

# ベイズモデルとABIC

田辺 國士

偶然に支配される現象やあいまいさを含んだ対象に関するデータから、その内に潜む「法則」や「構造」にかかわる情報をひき出すことは経験科学において日常的に求められる仕事である。このような非決定論的データを処理する方法にはいろいろなものがあるが、統計的方法、すなわち、データの発生機構を近似する確率モデル族を構成(統計的モデリング)し、この族の中からデータに最も適合するものを選ぶ(適合化)という過程を通して行なうことが最も自然であろう。確率モデルを用いる方法は、他のad hocな処理法に比べて、データ解析の結果の意味を吟味するうえでも、モデル族の妥当性と適合の度合いをはかる標準の良し悪しを客観的に検討するうえでも好ましい方法である。しかし、従来の統計的解析法においては、モデル族をパラメトリックモデルの形に表現することが要求され、ひとたび特定のパラメトリックモデル族が選ばれると、それをアプリアリに確定したものとして絶対視してしまう傾向があった。たとえばデータの平滑化において、多項式や三角関数を回帰関数に選んで“あてはめ”が行なわれることがよくある。しかし一般に、推定される関数がこれらの回帰関数でうまく表現できる保障はない。関数がなめらかであるということと、それを多項式や三角関数によってパラメトリックな形に表現することのあいだには、大きな論理の飛躍が

ある。実際、これらのパラメトリックモデルは、それぞれ固有の“くせ”をもっており、データにあてはめたときに元の「構造」を反映しない偏りが出てしまうことが多い。

一般に、統計的モデリングの根拠となる予備知識は、この平滑化の例のように漠然としたものであることが多く、措定されるべき確率モデル族は一意的には定まらない。統計的モデリングはきわめて主観的作業であり、教科書の例題を解くような具合にはいかない。したがって、この作業におけるモデル族設定のずれや誤りを許容し、それを補償する統計的方法が必要であることは明らかであろう。この必要に応じるものとして、確率モデル族を重層的に構成するベイズモデルによる方法がある。先の例でいえば、回帰関数をパラメトリックモデルの形に特定してしまわないで、回帰関数自身をさらに確率的に規定する方法である。こうすることにより、対象の「構造」に柔軟に追従できる可塑的なモデル族が構成され、従来の方法では考えられなかった統計解析が可能になる。本稿の目的は、データの平滑化を例にとり、この新しいデータ解析法を紹介することにある。

## 1. ベイズモデル

有限個の点  $\{x_i\} (i=1, 2, \dots, N)$  上で、あるなめらかな関数を観測してデータ  $\{y_i\} (i=1, 2, \dots, N)$  を得たとする。このデータを平滑化して元の関数を推定する問題を考えよう。話を簡単にすると

たなべ くに お 統計数理研究所

め、有界領域を等間隔の十分細かいメッシュ  $\{I_j\}$ , ( $j=1, 2, \dots, m$ )に刻み、各メッシュ  $I_j$  上で定数値  $f_j$  をとる区分的定数関数で元の関数を近似することにする。  $m$  をデータ数  $N$  に比べて十分大きくとれば実用上問題はない。以下、この近似関数を列ベクトル  $f=(f_1, f_2, \dots, f_m)^t$  で表現する。  $f$  が 2 変数のときには、長方形領域を十分細かい正方形メッシュ  $\{I_{ij}\}$ , ( $i=1, \dots, m_1, j=1, \dots, m_2$ )に刻み、  $I_{ij}$  上で定数値  $f_{ij}$  をとる関数  $f=(f_{ij})$  で近似する。

