

非正曲率象限空間における確率分布の対数凹密度近似について

東京大学・情報理工学系研究科 高澤 祐槻

東京大学・情報理工学系研究科 清 智也

Yuki Takazawa and Tomonari Sei

Graduate School of Information Science and Technology

The University of Tokyo

概要

象限空間は、いくつかのユークリッド空間の非負象限が貼り合わさった空間である。この空間の代表例は、生物の進化の歴史を表す系統樹の形全体のなす空間であり、この空間は非正の曲率を持つ。このような空間上での基本的な問題として、確率分布の推定の問題がある。一方で、空間の複雑さから、パラメトリックな分布族を構成することは容易ではない。対数凹密度はノンパラメトリックな密度のクラスであり、非正曲率を持つ象限空間上で定義可能である。系統樹空間においては一定の条件のもと、対数凹最尤推定量が存在することが知られている。

本研究では、まず対数凹最尤推定量の存在性条件を非正曲率を持つ象限空間の場合に一般化する。さらに、任意の確率分布の対数凹密度による近似である対数凹射影の概念を定義し、この性質を調べる。まず、対数凹射影の存在性についての十分条件を与えた後、その性質について調べることで対数凹最尤推定量の一致性の条件を導く。

1 はじめに

象限空間とは、ユークリッド空間の非負象限のいくつかを、共通の面で張り合わさった空間である。象限空間は自然に完備な測地距離空間となる。Billera et al. (2001) によって構成された系統樹の空間は、実応用上重要な象限空間の代表例である。系統樹空間における統計手法はここ数十年で多く開発されてきている。例えば、算術平均を一般化した量である Fréchet 平均の推定 (Bacák, 2014a), 主成分分析 (Nye, 2011; Nye et al., 2017), カーネル密度推定 (Weyenberg et al., 2014, 2017), クラスタリング (Yoshida et al., 2019), 信頼集合の構成 (Willis, 2019), 対数凹密度推定 (Takazawa and Sei, 2022) などが挙げられる。これらの統計的分析において重要な役割を果たしている系統樹空間の基本的性質は、空間が非正曲率を持つ (CAT(0) である) ということである。

そこで本研究では、非正曲率をもつ象限空間を対象にし、確率分布の推定という基本的な問題について考える。この問題に対するノンパラメトリックなアプローチとしては、先に言及した系統樹空間におけるカーネル密度推定の手法 (Weyenberg et al., 2014, 2017) を CAT(0) 象限空間の場合に一般化することが考えられるが、正規化定数の計算や、バンド幅の選択を含むカーネルの設定などに難点がある。一方で、象限空間のような複雑な空間においてパラメトリックな分布族を仮定することは容

易ではない。例えば, Weyenberg et al. (2014, 2017) におけるカーネル関数として用いられている, $f(x) \propto \exp(-d(x, x_0)^2/h)$ という形の球対称な形の「正規分布型」の分布など, 距離関数 $d(\cdot, \cdot)$ と 1 次元分布を用いれば分布の構成ができるが, このようにして表される分布は球対称なものだけであり, また正規化定数計算の難しさの問題は残ってしまう。また, 幾何的なブラウン運動を考えることでその推移密度として正規分布型の分布を構成する研究 (Nye and White, 2014; Nye, 2020) もあるが, これも「等方的」な分布であり, また推定手法も確立されていない。このような観点から, Takazawa and Sei (2022) により系統樹空間上で開発された対数凹密度最尤推定の手法は, より広いクラスの分布を表すことのできるノンパラメトリック手法である傍ら, 最尤推定量が存在するため, ハイパーパラメータの選択が必要ない自動的な推定過程を踏むことができる点でメリットがある。ただし, 実際には有限点の凸包の計算の難しさから, 3 次元以上の系統樹空間における計算手法は未だ確立されていない。本研究では, この対数凹最尤推定のアプローチに着目し, まず推定量の存在性の結果を CAT(0) 象限空間に拡張する。さらに, 任意の確率測度に対する対数凹密度による近似である対数凹射影についてのいくつかの性質を与える。具体的には, まず対数凹射影の存在性の十分条件の一つを与える。その後, 対数凹射影を与える写像の連続性の性質を調べることにより, 最尤推定量の一致性の条件を導く。Dümbgen et al. (2011) はこれらの性質をユークリッド空間上で与えている。

構成は以下のとおりである。まず 2 節では象限空間と CAT(0) 性について説明する。3 節では, 対数凹密度推定の問題を CAT(0) 象限空間において考え, 推定量の存在性についての結果を与える。4 節では対数凹射影の定義をし, その性質について述べる。最後に 5 節でまとめをする。

2 非正曲率象限空間

2.1 象限空間

まず, 本研究で考える空間である象限空間の定義を, Miller et al. (2015) に沿って与える。 \mathcal{E} を「軸」の集合を表す有限集合とし, $\Omega \subseteq 2^{\mathcal{E}}$ を単体複体とする。ここで, Ω が単体複体であるとは, 任意の $F \in \Omega$ と任意の部分集合 $G \subseteq F$ について, $G \in \Omega$ であることと定義する。また, 全ての 1 点集合 $\{e\}$ ($e \in \mathcal{E}$) は Ω の要素であるとする。今, 各 $F \in \Omega$ に対してユークリッド空間の非負象限 $\bar{O}_F := \mathbb{R}_{\geq 0}^F$ を対応づける。ここで, $\mathbb{R}_{\geq 0}^F$ は, F に入っている「軸」が生成する $|F|$ 次元ユークリッド空間の非負象限であるとする。ここで, 象限 F 内の座標 $x \in \mathbb{R}_{\geq 0}^{|F|}$ の点 (これを (x, F) とする) は, $\mathcal{E} \setminus F$ の各軸に対しては 0 という座標を持つと解釈する。この時, 任意の 2 点 $(x, F), (y, G)$ はその非ゼロの座標が \mathcal{E} 内で全て一致する時, 同一の点としてみなす。すると, ある 2 つの非負象限 \bar{O}_F, \bar{O}_G は, $F \cap G \neq \emptyset$ の時, $\bar{O}_{F \cap G}$ がなす象限によって貼りあわされているとみることができる。象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ とは, このようにして構成された非負象限が貼り合わさった空間である。

