統計多様体の幾何学と異常統計

松添 博　　名古屋工業大学

2017年9月29日

MACSセミナー

1 問題提起
2 擬加法的微分
3 統計モデルの幾何学
4 変形指数型分布族
5 変形指数型分布族の幾何学（前半）
6 変形指数型分布族の幾何学（後半）
7 独立性と最尤推定量の一般化

いろいろな共同研究を含んでいます.

和田達明氏（茨城大学），逸見昌之氏（統計数理研究所）
小原敦美氏（福井大学），甘利俊一氏（理化学研究所）

2　擬加法的微分

2.1　中心極限定理再考

Napier の数

$$e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

$$\begin{aligned}
a_1 &= \left(1+\frac{1}{1}\right) \\
a_2 &= \left(1+\frac{1}{2}\right)\left(1+\frac{1}{2}\right) \\
a_3 &= \left(1+\frac{1}{3}\right)\left(1+\frac{1}{3}\right)\left(1+\frac{1}{3}\right) \\
&\vdots \qquad\qquad \vdots \\
&\vdots \qquad\qquad \vdots
\end{aligned}$$

区間の縮小の速さとコピーの生成速度がうまく釣り合っている.

Napierの数

$$e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

区間の縮小の速さとコピーの生成速度がうまく釣り合っている.

中心極限定理

$X_1, \ldots, X_n, \cdots$: i.i.d.,

$$E[X_i] = \mu, \quad V[X_i] = \sigma^2$$

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \left(V[\bar{X}_n] = \frac{\sigma^2}{n}\right)$$

$$\Longrightarrow \qquad P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le t\right) \quad \to \quad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}dx$$

区間の縮小の速さとコピーの生成速度が不釣り合いなとき，何が起こるのか.
また，縮小の速さやコピーの生成が不均一なとき，何が起こるのか.

Napier の数

$$e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

区間の縮小の速さとコピーの生成速度がうまく釣り合っている.

中心極限定理

$$X_1,\ldots,X_n,\cdots : \text{i.i.d.},$$
$$E[X_i]=\mu,\quad V[X_i]=\sigma^2$$
$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i,\qquad \left(V[\bar{X}_n]=\frac{\sigma^2}{n}\right)$$
$$\Longrightarrow\qquad P\left(\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\le t\right)\quad\to\quad\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}dx$$

区間の縮小の速さとコピーの生成速度が不釣り合いなとき，何が起こるのか.
また，縮小の速さやコピーの生成が不均一なとき，何が起こるのか.

指数関数の特徴

$$\frac{d}{dx}e^x = e^x$$

2.2 擬加法的 q-微分

$$\frac{d}{dx}f(x) = \{f(x)\}^q : \text{エスコート描像}$$

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

q-指数関数，q-対数関数 $(q>0)$

$$\exp_q x := (1+(1-q)x)^{\frac{1}{1-q}} \qquad (1+(1-q)x>0) \qquad q\text{-指数関数}$$

$$\log_q y := \frac{y^{1-q}-1}{1-q} \qquad (y>0) \qquad q\text{-対数関数}$$

$q \to 1$ のとき，通常の指数関数と通常の対数関数に一致する．

q-和，q-積

$$x_1 \tilde{\oplus}^q x_2 \;:=\; x_1 + x_2 + (1-q)x_1x_2 \qquad q\text{-和}$$

$$y_1 \otimes_q y_2 \;:=\; \left[y_1^{1-q} + y_2^{1-q} - 1\right]^{\frac{1}{1-q}} \qquad q\text{-積}$$

ただし $\left(y_1^{1-q} + y_2^{1-q} - 1 > 0\right)$

q-指数関数，q-対数関数 $(q > 0)$

$$\exp_q x := (1 + (1-q)x)^{\frac{1}{1-q}} \qquad (1 + (1-q)x > 0) \qquad q\text{-指数関数}$$

$$\log_q y := \frac{y^{1-q} - 1}{1-q} \qquad (y > 0) \qquad q\text{-対数関数}$$

q-和，q-積

$$x_1 \tilde{\oplus}^q x_2 := x_1 + x_2 + (1-q)x_1 x_2 \qquad q\text{-和}$$

$$y_1 \otimes_q y_2 := \left[y_1^{1-q} + y_2^{1-q} - 1 \right]^{\frac{1}{1-q}} \qquad q\text{-積}$$

q-指数関数と q-対数関数の真数条件が成り立つとき

$$\exp_q(x_1 \tilde{\oplus}^q x_2) = \exp_q x_1 \cdot \exp_q x_2,$$
$$\log_q(y_1 \cdot y_2) = \log_q y_1 \tilde{\oplus}^q \log_q y_2,$$

$$\exp_q(x_1 + x_2) = \exp_q x_1 \otimes_q \exp_q x_2,$$
$$\log_q(y_1 \otimes_q y_2) = \log_q y_1 + \log_q y_2.$$

注意 2.1

q-和は定義域に作用する

q-積は値域（関数空間）に作用する

q-和，q-積

$$x_1 \tilde{\oplus}^q x_2 := x_1 + x_2 + (1-q)x_1x_2 \qquad q\text{-和}$$

$$y_1 \otimes_q y_2 := \left[y_1^{1-q} + y_2^{1-q} - 1\right]^{\frac{1}{1-q}} \qquad q\text{-積}$$

$$[-x] := \log_q\left(\frac{1}{\exp_q x}\right) = \frac{-x}{1+(1-q)x} : q\text{-和に関する逆元}$$

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

擬加法的 q-微分

$$\frac{d_q}{d_q x} f(x) := \lim_{x' \to x} \frac{f(x') - f(x)}{x' \tilde{\ominus}^q x} : \quad \text{擬加法的 } q\text{-微分}$$

$\dfrac{d_q}{d_q x} f(x) = f(x)$ とすると　（擬加法的描像）

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

$\dfrac{d}{dx} f(x) = \{f(x)\}^q$ とすると　（エスコート描像）

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

q-和，q-積

$$x_1 \tilde{\oplus}^q x_2 := x_1 + x_2 + (1-q)x_1x_2 \qquad q\text{-和}$$

$$y_1 \otimes_q y_2 := \left[y_1^{1-q} + y_2^{1-q} - 1\right]^{\frac{1}{1-q}} \qquad q\text{-積}$$

q-指数関数の無限積表現

命題 2.2 (Suyari)

任意の $n \in N$ に対して $n\left(1+\dfrac{x}{n}\right)^{1-q} - (n-1) > 0$

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

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

2.3 擬加法的 κ-微分

κ-指数関数, κ-対数関数 $(-1 < \kappa < 1)$

$$\exp_\kappa x := (\sqrt{1+\kappa^2 x^2} + \kappa x)^{\frac{1}{\kappa}}$$ κ-指数関数

$$\log_\kappa y := \frac{y^\kappa - y^{-\kappa}}{2\kappa} \qquad (y > 0)$$ κ-対数関数

$\kappa \to 0$ のとき，通常の指数関数と通常の対数関数に一致する.

