統計的推論に対する
情報幾何学的アプローチ

東北大学経済学部
澤谷一磨

今回の流れ

1. 情報幾何学とは何か
2. 拡張射影定理への道程
3. 情報幾何学的視点による統計的推定

今回の流れ

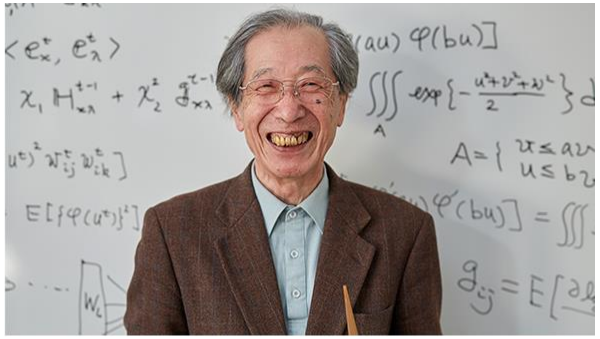
1. 情報幾何学とは何か

2. 拡張射影定理への道程

3. 情報幾何学的視点による統計的推定

情報幾何学の歴史

1922	Fisher情報行列の発見 (R.A. Fisher)
1929	Fisher情報行列をRiemann計量とする (H. Hotelling)
	→しかし論文として発表されず, 2007年にStiglerが明らかにするまで知られることはなかった
1945	Fisher情報行列をRiemann計量とする (C.R. Rao)
1972	Fisher計量が唯一の不変な計量である (N.N. Chentsov)
1975	曲率の導入, 高次漸近理論の考察, 計量的でない接続の考察 (Efron)
1978	自然パラメータと期待値パラメータの双対性 (Barndorff-Nielsen)
1982	双対性の解明, 双対平坦構造を持つ多様体の導入 (S. Amari)
1984	「情報幾何学」のはじまり (Cox卿, Workshop RSS150)
1996	グラフィカルモデルへの導入 (S. Lauritzen)
2019	甘利俊一氏が文化勲章を受章



甘利俊一氏（Riken）

情報幾何のモチベーション

- 情報要素の一つ一つを分離して考えるのではなく，つながった全体つまり多様体として考えてそこに豊かな構造を導入すれば，情報の分野に新しい方法論を提供できるに違いない．（甘利）

- 確率分布に基づくいろいろな分野，例えば統計学・情報理論・システム理論の問題を統一的に扱うことができ，既存の推定法を説明したり，異なる分野の関係を明らかにしたりできるようになる．そういう意味で，情報幾何は異分野間の共通言語的な役割をもつことができる可能性がある（赤穂）

今回の流れ

1. 情報幾何学とは何か
2. 拡張射影定理までの道程
3. 情報幾何学的視点による統計的推定

最低限の知識

- **可微分多様体(differentiable manifold)**

Hausdorff空間Mのn次元C^∞級座標近傍系$\{(\phi_\alpha, U_\alpha)\}_{\alpha\in A}$とは,

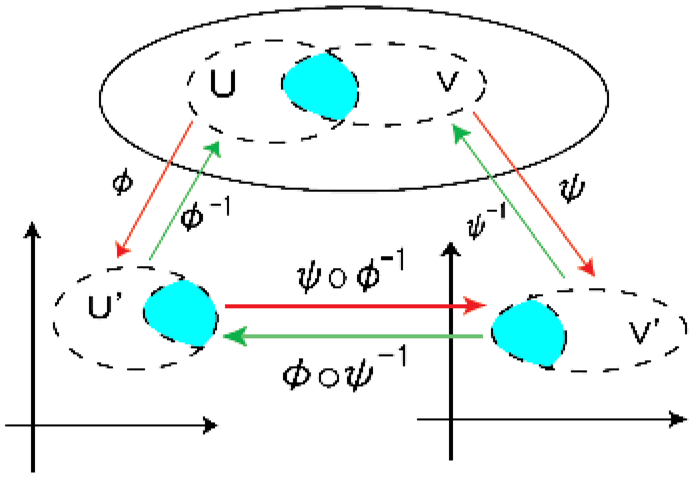
- U_αはMの開集合
- $\phi_\alpha\colon U_\alpha \to \mathbb{R}^n$

であり以下の条件を満たすことである.

- $M = \bigcup_{\alpha\in A} U_\alpha$
- 各$\alpha \in A$に対して$\phi_\alpha(U_\alpha)$は$\mathbb{R}^n$の開集合でありϕ_αは同相写像
- 任意の$\alpha, \beta \in A$で$U_\alpha \cap U_\beta \neq \emptyset$を満たすものに対して,

$$\phi_\alpha \circ (\phi_\alpha)^{-1}\Big|_{\phi_\alpha(U_\alpha\cap U_\beta)} : \phi_\alpha\left(U_\alpha \cap U_\beta\right) \to \phi_\beta(U_\alpha \cap U_\beta)$$

はC^∞級写像である.

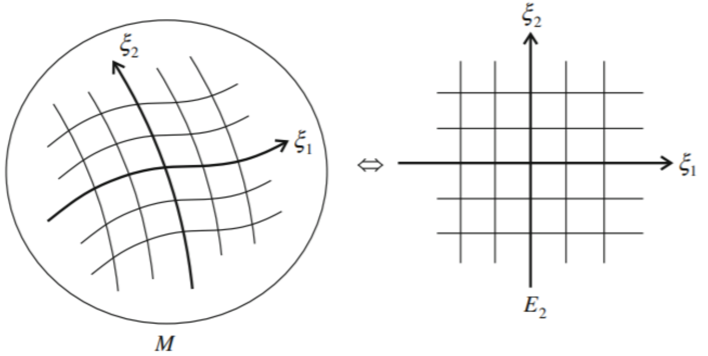


(Wikipediaより)

n次元座標近傍系が与えられたとき, Mを**n次元多様体**という.

