1 Introduction

情報幾何学とは、可微分多様体上の双対構造と呼ばれる幾何学的構造を扱う、幾何学の一分野である。情報幾何の重要な応用先のひとつは統計的推定であり、推定の問題を幾何学的に捉えなおすことで、より深い理解につながることが期待される。

情報幾何が統計的推定へどのように応用されるかをみるために、本稿では指数型分布族および混合型分布族について説明する。指数型/混合型分布族とは、ある可測空間上の確率測度からなる族の一種である。指数型分布族は、正規分布や Poisson 分布など基本的な確率分布が属するクラスであり、数学的にもよい性質を備えている [TODO] どのような？。混合型分布族は、いくつかの確率分布の重み付け和からなる族であり、応用上重要なクラスである。

指数型/混合型分布族は確率密度関数の形によって定義される。そこで本稿では、まず測度論の基本事項を復習した後、確率空間や確率変数などの確率論の基本事項を整理する。その後、指数型分布族と混合型分布族の定義および具体例を与える。

2 Radon-Nikodým の定理と Hölder の不等式

Radon-Nikodým の定理について復習する。

定義 2.1 (絶対連続). $(\mathcal{X}, \mathcal{B})$ を可測空間、μ, ν を $\mathcal{X}$ 上の測度とする。ν が μ に関し**絶対連続 (absolutely continuous)** であるとは、任意の $E \in \mathcal{B}$ に対し $\mu(E) = 0$ ならば $\nu(E) = 0$ が成り立つことをいう。

定理 2.2 (Radon-NIkodým の定理). $(X, \mathcal{B})$ を可測空間、μ を X 上の σ-有限測度、ν を X 上の測度とする。このとき、ν が μ に関して絶対連続であるための必要十分条件は、μ-a.e. $x \in X$ に対し定義された可積分関数 f が存在して

$$\nu(E) = \int_E f(x)\, d\mu(x) \quad (E \in \mathcal{B}) \tag{2.1}$$

が成り立つことである。この f を μ に関する ν の **Radon-Nikodým 微分 (Radon-Nikodým derivative)** といい、$\frac{d\nu}{d\mu}$ と書く。

Radon-Nikodým 微分のイメージとしては、μ の変化に対する ν の変化率を表している。

証明 関数族の sup として f を構成する。 □

Hölder の不等式について復習する。

命題 2.3 (Hölder の不等式). $(\mathcal{X}, \mathcal{B})$ を可測空間、μ を $\mathcal{X}$ 上の測度とする。$1 < p < \infty$、$q = p(p-1)^{-1}$、f を p 乗 μ-可積分関数、g を q 乗 μ-可積分関数とする。このとき、fg は μ-可積分であり、かつ

$$\int_X |fg| \mu(dx) \le \left(\int_X |f|^p \mu(dx)\right)^{\frac{1}{p}} \left(\int_X |g|^q \mu(dx)\right)^{\frac{1}{q}} \tag{2.2}$$

が成り立つ。

証明 Young の不等式を使う。 □

3 確率論の基本事項

確率論の基本事項を整理する。

A. 確率空間

定義 3.1 (確率空間). 測度空間 $(\Omega, \mathcal{F}, P)$ であって

(1) 各 $E \in \mathcal{F}$ に対し $P(E) \geq 0$

(2) $P(\Omega) = 1$

をみたすものを**確率空間 (probability space)** といい、P を $(\Omega, \mathcal{F})$ 上の**確率測度 (probability measure)** あるいは**確率分布 (probability distribution)** という。

定義 3.2 (確率変数). $(\Omega, \mathcal{F}, P)$ を確率空間、$(\mathcal{X}, \mathcal{A})$ を可測空間とする。可測関数 $X\colon (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ を $(\mathcal{X}, \mathcal{A})$ に値をもつ**確率変数 (random variable; r.v.)** という。

定義 3.3 (確率変数の確率分布). $(\Omega, \mathcal{F}, P)$ を確率空間、$X\colon (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ を確率変数とする。このとき、写像

$$P^X\colon \mathcal{A} \to [0, +\infty), \quad E \mapsto P(X^{-1}(E)) \quad (E \in \mathcal{A}) \tag{3.1}$$

