

# 変形指数型分布族の幾何学

- Geometry of deformed exponential families -

名古屋工業大学・大学院工学研究科 松添 博<sup>1</sup>

Hiroshi Matsuzoe  
Graduate School of Engineering  
Nagoya Institute of Technology

## 概要

指数型分布族は統計学では最も基本的な統計モデルの一つである。情報幾何学における研究で、指数型分布族は自然に統計多様体の構造を持ち、さらに双対平坦空間となることが知られている。変形指数型分布族は指数型分布族の一般化であるが、これまでの研究で 2 種類の異なる双対平坦構造が導入されることが知られている。

そこで本論文では、変形指数型分布族で最も基礎的なものの一つである  $q$ -指数型分布族の幾何学を考察する。 $q$ -指数型分布族が 2 種類の異なる双対平坦構造を持つことを確認したのち、推定関数からのダイバージェンス関数の構成、独立性と最尤推定量の一般化などについて解説を行う。

## 1 はじめに

指数型分布族は、統計学では最も基本的な統計モデルの一つであり、正規分布族やガンマ分布族、また離散標本空間上の確率分布など、主要な確率密度関数族、確率関数族が含まれる。統計的推論の幾何学的方法論の一つに情報幾何学があるが、指数型分布族は自然に統計多様体の構造を持ち、さらに双対平坦空間となることが知られている。この双対的な幾何学構造により、最尤推定法をはじめとする統計計算のアルゴリズムを明快に理解することができる (cf. [1])。

変形指数型分布族は指数型分布族の一般化であり、複雑系・強相関係の統計物理学において導入された (cf. [20], [16])。他の複雑系科学、強相関係の統計学と区別するために、本論文では変形指数型分布族に関する統計学を、異常統計とよぶことにする。これまでの異常統計における情報幾何学の研究から、変形指数型分布族には 2 種類の異なる双対平坦構造が導入されることが知られている (cf. [10], [11], [13])。さらに異常統計におけるエスコート確率分布や、変形した代数なども、推定関数やダイバージェンスの幾何学で説明することができる。

そこで本論文では、変形指数型分布族で最も基礎的なものの一つである  $q$ -指数型分布族に注目し、その幾何学を考察する。 $q$ -指数型分布族に 2 種類の異なる双対平坦構造が導入されることが確認したのち、推定関数からのダイバージェンス関数の構成、独立性と最尤推定量の一般化などについて解説を行う。なお本論文は、日本数学会における講演要旨 [9] を加筆、修正したものである。

<sup>1</sup>本研究の一部は科学研究費補助金(若手研究(B) 課題番号:23740047)の助成を受けたものである。

## 2 統計多様体

まずはじめに統計多様体の幾何学をまとめる. [12]にも統計多様体の解説がある.

$(M, h)$  を擬 Riemann 多様体,  $\nabla$  を  $M$  上の捩れの無いアフィン接続とする.  $\nabla h$  が対称な  $(0, 3)$ -テンソル場であるとき, すなわち

$$(\nabla_X h)(Y, Z) = (\nabla_Y h)(X, Z)$$

が成り立つとき,  $(M, \nabla, h)$  を**統計多様体**とよぶ [7], [8].

統計多様体  $(M, \nabla, h)$  に対して,  $\nabla$  の  $h$  に関する双対接続  $\nabla^*$  は捩れがなく,  $\nabla^* h$  は対称となる. したがって  $(M, \nabla^*, h)$  も統計多様体となるが, これを  $(M, \nabla, h)$  の**双対統計多様体**とよぶ. また, 統計多様体に対して対称な  $(0, 3)$ -テンソル場  $C(X, Y, Z) = (\nabla_X h)(Y, Z)$  が自然に定義される. この  $C$  を統計多様体  $(M, \nabla, h)$  の**3次形式**とよぶ. 逆に擬 Riemann 多様体  $(M, h)$  と対称な  $(0, 3)$ -テンソル場  $C$  が与えられると

$$\begin{aligned} h(\nabla_X Y, Z) &= h(\nabla_X^{(0)} Y, Z) - \frac{1}{2} C(X, Y, Z), \\ h(\nabla_X^* Y, Z) &= h(\nabla_X^{(0)} Y, Z) + \frac{1}{2} C(X, Y, Z) \end{aligned}$$

によって, 互いに双対的な捩れの無いアフィン接続  $\nabla, \nabla^*$  が定義できる. ただし  $\nabla^{(0)}$  は  $h$  の Levi-Civita 接続である. さらに,  $\nabla h$  と  $\nabla^* h$  はそれぞれ対称であり, 統計多様体  $(M, \nabla, h)$ ,  $(M, \nabla^*, h)$  が得られる.

統計多様体の双対接続を中心とした議論においては, 計量  $h$  の正值性の仮定は必要としない. 特に計量が正值, すなわち Riemann 計量であることを明示する場合には,  $h$  ではなく  $g$  と表記することにする.