κ-和，κ-積 (Kaniadakis)

$$\begin{aligned} x_1 \tilde{\oplus}^\kappa x_2 &:= \log_\kappa [\exp_\kappa x_1 \cdot \exp_\kappa x_2] \\ &= x_1\sqrt{1+\kappa^2 x_2^2} + x_2\sqrt{1+\kappa^2 x_1} \end{aligned}$$ κ-和

$$y_1 \otimes_\kappa y_2 := \exp_\kappa [\log_\kappa y_1 + \log_\kappa y_2]$$ κ-積

$$\exp_\kappa(x_1 \tilde{\oplus}^\kappa x_2) = \exp_\kappa x_1 \cdot \exp_\kappa x_2,$$
$$\log_\kappa(y_1 \cdot y_2) = \log_\kappa y_1 \tilde{\oplus}^\kappa \log_\kappa y_2,$$

$$\exp_\kappa(x_1 + x_2) = \exp_\kappa x_1 \otimes_\kappa \exp_\kappa x_2,$$
$$\log_\kappa(y_1 \otimes_\kappa y_2) = \log_\kappa y_1 + \log_\kappa y_2.$$

κ-和，κ-積 (Kaniadakis)

$$
\begin{aligned}
x_1 \tilde{\oplus}^{\kappa} x_2 &:= \log_\kappa [\exp_\kappa x_1 \cdot \exp_\kappa x_2] \\
&= x_1\sqrt{1+\kappa^2 x_2^2} + x_2\sqrt{1+\kappa^2 x_1} && \kappa\text{-和} \\
y_1 \otimes_\kappa y_2 &:= \exp_\kappa [\log_\kappa y_1 + \log_\kappa y_2] && \kappa\text{-積}
\end{aligned}
$$

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

擬加法的 κ-微分

$$\frac{d_\kappa}{d_\kappa x} f(x) := \lim_{x' \to x} \frac{f(x') - f(x)}{x' \tilde{\ominus}^{\kappa} x} : \quad \text{擬加法的 } \kappa\text{-微分}$$

$\dfrac{d_\kappa}{d_\kappa x} f(x) = f(x)$ とすると　（擬加法的描像）

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

$\dfrac{d}{dx} f(x) = \dfrac{1}{\sqrt{1+\kappa^2 x^2}} f(x)$ とすると　（エスコート描像）

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

κ-和，κ-積 (Kaniadakis)

$$
\begin{aligned}
x_1 \tilde{\oplus}^{\kappa} x_2 &:= \log_\kappa [\exp_\kappa x_1 \cdot \exp_\kappa x_2] \\
&= x_1\sqrt{1+\kappa^2 x_2^2} + x_2\sqrt{1+\kappa^2 x_1} && \kappa\text{-和} \\
y_1 \otimes_\kappa y_2 &:= \exp_\kappa [\log_\kappa y_1 + \log_\kappa y_2] && \kappa\text{-積}
\end{aligned}
$$

κ-指数関数の無限積表現

定理 1 (M and Wada) $x > -1$ のとき

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

Tsallis 統計とトロピカル幾何学

「トロピカル」はブラジルの計算機学者 Imre Simon 氏に由来する
Tsallis 氏もブラジル人なので「トロピカル統計？」でもある

$(\oplus_q, \otimes_q)$-半環

$$y_1 \oplus_q y_2 := \exp_q\left[\ln\left(\exp\ln_q y_1 + \exp\ln_q y_2\right)\right]$$ 値域上の q-和

$$y_1 \otimes_q y_2 := \exp_q\left[\ln_q y_1 + \ln_q y_2\right] = \left[y_1^{1-q} + y_2^{1-q} - 1\right]^{\frac{1}{1-q}}$$ q-積

$$y_1 \otimes_q y_2 \to \max(y_1, y_2, 1) \qquad (q \to -\infty)$$

可積分系理論における代数構造の一般化

$$f: \text{全単射} \Longrightarrow x \oplus y = f(f^{-1}(x) + f^{-1}(y)),\ x \otimes y = f(f^{-1}(x) f^{-1}(y))$$

$(+_h, \times_h)$-半環（脱量子化）

$$x +_h y := h \log\left[e^{\frac{x}{h}} + e^{\frac{y}{h}}\right]$$

$$x \times_h y := h \log\left[e^{\frac{x}{h}} \cdot e^{\frac{y}{h}}\right]$$

$$x_1 +_h y \to \max(x, y) \qquad (h \to 0)$$

超離散極限

$$x \times_h y \to x + y \qquad (h \to 0)$$

3　統計モデルの幾何学

定義 3.1
S が Ω 上の 統計モデル （またはパラメトリックモデル）
$\overset{\text{def}}{\Longleftrightarrow}$ S が $\xi \in \Xi$ をパラメータとする確率密度関数族で次を満たす：

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

S を $\{\Xi; \xi^1, \ldots, \xi^n\}$ を局所座標系とする多様体とみなす.

$g^F = (g^F_{ij})$ が S の Fisher 計量（Fisher 情報行列）

$$\overset{\text{def}}{\Longleftrightarrow} \qquad g^F_{ij}(\xi) := \int_\Omega \frac{\partial}{\partial \xi^i} \log p(x;\xi) \frac{\partial}{\partial \xi^j} \log p(x;\xi) p(x;\xi) dx$$

$$= \int_\Omega \partial_i p_\xi \left(\frac{\partial}{\partial \xi^j} \log p_\xi \right) dx = E_p[\partial_i l_\xi \partial_j l_\xi]$$

$\partial_i p_\xi$ $\overset{\text{def}}{\Longleftrightarrow}$ 混合型表現，m-表現

$\partial_i l_\xi = \left(\dfrac{\partial_i p_\xi}{p_\xi} \right)$ $\overset{\text{def}}{\Longleftrightarrow}$ 指数型表現，e-表現. （スコア関数）

$\nabla^{(\alpha)}$: α-接続 （$\alpha \in R$: 固定）

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

$$g^F(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = \Gamma^{(\alpha)}_{ij,k}$$

