論　文

事前分布が異なる場合のMDL原理に基づく符号とベイズ符号の符号長に関する解析

後藤　正幸† 　松嶋　敏泰† 　平澤　茂一†

A Study on Difference of Codelengths between Codes Based on MDL Principle and Bayes Codes for Given Prior Distributions

Masayuki GOTOH†, Toshiyasu MATSUSHIMA†, and Shigeichi HIRASAWA†

あらまし　J. Rissanenによって提案されたMDL（Minimum Description Length）原理は，符号長を最小化する確率モデルを選択するという点で一種のモデル推定機構を備えている．一方，確率モデルを特定せず，それらの混合をとって符号化関数を求めるのがベイズ符号である．MDL原理に基づく符号（MDL符号）はベイズ理論と関連深いことが既に指摘されているが，これは確率モデルの記述長の定義が暗に事前分布を仮定していることを意味するからである．本論文ではそれぞれが異なる事前分布をもつ場合を含めて，MDL符号とベイズ符号の符号長の差異を漸近解析する．その結果，モデル族が離散の場合には真のモデルに高い事前分布を仮定した方（つまり，有利な事前分布を仮定した方）の符号が優れるが，パラメトリックモデル族の場合には，MDL符号に有利な事前分布が仮定された場合でも，その差がある範囲内であれば，MDL符号よりベイズ符号の方が符号長が小さくなることを明らかとする．

キーワード　ベイズ符号，MDL原理，情報源符号化，事前分布

1. まえがき

J. Rissanenによって提案されたMDL原理[14]〜[17]は，情報源符号化やデータ解析，統計的学習理論の分野において広く研究されている．これはデータと確率モデルを同時に符号化することにより，データの一意復号可能な符号化を実現する方法である．その際に，合計の符号長を最小化するような確率モデルで符号化するという点で，一種のモデル推定機構を備えている．

一方，確率モデルを特定せず，それらの混合をとって符号化関数とするのがベイズ符号であり[3], [11]，現在ではFSMX情報源に対して効率的なアルゴリズムが研究されるに至っている[10], [12], [21]．MDL原理に基づく符号（MDL符号）はベイズ理論と関連深いことが既に指摘されている．なぜなら，最小記述長を計算する際に，確率モデルの記述長が定義されるが，これは確率モデルに暗に事前分布を仮定していることを意味するからである．RissanenはStochastic Complexityを提案するなど，MDLをベイズ的枠組みからも発展させた[16](注1)．更に，B.S. Clarke and A.R. Barronによってベイズ符号の冗長度の漸近的な解析が行われ[3]，T. Matsushimaらによって，この方法がベイズ最適であること，1パスと2パスの符号化でのロスは生じないこと，ミニマックス基準で事前分布を決定できることなどが示された[11]．

MDL符号，ベイズ符号の性質に対しては，様々な視点から研究が行われており，Rissanenによって，同じ事前分布のもとで，ベイズ符号の符号長がMDL符号の符号長を下界することが示されている[16]．また，漸近的解析に基づけば，両平均符号長は$o(\log n)$で一致することが知られている[8]．また，ベイズ符号を推定量として見ると，一般にもとのモデル族に含まれないことから，これをもとの族に射影した射影ベイ

† 早稲田大学理工学部経営システム工学科，東京都
School of Science and Engineering, Waseda University, 3–4–1 Ohkubo, Shinjyuku-ku, Tokyo, 169–8555 Japan

（注1）：Rissanenは過去の論文においてベイズ符号の符号長を統計的複雑さを測る基準として提案し，これをStochastic Complexityと定義している[16]．また現在では，あらゆるユニバーサル符号化の中での最小記述長[23]，あるいは最ゆう符号化の符号長[17]として定義されている．本論文ではこれらの符号についてはふれない．

ズ推定量と MDL 推定量との関係を解析した研究もある [19]．また，階層構造のモデル族に対して離散モデルとパラメータの両方で混合をとったベイズ符号の冗長度 [10] やベイズリスクの上界 [22] を解析した結果があり，いずれも興味深い結果が得られている．またこれらの結果を踏まえ，筆者らは先に，同じ事前分布のもとで，同じデータ系列を符号化したときの MDL 符号とベイズ符号の符号長の差を直接解析し，その差がたかだか定数オーダであることを示した [7]．しかし，異なる事前分布を仮定した場合に両符号長の差はどのようになるか，という点については解析がなされていない．

本論文では，異なる事前分布をもつ場合に対し，MDL 符号とベイズ符号の符号長の差異を解析し，両符号長の漸近的な大小関係を明らかにする．その結果，モデル族が離散の場合，漸近的には真のモデルに高い事前分布を仮定した方の符号が優れ，一方，パラメトリックの場合には，異なる事前分布をもつ場合でもその差がある値以内であれば，ベイズ符号の符号長の方が短くなることを明らかにする．

2. MDL 符号とベイズ符号

次に MDL 符号とベイズ符号を定義する．$\mathcal{X}$ を離散アルファベットとし，その要素である情報源記号を x で表す．$x^n = x_1 x_2 \cdots x_n$ を情報源からの長さ n の系列とする．$P(\cdot)$ は確率関数を，$f(\cdot)$ は確率密度関数を表すものとする．また，本論文を通じて対数の底は e とし,（理想）符号長の単位も nat とする．

1）離散的な場合：$P(x^n|m)$ を x^n を符号化するための符号化確率，$C(m)$ を確率モデル m を記述するのに必要な符号長とすれば，MDL 符号の符号長 $L^m_{MDL}(x^n)$ は，

$$L^m_{MDL}(x^n) = \min_m \left\{ -\log P(x^n|m) + C(m) \right\}, \tag{1}$$

となる．m は有限集合 $\mathcal{M}$ の要素で，$\sum_m e^{-C(m)} = 1$ を仮定する [13]（注 2）．この場合，確率モデル m の符号長 $C(m)$ の定義は m の事前確率を $P_M(m) = e^{-C(m)}$ と仮定することと等価である [13], [17]．このとき，

$$L^m_{MDL}(x^n) = \min_m \left\{ -\log P(x^n|m) - \log P_M(m) \right\}, \tag{2}$$