統計多様体  $(M, \nabla, h)$  において  $\nabla$  が平坦であるとき,  $(M, \nabla, h)$  を平坦統計多様体とよぶことにする. これは  $(\nabla, h)$  が  $M$  上の Hesse 構造 [18], または  $(M, h, \nabla, \nabla^*)$  が双対平坦空間 [1] であることと同値である. この場合,  $\nabla$  の接続係数を  $\{\Gamma_{ij}^k\}$  とすると適当な局所座標系  $\{\theta^i\}$  が存在して  $\Gamma_{ij}^k \equiv 0$  となる. このような座標系  $\{\theta^i\}$  を  $\nabla$  の**アフィン座標系**とよぶ. さらに, 双対接続  $\nabla^*$  のアフィン座標系  $\{\eta_i\}$  で, 次を満たすものが存在する.

$$h\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j}\right) = \delta_i^j.$$

この  $\{\eta_i\}$  を  $h$  に関するアフィン座標系  $\{\theta^i\}$  の**双対座標系**とよぶ.

**命題 2.1**  $(M, \nabla, h)$  を平坦統計多様体とし  $\{\theta^i\}$  を  $M$  の  $\nabla$  に関するアフィン座標系,  $\{\eta_i\}$  を  $\{\theta^i\}$  の双対座標系とする. このとき  $M$  上の関数  $\psi$  と  $\phi$  が存在して次が成り立つ.

$$\begin{aligned} \frac{\partial \psi}{\partial \theta^i} = \eta_i, \quad \frac{\partial \phi}{\partial \eta_i} = \theta^i, \quad \psi(p) + \phi(p) - \sum_{i=1}^n \theta^i(p) \eta_i(p) = 0, \quad (p \in M), \\ h_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}, \quad h^{ij} = \frac{\partial^2 \phi}{\partial \eta_i \partial \eta_j}. \end{aligned}$$

ただし  $(h_{ij})$  は擬 Riemann 計量  $h$  の  $\{\theta^i\}$  に関する成分行列,  $(h^{ij})$  は  $\{\eta_i\}$  に関する成分行列である. さらに

$$C_{ijk} = \frac{\partial^3 \psi}{\partial \theta^i \theta^j \theta^k}$$

は統計多様体  $(M, \nabla, h)$  の 3 次形式である (cf. [1]).

上述のように, 関数  $\psi$  および  $\phi$  の Hessian として計量  $h$  が与えられる. 関数  $\psi$  を  $\theta$ -ポテンシャル,  $\phi$  を  $\eta$ -ポテンシャルとよぶ.

命題 2.1 と同じ仮定の下で,  $M \times M$  上の関数  $D$  を次式で定義する.

$$D(p, r) = \psi(p) + \phi(r) - \sum_{i=1}^n \theta^i(p) \eta_i(r), \quad (p, r \in M).$$

この  $D$  を平坦統計多様体  $(M, \nabla, h)$  の正準ダイバージェンスとよぶ. 正準ダイバージェンスはアフィン座標系の取り方に依らず, 一意的に定まる.

次に実数  $\alpha$  を固定する. 2つの統計多様体  $(M, \nabla, h)$  と  $(M, \bar{\nabla}, \bar{h})$  が  $\alpha$ -共形同値であるとは, 適当な  $M$  上の関数  $\phi$  が存在して

$$\begin{aligned} \bar{h}(X, Y) &= e^\phi h(X, Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - \frac{1+\alpha}{2} h(X, Y) \text{grad}_h \phi + \frac{1-\alpha}{2} \{d\phi(Y) X + d\phi(X) Y\} \end{aligned}$$

が成り立つこととする. 統計多様体  $(M, \nabla, h)$  が局所的に平坦統計多様体と  $\alpha$ -共形同値であるとき,  $\alpha$ -共形平坦であるという [7].

### 3 統計モデルの幾何学

$(\Omega, \mathcal{F}, \mathcal{P})$  を確率空間とし,  $\Xi$  を  $\mathbf{R}^n$  の開集合とする.  $\Omega$  上の確率密度関数のなす集合  $S$  が統計モデルであるとは,  $\xi \in \Xi$  をパラメータとする次の集合である.

$$S = \left\{ p(x; \xi) \mid \int_{\Omega} p(x; \xi) dx = 1, p(x; \xi) > 0, \xi = (\xi^1, \dots, \xi^n) \in \Xi \subset \mathbf{R}^n \right\}.$$

本論文では  $S$  を  $\{\xi^i\}$  を一つの局所座標系とする多様体とみなす [1].

次に  $S$  に Riemann 計量を与える. まず, 次の式で  $S$  上に対称行列を定める.

$$\begin{aligned} g_{ij}^F(\xi) &= \int_{\Omega} \left( \frac{\partial}{\partial \xi^i} \log p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} \log p(x; \xi) \right) p(x; \xi) dx \\ &= E_p[\partial_i l_\xi \partial_j l_\xi]. \end{aligned}$$

ただし  $E_p[f] = \int_{\Omega} f(x) p(x; \xi) dx$  は確率変数  $f(x)$  の  $p(x; \xi)$  に関する期待値,  $\partial_i$  はパラメータ  $\xi$  に関する微分,  $l_\xi = l(x; \xi) = \log p(x; \xi)$  は  $p(x; \xi)$  の対数尤度である.