例 1. 最も簡単な象限空間は, Ω が 1 点集合と空集合のみからなる場合である。例えば, 各軸を自然数でインデックスづけて $\mathcal{E} = \{1, 2, 3\}$ と定義し, $\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}\}$ とすると, 対応する象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ は図 1 (左) のような空間となる。3 つの半直線が原点で繋がった形をしていることから, この空間は 3-spider と呼ばれ, 一般に k 本の半直線が繋がっている場合は k -spider と呼ばれる。あとに述べるように, 3-spider の空間は 3 種の系統樹の空間となっている。

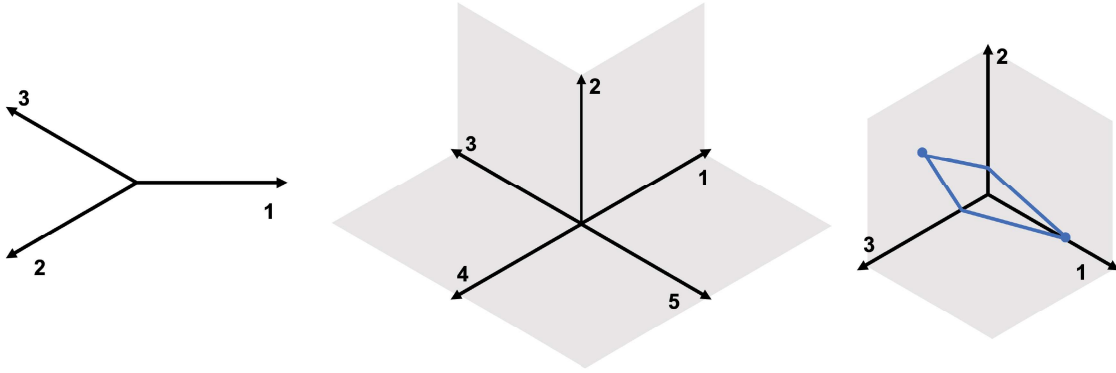


図1 象限空間の例. (左): 3-spider (例1). (中): 例2の2次元象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$. (右): 例2の2次元象限空間 $\mathcal{O}(\mathcal{E}', \Omega')$. 図上の2点を結ぶ測地線は2つ存在する.

例2. 軸の集合 $\mathcal{E} := \{1, 2, 3, 4, 5\}$ とする. ここで Ω を

$$\Omega := \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 1\}\} \quad (2.1)$$

と定義する. 対応する象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ を図示すると図1 (中) のような2次元の空間となる.

同様に, $\mathcal{E}' = \{1, 2, 3\}$,

$$\Omega' = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{3, 1\}\} \quad (2.2)$$

に対する象限空間 $\mathcal{O}'(\mathcal{E}', \Omega')$ を図示すると, 図1 (右) のような空間となる.

象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ に対して, $\max_{F \in \Omega} |F|$ をその次元と呼ぶこととする. また, 全ての座標が0である点を原点と呼ぶことにする. 例1の3-spiderは1次元の空間であり, 例2の2つの象限空間はどちらも2次元の象限空間である. また, p 次元の象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ に対して, $\Omega_p = \{F \in \Omega \mid |F| = p\}$ と定める.

応用で重要な象限空間の例としては, Billera et al. (2001) によって構成された系統樹がなす空間がある.

例3 (系統樹の空間). $n+1$ 個のラベル付きの葉を持つ木を n -tree と呼ぶ. n -tree は, 共通の祖先を表す一つの「根」と, n 個の現存する種を表す葉を持つ系統樹と考えることができる. 系統樹は, そのトポロジーと枝の長さで決定されるが, Billera et al. (2001) による系統樹空間の構成においては内部枝の長さのみを考慮する. ここで, 内部枝とは葉に直接接続されていない枝のことである. 内部枝は, $n+1$ 個の葉をどのように分割するかで特徴づけることができる. このことから, 内部枝の種類は全部で $\binom{n+1}{2}/2$ 個ある. この内部枝全体の集合が象限空間における軸の集合 ($\mathcal{E}^{(n)}$ とする) に対応する. ここで, 内部枝の集合 F が同じ木に共存できる時, $F \in \Omega^{(n)}$ であると定める. すると明らかに $\Omega^{(n)}$ は単体複体の構造をなす. このようにして作られた象限空間 $\mathcal{T}_n = \mathcal{O}(\mathcal{E}^{(n)}, \Omega^{(n)})$ が系統樹空間である.

ここで, 二分 n -tree の内部枝は $n-2$ 本であり, これが n -tree に共存できる枝の最大本数である. すなわち, この象限空間 \mathcal{T}_n の次元は $n-2$ である. また, 二分木の異なるトポロジーの数は $(2n-3)!!$ 個である (Felsenstein, 1978) ことから, この数だけの $n-2$ 次元非負象限が貼り合わさっている. つまり, $|\Omega_p| = (2n-3)!!$ である.

