α-平行事前分布とその性質

竹内 純一
RWCP[1)] 理論 NEC 研究室 [2)]
〒216 神奈川県川崎市宮前区宮崎 4-1-1

甘利 俊一
理化学研究所 国際フロンティア研究システム
〒351-01 埼玉県和光市広沢 2-1

[1)]RWCP: Real World Computing Partnership (新情報処理開発機構)

[2)]NEC C&C 研究所内

あらまし　Jeffreys 事前分布は統計的推定問題において重要な役割を果たすことが知られている．本稿では, 情報幾何の立場から Jeffreys 事前分布を一般化し, α-平行事前分布というワンパラメータの事前分布の族を導入する．これは α-接続に関して平行な体積要素として定義される．さらに, 曲指数型分布族の推定問題において, α-平行事前分布を用いた場合の射影 Bayes 推定量 (Bayes 推定量をもとのクラスに射影して得られる推定量) と MDL 推定量の漸近的性質を論じる．α-平行事前分布は $\alpha = 0$ のときに Jeffreys 事前分布に一致するが, Jeffreys 事前分布が必ず存在するのに対し, ($\alpha = 0$ 以外の) α-平行事前分布は一般には存在する保証がない．そこで本稿では, 一般の α-平行事前分布が存在するための条件の考察も行う．

キーワード　ジェフリーズの事前分布, 曲指数型分布族, 平行体積要素, ベイズ推定量, 記述長最小原理

The α-parallel prior and its properties

Jun-ichi Takeuchi
Theory NEC Laboratory[3)], RWCP[4)]
4-1-1 Miyazaki, Miyamae-ku, Kawasaki,
Kanagawa 216, Japan

Shun-ichi Amari
Frontier Research Program
The Institute of Physical and Chemical Research
2-1 Hirosawa, Wako-shi, Saitama 351-01, Japan

[3)] c/o C&C Research Laboratories, NEC Corporation

[4)] RWCP: Real World Computing Partnership

Abstract　It is known that the Jeffreys prior plays an important role in statistical inference. In this paper, we generalize the Jeffreys prior from the point of view of information geometry and introduce a one parameter family of prior distributons, which we named the α-parallel priors. The α-parallel prior is defined as the parallel volume element with respect to the α-connection and coincides with the Jeffreys prior when $\alpha = 0$. Moreover, we analyze asymptotic behaviors of the projected Bayes estimator (the estimator obtained by projecting the Bayes estimator onto the original class of distributions) and the MDL estimator for curved exponential families. Although the Jeffreys prior always exists, it is not guaranteed that the α-parallel prior with $\alpha \neq 0$ exists. Hence we consider conditions for the existence of the α-parallel prior.

key words　Jeffreys prior, curved exponential family, parallel volume element, Baeys estimator, MDL principle

1　まえがき

Jeffreys 事前分布は, 統計的推定問題において重要な役割を果たすことが知られている [3]. 本稿では, 情報幾何の立場から Jeffreys 事前分布を一般化し, **α-平行事前分布** (あるいは簡単に α-事前分布と呼ぶ. 欧文では **α-parallel prior**, または α-prior と表記) という, ワンパラメータの事前分布の族を導入する. さらに, 曲指数型分布族の推定問題において, α-事前分布を用いた場合の射影 Bayes 推定量 (Bayes 推定量をもとのクラスに射影して得られる推定量 [6, 9]) と MDL 推定量の漸近的性質を論じる. α-事前分布は $\alpha = 0$ のときに Jeffreys 事前分布に一致するが, Jeffreys 事前分布が必ず存在するのに対し, ($\alpha = 0$ 以外の) α-事前分布は一般には存在する保証がない. そこで本稿では, 一般の α-事前分布が存在するための条件の考察も行う.

α-事前分布は, α-接続 [1] に関して平行な体積要素として導入する[1]. これは, [6] における結果を曲指数型分布族の場合に拡張しようとするときに自然に導入される. [6] において, Takeuchi は指数型分布族に関して MDL 推定量と射影 Bayes 推定量の性質を調べ, 以下の様な結果を得ている: (a) Jeffreys 事前分布を用いた射影 Bayes 推定量と, 期待値パラメータについて一様な事前分布を用いた MDL 推定量のいずれもが, 自然パラメータについてバイアス補正した最尤推定量とほぼ一致する. (b) 自然パラメータについて一様な事前分布を用いた射影 Bayes 推定量と, Jeffreys 事前分布を用いた MDL 推定量のいずれもが, 最尤推定量 (期待値パラメータについて不偏) とほぼ一致する.

以上の結果は, 射影 Bayes 推定量と MDL 推定量の, 精度 $1/N$ (N は標本数) の近似式求め, それらを最尤推定量, および自然パラメータに対してバイアス補正した最尤推定量と比較することで得られたものである.

以上の結果を曲指数型分布族の場合に拡張するためには, 以下の四つの拡張を行う必要がある. すなわち, (i) 射影 Bayes 推定量の近似式の拡張. (ii) MDL 推定量の近似式の拡張. (iii) '自然パラメータ (および期待値パラメータ) について一様な事前分布' の概念の一般化. (iv) 自然パラメータ (および期待値パラメータ) に関するバイアスの概念の一般化. ((iii) と (iv) は, 曲指数型分布族については自然パラメータや期待値パラメータが一般には存在しないために生じる問題である.)

このうち (i) については, Komaki が Takeuchi の結果と独立に求めた曲指数型分布族に関する Bayes 推定量の近似式 [4] を利用出来る. すなわち, Komaki の近似式から簡単な考察により射影 Bayes 推定量の近似式が得られる.

(ii) については, 比較的簡単に拡張を行うことが出来る.

