書くことにする。このとき、 $X_n = x_n - H(x_n)g_n$ によって状態ベクトルを定義すると

$$\begin{aligned} X_{n+1} &= X_n + S_n v_n \\ z_n &= y_n^t X_n + w_n \end{aligned} \quad \begin{pmatrix} v_n \\ w_n \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q_n & S_n \\ S_n & R_n \end{pmatrix} \right)$$

$z_n = y_n^t x_n$, というモデルが考えられる。ここで、 v_n と w_n は厳密な二次曲面の場合には 0 となるが一般の関数では x_n に依存した値をとるので便宜的に雑音項とみなすことにする。このシステムの予測式は

$$\begin{aligned} \hat{X}_{n+1|n} &= \hat{X}_{n|n-1} + (V_{n|n-1} y_n + S_n S_n) (y_n^t V_{n|n-1} y_n + R_n)^{-1} (z_n - y_n^t \hat{X}_{n|n-1}) \\ V_{n+1|n} &= V_{n|n-1} - \frac{(V_{n|n-1} y_n + S_n S_n) (V_{n|n-1} y_n + S_n S_n)^t}{y_n^t V_{n|n-1} y_n + R_n} \end{aligned}$$

で与えられる。ここで、正規分布においてはヘシアン の逆行列と共分散行列が一致することから、 $V_{n|n-1}$ を H_n の推定値とみなすことにする。 Q_n , R_n , S_n の定め方によって、様々な最適化モデルと H_n の更新アルゴリズムが得られる。とくに

$$S_n = Q_n s_n^t y_n - 1, \quad R_n = S_n s_n^t y_n$$

という関係が成りたつとき、上記の更新公式はセカント条件をみたし、 $Q_n = 0$, $(s_n^t y_n)^{-1}$, ∞ とおくと、それぞれランク 1 公式、DFP 公式、BFGS 公式が得られる。最適化モデルの共分散構造の違いから、これらの公式の性質に対する一つの説明が得られる。また、 Q_n , R_n , S_n の様々な選び方によって、いろいろな新しい公式が導かれる。これらの公式のうち、いくつかは数値実験において極めて良い結果を示している。

第 6 研究部

海外学術調査実施の概要

鈴木 達 三

「比較文化研究における連鎖的調査手法の確立とその展開」(研究代表者 林 知己夫, 他共同研究者 8 名) が実施した海外調査 (科学研究費補助金) の概要をのべる。

<研究経過>

連鎖的調査手法および動態的調査手法を効果的に適用できるハワイ・ホノルル市において、市域在住の一般市民 (選挙人名簿登録者) を対象とする意識調査を面接聴取法により実施した。具体的な経過は次の通り。

58 年 4 月: 調査項目の選択, 原案決定, 英訳調査票原案作成。

58 年 5 月~6 月: ハワイ・ホノルル市において、プリテスト実施。プリテストの結果により本調査の調査票作成。調査項目は連鎖的調査手法の有効性を確認するため、日本、フランス、アメリカ本土調査との関連比較を考え選択し、同時に、ハワイにおける調査項目も継続し、経年変化の実態をとらえるようにした。調査項目は、意識実態項目 92 項目 (このうち「第 7 次国民性調査」の項目 50 項目)、この他、日系人を対象とする調査項目として 24 項目を追加した。

58 年 6 月: 選挙人登録名簿 (磁気テープ 2 巻) より単純無作為に標本抽出。調査対象地域は 71 年、78 年両度の調査と同一である。

58 年 6 月~10 月: 本調査実施。調査完了標本数 810 うち有効回答者数 807。調査実施上、ハワイ側共同研究者の多大な協力があつた。

58 年 11 月以降: 調査実施状況の点検確認。調査内容の点検確認をおこない、引続き調査票および関連資料の集積と整理を行ない、集計・分析作業中である。

<分析の方針>

連鎖的比較調査の具体例として、

- ① 同一時点におけるハワイと日本との比較、
- ② 日本；ハワイおよびフランスとの比較（21項目）、
- ③ ハワイとアメリカ本土の比較（30項目）が可能である。

さらに、動態的比較調査として、

- ④ 78年ハワイ調査との共通項目（48項目）、
- ⑤ ハワイ日系人について、71年、78年調査と今回の3回にわたる経年比較（35項目）、

ができることになる。

都市住民の環境意識の分析

大 隅 昇

近年、都市環境評価の一環として、都市住民の環境意識と関連させた生活環境指標の作成や住み易さ感の分析などが環境施策の上で重要な課題となっている。「都市住民の環境意識の計量にもとづく動的環境評価システムに関する研究」（環境庁公害防止等試験研究費）の目的も都市住民の地域環境に対する認知の程度、満足感、期待・願望などを意識調査を通じて測定し意識と環境の関連づけを図る実用的な方策を提案することにある。

