Wasserstein 統計学に向けて

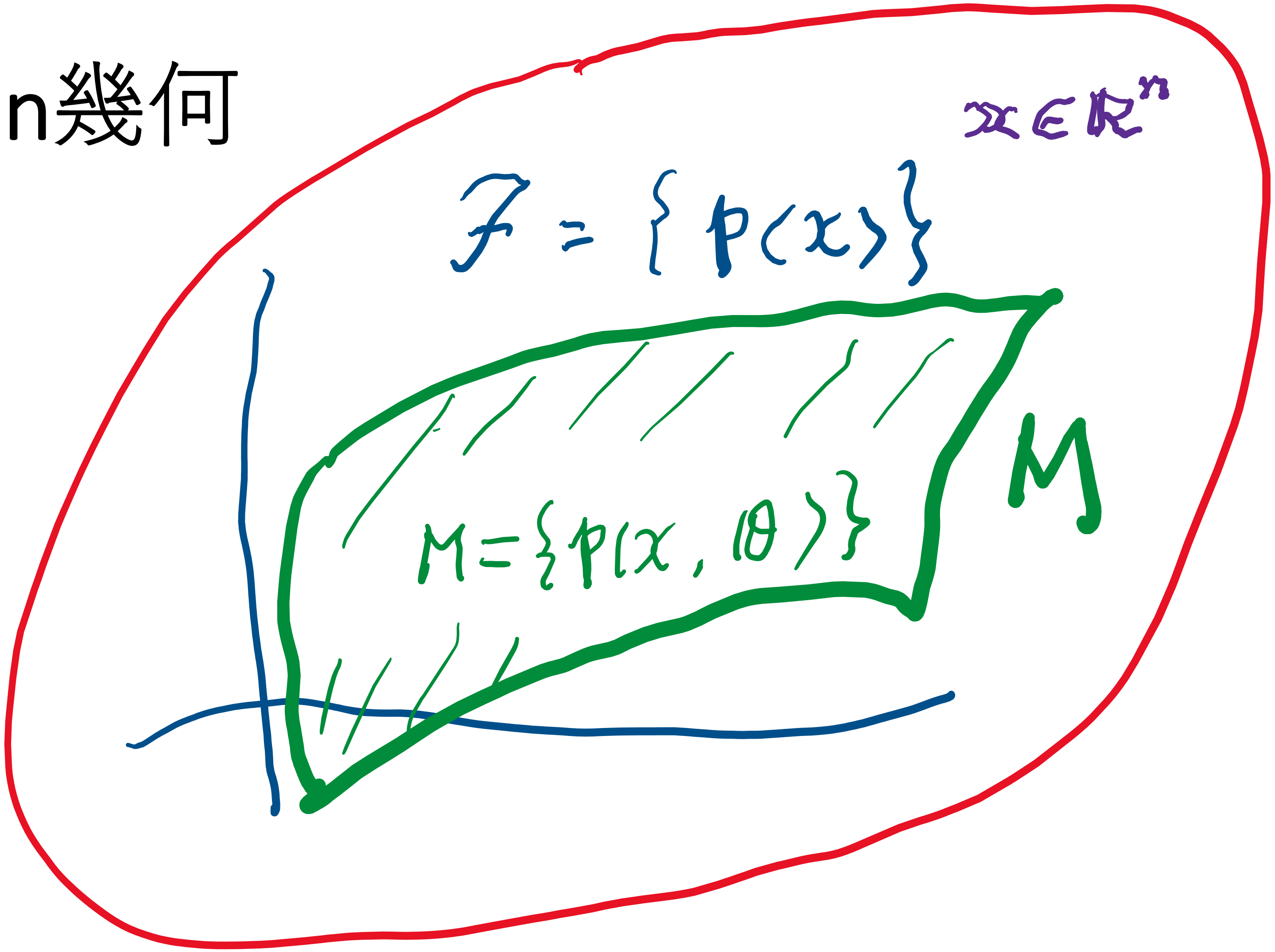
分布の形と変形

甘利俊一
帝京大学先端総合研究機構

研究協力者　松田 孟留（東大計数、理研脳科学）

情報幾何とWasserstein幾何

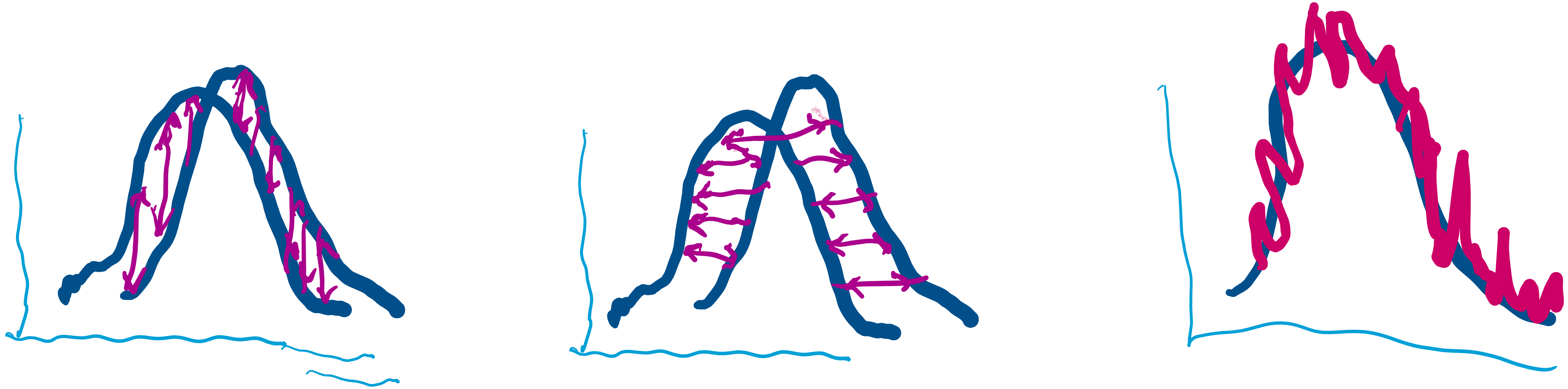
確率分布族の空間



幾何学はどう違うのか

divergence　D[p(x):q(x)]

不変なdivergence ： KL, α, Hellinger
輸送コスト：W-divergence



統計的推論

数理統計学　Fisher

推定、検定ー統計科学（Bayes）

情報幾何学

リーマン計量と双対接続

W幾何学:

輸送問題、物質輸送、非平衡熱力学