(iii) については, 自然パラメータが 1-接続に関するアファイン座標であることに注意すると, 自然パラメータについて一様な事前分布とは, 1-接続に関する平行な体積要素であることが理解出来る. ところが, 1-接続は曲指数型分布族にも定義されるので, 1-接続に関する平行な体積要素も曲指数型分布族に対して定義出来る. これを自然パラメータについて一様な事前分布の代わりに用いるのである. 同様に, 期待値パラメータについて一様な事前分布と Jeffreys 事前分布は, それぞれ, -1-接続と 0-接続に関する平行な体積要素と等価である.

以上のことに注意すると, 一般の α-接続に関する平行な体積要素として, α-平行事前分布を自然に導入することが出来る. ここで, Jeffreys 事前分布 ($\alpha = 0$) が必ず存在するのに対し, $\alpha \neq 0$ の場合には, α-事前分布は一般には存在する保証がないことを注意しておかなければならない. 本稿では, α-事前分布が存在するための十分条件を示す.

α-事前分布が存在する場合, $\alpha = \pm 1, 0$ 以外の場合も含めて, α-事前分布を用いた射影 Bayes 推定量と MDL 推定量の性質を調べることは興味深いと思われる. それゆえ, (iv) に関連しては, 自然パラメータや期待値パラメータに関してバイアス補正した最尤推定量の概念を, 一般の α-アファイン座標[2] に関してバイアス補正した最尤推定量の概念に拡張する. この目的のため, 曲指数型分布族を α-平坦な多様体に埋め込んで考える. (一般には可能でないことに注意.)

(i)-(iv) の拡張の結果, 次のことが明らかになる. すなわち, $(-(1+\alpha)/2)$-事前分布を用いた射影 Bayes 推定量と, $((1-\alpha)/2)$-事前分布を用いた MDL 推定量のいずれもが, α-アファイン座標に関して (拡張した意味で) バイアス補正した最尤推定量にほぼ等価である. これは, [6] における結果の一般化となっている.

2　準備

ν を $\Re^z$ の Borel 集合族の上の σ-有限な測度, $\mathcal{X}$ をその台とする. また, Θ を $\Re^n$ の凸集合とする. S で, 経数 $\theta \in \Theta$ が付いた, $\mathcal{X}$ 上の測度 ν に関する確率密度の集合を表す. すなわち, $S \equiv \{p(\cdot|\theta) | \theta \in \Theta\}$ とする. ただし, $\forall \theta \in \Theta, \int_{x \in \mathcal{X}} p(x|\theta)\nu(dx) = 1$ と $\forall \theta \in \Theta, \forall x \in \mathcal{X}, p(x|\theta) \geq 0$ が成り立つものとする. ここでさらに, $\forall \theta \in \Theta, \forall x \in \mathcal{X}, p(x|\theta) > 0$ であり, かつ $p(x|\theta)$ は Θ 上で 2 回微分可能と仮定する. また $\mathcal{P}(\mathcal{X})$ で, $\mathcal{X}$ 上の測度 ν に関する確率密度全体の集合を表す.

S を多様体とみなし, $T_p(S)$ で $p \in S$ における S の接空間を表す. $T_p(S)$ は, 微分演算子 $\partial_i \equiv \partial/\partial\theta^i$ $(i = 1, ..., n)$ で張られる線形空間である (一般に $i, j, k...$ を θ の添字として用いる). 次に, Amari[1] に従って Fisher 計量と α-接続を定義しよう. g で Fisher 計量を表し, $\langle X, Y \rangle$ で, g で定義される $X, Y \in T_p(S)$ の内積を表す. g_{ij} で, g の θ に関する成分 $\langle \partial_i, \partial_j \rangle$ を表すことにすると, g は

$$\begin{aligned} g_{ij} &\equiv E(\partial_i \ln p(x|\theta) \cdot \partial_j \ln p(x|\theta)) \\ &\equiv \int_{x \in \mathcal{X}} \partial_i \ln p(x|\theta) \cdot \partial_j \ln p(x|\theta) \cdot p(x|\theta)\nu(dx) \end{aligned}$$

[1] Riemann 多様体に接続が与えられているとき, その接続に関して平行な体積要素という概念は, **アファイン微分幾何学** [10] における重要な概念である. アファイン微分幾何学では, 情報幾何と同様に, 必ずしも計量的でない接続を扱う.

[2] 本稿では, α-接続に関するアファイン座標を 'α-アファイン座標' と呼ぶことにする.

で定義される．g_{ij} は θ に関する Fisher 情報行列に一致する．g^{ij} で g_{ij} の逆行列を表す．次に $\nabla^{(\alpha)}$ で α-接続 (すなわち α-共変微分) を表す．$\Gamma^{(\alpha)}_{ijk}$ で $\nabla^{(\alpha)}$ の θ に関する成分を表すことにすると，$\nabla^{(\alpha)}$ は $\Gamma^{(\alpha)}_{ijk} \equiv \Gamma^{(0)}_{ijk} - \alpha \cdot T_{ijk}/2$ で定義される．ただし，$\Gamma^{(0)}$ は g に関する Riemann 接続を表し，T は $T_{ijk} = E(\partial_i \ln p(x|\theta) \cdot \partial_j \ln p(x|\theta) \cdot \partial_k \ln p(x|\theta))$ で定義されるテンソルである．以下，Fisher 計量と α-接続が備わった S を統計的モデルと呼ぶ．ここで，$\int p(x|\theta)\nu(dx) = 1$ という制限を $\int p(x|\theta)\nu(dx) < \infty$ で置き換えても，同様に Fisher 計量と α-接続を定義出来る．そのような多様体を拡張統計的モデルと呼ぶ．統計的モデル S に対して $\bar{S} = \{q(\cdot|\theta,\tau)|\ q(x|\theta,\tau) = \tau \cdot p(x|\theta),\ \tau > 0,\ p \in S\}$ で多様体 $\bar{S}$ を定義する．これに Fisher 計量と α-接続を定義したものは拡張統計的モデルの例であり，S の拡張と呼ぶ．

n 次元の (拡張) 統計的モデル S に対して，α-接続に関する **Riemann-Christoffel 曲率テンソル** (以下では簡単のため α-Reimann 曲率テンソルと呼ぶ) を定義しよう．その成分を $R^{(\alpha)l}_{ijk}$ と書くと，

$$R^{(\alpha)l}_{ijk} \equiv \partial_i \Gamma^{(\alpha)l}_{jk} - \partial_j \Gamma^{(\alpha)l}_{ik} + \Gamma^{(\alpha)l}_{im}\Gamma^{(\alpha)m}_{jk} - \Gamma^{(\alpha)l}_{jm}\Gamma^{(\alpha)m}_{ik}$$

