数理解析研究所講究録
第2047巻 2017年 77-88

α-ダイバージェンス損失の下での Stein 現象

九州大学・経済学研究院　大西俊郎
Toshio Ohnishi
Faculty of Economics, Kyushu University

§1. Introduction

正規分布における予測問題を考える．すなわち，$\boldsymbol{X} \sim N(\boldsymbol{\theta}, v_x I_d)$ かつ $\boldsymbol{Y} \sim N(\boldsymbol{\theta}, v_y I_d)$ のとき，$\boldsymbol{X} = \boldsymbol{x}$ の下で $\boldsymbol{Y}$ の確率密度 $\phi(\boldsymbol{y} - \boldsymbol{\theta}; v_y)$ を推定する．ただし，$d \geq 3$ は次元，$\boldsymbol{\theta}$ は未知の平均パラメータ，$v_x, v_y > 0$ は既知定数であり，I_d は d 次単位行列を，$\phi(\boldsymbol{z} - \boldsymbol{\theta}; v)$ は正規分布 $N(\boldsymbol{\theta}, vI_d)$ の確率密度を表す．

予測の良さを α-ダイバージェンスで測ることにする．α-ダイバージェンスを定義するには次のような「べき乗」関数 u_r を用意しておくと便利である．

$$u_r(t) := \begin{cases} \log t & (r = 0), \\ \dfrac{t^r - 1}{r} & (r \neq 0). \end{cases} \tag{1.1}$$

ここで，$a := b$ または $b =: a$ は b によって a を定義することを意味する．後の議論を簡単にするために，逆関数 u_r^{-1} を求めておく．

$$u_r^{-1}(s) = \begin{cases} \exp(s) & (r = 0), \\ (1 + rs)^{\frac{1}{r}} & (r \neq 0). \end{cases} \tag{1.2}$$

$\alpha \in [-1, 1]$ とする．α-ダイバージェンスは次のように定義される．

Definition 1.1 (α-ダイバージェンス)**.** 確率密度 p から確率密度 q への α-ダイバージェ

ンスは,

$$\mathrm{D}_\alpha(p\|q) := \mathrm{E}_p\left[\tilde{u}_{1-\beta}\left(\frac{q}{p}\right)\right].$$

ただし, E_p は p の下での期待値を表し, $\beta := (1-\alpha)/2$ であり, 関数 $\tilde{u}_r$ は次のように定義される.

$$\tilde{u}_r(t) := \begin{cases} t\log t - (t-1) & (r=1), \\ \dfrac{u_r(t) - u_1(t)}{r-1} & (r \neq 1). \end{cases}$$

α-ダイバージェンスは Kullback-Leibelr ダイバージェンスの一般化であり, $\mathrm{D}_{-1}(p\|q) = \mathrm{KL}(p\|q)$ および $\mathrm{D}_{+1}(p\|q) = \mathrm{KL}(q\|p)$ が成り立つ. この 2 つは互いに双対と言われる. また, $\alpha \in [-1,1] \Leftrightarrow \beta \in [0,1]$ に注意されたい.

事前密度 $\pi(\boldsymbol{\theta})$ を仮定し, Bayes 予測問題を構成すると, 最適解が決まる. この最適解は Bayes 予測密度と呼ばれる. 本論文のねらいは, improper な一様事前密度 $\pi_{\mathrm{U}}(\boldsymbol{\theta}) \propto 1$ に基づく Bayes 予測密度を, 頻度主義の意味で改善するための十分条件を導くことである.

先行研究の結果を 2 つ補題の形で引用する.

