

確率統計学I

杉浦 誠

平成 22 年 7 月 28 日

目次

1	確率空間	3
1.1	集合	3
1.2	確率の定義	3
1.3	条件付確率と事象の独立	5
2	確率変数	7
2.1	確率変数の定義	7
2.2	可測関数と確率変数	8
2.3	分布関数	8
2.4	多次元確率変数	13
2.5	条件付き確率分布	15
2.6	確率変数の独立性	17
3	確率変数の変換	19
3.1	はじめに	19
3.2	離散型確率変数の変換	19
3.3	絶対連続型確率変数の変換	19
4	期待値	21
4.1	Lebesgue 積分	21
4.2	期待値の定義	23
4.3	積率 (モーメント)・分散	24
4.4	共分散と相関係数	27

これは

- 柳川 堯 統計数学 近代科学社
を教科書として作った講義ノートです。

1 確率空間

1.1 集合

各自読んでおくよう伝える。

1.2 確率の定義

定義 1.1 集合 Ω の部分集合からなる集合を Ω 上の集合族という。 $\Omega (\neq \emptyset)$ 上の集合族 \mathfrak{B} が次の条件

- (i) $\Omega \in \mathfrak{B}$
- (ii) $A \in \mathfrak{B} \implies A^c \in \mathfrak{B}$ (A^c は A の補集合を表す。)
- (iii) $A_1, A_2, \dots, A_n, \dots \in \mathfrak{B} \implies \bigcup_{n=1}^{\infty} A_n \in \mathfrak{B}$

を満たすとき、この \mathfrak{B} を Ω 上の σ 集合族という。また、 (Ω, \mathfrak{B}) を可測空間という。

確率論では、 σ 集合族 \mathfrak{B} の元を事象という。特に、 Ω を全事象と、 $\emptyset = \Omega^c$ を空事象という。

定理 1.1 (Ω, \mathfrak{B}) を可測空間、 $\{B_n\} \subset \mathfrak{B}$ とすると次が成り立つ。

- (0) $\emptyset \in \mathfrak{B}$,
- (1) $\bigcup_{i=1}^n B_i \in \mathfrak{B}$,
- (2) $\bigcap_{i=1}^n B_i \in \mathfrak{B}$,
- (3) $\bigcap_{n=1}^{\infty} B_n \in \mathfrak{B}$,
- (4) $B_1 \setminus B_2 := B_1 \cap B_2^c \in \mathfrak{B}$

証明: (0) 定義 1.1 (i), (ii) より $\emptyset = \Omega^c \in \mathfrak{B}$.

(1) $A_i = B_i (1 \leq i \leq n)$, $A_i = \emptyset (i \geq n+1)$ とおくと、 $\forall i$ に対し $A_i \in \mathfrak{B}$ であるから、定義 1.1 (iii) より $\bigcup_{i=1}^n B_i = \bigcup_{i=1}^{\infty} A_i \in \mathfrak{B}$ を得る。

(2) 定義 1.1 (ii) より $B_i^c \in \mathfrak{B}$ であるから、(1) と de Morgan の法則により $\bigcap_{i=1}^n B_i = \left(\bigcup_{i=1}^n B_i^c\right)^c \in \mathfrak{B}$.

(3) 定義 1.1 (iii) と de Morgan の法則により (2) と全く同様に証明できる。

(4) $A_1 = B_1, A_2 = B_2^c, n = 2$ として (2) を用いればよい。 \square

定義 1.2 (Ω, \mathfrak{B}) を可測空間とする。次の条件を満たす $P: \mathfrak{B} \rightarrow [0, 1]$ を (Ω, \mathfrak{B}) 上の確率測度という。

- (i) $P(\Omega) = 1$
- (ii) $A_n \in \mathfrak{B}, n = 1, 2, \dots$, が任意の $m \neq n$ なる組に対し $A_m \cap A_n = \emptyset$ を満たせば (このとき A_1, A_2, \dots は互いに排反であるという), $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

このとき、 $(\Omega, \mathfrak{B}, P)$ を確率空間という。また、 \mathfrak{B} の元を事象、 $A \in \mathfrak{B}$ に対し $P(A)$ を事象 A の確率という。

例 1.1 (古典的確率模型)

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\} \quad (\text{有限集合})$$

$$\mathfrak{B} = 2^\Omega \quad (\Omega \text{ のべき集合; } \Omega \text{ の部分集合すべてからなる集合族)}$$

$$P(A) = \frac{\#A}{\#\Omega}, A \in \mathfrak{B} \quad (\#A \text{ は集合 } A \text{ の元の個数})$$

この確率空間 $(\Omega, \mathfrak{B}, P)$ を古典的確率模型という。以下、その例をいくつか例示する。

a) サイコロを 1 回投じる: $\Omega = \{1, 2, 3, 4, 5, 6\}, P(\{\omega\}) = 1/6 (\forall \omega \in \Omega)$.