で定義される．ただし $\Gamma^{(\alpha)m}_{ik} = \Gamma^{(\alpha)}_{ikj} g^{jm} \equiv \sum_{j=1}^{n} \Gamma^{(\alpha)}_{ikj} g^{jm}$ である (本稿では，いわゆる和の規約を用いる.)．α-アファイン座標が存在する (拡張) 統計的モデルを α-平坦であるというが，α-平坦ならば α-Riemann 曲率は 0 になる．

次に指数型分布族を定義しておく．確率密度関数が $p(x|\theta) = \exp(\theta^i x_i - \psi(\theta))$ と書けるとき ($z = n$ と仮定．また x_i は x の第 i 成分.)，S を指数型分布族と呼ぶ．この場合，θ を自然パラメータまたは θ-座標と呼ぶ．θ は 1-アファイン座標である．また $\eta = E(x)$ で定義される座標を，期待値パラメータまたは η-座標と呼ぶ．これは -1-アファイン座標である．すなわち，指数型分布族は ± 1-平坦である．η の第 i 成分は η_i と書く．$\partial^i \equiv \partial/\partial\eta_i$ という記法を用いる．このとき $\partial^i = g^{ij}\partial_j$ となる．また，$\eta_i = \partial_i\psi(\theta)$, $g_{ij} = \partial_i\partial_j\psi(\theta)$, $T_{ijk} = \partial_i\partial_j\partial_k\psi(\theta)$ が成り立つ．

M を，S の m 次元の部分多様体とする ($m < n$)．(特に S が指数型分布族の場合 M は曲指数型分布族と呼ばれる.) 厳密には M は $M = \{p(\cdot|\theta(u))|u \in U\}$ で定義される．ただし U は $\Re^m$ の凸集合であり，$\theta(u)$ は $u \mapsto \theta(u) \in \Theta$ なる C^∞ 級の関数である．さらに $\theta(u)$ のヤコビアンのランクは全ての $u \in U$ について m に等しいものとする．以下，u^a で u の a 番目の成分を表す．一般に $a, b, c, ...$ を u の添字として用い，$a, b, c, ...$ の値域は $\{1, 2, ..., m\}$ であるとする．$\tilde{g}$ と $\tilde{\nabla}^{(\alpha)}$ はそれぞれ M の Fisher 計量と α-接続を表すものとする．また，例えば 2 階のテンソル K について，その u に関する成分は K_{ab} 等と書く．以下，他の座標系が定義されたときもこうしたルールに従う．

次に付随部分多様体を定義しよう．M の各点 $p(\cdot|u) \in M$ に対し，M に直交するような $n - m$ 次元の滑らかな S の部分多様体 $A(u)$ を付随させる．さらに各 $A(u)$ において座標系 v を定義する．ただし，$v = 0$ が $A(u)$ と M の交わりを表すものとする．$\kappa, \lambda, ...$ を v の添字として用い，それらの値域を $\{m+1, ..., n\}$ であるとする．ベクトル $\partial_\kappa = \partial/\partial v^\kappa$ は u と v について適当な回数だけ微分可能とする．さらに，族 $A = \{A(u)|u \in U\}$ は (少なくとも M の近傍において) S を滑らかに埋め尽くすものと仮定する．言い替えると，S の点を v と u によって指定できると仮定する．すなわち，$w = (u, v)$ を S の座標として用いることが出来る．以下，$\alpha, \beta, ...$ を w の添字として用い，それらの値域を $\{1, ..., n\}$ とする ($w^1 = u^1$,$w^2 = u^2$,..., $w^m = u^m$, $w^{m+1} = v^{m+1}$,..., $w^n = v^n$ とする)．

次に，M の S における埋め込み曲率テンソルを定義する．$\overset{(\alpha)}{H}$ で (X, Y) $(X, Y \in T_p(M))$ を $\nabla^{(\alpha)}_X Y - \tilde{\nabla}^{(\alpha)}_X Y \in T_p(A(p))$ に写す関数を表す．$\overset{(\alpha)}{H}$ は M 上の横断的ベクトル場であり，$\overset{(\alpha)}{H}(X, Y)$ は M に直交する．これを α-接続に関する埋め込み曲率テンソルと呼ぶ．以下，簡単のため α-曲率テンソルと呼ぶ．$\overset{(\alpha)}{H}$ の成分は $\overset{(\alpha)}{H}_{ab\kappa} \equiv \langle \overset{(\alpha)}{H}(\partial_a, \partial_b), \partial_\kappa \rangle$ のように書ける．M 上いたるところで $\overset{(\alpha)}{H} = 0$ となるとき，M は S において α-自己平行と云われる．本稿では，付随部分多様体 $A(p)$ は -1-自己平行であると仮定する．

次に標本に関する記法を述べる．x^N で，$\mathcal{X}^N$ の要素を表す．$x^N = x_1 ... x_N$ は $p(x^N|\theta) \equiv \prod_{t=1}^{N} p(x_t|\theta)$ に従うものとする (すなわち i.i.d. また，ここでの x_t は，x の第 t 成分を表すのではないことに注意.)．さらに，$\bar{x} \equiv \sum_{t=1}^{N} x_t/N$ なる記法を用いる．$\bar{x}$ は十分統計量である．また，x^∞ は無限列をさすものとする．

