忘却重み係数を含む MDL 基準に関する性質

正会員 松 代 信 人[†]

Considerations on the Weighted Minimum Description Length Criterion

Nobuhito Matsushiro[†]

Abstract A weighted minimum description length (WMDL) criterion which is employed in adaptive model selection for an observed symbol series has been derived. The criterion has been derived from a weighted log likelihood instead of the log likelihood which is the basis of the minimum description length (MDL) criterion. In this paper, characteristics of estimated parameters of models and characteristics of the predictive WMDL criterion against the WMDL criterion are proved. Further the accuracy of model selection by the predictive WMDL criterion depending on weighting coefficients is discussed using computer simulations.

1. ま え が き

算術符号^{1)~3)}が研究されて以来、データの符号化は、実際に符号を構成する機構と、観測シンボル系列から符号化モデルを推定する機構から成るという考え方を広くもたらした。算術符号では、符号を構成する機構と符号化モデルを推定する機構が完全に分離されていることから、それまではほとんど統計の分野で論じられてきたモデル推定の問題⁴⁾に関しても、多くの符号化研究者の関心が向けられるようになった。MDL (Minimum Description Length) 基準⁵⁾⁶⁾は、このような算術符号を中心とした研究の流れを背景に、J. Rissanen により提唱された。MDL 基準は、符号化、予測、推定等の種々の問題⁷⁾において、問題を記

述するのに考え得る複数のモデルのうちどのモデルを選択すればよいかということに対する、客観的な選択基準を与えるものである。

この MDL 基準は、観測シンボル系列全体に対する最適なモデルを選択する基準である。著者は文献 8) で、観測シンボル系列の局所的性質に適したモデルを適応的に選択する問題を想定して、そのために用いる基準として、WMDL 基準(忘却重み係数を含む MDL 基準: Weighted Minimum Description Length Criterion)、予測 WMDL 基準を提案し、画像符号化における応用例を通してその有効性を示した。

本論文では、この WMDL 基準、予測 WMDL 基準に関する性質を理論と実験の両面から検討する。まず 2 章で WMDL 基準について説明する。次に 3 章で、

キーワード: 情報量基準, MDL 基準, 符号化モデル選択

1995 年 10 月 6 日受付, 1995 年 12 月 14 日再受付

[†] 株式会社沖データ 研究所 (〒370 高崎市双葉町 3-1, TEL 0273-28-6172)

[†] Research Laboratory, Oki Data Corporation (3-1, Futabacho, Takasaki-shi, Gunma 370, Japan)

推定モデルパラメータの性質，WMDL 基準に対する予測 WMDL 基準の漸近特性に関する性質を証明する．4 章ではシミュレーションにより，忘却重み係数の値と予測 WMDL 基準に基づくモデル選択精度の関係に関する実験的考察を行う．最後に 5 章でむすびとする．

2. WMDL 基準

情報源の確率分布パラメータがわかっていれば，その分布による符号化によって最短の符号が構成できる．しかし未知の場合には，事前に記述するかもしくは推定しなければ復号化を可能とすることはできない．このモデルに関する記述長をデータの記述長に加えた合計の記述長を最小にするモデルを選択するという原理が MDL 原理である．WMDL 基準とは，MDL 基準がその出発点としているモデルの確率分布の対数尤度の代わりに，観測シンボル系列の局所性を考慮するための重み付き対数尤度を用いることで導いたものである．WMDL 基準を次式に示す． $\log$ は自然対数とする．

$$\begin{aligned} WMDL(S^n) &= \min_k \left[- \sum_{t=1}^n w_t \log Q_{k, \hat{\theta}}(s_t) + \frac{k}{2} \log W_n \right] \end{aligned} \quad (1)$$

ここで，

n : 観測シンボル数

$S^n : s_1 s_2 \cdots s_n$ ，長さ n の観測シンボル系列，

$s_t (t=1, 2, \dots, n)$ は独立な確率変数

k : パラメータ数

θ : k 次元パラメータ

$Q_{k, \theta}$: k 次元パラメータ θ によって表されるモデルの確率分布

$\hat{\theta}$: θ の最尤推定量

ξ : 減衰係数

w_t : 忘却重み係数， $w_{t-1} = \xi w_t (t=2, \dots, n)$ ，

$\xi < 1, w_n = 1$

$$W_n = \sum_{t=1}^n w_t$$

(1) 式の第 1 項 (符号なし) はモデルの確率分布の重み付き対数尤度すなわち (−) 符号を含めてデータの重み付き記述長，第 2 項はモデルパラメータの記述長である．WMDL 基準は， S^n を一度全部見て $\hat{k}, \hat{\theta}$ を求め記述しているから，あらためて S^n を $Q_{\hat{k}\hat{\theta}}$ で記述するものである．ここで $\hat{k}$ は (1) 式を最小にする k とする．これに対してモデルのパラメータを事前に記述せずに， $S^{t-1} (t=2, \dots, n)$ から s_t の確率分布

として最良であるものとして推定されたモデルパラメータ $\hat{k}(t-1), \hat{\theta}(t-1)$ を用い， s_t を $Q_{\hat{k}(t-1), \hat{\theta}(t-1)}$ で記述する場合のモデル選択基準として，予測 WMDL 基準を定式化している．予測 WMDL 基準 ($WMDL_p$) を次式に示す．

$$WMDL_p(S^n) = - \sum_{t=1}^n w_t \log Q_{\hat{k}(t-1), \hat{\theta}(t-1)}(s_t) \quad (2)$$

