Information Geometry and its Applications IV
June 12-17, 2016, Liblice, Czech

In honor of Professor Amari

Information geometry associated with two generalized means

Shinto Eguchi
Institute of Statistical Mathematics

A joint work with Osamu Komori and Atsumi Ohara
University of Fukui

Outline

- **Information geometry**

  (e-geodesic , m-geodesic, KL-divergence)

- **Generalized information geometry**

  Kolmogorov - Nagumo mean

  Generalized (e - geodesic, m-geodesic,KL-divergence)

  Quasi divergence

  The other generalized KL divergence

The core of information geometry

$$\mathscr{F} = \{ f : f(x) \geq 0, \int f(x) dP(x) = 1 \}$$

m-geodesic $\quad \mathrm{C}^{(\mathrm{m})}_{f,g} = \{ f^{(\mathrm{m})}_t(x) := (1-t) f(x) + t\, g(x) : t \in [0,1] \}$

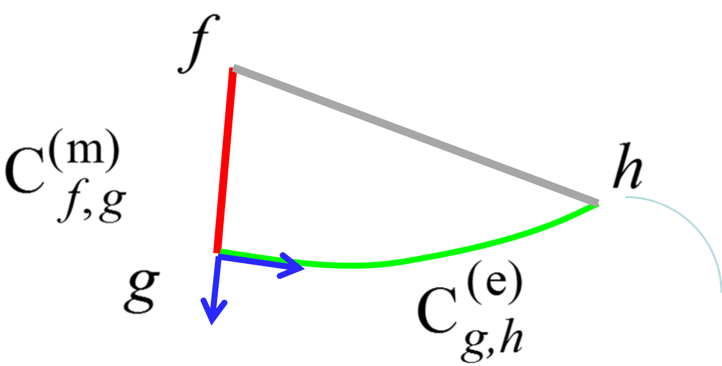
e-geodesic $\quad \mathrm{C}^{(\mathrm{e})}_{f,g} = \{ f^{(\mathrm{e})}_t(x) := e^{(1-t)\log f(x) + t \log g(x) - \kappa(t)} : t \in [0,1] \}$

KL divergence $\quad D(f, g) = \mathrm{E}_f (\log f - \log g)$

Pythagoras

$$\mathrm{C}^{(\mathrm{m})}_{f,g} \perp_g \mathrm{C}^{(\mathrm{e})}_{g,h} \Leftrightarrow D(f,h) = D(f,g) + D(g,h)$$

(Amari-Nagaoka, 2001)



Metric and connections

Let $M = \{ f_\theta(x) : \theta = (\theta_1, \cdots, \theta_d) \in \Theta \}$ with $\Theta \subseteq \mathbb{R}^d$

Information metric $$G_{ij}(\theta) = \int \frac{\partial f_\theta}{\partial \theta_i} \frac{\partial \log f_\theta}{\partial \theta_j} dP$$

m-connection $$\Gamma^{(m)}_{ij,k}(\theta) = \int \frac{\partial^2 f_\theta}{\partial \theta_i \partial \theta_j} \frac{\partial \log f_\theta}{\partial \theta_k} dP$$

e-connection $$\Gamma^{(e)}_{ij,k}(\theta) = \int \frac{\partial f_\theta}{\partial \theta_k} \frac{\partial^2 \log f_\theta}{\partial \theta_i \partial \theta_j} dP$$

Conjugate $$\frac{\partial}{\partial \theta_k} G_{ij}(\theta) = \Gamma^{(m)}_{ik,j}(\theta) + \Gamma^{(e)}_{jk,i}(\theta)$$

Rao (1945), Dawid (1975), Amari (1982)

Exponential model

Let us fix a canonical statistic $\boldsymbol{t} = (t_1, \cdots, t_K)$.

Exponential model $M^{(\mathrm{e})} = \{ f_{\boldsymbol{\theta}}^{(\mathrm{e})}(\boldsymbol{x}) := \exp\{\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{t}(\boldsymbol{x}) - \kappa(\boldsymbol{\theta})\} : \boldsymbol{\theta} \in \Theta \}$

Mean parameter $\boldsymbol{\eta} = \mathrm{E}_{f_{\boldsymbol{\theta}}^{(\mathrm{e})}}\{\boldsymbol{t}(\boldsymbol{X})\} = \frac{\partial}{\partial \boldsymbol{\theta}} \kappa(\boldsymbol{\theta})$

Remark 1 $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are affine parameters wrt $\Gamma^{(\mathrm{e})}$ and $\Gamma^{(\mathrm{m})}$.

($M^{(\mathrm{e})}$ is dually flat.) Amari (1982)

Remark 2 $M^{(\mathrm{e})}$ is totally e-gedesic.

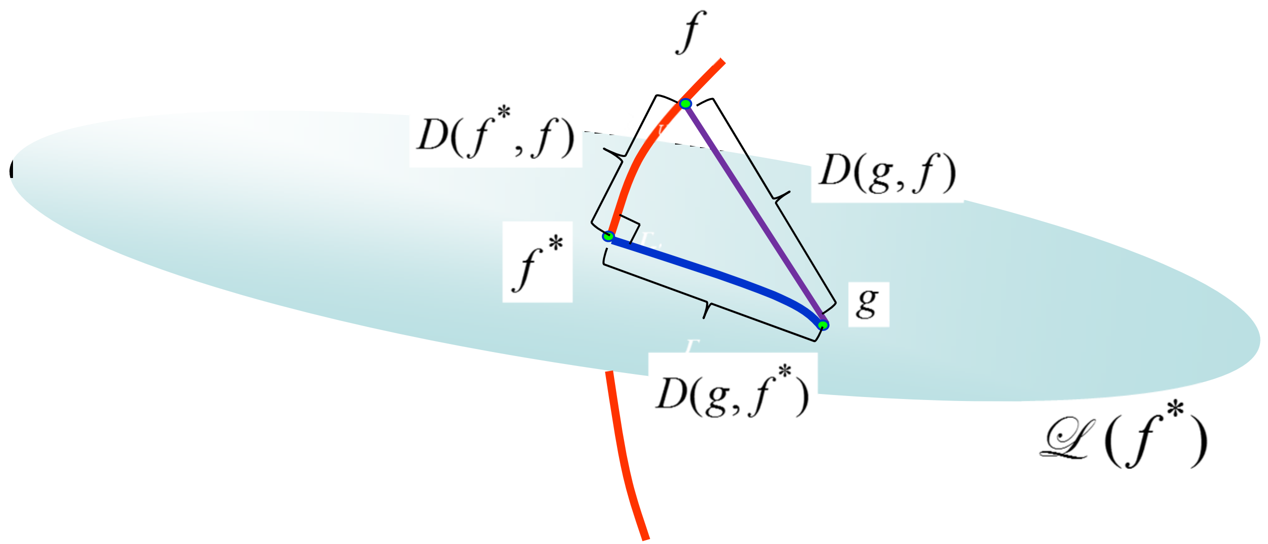
(The e-geodesic curve connecting any two densities in $M^{(\mathrm{e})}$ is always in $M^{(\mathrm{e})}$.)

