J Jpn Ind Manage Assoc **50**, 474–485, 2000

Original

An Analysis on Error Rate of Statistical Model Selection Based on Bayes Decision Theory

Masayuki GOTOH,[1] Toshiyasu MATSUSHIMA[1] and Shigeichi HIRASAWA[1]

Abstract

Statistical model selection is one of the most important problems in statistics, and many works have left essential results. The conventional information criteria for model selection, such as the Akaike information criterion(AIC), the Bayesian information criterion(BIC), and minimum description length(MDL) were derived from different viewpoints. Many other model selection criteria have also been reported from various viewpoints. On the other hand, if we specify the model class and assume prior probabilities, then we can acquire Bayes optimal model selection for a finite number of samples based on Bayes decision theory. Furthermore, we can assume the various loss function adapting the purpose of model selection for practical cases. In this paper, we analyze the asymptotic properties of stasistical model selection based on Bayes decision theory. At first, we formulate Bayes optimal solution based on Bayes decision theory. In this formulation, we introduce the general loss function for practical problems. Moreover, we analyze the upper limits of the error rate of the model selection.

Key words: statistical model selection, Bayes decision theory, error rate of model selection

[1]Waseda University
Received: July 12, 1999
Accepted: October 12, 1999

論　　文

ベイズ決定理論に基づく統計的モデル選択の選択誤り率に関する解析

後藤正幸[1], 松嶋敏泰[1], 平澤茂一[1]

本稿では，ベイズ決定理論に基づくモデル選択の漸近的性質について考察する．まず，モデル選択問題をベイズ決定理論に基づいて定式化する．その際，より現実的問題に対応できる形の一般的な損失関数を導入する．さらに，一般的なモデル族を仮定して，この選択方法の選択誤り確率について，その上界の解析を試みる．

キーワード： 統計的モデル選択，ベイズ決定理論，選択誤り確率

1. はじめに

統計的モデル選択問題〔10〕は，理論的に興味深いテーマであるだけでなく，非常に広範囲の応用を持つ極めて重要な問題の一つである．とくに候補であるモデル族が階層的な構造を持っている場合，次数の低いモデルはより次数の高い (表現能力の高い) モデルに含まれるという性質が問題 (あるいは問題設定) を難しくしている．モデル選択に関しては，AIC〔1〕, BIC〔14〕, MDL〔13〕といった情報量規準を代表として，様々な観点からの研究〔12],〔16],〔17〕がなされている．一方，もし事前分布が仮定できれば，ベイズ決定理論〔2],〔3〕に基づくモデル選択規準が定式化できる〔7〕．このベイズ規準では様々な損失関数に対し，有限長のデータに対してベイズ最適なモデル選択とパラメータ推定が可能となる．ベイズ決定理論に基づくモデル選択は，ベイズ統計の枠組みで統一した議論が行われるという点で一貫性があり，BIC や MDL でも事前分布が仮定されているように，モデル選択問題に対して非常に相性のよい解法の一つである．

本稿では，真の構造 (真のモデル:m^* で表す) の発見を目的とする統計的モデル選択問題に対し，ベイズ決定理論によって最適な推定を行うことを考える．ここで考慮に入れるべきことは，実際の応用ではデータからモデルを推定した後，この結果をもとにアクションをとることである．従って，損失関数は現実的応用を見据えたものであり，かつ一般性があることが望まれる．そこで本稿ではまず，0-1 損失を含み，問題に応じて設定できる一般化された損失関数を定義し，ベイズ決定理論に基づくモデル選択を定式化する．

一方，真のモデル m^* の発見を目的としたモデル選択基準の一つの合理的な評価基準は，

$$P_e^* = P^*\{\hat{m} \neq m^*\}, \tag{1}$$

で与えられる選択誤り率であろう．ここで，$\hat{m}$ はモデル選択によって選ばれたモデル (推定値) であり，P^* は真の確率測度である．情報量規準によるモデル選択に対しては，この選択誤り率の漸近的特性を解析した結果が示されており〔6],〔16〕，モデル選択基準の特性として選択誤り率を議論することは意味があると思われる．そこで本稿では，定式化したベイズ規準によるモデル選択に関して，この選択誤り率の漸近的上界を解析することにより，その漸近的性能について考察を与える．

2. 問題設定

情報源 (母集団) からの長さ n のデータ系列を x^n とし，確率変数 X^n の 1 つの実現値を表すものとする．また，確率モデルのクラスを $\mathcal{P} = \left\{p(\cdot|m,\theta^{k_m})\right\}$ とする．$p(\cdot)$ は確率分布であり，確率モデルに応じて確率あるいは確率密度のどちらかを表すものとする．m は 1 つのパラメトリッククラスを特定する離散ラベルで，モデルと呼び，可算有限集合 $\mathcal{M}$ の要素とする．すなわち，$m \in \mathcal{M}$ かつ $|\mathcal{M}| < \infty$ である．$\theta^{k_m} = (\theta_1^{k_m}, \theta_2^{k_m}, \cdots, \theta_{k_m}^{k_m})^T$ はモデル m の k_m次元パラメータであり，$\theta^{k_m} \in \Theta^{k_m}$, $\Theta^{k_m} \subseteq \mathcal{R}^{k_m}$ とする．ただし，上付きの T は転置を表す．データ x^n は真の分布 $p^*(\cdot) = p(\cdot|m^*, \theta^{k^*_{m^*}})$ に従って得られるものとする．m^* は真のモデル，$\theta^{k^*_{m^*}}$ は真のパラメータと呼ぶ．モデル選択の目的は，x^n に基づき $\mathcal{M}$ から m を選ぶこと，すなわち m の推定量 $\hat{m}$ を得ることである．

具体的なモデル選択問題の一例として，品質管理の分野で特によく研究されている母集団 (情報源) パラ

---

[1] 早稲田大学
受付：1999 年 7 月 12 日，再受付（2 回）
受理：1999 年 10 月 12 日

メータの変化位置検出問題を示す[1]〔18〕.

例 1 (〔7〕) ある特性値 (説明変数) が何水準 (カテゴリ) かに分かれており，目的変数 X の確率分布がそれぞれのカテゴリで条件付けられているモデル族を考える．そのカテゴリ間で分布のパラメータが変化している可能性があり，どのカテゴリ間でパラメータが変化しているかをデータから推定するものとする．この問題でパラメータに変化があるカテゴリ間の位置とそれぞれのパラメータの組は，一つの確率モデルととらえることができ，半順序を持つ階層構造モデル族からのモデル選択問題として定式化することができる．

r 水準のカテゴリを $C_1, C_2, \ldots, C_r$とする．X の確率分布は，クラス $\{p(x|\theta^k)|\theta^k \in \Theta^k\}$ に属するものとする．ここで，θ^k は k 次元パラメータであり，このパラメータは，カテゴリ間で変化する可能性がある．問題は，全ての水準から得られたデータに基づいて，パラメータの変化位置とそれぞれのパラメータ値を推定することである．

変化回数 $s-1$ の場合のパラメータの変化位置を $\tau_1, \tau_2, \cdots, \tau_{s-1}$ で表す．ただし，$s \in \{1, \cdots, r\}$ である．全ての変化位置の集合をモデルと呼び，インデックス m で表す．すなわち，$s \geq 2$ に対し，

$$m = \left\{\tau_1, \tau_2, \cdots, \tau_{s-1}\right\} \tag{2}$$

とし，$s=1$ のとき $m=\phi$ (変化なし) とする．変化回数 $s-1$ のモデルのパラメータ数は $k_m = sk$ である．m の集合を $\mathcal{M}$, s の集合を $S=\{0,1,\cdots,r\}$, s を固定したときの m の集合を $T(s)$, モデル m のパラメータ集合を Θ_m, モデル m の次元 (パラメータ数) を $k_m = sk$ とする．このとき，全モデルクラスは，

$$\mathcal{P} = \left\{p(\cdot|m, \theta^{sk}); s \in S, m \in T(s), \theta^{sk} \in \Theta_m\right\}. \tag{3}$$

