

目次

第 1 章	確率論	2
1	Radon-Nikodým の定理と Hölder の不等式	2
2	確率論の基本事項	2
1.2 A	確率空間	2
3	期待値と分散	3
第 2 章	Hessian	9
1	Hessian	9
2.1 A	U 上のアファイン接続	9
2.1 B	Hessian	10
第 3 章	指数型分布族	11
1	指数型分布族	11
2	対数分配関数	15
3	Fisher 計量	18
参考文献		20
記号一覧		21
索引		22

第1章 確率論

1 Radon-Nikodým の定理と Hölder の不等式

[TODO] σ -加法族は省略して書く

定義 1.1 (絶対連続). (X, \mathcal{B}) を可測空間、 μ, ν を X 上の測度とする。 ν が μ に関し**絶対連続 (absolutely continuous)** であるとは、任意の $E \in \mathcal{B}$ に対し $\mu(E) = 0$ ならば $\nu(E) = 0$ が成り立つことをいう。

定理 1.2 (Radon-Nikodým の定理). (X, \mathcal{B}) を可測空間、 μ を X 上の σ -有限測度、 ν を X 上の測度とする。このとき、 ν が μ に関して絶対連続であるための必要十分条件は、 μ -a.e. $x \in X$ に対し定義された可積分関数 f が存在して

$$\nu(E) = \int_E f(x) d\mu(x) \quad (E \in \mathcal{B}) \quad (1.1)$$

が成り立つことである。この f を μ に関する ν の **Radon-Nikodým 微分 (Radon-Nikodým derivative)** といい、 $\frac{d\nu}{d\mu}$ と書く。

証明 関数族の \sup として f を構成する。 □

命題 1.3 (Hölder の不等式). (X, \mathcal{B}) を可測空間、 μ を X 上の測度とする。 $1 < p < \infty$ 、 $q = p(p-1)^{-1}$ 、 f を p 乗 μ -可積分関数、 g を q 乗 μ -可積分関数とする。このとき、 fg は μ -可積分であり、かつ

$$\int_X |fg| \mu(dx) \leq \left(\int_X |f|^p \mu(dx) \right)^{\frac{1}{p}} \left(\int_X |g|^q \mu(dx) \right)^{\frac{1}{q}} \quad (1.2)$$

が成り立つ。

証明 Young の不等式を使う。 □

2 確率論の基本事項

A. 確率空間

定義 2.1 (確率空間). 測度空間 (Ω, \mathcal{F}, P) であって

- (1) 各 $E \in \mathcal{F}$ に対し $P(E) \geq 0$
- (2) $P(\Omega) = 1$

をみたすものを**確率空間 (probability space)** といい、 P を (Ω, \mathcal{F}) 上の**確率測度 (probability measure)** あるいは**確率分布 (probability distribution)** という。

定義 2.2 (確率変数). (Ω, \mathcal{F}, P) を確率空間、 (X, \mathcal{A}) を可測空間とする。可測関数 $X: (\Omega, \mathcal{F}) \rightarrow (X, \mathcal{A})$ を (X, \mathcal{A}) に値をもつ **確率変数 (random variable; r.v.)** という。

定義 2.3 (確率変数の確率分布). (Ω, \mathcal{F}, P) を確率空間、 $X: (\Omega, \mathcal{F}) \rightarrow (X, \mathcal{A})$ を確率変数とする。このとき、写像

$$P^X: \mathcal{A} \rightarrow [0, +\infty), \quad E \mapsto P(X^{-1}(E)) \quad (E \in \mathcal{A}) \quad (2.1)$$

は (X, \mathcal{A}) 上の確率測度となる。これを **X の確率分布 (probability distribution of X)** という。

X の確率分布が (X, \mathcal{A}) 上のある確率分布 ν に等しいとき、 X は ν に従う という。

定義 2.4 (確率密度関数). (X, \mathcal{A}) を可測空間、 μ を X 上の σ -有限測度、 ν を μ に関し絶対連続な (X, \mathcal{A}) 上の確率測度とする。このとき、 ν の μ に関する Radon-Nikodým 微分 $\frac{d\nu}{d\mu}$ を、 ν の **確率密度関数 (probability density function; PDF)** という。

3 期待値と分散

定義 3.1 (ベクトル値関数の積分). X を可測空間、 V を有限次元 \mathbb{R} -ベクトル空間、 p を X 上の確率測度、 $f: X \rightarrow V$ を可測写像とする。 V のある基底 e^1, \dots, e^m が存在して、この基底に関する f の成分 $f_i: X \rightarrow \mathbb{R}$ ($i = 1, \dots, m$) 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

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

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 $\int_X f(x) p(dx) = 0$ 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 $f_i, \tilde{f}_i: X \rightarrow \mathbb{R}$ ($i = 1, \dots, m$) とおく。示すべきことは「 \tilde{f}_i ($i = 1, \dots, m$) がすべて $L^1(X, p)$ に属するならば f_i ($i = 1, \dots, m$) もすべて $L^1(X, p)$ に属する」ということである。このことは、 $L^1(X, p)$ 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、

$$\left(\int_X \tilde{f}_j p(dx) \right) \tilde{e}^j = \left(\int_X \tilde{f}_j p(dx) \right) a_i^j e^i \quad (3.2)$$

