数理解析研究所講究録 1380 巻 2004 年 182-190

A Pythagorean relationship in the conjugate analysis

統計数理研究所　大西 俊郎 (Toshio Ohnishi)　柳本 武美 (Takemi Yanagimoto)
The Institute of Statistical Mathematics

1. イントロダクション

まず初めに指数型分布族における共役解析を概観する．密度が

$$p(x;\,\mu)=\exp\big\{\eta x-\psi(\eta)\big\}\,a(x) \tag{1.1}$$

で与えられる指数型分布族を考える．ここで，μ および η はそれぞれ平均パラメータおよび正準パラメータである．キュムラント関数 $\psi(\eta)$ と共役な凸関数を $\phi(\mu)$ とすると，$\mu=\psi'(\eta)$ および $\eta=\phi'(\mu)$ が成り立つ．

2 つの超パラメータ m と δ をもつ事前密度

$$\begin{aligned}\pi(\mu;\,m,\,\delta)&\propto\exp\big[\,\delta\big\{m\eta-\psi(\eta)\big\}\,\big]\,b(\mu)\\&=\exp\big[\,\delta\big\{m\phi'(\mu)-\psi(\phi'(\mu))\big\}\,\big]\,b(\mu)\end{aligned} \tag{1.2}$$

は標本密度 (1.1) の下で閉じている，すなわち，事前分布と事後分布が同じ分布族に属する．ここで，$b(\mu)$ は適当な非負関数である．特に $b(\mu)=|d\eta/d\eta|=\phi''(\mu)$ のとき，μ の事後平均はデータ x について線形になる．共役事前分布は事後平均の線形性に基づいて特徴づけられる (Diaconis and Ylvisaker, 1979) が，本研究では共役性の定義として「サンプリングの下で閉じていること」を採用する．これは，Raiffa and Schlaifer (1961) による初等的な定義である．

指数型分布族 (1.1) とその共役事前分布 (1.2) を少し別の角度から眺めてみる．指数型分布族に属する 2 つの密度 $p(x;\,\mu_1)$ および $p(x;\,\mu_2)$ の Kullback-Leibler 分離度は，

$$\mathrm{KL}(\mu_1,\,\mu_2)=\phi(\mu_1)+\psi(\phi'(\mu_2))-\mu_1\phi'(\mu_2)$$

である．これを用いと密度 (1.1) および (1.2) は

$$\begin{aligned}p(x;\,\mu)&=\exp\big\{-\mathrm{KL}(x,\,\mu)\big\}\,\tilde a(x),\\\pi(\mu;\,m,\,\delta)&\propto\exp\big\{-\delta\mathrm{KL}(m,\,\mu)\big\}\,b(\mu)\end{aligned}$$

と表現される．指数関数の肩に擬距離が乗っていることに注意する．この表現を用いた共役事前分布の拡張については，Yanagimoto and Ohnishi (2004a) を参照されたい．

本研究で取り扱う標本密度は

$$p(\boldsymbol{x};\,\boldsymbol{\mu})=\exp\big\{-d(\boldsymbol{x},\,\boldsymbol{\mu})\big\}\,a(\boldsymbol{x}) \tag{1.3}$$

である．確率変数 $\boldsymbol{x}$ およびパラメータ $\boldsymbol{\mu}$ の次元を p とし，$d(\boldsymbol{a}, \boldsymbol{t})$ はいくつかの正則条件を満たしているものとする．パラメータ $\boldsymbol{\mu}$ に対して仮定する事前密度は

$$\pi(\boldsymbol{\mu};\, \boldsymbol{m},\, \delta) \propto \exp\left[-\delta d(\boldsymbol{m},\, \boldsymbol{\mu})\}\right] b(\boldsymbol{\mu}) \tag{1.4}$$

である．ここで，$\boldsymbol{m}$ の次元も p とし，$b(\boldsymbol{\mu})$ は非負関数である．指数関数の肩に擬距離が乗っているのは指数型分布族とその共役事前分布のときと同様である．推定問題における損失関数としてこの擬距離 $d(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu})$ を採用する．

本研究の目的は，事前密度 (1.4) が標本密度 (1.3) に対して共役になるような関数の集合 $\mathcal{D}_p$ を導入し，ピタゴラス関係を用いて共役解析が意味するところを明らかにすることである．