最低限の知識

• 多様体をなす空間の例



局所座標系を $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ とする.

- **Euclid空間**

$$r = \sqrt{\xi_1^2 + \xi_2^2}, \theta = \tan^{-1}\frac{\xi_2}{\xi_1} \quad \Leftrightarrow \quad \xi_1 = rcos\theta, \xi_2 = rsin\theta$$

- **確率分布空間**

$$\boldsymbol{\xi} = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma}\right) \ \ s.t. \ \ p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$

- その他
  - 球面, 正測度空間, 正定値行列空間, …

最低限の知識

• ダイバージェンス(divergence)

多様体M上のある2点P,Qについて, そのダイバージェンス$D[P:Q]$, あるいは$D[\boldsymbol{\xi}_P:\boldsymbol{\xi}_Q]$とは,

1. $D[P:Q] \geq 0$
2. $D[P:Q] = 0$ $\textit{if and only if}$ $P = Q$
3. P,Qが十分に近いとき, $\boldsymbol{\xi}_Q = \boldsymbol{\xi}_P + d\boldsymbol{\xi}_P$ として, D のテイラー展開は

$$D[\boldsymbol{\xi}_P:\boldsymbol{\xi}_P + d\boldsymbol{\xi}_P] = \frac{1}{2}\sum g_{ij}(\boldsymbol{\xi}_P)d\xi_i d\xi_j + O(|d\boldsymbol{\xi}|^3).$$

$\boldsymbol{G} = \left(g_{ij}\right)$は正定値行列であり, **計量(metric)**と呼ばれる.

<u>※距離との違い：必ずしも対称性・三角不等式が成立しない</u>

最低限の知識

• ダイバージェンスの例

- Euclidian Divergence

$$D[P:Q] = \frac{1}{2}\Sigma\left(\xi_{P_i} - \xi_{Q_i}\right)^2$$

- **Kullback-Leibler Divergence**

$$D[p:q] = \int p(x) log \frac{p(x)}{q(x)} dx$$

- **α -Divergence**
- **Bregman Divergence**
- **f - Divergence**
- χ^2- Divergence
- γ - Divergence

- (α , β)-Divergence
- U – divergence
- (u, v)-Divergence
- Miscellaneous Divergence

etc.

最低限の知識

- **ダイバージェンスの例**
  - **α-Divergence**

$$D_\alpha[\boldsymbol{p}:\boldsymbol{q}] = \frac{4}{1-\alpha^2}\left(1-\sum p_i^{\frac{1-\alpha}{2}} q_i^{\frac{1+\alpha}{2}}\right), \alpha \neq \pm 1$$

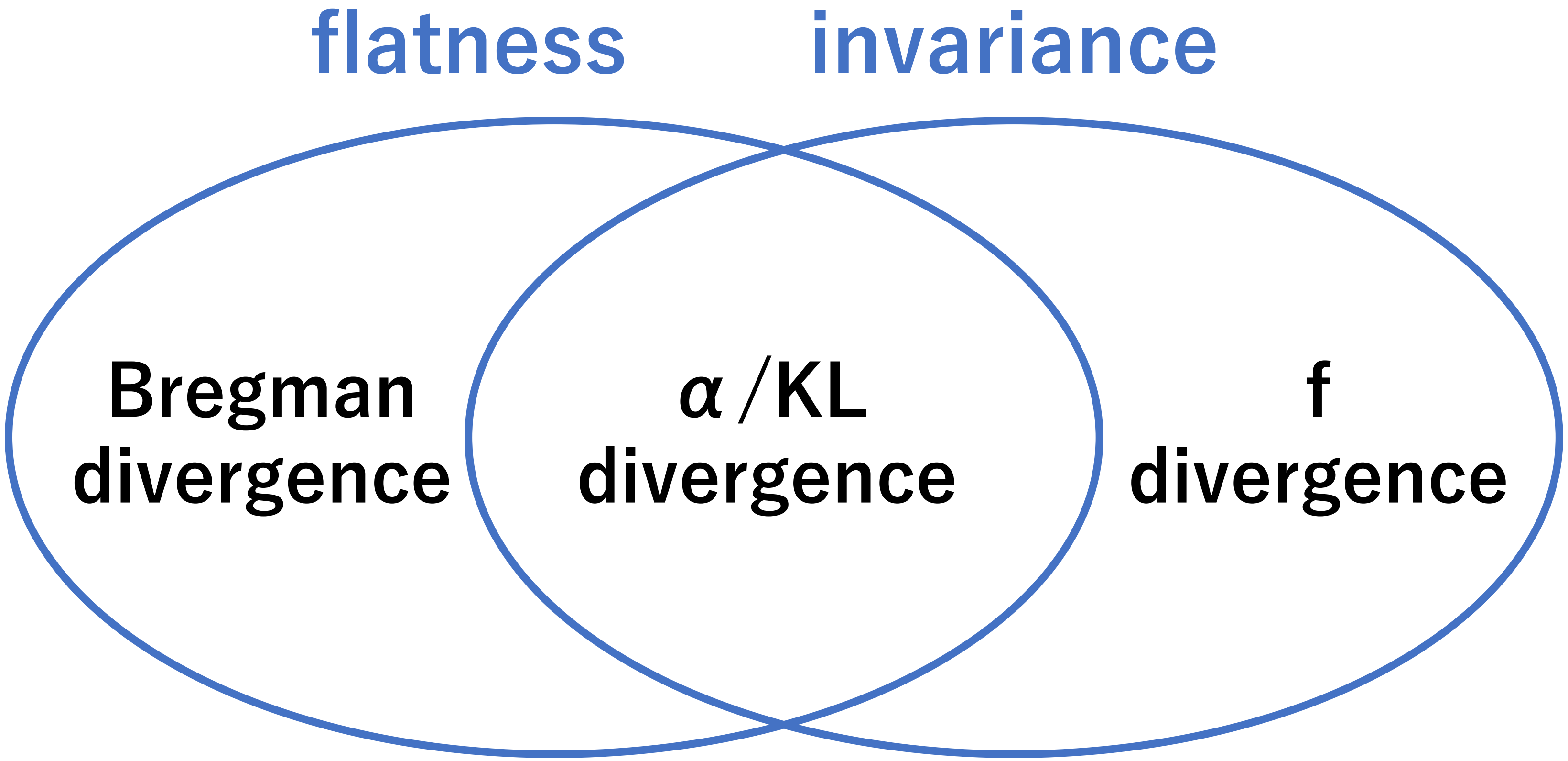
- α→-1　：KLダイバージェンスに相当
- α＝0　：Hellinger距離の二乗に相当
- α→1　：KLダイバージェンスの逆に相当

