

DeRobertis 分離度による全変動距離の上界

小林 景[†]

(受付 2011 年 2 月 23 日; 改訂 4 月 26 日; 採択 6 月 22 日)

要 旨

DeRobertis 分離度は確率測度間の差異を測る指標の一つであり、ベイズ事後分布や、分配関数の計算が困難な確率測度間の差異を調べる際に、応用上有効である。一方、密度関数の対が与えられたとき、この DeRobertis 分離度を用いて全変動距離の上界が得られることが知られている。本研究ノートでは、既存の上界を優越するような厳密な上界を導出し、さらにそれが最適であることを証明する。また、局所的な測度の違いを測る局所 DeRobertis 分離度を用いて得られる全変動距離の上界についても、既存のものを本質的に優越するような上界を示す。

キーワード: 全変動距離, ロバスト統計, ベイズ事後分布, ダイバージェンス, DeRobertis 分離度, 局所 DeRobertis 分離度.

1. 研究の背景 ~ DeRobertis 分離度と全変動距離

測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ 上の確率密度関数全体の集合を \mathcal{P} とする。このとき、 $f, g \in \mathcal{P}$ 間の DeRobertis 分離度 $d_{\text{DR}}: \mathcal{P} \times \mathcal{P} \rightarrow [0, \infty]$ を

$$d_{\text{DR}}(f, g) := \sup_{x \in \mathcal{X}} \sup_{x' \in \mathcal{X}} \frac{f(x) g(x')}{g(x) f(x')} - 1$$

で定義する。この分離度は

$$d_{\text{DR}}(f, g) = 0 \Leftrightarrow f = g \text{ } \mu\text{-a.e.}$$

を満たすが、三角不等式を満たさないため距離ではない。一方、スケール変換した

$$\tilde{d}_{\text{DR}} := \log(d_{\text{DR}} + 1)$$

は距離となるので、必要によりこちらに変換してから使用すればよい。

DeRobertis 分離度は DeRobertis (1978) で提案され、応用上重要な特徴として、以下に述べるように、ベイズ事後確率に関する分離度が事前分布のみによるという点と、分配関数が計算出来ないときにも、それを用いずに計算できるという点があげられる。なお、例の中に出てくる積分に関しては可積分性が保証されているものとする。

例 1.1. (ベイズ事後確率) 事前密度関数 π_1, π_2 に対するベイズ事後分布測度間の DeRobertis 分離度は、 $Z_1(f_n, x^n, \pi_1) := \int f_n(x^n | \theta) \pi_1(\theta) d\theta$, $Z_2(f_n, x^n, \pi_2) := \int f_n(x^n | \theta) \pi_2(\theta) d\theta$ を用いて、

$$d_{\text{DR}}(\pi_1(\cdot | x^n), \pi_2(\cdot | x^n)) = \sup_{\theta} \frac{f_n(x^n | \theta) \pi_1(\theta) / Z_1(f_n, x^n, \pi_1)}{f_n(x^n | \theta) \pi_2(\theta) / Z_2(f_n, x^n, \pi_2)} \sup_{\theta'} \frac{f_n(x^n | \theta') \pi_2(\theta') / Z_2(f_n, x^n, \pi_2)}{f_n(x^n | \theta') \pi_1(\theta') / Z_1(f_n, x^n, \pi_1)} - 1$$

[†] 統計数理研究所: 〒190-8562 東京都立川市緑町 10-3

$$= \sup_{\theta} \frac{\pi_1(\theta)}{\pi_2(\theta)} \sup_{\theta'} \frac{\pi_2(\theta')}{\pi_1(\theta')} - 1$$

と表される。よって、ベイズ事後分布測度間の DeRobertis 分離度は事前分布のみに依存する。これは標本や尤度関数によらないロバストな事後分布の理論評価をするときに有効な性質である。

