

第 1 章 指数型分布族

1 指数型分布族

定義 1.1 (指数型分布族). \mathcal{X} を可測空間、 $\emptyset \neq \mathcal{P} \subset \mathcal{P}(\mathcal{X})$ とする。 \mathcal{P} が \mathcal{X} 上の指数型分布族 (exponential family) であるとは、次が成り立つことをいう: $\exists (V, T, \mu)$ s.t.

- (E0) V は有限次元 \mathbb{R} -ベクトル空間である。
- (E1) $T: \mathcal{X} \rightarrow V$ は可測写像である。
- (E2) μ は \mathcal{X} 上の σ -有限測度であり、 $\forall p \in \mathcal{P}$ に対し $p \ll \mu$ をみたす。
- (E3) $\forall p \in \mathcal{P}$ に対し、 $\exists \theta \in V^\vee$ s.t.

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x) \rangle}{\int_{\mathcal{X}} \exp\langle \theta, T(y) \rangle \mu(dy)} \quad \mu\text{-a.e. } x \in \mathcal{X} \quad (1.1)$$

である。ただし $\langle \cdot, \cdot \rangle$ は自然なペアリング $V^\vee \times V \rightarrow \mathbb{R}$ である。

さらに次のように定める:

- (V, T, μ) を \mathcal{P} の実現 (representation) という。
 - V の次元を (V, T, μ) の次元 (dimension) という。
 - T を (V, T, μ) の十分統計量 (sufficient statistic) という。
 - μ を (V, T, μ) の基底測度 (base measure) という。
- 集合 $\Theta_{(V, T, \mu)}$

$$\Theta_{(V, T, \mu)} := \left\{ \theta \in V^\vee \mid \int_{\mathcal{X}} \exp\langle \theta, T(y) \rangle \mu(dy) < +\infty \right\} \quad (1.2)$$

を (V, T, μ) の自然パラメータ空間 (natural parameter space) という。

- 関数 $\psi: \Theta_{(V, T, \mu)} \rightarrow \mathbb{R}$,

$$\psi(\theta) := \log \int_{\mathcal{X}} \exp\langle \theta, T(y) \rangle \mu(dy) \quad (1.3)$$

を (V, T, μ) の対数分配関数 (log-partition function) という。

[TODO] 記号の修正 [TODO] full の追加

命題 1.2 (自然パラメータ空間は凸集合). $\Theta_{(T, \mu)}$ は \mathbb{R}^m の凸集合である。[TODO] V に修正

証明 表記の簡略化のため $\Theta := \Theta_{(T, \mu)}$ とおく。 $\theta, \theta' \in \Theta$, $t \in (0, 1)$ とし、 $(1-t)\theta + t\theta' \in \Theta$ を示せばよい。そこで $p := \frac{1}{1-t}$, $q := \frac{1}{t}$ とおくと、 $p, q \in (1, +\infty)$ であり、 $\frac{1}{p} + \frac{1}{q} = (1-t) + t = 1$ であり、 $e^{(1-t)\langle \theta, T(x) \rangle} \in L^p(\mathcal{X}, \mu)$ かつ $e^{t\langle \theta', T(x) \rangle} \in L^q(\mathcal{X}, \mu)$ だから、Hölder の不等式より

$$\int_{\mathcal{X}} e^{\langle (1-t)\theta + t\theta', T(x) \rangle} \mu(dx) = \int_{\mathcal{X}} e^{(1-t)\langle \theta, T(x) \rangle} e^{t\langle \theta', T(x) \rangle} \mu(dx) \quad (1.4)$$

$$\leq \left(\int_{\mathcal{X}} e^{(1-t)\langle \theta, T(x) \rangle} \mu(dx) \right)^{1/p} \left(\int_{\mathcal{X}} e^{t\langle \theta, T(x) \rangle} \mu(dx) \right)^{1/q} \quad (1.5)$$

$$= \left(\int_{\mathcal{X}} e^{\langle \theta, T(x) \rangle} \mu(dx) \right)^{1/p} \left(\int_{\mathcal{X}} e^{\langle \theta, T(x) \rangle} \mu(dx) \right)^{1/q} \quad (1.6)$$

$$< +\infty \quad (1.7)$$

が成り立つ。したがって $(1-t)\theta + t\theta' \in \Theta$ である。 \square

例 1.3 (有限集合上の確率分布). [TODO] V に修正 $\mathcal{X} = \{1, \dots, n\}$ 、 γ を \mathcal{X} 上の数え上げ測度とする。 \mathcal{X} 上の確率分布全体の集合 $\mathcal{P}(\mathcal{X})$ が \mathcal{X} 上の指数型分布族であることを確かめる。 δ^j ($j = 1, \dots, n$) を点 j での Dirac 測度とおく。任意の $P \in \mathcal{P}(\mathcal{X})$ に対し、

$$P(dk) := \sum_{j=1}^n a_j \delta^j(dk), \quad a_1, \dots, a_n \in \mathbb{R}_{>0}, \quad \sum_{j=1}^n a_j = 1 \quad (1.8)$$

が成り立つから、 δ_{jk} ($j, k = 1, \dots, n$) を Kronecker のデルタとして

$$P(dk) = \exp \left(\sum_{j=1}^n (\log a_j) \delta_{jk} \right) \gamma(dk) \quad (1.9)$$

$$= \exp \left(\sum_{j=1}^n \theta_j \delta_{jk} \right) \gamma(dk) \quad (1.10)$$

(ただし $\theta_j := \log a_j$) と表せる。したがって $T: \mathcal{X} \rightarrow \mathbb{R}^n$, $k \mapsto {}^t(\delta_{1k}, \dots, \delta_{nk})$ とおけば、 (T, γ) を実現として $\mathcal{P}(\mathcal{X})$ は指数型分布族となることがわかる。

例 1.4 (正規分布族). [TODO] V に修正 $\mathcal{X} = \mathbb{R}$ 、 λ を \mathbb{R} 上の Lebesgue 測度とする。 \mathcal{X} 上の確率分布の集合

$$\mathcal{P} := \left\{ P_{(\mu, \sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \mid \mu \in \mathbb{R}, \sigma^2 > 0 \right\} \quad (1.11)$$

を正規分布族 (family of normal distributions) という。このとき \mathcal{P} が \mathcal{X} 上の指数型分布族であることを確かめる。任意の $P_{(\mu, \sigma^2)} \in \mathcal{P}$ に対し

$$P_{(\mu, \sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \quad (1.12)$$

$$= \exp \left(-\frac{1}{2\sigma^2} (x^2 - 2\mu x + \mu^2) - \frac{1}{2} \log 2\pi\sigma^2 \right) \lambda(dx) \quad (1.13)$$

$$= \exp \left(\left[\frac{\mu}{\sigma^2} \quad -\frac{1}{2\sigma^2} \right] \begin{bmatrix} x \\ x^2 \end{bmatrix} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log 2\pi\sigma^2 \right) \lambda(dx) \quad (1.14)$$

$$= \exp \left(\begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} x \\ x^2 \end{bmatrix} + \frac{\theta_1^2}{4\theta_2} - \frac{1}{2} \log \left(-\frac{\pi}{\theta_2} \right) \right) \lambda(dx) \quad (1.15)$$

(ただし $\theta_1 := \frac{\mu}{\sigma^2}$, $\theta_2 := -\frac{1}{2\sigma^2}$) が成り立つから、 $T: \mathcal{X} \rightarrow \mathbb{R}^2$, $x \mapsto {}^t(x, x^2)$ とおけば、 (T, λ) を実現として \mathcal{P} は指数型分布族となることがわかる。

例 1.5 (Poisson 分布族). [TODO] V に修正 $X = \mathbb{N}$, γ を \mathbb{N} 上の数え上げ測度とする。 X 上の確率分布の集合

$$\mathcal{P} := \left\{ P_\lambda(dk) = \frac{\lambda^k}{k!} e^{-\lambda} \gamma(dk) \mid \lambda > 0 \right\} \quad (1.16)$$

を P_λ を **Poisson 分布族 (family of Poisson distributions)** という。このとき \mathcal{P} が X 上の指数型分布族であることを確かめる。任意の $P_\lambda \in \mathcal{P}$ に対し

$$P_\lambda(dk) = \frac{\lambda^k}{k!} e^{-\lambda} \gamma(dk) \quad (1.17)$$

$$= \exp(k \log \lambda - \lambda) \frac{1}{k!} \gamma(dk) \quad (1.18)$$

$$= \exp(\theta k - e^\theta) \frac{1}{k!} \gamma(dk) \quad (1.19)$$

(ただし $\theta := \log \lambda$) が成り立つから、 $T: X \rightarrow \mathbb{R}, k \mapsto k$ とおけば、 $\left(T, \frac{1}{k!} \gamma(dk)\right)$ を実現として \mathcal{P} は指数型分布族となることがわかる。

2 最小次元実現

[TODO] 節の内容を整理する

定義 2.1 (最小次元実現). 実現 (V, T, μ) が \mathcal{P} の実現のうちで次元が最小のものであるとき、 (V, T, μ) を \mathcal{P} の **最小次元実現 (minimal representation)** という。

定理 2.2 (「 θ が一意の実現」の存在). [TODO] 「単射性条件」の言葉に修正 X を可測空間、 $\mathcal{P} \subset \mathcal{P}(X)$ を X 上の指数型分布族とする。このとき、 \mathcal{P} の「 θ が一意の実現」が存在する。

証明 (V, T, μ) は \mathcal{P} の実現のうちで次元が最小のものであるとする。 (V, T, μ) の次元 (m とおく) が 0 ならば V^\vee は 1 点集合だから証明は終わる。

以下 $m \geq 1$ の場合を考え、 (V, T, μ) が「 θ が一意の実現」であることを示す。背理法のために (V, T, μ) が「 θ が一意の実現」でないこと、すなわちある $p_0 \in \mathcal{P}$ および $\theta_0, \theta'_0 \in V^\vee$, $\theta_0 \neq \theta'_0$ が存在して

$$\exp(\langle \theta_0, T(x) \rangle - \psi(\theta_0)) = \frac{dp_0}{d\mu}(x) = \exp(\langle \theta'_0, T(x) \rangle - \psi(\theta'_0)) \quad \mu\text{-a.e. } x \in X \quad (2.1)$$

が成り立つことを仮定する。証明の方針としては、次元 $m-1$ の実現 (V', T', μ) を具体的に構成することにより、 (V, T, μ) の次元 m が最小であることとの矛盾を導く。

さて、式 (2.1) を整理して

$$\langle \theta_0 - \theta'_0, T(x) \rangle = \psi(\theta_0) - \psi(\theta'_0) \quad \mu\text{-a.e. } x \in X \quad (2.2)$$

を得る。表記の簡略化のために $\theta_1 := \theta_0 - \theta'_0 \in V^\vee$, $r := \psi(\theta_0) - \psi(\theta'_0) \in \mathbb{R}$ とおけば

$$\langle \theta_1, T(x) \rangle = r \quad \mu\text{-a.e. } x \in X \quad (2.3)$$

を得る。ここで $V' := (\mathbb{R}\theta)^\perp = \{v \in V \mid \langle \theta, v \rangle = 0\}$ とおき、次の claim を示す。

Claim ある可測写像 $T': X \rightarrow V'$ および $v_0 \in V$ が存在して $T(x) = T'(x) + v_0$ (μ -a.e. x) が成り立つ。

⊙ いま背理法の仮定より $\theta_1 \neq 0$ であるから、 θ_1 を延長した V^\vee の基底 $\theta_1, \dots, \theta_m$ が存在する。このとき、 $\theta_1, \dots, \theta_m$ を双対基底に持つ V の基底 v_1, \dots, v_m が存在する。この基底 v_1, \dots, v_m に関する T の成分表示を $T(x) = \sum_{i=1}^m T^i(x)v_i$, $T^i: X \rightarrow \mathbb{R}$ とおくと、(2.3) より $T^1(x) = \langle \theta_1, T(x) \rangle = r$ (μ -a.e. x) が成り立つ。そこで $v_0 := rv_1 \in V$ とおくと $\langle \theta_1, T(x) - v_0 \rangle = 0$ (μ -a.e. x) が成り立つから、可測写像 $T': X \rightarrow V'$ を

$$T'(x) := \begin{cases} T(x) - v_0 & (\langle \theta_1, T(x) - v_0 \rangle = 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (2.4)$$

と定めることができる。この T, v_0 が求めるものである。 //

(V', T', μ) が \mathcal{P} の実現であることを示す。定義 1.1 の条件 (E0)-(E2) は明らかに成立しているから、あとは条件 (E3) を確認すればよい。そこで $p \in \mathcal{P}$ とする。いま (V, T, μ) が \mathcal{P} の実現であることより、ある $\theta \in V^\vee$ が存在して

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x) \rangle}{\int_X \exp\langle \theta, T(y) \rangle \mu(dy)} \quad \mu\text{-a.e. } x \in X \quad (2.5)$$

が成り立つ。 T', v_0 を用いて式変形すると、 μ -a.e. x に対し

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x) \rangle}{\int_X \exp\langle \theta, T(y) \rangle \mu(dy)} \quad (2.6)$$

$$= \frac{\exp\langle \theta, T'(x) \rangle + \langle \theta, v_0 \rangle}{\int_X \exp\langle \theta, T'(y) \rangle + \langle \theta, v_0 \rangle \mu(dy)} \quad (2.7)$$

$$= \frac{\exp\langle \theta, T'(x) \rangle}{\int_X \exp\langle \theta, T'(y) \rangle \mu(dy)} \quad (2.8)$$

が成り立つ。したがって (V', T', μ) は条件 (E3) も満たし、 \mathcal{P} の実現であることがいえた。 (V', T', μ) は次元 $m-1$ だから (V, T, μ) の次元 m の最小性に矛盾する。背理法より (V, T, μ) は \mathcal{P} の「 θ が一意の実現」である。 \square

定理 2.3 (極小実現の性質). [TODO] V に修正 X を可測空間、 $\mathcal{P} \subset \mathcal{P}(X)$ を X 上の指数型分布族、 (T, μ) を \mathcal{P} の次元 m の実現とする。このとき、 (T, μ) が極小実現ならば、 $\langle u, T(x) \rangle$ が μ -a.e. 定数であるような $u \in \mathbb{R}^m$ は $u = 0$ のみである。

