α-ダイバージェンス損失の下での予測問題における調和関数型事前分布の役割

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．

ここで v_x, v_y を導入している理由を述べる．例えば，$X_1, \ldots, X_m \sim N(\theta, I_d), \mathrm{iid}$ かつ $Y_1, \ldots, Y_n \sim N(\theta, I_d), \mathrm{iid}$ のとき，$\bar{X} \sim N(\theta, m^{-1} I_d)$ および $\bar{Y} \sim N(\theta, n^{-1} I_d)$ となり，上述の設定に含まれる．

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，$\alpha \in [-1, 1]$ および $\beta \in [0, 1]$ 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．標本分布 $\phi(\boldsymbol{w} - \boldsymbol{\theta}; v)$，事前分布 $\pi(\boldsymbol{\theta})$ に対応する周辺分布を m_π とおく．

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

本論文のねらいは, improper な一様事前分布 $\pi_{\mathrm{U}}(\boldsymbol{\theta}) = 1$ に基づく Bayes 予測分布を, 頻度主義の意味で改善することを考察し，improper な調和関数型事前分布 $\pi_{\mathrm{H}}(\boldsymbol{\theta}) = \|\theta\|^{2-d}$ が果たす役割を明らかにすることである．

先行研究の結果を 2 つ補題の形で引用する．

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

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

$\alpha = +1$ 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) $\beta = 1$ すなわち $\alpha = -1$ 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．

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

George *et al.* (2006) の結果が意味するところは，$\triangle\sqrt{m_\pi} \leq 0$ は改善のための十分条件であるが，もう少し弱い条件

$$\int_{v(1)}^{v_x}\left\{\frac{\triangle\sqrt{m_\pi(\boldsymbol{w};v)}}{\sqrt{m_\pi(\boldsymbol{w};v)}}\right\}\phi\big(\boldsymbol{w}-\boldsymbol{\theta};v\big)\mathrm{d}v \leq 0$$

でも改善されるということである．

§2. Preliminary results

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．$\beta=0$ すなわち $\alpha=1$ のとき幾何平均であり，$\beta=1$ すなわち $\alpha=-1$ のとき算術平均である．$\beta\neq 0$ のとき，

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

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．

Lemma 2.2 (Hardy *et al.*, 1934). Lemma 2.1 における平均では $u_\beta(t)=(t^\beta-1)/\beta$ および $u_\beta^{-1}(s)=(1+\beta s)^{1/\beta}$ をそれぞれ

$$t^\beta \text{ および } s^{1/\beta}$$

に置き換えても平均は不変である．

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 2.3 (Maruyama *et al.* (2019) の結果 1). 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 2.4 (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．

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

Lemma 2.3 の Bayes 予測密度は，proper な事前分布 $\theta \sim N(0_d, cI_d)$ に基づく Bayes 予測分布において $c \to \infty$ とすることによって得られる．したがって，Bayes 統計学の定理により，ミニマックス性が保証される．これを補題の形で述べておく．

Lemma 2.5 (Maruyama *et al.* (2019) の結果 2). Improper な一様事前分布 $\pi_{\mathrm{U}}(\boldsymbol{\theta})$ に基づく Bayes 予測分布 $\hat{p}_{\beta,\mathrm{U}}$ は，コンスタントリスクかつミニマックスである．

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 2.4 を利用すると，次の補題を得る．

Lemma 2.6 (Maruyama *et al.* (2019) の結果 3). 一般の事前分布 π に基づく Bayes 予測分布 $\hat{p}_{\beta,\pi}$ は，$\beta \neq 0$ のとき

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

であり，$\beta = 0$ のとき

$$\hat{p}_{0,\pi}(\boldsymbol{y};\boldsymbol{x}) \propto \phi\bigl(\boldsymbol{y} - \hat{\theta}_B; v_y\bigr),$$
$$\hat{\theta}_B := x - v_x \nabla \log m_\pi(\boldsymbol{x}; v_x)$$

である．

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 2.7 (Maruyama *et al.* (2019) の結果 4). $\hat{p}_{\beta,\mathrm{U}}(\boldsymbol{y};\boldsymbol{x})$ と $\hat{p}_{\beta,\pi}(\boldsymbol{y};\boldsymbol{x})$ のリスク差は，

