複雑系科学における統計的推論の幾何学

松添 博 (Hiroshi Matsuzoe) *

2011 年 4 月 14 日版

1 はじめに

地震の発生頻度や株価変動の分布など，複雑系科学に現れる確率分布は，確率の減衰が遅いものも多い．確率分布の「裾が長い」,「裾が重い」などの表現がされるが，このような確率分布では，確率変数の平均や分散が定義できないこともある．しかしながら平均や分散という概念は従来の指数型の確率分布に則した表現であり，冪分布をはじめとする非指数型の分布の表現には適さない．

そこで本論文では，微分幾何学を用いた非指数型分布の表現法や，統計推論の手法について解説する．

2 統計モデルの幾何学

初めに，簡単に統計モデルの幾何学を解説する．詳しくは情報幾何学に関する文献などを参照されたい [1]．$(\Omega, \mathcal{F}, \mathcal{P})$ を確率空間とし，Ξ を $\boldsymbol{R}^n$ の開集合とする．Ω 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．

$$S = \left\{ p(x;\xi) \,\middle|\, \int_\Omega p(x;\xi)dx = 1, p(x;\xi) > 0 \right\}.$$

適当な条件のもとで S を曲がった空間，すなわち多様体とみなし，$\{\xi^i\}$ をその局所座標系とする．

次に S に曲がった空間としての内積，すなわち多様体上の Riemann 計量を以下の式で定める．

$$g_{ij}^F(\xi) \quad := \quad E_\xi[\partial_i l_\xi \partial_j l_\xi].$$

ただし $E_\xi[f] = \displaystyle\int_\Omega f(x)p(x;\xi)dx$ は確率変数 $f(x)$ の $p(x;\xi)$ に関する期待値，$l_\xi = l(x;\xi) = \log p(x;\xi)$ は $p(x;\xi)$ の対数尤度，∂_i はパラメータ ξ に関する微分である．こうして定まる Riemann 計量を，特に Fisher 計量と呼ぶ．

さらに S の共変微分構造，すなわちアファイン接続を定める．$\alpha \in \boldsymbol{R}$ に対し

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

によって α-接続 $\nabla^{(\alpha)}$ を定義する．特に重要なものは 1-接続と (-1)-接続であり，$\nabla^{(e)} = \nabla^{(1)}$，$\nabla^{(m)} = \nabla^{(-1)}$ と書く．α-接続と β-接続の差から 3 次形式 C^F を

$$\frac{\alpha-\beta}{2}C_{ijk}^F = \Gamma_{ij,k}^{(\beta)} - \Gamma_{ij,k}^{(\alpha)}$$

と定義する．

有用な統計モデルとして，次の指数型分布族がある．

$$S = \left\{ p(x;\theta) \,\middle|\, p(x;\theta) = e^{Z(x)+\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)} \right\}.$$

ただし $\theta = \{\theta^1, \dots, \theta^n\}$ はパラメータ，$Z, F_1, \cdots, F_n$ は Ω 上の確率変数，ψ はパラメータ θ の関数である．以下，簡単のため $Z(x) = 0$ とするが Ω の座標変換によって一般性を失うことなく，このように置くことができる．

指数型分布族に対しては，1-接続と (-1)-接続は平坦な接続となる．すなわち，内積構造としては空間が曲がっているのであるが，微分構造としては空間が平らであるとみなすことができる．これは通常の Riemann 幾何学とは異なり直感も働かないが，応用上は目的に合わせて幾何構造をデザインすることも重要である．

$\{\theta^i\}$-座標系に関する Fisher 計量と 3 次形式 C は ψ をポテンシャルとして次の式で与えられる．

$$g_{ij}^F(\theta) \quad = \quad \partial_i\partial_j\psi(\theta), \tag{1}$$

$$C_{ijk}^F(\theta) \quad = \quad \partial_i\partial_j\partial_k\psi(\theta). \tag{2}$$

確率変数 $F_i(x)$ の期待値を η_i とすると $\{\eta_i\}$ も S の座標系となる．さらに関数 $\phi(\eta)$ を $\phi(\eta) = E[\log p(x;\eta)]$ で定めると $\{\eta_i\}$-座標系に関するポテンシャルとなり (1), (2) と同様な式が成り立つ．また，以下の関係式も得られる．

$$\frac{\partial\psi}{\partial\theta^i} = \eta_i, \quad \frac{\partial\phi}{\partial\eta_i} = \theta^i,$$

$$\psi(p) + \phi(p) - \sum_{i=1}^m \theta^i(p)\eta_i(p) = 0, \ (p(x) \in S).$$

また，S には正準ダイバージェンスと呼ばれる距離 2 乗型の関数も定まる．

$$\rho(p,q) \quad := \quad \psi(p) + \phi(q) - \sum_{i=1}^n \theta^i(p)\eta_i(q), \qquad p(x), q(x) \in S.$$

