情報幾何学

Yahata

Preface

本稿では情報幾何学について整理する。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、微分幾何学の一分野である。

情報幾何学で扱われる代表的な対象のひとつが、数理統計学におけるパラメトリックモデルを幾何学的に定式化した、パラメータ付き測度モデルである。パラメータ付き測度モデルとは、パラメータ全体の多様体 M と可測空間 $\mathcal{X}$、パラメータに測度を割り付ける滑らかな写像 $\mathbf{p}$ からなる 3 つ組 $(M, \mathcal{X}, \mathbf{p})$ のことである。ここで $\mathbf{p}$ の像は $\mathcal{X}$ 上の測度全体の多様体 $\mathcal{M}(\mathcal{X})$ に埋め込まれているとみなせるが、$\mathcal{M}(\mathcal{X})$ には自然に Fisher 計量や Amari-Chentsov テンソルなどの幾何学的構造が入っているから、M 上にこれらの構造を引き戻せば、M を介して $\mathcal{M}(\mathcal{X})$ を調べることができる。これが情報幾何の基本的なアイディアのひとつである。

さてこのような外在的 (extrinsic) な見方は幾何学の研究において重要であるが、一方で空間 M の内在的 (intrinsic) な性質を峻別することが難しくなるという問題もある。M の内在的性質としては、双対構造や Amari-Chentsov 構造、双対平坦性などの概念が基本的である。そこでパラメータ付き測度モデルから離れて、これらの内在的性質を研究するという方向性もある。

測度全体の多様体 $\mathcal{M}(\mathcal{X})$ についてもう少し詳しく考えてみる。ある可測空間 $(\mathcal{X}, \mathcal{A})$ 上の符号付き測度全体の空間は、全変動をノルムとして Banach 空間をなす。この空間を Banach 多様体とみれば、$(\mathcal{X}, \mathcal{A})$ 上の確率分布の全体は (Banach) 部分多様体として扱うことができる。[AJLS17] ではこのような枠組みが情報幾何学の基礎となっている。Banach 多様体の教科書は [Lan85] が有名であり、より一般的な無限次元多様体を論じた文献には [Sch22] がある。ただし本稿では Banach 多様体よりも一般の無限次元多様体を扱うことはしない。また (測度の空間 $\mathcal{M}(\mathcal{X})$ は無限次元であるが) パラメータの空間 M は主に有限次元の場合のみを扱う。

本稿を読み進めるための前提知識としては数学科の学部 3 年生レベルの知識を仮定する。ただし確率・統計については学部 1 年生レベルの知識があれば十分である。

目次

第 I 部

確率と統計

第 1 部では、情報幾何の議論に必要となる確率論と数理統計の前提知識を整理する。ただしここではごく基礎的な事項のみを述べ、それ以外の話題は第 2 部で情報幾何的な視点から扱うものとする。そもそも確率・統計を幾何学的に定式化しようとする際には、座標変換に関する不変性などを定義に繰り入れる必要がある。このとき定義が見かけ上煩雑になることで、確率・統計的な意味を一時的に見失ってしまうおそれがある。それを防ぐために、確率・統計的な意味を一旦明らかにしておくのもこの部の目的である。

第 1 章　確率論

確率論の基礎事項を整理する。

1.1　確率空間

定義 1.1.1 (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 1.1.2 (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。とくに $\mathcal{X} = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ (Borel 集合族) のとき、X を単に**確率変数**という。

定義 1.1.3 (確率変数の確率分布)**.** $X \colon (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ を確率変数とする。このとき、写像

$$P^X \colon \mathcal{A} \to [0, +\infty], \quad E \mapsto P(X^{-1}(E)) \quad (E \in \mathcal{A}) \tag{1.1.1}$$

は $(\mathcal{X}, \mathcal{A})$ 上の測度となる (このあと示す)。これを **X の確率分布 (probability distribution of 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 **ν に従う** という。

証明　[TODO] □

確率論には「**0-1 法則 (zero-one law)**」と呼ばれるいくつかの定理がある。次に述べる Borel-Cantelli の補題はそのひとつである。

定理 1.1.4 (Borel-Cantelli の補題)**.** [TODO]

証明　[TODO] □

定理 1.1.5 (Jensen の不等式)**.** $(\Omega, \mathcal{F}, P)$ を確率空間、$A \subset \mathbb{R}^n$ を凸集合、$\Psi \colon A \to \mathbb{R}$ を凸関数とする。このとき、$\Psi(f)$ が P-可積分となるような任意の P-可積分関数 f に対し

$$\Psi\left(\int_\Omega f(x)\, P(dx)\right) \leq \int_\Omega \Psi(f(x))\, P(dx) \tag{1.1.2}$$

が成り立つ。

証明 [TODO] cf. [Bog07, p.153] □

1.2 離散確率分布

定義 1.2.1 (離散確率分布)**.** $\mathcal{X}$ を高々可算集合、$\mathcal{A} := 2^{\mathcal{X}}$ とするとき、可測空間 $(\mathcal{X}, \mathcal{A})$ を**離散確率空間 (discrete probability space)** といい、$(\mathcal{X}, \mathcal{A})$ 上の確率分布を**離散確率分布 (discrete probability distribution)** という。

定義 1.2.2 (確率質量関数)**.** $(\mathcal{X}, \mathcal{A})$ を離散確率空間、μ を $(\mathcal{X}, \mathcal{A})$ 上の確率分布とする。このとき、$(\mathcal{X}, \mathcal{A})$ 上の数え上げ測度に関する μ の Radon-Nikodym 微分 [1)]、すなわち

$$\mu(E) = \sum_{x \in E} p(x) \quad (E \in \mathcal{A}) \tag{1.2.1}$$

なる関数 $p \colon \mathcal{X} \to [0, +\infty]$ を μ の**確率質量関数 (probability mass function; PMF)** という。

