振り返りと導入

前回は対数分配関数について調べた。本稿では次のことを行う:

- 分散の基本的な性質を調べる。
- Hessian を定義する。
- 対数分配関数から Fisher 計量を定める。

1 分散の性質

本節では $\mathcal{X}$ を可測空間、p を $\mathcal{X}$ 上の確率測度、V を m 次元 $\mathbb{R}$-ベクトル空間 ($m \in \mathbb{Z}_{\geq 0}$) とする。

以前 (0502_資料.pdf) の正規分布族の例では、十分統計量の分散が正定値対称であることをみた。ここでは分散の基本的な性質を調べる。

命題 1.1 (期待値・分散とペアリング). $f : \mathcal{X} \to V$ を可測写像とする。

(1) f が p に関する期待値を持つならば、任意の $\omega \in V^\vee$ に対し $E_p[\langle \omega, f(x) \rangle] = \langle \omega, E_p[f(x)] \rangle$ が成り立つ。

(2) f が p に関する分散を持つならば、任意の $\omega \in V^\vee$ に対し $\mathrm{Var}_p[\langle \omega, f(x) \rangle] = \langle \omega \otimes \omega, \mathrm{Var}_p[f(x)] \rangle$ が成り立つ。

証明 (1) V の基底をひとつ選んで固定し、この基底および双対基底に関する f, ω の成分をそれぞれ $f^i : \mathcal{X} \to \mathbb{R},\ \omega_i \in \mathbb{R}$ $(i = 1, \ldots, m)$ とおけば、

$$E[\langle \omega, f(x) \rangle] = E[\omega_i f^i(x)] = \omega_i E[f^i(x)] = \langle \omega, E[f(x)] \rangle \tag{1.1}$$

となる。

(2) 表記の簡略化のため $\alpha := E[f] \in V$ とおけば

$$\begin{aligned}
\mathrm{Var}[\langle \omega, f(x) \rangle] &= E[(\langle \omega, f(x) \rangle - \langle \omega, \alpha \rangle)^2] && (1.2)\\
&= E[\langle \omega, f(x) - \alpha \rangle^2] && (1.3)\\
&= E[\langle \omega \otimes \omega, (f(x) - \alpha)^2 \rangle] && (1.4)\\
&= \langle \omega \otimes \omega, E[(f(x) - \alpha)^2] \rangle && (1.5)\\
&= \langle \omega \otimes \omega, \mathrm{Var}[f(x)] \rangle && (1.6)
\end{aligned}$$

となる。 □

定理 1.2 (分散の半正定値対称性). $f : \mathcal{X} \to V$ を可測写像とし、f は p に関する分散を持つとする。このとき、$\mathrm{Var}_p[f] \in V \otimes V$ は対称かつ半正定値である。

証明 $\mathrm{Var}[f] = E[(f - E[f])^2]$ が対称であることは、写像 $(f - E[f])^2$ が $V \otimes V$ の対称テンソル全体からなるベクトル部分空間に値を持つことから従う。$\mathrm{Var}[f]$ が半正定値であることは、各 $\omega \in V^\vee$ に対し $\mathrm{Var}[f](\omega, \omega) = \langle \omega \otimes \omega, \mathrm{Var}[f] \rangle = \mathrm{Var}[\langle \omega, f(x) \rangle] \geq 0$ より従う。 □

分散が 0 であることの特徴づけを述べておく。

命題 1.3 (分散が 0 であるための必要十分条件). 可測写像 $f\colon \mathcal{X} \to V$ であって p に関する分散を持つものに関し、次は同値である:

(1) $\mathrm{Var}_p[f] = 0$

(2) f は p-a.e. 定数

証明には次の事実を用いる。

事実 1.4. $\mathcal{Y}$ を可測空間、μ を $\mathcal{Y}$ 上の測度とする。このとき、$g \in L^1(\mathcal{Y}, \mu)$ であって $g(y) \geq 0$ μ-a.e. をみたすものに関し、次は同値である:

(1) $\displaystyle\int_{\mathcal{Y}} g(y)\,\mu(dy) = 0$

(2) $g(y) = 0 \quad \mu$-a.e.

□

命題 1.3 の証明. V の基底 e_i $(i = 1, \ldots, m)$ をひとつ選んで固定し、$f, E[f]$ の成分表示をそれぞれ $f^i\colon \mathcal{X} \to \mathbb{R}$ および $a^i \in \mathbb{R}$ $(i = 1, \ldots, m)$ とおいておく。

(2) ⇒(1) f が a.e. 定数ならば、$f^i(x) = a^i$ a.e. $(i = 1, \ldots, m)$ したがって $(f^i(x)-a^i)(f^j(x)-a^j) = 0$ a.e. $(i, j = 1, \ldots, m)$ である。よって $\displaystyle\int_{\mathcal{X}} (f^i(x) - a^i)(f^j(x) - a^j)\, p(dx) = 0$ $(i, j = 1, \ldots, m)$ だから $\mathrm{Var}[f] = 0$ である。