Lemma 1.1 (Stein, 1981) $\beta = 0 \Leftrightarrow \alpha = +1$ のとき,

$$\begin{aligned} &(\pi_{\mathrm{U}} \text{ の下でのリスク}) - (\pi \text{ の下でのリスク}) \\ &= \frac{v_x^2}{2v_y}\int_{\mathbb{R}^d}\left\{-2\frac{\triangle m_\pi(\boldsymbol{w};v_x)}{m_\pi(\boldsymbol{w};v_x)} + \left\|\nabla \log m_\pi(\boldsymbol{w};v_x)\right\|^2\right\}\phi(\boldsymbol{w}-\boldsymbol{\theta};v_x)\mathrm{d}\boldsymbol{w}. \end{aligned}$$

ただし, m_π は標本密度 $\phi(\boldsymbol{w}-\boldsymbol{\theta};v)$, 事前密度 $\pi(\boldsymbol{\theta})$ に対応する周辺密度である.

$$m_\pi(\boldsymbol{w};v) := \int_{\mathbb{R}^d}\phi(\boldsymbol{w}-\boldsymbol{\theta};v)\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}.$$

$\alpha = +1$ のとき, Bayes 予測問題はパラメータの推定問題に帰着し, π_{U} に基づく予測は最尤推定になるにことに注意されたい.

ここで, 注意すべき等式を挙げる. 非負関数 f に対して,

$$-4\frac{\triangle\sqrt{f}}{\sqrt{f}} = -2\frac{\triangle f}{f} + \|\nabla \log f\|^2.$$

より一般には次が成り立つ.

$$\frac{\triangle f^r}{f^r} = r\frac{\triangle f}{f} + r(r-1)\|\nabla \log f\|^2 \quad (r \neq 0), \tag{1.3}$$

$$\triangle \log f = \frac{\triangle f}{f} - \|\nabla \log f\|^2.$$

Stein (1981) の結果は, $\sqrt{m_\pi}$ が優調和ならばリスクが改善されることを意味している.

Lemma 1.2 (George *et al.*, 2006) $\beta = 1 \Leftrightarrow \alpha = -1$ のとき,

$$\begin{aligned}&(\pi_{\mathrm{U}}\text{ の下でのリスク}) - (\pi\text{ の下でのリスク})\\ &= \frac{1}{2}\int_{\mathbb{R}^d}\int_{v(1)}^{v_x}\left\{-2\frac{\triangle m_\pi(\boldsymbol{w};v)}{m_\pi(\boldsymbol{w};v)} + \big\|\nabla \log m_\pi(\boldsymbol{w};v)\big\|^2\right\}\phi\big(\boldsymbol{w}-\boldsymbol{\theta};v\big)\mathrm{d}v\mathrm{d}\boldsymbol{w}.\end{aligned}$$

ただし, $v(1)$ は次式で定義され, $v(1) < v_x$ である.

$$\frac{1}{v(\beta)} := \frac{1}{v_x} + \frac{\beta}{v_y}. \tag{1.4}$$

本発表のねらいを具体的に言えば, 以下に引用する Brown (1979) の予想が正規分布における予測問題で正しいことを示すことである.

> "Decision-theoretic properties seem to depend on the general structure of the problem (the general type of problem (location, scale), and the dimension of the parameter space) and on the prior in a Bayesian-setup, but not on the loss function."

George *et al.* (2006) の結果 Lemma 1.2 と Stein (1981) の結果 Lemma 1.1 の関係に触れておく. $v_x/v_y \approx 0$ のとき, George *et al.* (2006) は Stein (1981) に帰着する. 実際, 近似式

$$v(1) = \frac{v_x v_y}{v_x + v_y} \approx v_x\left(1 - \frac{v_x}{v_y}\right)$$

を利用すると,

$$\begin{aligned}&\frac{1}{2}\int_{v(1)}^{v_x}\left\{-2\frac{\triangle m_\pi(\boldsymbol{w};v)}{m_\pi(\boldsymbol{w};v)} + \big\|\nabla \log m_\pi(\boldsymbol{w};v)\big\|^2\right\}\phi\big(\boldsymbol{w}-\boldsymbol{\theta};v\big)\mathrm{d}v\\ &\approx \frac{v_x^2}{2v_y}\left\{-2\frac{\triangle m_\pi(\boldsymbol{w};v_x)}{m_\pi(\boldsymbol{w};v_x)} + \big\|\nabla \log m_\pi(\boldsymbol{w};v_x)\big\|^2\right\}\phi\big(\boldsymbol{w}-\boldsymbol{\theta};v_x\big).\end{aligned}$$