$\nabla^{(0)}$ は Fisher 計量 g^F の Levi-Civita 接続

$\nabla^{(e)} := \nabla^{(1)}$: 指数型接続，e-接続

$\nabla^{(m)} := \nabla^{(-1)}$: 混合型接続，m-接続

(1) $\partial_i g(\partial_j, \partial_k) = g(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) + g(\partial_j, \nabla^{(-\alpha)}_{\partial_i}\partial_k)$

($\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は g^F に関して（互いに）双対的)

(2) $g(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = g(\nabla^{(0)}_{\partial_i}\partial_j, \partial_k) - \frac{\alpha}{2}C(\partial_i, \partial_j, \partial_k)$

$C_\xi(\partial_i, \partial_j, \partial_k) := E_p[(\partial_i l_\xi)(\partial_j l_\xi)(\partial_k l_\xi)]$: 3 次形式

(3) $(\nabla^{(\alpha)}_{\partial_i} g)(\partial_j, \partial_k) = (\nabla^{(\alpha)}_{\partial_j} g)(\partial_i, \partial_k) = \alpha C(\partial_i, \partial_j, \partial_k)$

$(S, \nabla^{(\alpha)}, g^F)$ を不変統計多様体 とよぶ.

統計モデル S_e が 指数型分布族

$$\overset{\text{def}}{\Longleftrightarrow} \qquad S_e = \left\{ p(x;\theta) \;\middle|\; p(x;\theta) = \exp[C(x) + \sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)] \right\},$$

$C, F_1, \cdots, F_n$: Ω 上の確率変数

ψ : パラメータ空間 Θ 上の関数

$[\theta^i]$ を 自然座標系とよぶ.

命題 3.2　指数型分布族に対し次が成り立つ

(1) $\nabla^{(1)}$ は平坦

(2) $[\theta^i]$ は $\nabla^{(1)}$ に関するアファイン座標系，すなわち $\Gamma_{ij}^{(1)\,k} \equiv 0$

簡単のため $C = 0$ を仮定する．★

$$\begin{aligned}
g_{ij}^F(\theta) &= E_p[(\partial_i \log p(x;\theta))(\partial_j \log p(x;\theta))] \\
&= E_p[-\partial_i\partial_j \log p(x;\theta)] \;=\; E_p[\partial_i\partial_j\psi(\theta)] \\
&= \partial_i\partial_j\psi(\theta) : \quad \text{Fisher 計量} \\
C_{ijk}^F(\theta) &= E_p[(\partial_i \log p(x;\theta))(\partial_j \log p(x;\theta))(\partial_k \log p(x;\theta))] \\
&= \partial_i\partial_j\partial_k\psi(\theta) : \quad 3 \text{ 次形式}
\end{aligned}$$

$(S_e, \nabla^{(e)}, g^F)$ と $(S_e, \nabla^{(m)}, g^F)$ は Hesse 多様体である.

命題 2　指数型分布族 S_e に対して以下が成り立つ.

(1) $(S_e, g^F, \nabla^{(e)}, \nabla^{(m)})$: 双対平坦空間

(2) $\{\theta^i\}$: S_e の $\nabla^{(e)}$-アファイン座標系

(3) $\psi(\theta)$: $\{\theta^i\}$ に関する g^F のポテンシャル

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

(4) 確率変数 $F_i(x)$ の期待値を $\eta_i = E_p[F_i(x)]$ とおく
$\Longrightarrow$ $\{\eta_i\}$ は $\{\theta^i\}$ の g^F に関する双対座標系

(5) $\phi(\eta) = E_p[\log p(x;\theta)]$ とおく
$\Longrightarrow$ $\phi(\eta)$ 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，Legendre 変換も成り立つ.

$$\frac{\partial \psi}{\partial \theta^i} = \eta_i, \quad \frac{\partial \phi}{\partial \eta_i} = \theta^i, \quad \psi(p) + \phi(p) - \sum_{i=1}^{m} \theta^i(p)\eta_i(p) = 0$$

$$g^F_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}, \quad C^F_{ijk} = \frac{\partial^3 \psi}{\partial \theta^i \partial \theta^j \partial \theta^k}$$

S の Kullback-Leibler ダイバージェンス（または相対エントロピー）

$$\overset{\text{def}}{\Longleftrightarrow} \quad D_{KL}(p,r) = \int_\Omega p(x)\log\frac{p(x)}{r(x)}dx$$
$$= E_p[\log p(x) - \log r(x)]$$
$$\left(= \psi(r) + \phi(p) - \sum_{i=1}^{n}\theta^i(r)\eta_i(p) = D(r,p) \right)$$

指数型分布族 S_e の場合，D_{KL} は平坦統計多様体 $(S_e, \nabla^{(m)}, g^F)$ のカノニカル・ダイバージェンスと一致する．

推定関数からのダイバージェンスの構成

$$s(x;\xi) = \begin{pmatrix} \partial/\partial\xi^1 \log p(x;\xi) \\ \vdots \\ \partial/\partial\xi^n \log p(x;\xi) \end{pmatrix}$$: $p(x;\xi)$ のスコア関数 （推定関数）

スコア関数をパラメータに関して積分し，期待値を考える．

$$d_{KL}(p,r) := \int_\Omega p(x;\xi)\log r(x;\xi')dx$$ S のクロス・エントロピー

クロス・エントロピーを用いて KL-ダイバージェンスは次で与えられる．

$$D_{KL}(p,r) = d_{KL}(p,p) - d_{KL}(p,r)$$

4　変形指数型分布族 (χ-指数型分布族)

$\chi : (0, \infty) \to (0, \infty)$ ： 狭義単調増加

定義 4.1 (χ-指数関数, χ-対数関数)

$$\log_\chi x := \int_1^x \frac{1}{\chi(t)} dt \qquad \chi\text{-対数関数}$$

$$\exp_\chi x := 1 + \int_0^x \lambda(t) dt \qquad \chi\text{-指数関数}$$

ただし　$\lambda(\log_\chi t) = \chi(t)$

通常，対数関数の変形には ϕ を用い ϕ-指数関数と呼ばれる．
情報幾何学ではϕは双対平坦空間における双対ポテンシャルとして用いる．

例 3 $\chi(t) = t^q$ のとき

$$\int_1^x \frac{1}{\chi(t)} dt = \int_1^x \frac{1}{t^q} dt = \frac{x^{1-q} - 1}{1-q} = \log_q x \qquad q\text{-対数関数}$$

$$\lambda(t) = (1 + (1-q)t)^{\frac{q}{1-q}}$$

$$1 + \int_0^x \lambda(t)\, dt = (1 + (1-q)x)^{\frac{1}{1-q}} \qquad q\text{-指数関数}$$