適当な条件のもとで  $g^F = (g_{ij}^F)$  は Riemann 計量となり, これを  $S$  の **Fisher 計量** とよぶ. Fisher 計量は以下の表示を持つ.

$$\begin{aligned} g_{ij}^F(\xi) &= \int_{\Omega} \left( \frac{\partial}{\partial \xi^i} \log p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} \log p(x; \xi) \right) p(x; \xi) dx \\ &= \int_{\Omega} \left( \frac{\partial}{\partial \xi^i} p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} \log p(x; \xi) \right) dx \end{aligned} \quad (1)$$

$$= \int_{\Omega} \frac{1}{p(x; \xi)} \left( \frac{\partial}{\partial \xi^i} p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} p(x; \xi) \right) dx. \quad (2)$$

なお, 対数尤度の微分

$$s^i(x; \xi) = \frac{\partial}{\partial \xi^i} \log p(x; \xi)$$

を  $p(x; \xi)$  の  $\xi$  に関する**スコア関数**とよぶ. また, スコア関数  $s^i(x; \xi)$  を確率密度関数  $p(x; \xi)$  の **e-表現**,  $\partial_i p(x; \xi)$  を **m-表現**とよぶこともある.

さらに  $S$  にアファイン接続を定める. 固定した  $\alpha \in \mathbf{R}$  に対し

$$\Gamma_{ij,k}^{(\alpha)}(\xi) = E_p \left[ \left( \partial_i \partial_j l_{\xi} + \frac{1-\alpha}{2} \partial_i l_{\xi} \partial_j l_{\xi} \right) (\partial_k l_{\xi}) \right]$$

によって  $\alpha$ -**接続**  $\nabla^{(\alpha)}$  を定義する. 1-接続と (-1)-接続は  $\nabla^{(e)} = \nabla^{(1)}$ ,  $\nabla^{(m)} = \nabla^{(-1)}$  と表示されることがあり, それぞれ**指数型接続**, **混合型接続**とよばれる.

$\nabla^{(\alpha)}$  は振れのないアファイン接続であり,  $\nabla^{(\alpha)} g^F$  は対称である. これらから得られる統計多様体  $(S, \nabla^{(\alpha)}, g^F)$  を**不変統計多様体**とよぶ. 不変統計多様体  $(S, \nabla^{(e)}, g^F)$  の3次形式  $C^F$  は  $C_{ijk}^F = \Gamma_{ij,k}^{(m)} - \Gamma_{ij,k}^{(e)}$  によって与えられる.

代表的な統計モデルとして, 次で定義される**指数型分布族**がある.

$$S_e = \left\{ p(x; \theta) \mid p(x; \theta) = \exp \left[ \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right], \theta \in \Theta \subset \mathbf{R}^n \right\}.$$

ただし  $F_1(x), \dots, F_n(x)$  は  $\Omega$  上の関数,  $\theta = \{\theta^1, \dots, \theta^n\}$  は確率密度関数のパラメータとし,  $\Theta$  は統計モデルのパラメータ空間で  $\mathbf{R}^n$  の開集合であると仮定する. また  $\psi(\theta)$  はパラメータ  $\theta$  に関する確率密度関数の規格化項である.

**命題 3.1** 指数型分布族  $S_e$  に対して以下が成り立つ.

1.  $(S_e, g^F, \nabla^{(e)}, \nabla^{(m)})$  は双対平坦空間である.
2.  $\{\theta^i\}$  は  $S_e$  の  $\nabla^{(e)}$ -アファイン座標系である.
3.  $\psi(\theta)$  は  $\{\theta^i\}$  に関する  $g^F$  のポテンシャルである. すなわち

$$g_{ij}^F(\theta) = \partial_i \partial_j \psi(\theta), \quad (\partial_i = \partial / \partial \theta^i).$$

4. 確率変数  $F_i(x)$  の期待値を  $\eta_i = E_p[F_i(x)]$  とおくと,  $\{\eta_i\}$  は  $S_e$  の  $\nabla^{(m)}$ -アファイン座標系で  $\{\theta^i\}$  の  $g^F$  に関する双対座標系である.
5.  $\phi(\eta) = E_p[\log p(x; \theta)]$  とおくと,  $\phi(\eta)$  は  $\{\eta_i\}$  に関する  $g^F$  のポテンシャルである.

$(S_e, g^F, \nabla^{(e)}, \nabla^{(m)})$  は双対平坦空間であるから、命題 2.1 と同様の関係式も成り立つ。さらに  $(S_e, \nabla^{(e)}, g^F)$  の 3-次形式は

$$C_{ijk}^F(\theta) = \partial_i \partial_j \partial_k \psi(\theta)$$

によって与えられる。

統計モデル  $S$  に **Kullback-Leibler ダイバージェンス** (または**相対エントロピー**) を次の式で定める。

$$\begin{aligned} D_{KL}(p, r) &= \int_{\Omega} p(x) \log \frac{p(x)}{r(x)} dx \\ &= E_p[\log p(x) - \log r(x)]. \end{aligned}$$

指数型分布族  $S_e$  の場合、 $D_{KL}$  は平坦統計多様体  $(S_e, \nabla^{(m)}, g^F)$  の正準ダイバージェンスと一致する。実際、次の式が成り立つ。

$$\begin{aligned} D_{KL}(p(\theta), p(\theta')) &= E_p[\log p(x; \theta) - \log p(x; \theta')] \\ &= \int_{\Omega} \left\{ \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) - \sum_{i=1}^n (\theta')^i F_i(x) + \psi(\theta') \right\} p(x; \theta) dx \\ &= \sum_{i=1}^n \theta^i \eta_i(x) - \psi(\theta) - \sum_{i=1}^n (\theta')^i \eta_i(x) + \psi(\theta') \\ &= \psi(\theta') + \phi(\eta) - \sum_{i=1}^n (\theta')^i \eta_i(x) = D(p(\theta'), p(\theta)). \end{aligned}$$

