損失関数を考慮した拡張事後密度の漸近正規性

後藤　正幸[†a)]　　松嶋　敏泰[††]　　平澤　茂一[††]

Asymptotic Normality of Extended Posterior Density Functions with Loss Functions

Masayuki GOTOH[†a)], Toshiyasu MATSUSHIMA[††], and Shigeichi HIRASAWA[††]

あらまし　漸近正規性は統計的推測における本質的な性質である．例えば，J. Rissanen は最ゆう推定量の漸近正規性のもとで，ユニバーサル符号化，ユニバーサル予測，ユニバーサルモデル化に関する重要な性質を導いている．本論文では，損失関数を考慮した場合のパラメータの拡張事後密度を定式化する．従来，ベイズルールに基づく事後確率密度や拡張確率的コンプレキシティ(ESC) のための拡張事後密度が示されているが，本論文で定義する拡張事後密度は，損失に対し単調減少，事前確率に対し単調増加のみを仮定しただけの密度関数である．更に，漸近正規性のために必要な密度関数のもつべき条件を示す．これにより漸近正規性という観点から，損失関数を考慮した場合の拡張事後密度として望ましい形が再認識される．本論文で議論する拡張事後密度の漸近正規性は，確率測度の法則収束の意味とは異なるが，適当な条件のもとでベイズ事後確率密度が漸近正規性を満たすことが知られているように，本質的な性質といえる．本論文ではこの漸近正規性がいくつかの重要な結果を導くことを示す．

キーワード　拡張事後密度，漸近正規性，拡張確率的コンプレキシティ

1. まえがき

統計的推測において，適当な次元のパラメータをもつパラメトリックな確率モデルを仮定することがしばしば行われる．そして実際に得られたデータ系列から，そのパラメータに関して推定を行ったり，次に得られるデータの予測を行うために用いられる．実際に有用な方法の一つは，データ系列から何らかの形でパラメータを推定し，それを目的に応じて利用することであろう．

確率モデルのパラメータ推定としては，点推定，区間推定など多くの定式化があるが，本論文では事前分布が仮定できる場合を考える．パラメータ空間上に事前分布が仮定できれば，ベイズ規則により事後分布を得ることができ，ベイズ統計理論に基づく定式化が可能となる [1]．これは理論的にも非常に明快な解答を与える．一方で，データ系列からパラメータの推定量として，分布の形で推定量を得る方法が考えられる [22]．その一つはベイズ規則に基づく事後確率密度である．本論文では，統計的推測において損失関数を考慮した場合について，密度関数の形式でのパラメータ推定を定式化する．ここで得られる密度を拡張事後密度と呼ぶが(注1)，本論文で定義する拡張事後密度は，損失に対し単調減少，事前確率に対し単調増加のみを仮定しただけの密度関数である．更に，この拡張事後密度が漸近的に正規分布に収束するための関数の満たすべき条件を示す．

漸近正規性は統計的推測において本質的であり，J. Rissanen は最ゆう推定量の漸近正規性のもとで，ユニバーサル符号化，ユニバーサル予測，ユニバーサルモデル化の問題に対する重要な性質を導いている [16], [18]．最ゆう推定量の場合は，真の確率測度に基づく法則収束の意味での漸近正規性が議論される．一方，ベイズ規則に基づいて得られる事後確率密度が，適当な条件の

† 東京大学大学院工学系研究科，東京都
School of Engineering, The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo, 133–8656 Japan
†† 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, Shinjuku-ku, Tokyo, 169–8555 Japan
a) E-mail: goto@naoe.t.u-tokyo.ac.jp

(注1)：本論文では，ベイズの定理に基づいて計算される事後確率密度に対してのみ確率という言葉を付けて用いる．本論文で定義する密度関数は，損失関数に基づいて計算されるもので厳密な意味での事後確率分布を表していないので，単に事後密度と呼んでいる．

もとで漸近正規性を満たすことも知られている (A.M. Walker [21], C.C. Heyde and I.M. Johnston [11] など)．事後確率密度は個々のデータ系列に対応して得られるので，確率収束，概収束，一様収束などの意味での漸近正規性が議論の対象となるが [13]，この性質からもベイズ統計に基づく方法の漸近的評価が可能であるなど [3], [9]，重要な性質を導くことができる [1], [5]．拡張事後密度の漸近正規性はこのような議論の一般化となっている．その結果，拡張事後密度が損失に対し指数減少以上であるときに漸近正規性が成り立つことを示す．従来から議論されている拡張事後密度は損失に対し指数減少の形のものであり，確率論的にはこれが自然である．しかし，より広いクラスの拡張事後密度の性質を議論することによって，漸近正規性という観点から論理的に損失関数を考慮した場合の拡張事後密度として望ましい形が再認識される．

本論文では補足として，前半の結果から K. Yamanishi により提案された拡張確率的コンプレキシティ [23], [24] の漸近評価が可能となり，その意味が漸近正規性の面から再認識されるということを示す．ESC に関する概収束と平均収束の結果は [9] の議論と同じであるが，前半で定義する拡張事後密度が漸近正規性を満たすための条件の境界に位置するものが ESC であるという観点から，ESC の正当性について考察する．更に，データ系列のある部分集合について成り立つ一様収束の意味での漸近式についても述べる．

2. 準　　備

2.1 1 本の密度関数列の漸近正規性

まず最初に 1 本の Θ^k 上の密度関数列 $f_n(\theta^k)$, $n = 1, 2, 3, \cdots$ を考える．ただし $\theta^k \in \Theta^k$ かつ $\int_{\theta^k} f_n(\theta^k) d\theta^k = 1$, $n = 1, 2, 3, \cdots$ とする．$\Theta^k \subset \mathcal{R}^k$ とし，$\mathcal{R}^k$ は k-次元ユークリッド空間とする．

$\tilde{\theta}^k$ は $K_n(\theta^k) = \log f_n(\theta^k)$ のある極大値とし，

$$K_n'(\tilde{\theta}^k) = \mathbf{0}, \tag{1}$$

を満たし，かつ

$$\Sigma_n = -\left(K_n''(\tilde{\theta}^k)\right)^{-1} \tag{2}$$

は正定値行列であるとする．ただし $K_n'(\tilde{\theta}^k)$ と $K_n''(\tilde{\theta}^k)$ は

$$K_n'(\tilde{\theta}^k) = \left.\frac{\partial K_n(\theta^k)}{\partial \theta^k}\right|_{\theta^k = \tilde{\theta}^k} \tag{3}$$

$$K_n''(\tilde{\theta}^k) = \left.\frac{\partial^2 K_n(\theta^k)}{\partial \theta^k (\partial \theta^k)^T}\right|_{\theta^k = \tilde{\theta}^k} \tag{4}$$

であり，T は行列またはベクトルの転値である．

［補題 1］[1], [5]　球 $B_\delta(\tilde{\theta}^k) = \left\{\theta^k \in \Theta^k \middle| \|\theta^k - \tilde{\theta}^k\| < \delta\right\}$ を定義すると，次の三つの条件が密度関数 $f_n(\theta^k)$ の漸近正規性の必要十分条件である．

(c.1)　*"Steepness"* $\lim_{n\to\infty} \bar{\sigma}_n^2 \to 0$, ただし $\bar{\sigma}_n^2$ は Σ_n の最大固有値である．

(c.2)　*"Smoothness"* 任意の $\epsilon > 0$ に対し，自然数 N と $\delta > 0$ が存在し，$\forall n > N$ と $\forall \theta^k \in B_\delta(\tilde{\theta}^k)$ に対し $K_n''(\theta^k)$ が存在して，かつ

$$I - A(\epsilon) \leqq K_n''(\theta^k)\left\{K_n''(\tilde{\theta}^k)\right\}^{-1} \leqq I + A(\epsilon) \tag{5}$$

が $\forall \theta^k \in B_\delta(\tilde{\theta}^k)$ に対し一様に成り立つような $A(\epsilon)$ が存在する．ただし，I は $k \times k$ 単位行列，$A(\epsilon)$ は $\epsilon \to 0$ のときのその最大固有値が 0 に収束するような $k \times k$ 次元の非負定値行列とする．