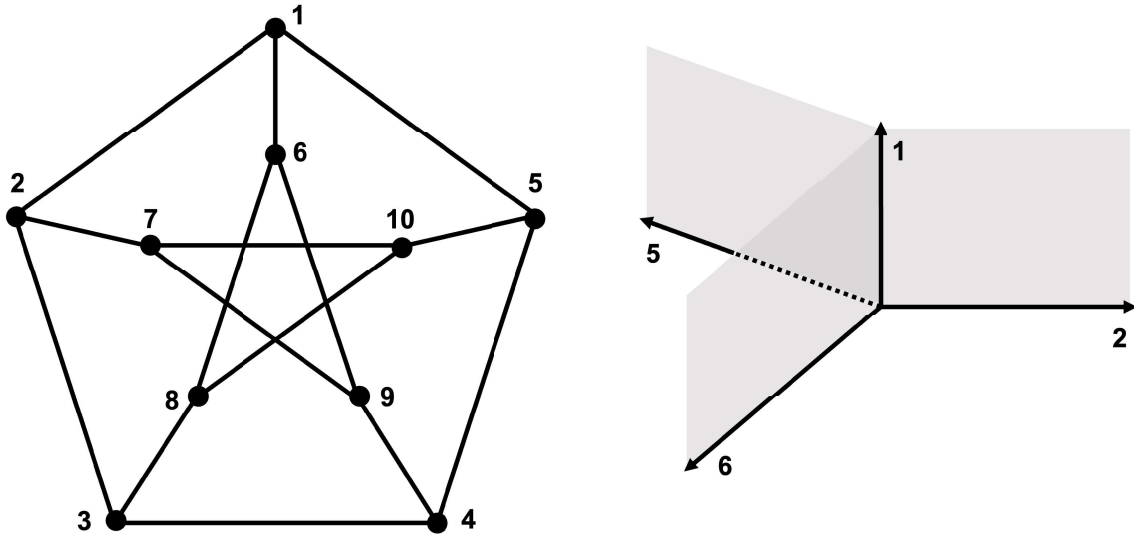


図2 (左): 系統樹空間 $\mathcal{T}_4 = \mathcal{O}(\mathcal{E}^{(4)}, \Omega^{(4)})$ に対する $\Omega^{(4)}$. $1, 2, \dots, 10$ とラベルが付けられているグラフの各頂点が Ω の 1 点集合に対応し, 線で繋がった 2 つの頂点の集合全体が Ω に含まれる 2 点集合に対応する. (右): \mathcal{T}_4 の一部.

$n = 3$ の場合は, 最大次元が 1 次元で二分木のトポロジーの数が 3 個であることから, 系統樹空間 \mathcal{T}_3 は例 1 で与えた 3-spider の空間と一致する. $n = 4$ の場合は, 空間は 2 次元で二分木のトポロジーの数が 15 個であることから, $\mathcal{T}_4 = \mathcal{O}(\mathcal{E}^{(4)}, \Omega^{(4)})$ は 15 個の 2 次元非負象限が貼り合わさった空間となる. $\Omega^{(4)}$ は図 2 (左) に示す通りとなる. \mathcal{T}_4 全体は図に書くことのできない複雑な繋がり方をしているが, $\Omega^{(4)}$ の形からわかるように局所的には 3 つの非負象限が共通の軸でつながった形をしている (図 2 右).

2.2 非正曲率の空間: CAT(0) 空間

本小節では, 非正曲率の空間である CAT(0) 空間の定義をする.

(\mathcal{X}, d) を任意の距離空間とする. 任意の \mathcal{X} の 2 点 x, y が, その間の距離の長さを持つようなパスによって繋がれる時, そのようなパスを測地線と呼び, (\mathcal{X}, d) は測地距離空間であるという. より正確には, 2 点 $x, y \in \mathcal{X}$ を結ぶ測地線とは連続写像 $\gamma_{x,y} : [0, 1] \rightarrow \mathcal{X}$ で, $\gamma_{x,y}(0) = x, \gamma_{x,y}(1) = y$ であり, 任意の $s, t \in [0, 1]$ に対して, $d(\gamma_{x,y}(s), \gamma_{x,y}(t)) = |s - t|d(x, y)$ となるものである.

測地距離空間が CAT(0) であることの定義は同値なものはいくつか考えられるが, ここでは次のように比較三角形を用いた定義を採用する.

定義 4. 任意の 3 点 $a, b, c \in \mathcal{X}$ に対し, その比較三角形 $\Delta(a', b', c') \in \mathbb{R}^2$ を, $\|a' - b'\| = d(a, b), \|b' - c'\| = d(b, c), \|c' - a'\| = d(c, a)$ を満たすような 2 次元ユークリッド空間上の点 $a', b', c' \in \mathbb{R}^2$ からなる三角形とする. 比較三角形は等長写像を除いて一意である.

この時, 任意の $s, t \in [0, 1]$ に対して $x = \gamma_{a,b}(s), y = \gamma_{a,c}(t)$ と定めた時,

$$d(x, y) \leq \|x' - y'\| \quad (2.3)$$

が成り立てば、測地距離空間は CAT(0) であるという。ここで、 $x' = (1-s)a + sb, y' = (1-t)a + tc$ は比較三角形上の x, y にそれぞれ対応する点である。

特に、完備な CAT(0) 空間を Hadamard 空間と呼ぶ。

CAT(0) 性の条件は、直感的には任意の三角形が \mathbb{R}^2 における三角形よりも「痩せている」という解釈ができる。

例 5. 例 1 で考えた 3-spider の空間を考える。考える 3 点が最大 2 つの象限に集中しているときは、 \mathbb{R} 上の 3 点と同一視できるため明らかに式 (2.3) が成立する。また、3 点が 3 つの象限に分散している場合も、式 (2.3) が成り立つことは明らかである。

Hadamard 空間上では、様々な性質が成り立つことが知られているが、最も基本的なのは、測地線の一意存在性である。Owen and Provan (2011) は系統樹空間の場合に多項式時間で測地線を求めるアルゴリズムを開発した。これを元にして、CAT(0) 象限空間 (Miller et al., 2015), 更には一般の CAT(0) 立方複体 (Hayashi, 2021) へのアルゴリズムの拡張が行われている。