#### 4 $q$ -指数型分布族

次に  $q$ -指数型分布族の定義を与え、その上の幾何学を考察する。まず指数関数と対数関数を拡張する。正の数  $q$  を固定し、 $q$ -指数関数と  $q$ -対数関数を次の式で定義する (cf. [16], [19]).

$$\begin{aligned} \exp_q x &:= (1 + (1 - q)x)^{\frac{1}{1-q}} \quad (1 + (1 - q)x > 0), \\ \log_q x &:= \frac{x^{1-q} - 1}{1 - q} \quad (x > 0). \end{aligned}$$

$q \rightarrow 1$  の極限を考えると、それぞれ通常の指数関数と対数関数である。

**定義 4.1** 次で定義される統計モデル  $S_q$  を  $q$ -指数型分布族とよぶ：

$$S_q := \left\{ p(x, \theta) \mid p(x; \theta) = \exp_q \left[ \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right], \theta \in \Theta \subset \mathbf{R}^n \right\},$$

ただし  $F_1(x), \dots, F_n(x)$  は確率変数、 $\theta = \{\theta^1, \dots, \theta^n\}$  はパラメータ、 $\psi(\theta)$  はパラメータに関する規格化項とする。

$q$ -指数型分布  $p(x; \theta) \in S_q$  に対し,

$$P(x; \theta) = \frac{1}{Z_q(\theta)} \int_{\Omega} p^q(x; \xi) dx, \quad \text{ただし } Z_q(\theta) = \int_{\Omega} \{p(x; \theta)\}^q dx$$

を  $p(x; \theta)$  のエスコート分布, また

$$E_{q,p}[f] = \int_{\Omega} f(x) P(x; \theta) dx$$

を  $f(x)$  の  $q$ -期待値とよぶ.

ここで統計モデルに対する主なダイバージェンスをまとめておく. ダイバージェンスや関連するコントラスト関数の幾何学は [4, 14, 15] など, Tsallis 相対エントロピーについては [19] を参照されたい.

$\alpha \in \mathbf{R}$  および  $q \in \mathbf{R}$  を固定する.  $S$  の  $\alpha$ -ダイバージェンス, および Csiszár-タイプの Tsallis 相対エントロピーを, それぞれ以下で定義する.

$$D^{(\alpha)}(p, r) = \frac{4}{1-\alpha^2} \left\{ 1 - \int_{\Omega} p(x)^{\frac{1-\alpha}{2}} r(x)^{\frac{1+\alpha}{2}} dx \right\},$$

$$D_q^C(p, r) = \frac{1}{1-q} \left\{ 1 - \int_{\Omega} p(x)^q r(x)^{1-q} dx \right\}.$$

$\alpha = 1 - 2q$  の関係で, これらは定数倍を除いて一致する. すなわち

$$D^{(\alpha)}(p, r) = \frac{1}{q} D_q^C(p, r)$$

が成り立つ. また  $q \rightarrow 1$  ( $\alpha \rightarrow -1$ ) のとき, これらは KL-ダイバージェンスとなる. なお, 本論文では  $D_q^C(p, r)$  ( $D^{(\alpha)}(p, r)$ ) を変形したもの

$$D_q^T(p, r) = \frac{1}{(1-q)Z_q(p)} \left\{ 1 - \int_{\Omega} p(x)^q r(x)^{1-q} dx \right\}$$

を, 正規化した Tsallis 相対エントロピーとよぶ.

次に  $\beta \in \mathbf{R}$  を固定し,  $S$  の  $\beta$ -ダイバージェンス, および Bregman タイプの Tsallis 相対エントロピーを, それぞれ以下で定義する.

$$D_{\beta}(p, r) = \int_{\Omega} \left\{ p(x) \frac{p(x)^{\beta} - r(x)^{\beta}}{\beta} - \frac{p(x)^{\beta+1} - r(x)^{\beta+1}}{\beta+1} \right\} dx$$

$$\left( = \frac{1}{\beta(\beta+1)} \int_{\Omega} p(x)^{\beta+1} dx - \frac{1}{\beta} \int_{\Omega} p(x)r(x)^{\beta} dx + \frac{1}{\beta+1} \int_{\Omega} r(x)^{\beta+1} dx \right),$$

$$D_q^B(p, r) = \int_{\Omega} \left\{ p(x) \frac{p(x)^{q-1} - r(x)^{q-1}}{q-1} - (p(x) - r(x))r(x)^{q-1} \right\} dx.$$

$\beta = q - 1$  の関係で, これらは定数倍を除いて一致する. また  $q \rightarrow 1$  ( $\beta \rightarrow 0$ ) のとき, これらは KL-ダイバージェンスとなる.