(c.3)　*"Concentration"* 任意の $\delta > 0$ に対し

$$\int_{\theta^k \in B_\delta(\tilde{\theta}^k)} f_n(\theta^k) d\theta^k \to 1 \tag{6}$$

条件 (c.1)，(c.2)，(c.3) は

$$f_n(\tilde{\theta}^k)\,(\det \Sigma_n)^{1/2} \to (2\pi)^{-k/2} \tag{7}$$

を意味する．更に，条件 (c.1)，(c.2)，(c.3) のもとで，$\phi_n^k = \Sigma_n^{-1/2}(\theta^k - \tilde{\theta}^k)$ の密度関数，$f_{\phi,n}(\phi_n^k)$, は

$$\int_R f_{\phi,n}(\phi_n^k) d\phi_n^k$$

$$\to \int_R (2\pi)^{-k/2} \exp\left\{-\frac{1}{2}(\phi_n^k)^T \phi_n^k\right\} d\phi_n^k \tag{8}$$

を満たす．ただし，R は任意の直方体であり，$f_{\phi,n}(\phi_n^k)$ は

$$f_{\phi,n}(\phi_n^k) = (\det \Sigma_n)^{1/2} f_n(\theta^k) \tag{9}$$

で与えられる．　□

この補題は 1 本の密度関数列に対する結果であり，一般に式 (8) の性質を漸近正規性という．ただし，以下では紙面の都合のため，後の議論に必要な式 (7) の形の収束のみを結果として示す．式 (7) が成り立てば，ほぼ同様の手順によって式 (8) も得られる [1]．また，本論文ではデータ系列から計算されるパラメータ空間上の密度関数を議論するが，これは 1 本の密度関数列を考えているわけではない．

2.2 データ系列により与えられる事後密度の漸近正規性

補題 1 は統計的推測に利用できる形に拡張できる．すなわち，長さ n のデータ系列 x^n に依存する密度関数列 $f_{x^n}(\theta^k)$ を考える．ここで，$x^n = x_1x_2\cdots x_n$ は確率変数 $X^n = X_1X_2\cdots X_n$ の実現値であり，真の確率分布 $p^*(x^n)$ に従う長さ n の i.i.d. 系列であるとする．$\mathcal{X}$ をデータ空間，$x_i \in \mathcal{X}$ とする．また，$x^n \in \mathcal{X}^n$, すなわち $\mathcal{X}^n$ はデータ系列 x^n の取り得る空間，すなわち確率変数 X^n の値域とする．補題 1 より，以下の補題が明らかに成り立つ．

［補題 2］ $\tilde{\theta}^k = \tilde{\theta}^k(x^n)$ を $K_{x^n}(\theta^k) = \log f_{x^n}(\theta^k)$ の極大値とし，

$$K'_{x^n}\left(\tilde{\theta}^k\right) = \mathbf{0} \tag{10}$$

かつ

$$\Sigma_{x^n} = -\left(K''_{x^n}(\tilde{\theta}^k)\right)^{-1} \tag{11}$$

は正定値であるとする．

(c.1') *"Steepness"* $\lim_{n\to\infty} \bar{\sigma}^2_{x^n} \to 0$ $a.s.$, ただし $\bar{\sigma}^2_{x^n}$ は Σ_{x^n} の最大固有値であり，$a.s.$ は概収束を意味する．

(c.2') *"Smoothness"* 任意の $\epsilon > 0$ に対し，$\delta > 0$ が存在して，任意の $\theta^k \in B_\delta(\tilde{\theta}^k)$ に対して $L''_{x^n}(\theta^k)$ が存在して $n \to \infty$ のとき

$$I - A(\epsilon) \leqq K''_{x^n}(\theta^k)\left\{K''_{x^n}(\tilde{\theta}^k)\right\}^{-1} \leqq I + A(\epsilon), \quad a.s. \tag{12}$$

が成り立つ(注 2)．

(c.3') *"Concentration"* 任意の $\delta > 0$ に対し

$$\int_{\theta^k \in B_\delta(\tilde{\theta}^k)} f_{x^n}(\theta^k) d\theta^k \to 1, \quad a.s. \tag{13}$$

条件 (c.1)，(c.2)，(c.3) は

$$f_{x^n}(\tilde{\theta}^k)\left(\det \Sigma_{x^n}\right)^{1/2} \to (2\pi)^{-k/2}, \quad a.s. \tag{14}$$

を意味する．

条件 (c.1')，(c.2')，(c.3') の収束を $p^*(\cdot)$ に対する確率収束で置き換えると，

$$f_{x^n}(\tilde{\theta}^k)\left(\det \Sigma_{x^n}\right)^{1/2} \to (2\pi)^{-k/2}, \quad in\ prob. \tag{15}$$

が成り立つ．ただし，$in\ prob.$ は確率収束を意味する．

条件 (c.1')，(c.2')，(c.3') の収束を $p^*(\cdot)$ に対する平均収束で置き換えると，

$$E^*\left[f_{x^n}(\tilde{\theta}^k)\left(\det \Sigma_{x^n}\right)^{1/2}\right] \to (2\pi)^{-k/2} \tag{16}$$

が成り立つ．ただし，$E^*[\cdot]$ は $p^*(\cdot)$ に基づく期待値を表す．

データ系列のある部分集合 $\overline{\mathcal{X}}^n \subset \mathcal{X}^n$ を考えたとき，条件 (c.1')，(c.2')，(c.3') が，データ系列 $x^n \in \overline{\mathcal{X}}^n$ に対して一様に成り立つ収束で置き換えられるものとすると，$x^n \in \overline{\mathcal{X}}^n$ に対して一様に

$$f_{x^n}(\tilde{\theta}^k)\left(\det \Sigma_{x^n}\right)^{1/2} \to (2\pi)^{-k/2} \tag{17}$$

が成り立つ． □

3. 主 結 果

3.1 損失関数に基づく拡張事後密度

補題 2 では密度関数 $f_{x^n}(\theta^k)$ はデータ系列 x^n に依存するものとしたが，実際にどのように計算されるかについては言及しなかった．ここでは，データ系列からパラメータ空間上の密度関数の形で推定量を得る問題を考え [22]，なるべく一般的な定式化を試みる．

パラメータ θ^k で特長づけられる確率モデルを $p_{\theta^k}(\cdot)$ とし，確率モデル族を $\{p_{\theta^k}(\cdot)|\theta^k \in \Theta^k\}$ と記述する．$\pi(\theta^k)$ を Θ^k 上の事前確率密度とする．事前確率密度 $\pi(\theta^k)$ のもとで，データ系列 x^n を得たときに推定量である $f_{x^n}(\theta)$ を計算する方法を考えよう．すなわち，分布の形でパラメータの推定量を得る方法について考える．

例えば，ベイズ規則から得られる事後確率密度 $f_B(\theta^k|x^n)$ は

$$f_B(\theta^k|x^n) = \frac{p_{\theta^k}(x^n)\pi(\theta^k)}{\displaystyle\int_{\theta^k} p_{\theta^k}(x^n)\pi(\theta^k)d\theta^k} \tag{18}$$

と与えられる．以下では，これをベイズ事後密度と呼ぶ．

次に，これを任意の損失関数を考慮できる場合に拡張しよう．

$\mathcal{D}$ はユークリッド空間の部分集合，あるいは X 上の確率分布の集合を表すものとする．$\mathcal{D}$ を決定空間と呼び，その要素を d とし決定と呼ぶ．統計的推測のためにパラメタライズされた仮説空間 $\mathcal{H} = \{d_{\theta^k}|\theta^k \in \Theta^k\}$ を準備する．$\mathcal{H} \subseteqq \mathcal{D}$ であり，要素である d_{θ^k} を一つの

(注 2)：本論文では，ある事象 A_n に対し $P^*(\cup_{n=1}^{\infty} \cap_{k=n}^{\infty} \mathcal{A}_k) = 1$ であるとき，"$n \to \infty$ のとき A_n，$a.s.$ である" と記述する．確率収束や一様収束に関しても同様．