§2. Bayes 予測密度

α-ダイバージェンス損失の下での Bayes 予測問題について，Corcuera & Giummole (1999) の結果が基本的である．

Lemma 2.1 (Corcuera & Giummole, 1999). 予測密度 $\hat{p}(\boldsymbol{y};\boldsymbol{x})$ のよさを α-ダイバージェンス損失 $\mathrm{D}_\alpha\Big(\phi(\boldsymbol{y}-\boldsymbol{\theta};v_y)\,\Big\|\,\hat{p}(\boldsymbol{y};\boldsymbol{x})\Big)$ によって評価するとき，Bayes 予測問題の最適解は，

$$\hat{p}_{\beta,\pi}(\boldsymbol{y};\boldsymbol{x}) \propto u_\beta^{-1}\left(\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\Big[u_\beta\Big(\phi(\boldsymbol{y}-\boldsymbol{\theta};v_y)\Big)\Big]\right).$$

ここで，u_β と u_β^{-1} は (1.1) と (1.2) で定義された「べき乗関数」とその逆関数であり，$\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}$ は事後期待値を意味する．

Bayes 予測密度 $\hat{p}_{\beta,\pi}$ は確率密度のさまざまな平均になっている．$\beta=0\Leftrightarrow\alpha=1$ のとき幾何平均であり，$\beta=1\Leftrightarrow\alpha=-1$ のとき算術平均である．$\beta\neq0$ のとき，

$$\hat{p}_{\beta,\pi}(\boldsymbol{y};\boldsymbol{x}) \propto \left\{\mathrm{E}_{\boldsymbol{\theta}|\boldsymbol{x}}\Big[\phi^\beta(\boldsymbol{y}-\boldsymbol{\theta};v_y)\Big]\right\}^{1/\beta}$$

のように書くこともできる．このことは，Hardy *et al.* (1952) の定理 83 から示される．すなわち，関数 $u_\beta(t)$ と t^β は

$$u_\beta(t)=\frac{t^\beta-1}{\beta}$$

のようにアファイン変換で結ばれており，このタイプの平均ではアファイン変換しても平均は不変だからである．

Improper な一様事前密度 π_{U} に基づく Bayes 予測密度を求める．Lemma 2.1 を適用すると次が得られる．

Lemma 2.2 (π_{U} に基づく Bayes 予測密度). Improper な一様事前密度 $\pi_{\mathrm{U}}(\boldsymbol{\theta})$ を仮定する．Bayes 予測密度 $\hat{p}_{\beta,\mathrm{U}}$ は，

$$\hat{p}_{\beta,\mathrm{U}}(\boldsymbol{y};\boldsymbol{x})=\phi(\boldsymbol{y}-\boldsymbol{x};\beta v_x+v_y).$$

これが改善のターゲットである．$\beta=0$ のとき，最尤推定になっている．$\beta\neq 0$ のときは分散が標本密度と異なっていて，推定問題の枠から外れることが分かる．

Lemma 2.2 の証明は，Lemma 2.1 および次の補題による．

Lemma 2.3 (平方完成による恒等式). $\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta}$ に関する次の恒等式が成り立つ．

$$
\begin{aligned}
&\phi^{\beta}(\boldsymbol{y}-\boldsymbol{\theta};v_y)\phi(\boldsymbol{x}-\boldsymbol{\theta};v_x)\\
&=\bigl\{L(\beta)\bigr\}^{\beta}\phi^{\beta}(\boldsymbol{y}-\boldsymbol{x};\beta v_x+v_y)\phi\bigl(\boldsymbol{x}+\gamma(\beta)(\boldsymbol{y}-\boldsymbol{x})-\boldsymbol{\theta};v(\beta)\bigr).
\end{aligned}
$$