2. 共役解析の再構成

事前密度 (1.4) の規格化因子を $\exp\{k(\boldsymbol{m}, \delta)\}$ とする，すなわち，

$$k(\boldsymbol{m},\, \delta) = -\log \int \exp\{-\delta d(\boldsymbol{m},\, \boldsymbol{\mu})\}\, b(\boldsymbol{\mu})\, d\boldsymbol{\mu} \tag{2.1}$$

とおく．本研究では 2 つの超パラメータが Fisher の意味で直交することを要請する．次の補題を得る．

補題 2.1　事前密度 (1.4) において超パラメータ $\boldsymbol{m}$ および δ が Fisher の意味で直交するならば，(2.1) で定義される $k(\boldsymbol{m}, \delta)$ は $\boldsymbol{m}$ に依存しない．

証明　事前分布を $\pi(\boldsymbol{\mu};\, \boldsymbol{m},\, \delta)$ と書く．簡単のために $\boldsymbol{m}$ がスカラーのときを考える．等式

$$\mathrm{E}\left[\frac{\partial^2}{\partial m \partial \delta} \log \pi(\mu;\, m,\, \delta)\right] = \frac{\partial^2}{\partial m \partial \delta} k(m,\, \delta) - \frac{1}{\delta} \frac{\partial}{\partial m} k(m,\, \delta)$$

に注意すると，Fisher 直交性は

$$k(m,\, \delta) = \delta k_1(m) + k_2(\delta)$$

を意味する．ここで，$k_1(m)$ および $k_2(\delta)$ は適当な関数である．関数 $k(m, \delta)$ の $\delta \to \infty$ での振舞いは

$$k(m,\, \delta) \sim \log b(m) + \frac{1}{2} \log \frac{2\pi}{\dfrac{\partial}{\partial m^2} d(m,\, m)} - \frac{1}{2} \log \delta$$

である．以上から，$k_1(m) \equiv 0$ を得る．□

補題 2.1 から，事前分布は Jørgensen (1997) によって提唱された proper な拡散モデルになる．

本研究では，関数 $d(\boldsymbol{a}, \boldsymbol{t})$ が集合

$$\mathcal{D}_p = \left\{ \sum_{j=1}^{p+1} h_j(\boldsymbol{a}) \{ f_j(\boldsymbol{t}) - f_j(\boldsymbol{a}) \} \;\middle|\; \text{正則条件 (1), (2), (3) および (4) を満たす} \right\}$$

に属する場合を扱う．正則条件を列挙する．

(1) 不等式 $d(\boldsymbol{a}, \boldsymbol{t}) \geq 0$ が成立し，等号成立は $\boldsymbol{a}=\boldsymbol{t}$ のときに限る．
(2) 変数 $\boldsymbol{a}$ に関する方程式 $\nabla_t d(\boldsymbol{a}, \boldsymbol{t}) = \boldsymbol{0}$，すなわち，

$$\begin{pmatrix} f_{1,1}(\boldsymbol{t}) & \cdots & f_{p+1,1}(\boldsymbol{t}) \\ \vdots & \ddots & \vdots \\ f_{1,p}(\boldsymbol{t}) & \cdots & f_{p+1,p}(\boldsymbol{t}) \end{pmatrix} \begin{pmatrix} h_1(\boldsymbol{a}) \\ \vdots \\ h_{p+1}(\boldsymbol{a}) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

は唯一つの解 $\boldsymbol{a}=\boldsymbol{t}$ をもつ．ただし，$f_{j,i}(\boldsymbol{a}) = (\partial/\partial a_i) f_j(\boldsymbol{a})$ である．
(3) 正則条件 (2) の行列は任意の $\boldsymbol{a}$ に対してフルランクである．
(4) 任意の $\boldsymbol{a}$ に対して行列

$$\begin{pmatrix} h_{1,1}(\boldsymbol{a}) & \cdots & h_{p,1}(\boldsymbol{a}) & h_{p+1,1}(\boldsymbol{a}) \\ \vdots & \ddots & \vdots & \vdots \\ h_{1,p}(\boldsymbol{a}) & \cdots & h_{p,p}(\boldsymbol{a}) & h_{p+1,p}(\boldsymbol{a}) \\ h_1(\boldsymbol{a}) & \cdots & h_p(\boldsymbol{a}) & h_{p+1}(\boldsymbol{a}) \end{pmatrix}$$