仮説（あるいは決定）といい，$\mathcal{H}$ は実数の集合，あるいは確率分布の集合である．Θ^k は k 次元パラメータ空間であり，k 次元ユークリッド空間 $\mathcal{R}^k$ の部分集合とする．

例えば確率分布 $p(x)$ を決定する問題の場合，$\mathcal{D}=\left\{p(x)\,\middle|\,\int_x dp(x)=1\right\}$ となる．特に $\mathcal{X}$ が離散集合の場合は $\mathcal{D}=\left\{p(x)\,\middle|\,\sum_x p(x)=1\right\}$ である．これに対して仮定される仮説空間が $\mathcal{H}=\{p_{\theta^k}(x)|\theta^k\in\Theta^k\}$ となる．

X の値を予測する問題なら，$\mathcal{D}=\mathcal{X}$ とすればよい. [23] では，二つの確率変数間の関係を議論する問題，すなわち X を得て Y の値や分布を予測するケースを扱っているが，本論文ではより単純に X のみを考えている．もしこのような場合を考えるならば，$\mathcal{X}\to\mathcal{Y}$ と写像する関数の集合を仮説空間とし $\mathcal{D}=\{f(Y|X)\}$, $\mathcal{H}=\{f_{\theta^k}(Y|X)\}$ などとすればよい．

$L(d:x):\mathcal{D}\times\mathcal{X}\to[0,\infty)$ を決定 d と $x\in\mathcal{X}$ 間の損失関数とする．例えば，決定空間 $\mathcal{D}$ が確率モデルのクラス $\{p(X)\}$ であるとき，すなわち $\mathcal{H}=\{p_{\theta^k}(X):\theta^k\in\Theta^k\}$ であるときには，損失関数は $L(p_{\theta^k}(\cdot):X)$ のように与えられる．例えば X が離散確率変数のとき，α-損失関数は

$$L(p_{\theta^k}(\cdot):X)=\left(1-p_{\theta^k}(X)\right)^{\alpha} \tag{19}$$

で与えられ(注 3)，対数損失関数は

$$L(p_{\theta^k}(\cdot):X)=-\log p_{\theta^k}(X) \tag{20}$$

で与えられる．その他の決定空間や損失関数については [23] を参照.

本論文では，次の形の拡張事後密度 $f_{x^n}(\theta^k)$ を考える．

$$f_{x^n}(\theta^k)=\frac{g\left(\sum_{t=1}^{n}L(d_{\theta^k}:x_t),\pi(\theta^k)\right)}{\int_{\theta^k} g\left(\sum_{t=1}^{n}L(d_{\theta^k}:x_t),\pi(\theta^k)\right)d\theta^k} \tag{21}$$

ここで，関数 $g(y,z)$ は $y\in[0,+\infty)$ に関して単調減少，$z\in[0,+\infty)$ に関して単調増加な非負関数とする．その意味するところは，事前確率密度が高く，損失が小さいパラメータには高い事後密度が割り当てられるということである．ここで，$g'(y,z)=\frac{\partial g(y,z)}{\partial y}$ と記述することにすると，$g(y,z)$ は非負であるから下に有界であるので，$y\to\infty$ のとき

$$g'(y,z)=O(g(y,z)) \tag{22}$$

であることに注意しておく(注 4)．なぜなら，平均値の定理より $\forall y_1,\forall y_2\in[0,+\infty)$ に対して

$$g(y_1,z)=g(y_2,z)+g'(y_3,z)(y_1-y_2)$$

かつ $y_1\geqq y_3\geqq y_2$ となる $y_3\in[0,+\infty)$ が存在するので，$y_2\to\infty$ を考えたとき，$\left|\frac{g'(y,z)}{g(y,z)}\right|\to\infty$ とすると $g(y,z)$ は有界となり得なくなってしまうからである．

もし $g(y,z)=\frac{\lambda}{y}+z$ であれば

$$f_{x^n}(\theta^k)=\frac{\frac{\lambda}{\sum_{t=1}^{n}L(d_{\theta^k}:x_t)}+\pi(\theta^k)}{\int_{\theta^k}\left(\frac{\lambda}{\sum_{t=1}^{n}L(d_{\theta^k}:x_t)}+\pi(\theta^k)\right)d\theta^k} \tag{23}$$

となる．類似の形として

$$f_{x^n}(\theta^k)=\frac{\sum_{t=1}^{n}\frac{\lambda}{L(d_{\theta^k}:x_t)}+\pi(\theta^k)}{\int_{\theta^k}\left(\sum_{t=1}^{n}\frac{\lambda}{L(d_{\theta^k}:x_t)}+\pi(\theta^k)\right)d\theta^k} \tag{24}$$

が考えられる．

もし $g(y,z)=z\exp\{-\lambda y\}$ であれば

$$f_{x^n}(\theta^k)=\frac{\exp\left(-\lambda\sum_{t=1}^{n}L(d_{\theta^k}:x_t)\right)\pi(\theta^k)}{\int_{\theta^k}\exp\left(-\lambda\sum_{t=1}^{n}L(d_{\theta^k}:x_t)\right)\pi(\theta^k)d\theta^k} \tag{25}$$

となる．これは K. Yamanishi [23] により提案された ESC に対して自然に導かれる事後密度である．

(注 3)：式 (19) は X が離散確率変数の場合の α-損失関数である．例えば $\mathcal{D}=\mathcal{X}$ の場合には，$|X-d|^{\alpha}$ と定義される．

(注 4)：本論文を通じて次の記法を用いる．$g(y)$, $h(y)$ を y の関数とし，$g(y)=O(h(y))$ は，$\forall y\geqq y_0$ に対し $|g(y)|\leqq c|h(y)|$ となるような正定数 c, y_0 が存在することを意味する．また，$g(y)=o(h(y))$ は $\lim_{y\to\infty}\frac{g(y)}{h(y)}=0$, $g(y)=\Omega(h(y))$ は $g(y)\neq o(h(y))$ であることを意味する．

ある損失関数 $L(d_{\theta^k}:x)$ に対し，$L_1(d_{\theta^k}:x)=\exp\{L(d_{\theta^k}:x)\}$ と定義すると，式 (25) は

$$f_{x^n}(\theta^k)=\frac{\prod_{t=1}^{n}\left(L_1(d_{\theta^k}:x_t)\right)^{-\lambda}\pi(\theta^k)}{\int_{\theta^k}\prod_{t=1}^{n}\left(L_1(d_{\theta^k}:x_t)\right)^{-\lambda}\pi(\theta^k)d\theta^k} \tag{26}$$

と書き換えられる．したがって，式 (23) と式 (25) はそれぞれ，損失の和と積に比例する形の密度関数を与えている．

もし $g(y,z)=z^{\lambda}\exp\{-\lambda y\}$ とすれば，式 (26) と類似した形として，

$$f_{x^n}(\theta^k)=\frac{\prod_{t=1}^{n}\Big\{L_1(d_{\theta^k}:x_t)\pi(\theta^k)\Big\}^{\lambda}}{\int_{\theta^k}\prod_{t=1}^{n}\Big\{L_1(d_{\theta^k}:x_t)\pi(\theta^k)\Big\}^{\lambda}d\theta^k} \tag{27}$$

が導かれる．$L_1(d_{\theta^k}:x_t)=p_{\theta^k}(x_t)$ であるとき，ボルツマン分布

$$f_{x^n}(\theta^k)=\frac{\Big\{p_{\theta^k}(x^n)\pi(\theta^k)\Big\}^{\lambda}}{\int_{\theta^k}\Big\{p_{\theta^k}(x^n)\pi(\theta^k)\Big\}^{\lambda}d\theta^k} \tag{28}$$

が得られる [22]．もし $\lambda=1$ であれば，ベイズ事後密度となる．$\lambda=\infty$ と考えれば，これは事後確率最大推定量と等価になる [22]．