一般に f で推定量 (標本の集合から $\mathcal{P}(\mathcal{X})$ への関数) を表す．ここで，$f[x^N]$ で f による x^N の像を表し，$f[x^N](x)$ で $f[x^N]$ の x における密度を表す．また $\hat{\theta}$ や $\hat{u}$, $\hat{g}_{ij}$ 等は，それらの，x^N のもとでの最尤推定値を表すものとする．指数型分布族については，$\hat{\eta} = \bar{x}$ が成り立つ (厳密にはある仮定がいる)．

3 射影ベイズ推定量の近似式

この節では，曲指数型分布族における Bayes 推定量の，Komaki による近似式 [4] を復習し，射影 Bayes 推定量の近似式を導く．

M を，n 次元の指数型分布族 S に埋め込まれた m 次元の曲指数型分布族とする．$\rho(u)du$ を M の経数 u の定義域 U 上の事前分布とする．ただし ρ は C^∞ 級とする．Bayes 推定量 f_ρ は $f_\rho[x^N] \equiv \int p(\cdot|u)\rho(u|x^N)du$ で定義される (ただし，$\rho(u|x^N)$ は x^N を与えられたもとでの u の事後分布.)．また，射影 Bayes 推定量 $\tilde{f}_\rho$ は $\tilde{f}_\rho[x^N] \equiv \arg\min_{p \in M} D(f_\rho||p)$ で定める．ただし $D(p||q)$ は q の p に対する Kullbuck Leibler 情報量を表す．

以下，この節では x^∞ を一つ固定し，x^N はその先頭の N 個からなる列とする．Komaki は以下の定理を示した．

定理 1 (Komaki '94) $f_\rho[x^N](x)$ は次のように展開される．

$$f_\rho[x^N](x) \quad = \quad p(x|\hat{u}) + \frac{\hat{g}^{ab}(\partial_a\partial_b p(x|u) - \hat{\Gamma}^{(-1)c}_{ab}\partial_c p(x|u))}{2N}$$

$$+\quad \frac{(\partial_a \ln\rho(\hat{u}) - \hat{\Gamma}^{(1)b}_{ab})\hat{g}^{ac}\partial_c p(x|\hat{u})}{N} + o(\frac{1}{N}).$$

ここで, $(\partial_a \ln\rho(\hat{u}) - \hat{\Gamma}^{(1)b}_{ab})\hat{g}^{ac}\partial_c p(x|\hat{u})/N$ なる項は, Bayes 推定値が, 最尤推定値に対して M に沿った方向にずれていることを表し, $\hat{g}^{ab}(\partial_a\partial_b p(x|u) - \hat{\Gamma}^{(-1)c}_{ab}\partial_c p(x|u))/2N$ なる項は, 同様に M に直交する方向[3] にずれていることを表す. これらのずれは, いずれも -1-測地線 (-1-接続に関する測地線) に沿った[4] ものである. Kullback Leibler 情報量による射影の足は, 最尤推定値から, M に垂直に -1-測地線を降ろした足に等しいことに注意すると, 次式が成り立つことが分かる.

$$u^c(\tilde{f}_\rho[x^N]) = \hat{u}^c + \frac{(\partial_a \ln\rho(\hat{u}) - \hat{\Gamma}^{(1)b}_{ab})\hat{g}^{ac}}{N} + o(\frac{1}{N}). \quad (1)$$

4 MDL 推定量の近似式の一般化

曲指数型分布族 M に関する MDL 推定量の近似式を与える. 事前分布 $\rho(u)du$ に関する MDL 推定量を f^ρ_{mdl} と書き, 次式で定義する [6].

$$u(f^\rho_{mdl}[x^N]) = \arg\min_u\Big(-\ln p(x^N|u) - \ln\frac{\rho(u)}{(\det|g_{ab}|)^{1/2}}\Big).$$

以下, 前節と同様に x^∞ を一つ固定し, x^N をその先頭の N 個からなる列とする. 今, $\bar{\rho}((u,0)) = \rho(u)$ かつ $\bar{\rho}(w) > 0$ を満たすような, w の滑らかな実数値関数 $\bar{\rho}$ を定義する. また, MDL 推定値に対応する θ 座標の値を $\tilde{\theta}$ と書き, M に対応する θ の値域を Θ_M と書く. すると, 次式のようになる.

$$\tilde{\theta} = \arg\min_{\theta\in\Theta_M}\Big(-\ln\frac{p(x^N|\theta)\bar{\rho}(w(\theta))}{(\det|g_{ab}|)^{1/2}}\Big).$$

最小化すべき関数を $G(\theta)$ とおき, v^κ を一定に保ったまま $G(\theta)$ を最小化する. すなわち未定係数 a_κ をおき, $G(\theta) + a_\kappa v^\kappa N$ の偏微分を 0 とおく. $V(\theta) = \ln(\bar{\rho}(w(\theta))/(\det|g_{ab}|)^{1/2})$ とおくと, $G(\theta) + a_\kappa v^\kappa N = -\ln p(x^N|\theta) - V(\theta) + a_\kappa v^\kappa N$ となる. これを θ^i で偏微分すると $(\bar{x}_i - \eta_i) + \partial_i V(\theta)/N = a_\kappa \partial_i v^\kappa$ となる. これに $\partial u^a/\partial\eta_i$ をかけて i で和をとると,

$$\frac{\partial u^a}{\partial\eta_i}(\bar{x}_i - \eta_i) + \frac{\partial u^a}{\partial\eta_i}\frac{\partial w^\alpha}{\partial\theta^i}\frac{\partial_\alpha V(\theta)}{N} = a_\kappa\frac{\partial v^\kappa}{\partial\theta^i}\frac{\partial u^a}{\partial\eta_i}$$

となる. ここで

$$\frac{\partial w^\alpha}{\partial\theta^i}\frac{\partial w^\beta}{\partial\eta_i} = \frac{\partial w^\alpha}{\partial\theta^i}\frac{\partial w^\beta}{\partial\theta^j}\frac{\partial\theta^j}{\partial\eta_i} = \frac{\partial w^\alpha}{\partial\theta^i}\frac{\partial w^\beta}{\partial\theta^j}g^{ij} = g^{\alpha\beta}$$

