

§15. 指数型分布族の場合

§13において扱ったように, 指数型分布族は Fisher 計量および e 接続, m 接続を考えると双対平坦空間となった. 指数型分布族の双対構造についてもう少し詳しく見てみよう.

S を指数型分布族とし,

$$S = \{p(x; \theta) | \theta \in \Theta\},$$

$$p(x; \theta) = \exp \left(C(x) + \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right) \quad (x \in \Omega, \theta = (\theta^1, \theta^2, \dots, \theta^n)) \quad (*)$$

と表しておく. このとき, §10において扱ったように, 自然座標系 θ は e 接続 $\nabla^{(e)} = \nabla^{(1)}$ に関するアファイン座標系であった.

まず, g を S の Fisher 計量とし, g に関する θ の双対座標系を求めよう.

$i = 1, 2, \dots, n$ とすると, $(*)$ より,

$$\frac{\partial}{\partial \theta^i} p(x; \theta) = \left(F_i(x) - \frac{\partial \psi}{\partial \theta^i}(\theta) \right) p(x; \theta)$$

だから,

$$\int_{\Omega} p(x; \theta) dx = 1$$

に注意して, 両辺を Ω 上で積分すると,

$$0 = E_{\theta}[F_i] - \frac{\partial \psi}{\partial \theta^i}.$$

更に,

$$g_{ij} = g \left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j} \right) \quad (i, j = 1, 2, \dots, n)$$

とおくと, $(*)$ より,

$$\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log p(x; \theta) = - \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\theta)$$

だから,

$$\int_{\Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x; \theta) \right) p(x; \theta) dx = 0$$

に注意すると,

$$\begin{aligned} g_{ij} &= \int_{\Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x; \theta) \right) \left(\frac{\partial}{\partial \theta^j} \log p(x; \theta) \right) p(x; \theta) dx \\ &= \int_{\Omega} \left(\frac{\partial}{\partial \theta^i} \log p(x; \theta) \right) \left(\frac{\partial}{\partial \theta^j} p(x; \theta) \right) dx \\ &= - \int_{\Omega} \left(\frac{\partial^2}{\partial \theta^i \partial \theta^j} \log p(x; \theta) \right) p(x; \theta) dx \\ &= \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}. \end{aligned}$$

よって,

$$\eta_i = E_{\theta}[F_i], \quad \eta = (\eta_1, \eta_2, \dots, \eta_n)$$

とおくと,

$$\frac{\partial \eta_i}{\partial \theta^j} = g_{ij}$$

だから, η は S 上の局所座標系である. 更に, θ と η は g に関して互いに双対的であることも分かる. η を期待値座標系という. $\nabla^{(e)}$ の双対接続は m 接続 $\nabla^{(m)} = \nabla^{(1)}$ だから, η は $\nabla^{(m)}$ に関するアファイン座標系となる.

次に, Legendre 変換

$$d\psi = \sum_{i=1}^n \eta_i d\theta^i, \quad d\varphi = \sum_{i=1}^n \theta^i d\eta_i, \quad \psi + \varphi = \sum_{i=1}^n \theta^i \eta_i$$

をあたえるポテンシャル ψ, φ を求めよう.

まず, 上の計算より, ポテンシャル ψ は (*) における ψ と同じものとして選ぶことができる.

このとき, (*) より,

$$\log p(x; \theta) = C(x) + \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \quad (**)$$

だから,

$$\begin{aligned} \varphi &= \sum_{i=1}^n \theta^i \eta_i - \psi \\ &= E_\theta[\log p - C] \end{aligned}$$

である.

更に, $p, q \in S$ とし, q は $q = p(x; \theta')$ のように表すことにすると, (**) より, 正準ダイバージェンス D は

$$\begin{aligned} D(p||q) &= \psi(\theta) + \varphi(\theta') - \sum_{i=1}^n \theta^i \eta'_i \\ &= \psi(\theta) + E_{\theta'}[\log q - C] - \sum_{i=1}^n \theta^i E_{\theta'}[F_i] \\ &= E_{\theta'}[\log q - \log p] \\ &= \int_{\Omega} q \log \frac{q}{p} dx \end{aligned}$$

と計算することができる.