画像: クラスター、分類、重心 --- Cuturi

情報幾何とW幾何の統合

GSI（Geometric Science of Information）
課題として掲げる

Amari-Karakida-Oizumi---regularized W幾何
（Cuturi）

Li-Zhao理論、
Wong理論(divergence空間上の輸送問題ー双対幾何)

W統計学----分布形と変形

推定　　　　検定 (今泉他）　　その他

consistency, efficiency

$$x_1, x_2, \ldots\ldots, x_n$$

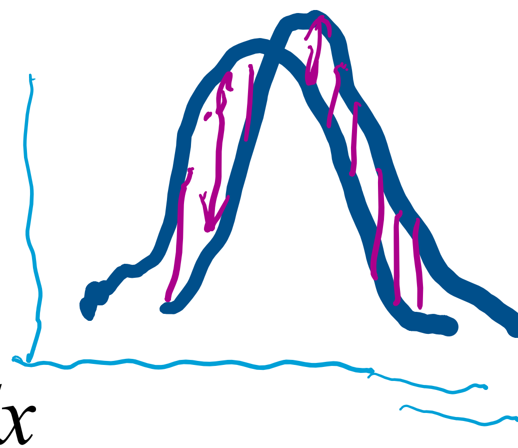
$$p_{emp}(x) = \frac{1}{n}\sum \delta(x - x_i)$$

