

情報幾何学講義ノート
— 統計的モデルをひもとく微分幾何 —

著者名

2026 年 2 月 12 日

目次

第 1 章	微分幾何の基礎知識	5
1.1	数学的準備：多様体論への橋渡し	5
1.2	微分可能多様体	6
1.3	接ベクトルと接空間	7
1.4	余接空間と 1 形式	10
1.5	ベクトル場とテンソル場	10
1.6	Riemann 計量	12
	【コラム】「接続」とは翻訳機である	15
1.7	アファイン接続と共変微分	18
	【コラム】アファイン接続と共変微分の直感	18
1.8	平坦性	20
1.9	自己平行性と部分多様体	21
1.10	Riemann 接続 (Levi-Civita 接続)	22
第 2 章	統計的モデルの幾何学的構造	25
2.1	統計的モデル	25
2.2	Fisher 情報行列の導出	25
2.3	α -接続の導出	27
第 3 章	双対接続の理論	35
3.1	接続の双対性	35
3.2	双対平坦空間	38
3.3	ダイバージェンス (正準ダイバージェンス)	38
3.4	一般化されたピタゴラスの定理	39
3.5	双対葉層構造	39
第 4 章	統計的推論の微分幾何	41
4.1	指数型分布族の幾何構造	41
4.2	最尤推定の幾何学的解釈 (m-射影)	42
4.3	曲指数型分布族と埋め込み曲率	42

4.4	推定の高次漸近理論	43
4.5	統計物理学と大偏差理論	43
4.6	統計的仮説検定の幾何	44
第 5 章	推定関数の理論とファイバーバンドル	47
5.1	推定関数と幾何学的要請	47
5.2	局所指数族バンドル	47
5.3	Hilbert バンドルと接続	48
5.4	最適推定関数の幾何	48
第 6 章	時系列と線形システムの幾何	51
6.1	システム空間と時系列の多様体	51
6.2	システム空間の計量と接続	51
6.3	安定システムと安定フィードバック	52
第 7 章	多元情報理論と統計的推論	53
7.1	多元情報の統計的推論	53
7.2	仮説検定と Stein の補題	53
7.3	Chernoff 情報量と α -ダイバージェンス	54
第 8 章	情報幾何のこれからの話題	55
8.1	凸解析と線形計画の内点法の幾何学	55
8.2	ニューロ多様体と特異モデル	55
8.3	Wasserstein 幾何と最適輸送	56
8.4	量子観測の情報幾何	56
8.5	情報幾何が提起する数学上の問題	57
付録 A	Python による計算情報幾何	59
A.1	Fisher 情報行列の自動微分による計算	59
A.2	Christoffel 記号と測地線方程式の数値解法	59
A.3	JAX による実装例 (概念コード)	60
付録 B	多様体上の最適化：自然勾配法	61
B.1	Riemann 勾配と自然勾配	61
B.2	Fisher 情報行列の近似と K-FAC	61
付録 C	高次元統計空間の幾何学	63
C.1	高次元空間における直交性と集中現象	63
C.2	高次元特有の双対構造	63

第 1 章

微分幾何の基礎知識

情報幾何学において、確率分布の族は微分可能多様体としての構造を持つ。本章では、統計モデルの幾何学的性質を議論するための言語である、現代微分幾何学の基礎概念を厳密に導入する。特に、後の章で展開される双対幾何学（dual geometry）の基礎となるアファイン接続と計量の内容に重点を置く。

1.1 数学的準備：多様体論への橋渡し

微分幾何学（および情報幾何学）は、曲がった空間（多様体）を「局所的には平らな空間（ユークリッド空間）とみなして解析する」学問である。本節では、その基礎となるユークリッド空間の性質と、非線形写像を線形近似するための道具（逆写像定理）を復習する。

1.1.1 ユークリッド空間と双対空間

n 次元ユークリッド空間 \mathbb{R}^n の元（ベクトル）を列ベクトルとして表記する。

$$x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \in \mathbb{R}^n \quad (1.1)$$

これに対し、 \mathbb{R}^n 上の線形関数（1 次形式）全体の集合を双対空間（dual space）と呼び、 $(\mathbb{R}^n)^*$ と記す。双対空間の元 $\omega \in (\mathbb{R}^n)^*$ は行ベクトルとして表現され、ベクトル x への作用は行列積（内積）となる。

$$\omega(x) = (\omega_1 \quad \dots \quad \omega_n) \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} = \sum_{i=1}^n \omega_i x^i \quad (1.2)$$

情報幾何学では、この「縦ベクトル（反変ベクトル）」と「横ベクトル（共変ベクトル）」の区別が、後の「パラメータ空間」と「期待値空間」の双対性へと繋がる重要な概念となる。

1.1.2 逆写像定理：非線形な世界への入り口

多様体の定義や座標変換の議論において、最も強力な道具となるのが逆写像定理 (Inverse Function Theorem) である。これは「微分の行列（ヤコビアン）が正則なら、その写像は局所的に可逆である」ことを保証する。

定理 1.1 (逆写像定理). $U \subset \mathbb{R}^n$ を開集合とし、 $f: U \rightarrow \mathbb{R}^n$ を C^1 級写像とする。点 $p \in U$ において、ヤコビ行列 $J_f(p) = \left(\frac{\partial f^i}{\partial x^j}(p) \right)$ が正則（行列式が非ゼロ）であるならば、点 p のある近傍 V と $f(p)$ のある近傍 W が存在して、制限写像 $f|_V: V \rightarrow W$ は C^1 級の逆写像 $f^{-1}: W \rightarrow V$ を持つ（すなわち、微分同相となる）。

この定理の幾何学的な意味は、**「顕微鏡で拡大して（微分して）世界が潰れていなければ、その周りでは座標の張り替えが可能である」**ということである。これにより、曲がった空間に局所座標系（地図）を自由に導入することが正当化される。

1.1.3 曲面から内在的幾何学へ

微分幾何学の歴史は、3次元ユークリッド空間内の曲面論 (Theorema Egregium など) から始まった。しかし、情報幾何学で扱う「確率分布の空間」は、必ずしも自然なユークリッド空間の中に埋め込まれているわけではない。そこで、空間を「外から見る」のではなく、「中に住む住人の視点」で幾何学を構築する必要がある。これを**内在的 (intrinsic) 幾何学**と呼ぶ。G. F. Riemann は、空間を「局所的にはユークリッド空間に見えるが、大域的には曲がっているかもしれない図形」として定義した。これが現代の多様体 (Manifold) の概念である。

1.2 微分可能多様体

定義 1.2 (C^∞ 級微分可能多様体). n 次元位相空間 \mathcal{M} が以下の条件を満たすとき、 n 次元 C^∞ 級微分可能多様体 (differentiable manifold) と呼ぶ。

1. \mathcal{M} は Hausdorff 空間であり、かつ第2可算公理を満たす。
2. \mathcal{M} は C^∞ 級座標近傍系 (atlas) $\{(U_\alpha, \psi_\alpha)\}_{\alpha \in A}$ を持つ。すなわち、
 - $\bigcup_{\alpha \in A} U_\alpha = \mathcal{M}$ (被覆性)
 - $\psi_\alpha: U_\alpha \rightarrow \psi_\alpha(U_\alpha) \subset \mathbb{R}^n$ は同相写像である。
 - $U_\alpha \cap U_\beta \neq \emptyset$ なる任意の α, β に対し、座標変換写像

$$\psi_\beta \circ \psi_\alpha^{-1}: \psi_\alpha(U_\alpha \cap U_\beta) \rightarrow \psi_\beta(U_\alpha \cap U_\beta)$$

は C^∞ 級である (C^∞ 適合性)。

注釈 1.3 (定義の直観的意味). 定義中の座標変換写像 $\psi_\beta \circ \psi_\alpha^{-1}$ は、異なる局所座標系

(地図) の間での「翻訳ルール」を表している。

- ψ_α^{-1} : 座標空間 \mathbb{R}^n (地図上の数値) から、多様体 M (実世界) への逆写像。
- ψ_β : 多様体 M (実世界) から、別の座標空間 \mathbb{R}^n (別の地図) への写像。

これらを合成することで、実世界を経由せずに「地図 A の座標」を「地図 B の座標」に直接変換する関数が構成される。この関数が滑らか (微分可能) であれば、微積分の計算が座標系の取り方によらず矛盾なく行えることが保証される。

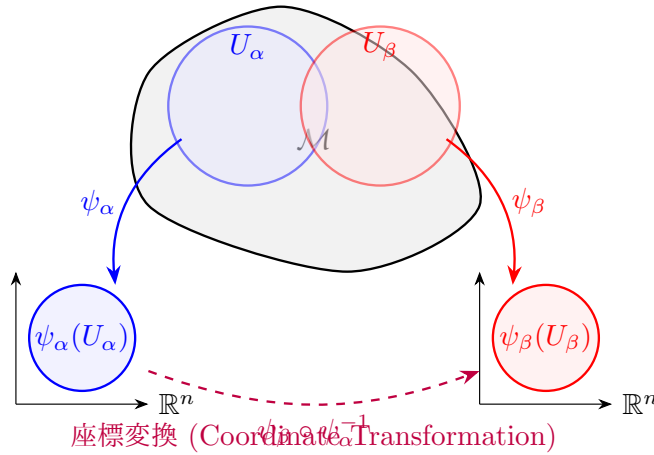


図 1.1 多様体と座標変換の直観的イメージ: 地球 (多様体) を複数の地図 (局所座標) で覆い、それらの間の翻訳 (座標変換) をつなぎ合わせる。

統計モデルにおいて、座標近傍 (U, ψ) はパラメータ空間に対応し、点 $p \in M$ の局所座標を $\xi = (\xi^1, \dots, \xi^n) = \psi(p)$ と表記する。

1.3 接ベクトルと接空間

導入: なぜ接ベクトルを「微分」と見なすのか

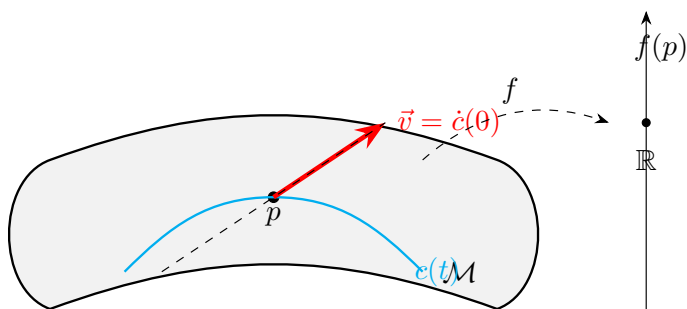
高校数学などのユークリッド空間 \mathbb{R}^n において、点 p におけるベクトル $\vec{v} = (v^1, \dots, v^n)$ は、「始点 p から伸びる矢印」として直観的に理解される。このベクトル \vec{v} は、滑らかな関数 f に対して、その方向への変化率 (方向微分) を与える作用素と見なすことができる。

$$D_{\vec{v}}f = \lim_{t \rightarrow 0} \frac{f(p + t\vec{v}) - f(p)}{t} = \sum_{i=1}^n v^i \frac{\partial f}{\partial x^i} \quad (1.3)$$

ここで、以下の対応関係に注目したい。

$$\text{幾何的な矢印 } \vec{v} \longleftrightarrow \text{代数的な微分演算子 } \sum v^i \frac{\partial}{\partial x^i}$$

平坦でない一般の多様体上では、離れた 2 点間の引き算 ($p + t\vec{v}$ のような操作) が定義できないため、「矢印」を直接定義することが困難である。しかし、右辺の「微分演算子」と



幾何学的視点: 曲線 $c(t)$ の速度ベクトル
 代数的視点: 関数 f の変化率 $\frac{d}{dt}f(c(t))$

図 1.2 接ベクトルの 2 つの側面: 曲線に対する接線 (幾何) と、関数に対する微分作用素 (代数)。

しての性質 (線形性やライプニッツ則) は、局所座標系を用いることで多様体上でも自然に定義可能である。

したがって、現代幾何学では、この**「関数 f を微分する作用素」そのものを接ベクトルと定義する**という一般化を行う。これにより、空間が曲がっているかどうかに関係せず、ベクトルを厳密に扱うことが可能になる。

1.3.1 接ベクトルの定義

多様体上のベクトルを、曲線に対する接線としてではなく、関数に対する「方向微分作用素 (derivation)」として定義する代数的なアプローチをとる。

定義 1.4 (接ベクトル). 点 $p \in \mathcal{M}$ における接ベクトル X とは、点 p の近傍で定義された C^∞ 関数全体の集合 $C^\infty(p)$ から \mathbb{R} への写像であり、以下の線形性と Leibniz 則を満たすものである。任意の $f, g \in C^\infty(p), \lambda, \mu \in \mathbb{R}$ に対して、

$$X(\lambda f + \mu g) = \lambda X(f) + \mu X(g) \quad (1.4)$$

$$X(fg) = X(f)g(p) + f(p)X(g) \quad (1.5)$$

点 p におけるすべての接ベクトルの集合は n 次元線形空間となり、これを接空間 (tangent space) と呼び $T_p\mathcal{M}$ で表す。

局所座標系 (ξ^1, \dots, ξ^n) が与えられたとき、座標曲線に沿った偏微分演算子

$$\left\{ \partial_i := \frac{\partial}{\partial \xi^i} \right\}_{i=1}^n$$

は $T_p\mathcal{M}$ の基底をなす。これを自然基底 (natural basis) と呼ぶ。任意の接ベクトル $X \in T_p\mathcal{M}$ は、成分 $X^i \in \mathbb{R}$ を用いて一意に

$$X = X^i \partial_i$$

と展開される (Einstein の縮約記法を用いる)。

導入：ヤコビアンと連鎖律の直観的理解

命題に入る前に、高校数学における「合成関数の微分公式（連鎖律）」を思い出そう。変数 x が t の関数であるとき、微分演算子は以下のように変換された。

$$\frac{d}{dt} = \frac{dx}{dt} \frac{d}{dx} \quad (1.6)$$

ここで係数 $\frac{dx}{dt}$ は、座標 t の目盛りが座標 x の目盛りに対してどれだけ「伸び縮み」しているかを表す拡大率（ヤコビアン）である。

多変数の場合も全く同様である。座標系 ξ を座標系 η に変換するとき、偏微分演算子 $\frac{\partial}{\partial \xi^i}$ は、 η 側のすべての成分 $\frac{\partial}{\partial \eta^j}$ の線形結合として表される。その「重み係数」がヤコビアン行列 $\frac{\partial \eta^j}{\partial \xi^i}$ である。

これを踏まえて、以下の一般的な変換則を定義する。

命題 1.5 (ベクトルの座標変換則). 別の局所座標系 $\eta = (\eta^1, \dots, \eta^n)$ を考える。座標変換の Jacobian を $J_j^i = \frac{\partial \eta^i}{\partial \xi^j}$ とすると、基底および成分は以下のように変換される。

$$\frac{\partial}{\partial \xi^i} = \frac{\partial \eta^j}{\partial \xi^i} \frac{\partial}{\partial \eta^j} \quad (1.7)$$

$$X(\xi)^i = \frac{\partial \xi^i}{\partial \eta^j} X(\eta)^j \quad (1.8)$$

すなわち、接ベクトルは反変ベクトル (contravariant vector) である。

Proof. 任意の滑らかな関数 $f: M \rightarrow \mathbb{R}$ を考える。連鎖律 (chain rule) により、偏微分演算子は以下のように変換される。

$$\frac{\partial f}{\partial \xi^i} = \frac{\partial f}{\partial \eta^j} \frac{\partial \eta^j}{\partial \xi^i} \quad (1.9)$$

この演算子の関係式から、直ちに基底の変換則 (1.4) が得られる。

次に、接ベクトル X 自体は座標系に依存しない実体であるため、

$$X = X(\xi)^i \frac{\partial}{\partial \xi^i} = X(\eta)^j \frac{\partial}{\partial \eta^j} \quad (1.10)$$