例 1.2. (分配関数未知の場合) 測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ 上の可測関数 H_f, H_g に関して、

$$f(x) := \frac{\exp(H_f(x))}{\int \exp(H_f(x)) d\mu},$$

$$g(x) := \frac{\exp(H_g(x))}{\int \exp(H_g(x)) d\mu}$$

とすると、 f と g 間の DeRobertis 分離度は

$$d_{\text{DR}}(f, g) = \sup_{x \in \mathcal{X}} \frac{\exp(H_f(x))}{\exp(H_g(x))} \sup_{x \in \mathcal{X}} \frac{\exp(H_g(x))}{\exp(H_f(x))} - 1$$

と表される。よって、分配関数 $\int \exp(H_f(x)) d\mu, \int \exp(H_g(x)) d\mu$ を計算することが困難な場合にも、DeRobertis 分離度を計算することができる。

最後に、本論文で主に扱う DeRobertis 分離度の性質として、全変動距離の上界(文末補足 A)になるということがあげられる。ただし、全変動距離としては以下の定義を用いる。

$$d_{\text{TV}}(f, g) := \int_{\mathcal{X}} |f(x) - g(x)| d\mu.$$

このとき、任意の確率密度関数 f, g に関して、次の不等式が成立する。

$$(1.1) \quad d_{\text{TV}}(f, g) \leq d_{\text{DR}}(f, g).$$

f と g の零点集合が一致しない場合は、右辺が発散するためこの不等式は明らかである。それ以外の場合でも、 $\sup_x \frac{f(x)}{g(x)} \geq 1, \sup_x \frac{g(x)}{f(x)} \geq 1$ を用いると次のようにして容易に確認できる。

$$\begin{aligned} d_{\text{TV}}(f, g) &= \int |f(x) - g(x)| d\mu \\ &= \int_{\{x|f(x) > g(x)\}} \left(\frac{f(x)}{g(x)} - 1 \right) g(x) d\mu + \int_{\{x|f(x) < g(x)\}} \left(\frac{g(x)}{f(x)} - 1 \right) f(x) d\mu \\ &\leq \left(\sup_x \frac{f(x)}{g(x)} \sup_{x'} \frac{g(x')}{f(x')} - 1 \right) \left(\int_{\{x|f(x) > g(x)\}} g(x) d\mu + \int_{\{x|f(x) < g(x)\}} f(x) d\mu \right) \\ &\leq \left(\sup_x \frac{f(x)}{g(x)} \sup_{x'} \frac{g(x')}{f(x')} - 1 \right) = d_{\text{DR}}(f, g). \end{aligned}$$

実際、Smith and Rigat (2009) では不等式(1.1)を用いて、二つの異なる事前分布に関するベイズ事後確率測度間の全変動の上界を求め、その上界をベイズ推定のロバスト性の指標として用いている。

しかし、Smith and Rigat (2009) では、不等式(1.1)の厳密(tight)さについては全く評価されていない。本論文では、まず2節で、この不等式より厳密な不等式を導出し、さらにそれがより厳密な不等式を持たない最適なものであることを示す。次に、3節では、Smith and Rigat (2009) で導入された局所 DeRobertis 分離度を用いた全変動の上界についても、より厳密な不等式を導出する。

2. DeRobertis 分離度による全変動距離の厳密上界

本節では、測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ 上の確率密度関数 f, g に対して、DeRobertis 分離度による全変動距離の上界 (1.1) より厳密な上界を導出する。また、さらにそれが最適な上界であることを示す。

定理 2.1.

$$(2.1) \quad d_{\text{TV}} \leq 2(1 + (d_{\text{DR}} + 1)^{-1/2})^{-1}(1 - (d_{\text{DR}} + 1)^{-1/2}).$$

よって,

$$(2.2) \quad d_{\text{TV}} \leq \frac{1}{2}d_{\text{DR}}.$$

証明. f, g の零点集合が共通でないときや, $f_g := \sup_x \frac{f(x)}{g(x)}$ もしくは $g_f := \sup_x \frac{g(x)}{f(x)}$ が発散するときは, (2.1) および (2.2) の右辺が発散するため定理の主張は成立する. よって以後は f と g の零点は共通であり, かつ $\sup_x \frac{f(x)}{g(x)} < \infty$ および $\sup_x \frac{g(x)}{f(x)} < \infty$ を仮定する.

