

Contrast from one probability measure to another : the associated geometry

広島大 理学部 江口真透

(Shinto Eguchi)

1. Introduction

情報, エントロピー, ダイバージェンス, エネルギー等。
考えは ランダム性を扱う数理の中で しばしば 重要な
役割をする。本稿の目的は 統計的推論の観点から, それら
の考えを定量化する “コントラスト汎関数” のクラスを考察
する事にある。二つのランダムな現象が 確率分布 P, Q で
表わされとせよ。この二現象間に生じる 上述の考えを
表わす尺度として $\rho(P, Q)$ に次の要請をする。

$$(1) \quad \rho(P, Q) \geq 0$$

$$(2) \quad \rho(P, Q) = 0 \iff P = Q$$

この ρ を コントラスト汎関数と呼ぶ。ここに $\rho(P, Q)$
は P から Q への尺度であることに注意する。即ち 対称性
 $\rho(Q, P) = \rho(P, Q)$ を課さない事が このクラスを多様なも
のにしている事があとで示される。

2. Contrast functionals

測度空間 $(\mathcal{X}, \mathcal{B})$ 上の σ 有限な測度 μ を固定する.

μ に同値な有限測度全体を $\mathcal{M}_f$, 確率測度全体を $\mathcal{P}$ で表わす. 関数 $W(0, \infty) \rightarrow \mathbb{R}$ が

$$(3) \quad W(t) > W(1) \quad \forall t > 0, t \neq 1$$

を満たすとする. 以後 $W(1) = 0$ とおく. (3) を満たす関数全体を $\mathcal{W}$ と書く. $P, Q, R \in \mathcal{P}$ に対して

$$\phi_W(P, Q, R) \equiv \int W\left(\frac{dQ}{dP}\right) dR$$

と定め, 更に

$$d_W(P, Q) = \phi_W(P, Q, R_0)$$

$$\delta_W(P, Q) = \phi_W(P, Q, Q)$$

$$\rho_W(P, Q) = \phi_W(P, Q, P) \quad (\text{see [2]})$$

と表わす. ここで dQ/dP は Q の P に関する R -N 導関数とする. 定義から d_W, δ_W, ρ_W は いづれも コントラスト汎関数である. 以後, 定数倍の自由性を除くために $W''(1) = 1$ と規格化する. 次の例がある.

(i) Kullback-Leibler $\rho_{KL}(P, Q) = \int \left(\log \frac{dQ}{dP}\right) dP$

(ii) Squared Hellinger $\rho_{H^2}(P, Q) = 4 \int \left(1 - \sqrt{\frac{dQ}{dP}}\right)^2 dP$

(iii) Chernoff of order α $\rho_\alpha(P, Q) = \frac{1}{1-\alpha^2} \int \left(1 - \left(\frac{dQ}{dP}\right)^{\frac{1+\alpha}{2}}\right) dP$

(iv) exponential $\rho_e(P, Q) = \frac{1}{2} \int \left(\log \frac{dQ}{dP}\right)^2 dP$

(v) Kagan $\rho_K(P, Q) = \frac{1}{2} \int \left(1 - \frac{dQ}{dP}\right)^2 dP$

(ここで積分の前の定数は $W'(1)=1$ のための規格化による.)

上例を関係づけるコントラストが [4] で導入された:

$$\rho_{\alpha, \beta}(P, Q) = \frac{2}{(1-\alpha)(1-\beta)} \int \left(1 - \left(\frac{dQ}{dP}\right)^{\frac{1-\alpha}{2}}\right) \left(1 - \left(\frac{dQ}{dP}\right)^{\frac{1-\beta}{2}}\right) dP.$$

命題 1. 次の関係がある.

$$\rho_{0,0} = \rho_{H^2}, \quad \rho_{-\alpha, \alpha} = \rho_{\alpha}$$

$$\lim_{\alpha \rightarrow 1} \rho_{\alpha, \alpha} = \rho_K, \quad \lim_{\alpha \rightarrow -1} \rho_{\alpha, \alpha} = \rho_e$$

命題 2. W_1 と W_2 が (1) をみたす解析関数と仮定する.

$$W_1 = W_2 \iff d_{W_1} = d_{W_2} \iff \delta_{W_1} = \delta_{W_2}$$

$$\iff \rho_{W_1} = \rho_{W_2}$$

$\mathcal{W}$ 上の変換 $\circ, *, +$ を

$$W^{\circ}(t) = t^{-1} W(t^{-1})$$

$$W^*(t) = t W(t^{-1})$$

$$W^+(t) = t W(t)$$

と定義する. この時

$$\delta_{W^{\circ}}(P, Q) = \delta_W(Q, P)$$

$$\rho_{W^*}(P, Q) = \rho_W(Q, P)$$

$$d_{W^+}(P, Q) = d_W(Q, P)$$

が成立する. 更に $\rho_{W^+} = \delta_W$ より ρ_W のクラス
は δ_W のクラス一致する.