が成り立つ。ここに (1.4) を代入すると、

$$X(\xi)^i \left(\frac{\partial \eta^j}{\partial \xi^i} \frac{\partial}{\partial \eta^j} \right) = X(\eta)^j \frac{\partial}{\partial \eta^j} \quad (1.11)$$

基底 $\{\frac{\partial}{\partial \eta^j}\}$ の一次独立性より、係数を比較して

$$X(\eta)^j = \frac{\partial \eta^j}{\partial \xi^i} X(\xi)^i \quad (1.12)$$

この両辺に $\frac{\partial \xi^k}{\partial \eta^j}$ を乗じて和をとれば、

$$\frac{\partial \xi^k}{\partial \eta^j} X(\eta)^j = \frac{\partial \xi^k}{\partial \eta^j} \frac{\partial \eta^j}{\partial \xi^i} X(\xi)^i = \delta_i^k X(\xi)^i = X(\xi)^k \quad (1.13)$$

となり、成分の変換則 (1.5) が得られる。 \square

1.4 余接空間と 1 形式

直観的理解：速度と勾配

接空間 $T_p\mathcal{M}$ の元（接ベクトル）が、多様体上を動く粒子の「速度」を表すのに対し、その双対空間である余接空間 $T_p^*\mathcal{M}$ の元（余接ベクトル）は、関数 f の「勾配」や「等高線」を表す幾何学的対象である。

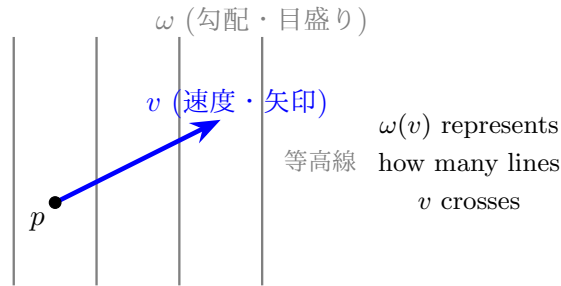


図 1.3 余接ベクトル (1 形式) の直観的イメージ：空間に置かれた「目盛り（等高線）」であり、接ベクトル（矢印）がそれをいくつ貫くかを計測する。

- 接ベクトル $X \in T_p\mathcal{M}$: 方向微分作用素であり、局所的には $\frac{\partial}{\partial \xi^i}$ の線形結合で書かれる。物理的な単位は「距離」に近い。
- 余接ベクトル $\omega \in T_p^*\mathcal{M}$: 接ベクトルを入力として実数を出力する線形写像（汎関数）である。局所的には全微分 $d\xi^i$ の線形結合で書かれる。物理的な単位は「1/距離」に近い。

もっとも基本的な余接ベクトルは、滑らかな関数 f の全微分（勾配） df である。

$$(df)_p(X) = X(f) \quad (1.14)$$

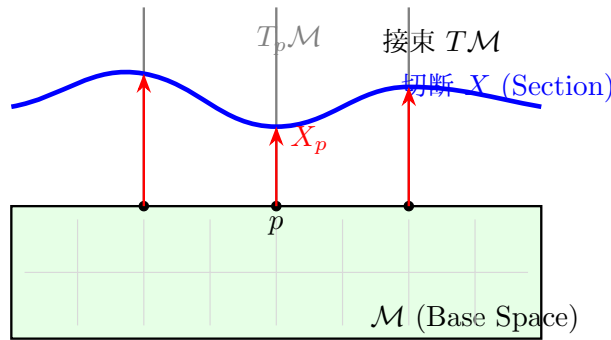
これは「速度 X で動いたときの、関数 f の瞬間変化率」を与える。すなわち、余接空間とは「多様体上の微小な定規（ものさし）」たちの集まりといえる。

1.5 ベクトル場とテンソル場

多様体全体で定義された滑らかなベクトルの分布をベクトル場と呼ぶ。

定義 1.6 (ベクトル場). \mathcal{M} 上のベクトル場 $\mathfrak{X}(\mathcal{M})$ とは、各点 $p \in \mathcal{M}$ に $X_p \in T_p\mathcal{M}$ を対応させる切断 (section) であり、局所的に $X = X^i(\xi)\partial_i$ と書いたとき、各成分 $X^i(\xi)$ が C^∞ 級関数となるものである。

注釈 1.7 (切断 (Section) の直観的意味). 「切断」という用語は、ファイバー束 (Fiber Bundle) の理論に由来する。多様体 \mathcal{M} の各点 p には、接空間 $T_p\mathcal{M}$ という線形空間が「植わって」いる。これらすべての接空間を束ねた全体集合 $T\mathcal{M} = \bigcup_{p \in \mathcal{M}} T_p\mathcal{M}$ を接束 (Tangent Bundle) と呼ぶ。ベクトル場 X とは、この大きな束の中から、各点 p ごとにベクトル X_p を1つずつ滑らかに選び出して (切り出して) 作った関数のことである。気象図において、地図上の各地点に「風向・風速ベクトル」を描き込んだものが、まさに「地表という多様体上のベクトル場の切断」である。



接束 (すべての可能性) の中から、各点 p ごとに
1つのベクトル (値) を選び取る = 「切断」

図 1.4 切断 (Section) のイメージ: 束 (Bundle) を横に切って、各点に1つの値を対応させる。

注釈 1.8 (フラクツール記号 \mathfrak{X} の読み方). ベクトル場全体の集合を表す記号 $\mathfrak{X}(\mathcal{M})$ は、ドイツ文字 (フラクツール体) の「X」であり、数学の慣習では「カイ (χ)」や「エックス」と読まれることが多い。主なフラクツール記号の読み方を以下に挙げる。

記号	元の文字	読み方 (慣用)	主な用途
\mathfrak{X}	X	カイ / エックス	ベクトル場
\mathfrak{g}	g	ゲー / ジー	Lie 代数
\mathfrak{S}	S	エス	対称群
\mathfrak{p}	p	ペー	素イデアル

接空間 $T_p\mathcal{M}$ の双対空間を余接空間 (cotangent space) と呼び、 $T_p^*\mathcal{M}$ と記す。その基底は座標微分形式 $\{d\xi^i\}_{i=1}^n$ であり、 $\langle d\xi^i, \partial_j \rangle = \delta_j^i$ を満たす。

一般に、 (r, s) 型テンソル場とは、各点において

$$T : \underbrace{T_p^* \mathcal{M} \times \cdots \times T_p^* \mathcal{M}}_r \times \underbrace{T_p \mathcal{M} \times \cdots \times T_p \mathcal{M}}_s \rightarrow \mathbb{R}$$

なる多重線形写像を与える場である。局所座標系では以下のように表現される。

$$T = T_{j_1 \dots j_s}^{i_1 \dots i_r} \partial_{i_1} \otimes \cdots \otimes \partial_{i_r} \otimes d\xi^{j_1} \otimes \cdots \otimes d\xi^{j_s}$$

1.6 Riemann 計量

多様体に幾何学的構造（長さ、角度）を与えるものが Riemann 計量である。

導入：リーマン計量の直感（歪んだ方眼紙）

ユークリッド空間 \mathbb{R}^2 において、微小距離の二乗は $ds^2 = dx^2 + dy^2$ で与えられる。これは「どこでも 1 目盛りの長さは同じ」かつ「軸は直交している」ことを前提としている。

しかし、ゴム膜のように歪んだ空間や、メルカトル図法のような地図上では、この前提は崩れる。ある地点では x 方向に引き伸ばされていたり、座標軸が斜めに交わっていたりするかもしれない。そこで、微小な移動ベクトル $(\Delta x, \Delta y)$ の「本当の長さ」を測るために、場所 ξ ごとに補正係数行列 $G(\xi)$ を導入する。

$$ds^2 = g_{11}dx^2 + 2g_{12}dxdy + g_{22}dy^2 \quad (1.15)$$

この補正係数のセット $g_{ij}(\xi)$ がリーマン計量である。つまり、リーマン計量とは「空間の各点に、それぞれ異なる縮尺を持つ定規を割り当てたもの」と理解できる。

定義 1.9 (Riemann 計量). Riemann 計量 g とは、 $(0, 2)$ 型テンソル場であり、各点 p において以下の性質を持つものである。

1. 対称性: $g(X, Y) = g(Y, X)$
2. 正定値性: $g(X, X) \geq 0$ であり、 $g(X, X) = 0 \iff X = 0$

局所成分を $g_{ij}(\xi) = g(\partial_i, \partial_j)$ と書くとき、行列 (g_{ij}) は正定値対称行列となる。

注釈 1.10 ($(0, 2)$ 型テンソル場の意味). テンソルの型 (r, s) は、その写像が受け取る引数の種類を表している。 $(0, 2)$ 型テンソル場 g とは、各点 p において 2 つの接ベクトル $X, Y \in T_p \mathcal{M}$ を引数に取り、実数 $g(X, Y) \in \mathbb{R}$ を出力する多重線形写像である。局所座標系では $g = g_{ij}d\xi^i \otimes d\xi^j$ と展開され、成分 g_{ij} は 2 次形式を規定する行列として振る舞う。Riemann 計量が $(0, 2)$ 型であることは、それが「2 つのベクトルの内積」を定義する道具であることを数学的に明示している。

2 つの接ベクトル X, Y の内積は $\langle X, Y \rangle_g = g_{ij}X^i Y^j$ で与えられる。また、計量 g は接空間 $T_p \mathcal{M}$ と余接空間 $T_p^* \mathcal{M}$ の間の同型（音楽同型）を誘導し、添字の上げ下げ（raising and lowering indices）に用いられる。

注釈 1.11 (音楽同型と添字の上げ下げ). Riemann 計量 g は、接ベクトル X (成分 X^i) と余接ベクトル ω (成分 ω_i) を一対一に対応させる写像 (線形同型) を引き起こす。これを音楽記号を用いて以下のように表す。

- フラット (下げ) : X^b

$$(X^b)_i = g_{ij} X^j$$

ベクトル X の添字を g で「下げて」、1 形式 (コベクター) に変換する操作。

- シャープ (上げ) : ω^\sharp

$$(\omega^\sharp)^i = g^{ij} \omega_j$$

1 形式 ω の添字を逆行列 g^{ij} で「上げて」、ベクトルに変換する操作。

この対応により、物理学では「速度ベクトル (v^i)」と「運動量ベクトル (p_i)」が、質量 (計量) を通じて等価なものとして扱われる。

【コラム】(0, 2) 型テンソルとは「2 人用の評価マシン」

定義に現れる「(0, 2) 型テンソル場」という言葉は、一言で言えば、「ベクトル (矢印) を 2 つ入力すると、それに応じた『スコア (数値)』を計算して返してくれる装置」のことである。

1. なぜ「(0, 2)」という名前なの？

テンソルの型を表す (r, s) という数字は、「何を何個食べるか」を表している。

- r (上付き) : 「余接ベクトル (目盛り)」を食べる数。
- s (下付き) : 「接ベクトル (矢印)」を食べる数。

つまり、(0, 2) 型テンソルは、「目盛りはいらないけど、矢印 (接ベクトル) を 2 個ください」という装置である。

2. 具体的に何をしているの？

Riemann 計量 g は (0, 2) 型テンソルの代表例である。2 つのベクトル X と Y を入力すると、内積 $\langle X, Y \rangle_g$ という数値を返す。

$$g(\underbrace{X}_{\text{矢印 1}}, \underbrace{Y}_{\text{矢印 2}}) = \text{数値 (スカラー)}$$

これは、高校数学で習った行列の計算そのものである。

$$\begin{pmatrix} X^1 & X^2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} Y^1 \\ Y^2 \end{pmatrix}$$

この真ん中にある「2 次形式を決める行列」が、(0, 2) 型テンソルの正体である。

3. 「場」がつくと何が変わる？

「テンソル」だけならただの行列だが、「テンソル場」になると、「場所によって行列の中身が変わってもいいよ」という意味になる。

- 地点 A: $g = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (ここでは普通の方眼紙)
- 地点 B: $g = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$ (ここでは 1cm が 5 倍重い)

このように、多様体の各点にそれぞれの行列が割り当てられている状態を「テンソル場」と呼ぶ。

まとめ

- (0, 2) 型: ベクトルを 2 つ受け取る。
- テンソル: 2 次形式 (行列) のような計算ルール。
- 場: 多様体の場所ごとにそのルールが定義されている。

計量 g が (0, 2) 型である理由は、「2 つの地点のズレ (ベクトル X と Y) から、距離や角度という 1 つの数値を導き出したいから」である。

【コラム】リーマン多様体：なぜ (M, g) のセットなのか？

リーマン多様体の定義は、多様体 M と計量 g のペア (M, g) として与えられる。なぜ「セット」でなければならないのか？それは一言で言えば、**「定規がないと長さや角度が決まらないから」**である。

1. 多様体 M だけでは「ふにゃふにゃ」：ただの微分可能多様体 M は、ゴム膜のように伸縮自在であり、そこには「長さ」や「角度」という概念が存在しない (トポロジーと微分構造のみ)。例えば、地球儀 (多様体) があっても、そこに「緯線・経線の目盛り (計量)」が描かれていなければ、東京と大阪の距離を測ることはできない。
2. リーマン計量 g = 「定規セット」：リーマン計量 g は、多様体の**各点に配られた定規 (内積) **である。これを与えることではじめて、多様体は「硬い形 (幾何学的構造)」を持つことができる。
 - **多様体 (M) **: 舞台となる空間。
 - **計量 (g) **: その空間での「距離のルール」。
 同じ M (例えば球面) でも、違う定規 g' (例えばデコボコの定規) を使えば、それは「違うリーマン多様体 (歪んだ球面)」になる。

日常の例で言えば、**Google マップ**がわかりやすい。

- **多様体 M^* : スマホ画面上の地図データ（座標データ）。
- **計量 g^* : 画面右下のスケールバー（「ここまで 500m」）。

もしスケールバーがなければ、画面上の「1cm」が現実の「100m」なのか「10km」なのか分からない。地図データ（ M ）とスケール（ g ）がセットになって初めて、ナビゲーション（距離計算）が可能になる。すなわち、

「リーマン多様体は (M, g) のセットである」 \iff 「空間だけじゃ距離が測れないから、専用の定規も一緒に持たせてね」という意味である。

例 1：2 次元球面（地球の表面）の計量

半径 R の球面の表面を、緯度 ϕ と経度 θ で表す。（ $\phi = 0$ を北極とする場合や、赤道を 0 とする場合がありますが、ここでは $z = R \cos \phi$ とする極座標系を用います）。

1. 3 次元空間での座標:

$$x = R \sin \phi \cos \theta, \quad y = R \sin \phi \sin \theta, \quad z = R \cos \phi$$

2. 微小変化（全微分）を計算: たとえば経度 θ を少し動かすと、横に $dx = -R \sin \phi \sin \theta d\theta$ だけ動く。

$$dx = R(\cos \phi \cos \theta d\phi - \sin \phi \sin \theta d\theta)$$

$$dy = R(\cos \phi \sin \theta d\phi + \sin \phi \cos \theta d\theta)$$

$$dz = -R \sin \phi d\phi$$

3. 距離の二乗 $ds^2 = dx^2 + dy^2 + dz^2$ に代入: 計算して整理すると、以下の式が得られる。

$$ds^2 = \underbrace{R^2}_{g_{\phi\phi}} d\phi^2 + \underbrace{R^2 \sin^2 \phi}_{g_{\theta\theta}} d\theta^2$$

4. 計量行列 G :

$$G(\phi, \theta) = \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \sin^2 \phi \end{pmatrix}$$

【この計算が教えること】

- 場所による変化: 北極（ $\phi = 0$ ）に行くと、 $\sin \phi = 0$ になり、 $g_{\theta\theta} = 0$ となる。これは「北極点では経度をいくら動かしても移動距離はゼロ」という事実を、計量が数値として表しているのである。

【コラム】「接続」とは翻訳機である

多様体論で最もつまづきやすい概念である「接続（Connection）」について、数式の前にその心（直観的イメージ）を解説する。

1. なぜ「接続」が必要なのか？（多様体の不便な点）

平らな空間（ユークリッド空間）では、ベクトルはどこでも自由に移動できる。例えば、東京にある「北向きの矢印」と、大阪にある「北向きの矢印」は、誰が見ても「同じベクトル（平行）」である。

