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:

- 双対構造を定義し、とくに指数型分布族の α-接続の性質を調べる。
- Legendre 変換を定義する。
- 指数型分布族の期待値パラメータを定義する。

1 双対構造

定義 1.1 (双対構造). M を多様体とする。M 上の Riemann 計量 g とアファイン接続 ∇, ∇^* の組 (g, ∇, ∇^*) が M 上の**双対構造 (dualistic structure)** 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。このとき、∇, ∇^* はそれぞれ g に関する ∇^*, ∇ の**双対接続 (dual connection)** であるという。

さらに ∇, ∇^* 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。

命題 1.2 (双対接続の存在と一意性). M を多様体、g を M 上の Riemann 計量、∇ 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 証明は付録に記した。 □

指数型分布族の α-接続について考える。以降、$\mathcal{P}$ を可測空間 $\mathcal{X}$ 上の open な指数型分布族、∇ を $\mathcal{P}$ 上の自然な平坦アファイン接続、g を $\mathcal{P}$ 上の Fisher 計量、S, A をそれぞれ $(0,3), (1,2)$ 型の Amari-Chentsov テンソル、$\nabla^{(\alpha)}$ $(\alpha \in \mathbb{R})$ を α-接続とする。

命題 1.3 (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、$R^{(\alpha)}$ の成分は

$${R^{(\alpha)}}_{ijk}{}^{l} = -\frac{1-\alpha^2}{4}\left({A_{jk}}^{m}{A_{im}}^{l} - {A_{ik}}^{m}{A_{jm}}^{l}\right) \tag{1.2}$$

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。

$$\begin{aligned}\partial_i {A_{jk}}^{l} - \partial_j {A_{ik}}^{l} &= \partial_i(g^{la} S_{jka}) - \partial_j(g^{la} S_{ika}) &(1.4)\\ &= \partial_i(g^{la}) S_{jka} + g^{la}\partial_i S_{jka} - \partial_j(g^{la}) S_{ika} - g^{la}\partial_j S_{ika} &(1.5)\end{aligned}$$

$$= \partial_i(g^{la})S_{jka} - \partial_j(g^{la})S_{ika} \tag{1.6}$$

である。右辺第 1 項について、$0 = \partial_i \delta^l_m = \partial_i(g^{la}g_{ma}) = \partial_i(g^{la})g_{ma} + g^{lb}\partial_i(g_{mb})$ より $\partial_i(g^{la}) = -g^{ma}g^{lb}\partial_i(g_{mb})$ だから

$$\begin{aligned} \partial_i(g^{la})S_{jka} &= -g^{ma}g^{lb}\partial_i(g_{mb})S_{jka} && (1.7)\\ &= -g^{ma}g^{lb}S_{imb}S_{jka} && (1.8)\\ &= -A_{im}{}^{l}A_{jk}{}^{m} && (1.9) \end{aligned}$$

同様にして

$$\partial_j(g^{la})S_{ika} = -A_{jm}{}^{l}A_{ik}{}^{m} \tag{1.10}$$

を得る。したがって $\partial_i A_{jk}{}^{l} - \partial_j A_{ik}{}^{l} = -A_{im}{}^{l}A_{jk}{}^{m} + A_{jm}{}^{l}A_{ik}{}^{m}$ だから

$$R^{(\alpha)}{}_{ijk}{}^{l} = \left(-\frac{1-\alpha}{2} + \left(\frac{1-\alpha}{2}\right)^2\right)\left(A_{jk}{}^{m}A_{im}{}^{l} - A_{ik}{}^{m}A_{jm}{}^{l}\right) = -\frac{1-\alpha^2}{4}\left(A_{jk}{}^{m}A_{im}{}^{l} - A_{ik}{}^{m}A_{jm}{}^{l}\right) \tag{1.11}$$

となる。 □

系 1.4.

(1) $\forall \alpha \in \mathbb{R}$ に対し $R^{(\alpha)} = (1-\alpha^2)R^{(0)} = R^{(-\alpha)}$.