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

ただし，

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

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

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

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

§3. Main results

主要な結果を述べる．まず最初に，improper な調和関数型事前分布を仮定したときに周辺分布が有界になることを示す．

Theorem 3.1 (有界性). $\pi_{\mathrm{H}}(\theta) := \|\theta\|^{2-d}$ に対する improper な周辺分布

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

は有界な非負関数である．

証明　証明の本質だけを記す．極座標変換を行う．$r = \|\theta\|$ とおくと，$d\theta = r^{d-1}drd\omega$ となる．ただし，$d\omega$ は角度要素を意味する．$r^{2-d} \times r^{d-1} = r$ に注意すれば所与の結果を得る．□

次に，improper な調和関数型事前分布を仮定したときの周辺分布が優調和であることを示す．

Theorem 3.2 (優調和性). Improper な周辺分布 $m_{\mathrm{H}}(\boldsymbol{w}; v)$ は優調和関数である．

証明　π_{H} は Laplace 方程式の Green 関数である：

$$\triangle\|\boldsymbol{\theta}\|^{2-d} = -(d-2)\omega_{d-1}\delta(\boldsymbol{\theta}).$$

ここで，ω_{d-1} は d 次元単位球の表面積であり，δ は Dirac の δ-関数である．次の Gauss の発散定理

$f(\boldsymbol{t})\nabla g(\boldsymbol{t}), \{\nabla f(\boldsymbol{t})\}g(\boldsymbol{t})$ が無限遠点でゼロに収束するとき，

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

を適用して，

$$\triangle m_{\mathrm{H}}(\boldsymbol{w}; v) = -(d-2)\omega_{d-1}\phi(\boldsymbol{w}; v). \quad \square$$

Improper な一様事前分布 π_{U} に摂動を与えることを考える．具体的には，$\varepsilon > 0$ を十分小さい正数とし，improper な一様事前分布 π_{U} に対して調和関数型事前分布 π_{H} を摂動として与える：

$$\pi(\theta) = 1 + \varepsilon \pi_{\mathrm{H}}(\theta).$$

これに対応する improper な周辺分布は

$$m_\pi(w; v) := 1 + \varepsilon m_{\mathrm{H}}(\boldsymbol{w}; v)$$

となる．このとき，リスクが改善される．

Theorem 3.3 (リスクの改善). $\varepsilon > 0$ とする．Improper な事前分布 $\pi(\theta) = 1+\varepsilon\|\theta\|^{2-d}$ を仮定するとき，ε^2 を無視する近似で，リスク差は

$$-\varepsilon \frac{\{L(\beta)\}^\beta}{\beta^2} \int_0^{\tilde{\gamma}(\beta)} S \, d\tau.$$

ただし，

$$S := 2\tau \int_{\mathbb{R}^d} \phi(\tilde{\boldsymbol{t}}) \triangle_{\boldsymbol{w}} m_{\mathrm{H}}\big(\boldsymbol{w} + \sqrt{2}\tau \tilde{\boldsymbol{t}}; v(\beta)\big) \mathrm{d}\tilde{\boldsymbol{t}}.$$

したがって，リスクは改善される．

証明　Lemma 2.7 により，リスク差は次の量に比例する：

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

ただし，Q は (2.1) および (2.2) によって定義される．(2.2) によって定義される R を近似計算する．m_{H} の有界性に注意して，ε について Taylor 展開し，ε^2 を無視する近似を行う．

$$\begin{aligned} R &= \left\{ \frac{m_\pi\big(\boldsymbol{w} + \tilde{\gamma}(\beta)(\boldsymbol{t} - \boldsymbol{s}); v(\beta)\big)}{m_\pi\big(\boldsymbol{w}; v(\beta)\big)} \right\}^{1/\beta} \\ &\approx 1 + \frac{\varepsilon}{\beta} \Big\{ m_{\mathrm{H}}\big(\boldsymbol{w} + \tilde{\gamma}(\beta)(\boldsymbol{t} - \boldsymbol{s}); v(\beta)\big) - m_{\mathrm{H}}\big(\boldsymbol{w}; v(\beta)\big) \Big\}. \end{aligned}$$

