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:

- 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。
- 分散の基本的な性質を調べる。

前回に引き続き、可測空間 $\mathcal{X}$ 上の確率測度全体の集合を $\mathcal{P}(\mathcal{X})$ と書くことにする。また、Einstein の記法を用いる。

2 対数分配関数

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

$$h(x,\theta) := e^{\langle \theta, T(x) \rangle} \qquad ((x,\theta) \in \mathcal{X} \times \Theta) \tag{2.1}$$

$$\lambda(\theta) := \int_{\mathcal{X}} h(x,\theta)\,\mu(dx) \quad (\theta \in \Theta) \tag{2.2}$$

と定める (つまり $\psi(\theta) = \log \lambda(\theta)$ である)。

本節の目標は次の定理を示すことである。

定理 2.1 (λ と ψ の C^∞ 性と積分記号下の微分)**.** $\varphi = (\theta_1, \dots, \theta_m)\colon \Theta^\circ \to \mathbb{R}^m$ を Θ° 上のチャートとする。このとき、任意の $k \in \mathbb{Z}_{\geq 1}$, $i_1, \dots, i_k \in \{1, \dots, m\}$ に対し、

$$\partial_{i_k} \cdots \partial_{i_1} \lambda(\theta) = \int_{\mathcal{X}} \partial_{i_k} \cdots \partial_{i_1} h(x,\theta)\,\mu(dx) \quad (\theta \in \Theta^\circ) \tag{2.3}$$

が成り立つ (∂_i は $\dfrac{\partial}{\partial \theta_i} \in \Gamma(T\Theta^\circ)$ の略記)。ただし、左辺の微分可能性および右辺の可積分性も定理の主張に含まれる。とくに λ および ψ は Θ° 上の C^∞ 関数である。

定理 2.1 の証明には次の事実を用いる。

事実 2.2 (積分記号下の微分)**.** $\mathcal{Y}$ を可測空間、ν を $\mathcal{Y}$ 上の測度、$I \subset \mathbb{R}$ を開区間、$f\colon \mathcal{Y} \times I \to \mathbb{R}$ を

(i) 各 $t \in I$ に対し $f(\cdot, t)\colon \mathcal{Y} \to \mathbb{R}$ が可測
(ii) 各 $y \in \mathcal{Y}$ に対し $f(y, \cdot)\colon I \to \mathbb{R}$ が微分可能

をみたす関数とする。このとき、f に関する条件

(1) 各 $t \in I$ に対し $f(\cdot, t) \in L^1(\mathcal{Y}, \nu)$ である。
(2) ある ν-可積分関数 $\Phi\colon \mathcal{Y} \to \mathbb{R}$ が存在し、すべての $t' \in I$ に対し $\left|\dfrac{\partial f}{\partial t}(y, t')\right| \leq \Phi(y)$ a.e.y である。

が成り立つならば、関数 $I \to \mathbb{R},\ t \mapsto \displaystyle\int_{\mathcal{Y}} f(y,t)\,\nu(dy)$ は微分可能で、

$$\frac{\partial}{\partial t} \int_{\mathcal{Y}} f(y,t)\,\nu(dy) = \int_{\mathcal{Y}} \frac{\partial f}{\partial t}(y,t)\,\nu(dy) \tag{2.4}$$

が成り立つ。 □

定理 2.1 の証明において最も重要なステップは、事実 2.2 の前提が満たされることの確認である。そのための補題を次に示す。

補題 2.3 (優関数の存在). e^i $(i = 1, \ldots, m)$ を $V^\vee$ の基底とし、この基底が定める Θ° 上のチャートを $\varphi = (\theta_1, \ldots, \theta_m)\colon \Theta^\circ \to \mathbb{R}^m$ とおく。このとき、任意の $k \in \mathbb{Z}_{\geq 1}$, $i_1, \ldots, i_k \in \{1, \ldots, m\}$ に対し、次が成り立つ:

(1) 任意の $\theta \in \Theta^\circ$ に対し、関数 $\partial_{i_k} \cdots \partial_{i_1} h(\cdot, \theta)\colon \mathcal{X} \to \mathbb{R}$ は $L^1(\mathcal{X}, \mu)$ に属する。