例えば、 $P(\text{奇数の目が出る}) = P(\{1, 3, 5\}) = \frac{\#\{1, 3, 5\}}{\#\Omega} = \frac{3}{6} = \frac{1}{2}$.

b) コインを 2 回投げる: $\Omega = \{HH, HT, TH, TT\}, P(\{\omega\}) = 1/4 (\forall \omega \in \Omega)$.

(H は表 (head), T は裏 (tail) を意味する。)

c) 箱の中に黒球 a 個と白球 b 個が入っている。この箱から球を 1 個ずつ取り出す復元抽出を n 回行う。この場合 $\Omega = \{B_1, \dots, B_a, W_1, \dots, W_b\}^n$ と考える。 $\#\Omega = (a+b)^n$ に注意する。

A で黒球がちょうど k 回出る事象とする。このとき、 $\#A = \binom{n}{k} a^k b^{n-k}$ であるから、 $P(A) = \binom{n}{k} \frac{a^k b^{n-k}}{(a+b)^n}$ 。

d) c) の例で非復元抽出 (取り出した球を戻さない) を n 回行う。このとき、 A で黒球が k 回出る事象とすると、 $\#A = \binom{n}{k} a P_k \cdot b P_{n-k}$ であるから、 $P(A) = \binom{n}{k} \frac{a P_k \cdot b P_{n-k}}{a+b P_n}$ 。

定理 1.2 $(\Omega, \mathfrak{B}, P)$ を確率空間とする。

(1) $P(\emptyset) = 0$

(2) $B_i \in \mathfrak{B}, i = 1, \dots, n,$ が互いに排反であれば、 $P\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n P(B_i)$ 。

(3) $A, B \in \mathfrak{B}, A \subset B \implies P(A) \leq P(B)$

(4) $B \in \mathfrak{B} \implies P(B^c) = 1 - P(B)$

(5) $A, B \in \mathfrak{B}$ に対し $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 。

証明: (1) 定理 1.1 に注意して、 $a = P(\emptyset)$ とおくと、 $0 \leq a \leq 1$ 。ここで、 $A_n = \emptyset (\forall n)$ とおくと、 $\bigcup_{n=1}^{\infty} A_n = \emptyset$ で、 $A_m \cap A_n = \emptyset (m \neq n)$ であるから定義 1.2 (ii) より $P(\emptyset) = \sum_{n=1}^{\infty} P(\emptyset)$ を得る。もし $a > 0$ であれば、これは $a = \infty$ を意味し、 $0 \leq a \leq 1$ に矛盾する。従って、 $P(\emptyset) = 0$ となる。

(2) $A_i = B_i (1 \leq i \leq n), A_i = \emptyset (i \geq n+1)$ とおくと、 $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^n B_i$ で、 $A_i \cap A_j = \emptyset (i \neq j)$ であるから、定義 1.2 (ii) と (1) より、 $P(\bigcup_{i=1}^n B_i) = \sum_{i=1}^n P(B_i) + \sum_{i=n+1}^{\infty} P(\emptyset) = \sum_{i=1}^n P(B_i)$ 。

(3) $B_1 = A, B_2 = B \setminus A$ とすると $B = B_1 \cup B_2, B_1 \cap B_2 = \emptyset$ より (2) から、 $P(B) = P(B_1) + P(B_2) \geq P(A)$ 。

(4) $B \cup B^c = \Omega, B \cap B^c = \emptyset$ より、定義 1.2 (i) と (2) により、 $P(B) + P(B^c) = P(\Omega) = 1$ となり従う。

(5) $B_1 = A \cap B, B_2 = A \cap B^c, B_3 = A^c \cap B$ とおくと、 B_1, B_2, B_3 は互いに排反であるから、(2) により $A = B_1 \cup B_2$ から $P(A) = P(B_1) + P(B_2)$ 、 $B = B_1 \cup B_3$ から $P(B) = P(B_1) + P(B_3)$ 、 $A \cup B = B_1 \cup B_2 \cup B_3$ から $P(A \cup B) = P(B_1) + P(B_2) + P(B_3)$ 。よって、与式は従う。 \square