は正則である．ただし，$h_{i,j}(\boldsymbol{a}) = (\partial/\partial a_j) f_i(\boldsymbol{a})$ である．

第 1 節で扱った指数型分布族の Kullback-Leibler 分離度は $d(\boldsymbol{a}, \boldsymbol{t})$ の例になっている．実際，$h_1(a) = a$, $h_2(a) = -1$, $f_1(t) = -\phi'(t)$ および $f_2(t) = -\psi(\phi'(t))$ とおくと，

$$\sum_{j=1}^{2} h_j(a)\{f_j(t) - f_j(a)\} = \phi(a) + \psi(\phi'(t)) - a\phi'(t) = \mathrm{KL}(a, t)$$

であるから，$\mathrm{KL}(a, t) \in \mathcal{D}_1$ であることが分かる．

以下本研究では，$d(\boldsymbol{a}, \boldsymbol{t}) = \sum_{j=1}^{p+1} h_j(\boldsymbol{a})\{f_j(\boldsymbol{t}) - f_j(\boldsymbol{a})\} \in \mathcal{D}_p$ であると仮定する．また，適当な $b(\boldsymbol{\mu})$ を選んで積分 $\int \exp\{-\delta d(\boldsymbol{m}, \boldsymbol{\mu})\} b(\boldsymbol{\mu})\, d\boldsymbol{\mu}$ が $\boldsymbol{m}$ に依存しないようにできるとし，この積分を $\exp\{-k(\delta)\}$ と書く．共役性が一般的な形で示される．

命題 2.2　事前分布

$$\pi(\boldsymbol{\mu}; \boldsymbol{m}, \delta) = \exp\{-\delta d(\boldsymbol{m}, \boldsymbol{\mu}) + k(\delta)\}\, b(\boldsymbol{\mu}) \tag{2.2}$$

は標本分布 $p(\boldsymbol{x}; \boldsymbol{\mu}) = \exp\{-d(\boldsymbol{x}, \boldsymbol{\mu})\}\, a(\boldsymbol{x})$ に対して共役である．

証明　推定量 $\hat{\boldsymbol{\mu}}_{smap}$ は $d(\boldsymbol{x}, \boldsymbol{\mu}) + \delta d(\boldsymbol{m}, \boldsymbol{\mu})$ を最小化する $\boldsymbol{\mu}$ の値とする．この推定量を定義する方程式は，

$$\begin{pmatrix} f_{1,1}(\boldsymbol{\mu}) & \cdots & f_{p+1,1}(\boldsymbol{\mu}) \\ \vdots & \ddots & \vdots \\ f_{1,p}(\boldsymbol{\mu}) & \cdots & f_{p+1,p}(\boldsymbol{\mu}) \end{pmatrix} \begin{pmatrix} h_1(\boldsymbol{x}) + \delta h_1(\boldsymbol{m}) \\ \vdots \\ h_{p+1}(\boldsymbol{x}) + \delta h_{p+1}(\boldsymbol{m}) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

である．正則条件 (2) および (3) から，任意の $j \in \{1, \ldots, p+1\}$ について

$$h_j(\boldsymbol{x}) + \delta h_j(\boldsymbol{m}) = c(\boldsymbol{x}, \boldsymbol{m}, \delta)\, h_j(\hat{\boldsymbol{\mu}}_{smap})$$

であるような $c(\boldsymbol{x},\,\boldsymbol{m},\,\delta)$ が存在することが分かる．式変形

$$\{d(\boldsymbol{x},\,\boldsymbol{\mu})+\delta d(\boldsymbol{m},\,\boldsymbol{\mu})\}-\{d(\boldsymbol{x},\,\hat{\boldsymbol{\mu}}_{smap})+\delta d(\boldsymbol{m},\,\hat{\boldsymbol{\mu}}_{smap})\}=c(\boldsymbol{x},\,\boldsymbol{m},\,\delta)\,d(\hat{\boldsymbol{\mu}}_{smap},\,\boldsymbol{\mu}) \quad (2.3)$$

に注意すると共役性が示される． □

命題 2.2 の証明で現れた推定量 $\hat{\boldsymbol{\mu}}_{smap}$ は Yanagimoto and Ohnishi (2004b) によって提案された標準化事後モード

$$\hat{\boldsymbol{\mu}}_{smap}=\underset{\boldsymbol{\mu}}{\mathrm{Argmin}}\{d(\boldsymbol{x},\,\boldsymbol{\mu})+\delta d(\boldsymbol{m},\,\boldsymbol{\mu})\}$$