$\mathcal{W}$ の部分クラス $\mathcal{W}_1 = \{W \in \mathcal{W} \mid W: \text{convex on } (0, \infty)\}$

上に 変換 $\oplus, \ominus$ を

$$W^\oplus(t) = tW(t) - 2 \int_1^t W(s) ds$$

$$W^\ominus(t) = t^2 W(t) + 2 \int_0^+ s^{-2} W(s) ds + 2 \int_1^t \int_1^s u^2 W(u) du ds$$

と定める.

命題 3. $\oplus: \mathcal{W}_1 \rightarrow \mathcal{W}_1$ の逆変換は $\ominus$ である.

$$\lim_{n \rightarrow \infty} W^{\oplus n}(t) = W_0(t)$$

$$\lim_{n \rightarrow \infty} W^{\ominus n}(t) = W_0^*(t)$$

ここで $W_0(t) = 0$ if $0 < t \leq 1$, ∞ otherwise. $\square$

同様の結果として

$$W^* \ominus \oplus = W^{\ominus} \oplus = W \quad \text{が得られる.}$$

ρ_W, δ_W は 定義域を $\mathcal{P} \times \mathcal{P}$ から $\mathcal{M} \times \mathcal{M}$ に拡張して考えられる. $\mathcal{M}$ 上の同値関係を

$$m_1 \sim m_2 \iff \exists C > 0 \text{ s.t. } m_1(B) = C m_2(B) \quad \forall B \in \mathcal{B}$$

と定める. この時,

定義. $m_1 \sim m_2$, $n_1 \sim n_2$ の時

$$\rho_W(m_1, m_2) = \rho_W(n_1, n_2)$$

を満たす時, ρ_W を スケール不変という.

定理 4. ρ_W が スケール不変である同値条件は

$$\exists \alpha \in \mathbb{R} \text{ s.t. } \rho_W = \rho_\alpha$$

ここで ρ_α は 例(iii)で定義されたもの.

3. The associated geometry

ρ を 与 へ る コ ン ト ラ ス ト 関 数 と す る . $\mathcal{P}$ 上 の
有 限 次 元 部 分 多 様 体 を M と す る . M が 座 標 系
 $(\theta^1, \dots, \theta^k)$ に 対 し て 局 所 的 に

$$M = \{ P_\theta \in \mathcal{P} : \theta \in \mathcal{H} \}$$

と 表 わ さ れ る と す る . こ こ で $\mathcal{H}$ は $\mathbb{R}^k$ の 開 部 分 集 合 .

こ の 時 は ρ は 次 で Riemann 計 量 $g^{(\rho)}$ と 一 組 の ア フ ィ ン
接 続 $\Gamma^{(\rho)}$ と $*\Gamma^{(\rho)}$ を 導 く : 座 標 系 $(\theta^i)_{i=1,2,\dots,k}$ に 対 し て

$$g_{ij}^{(\rho)}(\theta) = \left(-\frac{\partial^2}{\partial \theta^i \partial \theta^j} \rho(P_{\theta_1}, P_{\theta_2}) \right)_{\theta_1 = \theta_2 = \theta},$$

$$\Gamma_{ij,k}^{(\rho)}(\theta) = \left(-\frac{\partial^3}{\partial \theta^i \partial \theta^j \partial \theta^k} \rho(P_{\theta_1}, P_{\theta_2}) \right)_{\theta_1 = \theta_2 = \theta},$$

$$*\Gamma^{(\rho)}(\theta) = \left(-\frac{\partial^3}{\partial \theta^i \partial \theta^j \partial \theta^k} \rho(P_{\theta_2}, P_{\theta_1}) \right)_{\theta_1 = \theta_2 = \theta}.$$

と , そ の 各 々 を 成 分 と 定 め る . (see [3])

$\rho(P_{\theta_1}, P_{\theta_2})$ は $\theta_1 = \theta_2$ の 時 最 小 値 0 を 取 る
こ と か ら 容 易 に $g^{(\rho)}, \Gamma^{(\rho)}, *\Gamma^{(\rho)}$ が 座 標 変 換 の
規 則 を 満 た す こ と が 示 さ れ る .