例 4 $\chi(t) = \dfrac{2x}{x^{\kappa} + x^{-\kappa}}$ のとき

$$\int_1^x \frac{1}{\chi(t)}\,dt = \frac{x^{\kappa} - x^{-\kappa}}{2\kappa} = \log_{\kappa} x \qquad \kappa\text{-対数関数}$$

$$\lambda(t) = \frac{\left(\kappa x + \sqrt{1+\kappa^2 x^2}\right)^{\frac{1}{\kappa}}}{\sqrt{1+\kappa^2 x^2}}$$

$$1 + \int_0^x \lambda(t)\,dt = \left(\kappa x + \sqrt{1+\kappa^2 x^2}\right)^{\frac{1}{\kappa}} = \exp_{\kappa} x \qquad \kappa\text{-指数関数}$$

$\chi : (0, \infty) \to (0, \infty)$ ： 狭義単調増加

定義 4.1 — χ-指数関数, χ-対数関数

$$\log_\chi x := \int_1^x \frac{1}{\chi(t)} dt \qquad \chi\text{-対数関数}$$

$$\exp_\chi x := 1 + \int_0^x \lambda(t) dt \qquad \chi\text{-指数関数}$$

$$where \quad \lambda(\log_\chi t) = \chi(t)$$

$F_1(x), \ldots, F_n(x)$ ： Ω 上の関数

$\theta = \{\theta^1, \ldots, \theta^n\}$ ： パラメータ

$$S = \left\{ p(x, \theta) \;\middle|\; p(x; \theta) > 0, \int_\Omega p(x; \theta) dx = 1 \right\}$$ ： 統計モデル

定義 4.2

$S_\chi = \{p(x; \theta)\}$ ： χ-指数型分布族, 変形指数型分布族

$$\overset{\text{def}}{\Longleftrightarrow} S_\chi := \left\{ p(x, \theta) \middle| p(x; \theta) = \exp_\chi \left[\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right], p(x, \theta) \in S \right\}$$

定理 4.3 (有限集合上の確率分布族)
有限集合 S_n 上の確率分布の全体は，任意の χ に対して χ-指数型分布族である．

(証明) $\Omega = \{x_0, x_1, \ldots, x_n\}$

$$S_n = \left\{ p(x;\eta) \;\middle|\; \eta_i > 0,\ \sum_{i=0}^{n} \eta_i = 1,\ p(x;\eta) = \sum_{i=0}^{n} \eta_i \delta_i(x) \right\},$$

$$\eta_0 = 1 - \sum_{i=1}^{n} \eta_i$$

$\theta^i = \log_\chi p(x_i) - \log_\chi p(x_0) \;=\; \log_\chi \eta_i - \log_\chi \eta_0$ とおく．
このとき

$$\begin{aligned}
\log_\chi p(x) &= \log_\chi \left(\sum_{i=0}^{n} \eta_i \delta_i(x) \right) \\
&= \sum_{i=1}^{n} \left(\log_\chi \eta_i - \log_\chi \eta_0 \right) \delta_i(x) + \log_\chi(\eta_0)
\end{aligned}$$

$$\psi(\theta) = -\log_\chi \eta_0$$

例 4.4 (Student t-分布族 (q-正規分布族))

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

$$\theta^1 = \frac{2}{3-q}z_q^{q-1}\cdot\frac{\mu}{\sigma^2}, \qquad \theta^2 = -\frac{1}{3-q}z_q^{q-1}\cdot\frac{1}{\sigma^2}.$$

とおく．このとき

$$\begin{aligned}
\log_q p_q(x) &= \frac{1}{1-q}(p^{1-q}-1) = \frac{1}{1-q}\left\{\frac{1}{z_q^{1-q}}\left(1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right) - 1\right\} \\
&= \frac{2\mu z_q^{q-1}}{(3-q)\sigma^2}x - \frac{z_q^{q-1}}{(3-q)\sigma^2}x^2 - \frac{z_q^{q-1}}{3-q}\cdot\frac{\mu^2}{\sigma^2} + \frac{z_q^{q-1}-1}{1-q} \\
&= \theta^1 x + \theta^2 x^2 - \psi(\theta)
\end{aligned}$$

$$\psi(\theta) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{z_q^{q-1}-1}{1-q}$$

注意 4.5 (S_q, g^F) は負曲率 $\dfrac{q}{q-3}$ の定曲率空間.

5　変形指数型分布族の幾何学（前半）

S_χ : 変形指数型分布族
$\psi(\theta)$: Θ 上の狭義凸関数 (S_χ の規格化項)

$$s^\chi(x;\theta) = \left((s^\chi)^1(x;\theta),\ \ldots,\ (s^\chi)^n(x;\theta)\right)^T \quad \chi\text{-スコア関数}$$
$$\overset{\text{def}}{\Longleftrightarrow} \qquad (s^\chi)^i(x;\theta) = \frac{\partial}{\partial\theta^i}\log_\chi p(x;\theta), \quad (i=1,\ldots,n). \tag{1}$$

S_χ の統計多様体構造

Riemann 計量 g^M:

$$g^M_{ij}(\theta) = \int_\Omega \partial_i p(x;\theta)\partial_j \log_\chi p(x;\theta)\, dx$$

双対接続 $\nabla^{M(e)}, \nabla^{M(m)}$:

$$\Gamma^{M(e)}_{ij,k}(\theta) = \int_\Omega \partial_k p(x;\theta)\partial_i\partial_j \log_\chi p(x;\theta) dx$$
$$\Gamma^{M(m)}_{ij,k}(\theta) = \int_\Omega \partial_i\partial_j p(x;\theta)\partial_k \log_\chi p(x;\theta) dx$$

$(S_\chi, \nabla^{M(e)}, g^M)$ と $(S_\chi, \nabla^{M(m)}, g^M)$ は Hesse 多様体.

$$I_\chi(p_\theta) = -\int_\Omega \{V_\chi(p(x;\theta)) + (p(x;\theta)-1)V_\chi(0)\}\,dx,$$

一般化したエントロピー汎関数

$$\Psi(\theta) = \int_\Omega p(x;\theta)\log_\chi p(x;\theta)dx + I_\chi(p_\theta) + \psi(\theta),$$

一般化した Massieu ポテンシャル

ただし $\quad V_\chi(t) = \int_1^t \log_\chi(s)\,ds, \quad V_\chi(0) = \lim_{t\to+0} V_\chi(t) < \infty.$

命題 5 χ-指数型分布族 S_χ に対して以下が成り立つ.