は $(\mathcal{X}, \mathcal{A})$ 上の確率測度となる。これを $\boldsymbol{X}$ **の確率分布 (probability distribution of** $\boldsymbol{X}$**)** という。

X の確率分布が $(\mathcal{X}, \mathcal{A})$ 上のある確率分布 ν に等しいとき、X は $\boldsymbol{\nu}$ **に従う** という。

定義 3.4 (確率密度関数). $(\mathcal{X}, \mathcal{A})$ を可測空間、μ を $\mathcal{X}$ 上の σ-有限測度、ν を μ に関し絶対連続な $(\mathcal{X}, \mathcal{A})$ 上の確率測度とする。このとき、ν の μ に関する Radon-NIkodým 微分 $\dfrac{d\nu}{d\mu}$ を、ν の**確率密度関数 (probability density function; PDF)** という。

4 指数型分布族と混合型分布族

A. 指数型分布族

指数型分布族について述べる。

定義 4.1 (指数型分布族). $(\mathcal{X}, \mathcal{A})$ を可測空間、μ を $(\mathcal{X}, \mathcal{A})$ 上の σ-有限測度、$\Theta \subset \mathbb{R}^n$ を部分集合、$\mathcal{P} = (P_\theta)_{\theta \in \Theta}$ を $\mathcal{X}$ 上の確率分布族とする。ここで、ある

- $m \in \mathbb{Z}_{\geq 1}$
- $\mathcal{X}$ 上の可測関数 $g\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$
- $\mathcal{X}$ 上の可測関数 $T_i\colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \ldots, m)$
- Θ 上の関数 $\psi\colon \Theta \to \mathbb{R}$ および $a_i\colon \Theta \to \mathbb{R}$ $(i = 1, \ldots, m)$

が存在し、各 P_θ が μ に関し絶対連続で、確率密度関数が

$$\frac{dP_\theta}{d\mu}(x) = g(x)\exp\left(\sum_{i=1}^{m} a_i(\theta)T_i(x) - \psi(\theta)\right) \quad \mu\text{-a.e. } x \in \mathcal{X} \tag{4.1}$$

となるとき、$\mathcal{P}$ を**指数型分布族 (exponential family)** という。(4.1) を

$$P_\theta(dx) = g(x)\exp\left(\sum_{i=1}^{m} a_i(\theta)T_i(x) - \psi(\theta)\right)\mu(dx) \tag{4.2}$$

とも書く。

- $g\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$
- $a_i(\theta) = \theta_i$ $(i = 1, \dots, m)$ のとき**正準形 (canonical form)** という。$a_i(\theta)$ $(i = 1, \dots, m)$ を**自然パラメータ (natural parameter)** という。
- 十分統計量 $T_i\colon \mathcal{X} \to \Theta$
- 対数分配関数 $\psi\colon \Theta \to \mathbb{R}$

対数分配関数の形は g, a_i, T_i によって決まる。

命題 4.2 (対数分配関数の式)**.** 対数分配関数 ψ は

$$\psi(\theta) = \log\int_{\mathcal{X}} g(x)\exp\left(\langle a(\theta), T(x)\rangle\right)\mu(dx) \tag{4.3}$$

と表せる [1)]。ただし $\langle a(\theta), T(x)\rangle$ は $\displaystyle\sum_{i=1}^{m} a_i(\theta)T_i(x)$ の意味である。

証明 P_θ が確率測度であることより

$$1 = P_\theta(\mathcal{X}) = \int_{\mathcal{X}} g(x)\exp\left(\langle a(\theta), T(x)\rangle - \psi(\theta)\right)\mu(dx) \tag{4.4}$$

が成り立つ。$\exp(\psi(\theta))$ を両辺にかけて

$$\exp(\psi(\theta)) = \int_{\mathcal{X}} g(x)\exp\left(\langle a(\theta), T(x)\rangle\right)\mu(dx) \tag{4.5}$$