(2) 任意の $\theta \in \Theta^\circ$ に対し、Θ° における θ のある近傍 U と、ある μ-可積分関数 $\Phi\colon \mathcal{X} \to \mathbb{R}$ が存在し、すべての $\theta' \in U$ に対し $\left|\partial_{i_k} \cdots \partial_{i_1} h(x, \theta')\right| \leq \Phi(x)$ a.e.x が成り立つ。

証明 (1) は (2) より直ちに従うから、(2) を示す。そこで $\theta \in \Theta^\circ$ を任意とする。補題の主張は座標 $\theta_1, \ldots, \theta_m$ を平行移動して考えても等価だから、点 θ の座標は $\varphi(\theta) = 0 \in \mathbb{R}^m$ であるとしてよい。

Step 1: U の構成　$\varepsilon > 0$ を十分小さく選び、$\mathbb{R}^m$ 内の閉立方体

$$A_{2\varepsilon} := \prod_{i=1}^{m} [-2\varepsilon, 2\varepsilon] \quad A_\varepsilon := \prod_{i=1}^{m} [-\varepsilon, \varepsilon] \tag{2.5}$$

が $\varphi(\Theta^\circ)$ に含まれるようにしておく。すると $U := \varphi^{-1}(\operatorname{Int} A_\varepsilon) \subset \varphi(\Theta^\circ)$ は 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。

Step 2: h の座標表示　まず具体的な計算のために h の座標表示を求める。いま各 $\theta' \in U$ に対し

$$h(x, \theta') = \exp\langle \theta', T(x) \rangle = \exp\langle \theta_i(\theta') e^i, T(x) \rangle = \exp\left(\theta_i(\theta') T^i(x)\right) \tag{2.6}$$

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

$$\partial_{i_k} \cdots \partial_{i_1} h(x, \theta') = T^{i_1}(x) \cdots T^{i_k}(x) \exp\left(\theta_i(\theta') T^i(x)\right) \tag{2.7}$$

と表せることがわかる。

Step 3: Φ の構成　Φ を構成するため、式 (2.7) の絶対値を上から評価する。表記の簡略化のため $t' := (t'_1, \ldots, t'_m) := \varphi(\theta') \in \mathbb{R}^m$ とおいておく。まず $\dfrac{k+1}{\varepsilon} \dfrac{\varepsilon}{k+1} = 1$ より

$$\left| T^{i_1}(x) \cdots T^{i_k}(x) \exp\left(\sum_{i=1}^{m} t'_i T^i(x)\right) \right| = \left(\frac{k+1}{\varepsilon}\right)^k \left(\prod_{\alpha=1}^{k} \frac{\varepsilon}{k+1} |T^{i_\alpha}(x)|\right) \exp\left(\sum_{i=1}^{m} t'_i T^i(x)\right) \tag{2.8}$$

であり、$\prod$ の部分を評価すると

$$\prod_{\alpha=1}^{k} \frac{\varepsilon}{k+1} |T^{i_\alpha}(x)| \leq \prod_{\alpha=1}^{k} \left(\exp\left(\frac{\varepsilon}{k+1} T^{i_\alpha}(x)\right) + \exp\left(-\frac{\varepsilon}{k+1} T^{i_\alpha}(x)\right)\right) \quad (\because s \leq e^s + e^{-s} \ (s \in \mathbb{R})) \tag{2.9}$$

$$= \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^{k} \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x)\right) \quad (\because \text{式の展開}) \tag{2.10}$$

(ただし σ_α は σ の第 α 成分) となるから、式 (2.8) と式 (2.10) を合わせて

$$(2.8) \leq C \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^{k} \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x)\right) \exp\left(\sum_{i=1}^{m} t'_i T^i(x)\right) \tag{2.11}$$

$$= C \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{\alpha=1}^{k} \frac{\varepsilon}{k+1} \sigma_\alpha T^{i_\alpha}(x) + \sum_{i=1}^{m} t'_i T^i(x)\right) \tag{2.12}$$

となる。ただし $C := \left(\frac{k+1}{\varepsilon}\right)^k \in \mathbb{R}_{>0}$ とおいた。ここで最終行の exp の中身について、各 $i = 1, \ldots, m$ に対し $T^i(x)$ の係数を評価することで、ある $t'' \in A_{2\varepsilon}$ が存在して

$$(2.12) = C \sum_{\sigma \in \{\pm 1\}^k} \exp\left(\sum_{i=1}^m t''_i T^i(x)\right) = 2^k C \exp\left(\sum_{i=1}^m t''_i T^i(x)\right) \tag{2.13}$$