Minimum KL leaf

Exponential model $M^{(\mathrm{e})} = \{ f_{\boldsymbol{\theta}}^{(\mathrm{e})}(\boldsymbol{x}) : \boldsymbol{\theta} \in \Theta \}$

Mean equal space $\mathscr{L}(f) = \{ g \in \mathscr{F} : \mathrm{E}_g\{\boldsymbol{t}(\boldsymbol{X})\} = \mathrm{E}_f\{\boldsymbol{t}(\boldsymbol{X})\} \}$

$$f^* \in M^{(\mathrm{e})} \quad \Rightarrow \quad D(g,f) = D(g,f^*) + D(f^*,f) \;\; (\forall g \in \mathscr{L}(f^*),\ \forall f \in M^{(\mathrm{e})})$$

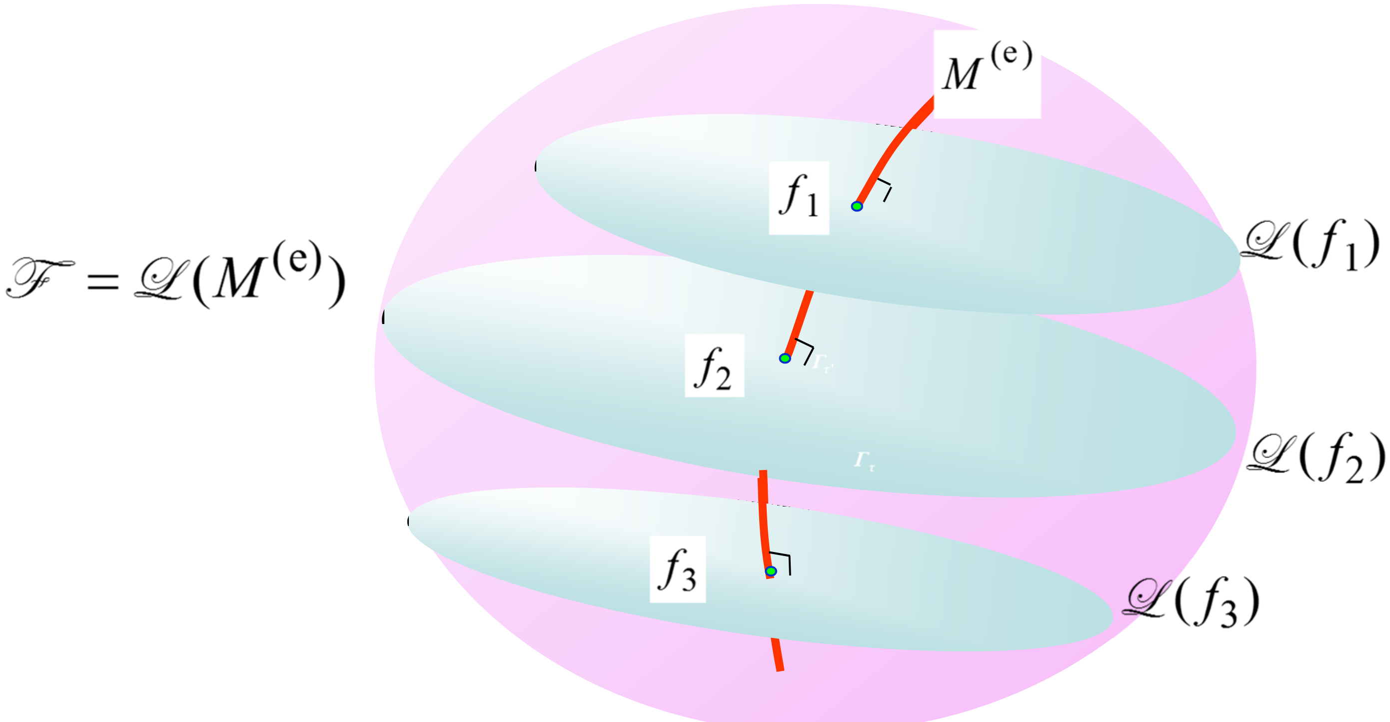


Pythagoras foliation

$\{\mathscr{L}(f) : f \in M^{(e)}\}$ is a foliation, i.e.

(i) $f_1 \neq f_2 \Rightarrow \mathscr{L}(f_1) \bigcap \mathscr{L}(f_2) = \phi$

(ii) $\mathscr{F} = \bigcup_{f \in M^{(e)}} \mathscr{L}(f)$



KL divergence (revisited)

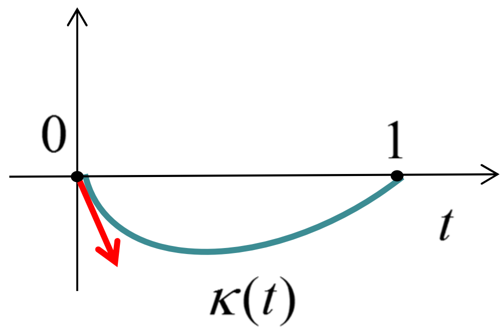
e-geodesic $f_t^{(e)}(x) = \exp\{(1-t)\log f(x) + t\log g(x) - \kappa(t)\}$

