情報幾何の基礎

熊谷　亘*

東北大学理学研究科数学専攻博士課程前期 2 年

情報幾何という言葉になじみのない人も多いと思うので、まずその起源である統計学に少しふれることにする。確率的事象がおきているとし、その確率的事象はどのような確率分布 p にもとずいているのかを推定したいとする。例えば偏りのあるコインの表の出る確率 θ_0 を知りたい、というような状況である。ここではパラメタづけられた確率分布の族 $S = \{p_\theta\}_{\theta\in\Theta}$ の中に、現在起きている確率的事象の確率分布 $p = p_{\theta_0}$ が入っているとする (このとき θ_0 を真のパラメータという)。コインの例では $\Theta =$ 開区間 $(0, 1)$ である。推定とは数学的には,p_{θ_0} から発生したサンプル $(x_1, \cdots x_n)$ に対し、パラメータ空間 Θ の点 θ を対応させることで記述される。コインの場合は例えばコインをふってでた列 (表、裏、表、表、裏、$\cdots$ 表、表) に (表がでた回数)/(コインを振った回数)$\in \Theta$ を対応させる、等である。この対応を与える関数を推定量とよぶ。これをもう少し正確に述べると次のようになる。

定義 0.1.
(X, dx):測度空間、$S = \{p_\theta|(X, dx)$ 上の確率分布 $\}_{\theta\in\Theta}$ $(\Theta \subset R^n : open)$ に対し、可測写像 $f : X \to \Theta$ を推定量という。

この推定量をひとつ固定して、サンプルに対しひとつパラメータを返すことを推定という。以下、推定量のひとつの良い性質—不偏性—を仮定する。

定義 0.2.
推定量 f が不偏 $\Leftrightarrow$ $E_\theta[f] = \theta \quad (\forall\theta \in \Theta)$.
ただしここで $E_\theta[f] = \int_X f(x)p_\theta(x)dx$; p_θ に関する f の平均とした

これは、任意のパラメータ θ に対し平均的にはそのパラメータが出てくるというものである。この平均という量は大数の法則によればたくさんのサンプルについての相加平均として実現できる。つまり推定量が不偏性をもつとは p_{θ_0} からのサンプルがたくさんあたえられれば、真のパラメータ θ_0 を相加平均として出力してくれるということである。推定論において、サンプルがたくさん与えられればそれを元に真のパラメータを当ててほしいと思うのは当然なので、この仮定はそれなりに自然である (しかし少々強い仮定ではある)。

さて推定量はたくさん考えられるが、どの推定量を用いて推定するのがよいのであろうか？ひとつの基準として分散と呼ばれる、その推定量の平均からのずれを示す量を考えよう。推定量には不偏性を仮定しているので、平均は真のパラメータと一致している。よって分散は真のパラメータからのずれを表しているともいえる。推定するときは真のパラメータからのずれがなるべく少ないほうがいいと思うのは自然であろう。そこで推定量のよさの基準として、この分散が小さいものほどよいとしてみよう。すなわち点 θ において推定量 f より推定量 g の方がよいとは $V_\theta[f] \geq V_\theta[g]$ のことをいう。ここで S 上のリーマン計量として以下のようなものを導入する。この計量こそ統計学と幾何の最初の重要な接点である。

* sa9m14@math.tohoku.ac.jp

定義 0.3.

確率分布族 $S = \{p_\theta \| (X, dx) \text{ 上の確率分布 }\}_{\theta \in \Theta} (\Theta \subset R^n : open)$ に対し、点 θ において

$$J_{ij}(\theta) = \int_X \partial_i log p_\theta(x)\ \partial_j log p_\theta(x)\ p_\theta(x) dx$$

を (i, j) 成分とする半正定値行列が定まる。各点 θ でこれが正定値の時、この J を Fisher 計量という。

Fisher 計量と分散との関係として以下の定理が知られている。

定理 0.4. (Cramer-Rao 不等式)

不偏推定量 f に対し以下の不等式が成り立つ。

$$V_\theta[f] \geq J_\theta^{-1}.$$

すなわち Fisher 計量の逆が各点 θ における不偏推定量の良さの限界を表しているのである。全ての点で $V_\theta[f] = J_\theta^{-1}$ をみたす不偏推定量を有効推定量という。これが分散を良さの基準としたときのもっとも良い推定量である。よって推定は有効推定量によって行えばよいことになる (注意しなければならないことに有効推定量は常に存在するわけではない。またこれは分散を基準に良さを考えた場合であり、他にも良いといわれる推定量はある)。

さて次に情報幾何学で最も重要な多様体のクラスを導入する。

定義 0.5.