ここで一つの疑問が生ずる．関数 $g(y,z)$ がどのような形であれば，これから計算される拡張事後密度 $f_{x^n}(\theta^k)$ が漸近正規性を満たすか，という問題である．次節では，漸近正規性が成り立つために関数 $g(y,z)$ が満たすべき条件を示す．

3.2 主 定 理

まず，次の情報行列 $I^*(\theta^k)$ と $J^*(\theta^k)$ を定義しておく．

$$I^*(\theta^k)=E^*\left[\frac{\partial^2 L(d_{\theta^k}:X)}{\partial\theta^k(\partial\theta^k)^T}\right] \tag{29}$$

$$J^*(\theta^k)=E^*\left[\frac{\partial L(d_{\theta^k}:X)}{\partial\theta^k}\frac{\partial L(d_{\theta^k}:X)}{(\partial\theta^k)^T}\right] \tag{30}$$

$I^*(\theta^k)$ と $J^*(\theta^k)$ は常に存在するとは限らないが，本論文では $\forall\theta^k\in\Theta^k$ に対しこれらが存在するものとする．

損失最小推定量 $\hat{\theta}^k$ と真の分布に対する最適パラメータ θ^{k^*} を

$$\hat{\theta}^k=\arg\min_{\theta^k}\frac{1}{n}\sum_{t=1}^{n}L(d_{\theta^k}:x_t) \tag{31}$$

$$\theta^{k*}=\arg\min_{\theta^k}E^*\left[L(d_{\theta^k}:X)\right] \tag{32}$$

とする．$\forall\delta>0$ に対し，球 $B_\delta(\theta^{k*})$ を $B_\delta(\theta^{k*})=\left\{\theta^k\in\Theta^k\Big|\|\theta^k-\theta^{k*}\|<\delta\right\}$ と定義する．同様に，$B_\delta(\tilde{\theta}^k)=\left\{\theta^k\in\Theta^k\Big|\|\theta^k-\tilde{\theta}^k\|<\delta\right\}$ としよう．

次に我々の解析に必要となる条件を示す．

［条件 1］ (1) Θ^k はコンパクト集合である．

(2) $\exists c_1<\inf_{\theta^k\in\Theta^k}\pi(\theta^k)$，かつ $\sup_{\theta^k\in\Theta^k}\pi(\theta^k)<\exists c_2$ である．ただし，c_1 と c_2 はある正定数である．更に，$\pi(\theta^k)$ は Θ^k 上で連続 2 回微分可能である．

(3) θ^{k*} は Θ^k の内部で一意に定まる．

(4) 非負関数 $g(y,z)$ は，$y,z\in[0,\infty)$ である y に関して単調減少，z に関して単調増加関数である．また $g(y,z)$ は連続 2 回微分可能である．

(5) $I^*(\theta^k)$ は Θ^k 内の任意の θ^k に関して連続微分可能，かつ $I^*(\theta^{k*})$ は正定値である．

(6) $\forall n\in\{1,2,\cdots\}$ に対し $\frac{1}{n}\sum_{t=1}^{n}L(d_{\theta^k}:x_t)$ は，ほとんど確実に Θ^k 上で連続である．

(7) $E^*\left[L(d_{\theta^k}:X)\right]$ は Θ^k 上で任意の θ^k に関して連続 2 回微分可能である．$\forall n\in\{1,2,\cdots\}$ に対し $\frac{1}{n}\sum_{t=1}^{n}L(d_{\theta^k}:x_t)$ は，ほとんど確実に Θ^k 上で連続 2 回微分可能である．

(8) Θ^k 上で一様に

$$\frac{1}{n}\sum_{t=1}^{n}L(d_{\theta^k}:x_t)\to E^*\left[L(d_{\theta^k}:X)\right],\quad a.s.$$

$$\frac{1}{n}\frac{\partial}{\partial\theta^k}\sum_{t=1}^{n}L(d_{\theta^k}:x_t)\to\frac{\partial}{\partial\theta^k}E^*\left[L(d_{\theta^k}:X)\right],\quad a.s.$$

$$\frac{1}{n}\frac{\partial^2}{\partial\theta^k(\partial\theta^k)^T}\sum_{t=1}^{n}L(d_{\theta^k}:x_t)\to I^*(\theta^k),\quad a.s.$$

(9) (中心極限定理) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^{k*}}:X_t)}{\partial\theta^k}$ は正規分布 $N\big(\mathbf{0},J^*(\theta^{k*})\big)$ に法則収束する． □

この条件は，先の論文 [9] の仮定と同一のものであり，パラメータ空間の境界などの特異点を除けば，広く成り立つ現実的な仮定といえる．同様の条件は [12], p.238 で扱われている．この条件を満足する例については [9], [12] を参照．

もし条件 (c.1’)，(c.2’)，(c.3’) が成り立てば，拡張事後密度 (21) は漸近正規性を満たす．以下では，条件 1のもとで，条件 (c.1’)，(c.2’)，(c.3’) が成り立つための関数 $g(y, z)$ の条件を議論しよう．

［定理 1］(注 5)　条件 1(1) ∼ (8) を仮定する．

もし $y \to \infty$ のとき $\log g(y, z) \neq O\left(\log \frac{1}{y}\right)$ であれば，

$$f_{x^n}(\tilde{\theta}^k)\,(\det \Sigma_{x^n})^{1/2} \to (2\pi)^{-k/2},\quad a.s. \tag{33}$$

$$E^*\left[f_{x^n}(\tilde{\theta}^k)\,(\det \Sigma_{x^n})^{1/2}\right] \to (2\pi)^{-k/2} \tag{34}$$

が成り立つ．漸近式 (34) は条件 1(8) の概収束を $p^*(\cdot)$ に対する平均収束に置き換えても成り立つ．

もし，条件 1(8) が確率収束で置き換えられたとすると，

$$f_{x^n}(\tilde{\theta}^k)\,(\det \Sigma_{x^n})^{1/2} \to (2\pi)^{-k/2},\quad in\, prob. \tag{35}$$

となる．

一方，もし $y \to \infty$ のとき $\log g(y, z) = O\left(\log \frac{1}{y}\right)$ であれば，$f_{x^n}(\tilde{\theta}^k)$ の漸近正規性は成り立たない．

（証明）　付録参照．　□

次に一様収束を議論するために，$\tilde{\theta}^k \in \overline{\Theta}^k$，かつ

$$\frac{1}{n}\sum_{t=1}^{n} L(d_{\tilde{\theta}^k} : x_t) < C \tag{36}$$

を満たす $x^n \in \mathcal{X}^n$ の部分集合 $\overline{\mathcal{X}}_C^n \subset \mathcal{X}^n$ を定義しよう．ここで $\overline{\Theta}^k$ は Θ^k の内部に含まれるある部分集合である．条件 1(8) の代わりに，$x^n \in \overline{\mathcal{X}}_C^n$ となる x^n を考える．

更に，条件 1(6)，(7) を次のように置き換える．

［条件 2］　(6’)　$\forall x^n \in \overline{\mathcal{X}}_C^n$ と $\forall n \in \{1, 2, \cdots\}$ に対し，$\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^k} : x_t)$ は Θ^k に関して連続である．

(7’)　$\forall x^n \in \overline{\mathcal{X}}_C^n$ と $\forall n \in \{1, 2, \cdots\}$ に対し，$\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^k} : x_t)$ は Θ^k 上で 2 回連続微分可能である．　□

［定理 2］　条件 1(1)，(2)，(4) と条件 2 を仮定する．

もし $y \to \infty$ のとき $\log g(y, z) \neq O\left(\log \frac{1}{y}\right)$ であれば，

$$f_{x^n}(\tilde{\theta}^k)\,(\det \Sigma_{x^n})^{1/2} \to (2\pi)^{-k/2} \tag{37}$$

が $x^n \in \overline{\mathcal{X}}_C^n$ に対して一様に成り立つ．　□

（概証明）　条件 2(5’) と $x^n \in \overline{\mathcal{X}}_C^n$ より，$\frac{1}{n}\sum_{t=1}^{n} L''(d_{\tilde{\theta}^k} : x_t)$ と $\frac{1}{n}\sum_{t=1}^{n} L''(d_{\tilde{\theta}^k} : x_t)$ は有界である．したがって，定理 1の証明と同様の展開により，定理が得られる [13].　□

