

統計数理研究所研究活動

1982年度研究発表会要旨

と き：1983年3月23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

確率分布の特徴づけとその安定性

清水 良一

1. 単調な hazard rate をもつ分布に関する不等式

正の値をとる確率変数 X につき， $\bar{F}(x) = \Pr\{X > x\}$ とおき， $f(x)$ をその p. d. f. とする．hazard rate $q(x) = f(x)/\bar{F}(x)$ が単調に増加（または単調に減少）することを仮定して，この分布と指数分布との近さを論じた．この分布のモーメントを $\mu_k = E(X^k)$ ， $k=1, 2, \dots$ ， $\mu = \mu_1$ ，とする．次の不等式関係はよく知られている（以下，増加の場合を述べる）．

$$(1) f(0) \leq 1/\mu,$$

$$(2) (k+1)\mu_k^2 \geq k\mu_{k+1} \cdot \mu_{k-2},$$

とくに，

$$(3) 1 \geq \frac{\mu_2}{2\mu^2} \geq \frac{\mu_3}{3!\mu^3} \geq \dots,$$

また $q = \lim_{x \rightarrow \infty} f(x)/\bar{F}(x)$ とおくと，

$$(4) q\mu \geq 1.$$

指数分布 $\bar{F}(x) = e^{-x/\mu}$ の場合を除けば，これらの不等式はいずれも strict，いいかえれば，上記不等号のどれか一つが等号になるなら，この分布は指数分布であり，他のすべての不等号も等号で置きかえられる．

問題は，等式が近似的になりたつとき，この分布が指数分布にどれだけ近いか，を明らかにすることである．次の諸不等式が，それに対する一つの解答を与える． $\varepsilon = 1 - \mu_2/2\mu^2 (\geq 0)$ とおくと

$$(5) |\bar{F}(x) - e^{-x/\mu}| \leq 1 - \mu f(0).$$

$$(6) -\varepsilon \leq \bar{F}(x) - e^{-x/\mu} \leq (1 - \varepsilon e)^{-1} \varepsilon.$$

$$(7) \int_0^\infty |\bar{F}(x) - e^{-x/\mu}| dx \leq 2\mu\varepsilon.$$

$$(8) e^{-qx} \leq \bar{F}(x) \leq q\mu e^{-x/\mu}$$

例えば，ガンマ分布 $f(x) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x} / \Gamma(\alpha)$ では， $\mu = \alpha/\lambda$ ， $\mu_2 = \alpha(\alpha+1)/\lambda^2$ ， $q = \lambda$ ．

したがって $\varepsilon = (\alpha-1)/2\alpha$ となるから，

$$(9) -\frac{\alpha-1}{2\alpha} \leq \bar{F}(x) - e^{-\lambda x/\alpha} \leq \frac{\alpha-1}{2\alpha - (\alpha-1)e},$$

$$(10) \int_0^\infty |\bar{F}(x) - e^{-\lambda x/\alpha}| dx \leq \frac{\alpha-1}{\lambda},$$

$$(11) e^{-\lambda x} \leq \bar{F}(x) \leq \alpha e^{-\lambda x/\alpha}.$$

2. Damage model におけるポアソン分布の特徴づけ定理の安定性

(要旨 省略)

分布の一樣近似に関する 修正情報量基準

松 縄 規

要旨

確率分布間の一樣距離 D に基づく近似を扱う際に， $K-L$ 情報量を modify した量を用いることにより，かなり広範囲の問題に対して，従来の関連する結果よりも良い近似が可能である事を得た．

記号

X, Y ; 抽象可測空間 (R, B) 上で定義された確率変数，

P^X (or P^Y); X (or Y) の確率分布，

μ ; (R, B) 上の σ -有限測度，

A ($\in B$); 近似主領域，

f (or g); P^X (or P^Y) の測度 μ に関する gpdf.

X, Y の隔りの測度として次の諸量を考える：

$$D(X, Y; B) = \sup_{E \in B} |P^X(E) - P^Y(E)|,$$

$$V(X, Y; A) = \int_A |f - g| d\mu,$$

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

$$I_a(X, Y; A) = \int_A f |\ln f / g| \, d\mu.$$

また,

$$\inf_A \sqrt{g/f} = \alpha, \quad \inf_{A^c} |1 - g/f| = \beta$$

$$\inf_{A^c} |1 - f/g| = \gamma,$$

$$A^c = R - A,$$

$$A^+ = A \cap \{x; f(x) \geq g(x), x \in R\},$$

$$A^- = A - A^+$$

と表わす。次の諸結果を得る:

定理 1

- (1) $D(X, Y; B) \leq I(X, Y; A^+) + P^X(A^c)$
 $\leq I_a(X, Y; A) + P^X(A^c)$
- (2) $D(X, Y; B) \geq \max[\alpha \cdot I_a(X, Y; A),$
 $-I(X, Y; A)] + \max$
 $[\beta \cdot P^X(A^c), \gamma \cdot P^Y(A^c)] (\geq 0).$

注

(i) 不等式 (1) で $I(X, Y; A^+)$ [or $I_a(X, Y; A)$] が $D(X, Y; B)$ と同じ Order であることが、種々の近似問題の誤差評価で役立つ。

(ii) 近似主領域としては、例えば $A = \{x; -\tau < \ln f(x)/g(x) < \kappa; \tau, \kappa > 0, x \in R\}$ などを考えることになる。

漸近理論への応用

定理 1 の不等式を用いて、以下の分布の漸近一様近似に対する基準を与え得る:

$\{X_n\}, \{Y_n\} (n=1, 2, \dots)$ を (R_n, B_n)

$(n=1, 2, \dots)$ 上で定義された確率変数列とする。

前述の関連諸量にも添字 n を施すものと約束する。次の結果を得る:

定理 2 $D(X_n, Y_n; B_n) \rightarrow 0, (n \rightarrow \infty)$

である為の必要十分条件は

$$(3) \quad \exists A_n \in B_n: P^{X_n}(A_n) \rightarrow 1, (n \rightarrow \infty)$$

かつ

$$(4) \quad I_a(X_n, Y_n; A_n) \rightarrow 0, (n \rightarrow \infty)$$

注

定理 2 は Matsunawa (1982), *Ann. Inst. Statist. Math.*, **34**, 1-17, 定理 3.1 の相補的な結果である。

都府県の大都市卸売依存率による地図化

川 合 伸 幸

1つのケース・スタディとして、各都府県の、東京・愛知・大阪の3大都市への卸売仕入れの数表を地図として表現し、卸売仕入れから見た各都府県と大都市との結びつきの構造を大局的に把握することを考える。依存率表は次の様に構成されている。

TOKYO AICHI OSAKA

i 県 $\dots e_{i1} \quad e_{i2} \quad e_{i3}$

ここで、 $e_{ij}; j=1, \dots, 3$ は各都市に対する卸売仕入れの全体の仕入れに対する百分率 (%) を表わす。

2次元平面上の地図を考え、各都市に対して与える座標を $x_{0j}; j=1, \dots, 3$ 各県に対して与える座標を $x_i; i=1, \dots, 46$ とする。その際、次の性質

- 1) $e_{ij}; j=1, 2, 3$ の大小関係が地図上からも読みとれること
- 2) 依存率合計の大きなものと小さなものとが区別できることが保障されていることが望ましい。

1つの方法として

$$(1) \quad L = \frac{\sum_i \sum_j e_{ij} \|x_i - x_{0j}\|^2}{\sum_i \sum_j e_{ij} \|x_i\|^2} \rightarrow \text{小}$$

により x_i, x_{0j} を決めることが考えられる。 L は、分散分析との類似で、 x_{0j} のまわりに x_i を分散 σ^2/e_{ij} に従ってばらつかせた時の級内分散と全分散の比になっている。しかし、この解は、

$$x_i = \lambda \sum_{j=1}^3 e_{ij} / w_i x_{0j}$$

という形をもつ。ここに、 λ はある定数、 $w_i = \sum_j e_{ij}$ である。

従って、相対的依存率 e_{ij}/w_i が同一な県は同じ点に配置される。これは 2) の観点から見てまずい。し

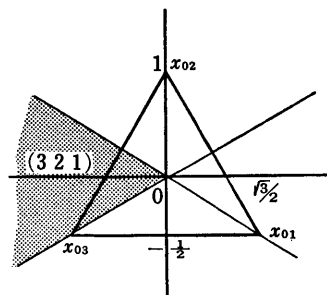


図 1

かしこの形で与えられる x_i は、三角形 $x_{01} x_{02} x_{03}$ を正三角形とすると、1) の性質を満たすことが示される。例えば、 $e_{i3} > e_{i2} > e_{i1}$ とすると x_i は図1の(321)領域に配され、逆に領域から e_{ij} の順序関係がわかる。従って、

$$x_i = \beta_i \sum_j e_{ij} / w_i x_{0j}$$

($x_{01} x_{02} x_{03}$ は図1のように与える)

とすると、これは性質1)を満たす。その上で、2)を満たすよう、 x_i と x_{0j} の距離が $d_{ij} = 1/e_{ij}$ に近くなるよう、相対誤差2乗和、

$$(2) \quad \sum_{i,j} (1 - k \|x_i - x_{0j}\|^2 / d_{ij}^2)^2 \rightarrow \text{小}$$

となるよう β_i を決める。ここで、 k は x_{0j} のスケールを d_{ij} のスケールに合わせるためのパラメータである。

$k^* = (\sum \hat{d}_{ij} / d_{ij}) / (\sum \hat{d}_{ij}^2 / d_{ij}^2)$, $\hat{d}_{ij} = \|x_i - x_{0j}\|$ で与えられ、結局

$$L = - (\sum \hat{d}_{ij} / d_{ij})^2 / (\sum \hat{d}_{ij}^2 / d_{ij}^2) \rightarrow \text{小}$$

となるよう β_i を決めることになる。なお、図1で x_{0j} をスカラー倍伸縮させても1)は保たれ、正三角形を伸縮させながら(2)を最小にするようにも考えられる。また、大都市の数が4以上になっても同様の考え方で方法を拡張することができる。

検定推定量について

平野 勝 臣

検定の結果に従って推定量を変えて使うということは、我々の直観によくアピールする。しかし、決定理論の枠組で考えて多くの検定推定量は、普通使われている損失関数の下で、非許容的であることが示されている。検定の結果が棄却か採択かの二者択一の選択のため、その規則に滑らかさがないことによる。規則に滑らかさを入れて考えればモデルを区別するという側面がうすれ、通例の推定理論の枠組の中で考察されるからである。

しかし、検定推定量を別の側面からみれば、これはモデルの選択を含む推定方式であるので、通例使われている損失関数で考察することになじめないところがある。複雑なモデルを選ぶ場合、単純なモデルよりも高いコストをかける損失関数を導入して考察することが望まれる。例えば、 $N(\mu, 1)$ の μ の検定推定量を標本平均の絶対値がある正の値 c より小さいとき0、そうでないとき標本平均とすると、これは許容的になることが示されている。この様な考え方はモデルを選ぶことと、その選ばれたモデルの母数を推定すること

を明確に区別して考えるわけで、この下で検定推定量の許容性を考察することが望まれる。

判定に使われる統計量と推定に使われるそれが同一とする積極的理由はない。例えば上の検定推定量で、平均値と中央値を組みあわせて作られる4つの検定推定量について調べてみる。ガウス分布をずらしてこれらの推定量のリスクを計算することは解析的に容易でない。標本数10、 $c = 1/\sqrt{5}$ (標本数に依存しこの決めかたは既報)で、ガウス、ラプラス、ロジスティック、コーシー分布の夫々からの乱数で4つの検定推定量の分散をシミュレーションで比較すると、規則に中央値、推定に中央値と、規則に中央値、推定に標本平均の2つの検定推定量はさほど悪くならない。初めの検定推定量は異常値の影響を受けやすいことを考えれば、この範囲の考察の中で特に後者は優れた検定推定量となっていることがわかる。

デザイン行列の選択に関する情報論的考察

久保木 久 孝

$$Y = X\beta + \varepsilon$$

なるモデルを考える。ここで、

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}, \text{rank}(X) = k,$$

$$Y = (Y_1, \dots, Y_n)', \beta = (\beta_1, \dots, \beta_k)' \\ \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$$

そして、確率ベクトル ε は、

$$E(\varepsilon) = 0$$

$$V(\varepsilon) = \Sigma: \text{正値対称行列}$$

であるとする。我々が、ここで問題とするのは、何らかの意味で良い実験を構成すること、すなわち、良いデザイン行列 X の選択である。

デザイン行列を選択するための規準はいろいろあるが、ここでは、情報の概念に基づいた新しい規準を導入する。今 β の最小二乗推定量を

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)'$$

とする。この統計量は損失関数

$$W(\beta, \hat{\beta}) = \sum_{i=1}^k c_i^2 (\hat{\beta}_i - \beta_i)^2,$$

$$c_i \neq 0, i = 1, \dots, k$$

に関して不偏である。そこで、以前に筆者が導入した discrimination rate を考える。それを $D(\beta; \hat{\beta}, W|h)$ とする。また、 Y の分布のもつ感度を $s(\beta|h)$ とす

る。これらは、それぞれ、パラメータ β の h 方向 ($h \in R^k$, $\|h\|=1$) への変動を検出するための、 $\hat{\beta}$ 及び Y の有する情報を評価した量である。この両者の間には、

$$D(\beta; \hat{\beta}, W|h) \leq s(\beta|h)$$

なる関係がある。我々は、 $\hat{\beta}$ を使って様々な推論を行なうわけであるから、 $\hat{\beta}$ が生のデータ Y に近い情報を持っているのが望ましい。そこで、我々は

$$e(\beta; \hat{\beta}, W|h) = D(\beta; \hat{\beta}, W|h) / s(\beta|h)$$

をできるだけ1に近づけるような X を選択することにする。すなわちここで導入する選択規準は

$$\max_X e(\beta; \hat{\beta}, W|h)$$

である。

定理 1. $\varepsilon \sim N(0, \Sigma)$ とする。

$$e(\beta; \hat{\beta}, W|h) = 1 \text{ for } \forall h$$

を達成するデザイン行列は

$$(*) \quad X' \Sigma^{-1} X = \begin{pmatrix} C_1^2 & & 0 \\ & \ddots & \\ 0 & & C_k^2 \end{pmatrix}$$

