統計モデルの幾何学と期待値系列

松添 博　　（名古屋工業大学）

2017年**10**月**15**日**-16**日

統計多様体の幾何学とその周辺

0　導入

1　統計モデルの幾何学

2　統計多様体と双対平坦空間

3　最尤推定量の幾何学

A　期待値系列と統計多様体

0　導入

そもそも情報幾何学がどのような分野を指すのかは未定義

- 狭義の情報幾何：　双対的なアファイン接続や，ダイバージェンスとよぶ距離2乗型関数を用いた幾何学
- 広義の情報幾何：　情報の持つ数理構造を幾何学を用いて表現する方法論全般（カーネル法や Wasserstein 幾何学も含む）

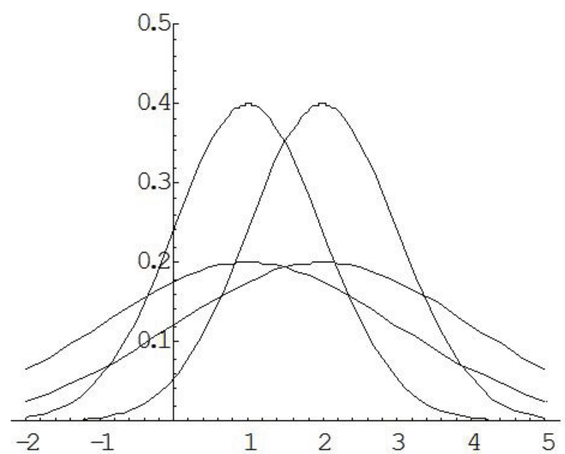
現在は，情報幾何学の研究の中心は欧州になりつつある.

1 統計モデル

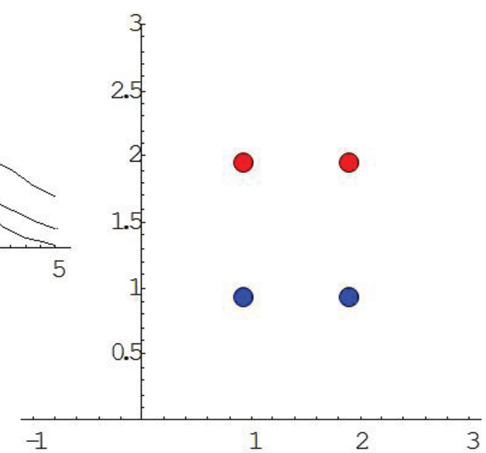
定義 1.1
S が Ω 上の 統計モデル （またはパラメトリックモデル）
$\overset{\text{def}}{\Longleftrightarrow}$ S が $\xi \in \Xi$ をパラメータとする確率密度関数族で

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

S を $\{\Xi; \xi^1, \ldots, \xi^n\}$ を局所座標系とする多様体（曲がった空間）とみなす.



多様体



局所座標系

統計モデルが多様体となるための仮定

(0) $p(x;\xi) > 0,\ \forall x \in \Omega,\ \xi \in \Xi$
（パラメータに関して確率分布のサポートが変化しない）

(1) パラメータ空間 Ξ は R^n の開集合

(2) $\xi \mapsto p(x;\xi)$ はほとんどすべての x に対して 1 対 1
（S は識別可能であると言う）

(3) $\xi \mapsto p(x;\xi)$ は（とりあえず）C^4 級

(4) 微分と積分の順序交換可能

$$\text{例えば} \int \frac{\partial}{\partial \xi^i} p(x;\xi)dx = \frac{\partial}{\partial \xi^i} \int p(x;\xi)dx$$
$$\left(= \frac{\partial}{\partial \xi^i} 1 = 0 \right)$$

$S = \{p(x;\xi) \mid \xi \in \Xi\}$ に対して

$$\phi : S \to R^n; \quad \phi(p(x;\xi)) = \xi$$
$$\xi = (\xi^1, \xi^2, \cdots, \xi^n)$$

とおく． $(\xi^1, \xi^2, \cdots, \xi^n)$ を一つの局所座標系と考え，S を多様体とみなす．

1.2 Fisher 計量

統計モデルの Riemann 計量

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

$$\overset{\text{def}}{\Longleftrightarrow} \quad g^F_{ij}(\xi) := E_p[\partial_i l_\xi \partial_j l_\xi] = \int_\Omega \frac{\partial}{\partial \xi^i} \log p(x;\xi) \frac{\partial}{\partial \xi^j} \log p(x;\xi) p(x;\xi) dx$$

Fisher 情報行列に対する仮定

(5) $^\forall \xi, {}^\forall i, j$ に対して $g^F_{ij}(\xi) < \infty$

(6) $g^F_{ij} : \Xi \to R$ は（とりあえず）C^4 級

命題 1.2

g^F は非負定値対称行列である．

（証明）　任意の $c \in R^n$ に対して

$$ {}^t c g^F c = \sum_{i,j=1}^n c^i c^j g^F_{ij}(\xi) = \int_\Omega \left\{ \sum_{i=1}^n c^i \partial_i l(x;\xi) \right\}^2 p(x;\xi) dx \geq 0$$

Fisher 計量に関する仮定

(7) g^F は正定値である．

命題 1.3　次の条件は同値

(1) g^F は正定値.

(2) $\{\partial_1 p_\xi, \ldots, \partial_n p_\xi\}$ は Ω 上の関数として線形独立.

(3) $\{\partial_1 l_\xi, \ldots, \partial_n l_\xi\}$ は Ω 上の関数として線形独立.

$g^F = (g_{ij})$: S の **Fisher 計量**

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

$$= \int_\Omega \left(\frac{\partial}{\partial \xi^i} p_\xi\right)\left(\frac{\partial}{\partial \xi^j} \log p_\xi\right) dx \tag{1}$$

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

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

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

($p(x;\theta)$ のスコア関数)（統計モデルの接ベクトル）

注意 1.4
p の識別可能性と g の正定値性は同値ではない.

(1) 統計モデル $S = \{p(x;u)\}$ を

$$p(x;u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-u)^2}{2}\right]$$

で定めると，S は正則である.

(2)

$$p(x;u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-u^2)^2}{2}\right] \quad (u \neq 0)$$

で定めると，g^F は正定値であるが p は識別可能ではない.

(3)

$$p(x;u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-u^3)^2}{2}\right]$$

で定めると，g^F は正定値ではないが p は識別可能である.