さらに Hadamard 空間上では、一意な測地線を用いて凸性を定義することができる。具体的には、Hadamard 空間 (\mathcal{H}, d) 上の集合 $S \subseteq \mathcal{H}$ が凸であることを、 S 内の任意の 2 点を結ぶ測地線がまた S に含まれることとして定義することができる。また、関数 $f: \mathcal{H} \rightarrow [-\infty, \infty]$ の凸性はそのエピグラフ $\{(x, \mu) \in \mathcal{H} \times \mathbb{R} \mid \mu \geq f(x)\}$ が凸であることで定義される。ユークリッド空間の場合同様に、関数 f が凹であることは $-f$ が凸であることである。Hadamard 空間上の凸解析に関する事実のいくつかは Bacák (2014b) にまとめられている。1 つの重要な事実は、距離関数の同時凸性である。

補題 6 (Bacák 2014b). (\mathcal{H}, d) を Hadamard 空間として、 $\gamma, \gamma': [0, 1] \rightarrow \mathcal{H}$ を 2 つの任意の測地線とする。この時、次の同時凸性が成り立つ：

$$d(\gamma(t), \gamma'(t)) \leq (1-t)d(\gamma(0), \gamma'(0)) + td(\gamma(1), \gamma'(1)). \quad (2.4)$$

特に、ある点 $y \in \mathcal{H}$ に対して、 $\gamma'(t) = y \ (\forall t \in [0, 1])$ とすれば、1 つ目の引数 (あるいは 2 つ目の引数) に対する凸性も導かれる。

統計学の観点からの 1 つの重要な帰結は、以下で定義される Fréchet 平均の一意存在性である。以下、 $\mathcal{P}^1(\mathcal{H})$ を、Hadamard 空間 \mathcal{H} 上の可分なボレル確率測度であって、ある点 $y \in \mathcal{H}$ に対して、 $\int d(x, y)dP < \infty$ であるもののクラスとする。任意の $z \in \mathcal{H}$ を固定する。この時、確率測度 $P \in \mathcal{P}^1(\mathcal{H})$ に対する Fréchet 関数 \mathbb{F}_P を次のように定義する：

$$\mathbb{F}_P(x) = \int (d(x, y)^2 - d(x, z)^2)dP(y). \quad (2.5)$$

Fréchet 関数 \mathbb{F}_P の最小解が存在すればそれを Fréchet 平均と呼ぶ。これが z の選び方に依存しないのは明らかである。Sturm (2003) は以下を示した。

定理 7 (Sturm 2003). (\mathcal{H}, d) を Hadamard 空間とする。この時、任意の確率測度 $P \in \mathcal{P}^1(\mathcal{H})$ に対して、Fréchet 平均は一意に存在する。

特に、サンプル X_1, \dots, X_n に対する経験測度 $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ は $\mathcal{P}^1(\mathcal{H})$ の要素である。ここで、 δ_{X_i} は点 X_i におけるディラック測度を表している。従って、与えられた Hadamard 空間上の

サンプルに対しては Fréchet 平均を求めて代表点とすることが考えられる。一方で、いくつかの空間においては、一定の条件のもと、サンプルサイズが十分に大きければ、Fréchet 平均が低次元空間に張り付く (*stickiness*) という問題点が知られている (Hotz et al., 2013; Huckemann et al., 2015; Barden et al., 2018).

2.3 象限空間の CAT(0) 性

象限空間は完備測地距離空間であるが CAT(0) とは限らない。一方で、CAT(0) 性の特徴づけは簡単な形で与えられている。ここで、 Ω が flag であるということは、次のように定義される：任意の $F \in 2^{\mathcal{E}}$ がその任意の 2 要素 $e, f \in F$ に対して $\{e, f\} \in \Omega$ という条件を満たす時（すなわち、 F の任意の 2 点部分集合が Ω の要素である時）、 $F \in \Omega$ となる。

定理 8 (Gromov 1987; Miller et al. 2015). 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ が CAT(0) であることの必要十分条件は、 Ω が flag であることである。

例 9. 例 2 で定義した 2 つの象限空間のうち $\mathcal{O}(\mathcal{E}, \Omega)$ は CAT(0) であるが、 $\mathcal{O}(\mathcal{E}', \Omega')$ は CAT(0) でない (図 1). Ω が flag であるのに対し、 Ω' が flag でないことは定義から簡単にわかる。また、図 1 (右) からわかるように、 $\mathcal{O}(\mathcal{E}', \Omega')$ における測地線が一意に定まらないことから後者が CAT(0) でないことが導かれる。

例 10. 例 3 で与えた系統樹空間は CAT(0) である。これは、系統樹空間の内部枝の集合 S がペアワイズに共存可能であるとき、 S 全体も共存可能であるということからわかる (Semple et al., 2003).

2.4 CAT(0) 象限空間上の基準測度

本稿では確率密度関数を考えるため、その基準測度が必要となる。一方、CAT(0) 象限空間はユークリッド空間（の非負象限）の貼り合わせでできているため、通常のルベーク測度を用いて自然な測度を定義することができる。

具体的には、 p 次元の CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 内のボレル集合 S の測度 ν を次のように定める：

$$\nu(S) = \sum_{F \in \Omega_p} \lambda_p(S \cap \bar{O}_F). \quad (2.6)$$

ここで、 λ_p は \mathbb{R}^p 上のルベーク測度である。 ν が測度になることは明らかである。この測度の完備化をしたものを改めて ν とおき、これを基準測度とする。このような測度は、系統樹空間におけるいくつかの研究 (Willis, 2019; Takazawa and Sei, 2022) によって採用されているものである。

3 CAT(0) 象限空間上の対数凹密度推定

本節では、CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ における確率密度推定の手法として対数凹最尤推定を考え、その推定可能性について議論する。

