

凸関数に基づく情報幾何学

— 擬似平面性・擬似球面性と諸性質との関係 —

平岡 和幸* 田中 勝†

目次

1	序論	3
2	ベクトル空間上の情報幾何	5
2.1	双対空間との非線形な対応づけ	6
2.2	接空間	9
2.3	計量	12
2.4	ダイバージェンス	14
2.5	双対接続	18
2.6	ダイバージェンス最小化	19
2.7	シフトとスケーリング	20
3	擬似平面および擬似球面上の情報幾何	21
3.1	部分多様体 P_1, L_1 上の情報幾何	21
3.2	擬似平面	22
3.2.1	平面と擬似平面	22
3.2.2	擬似平面上の接続	23
3.2.3	擬似平面内での平坦性	24
3.2.4	測地線に関する拡張 Pythagoras の定理	26
3.3	擬似球面	27
3.3.1	等ポテンシャル面と擬似球面	27
3.3.2	擬似球面上の接続	28
3.3.3	擬似球面内での平坦性	28
3.3.4	平坦性に関する補足	30
3.3.5	有限次元平面との交わり	32
3.3.6	ダイバージェンス最小化	34
3.3.7	線形制約のもとでの線形関数最小化	35
3.4	P_1, L_1 が共に擬似球面の場合	36
3.4.1	P_1, L_1 上の情報幾何	36

* 和歌山工業高等専門学校 総合教育科

† 福岡大学 理学部 応用数学科

3.4.2	ポテンシャル関数の特徴づけ	37
3.5	P_1 が擬似球面で L_1 が擬似平面の場合	39
3.5.1	P_1, L_1 上の情報幾何	39
3.5.2	ポテンシャル関数の特徴づけ	40
4	確率分布の情報幾何	41
4.1	関数空間上の情報幾何	41
4.2	確率分布のなす部分多様体	42
4.3	U -ダイバージェンス	43
4.4	α -ダイバージェンスと q -正規分布	45
4.5	η -ダイバージェンスと Kullback-Leibler ダイバージェンス	48
5	統計との関係 まだ不整合	50
5.1	Kullback ダイバージェンスと大偏差原理	50
5.2	Fisher 計量とクラメル-ラオの不等式	50
5.3	exponential family, mixture family	51
5.4	最尤推定と Kullback ダイバージェンス最小化	51
6	双対座標 まだ不整合	52
6.1	θ 座標と η 座標	52
6.2	θ 座標と η 座標の双対性	53
6.2.1	計量 G についての双対性	53
6.2.2	Legendre 変換としての双対性	54
6.3	座標で表現した情報幾何	55
6.3.1	θ 座標、 η 座標からはじめる情報幾何	55
6.3.2	フォーマルな情報幾何	56
付録 A	付録	59
A.1	Legendre 変換 (Young の不等式) まだ	59
A.2	相似直交な部分多様体	59
A.2.1	ベクトル場の相似性	59
A.2.2	相似直交な部分多様体	61
A.2.3	擬似平面と擬似球面	62
A.3	P_1, L_1 が共に擬似平面の場合	63
A.4	成分ごとの和に分解されないポテンシャル	65
A.4.1	P_1, L_1 が共に擬似球面の場合	65
A.4.2	P_1 が擬似球面で L_1 が擬似平面の場合	66
A.5	課題	66

1 序論

計量が入った多様体に対しては、通常、その計量に合った接続である Levi-Civita 接続を考えることが多い。しかし情報幾何 [1, 2, 3] においては、Levi-Civita 接続ではなく、互いに双対な 2 つの接続を活用する。このような接続の対が導入される源として、コントラスト関数 [5, 6, 7] およびアファイン微分幾何 [16] という二種類の枠組が挙げられる。

コントラスト関数は、多様体上の 2 点 p, q に対してその相異の度合 $\rho(p, q)$ を定めたものである。コントラスト関数の定義は非負性や $\rho(p, q) = 0$ iff $p = q$ などの距離に似た要請を含むが、対称性や三角不等式は仮定しない。コントラスト関数 ρ が与えられれば、ごく近い二点 p, q に対する ρ の振舞いに基き、

$$G_{ij}(r) = - \frac{\partial^2 \rho(p, q)}{\partial p^i \partial q^j} \Big|_{p=q=r} = \frac{\partial^2 \rho(p, q)}{\partial p^i \partial p^j} \Big|_{p=q=r} = \frac{\partial^2 \rho(p, q)}{\partial q^i \partial q^j} \Big|_{p=q=r} \quad (1)$$

$$\sum_j \Gamma_{km}^j(r) G_{jl}(r) = - \frac{\partial^3 \rho(p, q)}{\partial p^k \partial p^m \partial q^l} \Big|_{p=q=r} \quad (2)$$

$$\sum_j \Gamma_{km}^{*j}(r) G_{jl}(r) = - \frac{\partial^3 \rho(q, p)}{\partial p^k \partial p^m \partial q^l} \Big|_{p=q=r} = - \frac{\partial^3 \rho(p, q)}{\partial q^k \partial q^m \partial p^l} \Big|_{p=q=r} \quad (3)$$

のように計量 G と接続の対 Γ, Γ^* を導入することができる。統計的な議論を行う際には確率分布間の相異の度合が重要となるので、コントラスト関数に基いて幾何構造を定めるこの枠組は情報幾何の源にふさわしい。真の分布 q とそこから得られた経験分布 p との相異を測るという典型的な用途を考えると、両者の役割の違いから、非対称性 $\rho(p, q) \neq \rho(q, p)$ もむしろ自然である^{*1}[7]。

一方、アファイン微分幾何の理論は、 $(n+1)$ 次元アファイン空間内の n 次元部分多様体 M を扱う枠組を含む。そこでは、ベクトル場 ξ を一つ指定し、Euclid 空間における法線ベクトルの役割を ξ に代行させることによって、 M 上の接続や計量に相当するものが導入される。 M の接ベクトル v を微小平行移動するには、外側のアファイン空間でまず微小平行移動し、 M の接空間からはみ出した分を ξ 方向に射影して接ベクトルに戻せばよい。得られた接ベクトルにより接続 Γ が定められ、接ベクトルに戻す補正の大きさから計量に相当するものが定められる。また、もう一つの接続 Γ^* は、いまの計量に関する Γ の双対として得られる。確率密度関数の集合は関数空間内の部分多様体ととらえられるの

^{*1} 確率 1/2 で表が出るコインを 10 回投げて表が 0 回になる確率は、確率 0 で表が出るコインを 10 回投げて表が半分になる確率とは異なる。この極端な例からも非対称性が伺えるであろう。

で、アファインはめ込みの観点からその幾何を調べることもやはり自然なアプローチと言えよう。

情報幾何という言葉の指し示す範囲ははっきり定まったものではないが、実際の応用場で主に用いられる Kullback-Leibler ダイバージェンス、 α -ダイバージェンス、 U -ダイバージェンスなどに関する「情報幾何」は、上で挙げた二種類の枠組を兼ね備えた形をしている。つまり、アファイン空間 \mathcal{P} 上にコントラスト関数 ρ を与えておいて、 \mathcal{P} 内の部分多様体を取り扱う。ベクトル場 ξ としては、 ρ から導かれる計量に関する法線ベクトルを用いる。

ρ の定める接続 Γ をアファイン空間における平行移動と一致させるには、アファイン座標に関して $\Gamma_{km}^j = 0$ 、すなわち (2) の右辺が 0 でなくてはならない。 $\partial\rho(p, q)/\partial q^l$ が p に関して一次式ならこの要請は自動的に満たされるから、

$$\rho(p, q) = f(p) + \tilde{f}(q) - \langle p, \mathcal{F}(q) \rangle \quad (4)$$

という形のコントラスト関数を考えることにしよう。ただし記法の便宜上、原点 o を勝手に指定してアファイン空間 \mathcal{P} をベクトル空間とみなし、 \mathcal{P} の元 p と双対ベクトル空間 \mathcal{P}^* の元 l との内積を

$$\langle p, l \rangle = \sum_i p^i l_i \quad (5)$$

と書いた。この ρ がコントラスト関数の定義を満たすためには、 f は凸でなくてはならず、しかも

$$0 = \rho(q, q) = f(q) + \tilde{f}(q) - \langle q, \mathcal{F}(q) \rangle \quad (6)$$

$$0 = \left. \frac{\partial \rho(p, q)}{\partial p^j} \right|_{p=q} = \frac{\partial f(q)}{\partial q^j} - \mathcal{F}(q)_j \quad (7)$$

が要求される。以上をまとめれば、次のような枠組に至る。

- ベクトル空間 \mathcal{P} 上に凸関数 f が与えられたとき、
- f の微分を用いて双対空間との対応づけ $\mathcal{F} : \mathcal{P} \rightarrow \mathcal{P}^* : q \mapsto df(q)$ を定め、
- さらに f の Legendre 変換 $\tilde{f}(q) = \langle q, \mathcal{F}(q) \rangle - f(q)$ を用意した上で、
- $\rho(p, q) = f(p) + \tilde{f}(q) - \langle p, \mathcal{F}(q) \rangle$ により p, q の相異を測る。

短く言えば、凸関数 f を点 q のまわりで一次近似したときに、点 p における近似誤差が $\rho(p, q)$ である：

$$\rho(p, q) = f(p) - \{f(q) + \langle (p - q), df(q) \rangle\} \quad (8)$$

この枠組は f をポテンシャル関数とした Hesse 幾何学 [17] に他ならない。実際、いまの ρ から導かれる計量はアファイン座標に関して

$$G_{ij}(p) = \frac{\partial^2 f(p)}{\partial p^i \partial p^j} \quad (9)$$

となり、ポテンシャル関数の二階微分で与えられる。

前に挙げた Kullback-Leibler ダイバージェンス、 α -ダイバージェンス、 U -ダイバージェンスはいずれも、適当な凸関数 f を用いて上のように作ったコントラスト関数 ρ に一致する（ただし、 α -ダイバージェンスの場合は、 p, q を確率密度関数ではなくそのべき乗に対応づける）。本稿では、この枠組を簡明な描像で述べるとともに、情報幾何で多用される各種の性質が凸関数 f のどのような特性に起因しているかを整理する。また、情報幾何において特別な地位を与えられた Kullback-Leibler ダイバージェンスおよび α -ダイバージェンスが、 f の等位面に関する幾何学的な条件から特徴づけられることを説明する。さらに、非平衡系の物理で用いられる Tsallis エントロピー [12] に関連して、エスコート分布や q -正規分布 [13] についても触れる。事前知識としては多様体論の基礎を前提とする。

情報幾何の従来解説においてはしばしば「暗黙の同一視」が行われるので、数学書で学んだ微分幾何や多様体の知識をあてはめようとするとうる覚えを感じる場合がある。特に、ベクトル空間を多様体とみなした際の空間と接空間との同一視は混乱を招きやすい。本稿ではそれらの区別を陽に述べる。

2 ベクトル空間上の情報幾何

ベクトル空間上に凸関数を与えれば、計量、双対接続、ダイバージェンスなどからなる情報幾何の枠組を構成することができる。この構成法についてまず述べよう。なお、ベクトル空間を仮定するのは記法の便宜と対称性を考慮したからに過ぎず、本来はアファイン空間でも十分である。実際、原点のシフトは本質的でないことが 2.7 項で明らかになる。

本稿を通じ、無限次元に関しては素朴な直感的考察を述べるに留め、数学的に厳密な議論には立ち入らない。微分可能性などもいちいち断わらない。

2.1 双対空間との非線形な対応づけ

\mathcal{L} を一般の実ベクトル空間とし、その双対空間を \mathcal{P} とする*2。双対空間の定義により、 $l \in \mathcal{L}$ と $p \in \mathcal{P}$ との間には内積

$$\langle p, l \rangle \quad (10)$$

が定められる。有限次元の場合に成分で書けば、 l が縦ベクトル

$$l = (l^1, \dots, l^n)^T \quad (11)$$

のとき p は横ベクトル

$$p = (p_1, \dots, p_n) \quad (12)$$

と解釈され、

$$\langle p, l \rangle = pl = \sum_{k=1}^n p_k l^k \quad (13)$$

である。記号 T は転置を表す。反変ベクトルであることを明示するために、 l_k でなく上つき添字で l^k と書いた。

いま、ベクトル空間 \mathcal{L} 上に凸関数 $g: \mathcal{L} \rightarrow \mathbb{R}$ が与えられたとしよう。この凸関数をポテンシャル関数と呼ぶ。以下に述べる変換とその性質は、Legendre 変換の一般論としてよく知られている（付録 A.1）。

点 $l \in \mathcal{L}$ に対し、この点における g の微分 $dg(l)$ は双対空間 \mathcal{P} の元となる。有限次元の場合に成分で書けば、縦ベクトル $l = (l^1, \dots, l^n)^T$ に関する微分 $dg(l)$ は横ベクトル

$$dg(l) = \left(\frac{\partial g}{\partial l^1}, \dots, \frac{\partial g}{\partial l^n} \right) \quad (14)$$

である。一般の場合には、 $dg(l)$ とは次のようなものと思えばよい：

$l \in \mathcal{L}$ を固定したとき、 $\Delta l \in \mathcal{L}$ と $|\epsilon| \ll 1$ に対して

$$g(l + \epsilon \Delta l) = g(l) + \epsilon G_l(\Delta l) + o(\epsilon) \quad (15)$$

が成り立つ線形関数 $G_l: \mathcal{L} \rightarrow \mathbb{R}$ があったとしよう。 G_l は \mathcal{L} 上の線形関数だから双対空間 \mathcal{P} の元と解釈され、上の式は

$$g(l + \epsilon \Delta l) = g(l) + \epsilon \langle G_l, \Delta l \rangle + o(\epsilon) \quad (16)$$

*2 情報幾何の最も標準的な設定では、 \mathcal{P} は確率密度関数 $p(\cdot)$ の集合を含み、 \mathcal{L} はその対数 $l(\cdot) = \log p(\cdot)$ の集合を含む。この対応を想起しやすいように \mathcal{P}, \mathcal{L} という記号を用いることにした。なお、情報幾何における反変共変の慣習を多様体論とあわせるために、しばらくは \mathcal{L} が主、 \mathcal{P} が従という扱いで説明する。

と内積で書くこともできる。 $dg(l)$ とはこの $G_l \in \mathcal{P}$ のことである。

いずれにせよ、 dg は各点 $l \in \mathcal{L}$ を双対空間の点 $dg(l) \in \mathcal{P}$ へと対応づける。つまり、 dg は \mathcal{L} から \mathcal{P} への（非線形な）写像を定める。この写像を $\mathcal{G} = dg$ と書く^{*3}。

$$\mathcal{G} : \mathcal{L} \rightarrow \mathcal{P} : l \mapsto dg(l) \quad (17)$$

定義から、 $l, \Delta l \in \mathcal{L}$ に対する g の方向微分は

$$\left. \frac{d}{dt} g(l + t\Delta l) \right|_{t=0} = \langle \mathcal{G}l, \Delta l \rangle \quad (18)$$

と表される。また、 g が凸であるという前提から、写像 $\mathcal{G} : \mathcal{L} \rightarrow \mathcal{P}$ は単射となる。このことは背理法を用いて次のように証明される。もし仮に、異なる二点 $l, m \in \mathcal{L}$ において $\mathcal{G}l$ と $\mathcal{G}m$ が等しくなったとしよう。いま、 $0 \leq t \leq 1$ に対して関数 $h : [0, 1] \rightarrow \mathbb{R}$ を

$$h(t) = g(l + t\Delta l), \quad \Delta l = m - l \quad (19)$$

と定義すれば、 h は凸関数のはずである。したがってその微分 $h'(t)$ は狭義単調増加でなくてはならない。この $h'(t)$ は、先ほど述べた方向微分の性質から

$$h'(t) = \langle \mathcal{G}(l + t\Delta l), \Delta l \rangle \quad (20)$$

と書ける。すると $h'(0) = \langle \mathcal{G}l, \Delta l \rangle = \langle \mathcal{G}m, \Delta l \rangle = h'(1)$ だから、 h' の単調増加性に矛盾が生じる。よって、 $l \neq m$ なら $\mathcal{G}l \neq \mathcal{G}m$ が言える。

$\mathcal{G}l = p$ のとき、

$$f(p) + g(l) = \langle p, l \rangle \quad (21)$$

という関係により $f(p) \in \mathbb{R}$ を定義する。 \mathcal{G} が単射なのでこの $f(p)$ は well-defined である。 g の凸性から

$$\arg \min_{m \in \mathcal{L}} (g(m) - \langle p, m \rangle) = l \quad (22)$$

が成り立ち、

$$\min_{m \in \mathcal{L}} (g(m) - \langle p, m \rangle) = g(l) - \langle p, l \rangle = -f(p) \quad (23)$$

と表されることに注意しよう。逆にこれらを \mathcal{G} や f の定義とすることもできる。両式をまとめると、任意の $m \in \mathcal{L}$ と任意の $p = \mathcal{G}l \in \mathcal{P}$ に対して

$$f(p) + g(m) - \langle p, m \rangle \geq 0 \quad (24)$$

$$f(p) + g(m) - \langle p, m \rangle = 0 \quad \Leftrightarrow \quad m = l \quad (25)$$

^{*3} 小文字 g の微分を大文字で表記するのは慣習に反するが、後の議論において \mathcal{G} が演算子の役割を果たすことから、演算子らしく見える記法を採用した。

が言える (Young の不等式)。

f は \mathcal{P} 上の凸関数になり、やはりポテンシャル関数と呼ばれる。凸性を示すには、 $p \neq q$ と $0 < t < 1$ に対して

$$(1-t)f(p) + tf(q) > f((1-t)p + tq) \quad (26)$$

であることを次のように確かめればよい。

$$-(1-t)f(p) - tf(q) = (1-t) \min_{m \in \mathcal{L}} (g(m) - \langle p, m \rangle) + t \min_{m \in \mathcal{L}} (g(m) - \langle q, m \rangle) \quad (27)$$

$$= \min_{m \in \mathcal{L}} (1-t)(g(m) - \langle p, m \rangle) + \min_{m \in \mathcal{L}} t(g(m) - \langle q, m \rangle) \quad (28)$$

$$< \min_{m \in \mathcal{L}} \left\{ (1-t)(g(m) - \langle p, m \rangle) + t(g(m) - \langle q, m \rangle) \right\} \quad (29)$$

$$= \min_{m \in \mathcal{L}} \left\{ g(m) - \langle (1-t)p + tq, m \rangle \right\} \quad (30)$$

$$= -f((1-t)p + tq) \quad (31)$$

途中の不等号に $=$ が含まれないのは、二つの \min をそれぞれ実現する m が決して一致しないためである。

f が凸だと保証されたので、前と同様に f を微分することで単射

$$\mathcal{F} = df : \mathcal{P} \rightarrow \mathcal{L} \quad (32)$$

が定められる。 $\mathcal{F}p = l$ なら

$$\arg \min_{q \in \mathcal{P}} (f(q) - \langle q, l \rangle) = p \quad (33)$$

となることも同様である。このとき

$$\arg \min_{q \in \mathcal{P}} (f(q) + g(l) - \langle q, l \rangle) = p \quad (34)$$

でもあるから、Young の不等式と見比べれば $gl = p$ および $f(p) + g(l) - \langle p, l \rangle = 0$ でなくてはならない。前者から、 \mathcal{G} と \mathcal{F} とは互いに逆変換になっていることがわかる (Legendre 変換の対合性)。また、後者から

$$\arg \min_{q \in \mathcal{P}} (f(q) - \langle q, l \rangle) = f(p) - \langle p, l \rangle = -g(l) \quad (35)$$

も言える。

一般には \mathcal{G} は全射と限らないので、 \mathcal{F} の定義域 \mathcal{P}_0 は空間 \mathcal{P} の一部に限られる。また、 g の定義域 \mathcal{L}_0 自体が空間 \mathcal{L} の一部に限られる場合も扱えるようにしておくとお応用がしや

すい。したがって本来は、 g の定義域 \mathcal{L}_0 と \mathcal{G} によるその像 $\mathcal{P}_0 = \mathcal{G}\mathcal{L}_0$ を明示した上で

$$g : \mathcal{L}_0 \rightarrow \mathbb{R} \quad (\text{凸関数}) \quad (36)$$

$$f : \mathcal{P}_0 \rightarrow \mathbb{R} \quad (\text{凸関数}) \quad (37)$$

$$\mathcal{G} : \mathcal{L}_0 \rightarrow \mathcal{P}_0 \quad (\text{非線形な全単射}) \quad (38)$$

$$\mathcal{F} : \mathcal{P}_0 \rightarrow \mathcal{L}_0 \quad (\text{非線形な全単射}) \quad (39)$$

のように書くべきである。しかしこれをいちいち断わるのは煩雑なので、厳密な定義域は省き、単に「 \mathcal{L} 上の関数」や「 \mathcal{L} と \mathcal{P} との一対一対応」のように略称する。必要に応じて適宜読みかえられたい。

以上で、ベクトル空間 \mathcal{L} とその双対空間 \mathcal{P} との間に非線形な対応づけ \mathcal{G}, \mathcal{F} が得られた。情報幾何においては、この対応づけにより \mathcal{L} と \mathcal{P} とを同一視し単一の空間として扱うことが多い。なお、本節では凸関数 $g : \mathcal{L} \rightarrow \mathbb{R}$ をもとにすべてを構成したが、 g のかわりに $f : \mathcal{P} \rightarrow \mathbb{R}$ の方から出発しても同等の議論がもちろん可能である。 \mathcal{L}, \mathcal{P} のどちらを主と見るのが自然かは、確率分布との対応づけを想定した上での文脈に依存するため、一概には言えない。今後は、都合に応じて \mathcal{L}, \mathcal{P} の順で書いたり \mathcal{P}, \mathcal{L} の順で書いたりすることがある。

本項で述べた \mathcal{L} と \mathcal{P} との対応づけは、[4] の特別な場合とも解釈される。この点については 3.3.1 項で再度言及する。

2.2 接空間

ここからは、 \mathcal{L} や \mathcal{P} を単なるベクトル空間ではなく多様体として扱う。一般に、多様体 \mathcal{L} 上の点 l における接ベクトルは、方向微分として定義される [14]。すなわち、 l の近傍上で定義された実数値関数 h に対して実数を定める線形汎関数のうち、特定の条件を満たすものを接ベクトルと呼ぶのが、本来の抽象的な定義である。このような汎関数としての接ベクトル Δl は $(\Delta l)[h]$ という記号で書くことにする。局所座標のもとでは、

$$(\Delta l)[h] = \left. \frac{d}{dt} h(l + t\Delta l) \right|_{t=0} \quad (40)$$

