

異常統計における変形指数型分布族の 幾何的・代数的構造

名古屋工業大学・大学院工学研究科 松添 博¹

Hiroshi Matsuzoe

Graduate School of Engineering

Nagoya Institute of Technology

概要

複雑系科学における変形指数型分布族に関する統計学を，異常統計とよぶ。変形指数型分布族は統計学で用いられる指数型分布族の一般化である。しかしながら，ここで重要な役割を果たす変形指数関数と変形対数関数は，通常の指数法則を満たさない。そのために変形した代数構造が導入されるが，結果として確率変数の期待値や独立性などの概念，指数型分布族の場合に知られている情報幾何学的構造が修正を受ける。

そこで本論文は，変形指数型分布族の定義や基本的な性質をまとめたのち，変形指数関数と変形対数関数から自然に定義される代数構造，および変形指数型分布族の情報幾何学的構造などを考察する。さらに，代数構造や幾何構造の変形に対応した最尤推定量の一般化などを考える。変形指数型分布族においては，期待値や独立性の修正に対応した統計量が自然に構成され，さらに通常の指数型分布族の場合に知られている情報幾何学に関する性質も，変形指数型分布族の場合に拡張できることが示される。

1 はじめに

指数型分布族は，最も基本的な統計モデルの一つである。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，統計学における主要な確率密度関数族，確率関数族は指数型分布族である。また情報幾何学においては，指数型分布族は平坦な統計多様体，すなわち双対平坦空間となることが知られている [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・強相関系の統計物理学において導入された (cf. [27], [21])。複雑系科学は非常に多岐にわたる研究の総称であるため，本論文では複雑系科学，強相関系の統計学と区別するために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。異常統計では変形指数関数と変形対数関数が重要な役割を果たすが，通常の指数関数と対数関数が持つ性質である指数法則を満たさない。そこで，指数法則を拡張するために変形代数構造が議論されている (cf. [7], [9], [3])。情報幾何学の視点からは，変形指数型分布族には 2 種類の異なる双対平坦構造が構成されることが示されている (cf. [12], [13], [17])。

そこで本論文では，変形指数関数と変形対数関数から自然に定義される代数構造をまとめたのち，確率変数の期待値や独立性の修正について議論する。指数型分布

¹本研究の一部は MEXT 科研費 (15K04842, 26108003) の助成を受けたものである。

族の幾何学においては、確率変数の期待値は統計モデルに期待値座標系を指定することに対応した。変形指数型分布族の場合には、適切に期待値を修正することが自然であることを確かめる。また独立性の修正についても、最尤法の一般化が自然に定義され、ダイバージェンスによる射影定理など指数型分布族の情報幾何学で知られていた結果が、変形指数型分布族に対して直接的に拡張できることを確かめる。

なお本論文は、日本数学会における講演要旨 [11]、福岡大学微分幾何研究会の報告集 [15] などを加筆、修正したものである。

2 変形指数型分布族

初めに変形指数関数と変形指数型分布族の定義を与える。詳しくは [17], [19] または [21] などを参照されたい。変形指数型分布族の幾何学は、機械学習理論では U -幾何学と呼ばれている [20]。

χ を $\mathbf{R}_{>0}$ から $\mathbf{R}_{>0}$ への狭義単調増加関数とする。 χ -対数関数 (または変形対数関数) を次で定義する。

$$\ln_{\chi} s := \int_1^s \frac{1}{\chi(t)} dt.$$

$\ln_{\chi} s$ も狭義単調増加関数であり $\ln_{\chi} 1 = 0$ を満たす。

χ -対数関数の逆関数は χ -指数関数 (または変形指数関数) とよび

$$\exp_{\chi} t := 1 + \int_0^t \lambda(s) dx$$

によって与えられる。ただし関数 $\lambda(s)$ は関係式 $\lambda(\ln_{\chi} s) = \chi(s)$ で定義される。 $\chi(t)$ および $\lambda(s)$ をそれぞれ χ -対数関数, χ -指数関数の変形関数とよぶことにする。

例 2.1 変形関数 $\chi(s)$ が冪関数 $\chi(s) = s^q$ ($q > 0, q \neq 1$) の場合, χ -対数関数および χ -指数関数は次で与えられる。

$$\begin{aligned} \ln_q s &:= \frac{s^{1-q} - 1}{1-q}, & (s > 0), \\ \exp_q t &:= (1 + (1-q)t)^{\frac{1}{1-q}}, & (1 + (1-q)t > 0). \end{aligned}$$

$\ln_q s$ を q -対数関数, $\exp_q t$ を q -指数関数とよぶ (cf. [25], [27]). $q \rightarrow 1$ の極限で, それぞれ通常対数関数と通常指数関数に一致する. \triangle

例 2.2 変形関数 $\chi(s)$ が $\chi(s) = \frac{2s}{s^{\kappa} + s^{-\kappa}}$, ($-1 < \kappa < 1, \kappa \neq 0$) の場合, χ -指数関数と χ -対数関数は次で与えられる。

$$\begin{aligned} \ln_{\kappa} s &:= \frac{s^{\kappa} - s^{-\kappa}}{2\kappa}, & (s > 0), \\ \exp_{\kappa} t &:= (\kappa t + \sqrt{1 + \kappa^2 t^2})^{\frac{1}{\kappa}}. \end{aligned}$$

$\ln_{\kappa} s$ を κ -対数関数, $\exp_{\kappa} t$ を κ -指数関数とよぶ (cf. [8]). $\kappa \rightarrow 0$ の極限で, それぞれ通常対数関数と通常指数関数に一致する. \triangle

統計モデルを S_χ とする. パラメータに関する識別可能性など, 情報幾何学で一般的に要請させる仮定は, ここでも成り立っているとする. 統計モデル S_χ が次で与えられるとき, χ -指数型分布族 (または変形指数型分布族) という.