しかし、多様体（曲がった空間）では、これが通用しない。

- 点 A（東京）のベクトルは、東京の接空間 $T_A\mathcal{M}$ に住んでいる。
- 点 B（大阪）のベクトルは、大阪の接空間 $T_B\mathcal{M}$ に住んでいる。

これらは「住む世界が違う」ので、直接比べたり、引き算（微分）したりすることができない。「東京の 1 万円」と「ニューヨークの 100 ドル」を、そのままの数値で足し算できないのと同じである。

2. 接続＝「移動のルール」

そこで、離れた 2 点のベクトルを比べるために、「ベクトルを A から B へ移動させるルール」を決める必要がある。これが接続（Connection）である。

「東京のベクトル X を、そのまま大阪まで持って行ったら、大阪ではどのベクトル X' になるのか？」

この問いに対する答えを決めるのが接続であり、この移動操作を平行移動（Parallel Transport）と呼ぶ。接続とは、異なる接空間同士をつなぐ「通信ケーブル」のようなものである。

3. 接続係数 Γ = 「両替手数料」あるいは「ズレ」

接続係数（Christoffel 記号） Γ_{ij}^k は、移動したときに生じる「座標のズレ」を表している。

- 平らな空間（ $\Gamma = 0$ ）：
ベクトルをそのまま平行移動しても、成分は変わらない（東京の北向き＝大阪の北向き）。
- 曲がった空間・座標系（ $\Gamma \neq 0$ ）：
ベクトルを「真っ直ぐ」運んだつもりでも、座標の目盛り自体が歪んでいるため、到着地では成分が変わってしまう（東京で北を向いて真っ直ぐ進んだら、北極点では下を向いているかもしれない）。

この「座標の歪みによって生じるベクトルの成分変化」を補正して、「実質的には変化していない（平行である）」と言えるようにするための補正項が、接続 Γ である。

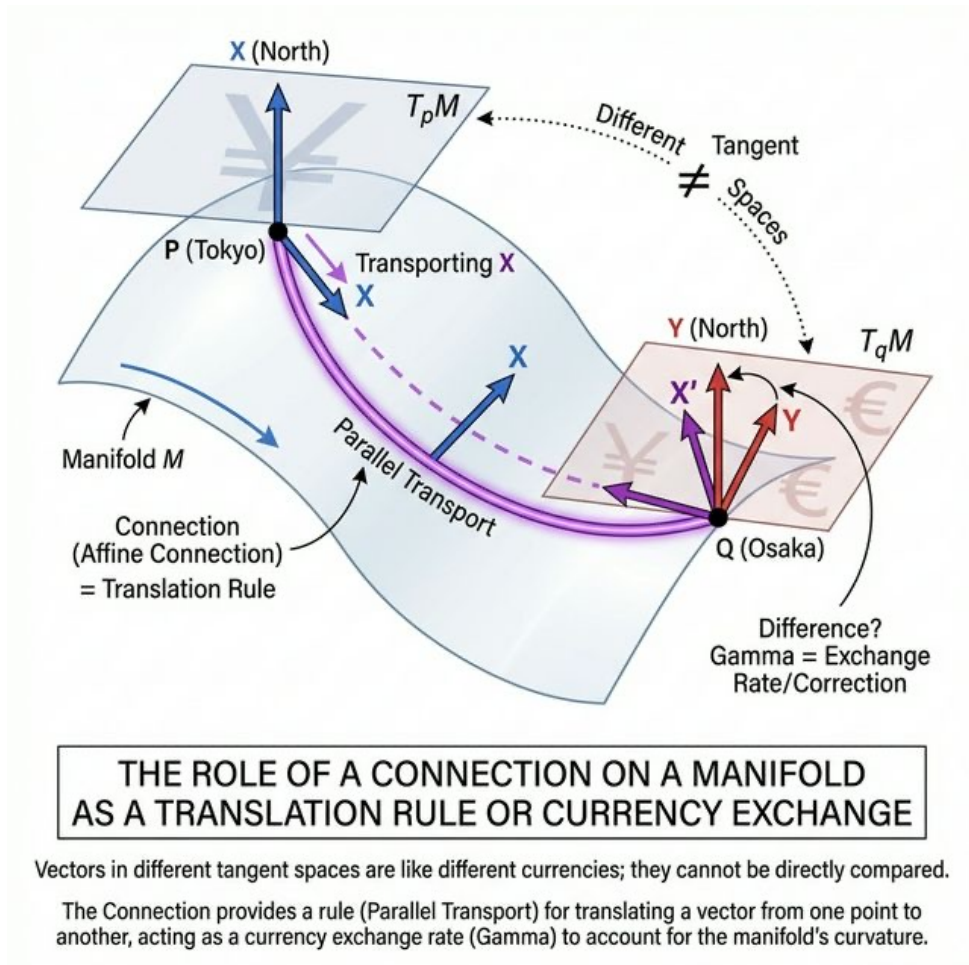


図 1.5 接続 (Connection) の直観的イメージ：異なる接空間 (通貨) をつなぐ翻訳機としての役割。

4. 共変微分 = 「為替レートを考慮した変化率」

接続を使って定義される共変微分 ∇ は、以下のような意味を持つ。

共変微分 = (見た目の成分変化) + (座標の歪みによる補正)

数式で書けば、以下のようになる。

$$\nabla_j X^i = \underbrace{\frac{\partial X^i}{\partial x^j}}_{\text{単なる成分微分}} + \underbrace{\Gamma_{jk}^i X^k}_{\text{座標のズレ補正}} \quad (1.16)$$

これがゼロになるとき、そのベクトルは「(空間の曲がりに逆らわずに) 真っ直ぐ移動している」とみなされる。すなわち、共変微分とは、単なる成分の引き算ではなく、「通貨の両替レート (接続) を考慮した上での、本質的な価値の変化率」を測る操作なのである。

1.7 アファイン接続と共変微分

Riemann 計量だけでは、異なる接空間に属するベクトル同士を比較（微分）することはできない。これを可能にする構造がアファイン接続である。

【コラム】アファイン接続と共変微分の直感

数学的な定義に入る前に、この2つの記号が「何をするための道具なのか」をイメージしましょう。

1. アファイン接続 $\nabla =$ 「平行移動マシン」

アファイン接続（Affine Connection）とは、一言で言えば「曲がった空間で『真っ直ぐ』を決めるルール」です。

- 直感イメージ: 多様体 M の上を移動する車を想像してください。「ハンドルを固定して真っ直ぐ走れ」と言われても、地面がデコボコしていたり、地球のように丸かったりすると、車は勝手に曲がってしまいます。そこで、「地面の曲がりに合わせてハンドルを微調整し、常に『真っ直ぐ』な状態を保つための操作」が必要です。この「ハンドルの操作ルール」こそがアファイン接続です。
- 機能: 点 p にあるベクトル Y を、点 q まで「平行移動（Parallel Transport）」させることができます。「接続」という名前は、「隣り合う接空間 T_pM と T_qM をスムーズにつなぐ（接続する）」ことに由来します。

2. 共変微分 $\nabla_X Y =$ 「補正付きの微分」

高校数学の微分は「変化率」でした。

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

しかし、曲がった空間でベクトル場の微分をしようとすると、問題が起きます。「場所が違うベクトル同士は、基準が違うので単純な引き算ができない」のです（通貨の違うお金を足せないのと同じ）。

そこで、共変微分（Covariant Derivative）が登場します。これは、「座標のズレを補正した上で、純粋な変化量だけを取り出す微分」です。

$$\nabla_X Y = \underbrace{X(Y)}_{\text{単純な微分}} + \underbrace{\text{補正項}(\Gamma)}_{\text{座標のズレ}}$$

- 直感イメージ: あなたは動く歩道（座標系）の上を歩いています。

- 単純な微分: 外から見たあなたの速度。
 - 共変微分: 動く歩道の上で、あなたが実際に足を使って歩いている速度。
- 「歩道が勝手に動く分 (補正項)」を差し引くことで、あなたの「正味の頑張り」を測るのが共変微分です。

3. ∇ が教えてくれること

共変微分を計算すると、幾何学的な意味がわかります。

- $\nabla_X Y = 0$ (共変微分がゼロ): ベクトル Y は、方向 X に沿って「平行移動」している。(変化していないように見えるのは、空間の曲がりに身を任せているから)
- $\nabla_X X = 0$ (自分自身の方向への共変微分がゼロ): その曲線は「測地線 (Geodesic)」である。(ハンドルを切らずに真っ直ぐ進んでいる状態)

定義 1.12 (アファイン接続). アファイン接続 ∇ とは、2つのベクトル場 $X, Y \in \mathfrak{X}(\mathcal{M})$ に対して、新たなベクトル場 $\nabla_X Y$ を対応させる写像 $\nabla: \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \rightarrow \mathfrak{X}(\mathcal{M})$ であり、以下の性質 (Koszul の公理) を満たすものである。ここで $f \in C^\infty(\mathcal{M})$ とする。

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$ (X に関する $C^\infty(\mathcal{M})$ -線形性)
2. $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$ (Y に関する \mathbb{R} -線形性)
3. $\nabla_X (fY) = (Xf)Y + f\nabla_X Y$ (Leibniz 則)

$\nabla_X Y$ を、 X 方向への Y の共変微分 (covariant derivative) と呼ぶ。

1.7.1 接続係数 (Christoffel 記号)

自然基底 $\{\partial_i\}$ に関する共変微分を考える。 $\nabla_{\partial_i} \partial_j$ は再びベクトル場であるから、基底の線形結合で表せる。

$$\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k(\xi) \partial_k \quad (1.17)$$

この係数関数 Γ_{ij}^k を接続の係数、あるいは (一般化された) Christoffel 記号と呼ぶ。情報幾何学において極めて重要な事実、接続係数 Γ_{ij}^k は必ずしも i, j に関して対称ではないということである。

命題 1.13 (共変微分の成分表示). ベクトル場 $Y = Y^j \partial_j$ の $X = X^i \partial_i$ 方向への共変微分 $\nabla_X Y$ の成分は以下で与えられる。

$$\begin{aligned} \nabla_X Y &= X^i \nabla_{\partial_i} (Y^j \partial_j) \\ &= X^i ((\partial_i Y^j) \partial_j + Y^j \nabla_{\partial_i} \partial_j) \\ &= X^i (\partial_i Y^k + Y^j \Gamma_{ij}^k) \partial_k \end{aligned} \quad (1.18)$$

例 1.14 (極座標における Christoffel 記号). 平面 \mathbb{R}^2 上の極座標 $(x^1, x^2) = (r, \theta)$ を

考える。計量行列は $G = \text{diag}(1, r^2)$ である。このとき、非ゼロの Christoffel 記号 (Levi-Civita 接続) は以下の3つのみである。

- $\Gamma_{\theta\theta}^r = -r$: 測地線方程式 $\ddot{r} - r\dot{\theta}^2 = 0$ に対応。これは遠心力を表す項である。
- $\Gamma_{r\theta}^\theta = \Gamma_{\theta r}^\theta = \frac{1}{r}$: 測地線方程式 $\ddot{\theta} + \frac{2}{r}\dot{r}\dot{\theta} = 0$ に対応。これはコリオリ力 (角運動量保存) を表す項である。

このように、空間自体は平坦であっても、曲がった座標系を用いることで Γ は非ゼロとなり、慣性力 (見かけの力) として現れる。

1.8 平坦性

定義 1.15 (曲率テンソル). 接続 ∇ の Riemann-Christoffel 曲率テンソル R は、ベクトル場 X, Y, Z に対して以下で定義される $(1, 3)$ 型テンソル場である。

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z \quad (1.19)$$

ここで $[X, Y] = XY - YX$ は Lie 括弧積である。

基底ベクトルに対する成分 $R(\partial_i, \partial_j)\partial_k = R_{ijk}^l \partial_l$ は、接続係数を用いて次のように計算される。

$$R_{ijk}^l = \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{jk}^m \Gamma_{im}^l - \Gamma_{ik}^m \Gamma_{jm}^l \quad (1.20)$$

定義 1.16 (平坦性). 接続 ∇ に対して曲率テンソル R が恒等的に 0 であるとき、空間は局所的に平坦 (locally flat) であるという。

直感：方眼紙にならせるか？

平坦な空間とは、局所的には「歪みのない方眼紙 (ユークリッド空間)」と同じ構造を持つ空間のことである。非平坦な空間 (曲がった空間) では、どんなに小さく切り取っても、方眼紙のように真っ直ぐなグリッドを敷くことができない。

- **平面・円柱 (平坦)** : 平面は当然平坦である。また、円柱の側面も、切り開けば平面になる (展開図が描ける) ため、幾何学的には「平坦」である。円柱上の三角形の内角の和は 180° である。
- **球面 (非平坦)** : 球面の皮を剥いて平面に広げようとすると、必ず裂け目ができるか、引き伸ばさなければならない (世界地図の歪み)。球面上の三角形 (大円で囲まれた領域) の内角の和は 180° より大きくなる。これは、どうあがいても平面とは同じ構造になれないことを示している。

数式の意味：曲率が 0 なら「座標でごまかせる」

「平坦性」の数学的な定義は「曲率テンソル R が 0」であることだが、これは次のように言い換えることができる。

「平坦な空間では、うまい座標系（アファイン座標系）を選べば、その座標系では接続係数 Γ_{ij}^k を局所的に全て 0 にできる」

例 1.17 (極座標の平坦性). 先ほどの例（極座標）では、 $\Gamma \neq 0$ であった（遠心力やコリオリ力が現れた）。しかし、この空間（平面 \mathbb{R}^2 ）の曲率テンソル R を成分計算すると、すべて 0 になる。これは、極座標 (r, θ) から直交座標 (x, y) へと座標変換すれば、

$$\Gamma_{ij}^k \equiv 0 \quad (\text{すべての成分がゼロ})$$

となることを意味している。つまり、極座標に現れた Γ は、単なる「座標の取り方の悪さ」に由来する見かけの力（慣性力）であり、座標変換によって消し去ることができる。

一方、球面のような「真に曲がった」空間では、どのような座標系を選んでも Γ を広範囲でゼロにすることはできない（曲率 $R \neq 0$ ）。これが「本質的な曲がり（重力）」と「見かけの曲がり（慣性力）」の違いである。

情報幾何における平坦性

情報幾何学では、1つの空間（統計的モデル）に対して、対となる2つの接続（ α -接続と $-\alpha$ -接続）を考える。特に重要なのが、 $\alpha = \pm 1$ に対応する **e-接続（指数型接続）** と **m-接続（混合型接続）** である。

多くの重要な統計モデル（指数型分布族）は、「Riemann 接続に関しては曲がっている（ $R^{(0)} \neq 0$ ）が、e-接続に関しては平坦（ $R^{(e)} = 0$ ）」という特殊な構造を持つ。この「双対平坦性」こそが、統計的推論や学習アルゴリズムを幾何学的に解くための鍵となる。（詳細は第3章で解説する）

1.9 自己平行性と部分多様体

定義 1.18 (自己平行部分多様体). 多様体 \mathcal{M} の部分多様体 $S \subset \mathcal{M}$ が接続 ∇ に関して自己平行 (autoparallel) であるとは、任意の $p \in S$ と任意の接ベクトル $X \in T_p S$ に対し、 \mathcal{M} における X 方向の測地線が常に S に含まれることをいう。

これは、部分多様体 S への接ベクトル場 $X, Y \in \mathfrak{X}(S)$ に対し、 \mathcal{M} の接続による共変微分 $\nabla_X Y$ が再び S の接空間 TS に留まることと同値である。すなわち、埋め込み曲率（Euler-Schouten 曲率）が消失することを意味する。統計的モデルにおいて、指数型分布族は（関数空間内の）e-接続に関して自己平行な部分多様体となる。

