数理解析研究所講究録
623巻 1987年 79-86

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，エネルギー等の考えは 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． この二つ現象間に生じる 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) $\rho(P,Q) \geq 0$

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

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 $\mathcal{M}$, 確率測度全体を $\mathcal{P}$ で表わす. 関数 $W(0,\infty) \to \mathbb{R}$ が

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

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 $\quad \rho_{KL}(P, Q) = \int \left(\log \frac{dP}{dQ}\right) dP$

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

(iii) Chernoff of order α $\quad \rho_\alpha(P, Q) = \frac{4}{1-\alpha^2} \int \left(1 - \left(\frac{dQ}{dP}\right)^{\frac{1+\alpha}{2}}\right) dP$

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

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

（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.

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

$$\lim_{\alpha\to 1}\rho_{\alpha,\alpha}=\rho_{K}, \qquad \lim_{\alpha\to -1}\rho_{\alpha\alpha}=\rho_{e}$$

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

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

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

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

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

$$w^{\oplus}(t) = t\,w(t) - 2\int_1^t w(s)\,ds$$

$$w^{\ominus}(t) = t^{-1}w(t) + 2\int_0^t s^{-2}w(s)\,ds + 2\int_1^t\int_1^s u^{-2}w(u)\,du\,ds$$

と定める.

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

$$\lim_{n\to\infty} w^{\oplus^n}(t) = w_0(t)$$

$$\lim_{n\to\infty} w^{\ominus^n}(t) = w_0^*(t)$$

ここで $w_0(t) = 0$ if $0 < t \leq 1$, ∞ otherwise. □

同様の結果として

$$w^{*\ominus*\ominus} = w^{\ominus*\ominus*} = w$$ が得られる.

ρ_w, δ_w は定義域を $\mathcal{P}\times\mathcal{P}$ から $\mathcal{M}\times\mathcal{M}$ へ拡張して考えられる. $\mathcal{M}$ 上の同値関係を

$$m_1 \sim m_2 \overset{\text{def}}{\Longleftrightarrow} \exists c > 0 \text{ s.t. } m_1(B) = c\,m_2(B) \ \forall B \in \mathfrak{B}$$

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

$\exists \alpha \in \mathbb{R}$ 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 かってな コントラスト汎関数とする. $\mathcal{P}$上の有限次元部分多様体を M とする. Mが 座標系 $(\theta^1, \cdots, \theta^k)$ に対して局所的に

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

と表わされたとする. ここで Θ 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 $(\theta^i)_{i=1,2,\cdots,k}$ に対して

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

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

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

と, それぞれを成分を定める. (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) $\frac{1}{2}(\Gamma^{(\rho)} + {}^*\Gamma^{(\rho)}) = \Gamma_0$

(ii) ${}^*\Gamma^{(\rho)} - \Gamma^{(\rho)}$ は対称テンソル

が成立する.

この事から $\Gamma^{(\rho)}$ と ${}^*\Gamma^{(\rho)}$ 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, $\Gamma^{(\rho)}$ と ${}^*\Gamma^{(\rho)}$ による平行移動とせよ. この時,

$$g^{(\rho)}(\Pi X, \Pi^* Y) = g^{(\rho)}(X, Y) \quad \forall X, Y \in T(M)$$

が成立する. 三つ一組にして

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

を M上の ρ の共役構造と呼ぶ.

注意. ρ が対称ならば 上の共役性は trivial となる.（即ち, $\Gamma^{(\rho)} = {}^*\Gamma^{(\rho)} = \Gamma_0$）. $\rho_0(P, Q)$ を P と Q を結ぶ測地線の長さの2乗と定義する. この時 $\Gamma^{(\rho_0)}$ は計量接続に還元される.

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

Amari [1] は統計的見地から Riemann g と一組のアフィン接続 $\overset{m}{\Gamma}$ と $\overset{e}{\Gamma}$ を次で導入した.

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

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

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

ここで $e_\theta = \partial_i \log \frac{dP_\theta}{d\mu}$, E_θ は P_θ での期待値を表わす.

以上まとめて, $\mathcal{C}_S = (g, \overset{m}{\Gamma}, \overset{e}{\Gamma})$ と書く. 更に

$\overset{m}{\Gamma}$ と $\overset{e}{\Gamma}$ を結ぶ ワンパラメーター族

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

に対して $\mathcal{C}_\alpha = (g, \Gamma_\alpha, \Gamma_{-\alpha})$ と置く. 定義より

$\mathcal{C}_{\alpha=-1} = \mathcal{C}_S$ 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

$$\mathcal{C}(\rho_W) = \mathcal{C}_\alpha , \quad \alpha = 2W'''(1)+3$$

$$\mathcal{C}(\delta_W) = \mathcal{C}_\beta , \quad \beta = 2W'''(1)+9$$

特に

$$\mathcal{C}(\rho_\alpha) = \mathcal{C}_\alpha , \quad \mathcal{C}(\rho_{KL}) = \mathcal{C}_S$$

が成立する. □

この命題より

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

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

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. 1 341-391

[4] Kobayashi, S & Nomizu, K. (1969) Foundations of differential geometry Vol 2, Interscience, New York.