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

ただし $F_1(x), \dots, F_n(x)$ は標本空間 Ω 上の関数, $\theta = \{\theta^1, \dots, \theta^n\}$ はパラメータで $\psi(\theta)$ はパラメータ θ に関する確率密度関数の正規化項である. また, 確率変数に関する正規化は不要となるように, Ω 上に適切な測度を仮定している.

χ -指数型分布族の変形指数関数として q -指数関数 \exp_q を取ったものを q -指数型分布族, κ -指数関数 \exp_κ を取ったものを κ -指数型分布族とよぶ.

通常の指数型分布族の場合と同様に, χ -指数型分布族 S_χ を $\{\theta^i\}$ を局所座標系とする多様体とみなす. χ -指数型分布族に対する詳細な仮定は [21] などを参照されたい. さらに本論文では ψ は狭義凸関数であることを仮定する.

例 2.3 (q -正規分布族) q を $1 < q < 3$ を満たす定数とする. $\Omega = \mathbf{R}$ 上の確率密度関数 $p_q(x; \mu, \sigma)$ が次で与えられるとき, $p_q(x; \mu, \sigma)$ を q -正規分布 (または自由度 $(3-q)/(q-1)$ の Student t -分布) とよぶ.

$$p_q(x; \mu, \sigma) := \frac{1}{Z_q(\sigma)} \left[1 - \frac{1-q}{3-q} \frac{(x-\mu)^2}{\sigma^2} \right]^{\frac{1}{1-q}},$$

ただし (μ, σ) は, それぞれ $-\infty < \mu < \infty$ および $0 < \sigma < \infty$ となるパラメータである. また $Z_q(\sigma)$ は確率分布の正規化項で

$$Z_q := \frac{\sqrt{3-q}}{\sqrt{q-1}} \text{Beta} \left(\frac{3-q}{2(q-1)}, \frac{1}{2} \right) \sigma$$

である. q -正規分布は $q \rightarrow 1$ の極限で通常の正規分布に一致する.

S_q を q -正規分布の全体とすると, これは q -指数型分布族である. 実際

$$\theta^1 := \frac{2}{3-q} \{Z_q(\sigma)\}^{q-1} \frac{\mu}{\sigma^2}, \quad \theta^2 := -\frac{1}{3-q} \{Z_q(\sigma)\}^{q-1} \frac{1}{\sigma^2}$$

とおき, q -正規分布の q -対数を考えると q -指数型分布であることがわかる.

$$\begin{aligned} \ln_q p_q(x) &= \frac{1}{1-q} (\{p_q(x)\}^{1-q} - 1) \\ &= \frac{1}{1-q} \left\{ \frac{1}{\{Z_q(\sigma)\}^{1-q}} \left(1 - \frac{1-q}{3-q} \frac{(x-\mu)^2}{\sigma^2} \right) - 1 \right\} \\ &= \frac{2\mu \{Z_q(\sigma)\}^{q-1}}{(3-q)\sigma^2} x - \frac{\{Z_q(\sigma)\}^{q-1}}{(3-q)\sigma^2} x^2 - \frac{\{Z_q(\sigma)\}^{q-1} \mu^2}{3-q} \frac{1}{\sigma^2} + \frac{\{Z_q(\sigma)\}^{q-1} - 1}{1-q} \\ &= \theta^1 x + \theta^2 x^2 - \psi(\theta), \end{aligned}$$

ただし $\psi(\theta) := -\frac{(\theta^1)^2}{4\theta^2} - \frac{\{Z_q(\sigma)\}^{q-1} - 1}{1-q}$ である. \triangle

q が $-\infty < q < 1$ の場合には $\text{Supp } p_q(x; \mu, \sigma)$ がパラメータによって変化するので、情報幾何学でいう統計モデルとはならないが、確率密度関数族は定義される。その場合の正規化項は

$$Z_q(\sigma) = \frac{\sqrt{3-q}}{\sqrt{1-q}} \text{Beta} \left(\frac{2-q}{1-q}, \frac{1}{2} \right) \sigma$$

となる。

例 2.4 (多変量 q -正規分布族) $\Omega = \mathbf{R}^d$ を標本空間とする。 q を $1 < q < 1 + 2/d$ を満たす定数とし、 $\nu = -d - 2/(1-q)$ とする。 Ω 上の確率密度関数 $p_q(x; \mu, \Sigma)$ が次で与えられるとき、 $p_q(x; \mu, \Sigma)$ を多変量 q -正規分布 (または自由度 ν の d -次元 Student t -分布とよぶ。

$$p_q(x; \mu, \Sigma) := \frac{\Gamma\left(\frac{1}{q-1}\right)}{(\pi\nu)^{\frac{d}{2}} \Gamma\left(\frac{\nu}{2}\right) \sqrt{\det(\Sigma)}} \left[1 + \frac{1}{\nu} {}^t(x - \mu) \Sigma^{-1} (x - \mu) \right]^{\frac{1}{1-q}},$$

