Fisher 情報計量の測地線と一般化平均

伊藤 光弘　(筑波大学 数理物質系)
佐藤 弘康　(日本工業大学 工学部)

2016 年 3 月 19 日

日本数学会 2016 年度年会

（筑波大学）

1. 正値確率測度の「平均」

- $\alpha \in \mathbb{R}$ に対し，写像 $\varphi^{(\alpha)} : \mathcal{P}^+(M) \times \mathcal{P}^+(M) \to \mathcal{P}^+(M)$ を

$$\varphi^{(\alpha)}(\mu_1, \mu_2) = \frac{1}{C} \left\{ 1 + \left(\frac{d\mu_2}{d\mu_1} \right)^{\alpha} \right\}^{1/\alpha} \mu_1$$

  と定める（ただし, C は確率測度となるための正規化定数）.
- $\varphi^{(\alpha)}(\mu_1, \mu_2)$ を μ_1 と μ_2 の**正規化 α-冪平均**とよぶ.

1. 正値確率測度の「平均」

- $\alpha \in \mathbb{R}$ に対し，写像 $\varphi^{(\alpha)} : \mathcal{P}^+(M) \times \mathcal{P}^+(M) \to \mathcal{P}^+(M)$ を

$$\varphi^{(\alpha)}(\mu_1, \mu_2) = \frac{1}{C} \left\{ 1 + \left(\frac{d\mu_2}{d\mu_1} \right)^{\alpha} \right\}^{1/\alpha} \mu_1$$

と定める（ただし, C は確率測度となるための正規化定数）.

- $\varphi^{(\alpha)}(\mu_1, \mu_2)$ を μ_1 と μ_2 の正規化 α-冪平均とよぶ.
- $\alpha = 1$ のときは算術平均, $\alpha = -1$ のときは調和平均とよばれる.
- $\alpha = 0$ のときは

$$\varphi^{(0)}(\mu_1, \mu_2) = \left(\int_M \sqrt{\frac{d\mu_2}{d\mu_1}}\, d\mu_1 \right)^{-1} \sqrt{\frac{d\mu_2}{d\mu_1}}\, \mu_1$$

となり, これを正規化幾何平均とよぶ.

2. Fisher 計量の測地線と幾何平均

- $\gamma(0) = \mu, \dot{\gamma}(0) = \tau$（ただし，$|\tau|_\mu = 1$）を満たす測地線 $\gamma(t)$ は

$$\gamma(t) = \left(\cos\frac{t}{2} + \frac{d\tau}{d\mu}\ \sin\frac{t}{2}\right)^2 \mu.$$

- $\gamma(\ell) = \mu'$ ならば，$\tau = \dfrac{1}{\sin(\ell/2)}\left(\sqrt{\dfrac{d\mu'}{d\mu}} - \cos\dfrac{\ell}{2}\right)\mu, \quad \ell = \ell(\mu, \mu')$.

2. Fisher 計量の測地線と幾何平均

- $\gamma(0) = \mu, \dot{\gamma}(0) = \tau$（ただし，$|\tau|_\mu = 1$）を満たす測地線 $\gamma(t)$ は

$$\gamma(t) = \left(\cos \frac{t}{2} + \frac{d\tau}{d\mu} \sin \frac{t}{2}\right)^2 \mu.$$

- $\gamma(\ell) = \mu'$ ならば，$\tau = \dfrac{1}{\tan(\ell/2)} \left(\varphi^{(0)}(\mu, \mu') - \mu\right), \quad \ell = \ell(\mu, \mu')$

2. Fisher 計量の測地線と幾何平均

- $\gamma(0) = \mu, \dot{\gamma}(0) = \tau$（ただし，$|\tau|_\mu = 1$）を満たす測地線 $\gamma(t)$ は

$$\gamma(t) = \left(\cos\frac{t}{2} + \frac{d\tau}{d\mu}\,\sin\frac{t}{2}\right)^2 \mu.$$