観測シンボル系列の局所的性質に適したモデルを適応的に選択する問題においては，事前にモデルパラメータの記述を行わないことから，実際の問題には予測 WMDL 基準が適用される．WMDL 基準は通常の MDL 基準との式の形の上での対応をとることと，忘却による見かけ上のシンボル数の減少といった特性を明らかにするために導出した．WMDL 基準，予測 WMDL 基準においては，忘却により見かけ上の観測シンボル数が n から W_n に減少し， $n \rightarrow \infty$ においても W_n は有限値 ($1/(1-\xi)$) をとる．WMDL 基準，予測 WMDL 基準のモデルパラメータの推定においても，観測シンボル系列に対する忘却係数による重み付けを行い，観測系の一貫性を確保する．

なお，通常の MDL 基準，予測 MDL 基準は，(1) (2) 式において $w_t=1, (t=1, 2, \dots, n)$ とした場合に相当する．予測 MDL 基準は，事前のモデルパラメータの記述を必要としない代わりに，過去に観測されたシンボル系列だけから $\hat{k}(t-1), \hat{\theta}(t-1)$ を推定することからそのための損失があって， $n \rightarrow \infty$ で MDL 基準と同じ記述長が必要となることが証明^{9)~11)} されている．

3. WMDL 基準に関する性質とその証明

本章では，WMDL 基準，予測 WMDL 基準について，推定モデルパラメータの性質，WMDL 基準に対する予測 WMDL 基準の漸近特性に関する証明を行う．証明は対数尤度の場合の文献 9) 12) に準ずるものであり， W_n が充分大の場合について θ の真値 θ^* の近傍において行う．以下で k 次元パラメータ θ に関する 1 階微分は k 次元列ベクトルを，2 階微分は $k \times k$ 行列を示す．また “ $'$ ” は転置を示す．

〔性質 1〕

W_n が大きくなるに従い， $\hat{\theta}(n)$ はパラメータ真値 θ^* に漸近的に近づく．

〔証明〕

$\sum_{s_t} Q_{\theta}(s_t) = 1$ の両辺を θ について 1 階微分し $\theta = \theta^*$ とすると $\sum_{s_t} \frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} = 0$ であり， $\frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} =$

$$\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} Q_{\theta^*}(s_t) \text{ を代入すると}$$

$$E\left[\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right] = \sum_{s_t} \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} Q_{\theta^*}(s_t) = 0 \quad (3\cdot a)$$

いま, $\frac{\partial \log Q_{\hat{\theta}(n)}(s_t)}{\partial \theta}$ を θ^* の回りで Taylor 展開すると

$$\sum_{t=1}^n w_t \frac{\partial \log Q_{\hat{\theta}(n)}(s_t)}{\partial \theta} = \sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} + \sum_{t=1}^n w_t \frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} (\hat{\theta}(n) - \theta^*) \quad (3\cdot b)$$

$$\hat{\theta} = \beta \hat{\theta}(n) + (1 - \beta) \theta^*, \quad 0 < \beta < 1$$

$\hat{\theta}(n)$ は重み付き対数尤度に関する最尤推定量であるから (3・b) 式の左辺は重み付き対数尤度の極値で 0 となり, W_n で正規化すると

$$\left(\left(\sum_{t=1}^n w_t \frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} \right) / W_n \right) (\hat{\theta}(n) - \theta^*) = - \left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \right) / W_n \quad (3\cdot c)$$

(3・c) 式の右辺の重み付き平均は, $s_t (t=1, 2, \dots, n)$ が独立であることと大数の法則から W_n が充分大のとき

$$\left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \right) / W_n \rightarrow E \left[\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \right] = 0 \quad (3\cdot d)$$

また (3・c) 式の左辺は W_n が十分大のとき,

$$\left(\sum_{t=1}^n w_t \frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} \right) / W_n \rightarrow E \left[\frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} \right] \neq 0 \quad (3\cdot e)$$

したがって W_n が充分大のとき, (3・c) 式で右辺が 0 に近づくことから左辺は 0 に近づく必要があり $\hat{\theta}(n)$ は θ^* に近づくことが示された. ただし, W_n は $n \rightarrow \infty$ においても有限値をとることから, $n \rightarrow \infty$ においても $\hat{\theta}(n)$ と θ^* の差は, 無限小まで近づくことはできない.