で与えられる．変化位置検出問題は，クラス $\mathcal{P}$ から一つの確率モデルを選択する問題と等価である． □

上の例にも示されたように，モデル選択が議論される現実的問題のほとんどは，全順序の階層構造，あるいは半順序の階層構造を持ったモデル族であることがわかる．例えば，AR モデルは全順序を持つ階層モデル族，重回帰モデルは半順序の階層モデル族に属する．以下にこれらを包含する広義の半階層モデル族を定義する．

[1]このような問題の例として，薬効問題がある．ある薬品の投入量に対して，その薬の効果がどのように変化するかを調べために，投入量を何段階かに分けて実験を行い解析する問題である．

定義 1 (半階層モデル族)

モデル m で構成される確率分布のクラスを $\mathcal{H}^{k_m}$ とする．すなわち，

$$\mathcal{H}^{k_m} = \left\{p(\cdot) = p(\cdot|m, \theta^{k_m})\middle|\theta^{k_m} \in \Theta^{k_m}\right\}. \tag{4}$$

とする．さらにそれぞれのモデルに関し，$k_{m1} < k_{m2}$ となる $m1, m2 \in \mathcal{M}$ に対し，

$$\mathcal{H}^{k_{m1}} \subset \mathcal{H}^{k_{m2}}, \tag{5}$$

という入れ子構造が存在してもかまわないものとする．この階層構造は，$\mathcal{M}$ の中で全順序構造でも半順序構造でもよい．また，入れ子構造を全く持たない分離的なモデル族，すなわち $\forall m1, \forall m2 \in \mathcal{M}$ に対して $\mathcal{H}^{k_{m1}} \cap \mathcal{H}^{k_{m2}} = \phi$ (空集合) の場合も含めて，半階層モデル族を $\mathcal{H}$

$$\mathcal{H} = \cup_{m \in \mathcal{M}} \mathcal{H}^{k_m}, \tag{6}$$

と定義する[2]． □

本稿の定義した半階層モデル族は，階層モデル族や分離型のモデル族を含む極めて一般的なモデル族を指している．$H(p_m^{\theta^{k_m}})$ を $\log p(\cdot|m, \theta^{k_m})$ の平均レート，すなわち

$$H(p_m^{\theta^{k_m}}) = \lim_{n\to\infty} \frac{1}{n} E^*\left[\log p(X^n|m, \theta^{k_m})\right], \tag{7}$$

と定義する．ここで，$E^*[\cdot]$ は真の確率測度 $p^*(\cdot) = p(\cdot|m^*, \theta^{k^*_{m^*}})$ に関する期待値である．また $D\left(p_{m^*}^{\theta^{k^*_{m^*}}}; p_m^{\theta^{k_m}}\right)$ を，

$$D\left(p_{m^*}^{\theta^{k^*_{m^*}}}; p_m^{\theta^{k_m}}\right) = \lim_{n\to\infty} \frac{1}{n} E^*\left[\log \frac{p(X^n|m^*, \theta^{k^*_{m^*}})}{p(X^n|m, \theta^{k_m})}\right], \tag{8}$$

で定義される $p(\cdot|m, \theta^{k_m})$ の $p(\cdot|m^*, \theta^{k_{m^*}})$ に対する KL 情報量とする[3]．さらに，$\forall m \in \mathcal{M}$ に対して，$\theta^{k^*_m}$ を

$$\theta^{k^*_m} = \arg\max_{\theta^{k_m}} \lim_{n\to\infty} \frac{1}{n} E^*\left[\log p(X^n|m, \theta^{k_m})\right],$$

[2]とくに階層関係が全順序のとき，入れ子型モデル族 (nasted model class) と呼ぶ．入れ子型モデル族の例としては AR モデル の他に有限マルコフモデルがある．重回帰モデル，FSMX モデルなどは半順序の階層構造を持つモデル族である．半階層モデル族を $\{p(\cdot|m, \theta^{k_m})|m \in \mathcal{M}, \theta^{k_m} \in \Theta^{k_m}\}$ とも記述する．

[3]$-\infty < H(p_{m^*}^{\theta^{k^*_{m^*}}}) < \infty$ かつ $-\infty < H(p_m^{\theta^{k_m}}) < \infty$ であり，確率測度 $p(\cdot|m, \theta^{k_m})$ のゼロ集合 (確率測度 0 を持つ集合) が $p(\cdot|m^*, \theta^{k^*_{m^*}})$ のそれと等しければ，$D\left(p_{m^*}^{\theta^{k^*_{m^*}}}; p_m^{\theta^{k_m}}\right) = H(p_{m^*}^{\theta^{k^*_{m^*}}}) - H(p_m^{\theta^{k_m}})$ が成り立つ．

(9)

と定義する．データ系列 x^n は真の分布 $p^*(x^n) = p(x^n|m^*, \theta^{k^*_{m^*}})$ に従って発生するものとするが，半階層モデル族では m^* を次のように定義する必要がある．すなわち，$\mathcal{H}^{k_{m^*}} \neq \mathcal{H}$ の場合に $p(x^n|m, \theta^{k_m}) = p(x^n|m^*, \theta^{k^*_{m^*}})$ をみたす $m(\neq m^*)$ と θ^{k_m} が存在してしまうので，これまでの研究と同様に次のように定義する〔8〕,〔9〕.

定義 2 (真のモデル m^*) 真のモデル m^* は

$$m^* = \arg_m \min\left\{k_m \middle| \exists\theta^{k_m}, D\left(p^*; p_m^{\theta^{k_m}}\right) = 0\right\}, \quad (10)$$

で定義する．すなわち，$\mathcal{H}^{k_m} \subset \mathcal{H}^{k_{m^*}}$ に対しては

$$\forall\theta^{k_m} \in \Theta^{k_m}, \quad D\left(p^*_{m^*}; p_m^{\theta^{k_m}}\right) > 0, \quad (11)$$

が成り立つ． □

一般的には，$D\left(p^*_{m^*}; p_m^{\theta^{k_m}}\right) = 0$ であるからといって，$\forall x^n, n = 1, 2, \cdots$ に対して必ずしも $p(x^n|m, \theta^{k_m}) = p^*(x^n)$ とは限らない．逆に，もし $\forall x^n$, $n = 1, 2, \cdots$ に対して $p(x^n|m, \theta^{k_m}) = p^*(x^n)$ であれば $D\left(p^*; p_m^{\theta^{k_m}}\right) = 0$ は成り立つ．しかし，現実的応用で必要となってくるモデル族に関しては，$D\left(p^*; p_m^{\theta^{k_m}}\right) = 0$ は $p(x^n|m, \theta^{k_m}) = p^*(x^n)$ と等価であるといえる．したがって，本稿では $D\left(p^*_{m^*}; p_m^{\theta^{k_m}}\right) = 0$ は $\forall x^n$, $n = 1, 2, \cdots$ に対して $p(x^n|m, \theta^{k_m}) = p^*(x^n)$ を意味するものとする．このとき，式 (5),(10) より，$\mathcal{H}^{k_{m^*}} \subset \mathcal{H}^{k_m}$ をみたす m に対し，

$$p(x^n|m, \theta^{k^*_m}) = p(x^n|m^*, \theta^{k^*_{m^*}}), \quad (12)$$

が成り立つ．

3. ベイズ決定理論による定式化

これまでモデル選択問題に対して主に適用されている 0-1 損失は，真のモデルに損失 0，それ以外のモデルに損失 1 を与えたものであり，ベイズ的に選択誤り率を最小化する規準を導くという点で一つの客観的な基準を与えている．しかし，適用場面によっては誤り率による評価だけでなく，モデル間に何らかの擬距離構造を考えることができる場合があり，なるべく真のモデルに近いモデルを選択するための規準も有用と考えられる〔11〕．また，真のモデルよりも次数を少なく見積もる誤りより，多く見積もる誤りを避けたいという状況や，この逆の状況も考えられる〔18〕．そこで本稿では，このような要求に応じたベイズ規準を検討し，これらのモデル間の距離構造も表現できる一般的な損失関数を提案，この損失関数に対するベイズ最適なモデル選択を定式化する．