3.1 対数凹関数

CAT(0) 象限空間 (\mathcal{X}, d) 上の関数 $f: \mathcal{X} \rightarrow [-\infty, \infty]$ が対数凹であるとは、 $\log f$ が凹関数となることである。特に対数凹関数 f が基準測度 ν に対する確率密度である時、すなわち $\int_{\mathcal{X}} f d\nu = 1$ である時、 f は対数凹密度であると呼ぶ。

ユークリッド空間における対数凹密度のクラスは、多変量正規分布、形状母数が 1 以上のガンマ分布、2つの形状母数がどちらも 1 以上のベータ分布など、実用的な分布を多く含むものとなっている。このクラスはノンパラメトリックなものである一方、Cule et al. (2010) は多次元ユークリッド空間 \mathbb{R}^p において、サンプルサイズ n が $p+1$ 以上の時、最尤推定量が確率 1 で存在することを示した。また、尤度の最大化はサンプルサイズ次元の凸最適化問題に帰着できることも示されている。

本研究では、CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ における以下のような対数凹関数のクラスを考える。まず、 Φ を $\mathcal{O}(\mathcal{E}, \Omega)$ 上の凹関数であり、上半連続で強圧的なものの集合とする。ただし、ここで関数 ϕ が強圧的であるとは、ある（任意の） $x_0 \in \mathcal{O}(\mathcal{E}, \Omega)$ に対して $d(x, x_0) \rightarrow \infty$ の極限で $\phi(x) \rightarrow -\infty$ となることと定める。この時、 $\exp(\Phi) = \{\exp(\phi) \mid \phi \in \Phi\}$ を対数凹関数のクラスとして考える。これらの制約はこれからの議論において制約的ではない。また、特に $\Phi_0 \subseteq \Phi$ を対数密度であるような Φ の要素の集合とし、 $\exp(\Phi_0)$ を対数凹密度の空間として考える。

測地線の凸性（補題 6）から、対数凹関数の典型的な例が導ける。

命題 11. $g: [0, \infty) \rightarrow [-\infty, \infty)$ を単調非増加な凹関数とする。この時、任意の $x_0 \in \mathcal{O}(\mathcal{E}, \Omega)$ に対して、関数 $\phi(x) = g(d(x, x_0))$ は $\mathcal{O}(\mathcal{E}, \Omega)$ から $[-\infty, \infty)$ への凹関数である。

3.2 CAT(0) 象限空間における対数凹最尤推定

Takazawa and Sei (2022) は系統樹空間における対数凹最尤推定量の存在性の十分条件を与えている。この十分条件はサンプルサイズが無限大に近づくにつれて 1 に近づいていく確率で満たされる。この結果の CAT(0) 象限空間への一般化は容易であり、以下の定理が成り立つ。

定理 12. (X_1, \dots, X_n) を CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上の密度 f からの独立サンプルとする。 C_n を $\{X_1, \dots, X_n\}$ の凸包とする。さらに、以下を仮定する：

(a) 任意の $x \in \text{cl}(C_n)$ に対し、ある $F \in \Omega_p$ が存在して、 $x \in \bar{O}_F$ であり、 $\text{cl}(C_n) \cap \bar{O}_F$ は p 次元の集合である。

この時、 $\exp(\Phi_0)$ 内の最尤推定量 \hat{f}_n が存在し、 ν に関してほとんど至る所で一意である。

定理の条件は、やはり $n \rightarrow \infty$ の極限で 1 に近づく確率で満たされる。

今、ユークリッド空間の場合同様、最尤推定量の対数は次の汎関数 $L_n(\phi)$ の凹関数の空間 Φ における最大解となる：

$$L_n(\phi) = \frac{1}{n} \sum_{i=1}^n \phi(X_i) - \int_{\mathcal{O}(\mathcal{E}, \Omega)} \exp(\phi) d\nu + 1. \quad (3.1)$$

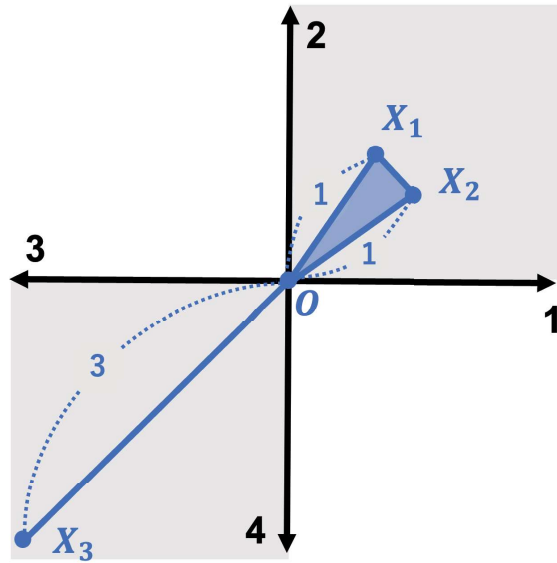


図3 最尤推定量が存在しない例.

Φ の要素は対数密度とは限らないことに注意されたい. すなわち, L_n は任意の Φ の要素を引数としてとるが, 最大解が存在すればそれは Φ_0 の要素に自動的になるということである. このことは4節において補題14によって確認される.

定理の条件が満たされない例としては, 以下のような状況が考えられる.

例13 (Takazawa and Sei (2022)). $\mathcal{E} = \{1, 2, 3, 4\}$, $\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{3, 4\}\}$ に対する象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ を考える. これは CAT(0) であり, 図3のような2つの2次元の非負象限が原点でのみつながった空間となる. 今, $X_1 \neq X_2$ を原点から距離1離れた象限 $O_{\{1,2\}}$ 上の2点, X_3 を原点から距離3離れた象限 $O_{\{3,4\}}$ 上の点とする. また, $C = \text{conv}(X_1, X_2, X_3)$ とする.