あるいは、 $|\epsilon| \ll 1$ に対して

$$h(l + \epsilon\Delta l) = h(l) + \epsilon(\Delta l)[h] + o(\epsilon) \quad (41)$$

と表される。ただし、 $l + t\Delta l$ は局所座標に関する和や実数倍と解釈する。有限次元の場合に成分で書けば、

$$l = (l^1, \dots, l^n)^T \quad (42)$$

$$\Delta l = (\Delta l^1, \dots, \Delta l^n)^T \quad (43)$$

のとき、点 l における接ベクトル Δl は微分演算子

$$\sum_{k=1}^n \Delta l^k \frac{\partial}{\partial l^k} \quad (44)$$

に対応する。点 l における接ベクトルの集合 $T_l \mathcal{L}$ は接空間と呼ばれ、自然な和と実数倍によりベクトル空間となる。

ベクトル空間である \mathcal{L} を多様体とみなしたときの著しい特徴は、各点 $l \in \mathcal{L}$ における接空間 $T_l \mathcal{L}$ がそれぞれ、空間 \mathcal{L} 自身と自然に同一視されることである。対応して、 $T_l \mathcal{L}$ の双対空間である共変接空間 $T_l^* \mathcal{L}$ はそれぞれ、 \mathcal{L} の双対空間 \mathcal{P} と同一視される。 $p \in \mathcal{P}$ についても同様である。

$$T_l \mathcal{L} \cong \mathcal{L} \cong T_p^* \mathcal{P} \quad (45)$$

$$T_l^* \mathcal{L} \cong \mathcal{P} \cong T_p \mathcal{P} \quad (46)$$

有限次元の場合に成分で書けば、これらは

$$(l^1, \dots, l^n)^T = (l^1, \dots, l^n)^T = (l^1, \dots, l^n)^T \quad (47)$$

$$(p_1, \dots, p_n) = (p_1, \dots, p_n) = (p_1, \dots, p_n) \quad (48)$$

となり表記上そもそも区別がつかない。一般の場合には、 l の近傍上で定義された実数値関数 h に対して点 $l \in \mathcal{L}$ における $\Delta l \in \mathcal{L}$ 方向の微分

$$D[h] = \left. \frac{d}{dt} h(l + t\Delta l) \right|_{t=0} \quad (49)$$

を与える線形汎関数 D を、接空間の本来の定義に基いて $\Delta l \in T_l \mathcal{L}$ と解釈すればよい。ただし、右辺の $l + t\Delta l$ はベクトル空間 \mathcal{L} における和や実数倍を表す。これで同一視 $\mathcal{L} \cong T_l \mathcal{L}$ が得られたので、対応して双対空間どうしの同一視 $\mathcal{P} \cong T_l^* \mathcal{L}$ も定められる。

空間と接空間の同一視は、情報幾何においてしばしば暗黙に用いられる。本稿では、同一視がどこで行われたかを見分けやすくするため、以下のように書く。

- 接ベクトルは $\Delta l, \Delta p$ などと Δ をつけた記号で表す。

- 空間の元 l を接空間の元とみなす際は \bar{l} 、接空間の元 Δl を空間の元とみなす際は $\underline{\Delta l}$ のように印をつける。

↑ 守られているか要確認 (まだ2章まで)

関数 $g: \mathcal{L} \rightarrow \mathbb{R}$ に対する多様体の意味での全微分 dg は \mathcal{L} 上の一次微分形式となり、各点 $l \in \mathcal{L}$ における共変ベクトル $dg(l) \in T_l^* \mathcal{L}$ を定める [14]。上の同一視によってこれを \mathcal{P} の元とみなせば、全微分は 2.1 項の \mathcal{G} に一致する。 f についても同様である。以上は要するに次のことを主張している：点 $l \in \mathcal{L}$ における接ベクトル $\Delta l \in T_l \mathcal{L}$ や点 $p \in \mathcal{P}$ における接ベクトル $\Delta p \in T_p \mathcal{P}$ に対して上の同一視を用いると、 g や f の方向微分は

$$(\Delta l)[g] = \langle \mathcal{G}l, \underline{\Delta l} \rangle \quad (50)$$

$$(\Delta p)[f] = \langle \underline{\Delta p}, \mathcal{F}p \rangle \quad (51)$$

と書ける。

多様体間の写像 $\mathcal{G}: \mathcal{L} \rightarrow \mathcal{P}$ は一般に、接空間どうしの写像 \mathcal{G}_* を引き起こす。点 $l \in \mathcal{L}$ における接空間 $T_l \mathcal{L}$ から対応する点 $p = \mathcal{G}l \in \mathcal{P}$ における接空間 $T_{\mathcal{G}l} \mathcal{P}$ へのこの線形写像を、

$$\mathcal{G}_*^{(l)}: T_l \mathcal{L} \rightarrow T_{\mathcal{G}l} \mathcal{P} \quad (52)$$

もしくは単に \mathcal{G}_* で表す。有限次元の場合に成分で書けば*4、

$$l = (l^1, \dots, l^n)^T \quad (53)$$

$$\Delta l = (\Delta l^1, \dots, \Delta l^n)^T \quad (54)$$

に対して

$$\mathcal{G}_*^{(l)} \Delta l = (\Delta p_1, \dots, \Delta p_n) \quad (55)$$

$$\Delta p_j = \sum_k \frac{\partial^2 g(l)}{\partial l^j \partial l^k} \Delta l^k \quad (56)$$

のように、接空間の対応づけ $\mathcal{G}_*^{(l)}$ はポテンシャル関数 g の二階微分から定まる。一般の場合も、 $|\epsilon| \ll 1$ に対して

$$\mathcal{G}(l + \epsilon \Delta l) = \mathcal{G}(l) + \epsilon \Delta p + o(\epsilon) \in \mathcal{P} \quad (57)$$

*4 このような言い方をするときには常に、ベクトル空間としての座標成分を表すという約束にする。つまり、何らかの基底ベクトル e_1, \dots, e_n を用いて $l = \sum l^k e_k$ と書かれるような線形な座標 (l^1, \dots, l^n) に限って成り立つ結果を述べる。実際、局所座標での単なる二階微分がテンソルにならないことは周知のとおりで、非線形な座標変換をすると (56) の形の表記は保存されない。

となるような接ベクトル $\Delta p \in T_{\mathcal{G}l}\mathcal{P}$ が $\mathcal{G}_*^{(l)}\Delta l$ だと思えばよい。なお、抽象的な本来の定義は、[14] など多様体論の適当な教科書を参照されたい*5。

前述の同一視によって点 $l \in \mathcal{L}$ を接ベクトル $\bar{l} \in T_l\mathcal{L}$ とみなしたとしても、 $\mathcal{G}l$ と $\mathcal{G}_*^{(l)}\bar{l}$ とは別物である。 $\mathcal{G} = dg$ は (14) のとおりポテンシャル関数 g の一階微分から定まり、 $\mathcal{G}_*^{(l)}$ は (56) のとおり g の二階微分から定まる。

逆向きも同様であり、 $T_p\mathcal{P}$ から $T_{\mathcal{F}p}\mathcal{L}$ への対応づけ $\mathcal{F}_*^{(p)}$ ないし \mathcal{F}_* が定められる。なお、 $\mathcal{G}\mathcal{F}p = p$ および $\mathcal{F}\mathcal{G}l = l$ から当然

$$\mathcal{G}_*\mathcal{F}_*\Delta p = \Delta p \quad (58)$$

$$\mathcal{F}_*\mathcal{G}_*\Delta l = \Delta l \quad (59)$$

が成り立つ。

2.3 計量

点 $l \in \mathcal{L}$ における接ベクトル $\Delta l, \Delta m \in T_l\mathcal{L}$ に対し、その内積を

$$(\Delta l, \Delta m)_l = \langle \underline{\mathcal{G}_*^{(l)}\Delta l}, \underline{\Delta m} \rangle \quad (60)$$

と定義する。上式の右辺では、前述の同型に基き接ベクトルを \mathcal{P}, \mathcal{L} 自身の元と同一視している。有限次元の場合に成分で書けば、

$$l = (l^1, \dots, l^n)^T \quad (61)$$

$$\Delta l = (\Delta l^1, \dots, \Delta l^n)^T \quad (62)$$

$$\Delta m = (\Delta m^1, \dots, \Delta m^n)^T \quad (63)$$

に対して

$$(\Delta l, \Delta m)_l = \sum_{j,k} \frac{\partial^2 g(l)}{\partial l^j \partial l^k} \Delta l^j \Delta m^k \quad (64)$$

となり、ポテンシャル関数 g の二階微分が計量テンソルの成分表示

$$G_{jk}(l) = \frac{\partial^2 g(l)}{\partial l^j \partial l^k} \quad (65)$$

*5 接ベクトルとは関数に対して値を定める線形汎関数だという本来の定義に基き、 \mathcal{P} 上の点 $\mathcal{G}l$ の近傍で定義された実数値関数 h に対して

$$(\Delta p)[h] = (\Delta l)[h \circ \mathcal{G}]$$

により $\mathcal{G}_*^{(l)}\Delta l = \Delta p \in T_{\mathcal{G}l}\mathcal{L}$ を定義する。

を与える。

この $(\cdot, \cdot)_l$ が内積の定義を満たすことは、次のように示される。作り方から双線形性は自明である。また、(57) および方向微分の性質 (18) から

$$(\Delta l, \Delta m)_l = \langle \mathcal{G}_*^{(l)} \Delta l, \Delta m \rangle \quad (66)$$

$$= \left. \frac{d}{dt} \langle \mathcal{G}(l + t\Delta l), \Delta m \rangle \right|_{t=0} \quad (67)$$

$$= \left. \frac{\partial}{\partial t} \frac{\partial}{\partial s} g(l + t\Delta l + s\Delta m) \right|_{t=s=0} \quad (68)$$

なので、対称性も成り立つ。さらに、この式から特に

$$(\Delta l, \Delta l)_l = \left. \frac{d^2}{dt^2} g(l + t\Delta l) \right|_{t=0} \quad (69)$$

が言えるので、 g の凸性から正值性も保証される。以上により、多様体 \mathcal{L} 上に計量が導入された。

多様体 \mathcal{P} 上にも同様にして計量が導入される。すなわち、点 $p \in \mathcal{P}$ における接ベクトル $\Delta p, \Delta q \in T_p \mathcal{P}$ の内積が

$$(\Delta p, \Delta q)_p = \langle \Delta p, \mathcal{F}_*^{(p)} \Delta q \rangle \quad (70)$$

で定義される。有限次元の場合に成分で書けば、

$$p = (p_1, \dots, p_n) \quad (71)$$

$$\Delta p = (\Delta p_1, \dots, \Delta p_n) \quad (72)$$

$$\Delta q = (\Delta q_1, \dots, \Delta q_n) \quad (73)$$

$$(74)$$

に対して

$$(\Delta p, \Delta q)_p = \sum_{j,k} \frac{\partial^2 f(p)}{\partial p_j \partial p_k} \Delta p_j \Delta q_k \quad (75)$$

と表される。 \mathcal{P}, \mathcal{L} の対称性に関する同種の説明は自明なので、今後は片方しか述べないこともある。

一対一対応 \mathcal{G}, \mathcal{F} による \mathcal{L}, \mathcal{P} の同一視は、これらの計量と辻褄が合っている。すなわち、

$$(\Delta l, \Delta m)_l = (\mathcal{G}_* \Delta l, \mathcal{G}_* \Delta m)_{\mathcal{G}l} \quad (76)$$

$$(\Delta p, \Delta q)_p = (\mathcal{F}_* \Delta p, \mathcal{F}_* \Delta q)_{\mathcal{F}p} \quad (77)$$

が成り立つことを容易に確かめられる。実際、定義から

$$(\mathcal{G}_* \Delta l, \mathcal{G}_* \Delta m)_{\mathcal{G}l} = \langle \mathcal{G}_* \Delta l, \mathcal{F}_* \mathcal{G}_* \Delta m \rangle = \langle \mathcal{G}_* \Delta l, \Delta m \rangle = (\Delta l, \Delta m)_l \quad (78)$$

$$(\mathcal{F}_* \Delta p, \mathcal{F}_* \Delta q)_{\mathcal{F}p} = \langle \mathcal{G}_* \mathcal{F}_* \Delta p, \mathcal{F}_* \Delta q \rangle = \langle \Delta p, \mathcal{F}_* \Delta q \rangle = (\Delta p, \Delta q)_p \quad (79)$$

と言える。

また、次の性質も有用であり、後に何度も利用される。点 $l \in \mathcal{L}$ における接ベクトル $\Delta l \in T_l \mathcal{L}$ と点 $p \in \mathcal{P}$ における接ベクトル $\Delta p \in T_p \mathcal{P}$ とが、

$$p = \mathcal{G}l, \quad \Delta p = \mathcal{G}_* \Delta l \quad (80)$$

$$l = \mathcal{F}p, \quad \Delta l = \mathcal{F}_* \Delta p \quad (81)$$

により対応していたとしよう。このとき、ポテンシャル関数 g, f の方向微分は

$$(\Delta l)[g] = \langle \mathcal{G}l, \underline{\Delta l} \rangle = (\bar{p}, \Delta p)_p \quad (82)$$

$$(\Delta p)[f] = \langle \underline{\Delta p}, \mathcal{F}p \rangle = (\Delta l, \bar{l})_l \quad (83)$$

と表される。左側の等号が成り立つことは (50), (51) ですでに述べた。右側の等号は内積の定義をあてはめたにすぎない。ちなみに、 $\langle \mathcal{G}l, \underline{\Delta l} \rangle$ は $(\bar{l}, \Delta l)_l = \langle \mathcal{G}_* \bar{l}, \underline{\Delta l} \rangle$ とは別物である ($\mathcal{G}l$ と $\mathcal{G}_* \bar{l}$ の違いは 2.2 項で注意した)。

単なる計量ベクトル空間との混同を避けるため、最後に次のことを注意しておく。内積が定義されるのはあくまで接ベクトルどうしに対してであり、点 $p, q \in \mathcal{P}$ に対してではない。 \mathcal{P} がベクトル空間なことから点と接ベクトルとを同一視することは可能だが、内積の値は、どの点における接ベクトルとみなすかによって変わる。なお、上の定義と一見よく似た $\langle p, \mathcal{F}q \rangle$ は、 $p, q \in \mathcal{P}$ に関して一般に対称でなく、 $\langle p, \mathcal{F}p \rangle \geq 0$ も保証されず、 q については線形ですらない。

2.4 ダイバージェンス

後に確率分布と対応づける際の役割を意識して、ここからは \mathcal{P} の方を主、 \mathcal{L} の方を従と扱う。

\mathcal{P} 上の二点 p, q に対し、ダイバージェンス $D(p||q)$ を

$$D(p||q) = f(p) + g(\mathcal{F}q) - \langle p, \mathcal{F}q \rangle \quad (84)$$

$$= f(p) - f(q) - \langle (p - q), \mathcal{F}q \rangle \quad (85)$$

$$= -g(\mathcal{F}p) + g(\mathcal{F}q) + \langle p, (\mathcal{F}p - \mathcal{F}q) \rangle \quad (86)$$

で定義する [1, 2, 3, 17]。2 番目と 3 番目の等号は、(21) から $g(\mathcal{F}q) = \langle q, \mathcal{F}q \rangle - f(q)$ と $f(p) = \langle p, \mathcal{F}p \rangle - g(\mathcal{F}p)$ により成り立つ。上の定義は

$$D(p\|q) = f(p) - \tilde{f}_q(p), \quad \tilde{f}_q(p) = f(q) + \langle (p - q), \mathcal{F}q \rangle \quad (87)$$

とも書き直せる。つまり、 f を点 q のまわりで一次近似した \tilde{f}_q について、点 p における近似誤差が $D(p\|q)$ である。このことは序論でも指摘した。

以下に述べる事実から、 $D(p\|q)$ は p, q 間の距離の自乗に相当する量と解釈される。ただし、一般に $D(p\|q) \neq D(q\|p)$ となることに注意を要する。

すでに述べた Legendre 変換の性質 (Young の不等式) から

$$D(p\|q) \geq 0 \quad (88)$$

であり、等号は $p = q$ のとき、そしてそのときに限り成立する。さらに、次の関係が成り立つ。

$$D(p\|q) + D(q\|r) = D(p\|r) + \langle (p - q), (\mathcal{F}r - \mathcal{F}q) \rangle \quad (89)$$

実際、

$$D(p\|q) + D(q\|r) \quad (90)$$

$$= \{f(p) - f(q) - \langle (p - q), \mathcal{F}q \rangle\} + \{f(q) - f(r) - \langle (q - r), \mathcal{F}r \rangle\} \quad (91)$$

$$= f(p) - f(r) - \langle (p - q), \mathcal{F}q \rangle - \langle (q - p + p - r), \mathcal{F}r \rangle \quad (92)$$

$$= (f(p) - f(r) - \langle (p - r), \mathcal{F}r \rangle) - \langle (p - q), \mathcal{F}q - \mathcal{F}r \rangle \quad (93)$$

$$= D(p\|r) + \langle (p - q), (\mathcal{F}r - \mathcal{F}q) \rangle \quad (94)$$

のように変形することができる。特に、 $\langle (p - q), (\mathcal{F}r - \mathcal{F}q) \rangle = 0$ のときには

$$D(p\|q) + D(q\|r) = D(p\|r) \quad (95)$$

である (拡張 Pythagoras の定理 [1, 2, 3, 8])。2.1 項で述べたように空間 \mathcal{P} と空間 \mathcal{L} を同一視して一枚の絵に重ねれば、拡張 Pythagoras の定理は図 1 のように描かれる。

なお、次の意味で計量 $(\cdot, \cdot)_q$ はダイバージェンス D の二次近似となっている。点 $q \in \mathcal{P}$ における接ベクトル $\Delta q \in T_q \mathcal{P}$ と $|\epsilon| \ll 1$ に対し、

$$D(q + \epsilon \Delta q\|q) = \frac{1}{2} \epsilon^2 (\Delta q, \Delta q)_q + o(\epsilon^2) \quad (96)$$

ただし左辺の $q + \epsilon \Delta q$ は局所座標についての演算と解釈する。この関係を示すには、左辺

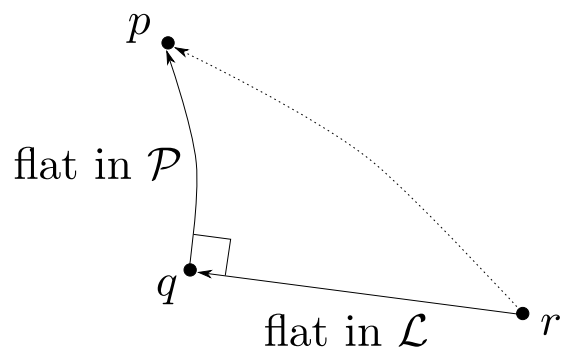


図 1 拡張 Pythagoras の定理

を展開して

$$D(q + \epsilon \Delta q \| q) = f(q + \epsilon \Delta q) - f(q) - \langle \epsilon \underline{\Delta q}, \mathcal{F}q \rangle \quad (97)$$

$$= \left\{ f(q) + \epsilon \frac{d}{d\delta} f(q + \delta \Delta q) \Big|_{\delta=0} + \frac{1}{2} \epsilon^2 \frac{d^2}{d\delta^2} f(q + \delta \Delta q) \Big|_{\delta=0} + o(\epsilon^2) \right\} - f(q) - \epsilon \langle \underline{\Delta q}, \mathcal{F}q \rangle \quad (98)$$

$$= \left\{ f(q) + \epsilon \langle \underline{\Delta q}, \mathcal{F}q \rangle + \frac{1}{2} \epsilon^2 (\Delta q, \Delta q)_q + o(\epsilon^2) \right\} - f(q) - \epsilon \langle \underline{\Delta q}, \mathcal{F}q \rangle \quad (99)$$

$$= \frac{1}{2} \epsilon^2 (\Delta q, \Delta q)_q + o(\epsilon^2) \quad (100)$$

となることを確かめればよい。途中の変形には、方向微分の性質 (83) や、(69) に相当する内積の性質を利用した^{*6}。

^{*6} 脚注*4 でも述べたとおり局所座標による単なる二階微分は一般にはテンソルにならない。しかしいまの場合は D の性質が効いて、座標に依存しない結果が得られる [5, 6, 7, 17]。実際、 q を固定したとき $D(p \| q)$ が $p = q$ で最小になることから、どんな（非線形の）局所座標でも

$$\frac{\partial D(p \| q)}{\partial p_i} \Big|_{p=q=r} = 0$$

であることが保証される。この両辺を r_j で偏微分すれば

$$\frac{\partial^2 D(p \| q)}{\partial p_j \partial p_i} \Big|_{p=q=r} + \frac{\partial^2 D(p \| q)}{\partial q_j \partial p_i} \Big|_{p=q=r} = 0$$

が得られる。左辺第 2 項は二変数関数 $D(p \| q)$ を各変数について一度ずつ偏微分しただけなので、座標変換に際して 2 階テンソルと同じ変換則に従う。右辺は恒等的に 0 だから、左辺第 1 項も同じ変換則に従わなくてはならない。序論の (1) も同様の事情による。

あるいはもっと直接的に、異なる座標 $(p_{i'})$ をとったときの変換則が

$$\begin{aligned} \frac{\partial^2 D(p \| q)}{\partial p_{j'} \partial p_{i'}} \Big|_{p=q=r} &= \sum_i \frac{\partial}{\partial p_{j'}} \left(\frac{\partial p_i}{\partial p_{i'}} \frac{\partial D(p \| q)}{\partial p_i} \right) \Big|_{p=q=r} \\ &= \sum_i \frac{\partial p_i}{\partial p_{i'}} \frac{\partial}{\partial p_{j'}} \frac{\partial D(p \| q)}{\partial p_i} \Big|_{p=q=r} + \sum_i \frac{\partial^2 p_i}{\partial p_{j'} \partial p_{i'}} \frac{\partial D(p \| q)}{\partial p_i} \Big|_{p=q=r} \\ &= \sum_{j,i} \frac{\partial p_j}{\partial p_{j'}} \frac{\partial p_i}{\partial p_{i'}} \frac{\partial^2 D(p \| q)}{\partial p_j \partial p_i} \Big|_{p=q=r} + 0 \end{aligned}$$