(1) $(S_\chi, g^M, \nabla^{M(e)}, \nabla^{M(m)})$ は双対平坦空間

(2) $\{\theta^i\}$ は S_χ の $\nabla^{M(e)}$-アファイン座標系

(3) $\Psi(\theta)$: $\{\theta^i\}$ に関する g^M のポテンシャル

$$g^M_{ij}(\theta) = \partial_i \partial_j \Psi(\theta).$$

(4) 関数 $F_i(x)$ の期待を $\eta_i = E_p[F_i(x)]$ とおく
$\Longrightarrow$ $\{\eta_i\}$ は $\{\theta^i\}$ の g^M に関する双対座標系

(5) $\Phi(\eta) = -I_\chi(p_\theta)$ とおく
$\Longrightarrow$ $\Phi(\eta)$ は $\{\eta_i\}$ に関する g^M のポテンシャルである.

$$I_\chi(p_\theta) = -\int_\Omega \{V_\chi(p(x;\theta)) + (p(x;\theta) - 1)V_\chi(0)\}\, dx,$$

一般化した entropy 汎関数

$$\Psi(\theta) = \int_\Omega p(x;\theta) \log_\chi p(x;\theta) dx + I_\chi(p_\theta) + \psi(\theta),$$

一般化した Massieu ポテンシャル

χ-ダイバージェンスの構成

$u_p^\chi(x;\theta)$: バイアス補正 χ-スコア関数

$$\overset{\text{def}}{\Longleftrightarrow} \quad u_p^\chi(x;\theta) = ((u_p^\chi)^1(x;\theta), \ldots, (u_p^\chi)^n(x;\theta))^T$$

$$(u_p^\chi)^i(x;\theta) = \frac{\partial}{\partial\theta^i}\log_\chi p(x;\theta) - E_p\left[\frac{\partial}{\partial\theta^i}\log_\chi p(x;\theta)\right].$$

バイアス補正したχ-スコア関数をパラメータに関して積分し，期待値を考える．

$d_\chi^M(p(x;\theta), p(x;\theta'))$ S_χ の Bregman タイプの χ-クロス・エントロピー

$$= -\int_\Omega p(x;\theta)\log_\chi p(x;\theta')dx + \int_\Omega U_\chi(\log_\chi p(x;\theta'))dx.$$

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

$D_\chi(p_\theta, p_\theta) = 0$ と調整して，χ-ダイバージェンス（または U-ダイバージェンス）

$$\begin{aligned} D_\chi(p_\theta, p_{\theta'}) &= -d_\chi^M(p_\theta, p_\theta) + d_\chi^M(p_\theta, p_{\theta'}) \\ &= \int_\Omega \{U_\chi(\log_\chi p(\theta')) - U_\chi(\log_\chi p(\theta)) \\ &\qquad -p(\theta)(\log_\chi p(\theta') - \log_\chi p(\theta))\}\,dx. \end{aligned}$$

6　変形指数型分布族の幾何学（後半）

定義 6.1

$P_\chi(x;\theta)$: $p(x;\theta)$ の エスコート分布

$$\overset{\text{def}}{\Longleftrightarrow} \quad P_\chi(x;\theta) = \frac{1}{Z_\chi(\theta)}\chi(p(x;\theta)), \quad Z_\chi(\theta) = \int_\Omega \chi(p(x;\theta))dx$$

$E_{\chi,p}[f(x)]$: $f(x)$ に対する χ-期待値

$\overset{\text{def}}{\Longleftrightarrow}$ $f(x)$ のエスコート分布 $P(x;\theta)$ に関する期待値:

$$E_{\chi,p}[f(x)] = \int f(x)P_\chi(x;\theta)dx = \frac{1}{Z_\chi(\theta)}\int f(x)\chi(p(x;\theta))dx$$

χ-指数関数, χ-対数関数

$$\log_\chi x := \int_1^x \frac{1}{\chi(t)}dt$$ χ-対数関数

$$\exp_\chi x := 1 + \int_0^x \lambda(t)dt$$ χ-指数関数

ただし　$\lambda(\log_\chi t) = \chi(t)$

定義 6.2 $S_\chi = \{p(x;\theta)\}$: χ-指数型分布族

$$g^\chi_{ij}(\theta) = \partial_i\partial_j\psi(\theta) \quad : \chi\text{-Fisher 計量}$$
$$C^\chi_{ijk}(\theta) = \partial_i\partial_j\partial_k\psi(\theta) : \chi\text{-3 次形式}$$

$\Gamma^{\chi(0)}_{ij,k}$: χ-Fisher 計量 g^χ に関する Levi-Civita 接続

$$\Gamma^{\chi(e)}_{ij,k} := \Gamma^{\chi(0)}_{ij,k} - \frac{1}{2}C^\chi_{ijk}, \qquad \Gamma^{\chi(m)}_{ij,k} := \Gamma^{\chi(0)}_{ij,k} + \frac{1}{2}C^\chi_{ijk},$$

$\nabla^{\chi(e)}$: χ-指数型接続
$\nabla^{\chi(m)}$: χ-混合型接続

補題 6.3
$S_\chi = \{p(x;\theta)\}$: χ-指数型分布族
$\Longrightarrow$ $(S_\chi, \nabla^{\chi(e)}, g^\chi)$ と $(S_\chi, \nabla^{\chi(m)}, g^\chi)$ は Hesse 多様体.
$\Longrightarrow$ $(S_\chi, g^\chi, \nabla^{\chi(e)}, \nabla^{\chi(m)})$ は双対平坦空間

命題 6　χ-指数型分布族 S_χ に対して以下が成り立つ.

(1) $(S_\chi, g^\chi, \nabla^{\chi(e)}, \nabla^{\chi(m)})$ は双対平坦空間

(2) $\{\theta^i\}$ は S_χ の $\nabla^{\chi(e)}$-アファイン座標系

(3) $\psi : \{\theta^i\}$ に関する g^χ のポテンシャル

$$g^\chi_{ij}(\theta) = \partial_i \partial_j \psi(\theta).$$

(4) 関数 $F_i(x)$ の χ-期待値を $\eta_i = E_{\chi,p}[F_i(x)]$ とおく
$\Longrightarrow$ $\{\eta_i\}$ は $\{\theta^i\}$ の g^χ に関する双対座標系

(5) $\phi(\eta) = E_{\chi,p}[\log_\chi p(x;\theta)]$ とおく
$\Longrightarrow$ $\phi(\eta)$ は $\{\eta_i\}$ に関する g^χ のポテンシャルである.