である．ヤコビアン因子 $b(\boldsymbol{\mu})$ が無視されていることに注意する．これによりパラメータ変換に対する不変性が保証される．彼らがこの推定量を提案した背景には，推定手続きは主要項 $d(\boldsymbol{a},\,\boldsymbol{t})$ に基づくべきであるというアイデアがあった．

推定量 $\hat{\boldsymbol{\mu}}_{smap}$ の最適性を示す等式を導く．

命題 2.3

(i) 有限の事後リスクをもつ任意の推定量 $\hat{\boldsymbol{\mu}}$ に対して，変形ピタゴラス関係

$$\begin{aligned}&\rho_{min}\,\mathrm{E}_{post}[\,d(\hat{\boldsymbol{\mu}},\,\boldsymbol{\mu})\,]-\rho_{min}\,\mathrm{E}_{post}[\,d(\hat{\boldsymbol{\mu}}_{smap},\,\boldsymbol{\mu})\,]\\&=\mathrm{KL}(\,\pi(\boldsymbol{\mu},\,\hat{\boldsymbol{\mu}}_{smap},\,\rho_{min}),\,\pi(\boldsymbol{\mu},\,\hat{\boldsymbol{\mu}},\,\rho_{min})\,)\end{aligned} \quad (2.4)$$

が成立する．ここで，E_{post} は事後平均を意味し，ρ_{min} は (2.3) の $c(\boldsymbol{x},\,\boldsymbol{m},\,\delta)$ である．

(ii) 標準化事後モード $\hat{\boldsymbol{\mu}}_{smap}$ は損失関数 $d(\hat{\boldsymbol{\mu}},\,\boldsymbol{\mu})$ の下で最適である．

(iii) 標準化事後モード $\hat{\boldsymbol{\mu}}_{smap}$ の事後リスクは

$$\mathrm{E}_{post}\left[\,d(\hat{\boldsymbol{\mu}}_{smap},\,\boldsymbol{\mu})\,\right]=k'(\rho_{min})$$

である．

証明 事後密度は

$$\exp\{-\rho_{min}d(\hat{\boldsymbol{\mu}}_{smap},\,\boldsymbol{\mu})+k(\rho_{min})\}\,b(\boldsymbol{\mu})$$

となる．この表現は変形ピタゴラス関係 (2.4) を与える．不等式 $\rho_{min}>0$ に注意すると，(ii) は変形ピタゴラス関係から従う．等式 $\int\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_{smap},\,\rho_{min})d\boldsymbol{\mu}=1$ の両辺を ρ_{min} について偏微分すると，(iii) が得られる． □

3. 具体例

具体例として共役解析が可能な location-dispersion family を考える．Ohnishi and Yanagimoto (2003) によれば，標本密度 $p_{\kappa,\gamma}(x-\mu;\,\tau)\propto\exp\{-\tau d_{\kappa,\gamma}(x-\mu)\}$ は共役解析が可能な location-dispersion density として特徴づけられる．ここで，

$$d_{\kappa,\gamma}(t)=\frac{1}{\kappa+\gamma}\left(\frac{e^{\kappa t}}{\kappa}+\frac{e^{-\gamma t}}{\gamma}-\frac{1}{\kappa}-\frac{1}{\gamma}\right)$$

であり，κ および γ は正定数である．密度 $p_{\kappa,\gamma}(x-\mu;\,\tau)$ は正規分布と対数変換されたガンマ分布を含む．実際，$\kappa=\gamma\to+0$ とすると正規分布になり，$\kappa=1,\gamma\to+0$ とすると対数変換され

たガンマ分布になる．この2つの密度はBhattacharyya boundを達成するlocation familyとして特徴づけられる(Tanaka, 2003)ことが知られている．

拡散関数$d_{\kappa,\gamma}(a-t)$が$\mathcal{D}_1$に属することを証明する．Ohnishi and Yanagimoto (2003)は加法公式

$$d_{\kappa,\gamma}(x+y)=d_{\kappa,\gamma}(x)\tilde{d}_{\kappa,\gamma}(y)+d'_{\kappa,\gamma}(x)d'_{\kappa,\gamma}(y)+d_{\kappa,\gamma}(y)$$