- $\gamma(\ell) = \mu'$ ならば，$\tau = \dfrac{1}{\tan(\ell/2)}\left(\varphi^{(0)}(\mu,\mu') - \mu\right), \quad \ell = \ell(\mu,\mu')$
- 2 点を結ぶ測地線は

$$\gamma(t) = a_1(t)\,\mu + a_2(t)\,\mu' + a_3(t)\,\varphi^{(0)}(\mu,\mu')$$

と表される（$a_i : [0,\ell] \to \mathbb{R}_{\geq 0}$ は，$\sum_i a_i(t) = 1$ を満たす）.

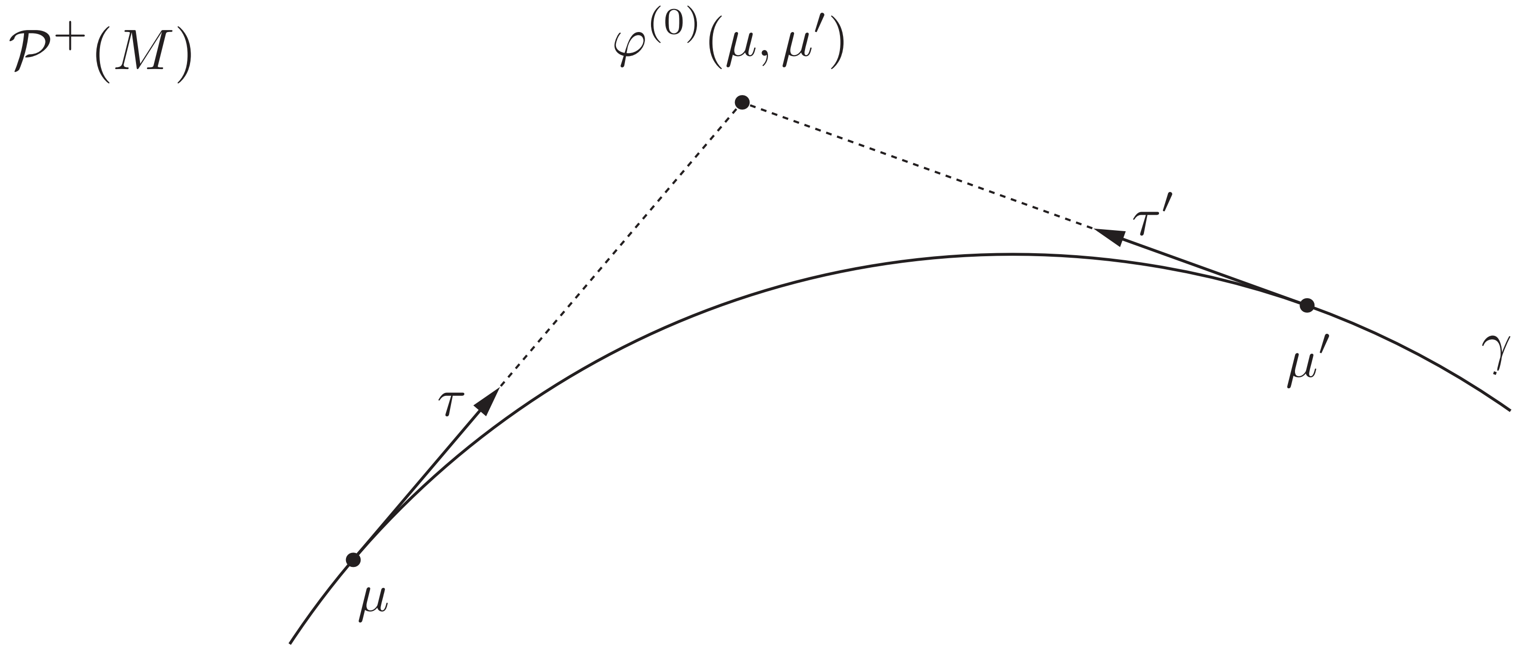
- 以上により，任意の 2 点は測地線で結ぶことができる ことがわかる.