周辺分布の pdf の差を

$$\Delta m_{\mathrm{H}} := m_{\mathrm{H}}\big(\boldsymbol{w}+\tilde{\gamma}(\beta)(\boldsymbol{t}-\boldsymbol{s}); v(\beta)\big) - m_{\mathrm{H}}\big(\boldsymbol{w}; v(\beta)\big).$$

のように定義すると，

$$\left[\int_{\mathbb{R}^d} R\phi(\boldsymbol{t})\mathrm{d}\boldsymbol{t}\right]^{\beta-1} \approx 1-\varepsilon\frac{1-\beta}{\beta}\int_{\mathbb{R}^d}\Delta m_{\mathrm{H}}\ \phi(\boldsymbol{t})\mathrm{d}\boldsymbol{t}.$$

Q の定義 (2.1) により，

$$Q \approx 1-\varepsilon\frac{1-\beta}{\beta}\int_{\mathbb{R}^{2d}}\Delta m_{\mathrm{H}}\ \phi(\boldsymbol{t})\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{t}d\boldsymbol{s}.$$

次の変数変換（直交変換）を行う：

$$\tilde{\boldsymbol{t}} := \frac{\boldsymbol{t}-\boldsymbol{s}}{\sqrt{2}}, \quad \tilde{\boldsymbol{s}} := \frac{\boldsymbol{t}+\boldsymbol{s}}{\sqrt{2}}.$$

ヤコビアンの絶対値が 1 であり，

$$\phi(\boldsymbol{t})\phi(\boldsymbol{s})\mathrm{d}\boldsymbol{t}d\boldsymbol{s} = \phi(\tilde{\boldsymbol{t}})\phi(\tilde{\boldsymbol{s}})\mathrm{d}\tilde{\boldsymbol{t}}d\tilde{\boldsymbol{s}}.$$

なので，

$$Q \approx 1-\varepsilon\frac{1-\beta}{\beta}\int_{\mathbb{R}^{2d}}\Delta m_{\mathrm{H}}\ \phi(\tilde{\boldsymbol{t}})\phi(\tilde{\boldsymbol{s}})\mathrm{d}\tilde{\boldsymbol{t}}d\tilde{\boldsymbol{s}}.$$

積分を用いて

$$\Delta m_{\mathrm{H}} := m_{\mathrm{H}}\big(\boldsymbol{w}+\tilde{\gamma}(\beta)(\boldsymbol{t}-\boldsymbol{s}); v(\beta)\big) - m_{\mathrm{H}}\big(\boldsymbol{w}; v(\beta)\big)$$

を表現すると，

$$\begin{aligned}\Delta m_{\mathrm{H}} &= \int_0^{\tilde{\gamma}(\beta)}\frac{\partial}{\partial\tau}m_{\mathrm{H}}\big(\boldsymbol{w}+\tau(\boldsymbol{t}-\boldsymbol{s}); v(\beta)\big)d\tau\\ &= (\boldsymbol{t}-\boldsymbol{s})\cdot\int_0^{\tilde{\gamma}(\beta)}\nabla_{\boldsymbol{w}}m_{\mathrm{H}}\big(\boldsymbol{w}+\tau(\boldsymbol{t}-\boldsymbol{s}); v(\beta)\big)d\tau.\end{aligned}$$

次の量を定義する：

$$S := \int_{\mathbb{R}^d}\sqrt{2}\,\tilde{\boldsymbol{t}}\cdot\nabla_{\boldsymbol{w}}m_{\mathrm{H}}\big(\boldsymbol{w}+\sqrt{2}\tau\tilde{\boldsymbol{t}}; v(\beta)\big)\phi(\tilde{\boldsymbol{t}})\mathrm{d}\tilde{\boldsymbol{t}}.$$