となることを確かめてもよい。これは二階テンソルの変換則そのものである。

2.5 双対接続

\mathcal{P}, \mathcal{L} はいずれももともと単なるベクトル空間なので、「平行移動」という操作が自然に可能である。これは明らかなことだが、念のため形式的にも書いておく。点 $p \in \mathcal{P}$ における接空間 $T_p\mathcal{P}$ は \mathcal{P} 自身と自然に同一視されるから、異なる二点 $p, q \in P$ の接空間に対しても

$$T_p\mathcal{P} \cong \mathcal{P} \cong T_q\mathcal{P} \quad (101)$$

という同一視ができる。つまり、点 p における接空間の元 $\Delta p \in T_p\mathcal{P}$ は点 q における接空間 $T_q\mathcal{P}$ の元へ自然に対応づけられる。この対応づけを点 p から点 q への接ベクトル Δp の平行移動と解釈する。このような接続 $\Pi_{\mathcal{P}}, \Pi_{\mathcal{L}}$ のもとで空間 \mathcal{P}, \mathcal{L} が平坦になること、すなわち平行移動が途中の経路によらないことは自明である。

ただし、 $\Pi_{\mathcal{P}}, \Pi_{\mathcal{L}}$ はそれぞれ単独では内積を保存しない。したがって、 $\Pi_{\mathcal{P}}, \Pi_{\mathcal{L}}$ は Levi-Civita 接続ではない。かわりに成り立つのは次の性質である。点 $p \in P$ における接ベクトル $\Delta p, \Delta q \in T_p\mathcal{P}$ を別の点 $p' \in P$ へそれぞれ移したものを $\Delta p', \Delta q' \in T_{p'}\mathcal{P}$ としよう。ただし、 Δp の方は $\Pi_{\mathcal{P}}$ で平行移動する。一方、 Δq の方は \mathcal{F}_* で $\mathcal{F}p \in \mathcal{L}$ の接ベクトルに移しておいて、 $\Pi_{\mathcal{L}}$ により $\mathcal{F}p' \in \mathcal{L}$ まで平行移動し、 \mathcal{G}_* で $p' \in P$ の接ベクトルに戻す。このとき、

$$(\Delta p, \Delta q)_p = (\Delta p', \Delta q')_{p'} \quad (102)$$

この意味で $\Pi_{\mathcal{P}}$ と $\Pi_{\mathcal{L}}$ は互いに双対接続 [1, 2, 3] と呼ばれる。

上式が成り立つことは、接続 $\Pi_{\mathcal{P}}, \Pi_{\mathcal{L}}$ の定義からほぼ自明であろう。接空間と空間自身とを同一視するとき、ベクトル空間 \mathcal{P}, \mathcal{L} の元として

$$\underline{\Delta p} = \underline{\Delta p'} \in \mathcal{P} \quad (103)$$

$$\underline{\mathcal{F}_* \Delta q} = \underline{\mathcal{F}_* \Delta q'} \in \mathcal{L} \quad (104)$$

なのだから、当然

$$(\Delta p, \Delta q)_p = \langle \underline{\Delta p}, \underline{\mathcal{F}_* \Delta q} \rangle \quad (105)$$

$$= \langle \underline{\Delta p'}, \underline{\mathcal{F}_* \Delta q'} \rangle \quad (106)$$

$$= (\Delta p', \Delta q')_{p'} \quad (107)$$

である。

2.6 ダイバージェンス最小化

n 次元部分多様体 $\mathcal{M} \subset \mathcal{P}$ と点 $\check{p} \in \mathcal{P}$ が与えられたとき、 \mathcal{M} 上で \check{p} に最も近い点 \hat{q} を求める問題を考える。近さは D で測るものとする。

$$\hat{q} = \arg \min_{q \in \mathcal{M}} D(\check{p} \| q) \quad (108)$$

拡張 Pythagoras の定理から、 \hat{q} において D が最小となるには、 $\hat{l} = \mathcal{F}\hat{q}$ が

$$\langle (\check{p} - \mathcal{G}\hat{l}), \underline{\Delta l} \rangle = 0 \quad \text{for all } \Delta l \in T_{\hat{l}}\mathcal{FM} \quad (109)$$

を満たさなくてはならない [1, 2, 3, 8]。 $\mathcal{FM} \subset \mathcal{L}$ は写像 \mathcal{F} による \mathcal{M} の像を表し、 $T_{\hat{l}}\mathcal{FM}$ は点 \hat{l} におけるその接空間を表す。

あるいは、次のような計算から (109) を直接導くこともできる。 \mathcal{FM} 上で定義された関数

$$h(l) = D(\check{p} \| \mathcal{G}l) = f(\check{p}) + g(l) - \langle \check{p}, l \rangle \quad (110)$$

に (82) を適用すれば、点 l における Δl 方向の方向微分が

$$(\Delta l)[h] = 0 + \langle \mathcal{G}l, \underline{\Delta l} \rangle - \langle \check{p}, \underline{\Delta l} \rangle = -\langle (\check{p} - \mathcal{G}\hat{l}), \underline{\Delta l} \rangle \quad (111)$$

となることがわかる。これが任意の接ベクトル $\Delta l \in T_{\hat{l}}\mathcal{FM}$ について 0 であるという要請を述べたのが (109) である。

特に、埋め込み $\mathcal{FM} \subset \mathcal{L}$ が $\Pi_{\mathcal{L}}$ に関して平坦^{*7}なときには、拡張 Pythagoras の定理から (109) が最小点の必要十分条件になる。よって特に、

$$\mathcal{M} = \mathcal{G}(\text{span}\{e_1, \dots, e_n\}) \quad (112)$$

と表される $\Pi_{\mathcal{L}}$ -平坦な \mathcal{M} 上での近似問題を解くには、

$$\langle \check{p}, e_j \rangle = \langle \hat{q}, e_j \rangle \quad (j = 1, 2, \dots, n) \quad (113)$$

を満たす点 $\hat{q} \in \mathcal{M}$ を求めればよい。

^{*7} \mathcal{FM} の接ベクトルを外側の接続 $\Pi_{\mathcal{L}}$ で平行移動してもやはり \mathcal{FM} の接ベクトルになるとき、埋め込み $\mathcal{FM} \subset \mathcal{L}$ は $\Pi_{\mathcal{L}}$ に関して平坦であると言う。あるいは、 \mathcal{FM} は自己平行であるとも言う。

2.7 シフトとスケーリング

最後に補足として、空間のシフトおよびスケーリングについて述べる。

$a > 0$ を正定数、 $e_0 \in \mathcal{L}$ を定ベクトルとし、 \mathcal{P} 上の新しいポテンシャル関数

$$\tilde{f}(p) = af(p) + \langle p, e_0 \rangle \quad (114)$$

を考えよう。この \tilde{f} から導かれる情報幾何の構造は、元の f による情報幾何の構造と次のように対応する。

$$\tilde{p} = p \in \mathcal{P} \quad (115)$$

$$\tilde{l} = al + e_0 \equiv \mathcal{A}l \in \mathcal{L} \quad (116)$$

$$\tilde{f}(\tilde{p}) = af(p) + \langle p, e_0 \rangle \quad (117)$$

$$\tilde{\mathcal{F}}\tilde{p} = d\tilde{f}(\tilde{p}) = \mathcal{A}\mathcal{F}p \quad (118)$$

$$\tilde{\mathcal{G}}\tilde{l} = \tilde{\mathcal{F}}^{-1}\tilde{l} = \mathcal{G}\mathcal{A}^{-1}\tilde{l} = \mathcal{G}l \quad (119)$$

$$\begin{aligned} \tilde{g}(\tilde{l}) &= \langle \tilde{\mathcal{G}}\tilde{l}, \tilde{l} \rangle - \tilde{f}(\tilde{\mathcal{G}}\tilde{l}) \\ &= \langle \mathcal{G}l, al + e_0 \rangle - (af(\mathcal{G}l) + \langle \mathcal{G}l, e_0 \rangle) \\ &= ag(l) \end{aligned} \quad (120)$$

$$\begin{aligned} \tilde{D}(\tilde{p}, \tilde{q}) &= \tilde{f}(\tilde{p}) - \tilde{f}(\tilde{q}) - \langle (\tilde{p} - \tilde{q}), \tilde{\mathcal{F}}\tilde{q} \rangle \\ &= (af(p) + \langle p, e_0 \rangle) - (af(q) + \langle q, e_0 \rangle) - \langle (p - q), (a\mathcal{F}q + e_0) \rangle \\ &= aD(p, q) \end{aligned} \quad (121)$$

$$\Delta\tilde{p} = \Delta p \quad (\text{同一視 } \Delta\tilde{p} \in T_{\tilde{p}}\mathcal{P} \cong \mathcal{P} \cong T_p\mathcal{P} \ni \Delta p \text{ のもとで}) \quad (122)$$

$$\Delta\tilde{l} = a\Delta l \quad (\text{同一視 } \Delta\tilde{l} \in T_{\tilde{l}}\mathcal{L} \cong \mathcal{L} \cong T_l\mathcal{L} \ni \Delta l \text{ のもとで}) \quad (123)$$

$$\tilde{\mathcal{F}}_*\Delta\tilde{p} = \mathcal{A}_*\mathcal{F}_*\Delta p = a\mathcal{F}_*\Delta p \quad (124)$$

$$\tilde{\mathcal{G}}_*\Delta\tilde{l} = \mathcal{G}_*\Delta l \quad (125)$$

$$(\Delta\tilde{p}, \Delta\tilde{q})_{\tilde{p}} = \langle \Delta\tilde{p}, \tilde{\mathcal{F}}_*\Delta\tilde{q} \rangle = \langle \Delta p, a\mathcal{F}_*\Delta q \rangle = a(\Delta p, \Delta q)_p \quad (126)$$

$$(\Delta\tilde{l}, \Delta\tilde{m})_{\tilde{l}} = \langle \tilde{\mathcal{G}}_*\Delta\tilde{l}, \Delta\tilde{m} \rangle = \langle \mathcal{G}_*\Delta l, a\Delta m \rangle = a(\Delta l, \Delta m)_l \quad (127)$$

この結果から特に次のことがわかる。

1. 空間 \mathcal{P}, \mathcal{L} のスケーリングやシフトは本質的ではない。それらはポテンシャル関数の簡単な変換により実現できる。
2. ダイバージェンスや計量における定数倍の違いも本質的ではない。この違いは空間 \mathcal{P}, \mathcal{L} のスケーリングにより吸収できる。

3 擬似平面および擬似球面上の情報幾何

本節では部分多様体上の情報幾何を論じる。特に、部分多様体の擬似平面性および擬似球面性という概念を導入し、それが情報幾何における様々な特性と密接に関係していることを述べる。なお、本節の「部分多様体」 P_1, L_1 は、通常の情報幾何における「全空間」（確率分布全体の集合）を指すことに注意されたい。外側の空間 \mathcal{P}, \mathcal{L} はその場合、正規化されていない不適切な分布まで含めた拡大空間になる。

アファイン微分幾何 [16] の用語で言えば、擬似平面はグラフはめ込みに、擬似球面は中心アファイン超曲面に相当する。しかし本稿では、両者の果す役割の対応を考慮して、二つの用語の釣合いがとれるよう端的な呼称を導入した。

3.1 部分多様体 P_1, L_1 上の情報幾何

P_1 を \mathcal{P} の連結な部分多様体とし、 \mathcal{F} によるその像を

$$L_1 = \mathcal{F}P_1 \subset \mathcal{L} \quad (128)$$

と置こう。このような P_1, L_1 上での情報幾何が応用上重要である。

P_1 上の点 p における接ベクトル $\Delta p \in T_p P_1$ を \mathcal{F}_* で移すと、得られる $\mathcal{F}_* \Delta p$ は対応する点 $l = \mathcal{F}p$ における L_1 の接ベクトルとなる。逆に、 L_1 の接ベクトルを \mathcal{G}_* で移せば P_1 の接ベクトルが得られる。この意味でも、一対一対応 \mathcal{F}, \mathcal{G} に基く $P_1 \subset \mathcal{P}$ と $L_1 \subset \mathcal{L}$ との同一視に破綻はない。

2.5 項までの議論は P_1, L_1 上でも当然成り立っている。ただし、拡張 Pythagoras の定理に現れる $(\mathcal{F}r - \mathcal{F}q)$ については、 $\mathcal{F}r$ と $\mathcal{F}q$ とを結ぶ \mathcal{L} 内の線分は一般に L_1 からはみ出してしまうことに注意を要する。 \mathcal{P} における $(p - q)$ の方も一般には同様である。また、点 $l \in L_1$ における (L_1 の) 接ベクトル $\Delta l \in T_l L_1$ を $\Pi_{\mathcal{L}}$ で平行移動した結果も、 L_1 の接空間からはみ出してしまう。すなわち、埋め込み $L_1 \subset \mathcal{L}$ は平坦ではない。 $P_1 \subset \mathcal{P}$ の方も一般には同様である。

性質の良い P_1, L_1 では、 $(\mathcal{F}r - \mathcal{F}q)$ や $(p - q)$ を射影して、拡張 Pythagoras の定理を P_1, L_1 上で記述することもできる*⁸。これについては 3.2.4 項で改めて述べる。

接続については一般に、はみ出た分を何らかの方向へ射影して P_1, L_1 の接空間へ戻し

*⁸ 射影すること自体はどんな P_1, L_1 でも可能だが、射影先で拡張 Pythagoras 型の命題が成り立つとは限らない。

てやることにより、 P_1, L_1 上の接続 Π_{P_1}, Π_{L_1} が定義される [16]。次のように考えると直感的に理解しやすいであろう：点 $p \in P_1$ における P_1 のベクトル $\Delta p \in T_p P_1$ をごく近くの点 $q \in P_1$ へ微小平行移動するには、まず $\Pi_{\mathcal{P}}$ で微小平行移動して $\Delta p \in T_q \mathcal{P}$ とみなし、さらにそれを線形部分空間 $T_q P_1 \subset T_q \mathcal{P}$ へ射影すればよい。離れた点への平行移動はこれを積分することで得られるが、結果は一般に経路に依存する。つまり、接続 Π_{P_1}, Π_{L_1} に関して多様体 P_1, L_1 自身は一般に平坦ではない。なお、射影の方向は、計量 $(\cdot, \cdot)_p$ に基いて P_1 の接空間へ直交射影を施すのが最も典型的である。今後は特に断らない限りこれに従う。そのとき、 P_1 上の共変微分 ∇ は、 \mathcal{P} における共変微分 D （つまり単なる方向微分）および P_1 の法ベクトル場 ξ を用いて

$$\nabla_X Y = D_X Y - h(X, Y)\xi \quad (129)$$

と表される。ここに X, Y は P_1 上の接ベクトル場であり、スカラー場 $h(X, Y)$ は $\nabla_X Y$ が P_1 に接するという条件から定まる。 h は対称な共変二階テンソルになり、 ξ に関するアファイン基本形式 (affine fundamental form) と呼ばれる [16]。

3.2 擬似平面

3.2.1 平面と擬似平面

ある $e_0 \in \mathcal{L}$ が存在し、部分多様体 $L_1 \subset \mathcal{L}$ 上の各点 l における法空間

$$N_l L_1 = \{\Delta k \in T_l \mathcal{L} \mid (\Delta k, \Delta l)_l = 0 \text{ for all } \Delta l \in T_l L_1\} \subset T_l \mathcal{L} \quad (130)$$

がいずれも（ベクトルとしての） $e_0 \in T_l \mathcal{L}$ で張られるとき、 L_1 を擬似平面と呼ぶことにする。部分多様体 $L_1 \subset \mathcal{L}$ が擬似平面であることは、対応する部分多様体 $P_1 = \mathcal{G}L_1 \subset \mathcal{P}$ が線形関数

$$\phi(p) = -\langle p, e_0 \rangle \quad (131)$$

の等位面であること^{*9}、すなわち C を定数として

$$P_1 = \{p \in \mathcal{P} \mid \langle p, e_0 \rangle = C\} \quad (132)$$

と書けることと同値になる。これは要するに、ベクトル空間 \mathcal{P} において P_1 が通常の意味での平面であることを主張している。

$$L_1 \text{ が擬似平面} \Leftrightarrow P_1 \text{ が平面} (\Leftrightarrow \text{埋め込み } P_1 \subset \mathcal{P} \text{ が } \Pi_{\mathcal{P}} \text{ に関して平坦}) \quad (133)$$

$$P_1 \text{ が擬似平面} \Leftrightarrow L_1 \text{ が平面} (\Leftrightarrow \text{埋め込み } L_1 \subset \mathcal{L} \text{ が } \Pi_{\mathcal{L}} \text{ に関して平坦}) \quad (134)$$

^{*9} マイナスをつけて定義しているのは、後に (199) で定義する ψ との対応を意識しているためである。

同値性が成り立つ理由は次のとおりである。 L_1 上の点 l における (\mathcal{L} の) 接ベクトルを Δl としよう。対応する点 $p = gl \in P_1$ における (\mathcal{P} の) 接ベクトル $\Delta p = g_* \Delta l$ は、内積や接ベクトルの定義から

$$(\Delta l, e_0)_l = \langle \Delta p, e_0 \rangle = \left. \frac{d}{dt} \langle p + t \Delta p, e_0 \rangle \right|_{t=0} = -(\Delta p)[\phi] \quad (135)$$

を満たす。特に、内積 $(\Delta l, e_0)_l$ が 0 になることは、接ベクトル Δp による ϕ の方向微分が 0 なことと等価である。したがって、 L_1 の法空間が e_0 で張られることは、 P_1 が ϕ の等位面であることを意味する。なぜなら、前者は点 l において e_0 に直交する接ベクトル $\Delta l \in T_l \mathcal{L}$ の集合が接空間 $T_l L_1$ に一致することを主張しているからである。上式によりその主張は、方向微分 $\Delta p[\phi]$ が 0 となる接ベクトル $\Delta p \in T_p \mathcal{P}$ の集合が接空間 $T_p P_1$ に一致すると言いかえられる。これは P_1 が ϕ の等位面であることを意味する。

3.2.2 擬似平面上の接続

いま示したとおり、 $L_1 \subset \mathcal{L}$ が擬似平面なら $P_1 \subset \mathcal{P}$ は平面である。したがってこのとき、 P_1 上の接続 Π_{P_1} は、単に $\Pi_{\mathcal{P}}$ の定義域を制限するだけで得られる。埋め込み $P_1 \subset \mathcal{P}$ が接続 $\Pi_{\mathcal{P}}$ に関して平坦なことも、 P_1 自身が接続 Π_{P_1} に関して平坦なことも、共に自明である。

一方、埋め込み $L_1 \subset \mathcal{L}$ は接続 $\Pi_{\mathcal{L}}$ に関して一般に平坦ではない。このため、 L_1 上の接続を定義する際には、3.1 節で述べたようにはみ出た分を何らかの方向へ射影する必要がある。すでに約束したとおり、擬似平面 L_1 に対して接続 Π_{L_1} と単に書いたら、直交射影に基く接続を指すことにしよう。

擬似平面 L_1 の場合は、この直交射影がどの点 $l \in L_1$ においても一定方向 $e_0 \in \mathcal{L}$ となる。射影の方向が経路上の全行程で常に一定なため、平行移動は経路に依存しない。つまり、 L_1 自身は Π_{L_1} に関して平坦である。実際、点 $l \in L_1$ における L_1 の接ベクトル $\Delta l \in T_l L_1$ を別の点 $l' \in L_1$ へ平行移動した結果は

$$\Delta l' = \Delta l - c e_0 \quad (136)$$

の形にしかなり得ず、実数 c は $\Delta l' \in T_{l'} L_1$ という条件、すなわち $(\Delta l', e_0)_{l'} = 0$ という条件から定められる。したがって

$$\Delta l' = \Delta l - \frac{(\Delta l, e_0)_{l'}}{(e_0, e_0)_{l'}} e_0 \quad (137)$$

という具体的な変換式が得られる。これは微小移動に限らず離れた l, l' 間でも成立し、 $\Delta l'$ は平行移動の経路には依存しない。

平坦性に関する以上の議論をまとめておこう。 L_1 が擬似平面のとき、

- P_1 自身は Π_{P_1} -平坦であり、埋め込み $P_1 \subset \mathcal{P}$ も $\Pi_{\mathcal{P}}$ -平坦
- L_1 自身は Π_{L_1} -平坦だが、埋め込み $L_1 \subset \mathcal{L}$ は一般に $\Pi_{\mathcal{L}}$ -平坦ではない

最後に、接続 Π_{P_1}, Π_{L_1} の双対性も容易に示される。点 $p \in P_1$ における接ベクトル $\Delta p, \Delta q \in T_p P_1$ を別の点 p' へ移すとしよう。 Δp の方は Π_{P_1} により平行移動する。このとき、得られる $\Delta p' \in T_{p'} P_1$ はベクトルとしては

$$\Delta p' = \Delta p \in P_1 \quad (138)$$

である。一方、 Δq については、対応する $\Delta l = \mathcal{F}_* \Delta q$ を点 $l = \mathcal{F}p$ から $l' = \mathcal{F}p'$ まで Π_{L_1} により平行移動し、得られた

$$\Delta l' = \Delta l - ce_0 \quad (139)$$

を \mathcal{G}_* で戻して

$$\Delta q' = \mathcal{G}_* \Delta l' \in T_{p'} P_1 \quad (140)$$

とする。規格化定数 c の具体的表記は不要なので省略した。移動後の内積は

$$(\Delta p', \Delta q')_{p'} = \langle \Delta p', \Delta l' \rangle \quad (141)$$

$$= \langle \Delta p, (\Delta l - ce_0) \rangle \quad (142)$$

$$= \langle \Delta p, \Delta l \rangle - c \langle \Delta p, e_0 \rangle \quad (143)$$

$$= (\Delta p, \Delta q)_p + 0 = (\Delta p, \Delta q)_p \quad (144)$$

となり、確かに移動前と等しい。

3.2.3 擬似平面内での平坦性

擬似平面 L_1 への埋め込み $\mathcal{M} \subset L_1$ が接続 Π_{L_1} に関して平坦なことは、 \mathcal{M} を拡張した多様体

$$\tilde{\mathcal{M}} = \{l + ce_0 \mid l \in \mathcal{M}, c \in \mathbb{R}\} \subset \mathcal{L} \quad (145)$$