[性質 2] ¹³⁾

WMDL 基準, 予測 WMDL 基準におけるシンボル数 m までの観測で, W_n が充分大のとき $(\hat{\theta}(n) - \theta^*)$ は平均 0, 分散行列 $cJ^{-1}(\theta^*)$, $c = \left(\sum_{t=1}^n w_t^2 \right) / W_n^2$ の正規分布 $N(0, cJ^{-1}(\theta^*))$ に従う. $J(\theta^*)$ は E を期待値とする以下の (4・a) 式で定義され, $\log Q_{\theta}$ は θ について 2 階連続微分可能とする.

$$J(\theta^*) = -E \left[\frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} \right] \quad (4\cdot a)$$

[証明]

$\sum_{t=1}^n w_t \frac{\partial \log Q_{\hat{\theta}(n)}(s_t)}{\partial \theta}$ を θ^* の回りで Taylor 展開す

る.

$$\sum_{t=1}^n w_t \frac{\partial \log Q_{\hat{\theta}(n)}(s_t)}{\partial \theta} = \sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} + \sum_{t=1}^n w_t \frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} (\hat{\theta}(n) - \theta^*) + R_n \quad (4\cdot b)$$

θ^* の近傍において (4・b) 式で R_n を無視する. $\hat{\theta}(n)$ は重み付き対数尤度に関する最尤推定量であるから (4・b) 式の左辺は重み付き対数尤度の極値で 0 となり

$$\left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \right) / W_n = - \left\{ \left(\sum_{t=1}^n w_t \frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} \right) / W_n \right\} (\hat{\theta}(n) - \theta^*) \quad (4\cdot c)$$

他方, (4・c) 式の右辺の $\frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'}$ について次式が成り立つ.

$$E \left[\frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} \right] = -E \left[\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta'} \right] \quad (4\cdot d)$$

なぜならば $\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} = \frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} Q_{\theta^*}^{-1}(s_t)$ より

$$\frac{\partial^2 \log Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} = \frac{\partial^2 Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} Q_{\theta^*}^{-1}(s_t) - \frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} \frac{\partial Q_{\theta^*}(s_t)}{\partial \theta'} Q_{\theta^*}^{-2}(s_t) \quad (4\cdot e)$$

ところで, $\sum_{s_t} Q_{\theta}(s_t) = 1$ であるから両辺を θ について

2 階微分し $\theta = \theta^*$ とすると $\sum_{s_t} \frac{\partial^2 Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} = 0$ であり,

(4・e) 式の右辺第 1 項の期待値は

$$E \left[\frac{\partial^2 Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} Q_{\theta^*}^{-1}(s_t) \right] = \sum_{s_t} \frac{\partial^2 Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} Q_{\theta^*}^{-1}(s_t) Q_{\theta^*}(s_t) = \sum_{s_t} \frac{\partial^2 Q_{\theta^*}(s_t)}{\partial \theta \partial \theta'} = 0 \quad (4\cdot f)$$

となる. また (4・e) 式の右辺の第 2 項は

$$\frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} \frac{\partial Q_{\theta^*}(s_t)}{\partial \theta'} Q_{\theta^*}^{-2}(s_t) = \left(\frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} Q_{\theta^*}^{-1}(s_t) \right) \left(\frac{\partial Q_{\theta^*}(s_t)}{\partial \theta'} Q_{\theta^*}^{-1}(s_t) \right) = \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta'} \quad (4\cdot g)$$

であるから, (4・e) 式の両辺の期待値をとって (4・d) 式が成り立つ.

(4・c) 式の左辺を ν_n とする.

$$\nu_n = \left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} \right) / W_n \quad (4\cdot h)$$

ところで, (3・a), (4・d)式より

$$E\left[\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right]=0 \quad (\text{平均}) \quad (4 \cdot i)$$

$$E\left[\left(\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right)\left(\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta'}\right)\right]=J(\theta^*) \quad (\text{分散}) \quad (4 \cdot j)$$

(4・i)式から

$$E[\nu_n]=\sum_{t=1}^n w_t E\left[\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right]=0 \quad (4 \cdot k)$$

$s_t(t=1, 2, \dots, n)$ が独立であることから $\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}$

($t=1, 2, \dots, n$)も独立であり, (4・j)式から

$$E[\nu_n \nu_n'] = \sum_{t=1}^n \frac{w_t^2}{W_n^2} E\left[\left(\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right) \cdot \left(\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta'}\right)\right] = cJ(\theta^*) \quad (4 \cdot l)$$

よって中心極限定理により見かけ上の観測シンボル数 W_n が充分大のとき

$$\nu_n \sim N(0, cJ(\theta^*)) \quad (4 \cdot m)$$

(4・c)式右辺 { } 内の重み付き平均を近似的に期待値 $J(\theta^*)$ とみなすと $\nu_n = J(\theta^*)(\hat{\theta}(n) - \theta^*)$ であり, (4・m)式から

$$(\hat{\theta}(n) - \theta^*) \sim N(0, cJ^{-1}) \quad (4 \cdot n)$$

〔性質3〕

$\hat{\theta}(n)$ は θ^* の不偏推定量であると仮定する. $\hat{\theta}(n)$ の推定誤差分散に関する以下の関係が成り立つ.

$$E[(\hat{\theta}(n) - \theta^*)(\hat{\theta}(n) - \theta^*)'] \geq cJ^{-1} \quad (5 \cdot a)$$

〔証明〕

$\hat{\theta}(s_t)$ を s_t による θ の推定値とする. $s_t(t=1, 2, \dots, n)$ が独立であることから以下の(5・b), (5・c)式が成り立つ. パラメータ θ の推定においても忘却係数による重み付けを行うことから, $\hat{\theta}(n)$ は以下のように書ける.

$$\hat{\theta}(n) = \left(\sum_{t=1}^n w_t \hat{\theta}(s_t)\right)/W_n \quad (5 \cdot b)$$