である．一方，ベイズ符号の符号長 $L^m_{Bayes}(x^n)$ は，

$$L^m_{Bayes}(x^n) = -\log \sum_m P(x^n|m) P_B(m), \tag{3}$$

で表される．ここで，$P_B(m)$ はベイズ符号における離散的なモデル m の事前確率である．

2）パラメトリックなモデルの場合：次に確率モデルが（連続の）k 次元未知パラメータで特定されるパラメトリックなクラスを考える．Rissanen が示したパラメータを量子化してそれらに符号長を割り当てるという操作も，暗にパラメータ上に事前分布を仮定していると見ることができる．$\{P(\cdot|\theta^k)|\theta^k \in \Theta^k\}$ をパラメトリックモデル族とし，パラメータに対するある事前分布 $f(\theta^k)$ はパラメータ集合 Θ^k 上に仮定されているものとする．ここで，Θ^k は $\mathcal{R}^k$ の有界な開部分集合とする．k 次元パラメータ θ^k の量子化後の代表点を $\bar{\theta}^k$ で表し，代表点 $\bar{\theta}^k$ で代表される直方体セルの j 番目の辺の長さを $\alpha_j(\bar{\theta}^k)$ とおく．すると，それらの量子化されたパラメータ $\bar{\theta}^k$ に対し，$-\log f_M(\bar{\theta}^k) - \sum_j \log \alpha_j(\bar{\theta}^k)$ の符号長で語頭条件を満足する符号を構成することができる [13]．ここで，$f_M(\theta^k)$ は MDL 符号における θ^k の事前確率密度であり，$\sum_{\bar{\theta}^k} f_M(\bar{\theta}^k) \prod_j \alpha_j(\bar{\theta}^k) = 1$ となるように $\bar{\theta}^k$ を選んでいるものと仮定する．したがってこの場合，離散化した $\bar{\theta}^k$ の事前確率を，$f_M(\bar{\theta}^k) \prod_j \alpha_j(\bar{\theta}^k)$ と定義して符号化していることと等価になる．このとき，MDL 符号の符号長は，

$$L^{\bar{\theta}^k}_{MDL}(x^n) = \min_{\bar{\theta}^k} \left\{ -\log P(x^n|\bar{\theta}^k) - \log f_M(\bar{\theta}^k) \prod_{j=0}^{k-1} \alpha_j(\bar{\theta}^k) \right\}, \tag{4}$$

で与えられる．ここで，量子化幅は大きすぎても小さすぎても符号長が伸びてしまい，最適な量子化幅が議論できることが知られている．その漸近的に最適な量子化セルは（最ゆう推定量の漸近正規性の仮定（後述の条件 2,（2））のもとで），フィッシャー情報行列 $I(\theta^k)$ の主軸方向を向き，

$$\prod_{j=0}^{k-1} \alpha^*_j(\bar{\theta}^k) = \frac{1}{\sqrt{n}^k \sqrt{\det I(\bar{\theta}^k)}}, \tag{5}$$

（注 2）：$\sum_m e^{-C(m)} > 1$ とすると語頭条件を満足する符号が存在しなくなるし，$\sum_m e^{-C(m)} < 1$ であればその分符号長にロスが生じる．

を満たすような量子化幅 $\alpha_j^*(\bar{\theta}^k)$，$j=0,1,\cdots,k-1$ で与えられる [8], [19], [20].

一方，ベイズ符号の符号長は

$$L_{Bayes}^{\theta^k}(x^n) = -\log\int_{\theta^k} P(x^n|\theta^k) f_B(\theta^k) d\theta^k, \tag{6}$$

となる．$f_B(\theta^k)$ はベイズ符号におけるパラメータの事前密度である．式 (6) は，事前分布 $f_B(\theta^k)$ に対してベイズ最適な符号を与えている．

以上のように，ベイズ符号は，符号化関数として，候補となるすべての確率モデルの混合をとることを許容する点が特徴である．

3. 符号長の差に関する解析

事前分布が同じ場合，ベイズ符号の符号長が MDL 符号の符号長を下界することが Rissanen によっても示されている [16]．以下では，異なる事前分布をもつ場合を含めて符号長の差を考察する．

3.1 離散的モデル族の場合

離散的に表現されるモデル族に対するベイズ符号長と MDL 符号長はそれぞれ式 (2)，(3) で与えられる．

ここで，離散有限モデル族 $\mathcal{M}$ について，以下の条件を仮定する．

［条件 1］

（1） すべてのモデル $m\in\mathcal{M}$ に対して，$P_M(m)>0$，$P_B(m)>0$ とする．

（2） 真のモデル m^* は $\mathcal{M}$ 内で唯一であるものとする．すなわち，$m \neq m^*$，$D(P_{m^*};P_m)=0$ を満たすモデル m は存在しない．ここで，$D(P_{m^*};P_m)$ は真の分布 $P(\cdot|m^*)$ からの $P(\cdot|m)$ の KL 情報量であり，次式で定義される．

$$D(P_{m^*};P_m) = \lim_{n\to\infty}\frac{1}{n}\sum_{x^n} P(x^n|m^*)\log\frac{P(x^n|m^*)}{P(x^n|m)}. \tag{7}$$

（3） 集合 D_1 を $D(P_{m^*};P_m)\neq 0$ を満たすモデル m，すなわち真のモデル以外のモデルの集合とする．このとき，$\forall m\in D_1$ に対して，$n\to\infty$ のとき，

$$\left|\frac{P(x^n|m)}{P(x^n|m^*)}\right| \to 0, \quad a.s. \tag{8}$$

が成り立つ(注3)． □

このとき，次の定理が成り立つ．

［定理 1］ 条件 1 のもとで，式 (2) に対して，以下の関係が成り立つ．

$$L_{MDL}^m(x^n) = -\log\sum_m P(x^n|m)P_M(m) - \log P_M(\hat{m}|x^n). \tag{11}$$

ただし，$\hat{m}$ は事前分布 $P_M(m)$ のもとでの事後確率 $P_M(m|x^n)$ を最大化する確率モデルである．

（証明）ベイズの定理より，

$$P(x^n|m)P_M(m) = P_M(m|x^n)P_M(x^n), \tag{12}$$

である．ここで，$P_M(x^n)$ は，

$$P_M(x^n) = \sum_m P(x^n|m)P_M(m), \tag{13}$$

である．よって，

$$-\log P(x^n|m) - \log P_M(m) = -\log P_M(m|x^n) - \log P_M(x^n), \tag{14}$$

であり，これを式 (2) に代入すると，$P_M(x^n)$ は MDL の最小化に無関係なことから，

$$\begin{aligned} L_{MDL}^m(x^n) &= \min_m\left\{-\log P_M(x^n) - \log P_M(m|x^n)\right\} \\ &= -\log P_M(x^n) - \log P_M(\hat{m}|x^n), \end{aligned}$$