定義 1.2.3 (確率母関数)**.** [TODO]

1.3 連続確率分布

$\mathbb{R}$ 上の確率分布は確率論や統計学において重要である。

定義 1.3.1 (連続確率分布)**.** $\mathcal{B}$ を $\mathbb{R}$ の Borel 集合族とするとき、可測空間 $(\mathbb{R}, \mathcal{B})$ を**連続確率空間 (continuous probability space)** といい、$(\mathbb{R}, \mathcal{B})$ 上の確率分布を**連続確率分布 (continuous probability distribution)** という。

定義 1.3.2 (絶対連続分布)**.** [TODO]

定義 1.3.3 (確率密度関数)**.** $(\mathbb{R}, \mathcal{B})$ を連続確率空間、μ を $(\mathbb{R}, \mathcal{B})$ 上の Lebesgue 測度に関し絶対連続な確率分布とする。このとき、μ の Radon-Nikodym 微分、すなわち

$$\mu(E) = \int_E p(x)\, dx \quad (E \in \mathcal{B}) \tag{1.3.1}$$

なる関数 $p \colon \mathbb{R} \to [0, +\infty]$ を μ の**確率密度関数 (probability density function; PDF)** という。

定義 1.3.4 (モーメント母関数)**.** [TODO]

モーメント母関数と異なり、特性関数は常に存在する。

定義 1.3.5 (特性関数)**.** [TODO]

1) 数え上げ測度に関する μ の Radon-Nikodym 微分はつねに存在する。なぜなら、$(\mathcal{X}, \mathcal{A})$ 上のいかなる確率測度も $(\mathcal{X}, \mathcal{A})$ 上の数え上げ測度に関して絶対連続だからである。

1.4 多変量分布

[TODO]

1.5 確率変数の収束

定理 1.5.1 (大数の法則). [TODO]

証明 [TODO] □

定理 1.5.2 (中心極限定理). [TODO]

証明 [TODO] □

第 2 章　推定

この章では、統計的推論の基本的な問題のひとつである推定について述べる。

2.1　十分統計量

標本 (=確率変数) [TODO] 標本とは確率変数のことなのか？Ω の元ではないのか？の実現値に対しその特徴を要約した値を割り当てる関数を統計量という。事象は統計量を通して「観測」される。もちろん標本それ自体も統計量である。

定義 2.1.1 (統計量). [TODO]

統計量を用いて母集団 (=確率分布) のパラメータを推定することを考える。

定義 2.1.2 (点推定). 標本 X から母集団のパラメータ θ を特定することを**推定 (estimate)** といい、とくに一意に推定することを**点推定 (point estimation)** という。点推定において、標本 X の実現値 x に対し θ の**推定値 (estimate)** $\widehat{\theta}(x)$ を割り当てる関数 $\widehat{\theta}$ を θ の**推定量 (estimator)** という。

注意 2.1.3. 今後、点推定のことを単に推定ということにする。

推定のために"十分"な情報を含んだ統計量を十分統計量という。

定義 2.1.4 (十分統計量). [TODO]

例 2.1.5 (Bernoulli 分布の例). [TODO]

次の定理はある統計量が十分統計量であるための必要十分条件を与え、具体的な判定に役立つ。

定理 2.1.6 (Fisher-Neyman の分解定理). [TODO]

証明　[TODO] □

例 2.1.7 (正規分布の例). [TODO]

2.2　指数型分布族

定義 2.2.1 (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、$\mathcal{P} = (P_\theta)_{\theta\in\Theta}$ を $(\mathcal{X}, \mathcal{A})$ 上の確率分布族とする。ここで、各 P_θ が μ に関し絶対連続で、Radon-Nikodym 微分が

$$\frac{dP_\theta}{d\mu}(x) = g(x)\exp\left(\sum_{i=1}^{m} a_i(\theta)T_i(x) - \psi(\theta)\right) = g(x)\exp(a(\theta)\cdot T(x) - \psi(\theta)) \quad \mu\text{-a.e. } x \in \mathcal{X} \tag{2.2.1}$$

の形に表せるとき、$\mathcal{P}$ を**指数型分布族 (exponential family)** という。ただし、T_i, g は $(\mathcal{X}, \mathcal{A})$ 上の可測関数、a_i, ψ は Θ 上の実数値関数である。とくに

$$\frac{dP_\theta}{d\mu}(x) = g(x)\exp\left(\sum_{i=1}^{m}\theta_i T_i(x) - \psi(\theta)\right) = g(x)\exp(\theta \cdot T(x) - \psi(\theta)) \quad \mu\text{-a.e. } x \in \mathcal{X} \tag{2.2.2}$$

の形を**正準形 (canonical form)** という [2]。

以下に指数型分布族の簡単な例を挙げる。

例 2.2.2. [TODO]

命題 2.2.3. 指数型分布族の定義の $T = (T_1, \ldots, T_m)$ は θ の十分統計量である。

証明 Fisher-Neyman の分解定理より従う。 □

2.3 混合型分布族

定義 2.3.1 (混合型分布族). [TODO]

2.4 点推定の手法

定義 2.4.1 (最尤推定). [TODO]

定義 2.4.2 (モーメント法). [TODO]

定義 2.4.3 (最小 2 乗法). [TODO]

2.5 推定量の性質

定義 2.5.1 (不偏性). [TODO]

定義 2.5.2 (一様最小分散不偏推定量; UMVUE). [TODO]

2.6 推定量の評価

定義 2.6.1 (スコア関数). [TODO]

Fisher 情報量を定義する。

2) 指数型分布族は常に正準形で書くことができる。実際、θ の代わりに $a(\theta)$ をパラメータとすればよい。

定義 2.6.2 (Fisher 情報量). [TODO]