が成立することを示した．ただし，

$$\tilde{d}_{\kappa,\gamma}(t)=\frac{\kappa\gamma}{\kappa+\gamma}\left(\frac{e^{\kappa t}}{\kappa}+\frac{e^{-\gamma t}}{\gamma}\right)$$

である．したがって，

$$d_{\kappa,\gamma}(a-t)=d_{\kappa,\gamma}(a)\left\{\tilde{d}_{\kappa,\gamma}(-t)-\tilde{d}_{\kappa,\gamma}(-a)\right\}+d'_{\kappa,\gamma}(a)\left\{d'_{\kappa,\gamma}(-t)-d'_{\kappa,\gamma}(-a)\right\}$$

となり，$d_{\kappa,\gamma}(a-t)\in\mathcal{D}_1$であることが分かる．

特に$\gamma=\kappa$のときを考える．密度関数は

$$p_{\kappa,\kappa}(x-\mu;\,\tau)=\frac{1}{2K_0(\tau\kappa^{-2})}\exp\left[-\frac{\tau}{\kappa^2}\cosh\{\kappa(x-\mu)\}\right]$$

となる．これをhyperbola familyと呼ぶことにする．命題2.2からhyperbola事前分布はhyperbola標本分布に対して共役である，すなわち，自己共役であることが分かる．正規分布および一様分布が自己共役であることは，hyperbola familyの自己共役性の特殊ケースとして理解できる．なぜなら，hyperbola familyは極限$\kappa\to+0$において正規分布$N(\mu,\,\tau^{-1})$に収束し，$\tau=\kappa^2/\cosh\kappa$の下で極限$\kappa\to+0$を考えると一様分布$U[\mu-1,\,\mu+1]$になるからである．

命題2.3(ii)を適用してみる．簡単のために$\kappa=1$とする．双曲三角関数の加法公式を用いると

$$\tau\cosh(x-\mu)+\delta\cosh(m-\mu)=R\cosh(\hat{\mu}_{smap}-\mu)$$

を得る．ここで，$R=\sqrt{\tau^2+\delta^2+2\tau\delta\cosh(x-m)}$である．したがって，Bayes推定量は

$$\hat{\mu}_B=\hat{\mu}_{smap}=\sinh^{-1}\left(\frac{\tau\sinh x+\delta\sinh m}{R}\right)$$

となる．この結果はvon Mises分布における結果と非常によく似ている．

4. ピタゴラス関係

共役解析が含意するものを明らかにするために，必ずしも共役ではない事前密度を仮定するBayes推定問題を考える．すなわち，標本密度$p(\boldsymbol{x};\,\boldsymbol{\mu})=\exp\{-d(\boldsymbol{x},\,\boldsymbol{\mu})\}\,a(\boldsymbol{x})$に対して，一般の事前密度$\pi(\boldsymbol{\mu})$を仮定する．損失関数はこれまでと同様に$d(\hat{\boldsymbol{\mu}},\,\boldsymbol{\mu})$を採用する．事前密度$\pi(\boldsymbol{\mu})$に対応する事後密度を$\pi(\boldsymbol{\mu}|\boldsymbol{x})$，(2.2)の共役事前密度を$\pi(\boldsymbol{\mu};\,\boldsymbol{m},\,\delta)$と書くことにする．

Bayes推定量に関して次の命題を得る．

命題4.1 評価関数$\mathrm{KL}\big(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\,\pi(\boldsymbol{\mu};\,\boldsymbol{\theta},\,\rho)\big)$を最小化する$(\boldsymbol{\theta},\,\rho)$の値を$(\boldsymbol{\theta}_{min},\,\rho_{min})$とする．損失関数$d(\hat{\boldsymbol{\mu}},\,\boldsymbol{\mu})$の下でのBayes推定量およびその事後リスクはそれぞれ$\boldsymbol{\theta}_{min}$および$k'(\rho_{min})$で与えられる．

証明　定義により，任意の推定量 $\hat{\mu}$ に対して

$$\mathrm{KL}\bigl(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\,\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\theta}}_{min},\,\rho_{min})\bigr)\le \mathrm{KL}\bigl(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\,\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}},\,\rho_{min})\bigr)$$