2. Fisher 計量の測地線と幾何平均

- $\gamma(0) = \mu, \dot\gamma(0) = \tau$（ただし，$|\tau|_\mu = 1$）を満たす測地線 $\gamma(t)$ は

$$\gamma(t) = \left(\cos\frac{t}{2} + \frac{d\tau}{d\mu}\,\sin\frac{t}{2}\right)^2 \mu.$$

- $\gamma(\ell) = \mu'$ ならば，$\tau = \dfrac{1}{\tan(\ell/2)}\left(\varphi^{(0)}(\mu,\mu') - \mu\right)$, $\ell = \ell(\mu,\mu')$



3. 測地線分の中点

[5] A. Ohara, *Geodesics for dual connections and means on symmetric cones*, Integr. Equat. Oper. Th. **50** (2004), 537-548.

- (Ω, g) : **対称錐 Ω 上のあるポテンシャル関数に関するヘッセ計量** g
- g **に関する** <u>**双対接続構造** $(\nabla^{(\alpha)}, \nabla^{(-\alpha)})$</u>, $-1 \leq \alpha \leq 1$ **を定義.**
- $\nabla^{(\alpha)}$ **に関する** <u>**測地線分（α-測地線分）の中点**</u> **が端点の** <u>α**-冪平均**</u> **であることを示した.**
- **ここでの** α**-冪平均とは, 関数** $\sigma^{(\alpha)}_{1/2}(t) = \left(\dfrac{1+t^{\alpha}}{2}\right)^{1/\alpha}$ **が生成する Ω の作用素平均である.**

3. 測地線分の中点

[5] A. Ohara, *Geodesics for dual connections and means on symmetric cones*, Integr. Equat. Oper. Th. **50** (2004), 537-548.

- (Ω, g) : 対称錐 Ω 上のあるポテンシャル関数に関するヘッセ計量 g
- g に関する 双対接続構造 $(\nabla^{(\alpha)}, \nabla^{(-\alpha)})$, $-1 \le \alpha \le 1$ を定義.
- $\nabla^{(\alpha)}$ に関する 測地線分（α-測地線分）の中点 が端点の α-冪平均 であることを示した.
- ここでの α-冪平均とは, 関数 $\sigma^{(\alpha)}_{1/2}(t) = \left(\dfrac{1+t^{\alpha}}{2}\right)^{1/\alpha}$ が生成する Ω の作用素平均である.

問題

- $(\mathcal{P}^{+}(M), G)$ においても同様のことが成り立つだろうか？
- $(\mathcal{P}^{+}(M), G)$ 上の（標準的な）双対接続構造 $(\nabla^{(\alpha)}, \nabla^{(-\alpha)})$ は？

（再）2. Fisher 計量の測地線と幾何平均

- $\gamma(0) = \mu, \dot{\gamma}(0) = \tau$（ただし，$|\tau|_\mu = 1$）を満たす 0-測地線 $\gamma(t)$ は

$$\gamma(t) = \left(\cos\frac{t}{2} + \frac{d\tau}{d\mu}\ \sin\frac{t}{2}\right)^2 \mu.$$

- $\gamma(\ell) = \mu'$ ならば，$\tau = \dfrac{1}{\sin(\ell/2)}\left(\sqrt{\dfrac{d\mu'}{d\mu}} - \cos\dfrac{\ell}{2}\right)\mu, \quad \ell = \ell(\mu, \mu')$.