1.10 Riemann 接続 (Levi-Civita 接続)

一般のアファイン接続 ∇ は計量 g と無関係に定義できるが、計量構造と整合する特別な接続が存在する。

定義 1.19 (計量整合性とねじれ率). 1. 接続 ∇ が計量 g と整合的 (metric compatible) であるとは、任意のベクトル場 X, Y, Z に対して次が成り立つことである。

$$X\langle Y, Z \rangle_g = \langle \nabla_X Y, Z \rangle_g + \langle Y, \nabla_X Z \rangle_g$$

すなわち、 $\nabla g = 0$ (計量の共変微分が 0) である。

2. 接続 ∇ のねじれ率テンソル (torsion tensor) T とは以下で定義される $(1, 2)$ 型テンソルである。

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$

$T \equiv 0$ のとき、接続はねじれがない (torsion-free) または対称であるという。

定理 1.20 (Riemann 幾何学の基本定理). Riemann 多様体 (M, g) において、計量と整合的かつねじれのないアファイン接続 $\nabla^{(0)}$ が唯一存在する。これを Riemann 接続または Levi-Civita 接続と呼ぶ。その接続係数は計量により以下のように一意に決定される。

$$\Gamma_{ij,k}^{(0)} = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}) \quad (1.21)$$

ここで、 $\Gamma_{ij,k} := g_{km} \Gamma_{ij}^m$ である。

直感：なぜ Levi-Civita 接続が「標準」なのか？

Levi-Civita 接続は、計量 g (定規・分度器) を持つ空間において、最も自然に決まる「標準の運搬ルール」である。

- ユークリッド空間の類推: 我々が普段馴れ親しんでいる平らな空間 (ユークリッド空間) では、ベクトルを平行移動しても長さや角度は変わらないし、長方形の経路を一周しても元の位置に戻る (ねじれがない)。Levi-Civita 接続は、曲がった空間においても「長さ・角度を変えない (計量整合性)」かつ「余計な回転を加えない (ねじれなし)」という、ユークリッド空間の素直な性質を可能な限り残した接続である。
- 地図アプリの方位補正: 曲面 (地球) の上で地図を持って移動するとき、現地の地形 (計量) に合わせて地図の向きを適切に補正しないと、自分が向いている方向がわからなくなる。Levi-Civita 接続は、その土地の距離感 (計量) に従って、最も無理のない方法で「北」を維持し続けるナビゲーションシステムのようなものである。

数式の導出：計量保存からの逆算

なぜ Christoffel 記号 Γ_{ij}^k があのような複雑な式 ((1.21)) になるのか、その理由は「計量整合性」の条件から直接導かれる。

まず、計量整合条件 $\nabla g = 0$ 、すなわち $Z\langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle$ を考える。ここで基底ベクトル場 $\partial_i, \partial_j, \partial_k$ を代入し、添字をサイクリックに入れ替えた 3 つの式を用意する。

$$\partial_i g_{jk} = \langle \nabla_{\partial_i} \partial_j, \partial_k \rangle + \langle \partial_j, \nabla_{\partial_i} \partial_k \rangle = \Gamma_{ij,k} + \Gamma_{ik,j} \quad (1.22)$$

$$\partial_j g_{ki} = \langle \nabla_{\partial_j} \partial_k, \partial_i \rangle + \langle \partial_k, \nabla_{\partial_j} \partial_i \rangle = \Gamma_{jk,i} + \Gamma_{ji,k} \quad (1.23)$$

$$-\partial_k g_{ij} = -\langle \nabla_{\partial_k} \partial_i, \partial_j \rangle - \langle \partial_i, \nabla_{\partial_k} \partial_j \rangle = -\Gamma_{ki,j} - \Gamma_{kj,i} \quad (1.24)$$

これら 3 式を足し合わせる ((1.22) + (1.23) + (1.24))。ねじれがない ($\Gamma_{ij}^k = \Gamma_{ji}^k$ 、すなわち $\Gamma_{ij,k} = \Gamma_{ji,k}$) ことを使うと、項が綺麗に打ち消し合い、最後に残るのは $2\Gamma_{ij,k}$ だけとなる。

$$\partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} = 2\Gamma_{ij,k} \quad (1.25)$$

これを $\Gamma_{ij,k}$ について解けば、定理の式が得られる。つまり、この定義式は「計量を保存し、かつ対称であるような接続はこれしかない」という要請の結果として必然的に導かれるものである。

情報幾何学では、この $\nabla^{(0)}$ を中心 ($\alpha = 0$) として、非対称な接続の族 (α -接続) を扱うことになる。

第 2 章

統計的モデルの幾何学的構造

本章では、確率分布の空間に Riemann 計量（Fisher 情報量）とアファイン接続（ α -接続）を導入する。これらの幾何学量は、定義として与えられるものではなく、確率分布の構造から自然に導出されるものである。その導出過程を詳細に追うことで、情報幾何の物理的意味を理解する。

2.1 統計的モデル

\mathcal{X} を確率変数のとる値の空間（標本空間）、 μ を \mathcal{X} 上の測度とする。パラメータ $\xi = (\xi^1, \dots, \xi^n) \in \Xi \subset \mathbb{R}^n$ によって指定される確率密度関数 $p(x; \xi)$ の集合

$$S = \{p(x; \xi) \mid \xi \in \Xi\}$$

を n 次元統計的モデルと呼ぶ。対数尤度関数を $l(x; \xi) = \log p(x; \xi)$ とおく。各点 ξ における接空間 $T_\xi S$ は、スコア関数 $\partial_i l(x; \xi)$ によって張られるベクトル空間とみなせる。

2.2 Fisher 情報行列の導出

Fisher 情報量 g_{ij} は、「確率分布 $p(x; \xi)$ と $p(x; \xi + d\xi)$ の距離」を測る量として導かれる。

2.2.1 KL ダイバージェンスの 2 次近似

2 つの近接した確率分布 $p(x; \xi)$ と $p(x; \xi')$ の間の Kullback-Leibler (KL) ダイバージェンスを考える。

$$D_{KL}[p_\xi \| p_{\xi'}] = \int p(x; \xi) \log \frac{p(x; \xi)}{p(x; \xi')} d\mu(x) \quad (2.1)$$

ここで $\xi' = \xi + \Delta\xi$ とし、 $\Delta\xi$ が微小であるとして Taylor 展開を行う。まず、対数項を展開する。

$$\begin{aligned}\log \frac{p(x; \xi)}{p(x; \xi + \Delta\xi)} &= -\log \frac{p(x; \xi + \Delta\xi)}{p(x; \xi)} \\ &= -\log \left(1 + \frac{1}{p} \partial_i p \Delta\xi^i + \frac{1}{2p} \partial_i \partial_j p \Delta\xi^i \Delta\xi^j + \dots \right)\end{aligned}$$

$\log(1+t) \approx t - t^2/2$ を用いて整理すると、

$$\begin{aligned}&\approx -\left(\frac{\partial_i p}{p} \Delta\xi^i + \frac{\partial_i \partial_j p}{2p} \Delta\xi^i \Delta\xi^j \right) + \frac{1}{2} \left(\frac{\partial_i p}{p} \Delta\xi^i \right) \left(\frac{\partial_j p}{p} \Delta\xi^j \right) \\ &= -\partial_i l \Delta\xi^i - \frac{1}{2p} (\partial_i \partial_j p) \Delta\xi^i \Delta\xi^j + \frac{1}{2} (\partial_i l) (\partial_j l) \Delta\xi^i \Delta\xi^j\end{aligned}$$

これを期待値 (p で重み付けした積分) に入れる。

$$\begin{aligned}D_{KL} &\approx - \underbrace{\int p(\partial_i l) dx \Delta\xi^i}_{E[\partial_i l]=0} - \frac{1}{2} \underbrace{\int \partial_i \partial_j p dx \Delta\xi^i \Delta\xi^j}_{\partial_i \partial_j \int p=0} + \frac{1}{2} E[(\partial_i l)(\partial_j l)] \Delta\xi^i \Delta\xi^j \\ &= \frac{1}{2} g_{ij}(\xi) \Delta\xi^i \Delta\xi^j\end{aligned}\tag{2.2}$$

ここで、 g_{ij} は以下で定義される Fisher 情報量 である。

定義 2.1 (Fisher 情報計量).

$$g_{ij}(\xi) = E[\partial_i l(x; \xi) \partial_j l(x; \xi)]\tag{2.3}$$

すなわち、Fisher 情報量は KL ダイバージェンスの Hessian (2 次の係数) として自然に導かれる。

例 2.2 (正規分布モデルの Fisher 計量). 平均 μ 、標準偏差 σ の正規分布族 $N(\mu, \sigma^2)$ を「2次元の多様体」とみなして、Fisher 情報計量を計算する。

1. 対数尤度関数:

$$l(\mu, \sigma) = \log \left(\frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \right) = -\log \sigma - \frac{(x-\mu)^2}{2\sigma^2} + \text{const.}$$

2. スコア関数 (1 階微分) :

$$\partial_\mu l = \frac{x-\mu}{\sigma^2}, \quad \partial_\sigma l = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

3. Fisher 情報量 $g_{ij} = E[\partial_i l \partial_j l]$ の計算:

- $g_{\mu\mu} = E[(\partial_\mu l)^2] = \frac{E[(x-\mu)^2]}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$
- $g_{\sigma\sigma} = E[(\partial_\sigma l)^2] = E \left[\left(-\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3} \right)^2 \right] = \frac{2}{\sigma^2}$
- $g_{\mu\sigma} = 0$ (期待値をとると消える)

4. 計量行列 G :

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$$

【この計算が教えること】

- 直感的な距離: σ (標準偏差) が分母にある。
 - σ が小さい (尖った分布) ととき: 計量は大きくなる。少し μ を動かすだけで「別の分布」だとハッキリ識別できる (= 距離が遠い)。
 - σ が大きい (平べったい分布) ととき: 計量は小さくなる。 μ を動かしても分布はあまり変わらない (= 距離が近い)。

2.2.2 恒等式 $g_{ij} = -E[\partial_i \partial_j l]$ の証明

Proof. 全確率の保存則 $\int p(x; \xi) dx = 1$ を ξ^i で偏微分する。

$$\int \partial_i p dx = \int p(\partial_i l) dx = E[\partial_i l] = 0$$

さらにもう一度 ξ^j で偏微分する。積の微分公式 $\partial_j(p \partial_i l) = (\partial_j p)(\partial_i l) + p(\partial_j \partial_i l)$ より、

$$\begin{aligned} 0 &= \partial_j \int p(\partial_i l) dx = \int \partial_j(p \partial_i l) dx \\ &= \int (\partial_j p)(\partial_i l) dx + \int p(\partial_j \partial_i l) dx \\ &= \int p(\partial_j l)(\partial_i l) dx + E[\partial_j \partial_i l] \\ &= g_{ij}(\xi) + E[\partial_i \partial_j l] \end{aligned}$$

よって、 $g_{ij} = -E[\partial_i \partial_j l]$ が得られる。これは Fisher 情報が「対数尤度の曲率」であることを示す。 \square

2.3 α -接続の導出

Fisher 計量だけでは Levi-Civita 接続 $\nabla^{(0)}$ しか定まらないが、統計モデルには 3 次のモーメント構造が存在する。これを用いて α -接続を導出する。

2.3.1 α -表現と接続係数

確率密度関数 p を「どのスケールで見るか」によって、自然な平坦性 (直線の定義) が変わる。関数 $L^{(\alpha)}(p)$ を以下のように定義する (α -埋め込み)。

$$L^{(\alpha)}(p) = \begin{cases} \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}} & (\alpha \neq 1) \\ \log p & (\alpha = 1) \end{cases} \quad (2.4)$$

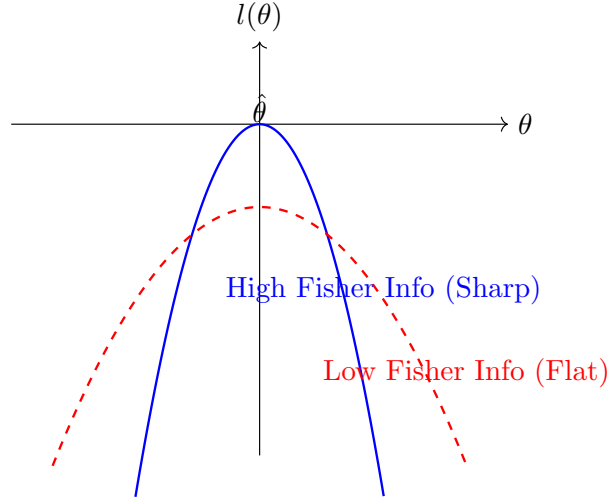


図 2.1 Fisher 情報量と対数尤度の曲率。尖った分布（青）は曲率が大きく、推定精度が高い（分散が小さい）。平坦な分布（赤）は曲率が小さく、推定精度が低い。

この関数の微分の期待値として、接続係数 $\Gamma_{ij,k}^{(\alpha)}$ を定義する。一般に、接続係数は「基底ベクトルの変化率」を計量で測ったものである。

$$\Gamma_{ij,k}^{(\alpha)} = E \left[\left(\partial_i \partial_j L^{(\alpha)} \right) \cdot \left(\partial_k L^{(-\alpha)} \right) \right] \quad (2.5)$$

（注：定義にはいくつかの流儀があるが、ここでは甘利の定義に整合するように記述する）
具体的に計算してみよう。 $\alpha \neq \pm 1$ のとき、

$$\begin{aligned} \partial_i L^{(\alpha)} &= p^{\frac{1-\alpha}{2}-1} \partial_i p = p^{-\frac{1+\alpha}{2}} \partial_i p \\ \partial_i \partial_j L^{(\alpha)} &= -\frac{1+\alpha}{2} p^{-\frac{3+\alpha}{2}} (\partial_i p)(\partial_j p) + p^{-\frac{1+\alpha}{2}} \partial_i \partial_j p \end{aligned}$$

一方、双対側 $(-\alpha)$ の 1 階微分は

$$\partial_k L^{(-\alpha)} = p^{-\frac{1-\alpha}{2}} \partial_k p$$

これらを掛け合わせて積分（期待値）をとる。

$$\begin{aligned} \Gamma_{ij,k}^{(\alpha)} &= \int \left(-\frac{1+\alpha}{2} p^{-\frac{3+\alpha}{2}} (\partial_i p)(\partial_j p) + p^{-\frac{1+\alpha}{2}} \partial_i \partial_j p \right) \left(p^{-\frac{1-\alpha}{2}} \partial_k p \right) dx \\ &= \int \left(-\frac{1+\alpha}{2} \frac{1}{p^2} (\partial_i p)(\partial_j p)(\partial_k p) + \frac{1}{p} (\partial_i \partial_j p)(\partial_k p) \right) dx \end{aligned}$$

ここで $\partial_i p = p \partial_i l$ および $\partial_i \partial_j p = p(\partial_i \partial_j l + \partial_i l \partial_j l)$ を代入する。

$$\begin{aligned} \Gamma_{ij,k}^{(\alpha)} &= \int \left(-\frac{1+\alpha}{2} p(\partial_i l)(\partial_j l)(\partial_k l) + (\partial_i \partial_j l + \partial_i l \partial_j l)(\partial_k l) p \right) dx \\ &= -\frac{1+\alpha}{2} T_{ijk} + E[(\partial_i \partial_j l)(\partial_k l)] + T_{ijk} \\ &= E[(\partial_i \partial_j l)(\partial_k l)] + \frac{1-\alpha}{2} T_{ijk} \end{aligned}$$

ゆえに、 α -接続の期待値表現は以下ようになる。

$$\Gamma_{ij,k}^{(\alpha)} = E[(\partial_i \partial_j l)(\partial_k l)] + \frac{1-\alpha}{2} T_{ijk} \quad (2.6)$$

ここで $T_{ijk} = E[(\partial_i l)(\partial_j l)(\partial_k l)]$ は 3 次の歪度テンソルである。

2.3.2 Levi-Civita 接続との関係

Levi-Civita 接続 ($\alpha = 0$) の場合、上式で $\alpha = 0$ とおくと

$$\Gamma_{ij,k}^{(0)} = E[(\partial_i \partial_j l)(\partial_k l)] + \frac{1}{2} T_{ijk}$$