ただし $X = {}^t(X_1, \dots, X_d)$ は Ω 上の確率変数ベクトル、 \mathbf{R}^d , $\mu = {}^t(\mu^1, \dots, \mu^d)$ は \mathbf{R}^d 上の擬平均ベクトル、 Σ は $\text{Sym}^+(d)$ 上の擬共分散行列である。 $d=1$ の場合、

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \text{Beta}(s, t) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$$

に注意すると、先述の q -正規分布に一致する。

簡単のため Σ は正則であると仮定する。 Σ が非正則な場合は $\text{Sym}^+(d)$ 上で適切な基底 $\{v^\alpha\}$ を選び $\Sigma = \sum_\alpha w_\alpha v^\alpha$ などと表示する必要がある。

S_q を多変量 q -正規分布の全体とすると、これは q -指数型分布族である。実際

$$z_q = \frac{(\pi\nu)^{\frac{d}{2}} \Gamma\left(\frac{\nu}{2}\right) \sqrt{\det(\Sigma)}}{\Gamma\left(\frac{1}{q-1}\right)}, \quad \tilde{R} = \frac{z_q^{q-1}}{(1-q)d+2} \Sigma^{-1}, \quad \text{and} \quad \theta = 2\tilde{R}\mu \quad (1)$$

とおくと

$$\begin{aligned} p_q(x; \mu, \Sigma) &= \frac{1}{z_q} \left[1 + \frac{1}{\nu} {}^t(x - \mu) \Sigma^{-1} (x - \mu) \right]^{\frac{1}{1-q}} \\ &= \left[\left(\frac{1}{z_q} \right)^{1-q} - \frac{1-q}{(1-q)d+2} \left(\frac{1}{z_q} \right)^{1-q} {}^t(x - \mu) \Sigma^{-1} (x - \mu) \right]^{\frac{1}{1-q}} \\ &= \exp_q \left[-{}^t(x - \mu) \tilde{R} (x - \mu) + \ln_q \frac{1}{z_q} \right] \\ &= \exp_q \left[\sum_{i=1}^d \theta^i x_i - \sum_{i=1}^d \tilde{R}_{ii} x_i^2 - 2 \sum_{i < j} \tilde{R}_{ij} x_i x_j - \frac{1}{4} {}^t \theta \tilde{R}^{-1} \theta + \ln_q \frac{1}{z_q} \right] \end{aligned}$$

となる。 $\theta \in \mathbf{R}^d$ と $\tilde{R} \in \text{Sym}^+(d)$ 、また Σ が正則であることの仮定から、多変量 q -正規分布の全体 S_q は $d(d+3)/2$ -次元の q -指数型分布族である。 \triangle

χ -指数関数と χ -対数関数の議論では指数関数を 1 関数分の自由度で拡張しているが、2 関数、またはそれ以上の関数で拡張することも可能である。2 関数分の自由度を用いた拡張では、たとえば [29, 30] などを参照されたい。

3 変形代数と擬加法的微分

次に変形指数関数を特徴付ける変形代数構造と擬加法的微分方程式を考える。変形指数関数は、そのままでは指数法則を満たさない。そこで代数構造の変形を行う。

まず q -指数関数と q -対数関数の場合を考える (cf. [3]). q -和 $\tilde{\oplus}^q$ を

$$\begin{aligned} x_1 \tilde{\oplus}^q x_2 &:= x_1 + x_2 + (1 - q)x_1 x_2 \\ &= \ln_q(\exp_q x_1 \cdot \exp_q x_2) \end{aligned}$$

によって定義する。ただし q -指数関数に関する真数条件 $1 + (1 - q)x_1 > 0$ および $1 + (1 - q)x_2 > 0$ を仮定する。一方 q -積 \otimes_q を

$$\begin{aligned} y_1 \otimes_q y_2 &:= [y_1^{1-q} + y_2^{1-q} - 1]^{\frac{1}{1-q}} \\ &= \exp_q(\ln_q y_1 + \ln_q y_2) \end{aligned}$$

と定義する。この場合 q -対数関数と q -指数関数の真数条件 $y_1 > 0$, $y_2 > 0$ および $y_1^{1-q} + y_2^{1-q} - 1 > 0$ を仮定する。

以上の定義と仮定のもと、 q -指数関数と q -対数関数には

$$\begin{aligned} \exp_q(x_1 \tilde{\oplus}^q x_2) &= \exp_q x_1 \cdot \exp_q x_2, \\ \ln_q(y_1 \cdot y_2) &= \ln_q y_1 \tilde{\oplus}^q \ln_q y_2, \\ \exp_q(x_1 + x_2) &= \exp_q x_1 \otimes_q \exp_q x_2, \\ \ln_q(y_1 \otimes_q y_2) &= \ln_q y_1 + \ln_q y_2 \end{aligned}$$

が成り立ち、指数法則の拡張が可能である。ここで q -和は定義域に作用し、 q -積は値域に作用していることに注意する。この作用の違いが、変形指数型分布族の理解には重要であると考えている。

なお、指数関数と対数関数の拡張には本論文で扱うもの以外にも、多くの手法が存在する。例えば代表的なものとして q -類似がある。Jackson の q -微分から指数関数を定義し、指数法則には通常のと積を用いるが、その結果、定義域の積が非可換となる代数構造を用いるものである。

q -指数関数に関する真数条件の仮定から q -和に関する逆元が存在し

$$[-x]_q := \ln_q \left(\frac{1}{\exp_q x} \right) = \frac{-x}{1 + (1 - q)x}$$

で与えられる。この逆元 $[-x]_q$ を用いて q -差 $\tilde{\ominus}^q$ を

$$x_1 \tilde{\ominus}^q x_2 := x_1 \tilde{\oplus}^q [-x_2]_q$$

と定義し、さらに q -差に関する極限から擬加法的 q -微分を次式で定義する.