メッシュ  $I_j$  に含まれる観測点  $x_i$  上のデータ  $y_i$  は誤差  $\varepsilon_i$  を含んだ  $f_j$  の観測値であるとする。すなわち  $y_i=f_j+\varepsilon_i$  とし  $\varepsilon_i$  ( $i=1, \dots, N$ )は独立に平均 0, 未知分散  $\sigma^2$  の正規分布にしたがうものと仮定する。いま、  $N \times m$  行列  $E=(e_{ij})$  を  $e_{ij}=1$  (観測点  $x_i$  がメッシュ  $I_j$  に属するとき),  $=0$  (その他のとき) と定義すると、データベクトル  $y=(y_1, \dots, y_N)^t$  の同時密度関数は

$$(1) \quad p(y|f, \sigma^2) \equiv (1/2\pi\sigma^2)^{N/2} \times \exp(-\|y-Ef\|^2/2\sigma^2)$$

と書ける。ただし  $\|x\|=\sqrt{x^t x}$  とする。

従来の平滑法は、たとえば  $f_k$  を多項式

$$(2) \quad f_k = \gamma_2 k^2 + \gamma_1 k + \gamma_0$$

などとパラメトライズし、最小 2 乗法などを用いてパラメータ  $\gamma_2, \gamma_1, \gamma_0$  を決めることにより行なわれる。しかし先にも述べたように、(2) のような決定論的なモデリングに十分な根拠があることは稀である。そこで、  $f$  がなめらかであるということ、次のように確率的に表現する。

$$(3) \quad \nabla^2 f_k = \eta_k, \quad (k=p+1, \dots, m)$$

ただし、  $\nabla$  は差分作用素で、たとえば  $\nabla f_k \equiv f_k - f_{k-1}$  で定義されるものとし、  $\eta_k$  は平均 0, 未知分散  $\sigma_f^2$  の正規分布に独立にしたがうものとする。(2) のモデルは、(3) のモデルにおいて  $p=3, \sigma_f^2 \rightarrow 0$  とした特別な場合と考えることができる。

以下、一変数のモデルとしては、  $p=2$ , すなわち

$$(4) \quad \nabla^2 f_k \equiv f_{k-2} - 2f_{k-1} + f_k = \eta_k$$

を考えることにする。このとき、  $f$  の尤度は

$$(5) \quad \pi(f|\sigma_f^2) \equiv (1/2\pi\sigma_f^2)^{l/2} \varphi \exp(-\|Df\|^2/\sigma_f^2)$$

ただし  $l=m-2, D$  は  $l \times m$  行列

$$(6) \quad D \equiv \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & & \dots & & \\ & & & & 1 & -2 & 1 \end{pmatrix}$$

$\varphi$  は  $D^t D$  の非零固有値の積の平方根とする。

$\pi(f|\sigma_f^2)$  を  $f$  の事前分布と考え、  $y$  の密度関数を  $p(y|f, \sigma^2)$  とすると、観測値  $y$  が与えられたもとの  $f$  の事後分布は、ベイズの定理により、

$$(7) \quad p(f|y, \sigma^2, \sigma_f^2) = \frac{p(y|f, \sigma^2)\pi(f|\sigma_f^2)}{\int p(y|f, \sigma^2)\pi(f|\sigma_f^2)df}$$

で与えられる。ここで事前分布  $\pi$  は、  $D^t D$  が非正則であるため、  $f$  で積分すると有限ではないので通常の密度関数とはならない。このような事前分布は変則(improper)と呼ばれる。事後分布(7)にもとづいて  $f$  (および  $\sigma^2, \sigma_f^2$ ) の推定値を得る方法は一般にベイズの方法と呼ばれる。

## 2. 情報量規準 ABIC

$f, \sigma^2, \sigma_f^2$  の推定値として、事後分布の平均やモードを用いることもできるが、赤池 [1] は  $f$  の安定した抵抗値を得るために、次のような 2 段階から成る推定法を提案した。まず、  $\sigma^2, \sigma_f^2$  に関する「ベイズモデルの尤度」を

$$(8) \quad L(\sigma^2, \sigma_f^2) \equiv \int p(y|f, \sigma^2)\pi(f|\sigma_f^2)df$$