このとき, $y \in \mathbb{R}$ を固定し, C 上に有効領域を持つ凹関数 $\psi_\Delta \in \Phi$ を

$$\psi_\Delta(X_3) = y + \Delta, \quad \psi_\Delta(X_1) = \psi_\Delta(X_2) = y - \Delta/3, \quad \psi_\Delta(O) = y.$$

を満たすようにとる. すなわち, ψ_Δ は C 内の任意の測地線上で線形な関数である. すると,

$$L_n(\psi_\Delta) = y + \Delta/9 - \int \exp(\psi_\Delta) d\nu + 1.$$

これは Δ に関して単調増加になる. ここで, 線分 OX_3 での関数値は L_n の第2項に直接影響を与えないことに注意されたい. さらに, $\Delta \rightarrow \infty$ の極限で L_n の値は ∞ へと近づく. したがって, この汎関数 L_n の最大解は存在しない. つまり, 最尤推定量は存在しない. 実際, この例では, $C \cap \bar{O}_{\{3,4\}}$ が1次元の集合であるため, 定理12の条件を満たしていないことが確認できる.

4 対数凹射影とその性質

対数凹射影とは, 確率分布の対数凹密度関数の空間への射影である. ユークリッド空間の場合には, その存在性や性質が Dümbgen et al. (2011) によって導かれている. 本節では, 同様の概念を

CAT(0) 象限空間上で考え、まずその存在性の十分条件を与える。また、確率測度から対数凹射影を与える写像の連続性について考察することで、最尤推定量の一致性について考察する。

4.1 対数凹射影の定義

$\phi \in \Phi$ と任意の CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上のボレル確率測度 Q について、以下のような汎函数 $L(\phi, Q)$ を考える：

$$L(\phi, Q) = \int \phi dQ - \int_{\mathcal{O}(\mathcal{E}, \Omega)} \exp(\phi) d\nu + 1. \quad (4.1)$$

この汎函数は、最尤推定量の特徴づけをする式 (3.1) における汎函数 L_n の一般化になっていることに注意されたい。すなわち、経験測度を \mathbb{P}_n とおけば、 $L_n(\phi) = L(\phi, \mathbb{P}_n)$ である。

$L(Q) = \sup_{\phi \in \Phi} L(\phi, Q)$ とする。 $L(Q)$ を達成するような凹関数 $\psi \in \Phi$ が存在する時、 $\exp(\psi)$ を対数凹射影と呼ぶ。したがって、経験測度に対する対数凹射影が最尤推定量である。次の補題は、 $L(\phi, Q)$ の最大解が存在すれば必ず対数密度になることを示している。すなわち、対数凹射影は必ず密度関数となる。

補題 14. CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上の確率測度 Q を固定する。ここで、 $L(Q) \in \mathbb{R}$ であり、ある $\psi \in \Phi$ が $L(Q)$ を達成すると仮定する：

$$L(\psi, Q) = L(Q) \equiv \sup_{\phi \in \Phi} L(\phi, Q).$$

この時、 ψ は密度の対数である：

$$\int \exp(\psi) d\nu = 1.$$

Proof. 任意の $\phi \in \Phi$ に対して、 $\int \exp(\phi) d\nu < \infty$ である。また、 $\int \exp(\phi) d\nu > 0$ と仮定して良い。なぜなら、この値が 0 であるときは $L(Q)$ は有限の値を取りえないからである。 $0 < c = \int \exp(\phi) d\nu$ とする。この時、 $\phi - \log c \in \Phi$ 、 $\int \exp(\phi - \log c) d\nu = 1$ であり、

$$\begin{aligned} L(\phi - \log c, Q) &= \int (\phi - \log c) dQ - \int \exp(\phi - \log c) d\nu + 1 \\ &= \int \phi dQ - \log c \\ &= L(\phi, Q) + c - 1 - \log c \\ &\leq L(\phi, Q). \end{aligned}$$

等号成立の必要十分条件は $c = 1$ であることである。 □

4.2 対数凹射影の一意存在性

確率測度 Q に対して、その対数凹射影が存在する時、これを $\exp(\psi_Q)$ と書くことにする。また、確率測度 Q の凸サポート $\text{csupp}(Q)$ を次のように定義する：

$$\text{csupp}(Q) = \bigcap \{C : C \subseteq \mathcal{O}(\mathcal{E}, \Omega); C \text{ は閉で凸, } Q(C) = 1\}. \quad (4.2)$$

$\text{csupp}(Q)$ は明らかに閉凸集合であり、また、空間の Lindelöf 性からこれもまた Q -確率 1 の集合となる。つまり、凸サポートは Q -確率 1 を持つ最小の閉凸集合である。

対数凹射影が存在する時、その一意性は ν に関してほとんど至る所で成り立つ。

定理 15. $CAT(0)$ 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上の任意の確率測度 Q の対数凹射影 $\exp(\psi_Q)$ は ν に関してほとんど至る所で一意である。

Proof. 2 つの凹関数 $\psi_1, \psi_2 \in \Phi$ が $L(\cdot, Q)$ を最大化するとする。この時、補題 14 より、 $\int \exp(\psi_1) d\nu = \int \exp(\psi_2) d\nu = 1$ である。したがって、 $\psi = \frac{1}{2}\psi_1 + \frac{1}{2}\psi_2 - \log \int \exp((\psi_1 + \psi_2)/2) d\nu$ とすれば、

$$\begin{aligned} L(\psi, Q) &= \frac{1}{2}(L(\psi_1, Q) + L(\psi_2, Q)) - \log \int \exp((\psi_1 + \psi_2)/2) d\nu \\ &\geq L(Q) - \log \int \frac{\exp(\psi_1) + \exp(\psi_2)}{2} d\nu \\ &= L(Q). \end{aligned}$$