(1) ⇒(2) $\mathrm{Var}[f] = 0$ とすると、すべての $i = 1, \ldots, m$ に対し $\displaystyle\int_{\mathcal{X}} (f^i(x) - a^i)^2\, p(dx) = 0$ が成り立つ。よって事実 1.4 より、すべての $i = 1, \ldots, m$ に対し $(f^i(x) - a^i)^2 = 0$ a.e. したがって $f^i(x) = a^i$ a.e. が成り立つ。よって f は a.e. 定数である。 □

2 Hessian

本節では W を m 次元 $\mathbb{R}$-ベクトル空間 $(m \in \mathbb{Z}_{\geq 0})$、$U \overset{\text{open}}{\subset} W$ を開部分集合とする。

本節では U 上の C^∞ 関数に対し Hessian を定義したい。そこで、まず U 上にアファイン接続を定義し、それを用いて Hessian を定義する。

A. U 上のアファイン接続

一般のアファイン接続の平坦性を定義しておく。

定義 2.1 (平坦アファイン接続). M を多様体、∇ を M 上のアファイン接続とする。

- M の開部分集合 $O \overset{\text{open}}{\subset} M$ 上の座標であって、それに関する ∇ の接続係数がすべて 0 となるものを、O 上の **∇-アファイン座標 (∇-affine coordinates)** という。
- 各 $p \in M$ に対し、p のまわりの ∇-アファイン座標が存在するとき、∇ は M 上**平坦 (flat)** であるという。

今考えている U 上には、次のような平坦アファイン接続が定まる。

命題-定義 2.2 (U 上の平坦アファイン接続). U 上のアファイン接続 $D\colon \Gamma(TU) \to \Gamma(T^\vee U \otimes TU)$ を、次の規則で well-defined に定めることができる:

- 各 $X \in \Gamma(TU)$ に対し、W の基底が定める U 上の座標 x^i $(i = 1, \dots, m)$ をひとつ選び、

$$DX \coloneqq dX^i \otimes \frac{\partial}{\partial x^i} \in \Gamma(T^\vee U \otimes TU) \tag{2.1}$$

  と定める。ただし、X の成分表示を $X = X^i \dfrac{\partial}{\partial x^i}$ とおいた。

さらに、この D は U 上のアファイン接続として平坦である。

証明 写像として well-defined であることを一旦認め、先に $\mathbb{R}$-線型性、Leibniz 則、平坦性を確かめる。D の $\mathbb{R}$-線型性と Leibniz 則は、外微分 d の $\mathbb{R}$-線型性と Leibniz 則から従う。平坦性は、式 (2.1) で用いた座標 x^i が D-アファイン座標となることから従う。最後に、D が写像として well-defined であることを示す。y^α $(\alpha = 1, \dots, m)$ を W の基底が定める U 上の座標とすると、

$$\begin{aligned} dX^i \otimes \frac{\partial}{\partial x^i} &= d\left(X^\alpha \frac{\partial x^i}{\partial y^\alpha}\right) \otimes \frac{\partial y^\alpha}{\partial x^i}\frac{\partial}{\partial y^\alpha} && (2.2)\\ &= \left(\frac{\partial x^i}{\partial y^\alpha} dX^\alpha + X^\alpha \underbrace{d\left(\frac{\partial x^i}{\partial y^\alpha}\right)}_{=0}\right) \otimes \frac{\partial y^\alpha}{\partial x^i}\frac{\partial}{\partial y^\alpha} && (2.3)\\ &= dX^\alpha \otimes \frac{\partial}{\partial y^\alpha} && (2.4) \end{aligned}$$

となる。ただし「$= 0$」の部分は x^i と y^α の間の座標変換がアファイン変換となることを用いた。これで well-defined 性も示された。 □

B. Hessian

U 上のアファイン接続 D により、$T^\vee U$ の接続が誘導される。これを用いて Hessian を定義する。

定義 2.3 (Hessian). C^∞ 関数 $f\colon U \to \mathbb{R}$ に対し、f の **Hessian** を

$$\operatorname{Hess} f \coloneqq Ddf \in \Gamma(T^\vee U \otimes T^\vee U) \tag{2.5}$$

と定義する。

D-アファイン座標を用いると、Hessian の成分表示は簡単な形になる。

命題 2.4 (Hessian の成分表示). x^i $(i = 1, \dots, m)$ を U 上の D-アファイン座標とする。このとき、座標 x^i に関する $\operatorname{Hess} f$ の成分表示は

$$\operatorname{Hess} f = \frac{\partial^2 f}{\partial x^i \partial x^j}\, dx^i \otimes dx^j \tag{2.6}$$

となる。とくに f の C^∞ 性より $\operatorname{Hess} f$ は対称テンソルである。

証明 $(\operatorname{Hess} f)(\partial_i, \partial_j) = \langle D_{\partial_i} df, \partial_j \rangle = \partial_i \langle df, \partial_j \rangle - \langle df, \underbrace{D_{\partial_i}\partial_j}_{=0} \rangle = \partial_i(\partial_j f) = \dfrac{\partial^2 f}{\partial x^i \partial x^j}$ より従う。 □

