情報幾何で見る機械学習

赤穂昭太郎

産業技術総合研究所
人間情報研究部門　情報数理研究グループ
（兼：人工知能研究センター機械学習研究チーム）

目次

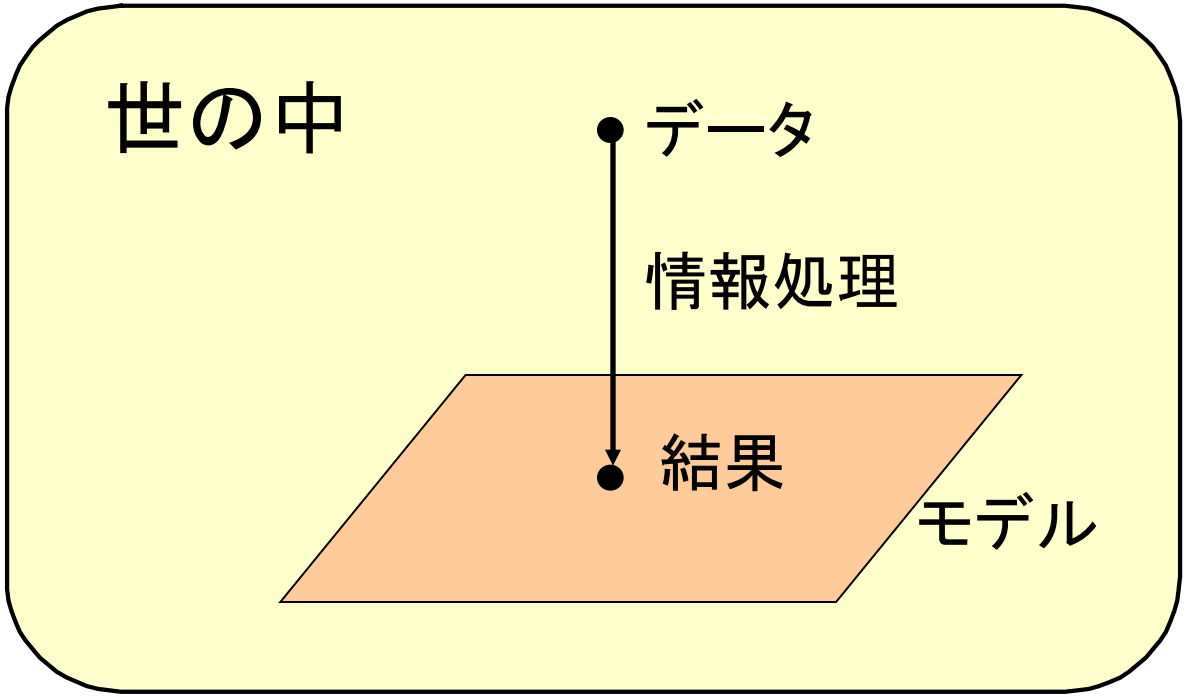
- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族 e と m
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

目次

- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族　ｅとｍ
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

情報幾何

情報処理を幾何的に（図で）理解する



情報幾何：曲がったものをまっすぐに

- 情報処理の空間は曲がった空間だが曲がった空間のままだと扱いが面倒
  →情報幾何を使うと平らな空間として扱える

- 多くのモデルは「平ら」である
- 多くのアルゴリズムは平らなモデルに「まっすぐ」射影を下ろしたものになっている

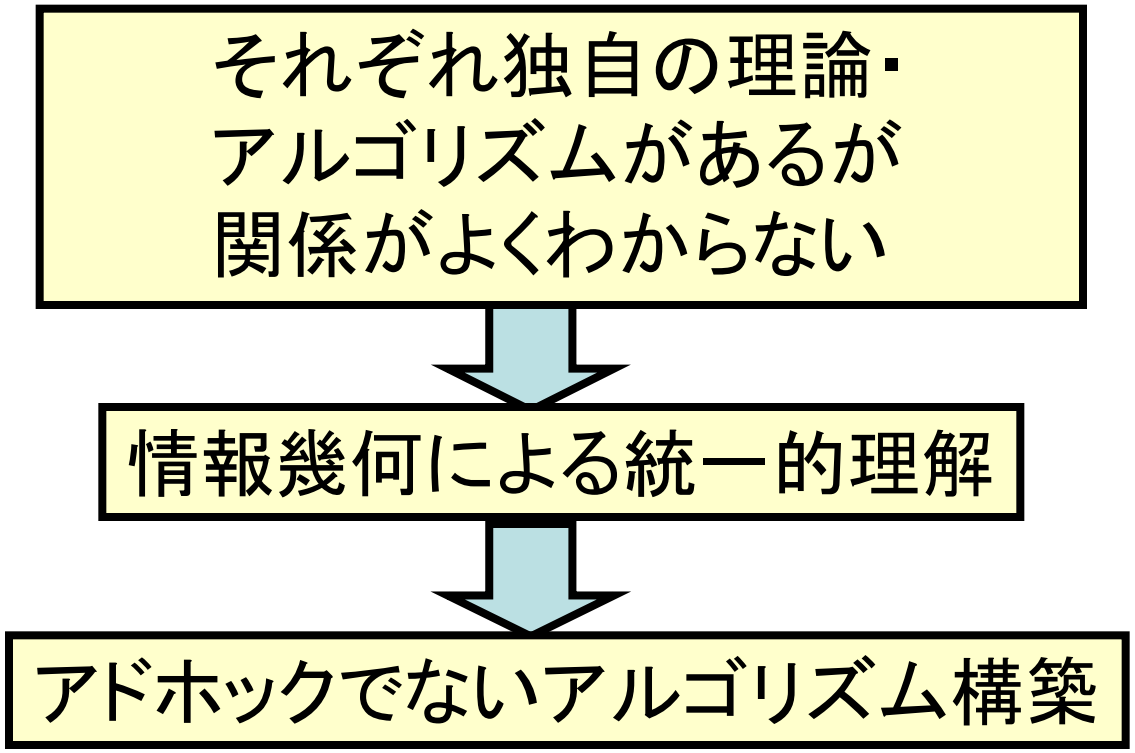
- ただし，「平ら」「まっすぐ」は普通と違って2種類ある（e と m：双対構造）

目次

- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族 e と m
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

情報幾何：異分野をつなぐ共通言語

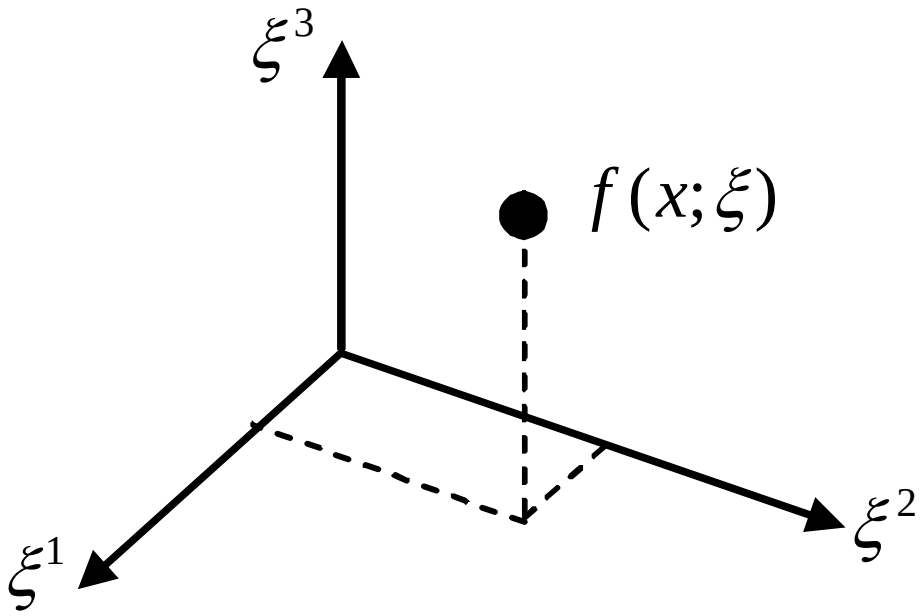
- 確率モデルやその周辺分野
  （そのほとんどに機械学習が関連）
  - 統計学
  - システム制御
  - 符号理論
  - 最適化理論
  - 統計物理