が成り立つ．この不等式を書き換えると

$$\rho_{min}\,\mathrm{E}\bigl[\,d(\hat{\boldsymbol{\mu}},\,\boldsymbol{\mu})-d(\boldsymbol{\theta}_{min},\,\boldsymbol{\mu})\,\bigm|\,\pi(\boldsymbol{\mu}|\boldsymbol{x})\bigr]\ge 0$$

となり，これは $\hat{\boldsymbol{\mu}}_B=\boldsymbol{\theta}_{min}$ を意味する．評価関数を ρ について偏微分し，$(\boldsymbol{\theta},\,\rho)=(\hat{\boldsymbol{\mu}}_B,\,\rho_{min})$ とおくと，

$$\mathrm{E}\bigl[\,k'(\rho_{min})-d(\hat{\boldsymbol{\mu}}_B,\,\boldsymbol{\mu})\,\bigm|\,\pi(\boldsymbol{\mu}|\boldsymbol{x})\bigr]=0$$

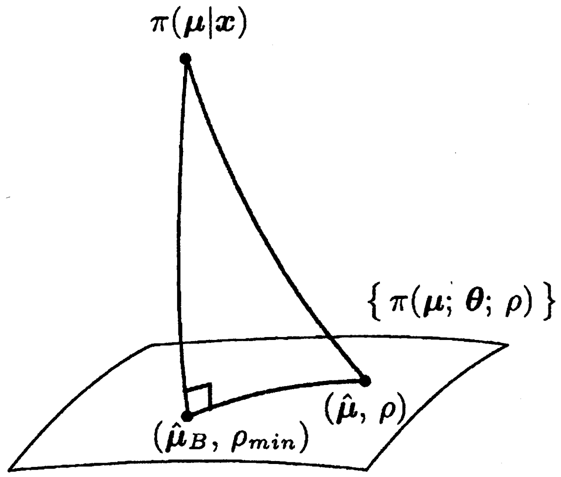
を得る．□

命題 4.1 は，事後分布を共役事前分布の‘平面’の中で最良近似すると，Bayes 推定量およびその事後リスクが得られることを示している．大雑把な表現をすれば，最良近似するということは事後密度から垂線を下ろすことであるから，垂線と平面は局所的に直交する．注意すべきは，この命題が proper な拡散モデルに属する事前分布に対して一般的に成立することである．つまり，局所的な直交性は非共役ときでも成立する．共役性が意味するものは大域的な直交性である．それが次の命題である．

命題 4.2　有限の事後リスクをもつ任意の推定量 $\hat{\mu}$ および任意の ρ に対して，ピタゴラス関係

$$\begin{aligned}&\mathrm{KL}\bigl(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\,\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}},\,\rho)\bigr)-\mathrm{KL}\bigl(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\,\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_B,\,\rho_{min})\bigr)\\&=\mathrm{KL}\bigl(\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_B,\,\rho_{min}),\,\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}},\,\rho_{min})\bigr)\end{aligned}\tag{4.1}$$

が成り立つ．



この命題は命題 2.3(i) の一般化である．事前密度が共役であるとき，事後密度は共役事前分布の平面上にある．一方，事前密度が共役でないときでも対応する共役解析が存在することを意味している．ここで，対応するとは Bayes 推定値とその事後リスクが共通という意味である．

証明　式変形

$$\log\frac{\pi(\boldsymbol{\mu}|\boldsymbol{x})}{\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}},\,\rho_{min})}-\log\frac{\pi(\boldsymbol{\mu}|\boldsymbol{x})}{\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_B,\,\rho_{min})}$$
$$=\sum_{j=1}^{p+1}\left\{\rho h_j(\hat{\boldsymbol{\mu}})-\rho_{min}h_j(\hat{\boldsymbol{\mu}}_B)\right\}f_j(\boldsymbol{\mu})+(\text{term constant in }\boldsymbol{\mu})$$

から分かるように，(4.1) の左辺の被積分関数は $f_1(\boldsymbol{\mu}),\ldots,f_{p+1}(\boldsymbol{\mu})$ を通してのみ $\boldsymbol{\mu}$ に依存する．任意の $j\in\{1,\ldots,p+1\}$ について

$$\mathrm{E}\left[\,f_j(\boldsymbol{\mu})\;\middle|\;\pi(\boldsymbol{\mu}|\boldsymbol{x})\right]=\mathrm{E}\left[\,f_j(\boldsymbol{\mu})\;\middle|\;\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_B,\,\rho_{min})\right]\tag{4.2}$$