で定義し、統計量

$$(9) \quad \text{ABIC} \equiv -2 \log L(\sigma^2, \sigma_f^2)$$

を最小化する  $\sigma^2 = \hat{\sigma}^2, \sigma_f^2 = \hat{\sigma}_f^2$  を求める。次に、事後尤度  $p(f|y, \hat{\sigma}^2, \hat{\sigma}_f^2)$  を最大にする  $f = \hat{f}$  を求める。

一見なんの変哲もないこの方法は、実は非常に強力な方法で、従来取り扱いが困難であったさまざまな実用上の問題に適用され、興味深い成果を生みつつある。

$\alpha \equiv \sigma/\sigma_f$  を固定するとき、ABIC を最小化する  $\sigma^2 = \hat{\sigma}^2$  は

$$(10) \quad \hat{\sigma}^2 = R^2(f_\alpha^*) / (N + l - m)$$

で与えられる。ただし、 $f_{\alpha^*}$  は  $f$  の関数

$$(11) R^2(f) \equiv \|y - Ef\|^2 + \alpha^2 \|Df\|^2$$

を最小にする、線形最小 2 乗問題の解である。したがって、ABIC は  $\alpha$  の関数となり

$$(12) \text{ABIC}(\alpha) = (N+l-m) \log R^2(f_{\alpha^*}) - 2l \log \alpha + \log \det(E^t E + \alpha^2 D^t D) + C$$

と書ける。ただし  $C$  は  $\alpha$  に無関係な定数である。

(12) を最小にする  $\alpha = \hat{\alpha}$  は解析的には求まらないので数値計算に頼らざるをえないが、高精度で求める必要はない。(11) の最小 2 乗問題を QR 分解法で解くならば、その過程で (12) 式の  $\det$  の中味が副次的に計算される [2]。こうして求められた  $\hat{\alpha}$  と  $\hat{\sigma}^2$  から  $\hat{\sigma}_f^2$  が得られる。ABIC 最小化法は、

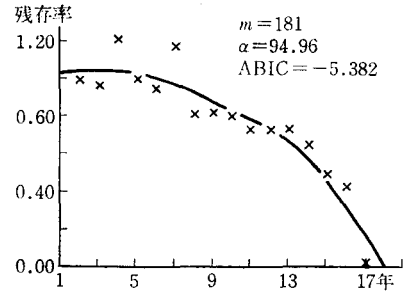


図 1

データにもとづいてモデルの分散  $\sigma_f^2$  と観測ノイズの分散  $\sigma^2$  の按配を決める方法であると考えられる。  $\hat{\sigma}$  と  $\hat{\alpha}$  を定めた後、 $p(f|y, \hat{\sigma}^2, \hat{\sigma}_f^2)$  を最大にする  $f = \hat{f}$  を求めるのであるが、実は  $\hat{f} = f_{\hat{\alpha}^*}$  であり、 $\hat{\alpha}$  を求める過程ですでに計算されている。

図 1 はビール箱の市場残存率を表わす  $N=16$  個のデータに上記の方法を適用して得たグラフである。×印はデータ、実線は推定曲線で、実際は階段関数であるが  $m$  が非常に大きいため ( $m=181$ ) なめらかに見える。ちなみに、 $\alpha$  が最小値  $\hat{\alpha}$  と異なるときに、それに対応する  $f_{\alpha^*}$  を図 2 に示す。式 (11) からわかるように、 $\alpha$  が大きいと直線回帰に近づき、 $\alpha$  が 0 に近いと  $f_{\alpha^*}$  はデータを補間する。図 3 は  $m=37$  と 381 の場合の ABIC による最適推定値を示したもので、ABIC 最小化法で得られるプロファイルは  $m$  にあまり依存しないことがわかる。