*名古屋工業大学大学院工学研究科・情報工学専攻／工学教育総合センター

今の場合はKullback-Leibler ダイバージェンス，または相対エントロピーと呼ばれる量と一致し

$$\rho(q,p) = E_p[\log p(x) - \log q(x)] \tag{3}$$

である．

3　q-指数型分布族とその双対平坦構造

この章では指数型分布族を拡張し[4)]，その幾何学構造を考える．まずはじめに，指数関数と対数関数の概念を拡張した q-指数関数と q-対数関数を定義する．正の定数 q を固定する．ここで $1+(1-q)x>0$ として

$$\exp_q x := \begin{cases} (1+(1-q)x)^{\frac{1}{1-q}}, & q \neq 1, \\ \exp x, & q = 1 \end{cases}$$

を q-指数関数,

$$\log_q x := \begin{cases} \frac{x^{1-q}-1}{1-q}, & q \neq 1,\ (x>0), \\ \log x, & q = 1 \end{cases}$$

を q-対数関数と呼ぶ．$q \to 1$ の極限を考えると，これらは通常の指数関数，対数関数である．一般には q-指数関数と q-対数関数は逆関数とはならないが，ここではパラメータに制限をつけることで，q-指数関数と q-対数関数が互いに逆関数の状況を考える．

統計モデルが q-指数関数に関して指数型分布と同じ表式を持つもの，すなわち

$$S_q := \left\{ p(x,\theta) \;\middle|\; p(x;\theta) = e^{\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)} \right\}$$

を q-指数型分布族と呼ぶ．q-指数関数は実際には冪型の関数であるので，q-指数分布は冪型の確率分布である．例えば q-正規分布

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

は q-指数型分布の典型例である．ただし Z_q は規格化定数である．通常の正規分布が Boltzmann- Gibbs-Shannon エントロピーの最大化によって特徴付けられるのに対し，q-正規分布は Tsallis エントロピーの最大化によって得られ，非加法的統計力学を中心とする分野で盛んに議論されている[5),6)]．

q-指数型分布族の幾何構造であるが，通常の指数型分布族の類推 (1) と (2) から

$$\begin{aligned} g_{ij}^q(\theta) &= \partial_i \partial_j \psi(\theta), \\ C_{ijk}^q(\theta) &= \partial_i \partial_j \partial_k \psi(\theta) \end{aligned}$$

を定義し，それぞれ q-Fisher 計量と q-3 次形式と呼ぶことにする．その結果 S_q には互いに双対的な平坦接続が定義できることがわかり，その平坦接続を $\nabla^{q(e)}$ および $\nabla^{q(m)}$ と書く．

S_q に定まる双対平坦構造を理解するために，q-エスコート分布と q-期待値を次の式で定義する．

$$\begin{aligned} P_q(x) &:= \frac{1}{Z_q(p)} p(x)^q,\ Z_q(p) \;:=\; \int_\Omega p(x)^q dx. \\ E_q[f(x)] &:= \int_\Omega f(x) P_q(x) dx \\ &= \frac{1}{Z_q(p)} \int_\Omega f(x) p(x)^q dx \end{aligned}$$

複雑系科学に現れる冪型の分布は裾確率が重いため，通常の期待値や分散が発散することも多い．もともとは q-期待値などの概念は，発散の困難を回避するために導入されたものである．

通常の指数型分布族の双対座標系 $\{\eta_i\}$ は確率変数 $F_i(x)$ の期待値によって定義されていたが，q-指数型分布族の場合には

$$\eta_i = \frac{\partial}{\partial \theta^i}\psi(\theta) = \int_\Omega F_i(x) P_q(x;\theta) dx.$$

と q-エスコート確率を用いた q-期待値で与えられる．S_q には正準ダイバージェンス (3) も定義されるが，

$$\rho_q(p(\theta'), p(\theta)) = E_{qp(\theta)}[\log_q p(\theta) - \log_q p(\theta')]$$

という関係式を満たす．指数型分布族の正準ダイバージェンスは (3) で見たように対数尤度の差の期待値で表されたが，q-指数型分布族の場合には q-対数尤度の差の q-期待値で表される．この結果から q-指数型分布族の場合には q-対数尤度という概念が有用であることが示唆されるが，次の章でさらにその意味を考える．