が得られ，式 (11) が成り立つ． □

この定理は，同じ事前分布のもとでは，ベイズ符号が MDL 符号より $-\log P_M(\hat{m}|x^n)$ だけ符号長を短く符号化することを意味し，有限長でのベイズ符号の優位性を示している．次に，事前分布が異なる場合の符号長の差を解析するためにまず，$-\log P_M(\hat{m}|x^n)$ が実際にどの程度のオーダになるかを考えてみる．

（注 3）：例えば KL 情報量に対する大数の強法則

$$\frac{1}{n}\log\frac{P(x^n|m^*)}{P(x^n|m)} \to D(P_{m^*};P_m), \quad a.s. \tag{9}$$

が成り立てば，式 (8) も成り立っている．式 (8) はすなわち

$$\sum_{n=1}^{\infty} P^*\left\{\frac{P(x^n|m)}{P(x^n|m^*)} \geqq \epsilon\right\} < \infty, \tag{10}$$

が成り立っていることを意味する（Borel-Cantelli の補題 [5]）．

［補題 1］ 条件 1 のもとで，任意の事前分布 $P_M(m)$ に対して，

$$-\log P_M(\hat{m}|x^n) = o^+(1), \quad a.s. \tag{15}$$

が成り立つ．ただし，$o^+(1)$ は正でかつ $\lim_{n\to\infty} o^+(1) = 0$, $a.s.$ となる項である．

（証明）

$$P_M(m|x^n) \propto P(x^n|m)P_M(m), \tag{16}$$

と条件 1,（1）及び式 (8) より，$n \to \infty$ のとき

$$\left|\frac{P_M(m|x^n)}{P_M(m^*|x^n)}\right| \to 0, \quad a.s. \tag{17}$$

が成り立つ．したがって，$n \to \infty$ のとき，

$$P_M(m^*|x^n) \to 1, \quad a.s. \tag{18}$$

も成り立ち，これより

$$-\log P_M(\hat{m}|x^n) = o^+(1), \quad a.s. \tag{19}$$

を得る． □

一方で，ベイズ符号の符号長に関しては以下が成り立つ．

［補題 2］ 条件 1 のもとで，任意の事前分布 $P_B(m)$ に対して，次式が成り立つ．

$$-\log \sum_m P(x^n|m)P_B(m) = -\log P(x^n|m^*) - \log P_B(m^*) - o^+(1), \ a.s. \tag{20}$$

（証明）付録 **1.** 参照． □

以上により，異なる事前分布をもつ場合の，MDL 符号とベイズ符号の符号長に関して，次の補題が成り立つ．

［補題 3］ 条件 1 のもとで，式 (2) と式 (3) に対して以下の式が成り立つ．

$$L^m_{MDL}(x^n) = L^m_{Bayes}(x^n) - \log\frac{P_M(m^*)}{P_B(m^*)} + o(1), \quad a.s. \tag{21}$$

$o(1)$ は $\lim_{n\to\infty} o(1) = 0$, $a.s.$ となる項である．

（証明）定理 1 より，

$$\begin{aligned} L^m_{MDL}(x^n) = L^m_{Bayes}(x^n) &- \log\sum_m P(x^n|m)P_M(m) \\ &+ \log\sum_m P(x^n|m)P_B(m) \\ &- \log P_M(\hat{m}|x^n), \ a.s. \end{aligned} \tag{22}$$

が得られるから，補題 1 と 2 を適用して式 (21) を得る． □

この補題と定理 1 より，漸近的な MDL 符号とベイズ符号の符号長の差について，次の定理が成り立つ．

［定理 2］ 条件 1 のもとで，$n \to \infty$ のとき，以下の不等式が成り立つ(注 4)．

$P_M(m^*) > P_B(m^*)$ のとき，

$$L^m_{MDL}(x^n) < L^m_{Bayes}(x^n), \quad a.s.$$

$P_M(m^*) < P_B(m^*)$ のとき，

$$L^m_{MDL}(x^n) > L^m_{Bayes}(x^n). \quad a.s.$$

$P_M(m^*) = P_B(m^*)$ のとき，両者は漸近的に等価である． □

最後の条件式 $P_M(m^*) = P_B(m^*)$ を，$\forall m \in \mathcal{M}$ に対して $P_M(m) = P_B(m)$ と置き換えれば，任意の x^n $(n = 1, 2, \cdots)$ に対して $L^m_{MDL}(x^n) \geqq L^m_{Bayes}(x^n)$ となる [16]．もし，$\forall m \in \mathcal{M}$ に対して $P(x^n|m) > 0$ かつ $P(m) > 0$ であれば，$L^m_{MDL}(x^n) \geqq L^m_{Bayes}(x^n)$ となることも容易にわかる．

この定理より，両符号化では，真のモデル m^* に高い事前分布を仮定した方の符号が漸近的に符号長が短くなる．これは，有利な事前分布を仮定した方が優れるという直感と一致する結果といえる．ただし，その符号長の差は $O(1)$ であるので，各々の符号長が $O(n)$ であることから，事前分布が異なる場合でも，1 記号当りの平均符号長は $O(1/n)$ で一致することがわかる．

同じ事前分布の場合には，データ数 n が有限の場合でもベイズ符号の方が符号長を短くできることが示せたが，このように事前分布が異なる場合には，漸近的な符号長の大小関係しか示すことができない．これは，事前分布が異なる場合には，ある有限長の x^n に対してはこの大小関係が逆転する可能性があるからである．

（注 4）：本論文では，事象 $\mathcal{A}_k$, $k = 1, 2, \cdots$ が真の確率測度 P^* に対して，$P^*\left(\limsup_{k\to\infty}(\mathcal{A}_k)^C\right) = 0$（すなわち，事象 $(\mathcal{A}_k)^C$ $k = 1, 2, \cdots$ のうち無限回が起こる確率は 0）を満たすことを，“$k \to \infty$ のとき $\mathcal{A}_k$ がほとんど確実に成り立つ” と述べることにする．ただし，$(\mathcal{A}_k)^C$ は $\mathcal{A}_k$ の補事象である．このことはよりシンプルに “十分大きな k に対して常に $\mathcal{A}_k$ は成り立つ” などと記述されることもある．

3.2 パラメトリックなモデル族の場合

パラメトリックモデル族に対する MDL 符号とベイズ符号は式 (4)，(6) で定義される．データ系列 x^n は真の分布 $P(\cdot|\theta^{k*})$ に従って発生するものとする．$\theta^{k*} \in \Theta^k$ は真のパラメータである．