(2) 次は同値:

   (a) すべての $\alpha \in \mathbb{R}$ に対し、$\nabla^{(\alpha)}$ は平坦である。

   (b) ある $\alpha \neq \pm 1$ が存在し、$\nabla^{(\alpha)}$ は平坦である。

証明 (1) 命題 1.3 より明らか。

(2) まず (1) より次は同値である:

(a)′ $\forall \alpha \in \mathbb{R}$ に対し $R^{(\alpha)} = 0$.

(b)′ $\exists \alpha \neq \pm 1$ s.t. $R^{(\alpha)} = 0$.

さらに α-接続はすべて torsion-free だから、曲率が 0 であることと平坦であることは同値である。 □

定理 1.5 (α-接続による双対構造)**.** 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。

証明 双対構造であることは、すべての $X, Y, Z \in \mathfrak{X}(\mathcal{P})$ に対し

$$\begin{aligned} g(\nabla^{(\alpha)}_X Y, Z) + g(Y, \nabla^{(-\alpha)}_X Z) &= g(\nabla^g_X Y, Z) - \frac{\alpha}{2}S(X,Y,Z) + g(Y, \nabla^g_X Z) + \frac{\alpha}{2}S(X,Z,Y) && (1.12)\\ &= g(\nabla^g_X Y, Z) + g(Y, \nabla^g_X Z) && (1.13)\\ &= X(g(Y,Z)) && (1.14) \end{aligned}$$

より従う。$\alpha = \pm 1$ で双対平坦となることは系 1.4 よりわかる。 □

2 Legendre 変換

本節では W を有限次元 $\mathbb{R}$-ベクトル空間とする。

定義 2.1 (Legendre 変換). $U \subset W$ を開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\nabla f\colon U \to W^\vee$ が単射であるものとする。関数

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

を f の **Legendre 変換 (Legendre transform)** という。

例 2.2 (Legendre 変換の例). 前回 (0704_資料.pdf) 扱った具体例について対数分配関数の Legendre 変換を計算してみる。

- Bernoulli 分布族 (i.e. 2 元集合上の full support な確率分布の族): 対数分配関数は $\psi\colon \mathbb{R} \to \mathbb{R}$, $\theta \mapsto \log(1+\exp\theta)$ であった。よって $\nabla\psi(\theta) = \dfrac{\exp\theta}{1+\exp\theta}$ であり、$(\nabla\psi)^{-1}(\eta) = \log\eta - \log(1-\eta)$ である。したがって $\psi^\vee(\eta) = \eta\log\eta + (1-\eta)\log(1-\eta)$ である。
- 正規分布族: 対数分配関数は $\psi\colon \mathbb{R}\times\mathbb{R}_{<0} \to \mathbb{R}$, $\theta \mapsto -\dfrac{(\theta^1)^2}{4\theta^2} - \dfrac{1}{2}\log(-\theta^2) + \dfrac{1}{2}\log\pi$ であった。よって $\nabla\psi(\theta) = \begin{pmatrix} -\dfrac{\theta^1}{2\theta^2} & \dfrac{(\theta^1)^2}{4(\theta^2)^2} - \dfrac{1}{2\theta^2}\end{pmatrix}$ であり、$(\nabla\psi)^{-1}(\eta) = \dfrac{1}{\eta_2 - (\eta_1)^2}\begin{pmatrix}\eta_1 \\ -1/2\end{pmatrix}$ である。よって $\psi^\vee(\eta) = -\dfrac{1}{2}\left(1 + \log 2\pi + \log(\eta_2 - (\eta_1)^2)\right)$ である。

本稿では、とくに次の状況を考えることになる。

命題 2.3. $U \subset W$ を凸開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\operatorname{Hess} f$ が U 上各点で (対称であることも含む意味で) 正定値であるものとする。このとき、次が成り立つ:

(1) ∇f は局所微分同相である。とくに $U' := \nabla f(U)$ は $W^\vee$ の開集合である。
(2) $\nabla f\colon U \to U'$ は微分同相である。とくに ∇f は単射である。

したがって $f^\vee$ が定義でき、$f^\vee$ は次をみたす:

(3) $f^\vee\colon U' \to \mathbb{R}$ は C^∞ 関数である。
(4) $\nabla f^\vee = (\nabla f)^{-1}$ が成り立つ。とくに $\nabla f^\vee$ は単射である。
(5) 各 $y \in U'$ に対し $(\operatorname{Hess} f^\vee)_y = ((\operatorname{Hess} f)_x)^{-1}$ が成り立つ (ただし $x := (\nabla f)^{-1}(y)$)。とくに $(\operatorname{Hess} f^\vee)_y$ は正定値である。