$E_{\chi,p}[f(x)]$: $f(x)$ に対する χ-期待値
$\overset{\text{def}}{\Longleftrightarrow}$ $f(x)$ のエスコート分布 $P(x;\theta)$ に関する期待値:

$$E_{\chi,p}[f(x)] = \int f(x) P_\chi(x;\theta) dx = \frac{1}{Z_\chi(\theta)} \int f(x) \chi(p(x;\theta)) dx$$

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）

$$D^{\chi}(p,r) \;=\; E_{\chi,p}[\log_{\chi} p(x) - \log_{\chi} r(x)].$$

χ-相対エントロピーの構成

$s^{\chi}(x;\theta)$: χ-スコア関数

$$\overset{\text{def}}{\Longleftrightarrow} \qquad (s^{\chi})^{i}(x;\theta) = \frac{\partial}{\partial\theta^{i}} \log_{\chi} p(x;\theta), \quad (i = 1,\ldots,n).$$

χ-スコア関数は χ-期待値に関して不偏，すなわち $E_{\chi,\theta}[(s^{\chi})^{i}(x;\theta)] = 0$.
$\Longrightarrow$ $s^{\chi}(x;\theta)$ を推定関数の一般化とみなす.

χ-スコア関数をパラメータに関して積分し，χ-期待値を考える.

$$d^{\chi}(p,r) = -\int_{\Omega} P(x) \log_{\chi} r(x) dx \quad S_{\chi} \text{ の } \chi\text{-クロス・エントロピー}$$

$D^{\chi}(p,p) = 0$ と調整して χ-相対エントロピーを得る

$$\begin{aligned} D^{\chi}(p,r) \;&=\; -d^{\chi}(p,p) + d^{\chi}(p,r) \\ &=\; E_{\chi,p}[\log_{\chi} p(x) - \log_{\chi} r(x)]. \end{aligned}$$

α-ダイバージェンス ($\alpha = 1 - 2q$)

$$D^{(1-2q)}\left(p(x;\theta), p(x;\theta')\right) = \frac{1 - \int p(x;\theta)^q p(x;\theta')^{1-q} dx}{q(1-q)}$$

$D^{(1-2q)}$ ($\alpha = 1 - 2q$) は不変統計多様体 $(S, \nabla^{(1-2q)}, g^F)$ を誘導する

Csiszar の f-ダイバージェンスと α-ダイバージェンスなどの関連

$$D^C(p(\theta), p(\theta')) = \int p(\theta') f\left(\frac{p(\theta)}{p(\theta')}\right) dx$$

$$\text{KL-divergence} \ : \ f(x) = x \log x$$

$$\alpha\text{-divergence} \ : \ f(x) = \frac{x^q - x}{q - 1} = x^q \log_q x$$

正規化した Tsallis 相対エントロピー（χ-相対エントロピー）

$$D_q^T\left(p(x;\theta), p(x;\theta')\right) = E_{q,p}\left[\log_q p(x;\theta) - \log_q p(x;\theta')\right]$$

$$= \frac{1 - \int p(\theta)^q p(\theta')^{1-q} dx}{(1-q) Z_q(\theta)}$$

正規化した Tsallis 相対エントロピーはHesse多様体 $(S_q, \nabla^{q(m)}, g^q)$ を誘導する

α-ダイバージェンスと正規化した Tsallis 相対エントロピーには次が成り立つ

$$D^{(1-2q)}\left(p(\theta), p(\theta')\right) = \frac{Z_q(\theta)}{q} D_q^T\left(p(x;\theta), p(x;\theta')\right).$$

q-指数型分布族 S_q の α-ダイバージェンスは共形的ダイバージェンスとよばれる

定理 6.4 (M. and Ohara (2011)) $S_q = \{p(x;\theta)\}$: q-指数型分布族

$\Longrightarrow$ (1) 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.

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

（正値測度を確率測度へ規格化することで，幾何構造が変化する.）

定義 6.5

統計多様体 (M, ∇, h) と $(M, \bar{\nabla}, \bar{h})$ が α-共形同値

$\overset{\text{def}}{\Longleftrightarrow}$ 関数 φ が存在して次が成り立つ.

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

射影変形 ((−1)-共形変形)

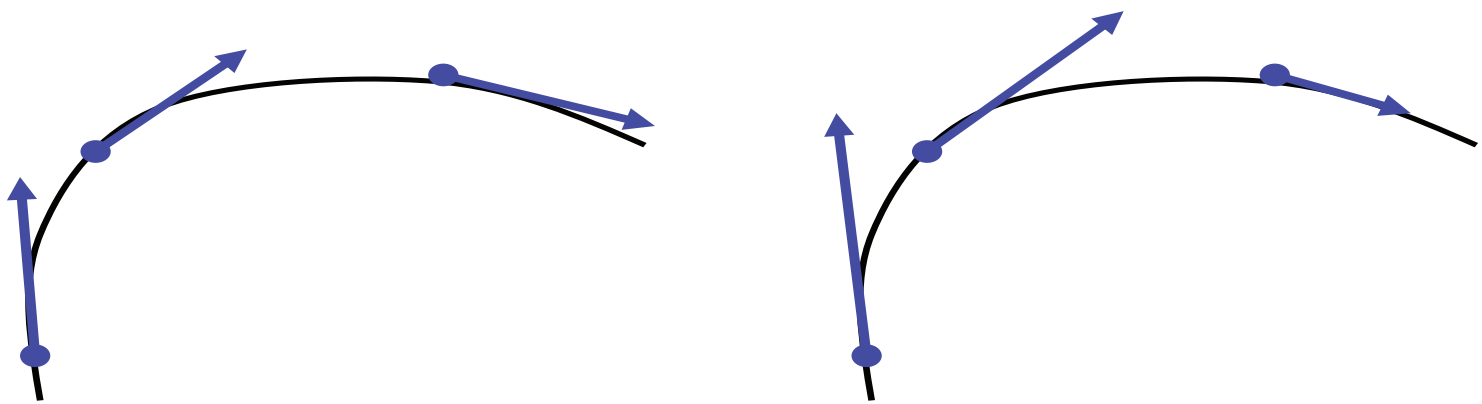
$c : I = (-\varepsilon, \varepsilon) \to M :$ 　M 上の曲線

$$c: \text{測地線} \iff \nabla_{\frac{d}{dt}} \dot{c} = 0$$

$$c: \text{プレ測地線} \iff \nabla_{\frac{d}{dt}} \dot{c} = \gamma(t)\dot{c}$$

射影変形はプレ測地線（前測地線，1次元自己平行部分多様体）を保つ

$$\nabla_{\frac{d}{dt}} \dot{c} = \beta(t)\dot{c} \iff \bar{\nabla}_{\frac{d}{dt}} \dot{c} = \bar{\gamma}(t)\dot{c}$$