$$\hat{\theta} = \mathrm{argmin}\, D[p_{emp}(x) : p(x,\theta)]$$

location-scale model

$$p(x,\theta) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

Divergence

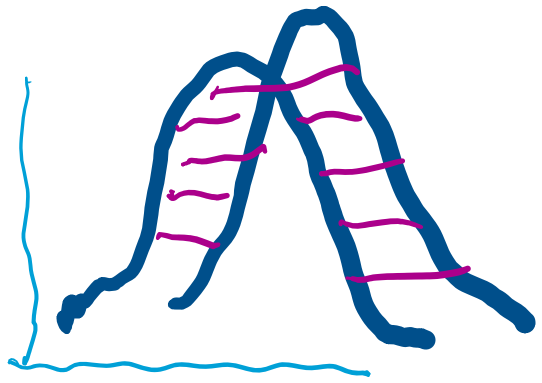


不変なdivergence

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

W-divergence

$$D_W[p(x):q(x)] = \min_{\pi}\int c(x,y)\pi(x,y)dxdy$$



$$\int \pi(x,y)dy = p(x)$$

$$\int \pi(x,y)dx = q(y)$$

$$c(x,y) = |x-y|^2$$

Riemann計量

$$\rho(x,t):\rho(x,0)=p(x),\ \rho(x,1)=q(x)$$

$$D_W[p(x):q(x)]=\inf\int_0^1|\mathbf{v}(x,t)|^2\,dt$$

$$\mathbf{v}(x,t)=\nabla_x\Phi(x,t)$$

Φ 速度ポテンシャル

$$ds^2=D[p(x):p(x)+\delta p(x)]$$

$$ds^2=\int\delta p(x)g\delta p(y)dxdy$$

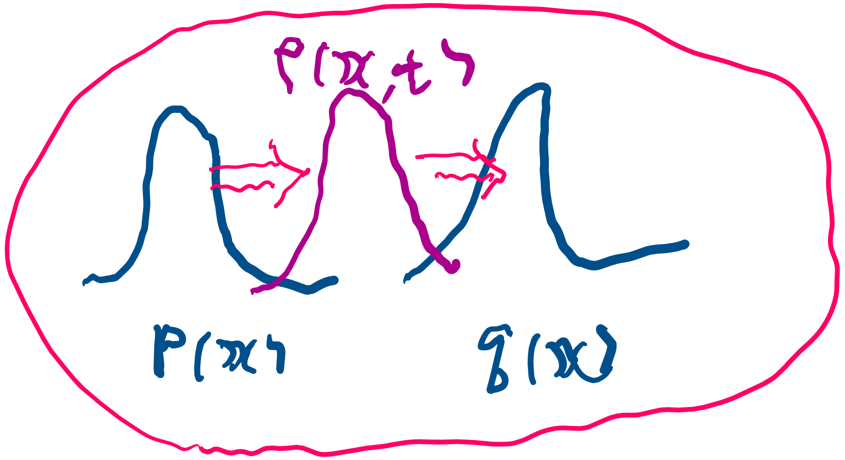
$$ds^2=\int\frac{(\delta p(x))^2}{p(x)}dx=E[(\delta\log p(x))^2]$$

$$ds^2=-\int\delta p(x)(\Delta_p)^{-1}\delta p(x)dx$$

$$\Delta_p\Phi(x)=\nabla_x(p(x)\nabla_x\Phi(x))$$

連続の方程式

$$\dot{\rho}(x,t)+div(\rho\mathbf{v})=0$$



スコア関数（速度potential)　確率モデル

$$s_i(x,\theta) = g \circ \partial_i p(x,\theta) \qquad \partial_i = \frac{\partial}{\partial \theta_i}$$

$$M = \{p(x,\theta)\}$$

$$s_i^F(x,\theta) = \frac{\partial_i p(x,\theta)}{p(x,\theta)} = \partial_i \log p(x,\theta)$$

$$s_i^W(x,\theta) = -\Delta_p[\partial_i p(x,\theta)]$$

$$\nabla_x \log p(x,\theta) \cdot \nabla_x s_i^W(x,\theta) + \nabla_x \cdot \nabla_x s_i^W(x,\theta) + \nabla_\theta \log p(x,\theta) = 0$$

$$E[s_i(x,\theta)] = 0$$

スコア関数と計量

$$g_{ij}(\theta) =< s_i(x,\theta), g^{-1} s_j(x,\theta) >$$

$$< a(x), b(x) >= \int p(x)a(x)b(x)dx$$

$$g^{F}{}_{ij}(\theta) =< s_i(x,\theta), g^{-1} s_j(x,\theta) >= \int p(x)\, s_i(x,\theta) s_j(x,\theta) dx$$

$$g^{W}{}_{ij}(\theta) =< s_i(x,\theta), g^{-1} s_j(x,\theta) >= \int p(x)\partial_x s_i(x,\theta)\partial_x s_j(x,\theta) dx$$