（再）2. Fisher 計量の測地線と幾何平均

- $\gamma(0)=\mu, \dot\gamma(0)=\tau$（ただし，$|\tau|_\mu=1$）を満たす 0-測地線 $\gamma(t)$ は

$$\gamma(t)=\left(\cos\frac{t}{2}+\frac{d\tau}{d\mu}\ \sin\frac{t}{2}\right)^2\mu.$$

- $\gamma(\ell)=\mu'$ ならば，$\tau=\dfrac{1}{\sin(\ell/2)}\left(\sqrt{\dfrac{d\mu'}{d\mu}}-\cos\dfrac{\ell}{2}\right)\mu, \quad \ell=\ell(\mu,\mu')$.

- 以上の式より，

$$\gamma(\ell/2)=\frac{\sin(\ell/4)}{\sin(\ell/2)}\left(1+\sqrt{\frac{d\mu'}{d\mu}}\right)^2\mu=\varphi^{(1/2)}(\mu,\mu')$$

を得る．つまり，Fisher 計量の測地線の中点は，(1/2)-冪平均である.

- 一般の α のときは？

4. ダイバージェンスと双対構造 (1)　（[1] を参照）

- 多様体 S 上の Riemann 計量 g と 2 つのアフィン接続 ∇, ∇^*
  - 任意のベクトル場 X, Y, Z に対して,

$$X\left(g(Y,Z)\right) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X, Z)$$

    を満たすとき, (g, ∇, ∇^*) を S 上の双対構造とよぶ.
  - 特に, ∇ と ∇^* がともに平坦接続のとき, 双対平坦構造とよぶ.

4. ダイバージェンスと双対構造 (1)　([1] を参照)

- 多様体 S 上の Riemann 計量 g と 2 つのアフィン接続 ∇, ∇^*
  - 任意のベクトル場 X, Y, Z に対して，

$$X\,(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X, Z)$$

を満たすとき，(g, ∇, ∇^*) を S 上の**双対構造**とよぶ.

  - 特に，∇ と ∇^* がともに平坦接続のとき，**双対平坦構造**とよぶ.

- $D(\cdot\|\cdot) : S \times S \to \mathbb{R}_{\geq 0}$ が「$D(p\|q) = 0 \iff p = q$」を満たすとき，D を S 上の**ダイバージェンス**とよぶ.

4. ダイバージェンスと双対構造 (1)　（[1] を参照）

- 多様体 S 上の Riemann 計量 g と 2 つのアファイン接続 ∇, ∇^*
  - 任意のベクトル場 X, Y, Z に対して,

$$X\left(g(Y,Z)\right) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X, Z)$$

を満たすとき，(g, ∇, ∇^*) を S 上の双対構造とよぶ.
  - 特に，∇ と ∇^* がともに平坦接続のとき，双対平坦構造とよぶ.

- $D(\cdot\|\cdot) : S \times S \to \mathbb{R}_{\geq 0}$ が「$D(p\|q) = 0 \iff p = q$」を満たすとき，D を S 上のダイバージェンスとよぶ.

- S が有限次元の場合,

（標準的な）ダイバージェンス D $\overset{\Longleftarrow}{\Longrightarrow}$ 双対（平坦）構造 (g, ∇, ∇^*)

4. ダイバージェンスと双対構造 (2)　([1] を参照)

- μ, μ' を確率測度とするとき,
  - f-ダイバージェンス D^f：凸関数 $f : \mathbb{R}_{>0} \to \mathbb{R}$, $f(1) = 0$ に対して,

$$D^f(\mu \| \mu') = \int_M f\left(\frac{d\mu'}{d\mu}\right) d\mu$$

  - α-ダイバージェンス $D^{(\alpha)}$：次の関数 $f^{(\alpha)}$ が定めるダイバージェンス；

$$f^{(\alpha)}(u) = \begin{cases} \frac{4}{1-\alpha^2}\left(1 - u^{(1+\alpha)/2}\right) & (\alpha \neq \pm 1) \\ u \log u & (\alpha = 1) \\ -\log u & (\alpha = -1) \end{cases}$$

  - $D^{(\alpha)}$ が誘導する接続 $\nabla^{(\alpha)}$ を α-接続とよぶ.

