1 振り返りと導入

前回は指数型分布族を定義した。本稿では、前回発表時の指摘を取り入れながら、指数型分布族と最小次元実現について簡単に整理しなおす。その後、新たなトピックとして期待値と分散の定義を行う。

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 指数型分布族と最小次元実現

本稿では、指数型分布族の定義は次のものを用いる。注意点として、前回は有限次元 $\mathbb{R}$-ベクトル空間として V ではなく $\mathbb{R}^m$ を用いていたが、より一般の状況に対応するため、本稿からは V を用いるように変更する。なお、この変更により前回の定理の証明は修正が必要となるが、その内容については本稿では扱わない (修正内容は `0425_コメント.pdf` に書きました)。

定義 2.1 (指数型分布族). $\mathcal{X}$ を可測空間、$\emptyset \neq \mathcal{P} \subset \mathcal{P}(\mathcal{X})$ とする。$\mathcal{P}$ が $\mathcal{X}$ 上の**指数型分布族 (exponential family)** であるとは、次が成り立つことをいう: $\exists\,(V, T, \mu)$ s.t.

- **(E0)** V は有限次元 $\mathbb{R}$-ベクトル空間である。
- **(E1)** $T\colon \mathcal{X} \to V$ は可測写像である。
- **(E2)** μ は $\mathcal{X}$ 上の σ-有限測度であり、$\forall\, p \in \mathcal{P}$ に対し $p \ll \mu$ をみたす。
- **(E3)** $\forall\, p \in \mathcal{P}$ に対し、$\exists\, \theta \in V^\vee$ s.t.

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x)\rangle}{\int_{\mathcal{X}} \exp\langle \theta, T(y)\rangle\, \mu(dy)} \quad \mu\text{-a.e. } x \in \mathcal{X} \tag{2.1}$$

である。ただし $\langle \cdot, \cdot \rangle$ は自然なペアリング $V^\vee \times V \to \mathbb{R}$ である。

さらに次のように定める:

- (V, T, μ) を $\mathcal{P}$ の**実現 (representation)** という。
  - m を (V, T, μ) の**次元 (dimension)** という。
  - T を (V, T, μ) の**十分統計量 (sufficient statistic)** という。
  - μ を (V, T, μ) の**基底測度 (base measure)** という。
- 集合 $\Theta_{(V,T,\mu)}$

$$\Theta_{(V,T,\mu)} := \left\{ \theta \in V^\vee \,\middle|\, \int_{\mathcal{X}} \exp\langle \theta, T(y)\rangle\, \mu(dy) < +\infty \right\} \tag{2.2}$$

  を (V, T, μ) の**自然パラメータ空間 (natural parameter space)** という。
- 関数 $\psi\colon \Theta_{(V,T,\mu)} \to \mathbb{R}$,

$$\psi(\theta) := \log \int_{\mathcal{X}} \exp\langle \theta, T(y)\rangle\, \mu(dy) \tag{2.3}$$

  を (V, T, μ) の**対数分配関数 (log-partition function)** という。

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 $\mathcal{P}$ の実現とする。

命題 2.2. (V, T, μ) に関する次の条件は同値である:

(1) $\langle \theta, T(x)\rangle$ 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。

(2) 各 $p \in \mathcal{P}$ に対し、定義 2.1 の条件 (E3) をみたす $\theta \in V^\vee$ はただひとつである。

証明 (2) ⇒(1) 前回示した。

(1) ⇒(2) $\theta, \theta' \in V^\vee$ が定義 2.1 の条件 (E3) をみたすとすると、

$$e^{\langle \theta, T(x) \rangle - \psi(\theta)} = \frac{dp}{d\mu}(x) = e^{\langle \theta', T(x) \rangle - \psi(\theta')} \qquad \mu\text{-a.e.}x \in \mathcal{X} \tag{2.4}$$

が成り立つ。式を整理して

$$\langle \theta - \theta', T(x) \rangle = \psi(\theta) - \psi(\theta') \qquad \mu\text{-a.e.}x \in \mathcal{X} \tag{2.5}$$

