情報幾何学入門
ー　幾何学者から見た情報幾何学　ー

松添 博　　（名古屋工業大学）
専門：　幾何学

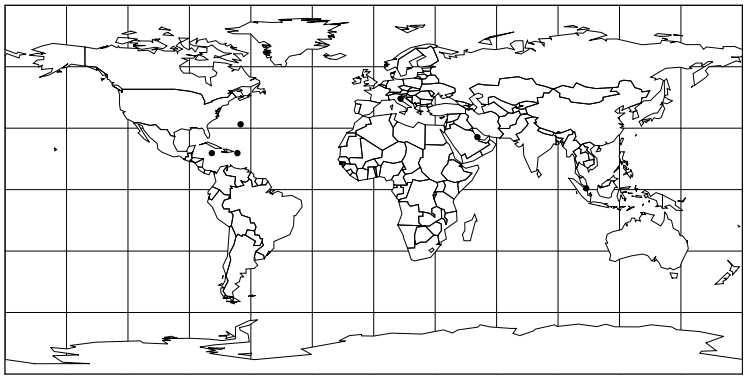
2017年9月28日-29日
MACSセミナー

1　多様体速習
2　統計モデルの幾何学
3　双対接続と双対平坦空間
4　最尤推定量の幾何学

A1　擬似スコア関数とプレ・コントラスト関数
A2　**Tsallis** 統計学

記号	用語	考え方
M	多様体	高次元の曲がった空間 局所座標系を持つ
T_pM $\frac{\partial}{\partial x^i}, X$	接（ベクトル）空間 接ベクトル，接ベクトル場	M を局所線形化したベクトル空間 微分作用素（微分演算子）が接ベクトル
g，h	リーマン計量	曲がった空間の内積
∇，Γ^k_{ij}	アファイン接続，共変微分	多様体上の微分規則
	測地線, geodesic	曲がった空間における「直線」

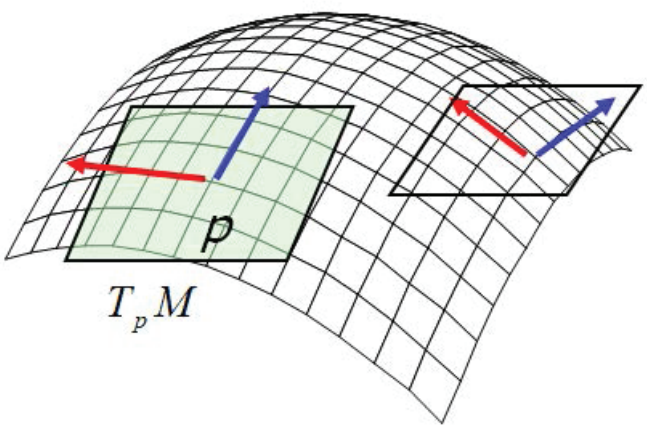
多様体



局所座標系（Euclid 空間への表現）

記号	用語	考え方
M	多様体	高次元の曲がった空間 局所座標系を持つ
T_pM $\frac{\partial}{\partial x^i}, X$	接（ベクトル）空間 接ベクトル，接ベクトル場	M を局所線形化したベクトル空間 微分作用素（微分演算子）が接ベクトル
g，h	リーマン計量	曲がった空間の内積
∇，Γ_{ij}^k	アファイン接続，共変微分	多様体上の微分規則
	測地線, geodesic	曲がった空間における「直線」

T_pM : $p \in M$ における接（ベクトル）空間
（多様体の接超平面）



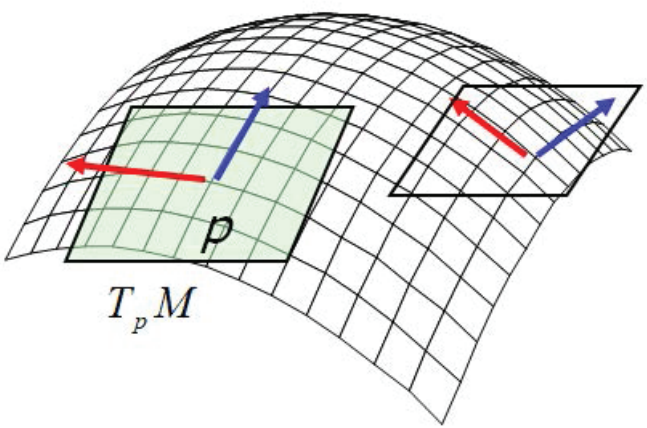
諸般の事情で接ベクトルは微分作用素からなる
$(x^1, x^2, \ldots, x^n)$ 局所座標系