となり、対数をとって命題の式が得られる。 □

以下に指数型分布族に関する具体例を挙げる。

例 4.3 (正規分布)**.** $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$、$\mu$ は $\mathbb{R}$ 上の Lebesgue 測度、$\Theta = \mathbb{R} \times \mathbb{R}_{>0}$ の場合を考える。$(\mu, \sigma^2) \in \Theta$ に対し、

$$P_{(\mu,\sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\mu(dx) \tag{4.6}$$

で定義される確率分布 $P_{(\mu,\sigma^2)}$ を**正規分布 (normal distribution)** という。このとき $\{P_{\mu,\sigma^2}\}_{(\mu,\sigma^2)\in\Theta}$ が指数型分布

1) $e^{\psi(\theta)} = \int_{\mathcal{X}} g(x)\exp\left(\langle a(\theta), T(x)\rangle\right)\mu(dx)$ は 0 より大きい実数だから ψ は確かに定義される。

族であることを確かめる。

$$P_{(\mu,\sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mu(dx) \tag{4.7}$$
$$= \exp\left(-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2) - \frac{1}{2}\log 2\pi\sigma^2\right) \mu(dx) \tag{4.8}$$
$$= \exp\left(\begin{bmatrix} -\frac{1}{2\sigma^2} & \frac{\mu}{\sigma^2} \end{bmatrix} \begin{bmatrix} x^2 \\ x \end{bmatrix} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log 2\pi\sigma^2\right) \mu(dx) \tag{4.9}$$

となるから、$g(x) = 1,\ a_1(\mu,\sigma^2) = -\dfrac{1}{2\sigma^2},\ a_2(\mu,\sigma^2) = \dfrac{\mu}{\sigma^2},\ T_1(x) = x^2,\ T_2(x) = x,\ \psi(\mu,\sigma^2) = \dfrac{\mu^2}{2\sigma^2} + \dfrac{1}{2}\log 2\pi\sigma^2$ とおけばよいことがわかる。[TODO] 十分統計量から自然パラメータを推定できる？

例 4.4 (Poisson 分布). $(\mathcal{X},\mathcal{A}) = (\mathbb{N}, 2^{\mathbb{N}})$、$\mu$ は $\mathbb{N}$ 上の数え上げ測度、$\Theta = \mathbb{R}_{>0}$ の場合を考える。

$$P_\lambda(dk) = \frac{\lambda^k}{k!}\exp(-\lambda)\mu(dk) \tag{4.10}$$

で定義される確率分布 P_λ を **Poisson 分布 (Poisson distribution)** という。このとき $\{P_\lambda\}_{\lambda>0}$ が指数型分布族であることを確かめる。

$$P_\lambda(dk) = \frac{\lambda^k}{k!}\exp(-\lambda)\mu(dk) \tag{4.11}$$
$$= \frac{1}{k!}\exp\left(k\log\lambda - \lambda\right)\mu(dk) \tag{4.12}$$

となるから、$g(k) = \dfrac{1}{k!},\ a_1(\lambda) = \log\lambda,\ T_1(k) = k,\ \psi(\lambda) = \lambda$ とおけばよいことがわかる。

例 4.5 (指数型分布族でない例). $(\mathcal{X},\mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$、$\mu$ は $\mathbb{R}$ 上の Lebesgue 測度、$\Theta = \mathbb{R}_{>0}$ の場合を考える。$\mathbb{R}$ 上の一様分布の族

$$\{P_\theta\}_{\theta>0}, \qquad \frac{dP_\theta}{d\mu}(x) = \begin{cases} \dfrac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases} \tag{4.13}$$

を考える。これは指数型分布族ではないことを確かめる。もし指数型分布族であったとすると $\dfrac{dP_\theta}{d\mu}(x) = g(x)e^{\varphi(x,\theta)}$ と表せる (g は θ によらない関数)。そこでとくに $\theta = 1, x = 2$ に対し

$$0 = \frac{dP_1}{d\mu}(2) = g(2)e^{\varphi(2,1)} \tag{4.14}$$

より $g(2) = 0$ となる。一方 $\theta = 2, x = 2$ に対し

$$\frac{1}{2} = \frac{dP_2}{d\mu}(2) = g(2)e^{\varphi(2,2)} \tag{4.15}$$

より $g(2) \neq 0$ となり矛盾が従う。よって (4.13) は指数型分布族ではない。

例 4.6 (Cauchy 分布). [TODO]