## 5 $q$ -指数型分布族の幾何学 (前半)

統計モデルの幾何学における接続の双対性は、確率密度関数の  $e$ -表現と  $m$ -表現の双対性に起因するものであった。この章では  $e$ -表現の一般化から  $q$ -指数型分布族の幾何学を考える。これに関連した統計学や、 $q$ -指数関数をさらに一般化した内容は [15, 17] などを参照されたい。

$q$ -対数関数によって  $e$ -表現を一般化し、Fisher 計量の (1) の表示の拡張として  $S_q$  上の Riemann 計量  $g^M$  を

$$g_{ij}^M = \int_{\Omega} \left( \frac{\partial}{\partial \xi^i} p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} \log_q p(x; \xi) \right) dx$$

と定める。ここで  $q$ -対数尤度の微分  $(s^q)^i(x; \theta) = (\partial/\partial \theta^i) \log_q p(x; \theta)$  を  $q$ -スコア関数とよぶことにする。また、エスコート分布を用いて別の Riemann 計量  $g^N$  を

$$g_{ij}^N = \int_{\Omega} \frac{1}{P(x; \xi)} \left( \frac{\partial}{\partial \xi^i} p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} p(x; \xi) \right) dx$$

と定める。これは Fisher 計量の (2) の拡張であるが、 $g^N$  は一般化した情報計量とよばれることもある [16]。

**命題 5.1**  $S_q$  を  $q$ -指数型分布族とすると  $g^M$  と  $g^N$  は共形的であり

$$g^M(p) = Z_q(p)g^N(p)$$

が成り立つ。ただし  $Z_q$  はエスコート分布の規格化関数である。

以下、この章では  $g^M$  を考察する。Riemann 計量  $g^M$  に関する双対的なアファイン接続を

$$\begin{aligned} \Gamma_{ij,k}^{M(e)}(\theta) &= \int_{\Omega} \partial_k p(x; \theta) \partial_i \partial_j \log_q p(x; \theta) dx, \\ \Gamma_{ij,k}^{M(m)}(\theta) &= \int_{\Omega} \partial_i \partial_j p(x; \theta) \partial_k \log_q p(x; \theta) dx \end{aligned}$$

によって定義する。 $q$ -指数型分布族と  $q$ -対数関数の定義から  $\Gamma_{ij,k}^{M(e)}$  は恒等的に消えていることがわかり、 $\nabla^{M(e)}$  は平坦である。したがって  $(\nabla^{M(e)}, g^M)$  は  $S_q$  上の Hesse 構造となる。ここで一般化した Massieu ポテンシャルを

$$\begin{aligned} \Psi(\theta) &= \phi(\theta) + \psi(\theta), \\ \text{ただし } \phi(\theta) &= \frac{1}{2-q} \int_{\Omega} (p(x; \theta)^{2-q} - p(x; \theta)) dx \end{aligned}$$

と定める。 $q \rightarrow 1$  のとき  $\phi \rightarrow 0$  である。

命題 5.2  $q$ -指数型分布族  $S_q$  に対して以下が成り立つ.

1.  $(S_q, g^M, \nabla^{M(e)}, \nabla^{M(m)})$  は双対平坦空間である.
2.  $\{\theta^i\}$  は  $S_q$  の  $\nabla^{M(e)}$ -アフィン座標系である.
3.  $\Psi(\theta)$  は  $\{\theta^i\}$  に関する  $g^M$  のポテンシャルである. すなわち

$$g_{ij}^M(\theta) = \partial_i \partial_j \Psi(\theta).$$

4. 確率変数  $F_i(x)$  の期待値を  $\eta_i = E_p[F_i(x)]$  とおくと,  $\{\eta_i\}$  は  $S_q$  の  $\nabla^{M(m)}$ -アフィン座標系で  $\{\theta^i\}$  の  $g^M$  に関する双対座標系である.
5.  $\Phi(\eta) = E_p[\frac{1}{2-q} \log_q p(x; \theta)]$  とおくと,  $\Phi(\eta)$  は  $\{\eta_i\}$  に関する  $g^M$  のポテンシャルである.

平坦統計多様体  $(S_q, \nabla^{M(m)}, g^M)$  のカノニカル・ダイバージェンス  $D$  は  $\beta$ -ダイバージェンス  $D_{1-q}$  ( $\beta = 1 - q$ ) に一致する.

推定関数から  $\beta$ -ダイバージェンスの構成は次の通りである. スコア関数に重みを付けた推定関数  $u_q(x; \theta)$  を次で定義する.

$$u_q^i(x; \theta) = p(x; \theta)^{1-q} s^i(x; \theta) - E_p[p(x; \theta)^{1-q} s^i(x; \theta)].$$

$q$ -対数関数の定義を用いて,  $u_q(x; \theta)$  は

$$\begin{aligned} u_q^i(x; \theta) &= \frac{\partial}{\partial \theta^i} \left\{ \frac{1}{1-q} p(x; \theta)^{1-q} - \frac{1}{2-q} \int_{\Omega} p(x; \theta)^{2-q} dx \right\} \\ &= \frac{\partial}{\partial \theta^i} \log_q p(x; \theta) - E_p \left[ \frac{\partial}{\partial \theta^i} \log_q p(x; \theta) \right] \end{aligned}$$

と与えられるから, この推定関数は  $q$ -スコア関数を不偏化したものにほかならない.

推定関数  $u_q(x; \theta)$  を積分することにより, クロス・エントロピーを

$$d_{1-q}(p, r) = -\frac{1}{1-q} \int_{\Omega} p(x; \theta) r(x; \theta)^{1-q} + \frac{1}{2-q} \int_{\Omega} r(x; \theta)^{2-q} dx$$

と定めると,  $\beta$ -ダイバージェンス ( $\beta = 1 - q$ ) が

$$\begin{aligned} D_{1-q}(p, r) &= -d_{1-q}(p, p) + d_{1-q}(p, r) \\ &= \frac{1}{(1-q)(2-q)} \int_{\Omega} p(x)^{2-q} dx \\ &\quad - \frac{1}{1-q} \int_{\Omega} p(x) r(x)^{1-q} dx + \frac{1}{2-q} \int_{\Omega} r(x)^{2-q} dx \end{aligned}$$

によって与えられる.