4　独立性概念の修正と複雑系科学における統計的推論

この章では確率変数の独立性の概念を修正し[3),5)]，q-指数型分布族の統計的推論を考える．

X, 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

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

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捉えることもでき，また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．つまり確率変数の独立性や情報量などの概念は指数型の確率分布の記述に適したものであり，

q-指数型分布族の場合にはその概念を修正することが自然であると考えられる．

まずは積に関する演算法則を修正しよう．q を正の定数として固定する．任意の正の数 x, y に対し $x^{1-q} + y^{1-q} - 1 > 0$ が成り立つとき，x と y の q-積 [2] を

$$x \otimes_q y := \left[x^{1-q} + y^{1-q} - 1\right]^{\frac{1}{1-q}}$$

によって定義する．すると

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

が成り立っていることがわかり，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．

さて，$S = \{p(x;\xi)\}$ を統計モデルとし，$\{x_1, \ldots, x_N\}$ を確率分布 $p(x;\xi) \in S$ から発生された N 個の独立な観測値とする．ここで 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．q-積が通常の積演算であれば，q-尤度関数は通常の尤度関数である．しかしながら，尤度関数はパラメータ空間上の確率密度関数であったのに対し，q-尤度関数は正値測度ではあるが確率密度関数にはならない．適切な規格化により確率密度とすることもできるが，尤度としての役割であればこのままでも十分である．

通常の正規分布は，独立な確率変数から中心極限定理によっても得られた．一方 q-正規分布は，q-積に関する独立な確率変数からの中心極限定理によって得られることも知られている [2]．

さて，尤度関数を最大にするパラメータが最尤推定量であった．そこで，q-尤度関数を最大にするパラメータを q-最尤推定量と定義する．すなわち

$$\hat{\xi} = \arg \max_{\xi \in \Xi} L_q(\xi)$$

である．q-対数関数は単調増加であるので q-対数尤度を考えても同値である．したがって q-尤度関数の定義と q-積の性質から

$$\begin{aligned} \log_q L_q(\xi) &= \sum_{i=1}^{N} \log_q p(x_i;\xi), \\ \hat{\xi} &= \arg \max_{\xi \in \Xi} \log_q L_q(\xi) \end{aligned}$$

が成り立つ．通常の対数尤度とは，独立な観測に対して，推定に関する情報が確率分布の対数程度に増加すると考えることもできる．一方で複雑系科学に現れる冪型の確率分布では，推定に関する情報は確率分布の q-対数程度で増加すると考えるのが自然であることを意味する．

最後に q-最尤法を幾何学の立場から考える．S_q を q-指数型分布族とし，M を S の部分モデル，すなわち S の曲 q-指数型分布族とし，真の分布 $p(x;u) = p(x;\theta(u)) \in M$ を含むとする．

S_q に定まる正準ダイバージェンスと q-対数尤度には

$$\rho_q(p(\theta(u), p(\hat{\eta})) \quad = \quad \phi(\hat{\eta}) - \frac{1}{N} \log_q L_q(u)$$

の関係式が成り立つ．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．さらにこれは，q-最尤推定量とは q-Fisher 計量と $\nabla^{q(m)}$-測地線に関する，データからのモデル分布への直交射影であることもわかる．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．

5 おわりに

複雑系科学における統計的推論と，その幾何学とのかかわりについて簡単に述べた．冪型分布や q-正規分布によって記述される現象は，例えば宇宙の大規模構造や株価の変動のように，個々が事象が独立に活動することができず，それぞれの結果が他の事象に何らかの相関を与えるようなものであると考えられる．このような現象では標本空間がユークリッド空間ではなく，ベクトル空間や多様体のように，ある種の数学的構造が内在していると思われる．

そもそも平均や分散などの概念自体が標本空間の座標系の取り方に依存する概念であるので，現象に応じた適切な確率分布の表現が必要であると考えられる．

参考文献

1) Shun-ichi Amari and Hiroshi Nagaoka, Methods of information geometry, Amer. Math. Soc., Providence, Oxford University Press, Oxford, 2000.
2) Ernesto P. Borgesa, *A possible deformed algebra and calculus inspired in nonextensive thermostatistics*, Phys. A, **340**(2004), 95–101.
3) 藤本 悠, 村田 昇, 独立性の一般化に基づく統計モデルの拡張, 第 12 回情報論的学習理論ワークショップ (IBIS2009) 講演要旨, 2009
4) Jan Naudts, *Generalised exponential families and associated entropy functions*, Entropy, **10**(2008), 131–149.
5) 須鎗 弘樹，複雑系のための基礎数理ーべき乗則とツァリスエントロピーの数理, 牧野書店, 2010.
6) Constantino Tsallis, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, Springer, New York, 2009.