1 振り返りと導入

前回は期待値と分散を定義した。本稿では次のことを行う:

- 分散の基本的な性質を調べる。
- 対数分配関数 ψ の C^∞ 性と、微分と積分の順序交換ができることを示す。
- Hessian を定義し、ψ の Hessian の正定値性を示す。

2 分散の性質

前回の正規分布族の例でみた $V_P[T]$ は正定値対称であった。一般に、分散は次の性質を持つ。

定理 2.1 (分散の半正定値対称性)**.** $V_p[f] \in V \otimes V$ は、対称かつ半正定値である。

証明 まず $V_p[f]$ が対称であることを示す。[TODO] $V^\vee$ の基底をとるべき? そこで V の基底 e^i $(i = 1, \ldots, m)$ をひとつ選んで固定し、$f, E_p[f]$ の成分表示をそれぞれ f_i, a_i $(i = 1, \ldots, m)$ とおく。すると

$$V_p[f] = E_p[(f - E_p[f])^2] \tag{2.1}$$

$$= \left(\int_X (f_i(x) - a_i)(f_j(x) - a_j)\, p(dx) \right) e^i e^j \tag{2.2}$$

となり、最終行の成分は添字 i, j の置換に関し不変である。したがって $V_p[f]$ は対称である。

つぎに $V_p[f]$ が半正定値であることを示す。示したいことは、$V_p[f]$ を $V^\vee$ 上の $\mathbb{R}$-双線型形式とみなして、各 $\theta \in V^\vee$ に対し $V_p[f](\theta, \theta) \geq 0$ が成り立つことであるが、これは

$$V_p[f](\theta, \theta) = \sum_{i,j} \left(\int_X (f_i(x) - a_i)(f_j(x) - a_j)\, p(dx) \right) \theta(e^i)\theta(e^j) \tag{2.3}$$

$$= \int_X \left(\sum_{i,j} \theta(e^i)(f_i(x) - a_i)\theta(e^j)(f_j(x) - a_j) \right) p(dx) \tag{2.4}$$

$$= \int_X \left(\sum_i \theta(e^i)(f_i(x) - a_i) \right)^2 p(dx) \tag{2.5}$$

$$\geq 0 \tag{2.6}$$

より従う。したがって $V_p[f]$ は半正定値である。 □

分散が 0 であることの特徴づけを与えておく。

命題 2.2 (分散が 0 であるための必要十分条件)**.** 分散を持つ可測写像 $f: \mathcal{X} \to V$ に関し、次は同値である:

(1) $V_p[f] = 0$

(2) f は p-a.e. 定数

証明には次の事実を用いる。

事実 2.3. $\mathcal{Y}$ を可測空間、μ を $\mathcal{Y}$ 上の測度とする。このとき、μ-可積分関数 $g\colon \mathcal{Y} \to \mathbb{R}$ であって $g(y) \geq 0$ μ-a.e. $y \in \mathcal{Y}$ をみたすものに関し、次は同値である。

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

(2) $g(y) = 0$ μ-a.e. $y \in \mathcal{Y}$

□

命題 2.2 の証明. ここでは「p-a.e.」を「a.e.」と略記する。[TODO] $V^\vee$ の基底をとるべき? V の基底 e^i $(i = 1, \ldots, m)$ をひとつ選んで固定し、この基底に関する f の成分を $f_i\colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \ldots, m)$、$E_p[f]$ の成分を $a_i \in \mathbb{R}$ $(i = 1, \ldots, m)$ とおいておく。

(⇐) f が a.e. 定数ならば、$f_i(x) = a_i$ a.e.x $(i = 1, \ldots, m)$ したがって $(f_i(x) - a_i)(f_j(x) - a_j) = 0$ a.e.x $(i, j = 1, \ldots, m)$ 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 $V_p[f] = 0$ である。

(⇒) 対偶を示すため、f は a.e. 定数ではないと仮定する。すると、f_i が a.e. 定数ではないようなある $i \in \{1, \ldots, m\}$ が存在する。このとき $(f_i - a_i)^2 = 0$ a.e. ではないから、事実 2.3 より $\displaystyle\int_{\mathcal{X}} (f_i(x) - a_i)^2\,p(dx) > 0$ である。したがって $V_p[f] \neq 0$ である。 □

3 対数分配関数

本節では、対数分配関数が自然パラメータ空間の内部において C^∞ であって、積分記号下の微分が可能であることを示したい。

[TODO] Θ は内部を持つだろうか?