## 6 $q$ -指数型分布族の幾何学（後半）

この章では、 $q$ -指数型分布族  $S_q$  は通常 Fisher 計量  $g^F$  と  $\alpha$ -接続  $\nabla^{(\alpha)}$  を持つと仮定する。

$S_q$  は  $q$ -対数関数を用いて線形化可能なモデルであるので、指数型分布族と同様に  $\psi$  から幾何構造を定めることができる。 $q$ -Fisher 計量  $g^q$  と  $q$ -3 次形式  $C^q$  を

$$g_{ij}^q(\theta) = \partial_i \partial_j \psi(\theta), \quad C_{ijk}^q(\theta) = \partial_i \partial_j \partial_k \psi(\theta) \quad (\partial_i = \partial / \partial \theta^i)$$

によって定義する [2], [14]. また  $q$ -指数型接続  $\nabla^{q(e)}$  と  $q$ -混合型接続  $\nabla^{q(e)}$  を

$$\begin{aligned} g^q(\nabla_X^{q(e)} Y, Z) &= g^q(\nabla_X^{q(0)} Y, Z) - \frac{1}{2} C^q(X, Y, Z), \\ g^q(\nabla_X^{q(m)} Y, Z) &= g^q(\nabla_X^{q(0)} Y, Z) + \frac{1}{2} C^q(X, Y, Z) \end{aligned}$$

によって定義する。ただし  $\nabla^{q(0)}$  は  $q$ -Fisher 計量  $g^q$  に関する Levi-Civita 接続である。

**命題 6.1**  $q$ -指数型分布族  $S_q$  に対して以下が成り立つ。

1.  $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$  は双対平坦空間である。
2.  $\{\theta^i\}$  は  $S_q$  の  $\nabla^{q(e)}$ -アフィン座標系である。
3. 確率変数  $F_i(x)$  の  $q$ -期待値を  $\eta_i = E_{q,p}[F_i(x)]$  とおくと、 $\{\eta_i\}$  は  $S_q$  の  $\nabla^{q(m)}$ -アフィン座標系で  $\{\theta^i\}$  の  $g^q$  に関する双対座標系である。
4.  $\phi(\eta) = E_{q,p}[\log_q p(x; \theta)]$  とおくと、 $\phi(\eta)$  は  $\{\eta_i\}$  に関する  $g^q$  のポテンシャルである。

$q$ -指数型分布族  $S_q$  の不変統計多様体と平坦統計多様体の関係について考える。

**命題 6.2**  $S_q = \{p(x; \theta)\}$  を  $q$ -指数型分布族,  $Z_q(p) = \int_{\Omega} \{p(x; \theta)\}^q dx$  とする。また不変統計多様体  $(S_q, \nabla^{(1)}, g^F)$  に定まる 3 次形式を  $C^F$  と書と、次が成り立つ。

$$\begin{aligned} g_{ij}^q(\theta) &= \frac{q}{Z_q(p)} g_{ij}^F(\theta), \\ C_{ijk}^q(\theta) &= \frac{q(2q-1)}{Z_q(p)} C_{ijk}^F - \frac{q}{(Z_q(p))^2} \{g_{ij}^F \partial_k Z_q(p) + g_{jk}^F \partial_i Z_q(p) + g_{ki}^F \partial_j Z_q(p)\}. \end{aligned}$$

この命題と統計多様体の  $\alpha$ -共形同値の定義から、次が得られる。

**定理 6.3**  $S_q = \{p(x; \theta)\}$  を  $q$ -指数型分布族とする。不変統計多様体  $(S_q, \nabla^{(2q-1)}, g^F)$  と平坦統計多様体  $(S_q, \nabla^{q(e)}, g^q)$  は 1-共形同値である。特に、次の関係式が成り立つ。

$$\begin{aligned} g^{q(e)}(X, Y) &= \frac{q}{Z_q(\theta)} g^F(X, Y), \\ \nabla_X^{q(e)} Y &= \nabla_X^{(2q-1)} Y + h(X, Y) \text{grad}_{g^F}(\log Z_q(\theta)). \end{aligned}$$

ただし  $\text{grad}_{g^F}(\log Z_q)$  は関数  $\log Z_q$  の  $g^F$  に関する勾配ベクトル場である [14].

$(S_q, \nabla^{q(e)}, g^q)$  は平坦統計多様体であるから、直ちに次の系が成り立つ。

**系 6.4** 不変統計多様体  $(S_q, \nabla^{(2q-1)}, g^F)$  は 1-共形平坦である。

不変統計多様体  $(S_q, \nabla^{(2q-1)}, g^F)$  は  $\alpha$ -ダイバージェンスで  $\alpha = 2q - 1$  とした  $D^{(2q-1)}$  から誘導される統計多様体である。平坦統計多様体  $(S_q, \nabla^{q(e)}, g^q)$  は Tsallis 相対エントロピーから誘導される統計多様体である。