0-1 損失は，モデル間の距離として，同じモデルには 0，そうでないときは 1 を考えていることになるが，これを含み，より一般的なモデル間の距離を表現できる形として，次の損失関数を定義する．

$$V(m, Am) = \begin{cases} 0 & ; Am = m \\ a(m, Am) & ; Am \neq m \end{cases}$$

ここで，$a(m, Am)$ は $a(m, Am) > 0$，かつ $\max_{m,Am} a(m, Am) < \infty$ をみたす正関数，Am は決定関数を表す．$a(m, Am)$ は有界正関数であれば任意であるので，様々な形をとることができる．例えば，真のモデルからパラメータ数が離れる程大きな損失を与えることを考えるならば，$a(m, Am) = |k_m - k_{Am}|^2$ などと定義することができる．また，$m \neq Am$ のとき $a(m, Am) = 1$ とすれば，0-1 損失となる．その他の具体的に損失を与える方法については，〔11〕,〔18〕などを参照．また本来は，パラメータについても決定のための損失を定義する必要があり，二乗誤差損失が実用上からも適当であると考えられるが，ここでは一般性を持たせるために特に限定せず，m の決定のみを議論の対象とする．

この損失関数のもとで，データ x^n に条件付けられたベイズ最適な決定は，

$$BV(Am) = \sum_{m'} V(m', Am) P(m'|x^n), \quad (13)$$

の最小化によって与えられ，$BV(Am)$ を最小化するベイズ最適なモデル m^{BD} は，

$$m^{BD} = \arg_{Am} \min \sum_{m'} V(m', Am) P(m'|x^n), \quad (14)$$

となる．ただし，$P(m|x^n)$ はモデル m の事後確率であり，

$$P(m|x^n) = \frac{p(x^n|m)P(m)}{\sum_{m \in \mathcal{M}} p(x^n|m)P(m)}, \quad (15)$$

$$p(x^n|m) = \int_{\theta^{k_m}} p(x^n|m, \theta^{k_m}) f(\theta^{k_m}|m) d\theta^{k_m}, \quad (16)$$

で与えられる．$P(m)$ はモデル m の事前確率，$f(\theta^{k_m}|m)$ は θ^{k_m} の事前密度である．$m \neq Am$ のとき $a(m, Am) = 1$ とした 0-1 損失を適用した場合には，

$$\sum_{m'} V(m', Am) P(m'|x^n) = 1 - P(Am|x^n) \quad (17)$$

となるので，事後確率 $P(m|x^n)$ を最大化する m がベイズ最適となる．結局，式 (15) より m^{BD} は

$$m^{BD} = \arg_m \max\left\{\log\int_{\theta^{k_m}} p(x^n|m,\theta^{k_m})f(\theta^{k_m}|m)d\theta^{k_m} + \log P(m)\right\}, \quad (18)$$

と与えられる．

例 2 例 1 のパラメータの変化時点検出問題において，シンボル 0 を確率 θ で，シンボル 1 を確率 $1-\theta$ で出力する二項分布族を考える $(k=1)$．n_i をカテゴリ i のサンプル数，y_i をそのうちのシンボル 0 の個数とする．$x_i=(y_i, n_i-y_i)$ である．このとき s 次元パラメータを持つモデル $m=\{\tau_1,\tau_2,\cdots,\tau_{s-1}\}$ の尤度関数は，

$$P(x^n|m,\theta^s) = \prod_{j=1}^{s}\theta_j^{\sum_{i=\tau_{j-1}+1}^{\tau_j} y_i}(1-\theta_j)^{\sum_{i=\tau_{j-1}+1}^{\tau_j}(n_i-y_i)}. \quad (19)$$

で与えられる．ただし，$\tau_{-1}=0$ である．ここで，θ_j の事前分布として共役なベータ分布

$$f(\theta_j) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \quad (20)$$

を仮定すれば，

$$\log\int_{\theta^s} P(x^n|m,\theta^s)f(\theta^s|m)d\theta^s + \log P(m) = \sum_{j=1}^{s}\log\frac{\Gamma\left(\sum_{i=\tau_{j-1}+1}^{\tau_j} y_i+\alpha\right)\Gamma\left(\sum_{i=\tau_{j-1}+1}^{\tau_j}(n_i-y_i)+\beta\right)}{\Gamma\left(\sum_{i=\tau_{j-1}+1}^{\tau_j} n_i+\alpha+\beta\right)} + s\log\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} + \log P(m). \quad (21)$$

となる．したがって，式 (21) を使って $BV(Am, A\theta_{Am})$ を計算すれば，ベイズ最適なモデル選択が可能となる．0-1 損失の場合は，式 (21) を最小化するモデルがベイズ最適である． □

例 3 例 2 のパラメータの変化時点検出問題において，最も簡単なモデルに対してその計算例を示そう．

カテゴリ数を 2 とし，このカテゴリ間にパラメータの変化がないとするモデルを m_0，変化があるとするモデルを m_1 とする．また，モデルの事前分布を $P(m_0)=P(m_1)=1/2$，パラメータの事前分布を $\alpha=\beta=1$ と，両方とも一様分布を仮定する．

それぞれのカテゴリに対し，10 ずつのサンプルを得て $(n_1=n_2=10)$，$y_1=3$, $y_2=5$ という結果を得たものとしよう．このとき，それぞれのモデルの事後確率は

$$P(m_0|x^n) \propto \frac{8!12!}{21!} \quad (22)$$

$$P(m_1|x^n) \propto \frac{3!7!}{11!}\frac{5!5!}{11!} \quad (23)$$

より，$P(m_0|x^n)=0.58$, $P(m_1|x^n)=0.42$ を得る．

0-1 損失の場合には事後確率の高いモデル m_0 がベイズ最適となる．実際，$BV(m_0)=0.42$, $BV(m_1)=0.58$ で，若干 m_0 の方が損失が小さい．一方，もし損失として $V(m_0,m_1)=1$, $V(m_1,m_0)=2$ と，m_1 が真であったときに m_0 を選んでしまう損失を大きくすると，$BV(m_0)=0.84$, $BV(m_1)=0.58$ となって m_1 の方がベイズ最適で選ばれる．これは，事後確率がほぼ同程度の値であるのに対し，m_1 が真であったときに m_0 を選んでしまう損失が大きいために生じた結果である．

このように，損失関数を適切に設定することにより，目的に合わせたモデル選択が可能となる． □

4. 漸近的性質の解析

本章では，広いクラスのモデル族に関して，前章で定式化したベイズ最適なモデル選択の漸近的な特性を解析する．そのために，まず 2 つの情報行列 $I(\theta^{k_m}|m)$, $J(\theta^{k_m}|m)$ を定義する[4]．

$$I(\theta^{k_m}|m) = \lim_{n\to\infty}\frac{1}{n}E^*\left[\frac{\partial\log p(X^n|m,\theta^{k_m})}{\partial\theta^{k_m}}\frac{\partial\log p(X^n|m,\theta^{k_m})}{(\partial\theta^{k_m})^T}\right], \quad (24)$$

$$J(\theta^{k_m}|m) = \lim_{n\to\infty} -\frac{1}{n}E^*\left[\frac{\partial^2\log p(X^n|m,\theta^{k_m})}{\partial\theta^{k_m}(\partial\theta^{k_m})^T}\right]. \quad (25)$$

4.1 仮　定

まず，ベイズ理論で設定される事前分布に関して以下を仮定する．

条件 1 (1) $\forall m\in\mathcal{M}$ に対して，$P(m)>0$.