$\hat{\theta}(n)$ は θ^* の不偏推定量であることから $\hat{\theta}(s_t)$ は θ^* の不偏推定量であり

$$E[\hat{\theta}(s_t)] = \theta^* \quad (5 \cdot c)$$

(5・c)式 $E[\hat{\theta}(s_t)] = \sum_{s_t} \hat{\theta}(s_t) Q_{\theta^*}(s_t) = \theta^*$ の両辺を θ^*

で微分すると $\sum_{s_t} \hat{\theta}(s_t) \frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} = I$, $\frac{\partial Q_{\theta^*}(s_t)}{\partial \theta} = \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} Q_{\theta^*}(s_t)$ であるから

$$\begin{aligned} \sum_{s_t} \hat{\theta}(s_t) \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} Q_{\theta^*}(s_t) \\ = E\left[\hat{\theta}(s_t) \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right] = I \end{aligned} \quad (5 \cdot d)$$

また $s_t(t=1, 2, \dots, n)$ が独立であることから $\hat{\theta}(s_t)$,

$\frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta} (t=1, 2, \dots, n)$ も互いに独立である. 任意のベクトルに対して, その共分散行列 R は半正定値行列であるから, (4・i)式, (5・b)式, (5・d)式を用い

$$\begin{aligned} R = E & \left[\begin{pmatrix} \left(\sum_{t=1}^n w_t \hat{\theta}(s_t)\right)/W_n - \theta^* \\ \left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right)/W_n \\ \left(\sum_{t=1}^n w_t \hat{\theta}(s_t)\right)/W_n - \theta^* \\ \left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right)/W_n \end{pmatrix} \right. \\ & \left. \cdot \begin{pmatrix} \left(\sum_{t=1}^n w_t \hat{\theta}(s_t)\right)/W_n - \theta^* \\ \left(\sum_{t=1}^n w_t \frac{\partial \log Q_{\theta^*}(s_t)}{\partial \theta}\right)/W_n \end{pmatrix} \right] \\ & = \begin{bmatrix} E[(\hat{\theta}(n) - \theta^*)(\hat{\theta}(n) - \theta^*)'] & cI \\ cI & cJ \end{bmatrix} \geq 0 \end{aligned} \quad (5 \cdot e)$$

半正定値行列 R とある行列 M に対して, $M'RM$ も半正定値行列となることを用いると

$$\begin{aligned} M'RM &= [c^{-1}I \quad -c^{-1}J^{-1}] \\ & \cdot \begin{bmatrix} E[(\hat{\theta}(n) - \theta^*)(\hat{\theta}(n) - \theta^*)'] & cI \\ cI & cJ \end{bmatrix} \\ & \cdot \begin{bmatrix} c^{-1} & I \\ -c^{-1} & J^{-1} \end{bmatrix} \\ & = \begin{bmatrix} c^{-2}E[(\hat{\theta}(n) - \theta^*)(\hat{\theta}(n) - \theta^*)'] - c^{-1}J^{-1} & 0 \\ 0 & c^{-1}J^{-1} \end{bmatrix} \geq 0 \end{aligned} \quad (5 \cdot f)$$

(5・f)式の行列を対角化しても半正定値行列であり, 以下の対角行列が得られる.

$$\begin{bmatrix} c^{-2}E[(\hat{\theta}(n) - \theta^*)(\hat{\theta}(n) - \theta^*)'] - c^{-1}J^{-1} & 0 \\ 0 & c^{-1}J^{-1} \end{bmatrix} \geq 0 \quad (5 \cdot g)$$

よって

$$E[\hat{\theta}(n) - \theta^*)(\hat{\theta}(n) - \theta^*)'] - cJ^{-1} \geq 0 \quad (5 \cdot h)$$

であり [性質3]は証明された.

[性質3]から, [性質2]のパラメータ $\hat{\theta}(n)$ の分散は下限に近づくことがわかる.

〔性質4〕¹³⁾

k を真値 k^* , θ を真値 θ^* としたときの(1)式すなわち $WMDL(S^n)$ の下限を $WMDL^*(S^n)$ とする.

$$\begin{aligned} E[WMDL_p(S^n)]/W_n \\ \leq E[WMDL^*(S^n)]/W_n + R_n^{(2)}/W_n \\ \leq E[WMDL(S^n)]/W_n + R_n^{(2)}/W_n \end{aligned} \quad (6 \cdot a)$$

の関係が成り立ち, W_n が充分大きいとき $R_n^{(2)}/W_n$ は小さくなり, $WMDL_p(S^n)$ の期待値は $WMDL^*(S^n)$ の期待値に近づく. また $WMDL(S^n)$ の期待値は

$WMDL^*(S^n)$ の期待値に近づく。すなわち、モデルに関する記述を事前に行っているか否かにかかわらず、同一の値に近づく。ここで k の値域は有限であり、 θ は W_n に対応した精度を取り得るものとする。 $\log Q_{k,\theta}$ は θ について 3 階連続微分可能とする。