$A := \{x | f(x) > g(x)\}$, $F(A) := \int_A f(x)d\mu$, $G(A) := \int_A g(x)d\mu$ と定義すると,

$$(2.3) \quad \begin{aligned} \frac{1}{2}d_{\text{TV}} &= F(A) - G(A) \\ &\leq G(A) \left(\sup_x \frac{f(x)}{g(x)} - 1 \right) = G(A)(f_g - 1), \end{aligned}$$

$$(2.4) \quad \begin{aligned} \frac{1}{2}d_{\text{TV}} &= G(A^c) - F(A^c) \leq G(A^c) \left(1 - \left(\inf_x \frac{f(x)}{g(x)} \right) \right) \\ &= (1 - G(A)) \left(1 - \left(\sup_x \frac{g(x)}{f(x)} \right)^{-1} \right) = (1 - G(A))(1 - g_f^{-1}). \end{aligned}$$

ただし, A^c は A の補集合 $\mathcal{X} \setminus A$ を表すとする. ここで, $G(A) \leq \frac{1-g_f^{-1}}{f_g-g_f^{-1}}$ のとき不等式 (2.3) を, $G(A) \geq \frac{1-g_f^{-1}}{f_g-g_f^{-1}}$ のとき不等式 (2.4) を用いて,

$$\begin{aligned} \frac{1}{2}d_{\text{TV}} &\leq \frac{(1-g_f^{-1})(f_g-1)}{f_g-g_f^{-1}} = \frac{(g_f-1)(f_g-1)}{f_g g_f - 1} \\ &\leq \frac{((f_g g_f)^{1/2} - 1)^2}{f_g g_f - 1} = \frac{1 - (f_g g_f)^{-1/2}}{1 + (f_g g_f)^{-1/2}}. \end{aligned}$$

これより (2.1) が従う.

また, (2.2) は, $t \in (0, \infty)$ に対して次の式が成り立つことから従う.

$$(2.5) \quad \frac{1 - (t+1)^{-1/2}}{1 + (t+1)^{-1/2}} = \frac{t+2 - 2(t+1)^{1/2}}{t} \leq \frac{t}{4}. \quad \square$$

定理 2.1 の不等式 (2.1) による全変動距離の上界は最適なものである. つまり, 任意の $d_{\text{TV}} \in [0, 2]$ の値に対して, (2.1) の等号を満たすような d_{TV} および d_{DR} を持つ f, g の組が存在する. そのことを確認するために, まず (0,1) 上の確率測度について以下の例をみる.

例 2.2. μ を $(0,1)$ 上のルベグ測度とする. $d_{DR} = \infty$ の時は, サポートが互いに素な f, g を選べば $d_{TV} = 2$ となり (2.1) の等号が成立するので, 以下では $d_{DR} < \infty$ とする. この時, $a \in [1, \infty)$ に対して, 以下の f, g は定理の条件を満たす.

$$f(x) = \begin{cases} a & \left(x \in \left(0, \frac{1}{a+1}\right]\right), \\ a^{-1} & \left(x \in \left(\frac{1}{a+1}, 1\right)\right), \\ g(x) \equiv 1 & (x \in (0, 1)). \end{cases}$$

実際, このとき,

$$d_{TV} = \frac{2(a-1)}{a+1}, \quad d_{DR} = a^2 - 1$$

となり, (2.1) の等号を満たす.

次に, 測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ が以下の条件を満たすとする.

$$(2.6) \quad \mu(A), \mu(B) \in (0, \infty) \text{ かつ } A \cap B = \phi \text{ なる } A, B \in \mathcal{F} \text{ が存在する.}$$

これは非常に弱い条件であるが, この条件のもとで, 例 2.2 と本質的に同じ確率密度関数の組が一般の測度空間で構成でき, 次の定理が成立する.

