

## 1 導入と振り返り

前回、指数型分布族を4つ組  $(\mathcal{P}, \mu, \Theta, X)$  として定義した ( $\mathcal{P}$  は確率分布の族、 $\mu$  は  $\sigma$ -有限測度、 $\Theta$  は添字集合、 $X$  は可測空間)。しかし発表時の指摘により、うまく定義すれば4つ組を持ち出さずとも  $\mathcal{P}$  のみで「指数型分布族」を定義できるらしいことがわかった。そこで本稿では指数型分布族を改めて定義し直し、新たな定義に基づいて指数型分布族の性質と具体例を調べる。定義は Dropbox の吉野先生の資料 (I01-... など) で始まるものと、そこで参考文献に挙げられていた [WJ07, Chap. 3] を参考にした。

## 2 指数型分布族の定義

指数型分布族を定義する。本稿では、可測空間  $X$  上の確率測度全体の集合を  $\mathcal{P}(X)$  と書くことにする。

**定義 2.1** (指数型分布族).  $X$  を可測空間、 $\emptyset \neq \mathcal{P} \subset \mathcal{P}(X)$  とする。 $\mathcal{P}$  が  $X$  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:  $\exists (T, \mu)$  s.t.

- (E1)  $T: X \rightarrow \mathbb{R}^m$  ( $m \in \mathbb{Z}_{\geq 0}$ ) は可測写像である。
- (E2)  $\mu$  は  $X$  上の  $\sigma$ -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、 $\exists \theta \in \mathbb{R}^m$  s.t.

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x) \rangle}{\int_X \exp\langle \theta, T(y) \rangle \mu(dy)} \quad \mu\text{-a.e. } x \in X \quad (2.1)$$

である。ただし  $\langle \cdot, \cdot \rangle$  は  $\mathbb{R}^m$  の標準的な内積である。

さらに次のように定める:

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

$$\Theta_{(T, \mu)} := \left\{ \theta \in \mathbb{R}^m \mid \int_X \exp\langle \theta, T(y) \rangle \mu(dy) < +\infty \right\} \quad (2.2)$$

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

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

$$\psi(\theta) := \log \int_X \exp\langle \theta, T(y) \rangle \mu(dy) \quad (2.3)$$

を  $(T, \mu)$  の対数分配関数 (log-partition function) という。

対数分配関数の性質は本稿ではまだ扱わない。自然パラメータ空間は次の性質を持つ。

**命題 2.2** (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。

**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、 $e^{(1-t)\langle \theta, T(x) \rangle} \in L^p(X, \mu)$

かつ  $e^{t\langle\theta', T(x)\rangle} \in L^q(X, \mu)$  だから、Hölder の不等式より

$$\int_X e^{\langle(1-t)\theta + t\theta', T(x)\rangle} \mu(dx) = \int_X e^{(1-t)\langle\theta, T(x)\rangle} e^{t\langle\theta', T(x)\rangle} \mu(dx) \quad (2.4)$$

$$\leq \left( \int_X e^{(1-t)\langle\theta, T(x)\rangle p} \mu(dx) \right)^{1/p} \left( \int_X e^{t\langle\theta', T(x)\rangle q} \mu(dx) \right)^{1/q} \quad (2.5)$$

$$= \left( \int_X e^{\langle\theta, T(x)\rangle} \mu(dx) \right)^{1/p} \left( \int_X e^{\langle\theta', T(x)\rangle} \mu(dx) \right)^{1/q} \quad (2.6)$$

$$< +\infty \quad (2.7)$$

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

### 3 指数型分布族の具体例

指数型分布族の具体例を挙げる。

**例 3.1** (有限集合上の確率分布).  $X = \{1, \dots, n\}$ 、 $\gamma$  を  $X$  上の数え上げ測度とする。 $X$  上の確率分布全体の集合  $\mathcal{P}(X)$  が  $X$  上の指数型分布族であることを確かめる。 $\delta^j$  ( $j = 1, \dots, n$ ) を点  $j$  での Dirac 測度とおく。任意の  $P \in \mathcal{P}(X)$  に対し、

$$P(dk) := \sum_{j=1}^n a_j \delta^j(dk), \quad a_1, \dots, a_n \in \mathbb{R}_{>0}, \quad \sum_{j=1}^n a_j = 1 \quad (3.1)$$

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

$$P(dk) = \exp \left( \sum_{j=1}^n (\log a_j) \delta_{jk} \right) \gamma(dk) \quad (3.2)$$

$$= \exp \left( \sum_{j=1}^n \theta_j \delta_{jk} \right) \gamma(dk) \quad (3.3)$$

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

**例 3.2** (正規分布族).  $X = \mathbb{R}$ 、 $\lambda$  を  $\mathbb{R}$  上の Lebesgue 測度とする。 $X$  上の確率分布の集合

$$\mathcal{P} := \left\{ P_{(\mu, \sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \mid \mu \in \mathbb{R}, \sigma^2 > 0 \right\} \quad (3.4)$$