The normalizing constant

$$\kappa(t) = \log\left[\int \exp\{(1-t)\log f(x) + t\log g(x)\} dP(x)\right]$$

We observe

$$-\frac{d\kappa(0)}{dt} = \mathrm{E}_f(\log f - \log g) = D(f, g)$$

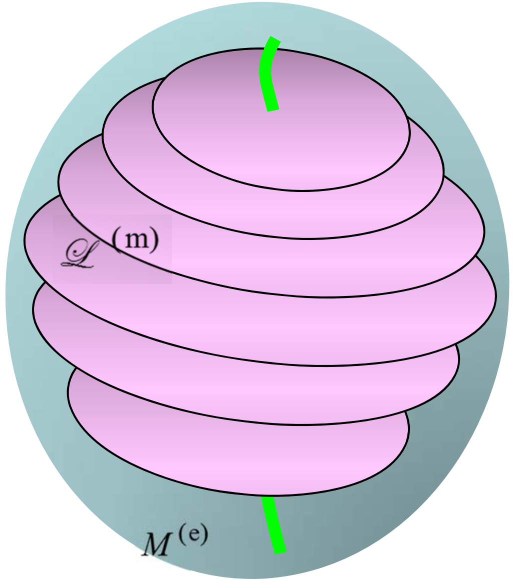


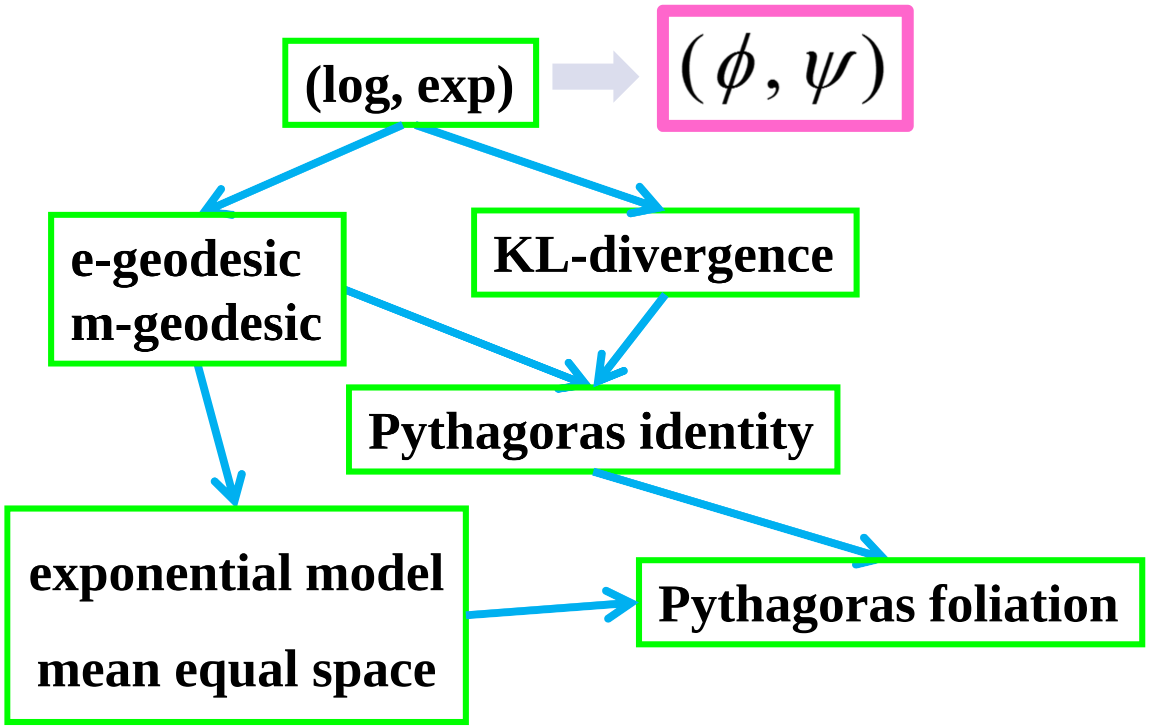
Cf. Ay-Amari (2015)

KL divergence is induced by e-geodesic

Cf**. the canonical divergence**, Amari-Nagaoka (2001)

(log , exp)





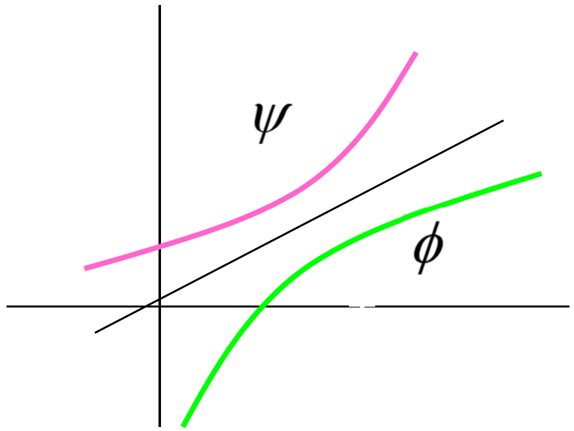
$$\exp((1-t)\log f(x) + t\log g(x) - \kappa_t)$$

$$f = \exp(\log(f))$$

$$(1-t)f(x) + t\,g(x)$$

$(\log, \exp) \longrightarrow (\phi, \psi)$

where ϕ is a concave function on $(0,\infty)$ and ψ is the inverse of ϕ.



$$\psi(\phi(f)) = f$$

$$\psi'(\phi(f))\phi'(f) = 1$$

Kolmogorov-Nagumo mean

K-N mean is $\boxed{\psi((1-t)\phi(f)+t\phi(g))}$ for positive numbers f and g.

Cf. Kolmogorov(1930), Nagumo (1930), Naudts (2009)

(1)	$\phi(s)=s$	arithmetic mean :	$(1-t)f+t\,g$
(2)	$\phi(s)=\log s$	geometric mean :	$\exp((1-t)f+tg)$
(3)	$\phi(s)=\dfrac{1}{s}$	harmonic mean :	$\dfrac{1}{(1-t)\dfrac{1}{f}+t\dfrac{1}{g}}$

Generalized e-geodesic

$$C_{f,g}^{(\mathrm{e}\phi)} = \{ f_t^{(\mathrm{e}\phi)}(x) : t \in [0,1] \}$$

$$f_t^{(\mathrm{e}\phi)}(x) := \psi((1-t)\phi(f(x)) + t\,\phi(g(x)) - \kappa^{(\phi)}(t))$$

where $\kappa^{(\phi)}(t)$ is a constant to satisfy

$$\int \psi((1-t)\phi(f) + t\phi(g) - \kappa^{(\phi)}(t))\, dP = 1$$

Remark Let $C_{f_1,\cdots,f_k}^{(\mathrm{e}\phi)} = \{ \psi(\pi_1\phi(f_1) + \cdots + \pi_k\phi(f_k) - \kappa(\pi)) : \pi \in S_{k-1} \}$.

If $f, g \in C_{f_1,\cdots,f_k}^{(\mathrm{e}\phi)}$, then $C_{f,g}^{(\mathrm{e}\phi)}$ is always in $C_{f_1,\cdots,f_k}^{(\mathrm{e}\phi)}$

($C_{f_1,\cdots,f_k}^{(\mathrm{e}\phi)}$ is totally eϕ-geodesic)

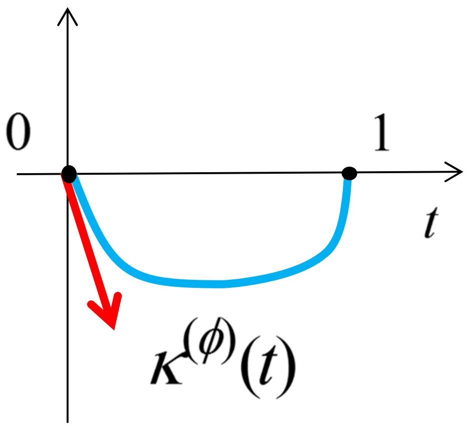
Generalized KL-divergence

$$D^{(\phi)}(f,g) = \mathrm{E}_{\xi(f)}\{\phi(f) - \phi(g)\}$$

$$\text{where} \quad \xi(f) = \frac{\dfrac{1}{\phi'(f)}}{\displaystyle\int \frac{dP}{\phi'(f)}} \qquad \left(\psi'(\phi(f)) = \frac{1}{\phi'(f)} \right)$$