ただし，$v(\beta)$ は (1.4) で定義された量であり，$\gamma(\beta)$ および $L(\beta)$ は次のように定義される．

$$
\begin{aligned}
\gamma(\beta)&:=\frac{\beta v_x}{\beta v_x+v_y},\\
L(\beta)&:=\begin{cases}\exp\left(-\frac{d}{2}\frac{v_x}{v_y}\right) & (\beta=0),\\ \left(\frac{v_y}{\beta v_x+v_y}\right)^{\frac{d}{2}\left(\frac{1}{\beta}-1\right)} & (0<\beta\leq 1).\end{cases}
\end{aligned}
$$

一般の事前密度 π に基づく Bayes 予測密度を求める．Lemma 2.1 と Lemma 2.3 を利用すると，次の補題を得る．

Lemma 2.4 (一般の π に基づく Bayes 予測密度). 事前密度 π に基づく Bayes 予測密度 $\hat{p}_{\beta,\pi}$ は，

$$
\hat{p}_{\beta,\pi}(\boldsymbol{y};\boldsymbol{x})\propto m_{\pi}^{1/\beta}\bigl(\boldsymbol{x}+\gamma(\beta)(\boldsymbol{y}-\boldsymbol{x});v(\beta)\bigr)\ \hat{p}_{\beta,\mathrm{U}}(\boldsymbol{y};\boldsymbol{x})\phi(\boldsymbol{z}-\boldsymbol{x};\beta v_x+v_z).
$$

Improper な一様事前密度に対する Bayes 予測密度と比較されたい．$\pi=\pi_{\mathrm{U}}$ のとき $m_{\pi}\propto 1$ であることに注意する．

以下，$\beta\in(0,1)$ を固定し，一般の事前密度に基づく Bayes 予測密度 $\hat{p}_{\beta,\pi}(\boldsymbol{y};\boldsymbol{x})$ と improper な一様な事前密度に基づく Bayes 予測密度 $\hat{p}_{\beta,\mathrm{U}}(\boldsymbol{y};\boldsymbol{x})$ のリスク差を計算する．リスク差を計算するとき，次の恒等式が有用である．

$$
\mathrm{D}_{\alpha}(p_1\|p_2)-\mathrm{D}_{\alpha}(p_1\|p_3)=\frac{1}{\beta(1-\beta)}\int_{\mathbb{R}^d}p_1^{\beta}(\boldsymbol{y})\bigl\{p_3^{1-\beta}(\boldsymbol{y})-p_2^{1-\beta}(\boldsymbol{y})\bigr\}\mathrm{d}\boldsymbol{y}.
$$

次の補題が得られる．

Lemma 2.5 (2 つの Bayes 予測密度のリスク差). $\hat{p}_{\beta,\mathrm{U}}(\boldsymbol{y};\boldsymbol{x})$ と $\hat{p}_{\beta,\pi}(\boldsymbol{y};\boldsymbol{x})$ のリスク差は，

$$\frac{\{L(\beta)\}^{\beta}}{\beta(1-\beta)}\left\{\int_{\mathbb{R}^d} Q\phi\big(\boldsymbol{w}-\boldsymbol{\theta};v(\beta)\big)\mathrm{d}\boldsymbol{w}-1\right\}.$$