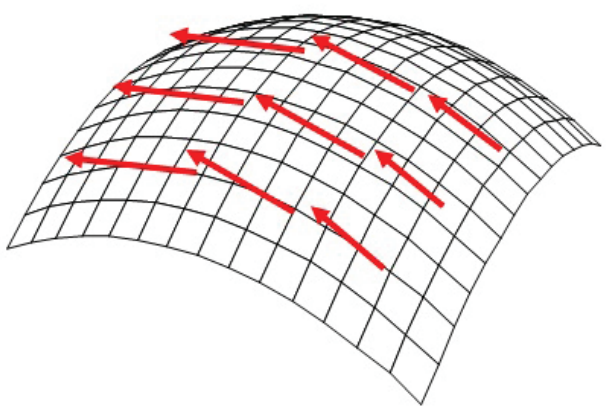
$$\Longrightarrow \left\{ \left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p \right\} : T_pM \text{ の基底}$$

$$X_p = \sum_{i=1}^{n} X^i \left(\frac{\partial}{\partial x^i}\right)_p \in T_pM$$

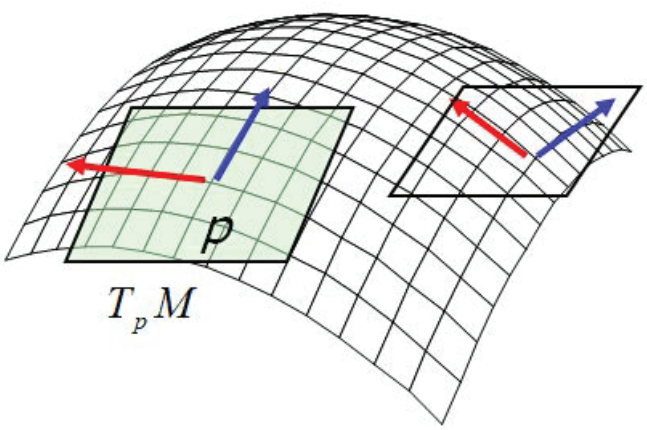
記号	用語	考え方
M	多様体	高次元の曲がった空間 局所座標系を持つ
T_pM $\frac{\partial}{\partial x^i}, X$	接（ベクトル）空間 接ベクトル，接ベクトル場	M を局所線形化したベクトル空間 微分作用素（微分演算子）が接ベクトル
g，h	リーマン計量	曲がった空間の内積
∇，Γ^k_{ij}	アファイン接続，共変微分	多様体上の微分規則
	測地線, geodesic	曲がった空間における「直線」

T_pM : $p \in M$ における接（ベクトル）空間
（多様体の接超平面）
多様体の各点ごとに（滑らかに変化する）接ベクトルを選んだものを（接）ベクトル場という.





記号	用語	考え方
M	多様体	高次元の曲がった空間 局所座標系を持つ
T_pM $\frac{\partial}{\partial x^i}, X$	接（ベクトル）空間 接ベクトル，接ベクトル場	M を局所線形化したベクトル空間 微分作用素（微分演算子）が接ベクトル
g，h	リーマン計量	曲がった空間の内積
∇，Γ_{ij}^k	アファイン接続，共変微分	多様体上の微分規則
	測地線, geodesic	曲がった空間における「直線」



g : M 上の Riemann 計量
（各点ごとに異なる内積が張り付いている）
$\overset{\text{def}}{\Longleftrightarrow}$ $X_p, Y_p \in T_pM$ に対し

$$g(X,Y) = \sum_{i,j=1}^{n} g_{ij} X^i Y^j$$

$$g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$$

（計量ベクトル空間の復習）
（一般の）内積 $\Longleftrightarrow$ 正値2次形式
$\Longleftrightarrow$ 正値対称行列

記号	用語	考え方
M	多様体	高次元の曲がった空間 局所座標系を持つ
T_pM $\frac{\partial}{\partial x^i}, X$	接（ベクトル）空間 接ベクトル，接ベクトル場	M を局所線形化したベクトル空間 微分作用素（微分演算子）が接ベクトル
g，h	リーマン計量	曲がった空間の内積
∇，Γ_{ij}^k	アファイン接続，共変微分	多様体上の微分規則
	測地線, geodesic	曲がった空間における「直線」