である。

定理 2. ε の分布にある正則条件を仮定する。(*) をみたま X で

$$e(\beta; \hat{\beta}, W|h) = 1 \text{ for } \forall h$$

となるのは、 $\varepsilon \sim N(0, \Sigma)$ の場合に限る。

(*) をみたまデザイン行列は、次の性質を持つことが知られている。

定理 (C.R. Rao) $X_i' \Sigma^{-1} X_i = c_i^2$, $i=1, \dots, k$ という制約の下で、各 $V(\beta_i)$ を最小にする X は、(*) をみたま。ここで、 X_i は X の第 i 列ベクトル。

対称な分布の位置母数のロバスト推定

安芸重雄

F_0 を 0 に関して対称な分布の分布関数とする。 M を $\{F(x): F(x) = F_0(x-\theta), -\infty < \theta < \infty\}$ なる分布関数の集合とする。 \mathcal{F} を M を含む或る分布関数の集合とする。 S を \mathcal{F} 上の location functional とする。即ち \mathcal{F} 上で定義された実数値汎関数で、対称な分布に対しては対称の中心を値にとり、 \mathcal{F} 上で shift invariant なものである。 ϕ を $[0, \infty)$ で定義された非減少な実数値関数で $\phi(0)=0$ を満たすものとする。このとき、 $G \in \mathcal{F}$ に対し、

$$T[S, \phi; G] = \left(\int_{-\infty}^{\infty} 1/(1+\phi((x-S[G])^2)) \right)$$

$$dG(x))^{-1} \int_{-\infty}^{\infty} x/(1+\phi((x-S[G])^2)) dG(x)$$

なる値をとる汎関数を考えると、 $T[S, \phi; \cdot]$ は対称な分布の集合上で、location functional となる。

$X_1, X_2, \dots, X_n, \dots$ を分布 F に従う独立確率変数列とし、 X_1, \dots, X_n から得られる経験分布関数を F_n と書く。 F の中心の推定を $T[S, \phi; F_n]$ によって行なう時、推定量 $T[S, \phi; F_n]$ は次のような漸近的な性質を持つ。

定理 或る正則条件 (省略) の下で、 $n \rightarrow \infty$ のとき、

$$\sqrt{n} (T[S, \phi; F_n] - T[S, \phi; F])$$

は平均 0、分散 v_T の正規分布へ法則収束する。

ここで、

$$v_T = \frac{1}{C^2} \int_{-\infty}^{\infty} \left(K \cdot IC_{F_0}[S; x] + \frac{x}{1+\phi(x^2)} \right)^2 dF_0(x),$$

$$C = \int_{-\infty}^{\infty} \frac{1}{1+\phi(x^2)} dF_0(x),$$

$$K = \int_{-\infty}^{\infty} \frac{2x^2 \phi'(x^2)}{(1+\phi(x^2))^2} dF_0(x),$$

$IC_{F_0}[S; x]$ は S の F_0 における influence curve である。

$T[S, \phi; F_n]$ がロバストである為には、 $|1/(1+\phi(x^2))|$ が有界で、かつ、 S の influence curve が有界であることが望ましい。実際にここで、 S を median にとり、 $\phi(t)=t$ と置いた推定量は、広い範囲で (漸近的な意味で) ロバストである。

また、上の形の推定量の小標本の場合の性質を調べるために次のような Monte Carlo 実験を行った。 S を median にとり、 ϕ としては、 $\phi(t)=t$, $\phi(t)=t^2$, $\phi(t)=t^3$, $\phi(t)=\sqrt{t}$ をとる。標本数が 10, 20, 30 の場合に、 F_0 =Gaussian, logistic, Cauchy, Laplace 分布に対して、それぞれ 200,000 回ずつ推定値を求め、推定値の分散を比較した。その結果、小標本の場合にも、これらの推定量はロバストであり、特に $\phi(t)=t$ と置いた推定量は、実験したすべての場合に、median よりも分散が小さいことがわかった。

統計的決定問題における情報価値の比較

村上征勝

状態数 2、行動数 2 の統計的決定問題の一つの例と

して、次の様な機械の取替に関する問題を考える。状態 θ_1 を正常、 θ_2 を故障、行動 a_1 を現状維持、 a_2 を取り替えとし、状態が θ の時に行動 a を取ることによって生じる損失を $L(\theta, a)$ とする (表 1)。今、この機械の状態を監視するモニターとして、 X, Y の 2 台のモニターがあり、モニター X は機械の状態を x_1, \dots, x_m の m 段階で、モニター Y は y_1, \dots, y_n の n 段階で表示でき、その確率は $P(x|\theta), P(y|\theta)$ で与えられる。 w_1, w_2 をそれぞれ θ_1, θ_2 の事前確率とするとモニター X とモニター Y をそれぞれ使って決定を行った時のベイズ期待危険度 $B_X(w_1, w_2), B_Y(w_1, w_2)$ は次のようになる。

$$B_X(w_1, w_2)$$

$$= \min_{d \in D_X} \left\{ \sum_{i=1}^2 w_i \sum_{j=1}^m L(\theta_i, d(x_j)) P(x_j|\theta_i) \right\}$$

$$B_Y(w_1, w_2)$$

$$= \min_{d \in D_Y} \left\{ \sum_{i=1}^2 w_i \sum_{j=1}^n L(\theta_i, d(y_j)) P(y_j|\theta_i) \right\}$$

ここで d は決定方式、 D_X, D_Y はモニター X , モニター Y を用いた時の非確率化決定方式の集合である。

w_1, w_2 が既知ならその時には $B_X(w_1, w_2), B_Y(w_1, w_2)$ の小さい方が情報の価値が高いというように X, Y の比較が可能となるが、 w_1, w_2 が未知の時

は一般には比較できない。しかしある場合 (例えば $P(x|\theta), P(y|\theta)$ が表 2, 表 3 のような場合) には、図 1 (この時の $L(\theta, a)$ は表 4) のようにすべての w_1, w_2 に対して $B_X(w_1, w_2) \leq B_Y(w_1, w_2)$ となる。この時 X の方が情報の価値が高いといえる。 $P(x|\theta), P(y|\theta)$ の間にどのような関係があれば、前述の不等式が成立するかを求め A Comparison of the Value of Information in 2-state 2-action Statistical Decision Problem...A Dominance Criterion...と題して Pacific Area Statistical Conference で発表した。

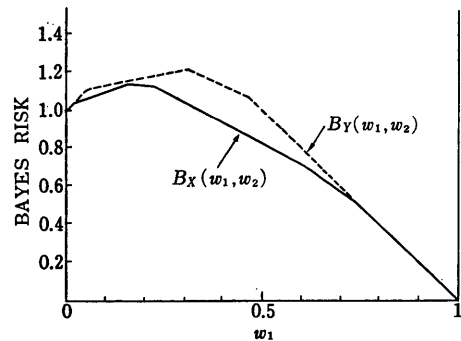


図 1

表 1

	a_1	a_2
θ_1	l_1	$l_1 + d_1$
θ_2	$l_2 + d_2$	l_2

表 2

	x_1	x_2	x_3	x_4	x_5	x_6
θ_1	0.40	0.24	0.10	0.16	0.06	0.04
θ_2	0.03	0.12	0.07	0.15	0.28	0.35

表 3

	y_1	y_2	y_3
θ_1	0.5	0.3	0.2
θ_2	0.1	0.4	0.5

表 4

	a_1	a_2
θ_1	0	3
θ_2	2	1

応力解放型点過程モデルの非正則漸近理論

尾形 良彦

今回報告する点過程モデル

$$(1) \lambda(t) = \exp(\alpha + \beta t - \gamma N_t)$$

は地震応力の解放モデル (Vere-Jones, 1978) から派生したものである。このモデルの尤度は危険度関数 (1) の対数線形性から単峰であることがわかり、最尤

推定値は常に一つだけ存在する。しかしながら、その漸近的な性質は複雑であり、通常の漸近理論 (Ogata, 1978; Kutoyants, 1979, 1980) の枠内には無い。

真のパラメタが $\beta, \gamma > 0$ であるときに話を明確にするために

$$(2) \lambda(t) = \exp(\alpha + \beta X_t - \gamma t/T)$$

なるパラメタ化を考える。ただし $X_t = t - N_t$ であり、 T は観測時間の長さである。 X_t はマルコフ過程であるが t が整数のときは必ず整数の値をとる。最近

の Tweedie (1982a, b) の結果を応用することによって X_t の定常分布に関する全モーメントの存在が証明できて、尤度の漸近展開に伴う、大数の法則が保証される。さらに Fisher score の中心極限定理にはマルテンゲールに基づいた Kutoyants (1979) の補題が最も有効である。

かくして、 $\beta, \gamma > 0$ のときには、通常の漸近理論(一致性, 正規性, X^2 性など)に帰着できることになった。

$\beta = \gamma = 0$ のときには上記の場合とは質量ともに全く異なった性質をもつ。これを調べるために

$$(3) \quad \lambda(t) = \exp(\alpha + \beta X_t / \sqrt{T} + \gamma t/T)$$

なるパラメタ化で考える。 X_t, T は上記と同じである。このとき N_t はポアソン過程であるので

$$W_T(\tau) = (N_{\tau T} - \lambda \tau T) / \sqrt{\tau T},$$

($0 \leq \tau \leq 1$) は標準 Wiener 過程に弱収束する。これに基づく Fisher score は正規分布でなく、Hessian は Fisher 情報行列(定数)に収束せず、分布をすることになる。

かくして、一致性が得られたものの、奇妙な漸近分布が得られた。特殊な場合には Cramér-von Mises 適合検定量の逆数の漸近分布になる。

Kutoyants, Y.A. (1979). Local asymptotic normality for processes of Poisson type, *Izv. Arm. Akad. Nauk, Seriya Matem.*, **14**, 3-20.

Kutoyants, Y.A. (1980). *Estimation of Parameters of Stochastic Processes* (in Russian). Arm. Akad. Nauk. Erevan.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimates for stationary point processes, *Ann. Inst. Stat. Math.*, **30**, 243-261.

Tweedie, R.L. (1982a). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory (to appear).

Tweedie, R.L. (1982b). The existence of moments for stationary Markov chains (submitted to JAP).

Vere-Jones, D. (1978). Earthquake prediction—a statistician's view, *J. Phys. Earth*, **26**, 129-146.

第2研究部

大気環境データの解析をめぐる

樋口 伊佐夫

大気環境のモニタリングは都道府県の仕事で、都道府県では固定測定局を要所要所に置いて、主な汚染物質濃度と気象要因を常時観測している。このデータの

蓄積を、大気汚染の実態の時空間的機構の解明に役立てられないものであろうか？ 事象が三次元的であるにもかかわらず、空の方向におけるデータがないことが、この問題を決定的に困難なものとしているが、部分的知見を加えてゆく上では役立つはずである。

こうした見解の下に、筆者は東京都大気保全部の昭和50年度～53年度の一般大気環境測定局のデータをもとにした範例的研究をはじめた。ちなみに東京都の場合、奥多摩山地と島をのぞいて、全都にほぼ一様に散在する36ヶ所の固定測定局で、 NO , NO_2 , CO , SO_2 , O_3 , ダスト等の濃度と気温、湿度、風向、風速が測定され、各一時間平均がセンターに送られる仕組みになっている。解析にあたっては、実態の把握と機構の解明に力を注ぎ、新しい解析法の開発や、従来の方法の改良洗練化などは眼中におかなかった。また使いたれた統計家だけが結果を理解し得るような、多次元解析のすゝんだ方法は、一般の人の考えている現象の解明にはつながらないので敢えて避け、単純な手段をつみ重ねてゆく方法をとった。その過程で発想した考えと結果のいくつかを述べる。

(1) 都平均からの偏差の利用: データを多くの地点の多次元時系列として見た場合、都全体の大きな変動に比べ、変動パターンの地域差が小さいため生データにもとづく限り、空間特性の面白い性質があらわれにくい。そこで各時間毎における都平均からの偏差を基礎データとして用いる。これによって、二点間の距離と濃度の相関との関係を見ると、コヒーレントな領域の広さが推定される。郊外では予想通り、都心部より広い。これは都心部と郊外部との相違の特徴である。

(2) 器差によらない方法: 各地点のピーク時の解析は測定器差によらない。

(3) 風の輸送効果: 風速風向を勘定に入れた二地点間のクロス相関を、それらを考慮しない相関と比べることにより輸送効果をみた。取扱った範囲では効果は認められない。すなわち汚染物質は“風によって運ばれる”のではなく“風によって吹き飛ばされる”と言った方が適切である。

(4) 高層気象データの利用: ラジオゾンデのデータから逆転層の高さを簡便に推定する方法を考案し、濃度との関連を考察中であるが有望である。

Spatial pattern の統計解析

種村 正美

生物個体群、地震の震央、都市などの配置パターンの統計解析法の必要性和重要性が次第に認識されてき

ている。その目的の一つは、与えられたパターンに関して、それが生成されたメカニズムや過程について統計モデルを構成し、モデルの母数を推定するなどして縮約された情報を得ることである。そして、得られた情報から、データと同一の条件下でパターンを再生したり、異なる条件下で生成されるパターンを予測したりすることが実用上必要となる。

相互作用ポテンシャルによるモデル化はそのような問題把握に対して一つの有効な手段となりうる。即ちポテンシャルが本来、各個体に備わったもので個体群密度や周囲の環境とは独立の特性であると考えれば、ポテンシャルが一旦データから得られれば上記の予測がシミュレーションなどを通じて可能となる。