以降、本節では $\mathcal{X}$ 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、$\psi\colon \Theta \to \mathbb{R}$ を対数分配関数とする。$V^\vee$ における Θ の内部を Θ° と書くことにする。さらに関数 $h\colon \mathcal{X} \times \Theta \to \mathbb{R}$ および $\lambda\colon \Theta \to \mathbb{R}$ を

$$h(x, \theta) := e^{\langle \theta, T(x) \rangle} \qquad ((x, \theta) \in \mathcal{X} \times \Theta) \tag{3.1}$$

$$\lambda(\theta) := \int_{\mathcal{X}} h(x, \theta)\,\mu(dx) \quad (\theta \in \Theta) \tag{3.2}$$

と定める (つまり $\lambda(\theta) = e^{\psi(\theta)}$ である)。

[TODO] [Bro86]

補題 3.1 (優関数の存在). $\varphi = (\theta_1, \ldots, \theta_m)\colon \Theta^\circ \to \mathbb{R}^m$ を Θ° 上のチャートとする。このとき、任意の $\theta \in \Theta^\circ$ および $k \in \mathbb{Z}_{\geq 0}$, $i_1, \ldots, i_k \in \{1, \ldots, m\}$ に対し、Θ° における θ のある近傍 U が存在して、$\mathcal{X}$ 上の関数族

$$\left\{ \frac{\partial^k h}{\partial \theta_{i_1} \ldots \partial \theta_{i_k}}(x, \theta') \right\}_{\theta' \in U} \tag{3.3}$$

はある μ-可積分関数 $\mathcal{X} \to \mathbb{R}$ により支配される。

証明 $\theta \in \Theta^\circ$, $k \in \mathbb{Z}_{\geq 0}$, $i_1, \ldots, i_k \in \{1, \ldots, m\}$ とし、θ の座標を $t = (t_1, \ldots, t_m) := \varphi(\theta) \in \mathbb{R}^m$ とおく。Θ° における θ の開近傍 U' をひとつ選ぶと、$\varphi(U')$ は $\mathbb{R}^m$ における t の開近傍となる。そこで、$\varepsilon > 0$ を十分小さく選び、$\mathbb{R}^m$ 内の開立方体 $A_{4\varepsilon} := \displaystyle\prod_{i=1}^{m} (t_i - 2\varepsilon, t_i + 2\varepsilon)$ が $\varphi(U')$ に含まれるようにしておく。さらに $\mathbb{R}^m$ 内の開立

方体 $A_{2\varepsilon}$ を $A_{2\varepsilon} \coloneqq \prod_{i=1}^{m}(t_i - \varepsilon, t_i + \varepsilon)$ と定める。すると $U \coloneqq \varphi^{-1}(A_{2\varepsilon}) \subset U'$ は Θ° における θ の近傍となるが、これが求める U の条件を満たすことを示す。

まず h の偏導関数の座標表示を求める。そこで、自然な同一視により $\left.\frac{\partial}{\partial\theta_i}\right|_\theta \in T_\theta\Theta^\circ$ $(i = 1, \ldots, m)$ を $V^\vee$ の基底とみなしたものを $e^i \in V^\vee$ $(i = 1, \ldots, m)$ とおいておく。すると各 $\theta' \in U$ に対し、θ' の座標を $t' = (t'_1, \ldots, t'_m) \coloneqq \varphi(\theta') \in A_{2\varepsilon}$ として

$$h(x, t') = e^{\langle t', T(x)\rangle} = e^{\langle t'_i e^i, T(x)\rangle} = e^{t'_i T^i(x)} \tag{3.4}$$

が成り立つ (ただし $T^i : \mathcal{X} \to \mathbb{R},\ x \mapsto \langle e^i, T(x)\rangle$ $(i = 1, \ldots, m)$ とおいた)。したがって

$$\frac{\partial^k h}{\partial\theta_{i_1} \ldots \partial\theta_{i_k}}(x, \theta') = T^{i_1}(x) \cdots T^{i_k}(x) e^{t'_i T^i(x)} \tag{3.5}$$

と表せることがわかる。

次に、式 (3.5) の絶対値を上から評価する。

$$|T^{i_1}(x) \cdots T^{i_k}(x) e^{t'_i T^i(x)}| = \left(\frac{k+1}{\varepsilon}\right)^k \exp\left(t'_i T^i(x)\right) \prod_{\alpha=1}^{k} \frac{\varepsilon}{k+1} |T^{i_\alpha}(x)| \tag{3.6}$$