定 理 5. $g^{(\rho)}$ に 関 す る 計 量 接 続 を Γ_0 と
表 わ す . こ の 時 ,

$$(i) \quad \frac{1}{2}(\Gamma^{(\rho)} + *\Gamma^{(\rho)}) = \Gamma_0$$

$$(ii) \quad *\Gamma^{(\rho)} - \Gamma^{(\rho)} \text{ は 対 称 ラ ン ソ ー ル}$$

が 成 立 す る .

この事から $\Gamma^{(p)}$ と $*\Gamma^{(p)}$ は 次の意味で共役性を持つことが示される. (cf. e.g. [4])

π と π^* を 各々, $\Gamma^{(p)}$ と $*\Gamma^{(p)}$ による平行移動とせよ. この時,

$$g^{(p)}(\pi X, \pi^* X) = g^{(p)}(X, X) \quad \forall X, X \in T(M)$$

が成立する. 三つ一組にして

$$\mathcal{C}(p) = (g^{(p)}, \Gamma^{(p)}, *\Gamma^{(p)})$$

を M 上の p の共役構造と呼ぶ.

注意. p が方称ならば 上の共役性は trivial とする.
(即ち, $\Gamma^{(p)} = *\Gamma^{(p)} = \Gamma_0$). $\rho(p, q)$ を p と q を結ぶ測地線の長さの二乗と定義する. この時 $\Gamma^{(p)}$ は 計量接続に還元される.

以上のように 与える p は $\mathcal{C}(p)$ を連想することを見たが, 逆に $\mathcal{C}(p)$ によって p 全体のクラスを分類すること試める. そのためには 次の準備が必要である.

Amari [1] は 統計的見地から Riemann g と一組のアソシエート接続 $\overset{m}{\Gamma}$ と $\overset{e}{\Gamma}$ を 次で導入した.

$$g_{ij}(\theta) = F_0 e_i e_j$$

$$\overset{e}{\Gamma}_{ij,k}(\theta) = F_0 \partial_i e_j e_k$$

$$\overset{m}{\Gamma}_{ij,k}(\theta) = \overset{e}{\Gamma}_{ij,k} + F_0 e_i e_j e_k$$

ここで $c_0 = \partial_i \log \frac{dP_0}{d\mu}$, E_0 は P_0 での期待値を表わす.

以上をまとめて, $C_S = (g, \Gamma^m, \Gamma^e)$ と書く. 更に

Γ^m と Γ^e を結合する ワン パラメータ-変換

$$\Gamma_\alpha = \frac{1-\alpha}{2} \Gamma^m + \frac{1+\alpha}{2} \Gamma^e$$

に対して $C_\alpha = (g, \Gamma_\alpha, \Gamma_\alpha)$ と置く. 定義より

$C_{\alpha=1} = C_S$ に注意する.

命題 6. W を (1) を満たす C^3 -関数とする. この時

$$C(P_W) = C_\alpha, \quad \alpha = 2W'''(1) + 3.$$

$$C(\delta_W) = C_\beta, \quad \beta = 2W'''(1) + 9$$

特に

$$C(P_\alpha) = C_\alpha, \quad C(P_{KL}) = C_S$$

が成立する. $\square$

この命題より

$$\{P_W : W \in \mathcal{W} \cap C^3\} \subset \{P : C(P) \in A\}$$

が言える. ここで $A = \{C_\alpha : \alpha \in \mathbb{R}\}$.

一方で d_W は少し異なる構造を持つ

$$g^{(d_W)}_{ij} = E_0 e_i e_j$$

$$\Gamma^{(d_W)}_{ij,k} = E_0 \partial_i e_j e_k - [W'''(1) + 3] E_0 e_i e_j e_k$$

ここで E_0 は P_0 での期待値を表わす.

参考文献

- [1] Amari, S. (1986) Differential-geometrical methods in Statistics. Lecture Note in Statistics 28 Springer, New York.
- [2] Csiszar (1967) On topological properties of f -divergences. Studia Math. Hung 2. 239-339.
- [3] Eguchi, S (1985) A differential Geometric Approach to Statistical Inference on the basis of contrast functionals. Hiroshima Math. J. 1341.-381
- [4] Kobayashi, S & Nomizu, K. (1969) Foundations of differential geometry Vol 2, Interscience, New York.