平面領域 V の中に N 個の個体 (座標 $X = \{x_i, i = 1, 2, \dots, N\}$) が存在し、各個体は二体相互作用ポテンシャル $\Phi_\theta(r_{ij})$ ($r_{ij} = |x_i - x_j|$) の下で平衡な配置をとっているとすれば、母数 θ の対数尤度関数 $L(\theta; X)$ は統計力学の Gibbs 分布を介して

$$L(\theta; X) = -\sum_{i < j}^N \Phi_\theta(r_{ij}) - \log \bar{Z}(\Phi_\theta; N, V) \quad (1)$$

と表わされる。ここで \bar{Z} は Gibbs 分布の規格化因子である。

ポテンシャル関数 $\Phi_\theta(r)$ を尤度法で推定するには (1) における $\log \bar{Z}$ を Φ_θ, N, V の関数として与えることが不可欠であるが、これは一般には困難である。

関数 $\Phi_\theta(r)$ として $\Phi_\theta(r) = \Phi_1(r/\theta)$ のタイプに限定すると N, V, θ は一つの母数 $\tau = N\theta^2/V$ にまとめられ、

$$(1/N) \log \bar{Z}(\tau, \Phi_1; N) = -\int_0^\tau \psi(t)/t \cdot dt$$

と表現できる。ここで ψ は $PV/N-1$ (P は圧力) に対応する関数である。

この場合には問題は結局、 $\psi(\tau)$ をいかに近似するかに集約される。方法はいくつかあるが、その中でも計算機シミュレーションによる方法の重要性を、パターン生成の計算機実験の重要性とともに強調しておきたい。それは任意の τ の値において有効であるからである。

τ の小さい領域ではビリアル展開

$$\psi(\tau) = B_2\tau + B_3\tau^2 + \dots + B_k\tau^{k-1} + \dots$$

(B_k ; k 次ビリアル係数)

が適当な Padé 近似の下で有効であり、計算機実験で得た $\psi(\tau)$ のチェックともなる。われわれは B_k を構成するブロックグラフを任意の k について得るプログ

ラムを作成し $k=8$ までのグラフを求め、一つのポテンシャルモデルについて上の方法で $\psi(\tau)$ の値をチェックした。本研究は尾形良彦氏との共同研究である。

集団構造・構造母数と集中解析

田 口 時 夫

統計学は今一つの岐路にあるように思われる。それは一般的財政危機により外的に機構の存続が問われている事と共に、内的に対象・方法論上の性格が問われる段階にあるように思われる。

統計学説史によると、本来数理統計学の名称は統計表の解析に由来するとされている。とすれば戦後の一時期における標本調査への関心やそれに続く統計推論への関心のたかまりは、本来の目的に対して序論の準備の段階にあったといえることになる。数理統計学が原点に立戻って統計表を通じて対象集団の構造そのものの究明を意図する時、単に確率論をもとにしたモデルの設定とその吟味のみでは不十分と思われる。集団構造・構造母数とは、そのような立場における一つの理論構成である。その統計表上の基礎は現行の家計調査統計に認められるフラクタル階層分割による統計表である。56年度は、この立場から集中曲面を解説し、又集中線形回帰方式を誘導した。57年度は、その回帰方式を更に整備し実証すると共に、統計的比率の解析法及び集団構造の測度論的公理化を試み、其等を日本数学会、日本統計学会等で報告した。ところで1980年に Kakwani により、ローレンツ曲線はマハラノビスによる一般化形式を通じてエンゲル曲線解析に有用であることが示されたが、此の種の解析は二次元化された集中曲面に於てより一層有効となることが予想出来る。以上述べた諸解析例は、従来の数理統計学では充分解析し得なかった分野であり、従って此等の分析を可能にする集中曲線・曲面の幾何学的特性量を特に構造母数とするのが適切であり、又それによる解析を一般に集中解析とすることを提案したい。

以上の趣旨による具体的解析例は、Mardia の二次元パレート分布を対象としたものであり、既に彙報に投稿されている (本号 1-32 頁参照)。

風の揺れと大気汚染濃度分布

岸 野 洋 久

大気汚染濃度を時間空間両側面から捉えようとする、時間変化する空間パターン解析に結びつく。そしてその要因は主に排出源、風、日射量その他の気象

条件や地形などに求められる。そこでここでは、風の揺れがパターンの変動にどのような影響を及ぼすかを見る。

分子レベルまで遡ると、各分子の位置 $x_t^{(i)}$ を支配する運動法則は、

$$(1) \quad dx_t^{(i)} = \{a(t, x_t^{(i)})dt + c(t, x_t^{(i)})d\tilde{B}_t\} + b(t, x_t^{(i)})dB_t^{(i)} \quad i = 1, 2, 3, \dots$$

と表現される。ここで、 $\tilde{B}_t, B_t^{(i)}, i=1, 2, \dots$ は独立なブラウン運動で、右辺第一項は風による移流項、第二項は分子拡散を表わし、大気の場合は後者の影響は微小である。 $a(t, x)dt$ は平均風速の、 $c(t, x)d\tilde{B}_t$ は風の揺れの場を示す。(さらに、空間相関に自由度をもった乱流場も考えられる。)

時刻 s 、地点 y における単位時間体積当りの排出量を $\lambda(s, y)$ とすると、時刻 t 、地点 x における濃度 $n(t, x)$ は (1) より次式を満たすことがわかる：

$$(2) \quad dn(t, x) = \{\mathcal{A}(t, x)n(t, x) + \lambda(t, x)\}dt + \mathcal{J}(t, x)n(t, x)d\tilde{B}_t$$

$$\text{但し, } \mathcal{A}(t, x)n(t, x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} \{ (b^2(t, x) + c^2(t, x))n(t, x) \} - \frac{\partial}{\partial x} (a(t, x)n(t, x))$$

$$\mathcal{J}(t, x)n(t, x) = -\frac{\partial}{\partial x} (c(t, x)n(t, x))$$

これは風の揺れが汚染濃度分布の揺れに伝播してゆく様子を表現しているが、特にその平均を見ると、時間平均濃度（即ちトレンド）は拡散方程式

$$(3) \quad \frac{\partial}{\partial t} E[n(t, x)] = \mathcal{A}(t, x)E[n(t, x)]$$

を満たす。つまり、平均濃度の動きは、平均風速を移流項、分子拡散と乱流拡散を加えたものを拡散項としたものより得られる。2 地点の濃度の共分散 $\text{Cov}(n(t, x), n(t, y))$ の満たす方程式も (2) から導かれる。

さらに、これを地点 x_1^0, \dots, x_L^0 で積算値の形で観測していたとする。

$$(4) \quad y_t^j = \int_0^t n(s, x_j^0)ds + \sigma \tilde{B}_t^j \quad j=1, \dots, L$$

ここで $\tilde{B}_t^j, j=1, \dots, L$ は $\tilde{B}_t, B_t^{(i)}$ と独立なブラウン運動。

現在まで得られた情報をもとに濃度分布の推定を逐次やっていくには、カルマンフィルターが適している。システム (2) は非線形な為、そのままでは閉じたアルゴリズムとならないが、修正することにより、擬似カルマンフィルターを構成することができる。

一方排出源からの排出量の揺れによる影響であるが、平均濃度 $E[n(t, x)]$ は排出量の揺れを捨象してトレンドを追えば求まることがわかるが、入り組んだ分散構造を持つ。

遺伝子構造データから系統樹を描く方法

長谷川 政 美

DNA や RNA の配列データから分子系統樹を描くのに従来、樹全体で必要最小限の塩基置換数が最小になるような樹を求める The Maximum Parsimony Method¹⁾を用いてきたが、この方法は系統ごとに置換速度が大きく異なる場合にはおかしな結果を与える。特に動物のミトコンドリア DNA は核 DNA にくらべて1けた速いので²⁾、ミトコンドリアの進化的起源をさぐるには、このような事実を考慮した方法が必要である。そのようなものとして Li³⁾の方法があるが、これは行列法であり配列データのもつ豊富な情報のうちの極くわずかしかり利用しておらず、また得られた結果に対して統計的評価ができないという欠点がある。そのような欠点を克服したものとして、Felsenstein⁴⁾の最尤法があるが、計算が大変なため従来具体的な問題に応用されたことがなかった。S-r RNA の配列データから Kuntzel と Köchel⁵⁾は、動物と菌類のミトコンドリアはどちらも共生説に従った由来をもつがその起源は別々であるとしているが、彼らの方法では置換速度一定の仮定がなされており、この場合明らかに不適当である。我々は、塩基座位、系統ごとに進化は独立に起こると仮定し、Felsenstein の最尤法によってS-rRNA の系統樹を求め、彼らとは逆に、動物と菌類のミトコンドリアが共通の起源をもつという結論を得た⁶⁾。更にこの方法を 5S rRNA のデータに適用することにより、従来系統学で問題となっていた様々の問題に対して、解答を与えることができた⁶⁾。

References

- [1] Hasegawa, M. (1978). Evolution of transfer RNA, Origin of Life, H. Noda (ed), *Proc. 2nd ISSOL Meeting*, Centre for Academic Publication Japan, 495-500; Komiya, H., Hasegawa, M. and Takemura, S. (1983). Nucleotide sequence of 5S rRNAs from sponge *Halichondria japonica* and tunicate *Halocynthia roretzi* and their phylogenetic positions, *Nucl. Acids Res.* 11, 1969-1974.
- [2] Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M. and Koike, K.

(1982). Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes, *J. Mol. Evol.*, **12**, 28-35.

- [3] Li, W. (1981). Simple method for constructing phylogenetic trees from distance matrices, *Proc. Nat. Acad. Sci. USA*, **78**, 1085-1089.
- [4] Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach, *J. Mol. Evol.*, **17**, 368-376.
- [5] Hasegawa, M., Kikuno, R., Miyata, T. and T. Yano (1983). The origin and the evolution of cellular organelles, *Endocytobiology II*, W. Schwemmler and H. Schenk (eds), Wachter de Gruyter (in press).
- [6] Hasegawa, M., Komiya, H. and Takemura, S. (1983). *Evolution of 5S rRNA* (in preparation).

多重四分表の解析における偏散分析

柳 本 武 美

1. 序

本年度の研究テーマとしては

- 安全性の評価と安全基準についての統計手法と
その実際.
- 比例ハザードモデルの生存時間モデル, 点過程
モデル, 分割表への適用について研究した. 特に全尤
度と部分尤度について詳細に検討した.
- 周辺尤度を最大にする規準によってモデルの適
合の診断をする方法を研究した.

2. 多重四分表

ここでは上記研究テーマの内2番目に関連して, 多重四分表の解析に関する新しい解析法を提案する.

データが複数の四分表として得られた場合多重四分表としてまとめられる. 多施設が参加した薬効試験(フェーズ3), あるいは独立になされた複数の疫学調査が典型的な例である. 表1はその1例である(Fleiss, Wiley (1981) より).

表1 複数の調査による分裂症の診断

調査	地 域	分裂症	その他
1	New York	81	24
	London	34	71
2	New York	118	74
	London	69	105
3	New York	82	63
	London	52	93

3. モデル

データを

$$n_{ij}^s \quad s = 1, \dots, S, \quad i = 1, 2, \quad j = 1, 2$$

と書く. 表1の場合 $s=3$ である. モデルとしてはこれらのデータが $2s$ の2項分布からの実現値とみなす. 言い換えると

$$n_{ij}^s \sim Bi(n_i^s, p_i^s)$$

とみなす. 通常以下のモデルが標準的である. $Z_1=1$, $Z_2=0$ として

$$(i) \quad H_S: \quad p_i^s \text{ は任意}$$

$$(ii) \quad H_0: \quad p_i^s = p$$

$$(iii) \quad H_{COM}: \quad p_i^s = \frac{\exp(\alpha + \beta z_i)}{1 + \exp(\alpha + \beta z_i)},$$

$$(iv) \quad H^{STR}: \quad p_i^s = \frac{\exp(\alpha_s)}{1 + \exp(\alpha_s)},$$

$$(v) \quad H_{COM}^{STR}: \quad p_i^s = \frac{\exp(\alpha_s + \beta z_i)}{1 + \exp(\alpha_s + \beta z_i)}.$$

4. 解析法

提案する解析法はモデル (i) ~ (iii) では

$$n_{ij}^s \rightarrow n_{ij}^s + \frac{n_{..}^s}{n_{..}^s - 1} \cdot \frac{1}{n_{..}^s - 1} \cdot \frac{n_i^s \cdot n_{.j}^s}{n_{..}^s}$$

とし, モデル (iv), (v) では

$$n_{ij}^s \rightarrow n_{ij}^s + \frac{1}{(n_{..}^s - 1)} \cdot \frac{n_i^s \cdot n_{.j}^s}{n_{..}^s}$$

として偏散分析を行うことである. 全尤度法と部分尤度法の長所を兼ねそなえることが期待される. 細かい点については更に検討が必要とされる.

確率的死亡モデルとその応用

鎌 倉 稔 成

通常の生存時間データの研究ではある個体の死亡あるいは故障までの時間 (一般的にはある事象が生起するまでの時間) をその解析対象とする. ここでは事象が生起するしないの2通りの場合しか考えられない. これに対して生起の仕方を0と1の間の連続量を取り得るように拡張した方が良い場合も考えられる.

腰痛の時間的な軽減現象では患者の最初の痛みを1として時間ごとに得られる痛みの強さから0と1の間の軽減曲線が規定される. むろん痛みが取れるまで観察して通常の生存時間データの解析方法を用いることもできるが, 観察時間が長くなったり患者の脱落があったりしてそのようなデータを得るのは困難である.

