

## 参 考 文 献

- Akaike, H. (1980). Likelihood and the Bayes procedure, *Bayesian Statistics*, (eds. J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith), Univ. Press, Valencia, Spain, 143-166.
- Breslow, N.E. (1974). Covariance analysis of censored survival data, *Biometrics*, **30**, 89-99.
- 市田浩三・吉本富士市 (1979). スプライン関数とその応用, 教育出版.
- Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *J. Amer. Statist. Assoc.*, **53**, 457-481.
- Klotz, J. (1982). Spline smooth estimate of survival, *Survival Analysis* (eds. J. Crowley and R.A. Johnson), Institute of Mathematical Statistics, Hayward, California, Vol. **2**, 14-25.
- Prentice, R.L. and J.D. Kalbfleisch (1979). Hazard rate models with covariates, *Biometrics*, **35**, 25-39.

## 討 論

田辺: 時間間隔が規則的でないのにこうしたベイズモデルを使ってもあまりうまくいかないのではないか.

鎌倉: Normalizeした  $B$  スプラインを用いているので良いと思います.

田辺: どのような Normalize か. 高さなのか. Volume なのか.

鎌倉: この式で示される意味です ((7) 式).

尾形: 積分形の分布関数でみるよりインテンシティィーで見た方が意味があるのでないか.

鎌倉: そうした要求があるのであれば, それは簡単に計算できます. ここでは生存時間分布関数の滑らかな推定値ということで議論を展開しました.

## 非定常モデルとエントロピー最大化

統計数理研究所 北 川 源 四 郎

### 1. ま え が き

エントロピー最大化原理によれば, モデルの良さはエントロピーによって評価する. しかし, 実際には, エントロピーは直接求めることができないので観測値から推定する必要がある. よく知られている様に, 対数尤度はこのエントロピーの推定を目ざしたものと考えることができる. したがって, たとえば, ある船のモデルがいくつか考えられるとき, その船の実際の運動の記録が得られると, 尤度を求めることによってモデルの良し悪しが比較できる. しかし, ここに一つの問題が生じる. データの採録中に, 海象や気象の変化や海流, 変針の影響などによって船体運動の様子が徐々にあるいは急激に変化することがある. この様な場合, 固定されたモデルの良さは時間とともに変化しており, 尤度はモデルの良さの平均的な値に対応しているにすぎない. もし, 各時刻におけるモデルの良さが何らかの形で評価できれば, それにもとづいて, 各時刻でモデル選択を行なうといったより細かい解析が行なえるようになるであろう.

第2節では, 時間とともに変化するエントロピーの様子をシミュレーションの結果を用いて例示する. 第3節では, 時間とともに変化するエントロピーを推定するために, 区分化尤度と平滑化尤度を提案する. 第4節では, これらを用いて非定常時系列の局所的モデル構成を行なっ

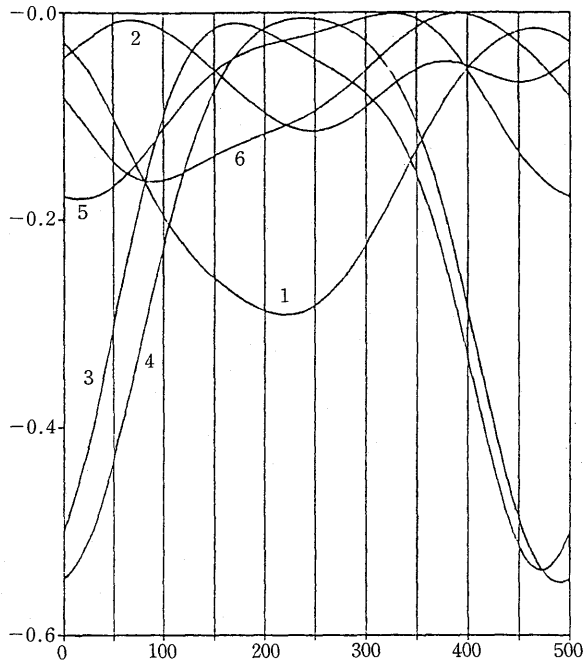


図1. モデル1~6のエントロピーの理論値

た例を示す.