ただし，

$$Q := \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} R\phi(\boldsymbol{t})\mathrm{d}\boldsymbol{t}\right]^{\beta-1}\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{s}, \tag{2.1}$$

$$R := \frac{m_\pi^{1/\beta}\big(\boldsymbol{w}+\tilde{\gamma}(\beta)(\boldsymbol{t}-\boldsymbol{s});v(\beta)\big)}{m_\pi^{1/\beta}\big(\boldsymbol{w};v(\beta)\big)}, \tag{2.2}$$

$$\phi(\boldsymbol{t}) := \ (d\ \text{変量標準正規分布の確率密度}),$$

$$\tilde{\gamma}(\beta) := \frac{\beta v_x}{\sqrt{\beta v_x + v_y}}.$$

§3. 主要な定理

主要な定理を述べるときに重要な役割を果たすのが，周辺密度の平方根が優調和であるという条件である．これを定義しておく．

Definition 3.1 (条件 MSH). 事前密度 π に対する周辺密度を m_π とする．
事前密度 π が条件 MSH を満たす $\overset{\mathrm{def}}{\Longleftrightarrow}$

$$\forall \boldsymbol{w}\in\mathbb{R}^d \text{および} \ \forall v>0\ \text{に対して} -2\frac{\triangle m_\pi(\boldsymbol{w};v)}{m_\pi(\boldsymbol{w};v)}+\|\nabla \log m_\pi(\boldsymbol{w};v)\|^2 \geq 0.$$

Lemma 1.1 の直後に述べたように，これは $\sqrt{m_\pi}$ の優調和性を意味する．

Lemma 2.5 から分かるように，不等式

$$Q\big(\boldsymbol{w},\tilde{\gamma}(\beta),v(\beta)\big)\geq 1,\quad \forall \boldsymbol{w}\in\mathbb{R}^d$$

が成り立つならば，リスクが改善される．ただし，

$$Q := Q(\boldsymbol{w}, \tau, v) := \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} R\phi(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} \right\}^{\beta-1} \phi(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}, \tag{3.1}$$

$$R := R(\boldsymbol{w}, \boldsymbol{t}-\boldsymbol{s}, \tau, v) := \frac{m_\pi^{1/\beta}\big(\boldsymbol{w} + \tau(\boldsymbol{t}-\boldsymbol{s}); v\big)}{m_\pi^{1/\beta}(\boldsymbol{w}; v)}. \tag{3.2}$$

これらは (2.1)，(2.2) で定義された関数を改めて定義したものである．

以下の計算における準備として，天下り的ではあるが，$a - b\tau^2 = v$ を満たすように定数 a, b を選ぶ．ただし，$a > 0$ とする．よって，$b > -v/\tau^2$ である．

(3.2) で定義された関数 R を次のように式変形する．

$$\begin{aligned} R &= \frac{m_\pi^{1/\beta}(\boldsymbol{w} - \tau\boldsymbol{s}; a)}{m_\pi^{1/\beta}(\boldsymbol{w}; v)} \cdot \frac{m_\pi^{1/\beta}\big(\boldsymbol{w} + \tau(\boldsymbol{t}-\boldsymbol{s}); v\big)}{m_\pi^{1/\beta}(\boldsymbol{w} - \tau\boldsymbol{s}; a)} \\ &= \frac{m_\pi^{1/\beta}(\boldsymbol{w} - \tau\boldsymbol{s}; v + b\tau^2)}{m_\pi^{1/\beta}(\boldsymbol{w}; v)} \frac{m_\pi^{1/\beta}(\boldsymbol{w} - \tau\boldsymbol{s} + \tau\boldsymbol{t}; a - b\tau^2)}{m_\pi^{1/\beta}(\boldsymbol{w} - \tau\boldsymbol{s}; a)}. \end{aligned}$$

関数 g, h を

$$\begin{aligned} g(\tau) &:= \frac{m_\pi^{1-1/\beta}(\boldsymbol{w} - \tau\boldsymbol{s}; v + b\tau^2)}{m_\pi^{1-1/\beta}(\boldsymbol{w}; v)}, \\ h(\tau_1, \tau_2) &:= \frac{m_\pi^{1/\beta}(\boldsymbol{w} - \tau_1\boldsymbol{s} + \tau_2\boldsymbol{t}; a - b\tau_2^2)}{m_\pi^{1/\beta}(\boldsymbol{w} - \tau_1\boldsymbol{s}; a)} \end{aligned}$$