$\log g(y, z) \neq O\left(\log \frac{1}{y}\right)$ という条件は，$g(y, z)$ が y に関して多項式オーダの減少関数では漸近正規性は成り立たず，指数オーダ以上になると成り立つことを示している．この $g(y, z)$ の条件より，式 (23) の拡張事後密度は漸近正規性を有さないことがわかる．同様に，式 (24) も漸近正規性を満たさない．一方，式 (25) は漸近正規性を満たす．しかし定理 1 より，$g(y, z)$ が y に関して指数オーダより早く減少するようなものであっても，漸近正規性を満たすこともわかる．

4. 拡張確率的コンプレキシティの評価

与えられた正定数 $\lambda > 0$ のもとで．拡張確率的コンプレキシティ(ESC) は次式で定義される [23], [24].

$$ESC(x^n) = -\frac{1}{\lambda}\log\int \exp\left(-\lambda\sum_{t=1}^{n} L(d_{\theta^k} : x_t)\right)\pi(\theta^k)d\theta^k \tag{38}$$

ESC はオンライン学習の the aggregating algorithm に適用されるなど，学習問題の性能を保証するために有用であることが示されている [23]．その精密な漸近式を求めることは，ESC を用いた学習アルゴリズムの漸近性能に対する保証を与える．

一方，本論文で与えた拡張事後密度の漸近正規性を応用することにより，ESC の漸近式が自然に導かれる [9]．概収束と平均収束の意味での ESC の評価は [9] で行っているが，ここでは一様収束の結果も示し，前節の漸近正規性の成立条件と併せて考察する．これにより，ESC の漸近式において漸近正規性が本質的であるだけでなく，ESC が漸近正規性という観点からも自然なものであることが再認識される．

まず，$h(x^n)$, $h(z^n|\theta^k)$, $\pi(\theta^k|x^n)$ を

$$h(x^n) = \int \exp\left(-\lambda\sum_{t=1}^{n} L(d_{\theta^k} : x_t)\right)\pi(\theta^k)d\theta^k \tag{39}$$

（注 5）：本論文では拡張事後密度最大推定量 $\tilde{\theta}^k$ を用いた展開のみを示す．$\tilde{\theta}^k$ を $\hat{\theta}^k$ で置き換えても，[9] と同様の議論により全く同様の結果が成り立つことが容易にわかる．

$$h(x^n|\theta^k) = \exp\left(-\lambda\sum_{t=1}^{n} L(d_{\theta^k} : x_t)\right) \tag{40}$$

$$\pi(\theta^k|x^n) = \frac{\exp\left(-\lambda\sum_{t=1}^{n} L(d_{\theta^k} : x_t)\right)\pi(\theta^k)}{\int \exp\left(-\lambda\sum_{t=1}^{n} L(d_{\theta^k} : x_t)\right)\pi(\theta^k)d\theta^k} \tag{41}$$

と定義する．ただし，$\pi(\theta^k|x^0) = \pi(\theta^k), L(d_{\theta^k} : x_0) = 0$ である．

このとき，次の補題が成り立つ．

［補題 3］[9]　$h(x^n), h(x^n|\theta^k), \pi(\theta^k|x^n)$, に対し次式の関係が成り立つ．

$$h(x^n) = \frac{h(x^n|\theta^k)\pi(\theta^k)}{\pi(\theta^k|x^n)} \tag{42}$$

□

式 (42) はベイズ規則

$$p(x^n) = \frac{p(x^n|\theta^k)\pi(\theta^k)}{f_B(\theta^k|x^n)} \tag{43}$$

の一般化である．式 (42) より，

$$ESC(x^n) = -\frac{1}{\lambda}\log h(x^n|\theta^k)\pi(\theta^k) + \frac{1}{\lambda}\log\pi(\theta^k|x^n) \tag{44}$$

となるから，$\log\pi(\theta^k|x^n)$ の漸近式が与えられれば，$ESC(x^n)$ の漸近式が得られることがわかる．

$\pi(\theta^k|x^n)$ は Θ^k 上の密度関数であるが，一般にはベイズ規則で計算されるベイズ事後確率密度ではない．もし対数損失と $\lambda = 1$ を適用すると，$\pi(\theta^k|x^n)$ はベイズ事後確率密度となり，ESC は確率的コンプレキシティ (SC) [17]

$$SC(x^n) = -\log\int_{\theta^k} p_{\theta^k}(x^n)\pi(\theta^k)d\theta^k \tag{45}$$

となる．

定理 1 の漸近正規性より，次の定理が得られる [9]．

［定理 3］[9]　条件 1(1)∼(8) のもとで，

$$\frac{\pi(\tilde{\theta^k}|x^n)}{\sqrt{n}^k} \to \left(\frac{\lambda}{2\pi}\right)^{k/2}\sqrt{\det I^*(\tilde{\theta^k})}, \quad a.s. \tag{46}$$

が成り立つ．　□

また，$\tilde{\theta^k}$ は $\hat{\theta^k}$ で置き換えることもできる [9]．定理 3 は，

$$(\Sigma_{x^n})^{-1} = -\lambda\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k} : x_t)}{\partial\theta^k(\partial\theta^k)^T} + \frac{\partial^2\log\pi(\theta^k)}{\partial\theta^k(\partial\theta^k)^T} \tag{47}$$

より得られる

$$-\frac{1}{n}(\Sigma_{x^n})^{-1} \to \lambda I^*(\tilde{\theta^k}), \quad a.s. \tag{48}$$

という式から導かれる．

この定理と補題 3 より，直ちに次の定理が得られる．

［定理 4］[9]　条件 1(1)∼(8) のもとで，$ESC(x^n)$ の漸近式は

$$ESC(x^n) = \sum_{t=1}^{n} L(d_{\hat{\theta}^k} : x_t) + \frac{k}{2\lambda}\log\frac{n\lambda}{2\pi} + \frac{1}{\lambda}\log\frac{\sqrt{\det I^*(\hat{\theta}^k)}}{\pi(\hat{\theta}^k)} + o(1), \quad a.s. \tag{49}$$

$$= \sum_{t=1}^{n} L(d_{\tilde{\theta}^k} : x_t) + \frac{k}{2\lambda}\log\frac{n\lambda}{2\pi} + \frac{1}{\lambda}\log\frac{\sqrt{\det I^*(\tilde{\theta}^k)}}{\pi(\tilde{\theta}^k)} + o(1), \quad a.s. \tag{50}$$

と与えられる．　□

上の定理は真の分布に対する概収束の意味で導かれている．すなわち，$\hat{\theta}^k$ と $\tilde{\theta}^k$ は確率変数である．

［補題 4］　[12], pp.240-241　条件 1のもとで，

$$E^*\left[\frac{\sum_{t=1}^{n} L(d_{\hat{\theta}^k} : X_t)}{\sum_{t=1}^{n} L(d_{\theta^{k*}} : X_t)}\right] \to \frac{TrJ^*(\theta^{k*})\{I^*(\theta^{k*})\}^{-1}}{2} \tag{51}$$

が成り立つ．ただし，Tr は行列のトレースを表す．

□

補題 4 より，$ESC(x^n)$ の期待値 $E^*[ESC(X^n)]$ の漸近評価が可能となる．

［定理 5］[9]　条件 1 のもとで，$E^*[ESC(X^n)]$ は

$$E^*[ESC(X^n)] = E^*[L(d_{\theta^{k*}} : X_t)] + \frac{k}{2\lambda}\log\frac{n\lambda}{2\pi}$$

$$
-\frac{TrJ^*(\theta^{k*})\{I^*(\theta^{k*})\}^{-1}}{2\lambda} + \frac{1}{\lambda}\log\frac{\sqrt{\det I^*(\theta^{k*})}}{\pi(\theta^{k*})} + o(1) \tag{52}
$$

を満たす. □

式 (49) と式 (52) より，SC の漸近式が

$$
SC(x^n) = -\log p_{\hat\theta^k}(x^n) + \frac{k}{2}\log\frac{n}{2\pi} + \log\frac{\sqrt{\det I^*(\hat\theta^k)}}{\pi(\hat\theta^k)} + o(1), \quad a.s. \tag{53}
$$

$$
\begin{aligned}
&E^*[SC(X^n)] \\
&= -E^*[\log p_{\theta^{k*}}(X^n)] + \frac{k}{2}\log\frac{n}{2\pi} \\
&\quad -\frac{TrJ^*(\theta^{k*})\{I^*(\theta^{k*})\}^{-1}}{2} + \log\frac{\sqrt{\det I^*(\theta^{k*})}}{\pi(\theta^{k*})} \\
&\quad + o(1)
\end{aligned} \tag{54}
$$

と与えられることもわかる.