※ $D_\alpha[p:q] = D_{-\alpha}[q:p]$

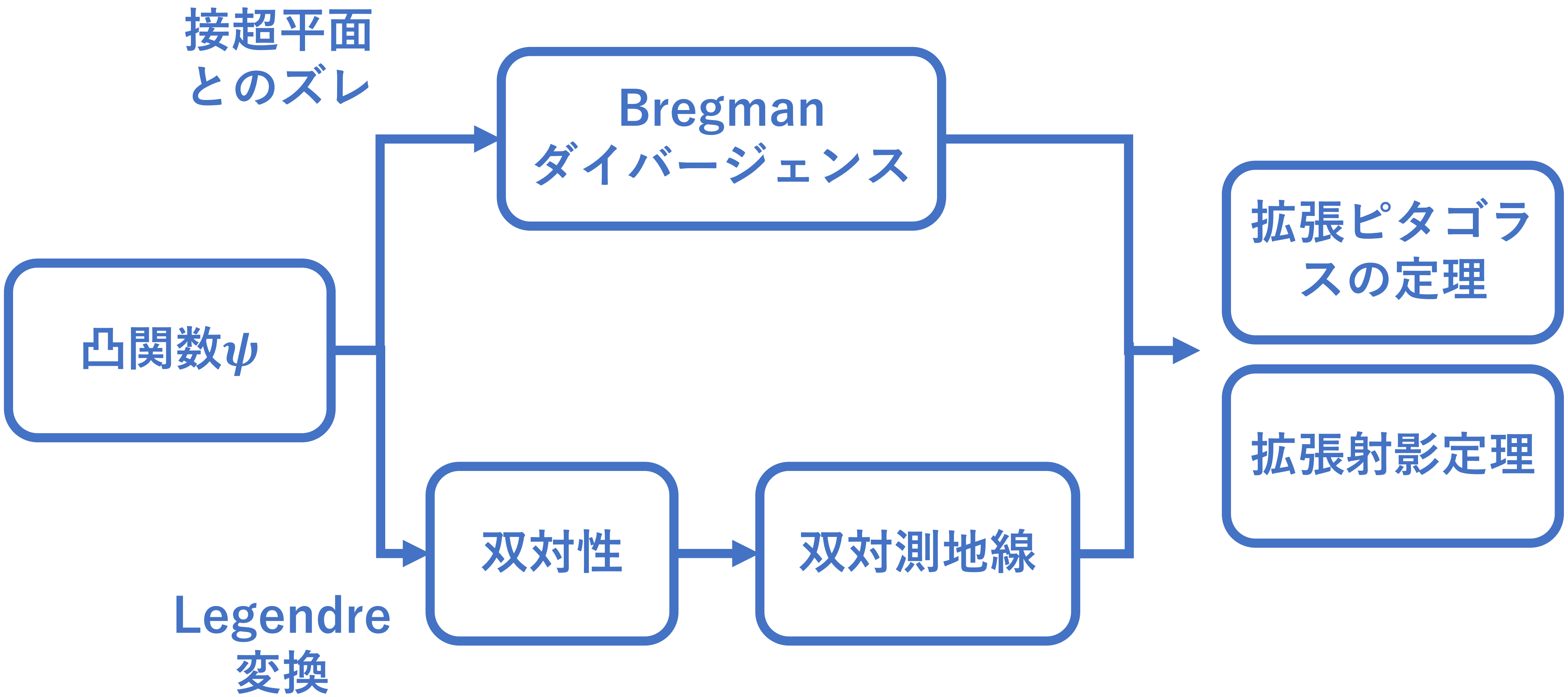
※ 正測度空間$\mathbb{R}^+$(確率分布の制約条件 $\sum p_i = 1$の削除)において不変で平坦な幾何構造をユニークに与える