研究内容は、まず環境意識の適切な測定・分析法を検討しその検証のための調査を実施することである。同時に、得られた意識調査データを用いて、そこに内在する環境と意識の関連性をカラーグラフィックスを活用して、調査対象地域全体の特徴として大まかに比較・観察する“領域クラスタリング手法”の開発を進めてきた。本年度はこの手法を実際の意識調査データ（千葉市市街部、千里ニュータウンの2地域で実施）に適用し、そこでみられた問題点を検討するとともに関連手法プログラムの改良を図った。

領域クラスタリングにより意識データの特徴を効果的に視覚化するという当面の目標からすると、個々の被調査者の意識を知るといよりも調査対象地域や地点に共通にみられる環境特性への関心の程度に興味がある。また多数の質問項目を用いた色表示はいたずらにイメージを複雑にするだけで、つまり意識データの種々の位相が混在して、領域クラスタリングの利点が生かされない。こうした点を考慮して次の事項の検討を進めながら両調査地域の住民の環境意識の比較分析を行った。

(1) まず質問項目を環境意識の種々の位相を代表するように分類要約する方式の検討が必要である。これは同時に、具体的な環境の諸要素（「緑が多いと感ずるか」、「自動車や電車の音がうるさい」など）、漠然とした総合的な住み易さ感（「住んでいるまちが気に入っているか」、「住んでいる地区は住みやすいか」など）、期待する住みよいまちのイメージ感（「住むまちについてどう考えるか」、「希望する住みよいまちとはどのような感じのまちか」など）各種位相を想定設計した質問項目の妥当性の検証でもある。；(2) 次に、分類した項目群を用いてそれぞれの特徴（緑の量、利便性、安全性など）を解釈が容易な形に尺度化・合成指標化する方式を検討する。これは各項目群にみられる地域差、地点差の有無や類似・差異を強調する操作でもある。；(3) そして(2)で統合した加工情報を種々の位相にあわせて、調査対象領域上の位置を持つ多変量特性値として集約化する方法とそのプログラムを作成する。このデータに領域クラスタリングを適用して“環境意識の色地図”を作成しこれらをもとに両調査地域にみられる質問項目群の特性や地点間の特徴の視察を進めている。

意識に関する調査および分析の問題

水 野 欽 司

本年度は、「第7回・日本人の国民性調査」（特別研究）と「都市環境の住みやすさに関する調査」（国立機関公害防止等試験研究費による）の二つの調査実施に従事した。

一般に人間の意識の調査測定では、対象サンプルの不協力をどう減らすか、地域事情に応じて調査員をどう起用するか、…など、実行面における多くの対策が重要である。広い意味の調査方法は、理論を基盤におくサンプリングのように合理的かつ整一的な対応がある程度可能な側面だけではない。調査実施にかかわる具体策のように、合理的対応のむずかしい非決定的な側面を多くもっている。調査の結果として得られる情報の良否は、‘意識’の場合、むしろ後者にその多くを負っているといつてよい。にもかかわらず、後者の改善に役立つ事前情報は常に不足がちであるというのが実状である。調査実施における、この種の側面は、時代の変化・社会の様相によって影響されやすく、対応を誤れば、いったん定着し確立されたかに見える調査方式や分析方法も、その有効性を失うことになりかねない。

今回は、意識調査の実施困難が増しているといわれる状況の下で、この種の問題に関係する最近の実態を、調査実施の各局面について、具体的に知ることができた。増大傾向にある多くの調査困難に対処するためには、従来の調査方式全体を抜本的に見直す必要があると考えられるが、詳細な検討は今後に残される。

コウホート分析の適用例

中 村 隆

コウホート分析とは、継続的な調査で得られる何らかの数量特性を年齢層×調査時点別に集計したデータから、年齢効果・時代効果・世代（コウホート）効果を分離しようとする方法である。ただし、コウホート分析に固有な識別問題のために3効果を分離することが困難であった。この困難を克服するために、パラメータの漸進的変化の条件を取り込んだベイズ型コウホート・モデルを開発し、種々のデータへの適用を行なっている。

〈齲蝕率のコウホート分析〉

「歯科疾患実態調査（厚生省）」による齲蝕率の推移を分析した。このデータは、昭和32年から昭和56年まで6年おきにとられ、年齢区分は5歳刻みである。このような調査間隔と年齢区分幅が一致しないデータの場合も分析することが可能である。