証明 (T, μ) を \mathcal{P} の極小実現とする。背理法のため、ある $u \neq 0$ が存在して $\langle u, T(x) \rangle$ が X 上 μ -a.e. 定数であると仮定しておく。 $p \in \mathcal{P}$ とし、定義 1.1 の条件 (E3) の $\theta \in \mathbb{R}^m$ をひとつ選ぶと、

$$\frac{dp}{d\mu}(x) = \frac{e^{\langle \theta, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} \mu(dy)} \quad (2.9)$$

$$= \frac{e^{\langle \theta, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} \mu(dy)} \cdot \frac{e^{\langle u, T(x) \rangle}}{e^{\langle u, T(x) \rangle}} \quad (2.10)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} e^{\langle u, T(y) \rangle} \mu(dy)} \quad (2.11)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} e^{\langle u, T(y) \rangle} \mu(dy)} \quad (2.12)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_X e^{\langle \theta+u, T(y) \rangle} \mu(dy)} \quad (2.13)$$

を得る。したがって $\theta + u$ も定義 1.1 の条件 (E3) を満たすが、いま $u \neq 0$ より $\theta + u \neq \theta$ だから、 (T, μ) が \mathcal{P} の極小実現であることに反する。背理法より定理が示された。 \square

例 2.4 (有限集合上の確率分布族). 例 1.3 の (T, γ) は $\mathcal{P}(X)$ の極小実現である。実際、任意の $P \in \mathcal{P}(X)$ に対し、 θ_j は $\theta_j = \log P(\{j\})$ ($j = 1, \dots, n$) として一意に決まる。

命題 2.5. [TODO] 上の命題とあわせる (V, T, μ) に関する次の条件は同値である:

- (1) $\langle \theta, T(x) \rangle$ が X 上 μ -a.e. 定数であるような $\theta \in V^\vee$ は $\theta = 0$ のみである。
- (2) 各 $p \in \mathcal{P}$ に対し、定義 1.1 の条件 (E3) をみたす $\theta \in V^\vee$ はただひとつである。

証明 (2) \Rightarrow (1) 前回示した。

(1) \Rightarrow (2) $\theta, \theta' \in V^\vee$ が定義 1.1 の条件 (E3) をみたすすると、

$$e^{\langle \theta, T(x) \rangle - \psi(\theta)} = \frac{dp}{d\mu}(x) = e^{\langle \theta', T(x) \rangle - \psi(\theta')} \quad \mu\text{-a.e. } x \in X \quad (2.14)$$

が成り立つ。式を整理して

$$\langle \theta - \theta', T(x) \rangle = \psi(\theta) - \psi(\theta') \quad \mu\text{-a.e. } x \in X \quad (2.15)$$

が成り立つ。したがって (1) より $\theta = \theta'$ である。 \square

本節の目標は、最小次元実現の間のアファイン変換の一意存在を述べた定理 2.17 の証明である。本節では、定理などのステートメントを簡潔にするために圏の言葉を用いる。

命題-定義 2.6. 次のデータにより圏が定まる:

- 対象: \mathcal{P} の実現 (V, T, μ) 全体
- 射: (V, T, μ) から (V', T', μ') への射は、 V から V' への全射アファイン写像 (L, b) ($L \in \text{Lin}(V, V')$, $b \in V'$) であって $T'(x) = L(T(x)) + b$ μ -a.e. x をみたすもの
- 合成: アファイン写像の合成 $(L, b) \circ (K, c) = (LK, Lc + b)$

この圏を $\mathbf{C}_{\mathcal{P}}$ と書く。

証明 示すべきことは、射の合成が射であること、恒等射の存在、結合律の 3 点である。射の合成が射であることは、全射と全射の合成が全射であることと、 μ と μ' が互いに絶対連続であることから従う。また、 (V, T, μ) の恒等射は明らかに恒等写像 $(\text{id}_V, 0)$ であり、結合律はアファイン写像の合成の結合律より従う。 \square

最小次元実現を特徴づける 2 つの条件を導入する。

命題-定義 2.7 (条件 A). \mathcal{P} の実現 (V, T, μ) に関する次の条件は同値である:

- (1) $P: \Theta \rightarrow \mathcal{P}(X)$ は単射である。
- (2) $\forall \theta \in V^\vee$ に対し「 $\langle \theta, T(x) \rangle = \text{const. } \mu\text{-a.e. } x \implies \theta = 0$ 」が成り立つ。
- (3) V の任意の真アファイン部分空間 W に対し、「 $T(x) \in W$ $\mu\text{-a.e. } x$ でない」が成り立つ。

これらの条件が成り立つとき、 (V, T, μ) は**条件 A** をみたすという。

証明 (1) \iff (2) は 0502_資料.pdf の命題 2.2 で示した。(2) \iff (3) は 0523_コメント.pdf の命題 0.4 に記した。 \square

定義 2.8 (条件 B). \mathcal{P} の実現 (V, T, μ) に関する条件

- (1) $\Theta^{\mathcal{P}}$ は V^\vee を affine span する。

が成り立つとき、 (V, T, μ) は**条件 B** をみたすという。

条件 A は射の一意性を保証する。

命題 2.9 (条件 A をみたす対象からの射の一意性). $(V, T, \mu), (V', T', \mu')$ を $\mathbf{C}_{\mathcal{P}}$ の対象とする。このとき、 (V, T, μ) が条件 A をみたすならば、 (V, T, μ) から (V', T', μ') への射は一意である。

証明 $(L, b), (K, c)$ を (V, T, μ) から (V', T', μ') への射とする。射の定義より

$$\begin{cases} T'(x) = L(T(x)) + b & \mu\text{-a.e. } x \\ T'(x) = K(T(x)) + c & \mu\text{-a.e. } x \end{cases} \quad (2.16)$$

が成り立つから、2 式を合わせて

$$(K - L)(T(x)) = b - c \quad \mu\text{-a.e. } x \quad (2.17)$$

となる。そこで基底を固定して成分ごとに (V, T, μ) の条件 A(2) を適用すれば、 $K = L$ を得る。よって上式で $K = L$ として $b = c$ $\mu\text{-a.e.}$ したがって $b = c$ を得る。以上より $(L, b) = (K, c)$ である。 \square

射が存在するための十分条件を調べる。

命題 2.10 (条件 A, B をみたす対象への射の存在). (V, T, μ) を $\mathbf{C}_{\mathcal{P}}$ の対象とする。このとき、 (V, T, μ) が条件 A と条件 B をみたすならば、任意の対象 (V', T', μ') から (V, T, μ) への射が存在する。

この命題の証明には次の補題を用いる。

補題 2.11. $(V, T, \mu), (V', T', \mu')$ を $\mathbf{C}_{\mathcal{P}}$ の対象とし、 $\theta: \mathcal{P} \rightarrow \Theta^{\mathcal{P}}$ および $\theta': \mathcal{P} \rightarrow \Theta^{\mathcal{P}}$ を P, P' の右逆写像とする。

このとき、任意の $p, q \in \mathcal{P}$ に対し、

$$\begin{aligned} & \langle \theta(p) - \theta(q), T(x) \rangle - \psi(\theta(p)) + \psi(\theta(q)) \\ &= \langle \theta'(p) - \theta'(q), T'(x) \rangle - \psi'(\theta'(p)) + \psi'(\theta'(q)) \end{aligned} \quad \mu\text{-a.e.}x \quad (2.18)$$

が成り立つ。

証明 $p, q \in \mathcal{P}$ を任意とすると、指数型分布族の定義と μ, μ' が互いに絶対連続であることより、 $\mu\text{-a.e.}x$ に対し

$$\begin{aligned} \frac{dp}{d\mu}(x) &= \exp(\langle \theta(p), T(x) \rangle - \psi(\theta(p))), & \frac{dp}{d\mu'}(x) &= \exp(\langle \theta'(p), T'(x) \rangle - \psi'(\theta'(p))) \\ \frac{dq}{d\mu}(x) &= \exp(\langle \theta(q), T(x) \rangle - \psi(\theta(q))), & \frac{dq}{d\mu'}(x) &= \exp(\langle \theta'(q), T'(x) \rangle - \psi'(\theta'(q))) \end{aligned} \quad (2.19)$$

が成り立つ。さらに p, q が互いに絶対連続であることから、 $\mu\text{-a.e.}x$ に対し

$$\frac{dp}{dq}(x) = \frac{dp}{d\mu}(x) \left/ \frac{dq}{d\mu}(x) \right. = \exp \{ \langle \theta(p) - \theta(q), T(x) \rangle - \psi(\theta(p)) + \psi(\theta(q)) \} \quad (2.20)$$

$$\frac{dp}{dq}(x) = \frac{dp}{d\mu'}(x) \left/ \frac{dq}{d\mu'}(x) \right. = \exp \{ \langle \theta'(p) - \theta'(q), T'(x) \rangle - \psi'(\theta'(p)) + \psi'(\theta'(q)) \} \quad (2.21)$$

が成り立つ。 \log をとって補題の主張の等式を得る。 \square

命題 2.10 の証明 Step 0: V, V^\vee の基底を選ぶ (V, T, μ) の条件 B より、 V^\vee のあるアフィン基底 $a^i \in \Theta^\mathcal{P}$ ($i = 0, \dots, m$) が存在する。そこで $e^i := a^i - a^0 \in V^\vee$ ($i = 1, \dots, m$) とおくとこれは V^\vee の基底である。さらに e^i の双対基底を V の元と同一視したものを $e_i \in V$ ($i = 1, \dots, m$) とおいておく。

Step 1: 射 (L, b) の構成 P, P' の右逆写像 $\theta: \mathcal{P} \rightarrow \Theta^\mathcal{P}$ および $\theta': \mathcal{P} \rightarrow \Theta^{\mathcal{P}'}$ をひとつずつ選んで $p^i := P(a^i) \in \mathcal{P}$ ($i = 0, \dots, m$) とおき、 (L, b) を次のように定める:

$$L: V' \rightarrow V, \quad t' \mapsto \langle \theta'(p^i) - \theta'(p^0), t' \rangle e_i \quad (2.22)$$

$$b := \{ \psi(\theta(p^i)) - \psi(\theta(p^0)) - \psi'(\theta'(p^0)) + \psi'(\theta'(p^i)) \} e_i \in V \quad (2.23)$$

示すべきことは、~~すべての $p \in \mathcal{P}$ に対し~~

$$T(x) = L(T'(x)) + b \quad \mu'\text{-a.e.}x \quad (2.24)$$

が成り立つことと、 (L, b) が全射となることである。

Step 2: $T(x) = L(T'(x)) + b$ の証明 各 $i = 1, \dots, m$ に対し、補題 2.11 より

$$\begin{aligned} & \langle \theta(p^i) - \theta(p^0), T(x) \rangle - \psi(\theta(p^i)) + \psi(\theta(p^0)) \\ &= \langle \theta'(p^i) - \theta'(p^0), T'(x) \rangle - \psi'(\theta'(p^i)) + \psi'(\theta'(p^0)) \end{aligned} \quad \mu'\text{-a.e.}x \quad (2.25)$$

となる。ここで (V, T, μ) の条件 A (1) より $\theta(p^i) = a^i$ が成り立つから、(2.25) より

$$\begin{aligned} \langle a^i - a^0, T(x) \rangle &= \langle \theta'(p^i) - \theta'(p^0), T'(x) \rangle \\ &\quad + \psi(\theta(p^i)) - \psi(\theta(p^0)) - \psi'(\theta'(p^i)) + \psi'(\theta'(p^0)) \end{aligned} \quad \mu'\text{-a.e.}x \quad (2.26)$$

したがって

$$T(x) = L(T'(x)) + b \quad \mu'\text{-a.e.}x \quad (2.27)$$

が成り立つ。

Step 3: (L, b) が全射であることの証明 L が全射であることをいえばよい。もし L が全射でなかったとすると、 $T(x) = L(T'(x)) + b \in \text{Im } L + b$ が $\mu'\text{-a.e.}x$ したがって $\mu\text{-a.e.}x$ に対し成り立つことになるが、 $\text{Im } L + b$ は V の真アファイン部分空間だから (V, T, μ) の条件 A (3) に反する。したがって L は全射である。 \square

各条件をみたさない場合にも、射が存在する。

補題 2.12 (条件 A をみたさない対象からの射の存在). (V, T, μ) を $\mathbf{C}_\mathcal{P}$ の対象とする。このとき、 (V, T, μ) が条件 A をみたさないならば、 (V, T, μ) よりも次元の小さいある対象 (V', T', μ') への射 $(V, T, \mu) \rightarrow (V', T', \mu')$ が存在する。

証明 (V, T, μ) が条件 A をみたさないという仮定から、ある $\theta \in V^\vee$, $\theta \neq 0$ および $r \in \mathbb{R}$ が存在して

$$\langle \theta, T(x) \rangle = r \quad \mu\text{-a.e.}x \quad (2.28)$$

が成り立つ。そこで $V' := (\mathbb{R}\theta)^\perp = \{v \in V \mid \langle \theta, v \rangle = 0\}$ とおくと、ある可測写像 $T': X \rightarrow V'$ および $v_0 \in V$ が存在して $T(x) = T'(x) + v_0$ ($\mu\text{-a.e.}x$) が成り立つ。このように定めた組 (V', T', μ) が \mathcal{P} の実現であることは一旦認めて最後に示すこととし、まず次元と射について確かめる。

まず (V', T', μ) の次元は $\dim V' = \dim V - 1 < \dim V$ より (V, T, μ) の次元よりも小さい。また、射影 $\pi: V \rightarrow V'$ をひとつ選べば、 $(\pi, 0)$ は明らかに (V, T, μ) から (V', T', μ) への射を与える。

あとは (V', T', μ) が \mathcal{P} の実現であることを示せばよい。指数型分布族の定義 (0502_資料.pdf) の条件 (E0), (E1), (E2) は明らかに成立しているから、あとは条件 (E3) を確認すればよい。そこで $p \in \mathcal{P}$ を任意とする。いま (V, T, μ) が \mathcal{P} の実現であることから、ある $\theta \in V^\vee$ が存在して

$$\frac{dp}{d\mu}(x) = \frac{\exp(\langle \theta, T(x) \rangle)}{\int_X \exp(\langle \theta, T(y) \rangle) \mu(dy)} \quad \mu\text{-a.e.}x \quad (2.29)$$

が成り立つ。 T', v_0 を用いて式変形すると、 $\mu\text{-a.e.}x$ に対し

$$\frac{dp}{d\mu}(x) = \frac{\exp(\langle \theta, T(x) \rangle)}{\int_X \exp(\langle \theta, T(y) \rangle) \mu(dy)} \quad (2.30)$$

$$= \frac{\exp(\langle \theta, T'(x) \rangle + \langle \theta, v_0 \rangle)}{\int_X \exp(\langle \theta, T'(y) \rangle + \langle \theta, v_0 \rangle) \mu(dy)} \quad (2.31)$$

$$= \frac{\exp(\langle \theta, T'(x) \rangle)}{\int_X \exp(\langle \theta, T'(y) \rangle) \mu(dy)} \quad (2.32)$$