1.3 α-接続

統計モデルのアファイン接続，共変微分

$\nabla^{(\alpha)}$: α-接続 （$\alpha \in \boldsymbol{R}$: 固定） （曲がった空間の微分）

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

$$g^F(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = g^F\left(\sum_{l=1}^{n}\Gamma^{(\alpha)l}_{ij}\partial_l, \partial_k\right) = \Gamma^{(\alpha)}_{ij,k}$$

$\Gamma^{(\alpha)l}_{ij}$: ベクトル場 $\dfrac{\partial}{\partial\xi^j}$ を $\dfrac{\partial}{\partial\xi^i}$ 方向に微分したときの $\dfrac{\partial}{\partial\xi^l}$ 成分

（補足）　ある種の変換性を満たす n^3 個の関数からアファイン接続が定義される．
(接ベクトルが共変微分でどのように変化するか指定すれば良い)

$\nabla^{(0)}$: g^F の Levi-Civita 接続

$\nabla^{(e)} := \nabla^{(1)}$: 指数型接続

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

統計モデルのアファイン接続，共変微分

$\nabla^{(\alpha)}$: α-接続 （$\alpha \in \boldsymbol{R}$: 固定） （曲がった空間の微分）

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

$$g^F(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k) = g^F\left(\sum_{l=1}^{n}\Gamma^{(\alpha)l}_{ij}\partial_l, \partial_k\right) = \Gamma^{(\alpha)}_{ij,k}$$

$\Gamma^{(\alpha)l}_{ij}$: ベクトル場 $\dfrac{\partial}{\partial \xi^j}$ を $\dfrac{\partial}{\partial \xi^i}$ 方向に微分したときの $\dfrac{\partial}{\partial \xi^l}$ 成分

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

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

($\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は g^F に関する双対接続)

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

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

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

1.4 指数型分布族

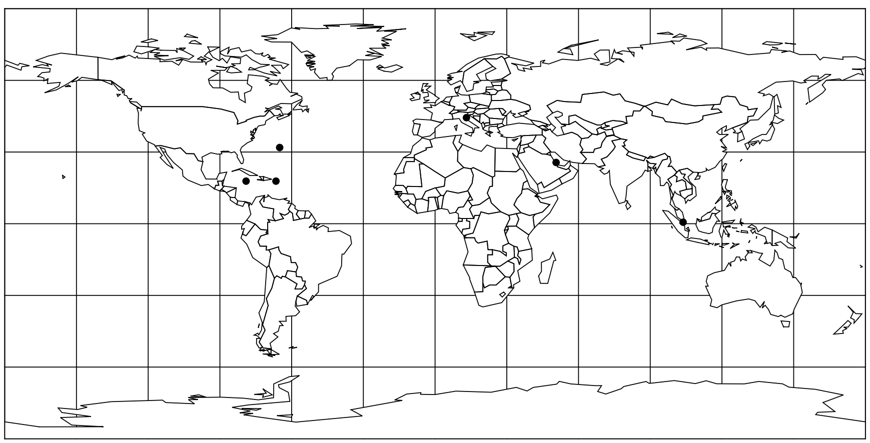
統計モデル S_e が 指数型分布族

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

$C, F_1, \cdots, F_n$: 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.



局所座標系は目的に合わせて選ぶ

正規分布族

$\Omega = R,\ n = 2,\ \xi = (\mu, \sigma) \in R^2_+$ （上半平面）

$$S = \left\{ p(x;\mu,\sigma) \;\middle|\; p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-u)^2}{2\sigma^2} \right] \right\}$$

Fisher 計量は次で与えられる.

$$(g^F_{ij}) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \left(S \text{ は曲率} -\frac{1}{2} \text{ の定曲率空間（双曲空間）} \right).$$

$\nabla^{(1)}$ と $\nabla^{(-1)}$ は平坦なアファイン接続.

$$\theta^1 = \frac{\mu}{\sigma^2},\ \theta^2 = -\frac{1}{2\sigma^2} \quad \psi(\theta) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2}\log\left(-\frac{\pi}{\theta^2}\right)$$

$$\Longrightarrow p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-u)^2}{2\sigma^2}\right] = \exp\left[x\theta^1 + x^2\theta^2 - \psi(\theta)\right]$$

$\{\theta^1, \theta^2\}$: 自然パラメータ. ($\nabla^{(1)}$-測地座標系)

$$\eta_1 = E[x] = \mu, \quad \eta_2 = E\left[x^2\right] = \sigma^2 + \mu^2.$$

$\{\eta_1, \eta_2\}$: 混合パラメータ. ($\nabla^{(-1)}$-測地座標系)

離散標本空間

$$\Omega_n = \{x_0, x_1, \cdots, x_n\},\ \dim S_n = n$$

$$p(x_i;\eta) = \begin{cases} \eta_i & (1 \le i \le n) \\ 1 - \sum_{j=1}^n \eta_j & (i = 0) \end{cases}$$

$$\Xi = \left\{ \{\eta_1, \cdots, \eta_n\} \;\middle|\; \eta_i > 0\ ({}^\forall i),\ \sum_{j=1}^n \eta_j < 1 \right\}$$

(n-次元確率単体)

Fisher 計量:

$$(g_{ij}) = \frac{1}{\eta_0} \begin{pmatrix} 1 + \frac{\eta_0}{\eta_1} & 1 & \cdots & 1 \\ 1 & 1 + \frac{\eta_0}{\eta_2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & \cdots & 1 + \frac{\eta_0}{\eta_n} \end{pmatrix},$$

ただし $\eta_0 = 1 - \sum_{j=1}^n \eta_j$.

$\left(S\ \text{は曲率}\ \dfrac{1}{4}\ \text{の定曲率空間（半径 2 の球面）}\right)$.

離散標本空間

$$\Omega_n = \{x_0, x_1, \cdots, x_n\},\ \dim S_n = n$$

$$p(x_i;\eta) = \begin{cases} \eta_i & (1 \le i \le n) \\ 1 - \sum_{j=1}^n \eta_j & (i = 0) \end{cases}$$

$$\Xi = \left\{ \{\eta_1, \cdots, \eta_n\} \,\middle|\, \eta_i > 0\ ({}^\forall i),\ \sum_{j=1}^n \eta_j < 1 \right\}$$

(n-次元確率単体)

$\{\theta^1, \cdots, \theta^n\}$: 自然パラメータ. ($\nabla^{(1)}$-測地座標系)

where $\theta^i = \log \dfrac{\eta_i}{1 - \sum_{j=1}^n \eta_j} = \log \dfrac{p(x_i)}{p(x_0)}$.

$\{\eta_1, \cdots, \eta_n\}$: 期待値パラメータ. ($\nabla^{(-1)}$-測地座標系)

Bernoulli 分布族

$\Omega = \{0,1\}, n = 1, \xi = \eta.$

$$C(x) = 0, \quad F(x) = x, \quad \theta = \log \frac{\eta}{1-\eta},$$
$$\psi(\theta) = -\log(1-\eta) \ = \ \log(1+e^{\theta})$$

確率分布は次のように表される.

$$\begin{aligned} p(x;\xi) \ &= \ \eta^x(1-\eta)^{1-x} \ = \ \exp\left[\log \eta^x(1-\eta)^{1-x}\right] \\ &= \ \exp\left[x\theta - \psi(\theta)\right]. \end{aligned}$$

これはベルヌーイ分布族が指数型分布族であることを意味している.
期待値パラメータは次のように与えられる.

$$E[x] \ = \ 1\cdot\eta + 0\cdot(1-\eta) \ = \ \eta$$

Fisher 計量は次となる.

$$g(\eta) = \frac{1}{\eta(1-\eta)}$$

1.5　Markov 埋め込みと Chenstov の定理

定義 1.5 (Markov 埋め込み)
S_n, S_l: 離散標本空間 Ω_n, Ω_l 上の確率分布全体（確率単体）　$(1 \leq n \leq l)$

このとき，以下で構成される写像 $f : S_n \to S_l$ をMarkov 埋め込みという.