$$\frac{d_q}{d_q x} f(x) := \lim_{x' \rightarrow x} \frac{f(x') - f(x)}{x' \ominus^q x}.$$

擬加法的 q -微分に関する微分方程式を考え、その固有関数を求める.

$$\frac{d_q}{d_q x} f(x) = f(x). \quad (2)$$

この変形常微分方程式を $f(0) = 1$ の初期条件のもとで解くと、 q -指数関数

$$f(x) = (1 + (1 - q)x)^{\frac{1}{1-q}} \quad (= \exp_q x)$$

が得られる.

一方 $q > 0$ を定数とし、次の非線形常微分方程式を考える.

$$\frac{d}{dx} f(x) = f(x)^q. \quad (3)$$

この常微分方程式を $f(0) = 1$ の初期条件のもとで解くと $1 + (1 - q)x > 0$ のとき、先程と同様に q -指数関数 $f(x) = \exp_q x$ が得られる.

変形常微分方程式 (2) と非線形常微分方程式 (3) のどちらからも q -指数関数が得られるが、(2) を q -指数関数の擬加法的描像に関する微分方程式、(3) をエスコート描像に関する微分方程式とよぶことにする.

通常の数関数は $x > 0$ のとき

$$\exp x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$$

と無限積で表示できた. q -指数関数の場合には次が成り立つ.

命題 3.1 (cf. [25]) $q > 0$ を定数とする. 任意の自然数 n に対して、真数条件

$$n \left(1 + \frac{x}{n}\right)^{1-q} - (n-1) > 0$$

が成り立つならば

$$\exp_q x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^{\otimes_q n}$$

となる. ただし

$$\left(1 + \frac{x}{n}\right)^{\otimes_q n} := \underbrace{\left(1 + \frac{x}{n}\right) \otimes_q \cdots \otimes_q \left(1 + \frac{x}{n}\right)}_{n \text{ 個}}$$

である. \square

つぎに κ -指数関数の場合を考える. 変形した代数構造 κ -和 $\tilde{\oplus}^\kappa$ と κ -積 \otimes_κ を

$$\begin{aligned} x_1 \tilde{\oplus}^\kappa x_2 &:= \ln_\kappa [\exp_\kappa x_1 \cdot \exp_\kappa x_2] \\ &= x_1 \sqrt{1 + \kappa^2 x_2^2} + x_2 \sqrt{1 + \kappa^2 x_1^2}, \\ y_1 \otimes_\kappa y_2 &:= \exp_\kappa [\ln_\kappa y_1 + \ln_\kappa y_2] \end{aligned}$$

と定義する. ただし κ -対数関数に対する真数条件 $y_1 > 0, y_2 > 0$ を仮定する. q -指数関数, q -対数関数の場合と同様に指数法則の拡張が成り立つ.

$$\begin{aligned} \exp_\kappa(x_1 \tilde{\oplus}^\kappa x_2) &= \exp_\kappa x_1 \cdot \exp_\kappa x_2, \\ \ln_\kappa(y_1 \cdot y_2) &= \ln_\kappa y_1 \tilde{\oplus}^\kappa \ln_\kappa y_2, \\ \exp_\kappa(x_1 + x_2) &= \exp_\kappa x_1 \otimes_\kappa \exp_\kappa x_2, \\ \ln_\kappa(y_1 \otimes_\kappa y_2) &= \ln_\kappa y_1 + \ln_\kappa y_2. \end{aligned}$$

κ -和に関する x の逆元は $-x$ となるので, κ -差 $\tilde{\ominus}^\kappa$ は

$$x_1 \tilde{\ominus}^\kappa x_2 := x_1 \tilde{\oplus}^\kappa (-x_2)$$

と定義する. この結果から擬加法的 κ -微分を次式で定義する.

$$\frac{d_\kappa}{d_\kappa x} f(x) := \lim_{x' \rightarrow x} \frac{f(x') - f(x)}{x' \tilde{\ominus}^\kappa x}.$$

擬加法的 κ -微分に関する次の微分方程式

$$\frac{d_\kappa}{d_\kappa x} f(x) = f(x) \quad (4)$$

を考え, この変形常微分方程式 (4) を初期条件 $f(0) = 1$ のもとで解くと, κ -指数関数

$$f(x) = (\sqrt{1 + \kappa^2 x^2} + \kappa x)^{\frac{1}{\kappa}} \quad (= \exp_\kappa x)$$

が得られる. この (4) が κ -指数関数の擬加法的描像に関する微分方程式である.

一方 κ ($-1 < \kappa < 1$) を定数とし, 次の非線形常微分方程式を考える.

$$\frac{d}{dx} f(x) = \frac{1}{\sqrt{1 + \kappa^2 x^2}} f(x). \quad (5)$$

この非線形常微分方程式を初期条件 $f(0) = 1$ のもとで解くと, 同じく κ -指数関数 $f(x) = \exp_\kappa x$ が得られる. (5) が κ -指数関数のエスコート描像に関する微分方程式である.

κ -指数関数の無限積表現は次で与えられる.

