Path connectedness on a space of probability density functions

Osamu Komori[1], Shinto Eguchi[2]
University of Fukui[1], Japan
The Institute of Statistical Mathematics[2], Japan

Ecole Polytechnique, Paris-Saclay (France)
October 28, 2015

Contents

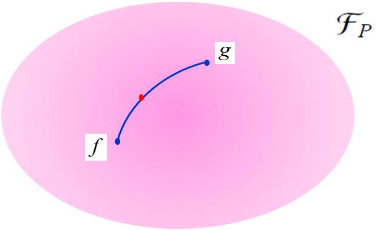
1. Kolmogorov-Nagumo (K-N) average
2. parallel displacement $A_t^{(\phi)}$ characterizing ϕ-path
3. U-divergence and its associated geodesic

Setting

Terminology

$\mathcal{X}$: data space
P : probability measure on $\mathcal{X}$
$\mathcal{F}_P$: space of probability density functions associated with P

We consider a path connecting f and g, where $f, g \in \mathcal{F}_P$, and investigate the property from a viewpoint of information geometry.



Kolmogorov-Nagumo (K-N) average

Let $\phi : (0, \infty) \to \mathbb{R}$ be an monotonic increasing and concave continuous function. Then for f and g in $\mathcal{F}_p$

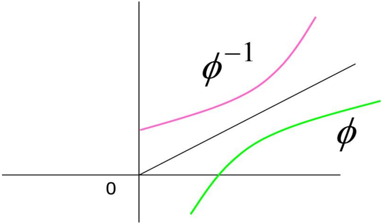
The Kolmogorov-Nagumo (K-N) average

$$\phi^{-1}\Big((1-t)\phi(f(x)) + t\phi(g(x))\Big)$$

for $0 \le t \le 1$.

Remark 1

ϕ^{-1} *is monotone increasing, convex and continuous on* $(0, \infty)$



ϕ-path

Based on K-N average, we consider ϕ-path connecting f and g in $\mathcal{F}_P$:

ϕ-path

$$f_t(x, \phi) = \phi^{-1}\Big((1 - t)\phi(f(x)) + t\phi(g(x)) - \kappa_t\Big),$$

where $\kappa_t \leq 0$ is a normalizing factor, where the equality holds if $t = 0$ or $t = 1$.

Existence of κ_t

Theorem 1

There uniquely exists κ_t such that

$$\int_X \phi^{-1}\Big((1-t)\phi(f(x)) + t\phi(g(x)) - \kappa_t\Big)dP(x) = 1$$

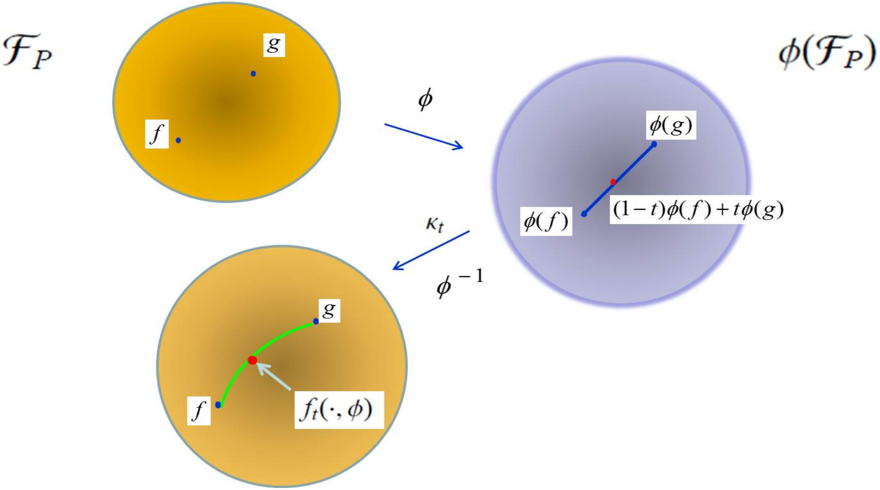
Proof

From the convexity of ϕ^{-1}, we have

$$0 \le \int \phi^{-1}\Big((1-t)\phi(f(x)) + t\phi(g(x))\Big)dP(x) \le \int \{(1-t)f(x) + tg(x)\}dP(x) \le 1$$