(1) $\Omega_l = \cup_{i=0}^{n} C_{(i)}$
ただし $C_{(i)} \subset \Omega_l$,　$C_{(i)} \neq \phi$,　$C_{(i)} \cap C_{(j)} = \phi$　$(i, j = 0, 1, \ldots n)$

(2) $Q_{(j)}$: $\mathrm{supp}\, Q_{(j)} = C_{(j)}$ となる Ω_l 上の確率分布　$(j = 0, 1, \ldots, n)$
すなわち $Q_{(j)} = \left(Q_{(j)}^0, Q_{(j)}^1, \ldots, Q_{(j)}^l\right)$ とすると

$$Q_{(j)}^k > 0 \ (x_k \in C_j), \quad Q_{(j)}^k = 0 \ (x_k \notin C_{(j)})$$

(3) $(y_0, y_1, \ldots y_l) = f(x_0, x_1, \ldots, x_n)$ を以下で定義する

$$y_k := \sum_{j=0}^{n} x_j Q_{(j)}^k$$

要請

S_n の幾何は S_l に埋め込まれた部分多様体 $f(S_n)$ の幾何と同等であるべき

Chentsov の定理

定理 1.6 (Markov 埋め込み)
S_n, S_l: 離散標本空間
$\{g^{[n]} : n \in \mathrm{N}\}$: S_n 上の $(0,2)$ 型テンソル場 $g^{[n]}$ からなる列
$\{C^{[n]} : n \in \mathrm{N}\}$: S_n 上の $(0,3)$ 型テンソル場 $C^{[n]}$ からなる列

$\Longrightarrow$ 任意の Markov 埋め込み $f : S_n \to S_l$ に関する不変性

$$g_p^{[n]}(X, Y) = g_{f(p)}^{[l]}(f_*X, f_*Y)$$
$$C_p^{[n]}(X, Y, Z) = C_{f(p)}^{[l]}(f_*X, f_*Y, f_*Z)$$

を満たすものは，定数倍を除いて

$$g_p^{[n]}(X, Y) = \sum_{i=0}^{n} p_\xi(x_i)(X \log p_\xi(x_i)(Y \log p_\xi(x_i))$$
$$C_p^{[n]}(X, Y, Z) = \sum_{i=0}^{n} p_\xi(x_i)(X \log p_\xi(x_i)(Y \log p_\xi(x_i))(Z \log p_\xi(x_i))$$

に限られる．

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

2　統計多様体と双対平坦空間

2.1　曲率テンソル場，捩率テンソル場

R : ∇ の曲率テンソル場 (curvature tensor field)

$$R(X,Y)Z := \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$$

$$[X,Y] := XY - YX = \sum_{i,j=1}^{n}\left\{X^i(\partial_i Y^j)\partial_j - Y^j(\partial_j X^i)\partial_i\right\}$$

$$R^l_{kij} = \partial_i\Gamma^l_{jk} - \partial_j\Gamma^l_{ik} + \sum_{m=1}^{n}\Gamma^l_{im}\Gamma^m_{jk} - \sum_{m=1}^{n}\Gamma^l_{jm}\Gamma^m_{ik}$$

$$\sum_{l=1}^{n} R^l_{kij}\partial_l = R(\partial_i,\partial_j)\partial_k$$

∂_k を ∂_j 方向に微小平行移動すると $\partial_k \mapsto \partial_k + \sum_{l=1}^{n}\Gamma^l_{jk}\partial_l$ と変化する.

さらにこのベクトル場を ∂_i 方向に微小平行移動すると，全体の変化は

$$\partial_k \mapsto \partial_k + \sum_{l=1}^{n}\Gamma^l_{jk}\partial_l + \left\{\sum_{l=1}^{n}\Gamma^l_{ik}\partial_l + \sum_{l=1}^{n}\left((\partial_i\Gamma^l_{jk}) + \sum_{m=1}^{n}\Gamma^l_{im}\Gamma^m_{jk}\right)\partial_l\right\}$$

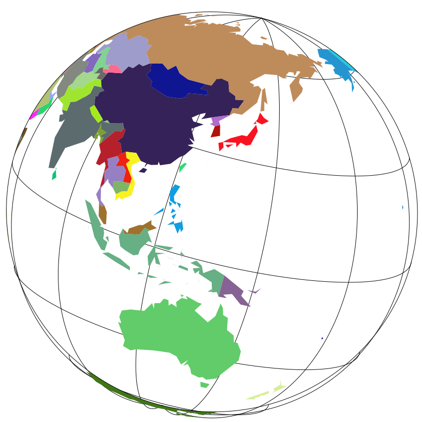
となる. ∂_i, ∂_j と平行移動の順序を入れ換え，差を考えると曲率テンソル場となる.

$\boldsymbol{T}$: $\boldsymbol{\nabla}$ の捩率テンソル場 (torsion tensor field)

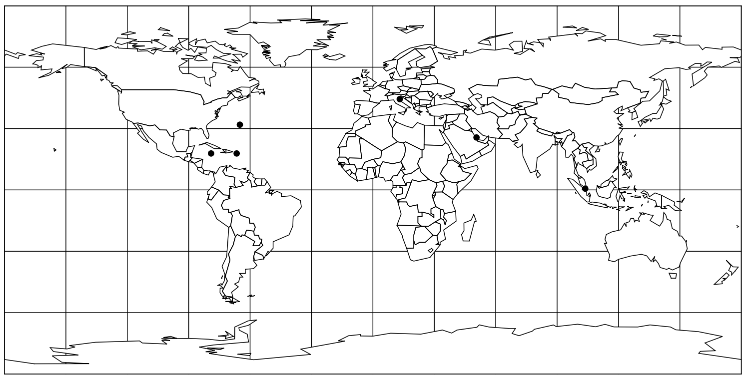
$$\boldsymbol{T}(\boldsymbol{X},\boldsymbol{Y}) := \boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{Y} - \boldsymbol{\nabla}_{\boldsymbol{Y}}\boldsymbol{X} - [\boldsymbol{X},\boldsymbol{Y}]$$

$$[\boldsymbol{X},\boldsymbol{Y}] := \boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X} = \sum_{i,j=1}^{n}\left\{\boldsymbol{X}^i(\partial_i\boldsymbol{Y}^j)\partial_j - \boldsymbol{Y}^j(\partial_j\boldsymbol{X}^i)\partial_i\right\}$$

$$\boldsymbol{T}_{ij}^{k} = \Gamma_{ij}^{k} - \Gamma_{ji}^{k}$$



球面に標準的に入る
Riemann 計量から定まる
Levi-Civita 接続を考える
$\boldsymbol{R} \neq 0,\ \boldsymbol{T} = 0$
（大円が測地線）