4. ダイバージェンスと双対構造 (2)　（[1] を参照）

- μ, μ' を確率測度とするとき，
  - f-ダイバージェンス D^f：凸関数 $f : \mathbb{R}_{>0} \to \mathbb{R}$, $f(1) = 0$ に対して，

$$D^f(\mu \| \mu') = \int_M f\left(\frac{d\mu'}{d\mu}\right) d\mu$$

  - α-ダイバージェンス $D^{(\alpha)}$：次の関数 $f^{(\alpha)}$ が定めるダイバージェンス；

$$f^{(\alpha)}(u) = \begin{cases} \frac{4}{1-\alpha^2}\left(1 - u^{(1+\alpha)/2}\right) & (\alpha \neq \pm 1) \\ u \log u & (\alpha = 1) \\ -\log u & (\alpha = -1) \end{cases}$$

  - $D^{(\alpha)}$ が誘導する接続 $\nabla^{(\alpha)}$ を α-接続とよぶ.
- $(G, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ は双対構造である.

[注] D^f が誘導する双対構造は，Fisher 計量（の定数倍）と α-接続である.

5. $\mathcal{P}^+(M)$ 上の双対（平坦）構造と測地線

- α-接続（[3]）：$\nabla^{(\alpha)}_{\tau_1}\tau_2(\mu) = -\dfrac{1+\alpha}{2}\left(\dfrac{d\tau_1}{d\mu}\dfrac{d\tau_2}{d\mu} - G_\mu(\tau_1,\tau_2)\right)\mu$
    - $\alpha = \pm 1$ のとき，$(\mathcal{P}^+(M), G, \nabla^{(+1)}, \nabla^{(-1)})$ は双対平坦である.

5. $\mathcal{P}^+(M)$ 上の双対（平坦）構造と測地線

- α-接続（[3]）：$\nabla^{(\alpha)}_{\tau_1}\tau_2(\mu) = -\dfrac{1+\alpha}{2}\left(\dfrac{d\tau_1}{d\mu}\dfrac{d\tau_2}{d\mu} - G_\mu(\tau_1,\tau_2)\right)\mu$
  - $\alpha = \pm 1$ のとき，$(\mathcal{P}^+(M), G, \nabla^{(+1)}, \nabla^{(-1)})$ は**双対平坦**である．
- $\nabla^{(\alpha)}$-測地線 $\gamma^{(\alpha)}(t) = p(t)\,d\theta$ は次の微分方程式の解；

$$\frac{d}{dt}\left(\frac{\dot p(t)}{p(t)}\right) + \frac{1-\alpha}{2}\left(\frac{\dot p(t)}{p(t)}\right)^2 + \frac{1+\alpha}{2}G_{\gamma^{(\alpha)}(t)}(\dot\gamma^{(\alpha)}(t), \dot\gamma^{(\alpha)}(t)) = 0$$