2 変数の場合にも同様の方法を適用することができる。2 変数の関数のなめらかさをモデリングするにも、いろいろな事前分布が考えられるが、図 4 で示

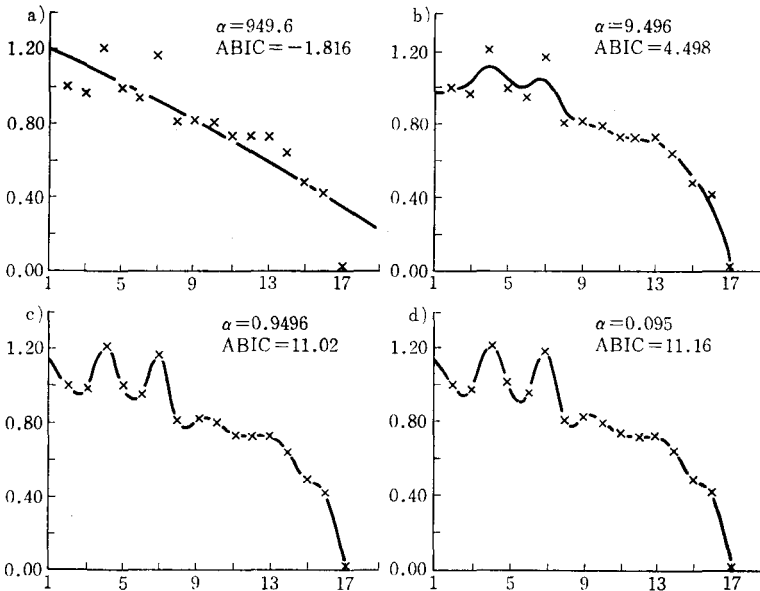


図 2

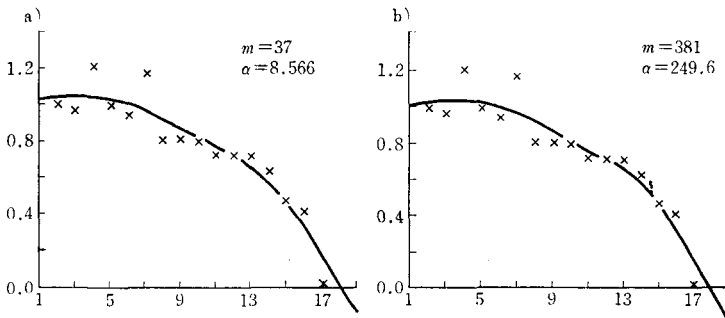


図 3

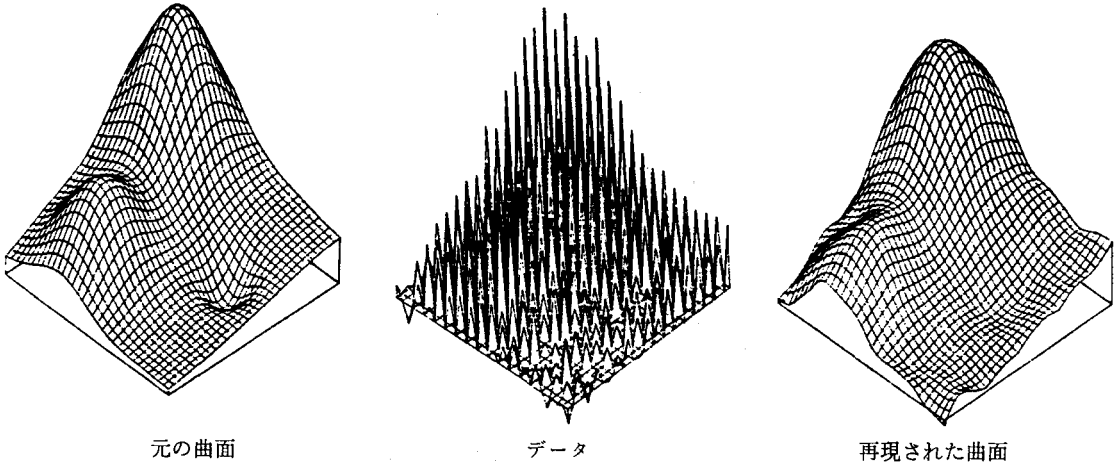


図 4

された曲面は、離散ラプラス作用素

(13)  $\Delta f_{ij} = f_{i+1,j} + f_{i,j+1} - 4f_{ij} + f_{i-1,j} + f_{i,j-1}$   
を用いて、(4)の代りに、モデル