(2) $\forall m\in\mathcal{M}$, $\forall\theta^{k_m}\in\Theta^{k_m}$ に対し，$f(\theta^{k_m}|m)>0$，かつ $f(\theta^{k_m}|m)$ は θ^{k_m} に関して 3 回微分可能とする．

[4] $p^*(\cdot)\in\mathcal{H}^{k_m}$ のとき $J(\theta^{k_m^*}_{}|m)=I(\theta^{k_m^*}|m)$ となる．

また対象とする半階層モデル族に関して次を仮定する．

条件 2 (1) (有界性) $\forall \theta^{k_m} \in \Theta^{k_m}$, $\forall m \in \mathcal{M}$ に対して，$0 < \det I(\theta^{k_m}|m) < \infty$ かつ $0 < \det J(\theta^{k_m}|m) < \infty$ とする．また，$\forall m \in \mathcal{M}$, $\forall \theta^{k_m} \in \Theta^{k_m}$ に対して，

$$0 \leq D\left(p_{m^*}^{\theta^{k^*_{m^*}}}; p_m^{\theta^{k_m}}\right) < \infty, \tag{26}$$

かつ，$|H(p_m^{\theta^{k_m}})| < \infty$ であるとする．

(2) (連続性) $\forall \theta^{k_m} \in \Theta^{k_m}$, $\forall m \in \mathcal{M}$ に対して，$I(\theta^{k_m}|m)$, $J(\theta^{k_m}|m)$ は Θ^{k_m} 上で θ^{k_m} について微分可能とする．

(3) (健全性) パラメータ空間上に半径 $\delta > 0$ の球 $B_\delta(\theta^{k^*_m}) = \left\{\theta^{k_m} \in \Theta^{k_m} \middle| \|\theta^{k_m} - \theta^{k^*_m}\| < \delta\right\}$ を定義する．このとき，$\forall m \in \mathcal{M}$ と $\forall \delta > 0$, 条件 *1* を満たす事前密度 $f(\theta^{k_m}|m)$ に対して，事後密度関数は

$$\int_{\theta^{k_m} \in B_\delta(\theta^{k^*_m})} f(\theta^{k_m}|x^n, m) d\theta^{k_m} \to 1, \quad a.s. \tag{27}$$

をみたす．

(4) (漸近正規性) $\forall m \in \mathcal{M}$ に対して，最尤推定量 $\hat{\theta}^{k_m} = \hat{\theta}^{k_m}(x^n)$ は漸近正規性をみたす〔*10*〕,〔*17*〕．すなわち，$\zeta^{k_m} = \sqrt{n}\left(\hat{\theta}^{k_m}(x^n) - \theta^{k^*_m}\right)$ の分布は平均 $\mathbf{0}$, 共分散 $\{K(\theta^{k^*_m}|m)\}^{-1}$ の k_m-次元正規分布に分布収束する[5]．ここで，$\{K(\theta^{k_m}|m)\}^{-1}$ は

$$\{K(\theta^{k_m}|m)\}^{-1} = \{J(\theta^{k_m}|m)\}^{-1} I(\theta^{k_m}|m) \{J(\theta^{k_m}|m)\}^{-1}. \tag{28}$$

で与えられる行列とする．

(5) (重複対数の法則) $\forall m \in \mathcal{M}$, $\forall \theta^{k_m} \in \Theta^{k_m}$ に対して，次の *3* つの推定量は重複対数の法則をみたす:

$$\hat{\theta}^{k_m} = \theta^{k^*_m} + O\left(\frac{(\log\log n)^{1/2}}{\sqrt{n}}\right), \quad a.s. \tag{29}$$

$$\frac{1}{n}\log p(x^n|m, \theta^{k_m}) = H(p_m^{\theta^{k_m}}) + O\left(\frac{(\log\log n)^{1/2}}{\sqrt{n}}\right), \quad a.s. \tag{30}$$

$$\frac{1}{n}\frac{\partial^2 \log p(x^n|m,\theta^{k_m})}{\partial\theta^{k_m}(\partial\theta^{k_m})^T} = -J(\theta^{k_m}|m) + O\left(\frac{(\log\log n)^{1/2}}{\sqrt{n}}\right), \quad a.s. \tag{31}$$

□

[5]厳密には，R_ζ を任意の直方体として，測度 $p^*(\zeta^{k_m})$ にもとづく確率が〔10〕,〔17〕

$$P^*\left\{\zeta^{k_m} \in R_\zeta\right\} = \int_{\zeta^{k_m} \in R_\zeta^n} p^*\left\{d\zeta^{k_m}\right\} \to$$

$$\frac{\sqrt{\det K(\theta^{k^*_m}|m)}}{(2\pi)^{k_m/2}} \int_{\zeta^{k_m} \in R_\zeta} e^{-\frac{1}{2}\|\zeta^{k_m}\|^2_{K(\theta^{k^*_m}|m)}} d\zeta^{k_m},$$

をみたすことを仮定している．

(3) の健全性は，n が大きくなると，事後確率は真のパラメータの近傍に集中してくることを述べており，例えば〔2〕,〔4〕において同類の仮定が示されている．Θ^{k_m} が有界で Fisher 情報行列が退化していなければ健全性は満足される．しかし，有界でない場合でも，正規分布族と共役事前分布などを当てはめればすぐ明らかなように，通常有用な分布族はこの条件を満たすものとなっている．例 1 の分布族が (3) を満たすことも明らかである．(3) の必要条件に関しては〔2〕参照．

(4) の最尤推定量の漸近正規性と (5) の重複対数の法則については，最も基本的な真のパラメータがパラメータ空間内に存在する場合について〔5〕で詳しく議論されている．モデル選択で必要となってくるように，真の分布を表現するパラメータが集合内にない場合 (真のモデルより次数の低いモデルでは，真の分布を表現できないので，このようなケースが生じる) については，〔10〕,〔17〕を参照．これらの仮定は，実務的に必要となる多くのモデル族で満足されることがわかる．

例えば，二項分布族は，そのパラメータ (生起確率) の範囲を (0,1) とすれば，上の条件 2 をみたすことがわかり〔5〕,〔10〕，例 2 で考察した二項分布に対するパラメータの変化時点検出問題はこの二項分布の組み合わせで与えられることから，明らかに条件 2 をみたす．正規分布に対するパラメータ変化時点検出問題においても，同様に条件 2 はみたされる．また，正規雑音を仮定した重回帰モデル，AR モデルも条件 2 はなりたつ〔5〕,〔10〕．

4.2 事後確率の漸近正規性

本節では，ベイズ規準に基づくモデル選択の漸近解析に本質的となるパラメータの事後確率密度の漸近正規性〔2〕を示す．具体的には条件1と2のもとで，以下の定理がなりたつ．

補題 1 条件1, 2のもとで，$\forall m \in \mathcal{M}$ に対し，以下の式が成り立つ．$\xi^{k_m} = \sqrt{n}\left(\theta^{k_m} - \hat{\theta}^{k_m}\right)$ の事後密度は次式をみたす．

$$f_\xi(\hat{\xi}^{k_m}|x^n, m) \to \frac{\sqrt{\det J(\hat{\theta}^{k_m}|m)}}{(2\pi)^{k_m/2}}, \quad a.s. \quad (32)$$