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

$$= \left(\int_X f_i p(dx) \right) e^i \quad (3.4)$$

が成り立つ。これで積分の well-defined 性も示せた。 \square

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

$$E_p[f] := \int_X f(x) p(dx) \in V \quad (3.5)$$

と定義する。

補題 3.4 (分散の存在条件). 可測写像 $f: X \rightarrow V$ 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 $f_i: X \rightarrow \mathbb{R} (i = 1, \dots, m)$ とおいておく。

(1) \Rightarrow (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 $a_i \in \mathbb{R} (i = 1, \dots, m)$ とおいておく。示すべきことは、すべての $i, j = 1, \dots, m$ に対し $f_i f_j \in L^1(X, p)$ 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 $1 \in L^1(X, p)$ である。
- (ii) f が p -可積分であることより $f_i \in L^1(X, p) (i = 1, \dots, m)$ である。
- (iii) $(f - a) \otimes (f - a)$ が p -可積分であることより $(f_i - a_i)(f_j - a_j) = f_i f_j - a_i f_j - a_j f_i + a_i a_j \in L^1(X, p) (i, j = 1, \dots, m)$ である。

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

(2) \Rightarrow (1) まず f が p -可積分であることを示す。そのためには、 $f_i \in L^1(X, p) (i = 1, \dots, m)$ 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、 $f_i f_j \in L^1(X, p) (i, j = 1, \dots, m)$ が成り立つ。とくにすべての $i = 1, \dots, m$ に対し $f_i \in L^2(X, p)$ が成り立つから、事実 3.5 とあわせて $f_i \in L^1(X, p)$ 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 $a_i \in \mathbb{R} (i = 1, \dots, m)$ とおいておく。示したいことは、 $(f_i - a_i)(f_j - a_j) = f_i f_j - a_i f_j - a_j f_i + a_i a_j \in L^1(X, p) (i, j = 1, \dots, m)$ 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 $1 \in L^1(X, p)$ である。
- (ii) f が p -可積分であることより $f_i \in L^1(X, p)$ ($i = 1, \dots, m$) である。
- (iii) $f \otimes f$ が p -可積分であることより $f_i f_j \in L^1(X, p)$ ($i, j = 1, \dots, m$) である。

したがって、 $L^1(X, p)$ が \mathbb{R} -ベクトル空間であることとあわせて $(f_i - a_i)(f_j - a_j) = f_i f_j - a_i f_j - f_i a_j + a_i a_j \in L^1(X, p)$ ($i, j = 1, \dots, m$) 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。 \square

この補題を踏まえて分散を定義する。

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

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

と定義する (補題 3.4 よりこれは存在する)。

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

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

と \mathcal{P} の実現 (V, T, μ) , $V = \mathbb{R}^2$, $T: X \rightarrow V$, $x \mapsto {}^t(x, x^2)$ 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 $V_P[T] \in V \otimes V$ を求めている。ただし、以下 x, \dots, 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

$$E_P[T] = E_P[x] e_1 + E_P[x^2] e_2 \quad (3.8)$$

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

$$E_P[T] = \mu e_1 + (\sigma^2 + \mu^2) e_2 \quad (3.9)$$

である。

次に分散を求める。求めるべきものは

$$V_P[T] = E_P[(T - E_P[T]) \otimes (T - E_P[T])] \quad (3.10)$$

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

$$V_P[T] = E_P[(x - \mu)^2] e_1 \otimes e_1 \quad (3.11)$$

$$+ E_P[(x - \mu)(x^2 - (\sigma^2 + \mu^2))] (e_1 \otimes e_2 + e_2 \otimes e_1) \quad (3.12)$$

$$+ E_P[(x^2 - (\sigma^2 + \mu^2))^2] e_2 \otimes e_2 \quad (3.13)$$

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

$$E_P[(x - \mu)^2] = \sigma^2 \quad (3.14)$$

$$E_P[(x - \mu)(x^2 - (\sigma^2 + \mu^2))] = a_3 - \mu(\sigma^2 + \mu^2) \quad (3.15)$$

$$E_P[(x^2 - (\sigma^2 + \mu^2))^2] = a_4 - (\sigma^2 + \mu^2)^2 \quad (3.16)$$

と求まる。したがって $V_P[T]$ は

$$V_P[T] = \sigma^2 e_1 \otimes e_1 \quad (3.17)$$

$$+ (a_3 - \mu(\sigma^2 + \mu^2))(e_1 \otimes e_2 + e_2 \otimes e_1) \quad (3.18)$$

$$+ (a_4 - (\sigma^2 + \mu^2)^2) e_2 \otimes e_2 \quad (3.19)$$

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 $P_0 := P_{(0,1)} \in \mathcal{P}$ 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