ここで，本論文で扱うパラメトリックモデルクラスに対して以下を仮定する．

［条件 2］

（1） θ_i を θ^k の第 i 要素，行列 $I(x^n, \theta^k)$ の $i-j$ 要素を

$$I_{ij}(x^n, \theta^k) = -\frac{1}{n}\left\{\frac{\partial^2 \log P(x^n|\theta^k)}{\partial\theta_i \partial\theta_j}\right\}, \qquad (23)$$

と定義し，この行列の $P(\cdot|\theta^k)$ による期待値の $n \to \infty$ の極限を $I(\theta^k)$ で表す．すなわち，$I(\theta^k)$ の要素は，

$$I_{ij}(\theta^k) = -\lim_{n\to\infty}\frac{1}{n}E\left[\frac{\partial^2 \log P(x^n|\theta^k)}{\partial\theta_i \partial\theta_j}\right], \quad (24)$$

で与えられる．このとき，$I(\theta^k)$ の要素は Θ^k 上で連続で，またすべての $\theta^k \in \Theta^k$ に対して $0 < C_0 \leqq \sqrt{\det I(\theta^k)} < \infty$ が成り立つものとする(注5)．$I(\theta^k)$ は Fisher 情報行列と呼ばれる．

（2） パラメータの最ゆう推定量 $\hat{\theta}^k = \hat{\theta}^k(x^n)$ は漸近正規性を満たす．すなわち，$\eta^k = \sqrt{n}\left(\hat{\theta}^k - \theta^{k*}\right)$ の事後分布は，原点を平均とし，共分散 $I^{-1}\left(\theta^{k*}\right)$ の正規分布に収束する．言い換えると $R_\eta(0)$ を原点を中心とする任意の直方体とすると，

$$P^*\{\eta^k \in R_\eta(0)\} \longrightarrow \frac{\sqrt{\det I(\theta^{k*})}}{(2\pi)^{k/2}}\int_{R_\eta(0)} \exp\left\{-\frac{1}{2}\parallel \eta^k \parallel^2_{I(\theta^{k*})}\right\} d\eta^k, \qquad (25)$$

が成り立つ．ここで，$P^*\{\cdot\}$ は真の分布 $P(\cdot|\theta^{k*})$ による確率測度である．

（3） パラメータに対する事後分布は漸近的に正規分布に収束するものとする．すなわち，$\xi^k = \sqrt{n}\left(\theta^k - \hat{\theta}^k\right)$ の事後分布は，平均ゼロベクトル，共分散 $I^{-1}\left(\hat{\theta}^k\right)$ の正規分布に収束する．すなわち，$R_\xi(0)$ を原点を中心とする任意の直方体とすると，

$$P\{\xi^k \in R_\xi(0)|x^n\} \longrightarrow \frac{\sqrt{\det I(\hat{\theta}^k)}}{(2\pi)^{k/2}}\int_{R_\xi(0)} \exp\left\{-\frac{1}{2}\parallel \xi^k \parallel^2_{I(\hat{\theta}^k)}\right\} d\xi^k, \quad a.s. \qquad (26)$$

が成り立つ(注6)．

（4） ゆう度関数 $P(x^n|\theta^k)$ は θ^k に関して単峰形であるとし，その最大値をとる最ゆう推定量 $\hat{\theta}^k$ は強一致推定量であるとする．すなわち，$\forall\epsilon > 0$ に対してある n_ϵ が存在して，$n > n_\epsilon$ のとき，

$$\left\|\hat{\theta}^k - \theta^{k*}\right\| \leqq \epsilon, \quad a.s.$$

が成り立つ．

（5） すべての $\theta^k \in \Theta^k$ に対して，$f(\theta^k) > 0$ とする．

（6） 事前密度関数の θ^k は次式を満たす．

$$\forall\theta^k \in \Theta^k, \quad \left\|\frac{\partial f_M(\theta^k)}{\partial\theta^k}\right\| < c_f, \qquad (27)$$

$$\forall\theta^k \in \Theta^k, \quad \left\|\frac{\partial f_B(\theta^k)}{\partial\theta^k}\right\| < c_f, \qquad (28)$$

ただし，c_f はある有限値である．更に，$\forall\theta^k \in \Theta^k$ に対して，

$$0 < C_1 \leqq \frac{f_M(\theta^k)}{f_B(\theta^k)} \leqq C_2 < \infty, \qquad (29)$$

であるとする．ただし，C_1，C_2 は正定数である．

（7） 量子化されたパラメータの各セルに番号を付ける．Θ^k_l を l 番目の量子化セル内のパラメータの集合とする．これらのセルの個数は，全体のパラメータ集合の境界における誤差を考慮しても $O(1/\min\prod_j \alpha_j)$ である．このとき，$\cup_l \Theta^k_l = \Theta^k$ かつ $i \neq j$，$\Theta^k_i \cap \Theta^k_j = \phi$ とする．□

パラメータの事後分布が漸近正規性をもつための十分条件については，例えば [4] などがある．これを満たす分布族としては，例えば事前分布にディレクレ分布を仮定した多項分布族，有限マルコフ情報源，FSMX

（注5）：$0 < C_0 \leqq \sqrt{\det I(\theta^k)}$ は通常の確率モデルでは一般に仮定できる．しかし，通常符号化に適用される分布族で，$\sqrt{\det I(\theta^k)} < \infty$ であるが，$\sqrt{\det I(\theta^k)} < C'_0 < \infty$ は成り立たないものは存在する．例えば 2 項分布では，パラメータ（シンボルの正規確率）θ を $(0, 1)$ 上に定義すれば，$\sqrt{\det I(\theta)} = \infty$ とはならないが，$\lim_{\theta\to 0 \text{ or } 1}\sqrt{\det I(\theta)} = \infty$ であるから有界ではない．

（注6）：$\forall R > 0$ が与えられたもとで，$\left|R_\xi(0)\right| < R$ となる $R_\xi(0)$ に対して一様に式 (26) が成り立つ．これは事前密度が仮定されたもとで，事後密度関数も正規密度関数に概収束することを示している．また，$I(\hat{\theta}^k)$ は Fisher 情報行列であるが，厳密には若干異なる情報行列を使う必要がある．しかし，θ^{k*} の近傍での議論になるので，漸近的に Fisher 情報行列で置き換えることができる．

