情報幾何と統計力学

大上由人

2024 年 9 月 29 日

1 情報幾何

1.1 双対平坦な多様体

狭い意味での情報幾何学として、双対アフィン接続の幾何を考える。

Def: 双対アフィン接続

アフィン接続を持つ Riemann 多様体 (M, g) に対して、双対アフィン接続 ∇^* を、

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \tag{1.1}$$

により定める。

例えば、Riemann 接続の双対アフィン接続は、Riemann 接続が計量的であることから、

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \quad (X, Y, Z \in \mathfrak{X}(M)) \tag{1.2}$$

となり、$\nabla^* = \nabla$ である。(自己双対)
このとき、双対アフィン接続は、一意に定まることを示すことができる。また、共変微分の公理を満たすことも示される。

Def: 双対構造

Riemann 多様体 (M, g) に対して、計量 g に対する双対性を満たすアフィン接続のペア (∇, ∇^*) が与えられたとき、(g, ∇, ∇^*) を M の双対構造という。

上で定義した双対接続を用いて、双対平坦な多様体の幾何を考える。特に、統計多様体においては、片方の接続が指数型分布族、もう片方の接続が混合型分布族に対応する。

Def: 双対平坦な多様体

双対構造 (g, ∇, ∇^*) を持つ多様体 (M, g) が双対平坦であるとは、∇-曲率と ∇^*-曲率がどちらも 0 かつ、∇-捩率と ∇^*-捩率がどちらも 0 であることをいう。

今回は二つの接続について考えているため、それぞれの接続について局所アフィン座標系をとることができる。さらに、それぞれのアフィン座標系が、アフィン変換に対する任意性を持つことを用いると、以下の定理を示すことができる。

Thm: 局所アフィン座標系の存在

双対構造 (g, ∇, ∇^*) に関して平坦な多様体 M では、各点の周りで

$$g(\partial_i, \partial_j) = \delta_{ij} \tag{1.3}$$

を満たす、局所 ∇-アフィン座標系 (x^i) および局所 ∇^*-アフィン座標系 (y^i) の組 $\{x^i, y^i\}$ をとることができる。このような組 $\{x^i, y^i\}$ を双対アフィン座標系という。

以下では、双対 affine 座標系を用いた局所的な話に限定するため、$U \cap V$ 自身を多様体 M とみなし、直交性をみたす大域的な ∇-affine 座標系を (θ^i)、 ∇^*-affine 座標系を (η_j) で表し、それぞれ θ-座標系、 η-座標系とよぶことにする。また、対応するベクトル場を

$$\partial_i := \frac{\partial}{\partial \theta^i}, \quad \partial^j := \frac{\partial}{\partial \eta_j} \tag{4.40}$$

と書くことにする。また、直交性は

$$g(\partial_i, \partial^j) = \delta_i^j \tag{1.4}$$

と表すことにする。
ダイバージェンスを定義する。

Def: ダイバージェンス

M を、双対構造 (g, ∇, ∇^*) に関する双対平坦多様体であるとする。このとき、二点 $p, q \in M$ に対して、定まる量

$$D(p||q) = \psi(\theta(p)) + \varphi(\eta(q)) - \theta^i(p)\eta_i(q) \tag{1.5}$$

を ∇-ダイバージェンスという。ただし、$\{(\theta^i), (\eta_i)\}$ は M の大域的な双対アフィン座標系であり、ψ, ϕ は、$\eta_i = \partial_i \psi \quad \theta^i = \partial^i \varphi$ を満たす。

定義にアフィン座標系を用いているが、結局、座標の取り方に依らないことが示される。(ここでは省略)