【証明】

$l(S^n) = -\sum_{t=1}^n w_t \log Q_{k^*, \hat{\theta}(t-1)}(s_t)$, パラメータ θ の推定誤差を $e_t = \hat{\theta}(t-1) - \theta^*$ とする。 $l(S^n)$ を θ^* の近傍で Taylor 展開すると

$$\begin{aligned} l(S^n) = & -\sum_{t=1}^n w_t \log Q_{k^*, \theta^*}(s_t) \\ & + \sum_{t=1}^n w_t \frac{-\partial \log Q_{k^*, \theta^*}(s_t)}{\partial \theta} e_t \\ & + \sum_{t=1}^n \frac{w_t}{2} e_t' \frac{-\partial^2 \log Q_{k^*, \theta^*}(s_t)}{\partial \theta \partial \theta'} e_t \\ & + \sum_{t=1}^n \frac{w_t}{6} r(s_t) \end{aligned} \quad (6\cdot b)$$

ここで

$$r(s_t) = e_t' \frac{-\partial^3 \log Q_{k^*, \theta^*}(s_t)}{\partial \theta \partial \theta' \partial \theta''} e_t e_t''$$

$$\hat{\theta} = \beta \hat{\theta}(t-1) + (1-\beta) \theta^*, \quad 0 < \beta < 1$$

(6・b) 式の右边第 2 項は、 s_t と $s_1 s_2 \cdots s_{t-1}$ から推定された $\hat{\theta}(t-1)$ が独立であること、(3・a) 式と同様にして $E\left[\frac{\partial \log Q_{k^*, \theta^*}(s_t)}{\partial \theta}\right] = 0$ であることから

$$\begin{aligned} E\left[\sum_{t=1}^n w_t \frac{-\partial \log Q_{k^*, \theta^*}(s_t)}{\partial \theta} e_t\right] \\ = -\sum_{t=1}^n w_t E\left[\frac{\partial \log Q_{k^*, \theta^*}(s_t)}{\partial \theta}\right] E[e_t] = 0 \end{aligned} \quad (6\cdot c)$$

また (6・b) 式の右边第 3 項は、(4・a)、(4・n) 式より

$$\begin{aligned} E\left[\sum_{t=1}^n \frac{w_t}{2} e_t' \frac{-\partial^2 \log Q_{k^*, \theta^*}(s_t)}{\partial \theta \partial \theta'} e_t\right] \\ = \frac{k^*}{2} \sum_{t=1}^n w_t / (\alpha W_{t-1}) + \frac{1}{2} \sum_{t=1}^n g_t w_t / (\alpha W_{t-1}) \end{aligned} \quad (6\cdot d)$$

ここで

$$\alpha = W_n / \sum_{i=1}^n w_i^2 = 1 / (c W_n)$$

g_t : t が小さい時点での e_t の正規分布の近似誤差および $n > t-1$ で α (=一定) としている誤差の補正項

$$\begin{aligned} \lim_{n \rightarrow \infty} \alpha &= \lim_{n \rightarrow \infty} \left(W_n / \sum_{i=1}^n w_i^2 \right) \\ &= (1 + \xi + \xi^2 + \cdots) / (1 + \xi^2 + \xi^4 + \cdots) \\ &= 1 + \xi \end{aligned}$$

であり α は有限値をとることから、 t が大きくなるに従い $g_t \rightarrow 0$ 。

さらに (6・b) 式右边第 4 項については、(4・n) 式か

ら $|e_t|$ は $W_t^{-1/2}$ のオーダー、 $|-\frac{\partial^3 \log Q_{k^*, \theta^*}(s_t)}{\partial \theta \partial \theta' \partial \theta''}|$ は仮定より有限値をとることから、 L を正の定数として

$$\begin{aligned} |r(s_t)| &\leq \left| \frac{-\partial^3 \log Q_{k^*, \theta^*}(s_t)}{\partial \theta \partial \theta' \partial \theta''} \right| \|e_t\| \|e_t'\| \|e_t''\| \\ &\leq L W_t^{-3/2} \end{aligned} \quad (6\cdot e)$$

よって (6・b) 式の両辺の期待値をとり、 W_n で正規化すると

$$\begin{aligned} E[l(S^n)] / W_n &\leq -E\left[\sum_{t=1}^n w_t \log Q_{k^*, \theta^*}(s_t)\right] / W_n \\ &\quad + \left(\frac{k^*}{2} \sum_{t=1}^n w_t / (\alpha W_{t-1})\right) / W_n + R_n / W_n \end{aligned} \quad (6\cdot f)$$

ここで

$$|R_n| \leq \sum_{t=1}^n (|g_t w_t| / (2\alpha W_{t-1}) + L w_t W_t^{-3/2} / 6)$$

(6・f) 式の右边第 2 項は、 $1/x$ を x の区間 $[0, \alpha W_n]$ において積分 $\int_0^{\alpha W_n} (1/x) dx = \log(\alpha W_n)$ し、これを分割幅 $w_t (t=1, 2, \dots, n)$ で離散化すると $\sum_{t=1}^n w_t / (\alpha W_t) = \log(\alpha W_n) = \log W_n + \log \alpha$ であるから、(6・f) 式は

$$\begin{aligned} E[l(S^n)] / W_n &\leq -E\left[\sum_{t=1}^n w_t \log Q_{k^*, \theta^*}(s_t)\right] / W_n \\ &\quad + \frac{k^*}{2} \log W_n / W_n + R_n^{(1)} / W_n \end{aligned} \quad (6\cdot g)$$