が成り立つことを示せば命題の証明が完成する．以下，等式 (4.2) を示す．命題 4.1 から $(p+1)$ 個の等式

$$\mathrm{E}\left[\,d_i(\hat{\boldsymbol{\mu}}_B,\,\boldsymbol{\mu})\;\middle|\;\pi(\boldsymbol{\mu}|\boldsymbol{x})\right]=0\quad(1\le i\le p),\tag{4.3}$$
$$\mathrm{E}\left[\,d(\hat{\boldsymbol{\mu}}_B,\,\boldsymbol{\mu})\;\middle|\;\pi(\boldsymbol{\mu}|\boldsymbol{x})\right]=k'(\rho_{min})\tag{4.4}$$

が得られる．ただし，$d_i(\boldsymbol{\theta},\,\boldsymbol{\mu})=(\partial/\partial\theta_i)d(\boldsymbol{\theta},\,\boldsymbol{\mu})$ である．これらの方程式系を行列表現すると，

$$\begin{pmatrix} h_{1,1}(\hat{\boldsymbol{\mu}}_B) & \cdots & h_{p,1}(\hat{\boldsymbol{\mu}}_B) & h_{p+1,1}(\hat{\boldsymbol{\mu}}_B)\\ \vdots & \ddots & \vdots & \vdots\\ h_{1,p}(\hat{\boldsymbol{\mu}}_B) & \cdots & h_{p,p}(\hat{\boldsymbol{\mu}}_B) & h_{p+1,p}(\hat{\boldsymbol{\mu}}_B)\\ h_1(\hat{\boldsymbol{\mu}}_B) & \cdots & h_p(\hat{\boldsymbol{\mu}}_B) & h_{p+1}(\hat{\boldsymbol{\mu}}_B)\end{pmatrix}\begin{pmatrix}\nu_1\\ \vdots\\ \nu_p\\ \nu_{p+1}\end{pmatrix}=\begin{pmatrix}\gamma_1\\ \vdots\\ \gamma_p\\ \gamma_{p+1}\end{pmatrix}$$

となる．ここで，

$$\nu_j=\mathrm{E}\left[\,f_j(\boldsymbol{\mu})\;\middle|\;\pi(\boldsymbol{\mu}|\boldsymbol{x})\right]\qquad(1\le j\le p+1),$$
$$\gamma_i=\sum_{j=1}^{p+1}\left\{f_j(\hat{\boldsymbol{\mu}}_B)h_{j,i}(\hat{\boldsymbol{\mu}}_B)+h_j(\hat{\boldsymbol{\mu}}_B)f_{j,i}(\hat{\boldsymbol{\mu}}_B)\right\}\qquad(1\le i\le p),$$
$$\gamma_{p+1}=k'(\rho_{min})+\sum_{j=1}^{p+1}h_j(\hat{\boldsymbol{\mu}}_B)h_j(\hat{\boldsymbol{\mu}}_B)$$

とおいた．この表現から値 $\nu_1,\ldots,\nu_{p+1}$ が唯一つに決まることが分かる．事後密度 $\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_B,\,\rho_{min})$ も同じモーメント条件 (4.3) および (4.4) を満たすので，任意の $j\in\{1,\ldots,p+1\}$ について

$$\nu_j=\mathrm{E}\left[\,f_j(\boldsymbol{\mu})\;\middle|\;\pi(\boldsymbol{\mu};\,\hat{\boldsymbol{\mu}}_B,\,\rho_{min})\right]$$

が成り立つ． □

5. 最小情報性

共役事前分布の最小情報性を示す．共役事前密度 $\pi(\boldsymbol{\mu};\ \boldsymbol{m},\ \delta)$ の超パラメータ $\boldsymbol{m}$ および δ を任意に固定する．データ $\boldsymbol{x}$ が与えられたという条件の下で，事前密度の集合

$$\mathcal{P}_{\boldsymbol{x}} = \left\{ \pi(\boldsymbol{\mu}) \;\middle|\; \begin{array}{l} \text{Bayes 推定値およびその事後リスクが} \\ \text{共役事前密度 } \pi(\boldsymbol{\mu};\ \boldsymbol{m},\ \delta) \text{ のそれらと共通である} \end{array} \right\}$$