に注意すると $\partial^i u^a(\bar{x}_i - \eta_i) + g^{a\alpha}\partial_\alpha V(\theta)/N = a_\kappa g^{a\kappa}$ となる. $g^{a\kappa} = 0$ により $\partial^i u^a(\bar{x}_i - \eta_i) + g^{ab}\partial_b V/N = 0$ となる. いま, $(\hat{u}, 0)$ に対応する η を $\bar{\eta}$ とおく. これは $\bar{x}$ から M におろした -1-射影の足に対応する η である. $|\bar{\eta} - \bar{x}| = o(1)$ という仮定をおく. 上記の式を u について $\hat{u}$ の周りで Taylor 展開すると

$$\begin{aligned}\big(\partial^i u^a(\hat{u}) + O(|u-\hat{u}|)\big)(\bar{x}_i - \bar{\eta}_i + \bar{\eta}_i - \eta_i)&\\ +\hat{g}^{ab}\partial_b\hat{V} + O(|u-\hat{u}|)/N &= 0\end{aligned}$$

となる. ここで $\partial^i u^a(\hat{u})(\bar{x}_i - \bar{\eta}_i) = 0$ であることと, 仮定 $|\bar{\eta} - \bar{x}| = o(1)$ を用いると

$$\partial^i u^a(\hat{u})(\bar{\eta}_i - \eta_i) + \frac{\hat{g}^{ab}\partial_b\hat{V} + O(|u-\hat{u}|)}{N} = o(|u-\hat{u}|)$$

となる. さらに $u^a - \hat{u}^a = \partial^i u^a(\hat{u})(\eta_i - \bar{\eta}_i) + O(|\eta - \bar{\eta}|^2)$ に注意すると $u^a - \hat{u}^a = \hat{g}^{ab}\partial_b\hat{V}/N + O(|u-\hat{u}|)/N + o(|u-\hat{u}|)$ となる. これより $\tilde{u}^a = \hat{u}^a + \hat{g}^{ab}\partial_b\hat{V}/N + o(1/N)$ を得る. $V = \ln\rho(u) - \ln(\det|g_{ab}|)^{1/2}$ であるから, $\partial_b V = \partial_b\ln\rho(u) - \Gamma^{(0)}_{bcd}g^{cd}$ となる. すなわち, 次の補題を得る.

補題 1 *曲指数型分布族 M における MDL 推定量について, $|\bar{x} - \bar{\eta}| = o(1)$ ならば, 次の式が成り立つ.*

$$u^a(f^\rho_{mdl}[x^N]) = \hat{u}^a + \frac{\hat{g}^{ab}(\partial_b\ln\rho(\hat{u}) - \hat{\Gamma}^{(0)}_{bcd}\hat{g}^{cd})}{N} + o(\frac{1}{N}).$$

5 α-平行事前分布

まず, S を n 次元の指数型分布族として, S 上の, θ について一様な事前分布をテンソル場として書き直す. 体積要素 (あるいは面積要素) とは, S 上の n 階交代テンソル場 (n 次微分形式) として定義されることに注意しよう ([11] 等参照). すなわち, このテンソル場は, S の各点において, n 個の接ベクトルの組 $(X_1, ..., X_n)$ に対し, それらが張る平行四辺形の体積を定義するようなテンソル場である. そのようなテンソル場のひとつを ϵ とおくと

$$\begin{aligned}&\epsilon(X_1, ..., X_n)\\ =\ & K\cdot\sum_\sigma \mathrm{sgn}(\sigma)d\theta^1(X_{\sigma(1)})\cdot ...\cdot d\theta^n(X_{\sigma(n)}) \quad (2)\end{aligned}$$

と書ける. ただし, K はスカラー密度であり, $d\theta^i$ は $d\theta^i(X) = \langle\partial^i, X\rangle$ で定まるテンソルとする. また, σ は $1, ..., n$ の置換, $\mathrm{sgn}(\sigma)$ は σ の符号を表し, 総和は全ての置換に渡ってとる. これを, これまで用いて来た測度の記法で表そう. $d\theta^i$ を, (テンソルではなく) θ^i の微小変化と考えたとき, $\rho(\theta)d\theta \equiv \rho(\theta)d\theta^1\cdot ...\cdot d\theta^n$ が点 θ の近傍の, 各辺が $d\theta^i\cdot(\partial/\partial\theta^i)$ である立方体の体積に対応することに注意する. すると $\rho(\theta)d\theta \equiv \epsilon(\partial/\partial\theta^1, ..., \partial/\partial\theta^n)d\theta$ とすればよいことが分かる. さらにこの右辺は $Kd\theta$ に等しいことに注意する. さて, θ に関する一様な測度とは, '$\rho(\theta) =$ 定数' の場合であるから, (2) において, K を定数とおいたものに相当する. 以下では, ϵ で $K \equiv 1$ とおいたものだけを表す.

今, θ が $\nabla^{(1)}$ に関するアファイン座標なので $\nabla^{(1)}d\theta^i = 0$ である. よって, $\nabla^{(1)}\epsilon = 0$ となることが分かる. また, 逆に $\nabla^{(1)}\epsilon = 0$ ならば, K は定数でなければならない. 以下ではこの事実に注目する.

[3] M の $\mathcal{P}(\mathcal{X})$ における -1-曲率の方向. 詳しくは [2, 4] を参照.

[4] M に直交したずれを $p_o(x)$, M に沿ったずれを $p_n(x)$ と書くと, $p(x|t) = p(x|\hat{u}) + t_1 p_n(x) + t_2 p_o(x) + o(1/N)$ は混合型分布族をなし ($f_\rho[x^N](x) = p(x|(1,1))$), t_1, t_2 はその η 座標である. また, このとき $\partial/\partial t_2$ は各 $\partial/\partial u^a$ に直交している.