のように定義すると，

$$Q = \int_{\mathbb{R}^d} g(\tau) \left\{ \int_{\mathbb{R}^d} h(\tau, \tau) \phi(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} \right\}^{\beta-1} \phi(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}.$$

関数 h が満たす積分方程式を導く．

Lemma 3.1 (h が満たす積分方程式). 次の積分方程式が成り立つ．

$$\begin{aligned} &\int_{\mathbb{R}^d} h(\tau_1, \tau_2) \phi(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} \\ &= 1 + \frac{1-\beta}{\beta^2} \int_0^{\tau_2} \left\{ \tau_2' \int_{\mathbb{R}^d} h(\tau_1, \tau_2') c_\pi(\boldsymbol{w} - \tau_1\boldsymbol{s} + \tau_2'\boldsymbol{t}; a - b\tau_2'^2) \phi(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} \right\} d\tau_2'. \end{aligned} \tag{3.3}$$

ただし,

$$c_\pi(\boldsymbol{w}; v) := -\frac{\beta(b-1)}{1-\beta}\frac{\triangle m_\pi(\boldsymbol{w}; v)}{m_\pi(\boldsymbol{w}; v)} + \|\nabla \log m_\pi(\boldsymbol{w}; v)\|^2. \tag{3.4}$$

略証： 積分形の平均値の定理により,

$$h(\tau_1, \tau_2) = 1 + \int_0^{\tau_2} \frac{\partial h(\tau_1, \tau_2')}{\partial \tau_2'} d\tau_2'.$$

周辺密度 $m_\pi(\boldsymbol{w}; v)$ は熱方程式の初期値問題

$$\frac{\partial}{\partial v} f(\boldsymbol{w}; v) = \frac{1}{2}\triangle f(\boldsymbol{w}; v), \quad f(\boldsymbol{w}; 0) = \pi(\boldsymbol{w})$$

の解である．Gauss の発散定理によれば，関数 f, g について，$f(\boldsymbol{t})\nabla g(\boldsymbol{t})$ が無限遠点でゼロに収束するとき,

$$\int_{\mathbb{R}^d} \nabla f(\boldsymbol{t}) \cdot \nabla g(\boldsymbol{t}) \mathrm{d}\boldsymbol{t} = -\int_{\mathbb{R}^d} f(\boldsymbol{t}) \triangle g(\boldsymbol{t}) \mathrm{d}\boldsymbol{t}.$$

$\boldsymbol{t}\phi(\boldsymbol{t}) = -\nabla\phi(\boldsymbol{t})$ に注意して，Gauss の発散定理を適用する．さらに (1.3) を適用すると，積分方程式 (3.3) が得られる．□

$\tau \approx 0$ の場合に Q の挙動を調べる．積分方程式 (3.3) の右辺において τ について 2 次近似すると,

$$\int_{\mathbb{R}^d} h(\tau, \tau)\phi(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \approx 1 + \frac{1-\beta}{2\beta^2}\tau^2 c_\pi(\boldsymbol{w}; v).$$

したがって,

$$\left\{\int_{\mathbb{R}^d} h(\tau, \tau)\phi(\boldsymbol{t})\mathrm{d}\boldsymbol{t}\right\}^{\beta-1} \approx 1 - \frac{(1-\beta)^2}{2\beta^2}\tau^2 c_\pi(\boldsymbol{w}; v)$$

であり，ゆえに,

$$Q \approx \left\{1 - \frac{(1-\beta)^2}{2\beta^2}\tau^2 c_\pi(\boldsymbol{w}; v)\right\} \int_{\mathbb{R}^d} g(\tau)\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{s}.$$