2. 例: 変化するエントロピー

$m$  次の AR 過程

$$(1) \quad y_n = \sum_{i=1}^m a_i y_{n-i} + u_n, \quad u_n \sim N(0, \sigma^2)$$

と  $l$  次のモデル

$$(2) \quad y_n = \sum_{i=1}^l b_i y_{n-i} + v_n, \quad v_n \sim N(0, \tau^2)$$

を考慮することにする. (1), (2) のモデルが規定する確率分布をそれぞれ,  $g, f$  とするとエントロピーは

$$(3) \quad B(g; f) = -\frac{1}{2} \left\{ \log \left( \frac{\tau^2}{\sigma^2} \right) + \tau^{-2} \left( \nu_0 - 2 \sum_{i=1}^l b_i \nu_i + \sum_{i=1}^l \sum_{j=1}^l b_i b_j \nu_{i-j} \right) - 1 \right\}$$

で与えられる. ただし,  $\nu_i (i=0, 1, \dots)$  はモデル (1) の自己共分散関数である.

ここで,  $m=l=2, \sigma^2=\tau^2=1$  とし, 真のモデルの自己回帰係数が

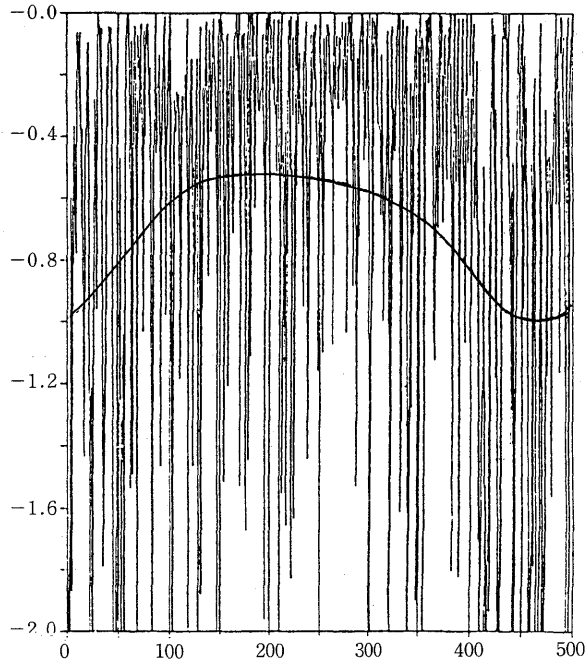


図2. 平均対数尤度と  $\log f(x_n)$

$$\begin{aligned}
 a_1 &= 2a_n \\
 a_2 &= -(a_n^2 + \beta_n^2) \\
 (4) \quad a_n &= 0.4802 + 0.15 \cos\left(\frac{2\pi n}{500}\right) \\
 \beta_n &= 0.4802 + 0.15 \sin\left(\frac{2\pi n}{500}\right)
 \end{aligned}$$

に従って変化するものと仮定する。これに対して6個の固定したモデル

$$\begin{aligned}
 (5) \quad \text{モデル 1:} & \quad b_1 = 1.386 \quad b_2 = -.64 \\
 \text{モデル 2:} & \quad b_1 = 1.132 \quad b_2 = -.64 \\
 \text{モデル 3:} & \quad b_1 = 0.8 \quad b_2 = -.64 \\
 \text{モデル 4:} & \quad b_1 = 0.6 \quad b_2 = -.36 \\
 \text{モデル 5:} & \quad b_1 = 0.8485 \quad b_2 = -.36 \\
 \text{モデル 6:} & \quad b_1 = 1.0392 \quad b_2 = -.36
 \end{aligned}$$