一般に, 多様体 S のアファイン接続 ∇ に対して, S 上いたるところ $\nabla\kappa=0$ となるような体積要素 κ が存在するとき, ∇ を等積接続 (equiaffine connection) という (または等積的という). また, そのような κ を ∇ について平行な体積要素 (parallel volume element) という[5]. すなわち, 指数型分布族において, 1-接続は等積的であり, θ について一様な体積要素 ϵ は, $\nabla^{(1)}$ について平行な体積要素として特徴付けられる. こうすると, θ について一様な測度という概念は, 指数型分布族でない場合でも自然に拡張出来る. すなわち我々は, 指数型分布族以外の統計的モデル (特に曲指数型分布族) M についても, $\tilde{\nabla}^{(1)}$ について平行な体積要素を θ について一様な事前分布の代わりに用いることにしよう. そしてそれを, 1-平行事前分布と呼ぶ. 同様に $\tilde{\nabla}^{(-1)}$ について平行な体積要素を, η について一様な事前分布の代わりに用いる. さらに, 一般に α-接続について平行な体積要素を **α-平行事前分布**と名付ける.

ここで導入した α-事前分布は, $\tilde{\nabla}^{(\alpha)}$ が等積接続でなければ存在しない. そこで次に, $\tilde{\nabla}^{(\alpha)}$ が (局所的に) 等積的になるための条件を考察する. 一般の接続に関する Riemann 曲率の成分を R^d_{abc} と書く. このとき, $R_{bc}\equiv R^a_{abc}$ で定まるテンソルは Ricci テンソルと呼ばれる. 特に本稿では, 接続が α-接続の場合の Riici テンソルを α-Ricci テンソルと呼び, $R^{(\alpha)}_{ij}$ と書く.

次の補題が成り立つ ([10],p.28, 命題 3.1 参照).

補題 2 捩れの無いアファイン接続 ∇ を持った多様体 S について, *(a)* ∇ が局所的に等積接続になること, *(b)* S 上で $R^k_{ijk}=0$ となること, および *(c)* S 上で $R_{ij}=R_{ji}$ となることは同値である.

例えば Riemann 接続 (0-接続) の場合は Ricci テンソルは必ず対称になるから, 0-平行事前分布は必ず存在し, 自然な体積要素 $(\det|g_{ij}|)^{1/2}d\theta$ を導く. これは Jeffreys 事前分布にほかならない.

補題 2 を α-接続に適用しよう. α-Riemann 曲率の定義から, $R^{(\alpha)k}_{ijk}=-\alpha(\partial_i T^k_{jk}-\partial_j T^k_{ik})/2$ となる. よって, 次の命題を得る.

命題 1 *(*拡張*)* 統計的モデル S について, $\partial_i T^k_{jk}-\partial_j T^k_{ik}=0$ ならば, 任意の $\alpha\in\Re$ について α-平行事前分布が存在する. そうでなければ, 0-平行事前分布のみが存在する.

これから, 次のような概念を定義することが出来る.

定義 1 (統計的等積) *(*拡張*)* 統計的モデル S について, S 上の各点で $\partial_i T^k_{jk}-\partial_j T^k_{ik}=0$ であるとき, S は *(*局所的に*)* 統計的等積 *(statistically equiaffine)* であるという.

すなわち, S が統計的等積であることは, 任意の α について α-平行事前分布が存在すること, および任意の α について $R^{(\alpha)k}_{ijk}=0$ であることと同値である.

次に統計的等積であるための十分条件を考察しよう. Lauritzen によって導入された **conjugate symmetry** という概念 [5] が有用である. (拡張) 統計的モデル S について $\forall\alpha\in\Re$, $R^{(\alpha)}_{ijkl}=-R^{(\alpha)}_{ijlk}$ であるとき, S は conjugate symmetric であるという. ただし, $R^{(\alpha)}_{ijkl}=R^{(\alpha)m}_{ijk}g_{lm}$ である. この概念を用いると, 補題 2 から次の命題を得る.

命題 2 S が *conjugate symmetric* であれば, S は統計的等積である.

Lauritzen はまた, S が conjugate symmetic になるための十分条件を与えた. すなわち, 次の命題が成り立つ.

命題 3 (Lauritzen) ある $\alpha_0\neq 0$ について α_0-平坦な *(*拡張*)* 統計的モデルは *conjugate symmetric* である.

次に, 一般に次のように書けることに注意する [7, 8].

$$R^{(\alpha)}_{abcd}=B^i_aB^j_bB^k_cB^l_dR^{(\alpha)}_{ijkl}+g^{\kappa\lambda}(H^{(-\alpha)}_{ad\kappa}H^{(\alpha)}_{bc\lambda}-H^{(\alpha)}_{ac\kappa}H^{(-\alpha)}_{bd\lambda}).$$

ただし $B^i_a=\partial_a\theta^i$ とおいた. ここで, 命題 1 に注意すると次を得る.

命題 4 ある $\alpha_0\neq 0$ について $R^{(\alpha_0)}_{ijkl}=-R^{(\alpha_0)}_{ijlk}$ を満たす *(*拡張*)* 統計的モデル S の中の, α_0-自己平行な部分多様体 M は統計的等積である. 特に *conjugate symmetric* な *(*拡張*)* 統計的モデル S の中の, α_0-自己平行な部分多様体 M $(\alpha_0\neq 0)$ は統計的等積である.

以上, 統計的モデル M について, α-接続が等積的になるための条件がある程度明らかになった. 次に, 統計的等積な曲指数型分布族 M について, α-平行事前分布の場合の射影 Bayes 推定量と MDL 推定量の近似式を求める. 以下, α-平行事前分布を用いた射影 Bayes 推定量を $\tilde{f}^{(\alpha)}$ と書く. また, α-平行事前分布を用いた MDL 推定量を $f^{(\alpha)}_{mdl}$ と書く.