また、痛みの軽減だけでなく時間とともに病態が悪化あるいは回復するような現象からもこのようなデータが得られる。

いま、 i 番目の患者について t_{ij} 時点での痛みの強さを v_{ij} ($i=1, \dots, n; j=0, \dots, m_i$) とする。痛みは相対的なものであるとして規準化し、

$$(1) \quad \begin{cases} v_{i0} = 1 \\ 0 \leq v_{ij} \leq 1 (i=1, \dots, n; j=1, \dots, m_i) \end{cases}$$

で与えられるものとする。 v_{ij} は一般的には j に関して単調に減少するとは限らないが、ここでは生存時間分布関数に対応したモデリングを行うために単調非増加であることを仮定する。つまり

$$(2) \quad 1 = v_{i0} \leq v_{i1} \leq \dots \leq v_{im_i}$$

を仮定するのである。

次に患者 i から得られる痛みの軽減曲線を、その生存時間分布関数との類似性により K 人から得られた生存時間分布（痛みが取れるまでの時間の分布）の Kaplan-Meier (1954) の推定値と見做して解析を行う。この解釈は、時刻 t_{ij} において患者 i の痛みを、同一の性質を持つ K 人の集団で考えると $K(1-v_{ij})$ 人は痛みがとれるという確率的な痛みの軽減を考えていることに対応する。

i 番目の患者に対して K 人からなる集団の痛みがとれるまでの時間 T_i の分布の密度関数を $f_i(t_i; \theta)$ 、その生存時間分布関数を $S_i(t_i; \theta)$ とすれば、尤度関数は

$$(3) \quad L' = \prod_{i=1}^n \left[\prod_{j=1}^{m_i} f_i(t_{ij}; \theta)^{k_{ij}} \right] S_i(t_{im_i}; \theta)^{k_{im_i}}$$

と表わすことができる。ただし、 $p_{ij} = v_{ij-1} - v_{ij}$ としてある。ここで (3) 式で K を取ったものを L と置けば $\log L' = (\log L)K$ となりパラメータの推定には関係がなくなる。また L を尤度と考えた方が良くとも Kullback-Leibler の情報量の意味からも示される。さらに f_i に比例ハザードモデルを仮定して痛みの時間的軽減に影響を与える要因が評価できる。

第3研究部

判別について

藤 本 照

潜在変数 θ と観測可能変数 X の対 (θ, X) の独立な系列を考える。 θ はランダム母数であって、未知の先験分布をもち、 θ の値 $\theta = \theta$ が正しいとき、尤度関数

$f(x|\theta)$ をもって X が観測される。これが経験的ベイズモデルの構造である。いま母集団 π からの個体の観測結果から、この個体を r 個の範疇 $\pi_1, \pi_2, \dots, \pi_r$ のいずれかへ類別しようとする問題を考えると、これを上のモデルの形に構成するのが便利である。すなわち、母数空間として、 π_i の替りにラベル i の集合 $\Omega = \{1, 2, \dots, r\}$ とすると、その上の先験分布 $w = (w_1, w_2, \dots, w_r)$ を考え、 $\theta = i$ のときの尤度関数を $f_i(x)$ とする。また $L(i, j) (\geq 0)$ を、 π_i からの個体を π_j へ判別するときこうむる損失とし、 $\max_j L(i, j) < \infty$ を仮定する。このとき、確率化決定法則 $\phi(x) = (\phi^{(1)}(x), \dots, \phi^{(r)}(x))$ は、 $\phi^{(j)}(x) \geq 0, \sum_{i=1}^r \phi^{(i)}(x) = 1$ である。ここに $\phi^{(j)}(x)$ は $X=x$ を観測するとき、 π_j からとしてその個体を判別する確率である。

X が $f_i(x)$ をもつとき、任意の $\phi(x)$ に対する危険度は、

$$R(i, \phi) = \int \sum_{j=1}^r L(i, j) \phi^{(j)}(x) f_i(x) d\mu(x),$$

$w = (w_1, w_2, \dots, w_r)$ に関する期待危険度は、

$$B(w, \phi) = \sum_{j=1}^r w_j R(i, \phi).$$

$h_j(x; w) = \sum_{i=1}^r w_i L(i, j) f_i(x)$ とするとき、与えられた w に対して、 $B(w, \phi)$ は、

$$\phi_w^{(j)}(x) = \begin{cases} 0 & (h_j(x; w) > \min_t h_t(x; w) \text{ のとき}) \\ \gamma_j(x) & (h_j(x; w) = \min_t h_t(x; w) \text{ のとき}) \end{cases}$$

とすれば最小になる。 $0 \leq \gamma_j(x) \leq 1, \sum_{j=1}^r \gamma_j(x) = 1, j=1, 2, \dots, r$ 。

π からの大きさ n の標本 (X_1, X_2, \dots, X_n) の各 $X_k (k=1, 2, \dots, n)$ は、 $f(x) = \sum_{i=1}^r w_i f_i(x)$ をもつから、 $f(x)$ に関して不偏な推定量 \hat{w} が求められ、上の決定法則に代入された $\phi_{\hat{w}}(x) = (\phi_{\hat{w}}^{(1)}(x), \dots, \phi_{\hat{w}}^{(r)}(x))$ は、ベイズ危険度 $B(w) = B(w, \phi_w)$ へ確率収束する。

ランダムな衝突モデルの集団遺伝学への応用

伊 藤 栄 明

箱のなかに1から m までの型の粒子が n 個あり、ランダムな衝突をくりかえしているとする。型 i 、型 j からなる2粒子が衝突する際に i が j より強い確率

が $1/2 + a_{ij}$, ($a_{ij} + a_{ji} = 0$, $|a_{ij}| \leq 1/2$), であるとし, 衝突により弱い方の 1 粒子が強い方の 1 粒子に変化する. 集団遺伝学における selection model はこのモデルに対応する. 特に $a_{ij} = 0$ がすべての i, j について成り立っている場合は, Wright model に対応する. すなわち遺伝子頻度の変化率の平均及び共分散がこれらのモデルと同様になるようなランダムな衝突モデルを考えることができるのである. 以上のような 2 体の衝突にもとづいたモデルだけでなく, 例えば 4 体の衝突にもとづいたモデルを考えると集団遺伝学における overdominance model に対応するものとなる. A_1, \dots, A_m なる m 個の異った遺伝子からなる系を考える. 2 倍体生物で N 個体からなるものとする. $2N$ 個の遺伝子のなかから先ず 2 個の遺伝子 A_i, A_j がランダムにえらばれ結合し $A_i A_j$ となる. 次に残りの $2N-2$ 個のなかから A_k, A_l がランダムにえらばれ結合し $A_k A_l$ となるとする. $A_i A_j$ と $A_k A_l$ が衝突し, 確率 $1/2 + S_{ij,kl}$ で 2 個の $A_i A_j$ に変化し, 確率 $1/2 + S_{kl,ij}$ で 2 個の $A_k A_l$ に変化し, ここで $S_{ij,kl} + S_{kl,ij} = 0$, $|S_{ij,kl}| \leq 1/2$, $i \neq j, k \neq l$ ならば $S_{ij,kl} = S$, $i = j, k \neq l$ ならば $S_{ij,kl} = -S$, $i \neq j, k = l$ または $i = j, k = l$ ならば $S_{ij,kl} = 0$ とする. このような衝突がくりかえされていくというモデルを考えると, 遺伝子頻度の平均及び共分散は $E(dx_i) = C_1 S ((\sum x_i^2) - x_i) dt$, $E(dx_i dx_j) = C_2 x_i (\delta_{ij} - x_j) (dt)^2$ となる. ここで C_1, C_2 は集団の大きさ及び世代交代の速度により定まる定数である. 集団遺伝学においては通常, 適応値という量にもとづいた別の記述がされていて, その場合共分散は同じであるが $E(dx_i) = C_1 S ((\sum x_i^2) - x_i) / (1 - S \sum x_i^2)$ となる. S が小さい時はランダムな衝突モデルと同じと考えてよい. 確率微分方程式による計算機実験は集団遺伝学において多く行われ始めた. ランダムな衝突モデルを計算機実験にもちいることができる. 拡散近似によらずに, より離散的なもので近似をするということも自然な方法であると思われる. ランダムな衝突モデルを考えることにより, たとえば大きさ 10^6 の集団のかわりに大きさ 10^2 の集団を考え overdominance model についての計算機実験を行うことができる.

非正規外乱に対するシステムの応答

岡崎 卓

1. システム応答の確率分布

風波を受けて動揺する船舶や雑音を含む電気回路な

ど外乱を受けるシステムの解析では, システムの応答 $V(t)$ (t は時間) の確率密度 $f(u, t)$

$$f(u, t) du = P_r\{u < V(t) < u + du\}$$

を見出すことが重要であるが, 外乱 $W(t)$ が正規確率過程から外れている場合, $f(u, t)$ を定める直截簡明な方法は未だ充分に開発されていない.

システムの時間発展を

$$\frac{dV}{dt} = M(v) + W (= \dot{V})$$

と表わし, $W(t)$ については白色正規過程 $u(t) (< u(t) u(t+\tau) = \sigma^2 \delta(\tau))$ で駆動される非線型モデル

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

を設定すれば, 全系 (V, W) の結合確率分布 $P(V, W)$ は Fokker-Planck 方程式

$$\frac{\partial p}{\partial t} = -iLP, \left(-iL = \frac{\partial}{\partial v} \dot{V} + \frac{\partial}{\partial w} N + \frac{\sigma^2}{2} \frac{\partial^2}{\partial w^2} \right)$$

を満たし, $f(u)$ はその周辺分布として与えられるわけであるが, この方程式を直接解くのは困難を伴うのみならず, 不要な情報をも同時に求めねばならぬ点で煩雑迂遠であり, 結合分布 $P(V, W)$ を介さずに $f(u, t)$ を見出す方法が望ましい.

2. 射影子法の援用

微粒子の集団を取扱う統計力学ではミクロの運動法則からエネルギーやエントロピーなど少数のマクロな量についての法則を抽出する射影子法が開発されつつある. ここでは (V, W) をミクロ量, V の特性関数 $\psi_u = \delta(v-u)$ をマクロ量とみなして射影子法を適用し, $f(u, t)$ を直接に定める方程式の構成を試みる.

先ず, ψ_u をそれ自身にうつす射影子 $P(P\psi_u = P, P\psi_u = \psi_u)$ を

$$PX(V, W) = \int du_1 \psi_{u_1} t_r \{g \psi_{u_1} X\}$$

のように構成する ($g(w)$ は外乱の確率密度, $t_r = \int (dV dW)$). このとき時間を進める作用素 $e^{i\hat{L}t}$ (\hat{L} は L の随伴作用素) が

$$\begin{aligned} e^{i\hat{L}t} &= e^{i\hat{L}t} P \\ &+ \int_0^t ds e^{i\hat{L}s} P i \hat{L} (1-P) e^{i\hat{L}(t-s)} \\ &+ (1-P) e^{i\hat{L}(t-P)t} \end{aligned}$$

と分解され, ミクロ量 $X(V, W)$ の時間変化をマクロ量 ψ_u によって表現することが可能となり, 最終的に $f(u, t) = t_r \{p e^{i\hat{L}t} \psi_u\}$ を決める “一般化 Fokker-Planck 方程式”

$$\frac{\partial}{\partial t} f(u, t) = - \frac{\partial}{\partial u} M(u) f(u, t) + \frac{\partial}{\partial u} \int_0^t ds \int du_1 D(u, u_1, t-s) \frac{\partial}{\partial u_1} f(u_1, s),$$

$D(u, u_1, t) = t_r [g w \psi_{u_1} (1-p) e^{i \mathcal{L}(1-p)t} w \psi_u]$ に達する。右辺に現われた時間積分は結合分布 $p(vw)$ から周辺分布 $f(u, t)$ に移行する際に失われた情報を過去の軌跡 $f(u, s)$, $0 < s < t$ の利用により必要なだけ回復する役割を担っている。外乱 $w(t)$ に比べてシステム応答 $v(t)$ の変化が遅ければ、近似的に

$$D(u, u_1, t) \doteq \frac{1}{2} K \delta(u - u_1),$$

$$K = \int_{-\infty}^{\infty} \langle w(t) w(t+s) \rangle ds$$

となり、上記方程式は同じ強度 K をもつ正規外乱に対して応答確率を与える通常の Fokker-Planck 式となる。

誤差分散の巾乗不均一性と部分情報の関係

鈴木 義一郎

“誤差分散が説明変数の $2g$ 乗に比例する”という回帰モデルに於て、回帰係数の切片 α が 0 に近い場合を考える。勾配 β の最良不偏推定量 $\hat{\beta}$ に対する競争相手として、 $\alpha=0$ を仮定したときの一般化最小 2 乗推定量 $\hat{\beta}_0$ を考える。それらは

$$\hat{\beta} = \frac{T(1-2g)S(-2g) - T(-2g)S(1-2g)}{S(2-2g)S(-2g) - S(1-2g)^2},$$

$$\hat{\beta}_0 = \frac{T(1-2g)}{S(2-2g)}$$

のように与えられる。ここで

$$S(\alpha) = \sum_{i=1}^n x_i \alpha, \quad T(\alpha) = \sum_{i=1}^n y_i x_i \alpha.$$

$\hat{\beta}_0$ の分散は $\hat{\beta}$ の分散より当然小さいけれども、 $\alpha=0$ でない限り β の不偏推定量にはならない。

そこで、 $\hat{\beta}_0$ の $\hat{\beta}$ に対する (最小 2 乗誤差の意味での) 効率を

$$\rho\left(\frac{\alpha}{\sigma}, g\right) = \frac{E\{(\hat{\beta}_0 - \beta)^2\}}{E\{(\hat{\beta} - \beta)^2\}} = \frac{V\{\hat{\beta}_0\} + (E\{\hat{\beta}_0\} - \beta)^2}{V\{\hat{\beta}\}}$$

と定義する。特に

$$\rho(0, g) = 1 - \frac{S(1-2g)^2}{S(2-2g)S(-2g)}, \quad \rho(A(g), g) = 1$$

となることが確かめられる。ここで

$$A(g) = \sqrt{\frac{S(2-2g)}{S(2-2g)S(-2g) - S(1-2g)^2}}$$

$x_i = i$ ($i=1, 2, \dots, u$) の場合について $\rho(0, g)$ の値を計算してみると、 $g=0$ のときで 4 分の 1、 $g=1$ のケースで 4 分の 3 くらいの値をとることが分かるから、 α の値が 0 に近いと予想されるケースでは $\hat{\beta}_0$ という推定量を積極的に利用すべきである。