を準備しておく。図1は時刻  $n$  が1から500まで変化するとき、各モデルのエントロピー  $B(g; f_i) (i=1, \dots, 6)$  がどのように変化するかを示したものである。個々のモデルに注目すると、ある時刻では0に近いが、多くの場合0から遠く離れていることがわかる。しかし、 $B(g; f_i) (i=1, \dots, 6)$  の各時刻  $n$  での最大値を結んでいくと0に近い値が常に得られることが分かる。これは、時間とともに変化していく系に対しても複数個のモデルを用意しておき、適当に切り換えることができれば、全体としてはかなり良いモデルが得られる可能性を示している。

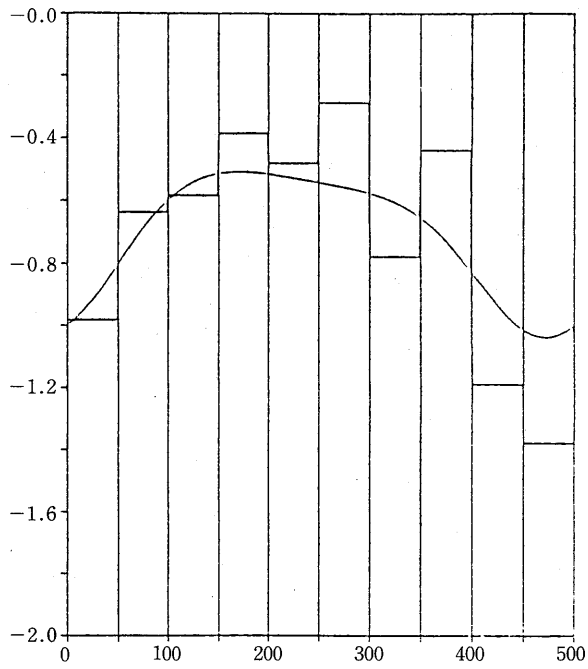


図3. 平均対数尤度と区分化対数尤度,  $M=50$

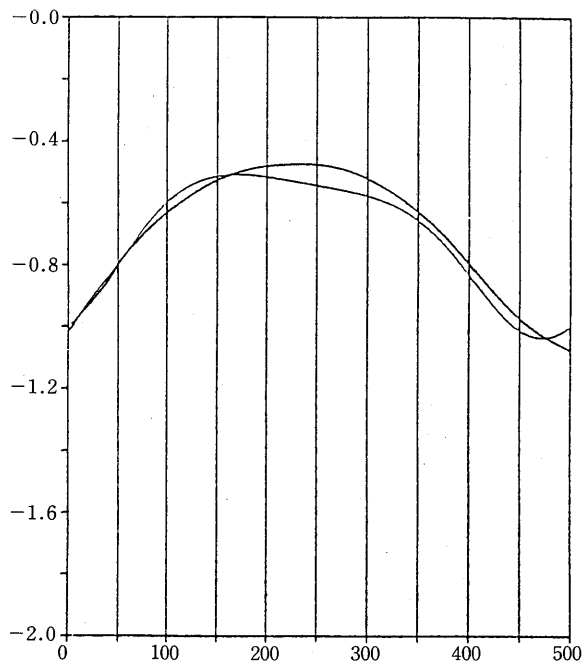


図4. 平均対数尤度と平滑化対数尤度,  $\tau^2=5 \times 10^{-4}$

### 3. エントロピーの局所的推定

前節で示した可能性の実現において問題となるのは、どのようにして最適なモデルを選択するかということである。エントロピー最大化原理に従えば、各時刻におけるエントロピー、あるいは平均対数尤度  $E \log f_i(X_n)$  が何らかの方法で推定できればよい。ただし、真の構造が時々刻々と変化する場合には、 $E \log f_i(X_n)$  の偏りのない推定値を与える観測値は  $x_n$  だけしかないことになる。図2には、 $i=3$  として  $E \log f_3(X_n)$  と  $\log f_3(x_n)$  ( $n=1, \dots, 500$ ) の値が示してある。 $\log f_3(x_n)$  の値は大きな変動を示し、そのままでは全く役に立たない事がわかる。同時に、 $\log f_3(x_n)$  の変動のしかたが  $E \log f_3(X_n)$  の大小を反映しており適当な方法によって、ある程度推定できることを示している。ここでは、区分化と局所化（平滑化）の二つの方法を示す。