以上の結果は，概収束と平均収束の意味での ESC の漸近式である．これに対し Yamanishi は平均的意味とともに，データ系列に対して一様に成り立つ ESC の漸近的上界を導いている [23] が，一様に成り立つ上界を考える代償として，厳しい漸近式は得られていない．平均収束の意味での ESC の漸近式 (52) は Yamanishi の上界がきついものであることを示しており，概収束の意味での ESC の漸近式 (49) は同様の収束がほとんど確実に得られる系列に対して成り立つことを述べている.

一方，データ系列に対して一様に漸近正規性が成り立つかどうかを考えると，一般に答えは否であり，データ系列の部分集合を考えなければならない．これに対する結果は以下のようになる.

［定理 6］ 条件 1(1)，(2)，(4) と条件 2 のもとで，

$$
ESC(x^n) = \sum_{t=1}^{n} L(d_{\tilde\theta^k} : x_t) + \frac{k}{2\lambda}\log\frac{n\lambda}{2\pi} + \frac{1}{\lambda}\log\frac{\sqrt{\det \hat I(\tilde\theta^k)}}{\pi(\hat\theta^k)} + o(1), \tag{55}
$$

が $x^n \in \overline{\mathcal{X}}_C^n$ に対して一様に成り立つ．ここで $\hat I(\theta^k)$ は経験情報行列であり，

$$
\hat I(\theta^k) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 L(d_{\theta^k} : x_t)}{\partial\theta^k(\partial\theta^k)^T} \tag{56}
$$

で与えられる. □

以上の結果は，ESC の漸近式の定数オーダの項が本質的に漸近正規性から導かれることを示している．また，定理 1 から，式 (41) の拡張事後密度は漸近正規性を満たすための条件の境界に位置するものであることがわかる．条件 1 のもとで，損失最小推定量を基準化した統計量は漸近的に正規分布 $N\big(\mathbf{0}, \{I^*(\theta^k)\}^{-1}J^*(\theta^k)\{I^*(\theta^k)\}^{-1}\big)$ に法則収束することが知られているが [12]，事後密度の共分散がこの $\{I^*(\theta^k)\}^{-1}J^*(\theta^k)\{I^*(\theta^k)\}^{-1}$ に比べて過度に小さいのは意味がない．なぜなら，共分散 $\{I^*(\theta^k)\}^{-1}J^*(\theta^k)\{I^*(\theta^k)\}^{-1}$ 以上にパラメータを精度良く特定するための情報は，確率的に生起するデータ系列には含まれていないからである．逆に関数 $g(y,z)$ を調整して事後密度の共分散を大きくしようとすると，いずれ漸近正規性が成り立たなくなる．それ以上共分散のオーダを大きくすると漸近正規性を満たさなくなるような境界に位置する場合が ESC で定義されている拡張事後密度であり，拡張事後密度の共分散と損失最小推定量の共分散が同じオーダになるようなケースになっている．その意味からも ESC は自然なものであることが再認識できる.

5. むすび

本論文では，損失関数を考慮した拡張事後密度による推定を考え，その漸近正規性の成り立つ条件を示した．更に，この結果を用いて確率的コンプレキシティ，拡張確率的コンプレキシティの漸近式を示した.

漸近正規性は，総計的推測において本質的な性質である．しかし，拡張確率的コンプレキシティの評価においては，個々のデータ系列 $\mathcal{X}^n$ に対して一様に成り立つ漸近式が重要視されている．漸近正規性から考察した場合，パラメータの境界領域などの経験情報行列 $\hat I(\tilde\theta^k)$ が発散する点において，一様性が保証できない．このようなケースの考察のためには，異なる視点が解析が必要である [20].

文 献

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

[2] G.E.P. Box and G.C. Tiao, Bayesian Inference in Statistical Analysis, John Wiley & Sons, Inc., 1992.

[3] B.S. Clarke, “Asymptotic normality of the posterior in relative entropy,” IEEE Trans. Inf. Theory, vol.45, no.1, pp.165–176, 1999.

[4] B.S. Clarke and A.R. Barron, “Information—Theoretic asymptotics of Bayes methods,” IEEE

Trans. Inf. Theory, vol.36, no.3, pp.453–471, 1990.
[5] C.-F. Chen, “On asymptotic normality of limiting density function with Bayesian implications,” J. R. Statist. Soc. B, vol.47, no.3, pp.540–546, 1988.
[6] W. Feller, An Introduction to Probability and Its Applications, vol.1 and 2, John Wiley & Sons, New York, 1957.
[7] T.S. Ferguson, Mathematical Statistics—A Decision Theoretic Approach, New York and London, Academic, 1967.
[8] M. Gotoh, T. Matsushima, and S. Hirasawa, “A generalization of B.S. Clarke and A.R. Barron’s asymptotics of Bayes codes for FSMX sources,” IEICE Trans. Fundamentals, vol.E81-A, no.10, pp.2123–2132, 1998.
[9] M. Gotoh, T. Matsushima, and S. Hirasawa, “Almost sure and mean convergence of extended stochastic complexity,” IEICE Trans. Fundamentals, vol.E82-A, no.10, pp.2129–2137, Oct. 1999.
[10] J.A. Hartigan, Bayes Theory, Springer-Verlag, 1983.
[11] C.C. Heyde and I.M. Johnstone, “On asymptotic posterior normality for stochastic process,” J. R. Statist. Soc. B, vol.41, no.2, pp.184–189, 1979.
[12] H. Linhart and W. Zucchini, Model Selection, John Wiley & Sons, 1986.
[13] 丸山英昭，後藤正幸，平澤茂一，“事後確率密度の漸近正規性に関する一考察，” 信学技報，IT99-28, 1999.
[14] N. Murata, S. Yoshizawa, and S. Amari, “Network information criterion—Determining the number of hidden units for an artificial neural network model,” IEEE Trans. Neural Networks, vol.5, no.6, pp.865–872, 1994.
[15] D.S. Poskitt, “Precision, complexity and Bayesian model determination,” J. R. Statist. Soc. B, vol.49, no.2, pp.199–208, 1987.
[16] J. Rissanen, “Universal conding, information, prediction, and estimation,” IEEE Trans. Inf. Theory, vol.IT-30, no.4, pp.629–636, 1984.
[17] J. Rissanen, “Stochastic complexity,” J. R. Statist. Soc. B, vol.49, pp.223–265, 1987.
[18] J. Rissanen, “Fisher information and stochastic complexity,” IEEE Trans. Inf. Theory, vol.42, no.1, pp.40–47, 1996.
[19] 竹内　啓，“情報量基準の分布とモデルの適切さの規準，” 数理科学，no.153, pp.12–18, 1976.
[20] 竹内純一，“確率的コンプレキシティと Jeffreys 混合予測戦略，” 1998 年情報論的学習理論ワークショップ予稿集, pp.9–16, 1998.
[21] A.M. Walker, “On the asymptotic behaviour of posterior distributions,” J. R. Statist. Soc. B, vol.31, pp.80–88, 1969.
[22] 渡辺澄夫，“ベイズ法による階層型統計モデルの推定誤差について，” 信学論（A），vol.J81-A, no.10, pp.1442–1452, Oct. 1998.
[23] K. Yamanishi, “A decision-theoretic extension of stochastic complexity and its applications to learning,” IEEE Trans. Inf. Theory, vol.44, no.4, pp.1424–1439, 1998.
[24] 山西健司，“拡張型確率的コンプレキシティと学習理論，” 1998 年情報論的学習理論ワークショップ予稿集, pp.33–40, 1998.