緯線と罫線が測地線になるように
アファイン接続を定める

$\boldsymbol{R} = 0,\ \boldsymbol{T} \neq 0$

2.2　統計多様体

M ： 多様体
h ： M 上の非退化 $(0,2)$-テンソル場　(気持ち: g と書くと Riemann 計量)
∇ ： M 上のアファイン接続

定義 2.1
∇^*: ∇ の h に関する双対接続

$$\overset{\text{def}}{\Longleftrightarrow} \quad Xh(Y,Z) = h(\nabla_X Y, Z) + h(Y, \nabla^*_X Z).$$

以下 h は対称（h: semi-Riemannian）, ∇ は捩れがない（$T=0$）と仮定

定義 1 (Kurose)
(M, ∇, h) が統計多様体

$$\overset{\text{def}}{\Longleftrightarrow} (\nabla_X h)(Y,Z) = (\nabla_Y h)(X,Z).$$

$C(X,Y,Z) := (\nabla_X h)(Y,Z)$, 3 次形式, Amari-Chentsov テンソル場
(M, ∇^*, h): (M, ∇, h) の双対統計多様体.

$\nabla^{(0)}$: h の Levi-Civita 接続 $\quad\Longleftrightarrow\quad \nabla^{(0)} = \frac{1}{2}(\nabla + \nabla^*)$

注意 2.2 (Lauritzen による定義)

(M,g) : Riemann 多様体

C : 対称 $(0,3)$-テンソル場

$\Longrightarrow$ (M,g,C) を 統計多様体とよぶ.

命題 2.3

(M,h) : 擬 Riemann 多様体

Levi-Civita 接続 $\nabla^{(0)}$ を持つ

C : 対称 $(0,3)$-テンソル場

$$h(\nabla^{(\alpha)}_X Y, Z) := h(\nabla^{(0)}_X Y, Z) - \frac{\alpha}{2} C(X,Y,Z),$$
$$h(\nabla^{(-\alpha)}_X Y, Z) := h(\nabla^{(0)}_X Y, Z) + \frac{\alpha}{2} C(X,Y,Z),$$

$\Longrightarrow$

(1) $\nabla^{(\alpha)}$ と $\nabla^{(-\alpha)}$ は互いに双対的な捩れのないアファイン接続

(2) $\nabla^{(\alpha)}h$ と $\nabla^{(-\alpha)}h$ は対称

$(M,\nabla^{(\alpha)},h)$ と $(M,\nabla^{(-\alpha)},h)$ は統計多様体

2.3　双対平坦空間

R : ∇ の曲率テンソル場

$$\Longrightarrow \quad h(R(X,Y)Z,V) + h(Z, R^*(X,Y)V) = 0$$

∇ が平坦 $\quad \Longleftrightarrow \quad$ ∇^* が平坦

(M, h, ∇, ∇^*) : 双対平坦空間 $\overset{\text{def}}{\Longleftrightarrow}$ ∇ （と ∇^*） が平坦なアファイン接続

∇ が平坦なアファイン接続
$\Longrightarrow$ M 上の局所座標系 $\{\theta^1, \theta^2, \ldots, \theta^n\}$ で次を満たすものが存在する.

$$\Gamma^{\nabla}_{ij}{}^{k} \equiv 0$$

このような局所座標系 $\{\theta^i\}$ を アファイン座標系という.

(M, h, ∇, ∇^*) ： 双対平坦空間
$\Longrightarrow$ ∇^*-アファイン座標系 $\{\eta_i\}$ 次を満たすものが存在する.

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

$\{\eta_i\}$ を ∇ の h に関する 双対座標系という

命題 2.4

(M, h, ∇, ∇^*) : 双対平坦空間

$\{\theta^i\}$: ∇-アファイン座標系

$\{\eta_i\}$: $\{\theta^i\}$ の双対アファイン座標系

$\Longrightarrow$ M 上の関数 ψ と ϕ が存在して次が成り立つ

$$\frac{\partial\psi}{\partial\theta^i} = \eta_i, \quad \frac{\partial\phi}{\partial\eta_i} = \theta^i, \quad \psi(p) + \phi(p) - \sum_{i=1}^{m} \theta^i(p)\eta_i(p) = 0. \tag{3}$$

さらに次も成り立つ

$$h_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}, \quad h^{ij} = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j}, \tag{4}$$

ただし

(h_{ij}) : 擬 Riemann 計量 h の成分行列

(h^{ij}) : (h_{ij}) の逆行列

ψ を θ-ポテンシャル 関数，ϕ を η-ポテンシャル 関数.

(3) の関係式を Legendre 変換とよぶ

(M, ∇, h)，(M, ∇^*, h) を Hesse 多様体とよぶ

命題 2.5

(M, h, ∇, ∇^*) : 双対平坦空間

$\{\theta^i\}$: ∇-アファイン座標系

$\psi(\theta)$: (M, h, ∇, ∇^*) の θ-ポテンシャル関数

M 上の $(0, 3)$-テンソル場を次で定める

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

このとき次が成り立つ.

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

ただし $\nabla^{(0)}$ は h の Levi-Civita 接続である.

注意 2.6　3 次形式から統計多様体を構成した場合と同じ

2.4　ダイバージェンスと拡張 Pythagoras の定理

正準ダイバージェンス

(M, h, ∇, ∇^*) : 双対平坦空間
$\{\theta^i\}$: ∇-アファイン座標系
$\{\eta_i\}$: $\{\theta^i\}$ の双対アファイン座標系
ψ : h の θ-ポテンシャル
ϕ : h の η-ポテンシャル
このとき $M \times M$ 上の関数 D が (M, h, ∇, ∇^*) の正準ダイバージェンス

$$\overset{\text{def}}{\Longleftrightarrow} \quad D(p, q) = \psi(p) + \phi(q) - \sum_{i=1}^{n} \theta^i(p)\eta_i(q)$$