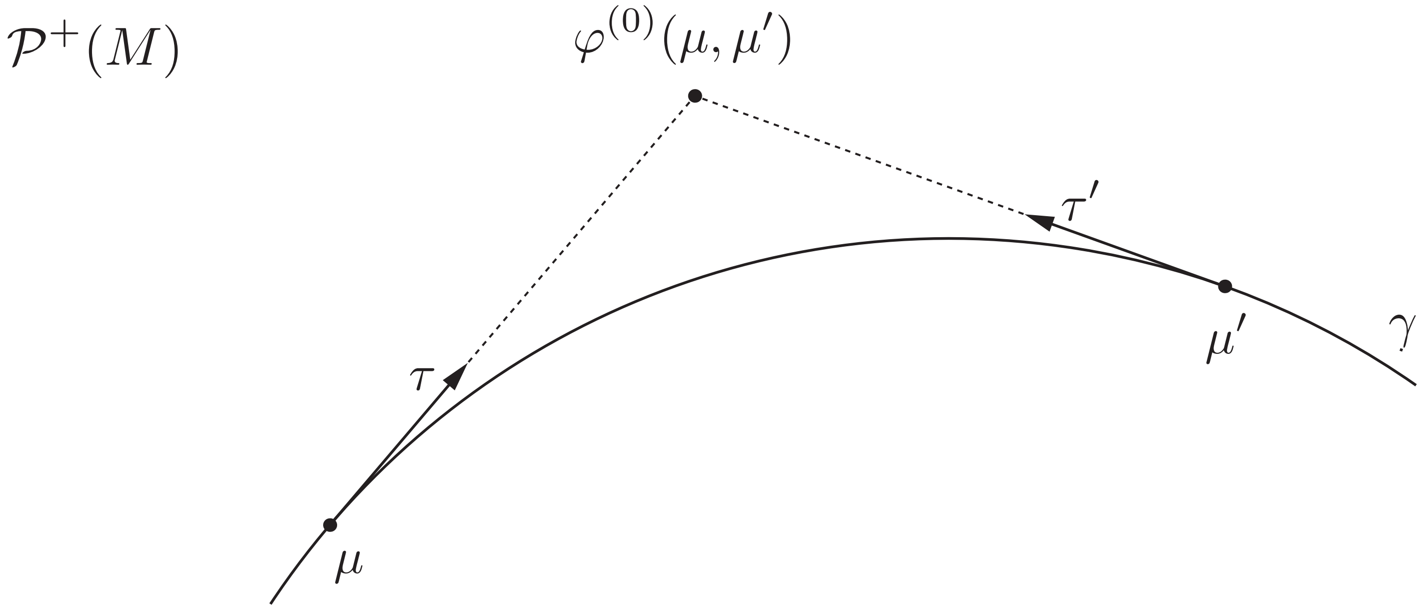
- 任意の $\mu, \mu' \in \mathcal{P}^+(M)$ を結ぶ $\nabla^{(\pm 1)}$-測地線 $\gamma^{(\pm 1)}(t)$ が存在；

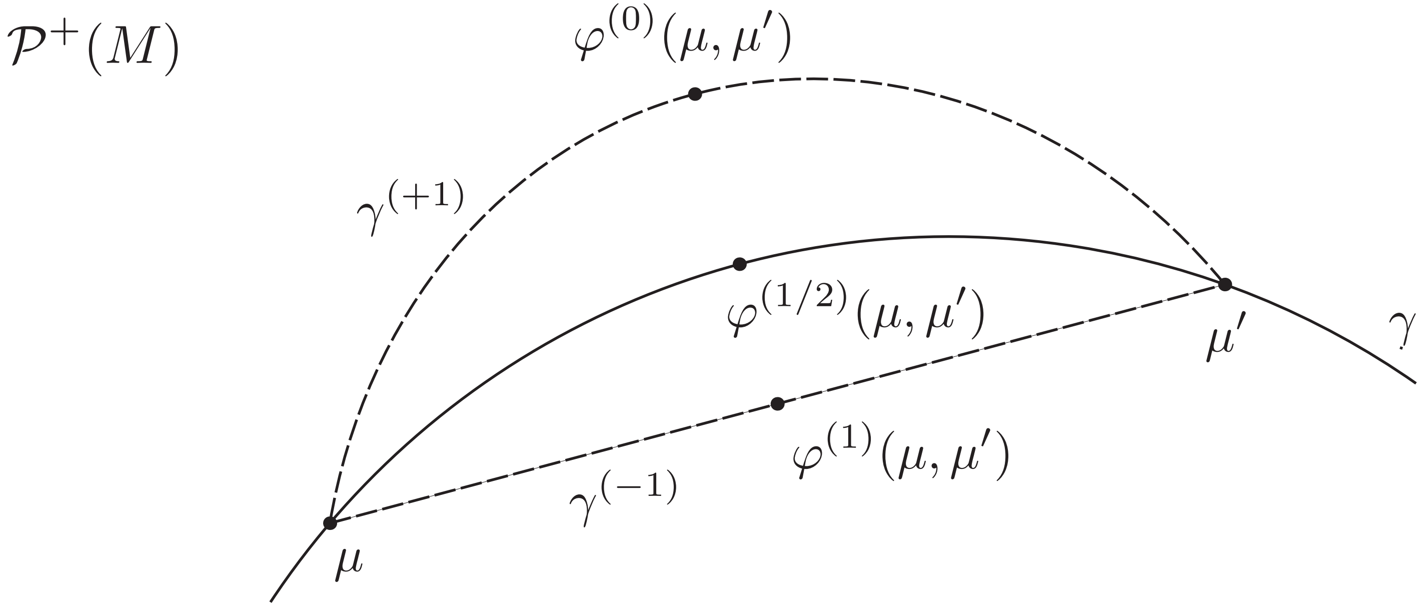
$$\gamma^{(-1)}(t) = \frac{1}{l_-}\left\{(l_- - t)\mu + t\mu'\right\}, \qquad \gamma^{(-1)}(l_-) = \mu'$$