命題 3.2 (cf. [19]) $-1 < \kappa < 1$ を定数とする. $x > -1$ のとき

$$\exp_{\kappa} x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^{\otimes_{\kappa} n}$$

となる. ただし

$$\left(1 + \frac{x}{n}\right)^{\otimes_{\kappa} n} := \underbrace{\left(1 + \frac{x}{n}\right) \otimes_{\kappa} \cdots \otimes_{\kappa} \left(1 + \frac{x}{n}\right)}_{n \text{ 個}}$$

である.

κ -変形代数は和に関する対称性が高いため, 真数条件が容易になる. 一般の変形指数関数の場合でも真数条件を記述できれば同様の議論ができるが, 一般の χ -関数の場合に真数条件を具体的に記述することは難しい.

4 期待値汎関数と擬加法的積分

前章で見たように, 変形指数関数を考える場合には標本空間の代数構造が変化していると捉えると自然である. この章では, 確率変数の期待値に関する修正を考える.

Ω を標本空間とし, 確率変数 X は確率密度関数 $p(x)$ で与えられる確率分布に従うとする. また $f(x)$ を Ω 上の関数とする. このとき

$$E_p[f(x)] := \int_{\Omega} f(x)p(x)dx$$

を $f(x)$ の単純期待値 (または (標準) 期待値) とよぶ.

指数型分布族など適切な条件のもとでは, 確率変数の期待値は統計モデルの座標系を与えることに相当する. 多様体の座標系は自由に取り替えることができるように, 単純期待値は幾何学的には特別な意味を持たない. そこで, 変形指数型分布族の場合には次の期待値を考える.

S_{χ} を χ -指数型分布族とし $p(x; \theta) \in S_{\chi}$ とする. このとき $p(x; \theta)$ の χ -エスコート分布 $P_{\chi}(x; \theta)$, および正規化した χ -エスコート分布 $P_{\chi}^{esc}(x; \theta)$ をそれぞれ

$$P_{\chi}(x; \theta) := \chi\{p(x; \theta)\},$$

$$P_{\chi}^{esc}(x; \theta) := \frac{1}{Z_{\chi}(\theta)} \chi\{p(x; \theta)\}, \quad \text{ただし } Z_{\chi}(\theta) := \int_{\Omega} \chi\{p(x; \theta)\} dx$$

で定義する. 関数 $f(x)$ のこれらエスコート分布に関する期待値を, χ -正準期待値, および正規化した χ -エスコート期待値とよび, それぞれ $E_{\chi, p}[f(x)]$, $E_{\chi, p}^{esc}[f(x)]$ で表す.

$$E_{\chi, p}[f(x)] := \int_{\Omega} f(x)P_{\chi}(x; \theta)dx = \int_{\Omega} f(x)\chi(p(x; \theta))dx,$$

$$E_{\chi, p}^{esc}[f(x)] := \int_{\Omega} f(x)P_{\chi}^{esc}(x; \theta)dx = \frac{1}{Z_{\chi}(\theta)} \int_{\Omega} f(x)\chi(p(x; \theta))dx,$$

特に q -指数型分布族 S_q に対する χ -正準期待値と正規化した χ -エスコート期待値を, それぞれ q -正準期待値, q -エスコート期待値とよび $E_{q,p}[f(x)], E_{q,p}^{esc}[f(x)]$ で表す. 一方, κ -指数型分布族 S_κ に対しては, それぞれ κ -正準期待値, κ -エスコート期待値とよび $E_{\kappa,p}[f(x)], E_{\kappa,p}^{esc}[f(x)]$ で表す.

次に変形代数との関連を見るために, 擬加法的積分を定義しよう. Ω 上の関数 $f(x)$ に対し, 重み付き積分の一つとして擬加法的 κ -積分を次の式で定義する.

$$\int f(x) d_\kappa x := \int \frac{f(x)}{\sqrt{1 + \kappa^2 x^2}} dx = \int f(x) w_\kappa(x) dx, \quad (6)$$

ただし $w(x)$ は積分の重み関数で $w_\kappa(x) = 1/\sqrt{1 + \kappa^2 x^2}$ である. Ω が離散集合 $\Omega = \{x_0, x_1, \dots, x_N\}$ の場合には擬加法的 κ -和を

$$\sum_{i=0}^N \kappa f(x_i) := \sum_{i=0}^N \frac{f(x_i)}{\sqrt{1 + \kappa^2 x_i^2}} = \sum_{i=0}^N f(x_i) w(x_i)$$

と定義する.

定理 4.1 $\chi(s)$ を κ -対数関数に対する変形関数 $\chi(s) = 2s/(s^\kappa + s^{-\kappa})$ とする. このとき $\chi(\exp_\kappa x)$ は擬加法的 κ -積分 (擬加法的 κ -和) の重み関数 $w(x)$ に一致する. すなわち

$$\chi(\exp_\kappa x) = \frac{1}{\sqrt{1 + \kappa^2 x^2}} = w_\kappa(x)$$

が成り立つ. \square

他の変形指数型分布族の場合にも, 真数条件等が成り立てば同様の議論ができるが, ここでは省略する. また, 本章の内容とは異なる考え方として, 標本空間の平行移動から正準期待値を定義する方法も報告されている [26].

5 χ -指数型分布族の幾何学

この章では χ -指数型分布族に自然に入る幾何学構造を考察する. 詳しくは [17], および [19]などを参照されたい. なお, 著者のこれまでの論文 [11]–[19] などとはエスコート分布の定義が異なるので注意されたい.