多様体上の微分の規則をアファイン接続，共変微分という

- $\nabla : (X, Y) \mapsto \nabla_X Y$ ： アファイン接続，共変微分
- $(\nabla_X Y)_p \in T_pM, \quad \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \sum_{k=1}^{n} \Gamma_{ij}^k \frac{\partial}{\partial \xi^k}$ （成分表示）

  Γ_{ij}^k ：第2種のChristoffel記号（ベクトル場jをi方向に微分したときのk成分）
- $\Gamma_{ij}^{(0)\,k} = \sum_{l=1}^{n} \frac{g^{kl}}{2} \left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \right)$ ： g の Levi-Civita 接続
- 曲線 c が測地線 $\overset{\text{def}}{\Longleftrightarrow}$ $\nabla_{\dot{c}} \dot{c} = 0$ （速度が曲線に沿って変化しない）
  (曲がった空間での直線（最短線ではない）)

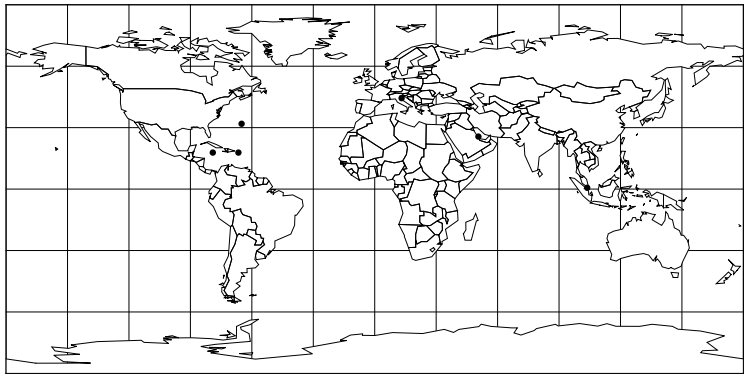
2　統計モデル

定義 2.1
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

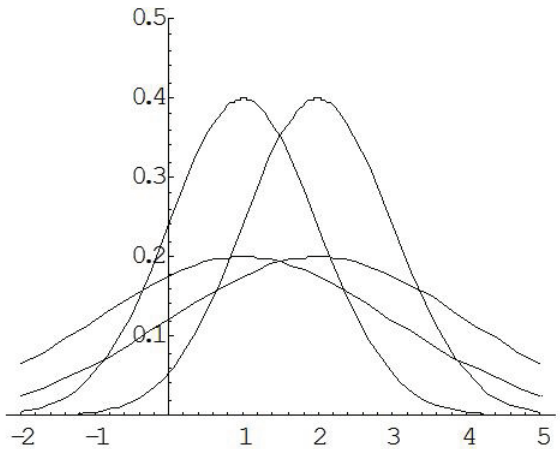
確率密度関数族とパラメータ空間の組 $\{S, \Xi\}$ を統計モデルとよぶこともある.

2　統計モデル

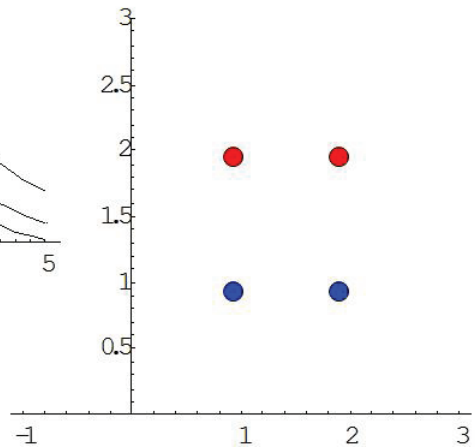
定義 2.1
S が Ω 上の 統計モデル （または パラメトリックモデル）
$\overset{\text{def}}{\Longleftrightarrow}$ S が $\xi \in \Xi$ をパラメータとする確率密度関数族で