付　　録

定理 1 の証明

紙面の都合上，概収束に関する証明のみを与える．他は収束の強さの違いであるが，ほぼ同様に証明できる．

漸近正規性を示すためには，補題 2 の条件 (c.1’)∼(c.3’) が成り立つことを示せばよい．

簡単のため，

$$g'(y,z) = \frac{\partial g(y,z)}{\partial y} \tag{A·1}$$

$$g''(y,z) = \frac{\partial^2 g(y,z)}{(\partial y)^2} \tag{A·2}$$

$$g = g\left(\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right) \tag{A·3}$$

$$g' = g'\left(\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right) \tag{A·4}$$

$$g'' = g''\left(\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right) \tag{A·5}$$

などと記述する．条件 1(4)，(7) から

$$\frac{\partial}{\partial\theta^k} K_{x^n}(\theta^k) = \frac{\partial \log g}{\partial\theta^k} = \frac{1}{g}\frac{\partial g}{\partial\theta^k} \tag{A·6}$$

$$\frac{\partial^2 K_{x^n}(\theta^k)}{\partial\theta^k(\partial\theta^k)^T} = \frac{1}{g}\frac{\partial^2 g}{\partial\theta^k(\partial\theta^k)^T} - \frac{1}{g^2}\frac{\partial g}{\partial\theta^k}\frac{\partial g}{(\partial\theta^k)^T} \tag{A·7}$$

が存在する．ここで，

$$\frac{\partial g}{\partial\theta^k} = g'\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k} : x_t)}{\partial\theta^k} \tag{A·8}$$

$$\begin{aligned}\frac{\partial^2 g}{\partial\theta^k(\partial\theta^k)^T} &= g''\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k} : x_t)}{\partial\theta^k}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k} : x_t)}{(\partial\theta^k)^T} \\ &\quad + g'\sum_{t=1}^{n}\frac{\partial^2 L(d_{\theta^k} : x_t)}{\partial\theta^k(\partial\theta^k)^T}\end{aligned} \tag{A·9}$$

であるから，

$$\frac{\partial^2 K_{x^n}(\theta^k)}{\partial\theta^k(\partial\theta^k)^T} = \frac{g'}{g}\sum_{t=1}^{n}\frac{\partial^2 L(d_{\theta^k}:x_t)}{\partial\theta^k(\partial\theta^k)^T} + \frac{g''g-(g')^2}{g^2}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k}:x_t)}{\partial\theta^k}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k}:x_t)}{(\partial\theta^k)^T} \quad \text{(A·10)}$$

が成り立つ．

一方，条件 (c.1') は

$$\det\left\{\frac{\partial^2 K_{x^n}(\theta^k)}{\partial\theta^k(\partial\theta^k)^T}\right\} \to \infty, \quad a.s. \quad \text{(A·11)}$$

を意味する．そこで式 (A·10) から式 (A·11) の成り立つための条件を考える．

まず，条件 1(8) より，

$$\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k}:x_t)}{\partial\theta^k}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k}:x_t)}{\partial\theta^k} = O(n^2), \quad a.s. \quad \text{(A·12)}$$

$$\sum_{t=1}^{n}\frac{\partial^2 L(d_{\theta^k}:x_t)}{\partial\theta^k(\partial\theta^k)^T} = O(n), \quad a.s. \quad \text{(A·13)}$$

である．また，g, g', g'' の中身である $\sum_{t=1}^{n} L(d_{\theta^k}:x_t)$ を考えても同様の議論から

$$\sum_{t=1}^{n} L(d_{\theta^k}:x_t) = O(n), \quad a.s. \quad \text{(A·14)}$$

が成り立つ．式 (A·10) より，式 (A·11) の成り立つための条件は，

$$\frac{g'}{g}\sum_{t=1}^{n}\frac{\partial^2 L(d_{\theta^k}:x_t)}{\partial\theta^k(\partial\theta^k)^T} \to \infty, \quad a.s.$$

または

$$\frac{g''g-(g')^2}{g^2}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k}:x_t)}{\partial\theta^k}\sum_{t=1}^{n}\frac{\partial L(d_{\theta^k}:x_t)}{(\partial\theta^k)^T} \to \infty, \quad a.s.$$

が成り立たなければならない．したがって，式 (A·12)～(A·14) と，$g(y,z)$ が y に関して単調減少であることから，(c.1') が成り立つためには，$y\to\infty$ のとき

$$\left|\frac{g''(y,z)}{g(y,z)}y\right| = \left|\frac{y}{g(y,z)}\frac{\partial^2 g(y,z)}{(\partial y)^2}\right| \to \infty \quad \text{(A·15)}$$

$$\left|\frac{g'(y,z)}{g(y,z)}y\right| = \left|\frac{y}{g(y,z)}\frac{\partial g(y,z)}{\partial y}\right| \to \infty \quad \text{(A·16)}$$

であることが必要十分である．したがって，$y\to\infty$ のとき $\log g(y,z) \neq O\left(\log\frac{1}{y}\right)$ でなければならない．逆にもし $\log g(y,z) = O\left(\log\frac{1}{y}\right)$ であれば，(c.1') は成り立たないので漸近正規性も成り立たないことがわかる．この条件は，$g(y,z)$ が y に関して多項式オーダの減少関数では (c.1') は成り立たず，指数オーダ以上になると (c.1') が成り立つこと，すなわち任意の正定数 $\gamma>0$ に対して $g(y,z)\neq O\left(n^{-\gamma}\right)$ のとき (c.1') が成り立つことを示している．一方，このとき条件 1(5),(8) より，(c.2') も成り立つ．

ここで，条件 (c.1'),(c.2') のもとでは，

$$\lim_{n\to\infty} f_{x^n}(\tilde{\theta}^k)(\det\Sigma_{x^n})^{1/2} \leqq (2\pi)^{-k/2}, \quad a.s. \quad \text{(A·17)}$$

が成り立ち，更に (c.3') も同時に成り立てば，

$$\lim_{n\to\infty} f_{x^n}(\tilde{\theta}^k)(\det\Sigma_{x^n})^{1/2} \to (2\pi)^{-k/2}, \quad a.s. \quad \text{(A·18)}$$

が成り立つ [1]．

よって，最後に $\log g(y,z)\neq O\left(\log\frac{1}{y}\right)$，$(y\to\infty)$ と式 (A·17) から，(c.3') が成り立つことを示そう．そのために

$$\begin{aligned} &K_{x^n}(\theta^k) - K_{x^n}(\theta^{k*}) \\ &= \log\frac{f_{x^n}(\theta^k)}{f_{x^n}(\theta^{k*})} \\ &= \log\frac{g\left(\sum_{t=1}^{n} L(d_{\theta^k}:x_t), \pi(\theta^k)\right)}{g\left(\sum_{t=1}^{n} L(d_{\theta^{k*}}:x_t), \pi(\theta^{k*})\right)} \end{aligned} \quad \text{(A·19)}$$

を評価したい．

$$\frac{1}{\beta_{\theta^k}(x^n)}\log\frac{g\left(\sum_{t=1}^{n} L(d_{\theta^k}:x_t), \pi(\theta^k)\right)}{g\left(\sum_{t=1}^{n} L(d_{\theta^{k*}}:x_t), \pi(\theta^{k*})\right)} = \log\frac{g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^k}:x_t), \pi(\theta^k)\right)}{g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^{k*}}:x_t), \pi(\theta^{k*})\right)}, \quad a.s. \quad \text{(A·20)}$$