平坦統計多様体  $(S_q, \nabla^{q(e)}, g^q)$  のカノニカル・ダイバージェンス  $D$  と、Tsallis 相対エントロピー  $D_q^T$ 、 $\alpha$ -ダイバージェンス  $D^{(2q-1)}$  ( $\alpha = 2q - 1$ ) には、次の関係が成り立つ。

$$D(p, r) = D_q^T(r, p) = \frac{q}{Z_q(r)} D^{(2q-1)}(p, r).$$

最後に、推定関数からの Tsallis 相対エントロピーの構成法を解説する。  $q$ -対数尤度の微分によって  $q$ -スコア関数を

$$(s^q)^i(x; \theta) = \frac{\partial}{\partial \theta^i} \log_q p(x; \theta)$$

と定める。  $q$ -スコア関数は前述の推定関数とはならないが、  $q$ -期待値に関して不偏性を持つ。すなわち  $E_{q,p}[(s^q)^i(x; \theta)] = 0$  が成り立つため、推定関数の一般化と考える。

$q$ -スコア関数のエスコート分布に関する積分から  $q$ -クロス・エントロピーを

$$d^q(p, r) = - \int_{\Omega} P(x) \log_q r(x) dx$$

と定めると、Tsallis 相対エントロピーが

$$\begin{aligned} D_q^T(p, r) &= -d^q(p, p) + d^q(p, r) \\ &= E_{q,p}[\log_q p(x) - \log_q r(x)] \end{aligned}$$

によって与えられる。

## 7 独立性と最尤推定量の一般化

この章では確率変数の独立性の概念を修正し、最尤推定量を一般化することで  $q$ -指数型分布族の統計的推論を考える。

$X$  と  $Y$  をそれぞれ確率分布  $p_1(x), p_2(y)$  に従う確率変数とする。この確率変数  $X, Y$  が独立であるとは、 $X, Y$  の同時確率分布  $p(x, y)$  が周辺確率分布の積として

$$p(x, y) = p_1(x)p_2(y)$$

と表されることであった。この式は  $p(x) > 0, p(y) > 0$  の場合には

$$p(x, y) = p_1(x)p_2(y) = \exp[\log p_1(x) + \log p_2(y)]$$

と書き換えることができる。したがって確率変数の独立性とは指数関数と対数関数の双対性に起因すると考えることができ、独立な確率変数に対する情報量の加法性という概念も内在していることがわかる (cf. [5], [6]).

$q$ -指数関数と  $q$ -対数関数を用いることで、独立性の一般化を考える。  $x > 0, y > 0$  とし  $x^{1-q} + y^{1-q} - 1 > 0$  ( $q > 0$ ) を仮定する。このとき  $x$  と  $y$  の  $q$ -積 [3] を次の式で定義する。

$$\begin{aligned} x \otimes_q y &:= [x^{1-q} + y^{1-q} - 1]^{\frac{1}{1-q}} \\ &= \exp_q [\log_q x + \log_q y]. \end{aligned}$$

この場合,

$$\begin{aligned} \exp_q x \otimes_q \exp_q y &= \exp_q(x + y), \\ \log_q(x \otimes_q y) &= \log_q x + \log_q y \end{aligned}$$

が成り立ち、指数法則がこの  $q$ -積によって修正されていることが確かめられる。

$X_i$  を  $\Omega_i$  上で  $p_i(x)$  ( $i = 1, 2, \dots, N$ ) に従う確率変数とする。  $\Omega_1, \Omega_2, \dots, \Omega_N$  の同時確率分布  $p(x_1, x_2, \dots, x_N)$  が

$$p(x_1, x_2, \dots, x_N) = \frac{p_1(x_1) \otimes_q p_2(x_2) \otimes_q \cdots \otimes_q p_N(x_N)}{Z_{p_1, p_2, \dots, p_N}}$$

と与えられるとき、  $X_1, X_2, \dots, X_N$  は ( $m$ -規格化の下で)  $q$ -独立とよぶ。ただし  $Z_{p_1, p_2, \dots, p_N}$  は

$$Z_{p_1, p_2, \dots, p_N} := \int \cdots \int_{\Omega_1 \cdots \Omega_N} p_1(x_1) \otimes_q p_2(x_2) \otimes_q \cdots \otimes_q p_N(x_N) dx_1 \cdots dx_N$$

によって定義される  $p_1(x_1) \otimes_q p_2(x_2) \otimes_q \cdots \otimes_q p_N(x_N)$  の規格化項である。

$q$ -指数関数は通常の和と積のもとでは

$$\begin{aligned} &\exp_q(x_1 + x_2 + \cdots + x_N) \\ &= \exp_q x_1 \otimes_q \exp_q x_2 \otimes_q \cdots \otimes_q \exp_q x_N \\ &= \exp_q x_1 \cdot \exp_q \left( \frac{x_2}{1 + (1-q)x_1} \right) \cdots \exp_q \left( \frac{x_N}{1 + (1-q) \sum_{i=1}^{N-1} x_i} \right) \end{aligned}$$

と表示されるので、通常の独立性のもとでは  $X_1, X_2, \dots, X_N$  は独立ではない。

さて、 $q$ -独立性のもとで最尤法の修正を考えよう。

$S_q = \{p(x; \xi) | \xi \in \Xi\}$  を  $q$ -指数型分布族とし、  $\{x_1, \dots, x_N\}$  を  $p(x; \xi) \in S_q$  から生成される  $N$  個の観測値とする。このとき  $q$ -尤度関数  $L_q(\xi)$  を

$$L_q(\xi) = p(x_1; \xi) \otimes_q p(x_2; \xi) \otimes_q \cdots \otimes_q p(x_N; \xi)$$