不偏推定量の分散の下界を Fisher 情報量の言葉で与えるのが Cramer-Rao 不等式である。[TODO] Cramer-Rao 不等式は Fisher "計量" に対しどのような意味を持つ？

命題 2.6.3 (Cramer-Rao 不等式). [TODO]

証明 [TODO] □

定義 2.6.4 (有効性). [TODO]

2.7 漸近理論

漸近理論について述べる。一般に最尤推定量は有効性をみたすとは限らないが、漸近的には有効性をみたすことが知られている[1)]。

[TODO]

1） ただし Neyman-Scott 問題においては最尤推定量は漸近有効とは限らない。

第 3 章　仮説検定

この章では、統計的推論の基本的な問題のひとつである仮説検定について述べる。

3.1　仮説検定

[TODO]

第 II 部

情報幾何学

第 2 部では、情報幾何の基本的な概念を整理する。まず第 1 章と第 2 章では具体例を扱う。第 1 章では有限集合上の測度の空間を考える。測度の空間を多様体とみなし、さらに統計的に自然な要請に基づいて計量や接続、ダイバージェンスなどの概念を導入する。第 2 章ではパラメトリックモデルを考える。

第 3 章と第 4 章では、パラメトリックモデルのパラメータ空間の内在的な幾何構造として、双対構造と双対平坦性を整理する。

第 5 章では無限集合上の測度の空間を考える。

第 6 章では正定値行列の空間について考える。

第 7 章では統計的推定について考える。

第 4 章　有限集合上の測度の空間

この章では有限集合上の確率分布について考える。有限集合上の確率分布を扱う目的は、これが情報幾何学の最も基本的な具体例だからというだけでなく、パラメータ空間の内在的な幾何や無限集合上の確率分布を扱う際に、幾何学的直感の基礎になると期待できるからである。

この章では全体を通して $n \in \mathbb{Z}_{\geq 0}$, $\mathcal{X} \coloneqq \{1, \dots, n+1\}$, $\mathcal{A} \coloneqq 2^{\mathcal{X}}$ とし、可測空間 $(\mathcal{X}, \mathcal{A})$ を考える。イメージとしては、$(\mathcal{X}, \mathcal{A})$ は何らかの確率変数の (定義域というよりも) 値域を表している。いまはそれが有限集合ということだから、たとえば「コインの裏表」「サイコロの出目」などといった、値が有限通りしかない確率変数を考えているという雰囲気である。

4.1　有限集合上の測度の空間

まず $\mathcal{X}$ 上の実数値関数全体の集合を考える。

定義 4.1.1. $\mathbb{R}$-代数 $\mathcal{F}(\mathcal{X})$ を

$$\mathcal{F}(\mathcal{X}) \coloneqq \{f \colon \mathcal{X} \to \mathbb{R}\} \tag{4.1.1}$$

と定め、$\mathcal{F}(\mathcal{X})$ の $\mathbb{R}$-ベクトル空間としての (代数的) 双対空間 $\mathcal{F}(\mathcal{X})^*$ を $\mathcal{S}(\mathcal{X})$ とおく。

先に注意しておくと、この章で主役となるのは $\mathcal{F}(\mathcal{X})$ よりもむしろ $\mathcal{S}(\mathcal{X})$ の方である。さて、$\mathcal{F}(\mathcal{X})$ は次の標準的な基底を持つ。

命題 4.1.2 ($\mathcal{F}(\mathcal{X})$ の標準基底). 各 $i \in \mathcal{X}$ に対して

$$e_i(j) \coloneqq \begin{cases} 1 & (j = i) \\ 0 & (j \neq i) \end{cases} \quad (j \in \mathcal{X}) \tag{4.1.2}$$

と定めると、$(e_1, \dots, e_{n+1})$ は $\mathcal{F}(\mathcal{X})$ の $\mathbb{R}$ 上の基底となる。

証明　$\mathbb{R}$-線型写像 $\mathbb{R}^{n+1} \to \mathcal{F}(\mathcal{X})$, $(x^1, \dots, x^{n+1}) \mapsto \sum_{i \in \mathcal{X}} x^i e_i$ が写像 $\mathcal{F}(\mathcal{X}) \to \mathbb{R}^{n+1}$, $f \mapsto (f(1), \dots, f(n+1))$ を逆写像として同型を与える。 □

そこで $(e_1, \dots, e_{n+1})$ の双対基底を $(\delta^1, \dots, \delta^{n+1})$ とおく。すると $\mathcal{S}(\mathcal{X})$ について次が成り立つ。

命題 4.1.3 ($\mathcal{S}(\mathcal{X})$ と符号付き測度の空間). $\mathcal{S}(\mathcal{X})$ は $\mathcal{X}$ 上の符号付き測度全体の空間とみなせる。より詳しく言えば、$\mathcal{X}$ 上の符号付き測度全体の空間は、要素ごとの和とスカラー倍で $\mathbb{R}$-ベクトル空間となり、写像

$$\Phi \colon \mathcal{S}(\mathcal{X}) \to \{\mathcal{X} \text{ 上の符号付き測度}\}, \quad \omega = \sum_{i \in \mathcal{X}} \omega_i \delta^i \mapsto \Big(A \mapsto \sum_{i \in A} \omega_i\Big) \tag{4.1.3}$$

が $\mathbb{R}$-線型同型を与える。

証明 Φ が $\mathbb{R}$-線型写像であることの証明はルーティンワークである。逆写像は

$$\Psi\colon \{\mathcal{X}\text{ 上の符号付き測度}\} \to \mathcal{S}(\mathcal{X}), \quad \mu \mapsto \sum_{i\in\mathcal{X}} \mu(\{i\})\delta^i \tag{4.1.4}$$