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

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．

2.2 Fisher 計量

（復習）　（一般の）内積 $\Longleftrightarrow$ 正値2次形式 $\Longleftrightarrow$ 正値対称行列

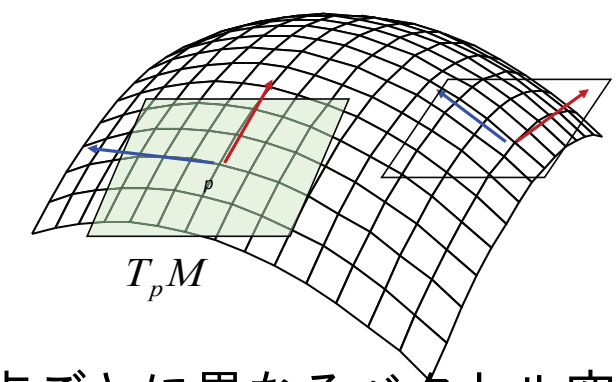
統計モデルの Riemann 計量

$g^F = (g^F_{ij})$ が S の**Fisher 計量**（**Fisher 情報行列**）（曲がった空間の内積）

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

g^F は正値で $g^F_{ij}(\xi) < \infty$ などを仮定

$$E_p[f] = \int_\Omega f(x) p(x;\xi) dx, \quad \partial_i l_\xi = \frac{\partial}{\partial \xi^i} \log p(x;\xi) : \text{ スコア関数.}$$



多様体の各点ごとに異なるベクトル空間が張り付いている

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 2.2
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 2.3 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）

2.3 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 （$\alpha \in R$: 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
微分幾何学で通常扱われるアファイン接続，一般相対性理論など

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

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

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

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

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

2.4 指数型分布族

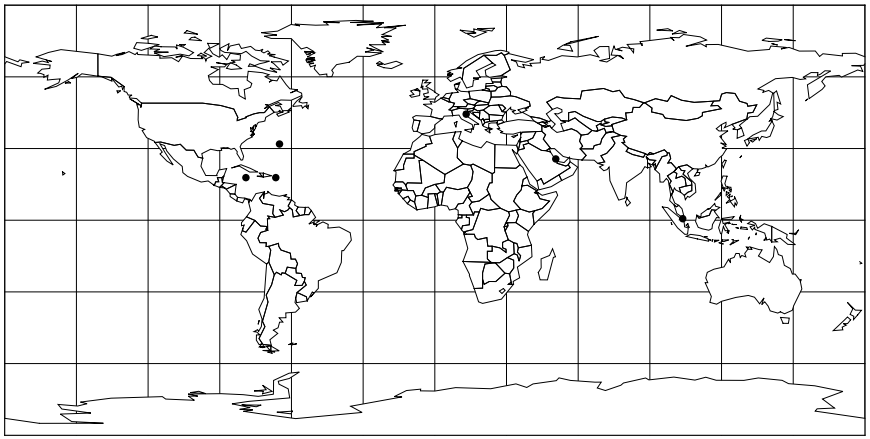
統計モデル S_e が 指数型分布族

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

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



局所座標系は目的に合わせて選ぶ

正規分布族

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

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

Fisher 計量は次で与えられる.

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

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

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

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

$\{\theta^1, \theta^2\}$: 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

$$\Omega = \{x_0, x_1, \cdots, x_n\},\ \dim S = n$$

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

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

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

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

離散標本空間

$$\Omega = \{x_0, x_1, \cdots, x_n\}, \ \dim S = n$$

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

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

(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.

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

3　双対接続と双対平坦空間

3.1　曲率テンソル場，捩率テンソル場

$R : \nabla$ 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．

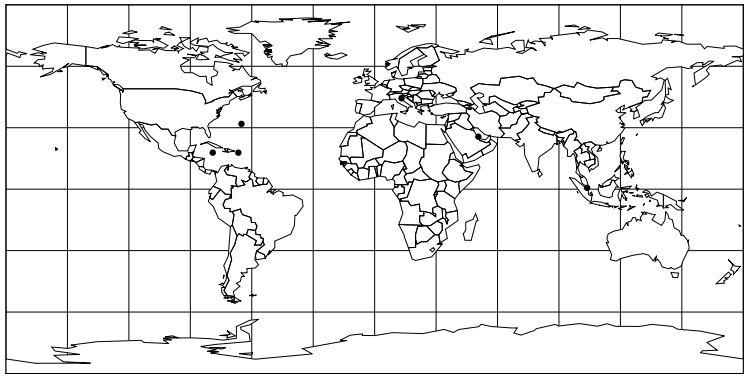
$\boldsymbol{T}$: $\boldsymbol{\nabla}$ の捩率テンソル場 (torsion tensor field)