ただし，$\hat{\xi}^{k_m} = 0$ は $\hat{\theta}^{k_m}$ に対応する ξ 空間上の原点，$f_\xi(\xi^{k_m}|x^n, m)$ は θ^{k_m} を変換した ξ^{k_m} 空間上の事後確率密度であり，

$$f_\xi(\xi^{k_m}|x^n, m) = \frac{1}{\sqrt{n}^{k_m}} f(\theta^{k_m}|x^n, m). \quad (33)$$

で与えられる．

(証明) 付録1参照． □

〔2〕, pp.287 - 294は，事後確率最大推定量を用いた漸近正規性の議論をしているが，定理1は最尤推定量を用いた漸近展開となっている．最尤推定量を用いた展開をすることにより，モデル選択問題の漸近的性質の解析において，すでによく知られている最尤推定量の漸近性質を用いた議論が可能となる．また，定理1は事後密度 $f_\xi(\hat{\xi}^{k_m}|x^n, m)$ の漸近式のみを示しているが，分布の収束についても同様に示すことが可能である．本稿の以下の議論では，$f_\xi(\hat{\xi}^{k_m}|x^n, m)$ の漸近式のみが使われる．

4.3 モデル選択の誤り率の上界

ここでは，条件1,2のもとで，$BV(Am, A\theta^{k_{Am}})$ を最小化するモデル選択の m に関する選択誤り率の上界を導出する．

定理 1 条件1,2のもとで，真のモデルの選択誤り確率 P_e^* は，次のように与えられる．

$$P_e^* \le O\left(\frac{1}{\left(1-\log n\right)^2}\right), \quad (34)$$

(証明) 付録2参照． □

定理1は，

$$P_e^* \le O\left(\frac{1}{(\log n)^2}\right), \quad (35)$$

ということを述べており，これは $BV(Am, A\theta^{k_{Am}})$ を最小化するモデル選択は漸近弱一致性を持つこと，すなわち，

$$m^{BD} \to m^*, \ in\ probability, \quad (36)$$

を示している．本稿では重複対数の法則が成り立つクラスを考えたが，〔8〕では同様の条件のもとで，BICやMDLが強一致性を持つことを示しており，本稿ではこのより強い結果を導くには至っていない．したがって，より厳しい上界式が存在する可能性があり，これは今後の課題である．

5. まとめ

本稿では統計的モデル選択をベイズ決定理論の面から考察し，一般的な損失関数に対するベイズ最適な決定方法を示した．また，そのような検出方法の漸近一致性と検出誤り率の上界を導出した．本稿で議論した半階層モデル族は，現実的応用で考えられる広いクラスの分布族を含むものとなっている．

参考文献

〔1〕 H. Akaike: "A New Look at the Statistical Model Identification", *IEEE Trans. on Auto. Contr.* AC-19, No.6, pp.716-722 (1974)

〔2〕 J.M.Bernardo and A.F.M.Smith: Bayesian Theory, John Wiley & Sons (1994)

〔3〕 G.E.P.Box and G.C.Tiao: Bayesian Inference in Statistical Analysis, John Wiley and Sons, Inc. (1992)

〔4〕 Chan-Fu Chen: "On Asymptotic Normality of Limiting Density Function with Bayesian Implications", *J. R. Statist. Soc.B*, 47, No.3, pp.540-546 (1985)

〔5〕 W.Feller: An Introduction to Probability and Its Applications, Volume 1 and 2, John Wiley & Sons, New York (1957, 1966)

〔6〕 M.Gotoh, T.Matsushima, and S.Hirasawa: "On Error Rates of Statistical Model Selection based on Information Criteria", *Proceeding of IEEE Interna. Symp. Inform. Theory* (1998)

〔7〕 M.Gotoh and S.Hirasawa: "Statistical Model Selection based on Bayes Decision Theory and Its Application to Change Detection Problem", *Production Economics*, Vol.60–61, pp.629–638 (1998)

〔8〕 E.J.Hannan and B.G.Quinn: "The Determination of the Order of an Autoregression", *J. R. Statist. Soc.* B, 41, No.2, pp.190-195 (1979)

〔9〕 韓 太舜, 小林欣吾: "情報と符号化の数理 (岩波講座応用数学 13), 岩波書店 (1994)

〔10〕 H.Linhart and W.Zucchini: Model Selection, John Wiley & Sons (1986)

〔11〕 中尾 峰, 後藤正幸, 松嶋敏泰, 平澤茂一：“統計的モデル選択問題における選択誤り確率に関する一考察”

信学技報, IT96 (1996)

〔12〕 D.S.Poskitt: "Precision, Complexity and Bayesian Model Determination", *J. R. Statist. Soc.* B, 49, No.2, pp.199-208 (1987)

〔13〕 J.Rissanen: "Modeling by Shortest Data Description", *Automatica*, Vol.46, pp.465-471 (1978)

〔14〕 C.Schwarz: "Estimating the Dimension of a Model", *Ann. Statist.*, 6, pp.461-464 (1978)

〔15〕 関 庸一, 橋本 巧: "MDL基準に基づく正規母集団変化時点検出に関する研究", 日本経営工学会誌, Vol.47, No.3, pp.192 - 198 (1996)

〔16〕 R.Shibata: "Consistency of Model Selection and Parameter Estimation", *Applied Probability Trust.*, pp.127-141 (1986) on State Decomposition", *IEICE Trans. Fundamentals*, Vol.E76-A, No.7, pp.1240-1251 (1993)

〔17〕 竹内 啓: "情報統計量の分布とモデルの適切さの基準", 数理科学, No.153, March, pp.12-18 (1976)

〔18〕 山下公子, 後藤正幸, 俵 信彦: "ベイズ的アプローチによるパラメータ変化時点の検出に関する一考察", 日本経営工学会 秋季大会予稿集, pp.105 -106 (1996)

付録 補題1の証明

まず, $L_n(\theta^{k_m}|m) = \log f(\theta^{k_m}|x^n, m)$, $B_\delta(\hat{\theta}^{k_m}) = \left\{\theta^{k_m} \in \Theta^{k_m} \middle| \|\theta^{k_m} - \hat{\theta}^{k_m}\| < \delta\right\}$,

$$\Sigma_n = -\left(L_n''(\hat{\theta}^{k_m}|m)\right)^{-1}$$
$$= -\left(\left.\frac{\partial^2 L_n(\theta^{k_m}|m)}{\partial\theta^{k_m}(\partial\theta^{k_m})^T}\right|_{\theta^{k_m}=\hat{\theta}^{k_m}}\right)^{-1}. \quad (37)$$

と定義する. $\forall\theta^{k_m}$ に対してテーラー展開より,

$$f(\theta^{k_m}|x^n, m)$$
$$= f(\hat{\theta}^{k_m}|x^n, m)\exp\left\{(\theta^{k_m} - \hat{\theta}^{k_m})^T L'(\hat{\theta}^{k_m}|m)\right\}$$
$$\cdot\exp\left\{\frac{1}{2}(\theta^{k_m} - \hat{\theta}^{k_m})^T(I+R_n)L''(\hat{\theta}^{k_m}|m)(\theta^{k_m} - \hat{\theta}^{k_m})\right\}, \quad (38)$$

を得る. ここで, R_n は

$$R_n = L_n''(\theta^{k_m+}|m)\{L_n''(\hat{\theta}^{k_m}|m)\}^{-1} - I, \quad (39)$$

で与えられ, θ^{k_m+} は θ^{k_m} と $\hat{\theta}^{k_m}$ でむすばれる直線上の点である. ここで, $L_n'(\hat{\theta}^{k_m}|m)$ は

$$L_n'(\hat{\theta}^{k_m}|m) = \left.\frac{\partial\log f(\theta^{k_m}|m)}{\partial\theta^{k_m}}\right|_{\theta^{k_m}=\hat{\theta}^{k_m}}, (40)$$

で与えられる.