そこで、一般化最小 2 乗推定によって

$$\hat{\alpha} = \frac{T(-2g)S(2-2g) - T(1-2g)S(1-2g)}{S(2-2g)S(-2g) - S(1-2g)^2}$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \frac{1}{\alpha_i^2 g} [y_i - (\hat{\alpha} + \hat{\beta} x_i)]^2$$

という推定値が得られるから

$$\hat{\beta} = w \hat{\beta}_0 + (1-w) \hat{\beta} \left(w = \frac{\hat{\sigma}^2 A(g)^2}{\hat{\alpha}^2 + \hat{\sigma}^2 A(g)^2} \right)$$

といった推定量の有用性が、数値実験を用いて確かめることができた。

級内相関係数の正規化変換

小西 貞則

0. はじめに

標本相関係数に対しては実用上いわゆる Fisher の z 変換が用いられる。これは変換によって

(a) 漸近分散がパラメータを含まない形にできる (分散安定化変換)

(b) 正規分布への収束速度を向上させることができる (正規化変換)

の二つの理由からである。ここでは k 変量正規母集団からの観測値に基づく級内相関係数に対して、上述の性質 (a), (b) を持つ変換の存在について考察する。

1. 級内相関係数と変換

平均ベクトル $\mu = [\mu_1, \mu_2, \dots, \mu_k]'_{(k \times 1)}$, Σ は同一の分散および共分散を持つものとする。すなわち $\Sigma = \sigma^2[(1-\rho)I + \rho e e']_{(k \times k)}$; $e = (1, 1, \dots, 1)'$. k 変量正規母集団 $N_k(\mu, \Sigma)$ からの観測値に基づく級内相関係数 ρ の最尤推定量を r_I とする。 r_I に対しては、一般に z 変換

$$(1) \quad Z_k(r_I) = \sqrt{\frac{k-1}{2k}} \log \frac{1 + (k-1)r_I}{1 - r_I}$$

が用いられる。しかし、この変換は「任意の次元 k に対して性質 (a) は満たすが、同時に性質 (b) を満たすのは $k=2$ の時だけである」。このことの証明と $k \geq 3$ の時の変換について検討する。

2. 変換の理論

一般に r_1 の関数 $f(r_1)$ を考え、性質 (a), (b) を持つ具体的な関数形を捜す。まず関数 $f(r_1)$ に対して次の漸近展開式が導かれる。

$$\begin{aligned} P_1[\sqrt{N}\{f(r_1) - f(\rho) - C/N\}/\tau < x] &= \Phi(x) \\ &- \frac{1}{\sqrt{N}} \left\{ \frac{-5k+7}{3\alpha k(k-1)} \right. \\ &\left. - \frac{C}{\tau} + a(\rho, f', f'')x^2 \right\} \phi(x) + O(N^{-1}) \end{aligned}$$

ここに $\Phi(x)$, $\phi(x)$ は各々正規分布関数とその密度関数, $\alpha = \{2/k(k-1)\}^{1/2}$, $\tau = \alpha(1-\rho)\{1+(k-1)\rho\}f'(\rho)$ とする。

性質 (a) の分散安定化変換は, $\tau=1$ の解として与えられ, それは (1) の z 変換である。一方性質 (b) の正規化変換を求めるには, 変換の一般論 (Konishi; 1981, Biometrika 68) に従うと $1/\sqrt{N}$ の項に含まれる $a(\rho, f', f'')$ に対して微分方程式 $a(\rho, f', f'')=0$ の解を必要とする。これは結局

$$(2) \quad \frac{df}{d\rho} = (1-\rho)^{-2(2k-2)/3k} \{1+(k-1)\rho\}^{-2(1+k)/3k}$$

の解を求める事に帰着される。解は $k=2$ の時 $Z_2(r_1) = (1/2) \log \{(1+r_1)/(1-r_1)\}$ となり, $c=1/2$ と置くと性質 (b) も同時に満たす。一方 $k \geq 3$ の時 (2) の解は

$$3/(k-2) \cdot \{[1+(k-1)r_1]/(1-r_1)\}^{(k-2)/3k}$$

となり性質 (a) は満たされなくなる。

3. z 変換の収束速度

(1) の z 変換 $Z_k(r_1)$ の収束の Order を評価すると

$$\begin{aligned} k=2: P_1 \left[\sqrt{N} \left\{ Z_2(r_1) - Z_2(\rho) + \frac{1}{2N} \right\} < x \right] \\ = \Phi(x) + O(N^{-1}) \end{aligned}$$

$$\begin{aligned} k \geq 3: P_1 \left[\sqrt{N} \left\{ Z_k(r_1) - Z_k(\rho) \right\} < x \right] \\ = \Phi(x) + O(N^{-1/2}) \end{aligned}$$

となり正規分布への収束は $k \geq 3$ になると $k=2$ の時ほど速くないことがわかる。以上のことは, $\mu = [\mu_1, \mu_2, \dots, \mu_p]'$ とした場合の最尤推定量の変換を考える時にも言えることである。

二相回帰のモデル選択について、
他

野田 一雄

本年度の主要な研究の一つとして、表記の課題につ

いて、伊藤政志 (統数研特別研究員・都公害研) と共同研究を行った。これまでに、ある点 $x=0\theta$ (未知) を境界として二相に分れる回帰

$$y_i = \tau_k(x_i, \theta) + \varepsilon_i;$$

$$\tau_1(x_i, \theta) = \theta_k + \sum_{j=1}^p \theta_j (x_i - 0\theta)^j \quad (i=1, 2, \dots, m),$$

$$\begin{aligned} \tau_2(x_i, \theta) &= \theta_0 + \sum_{j=1}^q \theta_j' (x_i - 0\theta)^j \\ &\quad (i=m+1, m+2, \dots, l), \end{aligned}$$

$\varepsilon_i \sim N(0, \sigma^2)$ において, $x_m \leq 0\theta \leq x_{m+1}$ となるものを一つのモデルとして設定し, AIC を基準とするモデル選択によってパラメータ θ, σ^2 を推定する方法を提起してきた (Behaviormetrika, 13(1983)). 今回, 二相回帰がクリティカルな状態にある場合の上記の推定法の安定性を調べた。折れ線の場合, θ_0, θ_0 のミニマム AIC 推定量 $\hat{\theta}_0, \hat{\theta}_0$ の漸近分散がともに $\sigma^2/(\theta_1 - \theta_1')$ および $(\bar{x} - 0\theta)^2 / \sum_i (x_i - \bar{x})^2$ に比例することから, これらの値の変化に対応して, AIC が真のモデル ($i=m$) ではなく, 異った区間 ($i \neq m$) に θ_0 が存在すると想定する偽のモデル, および通常の多項式回帰 (偽のモデル) を選択する状況をシミュレーションによって調べた。この傾向については, 竹内氏 (1976) が提案した情報量規準はより不安定であることが示され, AIC の改良になっていないことも同様に調べられた。また, 逆のことがら, すなわち真の回帰が通常の二次式や三次式であるときには, モデル選択が安定していることも確かめられた。これらの結果は, 回帰が二相をなすか否かという事前情報がこれらのモデル選択に大きく関係することを示している。

主要な他の研究として, 石井恵一 (阪大) と共に, 統計的決定関数の許容性に対する線型位相空間論的アプローチの研究を行った。より一般的な構造として, 決定関数の許容性ならびに改良可能性がそれぞれ凸錐による擬順序についてのミニマリティーおよびその否定として定義されることを示し, 一般的な制約条件下でこれらが成立するための必要十分条件を求めた。後者の問題は, その構造を線型位相空間における数値計画法問題の可能解の存在条件のそれに帰着させることによって解決される。具体的な適用としては, Lecam (1955) の考え方の拡張として意義をもつ必要十分条件, almost-admissibility についての James-Stein (1961) の定義をあらためてこれが成立するための必要十分条件, Hashimoto-Kūdo (1968) の \mathcal{S} -許容性のための必要十分条件の拡張, そして危険関数がパラメータについて連続な場合における必要十分条件を導出

した (Proceedings of Pacific Area Statistical Conference に投稿)。

第4研究部

標本相関係数の分布の高次漸近展開

仁木直人

相関係数 ρ を持つ2変量正規母集団から得られる標本相関係数 r の分布に関する漸近展開を標本サイズ (N) の (-4) 乗のオーダーまで求めた。この展開を用いることにより、 N が11程度でも、標本相関係数 r の分布関数を小数点以下5桁まで正確に計算することができる。 r の代わりに、その正規化変換 (および分散安定化変換) である $z = (1/2) \log \{(1+r)/(1-r)\}$ の分布に関する漸近展開を行なう方が、式の簡単さおよび収束の速さの両面から有利であるから、実際の計算にはこの方を用いる。分布関数の漸近展開からは、分布のパーセント点を計算するために便利な Cornish-Fisher 逆展開を Hill and Davis [AMS, 39 (1968)] の方法により同じオーダーまで求めることに成功し、その特殊な場合として、メディアンの値を求める漸近展開を得ることができた。漸近展開に基づく計算式は、 (N, ρ, r) のいろいろな組合せに対して、その「正確な値」と比較することにより、近似の評価を行なった。分布関数の「正確な値」を計算するためには、 Z の密度関数を Soper ら [Biometrika, 11 (1916)] の N に関する漸下式により順次求めていき、それらを必要な精度が得られるように数値積分を行った。パーセント点は、「正確な分布関数値」の3次逆補間して求めた。下の表は、分布関数値を求めるいくつかの方法について、 $N=11$ のときの $r=-0.995$ (0.005) 0.995 に対する最大近似誤差 ($\times 10^6$) を示す。

ρ	Ruben	Kraemer	Winterbottom	仁木・小西
・1	1671	70	158	4
・3	2739	623	166	5
・5	4001	1787	151	6
・7	5388	3655	169	8
・9	6782	6342	291	9

尚、本研究は、小西貞則氏との共同研究であり、多くの計算に数式処理用の言語である REDUCE-2 を活用した。

相関のある多重時系列データのスムージング

柏木宣久

本報告は「非定常マルコフ連鎖の遷移確率の推定」(N. Kashiwagi and H. Kishino, Res. Memo. No. 254) の研究過程で得られた成果の一部である。

1. はじめに

調査では、設問の解答として選択肢を用意し、その何れかを選ばせるという方法がよく採用される。こうした方法で得られた調査結果の解析では、多項分布を利用して結果に含まれる標本誤差が評価される。

一方、良く知られているように、調査結果には様々な理由によって生ずる系統誤差が混入しているのが現実である。しかも結果に対し標本誤差よりも深刻な影響を与えていると予想される。従って系統誤差の評価は重要である。

しかしながら、全く同じ環境のもとでの調査の繰り返しが現実的ではない為に、系統誤差の評価は識別不能であると考えられてきた。この問題に対する定式化された提案を知らない。

ところがこの識別の問題であるが、同様の調査が継続的に実施されている場合には多少事情が異なるように思われる。つまり、近接した時点での調査結果は何等かの意味で似た傾向を示し易い、という感覚的な先験情報が利用できるからである。

ここではこうした先験情報を用いた系統誤差の一評価法を提案する。

2. モデル

データに関するモデルとして、

$$y_i(t) = q_i(t) + S_i(t) + e_i(t),$$

$$i=1, 2, \dots, M, t=1, 2, \dots, T,$$

$$\sum_{i=1}^M y_i(t) = \sum_{i=1}^M q_i(t) = 1, \text{ for } \forall t,$$

を仮定する。但し $y_i(t)$, $q_i(t)$, $S_i(t)$, $e_i(t)$ は各々、 t 時点での i 番目のカテゴリの、調査結果による占有率、トレンド成分、系統誤差、標本誤差を表わす。更に $S_i(t)$, $e_i(t)$ は各々、時点間では独立に、しかし各時点内ではある共分散構造を持って確率分布を為しているものとする。

一方、先験情報によるモデルとしては、

$$q_i(t) = 2q_i(t-1) - q_i(t-2) + f_i(t),$$

を仮定する。データの平滑化でよく用いられる仮定で

はある。但しこの場合、残差 $f_i(t)$ は $S_i(t)$ 等と同様時点間では独立であるが、各時点内では独立ではない。なぜならば、 $\sum_i f_i(t) = 0$ (for $\forall t$) でなければならないからである。従って $f_i(t)$ の分布では何等かの共分散構造を仮定しなければならない。一般にはそれは未知である。

数量化理論第 III 類と応用の接点

駒澤 勉

医学の計量診断への応用として、多変量解析、数量化理論のデータ解析法は、外的基準があるばあい、自動診断へ積極的に応用できる。これに対して外的基準のないばあいには、現象を表現している個体に関する数多くの特性量をデータとして、個体間、特性間および両者のクロスの関係で似たものの類例化を行う。この種の分析法は医師の診断過程に対応する計量的診断、特に新しい疾病鑑別のようにあまりわかっていない未知の世界に探りを入れるのにも有用である。

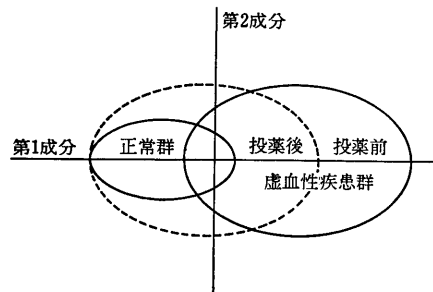
1. 成分的分析と判別・予測

主成分分析や数量化 III 類における個体の散布図は、既知のいくつかの群に布置できることがある。ここで判別だけを問題にするのであれば、この解析結果を参考に変数選択を行い、解析を判別分析でやり直せばよい。しかし、群別の定義にあいまいさが残るもの、新しい群と思われるものなどがあって、それらのものが既知の群とどんな類似性があるか等の点を成分的にそれぞれの布置情報から分析したばあいには、外的基準のない分析法の結果をそのまま利用して判別・予測するのも実際的である。分析基準量として、個体総合量の全分散 σ_T^2 、群間分散 σ_B^2 、群内分散 σ_W^2 、 $\sigma_T^2 = \sigma_B^2 + \sigma_W^2$ の関係において、潜在的にデータ群構造があるばあい、 $\eta^2 = \sigma_B^2 / \sigma_T^2$ が大になる傾向がある。この性質を利用した例として、図は心機能を評価する血圧、心拍、酸素消費量などの特性項目を用いて解析したときの個体総合量の布置である。重要成分、高精度での判別・予測ができることを分析した上で、次に同疾患群に心機能改善の薬を投与した後、再度同じ特性量を計測し、解析時のカテゴリー数量を用いて座標配置した布置が破線である。この図から使用した薬の効果は第 1 成分（酸素供給成分）を顕著に改善しているが、第 2 成分（末梢抵抗成分）を改善するにいたっていないことが推測される。