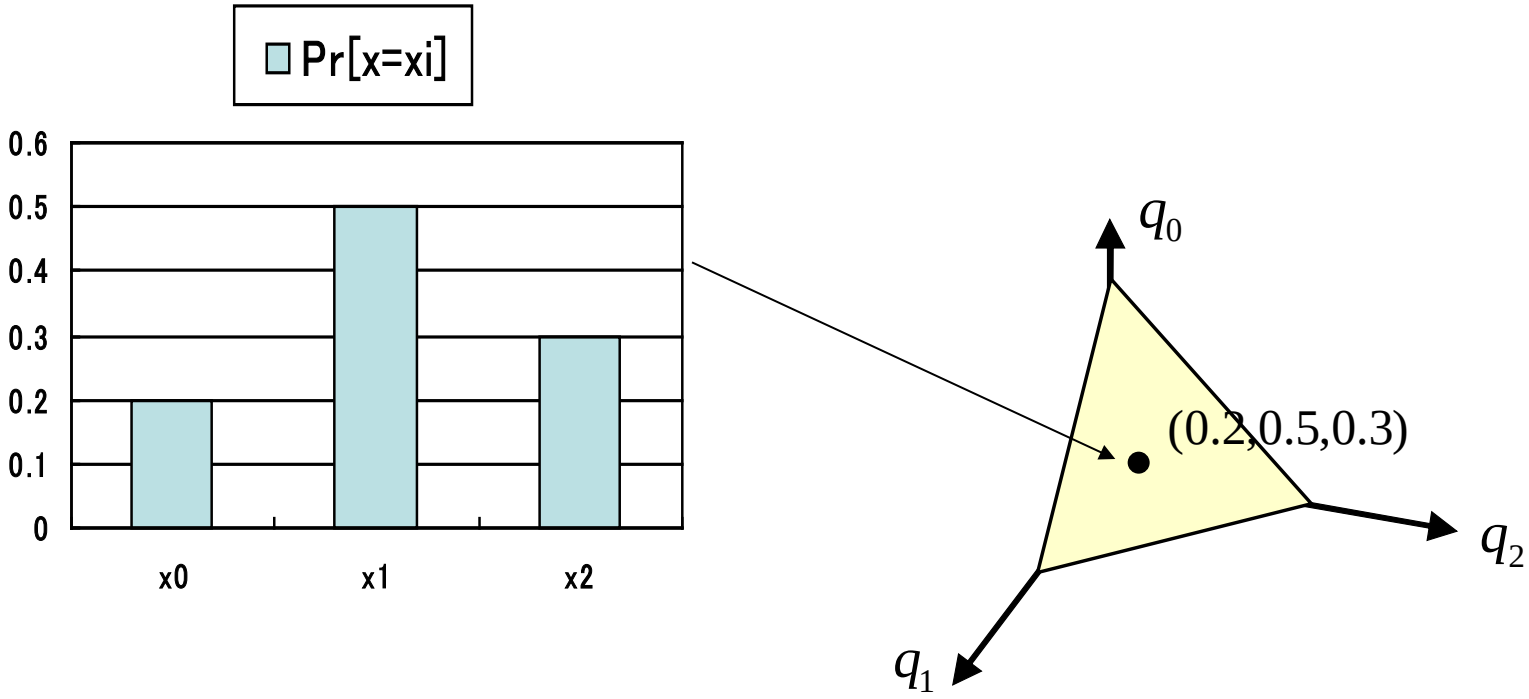
世の中＝確率モデル

- 情報幾何の出発点：
  確率モデル　$f(x;\xi)$　$\xi=(\xi^1,\xi^2,\cdots,\xi^n)$

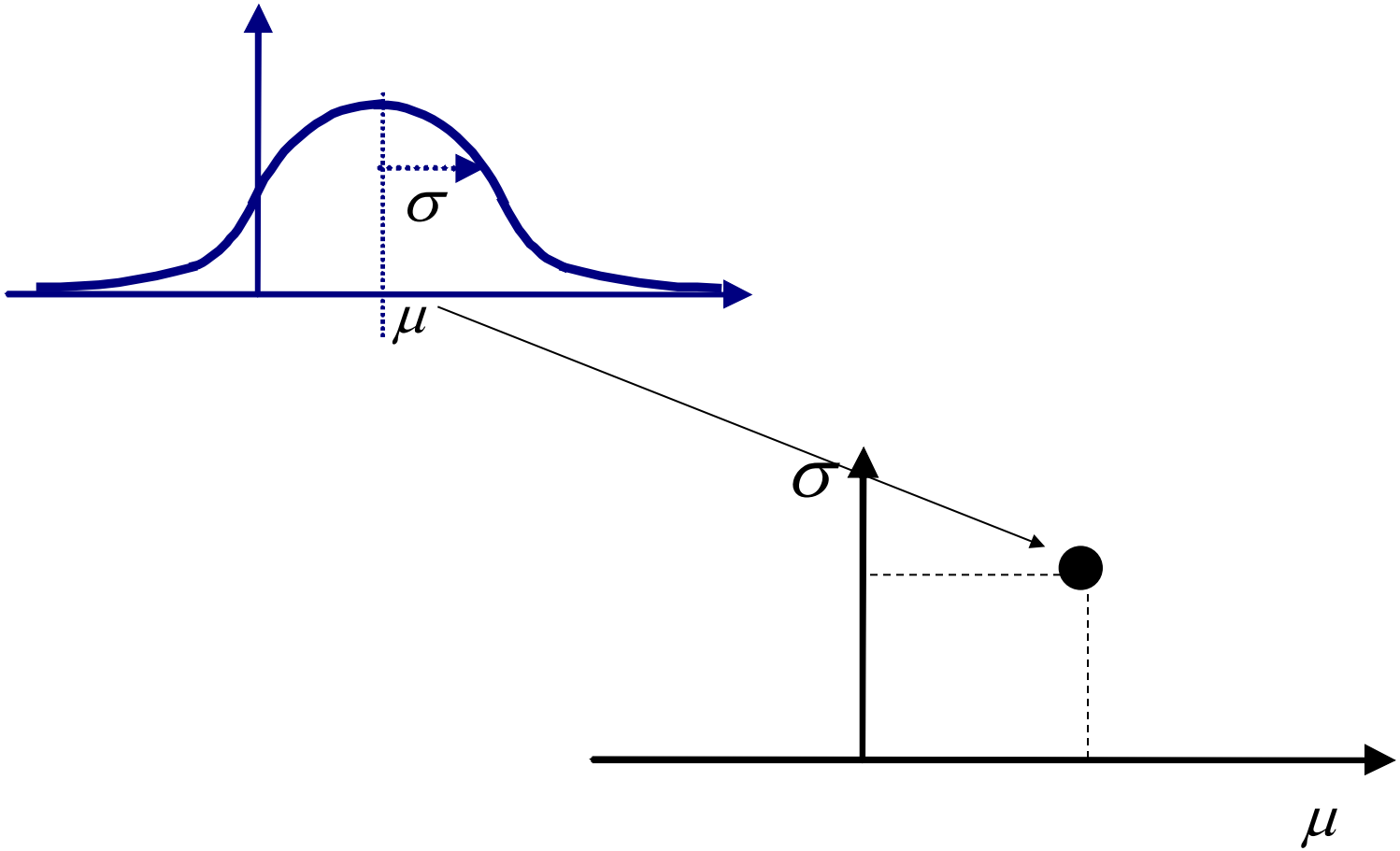
- 座標系



例1：離散分布

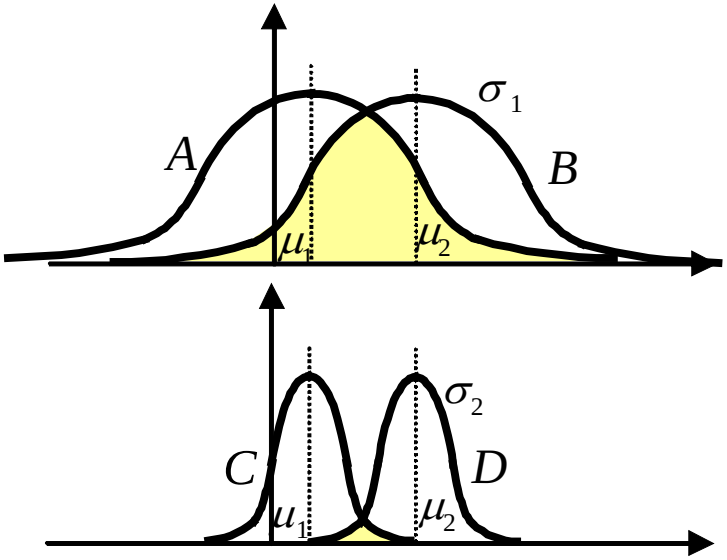


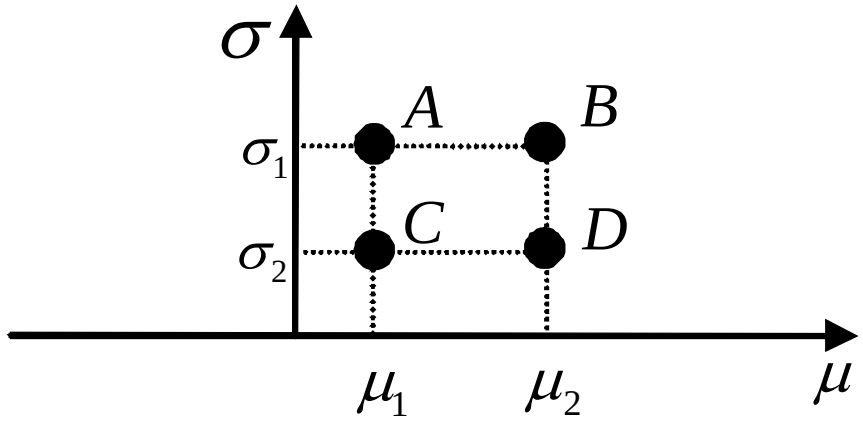
例2：正規分布



空間の構造

- ユークリッド空間ではダメ？

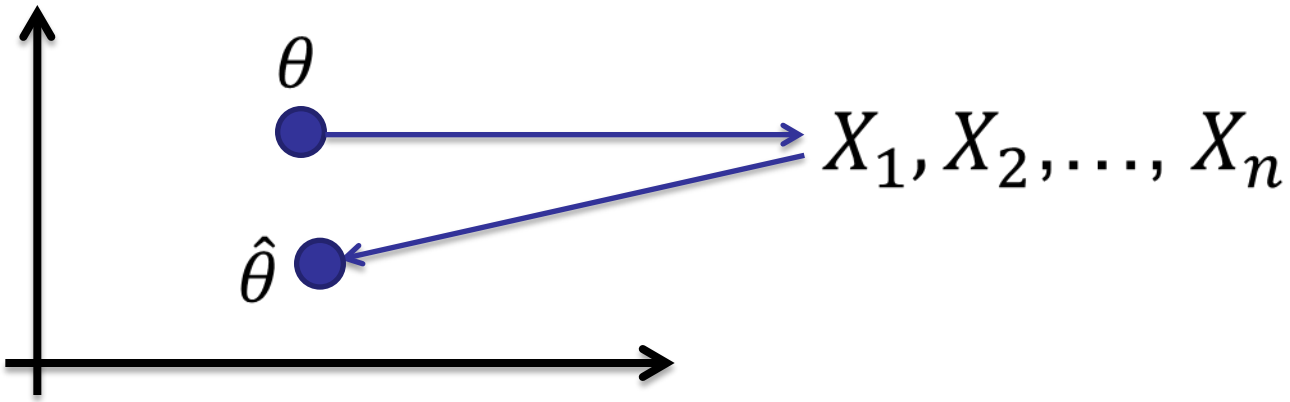




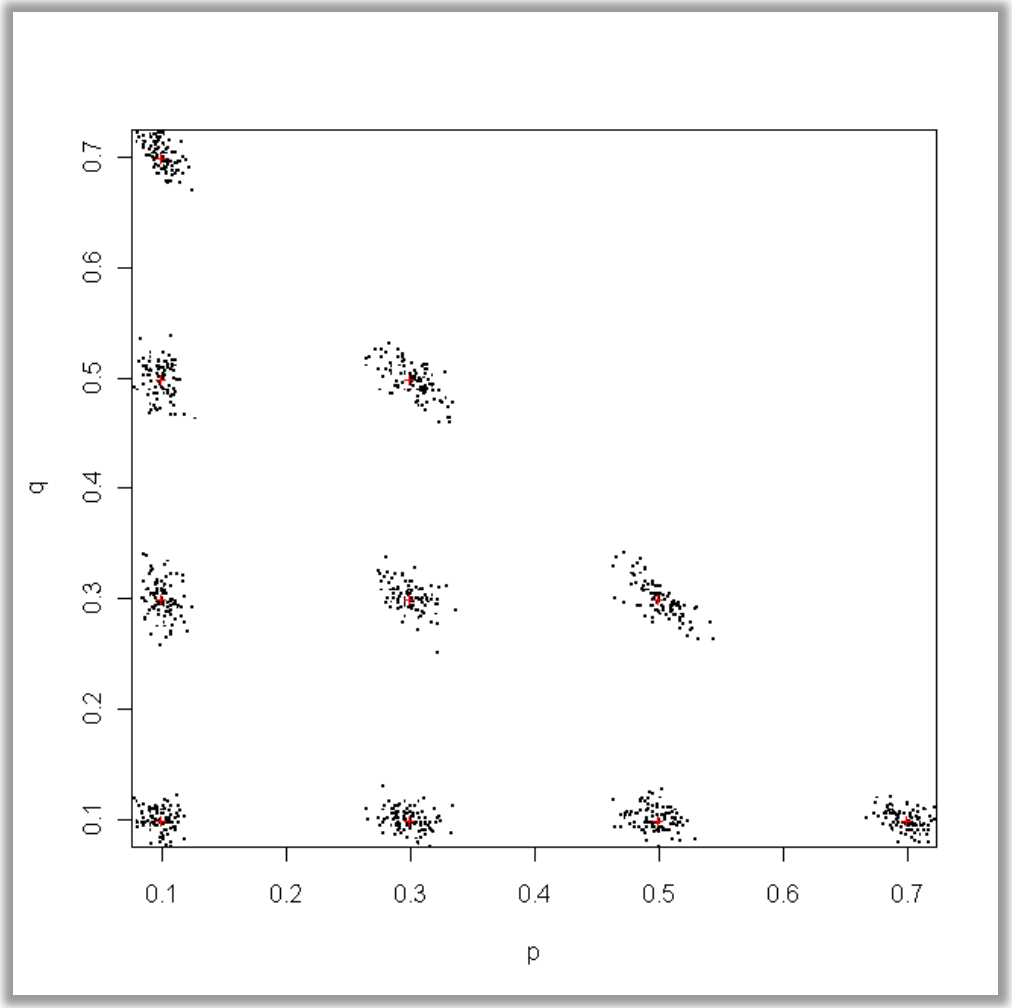
- ユークリッドではA-B と C-D の隔たりが同じになる

単純な実験

- Step 1: 確率分布$P(X;\theta)$のパラメータθを固定
- Step 2: $P(X;\theta)$に従うサンプルをたくさん生成
- Step 3: そのサンプルからθを推定
- Step 4: Step 2, 3 を繰り返す