$$\gamma^{(+1)}(t) = \left\{\int_M \left(\frac{d\mu'}{d\mu}\right)^{t/l_+} d\mu\right\}^{-1}\left(\frac{d\mu'}{d\mu}\right)^{t/l_+}\mu, \qquad \gamma^{(+1)}(l_+) = \mu'$$

$$\therefore\ \gamma^{(-1)}(l_-/2) = \frac{1}{2}(\mu + \mu') = \varphi^{(1)}(\mu, \mu'), \qquad \gamma^{(+1)}(l_+/2) = \frac{1}{C}\sqrt{\frac{d\mu'}{d\mu}}\mu = \varphi^{(1/2)}(\mu, \mu')$$

5. $\mathcal{P}^+(M)$ 上の双対（平坦）構造と測地線





6. ダイバージェンスと測地線

- 一般に，ダイバージェンスは対称ではない．
- $D^{(\alpha)}(p\|q) = D^{(-\alpha)}(q\|p)$
- $\mathcal{P}^+(M)$ 上の双対平坦構造を与える $D^{(-1)}$（または $D^{(1)}$）は Kullback-Leibler 情報量や情報ダイバージェンス，相対エントロピーなどとよばれる．

6. ダイバージェンスと測地線

- 一般に，ダイバージェンスは対称ではない．
- $D^{(\alpha)}(p\|q) = D^{(-\alpha)}(q\|p)$
- $\mathcal{P}^{+}(M)$ 上の双対平坦構造を与える $D^{(-1)}$（または $D^{(1)}$）は Kullback-Leibler 情報量や情報ダイバージェンス，相対エントロピーなどとよばれる．

命題

$\gamma^{(\pm 1)} : [0, l_{\pm}] \to \mathcal{P}^{+}(M)$ を，$\mu, \mu' \in \mathcal{P}^{+}(M)$ を結ぶ (± 1)-測地線分とする．このとき，

$$\int_0^{l_{\pm 1}} G(\dot{\gamma}^{(\pm 1)}(t), \dot{\gamma}^{(\pm 1)}(t))\, dt = \frac{1}{l_{\pm 1}} \left\{ D^{(-1)}(\mu\|\mu') + D^{(-1)}(\mu'\|\mu) \right\}$$

7. 参考文献に追加

[6] Nihat Ay and Shun-ichi Amari, *A Novel Approach to Canonical Divergences within Information Geometry*, Entropy **17**, Issue 12 (2015), 8111-8129.

[7] Shun-ichi Amari, Information Geometry and Its Applications, Applied Mathematical Sciences **194**, Springer, 2016.

[8] T. Kurose, *On the divergenced of 1-conformally Flat statistical manifolds*, Tohoku Math. J. **46** (1994), 427-433.