$\tilde{\epsilon}^{(\alpha)}$ を α-平行体積要素とする. $\tilde{\nabla}^{(\alpha)}\tilde{\epsilon}^{(\alpha)}=0$ に注意すると, 次式を示すことが出来る.

$$\begin{aligned}&\tilde{\nabla}^{(\alpha)}_{\partial_a}\big(\tilde{\epsilon}^{(\alpha)}(\partial/\partial u^1,...,\partial/\partial u^m)\big)\\ =\;&\Gamma^{(\alpha)b}_{ab}\tilde{\epsilon}^{(\alpha)}(\partial/\partial u^1,...,\partial/\partial u^m).\end{aligned}\tag{3}$$

次に, α-事前分布の測度としての表現を $\rho^{(\alpha)}(u)du$ と書く. すなわち, $\rho^{(\alpha)}(u)=\tilde{\epsilon}^{(\alpha)}(\partial/\partial u^1,...,\partial/\partial u^m)$ とおく. すると (3) より, $\partial_a\ln\rho^{(\alpha)}(u)=\Gamma^{(\alpha)b}_{ab}$ を得る. これと, (1) および $\Gamma^{(\alpha)}_{abc}=\Gamma^{(0)}_{abc}-\alpha T_{abc}/2$ より

$$u^c(\tilde{f}^{(\alpha)}[x^N])=\hat{u}^c+\frac{(1-\alpha)\hat{T}_{abd}\hat{g}^{bd}\hat{g}^{ac}}{2N}+o(1/N)\tag{4}$$

である. 同様に, 補題 1 により次式を得る.

$$u^c(f^{(\alpha)}_{mdl}[x^N])=\hat{u}^c-\frac{\alpha\hat{T}_{abd}\hat{g}^{bd}\hat{g}^{ac}}{2N}+o(1/N).\tag{5}$$

6 バイアス補正最尤推定量の一般化

曲指数型分布族 M が, α-平坦な多様体 L の部分空間であると仮定する. そして ξ を, L の α-アファイン座標である

[5] これらはアファイン微分幾何における概念である. 詳しくは [10] 参照.

とする[6]．また, ξ の添字には i' 等を用い, $\partial_{i'} = \partial/\partial\xi^{i'}$ とおく．ただし, i' 等の値域は $\{1, ..., \dim(L)\}$ とする．また, M を指数型分布族 S の部分空間と考えたときに, M の各点に $A(u)$ という M に直交する S の部分空間を考えたが, それと同様に, M を L の部分空間とみなし, M の各点に u において, M に直交する L の $\dim(L) - m$ 次元の部分空間 $B(u)$ を考える．そして $B(u)$ に座標 z をいれる．z の添字には κ' 等を用いる．また, (u, z) は L の座標とみなせるが, これを ζ と書く．ζ の添字には α' 等を用いる．

通常, 曲指数型分布族の推定問題についてバイアス補正を考えるときは, パラメータ u のバイアスを考えるが, ここでは次のように考える．$\hat{u}$ で指定される M の点を $p_{\hat{u}}$ とする．$p_{\hat{u}}$ を L の点とみなしたときに対応する ξ の値を $\xi(\hat{u})$ と書く．そして, $\xi(\hat{u})$ のバイアスの M の接空間方向への (Fisher 計量で分解した) 成分が $o(1/N)$ になるように補正したものを, ξ についてバイアス補正した最尤推定量と呼ぶ．以下, それを $\tilde{u}$ と書く．また, 以下では $\hat{\xi}$ は $\xi(\hat{u})$ を表す．

$\hat{\xi}$ を $\hat{u}$ の関数とみて, u の周りで Taylor 展開する．

$$\begin{aligned}\hat{\xi}^{i'} - \xi^{i'} &= \partial_a \xi^{i'} \delta u^a + \frac{\partial_b \partial_a \xi^{i'} \delta u^a \delta u^b}{2} + O(|\delta u|^3) \\ &= F_a^{i'} \delta u^a + \frac{\partial_b F_a^{i'} \delta u^a \delta u^b}{2} + O(|\delta u|^3)\end{aligned}$$

となる．ただし, $\delta u = \hat{u} - u$, $F_a^{i'} = \partial \xi^{i'}/\partial u^a$ とおいた．両辺の期待値をとると

$$E(\hat{\xi}^{i'} - \xi^{i'}) = F_a^{i'} E(\delta u^a) + (\partial_b F_a^{i'}) g^{ab}/2N + O(N^{-3/2})$$

を得る．$E(\delta u^a)$ は, 最尤推定量のバイアスで, $-\Gamma_{\alpha\beta}^{(-1)a} g^{\alpha\beta}/2N + O(N^{-3/2})$ となることが知られている．すなわち

$$E(\hat{\xi}^{i'} - \xi^{i'}) = -\frac{F_a^{i'} \Gamma_{\alpha\beta}^{(-1)a} g^{\alpha\beta} - (\partial_b F_a^{i'}) g^{ab}}{2N} + O(N^{-3/2})$$

が成り立つ．右辺の第一項の, ∂_c 方向の成分は

$$\begin{aligned} & -F_{i'}^c (F_a^{i'} \Gamma_{\alpha\beta}^{(-1)a} g^{\alpha\beta} - \partial_b F_a^{i'} g^{ab})/2N \\ = & -F_{i'}^c F_a^{i'} \Gamma_{\alpha\beta}^{(-1)a} g^{\alpha\beta} - F_{i'}^c (\partial_b F_a^{i'}) g^{ab}/2N \\ = & -\Gamma_{\alpha\beta}^{(-1)c} g^{\alpha\beta} - F_{i'}^c (\partial_b F_a^{i'}) g^{ab}/2N\end{aligned}$$