と表せることがわかる。そこで $|t''_i| \le 2\varepsilon$ $(i = 1, \ldots, m)$ より

$$(2.13) \le 2^k C \prod_{i=1}^m \left(\exp\left(2\varepsilon T^i(x)\right) + \exp\left(-2\varepsilon T^i(x)\right)\right) \tag{2.14}$$

$$= 2^k C \sum_{\tau \in \{\pm 1\}^m} \exp\left(\sum_{i=1}^m 2\varepsilon \tau_i T^i(x)\right) \tag{2.15}$$

を得る。この右辺は (t' によらないから) θ' によらない $\mathcal{X}$ 上の関数であり、また $\sum$ の各項が $2\varepsilon\tau \in A_{2\varepsilon}$ ゆえに μ-可積分だから式全体も μ-可積分である。したがってこれが求める優関数である。 □

目標の定理 2.1 を証明する。

定理 2.1 の証明. 定理 2.1 のステートメントで与えられているチャート $\varphi = (\theta_1, \ldots, \theta_m)$ は ($V^\vee$ の基底が定めるものとは限らない) 任意のものであるが、実は定理の主張を示すには、$V^\vee$ の基底をひとつ選び、その基底が定めるチャート $\widetilde{\varphi} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_m)$ に対して定理の主張を示せば十分である。その理由は次である:

- 式 (2.3) の左辺の微分可能性は、λ が C^∞ であればよいから、チャート $\widetilde{\varphi}$ で考えれば十分。
- 式 (2.3) の右辺の可積分性および式 (2.3) の等号の成立については、Leibniz 則より、λ の $\widetilde{\theta}_1, \ldots, \widetilde{\theta}_m$ に関する k 回偏導関数が、λ の $\theta_1, \ldots, \theta_m$ に関する k 回以下の偏導関数たちの (x によらない) $C^\infty(\Theta^\circ)$-係数の線型結合に書けることから従う。

そこで、以降 φ は $V^\vee$ の基底が定めるチャートとする。

補題 2.3 (1) より、式 (2.3) の右辺の可積分性はわかっている。よって、残りの示すべきことは

(i) 式 (2.3) の左辺の微分可能性
(ii) 式 (2.3) の等号の成立

の 2 点である。

まず $k = 1, i_k = 1$ の場合に (i), (ii) が成り立つことを示す。そのためには、$t = (t_1, \ldots, t_m) \in \varphi(\Theta^\circ)$ を任意に固定したとき、t_1 を含む $\mathbb{R}$ の十分小さな開区間 I が存在して、関数

$$g\colon \mathcal{X} \times I \to \mathbb{R}, \quad (x, s) \mapsto h(x, \varphi^{-1}(s, t_2, \ldots, t_m)) \tag{2.16}$$

が事実 2.2 の仮定 (1), (2) をみたすことをいえばよい。

いま $\varphi^{-1}(t) \in \Theta^\circ$ だから、補題 2.3(2) のいう Θ° における $\varphi^{-1}(t)$ の近傍 U と μ-可積分関数 $\Phi\colon \mathcal{X} \to \mathbb{R}$ が存在する。このとき $\varphi(U)$ は $\mathbb{R}^m$ における t の近傍となるから、t_1 を含む $\mathbb{R}$ の十分小さな開区間 I が存在して

$$I \times \{t_2\} \times \cdots \times \{t_m\} \subset \varphi(U) \tag{2.17}$$

が成り立つ。この I を用いて定まる関数 g が事実 2.2 の仮定 (1), (2) をみたすことを確認する。

まず補題 2.3 の結果 (1) より、g は事実 2.2 の仮定 (1) をみたす。また補題 2.3 の結果 (2) より、g は事実 2.2 の仮定 (2) をみたす。したがって $k = 1, i_k = 1$ の場合について (i),(ii) が示された。

同様にして $i_k = 2, \dots, m$ の場合についても示される。以降、k に関する帰納法で、すべての $k \in \mathbb{Z}_{\geq 1}$ および $i_1, \dots, i_k \in \{1, \dots, m\}$ に対して示される。これで定理の証明が完了した。 □