情報源(注7)などがある．これはディレクレ分布の共役性から [1]，pp.290–294 の議論より明らかである [7]．またこれらの分布族が条件 2 全体を満たすことは [5] より明らかである．

パラメータを量子化した後を考えれば，これらの仮定のもとで以下の補題が成り立つ．

［定理 3］ $\bar{\xi}^k = \sqrt{n}(\bar{\theta}^k - \hat{\theta}^k)$ と定義する．このとき条件 2 のもとで，式 (4) は次式で与えられる．

$$L_{MDL}^{\bar{\theta}^k}(x^n) = L_{Bayes}^{\theta^k}(x^n) - \max_{\bar{\xi}^k}\left\{\log\frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} + \log\frac{f_M\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}\right\}. \quad (30)$$

ただし，$f_{\xi,B}(\bar{\xi}^k|x^n)$ は $f_B(\cdot)$ を Θ^k 上の事前分布としたときの，$\xi^k = \sqrt{n}(\theta^k - \hat{\theta}^k)$ で形成される空間上の事後確率密度であり，$f_B(\theta^k|x^n)$ を θ^k の事後確率密度

$$f_B(\theta^k|x^n) = \frac{P(x^n|\theta^k)f_B(\theta^k)}{\int_{\theta^k} P(x^n|\theta^k)f_B(\theta^k)d\theta^k}, \quad (31)$$

として

$$f_{\xi,B}(\bar{\xi}^k|x^n) = \frac{1}{\sqrt{n}^k} f_B(\theta^k|x^n), \quad (32)$$

で与えられる(注8)．

（証明）MDL 符号とベイズ符号の定義より，

$$L_{MDL}^{\bar{\theta}^k}(x^n) = L_{Bayes}^{\theta^k}(x^n) - \max_{\bar{\theta}^k}\left\{\log\frac{P(x^n|\bar{\theta}^k)f_B(\bar{\theta}^k)}{\int_{\theta^k} P(x^n|\theta^k)f_B(\theta^k)d\theta^k} + \log\frac{f_M(\bar{\theta}^k)}{f_B(\bar{\theta}^k)} + \frac{1}{\sqrt{n}^k\sqrt{\det I(\bar{\theta}^k)}}\right\}. \quad (33)$$

であることから明らか． □

この定理より，式 (30) の右辺第 2 項によって，MDL 符号とベイズ符号の大小関係が決まることがわかる．以下では，この右辺第 2 項を漸近的に評価することにより，漸近的な両符号の符号長の差を解析する．

［補題 4］ $r > 0$ に対して，$\mathcal{A}(r)$ と $\bar{\mathcal{A}}(r)$ を以下のように定義する．

$$\mathcal{A}(r) = \left\{\bar{\xi}^k \,\middle|\, \left\|\bar{\xi}^k\right\|_{I(\hat{\theta}^k)} > r, \hat{\theta}^k + \frac{1}{\sqrt{n}}\bar{\xi}^k \in \Theta^k\right\}, \quad (34)$$

$$\bar{\mathcal{A}}(r) = \left\{\bar{\xi}^k \,\middle|\, \left\|\bar{\xi}^k\right\|_{I(\hat{\theta}^k)} \leqq r, \hat{\theta}^k + \frac{1}{\sqrt{n}}\bar{\xi}^k \in \Theta^k\right\}, \quad (35)$$

このとき，

$$\tilde{\xi}^k = \arg\max_{\bar{\xi}^k}\left\{\log\frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} + \log\frac{f_M\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}\right\}. \quad (36)$$

とすれば，条件 2 のもとで，$n \to \infty$ のとき

$$\tilde{\xi}^k \in \bar{\mathcal{A}}(r), \quad a.s. \quad (37)$$

を満たすような $r > 0$ が存在する．

（証明）付録 **2.** 参照． □

この定理は，式 (30) の右辺第 2 項の最大化は，$\bar{\mathcal{A}}(r)$ 内の $\bar{\xi}^k$ のみを考えればよいことを示している．この補題を用いると，次の結果が導かれる．

［補題 5］ 条件 2 のもとで，$\forall\epsilon > 0$ に対して，$n \to \infty$ のとき

$$\frac{k}{2}\log 2\pi - \log\frac{f_M(\theta^{k^*})}{f_B(\theta^{k^*})} - \epsilon \leqq L_{MDL}^{\bar{\theta}^k}(x^n) - L_{Bayes}^{\theta^k}(x^n) \leqq \frac{k}{2}\log 2\pi + \frac{1}{2} - \log\frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} + \epsilon, \quad a.s. \quad (38)$$

（注 7）：ただし，シンボルの正規確率を表すパラメータの真値が 0 や 1 の端点となる場合を除く．このような Fisher 情報行列が無限大となる点では特別な解析が必要となる．

（注 8）：$\xi^k = \sqrt{n}(\theta^k - \hat{\theta}^k)$ という関係より明らか．

が成り立つ．

（証明）付録 **3.** 参照．　□

ここで，$-\log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} < 0$ は，真のパラメータに対して，MDL 符号の方が高い事前分布を仮定していることを意味するが，漸近的には $-\log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})}$ が $\frac{k}{2}\log 2\pi$ をキャンセルするほど小さくなければ，ベイズ符号の符号長が MDL 符号の符号長よりも小さくなることがわかる．よって補題 5 と補題 6 より，MDL 符号とベイズ符号の符号長の差に関して，次の定理を得る．

［定理 4］ 条件 2 のもとで，もし $f_B(\theta^{k*}) < f_M(\theta^{k*})$ であっても，

$$\log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} < \frac{k}{2}\log 2\pi, \tag{39}$$

の範囲内であれば，$n \to \infty$ のときほとんど確実に，MDL 符号の符号長よりもベイズ符号の符号長が短くなる．　□

この定理から，真のパラメータに高い事前確率密度を仮定することが符号長に対して有利であるが，たとえこの意味でベイズ符号に不利な事前分布が仮定された場合でも，ある範囲内であればやはりベイズ符号の符号長が短くなるというベイズ符号化の優れた特性が理解できる．