を考えたときに、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ が接続 $\Pi_{\mathcal{L}}$ に関して平坦なことと同値になる。これを示すには次の補題を用いればよい：点 $\tilde{l} \in \tilde{\mathcal{M}}$ における \mathcal{L} の接ベクトル $\Delta \tilde{l}$ が $\tilde{\mathcal{M}}$ に接することは、

$$\tilde{l} = l + ce_0, \quad l \in \mathcal{M}, c \in \mathbb{R} \quad (146)$$

$$\Delta \tilde{l} = \Delta l + be_0, \quad \Delta l \in T_l \mathcal{M}, b \in \mathbb{R} \quad (147)$$

と表せることと同値である。

上の補題を用いて、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ の平坦性から埋め込み $\mathcal{M} \subset L_1$ の平坦性をまず導く。点 $l \in \mathcal{M}$ における \mathcal{M} の接ベクトル Δl を別の点 $l' \in \mathcal{M} \rightarrow \Pi_{\mathcal{L}}$ -平行移動して、 \mathcal{L} の接ベクトル $\Delta \tilde{l}'$ を得たとしよう。 $\Pi_{\mathcal{L}}$ とは「何もしない接続」だったから、ベクトルとしては単に $\Delta \tilde{l}' = \Delta l \in \mathcal{L}$ である。ここでもし埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ が接続 $\Pi_{\mathcal{L}}$ に関して平坦なら、 $\Delta \tilde{l}'$ は $\tilde{\mathcal{M}}$ の接ベクトルにならなくてははいけない。つまり、ベクトル Δl は点 l' においても $\tilde{\mathcal{M}}$ に接していなくてははいけない。補題から、このとき l' における \mathcal{M} の接ベクトル $\Delta l'$ と実数 b' が存在して、

$$\Delta \tilde{l}' = \Delta l' + b' e_0 \quad (148)$$

と表される。すると

$$\Delta l' = \Delta l - b' e_0 \quad (149)$$

の形の関係が導かれる。左辺の $\Delta l'$ は $\mathcal{M} \subset L_1$ の接ベクトルだから、当然 L_1 の接ベクトルでもある。前項の議論をふまえると、上式は、 Δl を接続 Π_{L_1} で平行移動したものが $\Delta l'$ であることを表している。こうして結局、 \mathcal{M} の接ベクトルを Π_{L_1} -平行移動しても \mathcal{M} の接ベクトルになること、すなわち、埋め込み $\mathcal{M} \subset L_1$ の Π_{L_1} -平坦性が示された。

逆に、埋め込み $\mathcal{M} \subset L_1$ の平坦性から埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ の平坦性が導かれることも示そう。 $\tilde{\mathcal{M}}$ 上の点 \tilde{l} における $\tilde{\mathcal{M}}$ の接ベクトルを $\Delta \tilde{l}$ とし、これを $\tilde{\mathcal{M}}$ 上の別の点 $\tilde{l}' \rightarrow \Pi_{\mathcal{L}}$ -平行移動したとする。 $\Pi_{\mathcal{L}}$ とは「何もしない接続」だったから、平行移動の結果はベクトルとしては $\Delta \tilde{l} \in \mathcal{L}$ のまま変わらない。さて、点 \tilde{l}, \tilde{l}' は、 \mathcal{M} 上の点 l, l' と実数 c, c' を用いて

$$\tilde{l} = l + c e_0 \quad (150)$$

$$\tilde{l}' = l' + c' e_0 \quad (151)$$

の形に表される。また、上の補題から、接ベクトル $\Delta \tilde{l}$ も

$$\Delta \tilde{l} = \Delta l + b e_0 \quad (152)$$

の形で書ける。ここに、 Δl は点 l における \mathcal{M} の接ベクトル、 b は実数である。いま、接ベクトル $\Delta l \in T_l \mathcal{M}$ を点 l から $l' \rightarrow \Pi_{L_1}$ -平行移動した結果を $\Delta l'$ とおこう。前項の議論から、ある実数 b' があって、

$$\Delta l' = \Delta l - b' e_0 \quad (153)$$

と書ける。そして、 $\mathcal{M} \subset L_1$ が Π_{L_1} -平坦なら、 $\Delta l'$ も \mathcal{M} の接ベクトルにならないといけな。すると

$$\Delta \tilde{l} = \Delta l' + (b + b') e_0, \quad \Delta l' \in T_{l'} \mathcal{M} \quad (154)$$

という関係が成り立つから、ふたたび補題により、 $\Delta \tilde{l}$ は点 \tilde{l}' における $\tilde{\mathcal{M}}$ の接ベクトルである。こうして、 $\tilde{\mathcal{M}}$ の接ベクトルを $\Pi_{\mathcal{L}}$ -平行移動しても $\tilde{\mathcal{M}}$ の接ベクトルになること、すなわち、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ の $\Pi_{\mathcal{L}}$ -平坦性が示された。

以上により、埋め込み $\mathcal{M} \subset L_1$ の平坦性は埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ の平坦性と等価である。特に、 L_1 上の点 l, l' を結ぶ線分

$$A = \{(1-t)l + tl' \mid 0 \leq t \leq 1\} \subset \mathcal{L} \quad (155)$$

に対して拡張

$$\tilde{A} = \{r + ce_0 \mid r \in A, c \in \mathbb{R}\} \subset \mathcal{L} \quad (156)$$

を作ると、 $\tilde{A} \cap L_1$ が l, l' を結ぶ Π_{L_1} -測地線になる。

3.2.4 測地線に関する拡張 Pythagoras の定理

L_1 が 擬似平面の場合には、測地線を用いることにより、拡張 Pythagoras の定理を P_1, L_1 上で言い換えることができる。

対応 \mathcal{F}, \mathcal{G} により P_1 と L_1 を同一視して、点 $p \in P_1$ と $q \in P_1$ を Π_{P_1} -測地線で結び、 q と $r \in P_1$ を Π_{L_1} -測地線で結んだとしよう^{*10}。このとき、 q において二本の測地線が直交する（それぞれの方向の接ベクトル $\Delta p, \Delta r \in T_q P_1$ が $(\Delta p, \Delta r)_q = 0$ となる）なら、

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

である。これを示すには、 $\langle (p - q), (\mathcal{F}r - \mathcal{F}q) \rangle = 0$ であることを確認すればよい。 Π_{P_1}, Π_{L_1} の定義から

$$p - q \propto \Delta p \quad (157)$$

$$\mathcal{F}r - \mathcal{F}q \propto \mathcal{F}_*^{(q)} \Delta r - ce_0 \quad (c \text{ は実数}) \quad (158)$$

なので、確かに

$$\langle (p - q), (\mathcal{F}r - \mathcal{F}q) \rangle \propto \langle (p - q), \mathcal{F}_*^{(q)} \Delta r \rangle - c \langle (p - q), e_0 \rangle \quad (159)$$

$$= (p - q, \Delta r)_q + 0 \quad (160)$$

$$\propto (\Delta p, \Delta r) = 0 \quad (161)$$

が成り立っている。途中の式変形には、 $p, q \in P_1$ に対して $\langle p, e_0 \rangle = \langle q, e_0 \rangle$ という擬似平面の特徴を利用した。

^{*10} $\mathcal{F}q \in L_1$ と $\mathcal{F}r \in L_1$ とを Π_{L_1} -測地線で結び、それを \mathcal{G} により P_1 上へ引き戻せば、 q, r を結ぶ曲線が得られる。この曲線を P_1 における Π_{L_1} -測地線とみなす。

3.3 擬似球面

3.3.1 等ポテンシャル面と擬似球面

部分多様体 $P_1 \subset \mathcal{P}$ 上の各点 p における法空間

$$N_p P_1 = \{\Delta k \in T_p \mathcal{P} \mid (\Delta k, \Delta p)_p = 0 \text{ for all } \Delta p \in T_p P_1\} \subset T_p \mathcal{P} \quad (162)$$

がいずれも (ベクトルとしての) $p \in T_p \mathcal{P}$ 自身で張られるとき、 P_1 を擬似球面と呼ぶことにする。部分多様体 $P_1 \subset \mathcal{P}$ が擬似球面であることは、対応する部分多様体 $L_1 = \mathcal{F}P_1 \subset \mathcal{L}$ がポテンシャル関数の等位面であることと同値になる。

$$P_1 \text{ が擬似球面} \Leftrightarrow L_1 \text{ が } g \text{ の等位面} \quad (163)$$

$$L_1 \text{ が擬似球面} \Leftrightarrow P_1 \text{ が } f \text{ の等位面} \quad (164)$$

同値性が成り立つ理由は次のとおりである。 P_1 上の点 p における (\mathcal{P} の) 接ベクトルを Δp としよう。対応する点 $l = \mathcal{F}p \in L_1$ における (\mathcal{L} の) 接ベクトル $\Delta l = \mathcal{F}_* \Delta p$ は、(82) から

$$(\Delta l)[g] = (p, \Delta p)_p \quad (165)$$

を満たす。特に、内積 $(p, \Delta p)_p$ が 0 になることは、接ベクトル Δl による g の方向微分が 0 なことと等価である。したがって、 P_1 の法空間が p で張られることは、 L_1 が g の等位面であることを意味する。なぜなら、前者は点 p においてベクトル p に直交する接ベクトル $\Delta p \in T_p \mathcal{P}$ の集合が接空間 $T_p P_1$ に一致することを主張しているからである。上式によりその主張は、方向微分 $\Delta l[g]$ が 0 となる接ベクトル $\Delta l \in T_l \mathcal{L}$ の集合が接空間 $T_l L_1$ に一致すると言いかえられる。これは L_1 が g の等位面であることを意味する。

P_1 が擬似球面のときには、対応する $L_1 = \mathcal{F}P_1$ が g の等位面となることから、 $p, q \in P_1$ のダイバージェンスは

$$D(p||q) = -g(\mathcal{F}p) + g(\mathcal{F}q) + \langle p, (\mathcal{F}p - \mathcal{F}q) \rangle \quad (166)$$

$$= \langle p, (\mathcal{F}p - \mathcal{F}q) \rangle \quad (167)$$

と書ける。これはもちろん $\langle p, \mathcal{F}p \rangle - \langle p, \mathcal{F}q \rangle$ と書いても良いが、 $\langle p, \mathcal{F}(p - q) \rangle$ とは一般に異なることを念のため注意しておく。

この場合、不等式 $D(p||q) \geq 0$ から

$$\langle p, \mathcal{F}p \rangle \geq \langle p, \mathcal{F}q \rangle \quad (168)$$

が導かれ、等号は $p = q$ のとき、そしてそのときに限り成り立つ。したがって、本稿の枠組による擬似球面上の情報幾何は、[4] の構成法により得られる統計多様体の構造に一致する。ただし [4] は、部分多様体上で (168) が成り立つことの方を前提としている点で、本稿とは逆向きの議論になっている。[4] においては、(168) は計量の正值性を保証する役割を持つ。一方、本稿の立場では、(168) は擬似球面性の単なる言い換えである。実際、 $l = \mathcal{F}p$ と置くとき (168) は

$$\langle p, l \rangle = \max_{m \in L_1} \langle p, m \rangle \quad (169)$$

を意味するので、 l における L_1 の任意の接ベクトル Δl に対して

$$\langle p, \Delta l \rangle = 0 \quad (170)$$

が導かれる。これを P_1 の側で書けば、 p における P_1 の任意の接ベクトル $\Delta p = \mathcal{G}_* \Delta l$ に対して

$$(p, \Delta p)_p = 0 \quad (171)$$

という擬似球面の定義式になる。

3.3.2 擬似球面上の接続

擬似球面という条件を課しても、埋め込み $P_1 \subset \mathcal{P}$ は $\Pi_{\mathcal{P}}$ に関して一般にはやはり平坦でない。そのような場合に P_1 上の接続 Π_{P_1} が射影に基いて定義されることや、射影方向は直交射影が最も典型的なことは、すでに述べた。 P_1 が擬似球面のとき、点 $p \in P_1$ の接空間におけるこの直交射影の方向は、ベクトルとしての p で与えられる。すでに約束したとおり、擬似球面 P_1 に対して接続 Π_{P_1} と単に書いたら、この直交射影に基く接続を指すことにする。 P_1 自身は Π_{P_1} に関して一般には平坦でない。

3.3.3 擬似球面内での平坦性

P_1 が擬似球面であることは、 P_1, L_1 自身の平坦性も、埋め込み $P_1 \subset \mathcal{P}$, $L_1 \subset \mathcal{L}$ の平坦性も導かない。その話題はここまでとし、ここからは、擬似球面 P_1 への埋め込み $\mathcal{M} \subset P_1$ について平坦性を議論しよう。

ある線形部分空間 $\tilde{\mathcal{M}} \subset \mathcal{P}$ が存在して

$$\mathcal{M} = \tilde{\mathcal{M}} \cap P_1 \quad (172)$$

と書けるときには、埋め込み $\mathcal{M} \subset P_1$ は上の接続 Π_{P_1} に関して平坦である。一般の埋め込み $\mathcal{M} \subset P_1$ が Π_{P_1} に関して平坦か否かは、 \mathcal{M} を拡張した多様体

$$\tilde{\mathcal{M}} = \{cp \mid p \in \mathcal{M}, c > 0\} \subset \mathcal{P} \quad (173)$$

により判定できる。埋め込み $\mathcal{M} \subset P_1$ が接続 Π_{P_1} に関して平坦なことは、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ が接続 $\Pi_{\mathcal{P}}$ に関して平坦なことと同値になる。これを示すには、次の補題を用いればよい：点 $\tilde{p} \in \tilde{\mathcal{M}}$ における \mathcal{P} の接ベクトル $\Delta\tilde{p}$ が $\tilde{\mathcal{M}}$ に接することは、

$$\tilde{p} = cp, \quad p \in \mathcal{M}, c > 0 \quad (174)$$

$$\Delta\tilde{p} = \Delta p + bp, \quad \Delta p \in T_p\mathcal{M}, b \in \mathbb{R} \quad (175)$$

と表せることと同値である。

上の補題を用いて、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ の平坦性から埋め込み $\mathcal{M} \subset P_1$ の平坦性をまず導く。点 $p \in \mathcal{M}$ における \mathcal{M} の接ベクトル Δp をごく近くの点 $p' \in \mathcal{M} \rightarrow \Pi_{\mathcal{P}}$ -微小平行移動して、 \mathcal{P} の接ベクトル $\Delta p'$ を得たとしよう^{*11}。 $\Pi_{\mathcal{P}}$ とは「何もしない接続」だったから、ベクトルとしては単に $\Delta p' = \Delta p \in \mathcal{P}$ である。ここでもし埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ が接続 $\Pi_{\mathcal{P}}$ に関して平坦なら、 $\Delta p'$ は $\tilde{\mathcal{M}}$ の接ベクトルにならなくてははいけない。つまり、ベクトル Δp は点 p' においても $\tilde{\mathcal{M}}$ に接していなくてははいけない。補題から、このとき p' における \mathcal{M} の接ベクトル $\Delta p'$ と実数 b' が存在して、

$$\Delta\tilde{p}' = \Delta p' + b'p' \quad (176)$$

と表される。すると

$$\Delta p' = \Delta p + b'p' \approx \Delta p + b'p \quad (177)$$

の形の関係が導かれる ($b'p'$ と $b'p$ との差は、 $\Delta p' - \Delta p = b'p'$ に比べて高次の微量なのでいまは無視してよい)。左辺の $\Delta p'$ は $\mathcal{M} \subset P_1$ の接ベクトルだから、当然 P_1 の接ベクトルでもある。 Π_{P_1} の定義をふまえると、上式は、 Δp を接続 Π_{P_1} で微小平行移動したものが $\Delta p'$ であることを表している。こうして結局、 \mathcal{M} の接ベクトルを Π_{P_1} -平行移動しても \mathcal{M} の接ベクトルになること、すなわち、埋め込み $\mathcal{M} \subset P_1$ の Π_{P_1} -平坦性が示された。

逆に、埋め込み $\mathcal{M} \subset P_1$ の平坦性から埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ の平坦性が導かれることも示そう。 $\tilde{\mathcal{M}}$ 上の点 \tilde{p} における $\tilde{\mathcal{M}}$ の接ベクトルを $\Delta\tilde{p}$ とし、これを $\tilde{\mathcal{M}}$ 上のごく近くの点 $\tilde{p}' \rightarrow \Pi_{\mathcal{P}}$ -平行移動したとする。 $\Pi_{\mathcal{P}}$ とは「何もしない接続」だったから、平行移動の結果はベクトルとしては $\Delta\tilde{p} \in \mathcal{P}$ のまま変わらない。さて、点 \tilde{p}, \tilde{p}' は、 \mathcal{M} 上の点 p, p' と実数 $c, c' > 0$ を用いて

$$\tilde{p} = cp \quad (178)$$

$$\tilde{p}' = c'p' \quad (179)$$

^{*11} P_1 自身が Π_{P_1} -平坦でないため、「微小な」平行移動を議論する必要がある。この点が擬似平面とは異なる。

の形に表される。また、上の補題から、接ベクトル $\Delta\tilde{p}$ も

$$\Delta\tilde{p} = \Delta p + bp \quad (180)$$

の形で書ける。ここに、 Δp は点 p における \mathcal{M} の接ベクトル、 b は実数である。いま、接ベクトル $\Delta p \in T_p\mathcal{M}$ を点 p から $p' \in \Pi_{P_1}$ -微小平行移動した結果を $\Delta p'$ とおこう。 Π_{P_1} の定義から、ある実数 b' があって、

$$\Delta p' \approx \Delta p - b'p \quad (181)$$

と書ける。そして、 $\mathcal{M} \subset P_1$ が Π_{P_1} -平坦なら、 $\Delta p'$ も \mathcal{M} の接ベクトルにならないといけない。すると

$$\Delta\tilde{p} \approx \Delta p' + (b + b')p \approx \Delta p' + (b + b')p', \quad \Delta p' \in T_{p'}\mathcal{M} \quad (182)$$

という関係が成り立つ。なお、 $(b + b')p$ と $(b + b')p'$ との差は、 $\Delta\tilde{p}' - \Delta\tilde{p}$ に比べて高次の微少量なのでいまは無視してよい。したがって、ふたたび補題により、 $\Delta\tilde{p}$ は点 \tilde{p}' における $\tilde{\mathcal{M}}$ の接ベクトルである。こうして、 $\tilde{\mathcal{M}}$ の接ベクトルを $\Pi_{\mathcal{P}}$ -平行移動しても $\tilde{\mathcal{M}}$ の接ベクトルになること、すなわち、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ の $\Pi_{\mathcal{P}}$ -平坦性が示された。

以上により、埋め込み $\mathcal{M} \subset L_1$ の平坦性は埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ の平坦性と等価である。特に、 P_1 上の点 p, q を結ぶ線分

$$A = \{(1 - t)p + tq \mid 0 \leq t \leq 1\} \subset \mathcal{P} \quad (183)$$

に対して拡張

$$\tilde{A} = \{cr \mid r \in A, c \in \mathbb{R}\} \subset \mathcal{P} \quad (184)$$

を作ると、 $\tilde{A} \cap P_1$ が、 p, q を結ぶ P_1 内の Π_{P_1} 測地線になる。

3.3.4 平坦性に関する補足

前項で示した同値性についての補足として、アファイン微分幾何の記法を用いたより正確な証明もここに記しておく。

3.1 節で述べたように、 P_1 上の共変微分 ∇ は、 \mathcal{P} における共変微分 D （つまり単なる方向微分）および P_1 の法ベクトル場 $\xi(p) = -p$ を用いて

$$\nabla_X Y = D_X Y - h(X, Y)\xi \quad (185)$$

と表される。ここに X, Y は P_1 上の接ベクトル場である。

いま、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ が $\Pi_{\mathcal{P}}$ -平坦だったとしよう。つまり、 $\tilde{\mathcal{M}}$ 上の任意の接ベクトル場 \tilde{X}, \tilde{Y} に対して、 $D_{\tilde{X}}\tilde{Y}$ がふたたび $\tilde{\mathcal{M}}$ の接ベクトル場になるとしよう。特に、 \mathcal{M} の接ベクトル場 X, Y に対しても $D_X Y$ が $\tilde{\mathcal{M}}$ に接することが保証されるので、補題から

$$D_X Y(p) = Z(p) + b(p)p, \quad Z(p) \in T_p \mathcal{M}, \quad b(p) \in \mathbb{R}, \quad p \in \mathcal{M} \quad (186)$$

と書ける。 $Z = D_X Y - b\xi$ の形で P_1 の接ベクトル場 Z が得られたことから、この Z は $\nabla_X Y$ に他ならない。よって $\nabla_X Y$ は \mathcal{M} の接ベクトル場である。すなわち、埋め込み $\mathcal{M} \subset P_1$ は Π_{P_1} -平坦である。

逆を示すために、実数 $c > 0$ を用いたスケーリング

$$\kappa : \mathcal{P} \rightarrow \mathcal{P} : \tilde{p} \rightarrow \frac{1}{c}\tilde{p} \quad (187)$$

を考える。 κ は全単射である。しかも κ は $\tilde{\mathcal{M}}$ を $\tilde{\mathcal{M}}$ 自身へ移す。よってその微分 κ_* は、 $\tilde{\mathcal{M}}$ の接ベクトルを $\tilde{\mathcal{M}}$ の接ベクトルへと移す。また、 κ_* で $\tilde{\mathcal{M}}$ の接ベクトルに移るものは $\tilde{\mathcal{M}}$ の接ベクトルに限られる。さらに、 κ は接続 $\Pi_{\mathcal{P}}$ を保つ：

$$\kappa_*(D_{\tilde{X}}\tilde{Y}) = D_{\kappa_*(\tilde{X})}\kappa_*(\tilde{Y}) \quad (188)$$

以上から、 $D_{\tilde{X}}\tilde{Y}$ が $\tilde{\mathcal{M}}$ に接することは、 $D_{\kappa_*(\tilde{X})}\kappa_*(\tilde{Y})$ が $\tilde{\mathcal{M}}$ に接することと同値になる。