ここで

$$\begin{aligned} |R_n^{(1)}| &\leq \sum_{t=1}^n (|g_t w_t| / (2\alpha W_{t-1}) + L w_t W_t^{-3/2} / 6) \\ &\quad + (k^* \log \alpha) / 2 \end{aligned}$$

(6・g) 式と $WMDL_p(S^n)$ には以下の関係式が成り立つ。

$$\begin{aligned} E[WMDL_p(S^n)] / W_n &= E[l(S^n) + C] / W_n \\ &\leq E\left[-\sum_{t=1}^n w_t \log Q_{k^*, \theta^*}(s_t) + \frac{k^*}{2} \log W_n\right] \\ &\quad / W_n + R_n^{(1)} / W_n + C / W_n \end{aligned} \quad (6\cdot h)$$

ここで、 C : $k(t-1)$ が k^* に一致するまでのオーバーヘッド。

すなわち (6・h) 式は

$$\begin{aligned} E[WMDL_p(S^n)] / W_n \\ &\leq E[WMDL^*(S^n)] / W_n + R_n^{(2)} / W_n \\ &\leq E[WMDL(S^n)] / W_n + R_n^{(2)} / W_n \end{aligned} \quad (6\cdot i)$$

ここで

$$\begin{aligned} |R_n^{(2)}| &\leq \sum_{t=1}^n (|g_t w_t| / (2\alpha W_{t-1}) + L w_t W_t^{-3/2} / 6) \\ &\quad + (k^* \log \alpha) / 2 + C \end{aligned}$$

ところで、WMDL 基準は観測シンボル系列に対する 1 回目のパスでパラメータ推定を、2 回目のパスで記述を行い、予測 WMDL 基準は 1 回のパスでパラメータ推定と記述を同時に行う。 $s_t (t=1, 2, \dots, n)$ と $\hat{\theta}(t-1)$ は独立であることから、パラメータ推定と記

述は独立であり、推定と記述を同時に行っても独立に行なった場合より全記述長が小さくなることはない。したがって、パラメータを真値とした $WMDL^*(S^n)$ は $WMDL_p(S^n)$ の下限である。ここで(6・i)式において W_n が充分大きいとき、不等式を成立させている余剰項 $R_n^{(2)}/W_n$ の絶対値は小さくなり、 $WMDL^*(S^n)$ は $WMDL_p(S^n)$ の下限であることを考慮すると $WMDL_p(S^n)$ の期待値は $WMDL^*(S^n)$ の期待値に近づく。また W_n が充分大きいとき、値域が有限である k の最適化において k は k^* に一致し、 $\hat{\theta}(n)$ は θ^* に近づき、 $WMDL(S^n)$ の期待値は $WMDL^*(S^n)$ の期待値に近づく。ただし、 W_n は $n \rightarrow \infty$ においても ξ に依存した有限値をとることから、 $WMDL^*(S^n)$ の期待値と $WMDL_p(S^n)$, $WMDL(S^n)$ の期待値の差は $n \rightarrow \infty$ においても無限小まで近づくことはできない。

4. シミュレーション実験

本章ではシミュレーションにより、忘却重み係数の値と予測 WMDL 基準に基づくモデル選択精度の関係に関する実験的考察を行う。予測 WMDL 基準は観測シンボル系列の局所的性質に適したモデルを適応的に選択するための基準であり、局所的性質に対する追従性を有する代わりに、見かけ上の観測シンボル数が減少することから、定常区間におけるモデル選択の精度を劣化させることになる。モデル選択においては、モデル選択基準値自身の精度より、モデル選択基準に基づき選択されるモデルパラメータに依存した全記述長の値が重要であり、本実験はこの評価を目的とするものである。実験に使用した観測シンボル系列は、3 次、5 次、7 次のマルコフモデルで、各々 3 種類のエントロピーを持つ 9 種類のモデルにより生成した。マルコフ次元を固定したことから、生成された観測シンボル系列は定常とみなすことができる。3 種類のエントロピーは大きい順に、0.449 (3 次)、0.423 (5 次)、0.438 (7 次) と、0.221 (3 次)、0.219 (5 次)、0.212 (7 次) と、0.110 (3 次)、0.117 (5 次)、0.101 (7 次) である。ここでは自然対数についてエントロピーを算出した。マルコフ状態分割された各予測状態の確率パラメータは、忘却重み係数で重み付けしたシンボル出現確率として算出した。忘却重み係数の減衰係数 ξ は 0.99, 0.995, 0.997, 0.999 の 4 種類に設定した。忘却重み係数は、定常とみなせる区間を限定するために導入したものであり、各系列に含まれるシンボル数は ξ の値に依存して設定した。ここでは、忘却重み係数が 1 から 10^{-2} までの範囲にあるシンボル数を観測シンボル系列長とした。 $\xi=0.99$ の場

表 1 忘却重み係数と予測 WMDL 基準に基づくモデル選択精度の関係

Accuracy of model selection by the predictive WMDL criterion depending on weighting coefficients.