Remenber $\kappa^{(\phi)}(t)$ defined to satisfy

$$\int \psi((1-t)\phi(f) + t\phi(g) - \kappa^{(\phi)}(t))\, dP = 1$$



$$\text{Then} \quad -\frac{d\kappa^{(\phi)}(0)}{dt} = \frac{\int \psi'(\phi(f))\{\phi(f) - \phi(g)\} dP}{\int \psi'(\phi(f))\, dP}$$

$$= \frac{\displaystyle\int \frac{\phi(f) - \phi(g)}{\phi'(f)}}{\displaystyle\int \frac{dP}{\phi'(f)}} \qquad (\psi'(\phi(f))\phi'(f) = 1)$$

Generalized m-geodesic

$$C_{f,g}^{(\mathrm{m}\phi)} = \{ f_t^{(\mathrm{m}\phi)}(x) : t \in [0,1] \}$$

where $f_t^{(\mathrm{m}\phi)} = \xi^{-1}((1-t)\xi(f) + t\,\xi(g)).$

Remark

(i) $\mathrm{E}_{\xi(f)}\{S(X)\} = \mathrm{E}_{\xi(g)}\{S(X)\} = \tau \Rightarrow \mathrm{E}_{\xi(f_t^{(m\phi)})}\{S(X)\} = \tau \ (\forall t \in (0,1))$

(ii) $f_t^{(\mathrm{m}\phi)} = \phi'^{-1}\left(\dfrac{1}{(1-t)\dfrac{1}{\phi'(f)} + t\dfrac{1}{\phi'(g)}}\right)$ (quasi-harmonic mean)

(iii) $C_{f_1,\cdots,f_k}^{(\mathrm{m}\phi)}$ is mϕ - totally geodesic

Metric and connections by $D^{(\phi)}$

$$D^{(\phi)}(f,g) \approx \int \xi(f)\phi(g)dP \qquad \text{cf. Eguchi (1992)}$$

Riemannian metric

$$G_{ij}^{(\phi)}(\theta) = \int \frac{\partial \xi(f_\theta)}{\partial \theta_i} \frac{\partial \phi(f_\theta)}{\partial \theta_j} dP$$

Generalized m-connection

$$\Gamma_{ij,k}^{(\mathrm{m}\phi)}(\theta) = \int \frac{\partial^2 \xi(f_\theta)}{\partial \theta_i \partial \theta_j} \frac{\partial \phi(f_\theta)}{\partial \theta_k} dP$$

Generalized e-connection

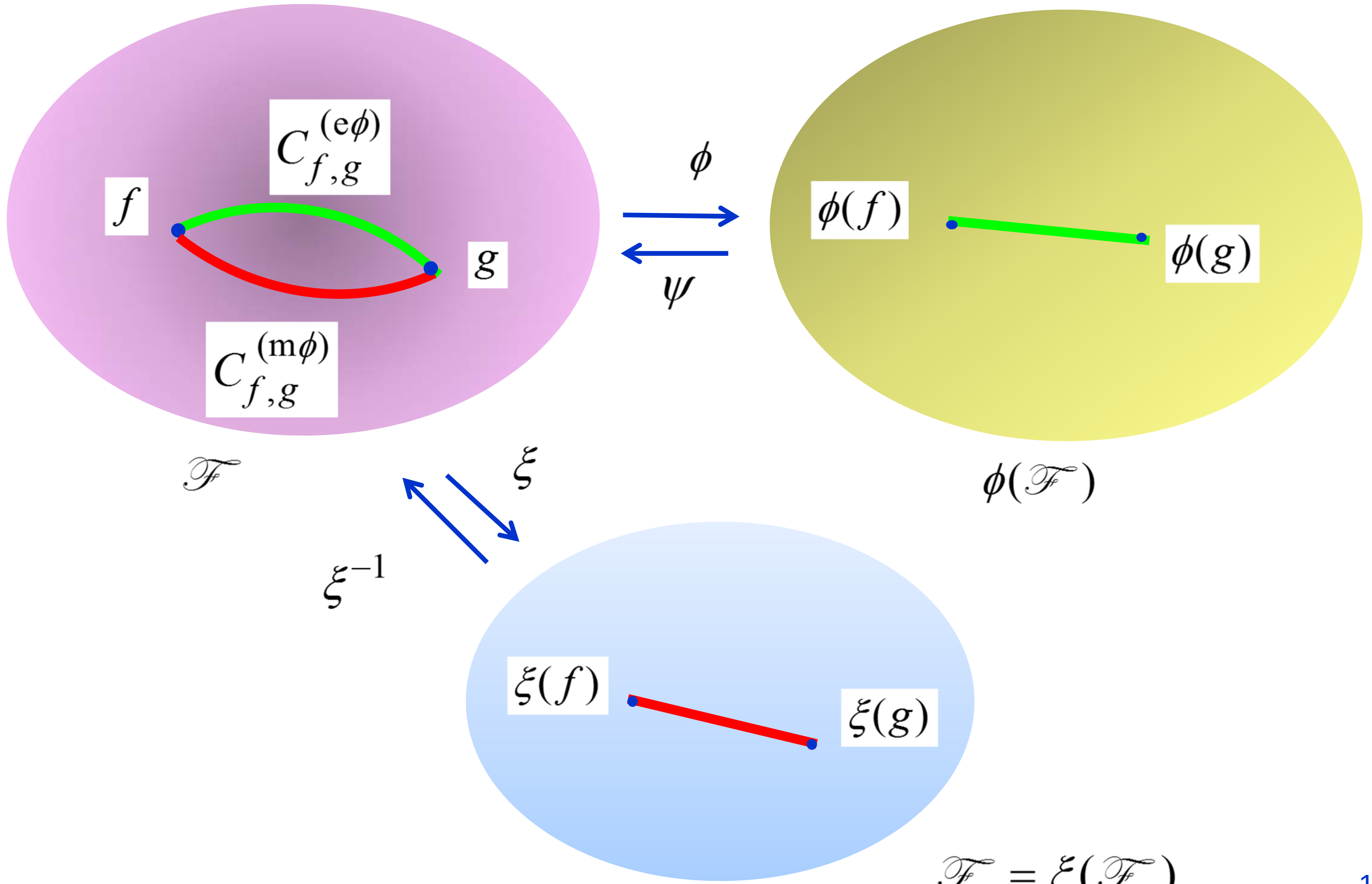
$$\Gamma_{ij,k}^{(\mathrm{e}\phi)}(\theta) = \int \frac{\partial \xi(f_\theta)}{\partial \theta_k} \frac{\partial^2 \phi(f_\theta)}{\partial \theta_i \partial \theta_j} dP$$

Remark

$$G_{ij}^{(\phi)}(\theta) \propto G_{ij}(\theta) \qquad (\text{ conformal })$$

$$\frac{\partial}{\partial \theta_k} G_{ij}^{(\phi)}(\theta) = \Gamma_{ik,j}^{(\mathrm{e}\phi)}(\theta) + \Gamma_{jk,i}^{(\mathrm{m}\phi)}(\theta) \qquad (\text{ conjugate })$$