簡単のための仮定

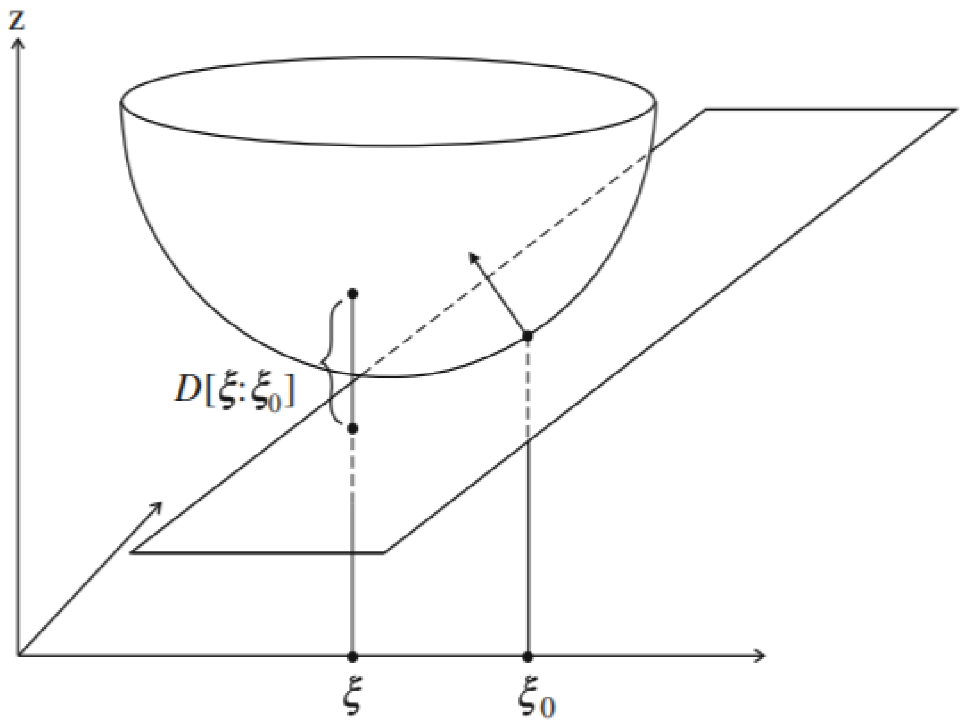
- 平坦性(flatness)



ここからの導出の流れ


接超平面
とのズレ
Bregman
ダイバージェンス
凸関数ψ
拡張ピタゴラスの定理
拡張射影定理
双対性
双対測地線
Legendre
変換

凸関数からの導入

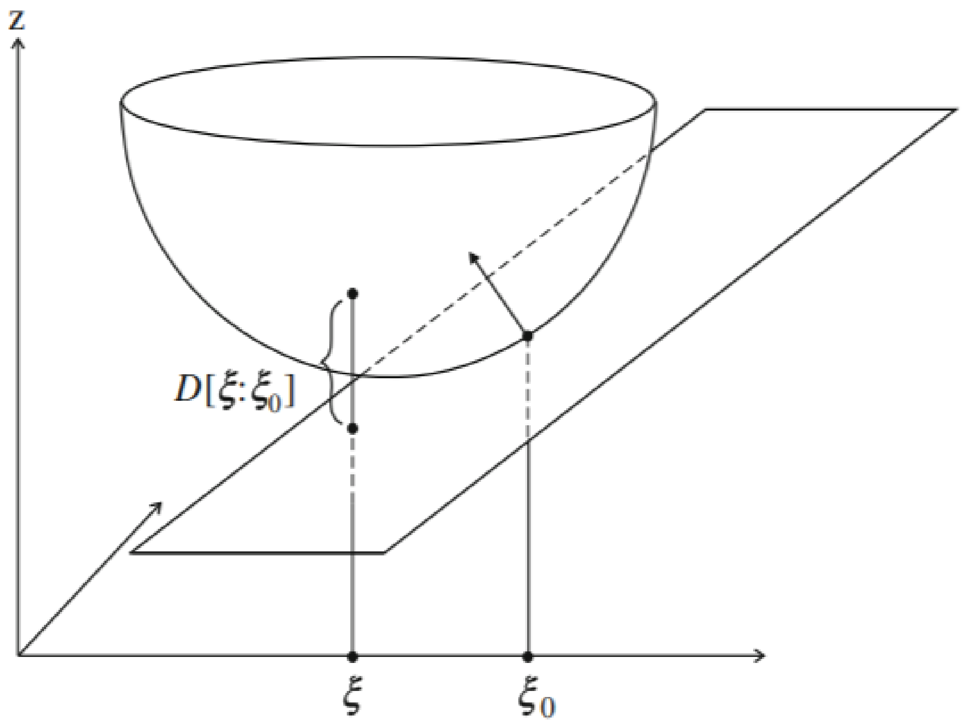


$\psi(\xi)$を凸関数とすると点ξ_0における接超平面は,

$$z = \psi(\xi_0) + \nabla\psi(\xi_0)\cdot(\xi - \xi_0)$$

と書ける.

Bregmanダイバージェンス



このとき, $\boldsymbol{\xi}$における凸関数とその接超平面とのずれを表すと,

$$D[\xi:\xi_0] = \psi(\xi) - \psi(\xi_0) - \nabla\psi(\xi_0)\cdot(\xi - \xi_0)$$

と書くことができ,

これを**Bregmanダイバージェンス**と呼ぶ.

Legendre変換

- 凸関数に対する微分, すなわちその関数の接超平面の勾配（法線）ベクトル $\boldsymbol{\xi}^{\star} = \nabla\psi(\boldsymbol{\xi})$ について, $\boldsymbol{\xi}$ からの変換は一対一で逆変換も存在する.
- よって多様体に2つの双対な座標系が得られる. この変換を**ルジャンドル変換**(Legendre Transformation)といい, すなわち,

$$\psi^{\star}(\boldsymbol{\xi}^{\star}) = \max_{\boldsymbol{\xi}}\{\boldsymbol{\xi}\cdot\boldsymbol{\xi}^{\star} - \psi(\boldsymbol{\xi})\}.$$

- このとき,

$$D^{\star}[\boldsymbol{\xi}^{\star}:\boldsymbol{\xi}_0^{\star}] = D[\boldsymbol{\xi}_0:\boldsymbol{\xi}].$$

- したがって,

$$D_{\psi}[P:Q] = \psi(\boldsymbol{\xi}_P) + \psi^{\star}\left(\boldsymbol{\xi}_Q^{\star}\right) - \boldsymbol{\xi}_P\cdot\boldsymbol{\xi}_Q^{\star}.$$

双対基底系の性質

- **接空間**

曲がったリーマン空間を1点 $\boldsymbol{\xi}$ の近傍において線形空間で近似したもの.

