振り返りと導入

前回は指数型分布族の具体例の計算を行った。本稿では次のことを行う:

- 一般化を念頭に置きながら、指数型分布族の双対構造の性質を整理する！！！
- [TODO]

1 双対構造

1.1 一般の多様体の場合

定義 1.1 (双対接続). (M,g) を Riemann 多様体、∇,∇' を M 上のアファイン接続とする。∇' が g に関する ∇ の**双対接続 (dual connection)** であるとは、すべての $X,Y,Z\in\mathfrak{X}(M)$ に対し

$$X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla'_X Z) \tag{1.1}$$

が成り立つことをいう。

命題 1.2. ∇ を M 上のアファイン接続とする。このとき、g に関する ∇ の双対接続がただひとつ存在する。

証明 [TODO] □

定義 1.3 (双対構造). (g,∇,∇^*) が M 上の**双対構造 (dualistic structure)** であるとは、∇ と ∇' が g に関し互いに双対接続であることをいう。

定義 1.4 (双対平坦). ∇,∇' がいずれも M 上平坦であるとき、(g,∇,∇') は**双対平坦 (dually flat)** であるという。双対平坦な双対構造を**双対平坦構造 (dually flat structure)** という。

たとえ双対平坦であっても、両方の接続係数が同一の座標のもとで消えるとは限らない。実際、それが成り立つためには次のような非常に強い条件が必要となる:

命題 1.5. (g,∇,∇') を双対構造、$x=(x_1,\ldots,x_m)$ を M の開集合 U 上の座標とする。このとき、x に関する ∇,∇' の接続係数が U 上つねに 0 ならば、g は U 上定数である。

証明 $\partial_i g_{jk} = \Gamma^l_{ij} g_{lk} + \Gamma^l_{ik} g_{lj}$ より従う。 □

1.2 指数型分布族の場合

指数型分布族の α-接続について考える。

命題 1.6 (曲率の AC テンソルによる表示). $\alpha\in\mathbb{R}$、$R^{(\alpha)}$ を $\nabla^{(\alpha)}$ の $(1,3)$-曲率テンソルとする。このとき、$\mathcal{P}$ の

任意の ∇-アファイン座標に関し [TODO] $\nabla^{(\alpha)}$ ではなく?、$R^{(\alpha)}$ の成分は

$$R^{(\alpha)}{}_{ijk}{}^{l} = -\frac{1-\alpha^2}{2}\left(A_{jk}{}^{m}A_{im}{}^{l} - A_{ik}{}^{m}A_{jm}{}^{l}\right) \tag{1.2}$$

となる。とくに $\alpha = \pm 1$ のとき $R^{(\alpha)} = 0$ となる。

証明 `0613_資料.pdf` 命題 1.12 の式

$$R^{(\alpha)}{}_{ijk}{}^{l} = \frac{1-\alpha}{2}\left(\partial_i A_{jk}{}^{l} - \partial_j A_{ik}{}^{l}\right) + \left(\frac{1-\alpha}{2}\right)^2 \left(A_{jk}{}^{m}A_{im}{}^{l} - A_{ik}{}^{m}A_{jm}{}^{l}\right) \tag{1.3}$$

を変形する。

$$\partial_i A_{jk}{}^{l} \tag{1.4}$$
$$= \partial_i (g^{ln} S_{jkn}) \tag{1.5}$$
$$= \partial_i (g^{ln}) S_{jkn} + g^{lm}\partial_i S_{jkm} \tag{1.6}$$
$$= -\partial_i (g_{mn}) g^{mn} g^{ln} S_{jkn} + g^{lm}\partial_i S_{jkm} \qquad (\partial_i (g_{nm} g^{lm}) = 0) \tag{1.7}$$
$$= -S_{imn} g^{mn} g^{ln} S_{jkn} + g^{lm}\partial_i S_{jkm} \qquad (\partial_i g_{mn} = S_{imn}) \tag{1.8}$$
$$= -A_{im}{}^{l} A_{jk}{}^{m} + g^{lm}\partial_i S_{jkm}. \tag{1.9}$$