さて、いま埋め込み $\mathcal{M} \subset P_1$ が Π_{P_1} -平坦だったとしよう。つまり、 \mathcal{M} 上の任意の接ベクトル場 X, Y に対して $\nabla_X Y$ がふたたび \mathcal{M} の接ベクトル場になるとしよう。このとき埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ が $\Pi_{\mathcal{P}}$ -平坦であることを示す。つまり、 $\tilde{\mathcal{M}}$ 上の任意の接ベクトル場 \tilde{X}, \tilde{Y} に対して、 $D_{\tilde{X}}\tilde{Y}$ がふたたび $\tilde{\mathcal{M}}$ の接ベクトル場になることを示す。

$\tilde{\mathcal{M}}$ 上の点 \tilde{p} は一般に、ある実数 $c > 0$ を用いて $\tilde{p} = cp$ と書けるのであった ($p \in \mathcal{M}$)。前述の写像 κ を施せば、 \tilde{p} における議論は $\kappa(\tilde{p}) = p \in \mathcal{M}$ へと移される。よって、 \mathcal{M} 上の点 p で $D_{\tilde{X}}\tilde{Y}(p)$ が $\tilde{\mathcal{M}}$ に接することさえ保証できればよい。

ここで、 \tilde{X}, \tilde{Y} について補題を適用し、

$$\tilde{X} = X - a\xi \quad (189)$$

$$\tilde{Y} = Y - b\xi \quad (190)$$

のように分解しておく。 a, b は $\tilde{\mathcal{M}}$ 上のスカラー場、 X, Y は $\tilde{\mathcal{M}}$ 上の接ベクトル場であって、 $p \in \mathcal{M}$ と $c > 0$ に対して

$$X(cp), Y(cp) \in T_p \mathcal{M} \quad (191)$$

という条件を満たす^{*12}。すると $D_{\tilde{X}}\tilde{Y}$ は

$$D_{\tilde{X}}\tilde{Y} = D_{\tilde{X}}Y - D_{\tilde{X}}(b\xi) \quad (192)$$

$$= D_XY - aD_\xi Y - D_{\tilde{X}}(b\xi) \quad (193)$$

と分解される。得られた右辺の各項がいずれも、 \mathcal{M} 上の点 p において $\tilde{\mathcal{M}}$ に接することを示そう。

1. \mathcal{M} 上では X も Y も \mathcal{M} に接するから、すでに述べたとおり

$$D_XY(p) = \nabla_XY(p) + h(X, Y)\xi(p) \quad (194)$$

と書ける。前提から $\nabla_XY(p)$ はふたたび \mathcal{M} の接ベクトルとなるので、補題より $D_XY(p)$ は $\tilde{\mathcal{M}}$ に接する。

2. 任意の $c > 0$ に対し $Y(cp) \in T_p\mathcal{M}$ なのだから、

$$D_\xi Y(p) = \left. \frac{\partial}{\partial t} Y(p + tp) \right|_{t=0} = \left. \frac{\partial}{\partial c} Y(cp) \right|_{c=1} \quad (195)$$

も $T_p\mathcal{M}$ に属する。よって特に $D_\xi Y(p)$ は $\tilde{\mathcal{M}}$ にも接する。

3. 共変微分（方向微分）の性質より、

$$D_{\tilde{X}}(b\xi) = \tilde{X}[b]\xi + bD_{\tilde{X}}\xi \quad (196)$$

が成り立つ。右辺の第一項は ξ に比例し、第二項は $\xi(p) = -p$ から $-b\tilde{X}$ に等しい。よって補題から $D_{\tilde{X}}(b\xi)(p)$ はやはり $\tilde{\mathcal{M}}$ に接する。

これで、 $D_{\tilde{X}}\tilde{Y}$ は $\tilde{\mathcal{M}}$ の接ベクトル場になることが示された。つまり、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{P}$ は $\Pi_{\mathcal{P}}$ -平坦である。

3.3.5 有限次元平面との交わり

引き続き P_1 を擬似球面としよう。本項では、 \mathcal{L} における有限次元平面と L_1 との交わりについて議論する。対応 \mathcal{F}, \mathcal{G} による同一視を意識して言い直せば、ある定数 C と線形独立なベクトル $l_0, e_0, e_1, \dots, e_n \in \mathcal{L}$ を用いて

$$\mathcal{FM} = \left\{ l = l_0 + \sum_{k=0}^n \theta^k e_k \mid g(l) = C, \theta^0, \dots, \theta^n \in \mathbb{R} \right\} \subset L_1 = \{ l \in \mathcal{L} \mid g(l) = C \} \quad (197)$$

^{*12} 点 cp における接ベクトル $X(cp)$ を点 p まで $\Pi_{\mathcal{P}}$ -平行移動したと解釈してもよいし、 $\kappa_*(X)$ の点 p における値と解釈してもよい。

と表されるような n 次元部分多様体 $\mathcal{M} \subset P_1$ について議論する^{*13}。このように

$$\mathcal{F}p = l = l_0 + \sum_{k=0}^n \theta^k e_k \quad (198)$$

と表したときの $\theta^0, \dots, \theta^n$ を、 $p \in \mathcal{M}$ ないし $l \in \mathcal{FM}$ の斉次 θ 座標と呼ぶことにする。

斉次 θ 座標には $g(l) = C$ という制約条件が課されるため、(縮退した場合を除けば) $\theta^0, \dots, \theta^n$ のうち n 個を指定すると残る 1 個も自動的に定まる。そこで、 $\theta = (\theta^1, \dots, \theta^n)$ を非斉次 θ 座標もしくは単に θ 座標と呼ぶことにして、 θ を指定したときの θ^0 の値を

$$\theta^0 = -\psi(\theta) \quad (199)$$

と置く。このときの $l \in \mathcal{L}$ および $p \in \mathcal{P}$ を

$$l(\theta) = l_0 + \sum_{k=1}^n \theta^k e_k - \psi(\theta) e_0, \quad p(\theta) = \mathcal{G}l(\theta) \quad (200)$$

と表そう^{*14}。すべての θ で恒等的に

$$g(l(\theta)) = g\left(l_0 + \sum_{k=1}^n \theta^k e_k - \psi(\theta) e_0\right) = C \quad (201)$$

が成り立つことから、両辺を θ^k で偏微分して

$$\langle p(\theta), e_k \rangle - \frac{\partial \psi(\theta)}{\partial \theta^k} \langle p(\theta), e_0 \rangle = 0 \quad (k = 1, \dots, n) \quad (202)$$

が得られる。なぜなら、 $g(l(\theta))$ を θ^k で偏微分することは、内側の偏微分

$$\frac{\partial l(\theta)}{\partial \theta^k} = e_k - \frac{\partial \psi(\theta)}{\partial \theta^k} e_0 \quad (203)$$

を点 $l(\theta)$ における接ベクトル Δl と解釈して方向微分 $(\Delta l)[g]$ を求めることに等しいからである。あとはこれに g の方向微分の性質 (82) を適用すればよい。

$$\frac{\partial g(l(\theta))}{\partial \theta^k} = (\Delta l)[g] = \langle p(\theta), \Delta l \rangle \quad (204)$$

$$= \langle p(\theta), e_k \rangle - \frac{\partial \psi(\theta)}{\partial \theta^k} \langle p(\theta), e_0 \rangle \quad (205)$$

^{*13} ただし、 $l_0 = o$ は特例として許し、その場合は残りの e_0, e_1, \dots, e_n が線形独立であることを要求する。

^{*14} このとき、ある $r^0 \in \mathcal{P}$ を用いて $\psi(\theta) = -\langle r^0, l(\theta) \rangle$ と表すことができる。この意味で ψ は (131) の ϕ と対応するものである。なお、 r^0 としては、 $\langle r^0, l_0 \rangle = \langle r^0, e_1 \rangle = \dots \langle r^0, e_n \rangle = 0$ かつ $\langle r^0, e_0 \rangle = 1$ であるような元をとればよい。

情報幾何の最も標準的な設定においては $\langle p(\theta), e_0 \rangle$ が常に一定値となり、その場合は関数 $\psi(\theta)$ に着目することが特に有用である (3.5.1 項および 4.5 項)。なお、関数 $\psi(\theta)$ もしばしばポテンシャル関数と呼ばれるが、本稿ではその呼称は f, g にのみ用いることにする。

3.3.6 ダイバージェンス最小化

3.3.5 項と同じ前提のもとで (197) のように表される部分多様体 $\mathcal{M} \subset P_1$ と点 $\check{p} \in P_1$ が与えられたとき、ダイバージェンスで測って \check{p} に最も近い点

$$\hat{q} = \arg \min_{q \in \mathcal{M}} D(\check{p} \| q) \quad (206)$$

を求めたいとしよう。擬似球面性から

$$D(\check{p} \| q) = \langle \check{p}, \mathcal{F}\check{p} \rangle - \langle \check{p}, \mathcal{F}q \rangle \quad (207)$$

なので、 D の最小化は $l = \mathcal{F}q$ に対する $\langle \check{p}, l \rangle$ の最大化に帰着される。そこで、Lagrange 未定係数法によりこれを解く。

まず、斉次座標

$$l = l_0 + \sum_{k=0}^n \theta^k e_k, \quad q = \mathcal{G}l \quad (208)$$

を用いて

$$J = \langle \check{p}, l \rangle - \lambda g(l) \quad (209)$$

と置く。そして、 θ^k による J の偏微分が 0 になることを要求すれば、3.3.5 項と同様に g の方向微分の性質 (82) から

$$\frac{\partial J}{\partial \theta^k} = \langle \check{p}, e_k \rangle - \lambda \langle q, e_k \rangle = 0 \quad (k = 0, \dots, n) \quad (210)$$

を得る。したがって結局、

$$\begin{pmatrix} \langle \check{p}, e_0 \rangle \\ \vdots \\ \langle \check{p}, e_n \rangle \end{pmatrix} \propto \begin{pmatrix} \langle \hat{q}, e_0 \rangle \\ \vdots \\ \langle \hat{q}, e_n \rangle \end{pmatrix} \quad (211)$$

を満たす $\hat{q} \in \mathcal{M}$ をみつけ出せばよい。

以上をふまえ、 $\langle \cdot, \cdot \rangle$ の双線形性に着目すれば、次のような手順で \hat{q} を求めることができる。

1. \mathcal{M} を拡張した多様体 $\tilde{\mathcal{M}}$ を作る。

$$\tilde{\mathcal{M}} = \{cq \mid q \in \mathcal{M}, c \in \mathbb{R}\} \subset \mathcal{P} \quad (212)$$

2. $\tilde{\mathcal{M}}$ 上の点 \tilde{q} で、

$$\langle \tilde{p}, e_k \rangle = \langle \tilde{q}, e_k \rangle \quad (k = 0, \dots, n) \quad (213)$$

を満たすものを探す。

3. \tilde{q} を正規化して \hat{q} を得る。すなわち、 $c\tilde{q} \in \mathcal{M}$ となるような実数 c を求め、 $\hat{q} = c\tilde{q} \in \mathcal{M}$ を答える。

3.3.7 線形制約のもとでの線形関数最小化

後ほど q -正規分布 [13] を導出するための準備として、擬似球面 $P_1 \subset \mathcal{P}$ に線形制約が与えられたときの線形関数最小化について最後に述べておく。

いま、線形独立な $e_0, e_1, \dots, e_n \in \mathcal{L}$ と定数 η_1, \dots, η_n が指定されたとしよう。このとき、 $p \in P_1$ について、制約

$$\langle p, e_k \rangle = \eta_k \quad (k = 1, \dots, n) \quad (214)$$

のもとで関数

$$\phi(p) = -\langle p, e_0 \rangle \quad (215)$$

を最小化する問題を考える^{*15}。

指定された $\eta = (\eta_1, \dots, \eta_n)$ に対して、制約を満たすような P_1 上の点の集合を $R(\eta)$ と置こう。点 $p \in R(\eta)$ における接ベクトル $\Delta p \in T_p \mathcal{P}$ が $R(\eta)$ の接ベクトルであるためには、 P_1 の接ベクトルであり、かつ制約式の左辺 $\langle p, e_k \rangle$ を変化させないことが必要十分である ($k = 1, \dots, n$)。 P_1 の接ベクトルであることは、 $l = \mathcal{F}p$, $\Delta l = \mathcal{F}_* \Delta p$ と置いて L_1 が擬似球面であることを用いれば

$$0 = (\Delta l, l)_l = \langle \Delta p, l \rangle = \langle \Delta p, \mathcal{F}p \rangle \quad (216)$$

と言いかえられる。また、 $\langle p, e_k \rangle$ を変化させないことは、関数 $h_k(p) = \langle p, e_k \rangle$ の方向微分が 0 であると言いかえられる。

$$0 = (\Delta p)[h_k] = \langle \Delta p, e_k \rangle \quad (k = 1, \dots, n) \quad (217)$$

^{*15} この関数は (131) と同じ形であり、マイナスをつけた理由も同じである。

したがって、点 $p \in R(\eta)$ における $R(\eta)$ の接空間は

$$T_p R(\eta) = \{\Delta p \in T_p \mathcal{P} \mid \langle \Delta p, \mathcal{F}p \rangle = 0 \text{ かつ } \langle \Delta p, e_k \rangle = 0, k = 1, \dots, n\} \quad (218)$$

と表される。さて、関数 $\phi(p) = -\langle p, e_0 \rangle$ が点 p で最小値をとるためには、任意の接ベクトル $\Delta p \in T_p R(\eta)$ に対して $\langle \Delta p, e_0 \rangle = 0$ でなくてはならない。したがって、

$$e_0 \in \text{span}\{\mathcal{F}p, e_1, \dots, e_n\} \quad (219)$$

を得る。 $e_0, e_1, \dots, e_n \in \mathcal{P}$ の線形独立性から、これは

$$\mathcal{F}p \in \text{span}\{e_0, e_1, \dots, e_n\} \quad (220)$$

を意味する。前提から $\mathcal{F}p \in L_1$ であることもあわせれば、結局

$$\mathcal{F}p \in \text{span}\{e_0, e_1, \dots, e_n\} \cap L_1 \equiv \mathcal{FM} \quad (221)$$

が言えた。

\mathcal{FM} は L_1 の部分多様体であり、3.3.5 項で議論した「 \mathcal{L} における有限次元平面と L_1 との交わり」の一例となっている。

3.4 P_1, L_1 が共に擬似球面の場合

3.4.1 P_1, L_1 上の情報幾何

P_1, L_1 が共に擬似球面であることは、 P_1, L_1 がそれぞれ f, g の等位面であることを意味する。このとき

$$f(p) + g(\mathcal{F}p) = \langle p, \mathcal{F}p \rangle \quad (222)$$

も P_1 上で一定値となる。また、 $p, q \in P_1$ に対するダイバージェンスは

$$D(p||q) = f(p) - f(q) - \langle (p - q), \mathcal{F}q \rangle \quad (223)$$

$$= \langle (q - p), \mathcal{F}q \rangle \quad (224)$$

$$= \text{定数} - \langle p, \mathcal{F}q \rangle \quad (225)$$

と表される。

さらに、接続 Π_{P_1}, Π_{L_1} の双対性が次のように示される。 $|\epsilon| \ll 1$ とし、点 $p \in P_1$ における接ベクトル $\Delta p, \Delta q, \Delta r \in T_p P_1$ をとり、 $\Delta p, \Delta q$ を $p' \sim p + \epsilon \Delta r$ へ微小移動するとしよう。 \sim は $o(\epsilon)$ を除いて等しいことを表す。 Δp を Π_{P_1} により平行移動すると、

$$\Delta p' \sim \Delta p - \epsilon a p \quad (226)$$

の形になる。規格化定数 a の具体的表記はいまは必要ない。一方、 Δq については、 $\Delta l = \mathcal{F}_* \Delta q$ を点 $l = \mathcal{F}p$ から $l' = \mathcal{F}p'$ まで Π_{L_1} により平行移動した

$$\Delta l' \sim \Delta l - \epsilon b l \quad (227)$$

を \mathcal{G}_* で戻して

$$\Delta q' = \mathcal{G}_* \Delta l' \quad (228)$$

とする。規格化定数 b の具体的表記はやはり必要ない。移動後の内積は

$$(\Delta p', \Delta q')_{p'} = \langle \Delta p', \Delta l' \rangle \quad (229)$$

$$\sim \langle \Delta p, \Delta l \rangle - \epsilon a \langle p, \Delta l \rangle - \epsilon b \langle \Delta p, l \rangle \quad (230)$$

$$= \langle \Delta p, \Delta l \rangle - \epsilon a (p, \Delta p)_p - \epsilon b (\Delta l, l)_l \quad (231)$$

$$= \langle \Delta p, \Delta l \rangle + 0 + 0 = (\Delta p, \Delta q)_p \quad (232)$$

となり、確かに移動前と等しい。交叉項が 0 となることは擬似球面性による。

3.3.5 項では、 \mathcal{L} における有限次元平面と L_1 との交わり、つまり (197) のように表される部分多様体 $\mathcal{M} \subset P_1$ について議論した。いまの場合は L_1 も擬似球面なため、3.3.3 項で述べた平坦性の判定法を $\mathcal{F}\mathcal{M} \subset L_1$ にも適用できる。すると、(197) のように表される部分多様体 \mathcal{M} に対して以下が同値なことがわかる。

1. $l_0 = o$ 。つまり、 $\mathcal{F}\mathcal{M}$ は \mathcal{L} の線形部分空間と L_1 との交わりで表される。
2. 埋め込み $\mathcal{F}\mathcal{M} \subset L_1$ が接続 Π_{L_1} に関して平坦となる。

その場合のダイバージェンス最小化については 3.3.6 項で述べたとおりである。また、3.3.7 項の線形関数最小化の解集合が Π_{L_1} -平坦となることもわかる。

3.4.2 ポテンシャル関数の特徴づけ

適当な制限のもとで、 P_1 と L_1 が共に擬似球面となるのはべき乗型の f に限られることが証明される。

\mathcal{P} を n 次元ベクトル空間とし、成分表示

$$p = (p_1, \dots, p_n)^T \quad (233)$$

が与えられているとしよう。ポテンシャル関数 f としては、

$$f(p) = \sum_i h_i(p_i) \quad (234)$$

という成分ごとの関数の和で書けるものを考える。この f から情報幾何を構築したときに、 P_1, L_1 が共に擬似球面となるための条件を本項で求める。すなわち、 $f(p)$ が一定値となるような $p \in \mathcal{P}$ については

$$g(\mathcal{F}p) = \sum_i p_i \frac{\partial f(p)}{\partial p_i} - f(p) \quad (235)$$

も一定値となることを要求して、 h_1, \dots, h_n の満たすべき条件を求める。

この要求は、 f の等位面上で

$$\sum_i p_i \frac{\partial f(p)}{\partial p_i} \quad (236)$$

が一定であることを主張している。言い換えれば、 f を変えない方向の $\Delta p = (\Delta p_1, \dots, \Delta p_n)^T$ に対しては

$$\sum_i (p_i + \epsilon \Delta p_i) \frac{\partial f(p + \epsilon \Delta p)}{\partial p_i} = \sum_i p_i \frac{\partial f(p)}{\partial p_i} + o(\epsilon) \quad (|\epsilon| \ll 1) \quad (237)$$

が成り立つこと、すなわち

$$\sum_j \Delta p_j \frac{\partial f(p)}{\partial p_j} + \sum_j \sum_i p_i \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \Delta p_j = 0 \quad (238)$$

を主張している。ここで、 f を変えない方向の Δp とは、方向微分が

$$(\Delta p)[f] = \sum_j \frac{\partial f(p)}{\partial p_j} \Delta p_j = 0 \quad (239)$$

を満たす Δp のことだから、最後の主張の左辺第一項は自動的に 0 である。したがって、要求は

$$\sum_j \sum_i p_i \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \Delta p_j = 0 \quad (240)$$

と等価になる。(239) を満たすすべての Δp に対してこれが成り立つには、

$$\frac{\partial f(p)}{\partial p_j} \propto \sum_i p_i \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \quad (j = 1, \dots, n) \quad (241)$$

でなくてはならない（左辺と右辺の比が j によらない^{*16}）。

^{*16} \mathbf{u} に直交するすべての \mathbf{v} について \mathbf{w} が \mathbf{v} と直交するなら、 \mathbf{u} と \mathbf{w} は平行である。

f が (234) の形なら^{*17}、この条件は

$$h'_j(p_j) \propto p_j h''_j(p_j) \quad (j = 1, \dots, n) \quad (242)$$

と単純化される。つまり、 j によらないある定数 c があって

$$(c - 1)h'_j(p_j) = p_j h''_j(p_j) \quad (j = 1, \dots, n) \quad (243)$$

が成り立つ。この微分方程式を解くことにより、

$$h_j(p_j) = a_j p_j^c + b_j \quad (j = 1, \dots, n) \quad (244)$$

が結論される (a_j, b_j は任意の定数)。

以上から、(234) の形のポテンシャル関数 f について P_1, L_1 が共に擬似球面となるのは、

$$f(p) = \sum_i a_i p_i^c + b \quad (245)$$

というべき乗型に限られることが示された。ここに $b = \sum_i b_i$ と置いた。

3.5 P_1 が擬似球面で L_1 が擬似平面の場合

3.5.1 P_1, L_1 上の情報幾何

P_1 が擬似球面であり、しかも L_1 は擬似平面だったとしよう。すなわち、 L_1 が g の等位面であり、対応する埋め込み $P_1 \subset \mathcal{P}$ が $\Pi_{\mathcal{P}}$ に関して平坦だったとしよう。このとき、 P_1 は擬似球面であると同時に、通常の意味での平面でもある。つまり、 $e_0 \in \mathcal{L}$ と定数 $C \in \mathbb{R}$ が存在して

$$P_1 = \{p \in \mathcal{P} \mid \langle p, e_0 \rangle = C\} \quad (246)$$

と表せる^{*18}。以上は擬似平面と擬似球面の一般論ですでに述べたとおりである。

3.3.5 項では、 \mathcal{L} における有限次元平面と L_1 との交わり、つまり (197) のように表される部分多様体 $\mathcal{M} \subset P_1$ について議論した。その際用いたベクトル e_0, \dots, e_n のうち、 e_0 は上式を満たすよう選ぶことにしよう。すると、埋め込み $\mathcal{FM} \subset L_1$ は接続 Π_{L_1} に関して平坦になる。さらに、最終的に得られる関係式は

$$\langle p(\theta), e_k \rangle = C \frac{\partial \psi(\theta)}{\partial \theta^k} \quad (k = 1, \dots, n) \quad (247)$$