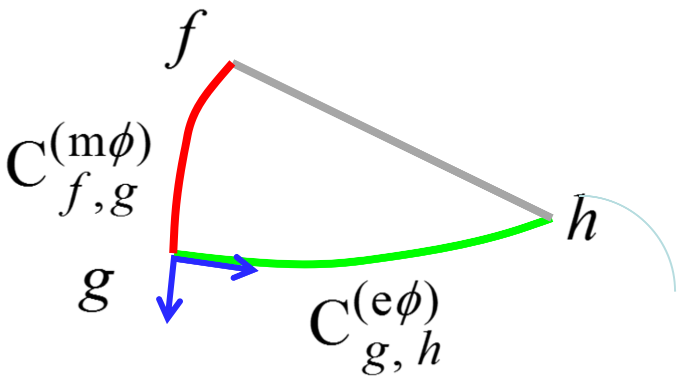
Generalized two geodesic curves



ϕ Pythagorean theorem

Pythagoras theorem

$$C^{(\mathrm{m}\phi)}_{f,g} \perp_g C^{(\mathrm{e}\phi)}_{g,h} \Leftrightarrow D^{(\phi)}(f,h) = D^{(\phi)}(f,g) + D^{(\phi)}(g,h)$$



- $\phi(s) = \log s \;\Rightarrow\; \psi(t) = \exp t,\;\; \xi(s) = s,$

$$(f_t^{(\mathrm{e}\phi)}, f_t^{(\mathrm{m}\phi)}, D^{(\phi)}) = (f_t^{(\mathrm{m})}, f_t^{(\mathrm{e})}, D\;) \quad .$$

Power-log function

- $$\phi(s) = \frac{s^{\beta} - 1}{\beta} \;\Rightarrow\; \psi(t) = (1 + \beta t)^{\frac{1}{\beta}}, \;\; \xi(s) = s^{1-\beta}$$

$$f_t^{(\mathrm{e}\phi)}(x) = ((1-t) f(x)^{\beta} + t\, g(x)^{\beta} - \kappa_t)^{\frac{1}{\beta}}$$

$$f_t^{(\mathrm{m}\phi)}(x) = ((1-\kappa_{\beta t}) f(x)^{1-\beta} + \kappa_{\beta t}\, g(x)^{1-\beta})^{\frac{1}{1-\beta}}$$

$$D^{(\phi)}(f, g) = \frac{1}{\beta \int f^{1-\beta} dP} \{1 - \int f^{1-\beta} g^{\beta} dP\}$$

Generalized exponential model

G-exponential model

$$M^{(\mathrm{e}\phi)} = \{ f_\theta^{(\mathrm{e}\phi)}(x) := \psi(\theta^{\mathrm{T}} t(x) - \kappa^{(\phi)}(\theta)) : \theta \in \Theta \}$$

where $\kappa^{(\phi)}(\theta)$ is a normalizing constant.

Mean parameter $\boldsymbol{\eta} = \mathrm{E}_{\xi(f_{\boldsymbol{\theta}}^{(\mathrm{e}\phi)})}\{\boldsymbol{t}(\boldsymbol{X})\} = \dfrac{\partial}{\partial \boldsymbol{\theta}} \kappa^{(\phi)}(\boldsymbol{\theta})$ Cf. Naudts (2010)

Remark 1 θ and η are affine parameters wrt $\Gamma^{(\mathrm{e}\phi)}$ and $\Gamma^{(\mathrm{m}\phi)}$.

($M^{(\mathrm{e}\phi)}$ is dually flat.) Cf. Matsuzoe-Henmi (2014)

Remark 2 $M^{(\mathrm{e}\phi)}$ is totally $(\mathrm{e}\,\phi)$ - geodesic.

(The generalized e-geodesic curve connecting any two densities in $M^{(\mathrm{e}\phi)}$ is always in $M^{(\phi\mathrm{e})}$.)

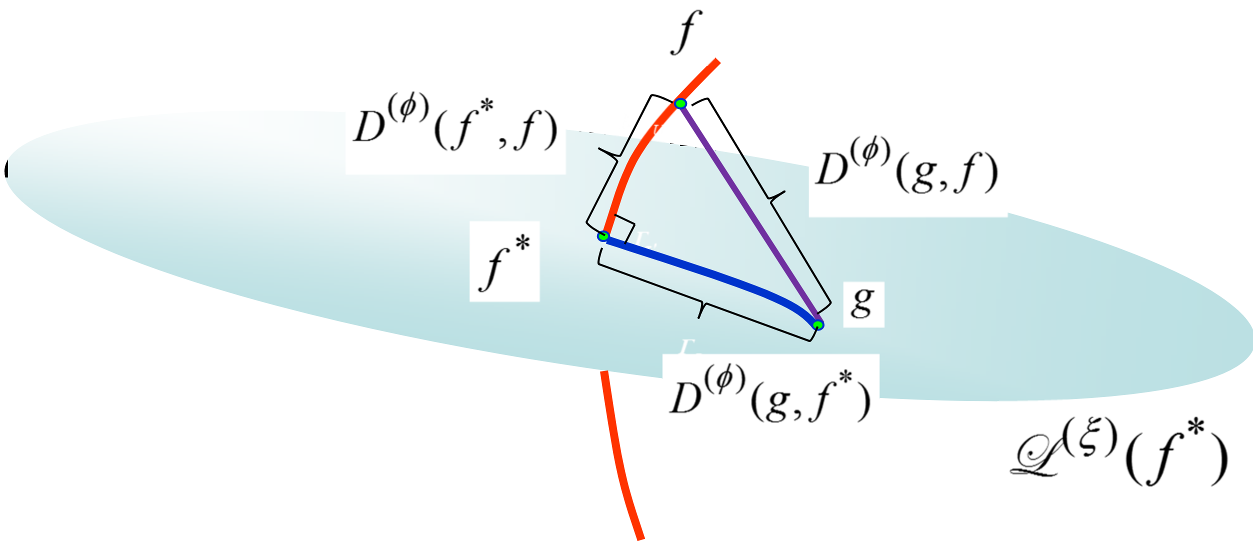
Minimum GKL leaf

G-exponential model $M^{(\mathrm{e}\phi)} = \{ f_{\boldsymbol{\theta}}^{(\mathrm{e}\phi)}(\boldsymbol{x}) : \boldsymbol{\theta} \in \Theta \}$

Mean equal space $\mathscr{Q}^{(\xi)}(f) = \{ g : \mathrm{E}_{\xi(g)}\{\boldsymbol{t}(\boldsymbol{X})\} = \mathrm{E}_{\xi(f)}\{\boldsymbol{t}(\boldsymbol{X})\} \}$

$$f^* \in M^{(\mathrm{e})} \Rightarrow$$

$$D^{(\phi)}(g,f) = D^{(\phi)}(g,f^*) + D^{(\phi)}(f^*,f) \quad (\forall g \in \mathscr{Q}^{(\xi)}(f^*), \forall f \in M^{(\mathrm{e})})$$