$$0 = E_P[(x - \mu)^3] \quad (3.20)$$

$$= E_P[x^3] - 3E_P[x^2]\mu + 3E_P[x]\mu^2 - \mu^3 \quad (3.21)$$

$$= a_3 - 3(\sigma^2 + \mu^2)\mu + 3\mu^3 - \mu^3 \quad (3.22)$$

$$= a_3 - 3\sigma^2\mu - \mu^3 \quad (3.23)$$

$$\therefore a_3 = 3\sigma^2\mu + \mu^3 \quad (3.24)$$

および

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

$$= E_P[x^4] - 4E_P[x^3]\mu + 6E_P[x^2]\mu^2 - 4E_P[x]\mu^3 + \mu^4 \quad (3.26)$$

$$= a_4 - 4a_3\mu + 6(\sigma^2 + \mu^2)\mu^2 - 4\mu^4 + \mu^4 \quad (3.27)$$

$$= a_4 - 6\sigma^2\mu^2 - \mu^4 \quad (3.28)$$

$$\therefore a_4 = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4 \quad (3.29)$$

を得る。これらを $V_P[T]$ の成分表示に代入して

$$V_P[T] = \sigma^2 e_1 \otimes e_1 \quad (3.30)$$

$$+ 2\sigma^2\mu(e_1 \otimes e_2 + e_2 \otimes e_1) \quad (3.31)$$

$$+ (4\sigma^2\mu^2 + 2\sigma^4) e_2 \otimes e_2 \quad (3.32)$$

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。

命題 3.8 (期待値・分散とペアリング). $f: \mathcal{X} \rightarrow V$ 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 $\text{Var}_p[\langle \omega, f(x) \rangle] = \langle \omega \otimes \omega, \text{Var}_p[f(x)] \rangle$ 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 $f^i: \mathcal{X} \rightarrow \mathbb{R}$, $\omega_i \in \mathbb{R}$ ($i = 1, \dots, m$) とおけば、

$$E[\langle \omega, f(x) \rangle] = E[\omega_i f^i(x)] = \omega_i E[f^i(x)] = \langle \omega, E[f(x)] \rangle \quad (3.33)$$

となる。

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

$$\text{Var}[\langle \omega, f(x) \rangle] = E[(\langle \omega, f(x) \rangle - \langle \omega, \alpha \rangle)^2] \quad (3.34)$$

$$= E[\langle \omega, f(x) - \alpha \rangle^2] \quad (3.35)$$

$$= E[\langle \omega \otimes \omega, (f(x) - \alpha)^2 \rangle] \quad (3.36)$$

$$= \langle \omega \otimes \omega, E[(f(x) - \alpha)^2] \rangle \quad (3.37)$$

$$= \langle \omega \otimes \omega, \text{Var}[f(x)] \rangle \quad (3.38)$$

となる。 □

定理 3.9 (分散の半正定値対称性). $f: \mathcal{X} \rightarrow V$ 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、 $\text{Var}_p[f] \in V \otimes V$ は対称かつ半正定値である。

証明 $\text{Var}[f] = E[(f - E[f])^2]$ 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。 $\text{Var}[f]$ が半正定値であることは、各 $\omega \in V^\vee$ に対し $\text{Var}[f](\omega, \omega) = \langle \omega \otimes \omega, \text{Var}[f] \rangle = \text{Var}[\langle \omega, f(x) \rangle] \geq 0$ 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 3.10 (分散が 0 であるための必要十分条件). 可測写像 $f: \mathcal{X} \rightarrow V$ 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) $\text{Var}_p[f] = 0$
- (2) f は p -a.e. 定数

証明には次の事実を用いる。

事実 3.11. \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) $\int_{\mathcal{Y}} g(y) \mu(dy) = 0$
- (2) $g(y) = 0$ μ -a.e. □

命題 3.10 の証明. V の基底 e_i ($i = 1, \dots, m$) をひとつ選んで固定し、 $f, E[f]$ の成分表示をそれぞれ $f^i: \mathcal{X} \rightarrow \mathbb{R}$ および $a^i \in \mathbb{R}$ ($i = 1, \dots, m$) とおいておく。

(2) \Rightarrow (1) f が a.e. 定数ならば、 $f^i(x) = a^i$ a.e. ($i = 1, \dots, m$) したがって $(f^i(x) - a^i)(f^j(x) - a^j) = 0$ a.e. ($i, j =$

$1, \dots, m$ である。よって $\int_X (f^i(x) - a^i)(f^j(x) - a^j) p(dx) = 0$ ($i, j = 1, \dots, m$) だから $\text{Var}[f] = 0$ である。

(1) \Rightarrow (2) $\text{Var}[f] = 0$ とすると、すべての $i = 1, \dots, m$ に対し $\int_X (f^i(x) - a^i)^2 p(dx) = 0$ が成り立つ。よって事実 3.11 より、すべての $i = 1, \dots, m$ 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。 \square

第2章 Hessian

1 Hessian

本節では W を m 次元 \mathbb{R} -ベクトル空間 ($m \in \mathbb{Z}_{\geq 0}$)、 $U \subset^{\text{open}} W$ 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 1.1 (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 $O \subset^{\text{open}} M$ 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 1.2 (U 上の平坦アファイン接続). U 上のアファイン接続 $D: \Gamma(TU) \rightarrow \Gamma(T^\vee U \otimes TU)$ 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、

$$DX := dX^i \otimes \frac{\partial}{\partial x^i} \in \Gamma(T^\vee U \otimes TU) \quad (1.1)$$