となる．ここで, $F_{i'}^{\gamma'} = \partial\zeta^{\gamma'}/\partial\xi^{i'}$ とおいた ($\gamma' \le m$ ならば $F_{i'}^{\gamma'} = \partial u^{\gamma'}/\partial\xi^{i'}$)．ところが

$$\begin{aligned}\Gamma_{bc\delta'}^{(\alpha)} &= \langle \nabla_{\partial b}^{(\alpha)} \partial_c, \partial_{\delta'} \rangle \\ &= \langle \nabla_{\partial_b}^{(\alpha)} (F_c^{j'} \partial_{j'}), F_{\delta'}^{k'} \partial_{k'} \rangle \\ &= (\partial_b F_c^{j'}) F_{\delta'}^{k'} \langle \partial_{j'}, \partial_{k'} \rangle \\ &= F_{\delta'}^{k'} (\partial_b F_c^{j'}) g_{j'k'}\end{aligned}$$

が成り立つ ($F_{\alpha'}^{i'} = \partial\xi^{i'}/\partial\zeta^{\alpha'}$)．ただし, ξ が α-アファイン座標であることを用いた．これから, $\partial_b F_a^{i'} = \Gamma_{ba\delta'}^{(\alpha)} g^{i'k'} F_{k'}^{\delta'}$ を得る．ゆえに, $F_{i'}^c \partial_b F_a^{i'} = \Gamma_{ba\delta'}^{(\alpha)} g^{i'k'} F_{k'}^{\delta'} F_{i'}^c = \Gamma_{ba\delta'}^{(\alpha)} g^{\delta' c} = \Gamma_{bad}^{(\alpha)} g^{dc}$ である ($g^{\kappa' c} = 0$ を用いた)．また, $g^{a\kappa} = 0$ より, $\Gamma_{\alpha\beta d}^{(-1)} g^{\alpha\beta} = \Gamma_{abd}^{(-1)} g^{ab} + \Gamma_{\kappa\lambda d}^{(-1)} g^{\kappa\lambda}$ であるが, $\Gamma_{\kappa\lambda d}^{(-1)}$ は $A(u)$ の -1-曲率を表し, 今の場合 0 である．よって, $\hat{\xi}$ のバイアスの ∂_c 方向の成分は

$$\begin{aligned} & -(\Gamma_{bad}^{(-1)} - \Gamma_{bad}^{(\alpha)}) g^{ba} g^{cd}/2N + O(N^{-3/2}) \\ = & -(1+\alpha) T_{bad} g^{ba} g^{cd}/4N + O(N^{-3/2})\end{aligned}$$

で与えられることが分かる．以上により, 次の補題を得る．

補題 3 曲指数型分布族 M が α-平坦な多様体 L に埋め込めるとする．ξ を L の α-アファイン座標とすると, $\bar{u}^c = \hat{u}^c + (1+\alpha)\hat{T}_{bad} \hat{g}^{ba} \hat{g}^{cd}/4N$ なる推定量について, $E(\xi(\bar{u}) - \xi)$ の $T_u(M)$ 方向の成分は $O(N^{-3/2})$ である．

以下では, 簡単のために $\mathcal{X}$ が有限集合の場合を考えて, 射影 Bayes 推定量と MDL 推定量の漸近的性質をまとめておこう．以下 $\mathcal{X}$ は有限集合と仮定する．このとき, $\mathcal{P}(\mathcal{X})$ の拡張 $\bar{\mathcal{P}}(\mathcal{X})$ は任意の α について α-平坦であることに注意する．この場合は, M が統計的等積であるという条件だけを仮定すれば, 補題 3, および $\tilde{f}^{(\alpha)}$, $f_{mdl}^{(\alpha)}$ の近似式 ((4) と (5)) を用いることが出来る．特に, $\mathcal{P}(\mathcal{X})$ は指数型であるから, $\mathcal{P}(\mathcal{X})$ 自体は統計的等積である．$f_{bc}^{(\alpha)}$ で, α-アファイン座標についてバイアス補正した最尤推定量を表す．すると, 次の定理が得られる．

定理 2 $\mathcal{X}$ を有限集合とする．M は $\mathcal{P}(\mathcal{X})$ の部分多様体であり, 統計的等積であると仮定する．このとき, $|\bar{x} - \bar{\eta}| = o(1)$ ならば $\tilde{f}^{((1-\alpha)/2)}[x^N]$, $f_{mdl}^{(-(1+\alpha)/2)}[x^N]$, および $f_{bc}^{(\alpha)}[x^N]$ の差は $o(1/N)$ である．

[6] 例えば, 確率変数 x のレンジ $\mathcal{X}$ が有限集合である場合は, 任意の α について成り立つ仮定である．また, $\alpha = \pm 1$ とすれば自明な仮定になる．

参考文献

[1] Amari, S. (1990). *Differential-geometrical methods in statistics,* Lecture Notes in Statistics, Vol.28, Springer-Verlag. Second Edition.

[2] Amari, S. (1994). Statistical curvature, in *Encyclopedia of Statistical Science Vol. 8,* pp. 642-646.

[3] Clarke, B. & Barron, A. (1994). Jeffreys prior is asymptotically least favorable under entropy risk. *J. Statistical Planning and Inference,* 41:37-60, 1994.

[4] Komaki, F. (1994). On asymptotic properties of predictive distributions. *METR94-21,* Univ. of Tokyo.

[5] Lauritzen, S. L. (1987). Statistical manifolds. *Differential geometry in statistical inference.* Chapter 4. IMS.

[6] Takeuchi, J. (1995). Characterization of the Bayes estimator and the MDL estimator for exponential families. The abstract appeard in *Proc. of 1995 IEEE ISIT.* To appear, *IEEE trans. on IT.*

[7] Vos, P. W. (1987). Dual geometries and their applications to generalized linear models. *Ph.D. Dissertation*, Univ. of Chicago.

[8] Vos, P. W. (1989). Fundamental equations for statistical submanifolds with applications to the Bartlett correction. *AISM*, 41, pp.429-450.

[9] 竹内 純一. (1995). **確率的知識表現の計算論的学習理論**, 学位請求論文, 東京大学工学部.

[10] 野水 克己, 佐々木 武. (1994). **アファイン微分幾何学.** 裳華房.

[11] 松島 与三. (1965). **多様体入門.** 裳華房.