まずは推定関数から双対平坦構造を構成する方法を解説する. S_χ を χ -指数型分布族とする. このとき $p(x; \theta) \in S_\chi$ の χ -スコア関数 $s^\chi(x; \theta) : S_\chi \rightarrow \mathbf{R}^n$, $s^\chi(x; \theta) = {}^t((s^\chi)^1(x; \theta), \dots, (s^\chi)^n(x; \theta))$ を次で定める.

$$(s^\chi)^i(x; \theta) := \frac{\partial}{\partial \theta^i} \ln_\chi p(x; \theta), \quad (i = 1, \dots, n). \quad (7)$$

χ -スコア関数は χ -正準期待値, および χ -エスコート期待値に関して不偏な推定関数である. すなわち

$$E_{\chi,p}[s^\chi(x; \theta)] = E_{\chi,p}^{esc}[s^\chi(x; \theta)] = 0$$

が成り立つ。一方, χ -スコア関数を単純期待値に関して不偏化し

$$(u_p^\chi)^i(x; \theta) := \frac{\partial}{\partial \theta^i} \ln_\chi p(x; \theta) - E_p \left[\frac{\partial}{\partial \theta^i} \ln_\chi p(x; \theta) \right] \quad (8)$$

によって定めた推定関数 $(u_p^\chi)(x; \theta)$ を, $p(x; \theta)$ のバイアス補正 χ -スコア関数 とよぶ。適当な正則条件のもとで χ -スコア関数から S_χ 上の Riemann 計量が定義できる。

$$\begin{aligned} g_{ij}^E(\theta) &:= \int_{\Omega} \partial_i \ln_\chi p(x; \theta) \partial_j \ln_\chi p(x; \theta) \chi\{p(x; \theta)\} dx \\ &= E_{\chi, p}[(s^\chi)^i(x; \theta)(s^\chi)^j(x; \theta)], \end{aligned} \quad (9)$$

$$g_{ij}^M(\theta) := \int_{\Omega} \partial_i p(x; \theta) \partial_j \ln_\chi p(x; \theta) dx, \quad (10)$$

$$g_{ij}^N(\theta) := \int_{\Omega} \frac{1}{\chi\{p(x; \theta)\}} \partial_i p(x; \theta) \partial_j p(x; \theta) dx. \quad (11)$$

通常の不変 Fisher 計量の場合と同様に, (9)–(11) で定まる Riemann 計量に対して次が成り立つ。

定理 5.1 χ -指数型分布族 S_χ 上の Riemann 計量 g^E, g^M, g^N は等しい, すなわち

$$g^E(\theta) = g^M(\theta) = g^N(\theta)$$

が成り立つ。□

これらの Riemann 計量の定義には, 正規化していない χ -エスコート分布を用いていることに注意されたい。正規化した χ -エスコート分布を用いた場合, 3種類の Riemann 計量は異なるものとなり, 互いに共形同値となる [13]。

さらに Riemann 計量 g^M の定義式 (10) の両辺を微分することで, S_χ 上の捩れない互いに双対的なアフライン接続 $\nabla^{M(e)}$ と $\nabla^{M(m)}$ が定義でき, 平坦であることが示される。したがって $(S_\chi, g^M, \nabla^{M(e)}, \nabla^{M(m)})$ は双対平坦空間となる [22, 17]。

上述の双対平坦構造を誘導するダイバージェンス (すなわち正準ダイバージェンス) は χ -ダイバージェンス (または U -ダイバージェンス) であり次で定義される。

$$D_\chi(p, r) = \int_{\Omega} \{U_\chi(\ln_\chi r(x)) - U_\chi(\ln_\chi p(x)) - p(x)(\ln_\chi r(x) - \ln_\chi p(x))\} dx.$$

ただし U_χ は

$$U_\chi(s) := \int_0^s \exp_\chi(t) dt.$$

によって定義される狭義凸関数である。

$\chi(s)$ が冪関数 $\chi(s) = s^q$ の場合には χ -ダイバージェンスは β -ダイバージェンス ($\beta = 1 - q$) に一致し, さらに $\chi(s) = s$ であれば Kullback-Leibler ダイバージェンスである。また, χ -ダイバージェンスはバイアス補正した χ -スコア関数から構成さ

れるダイバージェンスである [20], [17]. (なお, ダイバージェンスから統計構造, または双対平坦構造を誘導する方法については [4], [1] などを参照されたい.)

次に χ -指数型分布族 S_χ のパラメータに関する正規化項 ψ から, S_χ に双対平坦構造を構成する (cf. [2]). S_χ を χ -指数型分布族とし, $\psi(\theta)$ を χ -指数型分布族のパラメータに関する正規化項とする. このとき ψ の微分から S_χ 上に Riemann 計量と 3 次形式が定義される.

$$g_{ij}^\chi(\theta) := \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\theta), \quad C_{ijk}^\chi(\theta) := \frac{\partial^3 \psi}{\partial \theta^i \partial \theta^j \partial \theta^k}(\theta).$$

g^χ はその定義から S_χ 上の Hesse 計量である. よって Hesse 幾何学の標準的な手法で双対的な平坦接続 $\nabla^{\chi(e)}$ と $\nabla^{\chi(m)}$ が定義され [17, 24], 結果的に $(S_\chi, g^\chi, \nabla^{\chi(e)}, \nabla^{\chi(m)})$ が双対平坦空間となる.