が成り立つ。したがって (V', T', μ) は条件 (E3) も満たし、 \mathcal{P} の実現であることがいえた。 \square

補題 2.13 (条件 B をみたさない対象からの射の存在). (V, T, μ) を $\mathbf{C}_\mathcal{P}$ の対象とする。このとき、 (V, T, μ) が条件 B をみたさないならば、 (V, T, μ) よりも次元の小さいある対象 (V', T', μ') への射 $(V, T, \mu) \rightarrow (V', T', \mu')$ が存在する。

証明 (V, T, μ) が条件 B をみたさないとする。すると、ある真ベクトル部分空間 $W \subsetneq V^\vee$ および $\theta_0 \in \Theta^\mathcal{P}$ が存在して $\text{aspan } \Theta^\mathcal{P} = W + \theta_0$ が成り立つ。そこで $\tilde{V} := V/W^\perp$ と定め、 $\pi: V \rightarrow \tilde{V}$ を自然な射影として $\tilde{T} := \pi \circ T: \mathcal{X} \rightarrow \tilde{V}$ と定める。また、 \mathcal{X} 上の測度 $\tilde{\mu}$ を $\tilde{\mu} := \exp \langle \theta_0, T(x) \rangle \cdot \mu$ と定める。このように定めた組 $(\tilde{V}, \tilde{T}, \tilde{\mu})$ が \mathcal{P} の実現であることは一旦認めて最後に示すこととし、まず次元と射について確かめる。

まず $(\tilde{V}, \tilde{T}, \tilde{\mu})$ の次元は $\dim \tilde{V} = \dim V - \dim W^\perp = \dim W < \dim V^\vee = \dim V$ より (V, T, μ) の次元よりも小さい。また、 $(\pi, 0)$ は明らかに (V, T, μ) から $(\tilde{V}, \tilde{T}, \tilde{\mu})$ への射を与える。

あとは $(\tilde{V}, \tilde{T}, \tilde{\mu})$ が \mathcal{P} の実現であることを示せばよい。指数型分布族の定義 (0502_資料.pdf) の条件 (E0), (E1), (E3) の成立は簡単に確かめられるから、ここでは条件 (E3) だけ確かめる。そこで $p \in \mathcal{P}$ を任意とする。 (V, T, μ) が \mathcal{P} の実現であることから、ある $\theta \in V^\vee$ が存在して

$$p(dx) = \frac{\exp \langle \theta, T(x) \rangle}{\int_X \exp \langle \theta, T(x) \rangle d\mu(x)} \mu(dx) \quad (2.33)$$

が成り立つ。ここで線型写像 $\langle \theta - \theta_0, \cdot \rangle: V \rightarrow \mathbb{R}$ は $\text{Ker } \langle \theta_0, \cdot \rangle \supset W^\perp$ をみたすから、図式

$$\begin{array}{ccc} V & \xrightarrow{\langle \theta - \theta_0, \cdot \rangle} & \mathbb{R} \\ \downarrow \pi & \nearrow \tilde{\theta} & \\ \tilde{V} & & \end{array} \quad (2.34)$$

を可換にする線型写像 $\tilde{\theta}: \tilde{V} \rightarrow \mathbb{R}$ すなわち線型形式 $\tilde{\theta} \in \tilde{V}^\vee$ が存在する。この $\tilde{\theta}$ が条件 (E3) をみたすものであることを確かめればよいが、各 $x \in \mathcal{X}$ に対し

$$\langle \tilde{\theta}, \tilde{T}(x) \rangle = \langle \theta - \theta_0, T(x) \rangle \quad (2.35)$$

$$= \langle \theta, T(x) \rangle - \langle \theta_0, T(x) \rangle \quad (2.36)$$

が成り立つから

$$p(dx) = \frac{\exp \langle \theta, T(x) \rangle}{\int_X \exp \langle \theta, T(x) \rangle \mu(dx)} \mu(dx) \quad (2.37)$$

$$= \frac{\exp \langle \tilde{\theta}, \tilde{T}(x) \rangle \exp \langle \theta_0, T(x) \rangle}{\int_X \exp \langle \tilde{\theta}, \tilde{T}(x) \rangle \exp \langle \theta_0, T(x) \rangle \mu(dx)} \mu(dx) \quad (2.38)$$

$$= \frac{\exp \langle \tilde{\theta}, \tilde{T}(x) \rangle}{\int_X \exp \langle \tilde{\theta}, \tilde{T}(x) \rangle \tilde{\mu}(dx)} \tilde{\mu}(dx) \quad (2.39)$$

となる。したがって条件 (E3) の成立が確かめられた。以上より $(\tilde{V}, \tilde{T}, \tilde{\mu})$ は \mathcal{P} の実現である。これで証明が完了した。 \square

以上の補題を用いて最小次元実現の特徴づけが得られる。

定理 2.14 (最小次元実現の特徴づけ). \mathcal{P} の実現 (V, T, μ) に関する次の条件は同値である:

- (1) (V, T, μ) は \mathcal{P} の最小次元実現である。
- (2) (V, T, μ) は条件 A と条件 B をみたす。

証明 (1) \Rightarrow (2) 最小次元実現 (V, T, μ) が条件 A, B のいずれかをみたさなかったとすると、補題 2.12, 2.13 よりとくに (V, T, μ) よりも次元の小さい実現が存在することになり、矛盾が従う。

(2) \Rightarrow (1) (V, T, μ) が条件 A と条件 B をみたすとする。 \mathcal{P} の任意の実現 (V', T', μ') に対し、命題 2.10 より全射線型写像 $L: V' \rightarrow V$ が存在するから、 $\dim V \leq \dim V'$ である。したがって V は \mathcal{P} の最小次元実現である。 \square

例 2.15 (正規分布族の最小次元実現). 定理 2.14 により、0425_資料.pdf の例 3.2 でみた正規分布族の例は最小次元実現であることがわかる。実際、 $T(x) = {}^t(x, x^2)$ の像は \mathbb{R}^2 のいかなる真アフィン部分空間にも a.e. で含まれることはないから、条件 A (3) が成り立つ。また、 $\Theta^{\mathcal{P}} = \mathbb{R} \times \mathbb{R}_{<0}$ となることから条件 B も成り立つ。

本節の目標の定理を示す。

定理 2.16 (最小次元実現の間のアフィン変換). $(V, T, \mu), (V', T', \mu')$ がともに最小次元実現ならば、 (V, T, μ) から (V', T', μ') への射 (L, b) がただひとつ存在する。さらに、 L は線型同型写像である。

証明 命題 2.9, 2.10 より、射 $(L, b): (V, T, \mu) \rightarrow (V', T', \mu')$ はただひとつ存在する。また、命題 2.10 より存在する射 $(V', T', \mu') \rightarrow (V, T, \mu)$ をひとつ選んで (K, c) とおくと、合成射 $(K, c) \circ (L, b), (L, b) \circ (K, c)$ は命題 2.9 より恒等射 $(\text{id}_V, 0), (\text{id}_{V'}, 0)$ に一致する。したがって L は線型同型写像である。 \square

同じことを圏の言葉を使わずに言い換えると次のようになる。

定理 2.17 (最小次元実現の間のアフィン変換). $(V, T, \mu), (V', T', \mu')$ を $\mathbf{C}_{\mathcal{P}}$ の対象とする。このとき、 $(V, T, \mu), (V', T', \mu')$ がともに最小次元実現ならば、全射線型写像 $L: V \rightarrow V'$ とベクトル $b \in V'$ であって

$$\cancel{T(x) = L(T'(x)) + b} \quad T'(x) = L(T(x)) + b \quad \mu\text{-a.e. } x \quad (2.40)$$

をみたすものがただひとつ存在する。さらに、 L は線型同型写像である。 \square

系 2.18 (自然パラメータの変換). 上の定理の状況で、さらに $\theta^0 \in V^V$ であって

$$\cancel{\theta'(p) = L(\theta(p)) + \theta^0} \quad \theta(p) = {}^tL(\theta'(p)) + \theta^0 \quad (\forall p \in \mathcal{P}) \quad (2.41)$$

をみたすものがただひとつ存在する。ただし写像 $\theta: \mathcal{P} \rightarrow \Theta^{\mathcal{P}}$ および $\theta': \mathcal{P} \rightarrow \Theta^{\mathcal{P}}$ は P, P' の $\Theta^{\mathcal{P}}, \Theta^{\mathcal{P}}$ 上への制限の逆写像である。

証明 Step 1: 一意性 θ^0 が $(V, T, \mu), (V', T', \mu')$ に対し一意であることは L, θ, θ' の一意性より明らかである。

Step 2: 存在 $q \in \mathcal{P}$ をひとつ選んで $\theta^0 := -{}^tL(\theta(q)) + \theta'(q) \in V^V$ と定め、この θ^0 が (2.41) をみたすことを示せばよい。そこで $p \in \mathcal{P}$ を任意とすると、補題 2.11 より

$$\begin{aligned} & \langle \theta(p) - \theta(q), T(x) \rangle - \psi(\theta(p)) + \psi(\theta(q)) \\ &= \langle \theta'(p) - \theta'(q), T'(x) \rangle - \psi'(\theta'(p)) + \psi'(\theta'(q)) \end{aligned} \quad \mu\text{-a.e. } x \quad (2.42)$$

が成り立ち、さらに (2.40) より

$$\begin{aligned} & \langle \theta(p) - \theta(q), L(T(x)) + b \rangle - \psi(\theta(p)) + \psi(\theta(q)) \\ &= \langle \theta'(p) - \theta'(q), T'(x) \rangle - \psi'(\theta'(p)) + \psi'(\theta'(q)) \end{aligned} \quad \mu\text{-a.e. } x \quad (2.43)$$

が成り立つから、式を整理して

$$\begin{aligned} & \langle {}^tL(\theta(p) - \theta(q)) - (\theta'(p) - \theta'(q)), T'(x) \rangle \\ &= -\langle \theta(p) - \theta(q), b \rangle + \psi(\theta(p)) - \psi(\theta(q)) - \psi'(\theta'(p)) + \psi'(\theta'(q)) \end{aligned} \quad \mu\text{-a.e. } x \quad (2.44)$$

となる。この右辺は x によらないから、 (V', T', μ') の条件 A (2) より

$${}^tL(\theta(p) - \theta(q)) - \theta'(p) - \theta'(q) = 0 \quad (2.45)$$

$$\therefore {}^tL(\theta(p)) - \theta'(p) = {}^tL(\theta(q)) - \theta'(q) = -\theta^0 \quad (2.46)$$

$$\therefore {}^tL(\theta(p)) + \theta^0 = \theta'(p) \quad (2.47)$$

が成り立つ。 $p \in \mathcal{P}$ は任意であったから、(2.41) の成立が示された。 \square

3 対数分配関数

[TODO] 一般化した命題を使って証明を修正する

本節では \mathcal{X} を可測空間、 $\mathcal{P} \subset \mathcal{P}(\mathcal{X})$ を \mathcal{X} 上の指数型分布族、 (V, T, μ) を \mathcal{P} の次元 m の実現、 $\Theta \subset V^\vee$ を自然パラメータ空間、 $\psi: \Theta \rightarrow \mathbb{R}$ を対数分配関数とする。 V^\vee における Θ の内部を Θ° と書くことにする。さらに関数 $h: \mathcal{X} \times \Theta \rightarrow \mathbb{R}$ および $\lambda: \Theta \rightarrow \mathbb{R}$ を

$$h(x, \theta) := e^{\langle \theta, T(x) \rangle} \quad ((x, \theta) \in \mathcal{X} \times \Theta) \quad (3.1)$$

$$\lambda(\theta) := \int_{\mathcal{X}} h(x, \theta) \mu(dx) \quad (\theta \in \Theta) \quad (3.2)$$

と定める (つまり $\psi(\theta) = \log \lambda(\theta)$ である)。

本節の目標は次の定理を示すことである。

定理 3.1 (λ と ψ の C^∞ 性と積分記号下の微分). $\varphi = (\theta_1, \dots, \theta_m): \Theta^\circ \rightarrow \mathbb{R}^m$ を Θ° 上のチャートとする。このとき、任意の $k \in \mathbb{Z}_{\geq 1}$, $i_1, \dots, i_k \in \{1, \dots, m\}$ に対し、

$$\partial_{i_k} \cdots \partial_{i_1} \lambda(\theta) = \int_{\mathcal{X}} \partial_{i_k} \cdots \partial_{i_1} h(x, \theta) \mu(dx) \quad (\theta \in \Theta^\circ) \quad (3.3)$$

が成り立つ (∂_i は $\frac{\partial}{\partial \theta_i} \in \Gamma(T\Theta^\circ)$ の略記)。ただし、左辺の微分可能性および右辺の可積分性も定理の主張に含まれる。とくに λ および ψ は Θ° 上の C^∞ 関数である。

定理 3.1 の証明には次の事実を用いる。

事実 3.2 (積分記号下の微分). \mathcal{Y} を可測空間、 ν を \mathcal{Y} 上の測度、 $I \subset \mathbb{R}$ を开区間、 $f: \mathcal{Y} \times I \rightarrow \mathbb{R}$ を

- (i) 各 $t \in I$ に対し $f(\cdot, t): \mathcal{Y} \rightarrow \mathbb{R}$ が可測
- (ii) 各 $y \in \mathcal{Y}$ に対し $f(y, \cdot): I \rightarrow \mathbb{R}$ が微分可能

をみたす関数とする。このとき、 f に関する条件

- (1) 各 $t \in I$ に対し $f(\cdot, t) \in L^1(\mathcal{Y}, \nu)$ である。
- (2) ある ν -可積分関数 $\Phi: \mathcal{Y} \rightarrow \mathbb{R}$ が存在し、すべての $t' \in I$ に対し $\left| \frac{\partial f}{\partial t}(y, t') \right| \leq \Phi(y)$ a.e. y である。

が成り立つならば、関数 $I \rightarrow \mathbb{R}, t \mapsto \int_{\mathcal{Y}} f(y, t) \nu(dy)$ は微分可能で、

$$\frac{\partial}{\partial t} \int_{\mathcal{Y}} f(y, t) \nu(dy) = \int_{\mathcal{Y}} \frac{\partial f}{\partial t}(y, t) \nu(dy) \quad (3.4)$$