によって定義する. 同値な条件として,  $q$ -対数尤度関数  $l_q(\xi)$  を

$$l_q(\xi) := \log_q L_q(\xi) = \sum_{i=1}^N \log_q p(x_i; \xi)$$

によって定義する.  $q \rightarrow 1$  の極限では  $L_q$  は通常 of 尤度関数に収束する.  
 $q$ -尤度関数の最大値を与える引数  $\hat{\xi}$  を  $q$ -最尤推定量と定義する.

$$\hat{\xi} := \operatorname{argmax}_{\xi \in \Xi} L_q(\xi) \quad \left( = \operatorname{argmax}_{\xi \in \Xi} \log_q L_q(\xi) \right).$$

通常 of 指数型分布族における最尤推定量の計算と同様の計算手法で,  $q$ -最尤推定量は  $S_q$  の  $\{\eta\}$  座標系を用いて

$$\hat{\eta}_i = \frac{1}{N} \sum_{j=1}^N F_i(x_j)$$

と与えられる. さらに, 双対平坦空間  $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$  の正準ダイバージェンス  $D$  と正規化した Tsallis 相対エントロピー  $D^T$  を用いると

$$\begin{aligned} D_q^T(p(\hat{\eta}), p(\theta)) &= D(p(\theta), p(\hat{\eta})) \quad \left( = \psi(\theta) + \phi(\hat{\eta}) - \sum_{i=1}^n \theta^i \hat{\eta}_i \right) \\ &= \phi(\hat{\eta}) - \frac{1}{N} \log_q L_q(\theta). \end{aligned}$$

という関係式も成り立つ. すなわち,  $q$ -尤度の最大化は  $q$ -正準ダイバージェンスの最小化に等しい. この事実も, 指数型分布族における最尤法の直接的な一般化である. 正準ダイバージェンスは多様体の幾何学構造だけから決まる関数であるので,  $q$ -最尤原理は幾何学的に見ても自然な概念である.

ここまで, 変形指数型分布族で最も基礎的なものの一つである  $q$ -指数型分布族の幾何学と統計学を考察した. そもそも異常統計における観測は, 非復元抽出, または標本空間の小規模性に起因する影響が大きく, 通常 of 独立性を用いた統計学からの乖離が起こると予想している. 異常統計の理論構築には, 非復元抽出などの観測値を用いた統計学との関係の解明が必要であると考えている.

## 参考文献

- [1] S. Amari and H. Nagaoka, *Method of information geometry*, Amer. Math. Soc., Providence, Oxford University Press, Oxford, 2000.
- [2] S. Amari, A. Ohara and H. Matsuzoe, *Geometry of Deformed Exponential Families: Invariant, Dually-Flat and Conformal Geometry*, Physica A., **391**(2012), 4308-4319

- [3] E. P. Borges, *A possible deformed algebra and calculus inspired in nonextensive thermostatics*, Physica A, **340**(2004), 95–101.
- [4] S. Eguchi, *Geometry of minimum contrast*, Hiroshima Math. J., **22**(1992), 631–647.
- [5] Y. Fujimoto and N. Murata, *A Generalization of Independence in Naive Bayes Model*, Lecture Notes in Comp. Sci., **6283**(2010), 153–161.
- [6] Y. Fujimoto and N. Murata, *A generalisation of independence in statistical models for categorical distribution*, Int. J. Data Mining, Modelling and Management, No. 2, **4**(2012), 172–187.
- [7] T. Kurose, *On the divergences of 1-conformally flat statistical manifolds*, Tôhoku Math. J., **46**(1994), 427–433.
- [8] S. L. Lauritzen, *Statistical manifolds*, Differential Geometry in Statistical Inferences, IMS Lecture Notes Monograph Series 10, Institute of Mathematical Statistics, Hayward California, (1987), 96–163.
- [9] 松添 博, 統計多様体と推定関数の幾何学, 2013 年度日本数学会年会・幾何学分科会講演要旨, 2013
- [10] H. Matsuzoe, *Statistical manifolds and geometry of estimating functions*. Recent progress in differential geometry and its related fields, World Scientific, (2013), 187–202.
- [11] H. Matsuzoe, *Hessian structures on deformed exponential families and their conformal structures*, to appear in Diff. Geom. Appl.
- [12] 松添 博, 統計多様体とアファイン微分幾何学, 京都大学数理解析研究所講究録. (本論文と同一号に掲載)
- [13] H. Matsuzoe and M. Henmi, *Hessian structures and divergence functions on deformed exponential families*, Geometric Theory of Information, Signals and Communication Technology, Springer, (2014), 57-80.
- [14] H. Matsuzoe and A. Ohara, *Geometry for q-exponential families*, Proceedings of the 2nd International Colloquium on Differential Geometry and its Related Fields, World Sci. Publ., (2011), 55-71.
- [15] N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, *Information Geometry of U-Boost and Bregman Divergence*, Neural Comput., **16**(2004), 1437-1481.
- [16] J. Naudts, *Generalised Thermostatistics*, Springer-Verlag, 2011.
- [17] A. Ohara and T. Wada, *Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations*, J. Phys. A: Math. Theor., **43**(2010) No.035002.
- [18] H. Shima, *The Geometry of Hessian Structures*, World Scientific, 2007.
- [19] 須鎗弘樹, 複雑系のための基礎数理 — べき乗則とツァリスエントロピーの数理 —, 牧野書店. 2010.
- [20] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, New York, 2009.