モデル次元	ξ	Ent. 1 誤差比率 (%)	Ent. 2 誤差比率 (%)	Ent. 3 誤差比率 (%)
3	0.990	0.23	0.24	0.81
	0.995	0.15	0.16	0.51
	0.997	0.12	0.11	0.47
	0.999	0.05	0.05	0.21
5	0.990	0.39	0.90	1.1
	0.995	0.28	0.62	1.1
	0.997	0.21	0.53	0.91
	0.999	0.20	0.35	0.42
7	0.990	1.0	1.1	1.2
	0.995	0.67	0.71	0.79
	0.997	0.59	0.64	0.62
	0.999	0.38	0.34	0.46

合、シンボル数は 458, 見かけ上の観測シンボル数 W_n は 99, $\xi=0.995$ の場合、シンボル数は 918, W_n は 198, $\xi=0.997$ の場合、シンボル数は 1532, W_n は 330, $\xi=0.999$ の場合、シンボル数は 4602, W_n は 990 である。

表 1 は、観測シンボル系列の定常区間における予測 WMDL 基準によるモデル選択の精度算出を予測 MDL 基準との比較により行ったものである。精度算出の基準を予測 MDL 基準とし、予測 MDL 基準により選択されたマルコフ次元の予測 MDL 基準値 (mdl) と、予測 WMDL 基準により選択されたマルコフ次元の予測 MDL 基準値 ($mdl^{(w)}$) の間の誤差比率 $\left(\frac{|mdl - mdl^{(w)}|}{mdl} \times 100\right)$ (%) をモデル選択精度の評価値とした。誤差比率の算出には各条件毎の 10^3 個の観測シンボル系列を使用し結果を平均化した。表の Ent. 1, Ent. 2, Ent. 3 は、エントロピーが大きい順のインデックスである。概してマルコフ次元が高くなるほど、またエントロピーが小さくなるほど、誤差比率は大きくなっているが、最大値で 1.2% であり、 ξ の値が 0.997 以上では、いずれの場合においても 1% 以下になっている。これらの実験結果は、忘却重み係数を適切に設定すれば、定常区間における見かけ上の観測シンボル数の減少によるモデル選択精度の劣化は小さいことから、最適なモデルパラメータが局所的に

大きく変わるような問題⁸⁾においては、WMDL 基準はその追従性から MDL 基準より有効になることの裏付けとなるものである。

5. む す び

本論文では、WMDL 基準、予測 WMDL 基準に関する性質を理論と実験の両面から検討した。

まず、WMDL 基準、予測 WMDL 基準に関する性質を証明した。これらは、推定モデルパラメータの漸近特性、パラメータ推定誤差の分布特性、パラメータ推定誤差分散の下限に関するものであった。また、見かけ上の観測シンボル数の増加に伴い予測 WMDL の期待値は WMDL の期待値の下限に漸近的に近づくが、観測シンボル数無限大においても見かけ上の観測シンボル数は有限であることから、これらの差は、無限小まで近づくことはできないことを証明した。

次に、シミュレーションにより、観測シンボル系列の定常区間において、忘却重み係数の値と予測 WMDL 基準に基づくモデル選択精度の関係に関する実験的考察を行った。その結果、忘却重み係数を適切に設定すれば、定常区間における見かけ上の観測シンボル数の減少によるモデル選択精度の劣化は小さく、最適なモデルパラメータが局所的に大きく変わるような問題においては、WMDL 基準はその追従性から、MDL 基準より有効となることの裏付けが示された。

〔参 考 文 献〕

- 1) R. C. Pasco: "Source Coding Algorithm for Fast Data Compression", Ph. D. thesis, Dept. of Electrical Engineering, Stanford Univ. (May 1976)
- 2) G. G. Langdon and J. J. Rissanen: "Compression of Black-White Images with Arithmetic Coding", IEEE Trans. Commun., COM-29, pp. 858-867 (1981)
- 3) G. G. Langdon: "An Introduction to Arithmetic Coding", IBM J. Res. & Dev., 28, 2, pp. 135-149 (Mar. 1984)
- 4) 坂元ほか: "情報量統計学", 共立出版 (1983)
- 5) J. Rissanen: "Modeling by Shortest Data Description", Automatica, 14, pp. 465-471 (1978)
- 6) J. Rissanen: "An Universal Prior for Integers and Estimation by Minimum Description Length", Ann. Statist., 11, pp. 416-431 (1983)
- 7) J. Rissanen "Universal Coding, Information, Prediction and Estimation", IEEE Trans. Inf. Theory, IT-30, pp. 629-636 (July 1984)
- 8) 松代: "忘却重み係数を含む MDL 基準とその適応的符号化モデル選択への応用", テレビ誌, 49, 11, pp. 1513-1521 (1995)
- 9) J. Rissanen: "Stochastic Complexity and Modeling", Ann. Statist., 14, pp. 1080-1100 (1986)
- 10) 韓: "データ圧縮, 統計的推論, チューリング機械", 情報理論とその応用学会, データ圧縮ワークショップ, pp. 32-46 (Jan. 1988)