この双対平坦構造を誘導するダイバージェンスは χ -相対エントロピー (または一般化した相対エントロピー) である.

$$D^\chi(p, r) := E_{\chi, p}^{esc}[\ln_\chi p(x) - \ln_\chi r(x)].$$

D^χ は χ -スコア関数から正規化した χ -エスコート期待値に関して構成したダイバージェンスであることも知られている [17].

特に q -指数型分布族 S_q の場合には, q -スコア関数 (χ -スコア関数の変形対数関数として, q -対数関数を取ったもの) を推定関数として, 正準期待値から α -ダイバージェンス $D^{(1-2q)}$ ($\alpha = 1 - 2q$) が構成できる.

$$D^{(1-2q)}(p, r) := \frac{1}{q} E_{q, p}[\ln_q p(x) - \ln_q r(x)].$$

α -ダイバージェンス $D^{(1-2q)}$ からは, 不変統計多様体 $(S_q, \nabla^{(1-2q)}, g^F)$ が誘導される. ここで $\nabla^{(1-2q)}$ は S_q 上の α -接続 ($\alpha = 1 - 2q$), g^F は Fisher 計量である. 一般には不変統計多様体は平坦にはならないが, 確率変数の変数変換に対する不変性がある.

一方, 期待値の取り方を変更して, 正規化した q -エスコート期待値からは正規化した Tsallis 相対エントロピーが構成できる.

$$D_T^q(p, r) := E_{q, p}^{esc}[\ln_q p(x) - \ln_q r(x)].$$

このダイバージェンスからは双対平坦構造 $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ が誘導される.

α -ダイバージェンス $D^{(1-2q)}$ と正規化した Tsallis 相対エントロピー $D_T^q(p, r)$ の違いは, 本質的にはエスコート分布の正規化項だけであるが, 誘導される統計多様体の構造には違いがある. 特に不変統計多様体 $(S_q, \nabla^{(2q-1)}, g^F)$ と平坦統計多様体 $(S_q, \nabla^{q(e)}, g^q)$ には 1-共形同値という関係が成り立ち, $(S_q, \nabla^{(2q-1)}, g^F)$ は 1-共形平坦統計多様体である. (cf. [17]. 1-共形同値性の定義は [10] も参照されたい.)

6 独立性の一般化

この章では確率変数の独立性の概念を修正し、変形指数型分布族に適した構造を考える。

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。同時確率分布 $p(x, y)$ が周辺確率分布の積として

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

と表されるとき、確率変数 X, Y を独立とよんだ。 $p(x) > 0, p(y) > 0$ 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、確率変数の独立性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る。またこの式には、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], [25]).

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_i を Ω_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 -正規化のもとで) q -独立とよぶ。ただし Z_{p_1, p_2, \dots, p_N} は (m -正規化) 定数で

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

によって定義される。

一方、同時確率分布 $p(x_1, x_2, \dots, x_N)$ が

$$p(x_1, x_2, \dots, x_N) = p_1(x_1) \otimes_q p_2(x_2) \otimes_q \cdots \otimes_q p_N(x_N) \otimes_q (-c)$$

と与えられるとき、確率変数 X_1, X_2, \dots, X_N は e -正規化のもとで q -独立とよぶ。ただし c は e -正規化定数で

$$\int \cdots \int_{\text{Supp}\{p(x_1, x_2, \dots, x_N)\} \subset \Omega_1 \times \cdots \times \Omega_N} p_1(x_1) \otimes_q p_2(x_2) \otimes_q \cdots \otimes_q p_N(x_N) \otimes_q (-c) dx_1 \cdots dx_N = 1$$

によって定義される。 q -指数型分布の場合には、 m -正規化可能であることと e -正規化可能であることは同値である。

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. したがって q -独立な確率変数 X_1, X_2, \dots, X_N は, 通常の独立性のもとでは独立ではない.

さて, X と Y をそれぞれ \mathbf{R} 上の q -正規分布 $p_q(x)$ と $p_q(y)$ に従う確率変数とする. X と Y が独立であれば, 同時確率分布は $p(x, y) = p_q(x)p_q(y)$ で与えられるが, この場合 $p(x, y)$ は 2 変量 q -正規分布とはならない. 1 変量の q -正規分布と多変量の q -正規分布の関係を自然に記述するためには, q -独立性が必要である.

定理 6.1 (cf. [23]) q を $1 < q < 2$ を満たす定数とし, X と Y をそれぞれ \mathbf{R} 上の q -正規分布 $p_q(x)$ と $p_q(y)$ に従う確率変数とする. このとき X と Y が e -正規化のもとで q -独立 (または m -正規化のもとで q -独立) となるような 2 変量 q -正規分布 $p_q(x, y)$ が存在する. \square

この定理において e -正規化定数も陽に記述することもできるが, ここでは省略する.

通常の正規分布の場合には, 独立な確率変数を考えると多変量正規分布は次元の低い正規分布の積に分解ができた. 変形指数型分布族の場合にはこのような性質は成り立たず, 独立性の修正, すなわち関数空間での代数構造の修正が必要となる.

q -独立性には多くの定義が存在する. 本論文では確率変数間に相互作用が存在していることを想定している. 一方, 観測データの小標本性に起因する現象に対しては, 観測データ数の増加に伴い q の値が変化し, $q \rightarrow 1$ ($N \rightarrow \infty$) となる独立性が自然である.

q -中心極限定理に関する論文 [28] では q -Fourier 変換から特性関数の q -変形版を定義し, 特性関数の持つ性質の修正から q -独立性を定義している.