で与えられる。$\Psi \circ \Phi(\omega) = \omega$ の証明に ω の線型性を用いる。$\Phi \circ \Psi(\mu) = \mu$ の証明に μ の加法性を用いる。
[TODO] もう少しちゃんと書く □

上の命題の同型により、$\mathcal{F}(\mathcal{X})$ 上の線型形式 δ^i は、$\mathcal{X}$ 上の Dirac 測度 $\Phi(\delta^i)$ と同一視できる。そこで記号の濫用により、以後 $\Phi(\delta^i)$ も δ^i と書くことにする。

さて、$\mathcal{S}(\mathcal{X})$ の部分集合を定義する。

定義 4.1.4 ($\mathcal{S}(\mathcal{X})$ の部分集合)**.** $\mathcal{S}(\mathcal{X})$ の部分集合を次のように定義する:

$$\mathcal{M}(\mathcal{X}) := \left\{ \mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{S}(\mathcal{X}) \,\middle|\, \mu_i \geq 0 \right\}, \tag{4.1.5}$$

$$\mathcal{P}(\mathcal{X}) := \left\{ \mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{M}_+(\mathcal{X}) \,\middle|\, \sum_{i\in\mathcal{X}} \mu_i = 1 \right\} \tag{4.1.6}$$

すなわち、$\mathcal{M}_+(\mathcal{X})$ は $\mathcal{X}$ 上の有限測度全体の集合であり、$\mathcal{P}_+(\mathcal{X})$ は $\mathcal{X}$ 上の確率測度全体の集合である。また、次のように定義する:

$$\mathcal{M}_+(\mathcal{X}) := \left\{ \mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{S}(\mathcal{X}) \,\middle|\, \mu_i > 0 \right\}, \tag{4.1.7}$$

$$\mathcal{P}_+(\mathcal{X}) := \left\{ \mu = \sum_{i\in\mathcal{X}} \mu_i\delta^i \in \mathcal{M}_+(\mathcal{X}) \,\middle|\, \sum_{i\in\mathcal{X}} \mu_i = 1 \right\} \tag{4.1.8}$$

すなわち、$\mathcal{M}_+(\mathcal{X})$ は $\mathcal{X}$ 上の正の測度全体の集合であり、$\mathcal{P}_+(\mathcal{X})$ は $\mathcal{X}$ 上の正の確率測度全体の集合である。

$\mathcal{M}_+(\mathcal{X})$ および $\mathcal{P}_+(\mathcal{X})$ は、次の意味で部分多様体になっている。

命題 4.1.5. $\mathcal{M}_+(\mathcal{X})$ は $\mathcal{S}(\mathcal{X})$ の開部分多様体であり、$\mathcal{P}_+(\mathcal{X})$ は $\mathcal{M}_+(\mathcal{X})$ の n 次元部分多様体である。

証明 多様体としての同一視 $\mathcal{S}(\mathcal{X}) = \mathbb{R}^{n+1}$ のもとで $\mathcal{M}_+(\mathcal{X}) = \mathbb{R}^{n+1}_{>0} \overset{\text{open}}{\subset} \mathbb{R}^{n+1} = \mathcal{S}(\mathcal{X})$ が成り立つから、$\mathcal{M}_+(\mathcal{X})$ は $\mathcal{S}(\mathcal{X})$ の開部分多様体である。また、C^∞ 写像

$$F\colon \mathcal{M}_+(\mathcal{X}) = \mathbb{R}^{n+1}_{>0} \to \mathbb{R}, \quad (\mu_1, \ldots, \mu_{n+1}) \mapsto \mu_1 + \cdots + \mu_{n+1} \tag{4.1.9}$$

の Jacobi 行列は $JF_{(\mu_1,\ldots,\mu_{n+1})} = [1 \ldots 1]$ だから、微分 $dF_{(\mu_1,\ldots,\mu_{n+1})}$ は各点 $(\mu_1, \ldots, \mu_{n+1}) \in \mathbb{R}^{n+1}_{>0}$ で全射であり、したがってとくに F は 1 を正則値にもつ。よって $\mathcal{P}_+(\mathcal{X}) = F^{-1}(0)$ は $\mathcal{M}_+(\mathcal{X})$ の $\dim \mathcal{M}_+(\mathcal{X}) - \dim \mathbb{R} = n$ 次元部分多様体である。 □

それぞれの多様体の接空間は次のように表される。

命題 4.1.6 ($\mathcal{S}(\mathcal{X}), \mathcal{M}_+(\mathcal{X}), \mathcal{P}_+(\mathcal{X})$ の接空間).

$$T_\mu \mathcal{S}(\mathcal{X}) = \mathcal{S}(\mathcal{X}), \tag{4.1.10}$$
$$T_\mu \mathcal{M}_+(\mathcal{X}) = \mathcal{S}(\mathcal{X}), \tag{4.1.11}$$
$$T_\mu \mathcal{P}_+(\mathcal{X}) = \mathcal{S}_0(\mathcal{X}) \tag{4.1.12}$$

ただし

$$\mathcal{S}_0 := \left\{ a = \sum_{i \in \mathcal{X}} a_i \delta^i \in \mathcal{S}(\mathcal{X}) \,\middle|\, \sum_{i \in \mathcal{X}} a_i = 0 \right\} \tag{4.1.13}$$