ここで, $\{\boldsymbol{e}_i\}$を基底ベクトルとして以下のように書ける.

$$ds^2 = \langle d\boldsymbol{\xi}, d\boldsymbol{\xi}\rangle = \sum g_{ij} d\xi_i d\xi_j$$

ここで, **<u>2つの双対基底系は双直交系</u>**である. すなわち, $\langle e_i, e_j^\star\rangle = \delta_{ij}$.

証明

$d\boldsymbol{\theta} = \sum d\theta^i \boldsymbol{e}_i = \sum d\theta_i^\star \boldsymbol{e}^{\star i}$ であり, $\boldsymbol{\theta}^\star = \nabla\psi(\boldsymbol{\theta})$より, $d\theta_i^\star = \sum \partial_i\partial_j\psi d\theta^j = \sum g_{ij} d\theta^j$.

これを用いれば, $\boldsymbol{e}_i = \sum g_{ij}\boldsymbol{e}^{\star j}$, $\boldsymbol{e}^{\star j} = \sum g^{\star ij}\boldsymbol{e}_i$ より示された.■

測地線

- **測地線(geodesic)**とは向きを変えない（接ベクトルが平行に保たれる）曲線である. ここでは簡単のためにアファイン座標系 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$が存在して, tをパラメーターとした測地線を以下のようにあらわす.

$$\boldsymbol{\theta}(t) = \boldsymbol{a}t + \boldsymbol{b}.$$

※計量の定まった多様体Mにおいて測地線は局所的に2点間の**最短距離**を与える（M全体を見たときにはもっと短い測地線があっても構わない）

※測地線は直線の概念を曲がった空間に一般化したものと言え, 一般にアファイン接続を通じて次の方程式で与えられる.

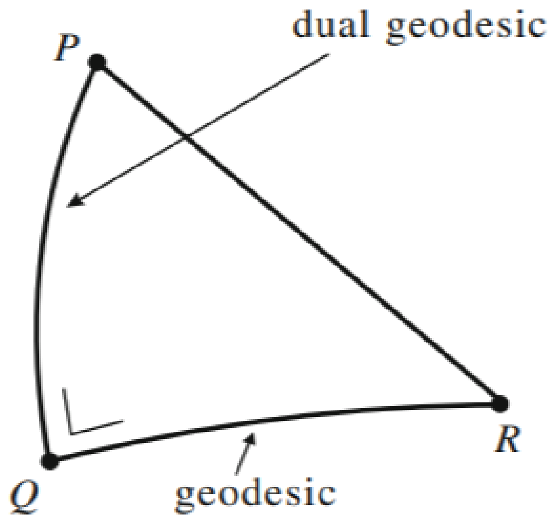
$$\ddot{\xi}^i(t) + \Gamma^i_{jk}\dot{\xi}^j(t)\dot{\xi}^k(t) = 0.$$

拡張ピタゴラスの定理

双対平坦空間における直角三角形において, **拡張ピタゴラスの定理**(Generalized Pythagorean Theorem)が成立する.

すなわち,

$$\boldsymbol{D}[\boldsymbol{P}:\boldsymbol{R}] = \boldsymbol{D}[\boldsymbol{P}:\boldsymbol{Q}] + \boldsymbol{D}[\boldsymbol{Q}:\boldsymbol{R}].$$



証明

P-Qの測地線とQ-Rの測地線の接線方向はそれぞれ,

$$\dot{\boldsymbol{\theta}}^\star(t) = \boldsymbol{\theta}_Q^\star - \boldsymbol{\theta}_P^\star,\ \dot{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}_{\boldsymbol{R}} - \boldsymbol{\theta}_{\mathrm{Q}}$$

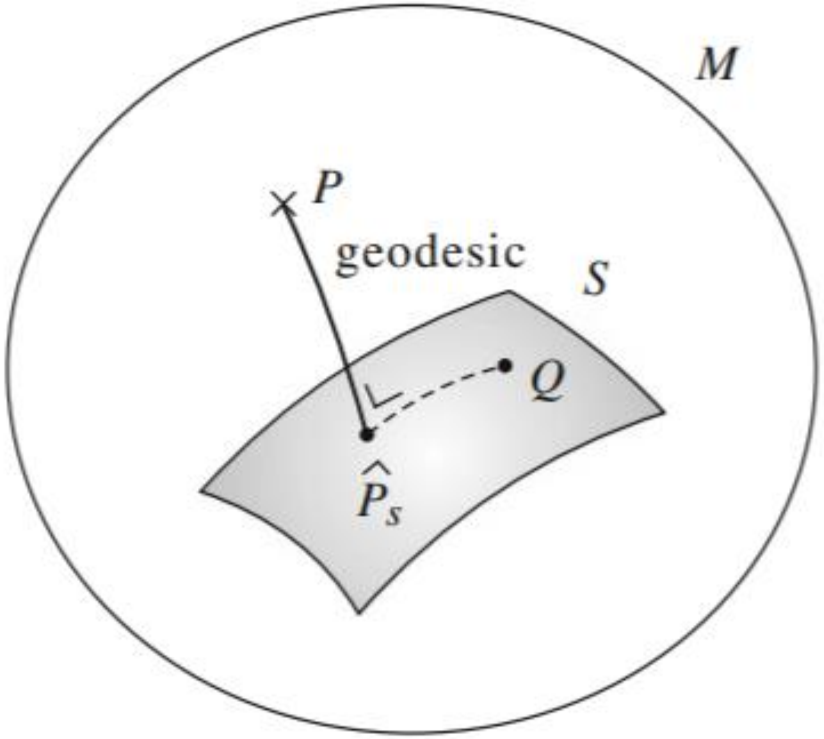
であり, この二つが直交することから積は0.