と定める。ただし、 X の成分表示を $X = X^i \frac{\partial}{\partial x^i}$ 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 (1.1) 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、

$$dX^i \otimes \frac{\partial}{\partial x^i} = d \left(X^\alpha \frac{\partial x^i}{\partial y^\alpha} \right) \otimes \frac{\partial y^\alpha}{\partial x^i} \frac{\partial}{\partial y^\alpha} \quad (1.2)$$

$$= \left(\frac{\partial x^i}{\partial y^\alpha} dX^\alpha + \underbrace{X^\alpha d \left(\frac{\partial x^i}{\partial y^\alpha} \right)}_{=0} \right) \otimes \frac{\partial y^\alpha}{\partial x^i} \frac{\partial}{\partial y^\alpha} \quad (1.3)$$

$$= dX^\alpha \otimes \frac{\partial}{\partial y^\alpha} \quad (1.4)$$

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ィン変換となることを用いた。これで well-defined 性も示された。 \square

B. Hessian

U 上のアフィン接続 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 1.3 (Hessian). C^∞ 関数 $f: U \rightarrow \mathbb{R}$ に対し、 f の **Hessian** を

$$\text{Hess } f := Ddf \in \Gamma(T^\vee U \otimes T^\vee U) \quad (1.5)$$

と定義する。

D -アフィン座標を用いると、Hessian の成分表示は簡単な形になる。

命題 1.4 (Hessian の成分表示). x^i ($i = 1, \dots, m$) を U 上の D -アフィン座標とする。このとき、座標 x^i に関する $\text{Hess } f$ の成分表示は

$$\text{Hess } f = \frac{\partial^2 f}{\partial x^i \partial x^j} dx^i \otimes dx^j \quad (1.6)$$

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

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

第3章 指数型分布族

1 指数型分布族

定義 1.1 (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) $T: \mathcal{X} \rightarrow V$ 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.

$$\frac{dp}{d\mu}(x) = \frac{\exp\langle \theta, T(x) \rangle}{\int_{\mathcal{X}} \exp\langle \theta, T(y) \rangle \mu(dy)} \quad \mu\text{-a.e. } x \in \mathcal{X} \quad (1.1)$$

である。ただし $\langle \cdot, \cdot \rangle$ は自然なペアリング $V^\vee \times V \rightarrow \mathbb{R}$ 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。
 - V の次元を (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 $\Theta_{(V, T, \mu)}$

$$\Theta_{(V, T, \mu)} := \left\{ \theta \in V^\vee \mid \int_{\mathcal{X}} \exp\langle \theta, T(y) \rangle \mu(dy) < +\infty \right\} \quad (1.2)$$

を (V, T, μ) の自然パラメータ空間 (natural parameter space) という。

- 関数 $\psi: \Theta_{(V, T, \mu)} \rightarrow \mathbb{R}$,

$$\psi(\theta) := \log \int_{\mathcal{X}} \exp\langle \theta, T(y) \rangle \mu(dy) \quad (1.3)$$

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

命題 1.2 (自然パラメータ空間は凸集合). $\Theta_{(T, \mu)}$ は \mathbb{R}^m の凸集合である。[TODO] V に修正

証明 表記の簡略化のため $\Theta := \Theta_{(T, \mu)}$ とおく。 $\theta, \theta' \in \Theta$, $t \in (0, 1)$ とし、 $(1-t)\theta + t\theta' \in \Theta$ を示せばよい。そこで $p := \frac{1}{1-t}$, $q := \frac{1}{t}$ とおくと、 $p, q \in (1, +\infty)$ であり、 $\frac{1}{p} + \frac{1}{q} = (1-t) + t = 1$ であり、 $e^{(1-t)\langle \theta, T(x) \rangle} \in L^p(\mathcal{X}, \mu)$ かつ $e^{t\langle \theta', T(x) \rangle} \in L^q(\mathcal{X}, \mu)$ だから、Hölder の不等式より

$$\int_{\mathcal{X}} e^{\langle (1-t)\theta + t\theta', T(x) \rangle} \mu(dx) = \int_{\mathcal{X}} e^{(1-t)\langle \theta, T(x) \rangle} e^{t\langle \theta', T(x) \rangle} \mu(dx) \quad (1.4)$$

$$\leq \left(\int_{\mathcal{X}} e^{(1-t)\langle \theta, T(x) \rangle p} \mu(dx) \right)^{1/p} \left(\int_{\mathcal{X}} e^{t\langle \theta', T(x) \rangle q} \mu(dx) \right)^{1/q} \quad (1.5)$$

$$= \left(\int_{\mathcal{X}} e^{\langle \theta, T(x) \rangle} \mu(dx) \right)^{1/p} \left(\int_{\mathcal{X}} e^{\langle \theta, T(x) \rangle} \mu(dx) \right)^{1/q} \quad (1.6)$$