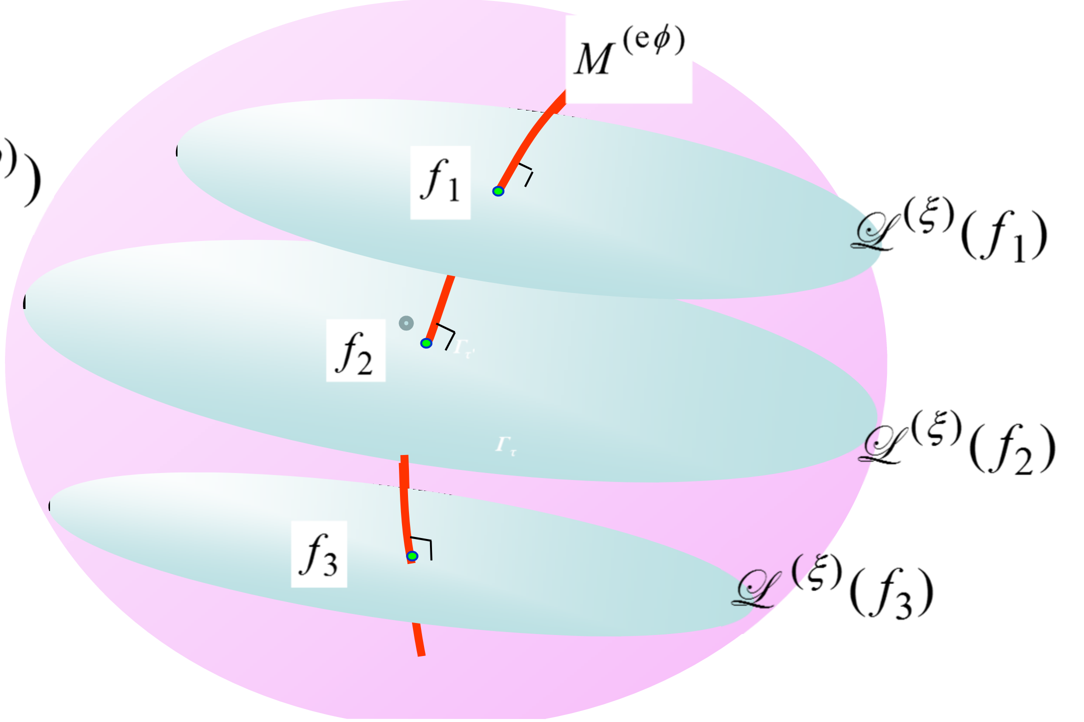
Generalized Pythagoras foliation

$\{\mathscr{L}^{(\xi)}(f) : f \in M^{(\mathrm{e}\phi)}\}$ is a foliation

(i) $f_1 \neq f_2 \Rightarrow \mathscr{L}^{(\xi)}(f_1) \cap \mathscr{L}^{(\xi)}(f_2) = \phi$

(ii) $\mathscr{F} = \bigcup_{f \in M^{(\mathrm{e}\phi)}} \mathscr{L}^{(\xi)}(f)$

$\mathscr{F} = \mathscr{L}^{(\xi)}(M^{(\mathrm{e}\phi)})$



(log , exp)

↓

$C^{(\mathrm{e})}$

$C^{(\mathrm{m})}$

D

$M^{(\mathrm{e})}$

$\mathscr{L}$

⇒

(ϕ , ψ)

↓

$C^{(\mathrm{e}\phi)}$

$C^{(\mathrm{m}\phi)}$

$D^{(\phi)}$

$M^{(\mathrm{e}\phi)}$

$\mathscr{L}^{(\phi)}$

If $(\phi,\psi) \neq (\log,\exp)$, then

(i) the geometry depends on the choice of P, Cf. Newton (2012)

(ii) the minimum $D^{(\phi)}$ estimation is not feasible.

Expected ϕ density

Consider $$\mathrm{E}_f\{-\phi(g)\}$$

Let $$g_{\mathrm{opt}} := \arg\min_{g\in\mathscr{F}} \mathrm{E}_f\{-\phi(g)\}$$

Then $$\frac{\partial}{\partial\varepsilon}\mathrm{E}_f\{\phi((1-\varepsilon)g_{\mathrm{opt}}+\varepsilon h)\}|_{\varepsilon=0}$$

$$= \int \{f\,\phi'(g_{\mathrm{opt}})(h-g_{\mathrm{opt}})\}\,dP = 0 \quad \text{for all } h \text{ of } \mathscr{F}$$

Therefore $$f(x)\,\phi'(g_{\mathrm{opt}}(x)) = c \quad \text{with a constant } c, \text{ that is,}$$

$$g_{\mathrm{opt}}(x) = \zeta(f(x)) \quad \text{where} \quad \zeta(f) = (\phi')^{-1}\left(\frac{c}{f}\right)$$

$$\mathrm{E}_f\{-\phi(\zeta(f)\} = \min_{g\in\mathscr{F}} \mathrm{E}_f\{-\phi(g)\}$$

Quasi divergence

Remark

$$\xi(\zeta(f)) = \frac{\frac{1}{\phi'(\zeta(f))}}{\int \frac{dP}{\phi'(\zeta(f))}} = \frac{\frac{f}{c}}{\int \frac{f}{c} dP} = f$$

because $\xi(f) = \frac{1}{\phi'(f)} / \int \frac{dP}{\phi'(f)}$ and $\zeta(f) = (\phi')^{-1}(\frac{c}{f})$

Let $D_0(f,g) = \mathrm{E}_f\{-\phi(g)\} - \min_{f \in \mathcal{F}} \mathrm{E}_f\{-\phi(g)\}.$

Then $D_0(f,g) = \mathrm{E}_f\{\phi(\xi^{-1}(f)) - \phi(g)\}$

$$D_0(f,g) \geq 0 \text{ with equality iff } g = \xi^{-1}(f)$$

Another generalization of KL-divergence

One adjustment

$$D^{(\phi)}(f,g) = D_0(\xi(f),g) = \mathrm{E}_{\xi(f)}\{\phi(f)-\phi(g)\}$$

The other adjustment

$$\Delta^{(\phi)}(f,g) = D_0(f,\xi^{-1}(g)) = \mathrm{E}_f\{\phi(\xi^{-1}(f))-\phi(\xi^{-1}(g))\}$$

Remark

(i) $\Delta^{(\phi)}(f,g) = D^{(\phi)}(\xi(f),\xi(g))$

(ii) $\Delta^{(\phi)}(f,g)$ is a unique divergence defined by the standard expectation in the class of corrected quasi divergence