定理 1.3 $A_1, A_2, \dots \in \mathfrak{B}$ に対し $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$ 。

証明: $B_1 = A_1, B_2 = A_2 \setminus A_1, \dots, B_n = A_n \setminus (\bigcup_{k=1}^{n-1} A_k)$ とおくと、 B_1, B_2, \dots は互いに排反で $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ 。よって、定義 1.2 (ii) により、

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)。$$

ここで、最後の不等式は $A_n \subset B_n$ と定理 1.2(3) を用いた。 \square

定理 1.4 $(\Omega, \mathfrak{B}, P)$ を確率空間、 $\{B_n\} \subset \mathfrak{B}$ とする。

(1) $B_1 \subset B_2 \subset \dots \subset B_n \subset \dots \implies P(\bigcup_{n=1}^{\infty} B_n) = \lim_{n \rightarrow \infty} P(B_n)$ 。

(2) $B_1 \supset B_2 \supset \dots \supset B_n \supset \dots \implies P(\bigcap_{n=1}^{\infty} B_n) = \lim_{n \rightarrow \infty} P(B_n)$ 。

証明: (1) $A_1 = B_1, A_n = B_n \setminus \bigcup_{k=1}^{n-1} B_k (n \geq 2)$ とおくと、 A_1, A_2, \dots は互いに排反で $\bigcup_{k=1}^n A_k = B_n (n \in \mathbb{N}), \bigcup_{k=1}^{\infty} A_k = \bigcup_{n=1}^{\infty} B_n$ 。よって、定義 1.2 (ii)、定理 1.2 (2) により、

$$P\left(\bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_k) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=1}^n A_k\right) = \lim_{n \rightarrow \infty} P(B_n)。$$

(2) $B_1^c \subset \dots \subset B_n^c \subset \dots$ であるから、定理 1.2 (4)、de Morgan の法則と前半より

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = 1 - P\left(\left(\bigcap_{n=1}^{\infty} B_n\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} B_n^c\right) = 1 - \lim_{n \rightarrow \infty} P(B_n^c) = \lim_{n \rightarrow \infty} (1 - P(B_n^c)) = \lim_{n \rightarrow \infty} P(B_n)。$$
 \square

問 1.1 (教科書 p.9 演習問題 A6) Ω を集合とし \mathcal{A} をその部分集合族とする。 \mathcal{A} を含む Ω 上の σ 集合族はいくつもある。これに *index* を付け、 \mathfrak{B}_λ ($\lambda \in \Lambda$) と表す。このとき、次が成り立つ。

(1) $\bigcap_{\lambda \in \Lambda} \mathfrak{B}_\lambda$ は σ 集合族である。

(2) $\bigcap_{\lambda \in \Lambda} \mathfrak{B}_\lambda$ は \mathcal{A} を含む σ 集合族の中で最小の σ 集合族である。

証明: (1) は演習問題とする。(2) まず、 \mathcal{A} を含む Ω 上の σ 集合族として、 Ω のべき集合があることに注意する (よって Λ は空ではない)。(1) より、 $\mathcal{F} := \bigcap_{\lambda \in \Lambda} \mathfrak{B}_\lambda$ が \mathcal{A} を含む σ 集合族の中で最小であることを示せばよい。今、 \mathcal{F} より小さい (真部分集合となる)、 \mathcal{A} を含む σ 集合族 \mathfrak{B}_0 があったとする。 \mathfrak{B}_0 は \mathcal{A} を含む σ 集合族なので、 \mathfrak{B}_λ のいずれかである。従って、 $\mathfrak{B}_0 \supset \bigcap_{\lambda \in \Lambda} \mathfrak{B}_\lambda = \mathcal{F}$ であるが、これは \mathfrak{B}_0 が \mathcal{F} の真部分集合であることに矛盾する。 \square