さて指数型分布族の定義では g, a_i, T_i, ψ などいくつかの関数が出てきた。ここではとくに ψ に着目してその性質を調べる。

命題 4.7 (対数分配関数の凸性). $\Theta \subset \mathbb{R}^n$ が凸集合で $a_i(\theta) = \theta_i$ $(i = 1, \ldots, n)$ ならば、対数分配関数 ψ は凸関数である。

証明 示したいことは

$$\psi(t\theta + (1-t)\theta') \le t\psi(\theta) + (1-t)\psi(\theta') \quad (\forall t \in (0,1),\ \theta, \theta' \in \Theta) \tag{4.16}$$

である。

$$\begin{aligned}
\psi(t\theta + (1-t)\theta') &= \log \int_{\mathcal{X}} g(x) \exp\langle t\theta + (1-t)\theta', T(x)\rangle \mu(dx) && (4.17)\\
&= \log \int_{\mathcal{X}} g(x)^t \exp t\langle \theta, T(x)\rangle\ g(x)^{1-t} \exp(1-t)\langle \theta', T(x)\rangle \mu(dx) && (4.18)\\
&\le \log \left(\int_{\mathcal{X}} g(x) \exp\langle \theta, T(x)\rangle \mu(dx)\right)^t \left(\int_{\mathcal{X}} g(x) \exp\langle \theta', T(x)\rangle \mu(dx)\right)^{1-t} \quad (\text{Hölder の不等式}) && (4.19)\\
&= t \log \int_{\mathcal{X}} g(x) \exp\langle \theta, T(x)\rangle \mu(dx) + (1-t) \log \int_{\mathcal{X}} g(x) \exp\langle \theta', T(x)\rangle \mu(dx) && (4.20)\\
&= t\psi(\theta) + (1-t)\psi(\theta') && (4.21)
\end{aligned}$$

よって示せた。 □

命題 4.8 (対数分配関数の微分可能性). $\Theta \subset \mathbb{R}^n$ が開集合で $a_i(\theta) = \theta_i$ $(i = 1, \ldots, n)$ ならば、対数分配関数 ψ は Θ 上で任意回微分可能である。

証明 証明の要点に集中するために、$n = 1$ の場合のみ示す。積分記号下での微分を行うため、次の claim を示す。

- 任意の $k \in \mathbb{Z}_{\ge 0}$ と $\theta \in \Theta$ に対し、$g(x)T(x)^k e^{\theta T(x)}$ は $\mathcal{X}$ 上 μ-可積分である。

$\because$ $k \in \mathbb{Z}_{\ge 0}$ に関する帰納法で示す。$k = 0$ で成り立つことは、すべての $\theta \in \Theta$ に対し $\psi(\theta)$ が定義されることからわかる。

与えられた $k \in \mathbb{Z}_{\ge 0}$ で成立を仮定し、$k+1$ での成立を示す。$\theta \in \Theta$ を任意に固定する。Θ が開集合であることから、ある $\theta' > 0$ が存在して $\theta \pm \theta' \in \Theta$ が成り立つ。$g(x)T(x)^{k+1}e^{\theta T(x)}$ が $\mathcal{X}$ 上 μ-可積分であることを示すため、$|g(x)T(x)^{k+1}e^{\theta T(x)}| \le \Phi_\theta(x)$ なる μ-可積分関数 $\Phi_\theta \colon \mathcal{X} \to \mathbb{R}$ を構成する。