同様にして

$$\partial_j A^{l}_{ik} = -A_{jm}{}^{l} A_{ik}{}^{m} + g^{lm}\partial_j S_{ikm}. \tag{1.10}$$

したがって命題の主張の式が得られた。 □

系 1.7. すべての $\alpha \in \mathbb{R}$ に対し

$$R^{(\alpha)} = R^{(-\alpha)} \tag{1.11}$$

が成り立つ。

証明 $\dfrac{1-\alpha}{2} = \dfrac{1+\alpha}{2} - \alpha$ および $\left(\dfrac{1-\alpha}{2}\right)^2 = \left(\dfrac{1+\alpha}{2}\right)^2 - \alpha$ が成り立つことより、命題から

$$R^{(\alpha)}{}_{ijk}{}^{l} = R^{(-\alpha)}{}_{ijk}{}^{l} - \alpha R^{(-1)}{}_{ijk}{}^{l} \tag{1.12}$$

となる。さらに $R^{(-1)} = 0$ だから系の主張が得られた。 □

定理 1.8. 任意の $\alpha \in \mathbb{R}$ に対し、3 つ組 $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ は $\mathcal{P}$ 上の双対構造となる。さらに、$\alpha = \pm 1$ ならば $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ は双対平坦である。[TODO] 逆はいえるか?

証明 双対構造であることは

$$g(\nabla^{(\alpha)}_X Y, Z) + g(Y, \nabla^{(-\alpha)}_X Z) = g(\nabla^{g}_X Y, Z) - \frac{1-\alpha}{2} S(X,Y,Z) + g(Y, \nabla^{g}_X Z) - \frac{1+\alpha}{2} S(X,Z,Y \tag{1.13}$$

$$
\begin{aligned}
&= g(\nabla^g_X Y, Z) + g(Y, \nabla^g_X Z) &(1.14)\\
&= X(g(Y,Z)) &(1.15)
\end{aligned}
$$

より従う。$\alpha = \pm 1$ で双対平坦となることは上の系よりわかる。 □

2 期待値パラメータ

補題 2.1. W を $\mathbb{R}$-ベクトル空間、$U \subset W$ を開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\operatorname{Hess} f$ が U 上各点で正定値であるものとする。このとき、$\nabla f\colon U \to W^\vee$ は単射である。

証明 [TODO] □

定義 2.2 (古典的な Legendre 変換)**.** W を $\mathbb{R}$-ベクトル空間、$U \subset W$ を開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\operatorname{Hess} f$ が U 上各点で正定値であるものとする。関数

$$
f^\vee\colon U' \to \mathbb{R}, \quad y \mapsto \langle y, (\nabla f)^{-1}(y)\rangle - f((\nabla f)^{-1}(y)) \quad \text{where} \quad U' := (\nabla f)(U) \tag{2.1}
$$

を f の**凸共役 (convex conjugate)** という。

例 2.3 (凸共役の例)**.**

- e^x (Poisson 分布族の実現の対数分配関数) $\to y \log y - y$
- $\log(1+e^x)$ (Bernoulli 分布族の実現の対数分配関数) $\to y\log y + (1-y)\log(1-y)$
- x^2 (分散既知の正規分布族の実現の対数分配関数) $\to y^2/4$

命題 2.4 (凸共役の性質)**.** [TODO]

証明 [TODO] 何が必要? □

命題 2.5. $(\nabla\psi)|_{\operatorname{Int}\widetilde{\Theta}}\colon \operatorname{Int}\widetilde{\Theta} \to V^{\vee\vee} = V$ は C^∞ 埋め込みかつ開写像である。