が成り立つ。 □

定理 3.1 の証明において最も重要なステップは、事実 3.2 の前提が満たされることの確認である。そのための補題を次に示す。

補題 3.3 (優関数の存在). e^i ($i = 1, \dots, m$) を V^\vee の基底とし、この基底が定める Θ° 上のチャートを $\varphi = (\theta_1, \dots, \theta_m): \Theta^\circ \rightarrow \mathbb{R}^m$ とおく。このとき、任意の $k \in \mathbb{Z}_{\geq 1}$, $i_1, \dots, i_k \in \{1, \dots, m\}$ に対し、次が成り立つ:

- (1) 任意の $\theta \in \Theta^\circ$ に対し、関数 $\partial_{i_k} \cdots \partial_{i_1} h(\cdot, \theta): X \rightarrow \mathbb{R}$ は $L^1(X, \mu)$ に属する。
- (2) 任意の $\theta \in \Theta^\circ$ に対し、 Θ° における θ のある近傍 U と、ある μ -可積分関数 $\Phi: X \rightarrow \mathbb{R}$ が存在し、すべての $\theta' \in U$ に対し $|\partial_{i_k} \cdots \partial_{i_1} h(x, \theta')| \leq \Phi(x)$ a.e. x が成り立つ。

証明 (1) は (2) より直ちに従うから、(2) を示す。そこで $\theta \in \Theta^\circ$ を任意とする。補題の主張は座標 $\theta_1, \dots, \theta_m$ を平行移動して考えても等価だから、点 θ の座標は $\varphi(\theta) = 0 \in \mathbb{R}^m$ であるとしてよい。

Step 1: U の構成 $\varepsilon > 0$ を十分小さく選び、 \mathbb{R}^m 内の閉立方体

$$A_{2\varepsilon} := \prod_{i=1}^m [-2\varepsilon, 2\varepsilon] \quad A_\varepsilon := \prod_{i=1}^m [-\varepsilon, \varepsilon] \quad (3.5)$$

が $\varphi(\Theta^\circ)$ に含まれるようにしておく。すると $U := \varphi^{-1}(\text{Int } A_\varepsilon) \subset \varphi(\Theta^\circ)$ は θ の近傍となるが、これが求める U の条件を満たすことを示す。

Step 2: h の座標表示 まず具体的な計算のために h の座標表示を求める。いま各 $\theta' \in U$ に対し

$$h(x, \theta') = \exp\langle \theta', T(x) \rangle = \exp\langle \theta_i(\theta') e^i, T(x) \rangle = \exp\left(\theta_i(\theta') T^i(x)\right) \quad (3.6)$$

が成り立っている。ただし $T^i: X \rightarrow \mathbb{R}, x \mapsto \langle e^i, T(x) \rangle$ ($i = 1, \dots, m$) とおいた。したがって

$$\partial_{i_k} \cdots \partial_{i_1} h(x, \theta') = T^{i_1}(x) \cdots T^{i_k}(x) \exp\left(\theta_i(\theta') T^i(x)\right) \quad (3.7)$$

と表せることがわかる。

Step 3: Φ の構成 Φ を構成するため、式 (3.7) の絶対値を上から評価する。表記の簡略化のため $t' := (t'_1, \dots, t'_m) := \varphi(\theta') \in \mathbb{R}^m$ とおいておく。まず $\frac{k+1}{\varepsilon} \frac{\varepsilon}{k+1} = 1$ より

$$\left| T^{i_1}(x) \cdots T^{i_k}(x) \exp\left(\sum_{i=1}^m t'_i T^i(x)\right) \right| = \left(\frac{k+1}{\varepsilon}\right)^k \left(\prod_{a=1}^k \frac{\varepsilon}{k+1} |T^{i_a}(x)|\right) \exp\left(\sum_{i=1}^m t'_i T^i(x)\right) \quad (3.8)$$

であり、 \prod の部分を評価すると

$$\prod_{\alpha=1}^k \frac{\varepsilon}{k+1} |T^{i_\alpha}(x)| \leq \prod_{\alpha=1}^k \left(\exp\left(\frac{\varepsilon}{k+1} T^{i_\alpha}(x)\right) + \exp\left(-\frac{\varepsilon}{k+1} T^{i_\alpha}(x)\right) \right) \quad (\because s \leq e^s + e^{-s} \ (s \in \mathbb{R})) \quad (3.9)$$

$$= \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^k \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x)\right) \quad (\because \text{式の展開}) \quad (3.10)$$

(ただし σ_α は σ の第 α 成分) となるから、式 (3.8) と式 (3.10) を合わせて

$$(3.8) \leq C \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^k \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x)\right) \exp\left(\sum_{i=1}^m t'_i T^i(x)\right) \quad (3.11)$$

$$= C \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^k \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x) + \sum_{i=1}^m t'_i T^i(x)\right) \quad (3.12)$$

となる。ただし $C := \left(\frac{k+1}{\varepsilon}\right)^k \in \mathbb{R}_{>0}$ とおいた。ここで最終行の \exp の中身について、各 $i = 1, \dots, m$ に対し $T^i(x)$ の係数を評価することで、ある $t'' \in A_{2\varepsilon}$ が存在して

$$(3.12) = C \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{i=1}^m t''_i T^i(x)\right) = 2^k C \exp\left(\sum_{i=1}^m t''_i T^i(x)\right) \quad (3.13)$$

と表せることがわかる。そこで $|t''_i| \leq 2\varepsilon$ ($i = 1, \dots, m$) より

$$(3.13) \leq 2^k C \prod_{i=1}^m \left(\exp\left(2\varepsilon T^i(x)\right) + \exp\left(-2\varepsilon T^i(x)\right) \right) \quad (3.14)$$

$$= 2^k C \sum_{\tau \in \{\pm 1\}^m} \exp\left(\sum_{i=1}^m 2\varepsilon \tau_i T^i(x)\right) \quad (3.15)$$

を得る。この右辺は (t' によらないから) θ' によらない \mathcal{X} 上の関数であり、また \sum の各項が $2\varepsilon \tau \in A_{2\varepsilon}$ ゆえに μ -可積分だから式全体も μ -可積分である。したがってこれが求める優関数である。 \square

目標の定理 3.1 を証明する。

定理 3.1 の証明. 定理 3.1 のステートメントで与えられているチャート $\varphi = (\theta_1, \dots, \theta_m)$ は (V^\vee の基底が定めるものとは限らない) 任意のものであるが、実は定理の主張を示すには、 V^\vee の基底をひとつ選び、その基底が定めるチャート $\tilde{\varphi} = (\tilde{\theta}_1, \dots, \tilde{\theta}_m)$ に対して定理の主張を示せば十分である。その理由は次である:

- 式 (3.3) の左辺の微分可能性は、 λ が C^∞ であればよいから、チャート $\tilde{\varphi}$ で考えれば十分。
- 式 (3.3) の右辺の可積分性および式 (3.3) の等号の成立については、Leibniz 則より、 λ の $\tilde{\theta}_1, \dots, \tilde{\theta}_m$ に関する k 回偏導関数が、 λ の $\theta_1, \dots, \theta_m$ に関する k 回以下の偏導関数たちの (x によらない) $C^\infty(\Theta^\circ)$ -係数の線型結合に書けることから従う。

そこで、以降 φ は V^\vee の基底が定めるチャートとする。

補題 3.3 (1) より、式 (3.3) の右辺の可積分性はわかっている。よって、残りの示すべきことは

- (i) 式 (3.3) の左辺の微分可能性

(ii) 式 (3.3) の等号の成立

の 2 点である。

まず $k = 1, i_k = 1$ の場合に (i), (ii) が成り立つことを示す。そのためには、 $t = (t_1, \dots, t_m) \in \varphi(\Theta^\circ)$ を任意に固定したとき、 t_1 を含む \mathbb{R} の十分小さな开区間 I が存在して、関数

$$g: X \times I \rightarrow \mathbb{R}, \quad (x, s) \mapsto h(x, \varphi^{-1}(s, t_2, \dots, t_m)) \quad (3.16)$$

が事実 3.2 の仮定 (1), (2) をみたすことをいばよい。

いま $\varphi^{-1}(t) \in \Theta^\circ$ だから、補題 3.3(2) のいう Θ° における $\varphi^{-1}(t)$ の近傍 U と μ -可積分関数 $\Phi: X \rightarrow \mathbb{R}$ が存在する。このとき $\varphi(U)$ は \mathbb{R}^m における t の近傍となるから、 t_1 を含む \mathbb{R} の十分小さな开区間 I が存在して

$$I \times \{t_2\} \times \dots \times \{t_m\} \subset \varphi(U) \quad (3.17)$$

が成り立つ。この I を用いて定まる関数 g が事実 3.2 の仮定 (1), (2) をみたすことを確認する。

まず補題 3.3 の結果 (1) より、 g は事実 3.2 の仮定 (1) をみたす。また補題 3.3 の結果 (2) より、 g は事実 3.2 の仮定 (2) をみたす。したがって $k = 1, i_k = 1$ の場合について (i), (ii) が示された。

同様に $i_k = 2, \dots, m$ の場合についても示される。以降、 k に関する帰納法で、すべての $k \in \mathbb{Z}_{\geq 1}$ および $i_1, \dots, i_k \in \{1, \dots, m\}$ に対して示される。これで定理の証明が完了した。 \square

定理 3.1 から次の系が従う。

系 3.4. $\varphi = (\varphi_1, \dots, \varphi_m): \Theta^\circ \rightarrow \mathbb{R}^m$ を V^\vee の基底が定めるチャートとする。また、各 $\theta \in \Theta$ に対し、 X 上の確率測度 P_θ を $P_\theta(dx) = e^{\langle \theta, T(x) \rangle - \psi(\theta)} \mu(dx)$ と定める。このとき、任意の $k \in \mathbb{Z}_{\geq 1}$, $i_1, \dots, i_k \in \{1, \dots, m\}$ に対し、

$$E_{P_\theta}[T^{i_k}(x) \cdots T^{i_1}(x)] = \frac{\partial_{i_k} \cdots \partial_{i_1} \lambda(\theta)}{\lambda(\theta)} \quad (\theta \in \Theta^\circ) \quad (3.18)$$

が成り立つ。ただし、左辺の期待値の存在も系の主張に含まれる。 \square

4 Fisher 計量

Fisher 計量を定義する。

命題-定義 4.1 (Fisher 計量). ψ を Θ° 上の C^∞ 関数とみなすと、各 $\theta \in \Theta^\circ$ に対し $(\text{Hess } \psi)_\theta \in T_\theta^{(0,2)} \Theta^\circ$ は $\text{Var}_{P_\theta}[T]$ に一致する。さらに (V, T, μ) が条件 A をみたすならば、 $\text{Hess } \psi$ は正定値である。

したがって (V, T, μ) が条件 A をみたすとき、 $\text{Hess } \psi$ は Θ° 上の Riemann 計量となり、これを ψ の定める **Fisher 計量 (Fisher metric)** という。

証明 まず $(\text{Hess } \psi)_\theta = \text{Var}_{P_\theta}[T]$ ($\theta \in \Theta^\circ$) を示す。 Θ° 上の D -アファイン座標 θ^i ($i = 1, \dots, m$) をひとつ選ぶと、??より、座標 θ^i に関する $\text{Hess } \psi$ の成分表示は $\text{Hess } \psi = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} d\theta^i \otimes d\theta^j$ となる。ここで前回 (0516_資料.pdf) の系 2.4 より

$$\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\theta) = \partial_i \partial_j \log \lambda(\theta) \quad (4.1)$$

$$= \partial_i \left(\frac{\partial_j \lambda(\theta)}{\lambda(\theta)} \right) \quad (4.2)$$

$$= \frac{\partial_i \partial_j \lambda(\theta)}{\lambda(\theta)} - \frac{\partial_i \lambda(\theta) \partial_j \lambda(\theta)}{\lambda(\theta)^2} \quad (4.3)$$

$$= E[T^i(x)T^j(x)] - E[T^i(x)]E[T^j(x)] \quad (4.4)$$

$$= E[(T^i(x) - E[T^i(x)])(T^j(x) - E[T^j(x)])] \quad (4.5)$$

を得る。ただし $E[\cdot]$ は P_θ に関する期待値 $E_{P_\theta}[\cdot]$ の略記である。したがって $\text{Hess}_\theta \psi = \text{Var}_{P_\theta}[T]$ が成り立つ。

次に、 (V, T, μ) が条件 A をみたすとし、 $\text{Hess} \psi$ が正定値であることを示す。すなわち、各 $\theta \in \Theta^\circ$ に対し $(\text{Hess} \psi)_\theta$ が正定値であることを示す。そのためには各 $u \in V^\vee$ に対し「 $(\text{Hess} \psi)_\theta(u, u) = 0$ ならば $u = 0$ 」を示せばよいが、上で示したことと??より

$$(\text{Hess} \psi)_\theta(u, u) = (\text{Var}_{P_\theta}[T])(u, u) = \langle u \otimes u, \text{Var}_{P_\theta}[T] \rangle = \text{Var}_{P_\theta}[\langle u, T(x) \rangle] \quad (4.6)$$

と式変形できるから、 $(\text{Hess} \psi)_\theta(u, u) = 0$ ならば??より $\langle u, T(x) \rangle$ は a.e. 定数であり、したがって条件 A より $u = 0$ となる。よって $(\text{Hess} \psi)_\theta$ は正定値である。したがって $\text{Hess} \psi$ は正定値である。□

5 Amari-Chentsov テンソルと α -接続

5.1 多様体構造と平坦アファイン接続

命題-定義 5.1 (\mathcal{P} が開であること). 指数型分布族 \mathcal{P} に関し、次は同値である:

- (1) ある最小次元実現 (V, T, μ) に対し、 $\Theta_{(V, T, \mu)}^\mathcal{P}$ は V^\vee で開である。
- (2) すべての最小次元実現 (V, T, μ) に対し、 $\Theta_{(V, T, \mu)}^\mathcal{P}$ は V^\vee で開である。

\mathcal{P} がこれらの同値な 2 条件をみたすとき、 \mathcal{P} は開 (open) であるという。

証明 (1) \Rightarrow (2) は、0606_資料.pdf 系 1.13 より、最小次元実現の真パラメータ空間がアファイン変換で写り合うことから従う。(2) \Rightarrow (1) は最小次元実現が存在することから従う。□

以降、本節では \mathcal{P} は開とする。

命題-定義 5.2 (\mathcal{P} の自然な多様体構造). \mathcal{P} 上の多様体構造 \mathcal{U} であって次をみたすものがただひとつ存在する:

- \mathcal{P} の任意の最小次元実現 (V, T, μ) に対し、 \mathcal{U} は全単射 $\theta_{(V, T, \mu)}$ により $\Theta_{(V, T, \mu)}^\mathcal{P}$ から \mathcal{P} 上に誘導された多様体構造に一致する。

この \mathcal{U} を \mathcal{P} の自然な多様体構造という。

証明 Step 1: \mathcal{U} の一意性 \mathcal{U} の存在を仮定すれば、最小次元実現をひとつ選ぶことで \mathcal{U} が決まるから、 \mathcal{U} は一意である。

Step 2: \mathcal{U} の存在 最小次元実現 (V, T, μ) をひとつ選び、 $\theta := \theta_{(V, T, \mu)}$ とおき、 θ により $\Theta_{(V, T, \mu)}^\mathcal{P}$ から \mathcal{P} 上に誘導された多様体構造を \mathcal{U} とおく。この \mathcal{U} が求めるものであることを示せばよい。示すべきことは、

(V', T', μ') を最小次元実現とし、 $\theta' := \theta_{(V', T', \mu')}$ とおき、 \mathcal{U}' を θ' により $\Theta_{(V', T', \mu')}^{\mathcal{P}}$ から \mathcal{P} 上に誘導された多様体構造とすると、恒等写像 $\text{id}: (\mathcal{P}, \mathcal{U}) \rightarrow (\mathcal{P}, \mathcal{U}')$ が微分同相となることである。これは図式

$$\begin{array}{ccc} (\mathcal{P}, \mathcal{U}) & \xrightarrow{\text{id}} & (\mathcal{P}, \mathcal{U}') \\ \theta \downarrow & & \downarrow \theta' \\ \Theta_{(V, T, \mu)}^{\mathcal{P}} & \xrightarrow{F} & \Theta_{(V', T', \mu')}^{\mathcal{P}} \end{array} \quad (5.1)$$

の可換性と、 θ, θ', F が微分同相であることから従う。ただし F とは、0606_資料.pdf 系 1.13 より一意に存在するアフィン変換 $V^{\vee} \rightarrow V'^{\vee}$ の制限である。□

以降、本節では \mathcal{P} に自然な多様体構造が定まっているものとする。

命題-定義 5.3 (\mathcal{P} 上の自然な平坦アフィン接続). \mathcal{P} 上の平坦アフィン接続 ∇ であって次をみたすものがただひとつ存在する:

- \mathcal{P} の任意の最小次元実現 (V, T, μ) に対し、 $\Theta_{(V, T, \mu)}^{\mathcal{P}}$ 上の標準的な平坦アフィン接続を $\tilde{\nabla}$ とおくと、 ∇ は $\nabla = \theta_{(V, T, \mu)}^* \tilde{\nabla}$ をみたす。

この ∇ を \mathcal{P} 上の**自然な平坦アフィン接続**という。

証明には次の補題を用いる。

補題 5.4 (アフィン変換によるアフィン接続の引き戻し). V, V' を有限次元 \mathbb{R} -ベクトル空間、 $F: V \rightarrow V'$ をアフィン変換、 ∇, ∇' をそれぞれ V, V' 上の標準的な平坦アフィン接続とする。このとき $F^* \nabla' = \nabla$ が成り立つ。

事実 5.5 (ベクトル場の押し出しと関数). M, N を (有限次元実 C^∞) 多様体、 $F: M \rightarrow N$ を微分同相写像とする。このとき、次が成り立つ:

- (1) 任意の $f \in C^\infty(M)$ に対し $F_*(fX) = f \circ F^{-1} F_* X$ が成り立つ。
- (2) 任意の $g \in C^\infty(N)$ に対し $((F_* X)g) \circ F = X(g \circ F)$ が成り立つ。

□

事実 5.6 (アフィン変換によるベクトル場の押し出し). V, V' を m 次元 \mathbb{R} -ベクトル空間、 ∂_i, ∂'_i ($i = 1, \dots, m$) をそれぞれ V, V' の基底をベクトル場とみなしたものの、 $F: V \rightarrow V'$ をアフィン変換とし、 ∂_i, ∂'_i に関する F の行列表示を $(a_j^i)_{i,j}$ とする。このとき、 $F_* \partial_j = a_j^i \partial'_i$ が成り立つ。

□

証明 ∂_i, ∂'_i ($i = 1, \dots, m$) をそれぞれ V, V' の基底をベクトル場とみなしたものとし、 ∂_i, ∂'_i に関する F の行列表示を $(a_j^i)_{i,j}$ とおき、その逆行列を $(\tilde{a}_j^i)_{i,j}$ とおく。任意の $X = X^i \partial_i, Y = Y^j \partial'_j \in \Gamma(TV)$ に対し

$$(F^* \nabla')_X Y = F_*^{-1} \left(\nabla'_{F_* X} F_* Y \right) \quad (5.2)$$

$$= F_*^{-1} \left(\nabla'_{F_* (X^i \partial_i)} F_* (Y^j \partial'_j) \right) \quad (5.3)$$

$$= F_*^{-1} \left(\nabla'_{X^i \circ F^{-1} F_* \partial_i} (Y^j \circ F^{-1} F_* \partial_j) \right) \quad (\text{事実 5.5 (1)}) \quad (5.4)$$

$$= F_*^{-1} \left(\nabla'_{X^i \circ F^{-1} a_i^k \partial'_k} (Y^j \circ F^{-1} a_j^l \partial'_l) \right) \quad (\text{事実 5.6}) \quad (5.5)$$

$$= F_*^{-1} \left(X^i \circ F^{-1} a_i^k a_j^l \nabla'_{\partial'_k} (Y^j \circ F^{-1} \partial'_l) \right) \quad (5.6)$$

$$= F_*^{-1} \left(X^i \circ F^{-1} a_i^k a_j^l \partial'_k (Y^j \circ F^{-1} \partial'_l) \right) \quad (\text{基底 } \partial'_i \text{ の定める座標は } \nabla' \text{-アファイン}) \quad (5.7)$$

$$= F_*^{-1} \left(X^i \circ F^{-1} a_i^k a_j^l ((F_*^{-1} \partial'_k) Y^j) \circ F^{-1} \partial'_l \right) \quad (\text{事実 5.5 (2)}) \quad (5.8)$$

$$= X^i a_i^k a_j^l (F_*^{-1} \partial'_k) (Y^j) F_*^{-1} \partial'_l \quad (\text{事実 5.5 (1)}) \quad (5.9)$$

$$= X^i a_i^k a_j^l \tilde{a}_k^m \partial_m (Y^j) \tilde{a}_l^n \partial_n \quad (\text{事実 5.6}) \quad (5.10)$$

$$= X^i \partial_i (Y^j) \partial_j \quad (5.11)$$

$$= \nabla_X Y \quad (5.12)$$

となる。よって $F^* \nabla' = \nabla$ が成り立つ。 \square

命題-定義 5.3 の証明 Step 1: ∇ の一意性 ∇ の存在を仮定すれば、最小次元実現をひとつ選ぶことで ∇ が決まるから、 ∇ は一意である。

Step 2: ∇ の存在 最小次元実現 (V, T, μ) をひとつ選び、 $\theta := \theta_{(V, T, \mu)}$ 、 $\Theta_{(V, T, \mu)}^{\mathcal{P}}$ 上の標準的な平坦アファイン接続を $\tilde{\nabla}$ 、 $\nabla := \theta^* \tilde{\nabla}$ と定める。この ∇ が求めるものであることを示せばよい。示すべきことは、 (V', T', μ') を最小次元実現とし、 $\theta' := \theta_{(V', T', \mu')}$ 、 $\Theta_{(V', T', \mu')}^{\mathcal{P}}$ 上の標準的な平坦アファイン接続を $\tilde{\nabla}'$ とおくと、 $\theta^* \tilde{\nabla} = \theta'^* \tilde{\nabla}'$ が成り立つことである。そこで、0606_資料.pdf 系 1.13 より一意に存在するアファイン変換 $V^V \rightarrow V'^V$ を F とおくと、

$$\theta'^* \tilde{\nabla}' = \theta^* F^* \tilde{\nabla}' \quad (F \text{ と } \theta, \theta' \text{ の関係}) \quad (5.13)$$

$$= \theta^* \tilde{\nabla} \quad (\text{補題 5.4}) \quad (5.14)$$

が成り立つ。したがって $\theta^* \tilde{\nabla} = \theta'^* \tilde{\nabla}'$ が示された。よって ∇ は命題-定義の主張の条件をみたす。 \square

以降、本節では \mathcal{P} に自然な平坦アファイン接続 ∇ が定まっているものとする。

5.2 Fisher 計量

命題-定義 5.7 (\mathcal{P} 上の Fisher 計量). \mathcal{P} 上の Riemann 計量 g であって次をみたすものがただひとつ存在する:

- \mathcal{P} の任意の最小次元実現 (V, T, μ) に対し、 $\Theta_{(V, T, \mu)}^{\mathcal{P}}$ 上の Fisher 計量を \tilde{g} とおくと、 $g = \theta_{(V, T, \mu)}^* \tilde{g}$ が成り立つ。

これを \mathcal{P} 上の **Fisher 計量** という。

証明には次の補題を用いる。

補題 5.8. $(V, T, \mu), (V', T', \mu')$ を \mathcal{P} の最小次元実現とし、 $\theta := \theta_{(V, T, \mu)}$ 、 $\theta' := \theta_{(V', T', \mu')}$ とおき、 $\Theta_{(V, T, \mu)}^{\mathcal{P}}$ 、 $\Theta_{(V', T', \mu')}^{\mathcal{P}}$ 上の Fisher 計量をそれぞれ g, g' とおき、0606_資料.pdf 定理 1.12 より一意に存在する線型同型写像 $V \rightarrow V'$ を L とおく。このとき、各 $p \in \mathcal{P}$ に対し $g_{\theta(p)} = (L \otimes L)(g'_{\theta'(p)})$ が成り立つ。

証明 L は $T'(x) = L(T(x)) + \text{const.}$ μ -a.e. x をみたし、また各 $p \in \mathcal{P}$ に対し $g_{\theta(p)} = \text{Var}_p[T]$, $g'_{\theta'(p)} = \text{Var}_p[T']$ が成り立つから、期待値と分散のペアリングの命題 (0523_資料.pdf 命題 1.1) と同様の議論により補題の主張の等式が成り立つ。 \square

命題-定義 5.7 の証明 Step 1: g の一意性 g の存在を仮定すれば、最小次元実現をひとつ選ぶことで g が決まるから、 g は一意である。

Step 2: g の存在 最小次元実現 (V, T, μ) をひとつ選び、 $\theta := \theta_{(V, T, \mu)}$ 、 $\Theta_{(V, T, \mu)}^{\mathcal{P}}$ 上の Fisher 計量を \tilde{g} とおき、 $g := \theta^* \tilde{g}$ と定める。この g が求めるものであることを示せばよい。示すべきことは、 (V', T', μ') を最小次元実現とし、 $\theta' := \theta_{(V', T', \mu')}$ 、 $\Theta_{(V', T', \mu')}^{\mathcal{P}}$ 上の Fisher 計量を \tilde{g}' とおいて、 $\theta^* g = \theta'^* g'$ が成り立つことである。そこで 0606_資料.pdf 定理 1.12 より一意に存在する線型同型写像 $V \rightarrow V'$ を L とおくと、各 $p \in \mathcal{P}$, $u, v \in T_p \mathcal{P}$ に対し

$$(\theta^* g)_p(u, v) = g_{\theta(p)}(d\theta_p(u), d\theta_p(v)) \quad (5.15)$$

$$= \langle g_{\theta(p)}, d\theta_p(u) \otimes d\theta_p(v) \rangle \quad (5.16)$$

$$= \langle (L \otimes L) g'_{\theta'(p)}, d\theta_p(u) \otimes d\theta_p(v) \rangle \quad (\text{補題 5.8}) \quad (5.17)$$

$$= \langle g'_{\theta'(p)}, {}^t L \circ d\theta_p(u) \otimes {}^t L \circ d\theta_p(v) \rangle \quad (5.18)$$

$$= \langle g'_{\theta'(p)}, d({}^t L \circ \theta)_p(u) \otimes d({}^t L \circ \theta)_p(v) \rangle \quad (5.19)$$

$$= \langle g'_{\theta'(p)}, d\theta'_p(u) \otimes d\theta'_p(v) \rangle \quad (L \text{ と } \theta, \theta' \text{ の関係}) \quad (5.20)$$

$$= g'_p(d\theta'_p(u), d\theta'_p(v)) \quad (5.21)$$

$$= (\theta'^* g')_p(u, v) \quad (5.22)$$

が成り立つ。したがって $\theta^* g = \theta'^* g'$ が示された。よって g は命題-定義の主張の条件をみたす。 \square

以降、本節では \mathcal{P} に Fisher 計量 g が定まっているものとする。

5.3 Amari-Chentsov テンソルと α -接続

定義 5.9 (Amari-Chentsov テンソル). \mathcal{P} 上の $(0, 3)$ -テンソル場 S を $S := \nabla g$ で定め、これを \mathcal{P} 上の **Amari-Chentsov テンソル (Amari-Chentsov tensor)** という。また、 \mathcal{P} 上の $(1, 2)$ -テンソル場 A を次の関係式により定める:

$$g(A(X, Y), Z) = S(X, Y, Z) \quad (\forall X, Y, Z \in \Gamma(T\mathcal{P})) \quad (5.23)$$

以降、「Amari-Chentsov テンソル」を「AC テンソル」と略記することがある。

以降、本節では \mathcal{P} に Amari-Chentsov テンソル S が定まっているものとする。

命題 5.10 (AC テンソルの成分). (V, T, μ) を \mathcal{P} の最小次元実現、 $\Theta^{\mathcal{P}} := \Theta_{(V, T, \mu)}^{\mathcal{P}}$, $\theta := \theta_{(V, T, \mu)}$ 、 (V, T, μ) の対数分配関数を ψ とおく。このとき、 \mathcal{P} 上の任意の ∇ -アファイン座標 $x := (x^1, \dots, x^m): \mathcal{P} \rightarrow \mathbb{R}^m$ に対し、

$\varphi := (\varphi^1, \dots, \varphi^m) := x \circ \theta^{-1}: \Theta^{\mathcal{P}} \rightarrow \mathbb{R}^m$ とおくと、 S の成分は

$$S_{ijk}(p) = \frac{\partial^3 \psi}{\partial \varphi^i \partial \varphi^j \partial \varphi^k}(\theta(p)) = E_p [(T_i - E_p[T_i])(T_j - E_p[T_j])(T_k - E_p[T_k])] \quad (5.24)$$