Theorem 3.1 における S の定義と異なっているように見えるが，実は相等しいことを示す．上で定義した S が $\tilde{\boldsymbol{s}}$ に依存しないことに注意すると，

$$
\begin{aligned}
Q &\approx 1 - \varepsilon\frac{1-\beta}{\beta}\int_0^{\tilde{\gamma}(\beta)} d\tau \int_{\mathbb{R}^d} S\ \phi(\tilde{\boldsymbol{s}})d\tilde{\boldsymbol{s}} \\
&= 1 - \varepsilon\frac{1-\beta}{\beta}\int_0^{\tilde{\gamma}(\beta)} S\ d\tau.
\end{aligned}
$$

標準正規分布の pdf の性質 $\tilde{\boldsymbol{t}}\phi(\tilde{\boldsymbol{t}}) = -\nabla\phi(\tilde{\boldsymbol{t}})$ に注意する．また，$\nabla_{\boldsymbol{w}}$ と $\nabla_{\tilde{\boldsymbol{t}}}$ の作用を考えると，

$$
\begin{aligned}
S &= -\sqrt{2}\int_{\mathbb{R}^d} \nabla_{\boldsymbol{w}} m_{\mathrm{H}}\big(\boldsymbol{w} + \sqrt{2}\tau\tilde{\boldsymbol{t}}; v(\beta)\big)\cdot\nabla\phi(\tilde{\boldsymbol{t}})\mathrm{d}\tilde{\boldsymbol{t}} \\
&= -\frac{1}{\tau}\int_{\mathbb{R}^d} \nabla_{\tilde{\boldsymbol{t}}}\, m_{\mathrm{H}}\big(\boldsymbol{w} + \sqrt{2}\tau\tilde{\boldsymbol{t}}; v(\beta)\big)\cdot\nabla\phi(\tilde{\boldsymbol{t}})\mathrm{d}\tilde{\boldsymbol{t}}.
\end{aligned}
$$

Theorem 3.2 の証明で言及した Gauss の発散定理を用いる．また，$\triangle_{\boldsymbol{w}}$ と $\triangle_{\tilde{\boldsymbol{t}}}$ の作用を考えると，

$$
\begin{aligned}
S &= -\frac{1}{\tau}\int_{\mathbb{R}^d} \nabla_{\tilde{\boldsymbol{t}}}\, m_{\mathrm{H}}\big(\boldsymbol{w} + \sqrt{2}\tau\tilde{\boldsymbol{t}}; v(\beta)\big)\cdot\nabla\phi(\tilde{\boldsymbol{t}})\mathrm{d}\tilde{\boldsymbol{t}} \\
&= \frac{1}{\tau}\int_{\mathbb{R}^d} \phi(\tilde{\boldsymbol{t}})\triangle_{\tilde{\boldsymbol{t}}}\, m_{\mathrm{H}}\big(\boldsymbol{w} + \sqrt{2}\tau\tilde{\boldsymbol{t}}; v(\beta)\big)\mathrm{d}\tilde{\boldsymbol{t}} \\
&= 2\tau\int_{\mathbb{R}^d} \phi(\tilde{\boldsymbol{t}})\triangle_{\boldsymbol{w}} m_{\mathrm{H}}\big(\boldsymbol{w} + \sqrt{2}\tau\tilde{\boldsymbol{t}}; v(\beta)\big)\mathrm{d}\tilde{\boldsymbol{t}}. \quad \square
\end{aligned}
$$

Theorem 3.3 の証明において，周辺分布の有界性および優調和性のみを使っていることに注意すると，次の定理を得る．

Theorem 3.4 (リスクの改善). $\varepsilon > 0$ とする．ある事前分布 π_0 に対する周辺分布が有界かつ優調和と仮定する．Improper な事前分布 $\pi(\theta) = 1 + \varepsilon\pi_0(\theta)$ を仮定するとき，ε^2 を無視する近似で，リスクは改善される．

Acknowledgement. This work was supported by the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University. It was also supported by KAKENHI 20K11711, Grant-in-Aid for Scientific Research (C).

REFERENCES

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

George, Liang & Xu (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, 34, 7891.

Hardy, Littlewood & Polya (1934). *Inequalities.* Cambridge.

Maruyama, Matsuda & Ohnishi (2019). Harmonic Bayesian prediction under α-divergence. *IEEE Transactions on Information Theory*, 65, 5352-5366.

Stein (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9, 1135-1151.