を正規分布族 (family of normal 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  $P_{(\mu, \sigma^2)} \in \mathcal{P}$  に対し

$$P_{(\mu, \sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \quad (3.5)$$

$$= \exp \left( -\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2) - \frac{1}{2} \log 2\pi\sigma^2 \right) \lambda(dx) \quad (3.6)$$

$$= \exp \left( \left[ \frac{\mu}{\sigma^2} \quad -\frac{1}{2\sigma^2} \right] \begin{bmatrix} x \\ x^2 \end{bmatrix} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log 2\pi\sigma^2 \right) \lambda(dx) \quad (3.7)$$

$$= \exp \left( \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} x \\ x^2 \end{bmatrix} + \frac{\theta_1^2}{4\theta_2} - \frac{1}{2} \log \left( -\frac{\pi}{\theta_2} \right) \right) \lambda(dx) \quad (3.8)$$

(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、 $(T, \lambda)$  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 3.3** (Poisson 分布族).  $\mathcal{X} = \mathbb{N}$ 、 $\gamma$  を  $\mathbb{N}$  上の数え上げ測度とする。 $\mathcal{X}$  上の確率分布の集合

$$\mathcal{P} := \left\{ P_\lambda(dk) = \frac{\lambda^k}{k!} e^{-\lambda} \gamma(dk) \mid \lambda > 0 \right\} \quad (3.9)$$

を  $P_\lambda$  を **Poisson 分布族 (family of Poisson 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  $P_\lambda \in \mathcal{P}$  に対し

$$P_\lambda(dk) = \frac{\lambda^k}{k!} e^{-\lambda} \gamma(dk) \quad (3.10)$$

$$= \exp(k \log \lambda - \lambda) \frac{1}{k!} \gamma(dk) \quad (3.11)$$

$$= \exp(\theta k - e^\theta) \frac{1}{k!} \gamma(dk) \quad (3.12)$$

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

## 4 極小実現

一般の  $(T, \mu)$  に対しては、各  $p \in \mathcal{P}$  に対し定義 2.1 の条件 (E3) をみたす  $\theta \in \Theta_{(T, \mu)}$  は一意とは限らないが<sup>1)</sup>、この一意性を持つ実現は特に重要である。

**定義 4.1** (極小実現).  $\mathcal{X}$  を可測空間、 $\mathcal{P} \subset \mathcal{P}(\mathcal{X})$  を  $\mathcal{X}$  上の指数型分布族とする。 $\mathcal{P}$  の実現  $(T, \mu)$  であって、各  $p \in \mathcal{P}$  に対し、定義 2.1 の条件 (E3) を満たす  $\theta$  がただひとつ定まるようなものを、 $\mathcal{P}$  の**極小実現 (minimal representation)** という<sup>2)</sup>。[TODO] できれば吉野先生に倣って「最小次元実現」と訳したいが、本稿でいう極小実現が必ずしも次元最小であるかどうかわからなかったの、ひとまず「極小実現」と訳すことにする。

**定理 4.2** (極小実現の存在).  $\mathcal{X}$  を可測空間、 $\mathcal{P} \subset \mathcal{P}(\mathcal{X})$  を  $\mathcal{X}$  上の指数型分布族とする。このとき、 $\mathcal{P}$  の極小実現が存在する。

**証明**  $(T, \mu)$  は  $\mathcal{P}$  の実現のうちで次元が最小のものであるとする。 $(T, \mu)$  が極小実現であることを示す。背理法のために  $(T, \mu)$  が極小実現でないこと、すなわちある  $p_0 \in \mathcal{P}$  および  $\theta_0, \theta'_0 \in \mathbb{R}^m$ ,  $\theta_0 \neq \theta'_0$  が存在して

$$\exp(\langle \theta_0, T(x) \rangle - \psi(\theta_0)) = \frac{dp_0}{d\mu}(x) = \exp(\langle \theta'_0, T(x) \rangle - \psi(\theta'_0)) \quad \mu\text{-a.e. } x \in \mathcal{X} \quad (4.1)$$

1) たとえば  $T$  の第  $i$  成分関数  $T_i$  が零写像ならば、 $\langle \theta, T(x) \rangle$  は  $\theta$  の第  $i$  成分の値によらず決まるため、 $\theta$  の第  $i$  成分を動かしたものはすべて定義 2.1 の条件 (E3) をみたす。

2) 極小実現は  $\mathcal{P}$  に対し一意とは限らない。実際、 $(T, \mu)$  が極小実現ならば、任意の直交行列  $A \in O(m)$  に対し  $\langle A\theta, AT(x) \rangle = \langle \theta, T(x) \rangle$  ゆえに  $(AT, \mu)$  も極小実現となる。

が成り立つことを仮定する。式を整理して

$$\langle \theta_0 - \theta'_0, T(x) \rangle = \psi(\theta_0) - \psi(\theta'_0) \quad \mu\text{-a.e. } x \in X \quad (4.2)$$