Gamma divergence

- $$\phi(f) = \frac{f^{\beta} - 1}{\beta} \quad \Rightarrow \quad \xi(g(x)) = \frac{g(x)^{1-\beta}}{\int g^{1-\beta} dP},$$

$$\xi^{-1}(g(x)) = \frac{g(x)^{\frac{1}{1-\beta}}}{\int g^{\frac{1}{1-\beta}} dP}$$

- $$\Delta^{(\phi)}(f, g) = -\frac{\mathrm{E}_f(g^{\frac{\beta}{1-\beta}})}{(\int g^{\frac{1}{1-\beta}} dP)^{\beta}} + (\int f^{\frac{1}{1-\beta}} dP)^{1-\beta}$$

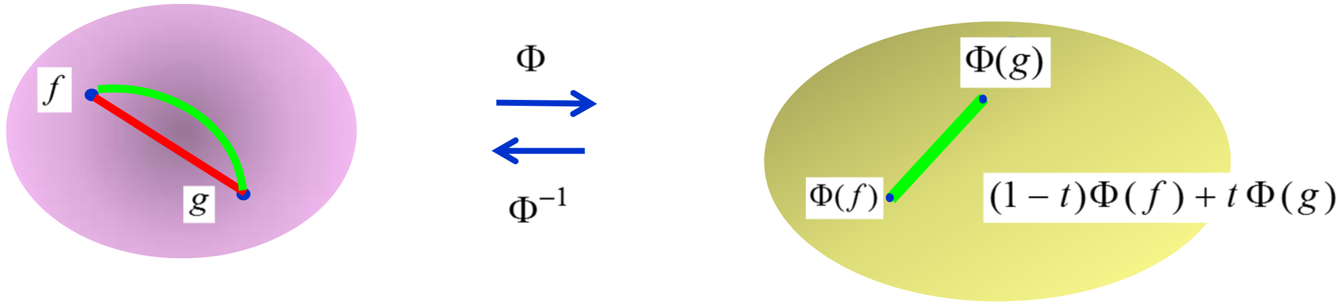
Remark $\Delta^{(\phi)}$ is γ-divengence with $\gamma = \frac{\beta}{1-\beta}$ Fujisawa-Eguchi (2008)

$\Delta^{(\phi)}$ -geometry

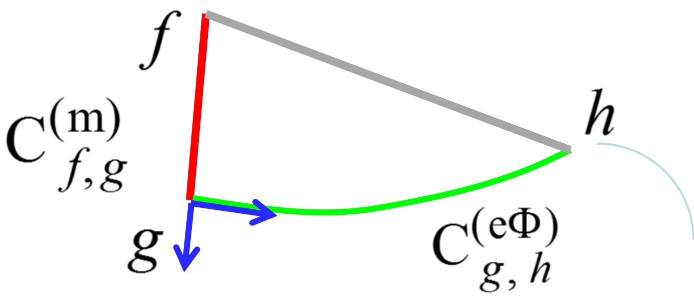
$$\text{Let } \Phi(f) = \phi(\xi^{-1}(f))$$

m-geodesic $C^{(\mathrm{m})} : \quad (1-t)f + t\,g$

G-e-geodesic $C^{(\mathrm{e}\Phi)} : \quad \Phi^{-1}((1-t)\Phi(g) + t\,\Phi(h) - \kappa^{(\Phi)}(t))$



Pythagoras $C^{(\mathrm{m})}_{f,g} \perp_g C^{(\mathrm{e}\Phi)}_{g,h} \Rightarrow \Delta^{(\phi)}(f,h) = \Delta^{(\phi)}(f,g) + \Delta^{(\phi)}(g,h)$



Metric and connections by $\Delta^{(\phi)}$

$$\Delta^{(\phi)}(f,g) \approx \int f\,\Phi(g)\,dP$$

Riemannian metric

$$G_{ij}^{(\Phi)}(\theta) = \int \frac{\partial f_\theta}{\partial \theta_i} \frac{\partial \Phi(f_\theta)}{\partial \theta_j} dP$$

Generalized m-connection

$$\Gamma_{ij,k}^{(\mathrm{m}\Phi)}(\theta) = \int \frac{\partial^2 f_\theta}{\partial \theta_i \partial \theta_j} \frac{\partial \Phi(f_\theta)}{\partial \theta_k} dP$$

Generalized e-connection

$$\Gamma_{ij,k}^{(\mathrm{e}\phi)}(\theta) = \int \frac{\partial f_\theta}{\partial \theta_k} \frac{\partial^2 \Phi(f_\theta)}{\partial \theta_i \partial \theta_j} dP$$

Remark

$$\Gamma_{ij,k}^{(\mathrm{m}\Phi)} = \Gamma^{(\mathrm{m})}$$

Minimum $\Delta^{(\phi)}$ estimation

Model $M = \{f_\theta : \theta \in \Theta\}$

Data set $\{X_i\} \underset{\text{i.i.d.}}{\sim} f_\theta(x)$

Loss function $L^{(\Phi)}(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\Phi(f_\theta(X_i))$

Proposed estimator $\hat{\theta}^{(\Phi)} = \underset{\theta \in \Theta}{\arg\min}\, L^{(\Phi)}(\theta)$

Expected loss $\mathbb{L}^{(\Phi)}(\theta) = \mathrm{E}_f\{-\Phi(f_\theta)\}$ if $\{X_i\} \sim f(x)$

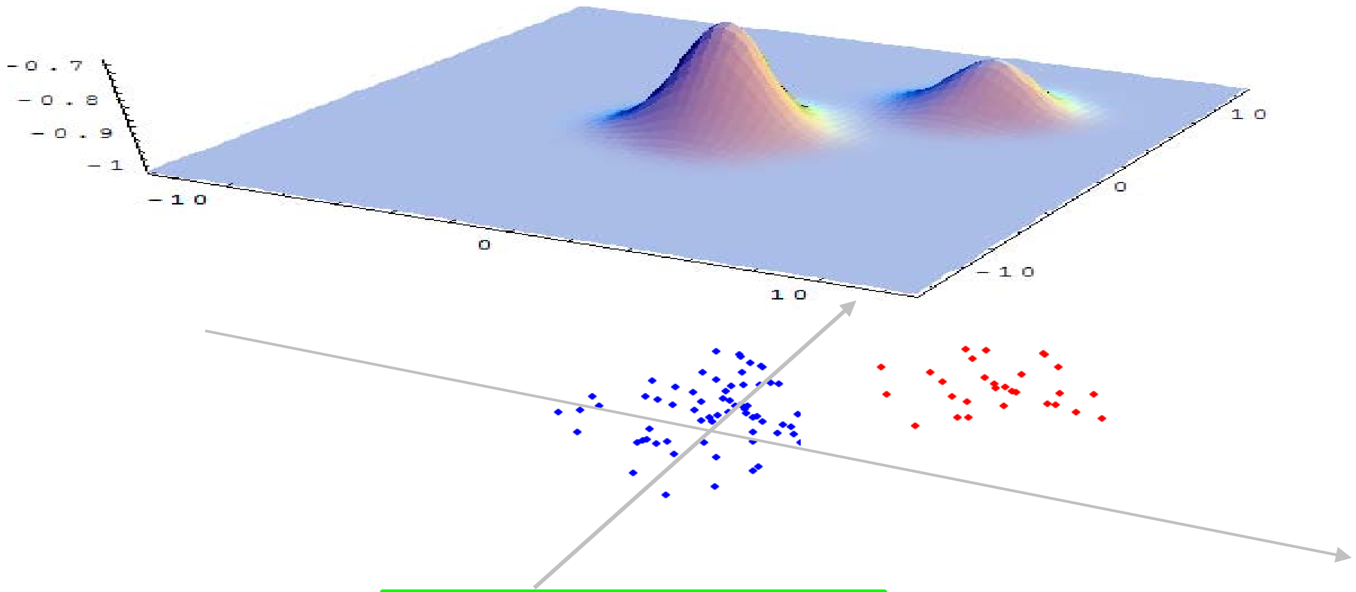
consistency $\hat{\theta}^{(\Phi)} \to \theta$ if $f = f_\theta$