一方, 条件1,(2) と条件2,(5), および

$$L_n''(\theta^{k_m}|m) = \frac{\partial^2\log p(x^n|m,\theta^{k_m})}{\partial\theta^{k_m}(\partial\theta^{k_m})^T} + \frac{\partial^2\log f(\theta^{k_m}|m)}{\partial\theta^{k_m}(\partial\theta^{k_m})^T}, \quad (41)$$

より, $\forall\theta^{k_m} \in \Theta^{k_m}$ に対して

$$-\frac{1}{n}L_n''(\theta^{k_m}|m) \to J(\theta^{k_m}|m),\ a.s. \quad (42)$$

が成り立つ. また, 最尤推定量に対する重複対数の法則より,

$$\hat{\theta}^{k_m} \to \theta^{k_m^*},\ a.s. \quad (43)$$

となるから, $J(\theta^{k_m}|m)$ は微分可能であるので、

$$L_n''(\theta^{k_m+}|m)\{L_n''(\hat{\theta}^{k_m}|m)\}^{-1}$$
$$\to J(\theta^{k_m+}|m)\{J(\hat{\theta}^{k_m}|m)\}^{-1},\ a.s. \quad (44)$$

も成り立っている. 一方, $J(\theta^{k_m}|m)$ の連続性から, $\hat{\theta}^{k_m}$ を固定した場合, $\forall\epsilon > 0$ に対してある正定数 $\delta > 0$ が存在して, $\forall\theta^{k_m} \in B_\delta(\hat{\theta}^{k_m})$ に対し,

$$I - A(\epsilon) \le J(\theta^{k_m}|m)\{J(\hat{\theta}^{k_m}|m)\}^{-1} \le I + A(\epsilon), \quad (45)$$

が成り立つ. ただし, I は $k_m \times k_m$ 基本行列, $A(\epsilon)$ は $\epsilon \to 0$ のときその最大固有値が 0 に収束する $k_m \times k_m$ 正定対称行列とする. ここで $J(\theta^{k_m}|m)$ の連続性から, $\epsilon \to 0$ とすると δ はいくらでも 0 に近づけられることに注意する. したがって, 式 (44), (45) と, $\hat{\theta}^{k_m} \to \theta^{k_m^*}$, $a.s.$ であることから, $\forall\epsilon > 0$ に対し正定数 $\delta > 0$ が存在して, $\forall\theta^{k_m} \in B_\delta(\hat{\theta}^{k_m})$ に対して $n \to \infty$ のとき

$$I - A(\epsilon) \le L_n''(\theta^{k_m}|m)\{L_n''(\hat{\theta}^{k_m}|m)\}^{-1}$$
$$\le I + A(\epsilon),\ a.s. \quad (46)$$

が成り立つ[6]. ここで, $\epsilon \to 0$ のとき正数 δ はいくらでも小さくとれる.

したがって, $n \to \infty$ のとき

$$P_n(\delta) = \int_{B_\delta(\hat{\theta}^{k_m})} f(\theta^{k_m}|x^n, m)d\theta^{k_m}, \quad (47)$$

はほとんど確実に

$$f(\hat{\theta}^{k_m}|x^n, m)\exp\{\bar{c}_f\delta\}$$
$$|\Sigma_n|^{1/2}|I - A(\epsilon)|^{-1/2}\int_{|z|<s_n}\exp\{-\frac{1}{2}z^Tz\}dz,$$

[6] 本稿では, ある事象 $\mathcal{A}_n$, $n = 1, 2, \cdots$ に対し, "$P^*(\cup_{n=1}^{\infty}\cap_{k=n}^{\infty}\mathcal{A}_k) = 1$" であることを, "$n \to \infty$ のとき $\mathcal{A}$, $a.s.$ が成り立つ", と記述する.

$$(48)$$

で上界され，

$$f(\hat{\theta}^{k_m}|x^n,m)\exp\{-\bar{c}_f\delta\}$$
$$|\Sigma_n|^{1/2}|I+A(\epsilon)|^{-1/2}\int_{|z|<t_n}\exp\{-\frac{1}{2}z^Tz\}dz, \quad (49)$$

で下界される．ここで，$\bar{c}_f$ は $\frac{\partial \log f(\theta^{k_m}|m)}{\partial\theta^{k_m}}$ の要素の最大絶対値であり，また s_n, t_n は，$\overline{a(\epsilon)}$ ($\underline{a(\epsilon)}$) と $\overline{l_n}$ ($\underline{l_n}$) をそれぞれ $A(\epsilon)$ と Σ_n の要素の最大固有値(最小固有値) として，$s_n=\delta(1-\overline{a(\epsilon)})^{1/2}/\underline{l_n}^{1/2}$ と $t_n=\delta(1-\underline{a(\epsilon)})^{1/2}/\overline{l_n}^{-1/2}$ で与えられる．

式 (44) と同様に

$$n\Sigma_n\to\{J(\hat{\theta}^{k_m}|m)\}^{-1},\quad a.s. \quad (50)$$

となるから，式 (43) より $\overline{l_n}\to 0, a.s.$, $\underline{l_n}\to 0, a.s.$ が得られ，これは $s_n\to\infty, a.s.$, $t_n\to\infty, a.s.$ を導く．よって，$n\to\infty$ のとき，

$$|I-A(\epsilon)|^{1/2}\exp\{-\bar{c}_f\delta\}\lim_{n\to\infty}P_n(\delta)$$
$$\le\lim_{n\to\infty}\frac{f_\xi(\hat{\xi}^{k_m}|x^n,m)(2\pi)^{k_m/2}}{\left\{J(\hat{\theta}^{k_m}|m)\right\}^{-1/2}}$$
$$\le|I+A(\epsilon)|^{1/2}\exp\{\bar{c}_f\delta\}\lim_{n\to\infty}P_n(\delta),\ a.s.(51)$$

が成り立つ．

以上より，$\epsilon\to 0$ のとき正数 δ はいくらでも小さくとれるので，$\forall\delta>0$ に対して，もし $\lim_{n\to\infty}P_n(\delta)=1$, $a.s.$ が成り立てば，

$$\lim_{n\to\infty}f_\xi(\hat{\xi}^{k_m}|x^n,m)\to\frac{\{\det J(\hat{\theta}^{k_m}|m)\}^{1/2}}{(2\pi)^{k_m/2}},\quad a.s. \quad (52)$$

は成り立つことが分かる．ここで，条件 2,(3) の健全性と $\hat{\theta}^{k_m}\to\theta^{k_m^*}$ $a.s.$ より，$\lim_{n\to\infty}P_n(\delta)=1$, $a.s.$ が成り立つので，証明が完結した．□

付録　定理1の証明

$BV(Am)=\sum_m V(Am,m)P(m|x^n)$ と定義すると，m^* が真でもあるにかかわらず，他のモデルを選択してしまう選択誤り確率 P_e^* は，

$$P_e^*=P^*\left[x^n\middle|\exists m\ne m^*\ \ BV(m^*)\ge BV(m)\right], \quad (53)$$

である．従って，$BV(m^*)\ge BV(m)$ となる事象の確率を評価して，ユニオンバウンドをとればよい．

ここで，$V_{max}=\max_{m,m'}V(m,m')=\max_{m,m'}a(m,m')$, $V_{min}=\min_{m,m'}a(m,m')$ と定義する．損失に関する前提より，$0<V_{min}\le V_{max}<\infty$ であることに注意．