スコア関数と推定

観測データ　　$x_1, \dots, x_n$

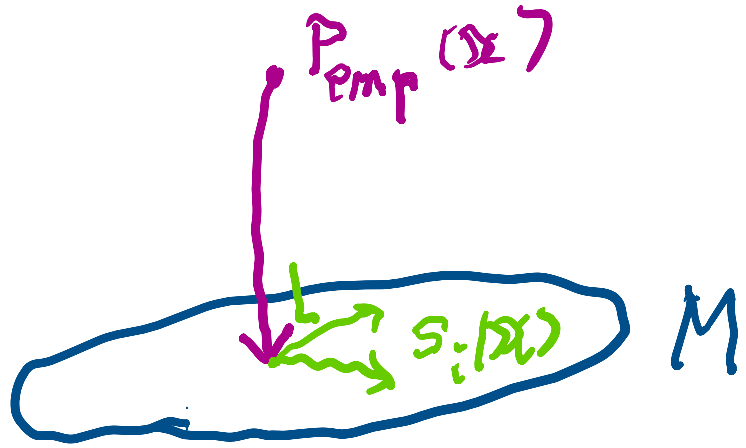
$$p_{emp}(x) = \frac{1}{n} \sum \delta(x - x_i)$$

推定関数
推定方程式　　$E[s_i(x, \theta)] = 0$

$$\sum_{j=1}^{n} s_i(x_j, \theta) = 0$$

射影
consistent,
efficient:直交

$$E_{emp}[s_i(x, \theta)] = 0$$



Cramer-Rao不等式—efficiency

$$Cov[\hat{\theta}-\theta] \geq \frac{1}{n} g_F^{-1}; \quad Cov[\hat{\theta}_F-\theta] \approx \frac{1}{n} g_F^{-1}$$

$$Cov_W[\hat{\theta}-\theta] \geq \frac{1}{n} g_W^{-1}; \quad Cov[\hat{\theta}_W-\theta] \approx \frac{1}{n} g_W^{-1}$$

$$Cov_W[\mathbf{T}(x)] = \int p(x) \nabla_x \mathbf{T}(x) \nabla_x \mathbf{T}(x)^T dx$$