2. アイテム・カテゴリーの配列順序を決める分析

カテゴリーの配列順は一般に内容が似たものを相隣

るように配列し、順次コード化する。しかし、時には一応データ解析をした上でデータ構造の特徴を調べてから、以後恒久的に使用するカテゴリー配列のコード化を行う方がよいときがある。アイテム・カテゴリーに順序があるデータ構造では、数量化理論第 III 類の結果、固有値 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$ に対応する固有ベクトル x_1, x_2, x_3, \dots について $(x_1, x_2), (x_1, x_3) \dots$ の対の要素の座標配置が 2 次曲線状、3 次曲線状、…の形状になる性質がある。この性質を利用して、たとえば、最大血圧と最小血圧を組み合わせた血圧項目に対するカテゴリーの配列順を決めたり、治療法やリハビリなどの進め方の手順を決定する応用法がある。



図A 群別個体総合量の布置

ベクトル変換を用いた数量化法

馬場 康 維

統計的データ解析において、データの視覚化はデータ構造の把握のための極めて有用な手段である。単なる数値を見るよりもグラフを見る方が、データの構造を直観的に理解しやすい。このような観点から、性別、職業などの質的な特性、あるいは 20 代、30 代という年齢区分のようにカテゴリー化された特性から目的変数の値を予測するという問題を考える。これは従来の方法でいえば数量化理論第 I 類の対象とする問題である。

まず目的変数 y を角度に変換し長さがアイテム数 R に等しいベクトル

$$Y = R \exp(i\phi) \quad (\phi = G(y))$$

に対応させる。個体 α には Y_α 、 ϕ_α が対応するものとする。次に

$$X_{j\alpha} = \exp\left\{i \sum_k \theta_{jk} \delta_\alpha(jk)\right\}$$

とする。ここで $\delta_{\alpha}(jk)$ は個体 α が j アイテムの k カテゴリーに該当したとき 1, 該当しないとき 0 となる変数であり, $X_{j\alpha}$ は j アイテム k カテゴリーに対応する方向を向く単位ベクトルである。 Y_{α} の予測値(ベクトル)を

$$Y_{\alpha} = \sum_j w_j X_{j\alpha}$$

と定義する。係数 w_j , アイテム・カテゴリーに与えられる角度 θ_{jk} は残差平方和

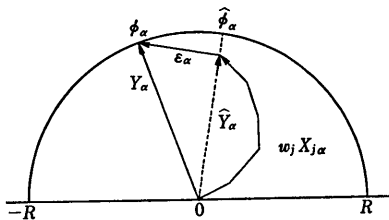
$$Q = \sum_{\alpha} |Y_{\alpha} - \hat{Y}_{\alpha}|^2$$

を最小にするように決める。 y の予測値 \hat{y} は合成ベクトル Y の角度 $\hat{\phi}$ から求める。即ち

$$\hat{y} = G^{-1}(\hat{\phi}) \quad (G^{-1} \text{ は } G \text{ の逆変換})$$

$w_j X_{j\alpha}$ を連結してできる径路によりアイテム・カテゴリーへの反応パターンを表わすことができる(図参照)。

詳細は、論文[馬場康維, 脇本和昌 (1983)「ベクトル変換を用いた数量化法」統計学集報第30巻2号]を参照されたい。



第5研究部

最小情報先験分布について

赤池弘次

ベイズ模型の実用を試みる時、ある段階まで先験分布の定式化が進んでもその先に不確定な部分が残ることが多い。このような場合によく利用される方法として、なるべく勝手な情報を導入しないような先験分布を、この不確定な部分に対して定義することがある。この方法に関する先駆的な業績として、H. Jeffreys の研究がある。

Jeffreys の研究の着想は、パラメータ空間の座標変換に対する先験分布の不変性に基礎を置くものであるから、その適用は連続パラメータに対する先験分布に限られる。これに対して、有限箇数の可能なデータ分布が与えられた場合に対する研究はほとんどなされて

いない。有限箇の分布のそれぞれに等確率を与えるような先験分布を用いることが妥当であろうなどと書かれている教科書も見られるが、その議論には何等の根拠も無い。

データ分布の全体を $\{f_k(\cdot)\} (k=1, 2, \dots, K)$ で表わし、現在と将来の観測値をそれぞれ x と y とで表わす。 k の先験確率を w_k とすると

$$p(y, x) = \sum_{k=1}^K f_k(y) f_k(x) w_k$$

$$p(y) = \sum_{k=1}^K f_k(y) w_k \quad p(x) = \sum_{k=1}^K f_k(x) w_k$$

によって y と x との同時分布、周辺分布が与えられる。最も強く情報を持込む先験分布は、 x の観測が y に関する推測に影響を持たなくなる様なものであるから、 $p(y, x) = p(y)p(x)$ が成立するであろう。そこで

$$I(w) = \iint p(y, x) \log \left(\frac{p(y, x)}{p(y)p(x)} \right) dy dx$$

によって $w = \{w_R\}$ によって定義される $p(y, x)$ の $p(y)p(x)$ からの離れ方を測り、これを最大にする w を最小情報先験分布と定義する。

最小情報先験分布は一意的に定まる。データ分布が互に完全に分離するときには等確率分布 $w_R = 1/K$ が最小情報先験分布となる。

最小情報先験分布の実用性には疑問がある。むしろ予測のリスクを考え

$$\text{Min}_{\{w\}} \text{Max}_k \int f_R(x) \int f_R(y) \log \left(\frac{f_R(y)}{p_w(y|x)} \right) dy dx$$

で与えられるミニマックス型先験分布を用いる方が自然のように思われる。ここに $p_w(y|x) = p(y, x)/p(x)$ である。この定義によれば互に近い分布の集りに対して、局所的に一様分布に近い先験分布が与えられる。最小情報先験分布では、互に極めて近いいくつかのデータ分布がある場合、その中のいくつかに対して先験確率として 0 が与えられてしまうのである。

ベイズ型大規模パラメトリックモデルの応用

石黒真木夫

I. 大規模パラメトリックモデル

統計的解析の目的はある現象に関するデータからその現象の特性を最大限かつ(再現性があるという意味で)安定的に抽出することにある。この目的を果すために導入されたのが、AIC によるモデル選択法であった。この方法は非常に強力であり、パラメータ数が

データ数の数分の1程度のモデルまで扱うことが可能になった。しかしながら現実の現象は限りなく複雑であり、さらに複雑なモデル、データ数の数倍のパラメータを含むモデルが要請される場面がある。たとえば経済時系列の季節調整がその典型的な例である。このようにデータ数と同程度あるいは、それ以上のパラメータを含むモデルを大規模パラメトリックモデルと呼ぶことにする。

II. ベイズ手法と ABIC

大規模パラメトリックをデータにあてはめるに際して、パラメータに事前分布を想定するベイズ手法が有効であることは昔から知られていたが事前分布をどうとるかが最大の難点であった。これを解決する一つの方法として、Akaike [文献省略] はパラメータを含む事前分布を想定してそのパラメータ (超パラメータと呼ぶ) をデータに合わせて調整する方法を提案した。超パラメータ選択の基準となるのが ABIC (Akaike's Bayesian Information Criterion) であり、モデル選択における AIC に相当する。

III. 応用のひろがり参考文献

1) 経済時系列の季節調整

[1] Akaike, Ishiguro (1980): Computer Science Monographs, No. 13

2) 地球潮汐データ解析

[2] Ishiguro et al. (1983): Proc. 9-th Intl. Symp. on Earth Tides

3) ベイズ型スプライン回帰

[3] 石黒, 荒畑 (1982): 統計研彙報, Vol. 29, No. 1

4) 2 値データ解析

[4] Ishiguro and Sakamoto (1983): Ann. Inst. Statist. Math., Vol. 35B

5) 密度関数の推定

[5] Ishiguro and Sakamoto (1982): Res. Memo, No. 239

6) 電波望遠鏡データの処理

翼の安定性の解析

濱田 義保

航空機の設計において、大気の流れに対する、翼の応答を調べることは、重要である。すなわち、翼の振動の安定性を調べる必要がある。一般に航空機の翼は、飛行速度が増加するに従い、気流の流れに対して大きな応答を示すようになる。速度がある臨界値を越えると、フラッターあるいはダイバージェンスと呼ばれる異常振動が生じ、翼の破壊が起こる。この臨界値を求めるために、航空機を翼が折れるまで飛ばして測定することが不可能なのは明白である。また、風洞実験でも、高価な模型を壊すことは、できる限り避けるべきである。

異常振動が生じる臨界値よりも十分小さい動圧で実施された風洞実験の結果から、臨界値を推定する方法について説明することが、本稿の目的である。ここで用いた実験データは、動圧をきわめてゆっくりと時間とともに大きくして得られた局所定常データであるので、本稿で述べる解析方法は、他の局所定常時系列データの解析にも用いることができる。

実験方法について説明する。風洞内に翼を固定し、2 点にゲージ計をとりつける。風洞内の動圧が定常になった後、動圧を $0.126 \text{ kg/cm}^2 \cdot S$ の割合で増加させて振動の変化をとらえる。図 1 に測定結果を示す。

解析方法の流れ図を図 2 に示す。データのチェックは視聴覚的情報検索システム¹⁾を用いて行う。定常性は、TIMSAC 74²⁾の中のプログラム NONST を用いて調べ、定常と考えても良いスパンの長さを求める。スムージングは、実験方法のために生じるトレンド及び振動の高周波成分を除去するためのバンドパス・フィルターを用いて行う。6 次の AR モデルをあてはめて解析を行った。この時、係数より作られる 5 次の行列式の値の変化より動圧の臨界値を求める。 3.91 kg/cm^2 の推定値を得たが、これは、実際の臨界値 3.92 kg/cm^2 に非常に近い。図 3 には、スペクトルの時間変化の様子を示す。なお、データは、航空宇宙技術研究所の安藤泰勝氏より提供して戴きました。

1) 濱田義保, 統計研彙報 29 巻 2 号

2) 赤池弘次他 TIMSAC 74

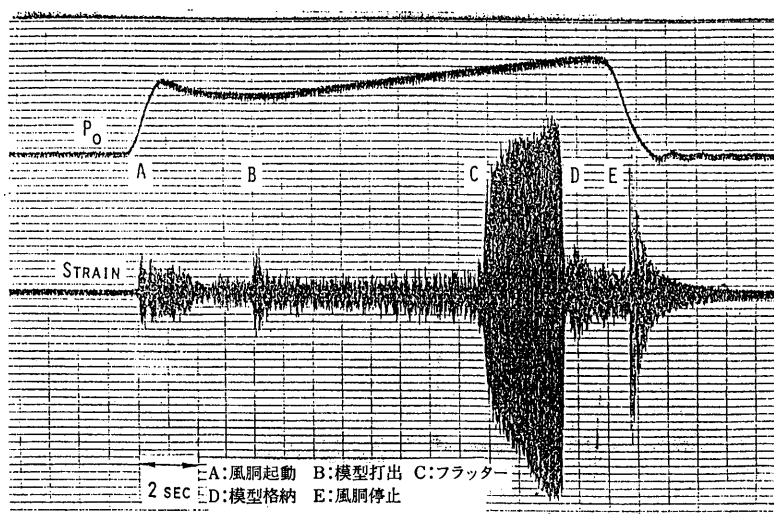


図 1

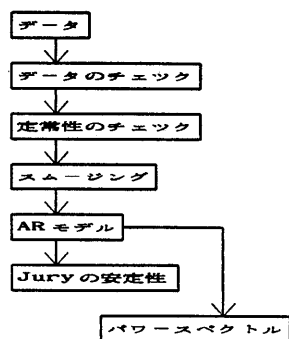


図 2

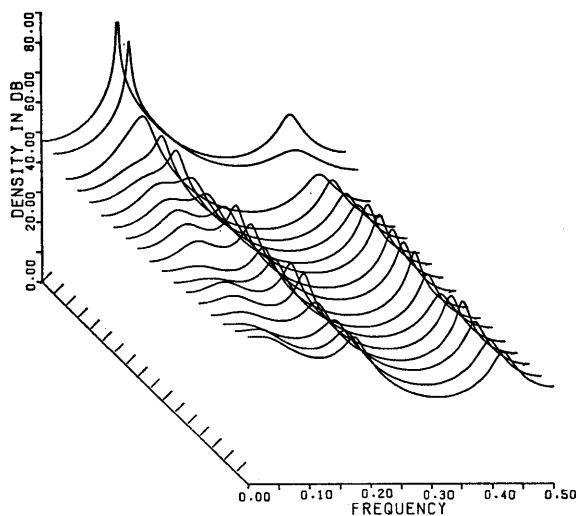


図 3

非線型時系列モデルとその応用

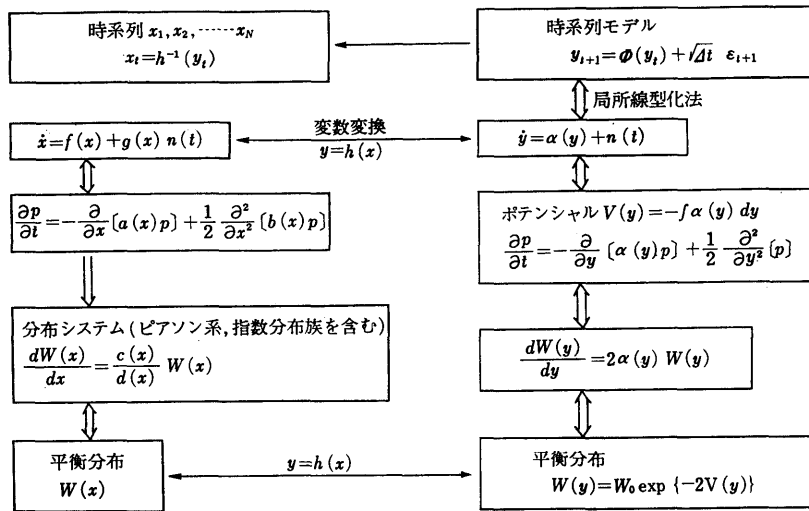
尾 崎 統

1. はじめに

工学、物理科学、生物科学などでよく用いられる数学モデルの一つに連続マルコフ過程（拡散過程）モデルがある。これと統計学の時系列解析で知られている

時系列モデルとの間に密接な関係があることが明らかとなった。連続マルコフ過程モデルは多くの場合、データ解析が問題とならない場面で用いられるが、なかには時系列の観測データが得られる場合もある。そのような場合にこの関係を用いれば、モデルのデータによる検証が可能となり、科学者が陥りがちな観念的モデル論争を回避し、より生産的なモデル論を展開していくことに役立つと思われる。