であり、末尾の部分は

$$\begin{aligned} \prod_{\alpha=1}^{k} \frac{\varepsilon}{k+1} |T^{i_\alpha}(x)| &\le \prod_{\alpha=1}^{k} \left(\exp\left(\frac{\varepsilon}{k+1} T^{i_\alpha}(x)\right) + \exp\left(-\frac{\varepsilon}{k+1} T^{i_\alpha}(x)\right)\right) && (3.7) \\ &= \sum_{(\sigma_1, \ldots, \sigma_k) \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^{k} \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x)\right) && (3.8) \end{aligned}$$

となるから、

$$\begin{aligned} (3.6) &\le \left(\frac{k+1}{\varepsilon}\right)^k \exp\left(t'_i T^i(x)\right) \sum_{(\sigma_1, \ldots, \sigma_k) \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^{k} \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x)\right) && (3.9) \\ &= \left(\frac{k+1}{\varepsilon}\right)^k \sum_{(\sigma_1, \ldots, \sigma_k) \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^{k} \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x) + \sum_{i=1}^{m} t'_i T^i(x)\right) && (3.10) \\ &= \left(\frac{k+1}{\varepsilon}\right)^k \sum_{\sigma = (\sigma_1, \ldots, \sigma_k) \in \{\pm 1\}^k} \exp\left(\sum_{i=1}^{m} t''_{\sigma,i} T^i(x)\right) && (3.11) \end{aligned}$$

となる。ただし最終行の $t''_{\sigma,i}$ は各 $i = 1, \ldots, m$ に対し

$$t''_{\sigma,i} \coloneqq t_i + \sum_{\substack{\alpha \in \{1, \ldots, k\} \\ i_\alpha = i}} \frac{\varepsilon}{k+1} \sigma_\alpha \in \mathbb{R} \tag{3.12}$$

とおいた。ここで、$t''_\sigma = (t''_{\sigma,1}, \ldots, t''_{\sigma,m}) \in \mathbb{R}^m$ は $A_{4\varepsilon}$ に属している。なぜならば、各 $i = 1, \ldots, m$ に対し

$$\begin{aligned} |t''_{\sigma,i} - t_i| &\le |t'_i - t_i| + \sum_{\substack{\alpha \in \{1, \ldots, k\} \\ i_\alpha = i}} \frac{\varepsilon}{k+1} && (3.13) \\ &< \varepsilon + \varepsilon && (3.14) \\ &= 2\varepsilon && (3.15) \end{aligned}$$

が成り立つからである。したがって (3.11) は $\mathcal{X}$ 上の μ-可積分関数の $\mathbb{R}$-線型結合だから μ-可積分であり、また (t' によらないから) θ' によらない。したがってこれが求める優関数である。 □

定理 3.2 (積分記号下の微分). λ に関し次が成り立つ。

(1) λ は Θ° 上 C^∞ 級である。

(2) 各 $\theta \in \Theta^\circ$ と θ の近傍上で定義された任意のベクトル場 $X^{(1)}, \dots, X^{(n)}$ $(n \in \mathbb{Z}_{\geq 1})$ に対し

$$(X^{(1)} \cdots X^{(n)} \lambda)(\theta) = \int_X (X^{(1)} \cdots X^{(n)} h)(x, \theta)\, \mu(dx) \tag{3.16}$$

が成り立つ。

証明 [TODO] □

4 Hessian

本節では、$\mathbb{R}^n$ の開部分集合上の関数に対する Hessian の概念を、一般の有限次元 $\mathbb{R}$-ベクトル空間 W の開部分集合 U 上の関数にまで拡張したい。そこで、まず U 上に平坦アファイン接続を導入しておく。

以降、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。

定義 4.1 ($W^\vee$ の基底が定める U 上の座標). $W^\vee$ の任意の基底 $(f^i)_{i=1}^m$, $f^i \colon W \to \mathbb{R}$ に対し、その U 上への制限 $(x^i)_{i=1}^m$, $x^i \colon U \to \mathbb{R}$ は U 上の座標となる。この $(x^i)_i$ を、本稿だけの呼び方として、**$W^\vee$ の基底 $(f^i)_i$ が定める U 上の座標**と呼ぶことにする。

命題-定義 4.2 (W の開部分多様体としての 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、$W^\vee$ の基底が定める U 上の座標 $(x^i)_{i=1}^m$ をひとつ選び、この座標に関する X の成分を $X^i \in C^\infty(U)$ とおいて

$$DX := dX^i \otimes \frac{\partial}{\partial x^i} \in \Gamma(T^\vee U \otimes TU) \tag{4.1}$$

と定める。

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。

以上により定まる D を、本稿だけの呼び方として、**W の開部分多様体としての U 上の平坦アファイン接続**と呼ぶことにする。

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、すなわち D の接続係数がすべて 0 になるような座標の存在は、(4.1) より明らかである。最後に、D が写像として well-defined であることを示す。[TODO] well-defined 性 □

[TODO] 1-form の共変微分をどう定義するかの remark?