D を (M, h, ∇, ∇^*) の ∇-ダイバージェンスともいう

$D(p, q)$: ∇-ダイバージェンス $\quad\Longleftrightarrow\quad$ $D(q, p)$: ∇^*-ダイバージェンス

命題 2.7　正準ダイバージェンス D はアファイン座標系の取り方に依存しない

$(\mathrm{R}^n, g^E, \nabla^E, \nabla^E)$: （通常の）Euclid 空間

$$D(p, q) = \frac{1}{2}||p - q||^2$$

拡張 Pythagoras の定理

定理 2.8

(M, h, ∇, ∇^*) : 双対平坦空間

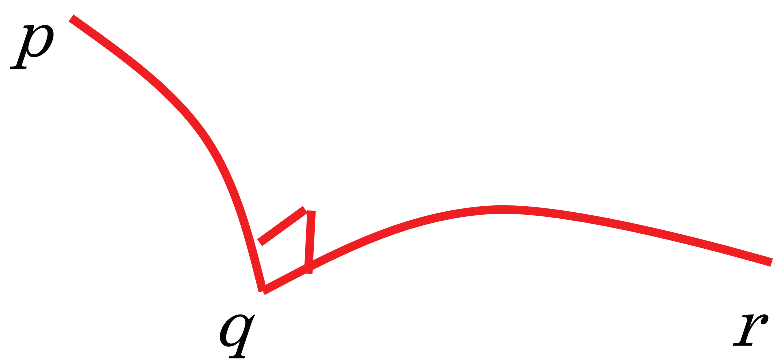
$p, q, r \in M$

γ_1 : p, q を結ぶ ∇-測地線

γ_2 : q, r を結ぶ ∇^*-測地線

q において γ_1 と γ_2 が h に関して直交する

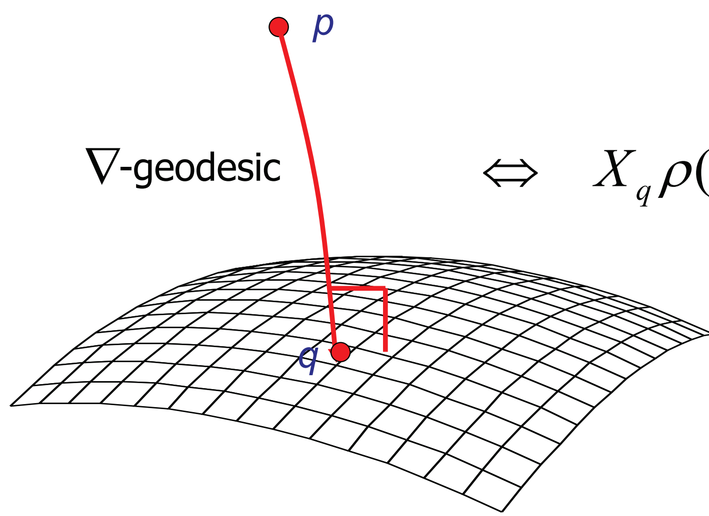
$$\Longrightarrow \quad D(p, r) = D(p, q) + D(q, r)$$



一般化した射影定理

定理 2.9

(M, h, ∇, ∇^*) : 双対平坦空間
$S \subset M$: 部分多様体
$p \in M,\ r \in S$
$D(p, r)$: (M, h, ∇, ∇^*) の ∇-ダイバージェンス
p を固定し $f(r) = D(p, r)$ によって S 上の関数 f を定める.
$f(r)$ が $q \in S$ において停留点となる
$\Longleftrightarrow$ $q \in S$ において p, q を結ぶ ∇-測地線が S と h に関して直交する



$$\Leftrightarrow \quad X_q \rho(p, q) = 0$$

2.5　指数型分布族の幾何学

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 2.10　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.

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

$(S_e, \nabla^{(e)}, g^F)$ 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

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

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

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．

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

3 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

$L(\xi)$: 尤度関数

$$\overset{\text{def}}{\iff} \quad L(\xi) = p(x_1;\xi)p(x_2;\xi)\cdots p(x_N;\xi)$$
$$\left(\iff \quad \log L(\xi) = \sum_{i=1}^{N} \log p(x_i;\xi) \right)$$

$\hat{\xi}$: 最尤推定量

$$\overset{\text{def}}{\iff} \quad \hat{\xi} = \arg\max_{\xi \in \Xi} L(\xi) \quad \left(= \arg\max_{\xi \in \Xi} \log p(x;\xi) \right).$$

尤度最大 $\iff$ **KL-ダイバージェンス最小**

S： 指数型分布族

M： S の曲指数型分布族（S 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

尤度関数は次で計算される：

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

対数尤度方程式は

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

したがって，S の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 **KL-ダイバージェンスは**

$$\begin{aligned} D_{KL}(p(\hat{\eta}), p(\theta(u))) &= D(p(\theta(u)), p(\hat{\eta})) \\ &= \psi(\theta(u)) + \phi(\hat{\eta}) - \sum_{i=1}^{n} \theta^i(u)\hat{\eta}_i \\ &= \phi(\hat{\eta}) - \frac{1}{N}\log L(u). \end{aligned}$$

となる.

尤度最大 $\Longleftrightarrow$ **KL-ダイバージェンス最小**

1 Deformed exponential family (χ-exp. family)

$\chi : (0, \infty) \to (0, \infty)$: strictly increasing (a deformation function)

χ-exponential, χ-logarithm

Definition 1.1

$$\frac{d}{dx} \exp_\chi x = \chi\{\exp_\chi x\} \qquad \chi\text{-exponential}$$

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

We define a function λ by $\lambda(\log_\chi t) = \chi(t)$

In the case $\chi(t) = t$, χ-exponential and χ-logarithm recover the standard exponential and the standard logarithm.

Example 1.2 *In the case $\chi(t) = t^q$, we have*

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

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

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

$\chi : (0, \infty) \to (0, \infty)$: strictly increasing (a deformation function)

χ-exponential, χ-logarithm

Definition 1.1

$$\frac{d}{dx} \exp_\chi x = \chi\{\exp_\chi x\} \qquad \chi\text{-exponential}$$

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

$F_1(x), \ldots, F_n(x)$: functions on Ω

$\theta = \{\theta^1, \ldots, \theta^n\}$: parameters

$$S = \left\{ p(x, \theta) \;\middle|\; p(x; \theta) > 0, \int_\Omega p(x; \theta) dx = 1 \right\}: \text{ statistical model}$$

Definition 1.4

$S_\chi = \{p(x; \theta)\}$: χ-exponential family, deformed exponential family

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

S_q: q-exponential family when the deformed exponential is $\exp_q$.

Example 1.5 (Student t-distribution (q-normal distribution))
$\Omega = \boldsymbol{R},\ n = 2,\ \xi = (\mu, \sigma) \in \boldsymbol{R}^2_+$ **(the upper half plane),** $q > 1$.

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

Set

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

Then

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

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

The set of Student t-distributions is a q-exponential family.

Example 1.6 (discrete distributions)

$\Omega = \{x_0, x_1, \ldots, x_n\}$

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

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

Set $\theta^i = \log_\chi p(x_i) - \log_\chi p(x_0) = \log_\chi \eta_i - \log_\chi \eta_0$

Then

$$\log_\chi p(x) = \log_\chi \left(\sum_{i=0}^{n} \eta_i \delta_i(x) \right)$$

$$= \sum_{i=1}^{n} \left(\log_\chi \eta_i - \log_\chi \eta_0 \right) \delta_i(x) + \log_\chi(\eta_0)$$

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

The set of discrete distributions is a χ-exponential family for any χ.