$$< +\infty \quad (1.7)$$

が成り立つ。したがって $(1-t)\theta + t\theta' \in \Theta$ である。 \square

例 1.3 (有限集合上の確率分布). [TODO] V に修正 $\mathcal{X} = \{1, \dots, n\}$ 、 γ を \mathcal{X} 上の数え上げ測度とする。 \mathcal{X} 上の確率分布全体の集合 $\mathcal{P}(\mathcal{X})$ が \mathcal{X} 上の指数型分布族であることを確かめる。 δ^j ($j = 1, \dots, n$) を点 j での Dirac 測度とおく。任意の $P \in \mathcal{P}(\mathcal{X})$ に対し、

$$P(dk) := \sum_{j=1}^n a_j \delta^j(dk), \quad a_1, \dots, a_n \in \mathbb{R}_{>0}, \quad \sum_{j=1}^n a_j = 1 \quad (1.8)$$

が成り立つから、 δ_{jk} ($j, k = 1, \dots, n$) を Kronecker のデルタとして

$$P(dk) = \exp \left(\sum_{j=1}^n (\log a_j) \delta_{jk} \right) \gamma(dk) \quad (1.9)$$

$$= \exp \left(\sum_{j=1}^n \theta_j \delta_{jk} \right) \gamma(dk) \quad (1.10)$$

(ただし $\theta_j := \log a_j$) と表せる。したがって $T: \mathcal{X} \rightarrow \mathbb{R}^n$, $k \mapsto {}^t(\delta_{1k}, \dots, \delta_{nk})$ とおけば、 (T, γ) を実現として $\mathcal{P}(\mathcal{X})$ は指数型分布族となることがわかる。

例 1.4 (正規分布族). [TODO] V に修正 $\mathcal{X} = \mathbb{R}$ 、 λ を \mathbb{R} 上の Lebesgue 測度とする。 \mathcal{X} 上の確率分布の集合

$$\mathcal{P} := \left\{ P_{(\mu, \sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \mid \mu \in \mathbb{R}, \sigma^2 > 0 \right\} \quad (1.11)$$

を **正規分布族 (family of normal distributions)** という。このとき \mathcal{P} が \mathcal{X} 上の指数型分布族であることを確かめる。任意の $P_{(\mu, \sigma^2)} \in \mathcal{P}$ に対し

$$P_{(\mu, \sigma^2)}(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \lambda(dx) \quad (1.12)$$

$$= \exp \left(-\frac{1}{2\sigma^2} (x^2 - 2\mu x + \mu^2) - \frac{1}{2} \log 2\pi\sigma^2 \right) \lambda(dx) \quad (1.13)$$

$$= \exp \left(\left[\frac{\mu}{\sigma^2} \quad -\frac{1}{2\sigma^2} \right] \begin{bmatrix} x \\ x^2 \end{bmatrix} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log 2\pi\sigma^2 \right) \lambda(dx) \quad (1.14)$$

$$= \exp \left(\begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} x \\ x^2 \end{bmatrix} + \frac{\theta_1^2}{4\theta_2} - \frac{1}{2} \log \left(-\frac{\pi}{\theta_2} \right) \right) \lambda(dx) \quad (1.15)$$

(ただし $\theta_1 := \frac{\mu}{\sigma^2}$, $\theta_2 := -\frac{1}{2\sigma^2}$) が成り立つから、 $T: \mathcal{X} \rightarrow \mathbb{R}^2$, $x \mapsto {}^t(x, x^2)$ とおけば、 (T, λ) を実現として \mathcal{P} は指数型分布族となることがわかる。

例 1.5 (Poisson 分布族). [TODO] V に修正 $X = \mathbb{N}$, γ を \mathbb{N} 上の数え上げ測度とする。 X 上の確率分布の集合

$$\mathcal{P} := \left\{ P_\lambda(dk) = \frac{\lambda^k}{k!} e^{-\lambda} \gamma(dk) \mid \lambda > 0 \right\} \quad (1.16)$$

を P_λ を **Poisson 分布族 (family of Poisson distributions)** という。このとき \mathcal{P} が X 上の指数型分布族であることを確かめる。任意の $P_\lambda \in \mathcal{P}$ に対し

$$P_\lambda(dk) = \frac{\lambda^k}{k!} e^{-\lambda} \gamma(dk) \quad (1.17)$$

$$= \exp(k \log \lambda - \lambda) \frac{1}{k!} \gamma(dk) \quad (1.18)$$

$$= \exp\left(\theta k - e^\theta\right) \frac{1}{k!} \gamma(dk) \quad (1.19)$$

(ただし $\theta := \log \lambda$) が成り立つから、 $T: X \rightarrow \mathbb{R}, k \mapsto k$ とおけば、 $\left(T, \frac{1}{k!} \gamma(dk)\right)$ を実現として \mathcal{P} は指数型分布族となることがわかる。

定義 1.6 (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。

定理 1.7 (「 θ が一意の実現」の存在). [TODO] 「単射性条件」の言葉に修正 X を可測空間、 $\mathcal{P} \subset \mathcal{P}(X)$ を X 上の指数型分布族とする。このとき、 \mathcal{P} の「 θ が一意の実現」が存在する。