証明 (1) 命題の仮定より $\operatorname{Hess} f$ は U 上各点で正定値だから、∇f の微分は各点で線型同型である。したがって ∇f は局所微分同相であり、とくに開写像である。よって $U' = \nabla f(U)$ は $W^\vee$ の開集合である。

(2) $u, \widetilde{u} \in U$, $u \neq \widetilde{u}$ を固定し、$[0,1]$ を含む $\mathbb{R}$ の開区間 I であって、すべての $t \in I$ に対し $(1-t)u + t\widetilde{u}$ が U に属するようなものをひとつ選ぶ (U は W の凸開集合だからこれは可能)。さらに $\varphi\colon I \to U$, $t \mapsto f((1-t)u + t\widetilde{u})$ と定めると、平均値定理より、ある $\tau \in (0,1)$ が存在して

$$\langle \nabla f(\widetilde{u}) - \nabla f(u), \widetilde{u} - u\rangle = \varphi'(1) - \varphi'(0) \tag{2.2}$$

$$= \varphi''(\tau) \qquad (\text{平均値定理}) \tag{2.3}$$

$$= \left\langle (\operatorname{Hess} f)_{(1-\tau)u+\tau\widetilde{u}}, (\widetilde{u}-u)^2 \right\rangle \tag{2.4}$$

$$> 0 \qquad (\operatorname{Hess} f \text{ は正定値}) \tag{2.5}$$

が成り立つ。よって $\nabla f(\widetilde{u}) \neq \nabla f(u)$ である。したがって ∇f は単射である。このことと (1) より $\nabla f : U \to U'$ は微分同相である。

(3) $\nabla f : U \to U'$ が微分同相ゆえに $(\nabla f)^{-1} : U' \to U$ は C^∞ だから、$f^\vee$ は C^∞ 関数である。

(4) $f^\vee$ の定義式を ∇ で微分すると、すべての $y \in U'$ に対し

$$(\nabla f^\vee)(y) = (\nabla f)^{-1}(y) + \left\langle y, \nabla(\nabla f)^{-1}(y) \right\rangle - \left\langle (\nabla f)((\nabla f)^{-1}(y)), \nabla(\nabla f)^{-1}(y) \right\rangle = (\nabla f)^{-1}(y) \tag{2.6}$$

が成り立つ。よって $(\nabla f)^{-1} = \nabla f^\vee$ である。

(5) (4) より

$$(\operatorname{Hess} f^\vee)_y = d(\nabla f^\vee)_y \tag{2.7}$$

$$= d((\nabla f)^{-1})_y \tag{2.8}$$

$$= (d(\nabla f)_x)^{-1} \tag{2.9}$$

$$= ((\operatorname{Hess} f)_x)^{-1} \tag{2.10}$$

となる。 □

系 2.4 (Legendre 変換の対合性). $f^{\vee\vee} = f$.

証明 Legendre 変換の定義より、すべての $x \in U$ に対し

$$f^{\vee\vee}(x) = \left\langle x, (\nabla f^\vee)^{-1}(x) \right\rangle - f^\vee((\nabla f^\vee)^{-1}(x)) \tag{2.11}$$

$$= \langle x, \nabla f(x) \rangle - f^\vee(\nabla f(x)) \qquad (\nabla f^\vee = (\nabla f)^{-1}) \tag{2.12}$$

$$= \langle x, \nabla f(x) \rangle - \left(\left\langle \nabla f(x), (\nabla f)^{-1}(\nabla f(x)) \right\rangle - f((\nabla f)^{-1}(\nabla f(x))) \right) \tag{2.13}$$

$$= f(x) \tag{2.14}$$

が成り立つ。よって $f^{\vee\vee} = f$ である。 □

3 期待値パラメータ

$\mathcal{P}$ を可測空間 $\mathcal{X}$ 上の open な指数型分布族、∇ を $\mathcal{P}$ 上の自然な平坦アファイン接続、g を $\mathcal{P}$ 上の Fisher 計量、S, A をそれぞれ $(0,3), (1,2)$ 型の Amari-Chentsov テンソル、$\nabla^{(\alpha)}$ $(\alpha \in \mathbb{R})$ を α-接続とする。