#### 3.1. 区分化による推定

時間軸を  $M$  個ずつの小区間に分割し、各小区間上では  $E \log f_i(X_n)$  は一定と仮定する。このとき

$$(6) \quad \frac{1}{M} \sum_{j=1}^M \log f_i(x_{kM+j})$$

が  $k+1$  番目の区間上での  $E \log f_i(X_n)$  の推定値を与える。区間長  $M$  を短くすると真のモデルの変化によく追従する様になるが、推定の安定性は悪くなる。したがって、この区分化の方法においては、適当な  $M$  の選定が成功の鍵となる。以下では、(6)によって推定されたものを区分化対数尤度と呼ぶことにする。

#### 3.2. 平滑化による推定

$E \log f_i(X_n)$  が時間的に滑らかに変化すると仮定できる場合には次の様な平滑化が有効な手段になると思われる。まず各  $\log f_i(x_n)$  が  $c_n x^2 + \text{const.}$  の形をしていることがわかる。 $y_n = c_n x^2$  について、 $z_m = 2^{-1}(y_{2m-1} + y_{2m})$  とし、 $t_m = \log z_m + \nu$  ( $\nu$  は Euler 定数) なる変数を考えると  $t_m$  は

$$(7) \quad \begin{aligned} E(t_m) &= \log c_m \\ \text{Var}(t_m) &= \pi/6 \end{aligned}$$

を満たす確率変数となる。この性質を利用して次の様な平滑化のためのモデルを導入してみる。

$$(8) \quad \begin{aligned} \Delta^k p_m &= v_m \\ t_m &= p_m + w_m \end{aligned}$$

ただし、 $k=1$  又は  $2$ 、 $v_m \sim N(0, \tau^2)$  である。ここで、 $w_m$  を (7) で示した平均、分散を持つ正規分布と見なすと、カルマンフィルタと固定区間平滑器によって  $p_m = \log c_m$  の推定値  $\hat{p}_m$  を求めることができる。 $w_m$  が密度関数  $p(w) = \exp\{w - e^w\}$  に従う確率変数であることを利用して非ガウスフィルターを適用することも考えられるが、ここでは省略する。ともかく、 $p_m = \log c_m$  の推定値が得られると  $\exp\{\hat{p}_m\} + \text{const.}$  によって  $E \log f_i(x_n)$  の推定値が得られる。これを平滑化対数尤度と呼ぶことにする。

## 例

図3および4は前節の例のモデル3の区分化対数尤度 ( $M=50$ ), 平滑化対数尤度 ( $\tau^2=5 \times 10^{-4}$ )を示したものである. 区分化あるいは平滑化の操作によって真の平均対数尤度がかなり良く推定できていることがわかる.

## 4. 局所的モデル構成とモデルの階層の深化

区分化, 平滑化の方法によって, 時間とともに確率構造自体が変化する現象に対して, 各時刻におけるモデルの良さをある程度評価できることがわかった. これらの局所的に求められた尤度を用いれば各時刻において, それぞれ最適なモデルを選択できることになる. ここでは, この様なモデルの構成のしかたを局所的モデル構成と呼ぶことにする. 図5は平滑化尤度によって, 6つのモデル間の選択を行なった一例である. 少なくともこの場合には, エントロピー最大化の目標はほぼ達成されていることがわかる.