4. 考　　察

離散モデル族に対する符号長は，両符号化のうち，真のモデルに高い事前確率を仮定した方の符号長が短くなったが，これは同じ事前分布を仮定した場合にはモデルを一つ選択する方法とすべてのモデルの混合を用いる方法による符号長の差が漸近的に 0 に収束することに起因する．つまり，同じ事前分布のときに MDL 符号とベイズ符号は漸近的等価になるので，事前分布が異なる場合には真のモデルに高い事前確率を仮定した方がより符号長を短くできるわけである．一方，パラメトリックなモデル族の場合には，真のモデルに MDL 符号よりも小さい事前密度を仮定したとしても，ある範囲内であれば漸近的にベイズ符号が優れることがわかる．これは，パラメータの事後分布の主軸方向の標準偏差と量子化幅が漸近的に等しいので，真のパラメータを含むセルの事後確率は常に 1 以下の定数で抑えられ，1 に収束しないこと，つまり真のセルというものを特定することができない(注 9)ことに起因する．この結果は，連続のパラメータを離散化する操作が必ずしも有効ではないことを示している．

（注 9）：そもそも真のセルというものは存在しない．

5. む す び

本論文では，事前分布の異なる場合について，MDL 符号とベイズ符号の差を符号長の面から解析した．その結果，離散モデル族の場合には，漸近的に真のモデルに高い事前分布を仮定した方が優れるという，極めて自然な結果が得られた．一方，パラメトリックなモデル族の場合には，真のパラメータに MDL 符号より低い事前分布を仮定したとしても，ある範囲内であれば，漸近的にベイズ符号が優れることが明らかとなった．

文　　献

[1] J.M. Bernardo and A.F.M. Smith, “Bayesian theory,” John Wiley & Sons, 1994.

[2] R.E. Blahut, “Principles and practice of information theory,” Addison Wesley, 1987.

[3] B.S. Clarke and A.R. Barron, “Jeffreys' prior is asymptotically least favorable under entropy risk,” JSPI 41, pp.37–60, 1994.

[4] Chan-Fu Chen, “On asymptotic normality of limiting density function with Bayesian implications,” J.R. Statist. Soc. B, vol.47, no.3, pp.540–546, 1985.

[5] W. Feller, “An introduction to probability and its applications,” John Wiley & Sons, New York, 1957.

[6] T.S. Ferguson, “Mathematical statistics, A decision theoretic approach,” New York and London, Academic, 1967.

[7] M. Gotoh, T. Matsushima, and S. Hirasawa, “An analysis on difference of codelengths between codes Based on MDL Principle and Bayes Codes,” submitted paper, 1998.

[8] 韓　太舜，小林欣吾，“情報と符号化の数理，” 岩波書店，1994.

[9] 伊藤秀一，“MDL のパターン認識への応用，” 人工知能誌，vol.7, no.4, pp.608–614, 1992.

[10] 川端　勉，“ベイズ符号と文脈木重み付け法，” 信学技報，IT93-121，1994.

[11] T. Matsushima, H. Inazumi, and S. Hirasawa, “A class of distortionless codes designed by Bayes decision theory,” IEEE Trans. Inform. Theory, vol.37, no.5, pp.1288–1293, 1991.

[12] T. Matsushima and S. Hirasawa, “A Bayes coding algorithm for Markov models,” 信学技報，IT95-1, 1995.

[13] G. Qian, G. Gabor, and R.P. Gupta, “On stochastic complexity estimation: A decision-theoretic approach,” IEEE Trans. Inform. Theory, vol.40, no.4, pp.1181–1191, 1994.

[14] J. Rissanen, “Modeling by shortest data description,” Automatica, vol.46, pp.465–471, 1978.

[15] J. Rissanen, “Universal modeling and coding,” IEEE

Trans. Inform. Theory, vol.27, no.1, pp.12–23, 1981.
[16] J. Rissanen, "Stochastic complexity," J. Roy, Statist., Soc. B, vol.49, pp.223–265, 1987.
[17] J. Rissanen, "Fisher information and stocastic complexity," IEEE Trans. Inform. Theory, vol.42, no.1, pp.40–47, 1996.
[18] C. Schwarz, "Estimating the dimension of a model," Ann. Statist., vol.6, pp.461–464, 1978.
[19] J. Takeuchi, "Characterization of the Bayes estimator and the MDL estimator for exponential families," IEEE Trans. Inform. Theory, vol.43, no.4, pp.1165–1174, 1997.
[20] C.S. Wallace and P.R. Freeman, "Estimation and inference by compact coding," J.R. Statist. Soc. B, 49, no.3, pp.240–265, 1987.
[21] F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens, "The context–tree weighting method: Basic properties," IEEE Trans. Inform. Theory, vol.41, no.3, pp.653–664, 1995.
[22] K. Yamanishi, "A loss bound model for on-line stochastic prediction algorithms," Information and Computation, vol.119, pp.39–54, 1995.
[23] 山西健司, "確率的コンプレキシティと学習理論," オペレーションズリサーチ, vol.41, no.7, pp.379–386, 1996.

付　　録

1. 補題 2 の証明

簡単な計算により,

$$
\begin{aligned}
&-\log \sum_{m} P(x^n|m)P_B(m)\\
&= -\log\left\{P(x^n|m^*)P_B(m^*) + \sum_{m\in D_1} P(x^n|m)P_B(m)\right\}\\
&= -\log P(x^n|m^*) - \log P_B(m^*)\\
&\quad -\log\left\{1+\frac{\sum_{m\in D_1}P(x^n|m)P_B(m)}{P(x^n|m^*)P_B(m^*)}\right\}.
\end{aligned} \tag{A·1}
$$

ここで, $m \neq m^*$ に対しては, $\frac{P(x^n|m)}{P(x^n|m^*)} \to +0, a.s.$ であるから,

$$
\frac{\sum_{m\in D_1}P(x^n|m)P_B(m)}{P(x^n|m^*)P_B(m^*)} = o^+(1),\ a.s. \tag{A·2}
$$

となり, よって,

$$
\log\left\{1+\frac{\sum_{m\in D_1}P(x^n|m)P_B(m)}{P(x^n|m^*)P_B(m^*)}\right\} = o^+(1),\ a.s. \tag{A·3}
$$

となって, 補題が示された. □

2. 補題 4 の証明

まず, $\bar{\xi}^k \in \mathcal{A}(r)$ の場合について考える. このとき $\left\|\bar{\xi}^k\right\|_{I(\hat{\theta}^k)} > r$ である. 条件 2,(3) より, ξ^k の事後密度 $f_{\xi,B}(\xi^k|x^n)$ は $R_\xi(0) \in \{|R_\xi(0)| < \forall R\}$ に対して一様に式 (26) を満たし, また条件 2,(4) よりゆう度関数は単峰形であり, 更に条件 2,(6) より事前密度の導関数は有界であることから, R を十分大きくとれば, $\forall \bar{\xi}^k \in \mathcal{A}(r)$ と $\forall \epsilon_1 > 0$, $\forall r > 0$ に対して, $n \to \infty$ のとき,

$$
f_{\xi,B}(\bar{\xi}^k|x^n) \leqq \frac{\sqrt{\det I(\hat{\theta}^k)}}{(2\pi)^{k/2}} e^{-\frac{r^2}{2}} + \epsilon_1,\ a.s. \tag{A·4}
$$