2. 関 係



3. 局所線型化法

$y_{t+\Delta t} = \Phi(y_t) + \sqrt{\Delta t} \varepsilon_{t+\Delta t}$: マルコフチェイン

$$\Phi(y_t) = y_t + \left\{ \frac{\partial \alpha(y)}{\partial y} \right\}_t^{-1} \left[\exp \left\{ \Delta t \left(\frac{\partial \alpha(y)}{\partial y} \right) \right\}_t - 1 \right] \alpha(y_t): \text{局所線型化法}$$

$$\Phi(y_t) = y_t + \Delta t \alpha(y_t) \quad : \text{Euler 法}$$

- i) 局所線型化法マルコフチェイン, Euler 法マルコフチェイン, Runge-Kutta 法マルコフチェイン

$$\xrightarrow{\Delta t \rightarrow 0} \dot{y} = \alpha(y) + n(t)$$

- ii) Δt を固定すると Euler 法マルコフチェイン, Runge-Kutta 法マルコフチェインは非定常で発散するが, 局所線型化法は定常マルコフチェインを定義する.

非定常時系列の分解

北 川 源四郎

$\{y_n\}$ は地震波の観測値とする. y_n は定常な常微動 (r_n) と信号 (S_n , 地震波) と観測ノイズ (w_n) の和と考えることができる. すなわち, 地震波のモデルとして

$$y_n = r_n + S_n + w_n$$

が考えられる. ここで, 次のような成分モデルを仮定する.

$$r_n = \sum_{i=1}^m a_i r_{n-i} + u_n, \quad u_n \sim N(0, \tau_1^2)$$

$$S_n = \sum_{i=1}^l b_i S_{n-i} + v_n, \quad v_n \sim N(0, \tau_2^2)$$

w_n は $N(0, \sigma^2)$ に従う白色雑音である. a_i, τ_1^2 および σ^2 は平常時の観測から既知と仮定できる. b_i はデータから推定するか, あるいは, いくつかの代表的な地震波のモデルの中から選択することになる. v_n の分散 τ_2^2 は未知で, しかも時間と共に変化するものと考えられる. ここで

$$F = \begin{pmatrix} a_1 a_2 \cdots a_m & & \\ & \ddots & \\ & & 1 \end{pmatrix}, \quad G = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$$

とおくと, y_n の状態空間表現

$$\begin{aligned} x_n &= F x_{n-1} + G \varepsilon_n \\ y_n &= H x_n + w_n \end{aligned}$$

が得られる. この状態空間表現と Kalman filter の利用によってモデルの尤度が計算できる. τ_2^2 は区分的には定数とみなすことにすると, 各区間で尤度の最大

化を行なうことにより時間とともに変化する τ_2^2 を推定することができる。

このようにして地震波のモデルが定められると、Smoother によって、各成分 r_n , S_n および w_n の最適な推定値を求めることができる。

[参考文献]

Kitagawa, G. (1983). State Space Modeling of Nonstationary Time Series and Smoothing of Unequally Spaced Data, *Proc. Symposium on Time Series Analysis of Irregularly Observed Data*, College Station, Texas, Feb. 10-13, 1983 (to appear).

第6研究部

調査データの二次分析と国際比較

鈴木 達三

すでに整理済みの調査データ（個票）が利用可能である場合、①新しい主題についての分析、②新しい分析手法を適用する等、広範囲の二次分析ができる。しかし、データを一定の形式で整理し、維持管理しているデータ資料館における調査データの蓄積が十分でないこと、目的に応じて適切な形式で調査データを整理、再編成して共通ファイルを作成するデータ処理システムが必ずしも容易には利用できないこと等のため、個票レベルのデータ利用を広範囲に実行するのは容易ではない。

これを補うものとして、各調査の集計表を基礎データとする二次的分析が考えられる。調査結果の基礎にある個票はすでに失われているか、保存されていても現在のデータ処理方式には不適である等の場合にも一定形式の集計表が存在することが多く、とくに単純な集計表ではその可能性が高い。主として利用できる素材は、単純集計表あるいはクロス集計表である。

このような調査データの二次分析は国際比較研究のデータ解析を進める上で有用な情報獲得手段となる。たとえば、①比較データを広くとり上げることが可能となり、現在計画している比較研究の対象相互間だけでなく、より広い社会の間の相互比較や大局的な位置づけができること、②比較の対象となった社会の環境の動向や、その社会に住む人びとの意識構造を動態的にとらえるため、それぞれの社会における継続調査データの活用を考え、これを調査計画にとり入れること等、比較研究の範囲が拡大し、豊かな情報が得られる。

<具体例>として、①国民性調査の「くらし方」に

関する質問項目について、13ヶ国価値意識調査(1979)の国別単純集計表と国民性調査の継続調査結果(1953-78)を合わせた分析により「くらし方」の質問の位置づけ、各国別の位置づけを大局的に掴むことができた。また、②国民性調査の「宗教を信じるか」の結果とアメリカ（ミシガン大）、オランダ（国勢調査）の「教会に行くか」の結果の各年齢別集計表を利用したコホート分析により、日本では年齢効果が、欧米ではコホート効果がより大きく見出され、各社会の宗教とのかかわり合い方の異なることが明らかにされた。

領域クラスタリングによる地域区分について

大 隅 昇

地域情報に関連したデータを用いて地図上のある地域を類似の領域に区分する方法として領域クラスタリングとその関連手法の研究を進めた（環境庁公害防止等試験研究費「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」の一環）。地域を区分するといっても、利用するデータは分析目的により多種多様である。ここでは主に、地理的情報、地域環境関連データ「国土・国勢数値情報など」、住民意識調査データなどを前提としたデータ解析への適用を考えている。

これらのデータが平面上の量子化された格子点（具体的には画素）の上に多変量特性データとして与えられる場合を考えて、これらをもとにして地域区分を行いその結果を色情報として図的表示することが当面の目標である。この種のデータとして良く知られたものにリモートセンシング・データやメッシュ・データがある。一方、住民意識調査データを初め多くの統計調査データは測定位置（又は地域代表点）が必ずしも規則的ではなく特定な位置（点又は面）の上でだけデータが与えられる。いずれの場合も、限られた一部の情報から分析対象地域の全体像を大まかなイメージとして把握することになる。しかも単なる多変量特性データによる分類に留まらず地域のもつ位置や地域的つながりなどの情報まで取り入れた地域区分でなければならない。これを効果的に行うにはそれを表現する道具も重要である。点の上に与えられた情報を面の上の情報に変え、しかもそれが多次元的であるということまで考えるとカラーグラフィックスを用いた色表示がきわめて効果的であろう。

試作を進めた領域クラスタリングとその関連手法の

各種プログラムは上記の点を配慮したもので、(1) 地域区分の結果を連続的な色調変化を伴う色のイメージとして表示する、(2) 観測点（又は小区域の集まり）上のデータから平滑化処理により面の情報として表示する、(3) 対象領域を大まかに分類したのち特定関心領域を抽出拡大して再区分する、(4) 地域区分で得た色情報の上に既存の情報（地図、地形など）をオーバーレイ表示する、(5) 色情報の編集機能を活用して即応的に地域の特徴を観察できる、などの特徴がある。これらの諸機能を効果的に利用するために地図作成、各種グラフィックス処理等の支援プログラムもあわせて開発した。今後は、実際の調査データについて解析を進め、同時に試用の結果みられた問題点（地域代表点の適切なサンプリング方式、区分結果の計量的評価の方法、色感と配色の関係など）についての検討を進める。

意識の測定と分析の基礎的問題

水 野 欽 司

1. 近年、諸方面で行われている意識調査の特徴として、(1) 調査で求めようとする情報が、把握のむずかしい内容（たとえば、防災態度、環境意識など）へと拡張されつつある、(2) 施策的な目的が背景にあり、調査で得る情報を、別質のデータと関連づけて利用しようとする、などの傾向がみられる。いきおい、調査は多様な対象層や多元的な内容を同時にカバーせざるを得なくなるが、他方で、実際に行使しうる調査手段には、質問内容の難易度や分量に自ずから制限が課される。加えて、調査対象者の協力が得にくいという事態や、プライバシー問題のため、調査の手法や内容が止むを得ず制約されるという状況が強まっている。このような現状においては、意識調査そのものの機能を、改めて見直す必要に迫られているといっても過言ではない。

2. 本年度に実施した、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）は、典型的にこれらの問題を抱えた調査であり、ここでは、まず問題点を具体的に検討するための試験的データを得ることを図った。対象地域（および対象人数）は、千里ニュータウン（1,800人）と千葉市市街部（1,440人）で、方法は留置法による。結果として、住民の地域全体への総合評価や各環境要素の評価に関して、同義的内容の質問を多く組入れた場合、内的整合性の意味で信頼性のある質問項目群が認められる一方、不適当とみられる質問内容や質問形式を見出すこ

とができた。また、回答者が各環境要素をどの範囲、どの場面で意識するかによって、回答傾向のちがいが大きいことが認められた。結局、この種の調査データから、内容的に粗くても妥当かつ信頼できる情報入手するためには、事前における回答者および回答の両面にわたる慎重なクラスタリングが必要であると考えられる。加えて、複数の類似質問への反応に基づく部分的な尺度化により、それら尺度値間の関連性あるいはパターンを調べるのが効果的とみられる。同様なことは、無視し得ない無回答バイアス（有効回収率は65%、50%）の処理問題にも当てはまるといえる。回収率は、明らかに環境条件、居住条件の良否による差が認められ、これを無視して地域全体の評価は成り立たない。問題は、どのような操作手順でこれを行うかにあるが、現在なお試験的な検討と細部の考察を続行中であり、その結果に基づいて再び同様の調査を実施し、検討を進める予定である。

コウホート分析について

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、コウホート分析には識別問題が存在し、3効果を分離することは原理的に不可能とされていた。

この識別問題を克服するために、パラメータの漸進的变化の条件を取り込んだベイズ型コウホート・モデルを開発した。さらに、データの形式や特性に対応するようモデルを拡張し、何種類かのデータの解析をつづけている。

解析対象の第1は、国民性調査データであり、これは年齢区分幅と調査間隔が一致する標準コウホート表形式の二項比率型データである。第2は、同じく標準コウホート表形式であるが、比率型ではなく、数量型のデータである。前者は二項分布をもとにしたモデル、後者は正規分布をもとにしたモデルで対応できる。

ところで現時点において標準コウホート表が得られるのは、まれなことといつてよい。毎年実施している調査も、蓄積は10年を越える程度であり、年齢区分が不規則・不統一であったりする。また、調査間隔6年、年齢区分幅5歳といった調査もある。

年齢区分幅と調査間隔が一致しないこれらの一般コウホート表に対しては、オーバーラップするコウホー

ト区分を想定することによって、標準表の場合とほとんど同様に解析をすすめることができる。

さて、過去に遡って調査することはできないが、内容によっては、調査対象者に過去を振り返ってもらう回顧調査が可能である。得られるのは、不完備なコウホート×年齢という特殊なデータ表となる。この場合も、デザイン行列を工夫することによって、解析が可能である。

以上まとめれば、ベイズ型コウホート・モデルは次のような形式・特性のデータに対応することになる。

現在、APL プログラムとして利用可能である。

「ベイズ型コウホート・モデル」	
<データ形式>	<データ特性>
標準表	比率型
一般表	数量型
特殊表	

附属統計技術員養成所

情報量統計学

坂 元 慶 行

統計学の目的は、観測されたデータに基づいて、不確実な現象の特性を確率によって表現し、将来の観測値の確率分布を推定し、予測や制御に資することにある。この目的を実現するためには、①データの特性を表現する確率分布（統計モデル）を、現実が生じるさまざまな分析目的に応じた的確に構成すること、②考え得るいくつかのモデルの良さを評価・比較しうる規準を提示すること、が不可欠である。

従来の統計的仮説検定は想定された数多くのモデルの評価・比較という第2の要請に対して余りにも無力であったばかりでなく、対立仮説が明示されず、そのためどのような仮説との対比において当該の仮説の当否が吟味されたのかさえ不明瞭な場合もあった。また、想定した仮説の当否を吟味するためには、その仮説に対応する標本分布を導く必要があり、それが統計的手法の実用性を著しく阻害していた。そこで、従来、推定や検定、あるいはデータの記述法として個々バラバラにとりあつたかわれてきた数理統計学の諸問題を、①個々の問題に対するモデルの構成、②情報量規準 AIC によるモデルの評価、という一貫した視点から見直すことによって、数理統計学の再構成を試み、[1] を公刊した。

一例を挙げると、従来統計データの情報縮約のための記述統計的な手法としてとりあつたかわれてきたヒス

トグラムも上の視点から見直すことが可能で、この視点からその自動描画のための方法を与えた。さらに、この考えを発展させ、ヒストグラムのもつ効率の悪さを改善するため、ベイズモデルを導入して、密度が滑かに変動する限り、任意の母集団分布の推定に有効な方法を[2]で与えた。

また、継続的に研究している分割表データの解析法に関しても、同様の視点から、説明変数がカテゴリカルな場合のみならず連続的な場合にも効率的な推定を得るための方法を考え、[3]で与えた。

【参考文献】

- [1] 坂元慶行・石黒真木夫・北川源四郎 (1983). 情報量統計学, 共立出版.
- [2] M. Ishiguro and Y. Sakamoto (1982). A Bayesian approach to probability density estimation, *Res. Memo.*, **239**.
- [3] M. Ishiguro and Y. Sakamoto (1983). A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, 115-137.