補題 2.6. $\operatorname{Int}\widetilde{\Theta}$ は $V^\vee$ の凸集合である。

証明 $\widetilde{\Theta}$ が $V^\vee$ の凸集合であることは `0425_資料.pdf` 命題 2.2 で確かめた。一般に、位相ベクトル空間の凸集合の内部は凸集合だから、補題が従う。 □

命題の証明 ψ は C^∞ だから $\nabla\psi$ も C^∞ である。また、$\nabla\psi$ $\operatorname{Hess}\psi$ は正定値だから $\nabla\psi$ の微分は全単射である。逆写像定理より $\nabla\psi$ は局所微分同相写像であり、とくに開写像である。また、補題より $\operatorname{Int}\widetilde{\Theta}$ は $V^\vee$ の凸集合だから、$\nabla\psi$ は単射である。したがって $\nabla\psi$ は埋め込みである。 □

命題-定義 2.7 (期待値パラメータ空間). $\mathcal{P}$ は開であるとし、(V, T, μ) を $\mathcal{P}$ の最小次元実現とする。このとき、集合

$$\mathcal{M} := \left\{ E_p[T] \in V \mid p \in \mathcal{P} \right\} \tag{2.2}$$

は V の開部分多様体となり、写像 $\eta \colon \mathcal{P} \to \mathcal{M},\ p \mapsto E_p[T]$ は微分同相写像となる。$\mathcal{M}$ を (V, T, μ) に関する $\mathcal{P}$ の**期待値パラメータ空間 (mean parameter space)** といい、η を**期待値パラメータ座標 (mean parameter coordinates)** という。

証明　まず $\mathcal{M} = (\nabla\psi)(\Theta)$ である。いま $\mathcal{P}$ は開だから Θ は $V^\vee$ の開集合である。このことと $\nabla\psi$ が開写像であることから $\mathcal{M}$ は V の開集合、したがって開部分多様体である。$\eta(p) = (\nabla\psi) \circ \theta(p)$ $(p \in \mathcal{P})$ が成り立つから、$(\nabla\psi), \theta$ が微分同相写像であることから η も微分同相写像である。 □

命題 2.8. $\mathcal{P}$ は開であるとし、ϕ を $\psi|_\Theta$ の凸共役とする。このとき次が成り立つ:

(1)

$$\frac{\partial\psi}{\partial\theta^i} = \eta_i, \qquad \frac{\partial\phi}{\partial\eta_i} = \theta^i. \tag{2.3}$$

(2)　θ-座標に関し

$$g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}, \qquad g^{ij} = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j}. \tag{2.4}$$

[TODO] g^{ij} は $T^\vee\mathcal{P}$ 上の計量を定める?

証明　(1)　$\dfrac{\partial\psi}{\partial\theta^i}(\theta(p)) = E_p[T^i] = \eta_i(p)$ である。[TODO]

(2)　$g_{ij}(\theta(p)) = \dfrac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\theta(p))$ は g の定義から明らか。[TODO] □

定理 2.9. 期待値パラメータ座標に $\mathcal{M}$ 上の任意の座標を合成したものは $\mathcal{P}$ 上の $\nabla^{(-1)}$-アファイン座標である。

証明

$$\Gamma^{(-1)}{}^{\alpha\beta}_{\gamma} = \frac{\partial\eta_\gamma}{\partial\theta^k}\left(\frac{\partial\theta^k}{\partial\eta_\alpha\partial\eta_\beta} + \Gamma^{(-1)}{}^k_{ij}\frac{\partial\theta^i}{\partial\eta_\alpha}\frac{\partial\theta^j}{\partial\eta_\beta}\right) \tag{2.5}$$

が 0 となることをいえばよい。 □

例 2.10 ($\nabla^{(-1)}$-測地線). [TODO]

今後の予定

- KL ダイバージェンス

参考文献

[Ama16] Shun-ichi Amari, **Information Geometry and Its Applications**, Applied Mathematical Sciences, vol. 194, Springer Japan, Tokyo, 2016 (en).

A 付録