これを用いると、任意の α に対する接続は次のように書ける。

定理 2.3 (α -接続の構造).

$$\Gamma_{ij,k}^{(\alpha)} = \Gamma_{ij,k}^{(0)} - \frac{\alpha}{2} T_{ijk} \quad (2.7)$$

この式は、 α -接続が Levi-Civita 接続を T_{ijk} の方向に「歪めた」ものであることを明示している。この「歪み」こそが、指数型分布族 ($\alpha = 1$) や混合分布族 ($\alpha = -1$) の平坦性を生み出す源泉である。

直感： α の値と分布族の関係

α の値には、確率分布を「どう混ぜるか（どう補間するか）」という幾何学的な意味が直接対応している。一言でいうと、以下の対応関係がある。

- $\alpha = 1$ (**e-接続**) : **対数 (log) **の世界で真っ直ぐ混ぜる \iff 指数型分布族
- $\alpha = -1$ (**m-接続**) : **確率 (probability) **の世界で真っ直ぐ混ぜる \iff 混合分布族

以前述べたように、 α は「座標の曲がり具合」を補正する項である。 $\alpha = 1$ の幾何学では、指数型分布族が持つ「曲がり」がちょうど打ち消されて平坦になる。 $\alpha = -1$ の幾何学では、混合分布族が持つ「曲がり」がちょうど打ち消されて平坦になる。

これがなぜなのか、具体的な「直線の引き方」の違いから見ていく。

1. $\alpha = -1$ と混合分布族 (m-接続)

確率分布 $p_1(x)$ と $p_2(x)$ があるとき、一番素直な混ぜ方は「足して 2 で割る」ことである。一般に、 $t \in [0, 1]$ として

$$p(x; t) = (1 - t)p_1(x) + tp_2(x)$$

と書ける分布の系列を考える。これは**確率の線形結合**であり、この混ぜ方で作られる直線を「m-測地線 (mixture geodesic)」と呼ぶ。

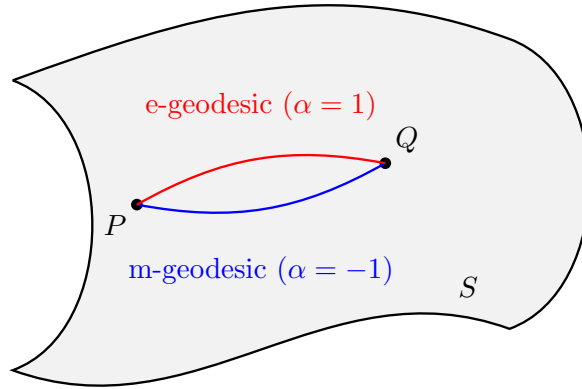


図 2.2 α -接続による測地線の違い。 $\alpha = -1$ (混合) では確率の線形結合が真っ直ぐに見え、 $\alpha = 1$ (指数) では対数尤度の線形結合が真っ直ぐに見える。

- **混合分布族 (Mixture Family)** : この「線形結合」で作られる分布の集まり (例 : 正規分布の重ね合わせなど)。
- **なぜ $\alpha = -1$ なのか?** : この分布族の座標系 (混合比率など) で見ると、 $\alpha = -1$ の接続係数が消失し、空間が平坦に見えるからである。つまり、「確率そのものを真っ直ぐとみなす幾何学」が $\alpha = -1$ である。

2. $\alpha = 1$ と指数型分布族 (e-接続)

一方、確率分布には「掛け算してルートを取る」ような混ぜ方もある。

$$p(x; t) \propto p_1(x)^{1-t} p_2(x)^t$$

両辺の対数 (log) をとると、

$$\log p(x; t) = (1-t) \log p_1(x) + t \log p_2(x) + (\text{定数})$$

となり、これは**対数尤度の線形結合**になっている。この混ぜ方で作られる直線を「e-測地線 (exponential geodesic)」と呼ぶ。

- **指数型分布族 (Exponential Family)** : $\log p$ がパラメータの線形結合で書ける分布の集まり (例 : 正規分布、ポアソン分布、ベルヌーイ分布など)。
- **なぜ $\alpha = 1$ なのか?** : この分布族の自然パラメータ (θ) で見ると、 $\alpha = 1$ の接続係数がゼロになる。つまり、「対数を真っ直ぐとみなす幾何学」が $\alpha = 1$ である。

3. α の意味： α -表現

一般に、確率分布を $f(p)$ という関数で変換してから「真っ直ぐ」とみなすことを考えたとき、 α -接続は以下の変換（ α -表現）に対応している。

$$L^{(\alpha)}(p) = \begin{cases} \log p & (\alpha = 1) \\ p & (\alpha = -1) \\ p^{(1-\alpha)/2} & (\text{その他}) \end{cases}$$

- $\alpha \rightarrow 1$ で対数 $\log p$ に近づく（指数型）
- $\alpha \rightarrow -1$ で確率 p そのものに近づく（混合型）
- $\alpha = 0$ は $p^{1/2} = \sqrt{p}$ に対応し、球面幾何（Fisher 計量そのもの）を表す。

統計学でこれらが重要なのは、多くの主要な確率モデルが「指数型分布族」か「混合モデル」のいずれかで記述されるため、それらを自然に扱うための幾何学が必要とされたからである。

具体例：ベルヌーイ分布の幾何

最も単純な例として、コイン投げ（ベルヌーイ分布）のモデル $S = \{p(x; \eta) = \eta^x(1 - \eta)^{1-x} \mid 0 < \eta < 1\}$ を考える。ここで η は「表が出る確率」そのものである（混合パラメータ）。

1. **Fisher 計量の計算**: 対数尤度 $l = x \log \eta + (1 - x) \log(1 - \eta)$ より、

$$\partial_\eta l = \frac{x}{\eta} - \frac{1-x}{1-\eta} = \frac{x-\eta}{\eta(1-\eta)}$$

これより $g_{11} = E[(\partial_\eta l)^2] = \frac{1}{\eta(1-\eta)}$ となる。

2. **3 次のテンソルの計算**:

$$T_{111} = E[(\partial_\eta l)^3] = \frac{E[(x-\eta)^3]}{[\eta(1-\eta)]^3} = \frac{1-2\eta}{\eta^2(1-\eta)^2}$$

（ここで中心 3 次モーメント $E[(x-\eta)^3] = \eta(1-\eta)(1-2\eta)$ を用いた）

3. **α -接続係数の計算**: Levi-Civita 接続係数 $\Gamma_{11,1}^{(0)} = \frac{1}{2} \partial_\eta g_{11} = -\frac{1-2\eta}{2\eta^2(1-\eta)^2}$ より、

$$\begin{aligned} \Gamma_{11,1}^{(\alpha)} &= \Gamma_{11,1}^{(0)} - \frac{\alpha}{2} T_{111} \\ &= -\frac{1-2\eta}{2\eta^2(1-\eta)^2} (1 + \alpha) \end{aligned}$$

これを計量で引き上げた $\Gamma_{11}^{1(\alpha)} = g^{11} \Gamma_{11,1}^{(\alpha)}$ は次のようになる。

$$\Gamma_{11}^{1(\alpha)} = -\frac{1-2\eta}{2\eta(1-\eta)} (1 + \alpha)$$

この結果から、以下のことがわかる。

- $\alpha = -1$ (m-接続) のとき: $\Gamma_{11}^{1(-1)} = 0$ となる。すなわち、座標 η (確率そのもの) はこの接続に関して**アファイン座標 (真っ直ぐな座標)**である。これは「確率は混合側 ($\alpha = -1$) で見ると真っ直ぐ」という直感と一致する。
- $\alpha = 1$ (e-接続) のとき: $\Gamma \neq 0$ であり、 η 座標は曲がっている。逆に、ロジット変換 $\theta = \log \frac{\eta}{1-\eta}$ (自然パラメータ) を行くと、この座標系では $\Gamma^{(1)}$ が消失し、アファイン座標となることが確かめられる。

定理 2.4 (双対性). α -接続と $-\alpha$ -接続は、Fisher 計量 g に関して互いに双対である。すなわち、任意のベクトル場 X, Y, Z に対して

$$Zg(X, Y) = g(\nabla_Z^{(\alpha)} X, Y) + g(X, \nabla_Z^{(-\alpha)} Y)$$

が成立する。

この定理は、統計的モデルにおいて α と $-\alpha$ の幾何学が表裏一体であることを示している。特に重要なのは $\alpha = 1$ (e-接続) と $\alpha = -1$ (m-接続) の対である。

【コラム】 α -接続の起源と必然性

なぜこれほど複雑に見える α -接続が導入されたのか、その背景には強い数学的必然性が存在する。

1. 歴史的背景：チェンツォフの定理

1980 年代に甘利俊一が情報幾何学を構築する際、拠り所となったのはロシアの数学者 N.N. チェンツォフ (Chentsov) による以下の定理である。

定理 2.5 (Chentsov の定理). 確率分布の空間において、マルコフ埋め込み (十分統計量による縮約など、情報の加工) に対して不変な幾何構造は、本質的に以下のものに限られる (一意性)。

1. 計量: Fisher 情報計量 g (の定数倍)
2. 接続: α -接続 $\nabla^{(\alpha)}$ の族 (およびその定数倍)

つまり、統計学的に意味のある「自然な接続」は、Levi-Civita 接続だけでなく、パラメータ α を持つこの一族しか存在しないことが証明されている。この定理が、情報幾何学における α -接続の正当性を保証している。

2. 定義の導出：確率変数の α -表現

α -接続の定義式 (期待値表現) は、恣意的に決められたものではなく、**「確率変数の α -表現」**の微分から自然に導かれるものである。

確率密度関数 $p(x; \xi)$ に対して、以下のような変換 $L^{(\alpha)}$ を考える。

$$L^{(\alpha)}(x; \xi) = \begin{cases} \frac{2}{1-\alpha} p(x; \xi)^{\frac{1-\alpha}{2}} & (\alpha \neq 1) \\ \log p(x; \xi) & (\alpha = 1) \end{cases}$$

これをパラメータ ξ で 2 階微分したものの期待値をとることで、接続係数が定義される。

$$\Gamma_{ij,k}^{(\alpha)} = E \left[\frac{\partial^2 L^{(\alpha)}}{\partial \xi^i \partial \xi^j} \cdot \frac{\partial L^{(-\alpha)}}{\partial \xi^k} \right]$$

これを計算すると、定義式 (2.6) が得られる。特に $\alpha = 1$ のときは対数尤度 $\log p$ の微分（指数型分布族の自然な構造）に対応し、 $\alpha = -1$ のときは密度関数 p そのものの微分（混合分布族の自然な構造）に対応していることがわかる。

3. 双対性の要請

α -接続の係数が現在の形 ($\pm \frac{\alpha}{2} T$) に落ち着いた決定的な理由は、**「双対性 (Duality)」** の発見にある。甘利は、統計的推論において「指数型分布族 (e-接続)」と「混合分布族 (m-接続)」が互いに逆の役割を果たしていることに着目し、 α と $-\alpha$ が Fisher 計量に関して互いに双対となるように理論を整備した。この美しい対称性 ($\alpha \leftrightarrow -\alpha$) が、情報幾何学の体系を支える柱となっている。

第 3 章

双対接続の理論

前章では、統計的モデル上に計量 g と 1 パラメータ接続族 $\nabla^{(\alpha)}$ を導入した。本章では、これらの幾何学的構造が織りなす「双対性 (duality)」に焦点を当てる。特に、双対平坦空間におけるピタゴラスの定理や射影定理は、統計的推論（最尤推定や情報量最小化）の幾何学的実体である。

3.1 接続の双対性

Riemann 多様体 (\mathcal{M}, g) 上の 2 つのアファイン接続 ∇ と ∇^* が双対であるという概念を厳密に定式化する。

導入：なぜ 2 つの接続が必要なのか？

通常の Riemann 幾何学では、計量 g と整合する接続はただ一つ (Levi-Civita 接続) である。しかし、統計的モデルにおける距離 (ダイバージェンス) は非対称性を持ち、1 つの接続だけではその構造を記述しきれない。

双対接続のアイデアは、「歪みを互いに打ち消し合う 2 つの接続 (∇, ∇^*) をペアで考える」ことにある。定義式 (3.1) は、内積の微分 (積の微分公式) を ∇ と ∇^* が分担して担うことを意味している。片方の接続が空間を「右に曲げる」なら、もう片方は「左に曲げる」ことで、トータルとしての整合性 (Goddard 条件) を保つのである。

定義 3.1 (双対接続). 多様体 \mathcal{M} 上の計量 g と 2 つのアファイン接続 ∇, ∇^* の組 $(\mathcal{M}, g, \nabla, \nabla^*)$ が与えられたとき、これらが g に関して互いに双対 (dual) であるとは、任意のベクトル場 $X, Y, Z \in \mathfrak{X}(\mathcal{M})$ に対して以下の Goddard 条件が成立することである。

$$Z\langle X, Y \rangle_g = \langle \nabla_Z X, Y \rangle_g + \langle X, \nabla_Z^* Y \rangle_g \quad (3.1)$$

ここで $\langle \cdot, \cdot \rangle_g$ は計量 g による内積を表す。

この定義は、内積の微分 (Leibniz 則) が ∇ と ∇^* によって分配されることを意味する。もし $\nabla = \nabla^*$ ならば、これは Levi-Civita 接続の計量整合条件 $\nabla g = 0$ に他ならない。

【コラム】なぜ2つの接続が必要なのか？（双対性の直感）

通常の幾何学（ユークリッド空間や球面の幾何）は、「対称な世界」である。A 地点から B 地点への距離は、行きも帰りも同じである。だから、「真っ直ぐな道（接続）」も1種類だけで済む。

しかし、統計モデルの世界は**「非対称な世界」**である。

1. 登山の例え：行きと帰りで「真っ直ぐ」が違う

急な山道をイメージしてほしい。

- **登り (∇) **: 「足元をしっかりと踏みしめて、最短距離で登るルート」が最適である。
- **下り (∇^*) **: 「勢いをつけて駆け下りるルート」が最適かもしれない。

統計モデルにおいて、2つの点は「真の分布 P 」と「モデル分布 Q 」である。KL ダイバージェンス（距離のようなもの）は非対称である。

$$D(P\|Q) \neq D(Q\|P)$$

- P から Q を見る視点 (∇)
- Q から P を見る視点 (∇^*)

この「行きと帰りの非対称性」を記述するために、2つの異なる「真っ直ぐの基準（接続）」を用意して、ペアで使う必要がある。

2. 積の微分の一般化

定義式 (3.1) をよく見ると、

$$Z\langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z^* Y \rangle$$

これは、高校数学の**積の微分公式 $(xy)' = x'y + xy'$

- 普通の幾何学（Levi-Civita 接続）では、 $\nabla = \nabla^*$ なので、

$$(xy)' = x'y + xy'$$

という「公平な分配」が行われる。

- 情報幾何学（双対接続）では、

$$(xy)' = (\text{前半分を微分})y + x(\text{後半分を微分})$$

という役割分担を、 ∇ と ∇^* で分け合っている。

片方 (∇) が歪んでいる分、もう片方 (∇^*) が逆方向に歪んでいれば、2つ合わせて「トータルでは整合性が取れる（内積を保存する）」という仕組みである。

「空間が非対称だから、1つの接続だけでは無理がある。歪みを打ち消し合う『2つの接続』をペアで使うことで、初めて計算が綺麗になる。」

これが双対接続を導入する理由である。

命題 3.2 (成分表示). 局所座標系における接続係数をそれぞれ $\Gamma_{ij}^k, \Gamma_{ij}^{*k}$ とすると、双対性の条件は以下と同値である。

$$\partial_k g_{ij} = \Gamma_{kij} + \Gamma_{kji}^* \quad (3.2)$$

ここで $\Gamma_{kij} = g_{jm} \Gamma_{ki}^m$ である。

命題 3.2 の証明. 双対接続の定義式 (3.1) において、ベクトル場として自然基底 $X = \partial_i, Y = \partial_j, Z = \partial_k$ を選ぶ。左辺は計量テンソルの偏微分となる。

$$\text{左辺} = \partial_k \langle \partial_i, \partial_j \rangle_g = \partial_k g_{ij} \quad (3.3)$$