を導入し，次の最小化問題の解を探す．

最小化問題　事前密度の集合 $\mathcal{P}_{\boldsymbol{x}}$ の中で，汎関数

$$G[\pi(\boldsymbol{\mu})] = \mathrm{KL}\big(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\ \pi(\boldsymbol{\mu};\ \boldsymbol{x},\ 1)\big) = \int \left\{ \log \frac{\pi(\boldsymbol{\mu}|\boldsymbol{x})}{\pi(\boldsymbol{\mu};\ \boldsymbol{x},\ 1)} \right\} \pi(\boldsymbol{\mu}|\boldsymbol{x}) d\boldsymbol{\mu}$$

を最小化する密度 $\pi_*(\boldsymbol{\mu})$ を求めよ．ここで，$\pi(\boldsymbol{\mu}|\boldsymbol{x})$ は $\pi(\boldsymbol{\mu})$ に対応する事後密度である．

汎関数 $G[\pi(\boldsymbol{\mu})]$ の意味を述べる．本研究の枠組みでは事前密度に現れる関数 $b(\boldsymbol{\mu})$ はヤコビアン因子にすぎず，推定手続きには無関係である．この意味において標本密度 $p(\boldsymbol{x};\ \boldsymbol{\mu})$ と事前密度 $\pi(\boldsymbol{\mu};\ \boldsymbol{x},\ 1)$ を同一視することができる．汎関数 $G[\pi(\boldsymbol{\mu})]$ は事前密度 $\pi(\boldsymbol{\mu})$ に含まれる情報量であるという解釈が成立するので，上記の最小化問題の解は最小情報事前密度である．

命題 4.2 のピタゴラス関係 (4.1) は共役事前密度の最小情報性を含意する．

命題 5.1　事前密度の集合 $\mathcal{P}_{\boldsymbol{x}}$ の中で汎関数 $G[\pi(\boldsymbol{\mu})]$ を最小化するのは，共役事前密度 $\pi(\boldsymbol{\mu};\ \boldsymbol{m},\ \delta)$ である．

証明　共役事前密度 $\pi(\boldsymbol{\mu};\ \boldsymbol{m},\ \delta)$ に対する Bayes 推定値およびその事後リスクをそれぞれ $\hat{\boldsymbol{\mu}}_B$ and $k'(\rho_{min})$ とおく．等式 (4.1) において $(\boldsymbol{\theta},\ \rho) = (\boldsymbol{x},\ 1)$ とすれば，共役事前密度 $\pi(\boldsymbol{\mu};\ \boldsymbol{m},\ \delta)$ の最小情報性を示すピタゴラス関係

$$\begin{aligned} &\mathrm{KL}\big(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\ \pi(\boldsymbol{\mu};\ \boldsymbol{x},\ 1)\big) - \mathrm{KL}\big(\pi(\boldsymbol{\mu}|\boldsymbol{x}),\ \pi(\boldsymbol{\mu};\ \hat{\boldsymbol{\mu}}_B,\ \rho_{min})\big) \\ &= \mathrm{KL}\big(\pi(\boldsymbol{\mu};\ \hat{\boldsymbol{\mu}}_B,\ \rho_{min}),\ \pi(\boldsymbol{\mu};\ \boldsymbol{x},\ 1)\big) \end{aligned}$$

を得る．　□

References

DIACONIS, P. AND YLVISAKER, D. (1979). Conjugate priors for exponential families, *Ann. Statist.*, **7**, 269–281.

JØRGENSEN, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.

OHNISHI, T. AND YANAGIMOTO, T. (2003). Conjugate location-dispersion families. Manuscript (Research memorandum of ISM No. 831).

RAIFFA, H. AND SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*, Graduate School of Business Administration, Harvard Univ., Boston.

TANAKA, H. (2003). On a location parameter family of distributions attaining the Bhattacharyya bound, In *Approximations to the Statistical Distributions, Kokyuroku No.1334 of RIMS* (Ed. M. Akahira), 158–174.

YANAGIMOTO, T. AND OHNISHI, T. (2004a). Extensions of a conjugate prior through the Kullback-Leibler separators, *J. Multivariate Anal.*, In press.

YANAGIMOTO, T. AND OHNISHI, T. (2004b). Standardized posterior mode for the flexible use of a conjugate prior, *J. Statist. Plann. Inference*, In press.