ここで、等号が成立する必要十分条件は、 $\exp(\psi_1) = \exp(\psi_2)$ が ν に関してほとんど至る所で成り立つことである。□

ここで、任意の対数凹射影 (の対数) ψ に対し、その $\text{csupp}(Q)$ への制限 $\psi|_{\text{csupp}(Q)}$ を考えると、 $L(\psi|_{\text{csupp}(Q)}, Q)$ は $L(\psi, Q)$ 以上の値をとることが簡単に確認できる。したがって、 Q の対数凹射影が存在する場合はそのサポートが $\text{csupp}(Q)$ に含まれるようなものを取ってくるができる。

ここからは Q が ν に関して密度 f_Q を持つ場合を考える。次の定理は、対数凹射影の存在性の十分条件を与える。

定理 16. $CAT(0)$ 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上の確率測度 Q が以下の条件を満たすとする：

- Q は ν に関して密度を持ち、
- ある (任意の) $x_0 \in \mathcal{O}(\mathcal{E}, \Omega)$ に対し、 $\int d(x, x_0) dQ < \infty$ となる。

この時、 $L(Q) \in \mathbb{R}$ である。これが成り立つ時、

$$\psi \in \arg \max_{\phi \in \Phi} L(\phi, Q)$$

が存在し、 ν に関してほとんど至る所で一意である。さらに、

$$\int \exp(\psi) d\nu = 1$$

であり、 ψ のあるバージョンは以下を満たす：

$$\text{int}(\text{csupp}(Q)) \subseteq \text{dom}(\psi) \subseteq \text{csupp}(Q).$$

証明は省略する。ここで、経験測度に対する対数凹射影は最尤推定量に対応することから、特に定理 12 は対数凹射影の存在性のための定理 16 とは別の十分条件を与えているということに注意されたい。

最後に、ユークリッド空間における結果との比較をする。Dümbgen et al. (2011) によると、ユークリッド空間における対数凹射影の存在性は、次の 2 条件で特徴付けられる：

$$\int_{\mathbb{R}^p} \|x\| dQ < \infty \quad \text{and} \quad \text{int}(\text{csupp}(Q)) \neq \emptyset. \quad (4.3)$$

したがって、CAT(0) 象限空間において対応する条件としては次のようなものが自然と考えられる： $x_0 \in \mathcal{O}(\mathcal{E}, \Omega)$ を固定して、

$$\int_{\mathcal{O}(\mathcal{E}, \Omega)} d(x, x_0) dQ < \infty \quad \text{and} \quad \text{int}(\text{csupp}(Q)) \neq \emptyset. \quad (4.4)$$

しかし、最尤推定量が存在しない例 13 は条件 (4.4) を満たしており、すなわち条件 (4.4) だけでは CAT(0) 象限空間の場合には対数凹射影の存在性を導くことができないことがわかる。

4.3 対数凹射影を与える写像の連続性と最尤推定量の一致性

本節では、確率測度 Q に対してその対数凹射影を与える写像 $Q \mapsto \exp(\psi_Q)$ の連続性についての結果を与えることで、最尤推定の一致性の結果を導く。

定理 17. CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上の確率測度の列 $(Q_n)_n$ と確率測度 Q が次の条件を満たすと仮定する：

- *Mallows-1* 距離で $(Q_n)_n$ は Q に収束する。
- ある x_0 に対して、 $\int d(x, x_0) dQ_n < \infty$ ($\forall n$), $\int d(x, x_0) dQ < \infty$.
- Q は ν に関して密度を持つ。
- ある $N \in \mathbb{N}$ が存在して、 $n \geq N$ ならば $L(\phi, Q_n)$ の最大解が存在する。

この時、 $\lambda := \lim_{n \rightarrow \infty} L(Q_n) = L(Q)$ となる。

さらに、 $\lim_{n \rightarrow \infty} \int |\exp(\psi_{Q_n}) - \exp(\psi_Q)| d\nu = 0$ である。

注意. *Mallows-1* 距離での収束は、次の 2 条件の組み合わせと等価である：

- $(Q_n)_n$ は Q に弱収束する。
- $\exists x_0, \int d(x, x_0) dQ_n \rightarrow \int d(x, x_0) dQ$.

特に、二つ目の条件は任意の x_0 について成立する。

定理の証明は省略する。この定理の系として、対数凹最尤推定量の一致性の十分条件を導くことができる。

系 18. $\{X_n\}_n$ を、ある CAT(0) 象限空間 $\mathcal{O}(\mathcal{E}, \Omega)$ 上の密度を持つ確率測度 P からの独立サンプルとし、 P は $\int d(x, x_0) dP < \infty$ を満たすとする。また、 $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ とする。

この時、任意の $\epsilon > 0$ に対し、ある $N \in \mathbb{N}$ が存在して、 $n \geq N$ であれば、 $1 - \epsilon$ よりも大きい確率で最尤推定量 $\exp(\psi_{\mathbb{P}_n})$ が存在する。さらに、確率 1 で次が成り立つ：

$$\lim_{n \rightarrow \infty} \int |\exp(\psi_{\mathbb{P}_n})(x) - \exp(\psi_P)(x)| d\nu = 0.$$

すなわち、最尤推定量は全変動距離で真の確率測度の対数凹射影へ収束する。

証明の概要. 系の1つ目の主張は定理12の主張である. Varadarajanの定理から, 確率1で経験測度 \mathbb{P}_n は測度 P へ弱収束する. また, 大数の強法則から, 任意の x_0 に対して確率1で $\int d(x, x_0)d\mathbb{P}_n \rightarrow \int d(x, x_0)dP$ が成り立つ. すなわち, \mathbb{P}_n は P へMallows-1距離で収束する. あとは定理17の結果を使う. \square

5 まとめ

本稿では, CAT(0)象限空間における確率分布の推定法として, 対数凹密度を用いた最尤法を考え, 推定量の存在性の十分条件を与えた. また, 確率分布の対数凹密度による近似である対数凹射影の概念を定義し, その存在性の十分条件を与えた. さらに, 確率測度に対しその対数凹射影を与える写像の連続性の結果を与えることにより, 最尤推定量の一致性の十分条件を導いた.