以降、$\mathcal{P}$ の最小次元実現 (V, T, μ) をひとつ固定し、この実現に関する対数分配関数を $\psi : \widetilde{\Theta} \to \mathbb{R}$ とおく。

命題-定義 3.1 (期待値パラメータ空間). 集合

$$\mathcal{M} := \left\{ E_p[T] \in V \mid p \in \mathcal{P} \right\} \tag{3.1}$$

は V の開部分多様体となり、写像 $\eta : \mathcal{P} \to \mathcal{M},\ p \mapsto E_p[T]$ 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 (V, T, μ) に関する $\mathcal{P}$ 上の**期待値パラメータ座標 (mean parameter coordinates)** という。

この証明には次の 2 つの事実を使う。

事実 3.2 (ψ の微分は十分統計量の期待値)**.** 写像 $\nabla\psi\colon \Theta \to V^{\vee\vee} = V$ は

$$(\nabla\psi)(\theta(p)) = \eta(p) \qquad (p \in \mathcal{P}) \tag{3.2}$$

をみたす。したがって $\mathcal{M} = \nabla\psi(\Theta)$ である。 □

事実 3.3. 位相ベクトル空間の凸集合の内部は凸集合である。 □

命題-定義 3.1 の証明 まず $\mathcal{M}$ が V の開部分多様体となることを示す。ψ を $\operatorname{Int}\widetilde{\Theta}$ 上の関数とみなすと、事実 3.3 とあわせて ψ は命題 2.3 の前提をみたすから、命題 2.3 (1) より $\nabla\psi\colon \operatorname{Int}\widetilde{\Theta} \to V^{\vee\vee} = V$ は局所微分同相、とくに開写像でもある。したがって $\nabla\psi(\operatorname{Int}\widetilde{\Theta})$ は V の開部分多様体となる。さらに Θ は $\operatorname{Int}\widetilde{\Theta}$ の開集合だから、$\nabla\psi(\Theta)$ は $\nabla\psi(\operatorname{Int}\widetilde{\Theta})$ の開部分多様体となる。このことと事実 3.2 より、$\mathcal{M} = \nabla\psi(\Theta)$ は $\nabla\psi(\operatorname{Int}\widetilde{\Theta})$ の開部分多様体となり、とくに V の開部分多様体となる。

次に η が微分同相写像であることを示す。命題 2.3 (2) より $\nabla\psi$ は $\operatorname{Int}\widetilde{\Theta}$ から $\nabla\psi(\operatorname{Int}\widetilde{\Theta})$ への微分同相だから、部分多様体への制限により $\nabla\psi$ は Θ から $\mathcal{M}$ への微分同相を与える。したがって写像 $\eta = (\nabla\psi) \circ \theta\colon \mathcal{P} \to \mathcal{M}$ は微分同相である。 □

以降、$\psi|_{\operatorname{Int}\widetilde{\Theta}}$ の Legendre 変換を $\mathcal{M}$ 上に制限したものを ϕ と書くことにする。

定理 3.4 (自然パラメータ座標と期待値パラメータ座標の関係)**.** 関数 $\psi\colon \Theta \to \mathbb{R}$ および $\phi\colon \mathcal{M} \to \mathbb{R}$ と、$\mathcal{P}$ 上の自然パラメータ座標 $\theta = (\theta^1, \ldots, \theta^n)$ および期待値パラメータ座標 $\eta = (\eta_1, \ldots, \eta_m)$ に関し次が成り立つ:

(1)

$$\frac{\partial\psi}{\partial\theta^i}(\theta(p)) = \eta_i(p), \qquad \frac{\partial\phi}{\partial\eta_i}(\eta(p)) = \theta^i(p) \qquad (p \in \mathcal{P}). \tag{3.3}$$

(2) g の θ-座標に関する成分は

$$g_{ij}(p) = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\theta(p)) = \frac{\partial\eta_j}{\partial\theta^i}(p), \qquad g^{ij}(p) = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j}(\eta(p)) = \frac{\partial\theta^i}{\partial\eta_j}(p) \qquad (p \in \mathcal{P}) \tag{3.4}$$