2 Expectation functionals

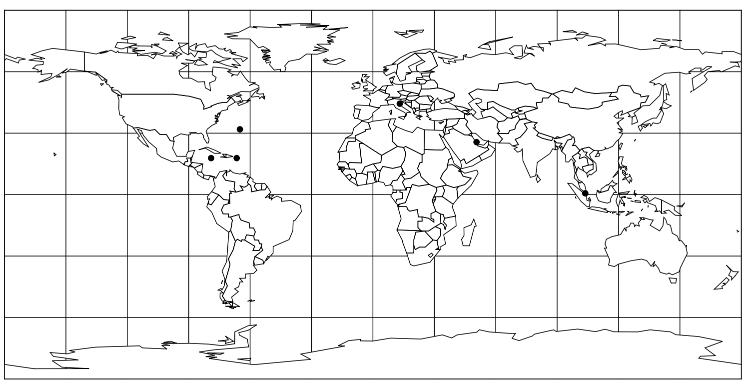
Example 2.1 (q-normal distributions)

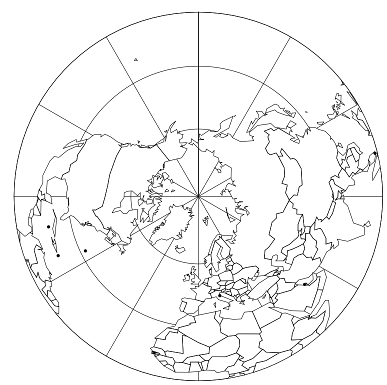
$$p(x;\mu,\sigma) = \frac{1}{z_q}\left[1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right]_+^{\frac{1}{1-q}}$$

q	distribution	mean	variance
-1	semi-circle	○	○
1	normal	○	○
$1+\dfrac{1}{n+1}$	student t	○	○
2	Cauchy	×	×

$q \geq \frac{5}{3}$
$V[X]$ does not exist.
$q \geq 2$
$E[X]$ does not exist.

Expectations give nothing but a local coordinate system.
$\Longrightarrow$ It is better to choose a good local coordinate (expectations).





Local coordinates should be chosen to the purposes.

Definition 2.2
$P_\chi(x)$, $P_\chi^{esc}(x)$: an escort distribution and a normalized escort distribution of $p(x)$

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

$E_{\chi,p}[f(x)]$: the χ-canonical expectation of $f(x)$

$$\overset{\text{def}}{\Longleftrightarrow} \quad E_{\chi,p}[f(x)] = \int_\Omega f(x)P_\chi(x;\theta)dx = \int_\Omega f(x)\chi(p(x;\theta))dx$$

$E_{\chi,p}^{esc}[f(x)]$: the normalized χ-escort expectation of $f(x)$

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

χ-exponential, χ-logarithm

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

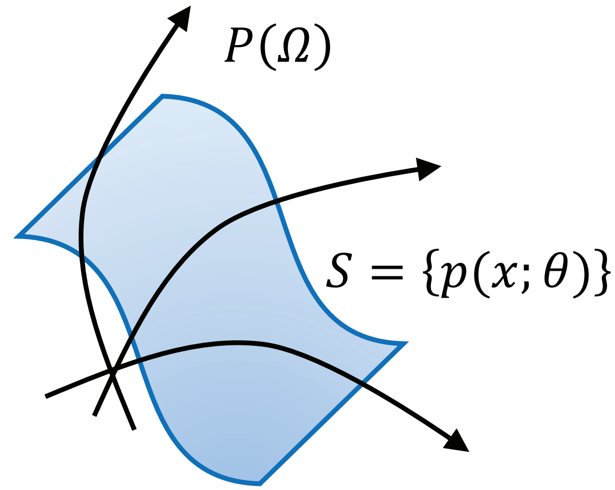
$$\exp_\chi x := 1 + \int_0^x \lambda(t)dt \quad (\text{where } \lambda(\log_\chi t) = \chi(t)) \qquad \chi\text{-exponential}$$

Why the escort expectation needs ?

$\boldsymbol{P(\Omega)}$: the set of positive functions on Ω

$\boldsymbol{S}$: a statistical model

$$S = \left\{ p(x;\theta) \,\middle|\, \int_\Omega p(x;\theta)dx = 1 \right\} \subset \mathcal{P}(\Omega)$$



Expectations should express the "embedding" of S into $\mathcal{P}(\Omega)$

In the case of normal distributions,

$$\frac{\partial}{\partial \mu}\int_{-\infty}^{\infty} p(x;\mu,\sigma)dx = \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sigma^2} p(x;\mu,\sigma)dx = 0$$

$$\frac{\partial}{\partial \sigma}\int_{-\infty}^{\infty} p(x;\mu,\sigma)dx = \int_{-\infty}^{\infty} \frac{(x-\mu)^2-\sigma^2}{\sigma^3} p(x;\mu,\sigma)dx = 0$$

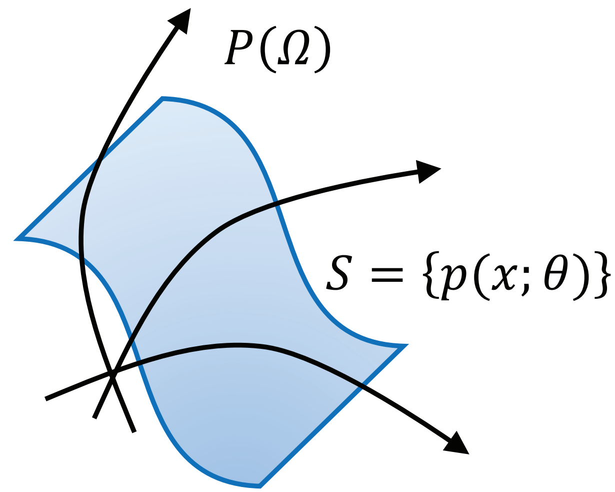
Therefore we obtain

$$E_p[X] = \mu, \qquad E_p[(X-\mu)^2] = V_p[X] = \sigma^2$$

In the case of q-normal distributions

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

$$S_q = \left\{ p(x;\theta) \,\middle|\, \int_\Omega p(x;\theta)dx = 1 \right\} \subset \mathcal{P}(\Omega)$$



$$\frac{\partial}{\partial\mu}\int_{-\infty}^{\infty} p(x;\mu,\sigma)dx$$
$$= \frac{2(z_q)^{q-1}}{(3-q)\sigma^2}\int_{-\infty}^{\infty}(x-\mu)\{p(x;\mu,\sigma)\}^q dx$$
$$= 0$$