定理 2.3. $(\mathcal{X}, \mathcal{F}, \mu)$ が条件 (2.6) を満たすとする. このとき, 定理 2.1 における全変動距離の DeRobertis 分離度による上界 (2.1) は最適である. つまり, 任意の $d_{TV} \in [0, 2]$ の値に対して, (2.1) の等号を満たすような d_{TV} および d_{DR} を持つ f, g の組が存在する. また, 不等式 (2.2) は d_{DR} の線形関数による上界として最適である.

証明. 測度の条件 (2.6) 中の $A, B \in \mathcal{F}$ と, $a \in (1, \infty)$ を用いて, μ 上の確率密度関数を以下のように定義する.

$$f(x) = \begin{cases} \frac{1}{\mu(A)} \frac{1}{a+1} & (\text{on } A), \\ \frac{1}{\mu(B)} \frac{a}{a+1} & (\text{on } B), \\ 0 & (\text{otherwise}), \end{cases}$$

$$g(x) = \begin{cases} \frac{1}{\mu(A)} \frac{a}{a+1} & (\text{on } A), \\ \frac{1}{\mu(B)} \frac{1}{a+1} & (\text{on } B), \\ 0 & (\text{otherwise}). \end{cases}$$

このとき, 例 2.2 と同一の d_{TV} と d_{DR} を持つことが確認でき, 不等式 (2.1) の最適性が示される.

また定理の後半は, 不等式 (2.5) の左辺が凹関数で, そのグラフが原点で右辺の線形関数のグラフと接することから従う. \square

3. 局所 DeRobertis 分離度による全変動距離の厳密上界

局所 DeRobertis 分離度は, 上記 DeRobertis 分離度を部分空間に制限したものであり, Smith and Rigat (2009) により提案された. 具体的には, 測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ 上の確率密度関数 f, g と

可測部分集合 $A \subset \mathcal{X}$ に対して,

$$d_{\text{DR}}^A(f, g) := \sup_{x \in A} \sup_{x' \in A} \frac{f(x)g(x')}{g(x)f(x')} - 1$$

で定義する. また, 全変動距離に対しても, 可測部分集合 A に制限したものを以下のように定義する.

$$d_{\text{TV}}^A(f, g) := \int_A |f - g| d\mu$$

このとき, $A \subset B \subset \mathcal{X}$ に対して,

$$d_{\text{DR}}^A(f, g) \leq d_{\text{DR}}^B(f, g)$$

および

$$d_{\text{TV}}^A(f, g) \leq d_{\text{TV}}^B(f, g)$$

が成り立つことは明らかである.

一方, A 上で一定の値をとる任意の f, g に関して $d_{\text{DR}}^A(f, g) = 0$ となることからわかるように, d_{TV}^A の自明でない上界を d_{DR}^A を用いて表すことは一般にはできない. しかし, d_{DR}^A を用いて, d_{TV} の上界を改良することは可能である. 実際, Smith and Rigat (2009) では任意の可測集合 $A \subset \mathcal{X}$ に関して,

$$(3.1) \quad d_{\text{TV}} \leq 2F(A^c)d_{\text{DR}} + F(A)d_{\text{DR}}^A$$

のように全変動距離の上界が導出された. この上界は, 密度関数の変動が小さく, かつ十分 $F(A)$ が大きいような A をうまく取れば, 大域的な DeRobertis 分離度 d_{DR} を用いた結果を改良し, 意味がある. 実際彼らは, ベイズ事後確率測度間の全変動距離に対して, サンプル数に依存させて A の領域を大きくしていき, その上界を実データ解析に応用している.

本節では, 不等式 (3.1) より厳密な上界を求める. これは, 2 節の結果の適用による右辺第二項の厳密化以外にも, 右辺第一項に関して $F(A^c)d_{\text{DR}}$ の値が大きいときに本質的な改良になっている.