よって, §12 において扱ったことより, D は Kullback-Leibler ダイバージェンス $D^{(1)}$ に一致する. 最後に, 統計学の基本定理である Cramér-Rao の不等式との関係について述べよう.

統計的モデルのパラメータを観測データから推定する問題を考えると, 推定される値が真の値からどの程度ばらつくのかを評価することができる. その評価式が Cramér-Rao の不等式であり, 推定の精度は Fisher 情報行列の逆行列が小さいほど良くなることが分かる.

Ω を高々可算集合または $\mathbf{R}^k$ とし, $\mathcal{F}$ を Ω が高々可算集合のときは Ω の部分集合全体で, Ω が $\mathbf{R}^k$ のときは $\mathbf{R}^k$ の Borel 集合族であるとする.

また,

$$S = \{p(x; \xi) | \xi \in \Xi\}$$

を Ω 上の n 次元統計的モデルとし, $p(x; \xi)$ が可測空間 $(\Omega, \mathcal{F})$ に値をとる確率変数 X_ξ に対する確率関数または密度関数であるとする.

更に, $\hat{\xi}$ を $\mathbf{R}^n$ に値をとる確率空間 $(\Omega, \mathcal{F}, \mu_{X_\xi})$ 上の確率変数とする.

統計的推定では $\hat{\xi}$ を観測することにより, 分布 $p(x; \xi)$ のパラメータ ξ を推定する. そこで, 次のように定める.

定義 任意の $\xi \in \Xi$ に対して

$$E_{\xi}[\hat{\xi}] = \xi,$$

すなわち $\hat{\xi}$ を

$$\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_n)$$

と表しておいたときに

$$\int_{\Omega} \hat{\xi}_i(x) p(x; \xi) dx = \xi_i \quad (i = 1, 2, \dots, n)$$

がなりたつとき, $\hat{\xi}$ を ξ の不偏推定量という.

$\hat{\xi}$ を ξ の不偏推定量とし, 確率関数または密度関数 $p(x; \xi)$ の定める $\hat{\xi}$ の分散行列を $\text{Var}_{\xi}[\hat{\xi}]$ と表すことにする. すなわち,

$$\text{Var}_{\xi}[\hat{\xi}] = E_{\xi} \left[{}^t (\hat{\xi} - E_{\xi}[\hat{\xi}]) (\hat{\xi} - E_{\xi}[\hat{\xi}]) \right]$$

である. このとき, 次がなりたつ.

Cramér-Rao の不等式 $G(\xi)$ を ξ における S の Fisher 情報行列とすると,

$$\text{Var}_{\xi}[\hat{\xi}] - G(\xi)^{-1}$$

は半正定値.

証明 まず, $v_1, v_2, \dots, v_n \in \mathbf{R}^n$ とすると, 分散行列は常に半正定値であるから,

$$\hat{\xi} - \sum_{i=1}^n v_i \frac{\partial}{\partial \xi_i} \log p(x; \xi)$$

の分散行列も半正定値であることに注意する.

特に, e_i を第 i 成分が 1 の単位ベクトルとし,

$$v_i = e_i G(\xi)^{-1}$$

とおくと, 上の分散行列は

$$\text{Var}_{\xi}[\hat{\xi}] - G(\xi)^{-1}$$

となることが分かる. □

Cramér-Rao の不等式を不等式という理由は, 一般に同じ型の 2 つの正方行列に対して $A - B$ が半正定値であることを $A \geq B$ と表すからである.