右辺の第1項は、接続係数の定義 $\nabla_{\partial_k} \partial_i = \Gamma_{ki}^m \partial_m$ を用いて、

$$\langle \nabla_{\partial_k} \partial_i, \partial_j \rangle_g = \langle \Gamma_{ki}^m \partial_m, \partial_j \rangle_g = \Gamma_{ki}^m g_{mj} = \Gamma_{kij} \quad (3.4)$$

となる。同様に第2項は、双対接続係数を用いて

$$\langle \partial_i, \nabla_{\partial_k}^* \partial_j \rangle_g = \langle \partial_i, \Gamma_{kj}^{*m} \partial_m \rangle_g = \Gamma_{kj}^{*m} g_{im} = \Gamma_{kji}^* \quad (3.5)$$

となる。これらを合わせることで成分表示の式 (3.2) が得られる。 \square

定理 3.3 (曲率の双対性). 双対接続 ∇, ∇^* の Riemann-Christoffel 曲率テンソルをそれぞれ R, R^* とすると、以下の関係が成り立つ。

$$\langle R(X, Y)Z, W \rangle_g = -\langle Z, R^*(X, Y)W \rangle_g \quad (3.6)$$

これより、 ∇ が平坦 ($R = 0$) であることと、 ∇^* が平坦 ($R^* = 0$) であることは同値である。

定理 3.3 の証明. 簡単のため $[X, Y] = 0$ と仮定して計算する。双対性の定義 $\langle \nabla_A B, C \rangle = A \langle B, C \rangle - \langle B, \nabla_A^* C \rangle$ を繰り返し用いて、 ∇ を ∇^* へと移し替える (部分積分に相当する操作)。

$$\begin{aligned} \langle \nabla_X \nabla_Y Z, W \rangle &= X \langle \nabla_Y Z, W \rangle - \langle \nabla_Y Z, \nabla_X^* W \rangle \\ &= X(Y \langle Z, W \rangle - \langle Z, \nabla_Y^* W \rangle) - (Y \langle Z, \nabla_X^* W \rangle - \langle Z, \nabla_Y^* \nabla_X^* W \rangle) \end{aligned} \quad (3.7)$$

同様に X, Y を入れ替えた式を作り、辺々を引くと、微分作用素 X, Y が掛かっている項はすべて相殺される ($XY = YX$ より)。残る項は

$$\langle (\nabla_X \nabla_Y - \nabla_Y \nabla_X)Z, W \rangle = \langle Z, (\nabla_Y^* \nabla_X^* - \nabla_X^* \nabla_Y^*)W \rangle \quad (3.8)$$

となる。右辺の括弧内は $-R^*(X, Y)W$ であるため、

$$\langle R(X, Y)Z, W \rangle = -\langle Z, R^*(X, Y)W \rangle \quad (3.9)$$

が成り立つ。 \square

3.2 双対平坦空間

双対接続 ∇, ∇^* が共に平坦であるとき、空間構造は極めて単純かつ強力になる。

定義 3.4 (双対平坦空間). 組 $(\mathcal{M}, g, \nabla, \nabla^*)$ において ∇ と ∇^* が双対であり、かつ ∇ (したがって ∇^*) の曲率テンソルが恒等的に 0 であるとき、 \mathcal{M} を双対平坦空間 (dually flat space) と呼ぶ。

双対平坦空間には、それぞれの接続に対応する特別な座標系が存在する。

定理 3.5 (双対アファイン座標系). \mathcal{M} が双対平坦であるならば、以下の性質を持つ 2 つの座標系 $\theta = (\theta^i)$ と $\eta = (\eta_i)$ が局所的に存在する。

1. θ は ∇ -アファイン座標系である。すなわち、 $\Gamma_{ij}^k(\theta) \equiv 0$ 。
2. η は ∇^* -アファイン座標系である。すなわち、 $\Gamma_{ij}^{*k}(\eta) \equiv 0$ 。
3. 両者は互いに Legendre 変換 (双対変換) の関係にある。すなわち、ある凸関数 $\psi(\theta)$ と $\varphi(\eta)$ が存在して、以下が成り立つ。

$$\eta_i = \frac{\partial \psi(\theta)}{\partial \theta^i}, \quad \theta^i = \frac{\partial \varphi(\eta)}{\partial \eta_i} \quad (3.10)$$

$$\psi(\theta) + \varphi(\eta) - \theta \cdot \eta = 0 \quad (3.11)$$

ここで θ を ∇ -アファイン座標、 η を ∇^* -アファイン座標 (または双対アファイン座標) と呼ぶ。また、計量 g_{ij} はポテンシャル関数 ψ の Hessian で与えられる。

$$g_{ij}(\theta) = \frac{\partial \eta_i}{\partial \theta^j} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} \quad (3.12)$$

逆計量 g^{ij} は双対ポテンシャル φ の Hessian となる。

3.3 ダイバージェンス (正準ダイバージェンス)

双対平坦空間上の 2 点間の「距離の二乗」に相当する非対称な量を、ポテンシャル関数を用いて自然に定義できる。

定義 3.6 (Bregman ダイバージェンス / 正準ダイバージェンス). 双対平坦空間 \mathcal{M} 上の 2 点 p, q に対し、以下の実数値関数 $D(p||q)$ を正準ダイバージェンス (canonical divergence) と呼ぶ。 p の θ 座標を $\theta(p)$ 、 q の θ 座標を $\theta(q)$ とするとき、

$$D(p||q) := \psi(\theta(p)) + \varphi(\eta(q)) - \theta(p) \cdot \eta(q) \quad (3.13)$$

この $D(p||q)$ は以下の性質を満たす。

1. 非負性: $D(p||q) \geq 0$ であり、等号成立は $p = q$ のときに限る (ψ の凸性による)。

2. 非対称性: 一般に $D(p\|q) \neq D(q\|p)$ 。
3. 局所距離: p と q が無限小に近いとき ($q = p + dp$)、

$$D(p\|p + dp) = \frac{1}{2} g_{ij}(p) d\theta^i d\theta^j + O(|d\theta|^3) \quad (3.14)$$

となり、計量 g を誘導する。

Kullback-Leibler ダイバージェンスは、指数型分布族（e-接続と m-接続に関して双対平坦）における正準ダイバージェンスの具体例である。

3.4 一般化されたピタゴラスの定理

双対平坦空間の幾何学的有用性を決定づけるのが、以下の定理である。これはユークリッド空間におけるピタゴラスの定理の拡張であり、射影（最適近似）の一意性と計算可能性を保証する。

定理 3.7 (一般化ピタゴラスの定理). 双対平坦空間 \mathcal{M} 内の 3 点 p, q, r を考える。点 p と q を結ぶ ∇ -測地線 (θ -座標で直線) と、点 q と r を結ぶ ∇^* -測地線 (η -座標で直線) が点 q において直交するとき、以下の等式が成立する。

$$D(p\|r) = D(p\|q) + D(q\|r) \quad (3.15)$$

Proof. p, q, r の座標をそれぞれ $\theta_p, \theta_q, \theta_r$ および η_p, η_q, η_r とする。定義より、

$$\begin{aligned} D(p\|q) + D(q\|r) - D(p\|r) &= (\psi_p + \varphi_q - \theta_p \cdot \eta_q) + (\psi_q + \varphi_r - \theta_q \cdot \eta_r) \\ &\quad - (\psi_p + \varphi_r - \theta_p \cdot \eta_r) \\ &= \psi_q + \varphi_q - \theta_p \cdot \eta_q - \theta_q \cdot \eta_r + \theta_p \cdot \eta_r \\ &= \theta_q \cdot \eta_q - \theta_p \cdot \eta_q - \theta_q \cdot \eta_r + \theta_p \cdot \eta_r \\ &= (\theta_q - \theta_p) \cdot (\eta_q - \eta_r) \end{aligned}$$

ここで、 ∇ -測地線 $\gamma(t)$ は $\theta(t)$ が線形、 ∇^* -測地線は $\eta(t)$ が線形である。接ベクトルはそれぞれ $\Delta\theta = \theta_q - \theta_p$ (反変) および $\Delta\eta = \eta_r - \eta_q$ (共変) に対応する。直交条件 $\langle \Delta\theta, \Delta\eta \rangle = 0$ より、上式は 0 となる。 \square

この定理は、情報射影（m-射影）を用いた最尤推定の解釈や、EM アルゴリズムの収束証明において中心的な役割を果たす。

3.5 双対葉層構造

双対平坦空間におけるピタゴラスの定理は、一点から一点への射影（点推定）だけでなく、より一般的な部分多様体への射影や、空間全体の分解構造（葉層構造）へと拡張される。これは、統計的推論において「局外パラメータ（nuisance parameter）」を含むモデルや、セミパラメトリックモデルを扱う際の幾何学的基礎となる。

3.5.1 葉層構造と混合座標系

n 次元多様体 \mathcal{M} 上の双対アファイン座標系 (θ, η) を考える。座標添字を 2 つのグループ $I = \{1, \dots, k\}$ と $J = \{k+1, \dots, n\}$ に分割する。このとき、パラメータを以下のように混合させた新しい座標系 ζ を導入する。

$$\zeta = (\theta^1, \dots, \theta^k, \eta_{k+1}, \dots, \eta_n) \quad (3.16)$$

この座標系 ζ は、以下の 2 つの部分多様体族（葉層, foliation）を自然に定義する。

- **e-葉層 $\mathcal{F}^{(e)}$ (η -固定) **: $\eta_{k+1}, \dots, \eta_n$ を定数に固定し、 $\theta^1, \dots, \theta^k$ だけを動かしたときに描かれる k 次元部分多様体の族。これらは ∇ に関して（e-接続に関して）平坦な部分多様体となる。

$$M_c^{(e)} = \{p \in \mathcal{M} \mid \eta_j(p) = c_j \ (\forall j \in J)\}$$

- **m-葉層 $\mathcal{F}^{(m)}$ (θ -固定) **: $\theta^1, \dots, \theta^k$ を定数に固定し、 $\eta_{k+1}, \dots, \eta_n$ だけを動かしたときに描かれる $(n-k)$ 次元部分多様体の族。これらは ∇^* に関して（m-接続に関して）平坦な部分多様体となる。

$$M_d^{(m)} = \{p \in \mathcal{M} \mid \theta^i(p) = d^i \ (\forall i \in I)\}$$

3.5.2 大域的な直交分解

重要な事実は、双対平坦空間において、これらの葉層が互いに**直交**することである。接ベクトル空間 $T_p\mathcal{M}$ の基底において、 $\mathcal{F}^{(e)}$ の接ベクトル（ θ^i 方向の微分）と、 $\mathcal{F}^{(m)}$ の接ベクトル（ η_j 方向の微分）の内積をとると、

$$\left\langle \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j} \right\rangle = \delta_i^j \quad (3.17)$$

であるため、 $i \in I, j \in J$ ならば内積は常に 0 となる。すなわち、多様体全体が「直交する 2 つの網目」によって綺麗に分解されている。これを**双対葉層構造（dual foliation structure）**と呼ぶ。

統計学的な文脈では、

- 関心のあるパラメータ θ^I （e-葉層上の移動）
- 局外パラメータ η_J （m-葉層上の移動）

が直交するということは、推定量が互いに無相関（漸近独立）であることを意味し、局外パラメータの影響を受けずに推論が可能であることを幾何学的に保証する（直交性による情報の分離）。

第 4 章

統計的推論の微分幾何

前章までの議論により、統計的モデルは双対平坦構造を持つ多様体として捉えられることが分かった。本章では、パラメータ推定や検定といった統計的推論の問題を、この幾何学的空間内での「射影 (Projection)」および「曲率 (Curvature)」の問題として再定式化する。

4.1 指数型分布族の幾何構造

統計学において最も重要なクラスである指数型分布族は、幾何学的には「双対平坦空間」の標準モデルである。

定義 4.1 (指数型分布族). 確率密度関数が以下の形式で書ける分布族 S_E を指数型分布族 (exponential family) という。

$$p(x; \boldsymbol{\theta}) = \exp \left(C(x) + \theta^i F_i(x) - \psi(\boldsymbol{\theta}) \right) \quad (4.1)$$

ここで $\boldsymbol{\theta} = (\theta^1, \dots, \theta^n)$ を自然パラメータ (natural parameter)、 $\psi(\boldsymbol{\theta})$ をキュムラント生成関数 (ポテンシャル関数) と呼ぶ。

命題 4.2 (e-平坦性と m-平坦性). 指数型分布族 S_E は、e-接続 $\nabla^{(1)}$ に関して平坦であり、 $\boldsymbol{\theta}$ はその e-アファイン座標系となる。また、双対座標 $\boldsymbol{\eta} = (\eta_i)$ は期待値パラメータ (expectation parameter) で与えられる。

$$\eta_i = E[F_i(x)] = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta^i} \quad (4.2)$$

S_E は m-接続 $\nabla^{(-1)}$ に関しても平坦であり、 (η_i) は m-アファイン座標系となる。ゆえに指数型分布族は $(\alpha = \pm 1)$ の意味で双対平坦空間をなす。

この空間の正準ダイバージェンスは Kullback-Leibler 情報量 となる。

$$D(p_{\boldsymbol{\theta}} \| p_{\boldsymbol{\theta}'}) = \psi(\boldsymbol{\theta}) + \varphi(\boldsymbol{\eta}') - \boldsymbol{\theta} \cdot \boldsymbol{\eta}' = \text{KL}(p_{\boldsymbol{\theta}} \| p_{\boldsymbol{\theta}'}) \quad (4.3)$$

4.2 最尤推定の幾何学的解釈 (m-射影)

経験分布（データから得られる分布）からモデル多様体への「あてはめ」である最尤推定法は、幾何学的には「m-射影」として理解される。

定義 4.3 (m-射影). 多様体 \mathcal{M} 上の点 p から部分多様体 $S \subset \mathcal{M}$ への m-射影 \hat{p} とは、m-測地線を用いて p から S へ下ろした垂線の足、すなわち以下の条件を満たす点である。

1. $\hat{p} \in S$
2. p と \hat{p} を結ぶ m-測地線が、 S に対して直交する (Fisher 計量に関して)。

双対平坦空間においては、これは正準ダイバージェンスの最小化点と一致する (ピタゴラスの定理より)。

$$\hat{p} = \arg \min_{q \in S} D(p \| q) \quad (4.4)$$

定理 4.4 (最尤推定の幾何). N 個の独立な観測データ x_1, \dots, x_N が得られたとする。経験分布 $\hat{p}_{emp}(x) = \frac{1}{N} \sum_{k=1}^N \delta(x - x_k)$ から統計モデル $S = \{p(x; \xi)\}$ への KL ダイバージェンス最小化は、最尤推定 (MLE) と等価である。

$$\hat{\xi}_{MLE} = \arg \max_{\xi} \sum_{k=1}^N \log p(x_k; \xi) \quad (4.5)$$

$$= \arg \min_{\xi} \text{KL}(\hat{p}_{emp} \| p_{\xi}) \quad (4.6)$$

幾何学的には、これは全確率分布の空間 (ノンパラメトリックな空間) にある経験分布 \hat{p}_{emp} から、モデル多様体 S への **m-射影** である。

4.3 曲指数型分布族と埋め込み曲率

一般の統計モデルは必ずしも平坦ではない。指数型分布族 S_E の部分多様体として定義されるモデル (曲指数型分布族) を考える。

定義 4.5 (曲指数型分布族). n 次元指数型分布族 S_E の内部に、 $m (< n)$ 個のパラメータ $\mathbf{u} = (u^1, \dots, u^m)$ で滑らかに埋め込まれた部分多様体 $M = \{p(x; \boldsymbol{\theta}(\mathbf{u}))\}$ を曲指数型分布族 (curved exponential family) と呼ぶ。