したがって, Bregmanダイバージェンスの定義より計算し,

$$D[P:Q] + D[Q:R] - D[P:R] = \left(\boldsymbol{\theta}_P^\star - \boldsymbol{\theta}_Q^\star\right)\cdot\left(\boldsymbol{\theta}_Q - \boldsymbol{\theta}_R\right)$$

であることから示された.■

拡張射影定理



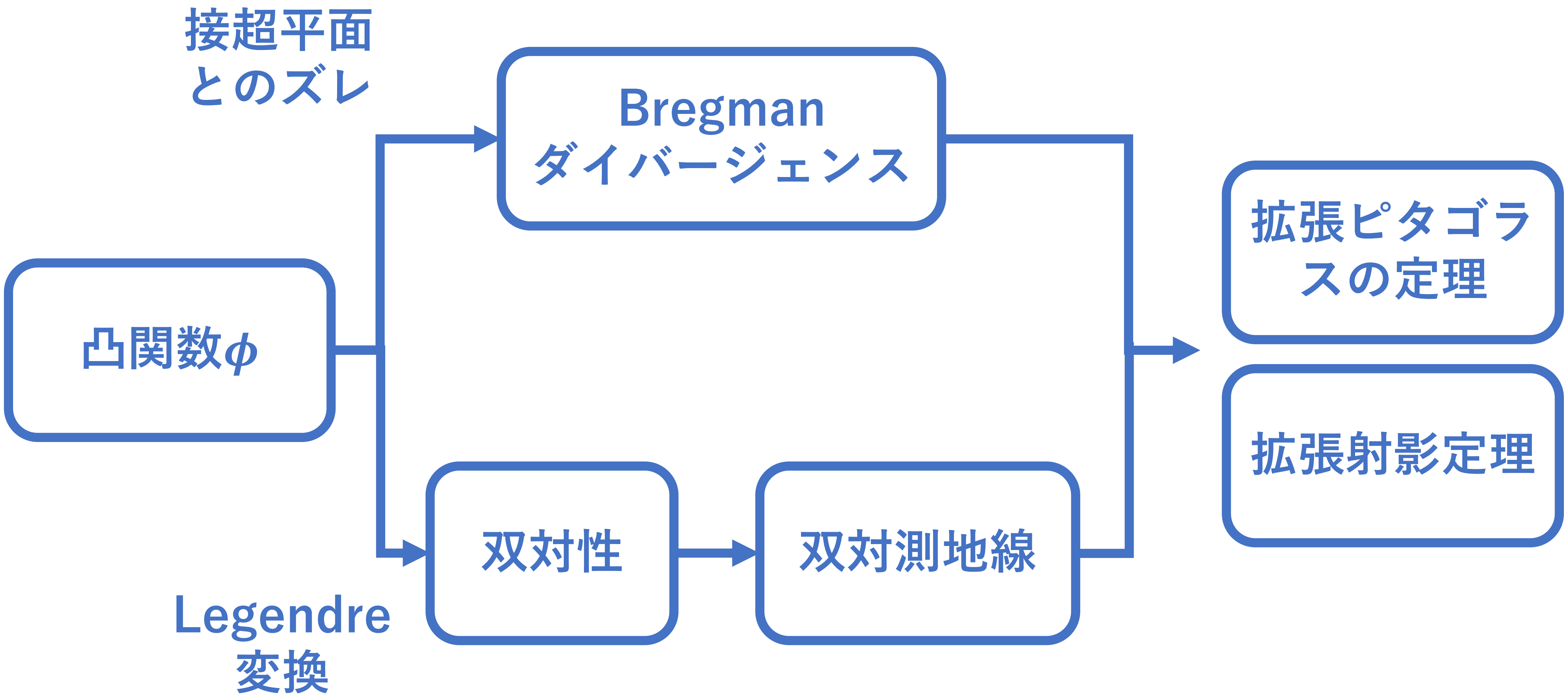
定理　$P \in M$と滑らかな部分多様体$S \subset M$があったとき，