不偏推定量 $\hat{\xi}$ は Cramér-Rao の不等式の等号をあたえるとき, すなわち

$$\text{Var}_{\xi}[\hat{\xi}] = G(\xi)^{-1}$$

がなりたつとき, 有効推定量という. 有効推定量が存在するための必要十分条件について次がなりたつ.

定理 統計的モデル

$$S = \{p(x; \xi) | \xi \in \Xi\}$$

が有効推定量をもつことと S が指数型分布族で ξ が期待値座標系であることは同値.

関連事項 15. α アファイン多様体

§6 や §9 において扱った Fisher 計量や α 接続は統計的モデルに対する条件を緩めても考えることができる.

Ω を高々可算集合または $\mathbf{R}^k$ とし, $\mathbf{R}^n$ の開集合 Ξ を用いて

$$S = \left\{ p(x; \xi) \mid \text{任意の } x \in \Omega \text{ と任意の } \xi \in \Xi \text{ に対して } p(x; \xi) > 0 \text{ で, } \int_{\Omega} p(x; \xi) dx < +\infty \right\}$$

と表される関数の族 S を考える.

S の元 $p(x; \xi)$ の ξ に関する微分可能性については, 統計的モデルの場合と同様に仮定しておこう. このとき, S に対しても Fisher 計量および α 接続を定めることができる. すなわち, $i = 1, 2, \dots, n$ に対して

$$l_{\xi}(x) = \log p(x; \xi), \quad \partial_i = \frac{\partial}{\partial \xi_i}$$

とおき,

$$g_{ij} = E_{\xi}[\partial_i l_{\xi} \partial_j l_{\xi}] \quad (i, j = 1, 2, \dots, n)$$

とおくと, Fisher 情報行列 $G = (g_{ij})$ は S の Fisher 計量 g を定める. また, $\alpha \in \mathbf{R}$ に対して

$$\Gamma_{ij,k}^{(\alpha)} = E_{\xi} \left[\left(\partial_i \partial_j l_{\xi} + \frac{1-\alpha}{2} \partial_i l_{\xi} \partial_j l_{\xi} \right) (\partial_k l_{\xi}) \right]$$

とおくと, $\Gamma_{ij,k}^{(\alpha)}$ は S の α 接続 $\nabla^{(\alpha)}$ を定める.

ここで,

$$L^{(\alpha)}(u) = \begin{cases} \frac{2}{1-\alpha} u^{\frac{1-\alpha}{2}} & (\alpha \neq 1), \\ \log u & (\alpha = 1) \end{cases}$$

とおき, $p(x; \xi) \in S$ に対して

$$l^{(\alpha)}(x; \xi) = L^{(\alpha)}(p(x; \xi))$$

とおくと,

$$\partial_i l^{(\alpha)} = p^{\frac{1-\alpha}{2}} \partial_i l_{\xi}, \quad \partial_i \partial_j l^{(\alpha)} = p^{\frac{1-\alpha}{2}} \left(\partial_i \partial_j l_{\xi} + \frac{1-\alpha}{2} \partial_i l_{\xi} \partial_j l_{\xi} \right)$$

が得られるから,

$$g_{ij} = \int_{\Omega} \partial_i l^{(\alpha)} \partial_j l^{(-\alpha)} dx$$

および

$$\Gamma_{ij,k}^{(\alpha)} = \int_{\Omega} \partial_i \partial_j l^{(\alpha)} \partial_k l^{(-\alpha)} dx$$

がなりたつ.

ある α に対して

$$\partial_i \partial_j l^{(\alpha)} = 0 \quad (i, j = 1, 2, \dots, n)$$

がなりたつとき, S を α アファイン多様体という. S が α アファイン多様体のとき, ξ は $\nabla^{(\alpha)}$ に関するアファイン座標系で,

$$l^{(\alpha)}(x; \xi) = C(x) + \sum_{i=1}^n \xi_i F_i(x)$$

と表すことができる. 更に, $(S, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ は双対平坦空間となることも分かる.