∇ と ∇^* の役割を入れ替えると、$D(p||q)$ における θ と η、ψ と φ の役割も入れ替わる。したがって、

$$D(p||q) = D^*(q||p) \tag{1.6}$$

が成り立つ。また、少し計算すると、

$$D(p||q) \geq 0 \tag{1.7}$$

かつ、

$$D(p||q) = 0 \Leftrightarrow p = q \tag{1.8}$$

が成り立つことがわかる。したがって、$D(p||q)$ は、M 上の距離のような役割を持つと考えられるが、実際には、対称性 $D(p||q) = D(q||p)$ は成り立たないし、三角不等式も成り立たない。したがって、$D(p||q)$ は、距離の公理を満たさない。

ex:Euclid 空間

Euclid 空間においては、$\nabla = \nabla^*$ である。したがって、双対平坦性はただの平坦性と帰着する。このとき、ポテンシャルは、

$$\psi(z) = \varphi(z) = \frac{1}{2}\sum_{i=1}^{n}(z^i)^2 \tag{1.9}$$

となる。したがって、ダイバージェンスは、

$$D(p||q) = \frac{1}{2}\sum_{i=1}^{n}(z^i(p))^2 + \frac{1}{2}\sum_{i=1}^{n}(z^i(q))^2 - \sum_{i=1}^{n} z^i(p)z^i(q) = \frac{1}{2}\sum_{i=1}^{n}(z^i(p) - z^i(q))^2 \tag{1.10}$$

となる。これは、Euclid 空間における距離の二乗に一致する。
ところで、Euclid 空間において、距離の二乗と結びつく定理として、Pythagoras の定理がある。これは、双対平坦な多様体へ拡張することができる。

Thm: 拡張 Pythagoras の定理

双対平坦多様体 (M, g, ∇, ∇^*) において、点 p, q を結ぶ ∇-測地線と q, r を結ぶ ∇^*-測地線が計量 g に関して直交しているなら、

$$D(p||r) = D(p||q) + D(q||r) \tag{1.11}$$

が成り立つ。

1.2 統計多様体

情報幾何では、確率分布全体の集合を多様体として考える。有限集合 Ω の要素である根元事象を自然数でラベル付けすることにして、サイズ n の根元事象系を

$$\Omega_n = \{1, 2, \cdots, n\} \tag{1.12}$$

と表すことにする。このとき、Ω_n 上の確率分布全体の集合を、

$$\mathcal{S}_n = \{p : \Omega_n \to \mathbb{R}_{++};\ \sum_{\omega \in \Omega_n} p(\omega) = 1\} \tag{1.13}$$

ただし、

$$\mathbb{R}_{++} = \{x \in \mathbb{R}; x > 0\} \tag{1.14}$$

Thm:Chentsov の定理

Markov 埋め込みの下での不変性を満たす Riemann 計量は、定数倍を除いて、

$$g_p(X, Y) = \sum_{\omega=1}^{n} p(\omega)(X \log p(\omega))(Y \log p(\omega)) \tag{1.15}$$

に限られる。一方、アフィン接続の不変性を満たすようなアフィン接続は、

$$g(\nabla_X^\alpha Y, Z) = g(\tilde{\nabla}_X^\alpha Y, Z) - \frac{\alpha}{2} S_p(X, Y, Z) \tag{1.16}$$

により、実数 α と一対一に対応する。ただし、aaa

$$S_p(X, Y, Z) = \sum_{\omega=1}^{n} p(\omega)(X \log p(\omega))(Y \log p(\omega))(Z \log p(\omega)) \tag{1.17}$$

である。

Def:Fisher 計量

1.15 を満たす計量 g を Fisher 計量といい、1.16 を満たすアフィン接続 ∇^α を α-接続という。

さらに、これは多様体とみなすことができ、接続および双対接続を与えることができる。この接続をそれぞれ ∇^e、∇^m とする。また、これに対応して双対アフィン座標系を導入することができる。これを $\{(\theta^i), (\eta_i)\}$ とする。
さらに、∇^* に対応するダイバージェンスを考えることができる。

Def.KL ダイバージェンス

$p, q \in \mathcal{S}_n$ に対して、

$$D^m(p||q) = \sum_{\omega=1}^{n} p(\omega) \log \frac{p(\omega)}{q(\omega)} \tag{1.18}$$