$$D_\psi[P:R], R \in S$$

を最小にする点$\hat{P}_S^\star$ はPからSへの双対測地線による射影となる．この双対も成立する．

※ P_Sはクリティカルポイントを与えるが，Sの形によって極大点や鞍点であったりする．

ここまでのまとめ


接超平面
とのズレ
Bregman
ダイバージェンス
凸関数ϕ
双対性
双対測地線
Legendre
変換
拡張ピタゴラスの定理
拡張射影定理

今回の流れ

1. 情報幾何学とは何か
2. 拡張射影定理までの道程
3. 情報幾何学的視点による統計的推定

- これ以降はホワイトボードで説明

指数型分布族の双対平坦構造

指数型分布族は一般に

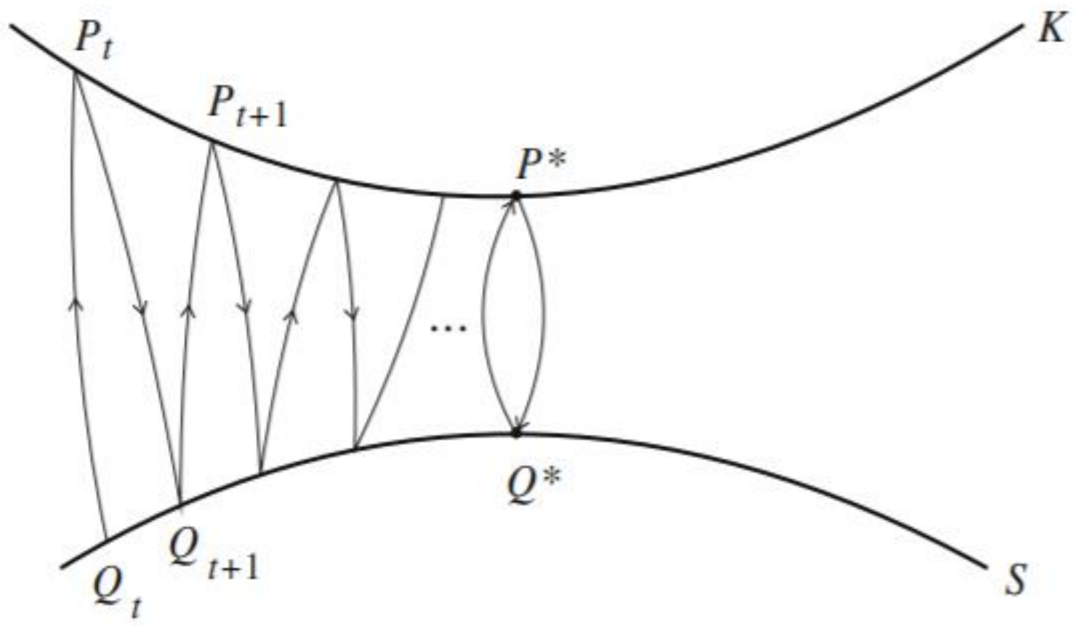
$$p(\boldsymbol{x};\boldsymbol{\theta})d\mu(\boldsymbol{x}) = \exp\{\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi(\boldsymbol{\theta})\}\,d\mu(\boldsymbol{x})$$

と表される.（ここで, $\boldsymbol{x}$の測度として$d\mu(\boldsymbol{x}) = \exp\{r(x)\}\,dx$を導入した）

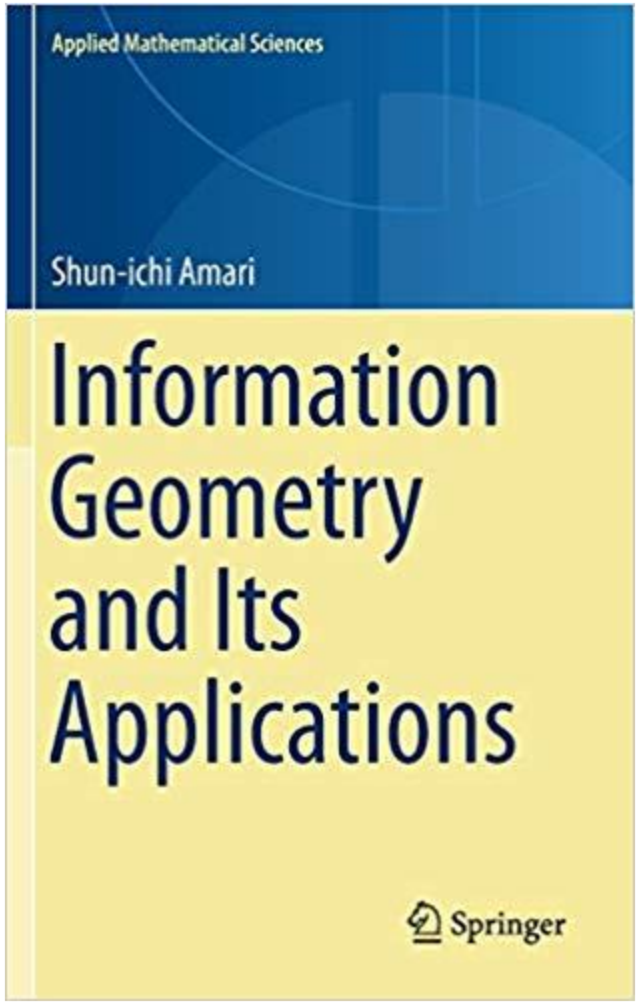
このとき, 正規化定数$\psi(\theta)$はキュムラント母関数であり, 凸関数である. ここから派生するBregmanダイバージェンスを幾何構造に持つ空間は元の指数型分布族に対応し, その双対は混合分布族を与える.