今後の研究課題として, 現在最尤推定の手法が2次元系統樹空間の場合までしか与えられていないため, 高次元の象限空間においても用いることのできる計算手法の開発が求められる. 最尤推定量はデータ点の凸包の閉包における区分的に線形な関数の形で与えられるため, このためには有限点の凸包の閉包(またはその近似)を求めるアルゴリズムの開発が必要である. また, 例えば Xu and Samworth (2021)は対数凹性に加え対称性に関する形状制約を導入している. このように, 必要な範囲で形状制約を追加したり, 別の形状制約をつけた推定方法についても開発することが考えられる. 特に, Nye and White (2014); Nye (2020)で提案されている, ブラウン運動の推移密度を用いて定義される立方複体上の確率分布や, 簡単な合祖過程から導かれる系統樹空間上の確率密度は境界で「折れ曲がる」挙動を示すことが指摘されている (Takazawa and Sei, 2022). このような状況に対応できるよう, 境界条件を追加した推定手法の研究も必要であると考えられる.

参考文献

- Bacák, M. (2014a), ‘Computing medians and means in Hadamard spaces’, *SIAM journal on optimization* **24**(3), 1542–1566.
- Bacák, M. (2014b), Convex analysis and optimization in Hadamard spaces, in ‘Convex Analysis and Optimization in Hadamard Spaces’, de Gruyter.
- Barden, D., Le, H. and Owen, M. (2018), ‘Limiting behaviour of Fréchet means in the space of phylogenetic trees’, *Annals of the Institute of Statistical Mathematics* **70**, 99–129.
- Billera, L. J., Holmes, S. P. and Vogtmann, K. (2001), ‘Geometry of the space of phylogenetic trees’, *Advances in Applied Mathematics* **27**, 733–767.
- Cule, M., Samworth, R. and Stewart, M. (2010), ‘Maximum likelihood estimation of a multi-dimensional log-concave density’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 545–607.
- Dümbgen, L., Dominic, R. and Samworth, S. (2011), ‘Approximation by log-concave distributions, with applications to regression’, *Annals of Statistics* **39**, 702–730.

- Felsenstein, J. (1978), ‘The number of evolutionary trees’, *Systematic Zoology* **27**, 27–33.
- Gromov, M. (1987), *Hyperbolic groups*, Springer.
- Hayashi, K. (2021), ‘A polynomial time algorithm to compute geodesics in CAT (0) cubical complexes’, *Discrete & Computational Geometry* **65**, 636–654.
- Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. and Skwerer, S. (2013), ‘Sticky central limit theorems on open books’, *Annals of Applied Probability* **23**, 2238–2258.
- Huckemann, S., Mattingly, J., Miller, E. and Nolen, J. (2015), ‘Sticky central limit theorems at isolated hyperbolic planar singularities’, *Electronic Journal of Probability* **20**(none), 1 – 34.
- Miller, E., Owen, M. and Provan, J. S. (2015), ‘Polyhedral computational geometry for averaging metric phylogenetic trees’, *Advances in Applied Mathematics* **68**, 51–91.
- Nye, T. M. (2011), ‘Principal components analysis in the space of phylogenetic trees’, *Annals of Statistics* **39**, 2716–2739.
- Nye, T. M. (2020), ‘Random walks and brownian motion on cubical complexes’, *Stochastic Processes and their Applications* **130**, 2185–2199.
- Nye, T. M., Tang, X., Weyenberg, G. and Yoshida, R. (2017), ‘Principal component analysis and the locus of the Fréchet mean in the space of phylogenetic trees’, *Biometrika* **104**(4), 901–922.
- Nye, T. M. and White, M. (2014), ‘Diffusion on some simple stratified spaces’, *Journal of mathematical imaging and vision* **50**, 115–125.
- Owen, M. and Provan, J. S. (2011), ‘A fast algorithm for computing geodesic distances in tree space’, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **8**, 2–13.
- Semple, C., Steel, M. et al. (2003), *Phylogenetics*, Vol. 24, Oxford University Press on Demand.
- Sturm, K.-T. (2003), Probability measures on metric spaces of nonpositive curvature, *in* ‘Heat kernels and analysis on manifolds, graphs, and metric spaces: lecture notes from a quarter program on heat kernels, random walks, and analysis on manifolds and graphs’, American Mathematical Society, pp. 357–390.
- Takazawa, Y. and Sei, T. (2022), ‘Maximum likelihood estimation of log-concave densities on tree space’, *arXiv preprint arXiv:2211.12037*.
- Weyenberg, G., Huggins, P. M., Schardl, C. L., Howe, D. K. and Yoshida, R. (2014), ‘KDE TREES: Non-parametric estimation of phylogenetic tree distributions’, *Bioinformatics* **30**, 2280–2287.
- Weyenberg, G., Yoshida, R. and Howe, D. (2017), ‘Normalizing kernels in the Billera-Holmes-Vogtmann treespace’, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **14**, 1359–1365.
- Willis, A. (2019), ‘Confidence sets for phylogenetic trees’, *Journal of the American Statistical Association* **114**, 235–244.
- Xu, M. and Samworth, R. J. (2021), ‘High-dimensional nonparametric density estimation via symmetry and shape constraints’, *The Annals of Statistics* **49**, 650–672.

Yoshida, R., Fukumizu, K. and Vogiatzis, C. (2019), 'Multilocus phylogenetic analysis with gene tree clustering', *Annals of Operations Research* **276**, 293–313.