^{*17} (234) を仮定しないと他の形の解がある。実際たとえば、 $f(p_1, p_2) = p_1^{a_1} p_2^{a_2}$ (a_1, a_2 は定数) や $f(p_1, p_2) = p_2(p_1 - p_2) \exp(p_1/p_2)$ は (241) を満たす。付録 A.4 を参照。

^{*18} 必要なら $(1/C)e_0$ を改めて e_0 ととり直せば、 $C = 1$ にすることもできる。

という簡単な形になる。 C は定数である。この両辺をさらに θ^j で偏微分することにより、 θ 座標に関する計量テンソルの具体的表示

$$G_{jk}(\theta) = \left(\frac{\partial p(\theta)}{\partial \theta^j}, \frac{\partial p(\theta)}{\partial \theta^k} \right)_{p(\theta)} = \left\langle \frac{\partial p(\theta)}{\partial \theta^j}, \mathcal{F}_* \frac{\partial p(\theta)}{\partial \theta^k} \right\rangle = \left\langle \frac{\partial p(\theta)}{\partial \theta^j}, \frac{\partial \mathcal{F}p(\theta)}{\partial \theta^k} \right\rangle \quad (248)$$

$$= \left\langle \frac{\partial p(\theta)}{\partial \theta^j}, e_k \right\rangle = C \frac{\partial^2 \psi(\theta)}{\partial \theta^j \partial \theta^k} \quad (249)$$

が得られる。

また、その場合のダイバージェンス最小化については 3.3.6 項で述べたとおりである。しかも今の場合 $\check{p}, \hat{q} \in P_1$ から $\langle \check{p}, e_0 \rangle = \langle \hat{q}, e_0 \rangle$ が保証され、(211) の比例係数が 1 に限定される。したがって、

$$\langle \check{p}, e_k \rangle = \langle \hat{q}, e_k \rangle \quad (k = 1, \dots, n) \quad (250)$$

を満たす $\hat{q} \in \mathcal{M}$ がダイバージェンス最小化の解となる。

3.5.2 ポテンシャル関数の特徴づけ

適当な制限のもとで、 P_1 が擬似球面かつ L_1 が擬似平面となるのは対数型の f に限られることが証明される。

3.4.2 項と同様に、 \mathcal{P} を n 次元ベクトル空間とし、ポテンシャル関数 f は (234) のように成分ごとの関数 h_1, \dots, h_n の和で書けるものを考えよう^{*19}。この f から情報幾何を構築したときに、 P_1 が平坦かつ L_1 が等ポテンシャル面となるための条件を求める。すなわち、

$$g(\mathcal{F}p) = \sum_i p_i \frac{\partial f(p)}{\partial p_i} - f(p) \quad (251)$$

の等位面がベクトル空間 \mathcal{P} における平面となることを要求して、 h_1, \dots, h_n が満たすべき条件を求める。

この要求は、関数 $\tilde{g}(p) = g(\mathcal{F}p)$ の勾配ベクトルが p によらず常に一定方向となることを意味する。言い換えれば、ある定ベクトル $c = (c_1, \dots, c_n)^T$ と実数値関数 $\lambda: \mathcal{P} \rightarrow \mathbb{R}$ が存在して、

$$\frac{\partial}{\partial p_j} \left(\sum_i p_i \frac{\partial f(p)}{\partial p_i} - f(p) \right) = \lambda(p) c_j \quad (j = 1, \dots, n) \quad (252)$$

^{*19} (234) を仮定しない議論は付録 A.4 を参照。

を満たさなくてはならない。この式を整理すると

$$\sum_i p_i \frac{\partial^2 f}{\partial p_i \partial p_j} = \lambda(p) c_j \quad (j = 1, \dots, n) \quad (253)$$

が得られ、(234) の場合には特に

$$p_j h_j''(p_j) = \lambda(p) c_j \quad (j = 1, \dots, n) \quad (254)$$

となる。左辺は p_j にしか依存しないから、 $\lambda(p)$ も p_j にしか依存できない。すべての j でそうなることから、次のいずれかが成り立っている。

1. c_1, \dots, c_n は一つを除いてすべて 0
2. $\lambda(p)$ は定数

前者は自明である。すなわち、一般性を失うことなく $c_2 = \dots = c_n = 0$ とすると $f(p) = C(p_1) + \sum_j a_j p_j$ の形で、 $g(\mathcal{F}p)$ が p_1 のみの関数となるから等位面は平面「 $p_1 = \text{一定}$ 」である (a_1, \dots, a_n は任意の定数、 C は任意の関数)。

後者の場合は、 $\lambda(p) c_j$ を改めて定数 c_j と置き直し、微分方程式

$$p_j h_j''(p_j) = c_j \quad (j = 1, \dots, n) \quad (255)$$

を解いて、対数型のポテンシャル関数

$$h_j(p_j) = c_j p_j \log p_j + b_j p_j + a_j \quad (j = 1, \dots, n) \quad (256)$$

$$f(p) = \sum_j h_j(p_j) = \sum_j c_j p_j \log p_j + \sum_j b_j p_j + a \quad (257)$$

を得る。 a_j, b_j, c_j および $a = \sum_j a_j$ は任意の定数である。ただし、 f が凸になるためには $c_j > 0$ でなくてはならない。

4 確率分布の情報幾何

4.1 関数空間上の情報幾何

いま、ある集合 X 上の実数値関数からなる適当な関数空間が \mathcal{P} であるとしよう。

$$\mathcal{P} = \{p \mid p : X \rightarrow \mathbb{R}\} \quad (258)$$

対応して双対空間 \mathcal{L} も、

$$\langle p, l \rangle = \int dx p(x) l(x) \quad (259)$$

と表されるような関数空間

$$\mathcal{L} = \{l | l: X \rightarrow \mathbb{R}\} \quad (260)$$

とみなす。さらに、点 $p \in \mathcal{P}$ における接ベクトル $\Delta p \in T_p \mathcal{P}$ も、 X 上の実数値関数と解釈する。これは形式的にはベクトル空間 \mathcal{P} と接空間 $T_p \mathcal{P}$ との自然な同一視によるものであるが、関数の摂動

$$p(x) + \epsilon \Delta p(x) \quad (|\epsilon| \ll 1) \quad (261)$$

を表していると考ええると理解しやすい。

関数空間の場合、 f, g は関数に対し数値を定める汎関数、 \mathcal{F}, \mathcal{G} は関数を別の関数へ変換する作用素となる。ここまでの議論のとおり、関数空間 \mathcal{P} 上に凸汎関数 f を一つ指定すれば、 $g, \mathcal{F}, \mathcal{G}$ が自動的に定まって情報幾何の枠組が誘導される。

4.2 確率分布のなす部分多様体

前項のような関数空間 \mathcal{P}, \mathcal{L} に対して部分多様体 $P_1 \subset \mathcal{P}$ をとり、 P_1 の元を何らかの方法で確率分布に対応づければ、ここまでの議論に確率としての意味を与えることができる。このとき、同一視 \mathcal{F}, \mathcal{G} に基いて定められるもう一方の部分多様体 $L_1 = \mathcal{F}P_1 \subset \mathcal{L}$ についても、 L_1 の各元がやはり確率分布に対応づけられる。

最も自明な対応づけは、積分が1であるような正值関数 $p \in \mathcal{P}$ をそのまま X 上の確率密度関数と解釈することである。ただし簡単のため境界は除いて

$$\left\{ p \in \mathcal{P} \mid \int dx p(x) = 1, p(x) > 0 \right\} \quad (262)$$

を P_1 と定義する。このとき、 Δp が P_1 の接ベクトルであることは

$$\int dx \Delta p(x) = 0 \quad (263)$$

と同値である。また、 $p \in P_1, l \in \mathcal{L}$ に対して

$$\langle p, l \rangle = \int dx p(x) l(x) = E_p[l(x)] \quad (264)$$

のように期待値としての意味づけがなされることを注意しておく^{*20}。

^{*20} 記号を節約するため、確率変数にもその実現値にも同じ記号 x を用いる。どちらを意味するかは文脈から明らかであろう。

自明でない対応づけの例としては、ある実数 β を決めておいて、

$$\int dx p(x)^\beta = 1 \quad (265)$$

となるような正値関数 p の集合を P_1 とするなどが挙げられる。この場合は

$$r(x) = p(x)^\beta \quad (266)$$

を p に対応する確率密度関数と解釈すればよい。

いずれにせよ、これまでに述べたようなダイバージェンス最小化問題を P_1 内でも考えることができる。すなわち、 P_1 内の n 次元部分多様体 \mathcal{M} と $\check{p} \in P_1$ が与えられたとき、ダイバージェンスで測って \check{p} に最も近い \mathcal{M} 上の点 \hat{q} を求める問題である^{*21}。

$$\hat{q} = \arg \min_{q \in \mathcal{M}} D[\check{p} \| q] \quad (267)$$

特に、 \check{p} として特定の $x_0 \in X$ が必ず出るという確率分布

$$\check{p}(x) = \delta(x - x_0) \quad (268)$$

を与えれば、この問題は観測されたデータ x_0 をモデル \mathcal{M} にあてはめるパラメトリック推定の一種とも解釈できる。 δ は Dirac のデルタ関数である。ただし、そのような \check{p} を \mathcal{P} 内にとることができるのかや $f(\check{p})$ を有限値として定められるのかといった数学的に厳密な議論には、ここでは踏み込まない。また、得られた \hat{q} が推定として良いものかどうかについても、幾何学とは別の議論が必要である。

4.3 U -ダイバージェンス

4.1 項でも述べたとおり、 \mathcal{P}, \mathcal{L} を関数空間としたときには、 f, g は関数に対し数値を定める汎関数、 \mathcal{F}, \mathcal{G} は関数を別の関数へ変換する作用素と解釈される。いま特に、実数から実数への関数 $U : \mathbb{R} \rightarrow \mathbb{R}$ を使って

$$g[l] = \int dx U(l(x)) \quad (269)$$

と表される場合を考えよう。その場合、 g が凸なことは U が凸なことと同値である [8]。この実数値凸関数 U を ELF(Envelope Function) と呼ぶことにする。

^{*21} 汎関数であることを見やすくするため、 $D(\cdot \| \cdot)$ のかわりに $D[\cdot \| \cdot]$ と書く。 f や g についても同様とする。

ELF を用いると、 $\mathcal{G}, \mathcal{F}, f$ は次のように表される。

$$(\mathcal{G}l)(x) = u(l(x)), \quad u(w) = \frac{dU(w)}{dw} \quad (270)$$

$$(\mathcal{F}p)(x) = \zeta(p(x)), \quad \zeta(v) = u^{-1}(v) \quad (271)$$

$$f[p] = \int dx L(p(x)), \quad L(v) = v\zeta(v) - U(\zeta(v)) \quad (272)$$

さらに、

$$\frac{dL(v)}{dv} = \zeta(v) \quad (273)$$

が成り立つ。 u を EPDF(Enveloped Probability Density Function) と呼ぶことにする。

$p, q \in \mathcal{P}$ に対するダイバージェンスは

$$D[p||q] = \int dx \left[L(p(x)) - L(q(x)) - \{p(x) - q(x)\}\zeta(q(x)) \right] \quad (274)$$

$$= \int dx \left[\{p(x)\zeta(p(x)) - U(\zeta(p(x)))\} - \{q(x)\zeta(q(x)) - U(\zeta(q(x)))\} \right. \\ \left. - \{p(x) - q(x)\}\zeta(q(x)) \right] \quad (275)$$

$$= \int dx \left[U(\zeta(q(x))) - U(\zeta(p(x))) + p(x)\zeta(p(x)) - p(x)\zeta(q(x)) \right] \quad (276)$$

$$= L[p||q] - L[p||p] \quad (277)$$

で与えられ、Bregman ダイバージェンスないし U -ダイバージェンスと呼ばれる [8]。ここに

$$L[p||q] = \int dx \{U(\zeta(q(x))) - p(x)\zeta(q(x))\} \quad (278)$$

と定義し、これを損失関数と呼ぶ。ポテンシャル f と損失関数 L との間には

$$f(p) = -L[p||p] \quad (279)$$

という関係がある。また、

$$S[p] = - \int dx p(x)\zeta(p(x)) \quad (280)$$

をエントロピーと呼ぶ。

さらに、接ベクトルの変換は

$$(\mathcal{G}_*^{(l)} \Delta l)(x) = u'(l(x))\Delta l(x) = U''(l(x))\Delta l(x) \quad (281)$$

$$(\mathcal{F}_*^{(p)} \Delta p)(x) = \zeta'(p(x))\Delta p(x) = L''(p(x))\Delta p(x) \quad (282)$$

となるので、 $l \in \mathcal{L}$, $p \in \mathcal{P}$ の接空間における内積はそれぞれ

$$(\Delta l, \Delta m)_l = \langle \mathcal{G}_*^{(l)} \Delta l, \Delta m \rangle \quad (283)$$

$$= \int dx u'(l(x)) \Delta l(x) \Delta m(x) = \int dx U''(l(x)) \Delta l(x) \Delta l(x) \quad (284)$$

$$(\Delta p, \Delta q)_p = \langle \Delta p, \mathcal{F}_*^{(p)} \Delta q \rangle \quad (285)$$

$$= \int dx \zeta'(p(x)) \Delta p(x) \Delta q(x) = \int dx L''(p(x)) \Delta p(x) \Delta q(x) \quad (286)$$

と表される。

4.4 α -ダイバージェンスと q -正規分布

実数 $\alpha \neq \pm 1$ に対して定義される α -ダイバージェンス

$$D^{(\alpha)}[r||s] = \frac{4}{1-\alpha^2} \int dx \left\{ \frac{1-\alpha}{2} r(x) + \frac{1+\alpha}{2} s(x) - r(x)^{\frac{1-\alpha}{2}} s(x)^{\frac{1+\alpha}{2}} \right\} \quad (287)$$

は、ある種の不変性を持つことなどから情報幾何において特別な地位を与えられている [1, 2, 3]。なお、 $\alpha = \pm 1$ に対する α -ダイバージェンスは後述の Kullback-Leibler ダイバージェンスで定義されるため、ここでは $\alpha = \pm 1$ は除外する。

確率密度関数 r を関数

$$p(x) = r(x)^{\frac{1-\alpha}{2}} \quad (288)$$

に対応づけて表現し、 U -ダイバージェンスにおいて

$$U(w) = \frac{1+\alpha}{2} w^{\frac{2}{1+\alpha}} \quad (289)$$

ととれば、 α -ダイバージェンスの定数倍が以下のように得られる^{*22}。実際このとき

$$u(w) = w^{\frac{1-\alpha}{1+\alpha}} \quad (290)$$

$$\zeta(v) = v^{\frac{1+\alpha}{1-\alpha}} \quad (291)$$

$$L(v) = \frac{1-\alpha}{2} v^{\frac{2}{1-\alpha}} \quad (292)$$

だから、 U -ダイバージェンスは

$$D[p||q] = \int dx \left\{ \frac{1-\alpha}{2} p(x)^{\frac{2}{1-\alpha}} + \frac{1+\alpha}{2} q(x)^{\frac{2}{1+\alpha}} - p(x)q(x)^{\frac{1+\alpha}{1-\alpha}} \right\} \quad (293)$$

^{*22} α -ダイバージェンスそのものが得られるよう係数を調節することもできるが、 u と ζ の対称性がかえって悪くなる。

となる。 p, q に対応する確率密度関数を r, s とすればこれは

$$\int dx \left\{ \frac{1-\alpha}{2} r(x) + \frac{1+\alpha}{2} s(x) - r(x)^{\frac{1-\alpha}{2}} s(x)^{\frac{1+\alpha}{2}} \right\} = \frac{1-\alpha^2}{4} D^{(\alpha)}[r||s] \quad (294)$$

に等しい^{*23}。

また、 $p \in \mathcal{P}$ における接空間の内積は、

$$(\Delta p, \Delta q)_p = \int dx \frac{1+\alpha}{1-\alpha} p(x)^{\frac{2\alpha}{1-\alpha}} \Delta p(x) \Delta q(x) \quad (295)$$

となり、確率密度関数 r とその摂動 $\Delta r, \Delta s$ で表せばこれは

$$\frac{1-\alpha^2}{4} \int dx \frac{\Delta r(x) \Delta s(x)}{r(x)} = \frac{1-\alpha^2}{4} \mathbb{E}_r \left[\frac{\Delta r(x)}{r(x)} \frac{\Delta s(x)}{r(x)} \right] = \frac{1-\alpha^2}{4} (\Delta r, \Delta s)_r^{\text{Fisher}} \quad (296)$$

に等しい。ここに

$$\Delta p(x) = \frac{1-\alpha}{2} r(x)^{-\frac{1+\alpha}{2}} \Delta r(x) \quad (297)$$

$$\Delta q(x) = \frac{1-\alpha}{2} r(x)^{-\frac{1+\alpha}{2}} \Delta s(x) \quad (298)$$

という関係を利用した。以上から、内積は Fisher 計量 [1, 2, 3] によるものと（定数倍を除いて）常に一致することがわかる。

いまの場合、確率密度関数に対応する部分空間は

$$P_1 = \left\{ p \in \mathcal{P} \mid \int dx p(x)^{\frac{2}{1-\alpha}} = 1, p(x) > 0 \right\} \quad (299)$$

$$L_1 = \left\{ l \in \mathcal{L} \mid \int dx l(x)^{\frac{2}{1+\alpha}} = 1, l(x) > 0 \right\} \quad (300)$$

^{*23} なお、Tsallis エントロピーの定義とあわせなければ、 U のとり方を次のように変えればよい [18].

$$\begin{aligned} U(w) &= \left(1 + \frac{1+\alpha}{2} w \right)^{\frac{2}{1+\alpha}} \\ u(w) &= \left(1 + \frac{1+\alpha}{2} w \right)^{\frac{1-\alpha}{1+\alpha}} \\ \zeta(v) &= \frac{2}{1+\alpha} \left(v^{\frac{1+\alpha}{1-\alpha}} - 1 \right) \\ L(v) &= \frac{2}{1+\alpha} \left(\frac{1-\alpha}{2} v^{\frac{2}{1-\alpha}} - v \right) \end{aligned}$$

であり、それぞれポテンシャル関数 f, g の等位面となっている。したがって 3.4 項の議論がすべて適用される。その際、 $p \in P_1$ に対応する確率密度関数 r を用いると、 $\langle p, l \rangle$ は

$$\langle p, l \rangle = \int dx p(x) l(x) = \int dx r(x)^{\frac{1-\alpha}{2}} l(x) \quad (301)$$

のように表されることを注意しておく。

さらに、実数値確率変数 ($X = \mathbb{R}$) のときには、

$$e_0(x) = 1, \quad e_1(x) = x, \quad e_2(x) = x^2 \quad (302)$$

に対して 3.3.7 節の結果を適用することにより q -正規分布 [13] が導出される。すなわち、3.3.7 節で求められた解集合

$$\text{span}\{1, x, x^2\} \cap L_1 \quad (303)$$

に対応する確率密度関数

$$r(x; \theta) = (\theta^1 x + \theta^2 x^2 - \psi(\theta))^{\frac{2}{1+\alpha}} \quad (304)$$

が q -正規分布である ($q = (1 - \alpha)/2$)。パラメータ $\theta = (\theta^1, \theta^2)$ に応じた関数 $\psi(\theta)$ の値は、規格化条件 $\int dx r(x; \theta) = 1$ から定められる。実際、 $\text{span}\{1, x, x^2\} \subset \mathcal{L}$ に属する l は

$$l(x) = \theta^0 + \theta^1 x + \theta^2 x^2 \quad (305)$$

の形に書けるので、対応する $p = \mathcal{G}l$ は

$$p(x) = u(l(x)) = (\theta^0 + \theta^1 x + \theta^2 x^2)^{\frac{1-\alpha}{1+\alpha}} \quad (306)$$

となり、これに相当する確率密度関数として

$$r(x) = p(x)^{\frac{2}{1-\alpha}} = (\theta^0 + \theta^1 x + \theta^2 x^2)^{\frac{2}{1+\alpha}} \quad (307)$$

が得られる。 $-\psi(\theta)$ は、 r が確率密度関数になるような θ^0 の値を、 θ_1, θ_2 の関数として表したものである。なお、 q -正規分布 $r(\cdot; \theta)$ に対応する $p \in \mathcal{P}$ の集合を \mathcal{M} と置けば、埋め込み $\mathcal{FM} \subset L_1$ は接続 Π_{L_1} について平坦になる。このことは 3.4.1 節ですでに述べた。

最後に、ポテンシャル関数

$$f[p] = \int dx L(p(x)) = \frac{1-\alpha}{2} \int dx p(x)^{\frac{2}{1-\alpha}} \quad (308)$$

が 3.4.2 項で述べたべき乗型の連続版になっていることも指摘しておく。

(補足) β -ダイバージェンスについて

[8] ではあくまで P_1 の元が直接そのまま確率密度関数だとして、べき乗型のダイバージェンスを

$$U(w) = \frac{1}{\beta+1}(\beta w + 1)^{\frac{\beta+1}{\beta}}, \quad u(w) = (\beta w + 1)^{\frac{1}{\beta}}, \quad \zeta(v) = \frac{v^\beta - 1}{\beta} \quad (309)$$

から導いている。これは β -ダイバージェンスと呼ばれ、 α -ダイバージェンスとは異なる。

このとき

$$L(v) = \frac{v^{\beta+1}}{\beta(\beta+1)} - \frac{v}{\beta} \quad (310)$$

なので、 $\int dx p(x) = \int dx q(x) = 1$ なら

$$D[p||q] = \frac{1}{\beta} \int dx \left\{ \frac{p^{\beta+1}}{\beta+1} + \frac{\beta q^{\beta+1}}{\beta+1} - p q^\beta \right\} \quad (311)$$

である。 p, q の意味づけの違いを無視すれば、(293) において $\beta = (1 + \alpha)/(1 - \alpha)$ と置くとこの D の定数倍になっている。その意義はまだ明らかでない。