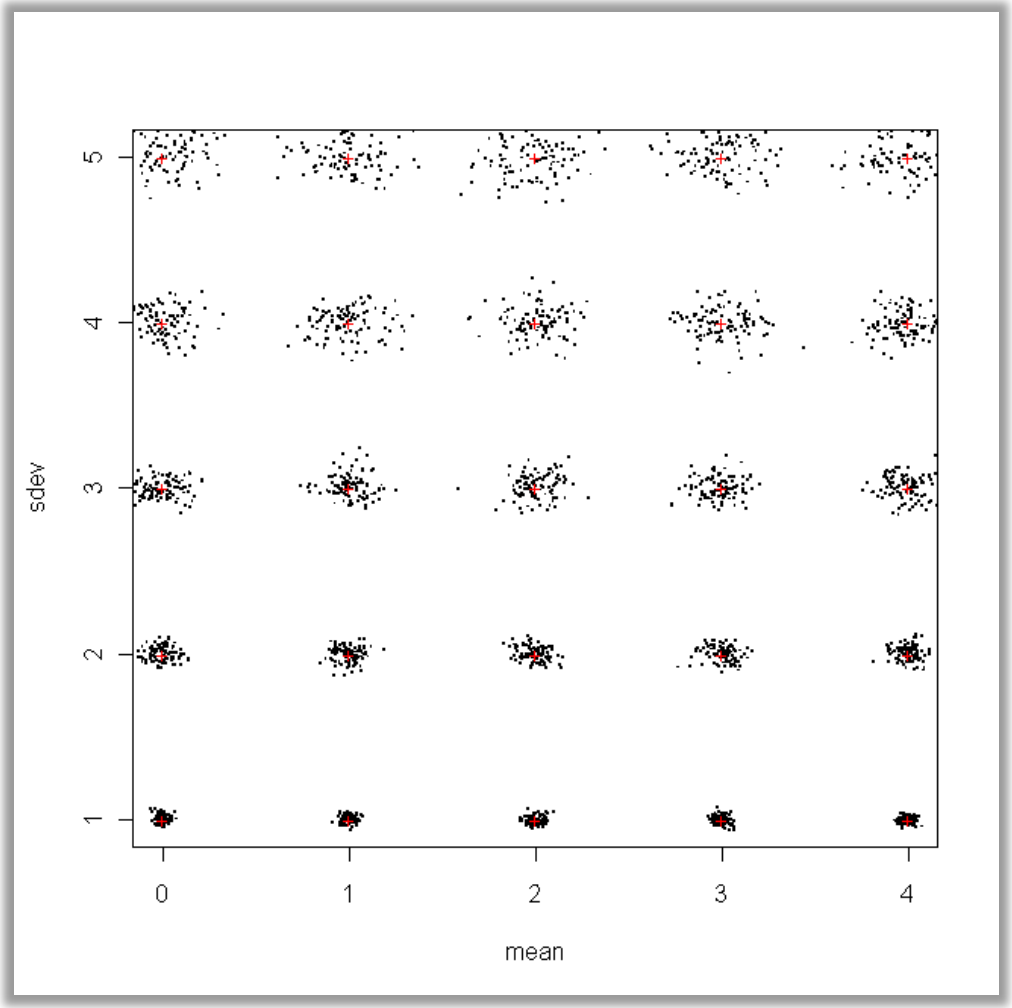


離散分布



- 3値の離散分布
  (2つの独立な
  パラメータ)
- パラメータの場所に
  よって分布が異なる

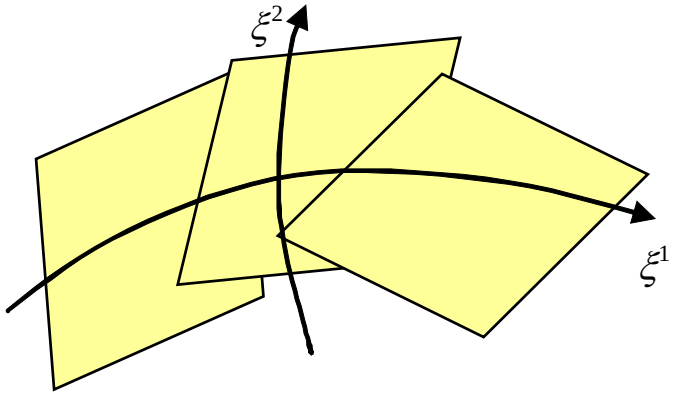
正規分布



- 平均-標準偏差
- 標準偏差が大きいところではパラメータ推定の分布がばらつく

空間の構造を決める

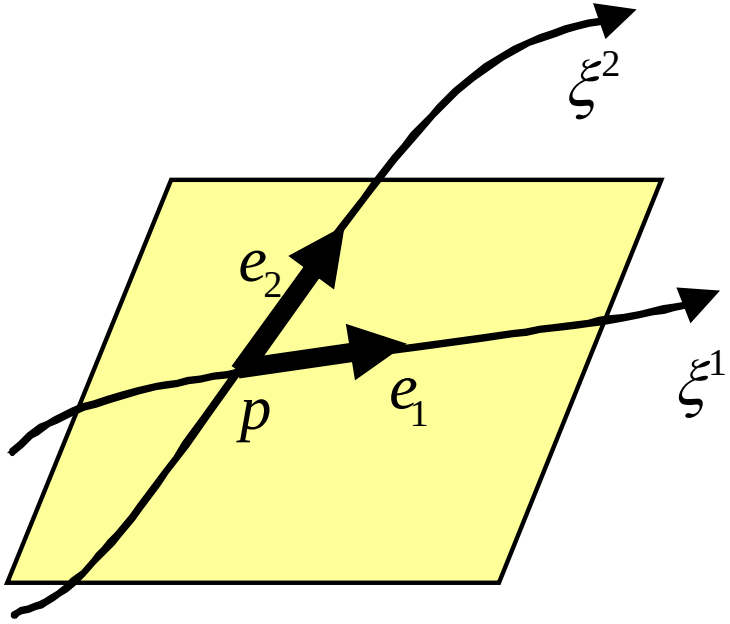
- 空間の構造は何で決まるか？
  - 点の近く: 線形空間（**計量**）
  - 空間全体: 線形空間のつながり方を決める（**接続**）



- 設計方針
  - 統計的に自然なもの
  - パラメータの取り方によらない
  - （結果的に）平さ・まっすぐさ

点の近くの構造：線形空間

- 線形空間（**接空間**）



- 接空間の構造は
  基底の間の内積で
  決まる（**リーマン計量**）

$$g_{ij} = \left\langle e_i, e_j \right\rangle_\xi$$

情報幾何での計量

- **統計的不変性⇒フィッシャー情報行列**

$$g_{ij}(\xi) = \mathrm{E}_{\xi}[\partial_i \log p(x,\xi) \partial_j \log p(x,\xi)]$$

$$\partial_i = \frac{\partial}{\partial \xi_i}$$

$$\mathrm{E}_{\xi}[f(x)] = \int f(x) p(x;\xi) dx$$

なぜフィッシャー情報量か？

- **クラメール･ラオの不等式**
  N個のサンプルからの ξ の推定量 $\hat{\xi}$ の分散の下限

$$\mathrm{Var}[\hat{\xi}] \geq \frac{1}{N} G^{-1}(\xi)$$

- G^{-1}が ξ のまわりでの散らばり具合を表す

  ⇔ G^{-1}が大きいところはきめが粗い

例1：離散分布

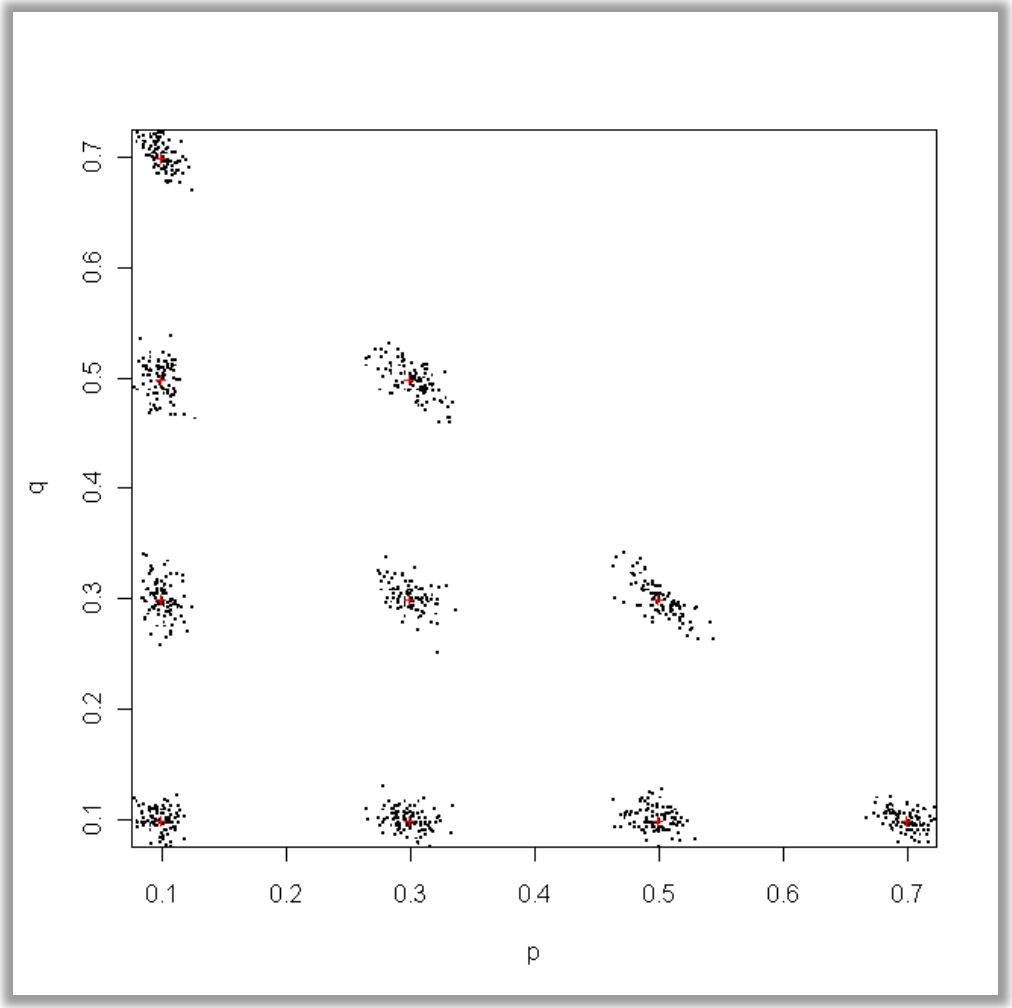
$$p(x;q_1,q_2)=q_0\delta(x)+q_1\delta(x-1)+q_2\delta(x-2)$$

$$q_0=1-q_1-q_2$$

$$G=\frac{1}{q_0}\begin{pmatrix}1+\dfrac{q_0}{q_1} & 1\\ 1 & 1+\dfrac{q_0}{q_2}\end{pmatrix}$$

- q_0,q_1,q_2 が 0 に近いところでは大きな値
  ⇒値の変化に敏感

離散分布



$$G = \frac{1}{q_0}\begin{pmatrix} 1+\dfrac{q_0}{q_1} & 1 \\ 1 & 1+\dfrac{q_0}{q_2} \end{pmatrix}$$

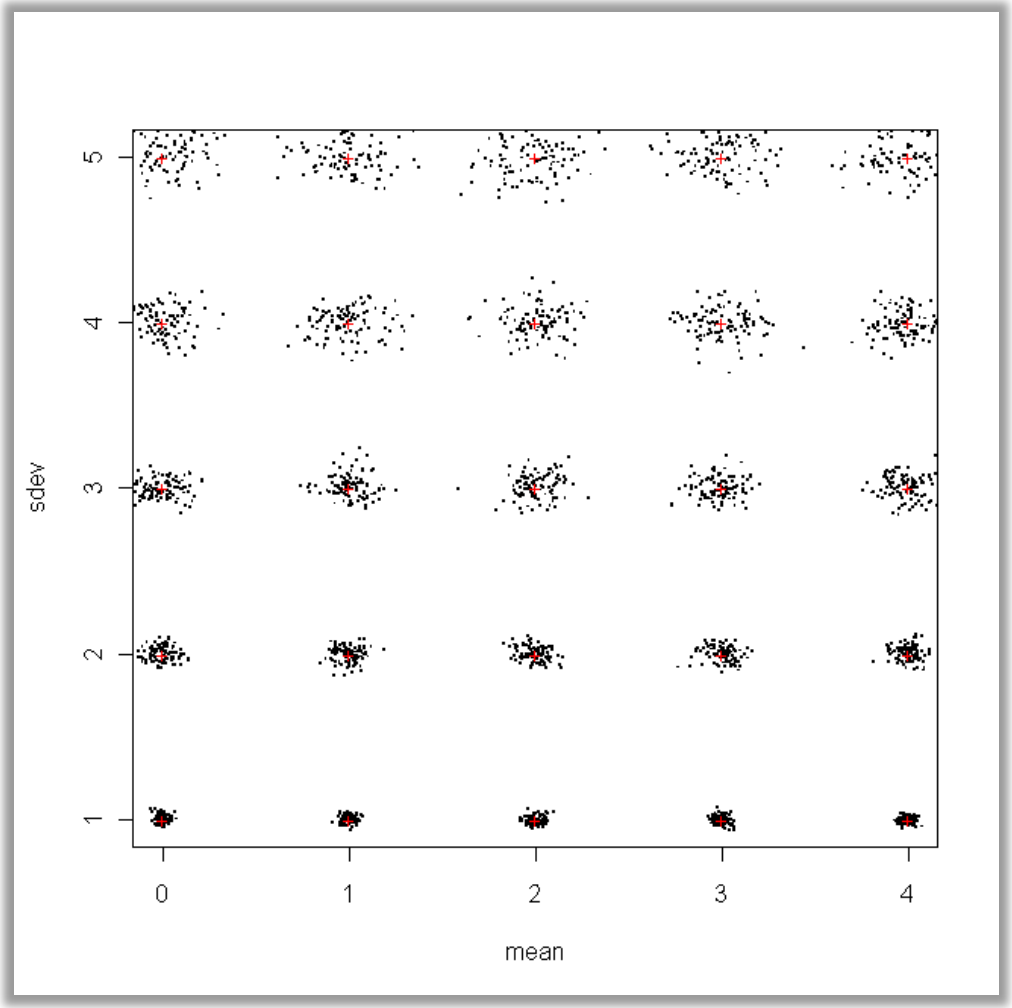
例2：正規分布

$$p(x;\mu,\sigma)=\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}-\frac{1}{2}\log 2\pi\sigma^2\right)$$

$$G=\frac{1}{\sigma^2}\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$$

- $d\mu, d\sigma$ だけ微小に動かしたときの変化は
  $(d\mu^2+2d\sigma^2)/\sigma^2$
  ⇒分散の小さいところは少し動かしただけで
  大きくずれる

正規分布



$$G = \frac{1}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

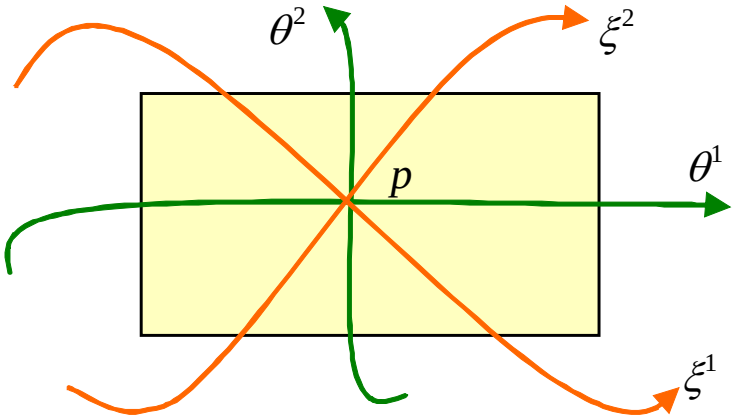
計量と座標変換

- 計量は(一般に非線形な)座標変換に対して線形に変換される(テンソル)

$$\xi = \left(\xi^i\right) \mapsto \theta = \left(\theta^a\right)$$

$$g_{ij} = \sum_{a,b} J_i^a J_j^b g_{ab}$$

$$J_i^a = \frac{\partial \theta^a}{\partial \xi^i}$$



目次

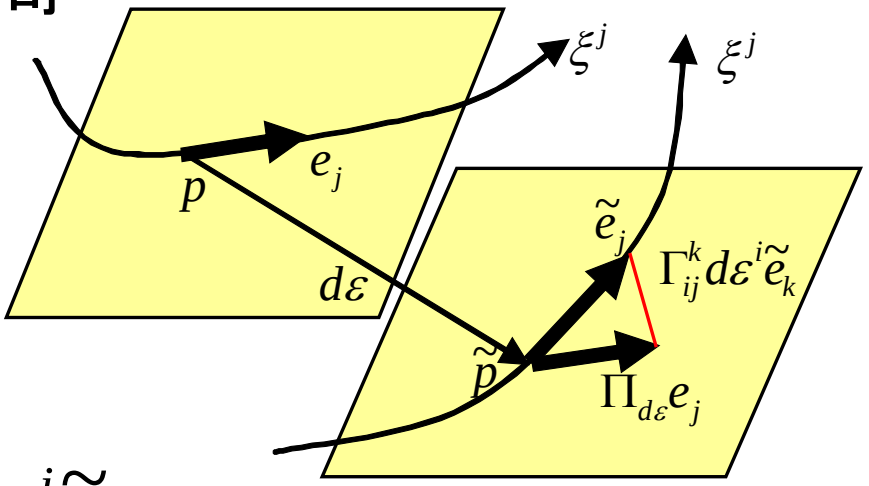
- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族 e と m
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