齲蝕率の推移は、年齢効果が支配的であるが、漸増する時代効果と、第2次大戦と戦後の高度成長が影響したとみられる世代ごとの違い（コウホート効果）が特徴である。

年齢効果は低年齢層での変化が相対的に大きいため、全年齢層にわたって均一の縛り方をすると結果全体が歪んでしまう。ABICの値をみながら、注意深く分析をすすめる必要がある。

〈主食パタンのコウホート分析〉

人口問題研究所の行なった回顧調査に基づき、朝昼夕食がそれぞれRRR, RNR, BRRという主食パタンの変化を、6地方都市について分析した（ここで、Rは米飯、Nは麺類、Bはパンを表す）。継続的な調査ではなく、一時点の回顧調査でも、コウホート分析を行なうことができる。

いずれの主食パタンの変化も、時代効果が顕著であり、コウホート効果もかなりみられる。これらに対し、年齢効果はほとんどみられない。以上の傾向は、6つの地方都市についてはほぼ同様であった。

〈その他のデータ〉

未婚婚率や死亡率など、人口統計の分野ではコウホート・データがそろっているが、3効果を分離するというコウホート分析はあまり行なわれていない。死因別の死亡率、有病率など疫学上のデータについても分析をすすめている。

附属統計技術員養成所

ベイズ型2次元2値回帰モデルとその応用

坂元慶行

ある事象の生起確率が1次元の説明変数に依存して滑かに変化する場合には、その生起確率を推定するためのベイズ手法を(Ishiguro・Sakamoto, 1983)で提案した。今年度は、この手法を説明変数が2次元の場合に拡張するとともに、その実データへの応用を行なった(石黒との共同研究による)。

2次元平面の任意の点 (x, y) における生起確率 $p(x, y)$ について、

$$(1) \quad p(x, y) = \frac{\exp\left\{\sum_{j=1}^J \sum_{k=1}^K q_{jk} S_j(x) S_k(y)\right\}}{1 + \exp\left\{\sum_{j=1}^J \sum_{k=1}^K q_{jk} S_j(x) S_k(y)\right\}}$$

なるモデルを想定する。ここで、 q_{jk} は (x, y) 平面を $J \times K$ 個のメッシュに分割したときの第 (j, k) 成分での2次直交スプライン関数 $S_j(x) S_k(y)$ の係数である。

いま、 $\bar{q} = (q_{11}, \dots, q_{1K}, q_{21}, \dots, q_{JK})^t$ とおき、 $p(x, y)$ を改めて $p(x, y | \bar{q})$ と書くと、データ (h_i, x_i, y_i) , $h_i = 1, 0, i = 1, \dots, n$ が与えられたときのパラメータ \bar{q} に関するモデル(1)の尤度は、

$$L(\bar{q}) = \prod_{i=1}^n p(x_i, y_i | \bar{q})^{h_i} \{1 - p(x_i, y_i | \bar{q})\}^{1-h_i}$$

となる。

\bar{q} の推定値の安定性を保障するために、 \bar{q} の成分 q_{jk} が局所的に1次式で近似することができるという想定の下に、 \bar{q} の主要部分 $\mathbf{q} = (q_{13}, q_{14}, \dots, q_{1K}, q_{22}, \dots, q_{2K}, q_{31}, \dots, q_{JK})^t$ に対して事前分布

$$\pi(\mathbf{q} | q_{11}, q_{12}, q_{21}, w^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^{JK-3} (\det D^t D / w^2)^{-1/2} \exp\left\{-\frac{1}{2w^2} |D\mathbf{q} + \mathbf{r}|^2\right\}$$

を仮定すれば、ABIC最小化法を介して、最終的には \bar{q} の推定値を求めることができる。ただし、 D と \mathbf{r} は次式を満すものとする。

$$|D\mathbf{q} + \mathbf{r}|^2 = \sum_{k=3}^K (q_{1k} - 2q_{1k-1} + q_{1k-2})^2 + \sum_{j=3}^J (q_{j1} - 2q_{j-1,1} + q_{j-2,1})^2 \\ + \sum_{j=2}^J \sum_{k=2}^K (q_{jk} - q_{j-1,k} - q_{j,k-1} + q_{j-1,k-1})^2.$$

手法の効率をシミュレーションによって検討した後、地図上における降雪確率の推定、世論調査における意見の支持率の推定等、実データへの適用を試みた。

[1] M. Ishiguro and Y. Sakamoto. (1983). *Ann. Inst. Statist. Math.*, 35-1.