定理 3.1. 測度空間 $(\mathcal{X}, \mathcal{F}, \mu)$ 上の任意の確率密度関数 f, g および可測集合 $A \subset \mathcal{X}$ に関して,

$$d_{\text{TV}} \leq 2 \frac{F(A^c)d_{\text{DR}}}{F(A^c)d_{\text{DR}} + 1} + 2F(A) \frac{1 - (d_{\text{DR}}^A + 1)^{-1/2}}{1 + (d_{\text{DR}}^A + 1)^{-1/2}}.$$

定理の証明の前に, 大域的な DeRobertis 分離度に関して以下の補題を示す.

補題 3.2. $S = S(f, g) := \{x \in \mathcal{X} \mid g(x) \geq f(x)\}$ としたとき,

$$d_{\text{TV}} \leq 2 \frac{(1 - F(S))F(S)d_{\text{DR}}}{F(S)d_{\text{DR}} + 1}.$$

補題 3.2 の証明. 定理 2.1 の証明と同様に, f と g の零点は共通であり, $f_g := \sup_{x \in \mathcal{X}} \frac{f(x)}{g(x)} < \infty$ および $g_f := \sup_{x \in \mathcal{X}} \frac{g(x)}{f(x)} < \infty$ としてよい. このとき, $d_{\text{DR}} = f_g g_f - 1$ に注意すると, 以下が成立する.

$$(3.2) \quad \begin{aligned} \frac{1}{2} d_{\text{TV}} &= \int_{\{x \mid g(x) > f(x)\}} g(x) - f(x) d\mu \\ &\leq F(S)(g_f - 1) = F(S)\{(d_{\text{DR}} + 1)f_g^{-1} - 1\}, \end{aligned}$$

$$(3.3) \quad \frac{1}{2} d_{\text{TV}} = \int_{\{x \mid f(x) > g(x)\}} f(x) - g(x) d\mu$$

$$\leq (1 - F(S))(1 - f_g^{-1}).$$

定理 2.1 の証明と同様にして, $f_g \geq F(s)d_{\text{DR}} + 1$ のとき不等式 (3.2) を, $f_g \leq F(s)d_{\text{DR}} + 1$ のとき (3.3) を用いることにより, 補題が従う. \square

定理 3.1 の証明. まず, $F(A) \geq G(A)$ のときを考える. このとき, $F(A^c) \leq G(A^c)$ であることから,

$$F(A^c) + F(B) = G(A^c) + G(B)$$

かつ B 上 $f \geq g$ となるような $B \subset A$ がとれる (文末補足 B).

このような B に対して,

$$(3.4) \quad d_{\text{TV}}^{A^c \cup B} \leq 2 \frac{F(A^c)d_{\text{DR}}^{A^c \cup B}}{F(A^c)d_{\text{DR}}^{A^c \cup B} + 1}$$

が成り立つことを以下のようにして示すことができる. まず,

$$\begin{aligned} \tilde{A}^c &:= A^c \cup \{x \in \mathcal{X} \mid g(x) \geq f(x)\} \\ \tilde{B} &:= (B \cup A^c) \setminus \tilde{A}^c \end{aligned}$$

と定義すると, $A^c \cup B = \tilde{A}^c \cup \tilde{B}$ かつ \tilde{B} 上で $f \geq g$, \tilde{A}^c 上で $g \geq f$ となる. これより, $\tilde{f} := f/F(\tilde{A}^c \cup \tilde{B})$ および $\tilde{g} := g/G(\tilde{A}^c \cup \tilde{B}) = g/F(\tilde{A}^c \cup \tilde{B})$ に対して補題 3.2 を適用すると,