まず，$V(m,m)=0$ であることから

$$BV(m^*)=\sum_{m'\ne m^*}V(m^*,m')P(m'|x^n)$$
$$=\sum_{m'\ne m^*,m}V(m^*,m')P(m'|x^n)+V(m^*,m)P(m|x^n) \quad (54)$$

$$BV(m)=\sum_{m'\ne m}V(m,m')P(m'|x^n)$$
$$=\sum_{m'\ne m,m^*}V(m,m')P(m'|x^n)+V(m,m^*)P(m^*|x^n) \quad (55)$$

が成り立つことが分かる．従って，

$$BV(m^*)-BV(m)$$
$$=V(m^*,m)P(m|x^n)-V(m,m^*)P(m^*|x^n)$$
$$\ge V_{min}P(m|x^n)-V_{max}P(m^*|x^n) \quad (56)$$

が導かれるので，$C=\frac{V_{min}}{V_{max}}$ として

$$CP(m|x^n)\ge P(m^*|x^n)\to BV(m^*)\ge BV(m)$$

である．

従って，

$$P^*\left[BV(m^*)\ge BV(m)\right]$$
$$\le P^*\left[CP(m|x^n)\ge P(m^*|x^n)\right]$$

と上界できるので，ユニオンバウンドをとって

$$P_e^*\le P^*\left[\exists m\ne m^*,\ \ CP(m|x^n)\ge P(m^*|x^n)\right]$$
$$\le\sum_{m\ne m^*}P^*\left[\log\frac{P(m^*|x^n)}{P(m|x^n)}\le\log C\right]$$

という形で誤り確率が上界できる．

一方，補題1より，

$$f(\hat{\theta}^{k_m}|x^n,m)=\frac{\sqrt{n}^{k_m}\sqrt{\det J(\hat{\theta}^{k_m}|m)}}{\sqrt{2\pi}^{k_m}}+o\left(\sqrt{n}^{k_m}\right),\ a.s. \quad (57)$$

が得られる．ベイズの定理

$$p(x^n|m)=$$

$$\frac{p(x^n|m,\hat{\theta}^{k_m})f(\hat{\theta}^{k_m}|m)}{f(\hat{\theta}^{k_m}|x^n,m)},\ a.s. \tag{58}$$

より，式(57)を式(58)に代入すると，$\forall m \in \mathcal{M}$ に対して，

$$P(m|x^n) \propto p(x^n|m)P(m) =$$

$$\frac{p(x^n|m,\hat{\theta}^{k_m})f(\hat{\theta}^{k_m}|m)P(m)\sqrt{2\pi}^{k_m}}{\sqrt{n}^{k_m}\sqrt{\det J(\hat{\theta}^{k_m}|m)}}\{1+o(1)\},\ a.s. \tag{59}$$

が得られる．式(29)より，$f(\hat{\theta}^{k_m}|m)=O(1)$, $a.s.$, $\sqrt{\det J(\hat{\theta}^{k_m}|m)}=O(1)$, $a.s.$ であるから，ある定数 C' が存在して，

$$P_e^* \le \sum_{m \ne m^*} P^*\left[\log\frac{P(x^n|m,\hat{\theta}^{k_m})n^{k_{m^*}/2}}{P(x^n|m^*,\hat{\theta}^{k_{m^*}})n^{k_m/2}} \ge C'\right], \tag{60}$$

が得られる．さらに，チェビシェフの不等式より，

$$P_e^* \le$$

$$\sum_{m \ne m^*}\frac{V^*\left[\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m^*,\hat{\theta}^{k_{m^*}})}+\frac{k_{m^*}-k_m}{2}\log n + C'\right]}{\left\{E^*\left[\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m^*,\hat{\theta}^{k_{m^*}})}+\frac{k_{m^*}-k_m}{2}\log n + C'\right]\right\}^2}, \tag{61}$$

$E^*[\cdot]$ と $V^*[\cdot]$ は $p(\cdot|m^*,\theta^{k^*_{m^*}})$ による期待値と分散である．

まず，$m \in \mathcal{M}$ に対する $\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\theta^{k^*_m})}$ を評価することから始める．テーラー展開により〔9〕，

$$-\log p(x^n|m,\theta^{k^*_m})$$

$$= -\log p(x^n|m,\hat{\theta}^{k_m})$$

$$-\frac{1}{2}(\hat{\theta}^{k_m}-\theta^{k^*_m})^T\left.\frac{\partial^2 \log p(x^n|m,\theta^{k_m})}{\partial\theta^{k_m}(\partial\theta^{k_m})^T}\right|_{\theta^{k_m}=\hat{\theta}^{k_m}}(\hat{\theta}^{k_m}-\theta^{k^*_m})$$

$$+O\left(\left\|\hat{\theta}^{k_m}-\theta^{k^*_m}\right\|^3 \log p(x^n|m,\theta^{k_m+})\right), \tag{62}$$

が得られる．ここで，θ^{k_m+} は $\theta^{k^*_m}$ と $\hat{\theta}^{k_m}$ の間にある点である．式(29)より，ある定数 $C_1>0$ が存在して，$n\to\infty$ のとき，

$$\sqrt{n}\left\|\hat{\theta}^{k_m}-\theta^{k^*_m}\right\| \le C_1(\log\log n)^{1/2},\ a.s. \tag{63}$$

が成り立つ．この式と式(31)より，ある定数 $C_2>0$ が存在して，$n\to\infty$ のとき，

$$\left|\frac{1}{2}\sqrt{n}(\hat{\theta}^{k_m}-\theta^{k^*_m})^TJ(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m}-\theta^{k^*_m}) +\frac{1}{2}(\hat{\theta}^{k_m}-\theta^{k^*_m})^T\left.\frac{\partial^2 \log p(x^n|m,\theta^{k_m})}{\partial\theta^{k_m}(\partial\theta^{k_m})^T}\right|_{\theta^{k_m}=\hat{\theta}^{k_m}}(\hat{\theta}^{k_m}-\theta^{k^*_m})\right|$$

$$\le \frac{(\log\log n)^{3/2}}{\sqrt{n}},\ a.s. \tag{64}$$

が成り立つ．一方，式(30)と，$\theta^{k_m+}\to\theta^{k^*_m}$, $a.s.$ であることから，$\forall\epsilon>0$ に対して正定数 $C_3>0$ が存在して $n\to\infty$ のとき，

$$\left|\log p(x^n|m,\theta^{k_m+})\right| \le nC_3+n\epsilon,\ a.s. \tag{65}$$

となり，したがって正定数 $C_4>0$ が存在して $n\to\infty$ のとき，

$$\left|\left\|\hat{\theta}^{k_m}-\theta^{k^*_m}\right\|^3\log p(x^n|m,\theta^{k_m+})\right|$$

$$\le C_4\frac{(\log\log n)^{3/2}}{\sqrt{n}},\ a.s. \tag{66}$$

が成り立つ．式(62)と式(64), 及び式(66)より，ある正定数 $C_5>0$ が存在して，$n\to\infty$ のとき，

$$\left|\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\theta^{k^*_m})} -\frac{1}{2}\sqrt{n}(\hat{\theta}^{k_m}-\theta^{k^*_m})^TJ(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m}-\theta^{k^*_m})\right|$$

