2012年6月15日　統計数理研究所　オープンハウス

DeRobertis分離度による全変動距離の上界

小林　景　数理・推論研究系　助教

【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$d_{\mathrm{DR}}:\mathcal{P}\times\mathcal{P}\to[0,\infty]$は

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

で定義され（DeRobertis(1978)），以下のような利点をもつ．

例 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，

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

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 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，

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

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．

DeRobertis分離度は全変動距離$d_{\mathrm{TV}}(f,g):=\int_{\mathcal{X}}|f(x)-g(x)|d\mu$の上界になる: 任意の確率密度関数$f,g$に関して，

$$d_{\mathrm{TV}}(f,g)\le d_{\mathrm{DR}}(f,g).\qquad(1)$$

実際，Smith・Rigat(2012)では不等式(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・Rigat(2012)では，不等式(1)の厳密（tight）さについては全く評価されていない．本研究では，この不等式より厳密な不等式を導出し，さらにそれが最良なものであることを示す．

【DeRobertis分離度による全変動距離の厳密上界】

本節では，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)より厳密な，最良の上界を示す．

定理 3

$$d_{\mathrm{TV}}\le 2\left(1+(d_{\mathrm{DR}}+1)^{-1/2}\right)^{-1}\left(1-(d_{\mathrm{DR}}+1)^{-1/2}\right).\qquad(2)$$

よって，$d_{\mathrm{TV}}\le\frac{1}{2}d_{\mathrm{DR}}$.

証明の概略） $A:=\{x|f(x)>g(x)\}$, $F(A):=\int_A f(x)d\mu$, $G(A):=\int_A g(x)d\mu$と定義し，$f_g:=\sup_x\dfrac{f(x)}{g(x)}$と表記すると，

$$\frac{1}{2}d_{\mathrm{TV}}=F(A)-G(A)\le G(A)\left(\sup_x\frac{f(x)}{g(x)}-1\right)=G(A)\left(f_g-1\right),\qquad(3)$$

$$\begin{aligned}\frac{1}{2}d_{\mathrm{TV}}&=G(A^c)-F(A^c)\le G(A^c)\left(1-\left(\inf_x\frac{f(x)}{g(x)}\right)\right)\\&=(1-G(A))\left(1-\left(\sup_x\frac{g(x)}{f(x)}\right)^{-1}\right)=(1-G(A))\left(1-g_f^{-1}\right).\qquad(4)\end{aligned}$$

ここで，$G(A)\le\frac{1-g_f^{-1}}{f_g-g_f^{-1}}$のとき(3)を，$G(A)\ge\frac{1-g_f^{-1}}{f_g-g_f^{-1}}$のとき(4)を用いて，

$$\begin{aligned}\frac{1}{2}d_{\mathrm{TV}}&\le\frac{\left(1-g_f^{-1}\right)(f_g-1)}{f_g-g_f^{-1}}=\frac{(g_f-1)(f_g-1)}{f_gg_f-1}\\&\le\frac{\left((f_gg_f)^{1/2}-1\right)^2}{f_gg_f-1}=\frac{1-(f_gg_f)^{-1/2}}{1+(f_gg_f)^{-1/2}}.\end{aligned}$$

これより(2)が従う．

一方，$t\in(0,\infty)$に対して$\frac{1-(t+1)^{-1/2}}{1+(t+1)^{-1/2}}=\frac{t+2-2(t+1)^{1/2}}{t}\le\frac{t}{4}$.　（証明終）

定理3の不等式(2)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$d_{\mathrm{TV}}\in[0,2]$の値に対して，(2)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．一般の場合の証明は省略するが，$(0,1)$上の確率測度についての以下の例と同様に示される．

例 4 μを$(0,1)$上のルベーグ測度とする．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．

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

$$g(x)\equiv 1\quad(x\in(0,1)).$$

このとき，$d_{\mathrm{TV}}=\frac{2(a-1)}{a+1}$, $d_{\mathrm{DR}}=a^2-1$となり，(2)の等号を満たす．

【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・Rigat(2012)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，

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

で定義する．d^A_{DR}を用いて，d_{TV}の上界を改良することが可能である．Smith・Rigat(2012)では任意の可測集合$A\subset\mathcal{X}$に関して，

$$d_{\mathrm{TV}}\le 2F(A^c)d_{\mathrm{DR}}+F(A)d^A_{\mathrm{DR}}\qquad(5)$$

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．

以下の定理では，不等式(5)より厳密な上界を求める．これは，定理3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$F(A^c)d_{\mathrm{DR}}$の値が大きいときに本質的な改良になっている．

定理 5 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，

$$d_{\mathrm{TV}}\le 2\frac{F(A^c)d_{\mathrm{DR}}}{F(A^c)d_{\mathrm{DR}}+1}+2F(A)\frac{1-(d^A_{\mathrm{DR}}+1)^{-1/2}}{1+(d^A_{\mathrm{DR}}+1)^{-1/2}}.$$

References

[1] DeRobertis, L. (1978). The use of partial prior knowledge in Bayesian inference, Ph.D. dissertation, Yale University.

[2] Smith, J.Q. and Rigat, F. (2012). Isoseparation and Robustness in Finite Parameter Bayesian Inference, *AISM*, Vol.64 , 3, 495 -519.