双対射影変形（1-共形変形）

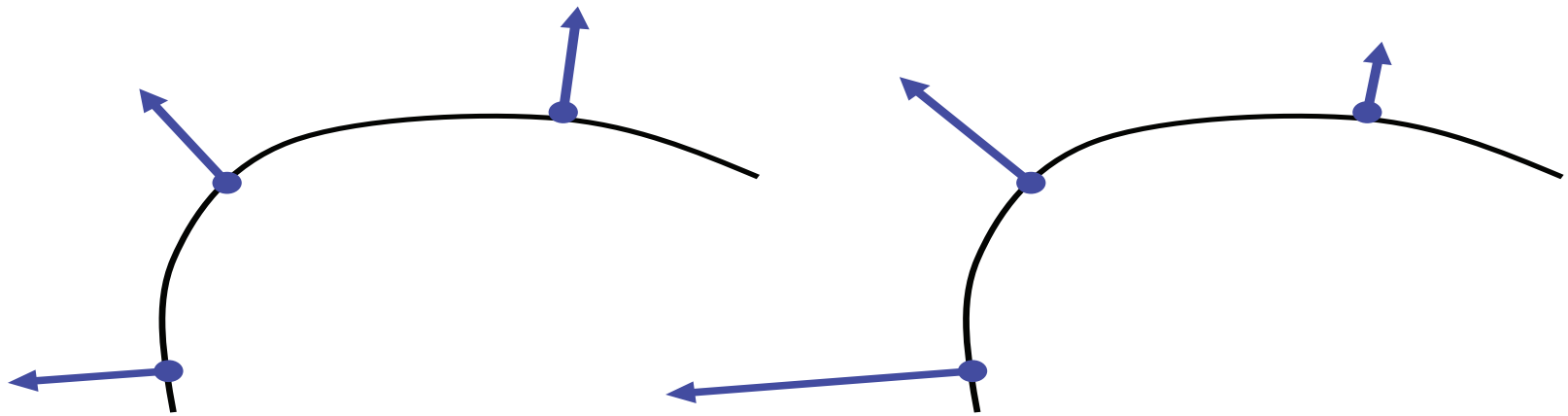
$c : I = (-\varepsilon, \varepsilon) \to M, \qquad \beta(X) = h(X, \dot{c}) :$ 接 1-次微分形式

$c :$ 双対測地線 $\iff \nabla_{\frac{d}{dt}} \beta = 0$

$c :$ プレ双対測地線 $\iff \nabla_{\frac{d}{dt}} \beta = \gamma(t)\beta$

双対射影変形はプレ双対測地線（前双対測地線）を保つ

$$\nabla_{\frac{d}{dt}} \beta = \gamma(t)\beta \iff \bar{\nabla}_{\frac{d}{dt}} \beta = \bar{\gamma}(t)\beta$$



7　独立性と最尤推定量の一般化

7.1　独立性の一般化

$X \sim p_1(x),\ Y \sim p_2(y)$
X と Y が 独立

$$\begin{aligned}
&\overset{\text{def}}{\Longleftrightarrow} && p(x,y) = p_1(x)p_2(y). \\
&\Longleftrightarrow && p(x,y) = \exp\left[\log p_1(x) + \log p_2(x)\right] \quad (p_1(x) > 0,\ p_2(y) > 0)
\end{aligned}$$

$x > 0, y > 0$ かつ $x^{1-q} + y^{1-q} - 1 > 0$ $(q > 0)$ とする.
$x \otimes_q y$: が X と Y の q-積

$$\begin{aligned}
\overset{\text{def}}{\Longleftrightarrow} \quad x \otimes_q y &:= \left[x^{1-q} + y^{1-q} - 1\right]^{\frac{1}{1-q}} \\
&= \exp_q\left[\log_q x + \log_q y\right]
\end{aligned}$$

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

7.1 独立性の一般化

$X \sim p_1(x),\ Y \sim p_2(y)$
X と Y が 独立

$$\overset{\text{def}}{\Longleftrightarrow} \quad p(x,y) = p_1(x)p_2(y).$$
$$\Longleftrightarrow \quad p(x,y) = \exp\left[\log p_1(x) + \log p_2(x)\right] \quad (p_1(x) > 0,\ p_2(y) > 0)$$

$x > 0, y > 0$ かつ $x^{1-q} + y^{1-q} - 1 > 0$ $(q > 0)$ とする.
$x \otimes_q y$: が x と y の q-積

$$\overset{\text{def}}{\Longleftrightarrow} \quad x \otimes_q y := \left[x^{1-q} + y^{1-q} - 1\right]^{\frac{1}{1-q}} = \exp_q\left[\log_q x + \log_q y\right]$$

X_1 と X_2 : m-規格化のもとで q-独立

$$\overset{\text{def}}{\Longleftrightarrow} \quad p_q(x_1, x_2) = \frac{p_1(x_1) \otimes_q p_2(x_2)}{Z_{p_1,p_2}}$$

ただし $\quad Z_{p_1,p_2} = \iint_{Supp\{p_q(x_1,x_2)\} \subset \mathcal{X}_1\mathcal{X}_2} p_1(x_1) \otimes_q p_2(x_2) dx_1 dx_2$

7.1 独立性の一般化

$X \sim p_1(x),\ Y \sim p_2(y)$
X と Y が 独立

$$\overset{\text{def}}{\Longleftrightarrow} \quad p(x,y) = p_1(x)p_2(y).$$
$$\Longleftrightarrow \quad p(x,y) = \exp\left[\log p_1(x) + \log p_2(x)\right] \quad (p_1(x) > 0,\ p_2(y) > 0)$$

$x > 0, y > 0$ かつ $x^{1-q} + y^{1-q} - 1 > 0$ $(q > 0)$ とする.
$x \otimes_q y$: が x と y の q-積

$$\overset{\text{def}}{\Longleftrightarrow} \quad x \otimes_q y := \left[x^{1-q} + y^{1-q} - 1\right]^{\frac{1}{1-q}} = \exp_q\left[\log_q x + \log_q y\right]$$

X_1 と X_2 : e-規格化のもとで q-独立

$$\overset{\text{def}}{\Longleftrightarrow} \quad p_q(x_1,x_2) = p_1(x_1) \otimes_q p_2(x_2) \otimes_q \exp_q(-c)$$

ただし c は次で定義される定数 $\displaystyle\iint_{Supp\{p_q(x_1,x_2)\}\subset \mathcal{X}_1\mathcal{X}_2} p_q(x_1,x_2)dx_1dx_2 = 1$

$p_q(x_1,x_2) = p_1(x_1) \otimes_q p_2(x_2)$ と同時分布のみを考え規格化をしない場合,
単に q-独立 とよぶ.