特別研究

広域的森林の動態とその管理に関する統計的研究

石 田 正 次

1. 研究の目的

この研究の目的は森林の生態系、地域特性、住民の意見を統計的に調査し、これを基礎データとして環境因子としての森林の有効配置とその管理方式を考察することである。研究は高田和彦、小林正吾(新潟大学)、北村昌美、今永正明(山形大学)、菅原聡(信州大学)の協力のもとにそれぞれ特徴を異にする次の三つの町をモデル地区として実施した。

1) 櫛引町(山形県)

大規模パイロット事業による地域開発地区。
森林の転用と今後の森林配置の問題。

2) 軽井沢町(長野県)

観光開発地域。別荘地分譲による自然環境の変化と保全の問題。

3) 大潟町(新潟県)

工場の誘致に伴う生活環境林の保続の問題。

2. 基礎調査

まず文献と聞き込みによって地域の特性と歴史を調べた。次に撮影年次の異なった航空写真(主に昭和38年、46年、55年の3回を利用)によって土地利用、林

相の変化を判読し、その結果にもとづいて宅地、水田、畑、果樹園、草地、森林を地形図上に区画した。更にこの土地利用図から森林プロットを抽出し森林生態系の現地調査を行った。住民の意識は各地区別に質問票を作成、有権者名簿から抽出した標本について面接及び郵送法によって調査した。

3. 櫛引町

この地域の農地ははじめ集落を中心に森林内に点在していたが時とともに次第にその面積を増大した。但し、江戸時代からの部落ではその集団に古くからかなりの面積の水田を保有している。昭和46年から55年の森林の減少は38年から46年に比較してはるかに少い。これは転換可能な場所がほぼ利用つくした上に、開発事業の経営不振によるものである。

過去における林地の転用に対して住民の54%が反対している。開発事業に関係の薄い層が特に反対の態度をとる傾向がみられる。これらの人々はかなり観念的な自然保護の考えを持っている。一方、開発事業の失敗を直接経験した層にも反対が多い。大面積皆伐によって荒廃した裸地の影響は表土の流出、流れの変化、保水機能の低下となって現れている。蓄産一辺倒による画一的な草地化に大きな問題があったと同時に森林の配置や保全について理論的な根拠を欠きその具体策を提示し得なかったことも責められるべきである。

4. 軽井沢町

牧場や農場の開発が目的で明治に国から払い下げられた土地が大正から昭和にかけて避暑地に変わり、これが更に観光地化してきた理由には軽井沢の宿場町としての特性、地域住民の土地に対する愛着心の低さ、大資本の介入などが考えられる。

軽井沢の森林は生活空間の中に広く分布していることが特徴で、森林面積の減少も開発の規模からすれば比較的小さい。地元の自然環境に対する意識は経済が先行する傾向が強い。別荘開発と土地所有形態の変化によって森林は細分化され、立木は過熟化している。このことは近い将来軽井沢の自然が大きく変化することを暗示する。

5. 大潟町

工場、宅地間に黒松を主とした森林が適度に配置されており、住民の「みどり」に関する満足度も高い。現在の森林をそのまま残すことを76%の住民が望んでいる。工場誘致と森林の保続のバランスを経済と環境保全から決定できる数少ない町である。

6. 結 論

土地利用、特に森林の配置についての住民の態度は

これと関係の薄い層ほど観念的であり、空想的でさえある。町当局、農林業関係者の判断は経済が優先し、住民の判断は住居に関する局所的な利害に左右される傾向がある。

森林を生活環境の一部として考え、そのありかたに関して具体的な意見を提示することは一般住民にも為政者にもかなり困難なことのように思われる。故に具体的な自然環境問題の解決法を単純に住民意見に求めることは危険であり、科学的と思われる施策を科学的という理由だけで住民に強要することにも無理がある。

世論調査による同時代史

西 平 重 喜

このたび(1983年4月1日)、私はこの研究所を辞職することになりました。ちょうど、1/4世紀をすぎました。研究所に入る1年前に、「日本人の読み書き能力調査」の結果のとりまとめに参加したのが、私のはじめての統計学とのコンタクトでした。そのときはこの調査を後援していた、連合軍総司令部民間情報教育局から給与をもらい、林知己夫氏の下でこの仕事のお手伝いをしました。「読み書き」の仕事で、はじめて与えられたことは、調査不能の影響の検証で、そのためにIBMカードの穴を肉眼で読みとり、コントロール・グループのデータとさしかえることでした。——IBMの機械は連合軍総司令部にあり、かんたんに頼める状況ではなかったのです。そして「読み書き」の関係者の多くが、後に国立国語研究所に入ったので、林、石田正次氏とともに言語学の社会心理学的研究に参加し、各種の調査方法を体験することができました。

「読み書き」の仕事が終わったあとは、同じ身分で、水野坦氏について仕事することになり、そのうちに研究所に採用されました。水野氏は当時、日本に世論調査を普及させることに熱心でした。まずはじめに木村等氏とともに、NHKのラジオ聴取者調査のためのサンプリング・デザインをおこないました。このデザインは「読み書き」のサンプリングとともに、その後の日本の社会調査、世論調査の雛型になったといえると思います。なお、NHKでは放送用語の理解度の研究もおこないました。

1949年には世論調査の妥当性を研究するため、水野、林氏等と港区長選挙の調査をおこない、さらに林、青山博次郎氏などに質問文の研究をしました。1953年からは林氏を中心に日本人の国民性の調査を5

年おきにおこなってきました。この間、5年おきの本調査のほか、サンプリング、質問文、調査法、データの分析のためにおこなった各種の付帯調査は、今日の世論調査の基礎をきづくものといえるでしょう。日本では世論調査や社会調査は多いけれど、そして最近では分析方法についての研究は進んだけど、調査法上の基礎研究がなおざりにされていることは、今後の調査データに心配を感じさせます。

国民性の調査と平行して、社会学関係——階層と移動、従業員のモラル——の研究にも参加しました。1954年3月からは林氏等と東京23区内で、年2回づつ——1975年以降1回——世論調査を実施してきました。この調査も、上記の基礎研究に多くの機会を提供しました。マス・コミや官公庁の世論調査は、時のホットな問題を一時的にしか扱わないが、この調査では世論の推移を研究する面で、ユニークなものといえます。

以上のような経験を積みかさねているうちに、民主主義社会では、歴史も世論という見地からの研究が必要ではないか、と考えるようになりました。欧米では歴史学者やジャーナリストで、「世論」という概念を導入した記述をしている人がめずらしくはありません。これに関連して、「世論」という概念の歴史も気になりましたが、この研究は長い時間が必要です。

そして世論調査により、同時代史を書くことを試み、10余りの論文を発表しました。その論文はアメリカの議会の対日問題公聴会でもとりあげられました。同じような見地から、欧米の世論調査から、宗教、教育などについて、彼等がどう考えているか、についてもとりまとめました。年度発表会では世論調査が本格化した1935年から太平洋戦争が始まるまでのアメリカ、イギリス、フランスなどの調査データについて述べましたが、機会をみてとりまとめたいと思っています。

先きののべた港区長選挙の調査以後、東京都知事選挙についても、水野、林氏等と研究し、さらに新聞社の選挙予測に協力してきました。これに関連して選挙の結果の分析が必要となり、さらに諸外国の選挙の分析、比較をするようになりました。その結果、日本でも比例代表制を採用すべきだと考え、具体的な提案をすることができました。

以上が、私がこの研究所でやってきたことの概略です。よき先輩と、よき関係者にめぐまれ、また世間一般の世論調査へ好意的な時期に、ここで研究できたことを幸せだと思っています。

非線型最適化法設計の新しい視点

田 辺 國 士

複雑で一見把えどころのない事象を、最適化過程あるいはその結果とみなすことにより理解しようという考えは、経験科学の諸分野において広く用いられている方法である。たとえば、自然法則の多くは常・偏微分方程式で表現されるが、これらはエネルギーやエントロピーなどの汎函数を最小（大）化する立場から導かれる。ミクロレベルの方程式をたてることが難しい確率的問題やある種の非線型問題の場合にも、理工学分野ではマクロな汎函数を定義することにより近似モデルを間接的に構成して問題を処理することが広く行なわれている。経営や管理の領域において漠然と与えられた問題を「輸送型モデル」とか「配分型モデル」などの最適化モデルに定式化することにより、経営・管理のポイントがより良く理解できるようになることもある。また生物の進化や社会的行動を、ある種の最適化過程として理解しようという試みもある。このように最適化（Optimization）は、数学的モデル構築（Mathematical Modeling）のための普遍的視点を与える認識論的方法概念である。

一方、線型計画法（Linear Programming）の例に見られるように、最適化法はきわめて実践的技術である。実際、LPはエネルギーシステムの解析、生産管理、資源配分、輸送計画などにおいて日常的に用いられており、高度に組織された今日の産業社会の運営に不可欠の手段となっている。また、工学における最適設計の分野では、制約条件の下で性能が最高になるように設計パラメータの値を決める最適化技術が広く利用されている。科学研究の分野においても、数学的モデルに含まれる諸定数の値を最適化法によって決定することがしばしば行われる。特に統計解析において用いられる確率的モデルは、最尤法を介して最適化モデルとなることが多く、パラメータ推定や予測に最適化法が必須となる。これらの最適値を具体的に構成する方法（アルゴリズム）は、コンピュータの発展と相まって、過去30年間に飛躍的な発展をとげており、近年、ますます大規模なモデルを最適化するアルゴリズムが期待されている。

Optimizationの認識論的側面としての最適化モデル（Optimization Model）とその操作技術としての最適化法（Optimization Algorithm）は、数理計画（Mathematical Programming）を構成する二大要素であり、両者はdialektischな関係にある。世間では「数理計画＝最適化法」と理解されているようであるが、数理計画の

発展の歴史をふりかえると、モデリングの技術にこそその本質があることがわかる。実際、最適化モデルをたてるだけでも現実の事象の概念的把握に役立つことがあり、最適パラメータ値を決めることは二の次ということも多い。一般に、指定されるモデルに現実性があるか否かは、それを操作する最適化法の有無によって決まる。もし想定された最適化モデルが、複雑かつ大規模すぎて最新の最適化法をもってしても解けないようなら、モデル自体を改変し単純化することが必要になる。逆に、複雑な最適化モデルは新しい最適化法の開発を促す。いずれにしても、意識するか否かにかかわらず、モデル構築はモデルの精密さと最適化の容易さとのバランスの上に行なわれている。モデルは逐次改良によって段階的に発展するものであるから、このバランスは常に変化・発展してやまない。

非線型実数値関数 $f(x)$ の最小値を与える $x \in R^n$ を求める‘制約なし最適化問題 Unconstrained Optimization Problem’や非線型方程式および不等式で表現される制約条件、 $g_1(x)=0, \dots, g_r(x)=0, g_{r+1}(x) \geq 0, \dots, g_m(x) \geq 0$ 、の下で目的関数 $f(x)$ の最小値を与える x を求める‘制約つき最適化問題 Constrained Optimization Problem’の数値解法は、非線型最適化法 (Nonlinear Optimization) あるいは**非線型計画法** (Nonlinear Programming) と呼ばれ、ここ 20 年の間に主として数値解析やオペレーションズ・リサーチの分野で研究が急速に進歩した。

もともと数理計画の分野は、線型計画を核として形成され発展しており、非線型計画法も LP の延長上に位置づけられ、これを拡張する立場から凸計画法の理論や非線型双対理論が展開されてきた。しかし、非線型計画問題の最適解に関するこれらの定性的理論の細緻に亘る展開にもかかわらず、数値解を具体的に構成するアルゴリズムの進歩には大きな影響を与えなかったように見える。

非線型計画法アルゴリズムにおいては、最適化問題

の組合せ論的側面の処理ばかりでなく、非線型方程式を能率よく解くという問題をも解決しなければならぬ。近年この分野において数値解析畑の人々の活躍が目立つのもこの理由からである。非線型最適化問題の幾何は線型計画や凸計画に比べて相当こみ入っており、局所的・解析的アプローチでは問題の解決はおぼつかない。問題の大域的幾何に正面から取り組む接近法が必要である。1960~70 年代を通じて罰金法 (Penalty Methods) や乗数法 (Multiplier Methods) が制約条件つき最適化法として流行したが、制約条件を目的関数に繰り入れて非制約最適化問題群に持ち込むこれらの方法に、筆者は上の観点から疑問を感じていた。これら拡大関数法 (Augmented Function Methods) と呼ばれる方法においては、数値解析や幾何学的配慮が乏しく、そのアルゴリズムは不自然なまでに多様であり単純性に欠けていたからである。筆者は、

1. Penalty Function などに対する古典的 **Lag-range** 関数の自然さ
2. 非線型計画法における幾何学的観点導入の必要性
3. 恣意的になりがちなアルゴリズム設計の統制原理としての **Newton-Raphson** 法の重要性

の認識の下に、制約つき非線型最適化法を理論的に設計する統一的視点を与え、性能が異なるいくつかのアルゴリズムを開発した。この最適化技術を背景として、偏微分方程式モデルを、変分法などを介してノンパラメトリックな**大規模ベイズモデル**化することが可能となり、統計的モデリングの領域に新しい方法論を築くことができる。これにより工学において大きな成功を収めている有限要素法技術 (数値解析) と確率論取り扱いの技術 (統計学) を総合した実用性の高い解析法が得られる。従来主としてパラメトリックなモデルによる推論に基礎を置いてきた統計的解析法に新しい次元を拓くものと思われる。