が成り立つ。したがって (1) より $\theta = \theta'$ である。 □

定義 2.3 (最小次元実現). 実現 (V, T, μ) が $\mathcal{P}$ の実現のうちで次元が最小のものであるとき、(V, T, μ) を $\mathcal{P}$ の**最小次元実現 (minimal representation)** という。

次の事実は前回示した。

事実 2.4. 最小次元実現は存在する。 □

事実 2.5. 最小次元実現は命題 2.2 の 2 条件をみたす。 □

3 期待値と分散

ここで一旦指数型分布族の話題から離れて、可測関数 (あるいは確率変数) の期待値と分散を定義しておく。これらの概念は、期待値パラメータ空間の定義や、対数分配関数の性質を調べる際に重要な役割を果たす。

ここでは期待値および分散の概念をベクトル値関数に対して定義したい。そこで、まずベクトル値関数の積分および可積分性を次のように定めておく。

定義 3.1 (ベクトル値関数の積分). $\mathcal{X}$ を可測空間、V を有限次元 $\mathbb{R}$-ベクトル空間、p を $\mathcal{X}$ 上の確率測度、$f\colon \mathcal{X} \to V$ を可測写像とする。V のある基底 $e^1, \ldots, e^m$ が存在して、この基底に関する f の成分 $f_i\colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \ldots, m)$ がすべて p-可積分であるとき、f は p に関し**可積分 (integrable)** であるという (well-defined 性はこのあと示す)。

f が p-可積分であるとき、f の p に関する**積分 (integral)** を

$$\int_X f(x)\, p(dx) := \left(\int_X f_i(x)\, p(dx) \right) e^i \in V \tag{3.1}$$

で定義する (well-defined 性はこのあと示す)。

ただし $\dim V = 0$ の場合は f は p-可積分で $\displaystyle\int_X f(x)\, p(dx) = 0$ と約束する。

注意 3.2. $V = \mathbb{R}$ の場合は $\mathbb{R}$-値関数の通常の積分に一致する。

well-defined 性の証明. f が p-可積分であるかどうかは V の基底の取り方によらないことを示す。そこで、$e^1, \dots, e^m$ および $\tilde{e}^1, \dots, \tilde{e}^m$ をそれぞれ V の基底とし、それぞれの基底に関する f の成分を f_i, $\tilde{f}_i \colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \dots, m)$ とおく。示すべきことは「$\tilde{f}_i$ $(i = 1, \dots, m)$ がすべて $L^1(\mathcal{X}, p)$ に属するならば f_i $(i = 1, \dots, m)$ もすべて $L^1(\mathcal{X}, p)$ に属する」ということである。このことは、$L^1(\mathcal{X}, p)$ が $\mathbb{R}$-ベクトル空間であることと、f_i たちが $\tilde{f}_i$ たちの $\mathbb{R}$-線型結合であることから従う。よって f が p-可積分であるかどうかは V の基底の取り方によらない。

次に、f の p に関する積分は V の基底の取り方によらないことを示す。e^i, $\tilde{e}^i$ をそれぞれ V の基底とする。いま、ある $a_i^j \in \mathbb{R}$ $(i, j = 1, \dots, m)$ が存在して $f_i = a_i^j \tilde{f}_j$ $(i = 1, \dots, m)$ および $\tilde{e}^j = a_i^j e^i$ $(j = 1, \dots, m)$ が成り立っているから、

$$\left(\int_{\mathcal{X}} \tilde{f}_j \, p(dx) \right) \tilde{e}^j = \left(\int_{\mathcal{X}} \tilde{f}_j \, p(dx) \right) a_i^j e^i \tag{3.2}$$

$$= \left(\int_{\mathcal{X}} a_i^j \tilde{f}_j \, p(dx) \right) e^i \quad (\text{積分の } \mathbb{R}\text{-線型性}) \tag{3.3}$$

$$= \left(\int_{\mathcal{X}} f_i \, p(dx) \right) e^i \tag{3.4}$$

が成り立つ。これで積分の well-defined 性も示せた。 □

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

ベクトル値関数の積分を用いて、期待値および分散を定義する。

定義 3.3 (期待値). f が p-可積分であるとき、f の p に関する**期待値 (expected value)** $E_p[f]$ を

$$E_p[f] \coloneqq \int_{\mathcal{X}} f(x) \, p(dx) \in V \tag{3.5}$$

と定義する。

つぎに分散を定義する。$\mathbb{R}$-値関数の通常の分散との類似で、ベクトル値関数 f の分散を"$V_p[f] \coloneqq E_p[(f - E_p[f]) \otimes (f - E_p[f])]$" と定義したい。その準備として次の補題を示しておく。