And we observe that $\lim_{c\to\infty} \phi^{-1}(c) = +\infty$ since ϕ^{-1} is monotone increasing. Hence the continuity of ϕ^{-1} leads to the existence of κ_t satisfying the equation above.

Illustration of ϕ-path



Examples of ϕ-path

Example 1

1. $\phi_0(x) = \log(x)$. *The* ϕ_0*-path is given by*

$$f_t(x, \phi_0) = \exp((1-t)\log f(x) + t \log g(x) - \kappa_t),$$

where $\kappa_t = \log \int \exp((1-t)\log f(x) + t\log g(x))dP(x)$.

2. $\phi_\eta(x) = \log(x+\eta)$ *with* $\eta \geq 0$. *The* ϕ_η*-path is given by*

$$f_t(x, \phi_\eta) = \exp\left[(1-t)\log\{f(x)+\eta\} + t\log\{g(x)+\eta\} - \kappa_t\right],$$

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

3. $\phi_\beta(x) = (x^\beta - 1)/\beta$ *with* $\beta \leq 1$. *The* ϕ_β*-path is given by*

$$f_t(x, \phi_\beta) = \{(1-t)f(x)^\beta + tg(x)^\beta - \kappa_t\}^{\frac{1}{\beta}},$$

where κ_t *does not have an explicit form.*

Contents

Extended expectation

For a function $a(x)$: $\mathcal{X} \to \mathbb{R}$, we consider

Extended expectation

$$\mathrm{E}_f^{(\phi)}\{a(X)\} = \frac{\displaystyle\int_{\mathcal{X}} \frac{1}{\phi'(f(x))} a(x) dP(x)}{\displaystyle\int_{\mathcal{X}} \frac{1}{\phi'(f(x))} dP(x)},$$

where ϕ: $(0, \infty) \to \mathbb{R}$ is a generator function.

Remark 2

If $\phi(t) = \log t$*, then* $\mathrm{E}^{(\phi)}$ *reduces to the usual expectation.*

Properties of extended expectation

We note that

1. $\mathrm{E}_f^{(\phi)}(c) = c$ for any constant c.
2. $\mathrm{E}_f^{(\phi)}\{ca(X)\} = c\mathrm{E}_f^{(\phi)}\{a(X)\}$ for any constant c.
3. $\mathrm{E}_f^{(\phi)}\{a(X) + b(X)\} = \mathrm{E}_f^{(\phi)}\{a(X)\} + \mathrm{E}_f^{(\phi)}\{b(X)\}$.
4. $\mathrm{E}_f^{(\phi)}\{a(X)^2\} \geq 0$ with equality if and only if $a(x) = 0$ for P-almost everywhere x in $\mathcal{X}$.

Remark 3

If we define $f^{(\phi)}(x) = 1/\phi'(f(x))/\int_{\mathcal{X}} 1/\phi'(f(x))dP(x)$, *then* $\mathrm{E}_f^{(\phi)}\{a(X)\} = \mathrm{E}_{f^{(\phi)}}\{a(X)\}$.