$$\begin{aligned} (3.5) \quad d_{\text{TV}}^{A^c \cup B} &\leq \frac{2 \left(1 - \frac{F(\tilde{A}^c)}{F(\tilde{A}^c \cup \tilde{B})}\right) \frac{F(\tilde{A}^c)}{F(\tilde{A}^c \cup \tilde{B})} d_{\text{DR}}^{A^c \cup B}}{\frac{F(\tilde{A}^c)}{F(\tilde{A}^c \cup \tilde{B})} d_{\text{DR}}^{A^c \cup B} + 1} F(\tilde{A}^c \cup \tilde{B}) \\ &= 2 \frac{F(\tilde{B})F(\tilde{A}^c)d_{\text{DR}}^{A^c \cup B}}{F(\tilde{A}^c)d_{\text{DR}}^{A^c \cup B} + F(\tilde{A}^c) + F(\tilde{B})} \\ &\stackrel{(*)1}{\leq} 2 \frac{(1 - F(\tilde{A}^c))F(\tilde{A}^c)d_{\text{DR}}^{A^c \cup B}}{F(\tilde{A}^c)d_{\text{DR}}^{A^c \cup B} + 1} \leq 2 \frac{F(\tilde{A}^c)d_{\text{DR}}^{A^c \cup B}}{F(\tilde{A}^c)d_{\text{DR}}^{A^c \cup B} + 1} \\ &\stackrel{(*)2}{\leq} 2 \frac{F(A^c)d_{\text{DR}}^{A^c \cup B}}{F(A^c)d_{\text{DR}}^{A^c \cup B} + 1}. \end{aligned}$$

ただし, 不等式 (*1) では左辺が $F(\tilde{B})$ に関する非減少関数であることを, 不等式 (*2) では左辺が $F(\tilde{A}^c)$ に関する非減少関数であることを用いた.

また, 不等式 (3.5) の最右辺は $F(A^c)$ の増加関数であり, さらに今 $F(A) \geq G(A)$ を仮定しているので,

$$d_{\text{TV}}^{A^c \cup B} \leq 2 \frac{G(A^c)d_{\text{DR}}^{A^c \cup B}}{G(A^c)d_{\text{DR}}^{A^c \cup B} + 1}$$

が従う. これより, 逆に $G(A) \geq F(A)$ の場合も, f と g を交換した後に上記の議論をなぞることにより, 不等式 (3.4) が示される.

以上と定理 2.1 の結果より,

$$\begin{aligned} d_{\text{TV}} &:= d_{\text{TV}}^{A^c \cup B} + d_{\text{TV}}^{A \setminus B} \\ &\leq 2 \frac{F(A^c)d_{\text{DR}}^{A^c \cup B}}{F(A^c)d_{\text{DR}}^{A^c \cup B} + 1} + 2F(A \setminus B) \frac{1 - (d_{\text{DR}}^{A \setminus B} + 1)^{-1/2}}{1 + (d_{\text{DR}}^{A \setminus B} + 1)^{-1/2}} \\ &\leq 2 \frac{F(A^c)d_{\text{DR}}}{F(A^c)d_{\text{DR}} + 1} + 2F(A) \frac{1 - (d_{\text{DR}}^A + 1)^{-1/2}}{1 + (d_{\text{DR}}^A + 1)^{-1/2}}. \end{aligned}$$

よって, 定理が示された. \square

4. 考察

全変動距離は他の Hellinger 距離, Prokhorov 距離, discrepancy 距離等, 距離やダイバージェンスの上界として用いられる (例えば Gibbs and Su, 2002, 参照). よって, 2 節での結果は, 間接的にそれらに対しても DeRobertis 分離度による上界の式を与える. また, そのようにして得られた上界の厳密性を評価する際に, 本論文で不等式の等号成立例を具体的に示したことは意味がある.

3 節で導出した局所 DeRobertis 分離度に関する上界は, Smith and Rigat (2009) において提案された上界と比べ, 単に厳密なだけでなく, $F(A^\circ)d_{DR}$ の値が大きいときには本質的な改良となっている. さらに, 局所 DeRobertis 分離度を用いた上界においては, 証明中に用いられた集合 B が果たす役割が本質的であるが, Smith and Rigat (2009) の証明ではそれが分かりづらい.

