α-Geodesical Skew Divergence

木村　正成　総合研究大学院大学　統計科学専攻　博士課程(5年一貫制)3年

1 Skew divergence

Definition 1. (Skew divergence) The skew divergence $D_S^{(\lambda)}[p\|q] : \mathcal{P}\times\mathcal{P}\to[0,\infty]$ is defined between two Radon–Nikodym densities p and q of μ-absolutely continuous probability measures by

$$D_S^{(\lambda)}[p\|q] \coloneqq D_{KL}[p\|(1-\lambda)p+\lambda q] = \int_{\mathcal{X}} p\ln\frac{p}{(1-\lambda)p+\lambda q}d\mu,$$

where $\lambda\in[0,1]$.

2 α-geodesic

Definition 2. (f-interpolation) For any $a,b,\in\mathbb{R}$, $\lambda\in[0,1]$ and $\alpha\in\mathbb{R}$, f-interpolation is defined as

$$m_f^{(\lambda,\alpha)}(a,b) = f_\alpha^{-1}\Big((1-\lambda)f_\alpha(a)+\lambda f_\alpha(b)\Big), \tag{1}$$

where

$$f_\alpha(x) = \begin{cases} x^{\frac{1-\alpha}{2}} & (\alpha\neq 1) \\ \ln x & (\alpha=1)\end{cases} \tag{2}$$

is the function that defines the f-mean.

The f-mean function satisfies

$$\lim_{\alpha\to\infty} f_\alpha(x) = \begin{cases}\infty & (|x|<1),\\ 1 & (|x|=1),\\ 0 & (|x|>1),\end{cases},\quad \lim_{\alpha\to-\infty} f_\alpha(x) = \begin{cases}0 & (|x|<1),\\ 1 & (|x|=1),\\ \infty & (|x|>1).\end{cases}$$

It is easy to see that this family includes various known weighted means including the e-mixture and m-mixture for $\alpha=\pm1$ in the literature of information geometry.

Definition 3. (α-representation) For some positive measure $m_i^{\frac{1-\alpha}{2}}$, the coordinate system $\boldsymbol{\theta}=(\theta^i)$ derived from the α-divergence is

$$\theta^i = m_i^{\frac{1-\alpha}{2}} = f_\alpha(m_i) \tag{3}$$

and θ^i is called the α-representation of a positive measure $m_i^{\frac{1-\alpha}{2}}$.

Definition 4. (α-geodesic) The α-geodesic connecting two probability vectors $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ is defined as

$$r_i(t) = c(t)f_\alpha^{-1}\Big\{(1-t)f_\alpha(p(x_i))+tf_\alpha(q(x_i))\Big\},\quad t\in[0,1] \tag{4}$$

where $c(t)$ is determined as $c(t)=\frac{1}{\sum_{i=1}^n r_i(t)}$.

3 α-geodesical skew divergence

Definition 5. (α-Geodesical Skew Divergence) The α-geodesical skew divergence $D_{GS}^{(\alpha,\lambda)} : \mathcal{P}\times\mathcal{P}\to[0,\infty]$ is defined between two Radon–Nikodym densities p and q of μ-absolutely continuous probability measures by:

$$\begin{aligned} D_{GS}^{(\alpha,\lambda)}\Big[p\|q\Big] &\coloneqq D_{KL}\Big[p\|m_f^{(\lambda,\alpha)}(p,q)\Big] \\ &= \int_{\mathcal{X}} p\ln\frac{p}{m_f^{(\lambda,\alpha)}(p,q)}d\mu, \end{aligned} \tag{5}$$

where $\alpha\in\mathbb{R}$ and $\lambda\in[0,1]$.

3.1 Properties

Proposition 6. (Non-negativity of the α-geodesical skew divergence) For $\alpha\geq-1$ and $\lambda\in[0,1]$, the α-geodesical skew divergence $D_{GS}^{(\alpha,\lambda)}[p\|q]$ satisfies the following inequality:

$$D_{GS}^{(\alpha,\lambda)}[p\|q]\geq 0. \tag{6}$$

Proposition 7. (Monotonicity of the α-geodesical skew divergence with respect to α) α-Geodesical skew divergence satisfies the following inequality for all $\alpha\in\mathbb{R},\lambda\in[0,1]$.

$$D_{GS}^{(\alpha,\lambda)}[p\|q]\geq D_{GS}^{(\alpha',\lambda)}[p\|q],\ (\alpha\geq\alpha').$$

Proposition 8. (Subadditivity of the α-geodesical skew divergence with respect to α) α-Geodesical skew divergence satisfies the following inequality for all $\alpha,\beta\in\mathbb{R},\lambda\in[0,1]$

Proposition 9. (Continuity of the α-geodesical skew divergence with respect to α and λ) α-Geodesical skew divergence has the continuity property.

Proposition 10. (Lower bound of the α-geodesical skew divergence) α-Geodesical skew divergence satisfies the following inequality for all $\alpha\in\mathbb{R},\lambda\in[0,1]$.

$$D_{GS}^{(\alpha,\lambda)}[p\|q]\geq\int_{\mathcal{X}} p\ln\frac{p}{\max\{p,q\}}d\mu. \tag{7}$$

Proposition 11. (Upper bound of the α-geodesical skew divergence) α-Geodesical skew divergence satisfies the following inequality for all $\alpha\in\mathbb{R},\lambda\in[0,1]$.

$$D_{GS}^{(\alpha,\lambda)}[p\|q]\leq\int_{\mathcal{X}} p\ln\frac{p}{\min\{p,q\}}d\mu. \tag{8}$$

Theorem 12. (Strong convexity of the α-geodesical skew divergence) α-Geodesical skew divergence $D_{GS}^{(\alpha,\lambda)}[p\|q]$ is strongly convex in p with respect to the total variation norm.

3.2 Function space

For an α-geodesical skew divergence $f_q^{(\alpha,\lambda)}(p) = D_{GS}^{(\alpha,\lambda)}[p\|q]$ with one side of the distribution fixed, let the entire set be

$$\mathcal{F}_q = \Big\{f_q^{(\alpha,\lambda)} \mid \alpha\in\mathbb{R},\lambda\in[0,1]\Big\}. \tag{9}$$

For $f_q^{(\alpha,\lambda)}\in\mathcal{F}_q$, its semi-norm is defined by

$$\Big\|f_q^{(\alpha,\lambda)}\Big\|_p \coloneqq \int_{\mathcal{X}}\Big(\Big|f_q^{(\alpha,\lambda)}\Big|^p d\mu\Big)^{\frac{1}{p}}. \tag{10}$$

Theorem 13. Let $\mathcal{N}$ be the kernel of $\|\cdot\|_p$ as follows:

$$\mathcal{N}\coloneqq ker(\|\cdot\|_p) = \Big\{f_q^{(\alpha,\lambda)} \mid f_q^{(\alpha,\lambda)}=0\Big\}. \tag{11}$$

Then the quotient space $\mathcal{V}\coloneqq(\mathcal{F}_q,\|\cdot\|_p)/N$ is a Banach space.