M は S_E から Fisher 計量と α -接続を誘導される。 M 自体は一般に平坦ではない。特に重要なのは、**e-埋め込み曲率 (exponential embedding curvature)** である。

定理 4.6 (推定量の一致性と曲率). モデル M が S_E 内で「e-平坦 (e-接続に関して測地線)」であるならば、最尤推定量は十分統計量 (sufficient statistic) の関数として陽に書ける。逆に、 M が e-平坦でない (e-埋め込み曲率を持つ) 場合、推定量は反復計算を必

要とし、情報損失 (information loss) が発生する。

4.4 推定の高次漸近理論

サンプル数 N が有限であるとき、推定量の良さ (有効性) は Fisher 情報量だけでは決まらない。漸近展開の第 2 項以降に現れる「高次の情報量」は、モデルの埋め込み曲率によって支配される。

推定量 $\hat{\mathbf{u}}$ の平均二乗誤差行列 $V[\hat{\mathbf{u}}]$ を $1/N$ のオーダーで展開する。

$$V[\hat{\mathbf{u}}] = \frac{1}{N} G^{-1}(\mathbf{u}) + \frac{1}{N^2} \left(\Gamma^{(m)2} + H_M^{(e)2} \right) + o(N^{-2}) \quad (4.7)$$

ここで、

- G^{-1} : Cram r-Rao の下界 (逆 Fisher 情報行列)。
- $\Gamma^{(m)2}$: パラメータ空間の m -接続係数に由来する項 (パラメータの取り方に依存)。
- $H_M^{(e)2}$: モデル M の e -埋め込み曲率 (多重曲率) の二乗ノルム。

定理 4.7 (漸近有効性の幾何). 1. 一致推定量の中で、1 次の項 ($1/N$) を最小化するのは最尤推定量 (およびそれに漸近等価な推定量) である。これらは有効推定量 (efficient estimator) と呼ばれる。2. 有効推定量の中で、2 次の項 ($1/N^2$) における情報損失 ΔI を最小化するものは、幾何学的には「バイアス補正された最尤推定量」である。3. その際の情報損失量は、モデルの e -埋め込み曲率 $H^{(e)}$ によって決定的かつ不可避に定まる。

$$\Delta I = \text{Tr} \left((H_M^{(e)})^2 \right) \quad (4.8)$$

すなわち、統計的推論の「難しさ」や「情報の失われやすさ」は、モデル多様体が確率分布の大空間の中でどのように「 e -接続の意味で曲がっているか」によって完全に幾何学的に記述される。

4.5 統計物理学と大偏差理論

情報幾何学の双対構造は、統計物理学における「最大エントロピー原理」や「平均場近似」といった手法に、明確な幾何学的意味を与える。さらには、確率変数の極限定理 (大偏差原理) と密接に関連している。

4.5.1 最大エントロピー原理と平均場近似

統計物理学における 2 つの主要なアプローチは、幾何学的には「2 つの射影」として統一的に理解できる。

1. **最大エントロピー原理 (MaxEnt)**: 観測された期待値 (エネルギーなど) η_{obs}

と整合する分布の中で、エントロピー $H(p)$ を最大化する分布 p^* を求める問題。

$$\text{maximize } H(p) \quad \text{subject to } E[F(x)] = \eta_{obs}$$

これは、一様分布 $u(x)$ からの KL ダイバージェンス $D(p\|u) = -H(p) + \text{const}$ の最小化と等価である。幾何学的には、制約条件が定める部分多様体 (m-平坦) へ、一様分布から下ろした **m-射影** である。解は Gibbs 分布 (e-平坦) となる。

2. **平均場近似 (変分推論)**: 複雑な真の分布 P を、扱いやすい簡単な分布族 $S = \{q_\theta\}$ (独立・平均場など) で近似する問題。通常、逆 KL ダイバージェンス $D(q_\theta\|P)$ を最小化する。

$$q^* = \arg \min_{q \in S} D(q\|P)$$

幾何学的には、真の分布 P からモデル多様体 S への **e-射影** に相当する。これにより、平均場近似の解は「 P の双対座標 (相互作用) を S 上に射影したもの」として特徴づけられる。

4.5.2 大偏差原理と Legendre 変換

大数の法則により、標本平均 \bar{X}_N は真の平均 μ に収束するが、そこからズレる確率 (大偏差) は指数関数的に減衰する。Cramer の定理によれば、その減衰率 (レート関数) $I(u)$ は以下ようになる。

$$P(\bar{X}_N \approx u) \sim \exp(-NI(u)) \quad (4.9)$$

ここで、レート関数 $I(u)$ は、対数モーメント生成関数 (情報幾何におけるポテンシャル $\psi(\theta)$) の Legendre 変換そのものである。

$$I(u) = \sup_{\theta} \{\theta \cdot u - \psi(\theta)\} = \varphi(u) \quad (4.10)$$

すなわち、情報幾何学における**双対ポテンシャル $\varphi(\eta)$ ** は、統計学・物理学における**大偏差レート関数 (またはエントロピー関数)** と物理的実体が同一である。この事実は、双対平坦空間の構造が、単なる形式的な類推ではなく、確率論の基本原理に根ざしていることを示している。

4.6 統計的仮説検定の幾何

最尤推定が「モデル多様体への射影」であったのと同様に、仮説検定 (Hypothesis Testing) もまた、情報幾何学的な解釈が可能である。

4.6.1 Neyman-Pearson の補題と m-平坦境界

単純仮説検定問題 $H_0 : p = p_0$ vs $H_1 : p = p_1$ を考える。Neyman-Pearson の補題によれば、最強力検定 (第一種過誤を固定した下で第二種過誤を最小化する検定) の棄却域

は、対数尤度比によって定まる。

$$\log \frac{p_1(x)}{p_0(x)} > T \quad (\text{閾値}) \quad (4.11)$$

指数型分布族において、対数尤度比は自然パラメータ θ に関して線形となる。

$$\log \frac{p(x; \theta_1)}{p(x; \theta_0)} = (\theta_1 - \theta_0) \cdot F(x) - (\psi(\theta_1) - \psi(\theta_0)) \quad (4.12)$$

したがって、棄却域の境界（決定境界）は、十分統計量 $F(x)$ の空間（期待値パラメータ η の空間）における**超平面（m-平坦な部分多様体）**となる。幾何学的には、検定問題とは「データ観測点から2つの仮説分布へのダイバージェンスの差（距離差）を測り、どちらに近いかを判定する」操作（Voronoi 分割）とみなせる。

4.6.2 Chernoff 情報量

検定の誤り確率（過誤率）の漸近的な指数減少率は、Chernoff 情報量によって与えられる。

$$C(p_0, p_1) = - \min_{t \in [0,1]} \log \int p_0(x)^{1-t} p_1(x)^t dx \quad (4.13)$$

幾何学的には、これは2点 p_0, p_1 を結ぶ**e-測地線**上の点で、両端からのダイバージェンスが釣り合う点（等距離点）までの距離に対応する。すなわち、仮説検定の性能限界は、分布空間内での「2点間の幾何学的な分離度（separation）」によって決定づけられる。

第 5 章

推定関数の理論とファイバーバンドル

前章までは、モデル S が真の分布を含む (Realizable case) か、少なくともパラメトリックなモデルが固定されている状況を扱った。しかし、実際の推論では「局所的な座標系 (パラメータ)」に依存しない普遍的な推定の枠組みが求められることがある。本章では、推定関数を「分布空間上のベクトル場」として幾何学的に定式化し、ファイバーバンドル (束) の言語を用いてその構造を解明する。

5.1 推定関数と幾何学的要請

観測データ x に基づいて、未知パラメータ $\theta \in \mathbb{R}^n$ を推定する問題を考える。推定量 $\hat{\theta}(x)$ を直接構成する代わりに、以下の推定方程式

$$y(x, \theta) = 0 \quad (5.1)$$

の解として推定量を与える関数 $y(x, \theta) = (y_1, \dots, y_n)$ を推定関数 (estimating function) と呼ぶ。最尤推定におけるスコア関数 $y = \nabla_{\theta} l(x, \theta)$ はその代表例である。

幾何学的な視点では、推定関数は以下の共変性 (covariance) を満たすべきである。パラメータ変換 $\theta \rightarrow \bar{\theta}(\theta)$ に対して、方程式の解は不変でなければならない。すなわち、推定関数自体はパラメータ変換に対して「ベクトル (あるいは 1 形式)」として変換されるべき対象である。

5.2 局所指数族バンドル

統計モデル $S = \{p(x; \theta)\}$ の各点 p_{θ} において、推定関数が住むべき自然な線形空間を構成する。

定義 5.1 (局所指数族). 点 $p_{\theta} \in S$ における接空間 $T_{\theta}S$ は、スコア関数 $\{\partial_i l(x; \theta)\}$ で張られる空間であった。これを拡張し、確率変数の関数空間全体 $L^2(p_{\theta})$ の中で、定数関数

を除き、かつ接空間を含むような「指数型分布族的な方向」を集めた空間を考える。これを局所指数族 (Local Exponential Family) あるいは接バンドルの拡張とみなす。

より厳密には、多様体 S 上のベクトル束 (vector bundle) E を構成する。

- 底空間 (Base space) : 統計モデル多様体 S
- ファイバー (Fiber) : 各点 $p \in S$ において、その点での「許容される推定関数」たちのなすヒルベルト空間 $\mathcal{H}_p \subset L^2(p)$ 。

推定関数 $\mathbf{y}(x, \boldsymbol{\theta})$ は、このバンドル E の切断 (section) として定義される。

$$\mathbf{y} \in \Gamma(E) \quad (\text{Section of bundle } E)$$

すなわち、各点 $\boldsymbol{\theta}$ に、確率変数ベクトル $\mathbf{y}(\cdot, \boldsymbol{\theta}) \in \mathcal{H}_{p_{\boldsymbol{\theta}}}$ を滑らかに対応させる写像である。

5.3 Hilbert バンドルと接続

推定関数の良さ (不偏性や有効性) を議論するためには、異なる点 $\boldsymbol{\theta}$ と $\boldsymbol{\theta} + d\boldsymbol{\theta}$ のファイバー間をつなぐ「接続」が必要となる。

定義 5.2 (不偏推定関数). 切断 \mathbf{y} が不偏 (unbiased) であるとは、すべての $\boldsymbol{\theta} \in S$ において

$$E_{\boldsymbol{\theta}}[\mathbf{y}(x, \boldsymbol{\theta})] = 0 \quad (5.2)$$

が成り立つことをいう。

この不偏条件を幾何学的に解釈するために、ファイバー束に共変微分 ∇ を導入する。確率密度関数 $p(x; \boldsymbol{\theta})$ 自体が $\boldsymbol{\theta}$ に依存して変化するため、期待値操作 $E_{\boldsymbol{\theta}}[\cdot]$ の微分には補正項が必要となる。

定理 5.3 (推定関数の共変微分). 推定関数ベクトル場 $\mathbf{y} = y_i e^i$ ($\{e^i\}$ はファイバーの基底) に対する共変微分は以下で与えられる。

$$\nabla_X \mathbf{y} = X(\mathbf{y}) - E_{\boldsymbol{\theta}}[X(\mathbf{y})] \quad (5.3)$$

あるいは、成分表示での微分則は

$$\partial_i y_j(x, \boldsymbol{\theta}) + E[(\partial_i l) y_j] = \dots \quad (5.4)$$

のような形式をとり、これが 0 になることが、平行移動 (parallel transport) の概念に対応する。

5.4 最適推定関数の幾何

Godambe の情報量 (Godambe information) を最大化する「最適推定関数」は、このバンドル幾何学において明確な解を持つ。

定理 5.4 (最適推定関数). 正則条件の下で、不偏推定関数のクラスの中で漸近分散を最小化 (Godambe 情報量を最大化) する最適な推定関数 y^* は、スコア関数そのものである。

$$y^*(x, \theta) = \nabla_{\theta} l(x, \theta) \quad (5.5)$$

しかし、モデルに局外母数 (nuisance parameter) が含まれる場合、あるいはセミパラメトリックな設定では、最適な推定関数はスコア関数を「局外母数の接空間に対して直交する方向」へ射影したものとなる。

この射影操作こそが、幾何学的な**接続 (Connection) による水平持ち上げ (Horizontal lift) **と解釈される。すなわち、有効スコア (efficient score) とは、局外母数の変動を吸収するように調整された、バンドル上の「水平な」ベクトル場である。

注釈 5.5 (数理物理との関連). この構造は、ゲージ理論 (Gauge Theory) における主束 (Principal Bundle) と接続の理論と全く同型である。パラメータの変換や局外母数のとりかえは「ゲージ変換」に対応し、推定量などの物理量は「ゲージ不変量」として記述される。この視点は、統計的推論における「不変性」の議論を微分幾何学の言葉で統一的に扱う強力な武器となる。

第 6 章

時系列と線形システムの幾何

確率過程、特に時系列信号を生成する線形システム全体がなす空間は、有限次元の微分可能多様体とみなせる。本章では、線形システムの安定性やフィードバック制御といった制御理論の概念を、情報幾何の視点から再構築する。

6.1 システム空間と時系列の多様体

離散時間の定常ガウス過程 $\{x_t\}_{t \in \mathbb{Z}}$ を生成する自己回帰モデル（AR モデル）を考える。次数 p の AR モデルは以下の差分方程式で記述される。

$$x_t - \sum_{i=1}^p a_i x_{t-i} = \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (6.1)$$

このモデルは、パラメータ $\mathbf{a} = (a_1, \dots, a_p) \in \mathbb{R}^p$ によって一意に定まる。したがって、AR モデル全体の集合 $M_{AR(p)}$ は \mathbb{R}^p の部分集合（多様体）とみなせる。

パワースペクトル密度関数 $S(\omega; \mathbf{a})$ は以下で与えられる。

$$S(\omega; \mathbf{a}) = \frac{1}{|A(e^{j\omega})|^2}, \quad A(z) = 1 - \sum_{i=1}^p a_i z^{-i} \quad (6.2)$$

時系列の幾何学において、各点 \mathbf{a} はこのスペクトル関数 $S(\omega; \mathbf{a})$ と同一視される。

6.2 システム空間の計量と接続

定常過程の確率分布間の「距離」は、観測時間が無限大の極限における単位時間あたりの KL ダイバージェンス（Kullback-Leibler rate）で定義される。

定義 6.1 (スペクトル空間の計量). 2 つのスペクトル密度 $S_1(\omega), S_2(\omega)$ の間のダイバージェンスレートは、以下のように周波数領域の積分で表される（Itakura-Saito 距離に関連）。

$$D(S_1 \| S_2) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{S_1(\omega)}{S_2(\omega)} - \log \frac{S_1(\omega)}{S_2(\omega)} - 1 \right) d\omega \quad (6.3)$$

これより誘導される Riemann 計量 (Fisher 計量) は、パラメータ空間上で以下のように計算される。

$$g_{ij}(\mathbf{a}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial \log S}{\partial a_i} \frac{\partial \log S}{\partial a_j} d\omega \quad (6.4)$$

この計量に関して、AR モデルの空間は平坦ではないが、一定の定曲率 (負の定曲率など) を持つ場合がある。これは双曲幾何学 (Lobachevsky 幾何学) との深い関連を示唆している。

6.3 安定システムと安定フィードバック

線形システムが有界入力・有界出力 (BIBO) 安定であるための条件は、特性多項式 $A(z) = 0$ のすべての根が複素単位円の内部にあることである。パラメータ空間 \mathbb{R}^p において、安定なシステム全体の集合 \mathcal{S}_{stable} は、ある種の多面体 (三角形や四面体の変形) のような領域を形成する。

定理 6.2 (安定領域の幾何). $p = 2$ の場合、安定領域は三角形となる (Jury の安定判別法)。高次元において、この領域の境界 (安定限界) は特異点となり、そこでは Riemann 計量が退化、あるいは発散する挙動を示す。

第 7 章

多元情報理論と統計的推論

シャノン理論における通信路符号化や、多端子情報理論（Multi-terminal Information Theory）の問題は、高次元空間における領域の体積評価の問題に帰着される。本章では、大偏差原理（Large Deviation Theory）と情報幾何学の融合について論じる。