7.2 q-最尤推定量の幾何学

$S = \{p(x;\xi) | \xi \in \Xi\}$: 統計モデル
$\{x_1, \ldots, x_N\}$: $p(x;\xi) \in S$ から得られる N-個の観測値

$L_q(\xi)$: q-尤度関数

$$\overset{\text{def}}{\Longleftrightarrow} \qquad L_q(\xi) = p(x_1;\xi) \otimes_q p(x_2;\xi) \otimes_q \cdots \otimes_q p(x_N;\xi)$$

$$\left(\Longleftrightarrow \quad \log_q L_q(\xi) = \sum_{i=1}^{N} \log_q p(x_i;\xi) \right)$$

$q \to 1$ のとき L_q は（通常の）尤度関数

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

測定ごとに，他の測定は影響を受ける．

7.3 q-最尤推定量の幾何学

$S = \{p(x;\xi) | \xi \in \Xi\}$: 統計モデル
$\{x_1, \ldots, x_N\}$: $p(x;\xi) \in S$ から得られる N-個の観測値

$L_q(\xi)$: q-尤度関数

$$\overset{\text{def}}{\Longleftrightarrow} \qquad L_q(\xi) = p(x_1;\xi) \otimes_q p(x_2;\xi) \otimes_q \cdots \otimes_q p(x_N;\xi)$$

$$\left(\Longleftrightarrow \quad \log_q L_q(\xi) = \sum_{i=1}^{N} \log_q p(x_i;\xi) \right)$$

$q \to 1$ のとき L_q は（通常の）尤度関数

$\hat{\xi}$: q-最尤推定量

$$\overset{\text{def}}{\Longleftrightarrow} \quad \hat{\xi} = \arg\max_{\xi \in \Xi} L_q(\xi) \qquad \left(= \arg\max_{\xi \in \Xi} \log_q L_q(\xi) \right).$$

定理 7.1

q-尤度最大 $\Longleftrightarrow$ 正規化した Tsallis 相対エントロピー最小

S_q : q-指数型分布族
M : S_q の q-曲指数型分布族（S_q の部分多様体）
$\{x_1, \ldots, x_N\}$: $p(x; u) = p(x; \theta(u)) \in M$ からの N 個の観測値

q-尤度関数は次で計算される：

$$\begin{aligned}\log_q L_q(u) &= \sum_{j=1}^{N} \log_q p(x_j; u) = \sum_{j=1}^{N} \left\{ \sum_{i=1}^{n} \theta^i(u) F_i(x_j) - \psi(\theta(u)) \right\} \\ &= \sum_{i=1}^{n} \theta^i(u) \sum_{j=1}^{N} F_i(x_j) - N\psi(\theta(u)).\end{aligned}$$

q-対数 q-尤度方程式は

$$\partial_i \log_q L_q(u) = \sum_{j=1}^{N} F_i(x_j) - N\partial_i \psi(\theta(u)) = 0.$$

したがって，S_q の q-最尤推定量は次で与えられる.

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

一方 正規化した Tsallis 相対エントロピー（カノニカル・ダイバージェンス）は

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

となる.

q-尤度最大 $\Longleftrightarrow$ 正規化した Tsallis 相対エントロピー最小

（補足1） q-尤度関数は任意の凸関数でさらに一般化される.

（補足2） 独立性の一般化は機械学習理論でも提案されている.
（というよりも，こちらが最初？）

推定量不変性

系 7.2 $1 < q < 3$ とし，q-正規分布族（t-分布族）を考える．

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

$\{x_1, \ldots, x_N\}$: $p(x; \mu, \sigma) \in S_q$ からの N-個の観測値

$\Longrightarrow$ q-最尤推定量は期待値座標系に関して次で与えられる．

$$\hat{\eta}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \hat{\eta}_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$N(\mu, \sigma^2) \quad \underset{\text{MLE}}{\Longrightarrow} \quad \hat{\eta}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \hat{\eta}_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

$$N_q(\mu, \sigma^2) \quad \underset{q\text{-MLE}}{\Longrightarrow} \quad \hat{\eta}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \hat{\eta}_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$$

推定量不変性

- 正規分布族の最尤推定量 $\Longrightarrow \hat{\eta}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \hat{\eta}_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$

- q-正規分布族 (Student's t-分布族)
  - 冪型変形エントロピー (Tsallis エントロピー) 最大化
  - 混合正規分布による表示

$$p_q(x;\mu,\sigma) = \int_0^\infty N\left(\mu, \frac{1}{t}\right) \mathrm{Gamma}\left(t; \frac{3-q}{2(q-1)}, \frac{q-1}{3-q}\cdot\frac{2}{\sigma^2}\right) dt$$

q-正規分布族の Bayes 表現

- エスコート分布，擬加法的代数
  微分幾何学の視点からは自然な数学構造

- q-正規分布族の q-最尤推定量 $\Longrightarrow \hat{\eta}_1 = \frac{1}{N}\sum_{i=1}^{N} x_i, \ \hat{\eta}_2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$

パラメータ空間の重みと標本空間の重みが
上手く釣り合っている．

まとめ

β-ダイバージェンス $(S_q, g^M, \nabla^{M(e)}, \nabla^{M(m)})$

推定関数 $u_q(x;\theta)$: $u_q^i(x;\theta) = \dfrac{\partial}{\partial\theta^i}\log_q p(x;\theta) - E_\theta\left[\dfrac{\partial}{\partial\theta^i}\log_q p(x;\theta)\right]$

Riemann 計量 g^M: $g_{ij}^M(\theta) = \displaystyle\int_\Omega \partial_i p(x;\theta)\partial_j \log_q p(x;\theta)dx$

双対座標系 $\{\eta_i\}$: $\eta_i = E_p[F_i(x)]$

Tsallis 相対エントロピー $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$

推定関数 $(s^q)(x;\theta)$: $(s^q)^i(x;\theta) = \dfrac{\partial}{\partial\theta^i}\log_q p(x;\theta)$ （q-期待値に関し不偏）

Riemann 計量 g^q: $g_{ij}^q(\theta) = \dfrac{\partial^2}{\partial\theta^i\theta^j}\psi(\theta)$

双対座標系 $\{\eta_i\}$: $\eta_i = E_{q,p}[F_i(x)]$

期待値，独立性は統計モデルに依存して決定される概念である．