とおいた。

証明　$T_\mu \mathcal{S}(\mathcal{X}) = \mathcal{S}(\mathcal{X})$ および $T_\mu \mathcal{M}_+(\mathcal{X}) = \mathcal{S}(\mathcal{X})$ は、一般にベクトル空間の開部分多様体の接空間がもとのベクトル空間に一致することから従う。また、命題 4.1.5 の証明の写像 F を用いると、部分多様体の接空間の一般論より $T_\mu \mathcal{P}_+(\mathcal{X}) = \operatorname{Ker} dF_\mu$ が成り立つが、ここで

$$dF_{(\mu_1,\dots,\mu_{n+1})} \colon \mathbb{R}^{n+1} \to \mathbb{R}, \quad \begin{pmatrix} a_1 \\ \vdots \\ a_{n+1} \end{pmatrix} \mapsto JF_{(\mu_1,\dots,\mu_{n+1})} \begin{pmatrix} a_1 \\ \vdots \\ a_{n+1} \end{pmatrix} = a_1 + \cdots + a_{n+1} \tag{4.1.14}$$

より $T_\mu \mathcal{P}_+(\mathcal{X}) = \operatorname{Ker} dF_\mu = \mathcal{S}_0(\mathcal{X})$ が従う。 □

4.2 Fisher 計量

$\mathcal{S}(\mathcal{X})$ に Riemann 計量を定義する。考えられる Riemann 計量は無数に存在するが、ここでは Fisher 計量と呼ばれる計量を導入する。$\mathcal{M}_+(\mathcal{X})$ および $\mathcal{P}_+(\mathcal{X})$ には誘導計量を入れる。

最適化理論において、自然勾配法で解の更新を行う際には、解きたい問題に適した計量を用いる必要がある。統計的推論を最適化問題として解く場合、Fisher 計量はその目的に最も適した計量であると考えられる。

[TODO] 下の定義をもっと統計的要請から自然に導入したい

定義 4.2.1 (Fisher 計量). 各 $\mu = \mu_i \delta^i \in \mathcal{M}_+(\mathcal{X})$, $a = a_i \delta^i, b = b_i \delta^i \in \mathcal{S}(\mathcal{X})$ に対し

$$\langle a, b \rangle_\mu := \int_{\mathcal{X}} \frac{da}{d\mu} \frac{db}{d\mu} \, d\mu = \sum_{i \in \mathcal{X}} \frac{1}{\mu_i} a_i b_i \tag{4.2.1}$$

と定義する。$\mathcal{M}_+(\mathcal{X})$ 上の Riemann 計量 $\mathfrak{g}$ を

$$\mathfrak{g}_\mu(a, b) := \langle a, b \rangle_\mu \quad (\mu \in \mathcal{M}_+(\mathcal{X}),\ a, b \in T_\mu \mathcal{M}_+(\mathcal{X})) \tag{4.2.2}$$

と定義し、これを **Fisher 計量 (Fisher metric)** と呼ぶ。

$\mathcal{P}_+(\mathcal{X})$ 上に誘導される Fisher 計量は次の形で与えられる。

命題 4.2.2 ($\mathcal{P}_+(\mathcal{X})$ 上の Fisher 計量).

$$g_{ij}(\mu) = \sum_{k=1}^{n} \frac{1}{\mu_k} \delta_{ki} \delta_{kj} + \frac{1}{\mu_{n+1}} \tag{4.2.3}$$

[TODO]

証明 [TODO] □

A. 等長変換群

[TODO]

4.3 Chentsov の定理

前節で導入した Fisher 計量が、統計的に自然な計量であることを示す。本節では $n' \in \mathbb{Z}_{\geq 0}$, $\mathcal{X}' := \{1, \dots, n'+1\}$, $\mathcal{A}' := 2^{\mathcal{X}'}$ とし、可測空間 $(\mathcal{X}', \mathcal{A}')$ を $(\mathcal{X}, \mathcal{A})$ とあわせて考える。まず、統計量と Markov カーネルの概念を導入する。

定義 4.3.1 (統計量). $\mathcal{X} \to \mathcal{X}'$ なる写像 κ を**統計量 (statistic)** という [4]。

定義 4.3.2 (Markov カーネル). $\mathcal{X} \to \mathcal{P}(\mathcal{X}')$ なる写像 K を **Markov カーネル (Markov kernel)** といい、ここだけの記法として $K(i)$ $(i \in I)$ の成分表示を

$$K(i) = \sum_{i' \in \mathcal{X}'} K(i)_{i'} \delta^{i'} \tag{4.3.1}$$

と書くことにする。

統計量 $\kappa \colon \mathcal{X} \to \mathcal{X}'$ が与えられると、$K^{\kappa}(i) := \delta^{\kappa(i)}$ $(i \in I)$ とおくことで Markov カーネル K^{κ} が定義される。ただし右辺は $\kappa(i)$ を台とする Dirac 測度である。この意味で、Markov カーネルは統計量の一般化になっている。

次に Markov カーネルの congruent 性を定義する。

定義 4.3.3 (congruent な Markov カーネル). [TODO]

例 4.3.4 (congruent な Markov カーネルの例). $\mathcal{X} := \{0, 1\}$, $\mathcal{X}' := \{1, 2, 3, 4, 5, 6\}$ とし、Markov カーネル $K \colon \mathcal{X} \to \mathcal{P}(\mathcal{X}')$ を [TODO]

次の定理により、$\mathcal{P}_+(\mathcal{X})$ 上の統計的に自然な計量は Fisher 計量に限られることがわかる。

定理 4.3.5 (Chentsov; Ченцов). すべての空でない有限集合 $\mathcal{X}$ に対し、$\mathcal{P}_+(\mathcal{X})$ 上の Riemann 計量 $h^{\mathcal{X}}$ が与えられているとする。このとき、任意の congruent な Markov カーネル $K \colon \mathcal{X} \to \mathcal{P}(\mathcal{X}')$ ($\mathcal{X}'$ は空でない有限集合) が不変ならば、ある正定数 $\alpha > 0$ が存在して、すべての $\mathcal{X}$ に対し $h^{\mathcal{X}} = \alpha g^{\mathcal{X}}$ が成り立つ。