Lemma 3.1 と同様にして次の補題が得られる．

Lemma 3.2 (g が満たす積分方程式). 次の積分方程式が成り立つ．

$$\int_{\mathbb{R}^d} g(\tau)\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \quad = 1 + \frac{1-\beta}{\beta^2}\int_0^\tau \tau' \left\{\int_{\mathbb{R}^d} g(\tau')\tilde{c}_\pi(\boldsymbol{w} - \tau'\boldsymbol{s}; v + b\tau'^2)\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right\} d\tau'.$$

ただし,

$$\tilde{c}_\pi(\boldsymbol{w};v) := -\beta(1+b)\frac{\triangle m_\pi(\boldsymbol{w};v)}{m_\pi(\boldsymbol{w};v)} + \|\nabla \log m_\pi(\boldsymbol{w};v)\|^2. \tag{3.5}$$

Lemma 3.2 を用いると,$\tau \approx 0$ のとき,

$$\int_{\mathbb{R}^d} g(\tau)\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \approx 1 + \frac{1-\beta}{2\beta^2}\tau^2\tilde{c}_\pi(\boldsymbol{w};v).$$

関数 $c_\pi, \tilde{c}_\pi$ の定義 (3.4),(3.5) を用いて計算すると,

$$Q \approx 1 + \frac{1-\beta}{2\beta}\tau^2\left\{-2\frac{\triangle m_\pi(\boldsymbol{w};v)}{m_\pi(\boldsymbol{w};v)} + \|\nabla \log m_\pi(\boldsymbol{w};v)\|^2\right\}.$$

したがって,Lemma 1.1 と組み合わせて,次の定理を得る.

Theorem 3.1 ($\tilde{\gamma}(\beta) \approx 0$ **の場合).** $\tilde{\gamma}(\beta) \approx 0$ とする.条件 MSH が満たされているならば,一様事前密度 π_{U} による Bayes 予測密度は改善される.

上の定理が示しているように,$\tilde{\gamma}(\beta) \approx 0$ ならば Stein (1981) の結果が現れる.これには $\beta \approx 0$ または $v_x/v_y \approx 0$ の場合が含まれる.

次に,$\beta \approx 1$ の場合を考える.$(t^{\beta-1}-1)/(1-\beta) \approx -\log t$ などに注意し,Lemma 1.2 を適用すると次の定理を得る.

Theorem 3.2 ($\beta \approx 1$ **の場合).** $\beta \approx 1$ のとき,

$$Q \approx 1 + (1-\beta) \times (\beta = 1 \text{ のときのリスク差})$$

と近似できる.条件 MSH が満たされているならば,一様事前密度 π_{U} による Bayes 予測密度は改善される.

次に,リスク差がゼロになる場合を考える.$b = 2/\beta - 1$ のとき,

$$c_\pi = \tilde{c}_\pi = -2\frac{\triangle m_\pi}{m_\pi} + \|\nabla \log m_\pi\|^2$$

であることに注意すると，次の定理を得る．

Theorem 3.3 (等号成立条件). 事前密度 π に対する周辺密度 m_π が

$$\forall \boldsymbol{w} \in \mathbb{R}^d \text{および} \forall v > 0 \text{に対して} -2\frac{\triangle m_\pi(\boldsymbol{w};v)}{m_\pi(\boldsymbol{w};v)} + \|\nabla \log m_\pi(\boldsymbol{w};v)\|^2 = 0$$

を満たしているならば，一様事前密度 π_{U} による Bayes 予測密度と同じリスクをもつ．

最後に，事前密度が scale family に属する，すなわち，

$$\pi(\boldsymbol{\theta}) = \frac{1}{v^{d/2}} f_0\left(\frac{\boldsymbol{\theta}}{\sqrt{v}}\right) \quad (f_0 \text{ は所与の関数})$$

