The 35th Symposium on Information Theory
and its Applications (SITA2012)
Beppu, Oita, Japan, Dec. 11–14, 2012

Bregman 情報量を損失としたベイズリスクに関する考察
A study of Bayes risk with Bregman divergence loss function

池田 思朗* 竹内 純一†
Shiro Ikeda Jun'ichi Takeuchi

Abstract—通信路の確率モデルと入力に関する制約が与えられたとすると，通信路容量は入力と出力との間の相互情報量の上限として定義される．一方で，通信路容量は KL 情報量を損失とするベイズリスクの max-min 問題の最適値として定義することもできる．本稿では，KL 情報量を Bregman 情報量に置き換えたベイズリスクを定義し，その max-min 問題を考える．スカラー通信路の通信路容量を達成する入力分布が連続分布となるのは AWGN に対して入力の平均パワーを制約した場合であり，そのときの最適な入力分布は正規分布となる．他の通信路，そして入力に関する制約に対しては，最適な入力分布が離散分布となることが多い．KL 情報量を Bregman 情報量に置き換えた場合も同様の結果が得られた．

1 はじめに

X を入力とし，Y を出力とする通信路の確率モデル $p(Y|X)$ が与えられたとする．通信路容量は X に関する与えられた制約の下で X と Y の相互情報量の上限として定義される [17]．最大値が存在するならば，最大値を与える X の分布はこの通信路への入力の最適な分布となる．この最適化問題はベイズ統計の立場では Kullback-Leibler (K-L) ダイバージェンスを損失とするベイズリスクに関する max-min 問題とみなすことができる．こうした max-min 問題に関しては様々な研究がなされている [6,8]．本研究では K-L ダイバージェンスを用いて定義された問題を一般化し，U-ダイバージェンス [13,19] を用いて定義する．Bregman ダイバージェンスの特殊なクラスである U-ダイバージェンスは K-L ダイバージェンスや β-ダイバージェンス [15] を含んでいる．

通信路容量の問題では，通信路が連続分布で定義され，制約を満たす入力の分布として連続分布を含める場合であっても，多くの場合，通信路容量を達成する分布が離散分布になる [1,5,7,11,18] ．連続分布が最適となる例は平均パワー制約下で Additive White Gaussian Noise (AWGN) 通信路を用いる場合の他はほとんど知られていない．本稿では一般化した問題に対して基本的な性質を調べた後，U-ダイバージェンスとして β-ダイバージェンスを用いたときに連続分布が最適となる例を調べる．その結果，正規分布が最適となるような通信路，および入力の制約が存在することがわかった．

* 〒 190-8562 東京都立川市緑町 10-3 統計数理研究所 The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo, 190-8562, Japan. E-mail: shiro@ism.ac.jp
† 〒 819-0395 福岡県福岡市西区元岡 744 九州大学 大学院システム情報科学研究院 情報学部門 Department of Informatics, Graduate School of Information Science and Electrical Engineering, Kyushu University. 744, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan.

2 本稿であつかう Bregman 情報量

2.1 U-ダイバージェンス

1 次元の実数値をとる確率変数 $\mathcal{Y} \subseteq \Re$ に関する 2 つの確率分布 $p(Y)$ と $q(Y)$ を考える．この場合，U-ダイバージェンスは次のように定義される [14,15].

$$
\begin{aligned}
D_U(p,q) &= H_U(p,q) - H_U(p) \\
H_U(p,q) &= \int_{\mathcal{Y}} \Big[U(\xi(q(y))) - p(y)\xi(q(y)) \Big] dy, \qquad (1) \\
H_U(p) &= H_U(p,p).
\end{aligned}
$$

ここで $U(t)$ は 1 次元の実数値 t を入力とする狭義凸関数，u は U の導関数，すなわち $u = U'$ である．また，ξ は u の逆関数，$\xi = (u)^{-1}$ だとする．以下で U は C^2 級であり，u は狭義増加関数だとする．