(14)  $\Delta f_{ij} = \eta_{ij}$  (領域の内部)

$\nabla_i f_{ij} = \eta_{ij}$  (境界上)

$\nabla_j f_{ij} = \eta_{ij}$  ( " )

を仮定して得たものである。ただし、 $\eta_{ij}$  は平均 0、未知分散  $\sigma_f^2$  の正規分布に独立にしたがうものとする。この例では、データは格子点上に  $17 \times 17 = 289$  点とり、メッシュは  $33 \times 33 = 1089$  個用いている。図中、左は元の曲面、中央は誤差を含んだデータ、右は上に述べた方法で再現された曲面である。

### 3. 非線形尤度モデルと ABIC 最小化法

前節の例では、尤度が(1)のように線形モデルで表現されていた。しかし ABIC 最小化法は非線形モデルにも適用することができる。これを具体的に示すために、なめらかな密度関数  $p(x)$  のノンパラメトリックな推定問題を考えよう。前節と同様に、議論を有限次元ですませるために、有界領域を細かい等形メッシュ  $\{I_j\} (j=1, \dots, m)$  に刻み、メッシュ  $I_j$  に落ちる標本数  $n_j (j=1, \dots, m)$  のデータから

$$(15) \quad p_j \equiv \int_{I_j} p(x) dx$$

を推定する問題を考える。ただし  $\sum_{j=1}^m p_j = 1$  と仮定し、 $m$  は  $N = \sum_{j=1}^m n_j$  より十分大きいとする。このとき  $n = (n_1, n_2, \dots, n_m)^t$  の尤度は

$$(16) \quad P(n|p) \equiv \frac{N!}{\prod_{j=1}^m n_j!} \prod_{j=1}^m p_j^{n_j}$$

と書ける。ただし、 $p = (p_1, \dots, p_m)^t$  とする。 $p_j$  に関するモデルを導入しないで、最尤法を(16)に適用すると  $p_j$  の最尤推定値  $\hat{p}_j = n_j/N$  を得るが、これは凹凸が激しく役に立たない。そこで  $\{p_j\}$  のなめらかさを確率的にモデリングする必要があるが、このままでは、 $p_j$  に対する制約条件

$$(17) \quad \sum_{j=1}^m p_j = 1, \quad p_j \geq 0$$

のとり扱いが容易ではない。そこで Leonard が導入したロジスティック密度変換

$$p_j = p_j(f) = \exp(f_j) / \sum_{i=1}^m \exp(f_i)$$

を用いて、条件(17)が満たされるように、 $p_j$  を  $f = (f_1, f_2, \dots, f_m)^t$  でパラメトライズする。 $p(f) = (p_1(f), \dots, p_m(f))^t$  とおくと任意の  $\delta$  と  $1 = (1, 1, \dots, 1)^t$  に対して  $p(f + \delta 1) = p(f)$  となるが、 $f$  が一意的に定まらなくても、推定したいものは  $p$  であるから問題はない。