1.3 条件付確率と事象の独立

確率空間 $(\Omega, \mathfrak{B}, P)$ において $C \in \mathfrak{B}$, $P(C) > 0$ のとき、次のように定める。

$$P(A|C) := \frac{P(A \cap C)}{P(C)} \quad (A \in \mathfrak{B})$$

定理 1.5 $P(\cdot | C)$ は可測空間 (Ω, \mathfrak{B}) 上の確率測度である。

証明: 定理 1.2 (3) より $0 \leq P(A \cap C) \leq P(C)$ となり、 $0 \leq P(A|C) \leq 1$ 。定義 1.2 (i) は明らか。(ii) は $\{A_n\} \subset \mathfrak{B}$ が互いに排反であれば、 $B_n = A_n \cap C$ とすると $B_n \in \mathfrak{B}$ で $B_m \cap B_n = A_m \cap A_n \cap C = \emptyset$ ($m \neq n$) より、

$$P\left(\bigcup_{n=1}^{\infty} A_n | C\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap C\right)}{P(C)} = \frac{P\left(\bigcup_{n=1}^{\infty} B_n\right)}{P(C)} = \frac{\sum_{n=1}^{\infty} P(B_n)}{P(C)} = \sum_{n=1}^{\infty} \frac{P(A_n \cap C)}{P(C)} = \sum_{n=1}^{\infty} P(A_n | C). \quad \square$$

定義 1.3 $P(A|C)$ を事象 C が与えられたときの $A \in \mathfrak{B}$ の条件付確率という。また、 $P(\cdot | C)$ を C が与えられたときの条件付確率測度という。

例 1.2 N 人のクラスで N_A 人が近視で、そのうち女子が N_{AB} 人であった。このクラスの女子は N_B 人であるが、無作為に女子を一人選び、その学生が近視である確率を考える。

A をクラスの一人を選び近視である事象、 B をクラスの一人を選び女子である事象とする。このとき、求める確率は $P(A|B)$ とあらわされる。ここで、 $P(A \cap B) = N_{AB}/N$, $P(B) = N_B/N$ であるから、

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{N_{AB}/N}{N_B/N} = \frac{N_{AB}}{N_B}.$$

定理 1.6 次が成り立つ。

(1) (乗法公式) $P(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$, $A_i \in \mathfrak{B}$ ($i = 1, 2, \dots, n$) のとき

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1})$$

(2) (全確率の公式) $\{A_n\} \subset \mathfrak{B}$ を Ω の分割、即ち、 $\{A_n\}$ は互いに排反で $P(A_n) > 0$ ($n \in \mathbb{N}$) かつ $\Omega = \bigcup_{n=1}^{\infty} A_n$ を満たすとする。このとき、

$$P(A) = \sum_{i=1}^{\infty} P(A_i)P(A|A_i) \quad (A \in \mathfrak{B})$$

(3) (Bayes の定理) $\{A_n\} \subset \mathfrak{B}$ を Ω の分割とする。このとき、 $B \in \mathfrak{B}$ で $P(B) > 0$ なるものに対して次が成立する。

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{n=1}^{\infty} P(A_n)P(B|A_n)}.$$

注意 1.1 定理 1.6 (2), (3) は Ω が有限個の $A_1, \dots, A_N \in \mathfrak{B}$ によって分割されている場合にも成立する。証明は、定義 1.2 (ii) のかわりに定理 1.2 (2) を用いることで、全く同様にできる。

証明: (1) $P(A \cap C) = P(A|C)P(C)$ に注意する。これを繰り返し用いて、

(右辺) $= P(A_1 \cap \dots \cap A_{n-1})P(A_n|A_1 \cap \dots \cap A_{n-1}) = P(A_1 \cap \dots \cap A_{n-2})P(A_{n-1}|A_1 \cap \dots \cap A_{n-2})P(A_n|A_1 \cap \dots \cap A_{n-1}) = \dots =$ (左辺)。

(2) (右辺) $= \sum_{n=1}^{\infty} P(A_n) \frac{P(B \cap A_n)}{P(A_n)} = P\left(\bigcup_{n=1}^{\infty} (B \cap A_n)\right) = P\left(B \cap \bigcup_{n=1}^{\infty} A_n\right) = P(B \cap \Omega) = P(B)$ 。

ここで、第 2 の等号は $m \neq n$ であれば $(B \cap A_m) \cap (B \cap A_n) = B \cap A_m \cap A_n = \emptyset$ より $\{B \cap A_n\}$ は互いに排反であることに注意し定義 1.2 (ii) を適用した。

(3) (右辺) $= \frac{P(A_k) \frac{P(B \cap A_k)}{P(A_k)}}{P(B)} = \frac{P(B \cap A_k)}{P(B)} = P(A_k|B)$ 。ここで、第 1 の等号は (2) を用いた。 \square