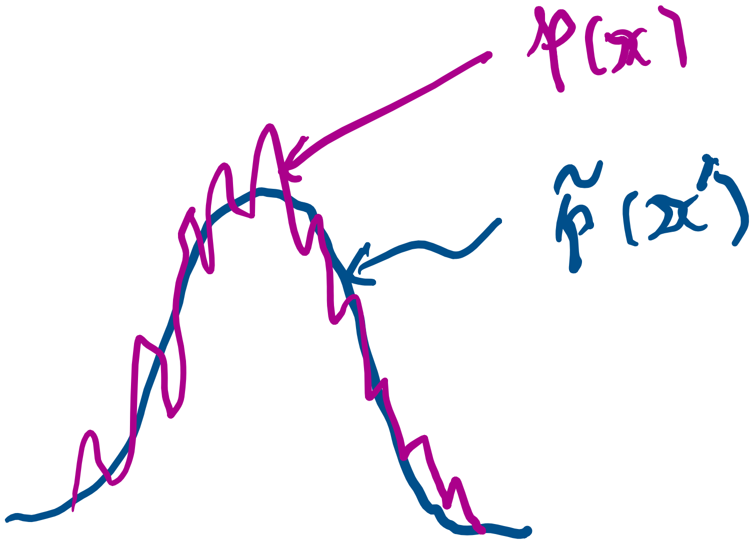
W-efficiencyの意味：波形安定性

$$x \to x' = x + n \qquad n \sim r(n)$$

$$\theta \to \theta + d\theta$$

$$p(x') = \int p(x'-n)r(n)dn$$

$$Cov_W[d\theta] = \mathbf{n}^T Cov_W \mathbf{n}$$



Affine-deformation model

標準波形の族

$\mathcal{F}_s = \{ f(z) \}$

$$\int f(z)dz = 1$$

$$\int \mathbf{z} f(z)dz = 0$$

$$\int \mathbf{z}\mathbf{z}^T f(z)dz = I$$

$$p(x,\theta) = |\Lambda| f(\Lambda(x-\mu)) \quad : M_f$$

$$\mathbf{z} = \Lambda(\mathbf{x}-\boldsymbol{\mu})$$

$$\Lambda = ODO^T$$

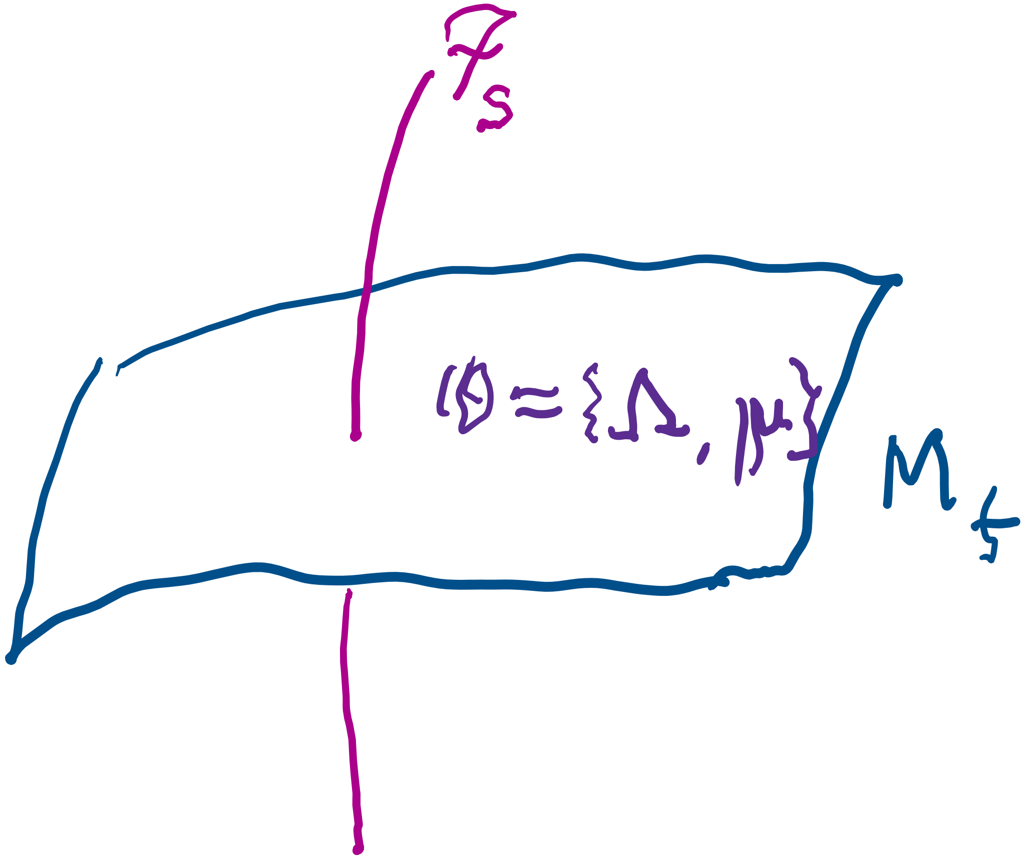
非等方的
拡大縮小

$$\int f(x)dx = 1$$

$$\int \mathbf{x} f(x)dx = \mu$$

$$\int \mathbf{x}\mathbf{x}^T f(x)dx = \Lambda^2 + \mu\mu^T$$