Therefore we obtain

$$E_{q,p}^{esc}[X] = \frac{\int_{-\infty}^{\infty} x\{p(x;\mu,\sigma)\}^q dx}{\int_{-\infty}^{\infty}\{p(x;\mu,\sigma)\}^q dx} = \mu$$

$$\left(E_{q,p}^{esc}[(X-\mu)^2] = \sigma^2 \text{ also holds.}\right)$$

$\exp_\chi^{(m)}(x)$: the m-th differential of deformed exponential function

Definition 2.3 (Higher order escort distributions)
$P_{\chi,(m)}(x;\theta)$, $P^{esc}_{\chi,(m)}(x;\theta)$: the m-th escort distribution and the normalized m-th escort distribution of $p(x;\theta)$

$$\overset{\text{def}}{\iff} \quad P_{\chi,(m)}(x;\theta) = \exp_\chi^{(m)}\left[\sum_{i=1}^{n}\theta^i F_i(x) - \Psi(\theta)\right],$$

$$P^{esc}_{\chi,(m)}(x;\theta) = \frac{1}{Z_{\chi,(m)}(\theta)} P_{\chi,(m)}(x;\theta), \; Z_{\chi,(m)}(\theta) = \int_\Omega P_{\chi,(m)}(x;\theta)dx$$

$E_{\chi,(m),p}[f(x)]$: the m-th χ-canonical expectation of $f(x)$

$$\overset{\text{def}}{\iff} \quad E_{\chi,(m),p}[f(x)] = \int_\Omega f(x) P_{\chi,(m)}(x;\theta)dx$$

$E^{esc}_{\chi,(m),p}[f(x)]$: the normalized m-th χ-escort expectation of $f(x)$

$$\overset{\text{def}}{\iff} \quad E^{esc}_{\chi,(m),p}[f(x)] = \int_\Omega f(x) P^{esc}_{\chi,(m)}(x;\theta)dx$$

In the case of q-exponential family $p_q \in S_q$,

$$P_{q,(m)}(x;\theta) := \{q(2q-1)\cdots((m-1)q-(m-2))\}\{p_q(x;\theta)\}^{mq-(m-1)}$$

3 Geometry of deformed exponential families

S_χ : a deformed exponential family

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

$P_{\chi,(m)}(x;\theta)$: the m-th escort distribution of $p(x;\theta) \in S_\chi$

Definition 3.1 (A sequential structure of statistical manifolds)
$g^{(m)}$: m-th Riemannian metric, and $C^{(m)}$: m-th cubic form

$$\begin{aligned} g_{ij}^{(m)}(\theta) &:= \int_\Omega (\partial_i \ln_\chi p_\theta)(\partial_j \ln_\chi p_\theta) P_{\chi,(m)}(x;\theta)dx, \\ C_{ijk}^{(m)}(\theta) &= \int_\Omega (\partial_i \ln_\chi p_\theta)(\partial_j \ln_\chi p_\theta)(\partial_j \ln_\chi p_\theta) P_{\chi,(m+1)}(x;\theta)dx. \end{aligned}$$

Then, we obtain statistical manifold structures

$$(S_\chi, g^{(1)}, C^{(1)}) \to (S_\chi, g^{(2)}, C^{(2)}) \to \cdots \to (S_\chi, g^{(m)}, C^{(m)}) \to \cdots$$

By setting
$$\Gamma_{ij,k}^{(m)(e)} := \Gamma_{ij,k}^{(m)LC} - \frac{1}{2} C_{ijk}^{(m)}, \quad \Gamma_{ij,k}^{(m)(m)} := \Gamma_{ij,k}^{(m)LC} + \frac{1}{2} C_{ijk}^{(m)},$$

where $\Gamma_{ij,k}^{(m)LC}$ is the connection coefficient of the Levi-Civita connection with respect to $g^{(m)}$, we obtain a pair of dual affine connections.

The first Riemannian metric for S_χ

$$g_{ij}^{(1)}(\theta) = \int_\Omega \partial_i \ln_\chi p(x;\theta) \partial_j \ln_\chi p(x;\theta) P_\chi(x;\theta)\, dx$$

$$= E_{\chi,p}[(\partial_i \ln_\chi p_\theta)(\partial_j \ln_\chi p_\theta)] \leftarrow \chi\text{-canonical expectation}$$

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

$$g_{ij}^{N}(\theta) = \int_\Omega \frac{1}{P_\chi(x;\theta)} \partial_i p(x;\theta) \partial_i p(x;\theta)\, dx$$

Theorem 3.2 ***For a deformed exponential family S_χ,***

$$g_{ij}^{(1)}(\theta) = g_{ij}^{M}(\theta) = g_{ij}^{N}(\theta).$$

Dual affine connections $\nabla^{(1)(e)}, \nabla^{(1)(m)}$:

$$\Gamma_{ij,k}^{(1)(e)}(\theta) = \int_\Omega \partial_k p(x;\theta) \partial_i \partial_j \ln_\chi p(x;\theta) dx$$

$$\Gamma_{ij,k}^{(1)(m)}(\theta) = \int_\Omega \partial_i \partial_j p(x;\theta) \partial_k \ln_\chi p(x;\theta) dx$$

The quadruplet $(S_\chi, g^{(1)}, \nabla^{(1)(e)}, \nabla^{(1)(m)})$ is a dually flat space.

Construction of β-divergence ($\beta = 1 - q$) for S_q

$u_q(x;\theta)$: a weighted score function

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

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

$$= \underbrace{\frac{\partial}{\partial\theta^i} \ln_q p(x;\theta)}_{\text{generalized score function}} - \underbrace{E_p\left[\frac{\partial}{\partial\theta^i} \ln_q p(x;\theta)\right]}_{\text{bias correction}}$$

By integrating $u_q(x;\theta)$, we obtain the β-divergence ($\beta = 1 - q$):

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

Remark 3.3

D_{1-q} induces a dually flat structure $(S_q, g^{(1)}, \nabla^{(1)(m)}, \nabla^{(1)(e)})$.
In the χ-exponential case, it is known as a U-divergence.

The second Riemannian metric for S_χ

$$g_{ij}^{(2)}(\theta) = \int_\Omega \partial_i \ln_\chi p(x;\theta)\partial_j \ln_\chi p(x;\theta) P_{\chi,(2)}(x;\theta)\, dx$$

the second Riemannian metric

$$g_{ij}^\chi(\theta) = \partial_i\partial_j\psi(\theta) \quad (\leftarrow \ \chi\text{-Fisher metric})$$

Theorem 3.4 *For a deformed exponential family S_χ,*

$$g_{ij}^{(2)}(\theta) = Z_{\chi,(1)}(\theta) g_{ij}^\chi(\theta),$$

$$\textit{where } Z_{\chi,(1)} = \int_\Omega P_\chi(x;\theta)dx$$

$$C_{ijk}^\chi(\theta) = \partial_i\partial_j\partial_k\psi(\theta) \quad (\leftarrow \ \chi\text{-Fisher metric})$$

By setting

$$\Gamma_{ij,k}^{(\chi)(e)} := \Gamma_{ij,k}^{(\chi)LC} - \frac{1}{2}C_{ijk}^{(\chi)}, \quad \Gamma_{ij,k}^{(\chi)(m)} := \Gamma_{ij,k}^{(\chi)LC} + \frac{1}{2}C_{ijk}^{(\chi)},$$

where $\Gamma_{ij,k}^{(\chi)LC}$ is the Levi-Civita connection with respect to $g^{(\chi)}$

$\Longrightarrow$ $(S_\chi, g^\chi, \nabla^{\chi(e)}, \nabla^{\chi(m)})$ is a dually flat space.

