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 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

1.1 一般の多様体の場合

定義 1.1 (双対構造)**.** M を多様体とする。3 つ組 (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)**.** ∇ 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。 □

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 1.3. (g, ∇, ∇') を双対構造、$x = (x_1, \dots, x_m)$ 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、$(U, g|_U)$ は $\mathbb{R}^m$ のある開集合と等長同型である。

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 1.4 (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 $\nabla^{(1)}$-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。

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

同様にして

$$\partial_j A^l_{ik} = -A_{jm}{}^{l} A_{ik}{}^{m} + g^{lm} \partial_j S_{ikm}. \tag{1.9}$$

したがって命題の主張の式が得られた。 □

系 1.5.

(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) 上の命題より明らか。

(2) まず、上の命題より次は同値である:

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

(2) $\exists \alpha \neq \pm 1$ 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 1.6. 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

$$\begin{aligned}
g(\nabla^{(\alpha)}_X Y, Z) + g(Y, \nabla^{(-\alpha)}_X Z) &= g(\nabla^g_X Y, Z) - \frac{1-\alpha}{2} S(X,Y,Z) + g(Y, \nabla^g_X Z) - \frac{1+\alpha}{2} S(X,Z,Y) && (1.10)\\
&= g(\nabla^g_X Y, Z) + g(Y, \nabla^g_X Z) && (1.11)\\
&= X(g(Y,Z)) && (1.12)
\end{aligned}$$

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 Legendre 変換

本節では W を有限次元 $\mathbb{R}$-ベクトル空間、∇ を W 上の標準的なアファイン接続とする。

命題-定義 2.1 (古典的な Legendre 変換). $U \subset W$ を開集合、$f\colon U \to \mathbb{R}$ を C^∞ 関数であって $\nabla f\colon U \to W^\vee$ が可逆であるものとする。このとき、関数 $f^\vee\colon U' \to \mathbb{R}$ (ただし $U' := (\nabla f)(U)$) であって次をみたすものがただひとつ存在する:

$$f(x) + f^\vee(y) = \langle x, y\rangle, \quad \text{where} \quad x := (\nabla f)^{-1}(y) \qquad (y \in U'). \tag{2.1}$$

$f^\vee$ を f の **Legendre 変換 (Legendre transform)** という。

証明 一意性は明らか。存在は $f^\vee(y) := \left\langle (\nabla f)^{-1}(y), y \right\rangle - f((\nabla f)^{-1}(y))$ と定義すればよい。 □

例 2.2 (Legendre 変換の例).

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

命題 2.3 (Legendre 変換の性質).

(1) $(\nabla f)^{-1} = \nabla f^\vee$

(2) $f^{\vee\vee} = f$

証明 (1)

$$(\nabla f^\vee)(y) = (\nabla f)^{-1}(y) + \left\langle y, (\nabla(\nabla f)^{-1})(y)\right\rangle - \left\langle (\nabla f)(\nabla f)^{-1}(y), (\nabla(\nabla f)^{-1})(y)\right\rangle = (\nabla f)^{-1}(y) \tag{2.2}$$

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

(2)

$$\begin{aligned} f^{\vee\vee}(x) &= \left\langle x, (\nabla f^\vee)^{-1}(x)\right\rangle - f^\vee((\nabla f^\vee)^{-1}(x)) && (2.3)\\ &= \langle x, (\nabla f)(x)\rangle - f^\vee((\nabla f)(x)) && (2.4)\\ &= \langle x, (\nabla f)(x)\rangle - \left(\left\langle (\nabla f)(x), (\nabla f)^{-1}(\nabla f)(x)\right\rangle - f((\nabla f)^{-1}(\nabla f)(x))\right) && (2.5)\\ &= f(x) && (2.6) \end{aligned}$$

よって $f^{\vee\vee} = f$ である。 □

本稿では、とくに次の状況下で Legendre 変換を考えることになる。

補題 2.4. W を有限次元 $\mathbb{R}$-ベクトル空間、$U \subset W$ 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 [1)]。

証明　$u, u' \in U,\ u \neq u'$ を固定する。$u_t := (1-t)u + tu'\ (t \in [0,1])$ とおき、$\varphi\colon [0,1] \to U,\ t \mapsto f(u_t)$ と定める (U は凸だから φ の像はたしかに U に属する)。平均値の定理より、ある $\tau \in (0,1)$ が存在して

$$\begin{aligned}
\langle (\nabla f)(u') - (\nabla f)(u), u' - u \rangle &= \varphi'(1) - \varphi'(0) && (2.7)\\
&= \varphi''(\tau) && (2.8)\\
&= \langle (\operatorname{Hess} f)(u_\tau)(u' - u), u' - u \rangle && (2.9)\\
&> 0 && (2.10)
\end{aligned}$$

が成り立つ。よって $(\nabla f)(u') \neq (\nabla f)(u)$ である。したがって ∇f は単射である。 □

3　期待値パラメータ

命題-定義 3.1 (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

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

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。

補題 3.2. 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。

証明　明らか。 □

補題 3.3. $(\nabla\psi)|_{\operatorname{Int}\widetilde{\Theta}}\colon \operatorname{Int}\widetilde{\Theta} \to V^{\vee\vee} = V$ は C^∞ 埋め込みかつ開写像である。