定理 2.1 から次の系が従う。

系 2.4. $\varphi = (\varphi_1, \dots, \varphi_m) \colon \Theta^\circ \to \mathbb{R}^m$ を $V^\vee$ の基底が定めるチャートとする。また、各 $\theta \in \Theta$ に対し、$\mathcal{X}$ 上の確率測度 P_θ を $P_\theta(dx) = e^{\langle \theta, T(x) \rangle - \psi(\theta)} \mu(dx)$ と定める。このとき、任意の $k \in \mathbb{Z}_{\geq 1}$, $i_1, \dots, i_k \in \{1, \dots, m\}$ に対し、

$$E_{P_\theta}[T^{i_k}(x) \cdots T^{i_1}(x)] = \frac{\partial_{i_k} \cdots \partial_{i_1} \lambda(\theta)}{\lambda(\theta)} \quad (\theta \in \Theta^\circ) \tag{2.18}$$

が成り立つ。ただし、左辺の期待値の存在も系の主張に含まれる。 □

例 2.5 (正規分布族における原点周りのモーメント). $\mathcal{P}$ が $\mathcal{X} = \mathbb{R}$ 上の正規分布族であるとき、任意の $P \in \mathcal{P}$ に対し、P に関する $x, x^2, \dots$ の期待値 $E_P[x], E_P[x^2], \dots$ が存在する。

3 分散の性質

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

前回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。一般に、分散は次の性質を持つ。

定理 3.1 (分散の半正定値対称性). $f \in L^2(\mathcal{X}, p; V)$ 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 まず $\mathrm{Var}_p[f]$ が対称であることを示す。そこで V の基底 e_i $(i = 1, \dots, m)$ をひとつ選んで固定し、$f, E_p[f]$ の成分表示をそれぞれ $f^i \colon \mathcal{X} \to \mathbb{R}$ および $a^i \in \mathbb{R}$ $(i = 1, \dots, m)$ とおく。すると

$$\mathrm{Var}_p[f] = E_p[(f - E_p[f])^2] = \left(\int_{\mathcal{X}} (f^i(x) - a^i)(f^j(x) - a^j)\, p(dx) \right) e_i e_j \tag{3.1}$$

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 $\mathrm{Var}_p[f]$ が半正定値であることを示す。示すべきことは、$\mathrm{Var}_p[f]$ 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 $\mathrm{Var}_p[f](\theta, \theta) \geq 0$ が成り立つことであるが、これは

$$\mathrm{Var}_p[f](\theta, \theta) = \sum_{i,j=1}^{m} \left(\int_{\mathcal{X}} (f^i(x) - a^i)(f^j(x) - a^j)\, p(dx) \right) \theta(e_i)\theta(e_j) \tag{3.2}$$

$$= \int_{\mathcal{X}} \left(\sum_{i=1}^{m} \theta(e_i)(f^i(x) - a^i) \right)^2 p(dx) \tag{3.3}$$

$$\geq 0 \tag{3.4}$$

より従う。したがって $\mathrm{Var}_p[f]$ 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 3.2 (分散が 0 であるための必要十分条件). $f \in L^2(\mathcal{X}, p; V)$ 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 3.3. $\mathcal{Y}$ 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) $\displaystyle\int_{\mathcal{Y}} g(y)\, \mu(dy) = 0$

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

□

命題 3.2 の証明. ここでは「p-a.e.」を「a.e.」と略記する。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 $f^i \colon \mathcal{X} \to \mathbb{R}$ および $a^i \in \mathbb{R}$ $(i = 1, \ldots, m)$ 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 $\mathrm{Var}_p[f] = 0$ 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 3.3 より $\displaystyle\int_{\mathcal{X}} (f^i(x) - a^i)^2\, p(dx) > 0$ である。したがって $\mathrm{Var}_p[f] \neq 0$ である。 □

4 今後の予定

- Hessian の定義と ψ の Hessian の正定値性を示す。
- KL ダイバージェンスを定義する。
- Fisher 計量を定義する。

5 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.

[Dud03] Richard Dudley, **18.466 Mathematical Statistics, Spring 2003**, 2003, https://dspace.mit.edu/handle/1721.1/103814, Last accessed on 2023-05-14.