(1) 関数 $C : X \to R, F_i : X \to R (i = 1, \cdots, n), \psi : \Theta \to R$ が存在し、

$$p_\theta(x) = exp(C(x) + \sum_i \theta_i F_i(x) - \psi(\theta))$$

の形で表される確率分布族 $S = \{p_\theta\}$ を指数型分布族という。

(2) 指数型分布族において、(1) で表される座標 θ を自然座標という。また $\eta(\theta) = (\eta_1, \cdots, \eta(\theta)), \eta_i(\theta) := E_\theta[F_i]$ を期待値座標という。

(3) 指数型分布族において、(1) で表される ψ をポテンシャルという。$\phi(\theta) = \sum \theta_i \eta_i(\theta) - \psi(\theta)$ を ψ の双対ポテンシャルという。

統計学で現れるパラメータづけられた確率分布族の多くのものが指数型分布族であることが知られている。

例 0.6.

(1) 有限集合上の確率分布全体

(2) 正規分布族全体

さらに次のような接続を導入する。

定義 0.7.

$\alpha \in R$ に対し、確率分布族 $S = \{p_\theta\}$ の上の α 接続を以下で定める。

$$J(\nabla_X Y, Z) = E_\theta[(XY log p_\theta + \frac{1-\alpha}{2} X log p_\theta Y log p_\theta) Z log p_\theta]$$

特に (-1) 接続、(+1) 接続を m-接続、e-接続という。

計量がある場合、リーマン幾何では Levi-Chivita 接続を用いるのが通例であるが、情報幾何ではその代わりに二つの接続の組を考える。ここでは二つの接続の組—双対接続—の一般論を少し見ることにする。

定義 0.8.

M:多様体、g:リーマン計量、とする。二つの接続 ∇, ∇' に対し、

(1)∇ と ∇' が g に関して双対 $\Leftrightarrow$ $Z(g(X,Y)) = g(\nabla_Z X, Y) + g(X, \nabla' Y)$ $(X, Y, Z$: ベクトル場).

(2) (M, g, ∇, ∇^*) が双対平坦 $\Leftrightarrow$ $tor\nabla = tor\nabla^* = 0, cur\nabla = cur\nabla^* = 0.$

ここで $tor\nabla, cur\nabla$ はそれぞれ ∇ の捩率、曲率を表す。

命題 0.9.

M:多様体、g:リーマン計量、∇:接続、とする。

(1)∇ の g に関する双対接続 ∇^* は一意的に存在する。

(2)∇ が Levi-Chivita 接続 $\Leftrightarrow$ $\nabla^* = \nabla$.

一般の確率分布族の上の α 接続と $(-\alpha)$ 接続は Fisher 計量に関して双対的である。特に指数型分布族の上の e 接続と m 接続は Fisher 計量に関して双対平坦である。一般に双対平坦構造がある場合、ダイバージェンスと呼ばれる概念が定義できるがここでは指数型分布族に対してのみ定義する。

定義 0.10.

$S = \{p_\theta = exp(C(x) + \sum_i \theta_i F_i(x) - \psi(\theta))\}$ を指数型分布族とする。自然座標 θ、期待値座標 η、ポテンシャル ψ、双対ポテンシャル ϕ を用いて

$$D(p_\theta, p_{\theta'}) = \psi(\theta) + \phi(\theta') - \sum \theta_i \eta(\theta')_i$$

と定め、これをダイバージェンスという。

ダイバージェンスは自然座標の取り方によらず定まる。これは距離 2 乗のような量であり、二点間の近さを表すものと考えられる (ただし距離関数と違い一般には対称性はない)。特に次の定理が重要である。

定理 0.11. (ピタゴラスの定理)

$S = \{p_\theta = exp(C(x) + \sum_i \theta_i F_i(x) - \psi(\theta))\}$ を指数型分布族とする。点 p から点 q への e 測地線と点 q から点 r への m 測地線が点 q で Fisher 計量に関して直交するとき、以下が成り立つ。

$$D(p, r) = D(p, q) + D(q, r).$$

ダイバージェンス等の性質を用いることで推定の理論に幾何学的手法を適用することができ、幾何学の視点から推定を見ることができる (部分がある) がここではこれ以上深くは述べない。