証明 [TODO] □

4.4 m-接続と e-接続

この節では m-接続と e-接続を導入する。まず $T_{\mu}\mathcal{M}_+(\mathcal{X})$ や $T_{\mu}\mathcal{M}_+^*(\mathcal{X})$ の定義から、m, e-平行移動が自ずと現れることをみる。m, e-平行移動から、それぞれに対応する m, e-接続の概念が、ひいては m, e-測地線、m, e-指数写像の概念が導かれる。

4) より一般には κ が可測写像であることも要求されるが、いまは $\mathcal{A} = 2^{\mathcal{X}}$ だから、この要求は自ずから満たされる。

[TODO] m, e-接続には統計的にどんな意味がある？

定義 4.4.1 (m, e-平行移動). [TODO]

定義 4.4.2 (m, e-接続). [TODO]

命題 4.4.3 (m, e-測地線と指数写像). [TODO]

証明 [TODO] □

A. アファイン群

[TODO]

4.5 Amari-Chentsov テンソル

[TODO] なんのためにある？ Skewness テンソルとも呼ぶから、3 次のキュムラント (歪度) と関係あり？

定義 4.5.1 (Amari-Chentsov テンソル). [TODO]

定義 4.5.2 (α-接続). [TODO]

A. アファイン群

[TODO]

4.6 ダイバージェンス

前節までに、$\mathcal{M}_+(\mathcal{X})$ 上の幾何学的構造として、Fisher 計量 $\mathfrak{g}$ や m, e-接続 $\nabla^{(m)}, \nabla^{(e)}$ を扱った。本節ではこれらの幾何学的構造がダイバージェンスと呼ばれるひとつの関数により誘導されることをみる。ダイバージェンスとは、空間の 2 点の相違性を表す関数であり、イメージとしては距離のようなものである。最適輸送理論における輸送コストにも似ているが、一般にはダイバージェンスの方が粗いものである [TODO] 本当に?。

定義 4.6.1 (KL ダイバージェンス). [TODO]

命題 4.6.2 (KL ダイバージェンスは Fisher 計量を誘導する). [TODO]

証明 [TODO] □

A. α-ダイバージェンス

Fisher 計量を誘導するダイバージェンスは、実は KL ダイバージェンスだけではない。α-ダイバージェンスと呼ばれる、より広いクラスのダイバージェンスが Fisher 計量を誘導する。

定義 4.6.3 (α-ダイバージェンス)**.**

$$D^{(\alpha)}(\mu\|\nu) := \frac{2}{1-\alpha}\sum_{i\in X}\nu_i + \frac{2}{1+\alpha}\sum_{i\in X}\mu_i - \frac{4}{1-\alpha^2}\sum_{i\in X}\mu_i^{\frac{1+\alpha}{2}}\nu_i^{\frac{1-\alpha}{2}} \tag{4.6.1}$$

[TODO]

命題 4.6.4 (α-ダイバージェンスは Fisher 計量を誘導する)**.** [TODO]

証明 [TODO] □

KL ダイバージェンスは α-ダイバージェンスの $\alpha = -1$ の場合である。

命題 4.6.5. KL ダイバージェンスは α-ダイバージェンスの $\alpha = -1$ の場合である。

証明 [TODO] □

B. f-ダイバージェンス

ダイバージェンスを最小化したいという最適化問題への応用を考えると、凸関数を用いて定義されるダイバージェンスのクラスは重要である [TODO] そうなの？。

定義 4.6.6 (f-ダイバージェンス)**.** [TODO]

4.7 指数型分布族

[TODO]

第 5 章　パラメトリックモデル

本章では、有限個のパラメータで特徴付けられるパラメトリックモデルについて述べる。パラメトリックモデルの定式化は [CU14] を参考にした。

5.1　パラメトリックモデル

[TODO]

第 6 章　双対構造

第 1 章では、有限集合上の測度の空間 $\mathcal{M}_+(\mathcal{X})$ について調べた。また $\mathcal{M}_+(\mathcal{X})$ に入る幾何学的構造としては、Fisher 計量 $\mathfrak{g}$ と m, e-接続 $\nabla^{(m)}, \nabla^{(e)}$ を扱ったのであった。本章ではこの構造を抽象化して整理する。すなわち Fisher 計量および m, e-接続という具体的対象から離れて、Riemann 計量 g と g に関し互いに双対的なアファイン接続 ∇, ∇^* の 3 つ組 (g, ∇, ∇^*) について考える。この 3 つ組は双対構造と呼ばれる。本章ではダイバージェンスにより双対構造が誘導されることをみる。

6.1　双対構造

双対構造とは、Riemann 計量 g と g に関し互いに双対的なアファイン接続 ∇, ∇^* の 3 つ組 (g, ∇, ∇) のことである。双対アファイン接続の概念は、計量を保つアファイン接続の概念の一般化である。

定義 6.1.1 (双対アファイン接続). (M, g) を擬 Riemann 多様体、∇, ∇^* を M のアファイン接続とする。∇^* が g に関する ∇ の**双対アファイン接続 (dual affine connection)** であるとは、

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \tag{6.1.1}$$

が成り立つことをいう。このとき (g, ∇, ∇^*) を M の**双対構造 (dualistic structure)** という。

命題 6.1.2 (双対アファイン接続の存在と一意性). [TODO]

証明　[TODO] 構成は [Lee18, p.96] を参照

一意性を示す。∇', ∇'' を g に関する ∇ の双対アファイン接続とすると、各 $X, Z \in \mathfrak{X}(M)$ に対して

$$0 = g(Y, \nabla'_X Z) - g(Y, \nabla''_X Z) = g(Y, \nabla'_X Z - \nabla''_X Z) \quad (\forall Y \in \mathfrak{X}(M)) \tag{6.1.2}$$