U-ダイバージェンスの定義は一般の Bregman 情報量の定義とは異なる．一般の Bregman 情報量は実関数 f, g に対して以下のように定義される．

$$
\begin{aligned}
D_{Bregman}(f,g) &= \int_{\Re} d(f(z), g(z)) dz, \\
d(f,g) &= U(g) - U(f) - u(f)(g-f).
\end{aligned}
$$

ここで U は U-ダイバージェンスで定義した狭義凸関数とする．図 1 から分るように $d(f,g)$ は常に正であるため，Bregman 情報量は常に非負であり，ほとんど至るところで $f = g$ が成り立つときにのみ 0 となる．

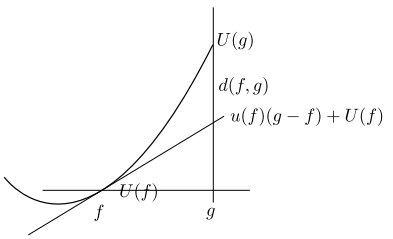

図 1: Bregman divergence.

U-ダイバージェンスは p と q を関数 ξ によって変換してから一般の Bregman 情報量に代入したものと考えることができる．したがって U-ダイバージェンスも

Bregman 情報量と同様に非負であり，2 つの確率分布がほとんど至るところで等しいときにのみ 0 となる．このため，2 つの分布の近さを表しているとみなせる．

以上から U-ダイバージェンスは Bregman 情報量のうち特殊なものであることがわかる．本稿では Bregman 情報量の中でも (1) 式の U-ダイバージェンスとして表されるもののみを扱う．

2.2 Bregman 情報量の例

ここでは，U-ダイバージェンスとして表現できる 2 つのダイバージェンスを示す．

K-L ダイバージェンス

$U(z) = \exp(z)$, $u(z) = \exp(z)$, $\xi(z) = \log(z)$ と定義する．すると (1) 式は以下のようになる．

$$D_{KL}(p,q) = \int_{\Re} p(y) \log \frac{p(y)}{q(y)} dy + \int_{\Re} (q(y) - p(y)) dy.$$

p と q が確率分布のときには以下のようになる．これは K-L ダイバージェンスである．

$$D_{KL}(p,q) = \int_{\Re} p(y) \log \frac{p(y)}{q(y)} dy,$$

β ダイバージェンス

β を正の実数として $U(z)$, $u(z)$, $\xi(z)$ を以下のように定義する．

$$U(z) = \frac{1}{\beta+1}(\beta z + 1)^{\frac{\beta+1}{\beta}}$$
$$u(z) = (\beta z + 1)^{\frac{1}{\beta}}, \qquad \xi(z) = \frac{z^\beta - 1}{\beta},$$

これらを用いて β-ダイバージェンスは以下のように定義される．

$$D_\beta(p,q) = \frac{1}{\beta+1}\int_{\Re} (q(y)^{\beta+1} - p(y)^{\beta+1}) dy - \frac{1}{\beta}\int_{\Re} p(y)(q(y)^\beta - p(y)^\beta) dy. \quad (2)$$

$\beta \downarrow 0$ とする極限では β-ダイバージェンスは K-L ダイバージェンスに収束する．

この β-ダイバージェンスは，統計学や機械学習の研究で用いられている [13, 19]．また，統計物理学で提案されている Tsallis 統計との関係が指摘されている [16]．

以下で用いるいくつかの記号を定義しておく．

$p(y|x)$ ：以下で扱う確率分布．情報理論では X を入力，Y を出力とする通信路と考えれば良い．統計では Y は注目している確率変数であり，X は母数である．ベイズ統計では X も確率変数とみなす．本稿では X も Y も実数で 1 次元のときのみを扱う．

F ：実数上で定義される X の累積分布．情報理論では入力の分布，ベイズ統計では母数の事前分布である．

$\mathcal{F}$ ： 以下では F として以下の条件を満たすものを考える．

$$\int_{\Re} m(x) dF(x) \le A, \quad m(x) \ge 0, \quad A > 0. \quad (3)$$

m は正値をとる X の関数であり，これは関数 m の期待値に関する制約となる．例えば，通信工学では入力の平均電力が問題となる．その場合には $m(x) = x^2$ とすれば良い．