4.5 η -ダイバージェンスと Kullback-Leibler ダイバージェンス

本項では、実数 η を一つ定めて

$$r(x) = p(x) + \eta \quad (312)$$

を確率密度関数に対応づけることにする。したがって部分空間 P_1 は

$$P_1 = \left\{ p \in \mathcal{P} \mid \int dx (p(x) + \eta) = 1, p(x) + \eta > 0 \right\} \quad (313)$$

となる。このとき、 U -ダイバージェンスにおいて

$$U(w) = (1 - \eta) \exp w \quad (314)$$

ととれば、 η -ダイバージェンス [9, 10, 11] が得られる。実際、

$$u(w) = (1 - \eta) \exp w \quad (315)$$

$$\zeta(v) = \log \frac{v}{1 - \eta} \quad (316)$$

$$L(v) = v \log \frac{v}{1 - \eta} - v \quad (317)$$

から $p, q \in P_1$ に対して

$$D[p||q] = \int dx \left\{ p(x) \log \frac{p(x)}{1-\eta} - p(x) \right\} + \int dx q(x) - \int dx p(x) \log \frac{q(x)}{1-\eta} \quad (318)$$

$$= \int dx p(x) \log \frac{p(x)}{q(x)} \quad (319)$$

であり、対応する確率密度関数 r, s で書けばこれは

$$\int dx (r(x) - \eta) \log \frac{r(x) - \eta}{s(x) - \eta} \quad (320)$$

に等しい。

P_1 は明らかに \mathcal{P} 内の平面をなす。しかも、 $p \in P_1$ に対して

$$g[\mathcal{F}p] = \int dx U(\zeta(p(x))) = \int dx p(x) = 1 - \eta \int dx \quad (321)$$

は定数となる。すなわち、 P_1 は擬似球面、 L_1 は擬似平面である。したがって、3.5 項の結果がすべて適用される。さらに、ポテンシャル関数 f が 3.5.2 項で述べた対数型の連続版になっていることも指摘しておく。

$$f[p] = \int dx L(p(x)) = \int dx \{ p(x) \log p(x) - (\log(1-\eta) + 1)p(x) \} \quad (322)$$

特に、 $\eta = 0$ ととれば、 $p, q \in P_1$ をそのまま確率密度関数と解釈できる。このとき D は Kullback-Leibler ダイバージェンス

$$D_{\text{KL}}[p||q] = \int dx p(x) \log \frac{p(x)}{q(x)} = D^{(-1)}[p||q] = D^{(1)}[q||p] \quad (323)$$

に一致する。Kullback-Leibler ダイバージェンスは確率分布間の相異を測る最も標準的な指標であり、様々な分野で広く利用されている。3.5 項の結果をこれに適用する際には、 $\langle \hat{q}, e_k \rangle$ が期待値 $E_{\hat{q}}[e_k(x)]$ に等しいことを念頭に置くと意味づけがしやすい。また、 $e_0 = 1$ という定数関数をとれば $P_1 = \{p \in \mathcal{P} \mid \langle p, e_0 \rangle = 1\}$ と表されるので、3.5.1 項の定数 C が 1 になることも注意しよう。さらに、 $p \in \mathcal{P}$ における接空間の内積は Fisher 計量によるものと一致する。

$$(\Delta p, \Delta q)_p = \int dx \frac{\Delta p(x) \Delta q(x)}{p(x)} = E_p \left[\frac{\Delta p(x)}{p(x)} \frac{\Delta q(x)}{p(x)} \right] = (\Delta p, \Delta q)_p^{\text{Fisher}} \quad (324)$$

なお、以上の導出は [8] とは見かけがやや異なる。本稿では、 L_1 がポテンシャル関数 g の等位面となるように対応づけを行った。一方 [8] では、 p, q 自体が確率密度関数である

とみなし、

$$U(w) = (1 - \eta) \exp w + \eta w, \quad u(w) = (1 - \eta) \exp w + \eta, \quad \zeta(v) = \log \frac{v - \eta}{1 - \eta} \quad (325)$$

および

$$L(v) = (v - \eta) \left(\log \frac{v - \eta}{1 - \eta} - 1 \right) \quad (326)$$

により η -ダイバージェンスを導出している。

5 統計との関係 まだ不整合

ここまでは単に、「ベクトル空間上に凸関数が一個与えられた」ら自動的に出てくる話。凸関数として特に $f(p) = \int (p \log p - p) dx$ (本質的にはエントロピー) をとった意義を、これから明らかにする。以下の4点が、情報幾何のキーポイントその3~6。

5.1 Kullback ダイバージェンスと大偏差原理

今、確率 q で表が出るコインを N 回投げたとき、表の割合が p になる確率を求めてみよう。 $N C_{Np}$ を Stirling の公式で評価すれば、

$$(\text{本当の確率が } q \text{ のとき、経験比率が } p \text{ になる確率}) \sim e^{-ND(p,q)} \quad (327)$$

がわかる^{*24}。

実は一般の分布でも、適当な意味で、「本当の分布が q のとき、経験分布が p になる確率」が、サンプル数 N に対し $e^{-ND(p,q)}$ で評価できる (大偏差原理)。この意味で、Kullback ダイバージェンス $D(p, q)$ は、分布 q と p の「見分けやすさ」を表している。たとえば、確率 0.5 と 0.6 は大差ないけど、確率 0 と 0.1 は無限に違う (\because 一つでも出たら、確率 0 ではないと断言できる)。そういうわけで、 D は検定論や情報理論 (typical sequence からはずれた符号列の出現確率の評価) で顔を出す。

5.2 Fisher 計量とクラメル-ラオの不等式

内積 G は、Fisher 計量と呼ばれる。この内積は、 D をローカルに 2 次形式で近似したものだから、確率分布の微小変化について「分布がどれくらい変わったか」を適切に評価

^{*24} 離散分布なので、これまで積分で書いてきた所は総和に読みかえてください。

した量になっている。このため、 G は推定論で顔を出す。具体的には、推定値 (正確には不偏推定量) の分散の下限 (どんなにうまい推定をしても、この限界より良くはできない) が G (の逆行列) を使って表される (クラメル-ラオの不等式)。「未知パラメータ θ をどれくらい正確に推定できるか」は、「 θ がちょっと変わったとき、データの分布がどれくらい変わるか」という感度で決まると考えれば、そうだろうなって気がする。

5.3 exponential family, mixture family

(これはポテンシャル関数の話ではないが…)

統計においては、

$$p(x; \theta) \propto \theta^1 p_1(x) + \cdots + \theta^n p_n(x) \quad (328)$$

$$p(x; \theta) \propto \exp\left(\theta^1 f_1(x) + \cdots + \theta^n f_n(x)\right) \quad (329)$$

のような形の集合 $\{p(\cdot; \theta) | \theta \in R^n\}$ を考えることが多い。前者を mixture family、後者を exponential family と呼ぶ。情報幾何で見ると、mixture family は m-flat、exponential family は e-flat である。

5.4 最尤推定と Kullback ダイバージェンス最小化

分布族 F 上でのパラメトリック推定問題とは、次のような問題である：

真の分布 q が、与えられた有限次元分布族 $F = \{q(\cdot; \theta) | \theta \in R^n\} \subset P$ に入っていることはわかっている。 q で生成されたデータ $\tilde{x}_1, \dots, \tilde{x}_N$ を見て、未知パラメータ θ を推定せよ。

その最尤推定は、

$$\hat{\theta} = \arg \max_{\theta} \sum_k \log q(\tilde{x}_k; \theta) \quad (330)$$

で与えられる。これは、情報幾何では、以下のように解釈される。データの経験分布を

$$p(x) = \frac{1}{N} \sum_k \delta(x - \tilde{x}_k) \quad (331)$$

という確率分布だと見るにより、観測されたデータを P 上の一点 p で表現する。実は、

$$\hat{\theta} = \arg \min_{\theta} D(p, q) \quad (332)$$

となっている ($\because D$ の定義をちょっと見ればわかる)。

特に、 F が exponential family の場合は、e-flat なので Pythagoras の定理から、「点 p を曲面 F へ m-射影^{*25} した点が最尤推定」となる。exponential family でなくても、局所的に exponential family で近似してやれば、「最尤推定は m-射影」と言える。

6 双対座標 まだ不整合

6.1 θ 座標と η 座標

ベクトル空間 \mathcal{L} に基底 $f_0(x) = 1, f_1(x), f_2(x), \dots$ をとり、確率分布 p を、

$$\log p(x) = \sum_{i=1}^{\infty} \theta^i(x) f_i - \psi(\theta) \quad (333)$$

と表示する。このパラメータ $\theta = (\theta^1, \theta^2, \dots)$ を、「 θ 座標」(あるいは、exponential family の自然パラメータ) と呼ぶ。関数 $\psi(\theta)$ は、 $\int p(x) dx = 1$ の条件から、

$$\psi(\theta) = \log \int e^{\sum_{i=1}^{\infty} \theta^i f_i(x)} dx \quad (334)$$

と決まる。

対応して、 \mathcal{P} の方に双対基底 $r^0(x) = p^0(x), r^1(x), r^2(x), \dots$ をとる (陽に求める必要はない) :

$$\int r^i(x) f_j(x) dx = \delta_j^i \quad (i, j = 0, 1, \dots) \quad (335)$$

特に、

$$\int p^0(x) dx = 1 \quad (336)$$

$$\int r^i(x) dx = 0 \quad (i = 1, 2, \dots) \quad (337)$$

$$\int p^0(x) f_i(x) = E_{p^0}[f_i(x)] = 0 \quad (i = 1, 2, \dots) \quad (338)$$

に注意。基底 $\{r^i\}$ を用いると、確率分布 p は、

$$p(x) = p^0(x) + \sum_{i=1}^{\infty} \eta_i r^i(x) \quad (339)$$

^{*25} m-測地線にそった直交射影

と表示することもできる。このパラメータ $\eta = (\eta_1, \eta_2, \dots)$ を「 η 座標」(あるいは、exponential family の期待値パラメータ) と呼ぶ。基底 $\{r^i\}$ を陽に求めておかなくても、次の定理を使えば、与えられた確率分布 p の η 座標を求められる。

Theorem 1

$$E_p[f_i(x)] = \eta_i \quad (340)$$

□

証明：左辺は $\int p(x)f_i(x)dx$ だから、基底の双対性より定理は明らか

■

6.2 θ 座標と η 座標の双対性

6.2.1 計量 G についての双対性

当然、接ベクトルについても、 θ 座標や η 座標で表示することができる。このとき、次の意味で、 θ 座標と η 座標は双対である。

Theorem 2 点 p における接ベクトル Δp_a の θ 座標を $(\Delta\theta^1, \Delta\theta^2, \dots)$ とし、接ベクトル Δp_b の η 座標を $(\Delta\eta_1, \Delta\eta_2, \dots)$ とする。このとき、

$$G(\Delta p_a, \Delta p_b) = \sum_{i=1}^{\infty} \Delta\theta^i \Delta\eta_i \quad (341)$$

□

証明：

$$G(\Delta p_a, \Delta p_b) = G(\Delta p_b, \Delta p_a) \quad (342)$$

$$= \langle \Delta\eta_1 r^1 + \Delta\eta_2 r^2 + \dots, \Delta\theta^1 f_1 + \Delta\theta^2 f_2 + \dots \rangle \quad (343)$$

■

6.2.2 Legendre 変換としての双対性

Theorem 3 確率分布 p の θ 座標がわかっているとき、 p の η 座標は

$$\eta_i = \frac{\partial \psi}{\partial \theta^i} \quad (344)$$

で求められる。すなわち、 θ 座標 $(\theta^1, \theta^2, \dots)$ から η 座標 (η_1, η_2, \dots) への変換は、 $\psi(\theta)$ をポテンシャル関数とする Legendre 変換である。

□

証明：式 (334) に書いたように

$$\psi(\theta) = \log \int e^{\sum_{i=1}^{\infty} \theta^i f_i(x)} dx$$

だから、

$$\frac{\partial \psi}{\partial \theta^k} = \frac{\int e^{\sum_{i=1}^{\infty} \theta^i f_i(x)} f_k(x) dx}{e^{\psi(\theta)}} = E[f_k(x)] = \eta_k \quad (345)$$

■

Theorem 4 確率分布 p の η 座標から θ 座標への変換もやはり Legendre 変換であり、ポテンシャル関数

$$\phi(\eta) = \int p(x) \log p(x) dx = -H(p) \quad (346)$$

により、

$$\theta^i = \frac{\partial \phi}{\partial \eta_i} \quad (347)$$

となっている。

□

証明：Legendre 変換の一般論より、順変換のポテンシャル ψ が与えられれば、逆変換のポテンシャル ϕ は

$$\phi(\eta) = \sum_{i=1}^{\infty} \theta^i \eta_i - \psi(\theta) \quad (348)$$

で求められる。計算すると、

$$\begin{aligned}\phi(\eta) &= \sum_{i=1}^{\infty} \theta^i \eta_i - \psi(\theta) \\ &= \sum_{i=1}^{\infty} \theta^i E[f_i(x)] - \psi(\theta)\end{aligned}\tag{349}$$

$$= E\left[\sum_{i=1}^{\infty} \theta^i f_i(x)\right] - \psi(\theta)\tag{350}$$

$$= \int p(x)(\log p(x) + \psi(\theta))dx - \psi(\theta)\tag{351}$$

■

6.3 座標で表現した情報幾何

6.3.1 θ 座標、 η 座標からはじめる情報幾何

「ベクトル空間に凸関数 (ポテンシャル関数) が 1 個与えられれば、Legendre 変換が定義されて、双対空間にもポテンシャル関数が決まって、…」となって、情報幾何が構築されたことを思い出してほしい。本稿では、「関数空間 $\{p\}$ とその双対空間で、ポテンシャル関数 f と g をとって…」という形で話をはじめた。しかし実は、「 θ 座標と η 座標で、ポテンシャル関数 ψ と ϕ をとって…」という所から話をはじめても、同じものが得られる。

Theorem 5 確率分布 p の η 座標を η^p 、確率分布 q の θ 座標を θ_q とする。このとき、

$$D(p, q) = \phi(\eta^p) + \psi(\theta_q) - \sum_{i=1}^{\infty} \eta_i^p \theta_q^i\tag{352}$$

□

証明： $\int (p \log p - p \log q)dx$ を計算すれば出るはず

■

混乱はしないと思うので、以下、同じ記号 D を使って

$$D(\eta, \theta) = \phi(\eta^p) + \psi(\theta_q) - \sum_{i=1}^{\infty} \eta_i^p \theta_q^i\tag{353}$$

と書くことにする。

Theorem 6 確率分布 p の θ 座標と η 座標の間には、

$$\eta = \arg \min_{\eta'} D(\eta', \theta) \quad (354)$$

$$\theta = \arg \min_{\theta'} D(\eta, \theta') \quad (355)$$

という関係が成り立つ。

証明：Legendre 変換に関する Young の不等式そのもの ■

6.3.2 フォーマルな情報幾何

基底を n 個で打ち切った集合 (exponential family)

$$E = \left\{ q(x) \mid q(x) = e^{\sum_{i=1}^n \theta^i f_i(x) - \psi(\theta)} \right\} \quad (356)$$

は、前述のように、 P の e-flat な部分多様体となる。確率分布 $p \in P$ に対し、 p を E へ m-射影した点 \hat{q} を考えよう。これまた前述のように、 E 上の点 q のうち $D(p, q)$ が最小となる点が \hat{q} である。 p の η 座標を (η_1, η_2, \dots) とすると、 \hat{q} の η 座標 $(\hat{\eta}_1, \hat{\eta}_2, \dots)$ は、

$$\hat{\eta}_i = \eta_i \quad (i = 1, \dots, n) \quad (357)$$

であること ($\because \eta$ 座標は m-flat。 θ 座標と η 座標は計量 G について双対) に注意しよう。

さて、exponential family E 上でのパラメトリック推定問題を再びとりあげる。前は、データ $\check{x}_1, \dots, \check{x}_N$ を経験分布 $p(x) = \frac{1}{N} \sum_k \delta(x - \check{x}_k)$ で表現して、「 p の E への m-射影 \hat{q} が最尤推定である」と言った。しかし外でしゃべるときには、こんなあやしい話 (デルタ関数まで含むような無限次元関数空間!) はできない。そこで、経験分布のかわりに、経験分布の η 座標 $\check{\eta}$ を使って、 $(\check{\eta}_1, \dots, \check{\eta}_n)$ でデータを表現する。これは、上述のように、m-射影 (=最尤推定) \hat{q} の η 座標の頭 n 個に一致する。

一般に、 E 上の点は、頭 n 個の η 座標 (η_1, \dots, η_n) を指定すれば決まってしまうことに注意しよう。最尤推定の η 座標 $(\check{\eta}_1, \dots, \check{\eta}_n)$ については、

$$\check{\eta}_i = E_p[f_i(x)] = \frac{1}{N} \sum_k f_i(\check{x}_k) \quad (358)$$

で具体的に計算できる。あとは、その θ 座標を求めてやれば、それが推定問題の答にな

る。 $\theta = \arg \min_{\theta'} D(\eta, \theta')$ という関係を思い出すと、最尤推定の θ 座標 $\hat{\theta}$ は、

$$\hat{\theta} = \arg \min_{\theta_1, \dots, \theta_n} D(\tilde{\eta}, \theta) = \arg \min_{\theta_1, \dots, \theta_n} \left(\phi(\tilde{\eta}) + \psi(\theta) - \sum_{i=1}^n \tilde{\eta}_i \theta^i \right) \quad (359)$$

となっていることがわかる。ただし、 $\phi(\tilde{\eta})$ は、 η 座標の頭 n 個が $(\tilde{\eta}_1, \dots, \tilde{\eta}_n)$ に一致しているような E 上の点の ϕ である。最尤推定は E に属しているので、 θ 座標は $n+1$ 番目から先は 0、よって和が有限になっているところにも注目。

さてさて今度は、exponential family E の部分多様体 E' 上での推定問題を考えよう。このような E' は、curved exponential family と呼ばれる。で、結論はやっぱり、

$$\hat{\theta}' = \arg \min_{\theta' \in E'} D(\tilde{\eta}, \theta') = \arg \min_{\theta' \in E'} \left(\phi(\tilde{\eta}) + \psi(\theta') - \sum_{i=1}^n \tilde{\eta}_i \theta'^i \right)$$

が最尤推定となる^{*26}。それはなぜかということ、経験分布 p を E へ m-射影した点を \hat{q} としたとき、 $q' \in E'$ に対して、

$$D(p, q') = D(q, q') + D(p, q) = D(q, q') + (\text{定数}) \quad (360)$$

(Pythagoras の定理) だから。

こんなふうには、 θ 座標・ η 座標を使って n 次元に「射影」して話をすれば、外でも発表できる理論になる。普通は情報幾何はこの形で記述される。

参考文献

- [1] S. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, Vol. 28, Springer-Verlag, New York, 1985.
- [2] 甘利俊一, 長岡浩司, 情報幾何の方法, 岩波講座応用数学 対象 12, 岩波書店, 1993.
- [3] S. Amari and H. Nagaoka, *Methods of Information Geometry*, Oxford Univ. Press, 2000.
- [4] S. L. Lauritzen, “Conjugate connection in statistical theory”, *Geometrization of Statistical Theory*, Proc. of the GST Workshop, C. T. J. Dodson ed., Univ. of Lancaster, pp. 33-51, 1987.
- [5] S. Eguchi, “A differential geometric approach to statistical inference on the basis of contrast functionals”, *Hiroshima math. J.*, vol. 15, pp. 341–391, 1985.

^{*26} $\theta' \in E'$ は、本当はもちろん $q(\cdot; \theta') \in E'$ と書かないといけないんだけど、まあ通じますよね。

- [6] S. Eguchi, “Geometry of minimum contrast”, Hiroshima math. J., vol. 22, pp. 631–647, 1992.
- [7] 江口真透, “情報幾何の新展開”, ISM Research Memo. 633, 1997.
http://www.geocities.jp/ismstats/past_report.html
<http://www.ism.ac.jp/eguchi/pdf/tenkai4.pdf>
- [8] N. Murata, T. Takenouchi, T. Kanamori, and S. Eguchi: “Information geometry of U-Boost and Bregman divergence”, Neural Computation, vol. 16, pp. 1437–1481, 2004.
- [9] 竹之内高志, “Eta-divergence を用いた AdaBoost のロバスト化”, 情報論的学習理論ワークショップ (IBIS02), pp.43–48, 2002.
- [10] M. Minami and S. Eguchi, “Robust blind source separation by beta-divergence”, Neural Computation, vol. 14, pp. 1859–1886, 2002.
- [11] T. Takenouchi, “Robust Boosting algorithm for multiclass classification by eta-divergence”, 2nd International Symposium on Information Geometry and its Applications, pp. 276–283, 2005.
- [12] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics”, J. Stat. Phys., Vol. 52, pp. 479–487, 1988.
- [13] 田中勝, “ q -正規分布に関する考察”, 信学論 (D-II), Vol. J85-D-II, No. 2, pp. 161–173, 2002.
- [14] 松島与三, 多様体入門, 裳華房, 1965.
- [15] 小林昭七, 接続の微分幾何とゲージ理論, 裳華房, 1989.
- [16] 野水克己, 佐々木武, アファイン微分幾何学—アファインはめ込みの幾何—, 裳華房, 1994.
- [17] 志磨裕彦, ヘッセ幾何学, 裳華房, 2001.
- [18] 田中勝, “ q -正規分布とエスコート分布—その情報幾何学的解釈” 第 54 回幾何学シンポジウム発表資料, 2007.