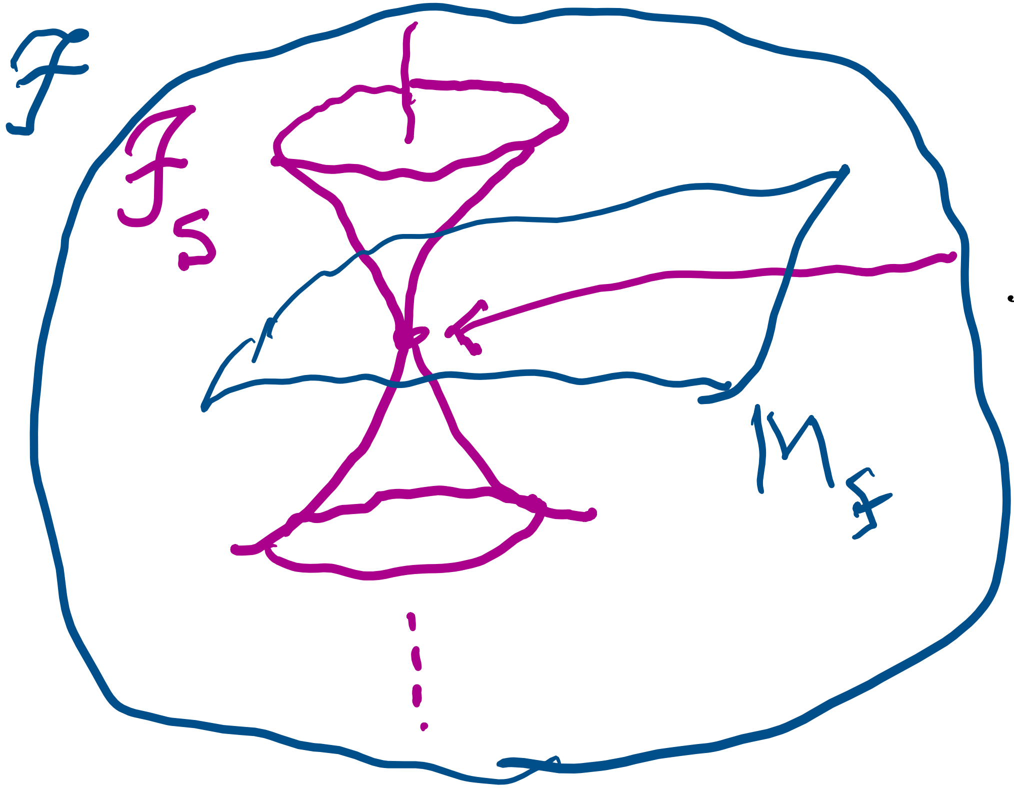
波形と変形：　W-直交



$$\int \delta f(z) s_i^W(z,\theta)\,dz = 0$$

標準波形 f の空間:　形と変形

波形の回転　$\tilde{f} = f(O\mathbf{z})$



$f = \tilde{f}$

W-Score Functions

$$s_i^W(x,\theta) = -\Delta_p^{-1}\partial_i p(x,\theta)$$

$$\partial_x \log p(x,\theta)\cdot\partial_x s_i^W(x,\theta) + \Delta s_i^W(x,\theta) + \partial_\theta \log p(x,\theta) = 0$$

W-Score Functions　　x の2次関数

$$s_\Lambda(\mathbf{x}, \Lambda, \boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}\Lambda\mathbf{x} + c(\Lambda, \boldsymbol{\mu})$$

$$s_\mu(\mathbf{x}, \Lambda, \boldsymbol{\mu}) = -\mathbf{x} + \boldsymbol{\mu}$$