ユークリッド空間をつなぐ

- 各点ごとにバラバラの接空間

$$\xi(\tilde{p}) = \xi(p) + d\varepsilon$$

⇒接空間をつなぐ（接続）



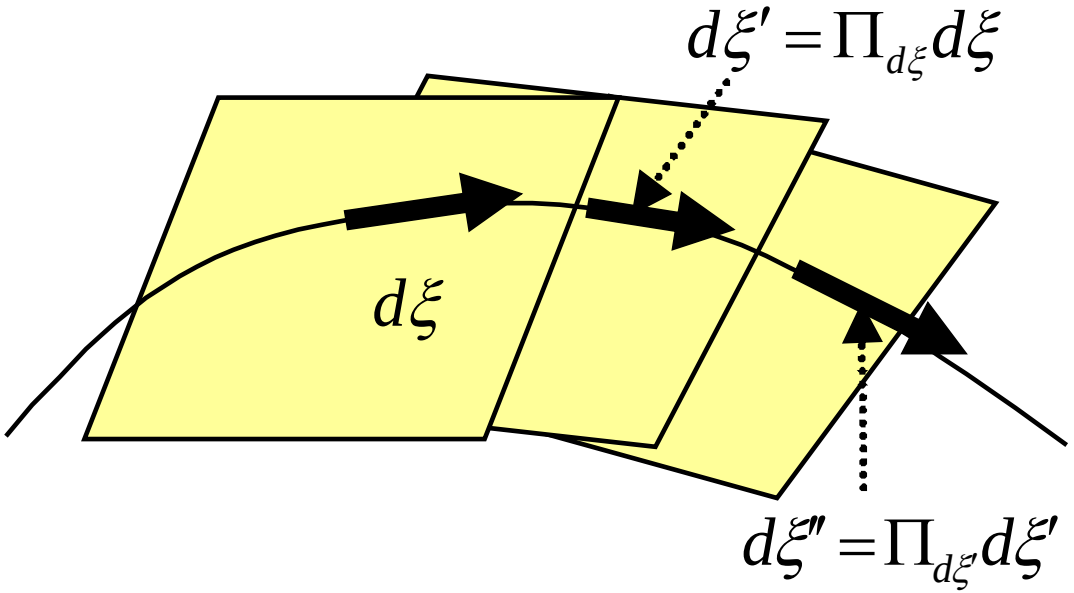
- 接ベクトルe_jの平行移動

$$\Pi_{d\varepsilon}\left[e_j\right] = \tilde{e}_j - \sum_{i,k} \Gamma_{ij}^k d\varepsilon^i \tilde{e}_k$$

- Γ_{ij}^kを（アファイン）接続係数と呼ぶ

測地線：まっすぐな線

- ある接ベクトルの方向 $d\xi$ の自分自身への平行移動 $\Pi_{d\xi}[d\xi]$ をつなげたものを**測地線**という（直線の概念の一般化）



接続をどう決めるか？

- 二つの接ベクトルを平行移動したとき，
  普通（物理等）はその間の内積を保存したい

$$\left\langle \Pi_{d\varepsilon}\left[d\xi_1\right], \Pi_{d\varepsilon}\left[d\xi_2\right]\right\rangle = \left\langle d\xi_1, d\xi_2 \right\rangle$$

- これを満たす接続は計量から一意的に決まってしまう⇒レビ・チビタ接続
  （ふつうの数学・物理ではこれで十分）
- ところが情報幾何ではそれ以外の接続も考える
  （平さ・まっすぐさと関係）

α接続

- 統計的な不変性⇒パラメータαをもつ接続係数に限られる

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

$$\partial_i l = \frac{\partial}{\partial \xi_i}\log p(x;\xi) \qquad \Gamma_{ij,k} = \sum_h \Gamma_{ij}^h g_{hk}$$

- 特にα＝0のときがレビ・チビタ接続
- 情報幾何ではα＝±1のときが最重要！

平坦な空間

- 接続はテンソルではない（座標系に依存）
- 逆に言えば，うまく座標系を取れば，$\Gamma=0$に
  できることがある (平坦な空間)
- このような座標系がもし存在するとき
  αアファイン座標系といい，その座標系に
  ついてα平坦であるという.
- 平坦な座標系の測地線（α測地線）はαアファイン座標系での直線になっている.

$$\xi = (1-t)\xi_0 + t\xi_1$$

目次

- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族 e と m
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

重要な分布族

- α＝±１ は特別な意味がある:

- 確率分布の分布族で, α平坦になるのは
  「指数分布族(exponential family)」
  と
  「混合分布族(mixture family)」
  の二つだけ！
- それぞれα＝±1に対応する

指数分布族

- 情報幾何で最も基本的な分布族

$$p(x;\theta) = \exp\left(\sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta) + C(x) \right)$$

- 指数分布族は θ をアファイン座標系として1-平坦
- 指数分布族は特別なので1-平坦や1-接続のことをe-平坦とかe-接続という(e=exponential)

混合分布族

- 確率分布の線形和

$$p(x;\theta) = \sum_{i=1}^{n} \theta^i F_i(x) + \theta^0 F_0(x)$$

$$\theta^0 = 1 - \sum_{i=1}^{n} \theta^i$$

- パラメータθをアファイン座標系として
  －1平坦
- 混合分布族は特別なので－1平坦，－1接続のことをm平坦，m接続という(m:mixture)

具体例1：離散分布は混合かつ指数

- 混合分布族としては

$$p(x;\xi)=\sum_{i=1}^{n} q_i \delta(x-i)+q_0\delta(x)$$

- 指数分布族としては

$$p(x;\xi)=\exp\left(\sum_{i=1}^{n} r_i \delta(x-i)-\psi(r)\right)$$

$$r_i=\log q_i-\log q_0 \qquad \psi(r)=-\log q_0$$

離散分布（続き）

- 混合分布族

$$\Gamma_{i,j;k} = \frac{1+\alpha}{2}\left(\frac{1}{{q_0}^2} - \frac{1}{{q_i}^2}\delta_{ijk}\right)$$

→ α＝－1で0

- 指数分布族

$$\Gamma_{i,j;k} = \frac{1-\alpha}{2} f(r)$$ という形になる

→ α＝1で0

具体例2：正規分布は指数分布族

$$p(x;\mu,\sigma)=\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}-\frac{1}{2}\log 2\pi\sigma^2\right)$$

$$p(x;\xi)=\exp\left(\sum_{i=1}^{n}\theta^i F_i(x)-\psi(\theta)+C(x)\right)$$

$$F_1(x)=x \qquad \theta^1=\mu/\sigma^2$$

$$F_2(x)=x^2 \qquad \theta^2=-1/(2\sigma^2)$$

$$\psi(\theta)=\frac{\mu^2}{2\sigma^2}+\frac{1}{2}\log 2\pi\sigma^2 \qquad C(x)=0$$

双対平坦と双対座標

- 実はα平坦なら，別の座標系が存在して
  ―α平坦 になる
- α平坦な座標系：θ，―α平坦な座標系：η
- ルジャンドル変換：ポテンシャル関数 ψ , φ

$$\psi(\theta)+\varphi(\eta)-\sum_i \theta^i \eta_i = 0$$

$$\frac{\partial \varphi(\eta)}{\partial \eta} = \theta \qquad \frac{\partial \psi(\theta)}{\partial \theta} = \eta$$

双対性

- θに対する計量: g_{ij}　ηに対する計量: g^{ij}

$$\frac{\partial \eta_i}{\partial \theta^j} = g_{ij} \qquad \frac{\partial \theta^i}{\partial \eta_j} = g^{ij}$$

- 計量が座標変換のヤコビ行列になっている
- θ座標での基底: e_i　η座標での基底: e^j

双対直交: $\left\langle e_i, e^j \right\rangle = \delta_i^j$

指数分布族の場合

- θ座標系は1平坦

$$p(x;\xi) = \exp\left(\sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta) + C(x) \right)$$

- 双対座標は $\eta_i = \mathrm{E}_\theta\left[F_i(x)\right]$
- ポテンシャルはψそのもの
- 混合分布族も双対平坦だが双対座標が単純な形で書けないので，結局指数分布族が唯一重要な分布族

例1：離散分布

（この形はすでに見た：離散分布は指数かつ混合）

- e座標系 $p(x;\xi)=\exp\left(\sum_{i=1}^{n} r_i \delta(x-i)-\psi(r)\right)$

$$\theta^i = r_i = \log q_i - \log q_0$$

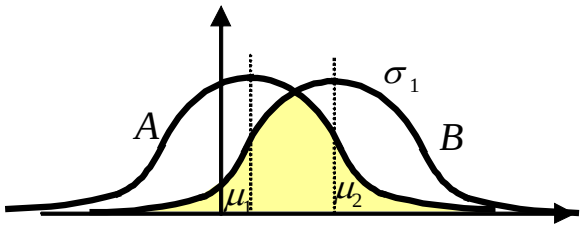
確率値の対数の線形空間

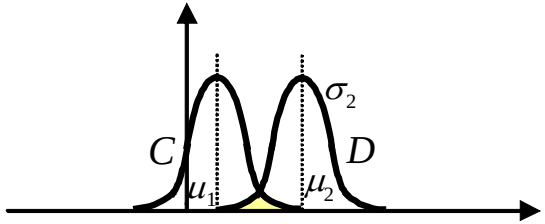
- m座標系

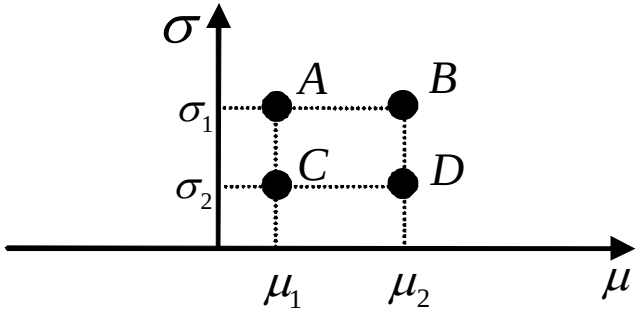
$$\eta_i = E_\theta[\delta(x-i)] = q_i$$

確率値の線形空間

例2：正規分布

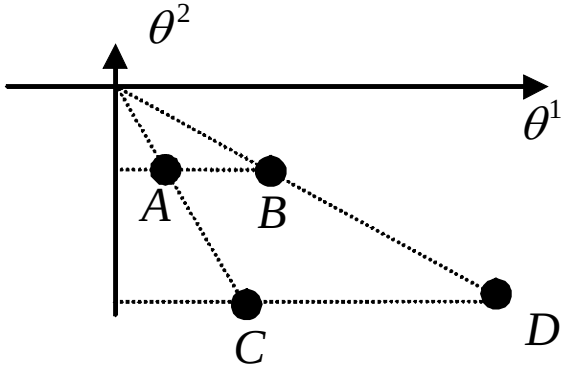






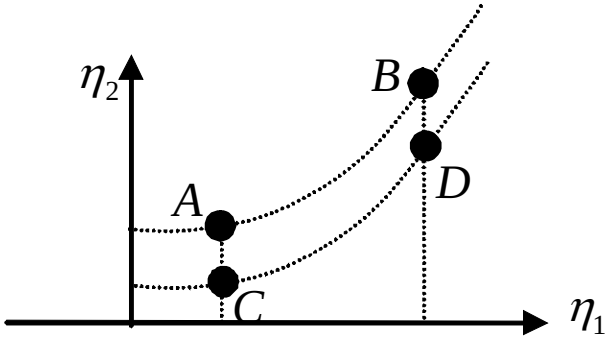
$$\theta^1 = \mu / \sigma^2$$

$$\theta^2 = -1 / (2\sigma^2)$$



$$\eta_1 = \mathrm{E}_\theta[x] = \mu$$

$$\eta_2 = \mathrm{E}_\theta[x^2] = \mu^2 + \sigma^2$$

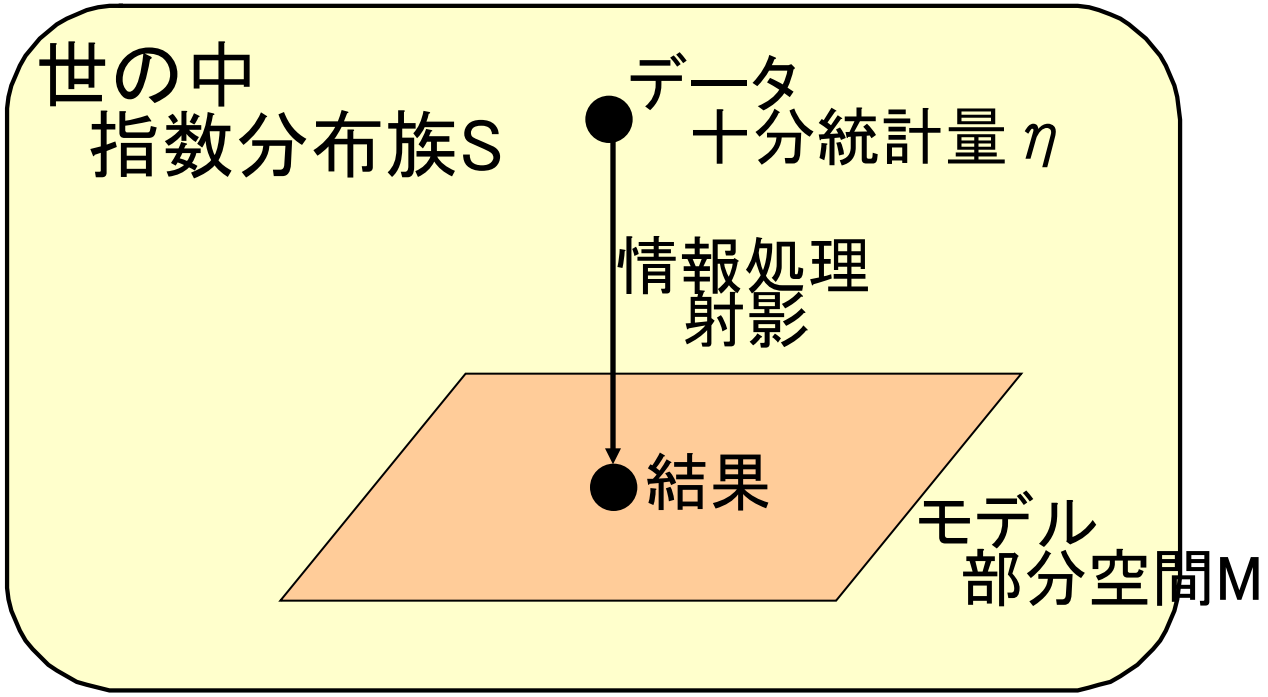


目次

- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族　ｅとｍ
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 最近の発展（IBIS2015の発表を中心に）