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

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

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

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

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

3.2　双対接続と双対平坦空間

∇ : M 上のアファイン接続

∇ : 曲率がない $\overset{\text{def}}{\Longleftrightarrow}$ $R = 0$

∇ : 捩れがない $\overset{\text{def}}{\Longleftrightarrow}$ $T = 0$

∇ : 平坦 $\overset{\text{def}}{\Longleftrightarrow}$ $R = 0,\ T = 0$

∇^*: ∇ の g に関する　双対接続（または共役接続）

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z).$$

R : ∇ の曲率テンソル場

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

∇ が平坦 $\Longleftrightarrow$ ∇^* が平坦

(M, g, ∇, ∇^*) : 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.

$$\Gamma_{ij}^{\nabla\ k} \equiv 0$$

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

命題 3.1
(M, g, ∇, ∇^*) : 双対平坦空間
$\{\theta^i\}$: ∇-アファイン座標系
$\Longrightarrow$ ∇^*-アファイン座標系 $\{\eta_i\}$ 次を満たすものが存在する.

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

$\{\eta_i\}$ を ∇ の g に関する 双対座標系という

命題 3.2
(M, g, ∇, ∇^*) : 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

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

ただし
(g_{ij}) : Riemann 計量 g の成分行列
(g^{ij}) : (g_{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, ∇, g)，(M, ∇^*, g) を Hesse 多様体とよぶ

命題 3.3
(M, g, ∇, ∇^*) : 双対平坦空間
$\{\theta^i\}$: ∇-アファイン座標系
$\psi(\theta)$: (M, g, ∇, ∇^*) 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.

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

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

定義 3.4
(M, g) : Riemaann 多様体
C : M 上の $(0,3)$-テンソル場
このとき (M, g, C) を 統計多様体とよぶ

3.3　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, g, ∇, ∇^*) : 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 : g の θ-ポテンシャル
ϕ : g の η-ポテンシャル
このとき $M \times M$ 上の関数 D が (M, g, ∇, ∇^*) 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, g, ∇, ∇^*) の ∇-ダイバージェンスともいう

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

命題 3.5　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 3.6

(M, g, ∇, ∇^*) : 双対平坦空間

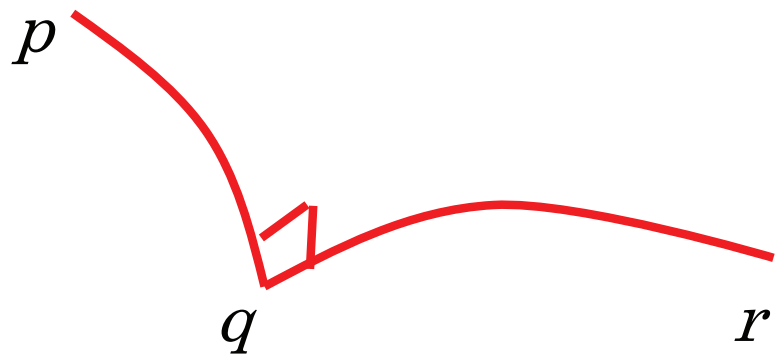
$p, q, r \in M$

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 g 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 3.7

(M, g, ∇, ∇^*) : 双対平坦空間

$S \subset M$: 部分多様体

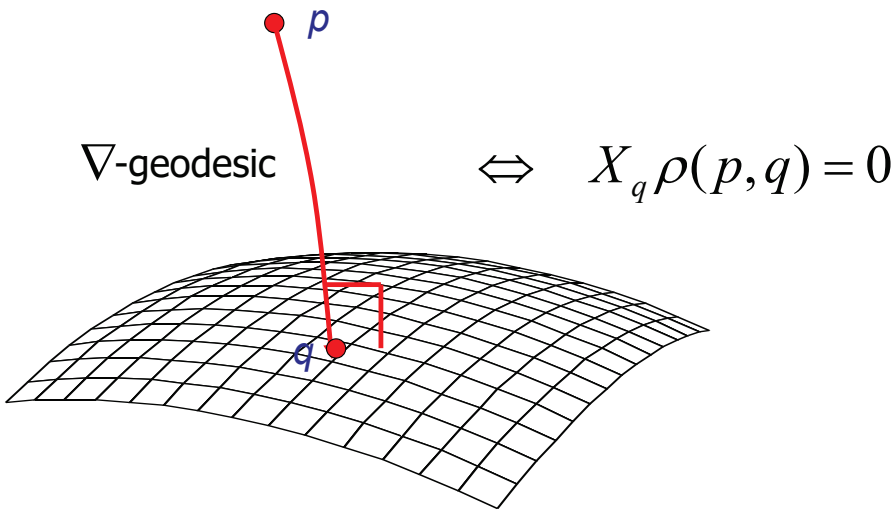
$p \in M,\ r \in S$

$D(p, r)$: (M, g, ∇, ∇^*) 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 g に関して直交する



3.4 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 3.8 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.