証明 (V, T, μ) は \mathcal{P} の実現のうちで次元が最小のものであるとする。 (V, T, μ) の次元 (m とおく) が 0 ならば V^\vee は 1 点集合だから証明は終わる。

以下 $m \geq 1$ の場合を考え、 (V, T, μ) が「 θ が一意の実現」であることを示す。背理法のために (V, T, μ) が「 θ が一意の実現」でないこと、すなわちある $p_0 \in \mathcal{P}$ および $\theta_0, \theta'_0 \in V^\vee$, $\theta_0 \neq \theta'_0$ が存在して

$$\exp(\langle \theta_0, T(x) \rangle - \psi(\theta_0)) = \frac{dp_0}{d\mu}(x) = \exp(\langle \theta'_0, T(x) \rangle - \psi(\theta'_0)) \quad \mu\text{-a.e. } x \in X \quad (1.20)$$

が成り立つことを仮定する。証明の方針としては、次元 $m-1$ の実現 (V', T', μ) を具体的に構成することにより、 (V, T, μ) の次元 m が最小であることとの矛盾を導く。

さて、式 (1.20) を整理して

$$\langle \theta_0 - \theta'_0, T(x) \rangle = \psi(\theta_0) - \psi(\theta'_0) \quad \mu\text{-a.e. } x \in X \quad (1.21)$$

を得る。表記の簡略化のために $\theta_1 := \theta_0 - \theta'_0 \in V^\vee$, $r := \psi(\theta_0) - \psi(\theta'_0) \in \mathbb{R}$ とおけば

$$\langle \theta_1, T(x) \rangle = r \quad \mu\text{-a.e. } x \in X \quad (1.22)$$

を得る。ここで $V' := (\mathbb{R}\theta)^\perp = \{v \in V \mid \langle \theta, v \rangle = 0\}$ とおき、次の claim を示す。

Claim ある可測写像 $T': X \rightarrow V'$ および $v_0 \in V$ が存在して $T(x) = T'(x) + v_0$ (μ -a.e. x) が成り立つ。

(\because) いま背理法の仮定より $\theta_1 \neq 0$ であるから、 θ_1 を延長した V^\vee の基底 $\theta_1, \dots, \theta_m$ が存在する。このとき、 $\theta_1, \dots, \theta_m$ を双対基底に持つ V の基底 v_1, \dots, v_m が存在する。この基底 v_1, \dots, v_m に関する

T の成分表示を $T(x) = \sum_{i=1}^m T^i(x)v_i$, $T^i: \mathcal{X} \rightarrow \mathbb{R}$ とおくと、(1.22) より $T^1(x) = \langle \theta_1, T(x) \rangle = r$ (μ -a.e. x) が成り立つ。そこで $v_0 := rv_1 \in V$ とおくと $\langle \theta_1, T(x) - v_0 \rangle = 0$ (μ -a.e. x) が成り立つから、可測写像 $T': \mathcal{X} \rightarrow V'$ を

$$T'(x) := \begin{cases} T(x) - v_0 & (\langle \theta_1, T(x) - v_0 \rangle = 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (1.23)$$

と定めることができる。この T, v_0 が求めるものである。 //

(V', T', μ) が \mathcal{P} の実現であることを示す。定義 1.1 の条件 (E0)-(E2) は明らかに成立しているから、あとは条件 (E3) を確認すればよい。そこで $p \in \mathcal{P}$ とする。いま (V, T, μ) が \mathcal{P} の実現であることより、ある $\theta \in V^\vee$ が存在して

$$\frac{dp}{d\mu}(x) = \frac{\exp(\langle \theta, T(x) \rangle)}{\int_{\mathcal{X}} \exp(\langle \theta, T(y) \rangle) \mu(dy)} \quad \mu\text{-a.e. } x \in \mathcal{X} \quad (1.24)$$

が成り立つ。 T', v_0 を用いて式変形すると、 μ -a.e. x に対し

$$\frac{dp}{d\mu}(x) = \frac{\exp(\langle \theta, T(x) \rangle)}{\int_{\mathcal{X}} \exp(\langle \theta, T(y) \rangle) \mu(dy)} \quad (1.25)$$

$$= \frac{\exp(\langle \theta, T'(x) \rangle + \langle \theta, v_0 \rangle)}{\int_{\mathcal{X}} \exp(\langle \theta, T'(y) \rangle + \langle \theta, v_0 \rangle) \mu(dy)} \quad (1.26)$$

$$= \frac{\exp(\langle \theta, T'(x) \rangle)}{\int_{\mathcal{X}} \exp(\langle \theta, T'(y) \rangle) \mu(dy)} \quad (1.27)$$

が成り立つ。したがって (V', T', μ) は条件 (E3) も満たし、 \mathcal{P} の実現であることがいえた。 (V', T', μ) は次元 $m-1$ だから (V, T, μ) の次元 m の最小性に矛盾する。背理法より (V, T, μ) は \mathcal{P} の「 θ が一意の実現」である。 \square

定理 1.8 (極小実現の性質). [TODO] V に修正 \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、 (T, μ) を \mathcal{P} の次元 m の実現とする。このとき、 (T, μ) が極小実現ならば、 $\langle u, T(x) \rangle$ が μ -a.e. 定数であるような $u \in \mathbb{R}^m$ は $u = 0$ のみである。