を得る。いま背理法の仮定より  $\theta_0 - \theta'_0 \neq 0$  だから、単位ベクトル  $u := (\theta_0 - \theta'_0) / \|\theta_0 - \theta'_0\|$  が定義でき、

$$\langle u, T(x) \rangle = \frac{1}{\|\theta_0 - \theta'_0\|} (\psi(\theta_0) - \psi(\theta'_0)) =: c \quad (4.3)$$

( $c$  は  $x \in X$  によらない定数) が成り立つ。ここで  $u$  を  $e_1 = {}^t(1, 0, \dots, 0)$  に写す直交行列  $A \in O(m)$  をひとつ選べば

$$c = \langle u, T(x) \rangle \quad (4.4)$$

$$= \langle Au, AT(x) \rangle \quad (4.5)$$

$$= \langle e_1, AT(x) \rangle \quad (4.6)$$

$$= (AT(x))_1 \quad (4.7)$$

(ただし  $(AT(x))_i$  は  $AT(x)$  の第  $i$  成分) が成り立つ。

写像  $T': X \rightarrow \mathbb{R}^{m-1}$  を  $T'(x) := {}^t((AT(x))_2, \dots, (AT(x))_m)$  と定め、 $(T', \mu)$  が  $\mathcal{P}$  の実現であることを示す。まず  $T'$  は定め方より可測写像だから定義 2.1 の条件 (E1) は満たされ、また  $\mu$  は元々条件 (E2) を満たしている。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  $(T, \mu)$  が  $\mathcal{P}$  の実現であることより、ある  $\theta \in \mathbb{R}^m$  が存在して

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x) \rangle}{\int_X \exp\langle \theta, T(y) \rangle \mu(dy)} \quad \mu\text{-a.e. } x \in X \quad (4.8)$$

が成り立つ。ここで  $\theta'' := ((A\theta)_2, \dots, (A\theta)_m) \in \mathbb{R}^{m-1}$  とおくと

$$\langle \theta, T(x) \rangle = \langle A\theta, AT(x) \rangle \quad (4.9)$$

$$= \sum_{i=1}^m (A\theta)_i (AT(x))_i \quad (4.10)$$

$$= (A\theta)_1 (AT(x))_1 + \sum_{i=2}^m (A\theta)_i (AT(x))_i \quad (4.11)$$

$$= \underbrace{(A\theta)_1 c}_{=: d} + \langle \theta'', T'(x) \rangle \quad (4.12)$$

( $d$  は  $x \in X$  によらない定数) が成り立つから

$$\frac{dp}{d\mu}(x) = \frac{\exp(d + \langle \theta'', T'(x) \rangle)}{\int_X \exp(d + \langle \theta'', T'(y) \rangle) \mu(dy)} = \frac{\exp\langle \theta'', T'(x) \rangle}{\int_X \exp\langle \theta'', T'(y) \rangle \mu(dy)} \quad \mu\text{-a.e. } x \in X \quad (4.13)$$

を得る。したがって  $(T', \mu)$  は定義 2.1 の条件 (E3) も満たし、 $\mathcal{P}$  の実現であることがいえた。 $(T', \mu)$  は次元  $m-1$  だから  $(T, \mu)$  の次元  $m$  の最小性に矛盾する。背理法より  $(T, \mu)$  は  $\mathcal{P}$  の極小実現である。□

**定理 4.3 (極小実現の性質).**  $X$  を可測空間、 $\mathcal{P} \subset \mathcal{P}(X)$  を  $X$  上の指数型分布族、 $(T, \mu)$  を  $\mathcal{P}$  の次元  $m$  の実現とする。このとき、 $(T, \mu)$  が極小実現ならば、 $\langle u, T(x) \rangle$  が  $\mu$ -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**  $(T, \mu)$  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  $\mu$ -a.e. 定数であると仮定しておく。  $p \in \mathcal{P}$  とし、定義 2.1 の条件 (E3) の  $\theta \in \mathbb{R}^m$  をひとつ選ぶと、

$$\frac{dp}{d\mu}(x) = \frac{e^{\langle \theta, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} \mu(dy)} \quad (4.14)$$

$$= \frac{e^{\langle \theta, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} \mu(dy)} \cdot \frac{e^{\langle u, T(x) \rangle}}{e^{\langle u, T(x) \rangle}} \quad (4.15)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_X e^{\langle \theta, T(y) \rangle} e^{\langle u, T(y) \rangle} \mu(dy)} \quad (4.16)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_X e^{\langle \theta+u, T(y) \rangle} \mu(dy)} \quad (4.17)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_X e^{\langle \theta+u, T(y) \rangle} \mu(dy)} \quad (4.18)$$

を得る。したがって  $\theta + u$  も定義 2.1 の条件 (E3) を満たすが、いま  $u \neq 0$  より  $\theta + u \neq \theta$  だから、 $(T, \mu)$  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 4.4** (有限集合上の確率分布族). 例 3.1 の  $(T, \gamma)$  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、 $\theta_j$  は  $\theta_j = \log P(\{j\})$  ( $j = 1, \dots, n$ ) として一意に決まる。

## 5 次回以降の予定

- 期待値と分散の定義および基本性質
- 対数分配関数の微分可能性およびその Hesse 行列の正定値性

## 6 参考文献

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