補題 3.4 (分散の存在条件). 可測写像 $f \colon \mathcal{X} \to V$ に関し次の 2 条件は同値である:

(1) f および $(f - E_p[f]) \otimes (f - E_p[f])$ が p-可積分

(2) $f \otimes f$ が p-可積分

この補題の証明には次の事実を用いる。

事実 3.5. $\mathcal{Y}$ を可測空間、μ を $\mathcal{Y}$ 上の有限測度とする。このとき、任意の実数 $1 < p < +\infty$ に対し $L^p(\mathcal{Y}, \mu) \subset L^1(\mathcal{Y}, \mu)$ が成り立つ。

補題 3.4 の証明. $\dim V = 0$ の場合は明らかに成り立つ。以後 $\dim V \geq 1$ の場合を考える。V の基底 $e^1, \dots, e^m$ をひとつ選んで固定し、この基底に関する f の成分を $f_i \colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \dots, m)$ とおいておく。

(1) ⇒(2)　f が p-可積分であることより $E_p[f] \in V$ が存在するから、これを $a := E_p[f]$ とおき、V の基底 e^i に関する a の成分を $a_i \in \mathbb{R}$ $(i = 1, \ldots, m)$ とおいておく。示すべきことは、すべての $i, j = 1, \ldots, m$ に対し $f_i f_j \in L^1(\mathcal{X}, p)$ が成り立つことである。そこで次のことに注意する:

(i) p が確率測度であることより $1 \in L^1(\mathcal{X}, p)$ である。
(ii) f が p-可積分であることより $f_i \in L^1(\mathcal{X}, p)$ $(i = 1, \ldots, m)$ である。
(iii) $(f - a) \otimes (f - a)$ が p-可積分であることより $(f_i - a_i)(f_j - a_j) = f_i f_j - a_i f_j - a_j f_i + a_i a_j \in L^1(\mathcal{X}, p)$ $(i, j = 1, \ldots, m)$ である。

したがって、$L^1(\mathcal{X}, p)$ が $\mathbb{R}$-ベクトル空間であることとあわせて $f_i f_j \in L^1(\mathcal{X}, p)$ $(i, j = 1, \ldots, m)$ が成り立つ。よって $f \otimes f$ は p-可積分である。

(2) ⇒(1)　まず f が p-可積分であることを示す。そのためには、$f_i \in L^1(\mathcal{X}, p)$ $(i = 1, \ldots, m)$ が成り立つことをいえばよい。いま $f \otimes f$ が p-可積分であるから、$f_i f_j \in L^1(\mathcal{X}, p)$ $(i, j = 1, \ldots, m)$ が成り立つ。とくにすべての $i = 1, \ldots, m$ に対し $f_i \in L^2(\mathcal{X}, p)$ が成り立つから、事実 3.5 とあわせて $f_i \in L^1(\mathcal{X}, p)$ が成り立つ。よって f は p-可積分である。

つぎに $(f - E_p[f]) \otimes (f - E_p[f])$ が p-可積分であることを示す。いま f が p-可積分であることより $E_p[f] \in V$ が存在するから、これを $a := E_p[f]$ とおき、V の基底 e^i に関する a の成分を $a_i \in \mathbb{R}$ $(i = 1, \ldots, m)$ とおいておく。示したいことは、$(f_i - a_i)(f_j - a_j) = f_i f_j - a_i f_j - f_i a_j + a_i a_j \in L^1(\mathcal{X}, p)$ $(i, j = 1, \ldots, m)$ が成り立つことである。そこで次のことに注意する:

(i) p が確率測度であることより $1 \in L^1(\mathcal{X}, p)$ である。
(ii) f が p-可積分であることより $f_i \in L^1(\mathcal{X}, p)$ $(i = 1, \ldots, m)$ である。
(iii) $f \otimes f$ が p-可積分であることより $f_i f_j \in L^1(\mathcal{X}, p)$ $(i, j = 1, \ldots, m)$ である。

したがって、$L^1(\mathcal{X}, p)$ が $\mathbb{R}$-ベクトル空間であることとあわせて $(f_i - a_i)(f_j - a_j) = f_i f_j - a_i f_j - f_i a_j + a_i a_j \in L^1(\mathcal{X}, p)$ $(i, j = 1, \ldots, m)$ が成り立つ。よって $(f - E_p[f]) \otimes (f - E_p[f])$ は p-可積分である。 □

この補題を踏まえて分散を定義する。

定義 3.6 (分散). $f \otimes f : \mathcal{X} \to V \otimes_{\mathbb{R}} V$ が p-可積分であるとき、f の p に関する**分散 (variance)** $V_p[f]$ を

$$V_p[f] := E_p[(f - E_p[f]) \otimes (f - E_p[f])] \in V \otimes V \tag{3.6}$$

と定義する (補題 3.4 よりこれは存在する)。

正規分布族の十分統計量について期待値と分散を求めてみる。

例 3.7 (正規分布族の十分統計量の期待値と分散). $\mathcal{X} = \mathbb{R}$、λ を $\mathbb{R}$ 上の Lebesgue 測度とし、正規分布族

$$\mathcal{P} := \left\{ P_{(\mu,\sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \,\middle|\, \mu \in \mathbb{R},\ \sigma^2 > 0 \right\} \tag{3.7}$$

と $\mathcal{P}$ の実現 (V, T, μ), $V = \mathbb{R}^2$, $T : \mathcal{X} \to V$, $x \mapsto {}^t(x, x^2)$ を考える。各 $P = P_{(\mu,\sigma^2)} \in \mathcal{P}$ に対し、T の期待値 $E_p[T] \in V$ と分散 $V_p[T] \in V \otimes V$ を求めてみる。ただし、以下 $x, \ldots, x^4$ の P に関する可積分性は仮定する (可積分性は次回示す)。

まず期待値を求める。求めるべきものは、$V = \mathbb{R}^2$ の標準基底を e_1, e_2 として

$$E_P[T] = E_P[x]\, e_1 + E_P[x^2]\, e_2 \tag{3.8}$$

である。各成分は $E_P[x] = \mu,\ E_P[x^2] = E_P[(x-\mu)^2] + E_P[x]^2 = \sigma^2 + \mu^2 \in \mathbb{R}$ と求まるから

$$E_P[T] = \mu\, e_1 + (\sigma^2 + \mu^2)\, e_2 \tag{3.9}$$

である。

次に分散を求める。求めるべきものは

$$V_P[T] = E_P[(T - E_P[T]) \otimes (T - E_P[T])] \tag{3.10}$$

である。これを $V \otimes V$ の基底 $e_i \otimes e_j\ (i, j = 1, 2)$ に関して成分表示すると

$$V_P[T] = E_P[(x-\mu)^2]\, e_1 \otimes e_1 \tag{3.11}$$
$$+ E_P[(x-\mu)(x^2 - (\sigma^2 + \mu^2))]\, (e_1 \otimes e_2 + e_2 \otimes e_1) \tag{3.12}$$
$$+ E_P[(x^2 - (\sigma^2 + \mu^2))^2]\, e_2 \otimes e_2 \tag{3.13}$$

と表される。そこで原点周りのモーメント $a_3 \coloneqq E_P[x^3],\ a_4 \coloneqq E_P[x^4] \in \mathbb{R}$ とおくと、各成分は

$$E_P[(x-\mu)^2] = \sigma^2 \tag{3.14}$$
$$E_P[(x-\mu)(x^2 - (\sigma^2 + \mu^2))] = a_3 - \mu(\sigma^2 + \mu^2) \tag{3.15}$$
$$E_P[(x^2 - (\sigma^2 + \mu^2))^2] = a_4 - (\sigma^2 + \mu^2)^2 \tag{3.16}$$

