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). $(\mathcal{X}, \mathcal{B})$ を可測空間、μ, ν を $\mathcal{X}$ 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

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

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、$\dfrac{d\nu}{d\mu}$ 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). $(\mathcal{X}, \mathcal{B})$ を可測空間、μ を $\mathcal{X}$ 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

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

が成り立つ。

証明　Young の不等式を使う。 □

2　確率論の基本事項

2.1　確率空間

定義 2.1 (確率空間). 測度空間 $(\Omega, \mathcal{F}, P)$ 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 $(\Omega, \mathcal{F})$ 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). $(\Omega, \mathcal{F}, P)$ を確率空間、$(\mathcal{X}, \mathcal{A})$ を可測空間とする。可測関数 $X\colon (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ を $(\mathcal{X}, \mathcal{A})$ 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). $(\Omega, \mathcal{F}, P)$ を確率空間、$X\colon (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ を確率変数とする。このとき、写像

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

は $(\mathcal{X}, \mathcal{A})$ 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 $(\mathcal{X}, \mathcal{A})$ 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). $(\mathcal{X}, \mathcal{A})$ を可測空間、μ を $\mathcal{X}$ 上の σ-有限測度、ν を μ に関し絶対連続な $(\mathcal{X}, \mathcal{A})$ 上の確率測度とする。このとき、ν の μ に関する Radon-NIkodým 微分 $\dfrac{d\nu}{d\mu}$ 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). $\mathcal{X}$ 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

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

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 $\displaystyle\int_{\mathcal{X}} f(x)\, p(dx) = 0$ 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、$e^1, \ldots, e^m$ および $\tilde{e}^1, \ldots, \tilde{e}^m$ 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, $\tilde{f}_i\colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \ldots, m)$ とおく。示すべきことは「$\tilde{f}_i$ $(i = 1, \ldots, m)$ がすべて $L^1(\mathcal{X}, p)$ に属するならば f_i $(i = 1, \ldots, m)$ 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 $a_i^j \in \mathbb{R}$ $(i, j = 1, \ldots, m)$ が存在して $f_i = a_i^j \tilde{f}_j$ $(i = 1, \ldots, m)$ および $\tilde{e}^j = a_i^j e^i$ $(j = 1, \ldots, m)$ 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。 □

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。

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

(1) f および $(f - E_p[f])^2$ が p-可積分
(2) f^2 が p-可積分

さらに V にノルム $\|\cdot\|$ が定義されているとき、次も同値である:

(3) $\|f\| \in L^2(\mathcal{X}, p)$

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。 □

上の事実を用いて補題を示す。

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 $e^1, \ldots, e^m$ をひとつ選んで固定し、この基底に関する f の成分を $f_i\colon \mathcal{X} \to \mathbb{R}$ $(i = 1, \ldots, m)$ 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 $a \coloneqq E_p[f]$ 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。

(2) ⇔(3) 有限次元ベクトル空間のノルムの同値性より、固定した基底に関する成分を用いた 2-ノルムを考えればよい。 □

この補題を踏まえて分散を定義する。

定義 3.6 (分散)**.** $f^2 \colon \mathcal{X} \to V \otimes_{\mathbb{R}} V$ が p-可積分であるとき、f の p に関する**分散 (variance)** $V_p[f]$ を

$$V_p[f] := E_p[(f - E_p[f])^2] \in V \otimes V \tag{3.6}$$

と定義する (補題 3.4 よりこれは存在する)。

例 3.7 (期待値と分散の例: 正規分布族の十分統計量)**.** $\mathcal{X} = \mathbb{R}$、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), $V = \mathbb{R}^2$, $T \colon \mathcal{X} \to V$, $x \mapsto {}^t(x, x^2)$ 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

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

$$
\begin{aligned}
&+ E_P[(x-\mu)(x^2-(\sigma^2+\mu^2))]\,(e_1\otimes e_2+e_2\otimes e_1) && (3.12)\\
&+ E_P[(x^2-(\sigma^2+\mu^2))^2]\,e_2\otimes e_2 && (3.13)
\end{aligned}
$$

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

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

と求まる。したがって $V_P[T]$ は

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

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

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

および

$$
\begin{aligned}
3\sigma^4 &= E_P[(x-\mu)^4] && (3.25)\\
&= E_P[x^4]-4E_P[x^3]\mu+6E_P[x^2]\mu^2-4E_P[x]\mu^3+\mu^4 && (3.26)\\
&= a_4-4a_3\mu+6(\sigma^2+\mu^2)\mu^2-4\mu^4+\mu^4 && (3.27)\\
&= a_4-6\sigma^2\mu^2-\mu^4 && (3.28)\\
\therefore\ a_4 &= 3\sigma^4+6\sigma^2\mu^2+\mu^4 && (3.29)
\end{aligned}
$$

を得る。これらを $V_P[T]$ の成分表示に代入して

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

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). [TODO] $L(E_p[f]) = E_p[L \circ f]$ の形の主張に修正する

$f\colon \mathcal{X} \to V$ 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 $f^i\colon \mathcal{X} \to \mathbb{R}$, $\omega_i \in \mathbb{R}$ $(i = 1, \ldots, m)$ とおけば、

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

となる。

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

$$\begin{aligned}
\mathrm{Var}[\langle \omega, f(x) \rangle] &= E[(\langle \omega, f(x) \rangle - \langle \omega, \alpha \rangle)^2] &(3.34)\\
&= E[\langle \omega, f(x) - \alpha \rangle^2] &(3.35)\\
&= E[\langle \omega \otimes \omega, (f(x) - \alpha)^2 \rangle] &(3.36)\\
&= \langle \omega \otimes \omega, E[(f(x) - \alpha)^2] \rangle &(3.37)\\
&= \langle \omega \otimes \omega, \mathrm{Var}[f(x)] \rangle &(3.38)
\end{aligned}$$

となる。 □

定理 3.9 (分散の半正定値対称性). $f\colon \mathcal{X} \to V$ 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 3.10 (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 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) $\displaystyle\int_{\mathcal{Y}} g(y)\,\mu(dy) = 0$

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

□

上の事実を用いて命題を示す。

命題 3.10 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 $\displaystyle\int_X (f^i(x) - a^i)(f^j(x) - a^j)\,p(dx) = 0$ $(i, j = 1, \ldots, m)$ 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 $\displaystyle\int_X (f^i(x) - a^i)^2\,p(dx) = 0$ が成り立つ。よって事実 3.11 より、すべての $i = 1, \ldots, m$ 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。 □