をみたす。

(3) δ_i^j を Kronecker のデルタとして

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_i^j \tag{3.5}$$

が成り立つ。

証明 (1) 事実 3.2 より $\nabla\psi \circ \theta = \eta$ であることと、命題 2.3 (4) より $\nabla\phi = (\nabla\psi)^{-1}$ であることから従う。

(2) g の定義および命題 2.3 (5) より従う。

(3)

$$g\left(\frac{\partial}{\partial\theta^i},\frac{\partial}{\partial\eta_j}\right)=g\left(\frac{\partial}{\partial\theta^i},\frac{\partial\theta^k}{\partial\eta_j}\frac{\partial}{\partial\theta^k}\right)=g_{ik}\frac{\partial\theta^k}{\partial\eta_j}=g_{ik}g^{kj}=\delta^j_i. \tag{3.6}$$

□

定理 3.5. 期待値パラメータ座標は $\mathcal{P}$ 上の $\nabla^{(-1)}$-アファイン座標である。

証明 $\partial_i=\dfrac{\partial}{\partial\theta^i}$, $\partial^i=\dfrac{\partial}{\partial\eta_i}$ と略記すれば、上の定理の (3) より

$$0=\partial^i\delta^j_k=g\left(\nabla^{(1)}_{\partial^i}\partial_k,\partial^j\right)+g\left(\partial_k,\nabla^{(1)}_{\partial^i}\partial^j\right) \tag{3.7}$$

だから

$$\begin{aligned}
{\Gamma^{(-1)}}^{ij}_{k}&=g\left(\partial_k,\nabla^{(-1)}_{\partial^i}\partial^j\right) &(3.8)\\
&=-g\left(\nabla^{(1)}_{\partial^i}\partial_k,\partial^j\right) &(3.9)\\
&=-\frac{\partial\theta^l}{\partial\eta_i}g\left(\nabla^{(1)}_{\partial_l}\partial_k,\partial^j\right) &(3.10)\\
&=-\frac{\partial\theta^l}{\partial\eta_i}{\Gamma^{(1)}}^{j}_{lk} &(3.11)\\
&=0 \qquad ({\Gamma^{(1)}}^{j}_{lk}=0) &(3.12)
\end{aligned}$$

となる。 □

今後の予定

- KL ダイバージェンス

参考文献

Legendre 変換については [NP18] を参考にした。期待値パラメータに関しては [WJ07] を参考にした。

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

[NP18] Constantin P. Niculescu and Lars-Erik Persson, **Convex Functions and Their Applications**, CMS Books in Mathematics, Springer International Publishing, Cham, 2018.

[WJ07] Martin J. Wainwright and Michael I. Jordan, **Graphical Models, Exponential Families, and Variational Inference**, Foundations and Trends in Machine Learning **1** (2007).

A 付録

命題 1.2 の証明 一意性は g の非退化性より明らか。以下、存在を示す。まず、$X, Z \in \mathfrak{X}(TM)$ を固定すると写像 $\mathfrak{X}(TM) \to C^\infty(M),\ Y \mapsto X(g(Y,Z)) - g(\nabla_X Y, Z)$ は $C^\infty(M)$-線型だから $\Omega^1(M)$ に属する。これを g で添字を上げて得られるベクトル場を $\nabla^*_X Z$ と書くことにすれば、$\nabla^*_X Z$ は目的の式をみたす。ここまでで、目的の式をみたす写像 $\nabla^*\colon \Gamma(TM) \to \mathrm{Map}(\Gamma(TM), \Gamma(TM))$ が得られた。∇^* の像が $\mathrm{Hom}_{C^\infty(M)}(\Gamma(TM), \Gamma(TM)) = \Gamma(T^\vee M \otimes TM)$ に属することは、各 $Z \in \mathfrak{X}(M)$ に対し $\nabla^* Z$ の $C^\infty(M)$-線型性を確かめればよく、すぐにわかる。あとは ∇^* の $\mathbb{R}$-線型性と Leibniz 則を確かめればよいが、これらも ∇^* の定め方から明らか。よって存在が示された。 □