と求まる。したがって $V_P[T]$ は

$$V_P[T] = \sigma^2\, e_1 \otimes e_1 \tag{3.17}$$
$$+ (a_3 - \mu(\sigma^2 + \mu^2))\, (e_1 \otimes e_2 + e_2 \otimes e_1) \tag{3.18}$$
$$+ (a_4 - (\sigma^2 + \mu^2)^2)\, e_2 \otimes e_2 \tag{3.19}$$

と表される。最後に原点周りのモーメント a_3, a_4 を具体的に求める。これは期待値周りのモーメントの計算に帰着される。そこで標準正規分布を $P_0 \coloneqq P_{(0,1)} \in \mathcal{P}$ とおくと、$E_P\left[\left(\frac{x-\mu}{\sigma}\right)^k\right] = E_{P_0}[x^k]\ (k = 3, 4)$ より $E_P[(x-\mu)^k] = \sigma^k E_{P_0}[x^k]\ (k = 3, 4)$ が成り立つ。ここで P_0 に関する期待値を部分積分などを用いて直接計算すると $E_{P_0}[x^3] = 0,\ E_{P_0}[x^4] = 3$ となるから、$E_P[(x-\mu)^3] = 0,\ E_P[(x-\mu)^4] = 3\sigma^4$ を得る。これらを用いて a_3, a_4 を計算すると

$$0 = E_P[(x-\mu)^3] \tag{3.20}$$
$$= E_P[x^3] - 3E_P[x^2]\mu + 3E_P[x]\mu^2 - \mu^3 \tag{3.21}$$
$$= a_3 - 3(\sigma^2 + \mu^2)\mu + 3\mu^3 - \mu^3 \tag{3.22}$$
$$= a_3 - 3\sigma^2\mu - \mu^3 \tag{3.23}$$
$$\therefore a_3 = 3\sigma^2\mu + \mu^3 \tag{3.24}$$

および

$$3\sigma^4 = E_P[(x-\mu)^4] \tag{3.25}$$

$$= E_P[x^4] - 4E_P[x^3]\mu + 6E_P[x^2]\mu^2 - 4E_P[x]\mu^3 + \mu^4 \tag{3.26}$$
$$= a_4 - 4a_3\mu + 6(\sigma^2 + \mu^2)\mu^2 - 4\mu^4 + \mu^4 \tag{3.27}$$
$$= a_4 - 6\sigma^2\mu^2 - \mu^4 \tag{3.28}$$
$$\therefore a_4 = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4 \tag{3.29}$$

を得る。これらを $V_P[T]$ の成分表示に代入して

$$V_P[T] = \sigma^2\, e_1 \otimes e_1 \tag{3.30}$$
$$+ 2\sigma^2\mu\, (e_1 \otimes e_2 + e_2 \otimes e_1) \tag{3.31}$$
$$+ (4\sigma^2\mu^2 + 2\sigma^4)\, e_2 \otimes e_2 \tag{3.32}$$

となる。行列表示は $\begin{bmatrix} \sigma^2 & 2\sigma^2\mu \\ 2\sigma^2\mu & 4\sigma^2\mu^2 + 2\sigma^4 \end{bmatrix} \in M_2(\mathbb{R})$ となり、これは対称かつ正定値である。

注意 3.8 (Fisher 情報行列に関する (あまり厳密でない) 補足). 上の例の $V_P[T]$ は座標変換 $\mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}^2$, $(\mu, \sigma) \mapsto (\theta_1, \theta_2) = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$ の Jacobi 行列 J による合同変換により ${}^tJV_P[T]J = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$ に写る。これは正規分布のパラメータ (μ, σ) に関する Fisher 情報行列に一致する (詳細は次回以降で Fisher 計量の話題のときに触れたい)。

上の例でみた $V_P[T]$ は正定値対称であった。より一般に分散は次の性質を持つ。

定理 3.9 (分散の半正定値対称性). $V_p[f] \in V \otimes_{\mathbb{R}} V$ は、$V^\vee$ 上の $\mathbb{R}$-双線型形式とみて対称かつ半正定値である。