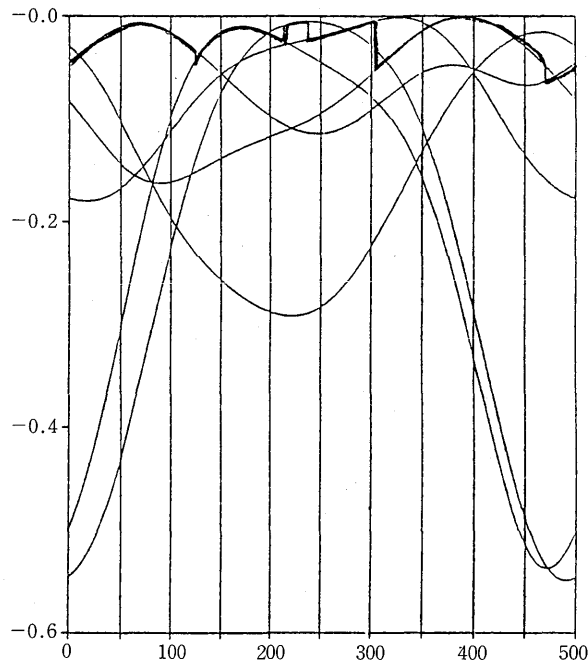


図5. 平滑化対数尤度によるモデル選択. 太線が選択されたモデルの平均対数尤度をあらわす.

## 例

## 適応的制御

典型的な状況に対応するいくつかのモデルとそれに対応する最適制御系を準備しておけば, 局所的モデル構成の方法により, 制御環境の変化にある程度適応できる制御システムを構成できる.

以上, 非定常な場合にも, 局所的モデル構成の方法により既存のモデル族から, 非定常なモデルが構成できることを示した. このモデルは(平均対数尤度の変化に関するモデル(8)は除い

て) 非定常性に関しては構造を仮定しない "記述的な" モデルとなっている。他方, 非定常性 (あるいは, 確率構造の変化のしかた) に何らかの構造を仮定してモデルを構成する方法がある。しかし, 両者は相反するものではない。構造を仮定するモデルは非定常性の程度を規定するパラメータを持っている。通常このパラメータは時間的に一定 (定常) と仮定されているが, この仮定も必ずしも満足できるものではない。したがって, より一段上での非定常性が問題となってくるがこの場合にも局所化によるモデル構成の方法は有効である。この方法は, モデルの階層を一段上げる実用的な方法であると同時に, 次のレベルでのモデルの開発を促す契機となるものと思われる。

例

季節調整モデル

$$y_n = t_n + s_n + w_n, \quad w_n \sim N(0, \sigma^2)$$

ただし,  $t_n$  はトレンド成分,  $s_n$  は季節成分でそれぞれ次のモデルに従う。

$$t_n = t_{n-1} + u_n, \quad u_n \sim N(0, \tau_1^2)$$

$$s_n = s_{n-1} + v_n, \quad v_n \sim N(0, \tau_2^2)$$

例

時変係数 AR モデル

$$y_n = \sum_{i=1}^m a_{in} y_{n-i} + w_n, \quad w_n \sim N(0, \sigma^2)$$

ただし,  $a_{in}$  は時変 AR 係数で次のモデルに従おう。

$$\Delta_n^k a_{in} = v_{in}, \quad v_{in} \sim N(0, \tau^2)$$

例

信号抽出モデル

$$y_n = r_n + s_n + w_n, \quad w_n \sim N(0, \sigma^2)$$

ただし,  $r_n$  は定常雑音,  $s_n$  は信号でそれぞれ次のモデルに従おう。

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

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

以上の例において,  $\sigma^2$  や  $\tau^2$  は定数と仮定されることが多いが局所的モデル構成の方法により時間的に変化する関数としてとらえることができる。

討 論

赤池: この方法は, モデル選択だけでなく, ベイズモデルの構成にも使えるのではないか?

答: たしかにその通りだが, 事前分布の選び方の問題もあるので, ここでは一番単純なモデル選択の形の問題だけを考えて。

大江: 地球回転や地殻変動などがどこで変わったかを見つけることに必死になっている。この方法はこの問題に適用できると思う。ただし, 細分化しすぎると, それに振り回されるので, それを何とかする基準が必要だと思っている。

答: 区分化する方法だと基本的なスパンの長さのとり方が重要になってくる。その点, 平滑化する方は変化した点を自動的にさがせるので良いのではないかと思う。