Tangent space of $\mathcal{F}_P$

Let H_f be a Hilbert space with the inner product defined by $\langle a, b\rangle_f = \mathrm{E}_f^{(\phi)}\{a(X)b(X)\}$, and the tangent space

Tangent space associated with extended expectation

$$T_f = \{a \in H_f : \langle a, 1\rangle_f = 0\}.$$

For a statistical model $M = \{f_\theta(x)\}_{\theta\in\Theta}$ we have

$$\mathrm{E}_{f_\theta}^{(\phi)}\{\partial_i \phi(f_\theta(X))\} = 0$$

for all θ of Θ, where $\partial_i = \partial/\partial\theta_i$ with $\theta = (\theta_i)_{i=1,\cdots,p}$. Further,

$$\mathrm{E}_{f_\theta}^{(\phi)}\{\partial_i\partial_j \phi(f_\theta(X))\} = \mathrm{E}_{f_\theta}^{(\phi)}\Big\{\frac{\phi''(f_\theta(X))}{\phi'(f_\theta(X))^2}\partial_i\phi(f_\theta(X))\partial_i\phi(f_\theta(X))\Big\}.$$

Parallel displacement $A_t^{(\phi)}$

Define $A_t^{(\phi)}(x)$ in T_{f_t} by the solution for a differential equation

$$\dot{A}_t^{(\phi)}(x) - \mathrm{E}_{f_t}^{(\phi)}\Big\{A_t^{(\phi)}\dot{f}_t\frac{\phi''(f_t)}{\phi'(f_t)}\Big\} = 0,$$

where f_t is a path connecting f and g such that $f_0 = f$ and $f_1 = g$. $\dot{A}_t^{(\phi)}(x)$ is the derivative of $A_t^{(\phi)}(x)$ with respect to t.

Theorem 2

The geodesic curve $\{f_t\}_{0\le t\le 1}$ by the parallel displacement $A_t^{(\phi)}$ is the ϕ-path.

Contents

U-divergence

Assume that $U(s)$ is a convex and increasing function of a scalar s and let $\xi(t) = \mathrm{argmax}_s\{st - U(s)\}$. Then we have

U-divergence

$$D_U(f,g) = \int \{U(\xi(g)) - f\xi(g)\}dP - \int \{U(\xi(f)) - f\xi(f)\}dP.$$

In fact, U-divergence is the difference of the cross entropy $C_U(f,g)$ with the diagonal entropy $C_U(f,f)$, where $C_U(f,g) = \int \{U(\xi(g)) - f\xi(g)\}dP$.

Connections based on U-divergence

For a manifold of finite dimension $M = \{f_\theta(x) : \theta \in \Theta\}$ and vector fields X and Y on M, the Riemannian metric is

$$G^{(U)}(X, Y)(f) = \int Xf\ Y\xi(f)dP$$

for $f \in M$ and linear connections $\nabla^{(U)}$ and $\nabla^{*(U)}$ are

$$G^{(U)}(\nabla_X^{(U)}Y, Z)(f) = \int XYf\ Z\xi(f)dP$$

and

$$G^{(U)}(\nabla_X^{*(U)}Y, Z)(f) = \int Zf\ XY\xi(f)dP.$$

See Eguchi (1992) for details.

Equivalence between ∇^*-geodesic and ξ-path

Let $\nabla^{(U)}$ and $\nabla^{*(U)}$ be linear connections associated with U-divergence D_U, and let $C^{(\phi)} = \{f_t(x,\phi) : 0 \le t \le 1\}$ be the ϕ path connecting f and g of $\mathcal{F}_P$. Then, we have

Theorem 3

A $\nabla^{(U)}$-geodesic curve connecting f and g is equal to $C^{(\mathrm{id})}$, where id denotes the identity function; while a $\nabla^{*(U)}$-geodesic curve connecting f and g is equal to $C^{(\xi)}$, where

$$\xi(t) = \mathrm{argmax}_s\{st - U(s)\}.$$

Summary

1. We consider ϕ-path based on Kolmogorov-Nagumo average.
2. The relation between U-divergence and ϕ-path was investigated (ϕ corresponds to ξ).
3. The idea of ϕ-path can be applied to probability density estimation as well as classification problems.
4. Divergence associated with ϕ-path can be considered, where a special case would be Bhattacharyya divergence.