(3) 式の制約を満たす F の集合を $\mathcal{F}$ と定義する．

$$\mathcal{F} = \Big\{ F \mid \int_{\Re} m(x) dF(x) \le A \Big\}.$$

この集合は凸集合である．また，一般には $\mathcal{F}$ は連続分布も離散分布も含む．

3 本稿で考える問題

3.1 通信路容量と max-min 問題

通信路容量は最適化問題の最適値として定義される．本稿ではその定義を拡張する．本節では，まず通信路容量の定義を確認する．

Problem 1. 通信路容量は以下の最適化問題の最適値として定義される．

$$C = \sup_{F \in \mathcal{F}} \int_{\Re} D_{KL}(p(y|x); p(y;F)) dF(x), \quad (4)$$

ここで $p(y;F)$ は以下で定義される周辺分布である，

$$p(y;F) = \int_{\Re} p(y|x) dF(x).$$

また，$\mathcal{F}$ の定義から，(4) 式は以下のように書き直せる．

$$C = \sup_{F} \int_{\Re} D_{KL}(p(y|x); p(y;F)) dF(x) \quad subject\ to \int_{\Re} m(x) dF(x) \le A. \quad (5)$$

この最適化問題は汎関数を最大とする X の確率分布を求めることである．K-L ダイバージェンスの積分値として表現されているこの汎関数は X と Y の間の相互情報量とも呼ばれる．相互情報量は以下の $I(X;Y)$ として定義される．

$$I(X;Y) = \int_{\Re}\int_{\Re} p(y|x) \log \frac{p(y|x)}{p(y;F)} dy dF(x) = \int_{\Re} D_{KL}(p(y|x); p(y;F)) dF(x).$$

$I(X;Y)$ を F の汎関数とみなすとき，$\mathcal{F}$ 上での上限が通信路容量である．

情報理論の観点では，最適値を与える F が存在するならば，その F は最適な入力分布となる．こうした最適な入力分布は理想的な変調方式と深く関係している [9,10]．

一方，ベイズ統計の立場では (4) 式の目的関数は K-L ダイバージェンスを損失とするベイズリスクと見なされる．したがってこの問題はベイズリスクの最悪値を評価したものである．最適値を与える F は reference 事前分布と呼ばれ，70 年代から研究されている [3,4]．reference 事前分布は無情報事前分布のひとつである．

(4) 式の問題は max-min 問題として書き直せる．まず，古くから知られている次の結果を示す [2]．

$$p(y;F)=\arg\min_{q}\int_{\Re}D_{KL}(p(y|x);q(y))dF(x).$$

この結果を用いると (4) 式は次のように書き直せる．

$$C=\sup_{F\in\mathcal{F}}\min_{q}\int_{\Re}D_{KL}(p(y|x);q(y))dF(x).$$

(4) 式にある最適化問題，あるいは上式の最適化問題において $\mathcal{F}$ 上で $I(X;Y)$ の最大値を与える分布を F^* と書くことにする．$\mathcal{F}$ は一般に連続分布も離散分布も含んでおり F^* が連続となるか離散となるかは通信路と X に関する制約 (3) 式との組み合わせで決まる．F^* が連続分布が最適となるのは $m(x)=x^2$ という平均パワー制約のもとで AWGN 通信路を考える場合以外ではほとんど知られていない．一方で多くの問題で F^* が離散分布となることが知られている [1,5,7,11,18] ．

本稿では連続分布が最適となる $m(x)$ と $p(y|x)$ との組み合わせに注目する．このため，まず，F^* を特徴づける Karush-Kuhn-Tucker (KKT) 条件について説明する．

以下では $\mathcal{F}$ がコンパクトで F に関して相互情報量 $I(X;Y)$ が連続で狭義凹関数であるとする[1] すなわち，最適値が存在し，それを達成する F^* がひとつだけ存在すると仮定する．まず，最適化問題をラグランジュの未定定数法によって書きなおす．

$$C=\max_{F\in\mathcal{F}}\Big[\int_{\Re}D_{KL}(p(y|x);p(y;F))dF(x)-\lambda\Big(\int_{\Re}m(x)dF(x)-A\Big)\Big].$$