が成り立つから、g の非退化性より $\nabla'_X Z = \nabla''_X Z$ となる。したがって $\nabla' = \nabla''$ である。 □

とくに ∇ が計量 g を保つ接続ならば ∇ 自身も ∇ の双対アファイン接続となるから、一意性より $\nabla^* = \nabla$ が成り立つ。

双対アファイン接続は次のような基本性質を持つ。

命題 6.1.3 (双対アファイン接続の基本性質). [TODO]

証明　[TODO] □

命題 6.1.4 (双対アファイン接続と平行移動). [TODO]

証明　[TODO] □

次の定理は情報幾何学の基本定理である [Nie20]。[TODO] なぜ？曲率一定だと測地線の表示がシンプルになるから？

命題 6.1.5 (双対アファイン接続と曲率)**.** [TODO] torsion-free の仮定は不要！

証明 [CU14, p.226] による。[TODO] □

ところで、双対アファイン接続の概念を Levi-Civita 接続と比較してみたとき、次の定義の意味での「強い」双対性が想起されるかもしれない。

定義 6.1.6 (強い双対アファイン接続)**.** [TODO]

しかし、実は強い双対アファイン接続は Levi-Civita 接続に限られるため、これは文字通り強すぎる制約である。

命題 6.1.7 (強い双対アファイン接続は Levi-Civita 接続)**.** [TODO]

証明 [TODO] □

6.2 統計的構造

有限集合上の測度の空間には、Fisher 計量 $\mathfrak{g}$ と Amari-Chentsov テンソル T の組 $(\mathfrak{g}, T)$ によっても幾何学的構造が定まるのであった。本節ではこれを抽象化して統計的構造と呼ばれるものを導入する。後で見るように統計的構造は、捩れのないアファイン接続の対からなる双対構造と等価である。

[TODO] 捩れなしであることはなぜ重要なのか？そもそもねじれとは何か？

定義 6.2.1 (統計的構造)**.** [TODO]

命題 6.2.2 (双対アファイン接続と Levi-Civita 接続)**.**

$$\frac{1}{2}(\nabla + \nabla^*) = \nabla^{\mathrm{LC}} \tag{6.2.1}$$

[TODO]

証明 [TODO] □

6.3 ダイバージェンス

ダイバージェンスの概念を定義する。ダイバージェンスは双対構造を誘導する方法の 1 つであり、またダイバージェンスにより誘導されるアファイン接続は捩れなしである。同じことだが、ダイバージェンスは統計的構造を誘導する。

定義 6.3.1 (ダイバージェンス)**.** M を多様体とする。C^∞ 関数

$$D\colon M \times M \to \mathbb{R}, \quad (p, q) \mapsto D(p\|q) \tag{6.3.1}$$

が**ダイバージェンス (divergence)** であるとは、次が成り立つことをいう:

(1) 任意の $p, q \in M$ に対し $D(p\|q) \geq 0$ である。

(2) $p, q \in M$ に関し $D(p\|q) = 0 \iff p = q$ である。

(3) [TODO]

ただし、以下の記法を用いる:

- [TODO]

命題 6.3.2. (1) $D[\partial_i\|\cdot] = D[\cdot\|\partial_i] = 0$ である。

(2) $D[\partial_i\partial_j\|\cdot] = D[\cdot\|\partial_i\partial_j] = -D[\partial_i\|\partial_j]$ である。

証明は [Egu92] によった。

証明 (1) [TODO]

(2) [TODO] □

定義 6.3.3 (ダイバージェンスから誘導される双対構造)**.** [TODO]

ダイバージェンスの分解可能性を定義する。

定義 6.3.4 (ダイバージェンスの分解可能性)**.** [TODO]

第 7 章　双対平坦性

情報幾何学では、双対構造がさらに双対平坦な場合を考えることが多い。[TODO] why?

7.1　平坦性とアファインチャート

[TODO] アファイン座標と正規座標はどう違う？正規座標は局所的には必ず存在するが…

定義 7.1.1 (アファインチャート [5]). M を多様体、∇ を M のアファイン接続とする。M のチャート (U, φ) が **∇-アファインチャート (∇-affine chart)** であるとは、(U, φ) により定まる TM の局所フレームに関する ∇ の接続形式が 0 であることをいう。

命題 7.1.2 (アファインチャートの存在). M を多様体、∇ を M のアファイン接続とする。このとき次は同値である:

(1)　M は ∇-平坦である。
(2)　各 $p \in M$ に対し、p のまわりのアファインチャートが存在する。

証明　[TODO] □

平坦性とアファインチャートのアナロジーで双対平坦性と双対アファインチャートを定義する。

定義 7.1.3 (双対平坦). [TODO]

定義 7.1.4 (双対アファインチャート). [TODO]

命題 7.1.5 (双対アファインチャートの存在). [TODO]

証明　[TODO] □

平行移動と測地線は、双対アファイン座標によって非常に簡単な形に書ける。

命題 7.1.6 (平行移動と測地線の双対アファイン座標表示). [TODO]

証明　[TODO] □

双対平坦な双対構造を誘導するダイバージェンスは平坦であるという。

5)　「アファインチャート」という名前は [藤 21a] で使われている「アファイン座標近傍」に由来する。

定義 7.1.7 (ダイバージェンスの平坦性). [TODO]

7.2 凸関数と Legendre 変換

[TODO] geodesically convex とはどう違う？

まず通常の凸関数の定義を復習する。

命題 7.2.1 (凸関数). A を $\mathbb{R}$-ベクトル空間の凸部分集合とする。関数 $f\colon A \to \mathbb{R}$ が**凸 (convex)** であるとは、任意の $x, y \in A$, $x \neq y$ および任意の $t \in (0, 1) \subset \mathbb{R}$ に対して