を KL ダイバージェンスという。

また、$\mathcal{S}_n$ の部分多様体として、確率関数が指数関数の形をとるものを考える。

Def: 指数型分布族

Ω 上の関数 $C(\omega), F_1(\omega), \cdots, F_k(\omega)$ 及び $\mathbb{R}^k$ 上の領域 Θ 上を動く k 次元パラメータ $\theta = (\theta^1, \cdots, \theta^k) \in \Theta$ を用いて

$$p_\theta(\omega) = \exp\left[C(\omega) + \sum_{i=1}^{k} \theta^i F_i(\omega) - \psi(\theta)\right] \tag{1.19}$$

と表される確率分属族 $\{p_{\theta;\theta\in\Theta}\}$ を指数型分布族という。ただし、$\psi(\theta)$ は、

$$\psi(\theta) = \log\left[\sum_{\omega\in\Omega} \exp\left[C(\omega) + \sum_{i=1}^{k} \theta^i F_i(\omega)\right]\right] \tag{1.20}$$

で定義される。

この量が、確率分布間の距離 (のようなもの) として、対応してくれる。

上の定義で与えた指数型分布族を M とする。また、指数型分布族から一旦視点を $\mathcal{S}$ に戻してあげて、

統計多様体 $(\mathcal{S}, g, \nabla^e, \nabla^m)$ の ∇^e-自己平行部分多様体である指数型分布族

$$M = \{p_\theta(\omega) \in \mathcal{S}; \log p_\theta(\omega) = C(\omega) + \sum_{i=1}^{k} \theta^i F_i(\omega) - \psi(\theta)\} \tag{1.21}$$

が与えられたとする。M の期待値座標系を固定したときに定まる $\mathcal{S}$ 確率分布族

$$\Gamma_\eta = \{q(\omega) \in \mathcal{S}; E_q[F_i] = \eta_i\} \tag{1.22}$$

を考える。

Thm:

M と Γ_η が共有点を持つならば、その点において、M と Γ_η は直交する。

2 統計力学

2.1 エントロピー最大原理

統計力学における各分布の構成方法として、シャノンエントロピーを適当な拘束条件の元で最大化する方法がある。

エネルギー期待値を一定に保った時の、シャノンエントロピー最大化問題を考える。シャノンエントロピーは確率分布関数の汎関数であるから、変分によって停留点を探す。

$$\tilde{S} = -k_B \sum_i p_i \log p_i - \lambda\left(\sum_i p_i - 1\right) - \rho\left(\sum_i p_i E_i - U\right) \tag{2.1}$$

である。ここで、λ と ρ は未定乗数である。微小な確率変分を考えると、

$$\delta\tilde{S} = -k_B \sum_i (p_i + \delta p_i)\log(p_i + \delta p_i) - \lambda\left(\sum_i (p_i + \delta p_i) - 1\right) - \rho\left(\sum_i (p_i + \delta p_i)E_i - U\right) - \tilde{S} \tag{2.2}$$

$$= -k_B \sum_i \left(\delta p_i \log p_i + (p_i + \delta p_i)\log\left(1 + \frac{\delta p_i}{p_i}\right)\right) - \lambda\left(\sum_i \delta p_i\right) - \rho\left(\sum_i \delta p_i E_i\right) \tag{2.3}$$

となる。ここで、$\log(1+x) = x + O(x^2)$ であることを用いて、

$$\delta\tilde{S} = -k_B \sum_i (\delta p_i \log p_i + \delta p_i) - \lambda\left(\sum_i \delta p_i\right) - \rho\left(\sum_i \delta p_i E_i\right) \tag{2.4}$$

$$= \sum_i \delta p_i (-k_B \log p_i - k_B - \lambda - \rho E_i) + O(\delta p_i^2) \tag{2.5}$$