- 11) 伊藤: "情報源モデルの推定と MDL 基準", 情報理論とその応用学会, データ圧縮ワークショップ, pp. 223-232 (Jan. 1988)
- 12) 添田ほか: "信号処理の基礎と応用", 日新出版 (1979)
- 13) 松代: "忘却重み係数を含む MDL 基準に関する 2, 3 の性質", テレビ誌, 49, 12, pp. 1626-1629 (1995)
- 14) 伊藤: MDL 原理の研究動向について, テレビ学技報, IT92-129, pp. 7-14 (Mar. 1993)

〈付録〉 WMDL 基準, 予測 WMDL 基準の導出⁸⁾¹⁴⁾

忘却重み係数を含む MDL 基準, 予測 MDL 基準は, MDL 基準がその出発点としているモデルの確率分布 $\prod_{t=1}^n Q_{k,\theta}(s_t)$ の対数尤度 $\log \prod_{t=1}^n Q_{k,\theta}(s_t) = \sum_{t=1}^n \log Q_{k,\theta}(s_t)$ の代わりに, 重み付き対数尤度 $\sum_{t=1}^n w_t \log Q_{k,\theta}(s_t)$ を用いる基準と定義する。この場合, 最良なモデルパラメータ $\hat{\theta}$ は, 重み付き対数尤度についての最尤推定量

$$\hat{\theta} = \arg \min_{\theta} \left[- \sum_{t=1}^n w_t \log Q_{k,\theta}(s_t) \right] \quad (\text{付} \cdot 1)$$

である。モデルの確率分布 $Q_{k,\theta}$ の記述は, $\hat{\theta}$ の表現を有限精度で行うこと, すなわち $\hat{\theta}$ の第 j 成分 ($[0, 1]$ 区間の数値に正規化) を e_j の精度で区分化することによって行う。

$$[\hat{\theta}] = \hat{\theta} + e \quad (\text{付} \cdot 2)$$

ここで,

$\hat{\theta}$: 重み付き対数尤度についてのパラメータ θ の最尤推定量

$[\hat{\theta}]$: $\hat{\theta}$ の有限精度表現

$$e = (e_1, e_2, \dots, e_k)$$

この場合, モデルの確率分布は $Q_{k,[\hat{\theta}]}$ であり, e_j の精度で区分化した $[\hat{\theta}]$ の第 j ($j=1, 2, \dots, k$) 成分の記述には精度の逆数の数値を区別できる必要があり, この区別の対数記述長は $\log(1/e_j)$ であるから全記述長 $l_w(s^n)$ は

$$l_w(s^n) = - \sum_{t=1}^n w_t \log Q_{k,[\hat{\theta}]}(s_t) + \sum_{j=1}^k \log(1/e_j) \quad (\text{付} \cdot 3)$$

となる。 $-\sum_{t=1}^n w_t \log Q_{k,\theta}(s_t)$ が $\theta = \hat{\theta}$ の回りで 2 階連続微分可能とし, (付・2) 式を (付・3) 式の第 1 項に代入し 2 階微分項までの Taylor 展開を行う。 $-\sum_{t=1}^n w_t \log Q_{k,\theta}(s_t)$ は (付・1) 式から $\theta = \hat{\theta}$ の極値 (最小値) をとることから, 1 階微分項は 0 となり, 次の近似式が得られる。

$$l_w(S^n) = - \sum_{t=1}^n w_t \log Q_{k,\hat{\theta}}(s_t) \\ + (1/2) e'(W_n M) e \\ + \sum_{j=1}^k \log(1/e_j) \quad (\text{付} \cdot 4)$$

ここで

$$W_n = \sum_{t=1}^n w_t \\ M = - \left(\sum_{t=1}^n w_t \frac{\partial^2 \log Q_{k,\hat{\theta}}(s_t)}{\partial \theta \partial \theta'} \right) / W_n$$

$I_w(S^n)$ を最小にする e_j すなわち全記述長を最小にするパラメータ区分化の精度は、(付・4)式の第2項、第3項の和を e_j について偏微分することで以下のように求まる。

$$e_j = d_j / \sqrt{W_n} \quad (\text{付} \cdot 5)$$

ここで

d_j : M によって決まる定数

(付・5)式を(付・4)式に代入し、MDL基準の場合と同様、(付・4)式の第2項と、第3項からくり出される定数項 $-\sum_{j=1}^k \log d_j$ を無視すると

$$l_w(S^n) = - \sum_{t=1}^n w_t \log Q_{k,\hat{\theta}}(s_t) \\ + (1/2) k \log W_n \quad (\text{付} \cdot 6)$$

となり、この式を k について最小化したものが、次式に示す WMDL 基準となる。

$$WMDL(S^n) = \min_k \left[- \sum_{t=1}^n w_t \log Q_{k,\hat{\theta}}(s_t) \right. \\ \left. + (1/2) k \log W_n \right] \quad (\text{付} \cdot 7)$$

予測 WMDL 基準は、予測 MDL 基準に重み付き対数尤度の定義を適用することで以下の式として得られる。

$$WMDL_p = - \sum_{t=1}^n w_t \log Q_{\hat{k}(t-1), \hat{\theta}(t-1)}(s_t) \quad (\text{付} \cdot 8)$$

まつしろ のぶひと
松代 信人 1980年、電気通信大学卒業。
1982年、同大学院修士課程修了。同年、沖
電気工業(株)に入社。以来、図形処理、画像
処理の研究開発に従事。現在、(株)沖データ
研究所所属。正会員。