をみたす。ただし T_i ($i = 1, \dots, m$) とは、同一視 $V = V^{\vee\vee} = T_{\theta(p)}^{\vee} \Theta^{\mathcal{P}}$ により $d\varphi^i$ ($i = 1, \dots, m$) を V の基底とみなしたときの T の成分である。

証明 左側の等号と右側の等号についてそれぞれ示す。

Step 1: 左側の等号 $\Theta^{\mathcal{P}}$ 上の標準的な平坦アフライン接続を $\tilde{\nabla}$ とおき、 ψ の定める $\Theta^{\mathcal{P}}$ 上の Fisher 計量を \tilde{g} とおくと、

$$S\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right) = \left(\nabla_{\frac{\partial}{\partial x^i}} g\right)\left(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right) \quad (5.25)$$

$$= \left(\left(\theta^* \tilde{\nabla}\right)_{\frac{\partial}{\partial x^i}} (\theta^* \tilde{g})\right)\left(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right) \quad (5.26)$$

$$= \left(\theta_*^{-1} \left(\tilde{\nabla}_{\theta_* \frac{\partial}{\partial x^i}} \tilde{g}\right)\right)\left(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right) \quad (5.27)$$

$$= \left(\tilde{\nabla}_{\theta_* \frac{\partial}{\partial x^i}} \tilde{g}\right)\left(d\theta\left(\frac{\partial}{\partial x^j}\right), d\theta\left(\frac{\partial}{\partial x^k}\right)\right) \quad (5.28)$$

$$= \left(\tilde{\nabla}_{\frac{\partial}{\partial \varphi^i}} \tilde{g}\right)\left(\frac{\partial}{\partial \varphi^j}, \frac{\partial}{\partial \varphi^k}\right) \quad (5.29)$$

$$= \left(\frac{\partial}{\partial \varphi^i} \left(\frac{\partial^2 \psi}{\partial \varphi^l \partial \varphi^n}\right) d\varphi^l d\varphi^n\right)\left(\frac{\partial}{\partial \varphi^j}, \frac{\partial}{\partial \varphi^k}\right) \quad (\varphi \text{ は } \tilde{\nabla}\text{-アフライン座標}) \quad (5.30)$$

$$= \frac{\partial^3 \psi}{\partial \varphi^i \partial \varphi^j \partial \varphi^k} \quad (5.31)$$

となるから、命題の主張の左側の等号が従う。

Step 2: 右側の等号 「 E_p 」の下付きの p を省略して書けば、直接計算より

$$E[(T_i - E[T_i])(T_j - E[T_j])(T_k - E[T_k])] \quad (5.32)$$

$$= E[T_i T_j T_k] - E[T_i]E[T_j T_k] - E[T_j]E[T_k T_i] - E[T_k]E[T_i T_j] + 2E[T_i]E[T_j]E[T_k] \quad (5.33)$$

が成り立つ。一方、 $\lambda := \exp \psi$ とおき、 $\frac{\partial}{\partial \varphi^i}$ を ∂_i と略記すれば、直接計算より

$$\frac{\partial^3 \psi}{\partial \varphi^i \partial \varphi^j \partial \varphi^k} = \partial_i \partial_j \partial_k \log \lambda \quad (5.34)$$

$$= \frac{\partial_i \partial_j \partial_k \lambda}{\lambda} - \frac{(\partial_i \lambda)(\partial_j \partial_k \lambda)}{\lambda^2} - \frac{(\partial_j \lambda)(\partial_k \partial_i \lambda)}{\lambda^2} - \frac{(\partial_k \lambda)(\partial_i \partial_j \lambda)}{\lambda^2} + 2 \frac{(\partial_i \lambda)(\partial_j \lambda)(\partial_k \lambda)}{\lambda^3} \quad (5.35)$$

が成り立つ。この右辺を 0516_資料.pdf 系 2.4 により期待値の形で表せば式 (5.33) に一致するから、命題の主張の右側の等号が従う。 \square

定義 5.11 (α -接続). $\alpha \in \mathbb{R}$ とする。 \mathcal{P} 上のアフライン接続 $\nabla^{(\alpha)}$ を次の関係式により定める:

$$g(\nabla_X^{(\alpha)} Y, Z) = g(\nabla_X^{(g)} Y, Z) - \frac{\alpha}{2} S(X, Y, Z) \quad (X, Y, Z \in \Gamma(T\mathcal{P})) \quad (5.36)$$

この $\nabla^{(\alpha)}$ を (g, S) の定める α -接続 (α -connection) という。とくに $\alpha = 1, -1$ の場合をそれぞれ **e-接続** (**e-connection**)、**m-接続** (**m-connection**) という。

命題 5.12 ($\nabla^{(g)}, \nabla^{(\alpha)}$ の AC テンソルによる表示). \mathcal{P} 上の任意の ∇ -アファイン座標に関し、 $\nabla^{(g)}$ および $\nabla^{(\alpha)}$ の接続係数は次をみたす:

$$(1) \quad \Gamma_{ij}^{(g)k} = \frac{1}{2} A_{ij}^k, \quad \Gamma_{ijk}^{(g)} = \frac{1}{2} S_{ijk} \quad (5.37)$$

$$(2) \quad \text{すべての } \alpha \in \mathbb{R} \text{ に対し} \quad \Gamma_{ij}^{(\alpha)k} = \frac{1-\alpha}{2} A_{ij}^k, \quad \Gamma_{ijk}^{(\alpha)} = \frac{1-\alpha}{2} S_{ijk} \quad (5.38)$$

とくに $\alpha = 1$ のとき $\Gamma_{ij}^{(1)k} = 0, \Gamma_{ijk}^{(1)} = 0$ である。

証明 (1) (5.37) の左側の等式は

$$\Gamma_{ij}^{(g)k} = \frac{1}{2} g^{kl} (\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij}) \quad (5.39)$$

$$= \frac{1}{2} g^{kl} (S_{ijl} + S_{jli} - S_{lij}) \quad (\text{命題 5.10}) \quad (5.40)$$

$$= \frac{1}{2} g^{kl} S_{ijl} \quad (5.41)$$

$$= \frac{1}{2} A_{ij}^k \quad (5.42)$$

より従う。 g で添字を下げて (5.37) の右側の等式も従う。

(2) α -接続の定義より $\Gamma_{ijk}^{(\alpha)} = \Gamma_{ijk}^{(g)} - \frac{\alpha}{2} S_{ijk}$ だから、(1) とあわせて (5.38) の左側の等式が従う。 g で添字を下げて (5.37) の右側の等式も従う。 \square

命題 5.13 (振率と曲率の AC テンソルによる表示). \mathcal{P} 上の任意の ∇ -アファイン座標に関し、 $\nabla^{(\alpha)}$ の振率テンソル $T^{(\alpha)}$ および (1,3)-曲率テンソル $R^{(\alpha)}$ の成分表示は次をみたす:

$$(1) \quad \text{すべての } \alpha \in \mathbb{R} \text{ に対し} \quad T_{ij}^{(\alpha)k} = 0 \quad (5.43)$$

$$(2) \quad \text{すべての } \alpha \in \mathbb{R} \text{ に対し} \quad R_{ijk}^{(\alpha)l} = \frac{1-\alpha}{2} (\partial_i A_{jk}^l - \partial_j A_{ik}^l) + \left(\frac{1-\alpha}{2} \right)^2 (A_{jk}^m A_{im}^l - A_{ik}^m A_{jm}^l) \quad (5.44)$$

とくに $\alpha = 1$ のとき $R_{ijk}^{(1)l} = 0$ である。

証明 (1)

$$T_{ij}^{(\alpha)} = \Gamma_{ij}^{(\alpha)k} - \Gamma_{ji}^{(\alpha)k} \quad (5.45)$$

$$= \frac{1-\alpha}{2} A_{ij}^k - \frac{1-\alpha}{2} A_{ji}^k \quad (\text{命題 5.12(2)}) \quad (5.46)$$

$$= 0 \quad (A_{ij}^k = A_{ji}^k) \quad (5.47)$$

より従う。

(2)

$$R^{(\alpha)l}_{ijk} = \partial_i \Gamma^{(\alpha)l}_{jk} - \partial_j \Gamma^{(\alpha)l}_{ik} + \Gamma^{(\alpha)m}_{jk} \Gamma^{(\alpha)l}_{im} - \Gamma^{(\alpha)m}_{ik} \Gamma^{(\alpha)l}_{jm} \quad (5.48)$$

$$= \frac{1-\alpha}{2} (\partial_i A^l_{jk} - \partial_j A^l_{ik}) + \left(\frac{1-\alpha}{2} \right)^2 (A^m_{jk} A^l_{im} - A^m_{ik} A^l_{jm}) \quad (\text{命題 5.12(2)}) \quad (5.49)$$

より従う。

□

6 指数型分布族の具体例

6.1 具体例: 有限集合上の full support な確率分布の族

本節では、有限集合上の full support な確率分布の族について、 α -接続に関する測地線方程式を求めている。

設定 6.1 (有限集合上の full support な確率分布の族). $\mathcal{X} := \{1, \dots, n\}$ ($n \in \mathbb{Z}_{\geq 1}$) とし、

$$\mathcal{P} := \left\{ \sum_{i=1}^n p_i \delta^i \in \mathcal{P}(\mathcal{X}) \mid 0 < p_i < 1 \ (i = 1, \dots, n) \right\} \quad (6.1)$$

とおく。ただし δ^i は 1 点 $i \in \mathcal{X}$ での Dirac 測度である。これが \mathcal{X} 上の指数型分布族であることは [0425_資料.pdf](#) 例 3.1 で確かめた。

命題 6.2 (最小次元実現の構成および \mathcal{P} が開であることの確認).

(1) (V, T, γ) を次のように定めると、これは \mathcal{P} の実現となる:

$$V := \mathbb{R}^{n-1}, \quad (6.2)$$

$$T: \mathcal{X} \rightarrow V, \quad k \mapsto {}^t(\delta_{1k}, \dots, \delta_{(n-1)k}), \quad (6.3)$$

$$\gamma: \text{数え上げ測度} \quad (6.4)$$

(2) この実現の対数分配関数 $\psi: \tilde{\Theta} \rightarrow \mathbb{R}$ は $\psi(\theta) = \log \left(1 + \sum_{i=1}^{n-1} \exp \theta^i \right)$ となる。

(3) 写像 $P := P_{(V, T, \gamma)}: \tilde{\Theta} \rightarrow \mathcal{P}(\mathcal{X})$ は次をみたす:

$$P(\theta) = \frac{1}{1 + \sum_{i=1}^{n-1} \exp \theta^i} \left(\sum_{i=1}^{n-1} (\exp \theta^i) \delta^i + \delta^n \right) \quad (6.5)$$

(4) $\Theta = \tilde{\Theta} = V^\vee$ が成り立つ。

(5) 次の写像 $\theta: \mathcal{P} \rightarrow \Theta$ は P の逆写像である:

$$\theta: \mathcal{P} \rightarrow \Theta, \quad \sum_{i=1}^n p_i \delta^i \mapsto \left(\log \frac{p_1}{p_n}, \dots, \log \frac{p_{n-1}}{p_n} \right) \quad (6.6)$$

(6) (V, T, γ) は最小次元実現である。とくに \mathcal{P} は開である。

証明 (1) (V, T, γ) が実現であることは 0425_コメント.pdf 演習問題 0.1 に記した。

(2) 対数分配関数の定義より

$$\psi(\theta) = \log \int_X \exp \langle \theta, T(k) \rangle \gamma(dk) \quad (6.7)$$

$$= \log \sum_{i=1}^n \exp \left(\sum_{j=1}^{n-1} \theta^j \delta_{ji} \right) \quad (6.8)$$

$$= \log \left(\sum_{i=1}^{n-1} \exp \theta^i + 1 \right) \quad (6.9)$$

である。

(3) P の定義より

$$P(\theta) = \exp(\langle \theta, T(k) \rangle - \psi(\theta)) \gamma \quad (6.10)$$

$$= \frac{1}{1 + \sum_{i=1}^{n-1} \exp \theta^i} \exp \left(\sum_{i=1}^{n-1} \theta^i \delta_{ik} \right) \gamma \quad (6.11)$$

$$= \frac{1}{1 + \sum_{i=1}^{n-1} \exp \theta^i} \left(\sum_{i=1}^{n-1} (\exp \theta^i) \delta^i + \delta^n \right) \quad (6.12)$$

である。

(4) 可積分性を考えると明らかに $\tilde{\Theta} = V^\vee$ である。また P が (3) のように表せることから $P(\tilde{\Theta}) \subset \mathcal{P}$ がわかる。したがって $V^\vee = \tilde{\Theta} \subset P^{-1}(\mathcal{P}) = \Theta$ である。よって $\Theta = \tilde{\Theta} = V^\vee$ である。

(5) $P \circ \theta, \theta \circ P$ を直接計算すれば確かめられる。

(6) 最小次元実現の特徴づけを確かめればよい。条件 A(3) が成り立つことは、いま V の任意のアファイン部分空間に対し「 $T(x) \in W$ γ -a.e. x 」と「 $T(x) \in W \forall x$ 」が同値であることから明らか。条件 B が成り立つことは $\Theta = V^\vee$ よりわかる。□

以降、 \mathcal{P} には自然な位相および多様体構造が入っているものとして扱い、 \mathcal{P} 上の自然な平坦アファイン接続を ∇ 、Fisher 計量を g 、 $(0, 3), (1, 2)$ 型の Amari-Chentsov テンソルをそれぞれ S, A とおく。また、 $\theta: \mathcal{P} \rightarrow \Theta$ は多様体 \mathcal{P} の座標とみなす。

注意 6.3 (\mathcal{P} の 2 通りの位相 & 多様体構造). \mathcal{P} 上の位相 & 多様体構造として、 X 上の符号付き測度全体のなすベクトル空間 $\mathcal{S}(X) \cong \mathbb{R}^n$ の部分多様体としてのものと、指数型分布族としての自然なものの 2 通りを考えられるが、これらは互いに一致する。なぜならば、いずれの位相 & 多様体構造に関しても写像 $\theta: \mathcal{P} \rightarrow \Theta$ は微分同相写像だからである。

命題 6.4 (Fisher 計量の成分). 座標 $\theta = (\theta^1, \dots, \theta^{n-1})$ に関する Fisher 計量 g の成分は

$$g_{ij}(p) = \delta_{ij} p_i - p_i p_j \quad (p \in \mathcal{P}, i, j = 1, \dots, n-1) \quad (6.13)$$