となる。したがって、

$$-k_B \log p_i - k_B - \lambda - \rho E_i = 0 \tag{2.6}$$

である。これを変形して、

$$p_i = \exp\left(-1 - \frac{\lambda}{k_B} - \frac{\rho E_i}{k_B}\right) \tag{2.7}$$

したがって、

$$p_i \propto \exp(-\beta E_i) \tag{2.8}$$

となる。ここで、$\beta = \frac{1}{k_B T}$ である。$(\rho = \frac{1}{T})$
あとは、規格化条件から、

$$p_i = \frac{1}{Z}\exp(-\beta E_i) \tag{2.9}$$

$$Z = \sum_i \exp(-\beta E_i) \tag{2.10}$$

であることがわかる。これがカノニカル分布である。

2.2 KL ダイバージェンス最小化

統計力学における最大エントロピー原理は、情報幾何の立場では、一様分布との KL ダイバージェンスを最小化することに対応する。

カノニカル分布の形を思い出しつつ、指数型分布族において、$k = 1$ とし、$F_1(\omega) = -H(\omega)$ とする。そして、$\theta = 0$ で一様分布

$$u = \left(\frac{1}{n}, \cdots, \frac{1}{n}\right) \tag{2.11}$$

を通る 1 次元指数型分布族 (∇^e-測地線) を考える。このとき、この測地線は、

$$p_\theta(\omega) = \exp\left[-\theta H(\omega) - \psi(\theta)\right] \tag{2.12}$$

となる。[*1] このとき、この測地線と、

$$\Gamma_\eta = \{q \in \mathcal{S}; E_q[-H] = \eta\} \tag{2.13}$$

は直交する。

一般化 Pythagoras の定理より、

$$\begin{aligned}
p_{\theta_*} &= \underset{q \in \Gamma_\eta}{\operatorname{argmin}}\{D^e(u||q)\} && (2.14)\\
&= \underset{q \in \Gamma_\eta}{\operatorname{argmin}}\{D^m(q||u)\} && (2.15)\\
&= \underset{q \in \Gamma_\eta}{\operatorname{argmin}} \sum_{\omega \in \Omega} q(\omega) \log \frac{q(\omega)}{u(\omega)} && (2.16)\\
&= \underset{q \in \Gamma_\eta}{\operatorname{argmin}}\{\log n - S(q)\} && (2.17)\\
&= \underset{q \in \Gamma_\eta}{\operatorname{argmax}}\{S(q)\} && (2.18)
\end{aligned}$$

となる。これは、確率変数 $F_1(\omega) = -H$ の期待値が一定であるもとで、Shanon エントロピー $S(q)$ を最大にする確率分布は、Γ_η から測地線に下した垂線の足 P_{θ_*} に一致することを示している。

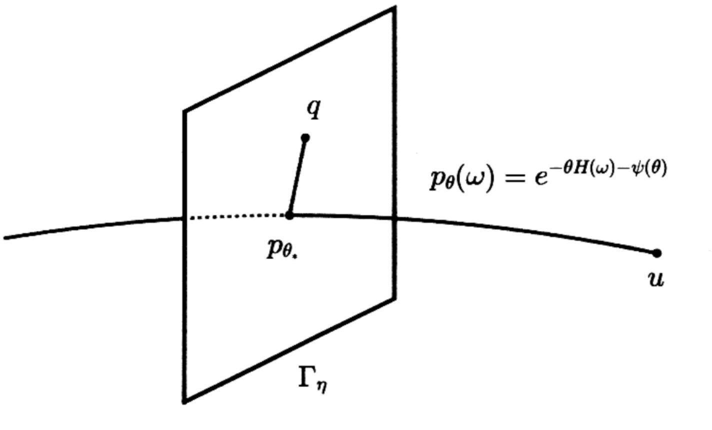

図 6.1　最大エントロピー原理

図 1　エネルギー期待値が一定の面の中で一様分布に最も近い分布が p_{θ_*} である。

[*1] $\theta = 0$ で一様分布になってほしいので、$C(\omega) = 0$ とする。