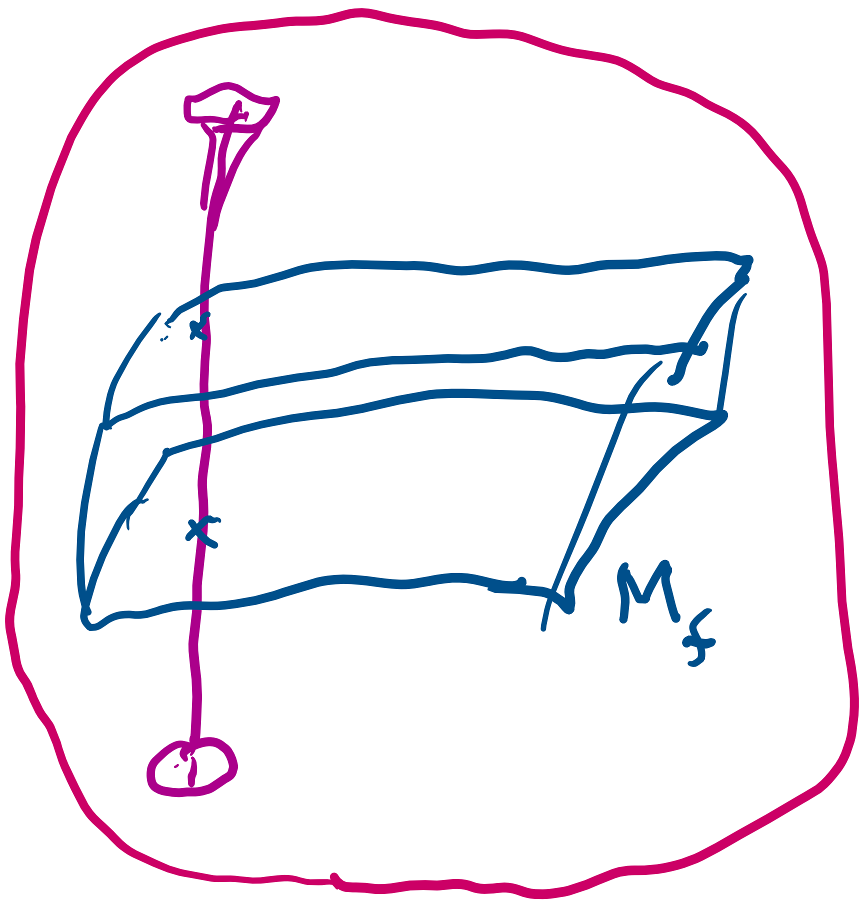
W一推定量 $\hat{\theta}^w$

定理　W推定量は2次モーメント推定量である。

$$\hat{\mu} = \frac{1}{n}\sum x_i$$

$$\hat{\Lambda} = \frac{1}{n}\sum (x_i - \hat{\mu})(x_i - \hat{\mu})^T$$

W-計量：Euclidean



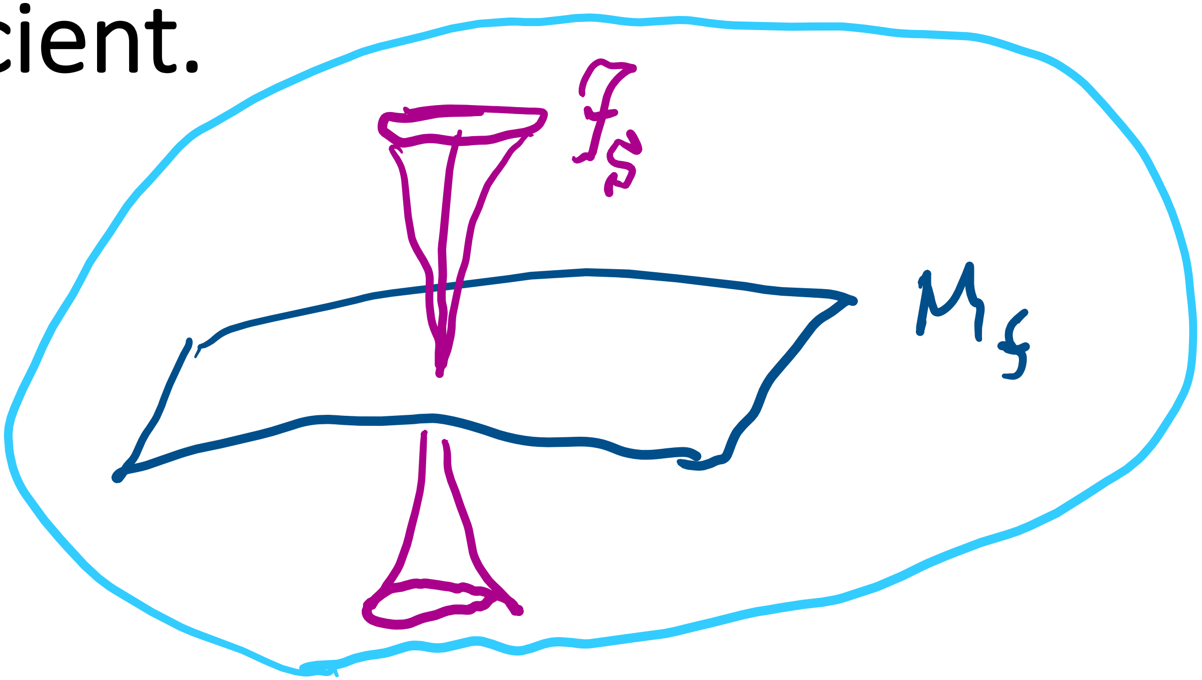
$$g_{ij}^{W}(\theta) = E[\partial_x s_i^W \cdot \partial_x s_j^W]$$

$$g^W = I$$ ：ユークリッド

定理　ガウス分布波形　　f

$$\hat{\theta}_F = \hat{\theta}_W$$

両者は一致し、ともにF-efficient,
W-efficient.



Ｆの高次キュミュラント

$g^F(\theta)-----f$に依存

$g^W(\theta)-----f$に寄らない

$\hat{\theta}^W$の情報量損失(*efficiecy*)

高次の*cumulant*

W統計学の建設に向けて

形の幾何、流れの幾何

Bayes統計学とWong理論
exponential familyとpriorの空間