上式を満たす $\lambda\geq 0$ が存在する．$I(X;Y)$ の F による方向微分 (ガトー微分) を考えると，最適な分布 F^* が満たす KKT 条件が得られる [1]．

Corollary 1 (KKT condition)**.** F^* が (4) 式の最適値を与えるとし，F^* の増加点の集合を E^* とおく．このとき，次の関係が成り立つ．

$$D_{KL}(p(y|x);p(y;F^*))\begin{cases}=C+\lambda(m(x)-A) & \text{for } x\in E^*\\ \leq C+\lambda(m(x)-A) & \text{for } x\notin E^*\end{cases}. \tag{6}$$

F^* の増加点は確率分布の台であり，F^* が連続分布であるならば，連続分布の台において (6) 式が成り立たなければならない．F^* が連続分布となる有名な例は AWGN 通信路における平均パワー制約のもとでの通信路容量である．このとき，制約は $\int_{\Re}x^2dF(x)\leq\sigma_S^2$, すなわち $m(x)=x^2$, $A=\sigma_S^2$ であり $p(y|x)=(1/\sqrt{2\pi\sigma_N^2})\exp(-(x-y)^2/2\sigma_N^2)$ である．最適な $F^*(x)$ は $F^*(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi\sigma_S^2}}\exp(-t^2/2\sigma_S^2)dt$ であり，(6) 式の左辺は以下のようになる．

$$D_{KL}(p(y|x);p(y;F^*))=\frac{1}{2}\log\Big(1+\frac{\sigma_N^2}{\sigma_S^2}\Big)+\frac{1}{2(\sigma_S^2+\sigma_N^2)}(x^2-\sigma_S^2).$$

このとき，通信路容量は $\log(1+\mathrm{SNR})/2$ であり，$\lambda=1/(2(\sigma_S^2+\sigma_N^2))$ とおけば (6) 式が成り立つことがわかる．

3.2 問題の拡張

本稿では 1 式中の K-L 情報量を Bregman 情報量で置き換える問題を考える．すなわち，以下の問題を考える．

Problem 2.

$$C_U=\sup_{F\in\mathcal{F}}\int_{\Re}D_U(p(y|x);p(y;F))dF(x), \tag{7}$$

この問題は以下のようにも書ける．

$$C_U=\sup_{F}\int_{\Re}D_U(p(y|x);p(y;F))dF(x)\quad subject\ to\int_{\Re}m(x)dF(x)\leq A. \tag{8}$$

この問題は情報理論の立場からは興味のある拡張では無いかもしれない．しかし，ベイズ統計の立場では Bregman 情報量を損失としたベイズリスクとみなすことができる．特殊な場合として reference 事前分布とも関係していることから reference 事前分布の一般化と考えられる．

通信路容量の問題と同様にここでも上記の問題が max-min 問題として書き直せることを示す．まず，以下の定理を示す [14]．

Theorem 1. *U*-ダイバージェンスを損失とする次のベイズリスク R を考える

$$R(q;F)=\int_{\Re}D_U(p(y|x),q(y))dF(x).$$

[1] 厳密には位相の定義から始めるべきだが，ここでは省略する．一般的な手続きについては以下を参照して欲しい [1,5,7,11,12,18].