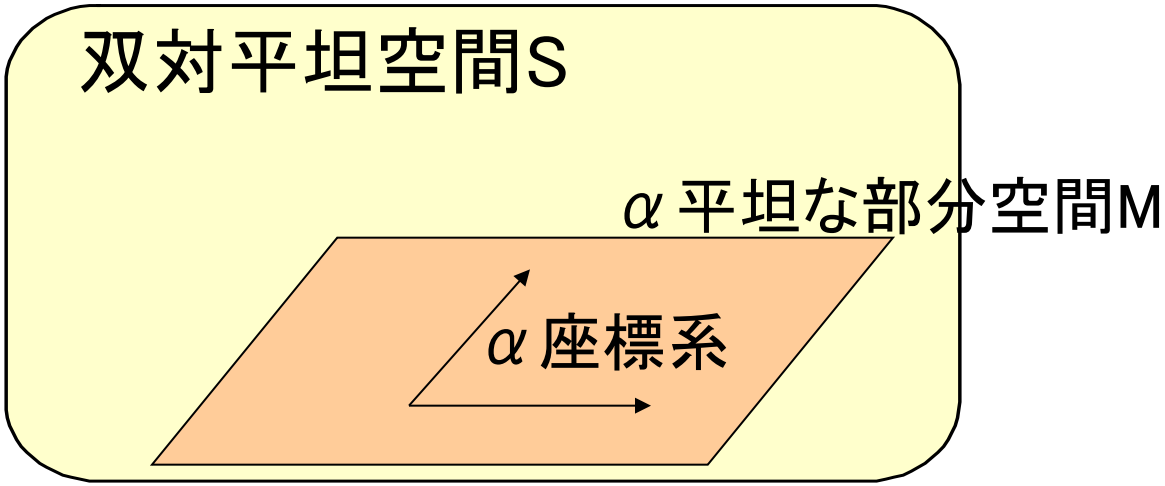
部分空間と射影

- 情報幾何的世界観



平坦な部分空間

- α平坦な線形部分空間：双対平坦な空間Sのα座標系での線形部分空間



- 注意：α平坦な部分空間は一α平坦な部分空間とは限らない　c.f. S自身はどちらも平坦

ダイバージェンス

- 射影を導入する前に. . .
- αダイバージェンス

$$D^{(\alpha)}(p\,\|\,q)=\psi(\theta(p))+\varphi(\eta(q))-\sum_i \theta^i(p)\eta_i(q)$$

c.f. ルジャンドル変換 $\psi(\theta)+\varphi(\eta)-\sum_i \theta^i\eta_i=0$

- 対称律以外は距離の性質を満たす
- p≒q なら距離に一致する
- 双対性 $D^{(\alpha)}(p\,\|\,q)=D^{(-\alpha)}(q\,\|\,p)$

指数分布族の場合

- α＝１（e接続）でのダイバージェンスは
  カルバックダイバージェンスに一致する

$$KL(f \| g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

- α＝－１（m接続）でのダイバージェンスは

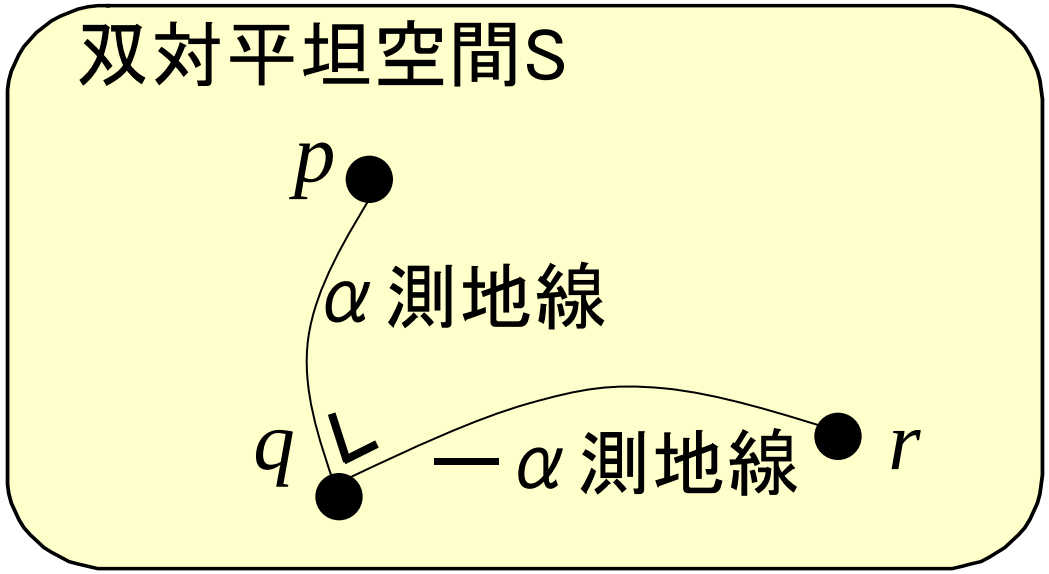
$$KL(g \| f)$$

距離の分解

- ユークリッド空間で部分空間への射影を取るのがなぜ簡単か？
- ある点から部分空間への距離が直交成分と水平成分に簡単に分解できるから（ピタゴラスの定理）

$$(x-y)^2 = (x-y^{\perp})^2 + (y-y^{\perp})^2$$

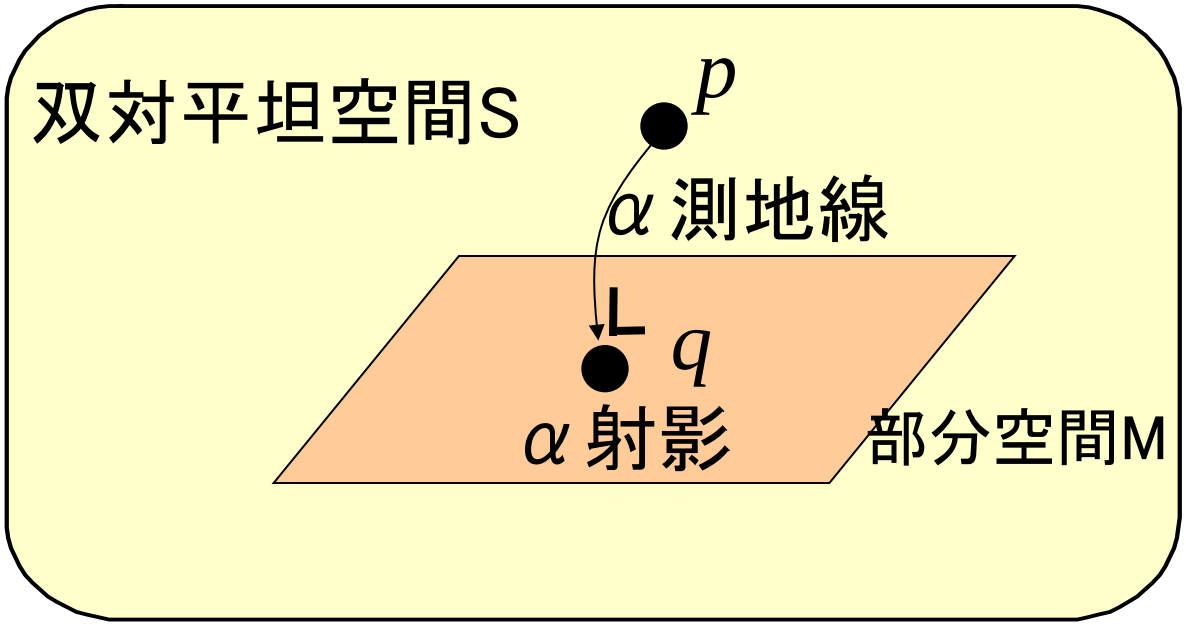
拡張ピタゴラスの定理



$$D^{(\alpha)}(p \| r) = D^{(\alpha)}(p \| q) + D^{(\alpha)}(q \| r)$$

射影定理

- α測地線で引いた直交射影は
  αダイバージェンス $D^{(\alpha)}(p\|q)$の停留点



- 特に*M*が一α平坦なら $\min_{q} D^{(\alpha)}(p\|q)$

部分空間と射影の組み合わせ

- e平坦な部分空間には m射影
- m平坦な部分空間には e射影
- ↑この組み合わせなら射影は一意的
  （それぞれのダイバージェンスの最小点）

目次

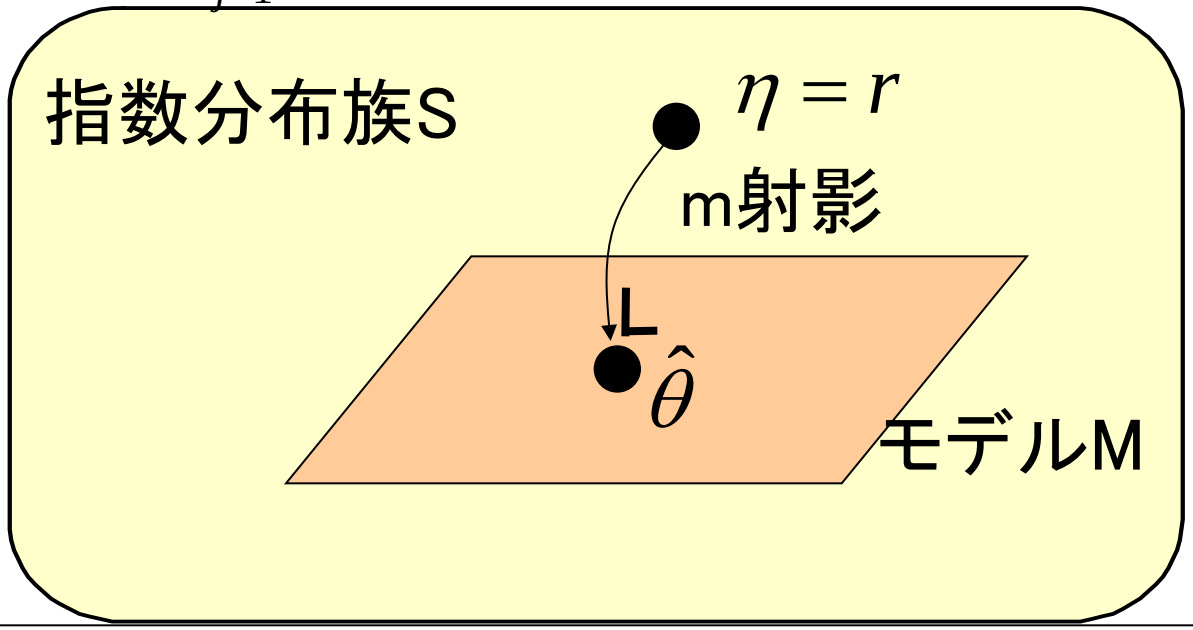
- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族　ｅとｍ
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

機械学習アルゴリズムの情報幾何的解釈

- 統計的推定
  - 最尤推定
  - 潜在変数モデルと em アルゴリズム
- 集団学習
- カーネル法
- グラフィカルモデル
  - 平均場近似
  - MCMC
- 分布パラメータの次元圧縮

統計的推定

- データは空間のどの点に配置するか？
- $\eta_i = \mathrm{E}_\theta\left[F_i(x)\right]$　なので，N個のデータの十分統計量 $r_i = \frac{1}{N}\sum_{j=1}^{N} F_i(x^{(j)})$をη座標とすればよい