However, $(S_\chi, g^{(2)}, C^{(2)})$ may not define a dually flat structure.

Construction of α-divergence ($\alpha = 1 - 2q$) for S_q

$(s^q)(x;\theta)$: q-score function

$$\overset{\text{def}}{\Longleftrightarrow} \quad \begin{aligned} (s^q)(x;\theta) &= ((s^q)^1(x;\theta), \ldots, (s^q)^n(x;\theta))^T \\ (s^q)^i(x;\theta) &= \frac{\partial}{\partial\theta^i} \ln_q p(x;\theta) \end{aligned}$$

A q-score function is unbiased w.r.t. the q-canonical expectation.

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

$\Longrightarrow$ We regard that $s^q(x;\theta)$ is a generalization of estimating function.
By integrating $s^q(x;\theta)$, we obtain the α-divergence ($\alpha = 1 - 2q$):

α-divergence ($\alpha = 1 - 2q$)

$$D^{(1-2q)}(p(x), r(x)) = \frac{1}{q}\int_\Omega p(x)^q \{\ln_q p(x) - \ln_q r(x)\} dx$$

$D^{(1-2q)}$ induces a non-flat invariant statistical manifold $(S_q, \nabla^{(1-2q)}, g^F)$.

α-divergence ($\alpha = 1 - 2q$)

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

$D^{(1-2q)}$ induces a non-flat invariant statistical manifold $(S_q, \nabla^{(1-2q)}, g^F)$, which coincides with $(S_q, \nabla^{(2)(e)}, g^{(2)})$.

normalized Tsallis relative entropy (χ-relative entropy)

$$D^q\left(p(x), r(x)\right) = E^{esc}_{q,p}\left[\ln_q p(x) - \ln_q r(x)\right]$$

$$= \int_\Omega \frac{p(x)^q}{Z_q(p)} \{\ln_q p(x) - \ln_q r(x)\} dx \quad \left(= \frac{q}{Z_q(p)} D^{(1-2q)}(p, r) \right)$$

D^q induces a dually flat structure $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$.

The difference is just a normalization $Z_q(\theta)$.

$$\underset{\text{pos. measure}}{\nu(x)} \quad \not\rightarrow \quad \underset{\text{prob. measure}}{\frac{\nu(x)}{Z_q(\nu)}}$$

Normalization of a positive measure to a probability measure is NOT a trivial problem.

Summary

Higher order escort distributions

$P_{\chi,(m)}(x;\theta)$, $P^{esc}_{\chi,(m)}(x;\theta)$: the m-th escort distribution and the normalized m-th escort distribution of $p(x;\theta)$

$$\overset{\text{def}}{\Longleftrightarrow} \quad P_{\chi,(m)}(x;\theta) = \exp_\chi^{(m)}\left[\sum_{i=1}^{n}\theta^i F_i(x) - \Psi(\theta)\right],$$

$$P^{esc}_{\chi,(m)}(x;\theta) = \frac{1}{Z_{\chi,(m)}(\theta)}P_{\chi,(m)}(x;\theta), Z_{\chi,(m)}(\theta) = \int_\Omega P_{\chi,(m)}(x;\theta)dx$$

$E_{\chi,(m),p}[f(x)]$: the m-th χ-canonical expectation of $f(x)$

$$\overset{\text{def}}{\Longleftrightarrow} \quad E_{\chi,(m),p}[f(x)] = \int_\Omega f(x)P_{\chi,(m)}(x;\theta)dx$$

$E^{esc}_{\chi,(m),p}[f(x)]$: the normalized m-th χ-escort expectation of $f(x)$

$$\overset{\text{def}}{\Longleftrightarrow} \quad E^{esc}_{\chi,(m),p}[f(x)] = \int_\Omega f(x)P^{esc}_{\chi,(m)}(x;\theta)dx$$

First escort distribution $(S_q, g^{(1)}, \nabla^{(1)(e)}, \nabla^{(1)(m)})$

Riemannian metric: $$g_{ij}^{(1)}(\theta) = \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) P_q(\theta) dx$$

$(S_q, g^{(1)}, \nabla^{(1)(e)}, \nabla^{(1)(m)})$: a **dually flat space**,
which is induced from **β-divergence**.

Second escort distribution $(S_q, g^{(2)}, \nabla^{(2)(e)}, \nabla^{(2)(m)})$

$$\begin{aligned} g_{ij}^{(2)}(\theta) &= \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) P_{q,(2)}(\theta) dx \\ &= (1/q) \int_\Omega (\partial_i \ln p_\theta)(\partial_j \ln p_\theta) p_\theta dx \\ &= (1/q)\, g_{ij}^F(\theta) \quad \text{((standard) Fisher metric)} \end{aligned}$$

$(S_q, g^{(2)}, \nabla^{(2)(e)}, \nabla^{(2)(m)})$: an **invariant statistical manifold structure**
which is induced from **α-divergence**.

Future problem

The meaning of sequential structure is not clear at this moment.

$$(S_\chi, g^{(1)}, C^{(1)}) \to (S_\chi, g^{(2)}, C^{(2)}) \to \cdots \to (S_\chi, g^{(m)}, C^{(m)}) \to \cdots$$

Probably, projective equivalence of statistical manifolds is important for this problem.

Markov埋め込み

Definition 3.5 (Markov 埋め込み)
S_n, S_l: 離散標本空間 Ω_n, Ω_l 上の確率分布全体（確率単体）　$(1 \leq n \leq l)$

このとき，以下で構成される写像 $f : S_n \to S_l$ をMarkov埋め込みという.

(1) $\Omega_l = \cup_{i=0}^{n} C_{(i)}$
　ただし $C_{(i)} \subset \Omega_l, \quad C_{(i)} \neq \phi, \quad C_{(i)} \cap C_{(j)} = \phi \quad (i, j = 0, 1, \ldots n)$

(2) $Q_{(j)} : \mathrm{supp}\, Q_{(j)} = C_{(j)}$ となる Ω_l 上の確率分布　$(j = 0, 1, \ldots, n)$
　すなわち $Q_{(j)} = \left(Q^0_{(j)}, Q^1_{(j)}, \ldots, Q^l_{(j)}\right)$ とすると

$$Q^k_{(j)} > 0 \ (x_k \in C_j), \quad Q^k_{(j)} = 0 \ (x_k \notin C_{(j)})$$

(3) $(y_0, y_1, \ldots y_l) = f(x_0, x_1, \ldots, x_n)$ を以下で定義する

$$y_k := \sum_{j=0}^{n} x_j Q^k_{(j)}$$

Tsallis 統計学では Markov 埋め込みの仮定は不自然？