7.1 多元情報の統計的推論

複数の関連した情報源 X, Y からのデータを圧縮・伝送する問題を考える。同時分布 $p(x, y)$ を持つモデル多様体上で、周辺分布の積 $p(x)p(y)$ （独立分布）からなる部分多様体への射影を考えることは、相互情報量 $I(X; Y)$ の幾何学的解釈を与える。

命題 7.1 (相互情報量の幾何). 相互情報量は、同時分布 p_{XY} から「独立分布のなす部分多様体 M_{indep} 」への KL ダイバージェンス（m-射影距離）に等しい。

$$I(X; Y) = \min_{q \in M_{indep}} D(p_{XY} \| q) = D(p_{XY} \| p_X p_Y) \quad (7.1)$$

7.2 仮説検定と Stein の補題

帰無仮説 $H_0 : p = p_0$ と対立仮説 $H_1 : p = p_1$ の検定において、サンプル数 $N \rightarrow \infty$ のときの誤り確率の指数的減少レート（Error exponent）は、幾何学的には「検定領域」の境界の幾何で完全に記述される。

定理 7.2 (Stein の補題の幾何). 第 1 種の誤り確率を一定値 ϵ 以下に抑える制約の下で、第 2 種の誤り確率 β_N は以下のように振る舞う。

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \log \beta_N = D(p_0 \| p_1) \quad (7.2)$$

これは、 p_0 を中心とする「検定領域」の境界が、双対平坦空間において p_0 と p_1 を結ぶ測地線に対して直交する超平面に漸近することから導かれる。

7.3 Chernoff 情報量と α -ダイバージェンス

ベイズ検定における誤り確率の指数レートは、KL ダイバージェンスではなく、Chernoff 情報量によって与えられる。幾何学的には、これは p_0 と p_1 を結ぶ「指数型測地線 (e-測地線)」上の最適な点 p_α における幾何で決定される。ここで α -ダイバージェンス $D^{(\alpha)}$ が自然に現れる。

$$C(p_0, p_1) = - \min_{\alpha \in [0, 1]} \log \int p_0(x)^{1-\alpha} p_1(x)^\alpha dx \quad (7.3)$$

この最適化問題は、双対平坦空間における凸集合間の距離問題として定式化でき、情報幾何学が漸近理論に対して強力な直観を提供する好例である。

第 8 章

情報幾何のこれからの話題

これまでの章では、Fisher 計量と双対アファイン接続に基づく古典的な（正則な）情報幾何学を体系化した。本章では、これらの枠組みを拡張し、現代的な課題（特異モデル、最適輸送、量子系）へと接続する最先端の話題を概観する。

8.1 凸解析と線形計画の内点法の幾何学

線形計画問題（LP）や半正定値計画問題（SDP）を解くための内点法（Interior Point Method）は、幾何学的には「許容領域（凸多面体）の内側に Riemann 計量を定義し、その測地線に沿って最適解へ向かう」アルゴリズムとして解釈される。

定義 8.1 (自己整合障壁関数と計量). 凸領域 $\Omega \subset \mathbb{R}^n$ 上の自己整合障壁関数 (self-concordant barrier function) を $F(\mathbf{x})$ とする。これの Hessian $\nabla^2 F(\mathbf{x})$ によって定義される計量

$$g_{ij}(\mathbf{x}) = \frac{\partial^2 F}{\partial x^i \partial x^j} \quad (8.1)$$

は、 Ω 上に双対平坦構造を誘導する。この計量に関する測地線（特に中心パス central path）を追跡することが、主双対内点法の本質である。

8.2 ニューロ多様体と特異モデル

多層パーセプトロンなどの階層型モデル（ニューロ多様体）は、パラメータ空間において重大な縮退（singularity）を持つ。例えば、中間層の重みが 0 になると、入力層の重みが何であっても出力が変わらない。このような点では Fisher 情報行列のランクが落ち、正則性が破れる。

定理 8.2 (特異点の幾何と代数幾何). 特異点近傍における KL ダイバージェンスの挙動は、もはや 2 次形式（Riemann 計量）では記述できず、より高次の多項式オーダーとなる。渡辺澄夫（Watanabe, 2009）は、特異点解消定理（Hironaka's theorem）を用いて、

分配関数の漸近挙動が「実対数閾値 (real log canonical threshold)」という双有理不変量によって支配されることを示した。これは情報幾何学と代数幾何学の深遠な融合領域 (Singular Learning Theory) である。

8.3 Wasserstein 幾何と最適輸送

確率分布の空間には、Fisher 計量とは異なる自然な計量構造として、最適輸送コストに基づく Wasserstein 距離 (Earth Mover's Distance) が存在する。

定義 8.3 (Wasserstein 計量). 確率密度空間上の接ベクトル (変分) δp に対し、連続の方程式 $\delta p + \nabla \cdot (p \nabla \phi) = 0$ を満たすポテンシャル ϕ を用いて、Wasserstein 計量 (L^2 -Wasserstein metric) は以下のように定義される。

$$g_p^W(\delta p, \delta p) = \int p(x) \|\nabla \phi(x)\|^2 dx \quad (8.2)$$

これは Benamou-Brenier の公式として知られ、分布空間を無限次元 Riemann 多様体とみなす視点を与える。

近年、情報幾何 (Fisher-Rao 幾何) と最適輸送 (Wasserstein 幾何) を補間・統合する試みが盛んに行われており、エントロピー正則化された最適輸送問題において、両者の幾何学的構造が交錯する様子 (Sinkhorn アルゴリズムの幾何など) が明らかになりつつある。

8.4 量子観測の情報幾何

古典確率分布 $p(x)$ を密度行列 ρ (トレース 1 の正定値エルミート行列) に置き換えることで、量子情報幾何学 (Quantum Information Geometry) が構築される。これは量子推定 (量子センシング) や量子通信の限界を記述する数学的言語である。

8.4.1 量子状態と測定

n 次元 Hilbert 空間 \mathcal{H} 上の量子状態は、密度行列の集合 $\mathcal{S}(\mathcal{H}) = \{\rho \in \mathcal{L}(\mathcal{H}) \mid \rho \geq 0, \text{Tr} \rho = 1\}$ で表される。これは $(n^2 - 1)$ 次元の実多様体をなす。量子系に対する測定は、POVM (正作用素値測定) $M = \{\Pi_x\}$ によって記述され、測定値 x が得られる確率は $p(x) = \text{Tr}(\rho \Pi_x)$ で与えられる。

8.4.2 量子 Fisher 計量と作用素単調関数

古典系とは異なり、量子系においては「Fisher 計量の自然な拡張」が一意には定まらない。Petz (1996) は、量子 Fisher 計量は**作用素単調関数 (operator monotone

function) $**f$ と一対一に対応することを示した (Petz の分類定理)。単調性 (CPTP 写像による情報の非増大) を満たす計量 g^f は以下のように書ける。

$$g_p^f(X, Y) = \text{Tr} \left(X (\mathbf{R}_\rho^{1/2} f(\mathbf{L}_\rho \mathbf{R}_\rho^{-1}) \mathbf{R}_\rho^{1/2})^{-1} (Y) \right) \quad (8.3)$$

ここで $\mathbf{L}_\rho, \mathbf{R}_\rho$ はそれぞれ ρ の左・右乗算作用素である。

代表的な量子計量として以下がある。

- **SLD 計量 (対称対数微分)**: $f(t) = \frac{1+t}{2}$ に対応。推定理論で標準的に用いられ、最小の量子 Cram é r-Rao 下界を与える。
- **RLD 計量 (右対数微分)**: $f(t) = \frac{2t}{1+t}$ に対応。非可換性が強い系で重要となる。
- **Bogoliubov-Kubo-Mori (BKM) 計量**: $f(t) = \frac{t-1}{\log t}$ に対応。量子統計力学 (線形応答理論) で自然に現れる。

8.4.3 量子推定と有限量子状態空間の双対構造

未知パラメータ θ を持つ量子状態 ρ_θ を推定する際、測定手段 M も最適化の対象となる。量子 Cram é r-Rao 不等式は、任意の不偏推定量と任意の測定に対して、誤差共分散行列が量子 Fisher 情報行列の逆行列以上になることを主張する。

$$V[\hat{\theta}] \geq (G^{SLD}(\theta))^{-1} \quad (8.4)$$

また、量子状態空間においても、古典系と同様に「指数型分布族 (Gibbs 状態)」と「混合型分布族」の双対平坦構造を定義することができる。ただし、演算子の非可換性に起因して、古典のような「曲率ゼロ」の性質が完全には成り立たない場合があるため、Nagaoka (長岡) らによって拡張された「量子双対幾何」の枠組みが必要となる。

8.5 情報幾何が提起する数学上の問題

最後に、今後の発展が期待される未解決・重要課題を挙げる。

1. 無限次元情報幾何: ノンパラメトリックモデルにおける Banach 多様体としての厳密な構成。
2. 離散構造の幾何: グラフやネットワークなどの離散対象に対する、曲率やダイバージェンスの自然な定義 (離散 Ricci 曲率など)。
3. 深層学習のダイナミクス: 超多数のパラメータを持つニューラルネットの学習過程を、異方的な計量を持つ空間上の運動として解析する (自然勾配法の一般化)。

これらの問題は、情報幾何学が単なる「統計の幾何学的解釈」にとどまらず、複雑系科学全体の基盤言語へと進化する可能性を秘めている。

付録 A

Python による計算情報幾何

本章では、前章までに定義した抽象的な幾何学量（計量、接続、曲率）を、具体的な確率モデルに対して数値的に計算するためのアルゴリズムと実装指針を示す。現代の深層学習フレームワーク（PyTorch, JAX）の自動微分機能は、これらの計算を劇的に簡略化する。

A.1 Fisher 情報行列の自動微分による計算

確率モデル $p(x; \theta)$ の対数尤度 $l(\theta) = \log p(x; \theta)$ に対し、Fisher 情報行列 $G(\theta)$ は期待値 $\mathbb{E}[\nabla l(\nabla l)^T]$ で定義される。モンテカルロ積分と自動微分を組み合わせることで、任意の複雑なモデルに対して G を近似計算できる。

アルゴリズム A.1 (Fisher 行列の計算). 1. パラメータ θ におけるモデルから N 個のサンプル x_1, \dots, x_N を生成する。
2. 各サンプルに対し、逆伝播法（Backpropagation）を用いて勾配ベクトル $\mathbf{g}_k = \nabla_{\theta} \log p(x_k; \theta)$ を計算する。
3. サンプル共分散行列として G を推定する。

$$\hat{G}(\theta) = \frac{1}{N} \sum_{k=1}^N \mathbf{g}_k \mathbf{g}_k^T$$

A.2 Christoffel 記号と測地線方程式の数値解法

測地線方程式 $\ddot{\theta}^k + \Gamma_{ij}^k \dot{\theta}^i \dot{\theta}^j = 0$ は、2 階の常微分方程式（ODE）である。接続係数 Γ_{ij}^k は計量 G の微分を含み、計算コストが高い。しかし、ラグランジュ形式（Lagrangian Mechanics）を用いることで、 Γ を陽に計算せずに測地線を求めることができる。

$$L(\theta, \dot{\theta}) = \frac{1}{2} \dot{\theta}^T G(\theta) \dot{\theta} \quad (\text{A.1})$$

このラグランジアンに対する Euler-Lagrange 方程式を解くことは、測地線方程式を解くことと等価である。自動微分を用いて $\partial L / \partial \theta$ と $\partial L / \partial \dot{\theta}$ を計算し、Symplectic Euler 法

などの数値積分法を適用する。

A.3 JAX による実装例（概念コード）

```
import jax.numpy as jnp
from jax import grad, jacobian, vmap

def log_likelihood(theta, x):
    # 任意のモデルの対数尤度定義
    ...

def fisher_metric(theta, samples):
    # per-sample gradients
    grads = vmap(grad(log_likelihood, argnums=0), (None, 0))(theta, samples)
    #  $G = E[\text{grad} * \text{grad}.T]$ 
    return jnp.mean(jnp.einsum('ni,nj->nij', grads, grads), axis=0)

# Christoffel symbols via automatic differentiation of Metric
def christoffel_symbols(theta, samples):
    G_func = lambda t: fisher_metric(t, samples)
    dG = jacobian(G_func)(theta) # Partial derivatives of G
    G_inv = jnp.linalg.inv(G_func(theta))

    #  $\Gamma_{ij}^k = 1/2 * g^{kl} * (d_i g_{jl} + d_j g_{il} - d_l g_{ij})$ 
    # (Tensor operations implementation...)
    return Gamma
```

付録 B

多様体上の最適化：自然勾配法

ユークリッド空間上の通常の勾配法 $\theta_{t+1} = \theta_t - \eta \nabla L$ は、パラメータ空間の幾何学的構造（単位）を無視しているため、プラトー（平坦な領域）での停滞や振動の問題が生じる。これに対し、Riemann 多様体上の最急降下方向を用いるのが自然勾配法である。

B.1 Riemann 勾配と自然勾配

損失関数 $L(\theta)$ の最急降下方向は、パラメータの変化量 $\|\delta\theta\|$ を計量 G で測ったノルム $\|\delta\theta\|_G^2 = \delta\theta^T G \delta\theta$ が一定という制約の下で、 L を最も減少させる方向である。

定理 B.1 (自然勾配). Riemann 多様体 (\mathcal{M}, G) 上の関数 L の最急降下方向（反変ベクトル）は以下で与えられる。

$$\tilde{\nabla} L = G^{-1}(\theta) \nabla L(\theta) \quad (\text{B.1})$$

これを用いた更新則

$$\theta_{t+1} = \theta_t - \eta G^{-1}(\theta_t) \nabla L(\theta_t) \quad (\text{B.2})$$

を自然勾配学習法（Natural Gradient Descent, NGD）と呼ぶ（Amari, 1998）。

B.2 Fisher 情報行列の近似と K-FAC

深層学習においてパラメータ数 n が数百万に達する場合、 $n \times n$ の Fisher 行列 G の逆行列計算は $O(n^3)$ の計算量を要し、実行不可能である。実用上は、以下のような構造的近似が用いられる。

- 対角近似: G の対角成分のみ用いる（AdaGrad, RMSProp のヒューリスティクスに近い）。
- **K-FAC (Kronecker-Factored Approximate Curvature)**: 各層の重み行列に対する Fisher 行列を、入力共分散と出力勾配共分散のクロネッカー積で近似する。

$$G \approx A \otimes S$$

これにより逆行列計算を $O(n)$ 程度まで高速化できる。

付録 C

高次元統計空間の幾何学

次元数 d がサンプル数 N よりも大きい、あるいは同程度のオーダーで増大する高次元設定 (High-Dimension, Low-Sample-Size; HDLSS) において、古典的な情報幾何 (固定次元 d) の直観は破綻する。本章では、高次元空間特有の「集中の幾何 (Geometry of Concentration)」について論じる。

C.1 高次元空間における直交性と集中現象

高次元球面上からランダムに選ばれた 2 つのベクトルは、確率 1 でほぼ直交する (Concentration of Measure)。この現象は、統計モデルの Fisher 幾何にも影響を与える。

命題 C.1 (高次元における距離の収束). $d \rightarrow \infty$ のとき、ノイズを含む 2 つのデータ点間のユークリッド距離やダイバージェンスは、決定論的な値に確率収束する (Hall et al., 2005; Aoshima et al., 2018)。この「幾何学的表現 (Geometric Representation)」を利用することで、次元の呪いを逆手に取った分類器や検定法が構築可能となる。

C.2 高次元特有の双対構造

古典的な双対平坦性は固定次元での議論であるが、高次元極限においては、標本共分散行列の固有値分布 (Marcenko-Pastur 則など) が支配的となる。スパース推定 (Lasso など) における L_1 正則化項は、幾何学的には「パラメータ空間における特異な計量 (原点で尖った単位球)」を導入することと解釈でき、これはある種の Finsler 幾何学への拡張を示唆している。