が成り立つ. 更に, 条件 2,(1),(6) より, $\forall \theta^k \in \Theta^k$ に対して,

$$
0 < C_0 \leqq \sqrt{\det I(\theta^k)} < \infty, \tag{A·5}
$$

$$
0 < C_1 \leqq \frac{f_M(\theta^k)}{f_B(\theta^k)} \leqq C_2 < \infty, \tag{A·6}
$$

であることから, $\forall \bar{\xi}^k \in \mathcal{A}(r)$ と $\forall \epsilon_1 > 0$, $\forall r > 0$ に対して, $n \to \infty$ のとき

$$
\begin{aligned}
&\frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} \frac{f_M\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}\\
&\quad \leqq \frac{C_2}{C_0(2\pi)^{k/2}}\left\{\sqrt{\det I(\hat{\theta}^k)}e^{-\frac{r^2}{2}} + \epsilon_1(2\pi)^{k/2}\right\},\\
&\qquad a.s.
\end{aligned} \tag{A·7}
$$

が成り立つ. $r > 0$ と $\epsilon_1 > 0$ は任意だったので, この右辺は任意に小さくできることがわかる.

次に, $\bar{\xi}^k \in \bar{\mathcal{A}}(r)$ の場合を考える. $\sqrt{\det I(\theta^k)}$ の θ^k に関する微係数が存在するので, $\bar{\xi}^k \in \bar{\mathcal{A}}(r)$ に対して,

$$
\left|\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)} - \sqrt{\det I(\hat{\theta}^k)}\right| \leqq \frac{K_r(\hat{\theta}^k)}{\sqrt{n}}, \tag{A·8}
$$

を満たす $K_r(\hat{\theta}^k)$ が存在する. 一般に, $K_r(\hat{\theta}^k) > 0$ は $\hat{\theta}^k$ に依存する.

もし $n \to \infty$ のとき $\sqrt{\det I(\hat{\theta}^k)} \to \infty$ であるとすると，一般に $K_r(\hat{\theta}^k) \to \infty$ となってしまう．しかし条件 2,（4）より，$\theta^{k*} \in \Theta^k$ と $\forall \epsilon_2 > 0$ に対して，$n \to \infty$ のとき

$$\left\|\hat{\theta}^k - \theta^{k*}\right\|_{I(\theta^{k*})} < \epsilon_2, \quad a.s. \tag{A$\cdot$9}$$

が成り立つ．したがって，$\forall \epsilon_3 > 0$ に対して $n \to \infty$ のとき，

$$K_r(\hat{\theta}^k) \leqq K_r(\theta^{k*}) + \epsilon_3, \quad a.s. \tag{A$\cdot$10}$$

が成り立つ．ここで，$K_r(\theta^{k*})$ は確率変数である $\hat{\theta}^k$ には依存せず，真のパラメータ θ^{k*} に依存する(注10)．

したがって式 (A·6) とから，$\forall \epsilon_3 > 0$ に対して $n \to \infty$ のとき，

$$\frac{C_1 f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I(\hat{\theta}^k)} + \dfrac{K_r(\theta^{k*})+\epsilon_3}{\sqrt{n}}} \leqq \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}} \frac{f_M\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}, \quad a.s. \tag{A$\cdot$11}$$

が成り立ち，このとき同様に，

$$\frac{C_1 \max_{\bar{\xi}^k \in \bar{\mathcal{A}}(r)} f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I(\hat{\theta}^k)} + \dfrac{K_r(\theta^{k*})+\epsilon_3}{\sqrt{n}}} \leqq \max_{\bar{\xi}^k \in \bar{\mathcal{A}}(r)} \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}} \frac{f_M\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}, \quad a.s. \tag{A$\cdot$12}$$

が成り立つ．

ここで，$\bar{\xi}^k \in \bar{\mathcal{A}}(r)$ に対しては，条件 2 の漸近正規性の仮定と事後確率密度の連続性より，$\forall \epsilon_4 > 0$ に対して $n \to \infty$ のとき，

$$\left| f_{\xi,B}(\bar{\xi}^k|x^n) - \frac{\sqrt{\det I(\hat{\theta}^k)}}{(2\pi)^{k/2}} e^{-\frac{\|\bar{\xi}^k\|^2_{I(\hat{\theta}^k)}}{2}} \right| < \epsilon_4, \quad a.s. \tag{A$\cdot$13}$$

が成り立つ．一方，ξ^k 空間上での MDL の量子化幅 $\alpha^*_{\xi_j}(\bar{\xi}^k)$, $j = 0, 1, \cdots, k-1$ は，フィッシャー情報行列の主軸方向を向き，かつ

$$\prod_{j=0}^{k-1} \alpha^*_{\xi_j}(\bar{\xi}^k) = \frac{1}{\sqrt{\det I\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}} \tag{A$\cdot$14}$$

を満たすこと，及び式 (A·8)，(A·13) とから，$\forall \epsilon_5 > 0$ に対して，$n \to \infty$ のとき，

$$\begin{aligned}\frac{\sqrt{\det I(\hat{\theta}^k)}}{(2\pi)^{k/2}} e^{-\frac{1}{2}} - \epsilon_5 &\leqq \max_{\bar{\xi}^k \in \bar{\mathcal{A}}(r)} f_{\xi,B}(\bar{\xi}^k|x^n) \\ &\leqq \frac{\sqrt{\det I(\hat{\theta}^k)}}{(2\pi)^{k/2}} + \epsilon_5, \quad a.s.\end{aligned} \tag{A$\cdot$15}$$