付録 A 付録

A.1 Legendre 変換 (Young の不等式) まだ

A.2 相似直交な部分多様体

本項は、擬似平面および擬似球面という概念を統一的に導入し特徴づける試みである。参考のため付録として掲載する。

A.2.1 ベクトル場の相似性

部分多様体が特定の条件を満たす場合には、いくつかの有用な性質が保証される。そのことを述べる準備として、ベクトル場の相似性という概念をまず導入する。

定数 $a \in \mathbb{R}$ と定ベクトル $e_0 \in \mathcal{L}$ が与えられたとき、 \mathcal{L} 上の次のような接ベクトル場 Δn を考えよう。

$$\Delta n(l) = al + e_0 \in T_l \mathcal{L}, \quad l \in \mathcal{L} \quad (361)$$

ただし $(a, e_0) = (0, o)$ の場合は除くことにする。また、ここですでに接空間 $T_l \mathcal{L}$ と空間自身 \mathcal{L} との同一視が用いられていることを注意しておく。

ベクトル場 Δn は、 \mathcal{L} 上の 1 助変数変換群 $\nu_t = \exp t \Delta n$ を生成する。直感的に言えば、 $\Delta n(l)$ を各位置 l における流速とみなし、時刻 0 に位置 $l_0 \in \mathcal{L}$ へ浮かべた粒子がそれによって流れると考えればよい。このとき時刻 $t \in \mathbb{R}$ における粒子の位置が $\nu_t(l_0) \in \mathcal{L}$ である。具体的には、各実数 $t \in \mathbb{R}$ に対して $\nu_t : \mathcal{L} \rightarrow \mathcal{L}$ は

$$\nu_t(l) = \begin{cases} l + te_0 & (a = 0) \\ (\exp at) \left(l + \frac{1}{a} e_0 \right) - \frac{1}{a} e_0 & (a \neq 0) \end{cases} \quad (362)$$

となる。前者はシフトに、後者は $-e_0/a$ を中心とした拡大縮小に対応する。これらが実際に $\exp t \Delta n$ であること、すなわち

$$\frac{d}{dt} \nu_t(l_0) = \Delta n(\nu_t(l_0)), \quad \nu_0(l_0) = l_0 \quad (363)$$

を満たすことは、直接微分してみれば確認できる。

ベクトル場 Δn から生成された 1 助変数変換群 ν_t は、さらに接空間どうしの線形変換 $\nu_{t*} : T_l \mathcal{L} \rightarrow T_{\nu_t(l)} \mathcal{L}$ を引き起こす。そして、得られた変換 ν_{t*} は

$$\nu_{t*}(\Delta l) \propto \Delta l \quad (364)$$

を満たす^{*27}。実際、 $|\epsilon| \ll 1$ に対して

$$\nu_t(l + \epsilon \Delta l) - \nu_t(l) \propto \epsilon \Delta l + o(\epsilon) \quad (365)$$

が容易に確かめられる^{*28}。つまり、点 l における任意の接ベクトル Δl をベクトル場 Δn にのせて流しても、大きさが変化するのみで方向は常に一定のままである。この性質を相似性と呼ぶことにする^{*29}。

相似性を持つベクトル場は実は (361) の形に限られることが、有限次元の場合には次のように示される。ベクトル空間としての \mathcal{L} の基底を用いて成分で書けば、ベクトル場 $\Delta n = (\Delta n^1, \dots, \Delta n^k)^T$ が相似性を満たすことは、任意の点 $l = (l^1, \dots, l^k)^T$ およびベクトル $\Delta l = (\Delta l^1, \dots, \Delta l^k)^T$ に対して

$$\sum \Delta l^i \frac{\partial}{\partial l^i} \Delta n(l) = a(l) \Delta l \quad (367)$$

が成り立つことと同値になる。ここに $a(\cdot)$ は \mathcal{L} 上の実数値関数である。第 i 成分のみが 1 で他の成分は 0 であるベクトル u_i を用いれば、上式の右辺は $\sum \Delta l^i a(l) u_i$ と書き換えられることに注意しよう。したがって、上式が任意の Δl で成り立つには、

$$\frac{\partial}{\partial l^i} \Delta n(l) = a(l) u_i \quad (i = 1, \dots, k) \quad (368)$$

すなわち

$$\frac{\partial \Delta n^j}{\partial l^i} = \begin{cases} a(l) & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (i, j = 1, \dots, k) \quad (369)$$

でなくてはならない。特に、 $i \neq j$ のとき $\partial \Delta n^j / \partial l^i = 0$ であることから、 Δn^j は l^j のみで定まることがわかる。すると $\partial \Delta n^i / \partial l^i = a(l)$ の左辺は l^i にしか依存しないので、右

^{*27} (364) の両辺は異なる点における接空間に属するので、一般の多様体ではこの式は意味をなさない。いまは \mathcal{L} がベクトル空間なので、接空間を空間 \mathcal{L} 自身と同一視することにより、すなわち平坦な接続 $\Pi_{\mathcal{L}}$ を用いることにより比較が可能となっている。

^{*28} (361) の場合は ν_t がベクトル空間 \mathcal{L} 上の線形変換になるので、実は $|\epsilon| \ll 1$ という制限や高次項 $+o(\epsilon)$ を外しても成り立つ： $\nu_t(l + \Delta l) - \nu_t(l) \propto \Delta l$

^{*29} 相似性は次のようにも言い換えられる：接続 $\Pi_{\mathcal{L}}$ に基づく共変微分 $\nabla(\Delta l)$ が o であるような \mathcal{L} 上のすべての接ベクトル場 Δl に対して、 Δn による Lie 微分が

$$[\Delta n, \Delta l] = h \Delta l \quad (366)$$

を満たす。ここに h は \mathcal{L} 上の実数値関数である。理由は、 \mathcal{L} が $\Pi_{\mathcal{L}}$ -平坦なことと Lie 微分の定義および意味から自明であろう。(ここで言う共変微分とは、線形写像 $\nabla : \Gamma(T\mathcal{L}) \rightarrow \Gamma(T^*\mathcal{L} \otimes T\mathcal{L})$ のうちある条件を満たすものと定義される [15]. Γ はベクトルバンドルの切断全体を表す.)

辺 $a(l)$ も l^i にしか依存しない。 $i = 1, \dots, k$ のすべてで同じことが言えるのだから、結局 $a(l)$ は l に依存しない定数である。以上により常微分方程式

$$\frac{d\Delta n^i(l^i)}{dl^i} = a \quad (i = 1, \dots, k) \quad (370)$$

が得られ、これを解くと (361) が求まる^{*30}。

A.2.2 相似直交な部分多様体

(361) のベクトル場を法線ベクトルとするような部分多様体を、相似直交な部分多様体と言うことにする。つまり、部分多様体 $L_1 \subset \mathcal{L}$ 上の各点 l において法空間

$$N_l L_1 = \{\Delta k \in T_l \mathcal{L} \mid (\Delta k, \Delta l)_l = 0 \text{ for all } \Delta l \in T_l L_1\} \subset T_l \mathcal{L} \quad (371)$$

がベクトル $\Delta n(l) \in T_l \mathcal{L}$ によって張られるとき、 L_1 は相似直交であると言う^{*31}。アファイン微分幾何 [16] の用語に従えば、内積 $(\cdot, \cdot)_l$ に基く法線ベクトルに関して、 $L_1 \subset \mathcal{L}$ がグラフはめ込み ($a = 0$) ないし中心アファイン超曲面 ($a \neq 0$) であることをこれは意味する。なお、 Δn が法線ベクトルなら当然 $-\Delta n$ も法線ベクトルになる。そこで、もし $a < 0$ の場合には $-\Delta n$ を改めて Δn と置き直し、以後は必ず $a \geq 0$ であるものとする。この約束は 2.7 項の結果を後ほど利用することを意図している。

L_1 が相似直交なときには、部分多様体 $\mathcal{M} \subset L_1$ の平坦性を次のように判定することができる：埋め込み $\mathcal{M} \subset L_1$ が接続 Π_{L_1} に関して平坦なことは、 \mathcal{M} を拡張した多様体

$$\tilde{\mathcal{M}} = \bigcup_{t \in \mathbb{R}} \nu_t(\mathcal{M}) = \{\nu_t(l) \mid l \in \mathcal{M}, t \in \mathbb{R}\} \subset \mathcal{L} \quad (372)$$

を考えたときに、埋め込み $\tilde{\mathcal{M}} \subset \mathcal{L}$ が接続 $\Pi_{\mathcal{L}}$ に関して平坦なことと同値になる。

実際、 $\tilde{\mathcal{M}} \subset \mathcal{L}$ が $\Pi_{\mathcal{L}}$ -平坦なら $\mathcal{M} \subset L_1$ が Π_{L_1} -平坦なことは、直交射影に基く接続 Π_{L_1} の作り方から当然である。すなわち、 \mathcal{M} の接ベクトルを \mathcal{M} 上の別の点へ Π_{L_1} -平行移動してもやはり \mathcal{M} の接ベクトルとなることが、定義からほぼ直接に示される。逆に

^{*30} 次の $\tilde{\mathcal{M}}$ の話をするためには、実はもっと弱く

$$\sum \Delta l^i \frac{\partial}{\partial l^i} \Delta n(l) = a(l) \Delta l + b(l, \Delta l) \Delta n(l)$$

でも十分。他の項が Δl について線形なことから、実は $b(l, \Delta l) = \sum c_i(l) \Delta l^i$ の形。だから微分方程式はこうなる。

$$\frac{\partial}{\partial l^i} \Delta n(l) = a(l) u_i + c_i(l) \Delta n(l)$$

^{*31} 無限次元のときを議論できていないので、「相似性を持つ Δn 」とは言わず、(361) を使って定義した。

$\mathcal{M} \subset L_1$ が Π_{L_1} -平坦なら $\tilde{\mathcal{M}} \subset \mathcal{L}$ が $\Pi_{\mathcal{L}}$ -平坦なことを示すには、ベクトル場 Δn の相似性を用いる。まず定義から、 $\tilde{\mathcal{M}}$ 上の各点 \tilde{l} は、 \mathcal{M} 上の点 l と実数 t を用いて $\tilde{l} = \nu_t(l)$ と書ける。そして相似性から、点 \tilde{l} における接ベクトル $\Delta \tilde{l} \in T_{\tilde{l}}\mathcal{L}$ が $\tilde{\mathcal{M}}$ に接することは、点 l において $\Delta \tilde{l}$ が $\tilde{\mathcal{M}}$ に接することと同値である。したがって、点 $l \in \mathcal{M}$ における $\tilde{\mathcal{M}}$ の接ベクトル $\Delta \tilde{l}$ をごく近くの点 $l' \in \mathcal{M} \rightarrow \Pi_{\mathcal{L}}$ -微小平行移動してもやはり $\tilde{\mathcal{M}}$ に接することさえ示せばよい。ここで、 $\Delta \tilde{l}$ は \mathcal{M} に平行な成分と垂直な成分との和で書けることに注意しよう。平行な成分は \mathcal{M} の接ベクトルだから、これを Π_{L_1} -微小平行移動しても \mathcal{M} に接するという前提であった。すると $\Pi_{\mathcal{L}}$ -微小平行移動しても $\tilde{\mathcal{M}}$ に接することが、 Π_{L_1} の作り方から明らかに言える。一方、垂直な成分は、ある実数 c を用いて $c\Delta n(l)$ と書ける。これを $\Pi_{\mathcal{L}}$ -微小平行移動したものと $c\Delta n(l')$ との差は、ふたたび相似性から $l - l'$ に比例する。すなわち、垂直な成分を微小平行移動した結果は、 \mathcal{M} の接ベクトルと $\Delta n(l')$ との線形結合で表される。したがってやはり $\tilde{\mathcal{M}}$ の接ベクトルとなっている。

相似直交な部分多様体 $L_1 \subset \mathcal{L}$ の持つもう一つの有用な特徴は、対応する部分多様体 $P_1 = \mathcal{G}L_1 \subset \mathcal{P}$ が関数

$$\tilde{f}(p) = af(p) + \langle p, e_0 \rangle \quad (373)$$

の等位面となることである。逆に、 \tilde{f} の等位面が必ず相似直交になることも言える。

これらを示すための補題をまず述べよう。 L_1 上の点 l における接ベクトル $\Delta l \in T_l L_1$ について、対応する接ベクトル $\Delta p = \mathcal{G}_* \Delta l \in T_{\mathcal{G}l} P_1$ による関数 \tilde{f} の方向微分は、

$$(\Delta p)[\tilde{f}] = (\Delta l, \Delta n(l))_l \quad (374)$$

と表される。実際、(83) や内積の定義を用いて

$$(\Delta p)[\tilde{f}] = a(\Delta p)[f] + \langle \Delta p, e_0 \rangle \quad (375)$$

$$= a(\Delta l, l)_l + (\Delta l, e_0)_l \quad (376)$$

$$= (\Delta l, \Delta n(l))_l \quad (377)$$

と変形することができる。

この補題により、もし L_1 が相似直交なら、 P_1 のすべての接ベクトルについて \tilde{f} の方向微分が 0 であることが直ちにわかる。逆に、方向微分が 0 になる接ベクトルはすべて、 \mathcal{L} の側で見ると Δn と直交していることもわかる。したがって、 L_1 が相似直交なことは P_1 が \tilde{f} の等位面になることと同値である。

A.2.3 擬似平面と擬似球面

相似直交な部分多様体のうち、 $a = 0$ であるものを擬似平面 (planaresque)、 $e_0 = o$ であるものを擬似球面 (spheresque) と呼ぶことにする。前にも述べたとおり、アファイン

微分幾何 [16] の用語で言えばそれぞれがグラフはめ込みおよび中心アファイン超曲面に相当する。しかし本稿では、両者の果す役割の対応を考慮し、二つの用語の釣合いがとれるように端的な呼称を導入した。擬似平面および擬似球面についてはこの後詳しく議論する。

残る $a > 0$ かつ $e_0 \neq o$ の場合は、いわば中心が原点でない擬似球面に相当する。その場合、2.7 項で述べた変換により空間 \mathcal{L} 上のシフトとスケーリングを誘導すれば、変換後の部分多様体

$$\tilde{L}_1 = \mathcal{A}L_1 \subset \mathcal{L} \quad (378)$$

は擬似球面になる。実際、変換後のポテンシャル関数

$$\tilde{f}(p) = af(p) + \langle p, e_0 \rangle \quad (379)$$

に関して、部分多様体 $\tilde{P}_1 = P_1 \subset \mathcal{P}$ は関数

$$\tilde{h}(\tilde{p}) = 1\tilde{f}(\tilde{p}) + \langle \tilde{p}, o \rangle = \tilde{f}(\tilde{p}) \quad (380)$$

の等位面である。すると、これに対応する部分多様体

$$\tilde{\mathcal{F}}\tilde{P}_1 = \mathcal{A}\mathcal{F}P_1 = \mathcal{A}L_1 = \tilde{L}_1 \subset \mathcal{L} \quad (381)$$

は擬似球面になることがわかる。

このように、残りの場合は適当な変換によって擬似球面に帰着されるので、以後は擬似平面と擬似球面だけに専念する。

A.3 P_1, L_1 が共に擬似平面の場合

3.4.2 項や 3.5.2 項にならって、 P_1, L_1 が共に擬似平面の場合についてもポテンシャル関数の特徴づけを与えておく。適当な制限のもとで、 P_1 と L_1 が共に擬似平面となるのは二次型の f に限られることが証明される。

3.4.2 項と同様に、 \mathcal{P} を n 次元ベクトル空間とし、ポテンシャル関数 f は (234) のように成分ごとの関数 h_1, \dots, h_n の和で書けるものを考えよう。この f から情報幾何を構築したときに、 P_1, L_1 が共に擬似平面となるための条件を本項で求める。

P_1, L_1 が共に擬似平面になることは、次の性質と等価である：定ベクトル $b = (b_1, \dots, b_n) \in \mathcal{P}$ と $a = (a^1, \dots, a^n)^T \in \mathcal{L}$ が存在して、

$$\langle p, a \rangle = \sum_i a^i p_i \quad (382)$$

が一定値となるような $p \in \mathcal{P}$ については

$$\langle b, \mathcal{F}p \rangle = \sum_i b_i \frac{\partial f(p)}{\partial p_i} \quad (383)$$

も一定値になる。この性質を要求して、 h_1, \dots, h_n の満たすべき条件を求める。

この要求は、 $\langle p, a \rangle$ を変えない方向の接ベクトル $\Delta p = (\Delta p_1, \dots, \Delta p_n)$ に対して

$$\langle b, \mathcal{F}(p + \epsilon \Delta p) \rangle = \langle b, \mathcal{F}p \rangle + o(\epsilon) \quad (|\epsilon| \ll 1) \quad (384)$$

が成り立つこと^{*32}、すなわち

$$\langle b, \mathcal{F}_*^{(p)} \Delta p \rangle = \sum_i \sum_j b_i \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \Delta p_j = \sum_j \left(\sum_i b_i \frac{\partial^2 f(p)}{\partial p_i \partial p_j} \right) \Delta p_j = 0 \quad (385)$$

を主張している。ここで、 $\langle p, a \rangle$ を変えない方向の Δp とは

$$\langle \Delta p, a \rangle = \sum_j a^j \Delta p_j = 0 \quad (386)$$

を満たす Δp のことだから、要求は

$$a^j \propto \sum_i b_i \frac{\partial^2 f(p)}{\partial p_i \partial p_j}, \quad (j = 1, \dots, n) \quad (387)$$

と等価になる（左辺と右辺の比が j によらない）。

f が (234) の形なら、この条件は

$$a^j \propto b_j h_j''(p_j) \quad (j = 1, \dots, n) \quad (388)$$

と単純化される。 a^j も b_j も定数なのだから、結局 $h_j''(p_j)$ も定数、つまり h_j は二次式であることが導かれた。

以上から、(234) の形のポテンシャル関数 f について P_1, L_1 が共に擬似平面となるのは、

$$f(p) = \sum_i (\alpha_i p_i^2 + \beta_i p_i + \gamma_i) \quad (389)$$

という二次型に限られることが示された。

^{*32} $\langle p, a \rangle$ が p について線形なので、実際には誤差項 $+o(\epsilon)$ や条件 $|\epsilon| \ll 1$ は不要である。

A.4 成分ごとの和に分解されないポテンシャル

A.4.1 P_1, L_1 が共に擬似球面の場合

3.4.2 項の補足として、(234) を仮定しない場合について述べる。この場合、(241) は単純なべき乗型でない解も持つことを示そう。

(241) の比例定数を c とおけば、

$$c \frac{\partial f(p)}{\partial p_j} = \sum_i p_i \frac{\partial}{\partial p_i} \left(\frac{\partial f(p)}{\partial p_j} \right) \quad (390)$$

より

$$\sum_i p_i \frac{\partial}{\partial p_i} \log \frac{\partial f(p)}{\partial p_j} = c \quad (391)$$

特に $j = 1$ の式に着目して $\log \frac{\partial f(p)}{\partial p_1} = \xi(p)$ とおくと、

$$p \cdot \nabla \xi(p) = c \quad (392)$$

特解はたとえば $\xi(p) = c \log p_1$ 。また、 $c = 0$ の場合の一般解は、 $p \perp \nabla \xi(p)$ より「 $\xi(p)$ は p の方向のみで定まる」。

以下、簡単のため $n = 2$ としよう。上の議論により、 R_1, R_2 を任意の関数として

$$\frac{\partial f(p)}{\partial p_1} = p_1^c R_1(p_2/p_1) \quad (393)$$

$$\frac{\partial f(p)}{\partial p_2} = p_2^c R_2(p_1/p_2) \quad (394)$$

が得られる。ただし、これらが両立するためには

$$\frac{\partial^2 f(p)}{\partial p_1 \partial p_2} = p_1^{c-1} R_1'(p_2/p_1) = p_2^{c-1} R_2'(p_1/p_2) \quad (395)$$

が要求される。

ポテンシャル関数 $f(p)$ を具体的に構成するには、 c, R_1 を任意に設定して $\frac{\partial f(p)}{\partial p_1}$ を定め、それを p_1 で積分すればよい。たとえば、 $\frac{\partial f(p)}{\partial p_1} = p_1 \exp(p_1/p_2)$ とおけば $f(p_1, p_2) = p_2(p_1 - p_2) \exp(p_1/p_2)$ という例が得られる。

凸関数になる例におきかえるほうがベター

A.4.2 P_1 が擬似球面で L_1 が擬似平面の場合

3.5.2 項の補足として、(234) を仮定しない場合について述べる。簡単のため $n = 2$ として解を構成する。

\mathcal{P} を適当に線形変換して座標をとり直せば、3.5.2 項で述べた「 \mathcal{P} における平面」の法線ベクトルを $\mathbf{c} = (c_1, c_2) = (1, 0)$ とおいて一般性を失わない。そのとき (253) から

$$\left(p_1 \frac{\partial}{\partial p_1} + p_2 \frac{\partial}{\partial p_2} \right) \frac{\partial f}{\partial p_2} = 0 \quad (396)$$

であり、前述のとおりこれは $\frac{\partial f}{\partial p_2}$ が p_2/p_1 の関数であることを意味する。したがって、

$$f(p) = p_1 R \left(\frac{p_2}{p_1} \right) + C(p_1) \quad (397)$$

という関数形を得る (R, C は任意の関数)。

実際このとき $g(\mathcal{F}p)$ は、

$$\begin{aligned} g(\mathcal{F}p) &= p_1 \frac{\partial f(p)}{\partial p_1} + p_2 \frac{\partial f(p)}{\partial p_2} - f(p) \end{aligned} \quad (398)$$

$$\begin{aligned} &= p_1 \left\{ R \left(\frac{p_2}{p_1} \right) - \frac{p_2}{p_1} R' \left(\frac{p_2}{p_1} \right) + C'(p_1) \right\} + p_2 R' \left(\frac{p_2}{p_1} \right) - \left\{ p_1 R \left(\frac{p_2}{p_1} \right) + C(p_1) \right\} \\ &= p_1 C'(p_1) - C(p_1) \end{aligned} \quad (399)$$

$$= p_1 C'(p_1) - C(p_1) \quad (400)$$

という p_1 のみの関数となる。よってその等位面は平面「 $p_1 = \text{一定}$ 」である。

A.5 課題

これから取り組むべき課題などを挙げる。

1. 擬似平面と異なり、擬似球面では拡張 Pythagoras の定理の測地線版が得られないことを、反例で明示する。
2. (211) の解は唯一か？
3. ポテンシャル関数の特徴づけについて
 - (a) 擬似平面と擬似球面を統一して扱えないか。
 - (b) (234) の制限を外した一般の f だとどうか？
 - (c) 無限次元の場合は？

4. ダイバージェンス D で遠近を測る理由は？
5. (301) をエスコート分布の根拠づけに結びつけられないか？
6. 4.4 項で言及した β -ダイバージェンスについて。(293) との関係、およびその意義は？ β -ダイバージェンス自体の動機や利点も要確認。
7. q -正規分布についてもっと丁寧な説明を。
8. η -ダイバージェンスの式や動機を文献で確認する必要あり。

更新履歴

- [2007-09-14 17:35] 脚注*23 を追加