統計的推定（つづき）

- 最尤推定 $\max_{\theta} p(x^{(1)} \cdots, x^{(N)}; \theta)$

$$\Leftrightarrow \max_{\theta \in M} \sum_{j=1}^{N} \log p(x^{(j)}; \theta)$$

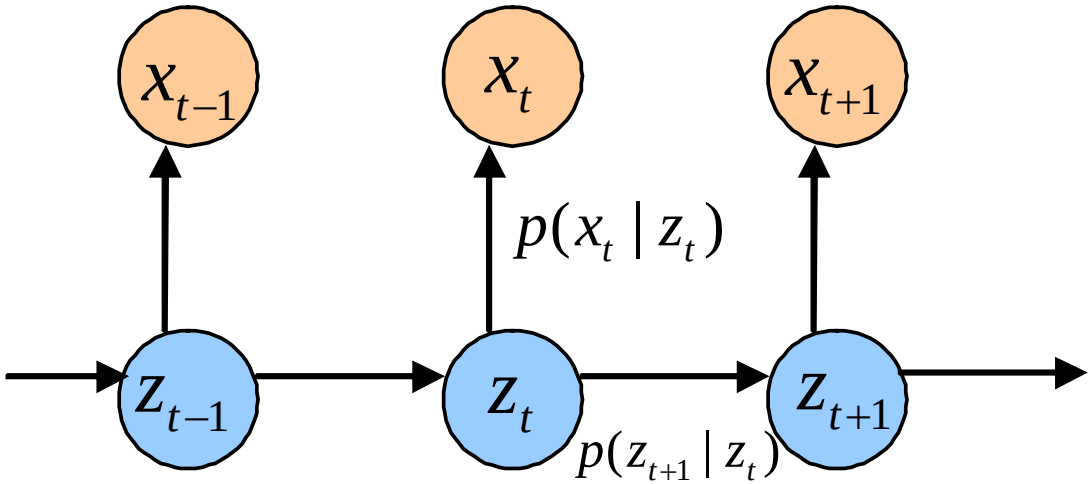
- 最尤推定はm射影と等価

$$KL(q(x) \| p(x;\theta)) = \int q(x) \log \frac{q(x)}{p(x;\theta)} dx \to \min_{\theta \in M}$$

- モデルが平らなときは推定が易しい.
  推定の質についてはモデルの曲がり具合（曲率）に関係⇒統計的漸近理論

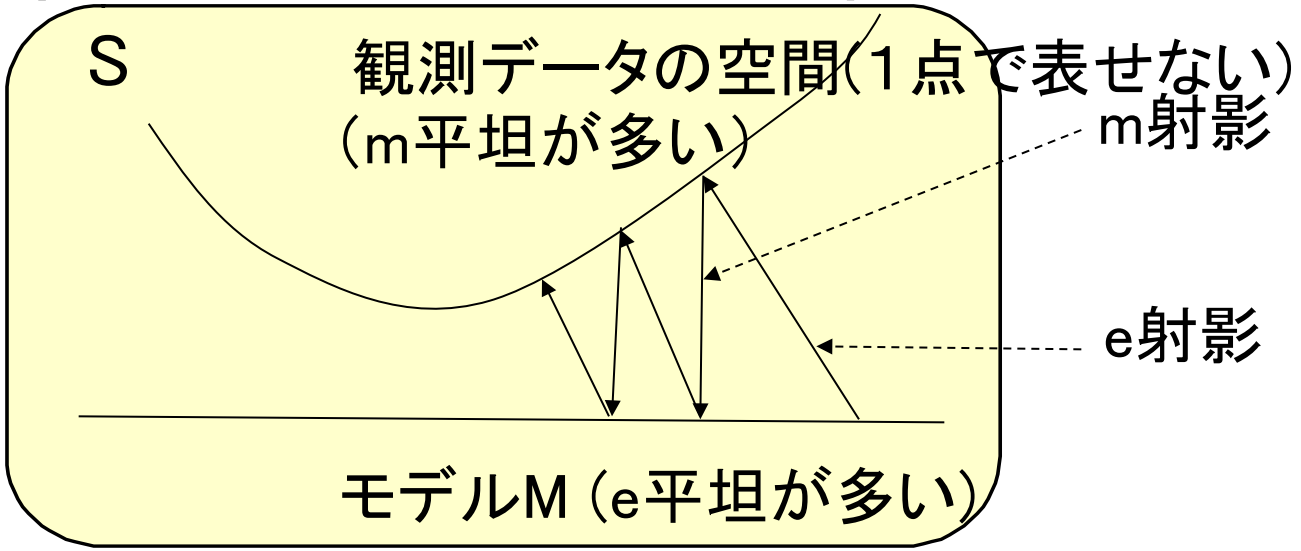
潜在変数モデル

- x だけが観測される　$p(x, z; \xi)$
  例: 隠れマルコフモデル(HMM)



em アルゴリズム

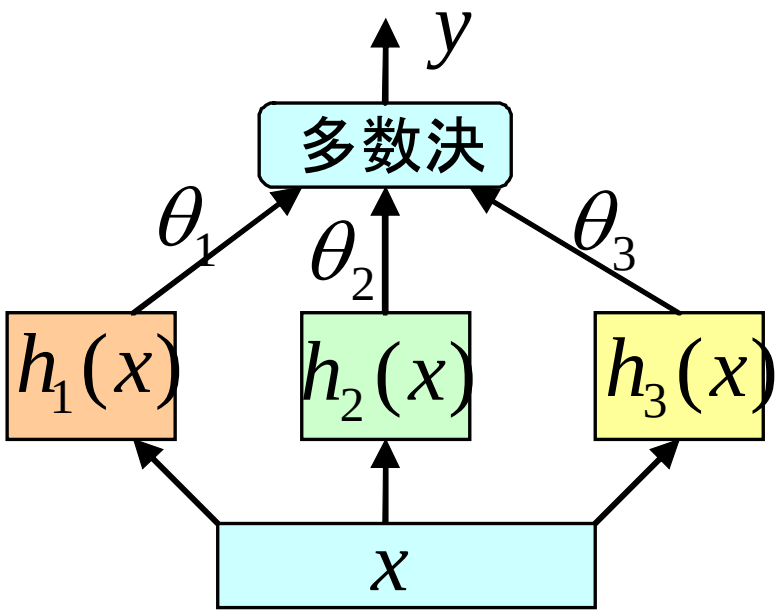
- em (exponential and mixture)



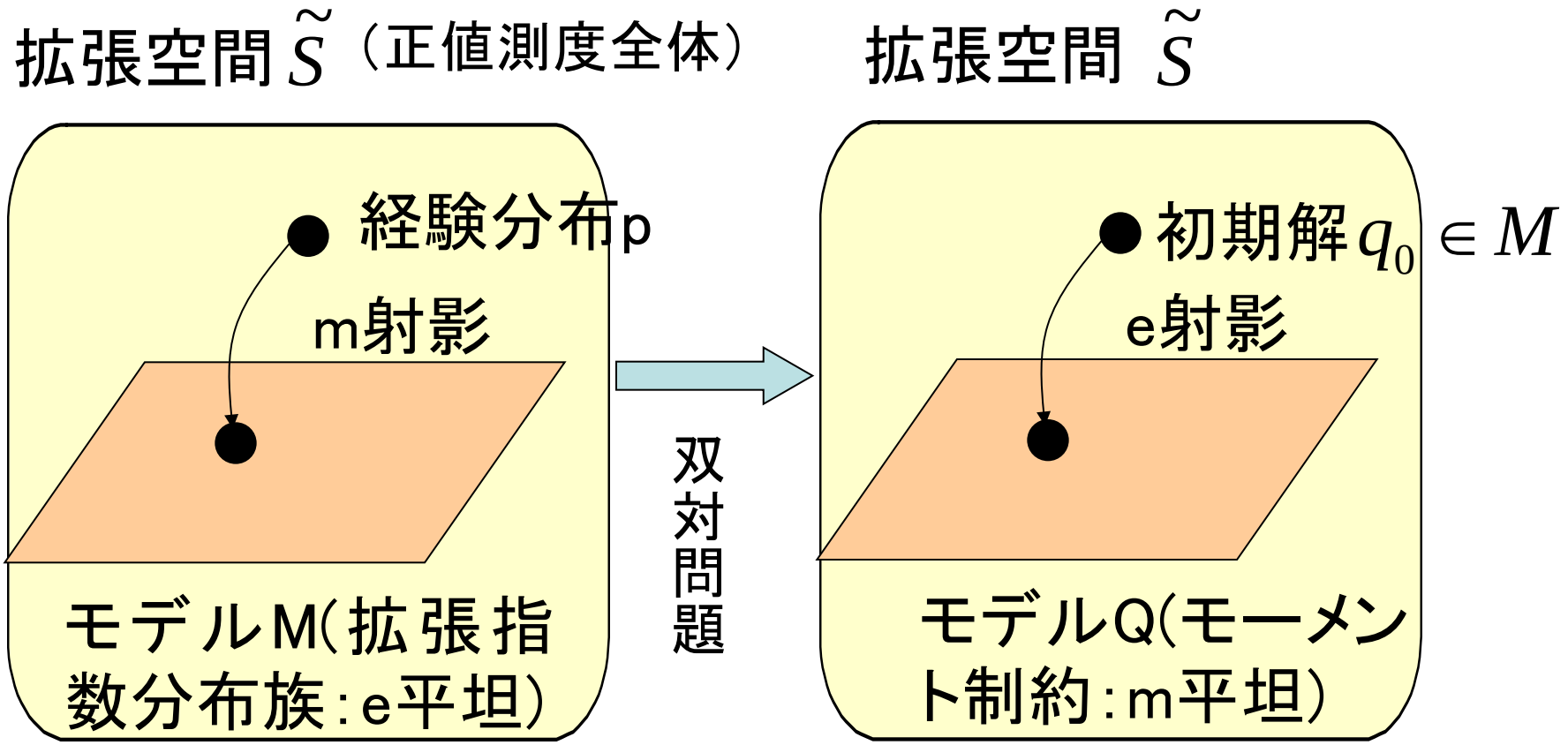
- 実はこれがEMアルゴリズム(Expectation-Maximization/Baum-Welch) とほぼ等価 (Amari 1995)

集団学習

- 三人寄れば文殊の知恵？
- バギング・ブースティング



集団学習(つづき)



ブースティングアルゴリズムの幾何的描像(Murata et al2004)

カーネルの情報幾何

- カーネル法：サポートベクトルマシンに代表されるパターン認識やデータ解析の重要なツール（赤穂：カーネル多変量解析，岩波2008参照）
- カーネル行列（正定値行列）が重要な役割を果たす
- 正規分布の分散をカーネル行列とみなす

$$p(x;\xi) = c\exp\left(-\frac{1}{2}x^T V^{-1}x - \frac{1}{2}\log\det V\right)$$

カーネルの情報幾何（つづき）

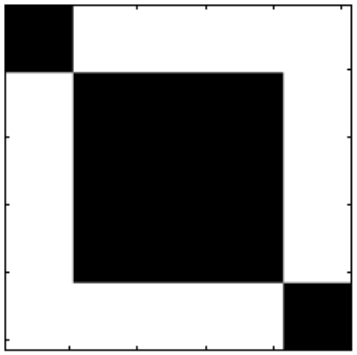
- 指数分布族 $p(x;\xi) = c\exp\left(-\frac{1}{2}x^T V^{-1}x - \frac{1}{2}\log\det V\right)$
  双対座標: $\theta = V^{-1}$ $\eta = V$

$$KL(V_1, V_2) = tr(V_1^{-1}V_2) + \log\det V_1 - \log\det V_2 - n$$

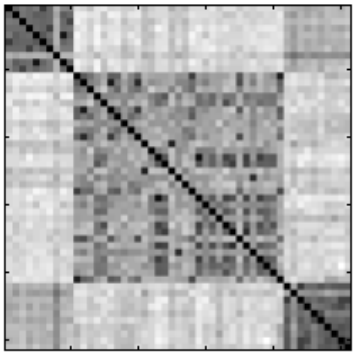
- 応用:
  - 制御系の安定性解析
  - カーネル行列の補完
  - 複数のカーネル行列の統合

カーネル行列の補間

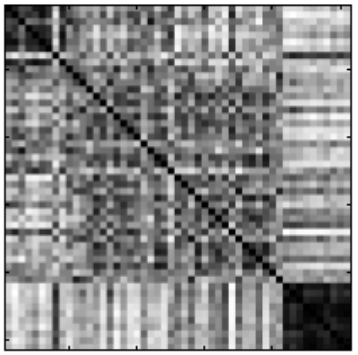
- バクテリア遺伝子の分類 (Tsuda,Akaho et al2003)



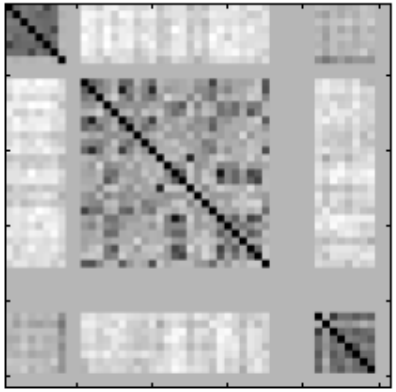
(a) Ideal



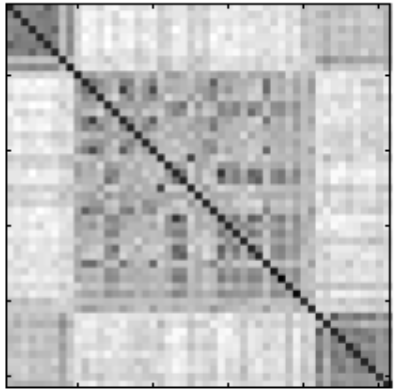
(b) gyrB (complete)



(c) 16S (complete)



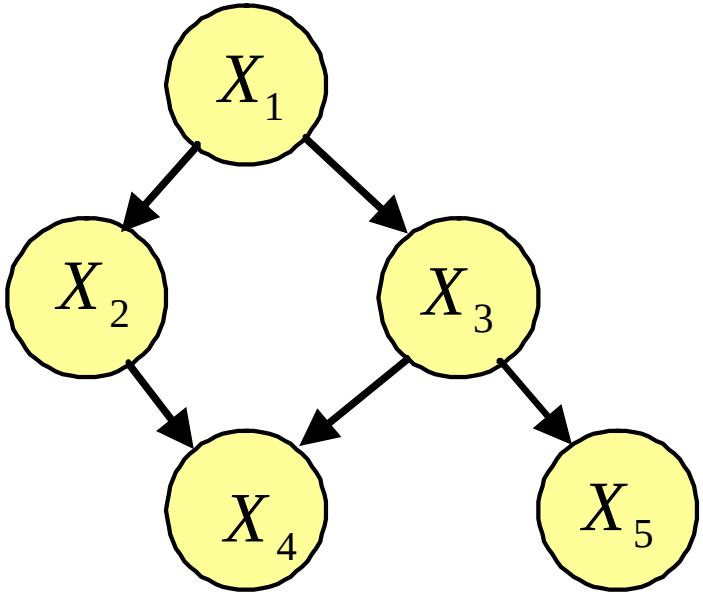
(d) gyrB (20% missing)