となる。

証明 微分同相写像 θ により g を Θ 上のテンソル場とみなして計算すれば、各 $p \in \mathcal{P}$ に対し

$$g_{ij}(p) = (\text{Var}_p[T])(e^i, e^j) \quad (6.14)$$

$$= E_p[(T^i - E_p[T^i])(T^j - E_p[T^j])] \quad (6.15)$$

$$= \sum_{k=1}^n (\delta_{ik} - p_i)(\delta_{jk} - p_j)p_k \quad (6.16)$$

$$= \delta_{ij}p_i - p_i p_j \quad (6.17)$$

が成り立つ。 \square

命題 6.5 (AC テンソルの成分). 座標 θ に関する AC テンソル S の成分は

$$S_{ijk}(p) = p_i \delta_{ij} \delta_{jk} - p_i p_k \delta_{ij} - p_i p_j \delta_{jk} - p_j p_k \delta_{ik} + 2p_i p_j p_k \quad (p \in \mathcal{P}, i, j, k = 1, \dots, n-1) \quad (6.18)$$

となる。

証明 前回 (0613_資料.pdf) の命題 1.9 を用いると

$$S_{ijk}(p) = E_p[(T^i - E_p[T^i])(T^j - E_p[T^j])(T^k - E_p[T^k])] \quad (6.19)$$

となるから、命題 6.4 と同様に直接計算して命題の主張の等式が得られる。 \square

以降、 $n = 3$ の場合を考える。

命題 6.6 ($n = 3$ での g, S, A の計算). 座標 θ に関し、 g の行列表示は

$$(g_{ij})_{i,j} = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 \\ -p_1 p_2 & p_2(1-p_2) \end{pmatrix}, \quad (g^{ij})_{i,j} = \frac{1}{p_3} \begin{pmatrix} \frac{p_3}{p_1} + 1 & 1 \\ 1 & \frac{p_3}{p_2} + 1 \end{pmatrix} \quad (6.20)$$

となる。 S の成分は

$$S_{111} = p_1 - 3p_1^2 + 2p_1^3, \quad (6.21)$$

$$S_{112} = S_{121} = S_{211} = -p_1 p_2 + 2p_1^2 p_2, \quad (6.22)$$

$$S_{122} = S_{212} = S_{221} = -p_1 p_2 + 2p_1 p_2^2, \quad (6.23)$$

$$S_{222} = p_2 - 3p_2^2 + 2p_2^3 \quad (6.24)$$

となる。 A の成分は

$$A_{11}^1 = 1 - 2p_1, \quad A_{11}^2 = 0 \quad (6.25)$$

$$A_{12}^1 = A_{21}^1 = -p_2, \quad A_{12}^2 = A_{21}^2 = -p_1 \quad (6.26)$$

$$A_{22}^1 = 0, \quad A_{22}^2 = 1 - 2p_2 \quad (6.27)$$

となる。

証明 g の行列表示は命題 6.4 よりわかる。その逆行列は直接計算よりわかる。 S の成分は命題 6.5 よりわかる。 A の成分は「 $A_{ij}^k = g^{kl} S_{ijl}$ 」を用いて求める。具体的には以下の行列を直接計算すればわかる:

$$\begin{pmatrix} A_{11}^1 & A_{12}^1 & A_{22}^1 \\ A_{11}^2 & A_{12}^2 & A_{22}^2 \end{pmatrix} = \frac{1}{p_3} \begin{pmatrix} \frac{p_3}{p_1} + 1 & 1 \\ 1 & \frac{p_3}{p_2} + 1 \end{pmatrix} \begin{pmatrix} S_{111} & S_{121} & S_{221} \\ S_{112} & S_{122} & S_{222} \end{pmatrix} \quad (6.28)$$

□

命題 6.7 ($n = 3$ での測地線方程式). 各 $\alpha \in \mathbb{R}$ に対し、座標 θ に関する $\nabla^{(\alpha)}$ -測地線の方程式は

$$\ddot{\theta}^1 = -\frac{1-\alpha}{2} \left(\left(1 - \frac{2 \exp \theta^1}{1 + \exp \theta^1 + \exp \theta^2} \right) (\dot{\theta}^1)^2 - \frac{2 \exp \theta^2}{1 + \exp \theta^1 + \exp \theta^2} \dot{\theta}^1 \dot{\theta}^2 \right) \quad (6.29)$$

$$\ddot{\theta}^2 = -\frac{1-\alpha}{2} \left(-\frac{2 \exp \theta^1}{1 + \exp \theta^1 + \exp \theta^2} \dot{\theta}^1 \dot{\theta}^2 + \left(1 - \frac{2 \exp \theta^2}{1 + \exp \theta^1 + \exp \theta^2} \right) (\dot{\theta}^2)^2 \right) \quad (6.30)$$

となる。とくに $\alpha = 1$ のとき

$$\ddot{\theta}^1 = 0, \quad \ddot{\theta}^2 = 0 \quad (6.31)$$

である。

証明 測地線の方程式

$$\ddot{\theta}^k = -\Gamma_{ij}^k \dot{\theta}^i \dot{\theta}^j \quad (6.32)$$

に、前回 (0613_資料.pdf) の命題 1.11 の等式 $\Gamma_{ij}^{(\alpha)k} = \frac{1-\alpha}{2} A_{ij}^k$ を代入して得られる。 □

6.2 具体例: 正規分布族

本節では、正規分布族について、 α -接続に関する測地線方程式を求めてみる。

設定 6.8 (正規分布族). $\mathcal{X} := \mathbb{R}$ とし、

$$\mathcal{P} := \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \in \mathcal{P}(\mathcal{X}) \mid (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_{>0} \right\} \quad (6.33)$$

とおく。これが \mathcal{X} 上の指数型分布族であることは 0425_資料.pdf 例 3.2 で確かめた。

以降、次の事実をしばしば用いる:

事実 6.9. 次の 2 つの写像は互いに逆な C^∞ 写像である:

$$\mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R} \times \mathbb{R}_{<0}, \quad (\mu, \sigma) \mapsto \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right), \quad (6.34)$$

$$\mathbb{R} \times \mathbb{R}_{<0} \rightarrow \mathbb{R} \times \mathbb{R}_{>0}, \quad (\theta^1, \theta^2) \mapsto \left(-\frac{\theta^1}{2\theta^2}, \sqrt{-\frac{1}{2\theta^2}} \right) \quad (6.35)$$

□

命題 6.10 (最小次元実現の構成および \mathcal{P} が開であることの確認).

(1) (V, T, λ) を次のように定めると、これは \mathcal{P} の実現となる:

$$V = \mathbb{R}^2, \quad (6.36)$$

$$T: \mathcal{X} \rightarrow V, \quad x \mapsto {}^t(x, x^2), \quad (6.37)$$

$$\lambda: \text{Lebesgue 測度}. \quad (6.38)$$

(2) この実現の対数分配関数 $\psi: \tilde{\Theta} \rightarrow \mathbb{R}$ は $\psi(\theta) = -\frac{(\theta^1)^2}{4\theta^2} - \frac{1}{2} \log(-\theta^2) + \frac{1}{2} \log \pi$ となる。

(3) $\Theta = \tilde{\Theta} = \mathbb{R} \times \mathbb{R}_{<0}$ が成り立つ。

(4) 次の写像 $\theta: \mathcal{P} \rightarrow \Theta$ は $P := P_{(V, T, \lambda)}$ の逆写像である:

$$\theta: \mathcal{P} \rightarrow \Theta, \quad p \mapsto \left(\frac{E_p[x]}{\text{Var}_p[x]}, -\frac{1}{2 \text{Var}_p[x]} \right) \quad (6.39)$$

(5) (V, T, λ) は最小次元実現である。とくに \mathcal{P} は開である。

証明 (1) 実現であることは [0425_資料.pdf](#) 例 3.2 で確かめた。

(2) 対数分配関数の定義から直接計算よりわかる。

(3) $\theta^2 \geq 0$ だと $\exp(\theta^1 x + \theta^2 x^2 - \psi(\theta))$ は積分可能でないから $\Theta \subset \tilde{\Theta} \subset \mathbb{R} \times \mathbb{R}_{<0}$ である。逆に写像 $P := P_{(V, T, \lambda)}$ について、すべての $p \in P(\mathbb{R} \times \mathbb{R}_{<0})$ は $p(dx) = \exp(\theta^1 x + \theta^2 x^2 - \psi(\theta)) \lambda(dx)$ ($\exists (\theta^1, \theta^2) \in \mathbb{R} \times \mathbb{R}_{<0}$) と表せるから、 $(\mu, \sigma) := \left(-\frac{\theta^1}{2\theta^2}, \sqrt{-\frac{1}{2\theta^2}} \right) \in \mathbb{R} \times \mathbb{R}_{>0}$ とおけば $p(dx) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \lambda(dx)$ と表せることになり $p \in \mathcal{P}$ がわかる。したがって $P(\mathbb{R} \times \mathbb{R}_{<0}) \subset \mathcal{P}$ をみたすから $\mathbb{R} \times \mathbb{R}_{<0} \subset P^{-1}(\mathcal{P}) = \Theta$ である。よって $\Theta = \tilde{\Theta} = \mathbb{R} \times \mathbb{R}_{<0}$ である。

(4) $(\theta^1, \theta^2) \in \mathbb{R} \times \mathbb{R}_{<0}$ と $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_{>0}$ の対応に注意すれば直接計算よりわかる。

(5) 最小次元実現の特徴づけの条件 A(3) と条件 B が成り立つことから、最小次元実現であることがわかる。 \square

以降、 \mathcal{P} には自然な位相および多様体構造が入っているものとして扱い、 \mathcal{P} 上の自然な平坦アファイン接続を ∇ 、Fisher 計量を g 、 $(0, 3), (1, 2)$ 型の Amari-Chentsov テンソルをそれぞれ S, A とおく。また、 $\theta: \mathcal{P} \rightarrow \Theta$ は多様体 \mathcal{P} の座標とみなす。

命題 6.11. 座標 (μ, σ) に関する g の行列表示は

$$(g_{ij})_{i,j} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}, \quad (g^{ij})_{i,j} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix} \quad (6.40)$$

となる。

証明 微分同相写像 θ により g を Θ 上のテンソル場とみなして計算する。座標 (θ^1, θ^2) と座標 (μ, σ) の間の

座標変換が $\theta^1 = \frac{\mu}{\sigma^2}$, $\theta^2 = -\frac{1}{2\sigma^2}$ および $\mu = -\frac{\theta^1}{2\theta^2}$, $\sigma = \sqrt{-\frac{1}{2\theta^2}}$ であることに注意すると

$$d\mu = -\frac{1}{2\theta^2}d\theta^1 + \frac{\theta^1}{2(\theta^2)^2}d\theta^2, \quad d\sigma = \frac{1}{2\sqrt{2}}(-\theta^2)^{-3/2}d\theta^2, \quad (6.41)$$

$$d\theta^1 = \frac{1}{\sigma^2}d\mu - \frac{2\mu}{\sigma^3}d\sigma, \quad d\theta^2 = \frac{1}{\sigma^3}d\sigma, \quad (6.42)$$

さらに

$$(d\theta^1)^2 = \frac{1}{\sigma^4}(d\mu)^2 - \frac{\mu}{\sigma^5}d\mu d\sigma + \frac{4\mu^2}{\sigma^6}(d\sigma)^2, \quad (6.43)$$

$$d\theta^1 d\theta^2 = \frac{1}{\sigma^5}d\mu d\sigma - \frac{2\mu}{\sigma^6}(d\sigma)^2, \quad (6.44)$$

$$(d\theta^2)^2 = \frac{1}{\sigma^6}(d\sigma)^2 \quad (6.45)$$

である。したがって、 Θ 上の標準的な平坦アファイン接続を D とおくと

$$Dd\mu = \frac{1}{(\theta^2)^2}d\theta^1 d\theta^2 - \frac{\theta^1}{(\theta^2)^3}(d\theta^2)^2 = \frac{4}{\sigma}d\mu d\sigma, \quad (6.46)$$

$$Dd\sigma = \frac{3}{4\sqrt{2}}(-\theta^2)^{-5/2}(d\theta^2)^2 = \frac{3}{\sigma}(d\sigma)^2 \quad (6.47)$$

である。よって

$$d\psi = \frac{\mu}{\sigma^2}d\mu + \left(-\frac{\mu^2}{\sigma^3} + \frac{1}{\sigma}\right)d\sigma, \quad (6.48)$$

$$\text{Hess } \psi = Dd\psi \quad (6.49)$$

$$= d\left(\frac{\mu}{\sigma^2}\right)d\mu + \frac{\mu}{\sigma^2}Dd\mu + d\left(-\frac{\mu^2}{\sigma^3} + \frac{1}{\sigma}\right)d\sigma + \left(-\frac{\mu^2}{\sigma^3} + \frac{1}{\sigma}\right)Dd\sigma \quad (6.50)$$

$$= \frac{1}{\sigma^2}(d\mu)^2 + \frac{2}{\sigma^2}(d\sigma)^2 \quad (6.51)$$

である。これより命題の主張が従う。 \square

命題 6.12 (AC テンソルの成分). 座標 (μ, σ) に関する AC テンソル S の成分は

$$S_{111} = 0 \quad (6.52)$$

$$S_{112} = S_{121} = S_{211} = \frac{2}{\sigma^3} \quad (6.53)$$

$$S_{122} = S_{212} = S_{221} = 0 \quad (6.54)$$

$$S_{222} = \frac{8}{\sigma^3} \quad (6.55)$$

である。座標 (μ, σ) に関する A の成分は

$$A_{11}^1 = 0, \quad A_{11}^2 = \frac{1}{\sigma}, \quad (6.56)$$

$$A_{12}^1 = A_{21}^1 = \frac{2}{\sigma}, \quad A_{12}^2 = A_{21}^2 = 0, \quad (6.57)$$

$$A_{22}^1 = 0, \quad A_{22}^2 = \frac{4}{\sigma} \quad (6.58)$$

である。

証明 微分同相写像 θ により S, A を Θ 上のテンソル場とみなして計算する。 Θ 上の標準的な平坦アファイン接続を D とおくと

$$DDd\psi = D\left(\frac{1}{\sigma^2}(d\mu)^2 + \frac{2}{\sigma^2}(d\sigma)^2\right) \quad (6.59)$$

$$= -\frac{2}{\sigma^3}(d\mu)^2 d\sigma + \frac{1}{\sigma^2}D(d\mu)^2 - \frac{4}{\sigma^3}(d\sigma)^3 + \frac{2}{\sigma^2}D(d\sigma)^2 \quad (6.60)$$

ここで

$$D(d\mu)^2 = 2d\mu Dd\mu = \frac{8}{\sigma}(d\mu)^2 d\sigma, \quad (6.61)$$

$$D(d\sigma)^2 = 2d\sigma Dd\sigma = \frac{6}{\sigma}(d\sigma)^3 \quad (6.62)$$

だから

$$DDd\psi = \frac{6}{\sigma^3}(d\mu)^2 d\sigma + \frac{8}{\sigma^3}(d\sigma)^3 \quad (6.63)$$

である。これより命題の主張の式が得られる。 A の成分は「 $A_{ij}{}^k = g^{kl}S_{ijl}$ 」を用いて直接計算より得られる。

□

命題 6.13 (接続係数).