$p_i$  のなめらかさを表現するモデルとして  $f$  に次のような事前分布を導入することができる。

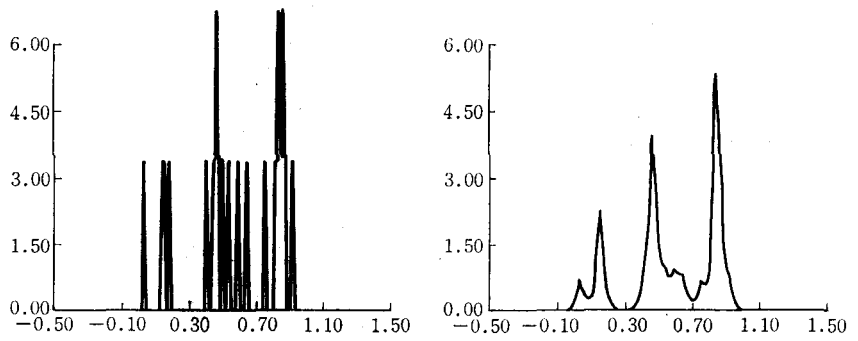
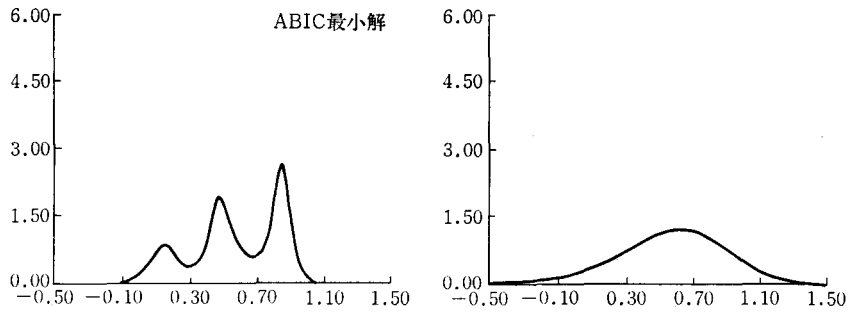


図 5 ▶



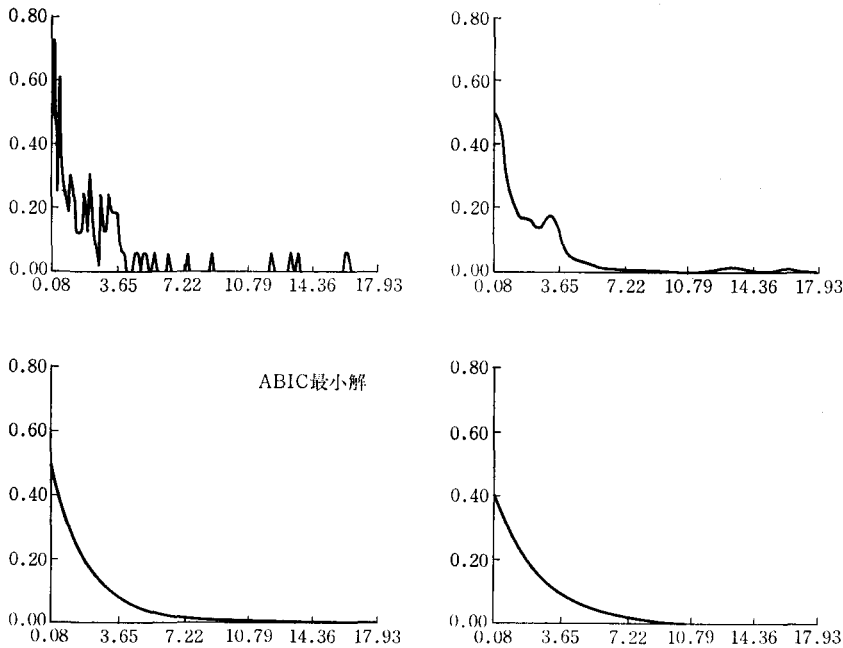
(18)  $\Pi'(f|\sigma_f^2, s^2) = \Pi(f|\sigma_f^2) \times$