3 Fisher 計量

前回までと同様に、$\mathcal{X}$ を可測空間、$\mathcal{P} \subset \mathcal{P}(\mathcal{X})$ を $\mathcal{X}$ 上の指数型分布族、(V, T, μ) を $\mathcal{P}$ の次元 m の実現、$\Theta \subset V^\vee$ を自然パラメータ空間、$\psi\colon \Theta \to \mathbb{R}$ を対数分配関数、Θ° を $V^\vee$ における Θ の内部とし、$\lambda\colon \Theta \to \mathbb{R},\ \theta \mapsto \displaystyle\int_{\mathcal{X}} \exp\langle \theta, T(x) \rangle\, \mu(dx)$ とおく。

本節では、対数分配関数 ψ から Θ° 上に計量が定まることをみる。まず、`0502_資料.pdf` の命題 2.2 でも触れた次の条件に名前をつけておく。

定義 3.1 (条件 A). $\mathcal{P}$ の実現 (V, T, μ) に関する次の条件を、**条件 A** と呼ぶことにする。

(条件 A) $\langle \theta, T(x) \rangle$ が $\mathcal{X}$ 上 μ-a.e. 定数であるような $\theta \in V^\vee$ は $\theta = 0$ のみである。

Fisher 計量を定義する。

命題-定義 3.2 (Fisher 計量). ψ を Θ° 上の C^∞ 関数とみなすと、各 $\theta \in \Theta^\circ$ に対し $(\operatorname{Hess}\psi)_\theta \in T_\theta^{(0,2)}\Theta^\circ$ は $\operatorname{Var}_{P_\theta}[T]$ に一致する。さらに (V, T, μ) が条件 A をみたすならば、$\operatorname{Hess}\psi$ は正定値である。

したがって (V, T, μ) が条件 A をみたすとき、$\operatorname{Hess}\psi$ は Θ° 上の Riemann 計量となり、これを ψ の定める **Fisher 計量 (Fisher metric)** という。

証明 まず $(\operatorname{Hess}\psi)_\theta = \operatorname{Var}_{P_\theta}[T]$ $(\theta \in \Theta^\circ)$ を示す。Θ° 上の D-アファイン座標 θ^i $(i = 1, \ldots, m)$ をひとつ選ぶと、命題 2.4 より、座標 θ^i に関する $\operatorname{Hess}\psi$ の成分表示は $\operatorname{Hess}\psi = \dfrac{\partial^2 \psi}{\partial\theta^i \partial\theta^j} d\theta^i \otimes d\theta^j$ となる。ここで前回 (`0516_資料.pdf`) の系 2.4 より

$$
\begin{aligned}
\frac{\partial^2 \psi}{\partial\theta^i \partial\theta^j}(\theta) &= \partial_i \partial_j \log \lambda(\theta) && (3.1)\\
&= \partial_i \left(\frac{\partial_j \lambda(\theta)}{\lambda(\theta)} \right) && (3.2)\\
&= \frac{\partial_i \partial_j \lambda(\theta)}{\lambda(\theta)} - \frac{\partial_i \lambda(\theta) \partial_j \lambda(\theta)}{\lambda(\theta)^2} && (3.3)\\
&= E[T^i(x) T^j(x)] - E[T^i(x)] E[T^j(x)] && (3.4)\\
&= E[(T^i(x) - E[T^i(x)])(T^j(x) - E[T^j(x)])] && (3.5)
\end{aligned}
$$

を得る。ただし $E[\cdot]$ は P_θ に関する期待値 $E_{P_\theta}[\cdot]$ の略記である。したがって $\operatorname{Hess}_\theta \psi = \operatorname{Var}_{P_\theta}[T]$ が成り立つ。

次に、(V, T, μ) が条件 A をみたすとし、$\operatorname{Hess}\psi$ が正定値であることを示す。すなわち、各 $\theta \in \Theta^\circ$ に対し $(\operatorname{Hess}\psi)_\theta$ が正定値であることを示す。そのためには各 $u \in V^\vee$ に対し「$(\operatorname{Hess}\psi)_\theta(u, u) = 0$ ならば $u = 0$」を示せばよいが、上で示したことと命題 1.1 より

$$
(\operatorname{Hess}\psi)_\theta(u, u) = (\operatorname{Var}_{P_\theta}[T])(u, u) = \langle u \otimes u, \operatorname{Var}_{P_\theta}[T] \rangle = \operatorname{Var}_{P_\theta}[\langle u, T(x) \rangle] \tag{3.6}
$$

と式変形できるから、$(\operatorname{Hess}\psi)_\theta(u, u) = 0$ ならば命題 1.3 より $\langle u, T(x) \rangle$ は a.e. 定数であり、したがって条件 A より $u = 0$ となる。よって $(\operatorname{Hess}\psi)_\theta$ は正定値である。したがって $\operatorname{Hess}\psi$ は正定値である。 □

今後の予定

- Fisher 計量の具体例として正規分布族を考える。
- Amari-Chentsov テンソルを定義する。

参考文献

[Ama16] Shun-ichi Amari, **Information Geometry and Its Applications**, Applied Mathematical Sciences, vol. 194, Springer Japan, Tokyo, 2016 (en).