例題 1.1 ある病気の判定薬があり、病気の人に対して 98%の確率で陽性反応 (病気であると判定) を示し、健康者に対して 2%の確率で陽性反応を示す。この病気の罹病率は 1%であるとする。ある人がこの判定薬で陽性反応がでたとき、この人が本当にその病気に罹っている確率を求めよ。

解: A をその病気に罹っている事象、 B を陽性反応を示す事象とすると、病気の人に対して 98%の確率で陽性反応を示すから $P(B|A) = 0.98$ 。健康者に対して 2%の確率で陽性反応から $P(B|A^c) = 0.02$ 。また、この病気の罹病率は 1%より $P(A) = 0.01$ 。求める確率は $P(A|B)$ である。以上の式より、

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{0.01 \cdot 0.98}{0.01 \cdot 0.98 + 0.99 \cdot 0.02} = \frac{49}{49 + 99} = \frac{49}{148} (= 0.331\dots) \quad \square$$

定義 1.4 (1) $P(A \cap B) = P(A)P(B)$ ($A, B \in \mathfrak{B}$) のとき、事象 A と B は独立であるといい、 $A \perp B$ と表わす。

(2) 事象の列 $\{B_n\} \subset \mathfrak{B}$ (有限個でも無限個でもよい) に対し、その任意の有限部分列 $B_{n_1}, B_{n_2}, \dots, B_{n_l}$ が、

$$P(B_{n_1} \cap B_{n_2} \cap \dots \cap B_{n_l}) = P(B_{n_1})P(B_{n_2}) \cdots P(B_{n_l}) \quad (1.1)$$

を満たすとき、事象の列 $\{B_n\}$ は独立であるという。

定理 1.7 (1) $A, B \in \mathfrak{B}$ とし $P(B) > 0$ のとき、 $A \perp B \iff P(A|B) = P(A)$ 。

(2) 事象の列 B_1, B_2, \dots, B_n が独立であれば、任意の $k = 1, \dots, n$ に対して、 $B_1^c, \dots, B_k^c, B_{k+1}, \dots, B_n$ も独立となる。

証明: (1) 独立の定義から明らか。(2) B_1, B_2, \dots, B_n が独立ならば、 B_1^c, B_2, \dots, B_n も独立となることを示せばよい。実際、これより $B_2, B_3, \dots, B_n, B_1^c$ が独立となるから、 $B_2^c, B_3, \dots, B_n, B_1^c$ は独立となり、これを繰り返すことで主張を得る。

このため任意の $1 \leq n_1 < n_2 < \dots < n_l \leq n$ に対し (1.1) に相当する式を示せばよが、 $n_1 \geq 2$ の場合は B_1^c が関与しないため (1.1) は明らかに成り立つ。 $n_1 = 1$ の場合:

$$P(B_1^c \cap B_{n_2} \cap \dots \cap B_{n_l}) = P(B_1^c)P(B_{n_2}) \cdots P(B_{n_l}) \quad (1.2)$$

を示せばよい。まず、

$$P(B_1 \cap B_{n_2} \cap \dots \cap B_{n_l}) + P(B_1^c \cap B_{n_2} \cap \dots \cap B_{n_l}) = P(B_{n_2} \cap \dots \cap B_{n_l})$$

に注意する。ここで、 B_1, B_2, \dots, B_n の独立性を用いると

$$P(B_1 \cap B_{n_2} \cap \dots \cap B_{n_l}) = P(B_1)P(B_{n_2}) \cdots P(B_{n_l}), \quad P(B_{n_2} \cap \dots \cap B_{n_l}) = P(B_{n_2}) \cdots P(B_{n_l})$$

となる。よって、

$$P(B_1^c \cap B_{n_2} \cap \dots \cap B_{n_l}) = \{1 - P(B_1)\}P(B_{n_2}) \cdots P(B_{n_l}) = P(B_1^c)P(B_{n_2}) \cdots P(B_{n_l}). \quad \square$$

注意 1.2 (1) 事象 A, B, C に対し、2つの事象の組すべてが独立であっても、 A, B, C が独立であるとは限らない。

実際、 $\Omega = \{1, 2, 3, 4\}$, $P(\{k\}) = 1/4$ ($k = 1, 2, 3, 4$) において、 $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{2, 3\}$ とおくと、 $P(A) = P(B) = P(C) = 1/2$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$ となるから、どの2つの事象も独立となる。