$R(q;F)$ の最小値は $q(y) = p(y;F)$ のときに達成される．すなわち，以下の式が成り立つ．

$$\min_{q\in\mathcal{P}} R(q;F) = \int_{\Re} D_U(p(y|x), p(y;F))dF(x).$$

Proof. U-ダイバージェンスの定義から $H_U(p(x|\theta))$ は q を含んでいない．このことから以下の関係が成り立つ．

$$\begin{aligned}\arg\min_{q\in\mathcal{P}} R(q;F) &= \arg\min_{q\in\mathcal{P}} \int_{\Re} D_U(p(y|x), q(y))dF(x)\\ &= \arg\min_{q\in\mathcal{P}} \int_{\Re} H_U(p(y|x), q(y))dF(x).\end{aligned}$$

さらに次の式が成り立つ．

$$\begin{aligned}&\int_{\Re} H_U(p(y|x), q(y))dF(x)\\ =&\int_{\Re}\big[U(\xi(q(y)))-p(y;F)\xi(q(y))\big]dy=H_U(p(y;F), q(y)).\end{aligned}$$

したがって

$$\begin{aligned}\arg\min_{q\in\mathcal{P}} R(q;F) &= \arg\min_{q\in\mathcal{P}} H_U(p(y;F), q(y))\\ &= \arg\min_{q\in\mathcal{P}} D_U(p(y;F), q(y)) = p(y;F),\end{aligned}$$

このとき最適値は $\int_{\Re} D_U(p(y|x), p(y;F))dF(x)$ となる．

□

この結果から Problem 2 は以下の max-min 問題となることがわかる．

$$C_U = \sup_{F\in\mathcal{F}} \min_{q\in\mathcal{P}} R(q;F).$$

以下では，新しく定義した問題の最適値を与える F^* の満たす KKT 条件をみていく．

Lemma 1. 以下のように F の汎関数 $R(F)$ を定義する．

$$R(F) = \min_{q\in\mathcal{P}} R(q,F) = \int_{\mathcal{X}} D_U(p(y|x); p(y;F))dF(x). \tag{9}$$

この定義を用いれば (8) 式の問題は 次のように書ける．

$$C_U = \sup_{F\in\mathcal{F}} R(F).$$

(3) 式のように定義される $\mathcal{F}$ が汎弱位相に関してコンパクトだとする．このとき $R(F)$ は $F \in \mathcal{F},$ に関して凹関数である．

Proof. 任意の $F_0, F_1 \in \mathcal{F}$, $(F_0 \neq F_1)$ と $0 < \eta < 1$ に対して以下の関係が成り立つことを示せば良い．

$$R((1-\eta)F_0 + \eta F_1) > (1-\eta)R(F_0) + \eta\ R(F_1). \tag{10}$$

まず，$F_\eta = (1-\eta)F_0 + \eta F_1$ とおき，(10) 式の左辺を書き直す．

$$\begin{aligned}R(F_\eta) &= \int_{\mathcal{X}} \Big[H_U(p(y|x), p(y;F_\eta)) - H_U(p(y|x))\Big]dF_\eta(x)\\ &= H_U(p(y;F_\eta)) - \int_{\mathcal{X}} H_U(p(y|x))dF_\eta(x).\end{aligned}$$

(10) 式の右辺は以下のように書ける．

$$\begin{aligned}&(1-\eta)R(F_0) + \eta\ R(F_1)\\ &\quad=(1-\eta)H_U(p(y;F_0)) + \eta\ H_U(p(y;F_1))\\ &\qquad\qquad - \int_{\mathcal{X}} H_U(p(y;\theta))dF_\eta(x).\end{aligned}$$

したがって

$$\begin{aligned}&\text{左辺} - \text{右辺} = H_U(p(y;F_\eta)) -\\ &\qquad\Big[(1-\eta)H_U(p(y;F_0)) + \eta\ H_U(p(y;F_1))\Big].\end{aligned} \tag{11}$$

次の関係に注意すると，

$$\begin{aligned}H_U(p(y;F_\eta)) &= (1-\eta)H_U(p(y;F_0), p(y;F_\eta))\\ &\quad + \eta\ H_U(p(y;F_1), p(y;F_\eta)).\end{aligned}$$

(11) 式は以下のように書き直せる．

$$\begin{aligned}\text{左辺} - \text{右辺} =&(1-\eta)D_U(p(y;F_0), p(y;F_\eta))\\ &+ \eta\ D_U(p(y;F_1), p(y;F_\eta)) > 0.\end{aligned}$$