しかし, 定理 3.1 での上界においても, 定理 2.3 において大域的な DeRobertis 分離度による上界に対して示されたような厳密さは持たない. この意味で厳密な局所 DeRobertis 分離度による全変動距離の上界については, 未解決である.

また, 一般の測度空間に対する定理 2.3 の前に, 例 2.2 において $(0,1)$ 上のルベグ測度に関する確率密度関数の例を示した. 実は, Hardy et al. (1988) で紹介されている密度関数の rearrangement の技術を一般の測度空間に拡張することにより, 一般の測度空間上の確率密度関数の組 f, g に対して, $(0,1)$ 上のルベグ測度に関する確率密度関数の組 \tilde{f}, \tilde{g} で, 全変動距離 d_{TV} と DeRobertis 分離度 d_{DR} および局所 DeRobertis 分離度を保存するようなものを構成することができる. よって, 一般の測度空間に対する DeRobertis 分離度と全変動距離の上界の議論は, 本質的に $(0,1)$ 上のルベグ測度に対する議論に帰着させられることを付記しておく.

補 足

A. 用語「上界」について

ここで, 各密度関数対 (f, g) に関して, $\psi(f, g) \leq \psi'(f, g)$ が成り立つとき, 汎関数 ψ' を汎関数 ψ の「上界」と呼ぶ. この「上界」の用法は数学本来の定義とは異なるが, 応用分野では慣例的に用いられるため, 本論文ではこの意味で用いる. あえて言うなら汎関数空間に上記の意味で半順序を導入したときに ψ' は $\{\psi\}$ の上界である.

B. 集合 B の存在

このような B がとれるためには, $\int_{\{x|f-g \geq \alpha\}} (f-g)d\mu = G(A^\circ) - F(A^\circ) = F(A) - G(A)$ なる $\alpha \in [0, \infty)$ が存在することが十分であるが, これは基準測度 μ が atomic であるときには必ずしも成り立たない. しかしその場合であっても, B の「境界の適当な重み付け」 B_t を以下のように形式的に導入し, B の代わりに用いることにより, 同様な議論が可能である. つまり, $\alpha' := \sup\{\alpha \in [0, \infty) \mid \int_{\{x|f-g \geq \alpha\}} (f-g)d\mu \geq F(A) - G(A)\}$ を用いて, $B := \{x \mid f-g \geq \alpha'\}$, $B^\circ := \{x \mid f-g > \alpha'\}$ かつ $\partial B := B \setminus B^\circ$ としたとき, 各 $t \in [0, 1]$ に対しての B_t 上の各種分値を例えば $F(B_t) = F(B^\circ) + tF(\partial B)$ かつ $F(B_t^c) = F(B^c) + (1-t)F(\partial B)$ のように計算すれば良い.

参 考 文 献

- DeRobertis, L. (1978). The use of partial prior knowledge in Bayesian inference, Ph.D. dissertation, Yale University, New Haven.
- Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics, *International Statistical Review*, **70**(3), 419–435.
- Hardy, G. H., Littlewood, J. E. and Pólya, G. (1988). *Inequalities*, Cambridge University Press, Cambridge.
- Smith, J. Q. and Rigat, F. (2009). Iseparation and robustness in finite parameter Bayesian inference, *Annals of the Institute of Statistical Mathematics* (to appear).

Upper Bounds on Total Variation Distance by DeRobertis Separation

Kei Kobayashi

The Institute of Statistical Mathematics

The DeRobertis separation is a function of two probability distributions and it measures the difference between them. Because of some properties of the DeRobertis separation, it works well for Bayesian posterior distributions and probability distributions whose partial function is hard to compute. Given a pair of probability density functions, the DeRobertis separation is known to be an upper bound on the total variation distance, though the tightness of the bound has not been studied. In the letter, a tighter upper bound is derived and the bound is proved to be optimal. Furthermore, another upper bound is proved by using the local DeRobertis separation, which is essentially tighter than the known upper bounds.