証明 (T, μ) を \mathcal{P} の極小実現とする。背理法のため、ある $u \neq 0$ が存在して $\langle u, T(x) \rangle$ が \mathcal{X} 上 μ -a.e. 定数であると仮定しておく。 $p \in \mathcal{P}$ とし、定義 1.1 の条件 (E3) の $\theta \in \mathbb{R}^m$ をひとつ選ぶと、

$$\frac{dp}{d\mu}(x) = \frac{e^{\langle \theta, T(x) \rangle}}{\int_{\mathcal{X}} e^{\langle \theta, T(y) \rangle} \mu(dy)} \quad (1.28)$$

$$= \frac{e^{\langle \theta, T(x) \rangle}}{\int_{\mathcal{X}} e^{\langle \theta, T(y) \rangle} \mu(dy)} \cdot \frac{e^{\langle u, T(x) \rangle}}{e^{\langle u, T(x) \rangle}} \quad (1.29)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_{\mathcal{X}} e^{\langle \theta, T(y) \rangle} e^{\langle u, T(y) \rangle} \mu(dy)} \quad (1.30)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_{\mathcal{X}} e^{\langle \theta, T(y) \rangle} e^{\langle u, T(y) \rangle} \mu(dy)} \quad (1.31)$$

$$= \frac{e^{\langle \theta+u, T(x) \rangle}}{\int_{\mathcal{X}} e^{\langle \theta+u, T(y) \rangle} \mu(dy)} \quad (1.32)$$

を得る。したがって $\theta + u$ も定義 1.1 の条件 (E3) を満たすが、いま $u \neq 0$ より $\theta + u \neq \theta$ だから、 (T, μ) が \mathcal{P} の極小実現であることに反する。背理法より定理が示された。 \square

例 1.9 (有限集合上の確率分布族). 例 1.3 の (T, γ) は $\mathcal{P}(X)$ の極小実現である。実際、任意の $P \in \mathcal{P}(X)$ に対し、 θ_j は $\theta_j = \log P(\{j\})$ ($j = 1, \dots, n$) として一意に決まる。

命題 1.10. [TODO] 上の命題とあわせる (V, T, μ) に関する次の条件は同値である:

- (1) $\langle \theta, T(x) \rangle$ が 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 1.1 の条件 (E3) をみたす $\theta \in V^\vee$ はただひとつである。

証明 $(2) \Rightarrow (1)$ 前回示した。

$(1) \Rightarrow (2)$ $\theta, \theta' \in V^\vee$ が定義 1.1 の条件 (E3) をみたすとする、

$$e^{\langle \theta, T(x) \rangle - \psi(\theta)} = \frac{dp}{d\mu}(x) = e^{\langle \theta', T(x) \rangle - \psi(\theta')} \quad \mu\text{-a.e. } x \in X \quad (1.33)$$

が成り立つ。式を整理して

$$\langle \theta - \theta', T(x) \rangle = \psi(\theta) - \psi(\theta') \quad \mu\text{-a.e. } x \in X \quad (1.34)$$

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

2 対数分配関数

[TODO] 一般化した命題を使って証明を修正する

本節では X を可測空間、 $\mathcal{P} \subset \mathcal{P}(X)$ を 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、 $\psi: \Theta \rightarrow \mathbb{R}$ 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 $h: X \times \Theta \rightarrow \mathbb{R}$ および $\lambda: \Theta \rightarrow \mathbb{R}$ を

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

$$\lambda(\theta) := \int_X h(x, \theta) \mu(dx) \quad (\theta \in \Theta) \quad (2.2)$$

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). $\varphi = (\theta_1, \dots, \theta_m): \Theta^\circ \rightarrow \mathbb{R}^m$ 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、

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

が成り立つ (∂_i は $\frac{\partial}{\partial \theta_i} \in \Gamma(T\Theta^\circ)$ 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、 $f: \mathcal{Y} \times I \rightarrow \mathbb{R}$ を

- (i) 各 $t \in I$ に対し $f(\cdot, t): \mathcal{Y} \rightarrow \mathbb{R}$ が可測
- (ii) 各 $y \in \mathcal{Y}$ に対し $f(y, \cdot): I \rightarrow \mathbb{R}$ 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 $\Phi: \mathcal{Y} \rightarrow \mathbb{R}$ が存在し、すべての $t' \in I$ に対し $\left| \frac{\partial f}{\partial t}(y, t') \right| \leq \Phi(y)$ a.e. y である。

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

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

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 ($i = 1, \dots, m$) を V^\vee の基底とし、この基底が定める Θ° 上のチャートを $\varphi = (\theta_1, \dots, \theta_m): \Theta^\circ \rightarrow \mathbb{R}^m$ とおく。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、次が成り立つ:

- (1) 任意の $\theta \in \Theta^\circ$ に対し、関数 $\partial_{i_k} \cdots \partial_{i_1} h(\cdot, \theta): X \rightarrow \mathbb{R}$ は $L^1(X, \mu)$ 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 $\Phi: X \rightarrow \mathbb{R}$ が存在し、すべての $\theta' \in U$ に対し $|\partial_{i_k} \cdots \partial_{i_1} h(x, \theta')| \leq \Phi(x)$ 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 $\theta_1, \dots, \theta_m$ 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

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

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

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

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

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

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

$t' := (t'_1, \dots, t'_m) := \varphi(\theta') \in \mathbb{R}^m$ とおいておく。まず $\frac{k+1}{\varepsilon} \frac{\varepsilon}{k+1} = 1$ より

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

であり、 \prod の部分を評価すると

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

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

(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

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

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

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 $i = 1, \dots, m$ 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

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

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

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

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

を得る。この右辺は (t' によらないから) θ' によらない 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。 \square

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 $\varphi = (\theta_1, \dots, \theta_m)$ 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 $\tilde{\varphi} = (\tilde{\theta}_1, \dots, \tilde{\theta}_m)$ 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 $\tilde{\varphi}$ 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 $\tilde{\theta}_1, \dots, \tilde{\theta}_m$ に関する k 回偏導関数が、 λ の $\theta_1, \dots, \theta_m$ 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、 $t = (t_1, \dots, t_m) \in \varphi(\Theta^\circ)$ 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

$$g: X \times I \rightarrow \mathbb{R}, \quad (x, s) \mapsto h(x, \varphi^{-1}(s, t_2, \dots, t_m)) \quad (2.16)$$

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 $\Phi: X \rightarrow \mathbb{R}$ 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

$$I \times \{t_2\} \times \dots \times \{t_m\} \subset \varphi(U) \quad (2.17)$$

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 $i_k = 2, \dots, m$ 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。 \square

定理 2.1 から次の系が従う。

系 2.4. $\varphi = (\varphi_1, \dots, \varphi_m): \Theta^\circ \rightarrow \mathbb{R}^m$ 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、 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、

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

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。 \square

3 Fisher 計量

定義 3.1 (条件 A). [TODO] 単射性条件の言葉に修正 \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 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 $(\text{Hess } \psi)_\theta \in T_\theta^{(0,2)}\Theta^\circ$ は $\text{Var}_{P_\theta}[T]$ に一致する。さらに (V, T, μ) が条件 A をみたすならば、 $\text{Hess } \psi$ 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、 $\text{Hess } \psi$ 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 $(\text{Hess } \psi)_\theta = \text{Var}_{P_\theta}[T]$ ($\theta \in \Theta^\circ$) を示す。 Θ° 上の D -アファイン座標 θ^i ($i = 1, \dots, m$) をひとつ選ぶと、命題 1.4 より、座標 θ^i に関する $\text{Hess } \psi$ の成分表示は $\text{Hess } \psi = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j} d\theta^i \otimes d\theta^j$ となる。ここで前回 (0516_資料.pdf) の系 2.4 より

$$\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}(\theta) = \partial_i \partial_j \log \lambda(\theta) \quad (3.1)$$

$$= \partial_i \left(\frac{\partial_j \lambda(\theta)}{\lambda(\theta)} \right) \quad (3.2)$$

$$= \frac{\partial_i \partial_j \lambda(\theta)}{\lambda(\theta)} - \frac{\partial_i \lambda(\theta) \partial_j \lambda(\theta)}{\lambda(\theta)^2} \quad (3.3)$$

$$= E[T^i(x)T^j(x)] - E[T^i(x)]E[T^j(x)] \quad (3.4)$$

$$= E[(T^i(x) - E[T^i(x)])(T^j(x) - E[T^j(x)])] \quad (3.5)$$

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 $\text{Hess}_\theta \psi = \text{Var}_{P_\theta}[T]$ が成り立つ。

次に、 (V, T, μ) が条件 A をみたすとし、 $\text{Hess } \psi$ が正定値であることを示す。すなわち、各 $\theta \in \Theta^\circ$ に対し $(\text{Hess } \psi)_\theta$ 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「 $(\text{Hess } \psi)_\theta(u, u) = 0$ ならば $u = 0$ 」を示せばよいが、上で示したことと命題 3.8 より

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

と式変形できるから、 $(\text{Hess } \psi)_\theta(u, u) = 0$ ならば命題 3.10 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 $(\text{Hess } \psi)_\theta$ は正定値である。したがって $\text{Hess } \psi$ は正定値である。 \square

参考文献

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

記号一覧

(V, T, μ) 指数型分布族. 11

索引

F

Fisher 計量 19

H

Hessian 10

P

Poisson 分布族 13

R

Radon-Nikodým 微分 2

ア

アフライン座標 9

カ

確率空間 2

確率測度 2

確率分布 2

 確率変数の— 3

確率分布に従う 3

確率変数 3

確率密度関数 3

可積分

 ベクトル値関数の— 3

キ

期待値 4

基底測度 11

サ

最小次元実現 13

シ

次元 11

指数型分布族 11

自然パラメータ空間 11

実現 11

十分統計量 11

セ

正規分布族 12

積分 3

絶対連続 2

タ

対数分配関数 11

フ

分散 5

ヘ

平坦 9