□

次に KKT 条件を導く．

Lemma 2 (方向微分 (ガトー微分))**.** 次の方向微分を考える．

$$R'_{F_0}(F_1) = \lim_{\eta\downarrow 0} \frac{R(F_\eta) - R(F_0)}{\eta},$$

ここで $F_\eta = (1-\eta)F_0 + \eta F_1$ である．方向微分は次のようになる．

$$R'_{F_0}(F_1) = \int_{\mathcal{X}} D_U(p(y|x), p(y;F_0))dF_1(x) - R(F_0).$$

Proof. まず η が微小だとして $R(F_\eta)$ を書き下す．

$$\begin{aligned}R(F_\eta) &= R(F_0)\\ &\quad + \eta\int_{\mathcal{X}} D_U(p(y|x), p(y;F_0))(dF_1(x) - dF_0(x))\\ &\quad + \eta\int_{\mathcal{X}} \partial_\eta H_U(p(y|x), p(y;F_\eta))\Big|_{\eta=0} dF_0(x) + O(\eta^2).\end{aligned}$$

ξ が $u=U'$ の逆関数であることを用いると次の関係が成り立つ．

$$\begin{aligned}
&\eta \int_{\Re} \partial_\eta H_U(p(y|x),p(y;F_\eta))\Big|_{\eta=0} dF_0(x)\\
=&\eta \int_{\Re} u\big(\xi(p(y;F_0))\big)\xi'(p(y;F_0))\big(p(y;F_1)-p(y;F_0)\big)dy\\
-&\eta \int_{\mathcal{X}}\int_{\Re} p(y|x)\xi'(p(y;F_0))\big(p(y;F_1)-p(y;F_0)\big)dydF_0(x)\\
=&0.
\end{aligned}$$

したがって，次の結果が得られる．

$$\begin{aligned}
R'_{F_0}(F_1) &= \lim_{\eta\downarrow 0}\frac{R(F_\eta)-R(F_0)}{\eta}\\
&= \int_{\mathcal{X}} D_U(p(y|x),p(y;F_0))dF_1(x) - R(F_0).
\end{aligned}$$

□

Corollary 2 (KKT 条件)．(8) 式の最適値を与える F を F^* と定義する．すなわち，$F^*=\arg\max_{F\in\mathcal{F}} R(F)$ である．このとき E^* を F^* の増加点の集合とする．このとき，以下の式が成り立つ．

$$D_U(p(y|x);p(y;F^*))\begin{cases} =C_U & \text{for } x\in E^*\\ \le C_U & \text{for } x\notin E^*\end{cases}$$

Proof. F^* における $F\in\mathcal{F}$ への方向微分は，$F\neq F^*$ ならば，どの方向に対しても $R'_{F^*}(F)<0$ となる．したがって，Lemma 2 から次の関係が成り立つ．

$$\int_{\Re} D_U(p(y|x),p(y;F^*))dF(x) < C_U, \quad \text{for} \quad F\neq F^*.$$

ここで $F\in\mathcal{F}$ が $\mathcal{F}$ の任意の確率測度であることから，$D_U(p(y|x),p(y;F^*))\le C_U$ が成り立つ．また，$x\in E^*$ で $D_U(p(y|x),p(y;F^*))\neq C_U$ とすると，方向微分の条件 $\int_{\Re} D_U(p(y|x);p(y;F^*))dF\le C_U$ と矛盾する． □

$\int_{\mathcal{X}} m(x)dF(x)\le A$ で示される F に関する条件を考えると，上の結果は以下のように書き直せる．

Corollary 3 (KKT 条件)．$\lambda>0$ としてラグランジュ未定定数法を用いると，

$$\begin{aligned}
C_U =& \max_{F\in\mathcal{F}}\Big[\int_{\Re} D_U(p(y|x);p(y;F))dF(x)\\
&-\lambda\Big(\int_{\Re} m(x)dF(x)-A\Big)\Big].
\end{aligned}$$