$$\frac{1}{2\pi s^2} \exp\left(-\frac{f_1^2 + f_2^2}{2s^2}\right)$$

右辺の第2項は、 $\Pi'$ を proper な事前分布とす

るために附加したもので、 $s^2$ は十分大きな値に固定する。このとき、ベイズモデルの尤度は

(19)  $L(\sigma_f^2, s^2) = \int P(n|p(f)) \Pi'(f|\sigma_f^2, s^2) df$



◀ 図 6

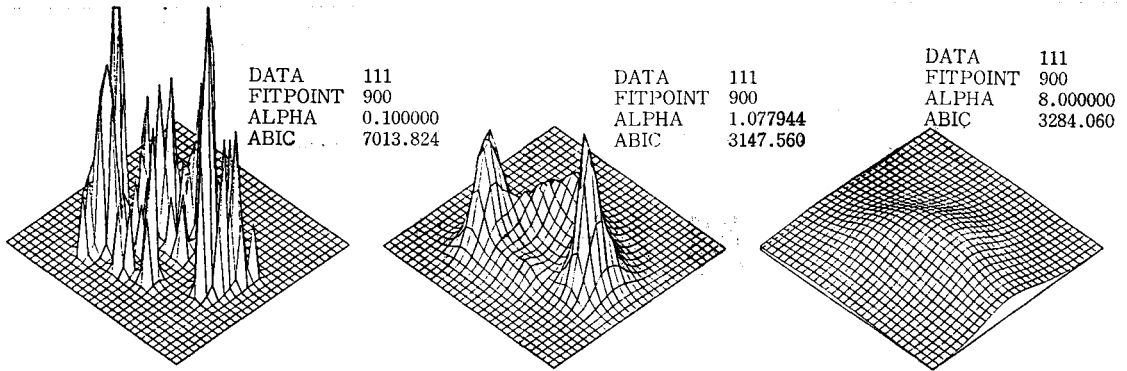


図 7

で与えられる。統計量

$$ABIC(\sigma_f^2, s^2) = -2 \log L(\sigma_f^2, s^2)$$

を定義し、これを最小にする  $\sigma_f^2 = \hat{\sigma}_f^2$  を求め、この条件のもとに、 $f$  の事後分布

$$p(f|n) = P(n|p(f)) \prod (f|\hat{\sigma}_f^2, s^2) / L(\hat{\sigma}_f^2, s^2)$$

のモード(最大値を与えるもの)  $\hat{f}$  を求める。この  $\hat{f}$  を介して  $p$  の推定値  $\hat{p} \equiv p(\hat{f})$  が得られる。変数  $s$  の値を十分大きな値に固定すると、 $ABIC(\sigma_f^2, s^2)$  の最小化と  $\hat{f}$  の計算過程において、 $s^2$  はまったく現われないようにすることができる。(19)の計算はモンテカルロ法を用いて行なうが、紙数の都合で割愛する。

図5は、隕石に含まれる無水珪酸量の22個のデータに  $m=150$  としてこの方法を適用したものである。4つの異なる  $\sigma_f$  の推定値に対応する密度のグラフのうち  $ABIC$  最小化法で選ばれたものは左下の密度である。

図6は、1875年から1951年のあいだに英国でおきた重大炭鉱爆発事故の間隔の109個のデータに、 $m=120$  として適用して得たものである。この場合も、 $ABIC$  最小解は左下のものである。横軸の単位は100日としている。

2変数の密度関数の推定にも同様の方法が適用できる。図7は、(14)の事前分布を用いて  $ABIC$  最小化法を、三陸沖で発生した地震源の位置111個のデータに適用したものである。メッシュは  $30 \times 30 = 900$  個用いている。中央のものが  $ABIC$  最小解で、左と右のものはそれぞれ平滑化不足と過

剰のケースを示している。

### 参考文献

- [1] Akaike, H.: Likelihood and Bayes procedure, In Bayesian Statistics, Bernardo et al. eds., University Press, Valencia, Spain, 143-166(1980)
- [2] 田辺國士, 田中輝雄: ベイズモデルによる曲線・曲面のあてはめ, 月刊地球5巻3号, 179-186(1983)
- [3] Tanabe, K., and Sagae, M.: A Bayes method for univariate and bivariate nonparametric density estimation, to appear.(1984)

### 次号予告

#### 特集 地理情報のOR

兵庫県における地理情報システム	宮崎 秀紀
地理的情報と都市・地域解析	腰塚 武志
勢力圏図と地理的最適化問題	鈴木敦夫・浅野孝夫
東京ガスにおける コンピュータ・マッピング・システム	東明 佐久良
地図と数値情報	塚原 弘一
地理情報システムの現状と課題	兼堀 文博
講座 Karmarkar の新 LP 解法(2)	刀根 薫
事例研究 指数平滑法を用いた生産指示 方式の解析	小谷重徳・田中吉弘