$$\le C_5\frac{(\log\log n)^{3/2}}{\sqrt{n}},\ a.s. \tag{67}$$

が成り立つ．上の展開より，ある正定数 $C_6>0$ が存在して，$n\to\infty$ のとき，

$$\left|\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m^*,\hat{\theta}^{k_{m^*}})}-\log\frac{p(x^n|m,\theta^{k^*_m})}{p(x^n|m^*,\theta^{k^*_{m^*}})} -\frac{1}{2}\sqrt{n}(\hat{\theta}^{k_m}-\theta^{k^*_m})^TJ(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m}-\theta^{k^*_m}) +\frac{1}{2}\sqrt{n}(\hat{\theta}^{k_{m^*}}-\theta^{k^*_{m^*}})^TJ(\theta^{k^*_{m^*}}|m)\sqrt{n}(\hat{\theta}^{k_{m^*}}-\theta^{k^*_{m^*}})\right|$$

$$\le C_6\frac{(\log\log n)^{3/2}}{\sqrt{n}},\ a.s. \tag{68}$$

が成り立つことがわかる．

まず，$p(\cdot|m,\theta^{k^*_m}) \neq p^*(\cdot)$ の場合を考える．この場合には，式 (31) と式 (11) から，正定数 $D_1 > 0$, $C_7 > 0$ が存在して，$n \to \infty$ のとき，

$$\left|\log \frac{p(x^n|m^*,\theta^{k^*_{m^*}})}{p(x^n|m,\theta^{k^*_m})} - nD_1\right| \leq C_7\sqrt{n}(\log\log n)^{1/2}, \quad a.s. \tag{69}$$

が成り立つ．また式 (29), より，定数 $C_8 > 0$ が存在して，$n \to \infty$ のとき，

$$\left|\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})^T J(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})\right| \leq C_8 \log\log n, \quad a.s. \tag{70}$$

が成り立つ．式 (69) と式 (70) より，正定数 $D_2 > 0$, $C_9 > 0$ が存在して，$n \to \infty$ のとき，

$$nD_2 + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n - C_9\sqrt{n}(\log\log n)^{1/2}$$
$$\leq \log \frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})} + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n + C'$$
$$\leq nD_2 + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n + C_9\sqrt{n}(\log\log n)^{1/2}, \quad a.s. \tag{71}$$

が成り立つことがわかる．これまでの議論が，確率 1 での収束を扱っていることから，これらの期待値を取ることができて，$D_2 > 0$ に対して，$n \to \infty$ のとき，

$$nD_2 + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n - C_9\sqrt{n}(\log\log n)^{1/2}$$
$$\leq E^*\left[\log \frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})} + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n + C'\right]$$
$$\leq nD_2 + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n + C_9\sqrt{n}(\log\log n)^{1/2}, \tag{72}$$

が成り立つ．式 (71) より，正定数 $C_{10} > 0$ が存在して，$\forall\epsilon > 0$ に対して $n \to \infty$ のとき

$$V^*\left[\log \frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})} + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n + C'\right] \leq C_{10} n \log\log n. \tag{73}$$

が成り立つ．式 (61), (72), (73) より，$p(\cdot|m,\theta^{k^*_m}) \neq p^*(\cdot)$ をみたす m を選択してしまう誤り確率 $P^*_{e1}[m]$ は，

$$P^*_{e1}[m] \leq O\left(\frac{\log\log n}{n}\right), \tag{74}$$

とバウンドされる．

次に，$p(\cdot|m,\theta^{k^*_m}) = p^*(\cdot)$ の場合を考える．このときは

$$\log \frac{p(x^n|m,\theta^{k^*_m})}{p(x^n|m,\theta^{k^*_{m^*}})} = 0, \tag{75}$$

が成り立つので，ある正定数 $C_{11} > 0$ が存在して，$n \to \infty$ のとき，

$$\left|\log \frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})} - \frac{1}{2}\sqrt{n}(\hat{\theta}^{k_{m^*}} - \theta^{k^*_{m^*}})^T J(\theta^{k^*_{m^*}}|m)\sqrt{n}(\hat{\theta}^{k_{m^*}} - \theta^{k^*_{m^*}}) + \frac{1}{2}\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})^T J(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})\right|$$
$$\leq C_{11}\frac{(\log\log n)^{3/2}}{\sqrt{n}}, \quad a.s. \tag{76}$$

ここで，$\lim_{n\to\infty} \frac{(\log\log n)^{3/2}}{\sqrt{n}} = 0$ である．一方，$\hat{\theta}^{k_m}$ の漸近正規性より，

$$\lim_{n\to\infty} E^*\left[\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})^T J(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})\right] = trace J^{-1}(\theta^{k^*_m}|m) I(\theta^{k^*_m}|m), \tag{77}$$

が成り立つ．また，$p(\cdot|m,\theta^{k^*_m}) = p^*(\cdot)$ のときは，$J(\theta^{k^*_m}|m) = I(\theta^{k^*_m}|m)$ が成り立つので，$trace J^{-1}(\theta^{k^*_m}|m)I(\theta^{k^*_m}|m) = k_m$ となり，

$$\lim_{n\to\infty} E^*\left[\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})^T J(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})\right] = k_m, \tag{78}$$

が得られる．同様に，

$$\lim_{n\to\infty} V^*\left[\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})^T J(\theta^{k^*_m}|m)\sqrt{n}(\hat{\theta}^{k_m} - \theta^{k^*_m})\right] = k_m, \tag{79}$$

となる．式 (76) は確率 1 で成り立つので，ある正定数 $C_{12} > 0$ が存在して，$n \to \infty$ のとき，

$$\left|E^*\left[\log \frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})} + \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n + C'\right] + \frac{k_{m^*}}{2} - \frac{k_m}{2} - \left(\frac{k_{m^*}}{2} - \frac{k_m}{2}\right)\log n\right|$$
$$\leq C_{12}\frac{(\log\log n)^{3/2}}{\sqrt{n}}, \tag{80}$$

が成り立つ．このことは，$n \to \infty$ のときに

$$\left\{\left(\frac{k_{m^*}}{2}-\frac{k_m}{2}\right)\left(1-\log n\right)-C_{12}\frac{(\log\log n)^{3/2}}{\sqrt{n}}\right\}^2$$

$$\leq \left\{E^*\left[\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})}\right.\right.$$

$$\left.\left.+\left(\frac{k_{m^*}}{2}-\frac{k_m}{2}\right)\log n+C'\right]\right\}^2$$

$$\leq \left\{\left(\frac{k_{m^*}}{2}-\frac{k_m}{2}\right)\left(1-\log n\right)+C_{12}\frac{(\log\log n)^{3/2}}{\sqrt{n}}\right\}^2, \quad a.s. \qquad (81)$$

であることを意味している．一方，

$$V^*\left[\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})}+\left(\frac{k_{m^*}}{2}-\frac{k_m}{2}\right)\log n+C'\right]$$

$$=V^*\left[\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})}\right]. \qquad (82)$$

であることから，$C_{13}>0$ が存在して $n \to \infty$ のとき，

$$V^*\left[\log\frac{p(x^n|m,\hat{\theta}^{k_m})}{p(x^n|m,\hat{\theta}^{k_{m^*}})}+\left(\frac{k_{m^*}}{2}-\frac{k_m}{2}\right)\log n+C'\right]$$

$$\leq k_m+k_{m^*}+2k_mk_{m^*}+C_{13}\frac{(\log\log n)^{3/2}}{\sqrt{n}}, (83)$$

が成り立つ．これは，

$$V[X-Y]\leq V[X]+V[Y]+2V[X]V[Y], (84)$$

から導かれる．式(61), (81), (83) より，$p(\cdot|m,\theta^{k^*_m}) = p^*(\cdot)$ をみたす m^* 以外のある m を選択してしまう誤り確率 $P^*_{e2}[m]$ は

$$P^*_{e2}[m]\leq O\left(\frac{1}{\left(1-\log n\right)^2}\right), \qquad (85)$$

が導かれる．式(74)と式(85)より，式(61)のようにユニオンバウンドをとれば定理が得られる． □