この関数の方向微分を考えると，KKT条件が求まる．すなわち，E^* を F^* の増加点の集合とする．最適な F^* に関して次の関係が成り立つ．

$$D_U(p(y|x);p(y;F^*))\begin{cases} =C_U+\lambda(m(x)-A) & \text{for } x\in E^*\\ \le C_U+\lambda(m(x)-A) & \text{for } x\notin E^*.\end{cases}$$

3.3 最適な入力分布が連続分布となる例

通信路容量，そして U-ダイバージェンスに拡張した C_U は X に関する制約と確率モデル $p(y|x)$ の組み合わせによって定まる量である．仮りにそうした最適値が求まったとして，それを達成する入力の分布 $F(X)$ がどのような分布となるかは興味ある問題である．我々は特に最適な分布が連続分布となる場合に興味がある．1 次元の X を入力とする現実的な通信路の場合で通信路容量を達成する分布が連続分布となることが知られているのは AWGN に対して平均パワーを制約した場合のみであろう．同様の例は U-ダイバージェンスに拡張した C_U では存在するのだろうか．

ここでは U-ダイバージェンスの例として (2) 式の β-ダイバージェンスを考える．C_U を達成る入力の分布が連続分布となる $p(y|x)$ と $m(x)$ の例を以下で示す．

Lemma 3. 次の正規分布を考える．

$$p(y|x)=\frac{1}{\sqrt{2\pi\sigma_N^2}}\exp\Big(-\frac{(y-x)^2}{2\sigma_N^2}\Big). \tag{12}$$

また，X に関する以下の制約を満たす F を考える．

$$\mathcal{F}=\Big\{F \;\Big|\; \int_{\Re} m(x)dF(x)\le A\Big\} \tag{13}$$

$$\begin{aligned}
m(x)=&-\sqrt{\frac{(\beta+1)(\sigma_N^2+\sigma_s^2)}{(\beta+1)\sigma_N^2+\sigma_s^2}}\exp\Big(\frac{-\beta x^2}{2((\beta+1)\sigma_N^2+\sigma_s^2)}\Big),\\
A=&-\frac{1}{\sqrt{\beta+1}}.
\end{aligned}$$

(13) の条件の下，次の結果が得られる．

$$\begin{aligned}
C_U &= \max_{F\in\mathcal{F}}\int_{\Re} D_U(p(y|x);p(y;F))dF(x)\\
&= \frac{1}{(2\pi)^{\beta/2}\beta(\beta+1)^{3/2}}\Big(\frac{1}{(\sigma_N^2)^{\beta/2}}-\frac{1}{(\sigma_N^2+\sigma_S^2)^{\beta/2}}\Big).
\end{aligned}$$

また，このとき最適値を与える F^* は次のようになる．

$$F^*(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi\sigma_S^2}}\exp\Big(-\frac{t^2}{2\sigma_S^2}\Big)dt. \tag{14}$$

したがって，最適な X の分布は正規分布である．

Proof. (12) 式の $p(y|x)$ に対して (14) の F を X の分布として $D_U(p(y|x);p(y;F))$ を計算する．

$$\begin{aligned}
&D_U(p(y|x);p(y;F))\\
=&\frac{1}{(2\pi)^{\beta/2}\beta(\beta+1)^{3/2}}\Big(\frac{1}{(\sigma_N^2)^{\beta/2}}-\frac{1}{(\sigma_N^2+\sigma_S^2)^{\beta/2}}\Big)\\
&+\lambda(m(x)-A).
\end{aligned}$$

ただし $\lambda=1/(\beta(2\pi)^{\beta/2}(\sigma_N^2+\sigma_S^2)^{\beta/2})$ である．Corollary 3 の結果から (14) の F が最適であることが分る． □

この結果，U-ダイバージェンスを損失とするベイズリスクでも $p(y|x)$ が正規分布であるとき，ある制約の下では正規分布が最適な分布となることが示された．証明の手続きを見ると，通信路容量の場合と同様に他のモデルと条件下で連続分布が最適解を与える組み合わせを探すのは困難であるように見える．

4 まとめ

本稿では，通信路容量，あるいは K-L ダイバージェンスを損失とするベイズリスクの問題を拡張し，Bregman 情報量，より正確には U-ダイバージェンスを損失とするベイズリスクについて考察した．ベイズ統計の立場からみれば，これは reference 事前分布の拡張になっている．