(e) completed D

グラフィカルモデルとベイズ推定

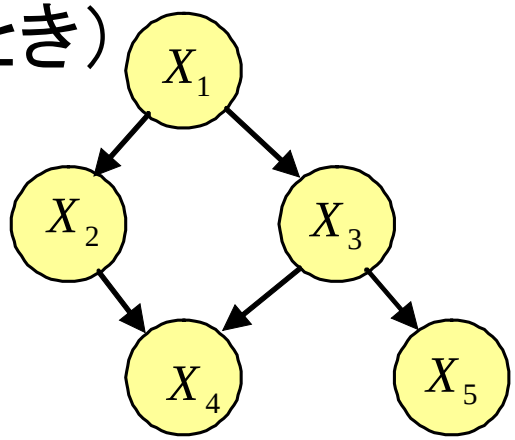
- 変数間の依存関係をグラフであらわす
- HMM, カルマンフィルタもその一種



$$p(X) = p(X_1)$$
$$p(X_2 \mid X_1)p(X_3 \mid X_1)$$
$$p(X_4 \mid X_2, X_3)p(X_5 \mid X_3)$$

ベイズ推定

- 一部が観測されたときに残りの変数を推定
  事後分布 $$p(X_1, X_2, X_3 \mid X_4, X_5) = \frac{p(X)}{p(X_4, X_5)} = \frac{p(X)}{\sum_{X_1, X_2, X_3} p(X)}$$
- ノード数が増えると総和計算
  （or 積分）が大変！（特に木でないとき）
- ⇒近似計算
  （平均場近似・変分ベイズ）
  （マルコフ連鎖モンテカルロ・
  パーティクルフィルタ）



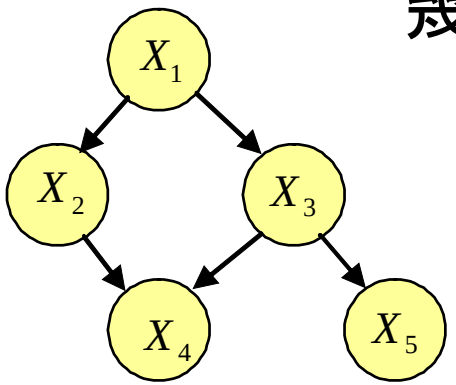
平均場近似・変分ベイズ法

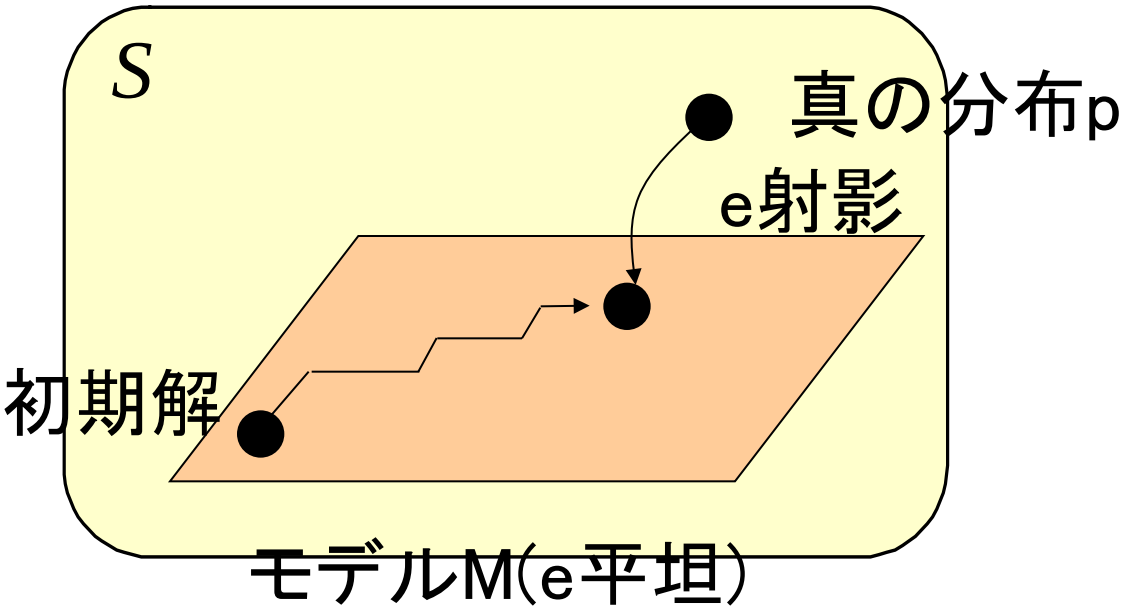
(Tanaka1999)

$$p(X_1, X_2, X_3 \mid X_4, X_5) \cong q_1(X_1) q_2(X_2) q_3(X_3)$$ モデルM(e平坦)

$$\min KL\left[q_1(X_1) q_2(X_2) q_3(X_3) \,\|\, p(X_1, X_2, X_3 \mid X_4, X_5)\right]$$ e射影

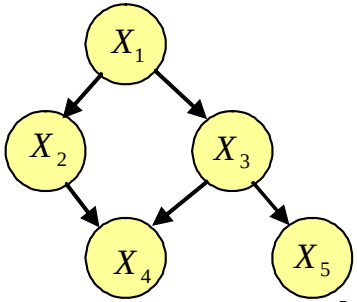
幾何的には相性の良くない組み合わせ！





マルコフ連鎖モンテカルロ

- 乱数発生により事後分布からのサンプルを生成する
- ギブスサンプラー



$$p(X_1^{(t+1)} \mid X_2^{(t)}, X_3^{(t)}; X_4, X_5)$$

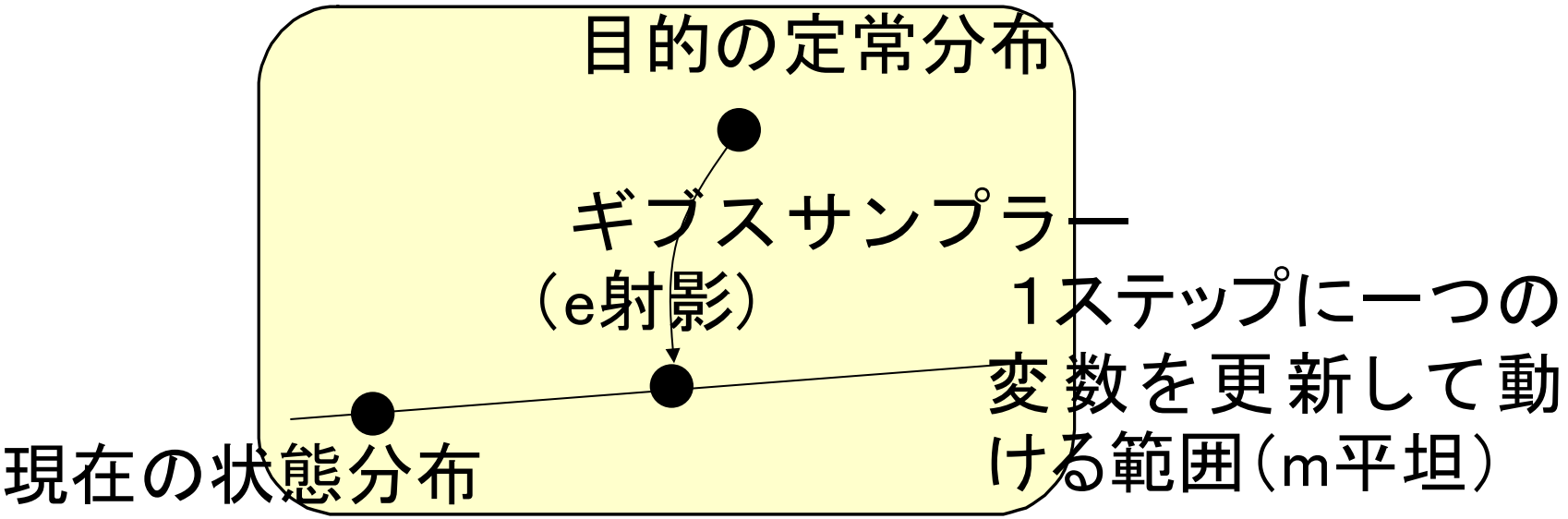
$$p(X_2^{(t+1)} \mid X_3^{(t)}, X_1^{(t+1)}; X_4, X_5)$$

$$p(X_3^{(t+1)} \mid X_1^{(t+1)}, X_2^{(t+1)}; X_4, X_5)$$

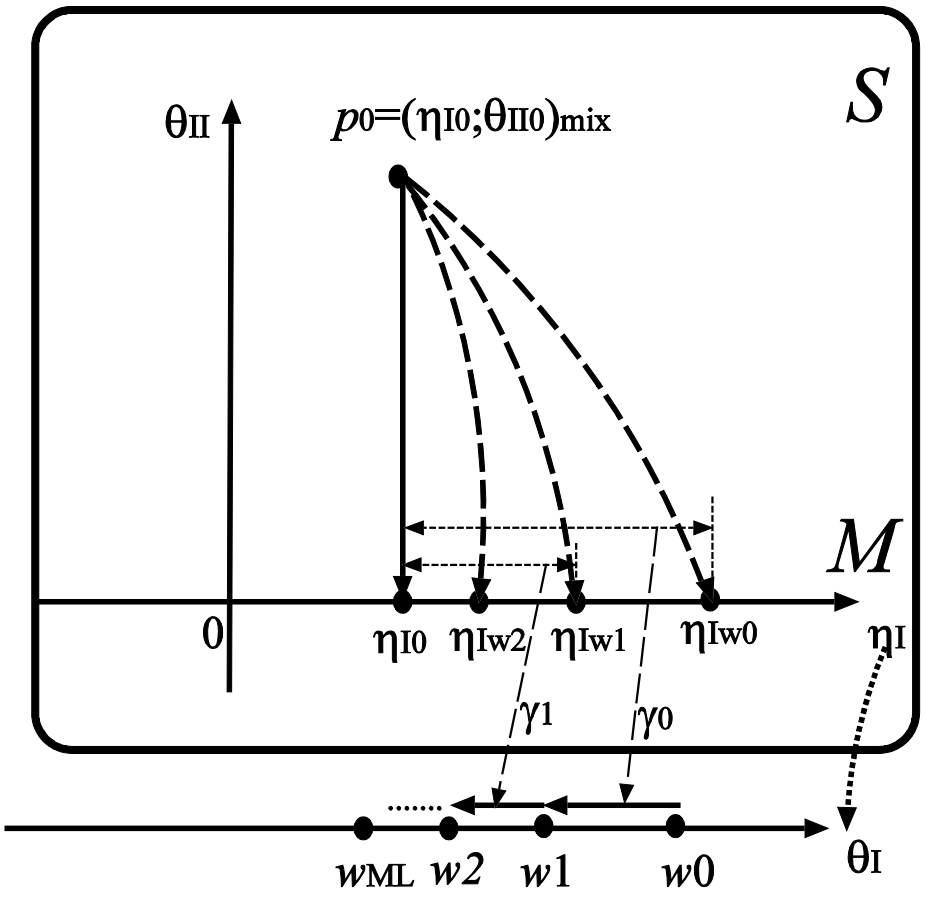
- どのような初期値から始めても，
  $p(X_1, X_2, X_3 \mid X_4, X_5)$ に分布収束する
- パーティクルフィルタなどもこの一種

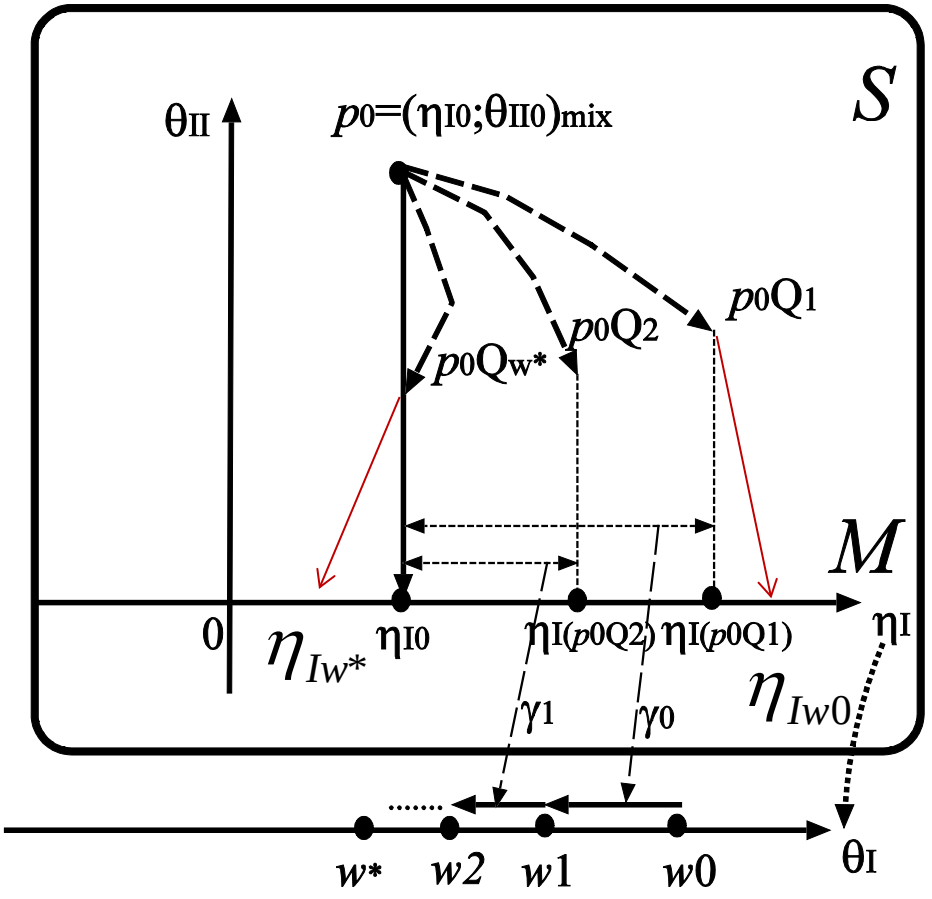
ギブスサンプラーの幾何
(Takabatake,Akaho2008)

- 1ステップに一つの変数を更新するマルコフ連鎖モンテカルロを考える.