の場合を考える．このとき，簡単な計算により，周辺密度も scale family に属することが分かる．実際，

$$m_\pi(\boldsymbol{y};v) = \frac{1}{v^{d/2}} f_1\left(\frac{\boldsymbol{y}}{\sqrt{v}}\right).$$

ただし，

$$f_1(\boldsymbol{w}) := \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\|\boldsymbol{w} - \boldsymbol{\theta}'\|^2\right\} f_0(\boldsymbol{\theta}')\mathrm{d}\boldsymbol{\theta}'.$$

関数 $Q(\boldsymbol{w}, \tau, v)$ の定義 (3.1) を思い出すと，事前密度が scale family に属するとき，

$$Q(\boldsymbol{w}, \tau, v) = \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \frac{f_1^{1/\beta}\left(\frac{\boldsymbol{w}+\tau(\boldsymbol{t}-\boldsymbol{s})}{\sqrt{v}}\right)}{f_1^{1/\beta}\left(\frac{\boldsymbol{w}}{\sqrt{v}}\right)} \phi(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \right\}^{\beta-1} \phi(\boldsymbol{s})\mathrm{d}\boldsymbol{s} =: \tilde{Q}\left(\frac{\boldsymbol{w}}{\sqrt{v}}, \frac{\tau}{\sqrt{v}}\right).$$

つまり，関数 Q は実質的に 2 変数関数となる．

次の定理が得られる．

Theorem 3.4 (事前密度が scale family の場合における改善). 事前密度が scale family に属するとする．条件 MSH が満たされているならば，

$$\tilde{Q}\left(\frac{\boldsymbol{w}}{\sqrt{v(\beta)}}, \frac{\tilde{\gamma}(\beta)}{\sqrt{v(\beta)}}\right) \geq 1, \quad \forall \boldsymbol{w} \in \mathbb{R}^d.$$

すなわち，一様事前密度 π_{U} に基づく Bayes 予測密度は改善される．

証明：Theorem 3.1 で示されたことは次のとおりである．条件 MSH が満たされているならば，十分小さい $\varepsilon > 0$ に対して

$$Q(\boldsymbol{w}', \varepsilon, v) \geq 1, \quad \forall \boldsymbol{w}' \in \mathbb{R}^d \text{および} \, \forall v > 0.$$

事前密度 π が scale family に属するならば，十分小さい $\varepsilon > 0$ に対して

$$\tilde{Q}\left(\frac{\boldsymbol{w}'}{\sqrt{v}}, \frac{\varepsilon}{\sqrt{v}}\right) \geq 1, \quad \forall \boldsymbol{w}' \in \mathbb{R}^d \text{および} \, \forall v > 0.$$

与えられた $\boldsymbol{w}, \tilde{\gamma}(\beta), v(\beta)$ に対して，

$$\frac{\boldsymbol{w}'}{\sqrt{v}} = \frac{\boldsymbol{w}}{\sqrt{v(\beta)}} \quad \text{かつ} \quad \frac{\varepsilon}{\sqrt{v}} = \frac{\tilde{\gamma}(\beta)}{\sqrt{v(\beta)}}$$

のように $\boldsymbol{w}$ および v を選ぶと，

$$\tilde{Q}\left(\frac{\boldsymbol{w}}{\sqrt{v(\beta)}}, \frac{\tilde{\gamma}(\beta)}{\sqrt{v(\beta)}}\right) \geq 1.$$

この不等式は改善を意味する．□

REFERENCES

Brown, L. D. (1979). A heuristic method for determining admissibility of estimators —with applications. *Annal of Statistic*, **7**, 960-994.

Corcuera, J. M. & Giummole, F. (1999). A generalized Bayes rule for prediction. *Scandinavian Journal of Statistics*, **26**, 265-279.

George, E. I., Liang, F. & Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annal of Statistic*, **34**, 78-91.

Hardy, G. H., Littlewood, J. E., & Polya, G. (1952). *Inequalities*, Cambridge University Press.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annal of Statistic*, **9**, 1135-1151.