Estimation selection, Robustness, Spontaneous data learning

Non-convex learning

Model $$f_\theta(x) = \frac{1}{2\pi^{d/2}} \exp\{-\frac{1}{2}(x-\theta)^T(x-\theta)\}$$

Loss function $$L^{(\Phi)}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} \exp\{-\frac{\gamma}{2}(x_i-\theta)^{\mathrm{T}}(x_i-\theta)\}$$



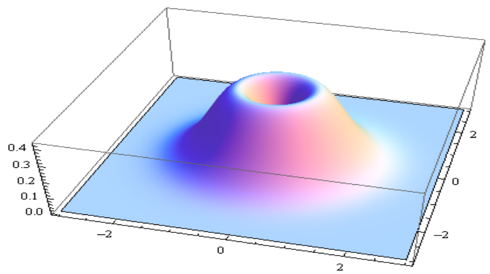
Selection of γ

Spontaneous data learning

The consistent class of estimators $\{\hat{\theta}^{(\Phi)} \text{ with } L^{(\Phi)}(\theta)\}_{\Phi}$

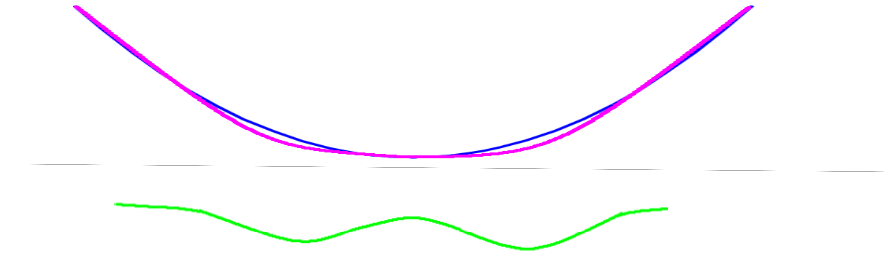
Super robustness

Redescending Influence function



Multi modal detection

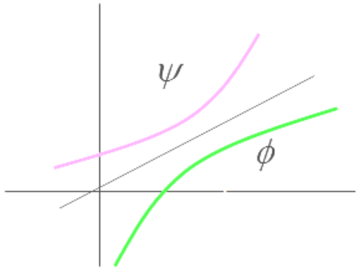
Difference of convex functions



two local minima

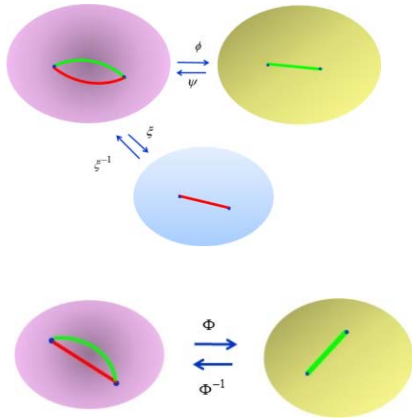
Estimation selection than **model selection**

Concluding remarks



$(\phi\,,\,\psi)$ → $D^{(\phi)}(f,g)$

$(\phi\,,\,\psi)$ → $\Delta^{(\phi)}(f,g)$



	expectation	metric
$D^{(\phi)}$	$\mathrm{E}_{\xi(f)}$	$\propto G$
$\Delta^{(\phi)}$	E_{f}	$\not\propto G$

$$(\phi,\psi)=(\log,\exp)\ \Rightarrow \Delta^{(\phi)}=D^{(\phi)}=D$$

Τηανκ ψου

Welcome to any comments:

eguchi@ism.ac.jp

U-divergence

Let U be a function satisfying $U(s)=\int_0^s \psi(t)dt$

_U_-cross-entropy $C_U(f,g)=\int \{-f\,\phi(g)+U(\phi(g))\}d\mu$

_U_-entropy $H_U(f)=C_U(f,f)$

_U_-divergence $D_U(f,g)=C_U(f,g)-H_U(f)$

Note $C_U(f,g)\geq H_U(f)$ or $D_U(f,g)\geq 0$

Exm Let $U_\beta(s)=\frac{1}{1+\beta}(1-\beta s)^{\frac{1+\beta}{\beta}}.$

Then power entropy $C_{U_\beta}(f,g)=-\frac{1}{\beta}\int fg^\beta \mathrm{d}\mu+\frac{1}{\beta+1}\int g^{\beta+1}\mathrm{d}\mu$

beta divergence $D_{U_\beta}(f,g)=\frac{1}{\beta}\int f(f^\beta-g^\beta)\mathrm{d}\mu+\frac{1}{\beta+1}\int (f^{\beta+1}-g^{\beta+1})\mathrm{d}\mu$

Generalized exponential model (ver. 2)

Let us fix a canonical statistic $\boldsymbol{t} = (t_1, \cdots, t_K)$.

G-exponential model

$$M^{(e\Phi)} = \{ f_\theta^{(e\Phi)}(x) := \Phi^{-1}(\theta^{\mathrm{T}} t(x) - \kappa^{(\Phi)}(\theta)) : \theta \in \Theta \}$$

where $\kappa^{(\Phi)}(\theta)$ is a normalizing constant.

Mean parameter $\boldsymbol{\eta} = \mathrm{E}_{f_\theta^{(e\Phi)}} \{ \boldsymbol{t}(\boldsymbol{X}) \} = \frac{\partial}{\partial \boldsymbol{\theta}} \kappa^{(\Phi)}(\boldsymbol{\theta})$

Remark 1 θ and η are affine parameters wrt $\Gamma^{(e\Phi)}$ and $\Gamma^{(m)}$.

($M^{(e\Phi)}$ is dually flat.)

Remark 2 $M^{(e\Phi)}$ is totally gedesic wrt $\Gamma^{(e\Phi)}$.