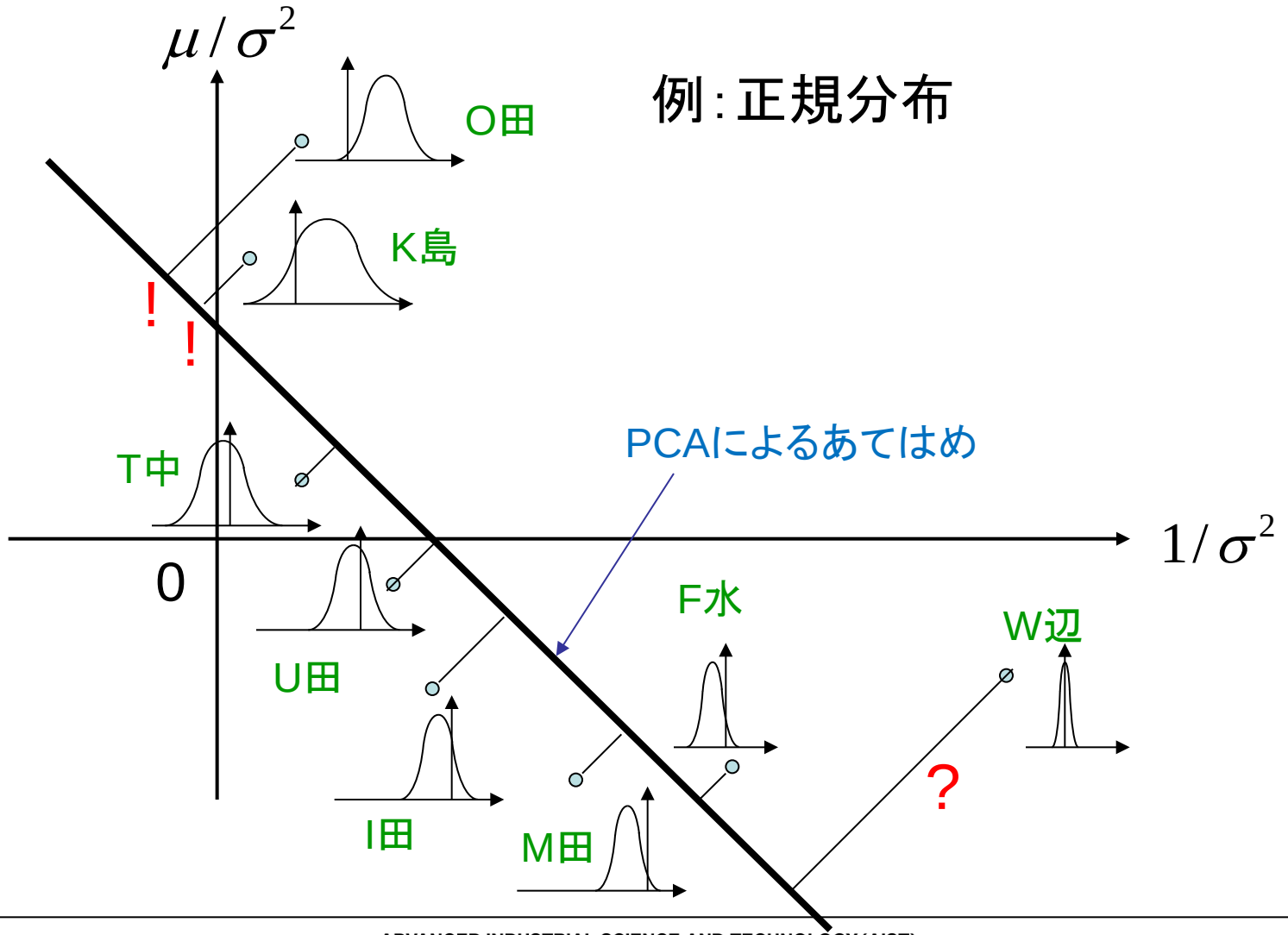


コントラスティブダイバージェンスの情報幾何（ディープラーニング）





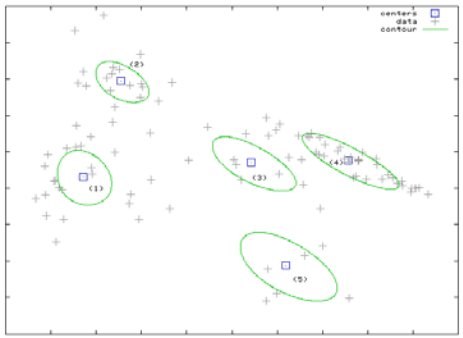
分布パラメータの次元圧縮
（Akaho2004）

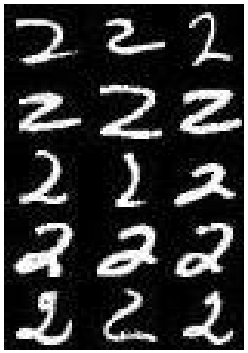


例：正規分布

分布パラメータの次元圧縮（つづき）

- 双対座標系に応じて2種類の次元圧縮法がある: e-PCA, m-PCA
  射影は必ず中に入る・距離は自然なダイバージェンス
- 次元圧縮だけでなくクラスタリングなどいろいろなデータ解析法に適用可能





手書き文字認識の e-PCA による次元圧縮とクラスタリング結果
（Watanabe, Akaho, Omachi, Okada 2008)

目次

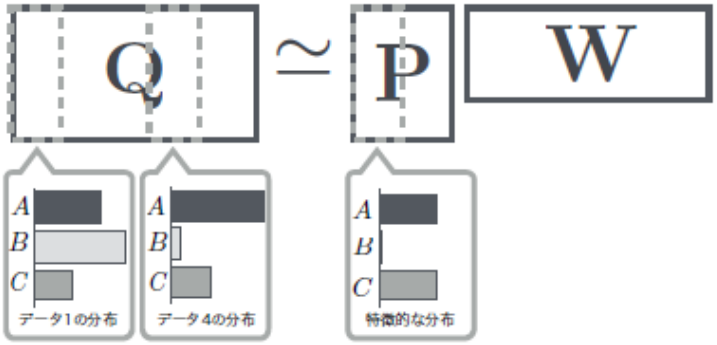
- 情報幾何とは
- 確率分布の距離と曲がった空間
- 双対平坦性
- 指数分布族　ｅとｍ
- 部分空間と射影
  ピタゴラスの定理とダイバージェンス
- 機械学習アルゴリズムの情報幾何的解釈
- 解釈を越えて（IBIS2015の発表を中心に）

T-09: 非負値行列分解の情報幾何

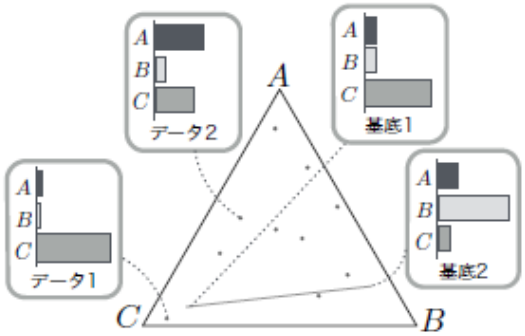
奈良 寧々花[1]・高野 健[1]・日野 英逸[2]・赤穂 昭太郎[3]・村田 昇[1]　学生優秀プレゼンテーション賞対象
[1]早稲田大学大学院先進理工学研究科　[2]筑波大学システム情報工学研究科　[3]産業技術総合研究所

情報幾何の観点から，新しいNMFのアルゴリズムを構成する

行列分解 $\mathbf{X} \simeq \mathbf{DC}$ における各行列に
列和を1とする制約を入れた行列 $\mathbf{Q}$，$\mathbf{P}$，$\mathbf{W}$ を作る．



確率密度関数の空間で
データを扱うことが可能となる．

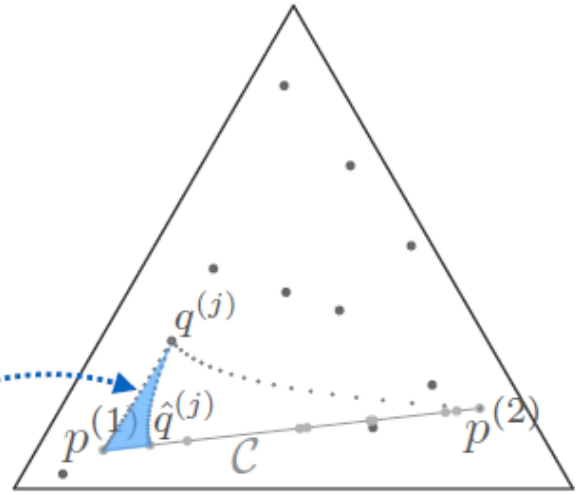


確率密度関数の空間で拡張ピタゴラスの定理を
成り立たせることで，間接的に目的関数である
近似値 $\hat{q}^{(j)}$ とデータ点 $q^{(j)}$ とのKL距離

$$D_{KL}(\hat{q}^{(j)}, q^{(j)})$$

の最小化を行うことができる．

拡張ピタゴラスの定理が
成り立つような $\hat{q}^{(j)}$ を求める．



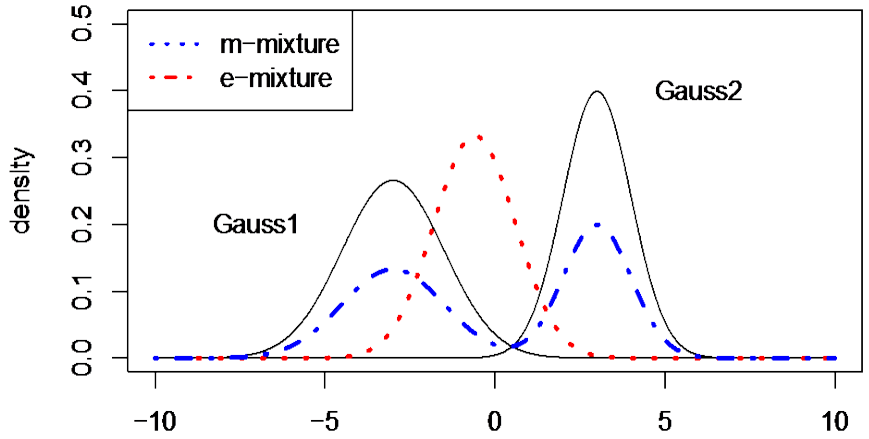
非負値行列分解の情報幾何（補足）

- トピックモデル pLSA, LDA と関係
- 実はトピックモデルは m 平坦と m 射影の組み合わせ
- 情報幾何的NMFは m 平坦と e 射影の組み合わせで幾何的により自然！

T-08:ノンパラメトリックモデルのe混合推定とその応用

高野　健(発表者・早大)・日野英逸(筑波大)・赤穂昭太郎(産総研)・村田　昇(早大)

情報幾何ではe混合とm混合という2つの混合モデルを考えることができる.

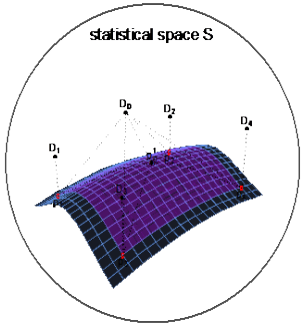


$$m\text{-mixture: } p^m = \sum_{i=1}^{N} \theta_i p_i$$

$$e\text{-mixture: } p^e = \exp\left(\sum_{i=1}^{N} \theta_i \log p_i - b(\boldsymbol{\theta})\right)$$

$$p^e(x;\boldsymbol{\theta}) = \exp\left\{\sum_{i=1}^{N} \theta_i \log \frac{1}{n_i}\sum_{k=1}^{n_i} \delta(x - x_k^{(i)}) - b(\boldsymbol{\theta})\right\}$$
$$\simeq \sum_{k=1}^{K} w_k \delta(x - y_k)$$

1.幾何学的な観点からアルゴリズムを構成
2.転移学習のようなアプローチで応用



ノンパラメトリックモデルのe混合

参考文献

- 赤穂：情報幾何と機械学習
（「計測と制御」2005年5月号）
- 甘利・長岡：情報幾何の方法，岩波講座応用数学，1993
- 公文：推定と検定への幾何学的アプローチ，
（「統計科学のフロンティア 2
統計学の基礎II」，岩波書店），2003
- 村田：新版 情報理論の基礎（SGC Books)，サイエンス社，2008
- 甘利：情報幾何の新展開，サイエンス社，2014
- 藤原：情報幾何学の基礎，牧野書店，2015

Thank you for your attention!