ここで各 $x \in \mathcal{X}$ に対し $T(x)$ の符号で場合分けすると次が成り立つ。$T(x) > 0$ のとき

$$\begin{aligned}
|g(x)T(x)^{k+1}e^{\theta T(x)}| &= |g(x)T(x)^k| T(x) e^{\theta T(x)} && (4.22)\\
&= \frac{1}{\theta'} |g(x)T(x)^k| \theta' T(x) e^{\theta T(x)} && (4.23)\\
&\le \frac{1}{\theta'} |g(x)T(x)^k| e^{(\theta+\theta')T(x)} \quad (\because\ \theta' T(x) \le e^{\theta' T(x)}) && (4.24)\\
&= \frac{1}{\theta'} |g(x)T(x)^k e^{(\theta+\theta')T(x)}| && (4.25)
\end{aligned}$$

$T(x) \leq 0$ のとき

$$\begin{aligned}
|g(x)T(x)^{k+1}e^{\theta T(x)}| &= |g(x)T(x)^k|(-T(x))e^{\theta T(x)} && (4.26)\\
&= \frac{1}{\theta'}|g(x)T(x)^k|(-\theta' T(x))e^{\theta T(x)} && (4.27)\\
&\leq \frac{1}{\theta'}|g(x)T(x)^k|e^{(\theta-\theta')T(x)} \quad (\because -\theta' T(x) \leq e^{-\theta' T(x)}) && (4.28)\\
&= \frac{1}{\theta'}|g(x)T(x)^k e^{(\theta-\theta')T(x)}| && (4.29)
\end{aligned}$$

そこで $\Phi_\theta(x)$ を上の 2 つの不等式の右辺の和と定めれば、帰納法の仮定よりそれぞれは $\mathcal{X}$ 上 μ-可積分だから、Φ_θ も $\mathcal{X}$ 上 μ-可積分である。したがって $g(x)T(x)^{k+1}e^{\theta T(x)}$ も $\mathcal{X}$ 上 μ-可積分である。これで帰納法が完成した。 //

$\psi(\theta) = \log \int g(x)e^{\theta T(x)}\mu(dx)$ の被積分関数は θ に関し微分可能で、導関数 $g(x)T(x)e^{\theta T(x)}$ は補題より $\mathcal{X}$ 上 μ-可積分だから、積分記号下の微分ができる。したがって

$$\begin{aligned}
\psi'(\theta) &= \frac{\int_{\mathcal{X}} g(x)T(x)e^{\theta T(x)}\mu(dx)}{\int_{\mathcal{X}} g(x)e^{\theta T(x)}\mu(dx)} && (4.30)\\
&= \frac{e^{-\psi(\theta)}\int_{\mathcal{X}} g(x)T(x)e^{\theta T(x)}\mu(dx)}{\int_{\mathcal{X}} g(x)e^{\theta T(x)-\psi(\theta)}\mu(dx)} && (4.31)\\
&= e^{-\psi(\theta)}\int_{\mathcal{X}} g(x)T(x)e^{\theta T(x)}\mu(dx) && (4.32)
\end{aligned}$$

を得る。上の claim より、右辺に現れる積分も積分記号下の微分ができるから、帰納的に ψ は任意回微分可能であることがわかる。 □

B. 混合型分布族

混合型分布族について述べる。

定義 4.9 (混合型分布族). $(\mathcal{X}, \mathcal{A})$ を可測空間、μ を $(\mathcal{X}, \mathcal{A})$ 上の σ-有限測度、$H \subset (0,1)^n$ を部分集合、$\mathcal{P} = (\mathcal{P}_\eta)_{\eta \in H}$ を $\mathcal{X}$ 上の確率分布族とする。ここで、ある

- $m \in \mathbb{Z}_{\geq 1}$
- $\mathcal{X}$ 上の確率分布 q_i $(i = 1, \ldots, k)$

が存在し、各 P_η が μ に関し絶対連続で、確率密度関数が

$$\frac{dP_\eta}{d\mu}(x) = \sum_{i=1}^{m} \eta_i q_i(x) \quad \mu\text{-a.e. } x \in \mathcal{X}, \qquad \eta_i \in H, \qquad \eta_1 + \cdots + \eta_m = 1 \tag{4.33}$$

の形に表されるとき、$\mathcal{P}$ を**混合型分布族 (mixture family)** という。[TODO] 1 次独立性は何に使う？

命題 4.10 (負のエントロピー). [TODO]

証明 [TODO] □

5 今後の予定

- [TODO]