事実 3.4. 位相ベクトル空間の凸集合の内部は凸集合である。 □

補題の証明　ψ は C^∞ だから $\nabla\psi$ も C^∞ である。また、$\operatorname{Hess}\psi$ は正定値だから $\nabla\psi$ の微分は可逆である。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、$\widetilde{\Theta}$ が $V^\vee$ の凸集合であること (`0425_資料.pdf` 命題 2.2) と上の事実より $\operatorname{Int}\widetilde{\Theta}$ 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 3.1 の証明　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。$\eta(p) = (\nabla\psi) \circ \theta(p)\ (p \in \mathcal{P})$

1)　逆は成立しない。すなわち、f' が単射であっても $\operatorname{Hess} f$ が正定値であるとは限らない。$f(x) = x^4$ が反例となる。

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 3.5. $\mathcal{P}$ は開であるとし、ϕ を $\psi|_\Theta$ の Legendre 変換とする。このとき次が成り立つ:

(1)
$$\frac{\partial\psi}{\partial\theta^i} = \eta_i, \qquad \frac{\partial\phi}{\partial\eta_i} = \theta^i. \tag{3.3}$$

(2) θ-座標に関し
$$g_{ij} = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j} = \frac{\partial\eta_j}{\partial\theta^i}, \qquad g^{ij} = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j} = \frac{\partial\theta^i}{\partial\eta_j}. \tag{3.4}$$

(3)
$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta^j_i. \tag{3.5}$$

証明 (1) $\dfrac{\partial\psi}{\partial\theta^i}(\theta(p)) = E_p[T^i] = \eta_i(p)$ である。また、Legendre 変換の定義より
$$\phi(\eta) = \left\langle \eta, (\nabla\psi)^{-1}(\eta)\right\rangle - \psi((\nabla\psi)^{-1}(\eta)) \tag{3.6}$$
だから
$$(\nabla\phi)(\eta) = (\nabla\psi)^{-1}(\eta) + \left\langle \eta, \nabla(\nabla\psi)^{-1}(\eta)\right\rangle - \left\langle (\nabla\psi)(\nabla\psi)^{-1}(\eta), \nabla(\nabla\psi)^{-1}(\eta)\right\rangle \tag{3.7}$$
$$= (\nabla\psi)^{-1}(\eta) + \left\langle \eta, \nabla(\nabla\psi)^{-1}(\eta)\right\rangle - \left\langle \eta, \nabla(\nabla\psi)^{-1}(\eta)\right\rangle \tag{3.8}$$
$$= (\nabla\psi)^{-1}(\eta) \tag{3.9}$$
である。したがって $(\nabla\phi)(\eta(p)) = (\nabla\psi)^{-1}(\eta(p)) = \theta(p)$ よって $\dfrac{\partial\phi}{\partial\eta_i}(\eta(p)) = \theta^i(p)$ である。

(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。また $\delta^j_k = \dfrac{\partial\eta_i}{\partial\theta^k}\dfrac{\partial\theta^j}{\partial\eta_i} = g_{ik}\dfrac{\partial\theta^j}{\partial\eta_i}$ より $g^{ij} = \dfrac{\partial\theta^j}{\partial\eta_i} = \dfrac{\partial^2\phi}{\partial\eta_i\partial\eta_j}$ である。

(3)
$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial\theta^k}{\partial\eta_j}\frac{\partial}{\partial\theta^k}\right) = g_{ik}\frac{\partial\theta^k}{\partial\eta_j} = g_{ik}g^{kj} = \delta^j_i. \tag{3.10}$$
□

定理 3.6. 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 $\partial_i = \dfrac{\partial}{\partial\theta^i}$, $\partial^i = \dfrac{\partial}{\partial\eta_i}$ と略記すれば、上の命題の (3) より
$$0 = \partial^i\delta^j_k = g\left(\nabla^{(1)}_{\partial^i}\partial_k, \partial^j\right) + g\left(\partial_k, \nabla^{(1)}_{\partial^i}\partial^j\right) \tag{3.11}$$
だから
$$\Gamma^{(-1)}{}^{ij}_{k} = g\left(\partial_k, \nabla^{(-1)}_{\partial^i}\partial^j\right) \tag{3.12}$$

$$
\begin{aligned}
&= -g\left(\nabla^{(1)}_{\partial^i}\partial_k, \partial^j\right) && (3.13)\\
&= -\frac{\partial\theta^l}{\partial\eta_i} g\left(\nabla^{(1)}_{\partial_l}\partial_k, \partial^j\right) && (3.14)\\
&= -\frac{\partial\theta^l}{\partial\eta_i}\Gamma^{(1)}{}^{j}_{lk} && (3.15)\\
&= 0 \qquad (\Gamma^{(1)}{}^{j}_{lk} = 0) && (3.16)
\end{aligned}
$$

となる。 □

例 3.7 ($\nabla^{(-1)}$-測地線)**.** [TODO] 有限集合上の場合の $\nabla^{(-1)}$-測地線を自然パラメータ座標で表すとどうなる?

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

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。これで写像 $\nabla' \colon \Gamma(TM) \to \mathrm{Map}(\Gamma(TM), \Gamma(TM))$ 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。 □