(1) 座標 (μ, σ) に関する ∇^g の接続係数は

$$\Gamma_{11}^g = 0, \quad \Gamma_{12}^g = \Gamma_{21}^g = -\frac{1}{\sigma}, \quad \Gamma_{22}^g = 0, \quad (6.64)$$

$$\Gamma_{11}^g = \frac{1}{2\sigma}, \quad \Gamma_{12}^g = \Gamma_{21}^g = 0, \quad \Gamma_{22}^g = -\frac{1}{\sigma} \quad (6.65)$$

である。

(2) 座標 (μ, σ) に関する $\nabla^{(\alpha)}$ の接続係数は

$$\Gamma_{11}^{(\alpha)} = 0, \quad \Gamma_{12}^{(\alpha)} = \Gamma_{21}^{(\alpha)} = -\frac{1+\alpha}{\sigma}, \quad \Gamma_{22}^{(\alpha)} = 0, \quad (6.66)$$

$$\Gamma_{11}^{(\alpha)} = \frac{1-\alpha}{2\sigma}, \quad \Gamma_{12}^{(\alpha)} = \Gamma_{21}^{(\alpha)} = 0, \quad \Gamma_{22}^{(\alpha)} = -\frac{1+2\alpha}{\sigma} \quad (6.67)$$

である。

証明 Γ^g は $\Gamma_{ij}^g = \frac{1}{2}g^{kl}(\partial_i g_{jl} + \partial_j g_{li} - \partial_l g_{ij})$ を直接計算することで得られる。 $\Gamma^{(\alpha)}$ は $\Gamma_{ij}^{(\alpha)} = \Gamma_{ij}^g - \frac{\alpha}{2}A_{ij}{}^k$ より得られる。

□

命題 6.14 (測地線方程式). (μ, σ) 座標に関する $\nabla^{(\alpha)}$ -測地線の方程式は

$$\begin{cases} \ddot{\mu} - \frac{2(1+\alpha)}{\sigma} \dot{\mu} \dot{\sigma} = 0, \\ \ddot{\sigma} + \frac{1-\alpha}{2\sigma} \dot{\mu}^2 - \frac{1+2\alpha}{\sigma} \dot{\sigma}^2 = 0 \end{cases} \quad (6.68)$$

である。とくに $\alpha = 0$ のとき

$$\begin{cases} \ddot{\mu} - \frac{2}{\sigma} \dot{\mu} \dot{\sigma} = 0, \\ \ddot{\sigma} + \frac{1}{2\sigma} \dot{\mu}^2 - \frac{1}{\sigma} \dot{\sigma}^2 = 0 \end{cases} \quad (6.69)$$

である。

証明 測地線の方程式「 $\ddot{x}^k = -\Gamma_{ij}^k \dot{x}^i \dot{x}^j$ 」に接続係数を代入して得られる。

□

命題 6.15. ∇^g -測地線の像は、楕円

$$\left(\frac{x - x_0}{\sqrt{2}} \right)^2 + y^2 = r^2 \quad (x_0 \in \mathbb{R}, r \in \mathbb{R}_{>0}) \quad (6.70)$$

の一部または y 軸に平行な直線の一部である。

証明¹⁾ 測地線の方程式

$$\ddot{\mu} - \frac{2}{\sigma} \dot{\mu} \dot{\sigma} = 0, \quad (6.71)$$

$$\ddot{\sigma} + \frac{1}{2\sigma} \dot{\mu}^2 - \frac{1}{\sigma} \dot{\sigma}^2 = 0 \quad (6.72)$$

を変形していく。

$\dot{\mu} = 0$ の場合は $\mu = \text{const.}$ ゆえに測地線は y 軸に平行な直線の一部である。

以下、 $\dot{\mu} \neq 0$ の場合を考える。(6.71) の両辺を $\dot{\mu}$ で割って

$$\frac{\ddot{\mu}}{\dot{\mu}} - 2 \frac{\dot{\sigma}}{\sigma} = 0 \quad (6.73)$$

これより $\log \dot{\mu} = 2 \log \sigma + \text{const.}$ したがって

$$\dot{\mu} = k \sigma^2 \quad (k \in \mathbb{R}) \quad (6.74)$$

である。一方、 ∇^g は g の Levi-Civita 接続であるから、測地線の速度ベクトルの g に関する大きさは一定、すなわち

$$\frac{\dot{\mu}^2 + 2\dot{\sigma}^2}{\sigma^2} = r^2 \quad (a \in \mathbb{R}) \quad (6.75)$$

である。(6.75) に (6.74) を代入して

$$\frac{k^2 \sigma^4 + 2\dot{\sigma}^2}{\sigma^2} = a^2 \quad (6.76)$$

$$\dot{\sigma} = \pm \sigma \sqrt{\frac{a^2 - k^2 \sigma^2}{2}} \quad (6.77)$$

を得る。これと (6.74) より

$$\frac{d\mu}{d\sigma} = \frac{\dot{\mu}}{\dot{\sigma}} = \frac{k\sigma^2}{\pm \sigma \sqrt{\frac{a^2 - k^2 \sigma^2}{2}}} \quad (6.78)$$

$$= \mp \frac{\sqrt{2}|a|}{k} \frac{\left(\frac{k}{a}\right)^2 \sigma}{\sqrt{1 - \left(\frac{k}{a}\right)^2 \sigma^2}} \quad (6.79)$$

$$\therefore \mu = \mp \frac{\sqrt{2}|a|}{k} \sqrt{1 - \left(\frac{k}{a}\right)^2 \sigma^2} + \mu_0 \quad (\mu_0 \in \mathbb{R}) \quad (6.80)$$

を得る。よって

$$(\mu - \mu_0)^2 = \frac{2a^2}{k^2} - 2\sigma^2 \quad (6.81)$$

$r := \frac{a}{k}$ とおいて整理すれば

$$\left(\frac{\mu - \mu_0}{\sqrt{2}}\right)^2 + \sigma^2 = r^2 \quad (6.82)$$

が得られる。

□

7 α -接続

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ライン接続、 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 7.1 (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 ∇ -アフライン座標に関し、 $R^{(\alpha)}$ の成分は

$$R^{(\alpha)}_{ijk}{}^l = -\frac{1-\alpha^2}{4} \left(A_{jk}{}^m A_{im}{}^l - A_{ik}{}^m A_{jm}{}^l \right) \quad (7.1)$$

となる。

証明 0613_資料.pdf 命題 1.12 の式

$$R^{(\alpha)}_{ijk}{}^l = \frac{1-\alpha}{2} \left(\partial_i A_{jk}{}^l - \partial_j A_{ik}{}^l \right) + \left(\frac{1-\alpha}{2} \right)^2 \left(A_{jk}{}^m A_{im}{}^l - A_{ik}{}^m A_{jm}{}^l \right) \quad (7.2)$$

を変形する。

$$\partial_i A_{jk}{}^l - \partial_j A_{ik}{}^l = \partial_i (g^{la} S_{jka}) - \partial_j (g^{la} S_{ika}) \quad (7.3)$$

$$= \partial_i (g^{la}) S_{jka} + g^{la} \partial_i S_{jka} - \partial_j (g^{la}) S_{ika} - g^{la} \partial_j S_{ika} \quad (7.4)$$

$$= \partial_i (g^{la}) S_{jka} - \partial_j (g^{la}) S_{ika} \quad (7.5)$$

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

$$\partial_i (g^{la}) S_{jka} = -g^{ma} g^{lb} \partial_i (g_{mb}) S_{jka} \quad (7.6)$$

$$= -g^{ma} g^{lb} S_{imb} S_{jka} \quad (7.7)$$

1) 証明の流れは [?, Chap.3 14.4] を参考にした。

$$= -A_{im}^l A_{jk}^m \quad (7.8)$$

同様にして

$$\partial_j(g^{la})S_{ika} = -A_{jm}^l A_{ik}^m \quad (7.9)$$

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

$$R^{(\alpha)}_{ijk}{}^l = \left(-\frac{1-\alpha}{2} + \left(\frac{1-\alpha}{2} \right)^2 \right) (A_{jk}^m A_{im}^l - A_{ik}^m A_{jm}^l) = -\frac{1-\alpha^2}{4} (A_{jk}^m A_{im}^l - A_{ik}^m A_{jm}^l) \quad (7.10)$$

となる。

□

系 7.2.

- (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 7.1 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 7.3 (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

$$g(\nabla_X^{(\alpha)} Y, Z) + g(Y, \nabla_X^{(-\alpha)} Z) = g(\nabla_X^g Y, Z) - \frac{\alpha}{2} S(X, Y, Z) + g(Y, \nabla_X^g Z) + \frac{\alpha}{2} S(X, Z, Y) \quad (7.11)$$

$$= g(\nabla_X^g Y, Z) + g(Y, \nabla_X^g Z) \quad (7.12)$$

$$= X(g(Y, Z)) \quad (7.13)$$

より従う。 $\alpha = \pm 1$ で双対平坦となることは系 7.2 よりわかる。

□

8 期待値パラメータ

命題-定義 8.1 (期待値パラメータ空間). 集合

$$\mathcal{M} := \{E_p[T] \in V \mid p \in \mathcal{P}\} \quad (8.1)$$

は V の開部分多様体となり、写像 $\eta: \mathcal{P} \rightarrow \mathcal{M}$, $p \mapsto E_p[T]$ 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 8.2 (ψ の微分は十分統計量の期待値). 写像 $\nabla\psi: \Theta \rightarrow V^{\vee\vee} = V$ は

$$(\nabla\psi)(\theta(p)) = \eta(p) \quad (p \in \mathcal{P}) \quad (8.2)$$

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

事実 8.3. 位相ベクトル空間の凸集合の内部は凸集合である。 \square

命題-定義 8.1 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 $\text{Int}\tilde{\Theta}$ 上の関数とみなすと、事実 8.3 とあわせて ψ は ?? の前提をみたすから、?? (1) より $\nabla\psi: \text{Int}\tilde{\Theta} \rightarrow V^{\vee\vee} = V$ は局所微分同相、とくに開写像でもある。したがって $\nabla\psi(\text{Int}\tilde{\Theta})$ は V の開部分多様体となる。さらに Θ は $\text{Int}\tilde{\Theta}$ の開集合だから、 $\nabla\psi(\Theta)$ は $\nabla\psi(\text{Int}\tilde{\Theta})$ の開部分多様体となる。このことと事実 8.2 より、 $\mathcal{M} = \nabla\psi(\Theta)$ は $\nabla\psi(\text{Int}\tilde{\Theta})$ 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。?? (2) より $\nabla\psi$ は $\text{Int}\tilde{\Theta}$ から $\nabla\psi(\text{Int}\tilde{\Theta})$ 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 $\eta = (\nabla\psi) \circ \theta: \mathcal{P} \rightarrow \mathcal{M}$ は微分同相である。 \square

以降、 $\psi|_{\text{Int}\tilde{\Theta}}$ 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 8.4 (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 $\psi: \Theta \rightarrow \mathbb{R}$ および $\phi: \mathcal{M} \rightarrow \mathbb{R}$ と、 \mathcal{P} 上の自然パラメータ座標 $\theta = (\theta^1, \dots, \theta^n)$ および期待値パラメータ座標 $\eta = (\eta_1, \dots, \eta_m)$ に関し次が成り立つ:

$$(1) \quad \frac{\partial\psi}{\partial\theta^i}(\theta(p)) = \eta_i(p), \quad \frac{\partial\phi}{\partial\eta_i}(\eta(p)) = \theta^i(p) \quad (p \in \mathcal{P}). \quad (8.3)$$

(2) g の θ -座標に関する成分は

$$g_{ij}(p) = \frac{\partial^2\psi}{\partial\theta^i\partial\theta^j}(\theta(p)) = \frac{\partial\eta_j}{\partial\theta^i}(p), \quad g^{ij}(p) = \frac{\partial^2\phi}{\partial\eta_i\partial\eta_j}(\eta(p)) = \frac{\partial\theta^i}{\partial\eta_j}(p) \quad (p \in \mathcal{P}) \quad (8.4)$$

をみたす。

(3) δ_i^j を Kronecker のデルタとして

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_i^j \quad (8.5)$$

が成り立つ。

証明 (1) 事実 8.2 より $\nabla\psi \circ \theta = \eta$ であることと、?? (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?? (5) より従う。

(3)

$$g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = g\left(\frac{\partial}{\partial\theta^i}, \frac{\partial\theta^k}{\partial\eta_j} \frac{\partial}{\partial\theta^k}\right) = g_{ik} \frac{\partial\theta^k}{\partial\eta_j} = g_{ik} g^{kj} = \delta_i^j. \quad (8.6)$$

□

定理 8.5. 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 $\partial_i = \frac{\partial}{\partial\theta^i}$, $\partial^i = \frac{\partial}{\partial\eta_i}$ と略記すれば、上の定理の (3) より

$$0 = \partial^i \delta_k^j = g\left(\nabla_{\partial^i}^{(1)} \partial_k, \partial^j\right) + g\left(\partial_k, \nabla_{\partial^i}^{(1)} \partial^j\right) \quad (8.7)$$

だから

$$\Gamma^{(-1)ij}_k = g\left(\partial_k, \nabla_{\partial^i}^{(-1)} \partial^j\right) \quad (8.8)$$

$$= -g\left(\nabla_{\partial^i}^{(1)} \partial_k, \partial^j\right) \quad (8.9)$$

$$= -\frac{\partial\theta^l}{\partial\eta_i} g\left(\nabla_{\partial^i}^{(1)} \partial_k, \partial^j\right) \quad (8.10)$$

$$= -\frac{\partial\theta^l}{\partial\eta_i} \Gamma^{(1)j}_{lk} \quad (8.11)$$

$$= 0 \quad (\Gamma^{(1)j}_{lk} = 0) \quad (8.12)$$

となる。

□