本論文の最後に, q -独立性のもとでの最尤法の修正を考察する.

$S_q = \{p(x; \theta) | \theta \in \Theta\}$ を q -指数型分布族とし, $\{x_1, \dots, x_N\}$ を $p(x; \theta) \in S_q$ 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 $L_q(\theta)$ を

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

によって定義する. ただし q -積が定義できるための真数条件等は成り立っていると仮定する. $q \rightarrow 1$ の極限では L_q は通常の尤度関数に収束するが, $q \neq 1$ の場合には L_q を同時確率分布と解釈することはできない.

q -対数関数の単調性から, L_q の同値な条件として q -対数尤度関数 $l_q(\theta)$

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

を考えることもできる.

q -尤度関数の最大値を与える引数 $\hat{\theta}$ を q -最尤推定量と定義する.

$$\hat{\theta} := \operatorname{argmax}_{\theta \in \Theta} L_q(\theta) \quad \left(= \operatorname{argmax}_{\theta \in \Theta} \ln_q L_q(\theta) \right).$$

定理 6.2 (cf. [16, 17]) S_q を q -指数型分布族とし, $\{x_1, \dots, x_N\}$ を $p(x; \theta) \in S_q$ から生成される N 個の観測値とする. また $\{\eta_i\}$ を双対平坦構造 $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ の $\nabla^{q(m)}$ に関するアファイン座標系とする. このとき q -最尤推定量は S_q の η -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.

q の値が変わると統計モデルが変化し, 期待値や推定量の定義も変化するが, 標本の実現値が同じであれば推定値は一定である.

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, ϕ を $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ の η -座標系に関するポテンシャルとして

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

が成り立つ. したがって 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く, この事実も指数型分布族における最尤推定量に対する幾何学的解釈の一般化である. 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 -最尤法は幾何学的に見ても自然な概念である.

7 おわりに

本論文では指数関数, および対数関数を一般化し, 変形指数型分布族の持つ幾何学的, 統計学的性質を考察した. 変形指数関数と変形対数関数には指数法則が成り立たないため, 確率変数の定義域となる標本空間, および確率分布の値域となる関数空間の代数構造を変化させたが, この修正が期待値の定義や, 確率変数の独立性の定義に影響を与えた.

しかしながら, そもそも異常統計において変形指数型分布族が必要となる理由を複数考えることができ, 本論文は主に確率変数間に相互作用があると思われる現象を想定している. 一方, 観測データの少数性, 標本空間の小規模性に起因することも考えられる. このような場合には, 観測データの数の増加に対して独立性の定義を修正することが自然であると考えており, 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.

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] G. Kaniadakis, *Non-linear kinetics underlying generalized statistics*, Physica A, **296**(2001), 405–425.
- [8] G. Kaniadakis, *Theoretical foundations and mathematical formalism of the power-law tailed statistical distributions*, Entropy, **15**(2013), 3983-4010.
- [9] G. Kaniadakis and A.M. Scarfone, *A new one-parameter deformation of the exponential function*, Physica A, **305**(2002), 69–75.
- [10] T. Kurose, *On the divergences of 1-conformally flat statistical manifolds*, Tôhoku Math. J., **46**(1994), 427–433.
- [11] 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
- [12] H. Matsuzoe, *Statistical manifolds and geometry of estimating functions*. Prospects of Differential Geometry and Its Related Fields, World Scientific, (2013), 187–202.
- [13] H. Matsuzoe, *Hessian structures on deformed exponential families and their conformal structures*, Diff. Geom. Appl., **35**(2014), 323–333.
- [14] 松添 博, 変形指数型分布族の幾何学, 京都大学数理解析研究所講究録. **1916** (2014), 143–155.
- [15] 松添 博, 異常統計における Hesse 構造, 福岡大学微分幾何研究会報告集, (2014), 29–38.
- [16] H. Matsuzoe, *Information geometry of Bayesian statistics*, AIP conference proceedings, **1641**(2015), 279–286.
- [17] 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.

- [18] 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.
- [19] H. Matsuzoe and T. Wada, *Deformed algebras and generalizations of independence on deformed exponential families*, preprint.
- [20] N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, *Information Geometry of U-Boost and Bregman Divergence*, Neural Comput., **16**(2004), 1437-1481.
- [21] J. Naudts, *Generalised Thermostatistics*, Springer-Verlag, 2011.
- [22] 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.
- [23] M. Sakamoto and H. Matsuzoe *A generalization of independence and multivariate Student's t -distributions*, preprint.
- [24] H. Shima, *The Geometry of Hessian Structures*, World Scientific, 2007.
- [25] 須鎗弘樹, 複雑系のための基礎数理 — ベキ乗則とツァリスエントロピーの数理 —, 牧野書店. 2010.
- [26] 田中 勝, ゲージ理論としての情報幾何学, 福岡大学微分幾何研究会報告集, (2014), 39-54.
- [27] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, New York, 2009.
- [28] S. Umarov, C. Tsallis and S. Steinberg, *On a q -central limit theorem consistent with nonextensive statistical mechanics*, Milan J. Math., **76**(2008), 307-328.
- [29] J. Zhang, *Referential duality and representational duality on statistical manifolds*, in Proceedings of the Second International Symposium on Information Geometry and Its Applications, 2005, 58-67.
- [30] J. Zhang, *On monotone embedding in information geometry*, preprint.