定義 4.3 (Hessian). [TODO] Fréchet 微分で定義すべき?しかしノルムは入ってない。C^∞ 関数 $f \colon U \to \mathbb{R}$ に対し、f の **Hessian** $\operatorname{Hess} f \in \Gamma(T^\vee U \otimes T^\vee U)$ を

$$\operatorname{Hess} f := Ddf \tag{4.2}$$

と定義する。

命題 4.4. Hessian は対称テンソルである。

証明 $W^\vee$ の基底が定める U 上の座標 $(x^i)_{i=1}^m$ をひとつ選び、この座標に関する $\mathrm{Hess}\, f$ の成分表示を具体的に計算する。そこで $\dfrac{\partial}{\partial x^i}$ を ∂_i と略記することにすれば

$$
\begin{aligned}
(\mathrm{Hess}\, f)(\partial_i, \partial_j) &= (D_{\partial_i} df)(\partial_j) && (4.3)\\
&= \partial_i(df(\partial_j)) - df(\underbrace{D_{\partial_i}\partial_j}_{=0}) && (4.4)\\
&= \partial_i(\partial_j f) && (4.5)\\
&= \frac{\partial^2 f}{\partial x^i \partial x^j} && (4.6)
\end{aligned}
$$

となり、いま f が C^∞ であることよりこれは添字 i, j の置換に関し対称である。したがって $\mathrm{Hess}\, f$ は対称テンソルである。 □

命題 4.5. U がさらに W の凸集合であるとする。このとき、$\mathrm{Hess}\, f$ が正定値ならば f は U 上で狭義凸である。逆は成立しない。

証明 [TODO] □

定理 4.6 (対数分配関数の Hessian の正定値性)**.** すべての $\theta \in \mathrm{Int}\,\Theta$ に対し、$\mathrm{Hess}_\theta\, \psi = \left(\dfrac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}(\theta)\right)_{i,j}$ は非負定値かつ対称である。さらに (T, μ) が最小表現ならば、$\mathrm{Hess}_\theta\, \psi$ は正定値である。

証明 $\psi(\theta) = \log \displaystyle\int_X e^{\langle \theta, T(x) \rangle} \mu(dx)$ だから、直接計算により

$$
\begin{aligned}
\frac{\partial \psi}{\partial \theta_i}(\theta) &= \int_X T_i(x) e^{\langle \theta, T(x)\rangle - \psi(\theta)}\, \mu(dx), && (4.7)\\
\frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}(\theta) &= \int_X T_i(x) T_j(x) e^{\langle \theta, T(x)\rangle - \psi(\theta)}\, \mu(dx)\\
&\quad - \left(\int_X T_i(x) e^{\langle \theta, T(x)\rangle - \psi(\theta)}\, \mu(dx)\right)\left(\int_X T_j(x) e^{\langle \theta, T(x)\rangle - \psi(\theta)}\, \mu(dx)\right) && (4.8)\\
&= E_\theta[T_i(x) T_j(x)] - E_\theta[T_i(x)] E_\theta[T_j(x)] && (4.9)\\
&= E_\theta[T_i(x) T_j(x)] - \alpha_i \alpha_j && (4.10)\\
&= E_\theta[(T_i(x) - \alpha_i)(T_j(x) - \alpha_j)] && (4.11)
\end{aligned}
$$

を得る。ただし $\alpha_k := E_\theta[T_k(x)]$ $(k = 1, \ldots, m)$ とおいた。よって $\mathrm{Hess}_\theta\, \psi = \left(\dfrac{\partial^2 \psi}{\partial \theta_i \partial \theta_j}(\theta)\right)_{i,j}$ は θ に関する T の分散 $V_\theta[T(x)]$ に一致する。したがって**??** より $\mathrm{Hess}_\theta\, \psi$ は非負定値かつ対称である。[TODO] □

系 4.7 (対数分配関数の凸性)**.** [TODO] 正定値性より従う。あるいは Hölder

証明 [TODO] □

5 今後の予定

- KL ダイバージェンス
- Fisher 計量
- アファイン接続

6 参考文献

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

[Bro86] L. D. Brown, **Fundamentals of statistical exponential families: with applications in statistical decision theory**, Institute of Mathematical Statistics, 1986.

[Yos] Taro Yoshino, **bn1970.pdf**, Dropbox.