が成り立つ．したがって，式 (A·12) と式 (A·15) より，$\forall \bar{\xi}^k \in \bar{\mathcal{A}}(r)$ と $\forall \epsilon_3, \epsilon_5 > 0$ に対して，$n \to \infty$ のとき，

$$\begin{aligned}&\frac{C_1}{\sqrt{\det I(\hat{\theta}^k)} + \dfrac{K_r(\theta^{k*})+\epsilon_3}{\sqrt{n}}} \cdot \left\{ \frac{\sqrt{\det I(\hat{\theta}^k)}}{(2\pi)^{k/2}} e^{-\frac{1}{2}} - \epsilon_5 \right\} \\ &\leqq \max_{\bar{\xi}^k \in \bar{\mathcal{A}}(r)} \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}} \cdot \frac{f_M\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}, \quad a.s.\end{aligned} \tag{A$\cdot$16}$$

が成り立つ．

$\epsilon_3, \epsilon_5 > 0$ は任意でかつ，条件 2,（4）より $n \to \infty$ で $\hat{\theta}^k \to \theta^{k*}$, $a.s.$ であるから，$\forall \epsilon_6 > 0$ に対して，$n \to \infty$ のとき，

$$\begin{aligned}&\frac{C_1 e^{-\frac{1}{2}}}{(2\pi)^{k/2}} - \epsilon_6 \\ &\leqq \max_{\bar{\xi}^k \in \bar{\mathcal{A}}(r)} \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}} \frac{f_M\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \dfrac{\bar{\xi}^k}{\sqrt{n}}\right)}, \quad a.s.\end{aligned} \tag{A$\cdot$17}$$

(注 10)：もちろん，もし $0 < C_0 < \sqrt{\det I(\theta^k)} < C_0' < \infty$，が $\forall \theta^k \in \Theta^k$ に対して仮定できれば，$\forall \hat{\theta}^k \in \Theta^k$ に対して $K_r(\hat{\theta}^k) < K_r$ となる $K_r > 0$ が存在する．

が成り立つ．

r は任意であったこと，式 (A·7) において正数 ϵ_1 も任意であること，及び式 (A·17) の左辺は正定数で下界されることから，十分大きい r を選べば，$n \to \infty$ のとき，

$$\tilde{\xi}^k \in \bar{\mathcal{A}}(r), \quad a.s. \tag{A·18}$$

が成り立つ．□

3. 補題5の証明

補題4より次式が成り立つ．

$$\max_{\bar{\xi}^k} \log \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} \frac{f_M\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}$$
$$= \max_{\bar{\xi}^k \in \bar{\mathcal{A}}(r)} \log \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} \cdot \frac{f_M\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)} \quad a.s. \tag{A·19}$$

$f_M(\theta^k)$，$f_B(\theta^k)$ の連続性と r が定数であること，及び $\hat{\theta}^k \to \theta^{k*}$，$a.s.$ より，$\forall\epsilon_7 > 0$ と $\bar{\xi}^k \in \bar{\mathcal{A}}(r)$ に対して，$n \to \infty$ のとき，

$$\left| \frac{f_M\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}{f_B\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)} - \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} \right| \leqq \epsilon_7, \quad a.s.$$

が成り立ち，このとき同時に

$$-\max_{\bar{\xi}^k} \log \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} - \log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} - \epsilon_7$$
$$\leqq L_{MDL}^{\bar{\theta}^k}(x^n) - L_{Bayes}^{\theta^k}(x^n)$$
$$\leqq -\max_{\bar{\xi}^k} \log \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} - \log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} + \epsilon_7, \quad a.s. \tag{A·20}$$

が成り立つ．補題4より $n \to \infty$ のとき

$$\tilde{\xi}^k \in \bar{\mathcal{A}}(r), \quad a.s. \tag{A·21}$$

であることと，$\bar{\xi}^k \in \bar{\mathcal{A}}(r)$ に対しては式 (A·15) が成り立つことから，$\forall\epsilon_8 > 0$ に対して $n \to \infty$ のとき，

$$\frac{k}{2}\log 2\pi - \epsilon_8 \leqq -\max_{\bar{\xi}^k} \log \frac{f_{\xi,B}(\bar{\xi}^k|x^n)}{\sqrt{\det I\left(\hat{\theta}^k + \frac{\bar{\xi}^k}{\sqrt{n}}\right)}} \leqq \frac{k}{2}\log 2\pi + \frac{1}{2} + \epsilon_8, \quad a.s. \tag{A·22}$$

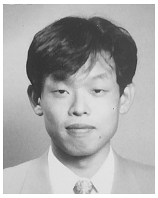
が成り立つ．

したがって，$\forall\epsilon_9 > 0$ に対して，$n \to \infty$ のとき，

$$\frac{k}{2}\log 2\pi - \log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} - \epsilon_9 \leqq L_{MDL}^{\bar{\theta}^k}(x^n) - L_{Bayes}^{\theta^k}(x^n) \leqq \frac{k}{2}\log 2\pi + \frac{1}{2} - \log \frac{f_M(\theta^{k*})}{f_B(\theta^{k*})} + \epsilon_9, \quad a.s. \tag{A·23}$$

が成り立ち，補題が得られた．□

（平成10年9月14日受付）



後藤　正幸　（正員）

平4武蔵工大・工・経営卒．平6同大大学院修士課程了．平6早大・理工学研究科・博士後期課程入学．平8同大・理工学部・経営システム工学科助手．現在，同大・メディアネットワークセンター・非常勤講師．情報源符号化，統計的学習理論，統計的モデル選択，ベイズ統計応用などの研究に従事．IEEE，情報理論とその応用学会，人工知能学会，日本経営工学会各会員．

松嶋　敏泰　（正員）

昭53早大・理工・工業経営卒．昭55同大大学院修士課程了．同年，日本電気（株）入社．昭61早大・理工学研究科・博士後期課程入学．平1横浜商科大学講師．平3同大助教授．平4早大・理工・工業経営学科（現在経営システム工学科）助教授，現在に至る．知識情報処理及び情報理論とその応用に関する研究に従事．工博．IEEE，情報理論とその応用学会，人工知能学会，情報処理学会，OR学会，日本経営工学会等各会員．

平澤　茂一　（正員）

昭36早大・理工・数学卒．昭38同大電気通信卒．同年三菱電機（株）入社．昭56早大・理工・工業経営学科（現在経営システム工学科）教授，現在に至る．情報理論とその応用，データ伝送方式，並びに計算機応用システムの開発などの研究に従事．工学博士．昭54UCLA計算機科学科客員研究員．昭60ハンガリー科学アカデミー，昭61伊トリエステ大学客員研究員．平5電子情報通信学会小林記念特別賞，業績賞受賞．平8情報理論とその応用学会会長．IEEE，情報理論とその応用学会，人工知能学会，情報処理学会，OR学会，日本経営工学会等各会員．