例：EMアルゴリズムの幾何学

- emアルゴリズム（≃EMアルゴリズム）



書籍紹介



- **Information Geometry and Its Applications**

平坦な場合
→不変な場合
→一般理論(双対アファイン接続)
→統計学・機械学習等への応用
と進んでいく

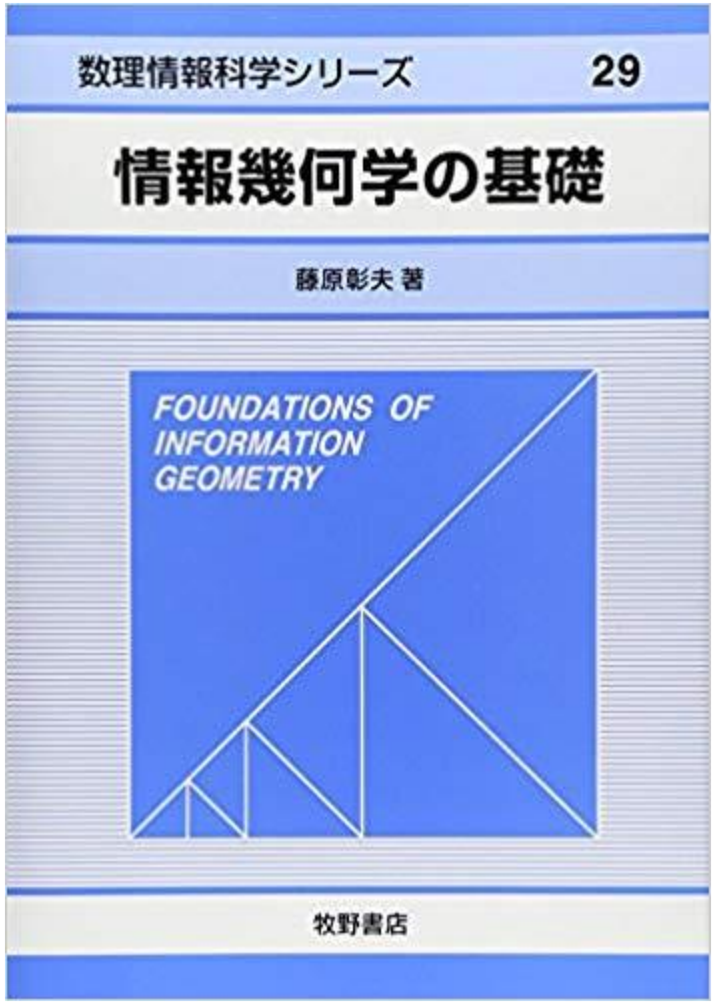
※Springer linkから無料で入手可能
※今回, 特に断りがない場合, この本から図表を拝借した

書籍紹介

• 情報幾何学の新展開

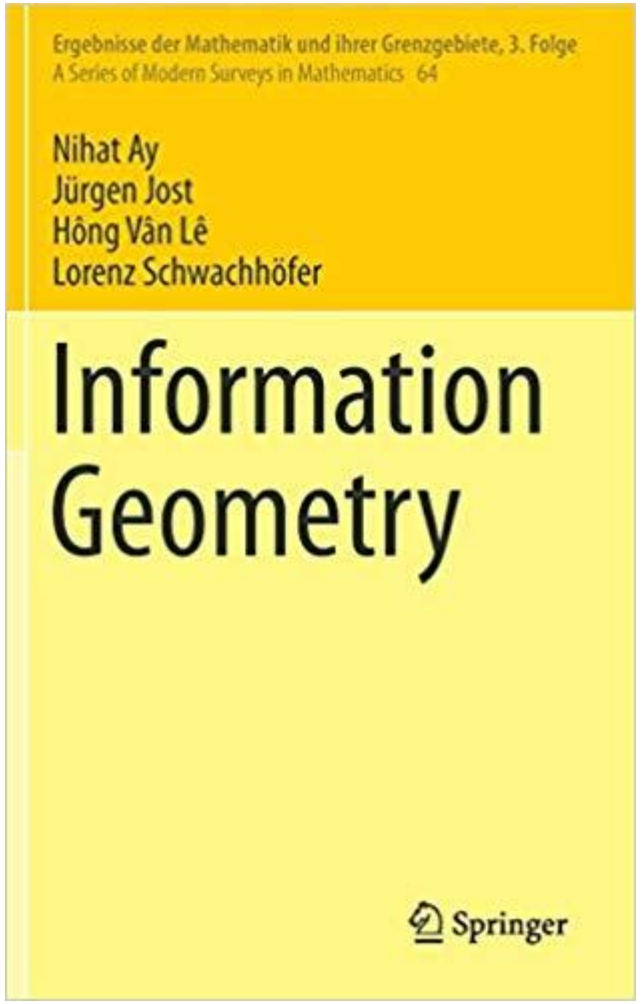
- 先ほど同様甘利先生の著書であり, 説明の仕方が多少異なる点はあれ内容は重複している部分が多い.
- 所によって式変形や説明が粗雑・冗長なことがあり, 誤植も見られる
- 近々改訂されることが決まっているらしい(『数理科学11月号』).

書籍紹介



- **情報幾何学の基礎**
  - 基礎的な微分幾何学の説明にはじまり, 核心である双対アファイン接続の幾何学を丁寧に導く.
  - 先述の新展開に挫折してこちらに挑戦し感動する人多数（？）

書籍紹介



- Information geometry
  - 読んでいない
  - 相当の数学的な素地がないときつそう（？）
  - 新しめの応用についても取り上げられている様子

参考

1. S. Amari “Information geometry and its applications”, Springer（2016）
2. 甘利俊一『情報幾何学の新展開』サイエンス社（2014）
3. 藤原彰夫『情報幾何学の基礎』牧野書店（2015）
4. 小林昭七『曲線と曲面の微分幾何』裳華房（1977）
5. F. Nielsen “An elementary introduction to information geometry”(2018)
   - https://project.inria.fr/gudhi/files/2018/10/An-elementary-introduction-to-information-geometry.pdf
6. 長岡浩司「情報幾何の基礎概念」(2006)
   - http://www.sci.osaka-cu.ac.jp/~ohnita/2006/inf_geom/sos_dan/01_nagaoka_0403.pdf
7. 赤穂昭太郎「情報幾何と機械学習」計測と制御 (2005)
   - https://staff.aist.go.jp/s.akaho/papers/infogeo-sice.pdf
8. 赤穂昭太郎「情報幾何で見る機械学習」（2015?）
   - https://www.airc.aist.go.jp/seminar_detail/docs/seminar02-akaho.pdf
9. 田中利幸「確率モデルに基づく推論の情報幾何」(2006)
   - http://www.sci.osaka-cu.ac.jp/~ohnita/2006/inf_geom/sos_dan/T_Tanaka_2.pdf
10. 甘利俊一「講座・情報幾何とその応用」Ⅰ～IX
    - 一覧（Ⅰ～IX）
11. S. Eguchi ”Information Geometry”(2013)
    - https://www.ism.ac.jp/shikoin/summer-school/2013/2013SummerSchool.pdf