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \tag{7.2.1}$$

が成り立つことをいう。

注意 7.2.2. 関数の凸性は座標に依存する。この意味を理解するため、多様体 $M = \mathbb{R}_{>0}$ 上の関数 $f\colon \mathbb{R}_{>0} \to \mathbb{R}$ の座標表示について考えてみよう。実は、f がある座標で凸であったとしても別の座標でも凸とは限らない。たとえば $f\colon \mathbb{R}_{>0} \to \mathbb{R}$, $x \mapsto x$ は座標 $\varphi\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $x \mapsto \sqrt{x}$ のもとで $f \circ \varphi(t) = f(t^2) = t^2$ だから凸だが、座標 $\varphi'\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $x \mapsto x$ のもとでは $f \circ \varphi'(t) = f(t) = t$ だから凸ではない。

命題 7.2.3 (凸性はアファイン変換で不変). [TODO]

証明 [TODO] □

定義 7.2.4 (Legendre 変換). [TODO]

7.3 Bregman 幾何

凸関数をポテンシャルとして双対平坦構造が誘導される。

定義 7.3.1 (Bregman ダイバージェンス). [TODO]

7.4 一般化 Pythagoras の定理

定理 7.4.1 (一般化 Pythagoras の定理). [TODO]

証明 [TODO] □

7.5 射影定理

定義 7.5.1 (測地射影). [TODO]

定理 7.5.2 (射影定理)**.** M を多様体、D を M 上のダイバージェンス、(g, ∇, ∇^*) を D により誘導された双対構造、$S \subset M$ を部分多様体とする。このとき次が成り立つ:

(1) 任意の $p \in M$ に対し、$D(p \| \square)$ を最小化する S 上の点 $q \in S$ は p の S への ∇^*-射影である。
(2) 任意の $p \in M$ に対し、$D^*(p \| \square)$ を最小化する S 上の点 $q \in S$ は p の S への ∇-射影である。

証明 [TODO] □

定理 7.5.3 (一意性)**.** S が平坦のとき [TODO]

証明 [TODO] □

第 8 章　無限集合上の測度の空間

8.1　無限集合上の測度の空間

[TODO]

8.2　パラメータ付き測度モデル

定義 8.2.1 (パラメータ付き測度モデル). [TODO]

パラメトリックモデルとは、M が有限次元で $\mathbf{p}$ の値が確率分布の場合である。

定義 8.2.2 (パラメトリックモデル). [TODO]

第 9 章　正定値行列の空間

[TODO] $n \times n$ 行列を n 元集合上の測度とみなす？

[TODO] Stiefel 多様体と最適化についても触れる？

9.1　正定値対称行列の空間

定義 9.1.1 (正定値対称行列の空間)**.** [TODO]

9.2　ダイバージェンス

第 10 章　統計的推定と情報幾何

10.1　漸近理論 (1 次オーダー)

測地射影を用いて漸近理論を幾何学的に考察する。

[TODO]

10.2　漸近理論 (高次オーダー)

高次オーダーの漸近理論には曲率が現れる。

[TODO]

10.3　ノンパラメトリック統計

定義 10.3.1 (Estimating function)**.** [TODO]

演習問題の解答

参考文献

扱う内容は [Ama16] をベースとしたが、同じ内容を同じ配列で並べたわけではない。情報幾何学の内容については [AN07] や [AJLS17] も参考にした。日本語の文献では [甘 19]、[藤 21b] や [藤 21a] を参考にした部分もある。サーベイ論文の [Nie20] も参考にした。関数解析は [BS20]、Banach 多様体は [Lan85]、数理統計は [吉 06] が詳しいと思う。

[AJLS17] Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer, **Information Geometry**, Ergebnisse der Mathematik und ihrer Grenzgebiete 34, vol. 64, Springer International Publishing, Cham, 2017.

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

[AN07] Shun-ichi Amari and Hiroshi Nagaoka, **Methods of Information Geometry**, Translations of Mathematical Monographs, vol. 191, American Mathematical Society, Providence, Rhode Island, April 2007 (en).

[Bog07] Vladimir I. Bogachev, **Measure theory**, Springer, 2007.

[BS20] Vladimir I. Bogachev and Oleg G. Smolyanov, **Real and Functional Analysis**, Moscow Lectures, vol. 4, Springer International Publishing, Cham, 2020 (en).

[CU14] Ovidiu Calin and Constantin Udrite, **Geometric Modeling in Probability and Statistics**, Springer International Publishing, Cham, 2014 (en).

[Egu92] Shinto Eguchi, **Geometry of minimum contrast**, Hiroshima Mathematical Journal **22** (1992), no. 3, 631–647, Publisher: Hiroshima University, Mathematics Program.

[Lan85] Serge Lang, **Differential Manifolds**, Springer US, New York, NY, 1985 (en).

[Lee18] John. M. Lee, **Introduction to riemannian manifolds**, Springer, 2018.

[Nie20] Frank Nielsen, **An elementary introduction to information geometry**, Entropy **22** (2020), no. 10, 1100, arXiv:1808.08271 [cs, math, stat].

[Sch22] Alexander Schmeding, **An Introduction to Infinite-Dimensional Differential Geometry**, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2022.

[吉 06] 明広 吉田, **数理統計学**, 朝倉書店, 2006.

[甘 19] 俊一 甘利, **新板 情報幾何学の新展開**, サイエンス社, 2019.

[藤 21a] 彰夫 藤原, **情報幾何学の基礎**, 共立出版, 2021.

[藤 21b] 敦 藤岡, **入門 情報幾何**, 共立出版, 2021.