となるような x^n の関数 $\beta_{\theta^k}(x^n)$ を考える．条件 1(4),(8) から，$\forall\theta^k\in\Theta^k$ に対し一様に，

$$\begin{aligned} &g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^k}:x_t), \pi(\theta^k)\right) \\ &\to g\left(E^*[L(d_{\theta^k}:X)], \pi(\theta^k)\right), \quad a.s. \\ &= O(1), \quad a.s. \end{aligned} \quad \text{(A·21)}$$

$$\sum_{t=1}^{n} L(d_{\theta^k} : x_t) = O(n), \quad a.s. \tag{A·22}$$

が成り立つ．よって，$\beta_{\theta^k}(x^n)$ は漸近的にはほとんど確実に x^n に依存せず，n だけに依存する関数に置き換えることができる．したがって，$n \to \infty$ のとき

$$\frac{1}{\beta_{\theta^k}(n)} \log \frac{g\left(\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right)}{g\left(\sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t), \pi(\theta^{k*})\right)}$$
$$= \log \frac{g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right)}{g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t), \pi(\theta^{k*})\right)}, \quad a.s. \tag{A·23}$$

となるような $\beta_{\theta^k}(n)$ が存在する．一方，

$$\frac{1}{n}\left\{\sum_{t=1}^{n} L(d_{\theta^k} : x_t) - \sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t)\right\}$$
$$\to E^*[L(d_{\theta^k} : X)] - E^*[L(d_{\theta^{k*}} : X)], \quad a.s. \tag{A·24}$$

であり，$\forall\delta > 0$, $\forall\theta^k \notin B_\delta(\theta^{k*})$ に対し $E^*[L(d_{\theta^k} : X)] - E^*[L(d_{\theta^{k*}} : X)] > C$ となる正定数 $C > 0$ が存在するから，

$$\sum_{t=1}^{n} L(d_{\theta^k} : x_t) - \sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t) = O(n), \quad a.s.$$
$$\to \infty, \quad a.s. \tag{A·25}$$

が成り立つ．したがって，式 (A·22), (A·25) と $\log g(y, z) \neq O\left(\log \frac{1}{y}\right)$ $(y \to \infty)$ であることから，$\forall\delta > 0$, $\forall\theta^k \notin B_\delta(\theta^{k*})$ に対して

$$\log \frac{g\left(\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right)}{g\left(\sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t), \pi(\theta^{k*})\right)} \neq O\left(\log \frac{1}{n}\right)$$
$$\log \frac{g\left(\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right)}{g\left(\sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t), \pi(\theta^{k*})\right)} \to -\infty, \quad a.s. \tag{A·26}$$

となる(注6)．一方，式 (A·23) の右辺は，条件 1(8) より，θ^k に関して一様に

$$\log \frac{g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^k} : x_t), \pi(\theta^k)\right)}{g\left(\frac{1}{n}\sum_{t=1}^{n} L(d_{\theta^{k*}} : x_t), \pi(\theta^{k*})\right)}$$
$$\to \log \frac{g\left(E^*[L(d_{\theta^k} : X)], \pi(\theta^k)\right)}{g\left(E^*[L(d_{\theta^{k*}} : X)], \pi(\theta^{k*})\right)} < 0, \quad a.s.$$

と収束するから，$\forall\theta^k \notin B_\delta(\theta^{k*})$ に対して $\beta_{\theta^k}(n) \neq O(\log n)$, $\beta_{\theta^k}(n) \to \infty$ が成り立つ．これは，任意の $\gamma > 0$ に対して $\exp\{\beta_{\theta^k}(n)\} \neq O(n^\gamma)$ であることを意味する．同時に，$\forall\delta > 0$ に対して正定数 $\exists C_\delta > 0$ が存在し，$n \to \infty$ のとき $\forall\theta^k \notin B_\delta(\theta^{k*})$ に対して一様に

$$\frac{1}{\beta_{\theta^k}(n)} \log \frac{f_{x^n}(\theta^k)}{f_{x^n}(\theta^{k*})} < -C_\delta, \quad a.s. \tag{A·27}$$

が成り立つこともわかる．

$\beta(n)$ を $\beta(n) = \inf_{\theta^k \notin B_\delta(\theta^{k*})} \beta_{\theta^k}(n)$ のように定めると，$\forall\delta > 0$ に対し，$n \to \infty$ のとき，$\forall\theta^k \notin B_\delta(\theta^{k*})$ に対して一様に

$$\frac{f_{x^n}(\theta^k)}{f_{x^n}(\theta^{k*})} < \exp\left\{-\beta(n) C_\delta\right\}, \quad a.s. \tag{A·28}$$

が成り立つ．

一方，式 (A·17) から，$n \to \infty$ のとき，

$$f_{x^n}(\theta^{k*}) \leqq (2\pi)^{-k/2} \left(\det K''_{x^n}(\theta^k)\right)^{1/2}, \quad a.s. \tag{A·29}$$

が成り立っている．条件 1(8) と式 (A·10) より，

$$\det K''_{x^n}(\theta^k) = O\left(\max\left\{\left(\frac{ng'}{g}\right)^k, \left(\frac{ng''}{g}\right)^k\right\}\right)$$
$$a.s. \tag{A·30}$$

であるので，不等式 (A·28) から，$\forall\delta > 0$ に対し，$\theta^k \notin B_\delta(\theta^{k*})$ に対して一様に

$$f_{x^n}(\theta^k) < f_{x^n}(\theta^{k*}) \exp\{-\beta(n) C_{\delta_\epsilon}\}$$
$$= O\left(\frac{\max\left\{\left(\frac{ng'}{g}\right)^k, \left(\frac{ng''}{g}\right)^k\right\}}{\exp\{\beta(n) C_\delta\}}\right)$$
$$\to 0, \quad a.s. \tag{A·31}$$

が成り立つ．最後の収束は，$\forall\gamma > 0$ に対し $\exp\{\beta_{\theta^k}(n)\} = \Omega(n^\gamma)$ である(注7)ことと，式 (22) の $\frac{g'}{g} = O(1)$ から明らかである．

条件 1(1) より，Θ^k はコンパクトであるから，$\forall\delta > 0$ に対し

(注6)：n の関数 $y_1(n)$, $y_2(n)$ が，$y_1(n) = O(n)$, $y_2(n) = O(n)$, $y_1(n) - y_2(n) = O(n)$ をみたしているとき，$n \to \infty$ において $\frac{\log g(n,z)}{\log \frac{1}{n}} \to \infty$ であれば，$\frac{\log g(y_1(n),z) - \log g(y_2(n),z)}{\log \frac{1}{n}} \to \infty$ が成り立つため．

(注7)：$\exp\{\beta_{\theta^k}(n)\} \neq O(n^\gamma)$ より明らかに $\exp\{\beta_{\theta^k}(n)\} \neq o(n^\gamma)$ であるから，$\exp\{\beta_{\theta^k}(n)\} = \Omega\left(n^\gamma\right)$ である．

$$\int_{\theta^k \notin B_\delta(\theta^{k*})} f_{x^n}(\theta^k) d\theta^k \to 0, \quad a.s. \qquad (\text{A·32})$$

となり，これは

$$\int_{\theta^k \in B_\delta(\theta^{k*})} f_{x^n}(\theta^k) d\theta^k \to 1, \quad a.s. \qquad (\text{A·33})$$

であることを意味する．

$\tilde{\theta}^k \to \theta^{k*}$ $a.s.$ であることから，$0 < \forall\delta' < \forall\delta''$ であれば，$n \to \infty$ のとき，$B_{\delta'}(\theta^{k*}) \subset B_{\delta''}(\hat{\theta}^k)$，$a.s.$ したがって (c.3') が成り立ち，式 (33) が証明された．

式 (34) は，式 (33) と有界収束定理から直ちに導かれる． □

（平成 11 年 10 月 22 日受付，12 年 1 月 11 日再受付）

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 8〜11 同大理工学部・経営システム工学科助手．平 11 同大メディアネットワークセンター等・非常勤講師．現在，東大大学院・工学系研究科・環境海洋工学専攻・助手．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．平 12 より，ビジネスモデル，コストモデル等の研究にも着手．工博．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・理工学部・工業経営学科（現在経営システム工学科）助教授，現在教授．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．工博．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(Fellow)，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．