注意 3.10 (双線型形式としてのテンソル). 一般に、$\omega \in V \otimes V$ を $V^\vee$ 上の $\mathbb{R}$-双線型形式とみるとは、V の基底 e^i $(i = 1, \dots, m)$ をひとつ選んで ω の成分表示を $\omega = \omega_{ij} e^i \otimes e^j$, $\omega_{ij} \in \mathbb{R}$ とおくとき、

$$\omega(\theta, \theta') := \omega_{ij}\theta(e^i)\theta'(e^j) \in \mathbb{R} \tag{3.33}$$

と定めるということである (これは well-defined である)。

証明　まず $V_p[f]$ 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, m_i $(i = 1, \dots, m)$ とおく。すると

$$V_p[f] = E_p[(f - E_p[f]) \otimes (f - E_p[f])] \tag{3.34}$$
$$= \left(\int_X (f_i(x) - m_i)(f_j(x) - m_j)\, p(dx) \right) e^i \otimes e^j \tag{3.35}$$

となり、最後の行の $\int_X (f_i(x) - m_i)(f_j(x) - m_j)\, p(dx)$ は添字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、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

$$V_p[f](\theta,\theta) = \sum_{i,j}\left(\int_{\mathcal{X}}(f_i(x)-m_i)(f_j(x)-m_j)\,p(dx)\right)\theta(e^i)\theta(e^j) \tag{3.36}$$

$$= \int_{\mathcal{X}}\left(\sum_{i,j}\theta(e^i)(f_i(x)-m_i)\theta(e^j)(f_j(x)-m_j)\right)p(dx) \tag{3.37}$$

$$= \int_{\mathcal{X}}\left(\sum_{i}\theta(e^i)(f_i(x)-m_i)\right)^2 p(dx) \tag{3.38}$$

$$\geq 0 \tag{3.39}$$

より従う。したがって $V_p[f]$ は半正定値である。 □

最後に、分散が 0 であることの特徴づけを与えておく。

命題 3.11 (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 $f\colon \mathcal{X}\to V$ に関し、$V_p[f]=0$ であることと f が p-a.e. 定数であることは同値である。

証明には次の事実を用いる。

事実 3.12. $\mathcal{Y}$ を可測空間、μ を $\mathcal{Y}$ 上の測度とする。このとき、$g(x)\geq 0$ μ-a.e.x なる μ-可積分関数 $g\colon\mathcal{Y}\to\mathbb{R}$ に関し、$\int_{\mathcal{Y}} g(x)\,\mu(dx)=0$ であることと $g=0$ μ-a.e.x であることは同値である。

命題 3.11 の証明. ここでは「p-a.e.」を「a.e.」と略記する。V の基底 e^i $(i=1,\dots,m)$ をひとつ選んで固定し、この基底に関する f の成分を $f_i\colon\mathcal{X}\to\mathbb{R}$ $(i=1,\dots,m)$、$E_p[f]$ の成分を $a_i\in\mathbb{R}$ $(i=1,\dots,m)$ とおいておく。

<u>(⇐)</u> f が a.e. 定数ならば、$f_i(x)=a_i$ a.e.x $(i=1,\dots,m)$ したがって $(f_i(x)-m_i)(f_j(x)-m_j)=0$ a.e.x $(i,j=1,\dots,m)$ である。よって $\int_{\mathcal{X}}(f_i(x)-m_i)(f_j(x)-m_j)\,p(dx)=0$ $(i,j=1,\dots,m)$ だから $V_p[f]=0$ である。

<u>(⇒)</u> 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 $i\in\{1,\dots,m\}$ が存在する。このとき $(f_i-m_i)^2=0$ a.e. ではないから、事実 3.12 より $\int_{\mathcal{X}}(f_i(x)-m_i)^2\,p(dx)>0$ である。したがって $V_p[f]\neq 0$ である。 □

4 今後の予定

- 期待値パラメータ空間
- 対数分配関数の微分可能性および Hessian の正定値性
- 対数分配関数から誘導されるダイバージェンス

[TODO] このような流れでよい?

5 参考文献

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

[WJ07] Martin J. Wainwright and Michael I. Jordan, **Graphical Models, Exponential Families, and Variational Inference**, Foundations and Trends in Machine Learning **1** (2007).

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