3.3 節では U-ダイバージェンスのひとつである β-ダイバージェンスを例にとり，X の最適な分布が連続分布となる場合について調べた．その結果，確率モデルを正規分布にとり，正規分布の形をした関数の期待値によって制約をつけた場合に最適な分布が正規分布となることを示した．β-ダイバージェンスは Tsallis 統計と結びついている．Tsallis 統計では一般化エントロピー (Tsallis エントロピー) の最大化の結果，q-正規分布と呼ばれる分布が現れ，これが重要な役割を果している．しかし，β-ダイバージェンスの最大化では q-正規分布が表われるような確率モデルと制約の組み合わせは発見できなかった．

謝辞

本研究は日本学術振興会の科研費補助金 基盤 C 24560490 および 基盤 C 24500018 の助成を受けたものである．

参考文献

[1] I. C. Abou-Faycal, M. D. Trott, and S. Shamai(Shitz). The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inf. Theory*, 47(4), pp. 1290–1301, 2001.

[2] J. Aitchison. Goodness of prediction fit. *Biometrika*, 62(3), pp. 547–554, 1975.

[3] J. O. Berger and J. M. Bernardo. On the development of reference priors. *Bayesian Statistics*, 4, pp. 35–60, 1992.

[4] J. M. Bernardo. Reference posterior distributions for Bayesian inference. *J. R. Stat. Soc., B*, 41(2), pp. 113–147, 1979.

[5] T. H. Chan, S. Hranilovic, and F. R. Kschischang. Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs. *IEEE trans. Inf. Theory*, 51(6), pp. 2073–2088, 2005.

[6] P. D. Grünwald and A. P. Dawid. Game thoery, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Stat.*, 32(4), pp. 1367–1433, 2004.

[7] M. C. Gursoy, V. Poor, and S. Verdú. The noncoherent Rician fading channel–part I: Structure of the capacity-achieving input. *IEEE trans. on Wireless Communi.*, 4(5), pp. 2193–2206, 2005.

[8] D. Haussler. A general minimax result for relative entropy. *IEEE Trans. Inf. Theory*, 43(4), pp. 1276–1280, 1997.

[9] 池田. 通信路容量と確率測度の最適化–最適な変調方式のために–. 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, 5(3), pp. 230–238, 2012.

[10] 池田, 林, 田中. 最大出力制約下での通信路容量と変調方式. SITA 2010 予稿集. 第 33 回情報理論とその応用シンポジウム, 2010.

[11] S. Ikeda and J. H. Manton. Capacity of a single spiking neuron channel. *Neural Comput.*, 21(6), pp. 1714–1748, 2009.

[12] D. G. Luenberger. *Optimization by Vector Space Method.* John Wiley & Sons, Inc., 1969.

[13] M. Minami and S. Eguchi. Robust blind source separation by beta divergence. *Neural Comput.*, 14(8), pp. 1859–1886, 2002.

[14] N. Murata and Y. Fujimoto. Bregman divergence and density integration. *J. of Math-for-industry*, 1(2009B-3), pp. 97–104, 2009.

[15] N. Murata, T. Takenouchi, T. Kanamori, and S. Eguchi. Information geometry of u-boost and Bregman divergence. *Neural Comput.*, 16(7), pp. 1437–1481, 2004.

[16] A. Ohara and T. Wada. Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations. *J. of Phys. A: Math. and Theoretical*, 43(3), p. 035002, 2010.

[17] C. E. Shannon. A mathematical theory of communication. *The Bell Sys. Tech. J.*, 27, pp. 379–423 and 623–656, 1948.

[18] J. G. Smith. The information capacity of amplitude- and variance-constrained scalar Gaussian channels. *Inform. and Control*, 18, pp. 203–219, 1971.

[19] T. Takenouchi and S. Eguchi. Robustifying adaboost by adding the naive error rate. *Neural Comput.*, 16(4), pp. 767–787, 2004.