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

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

命題 1　指数型分布族 S_e に対して以下が成り立つ.

(1) $(S_e, g^F, \nabla^{(e)}, \nabla^{(m)})$: 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．

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

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．

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

Kullback-Leibler ダイバージェンス，相対エントロピー

$$
\begin{aligned}
D_{KL}\left(p(x;\theta), r(x;\theta')\right) &= \int_\Omega p(x;\theta)\log\frac{p(x;\theta)}{r(x;\theta')}dx \\
&= E_\theta\left[\log p(x;\theta) - \log r(x;\theta')\right]
\end{aligned}
$$

$$
\begin{aligned}
D_{KL}[\partial_i|\partial_j] &= -\int_\Omega \partial_i p(\theta)\partial'_j \log p(\theta')dx\Big|_{\theta=\theta'} \quad \left(\partial_i = \frac{\partial}{\partial\theta^i},\ \partial'_j = \frac{\partial}{\partial\theta'^j}\right) \\
&= -\int_\Omega \partial_i \log p(\theta)\partial'_j \log p(\theta')p(\theta)dx\Big|_{\theta=\theta'} \\
&= -g^F_{ij} \qquad \text{Fisher 計量} \\
D_{KL}[\partial_i\partial_j|\partial_k] &= -\int_\Omega \left(\partial_i\partial_j l(\theta)\partial'_k l(\theta') + \partial_i l(\theta)\partial_j l(\theta)\partial'_k l(\theta')\right)p(\theta)dx\Big|_{\theta=\theta'} \\
&= -\Gamma^{(m)}_{ij,k} \qquad \text{m-接続}
\end{aligned}
$$

KL-ダイバージェンスは不変統計多様体 $(S, \nabla^{(m)}, g^F)$ を誘導する.

KL-ダイバージェンス（幾何学的意味）

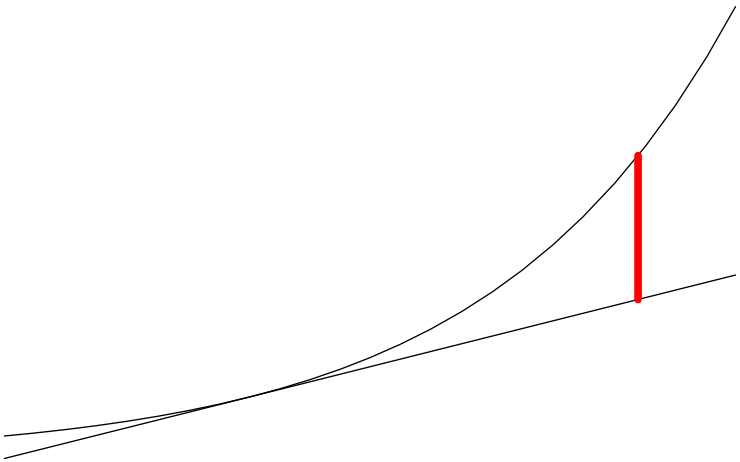
S_e を指数型分布族とし

$\psi(\theta)$: $p(\theta)$ のポテンシャル関数 （パラメータに関する規格化項）

l_θ : ψ の $(\theta, \psi(\theta))$ における接超平面.

$$D_{KL}(p(\theta), p(\theta')) = \psi(\theta') - l_\theta(\theta') \quad + c \quad (c : \text{定数})$$

Kullback-Leibler ダイバージェンスは接超平面 $l_\theta(\theta')$ とポテンシャル関数 $\psi(\theta')$ の差に一致する.



$\{\theta^i\}$: 自然座標系

$$\eta_i = E[F_i(x)] = \frac{\partial \psi(\theta)}{\partial \theta^i}$$

$\{\eta_i\}$: $\{\theta^i\}$ の双対座標系

4　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

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

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

$\hat{\xi}$: 最尤推定量

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

尤度最大 $\Longleftrightarrow$ 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：

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

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