一方、 $A \cap B \cap C = \emptyset$ であるから、 $0 = P(A \cap B \cap C) \neq P(A)P(B)P(C) = 1/8$ となり、 A, B, C は独立とならない。 □

(2) $P(A \cap B \cap C) = P(A)P(B)P(C)$ であっても、 $A \perp B$, $B \perp C$, $C \perp A$ は必ずしも成立しない。

実際、 $\Omega = \{1, 2, 3, 4\}$, $P(\{1\}) = \frac{1}{\sqrt{2}} - \frac{1}{4}$, $P(\{2\}) = \frac{3}{4} - \frac{1}{\sqrt{2}}$, $P(\{3\}) = P(\{4\}) = 1/4$ において、 $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{2, 4\}$ とおくと、 $P(A \cap B \cap C) = P(\{2\}) = \frac{3}{4} - \frac{1}{\sqrt{2}} = P(A)P(B)P(C)$ となるが、 $P(A \cap B) = P(\{2\}) = \frac{3}{4} - \frac{1}{\sqrt{2}}$, $P(A)P(B) = \frac{1}{2} \left(1 - \frac{1}{\sqrt{2}}\right)$ となる。 $P(B \cap C) \neq P(B)P(C)$, $P(A \cap C) \neq P(A)P(C)$ も同様。 □

2 確率変数

2.1 確率変数の定義

可測空間 (Ω, \mathfrak{B}) において、 Ω 上の実数値関数 $X : \Omega \rightarrow \mathbf{R}$ が $\forall a \in \mathbf{R}$ に対して

$$\{\omega; X(\omega) \leq a\} \in \mathfrak{B}$$

を満たすとき X を (Ω, \mathfrak{B}) 上の確率変数という。

例 2.1 1回のコイン投げでは、 $\Omega = \{H, T\}$, $\mathfrak{B} = 2^\Omega = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ であった。このとき、

$$X(\omega) = \begin{cases} 1 & (\omega = H) \\ 0 & (\omega = T) \end{cases} \quad \text{とすると、} \quad \{\omega; X(\omega) \leq a\} = \begin{cases} \Omega & (1 \leq a) \\ \{T\} & (0 \leq a < 1) \\ \emptyset & (a < 0) \end{cases} \quad \text{となるので、}$$

$\forall a \in \mathbf{R}$ に対して $\{\omega; X(\omega) \leq a\} \in \mathfrak{B}$ がわかる。従って、 X は確率変数である。

定理 2.1 X は (Ω, \mathfrak{B}) 上の確率変数 $\iff \forall a \in \mathbf{R}$ に対して $\{\omega; X(\omega) < a\} \in \mathfrak{B}$.

証明: (\implies) $\{\omega; X(\omega) < a\} = \bigcup_{n=1}^{\infty} \{\omega; X(\omega) \leq a - \frac{1}{n}\}$ より従う。

(\impliedby) $\{\omega; X(\omega) \leq a\} = \bigcap_{n=1}^{\infty} \{\omega; X(\omega) < a + \frac{1}{n}\}$ と定理 1.2 (2) より従う。 □

ここで、 σ 集合族の定義から

$$\begin{aligned} \{\omega; X(\omega) \leq a\} \in \mathfrak{B} &\iff \{\omega; X(\omega) > a\} \in \mathfrak{B}, \\ \{\omega; X(\omega) < a\} \in \mathfrak{B} &\iff \{\omega; X(\omega) \geq a\} \in \mathfrak{B}. \end{aligned}$$

これと定理 2.1 より、一般に次が成立する。

定理 2.2 X は (Ω, \mathfrak{B}) 上の確率変数 \iff 任意の区間 I に対して $\{\omega; X(\omega) \in I\} \in \mathfrak{B}$.

証明: (\impliedby) は明らか。 (\implies) について、 $I = (-\infty, b)$, $(-\infty, b]$, (a, ∞) , $[a, \infty)$ の場合はすでに示した。 $I = (a, b)$, $[a, b]$, $(a, b]$, $[a, b)$ の場合に $\{\omega; X(\omega) \in I\} \in \mathfrak{B}$ を示せばよい。例えば、 $I = (a, b)$ のときは、

$$\{\omega; X(\omega) \in (a, b)\} = \{\omega; X(\omega) < b\} \cap \{\omega; X(\omega) > a\}$$

より従う。他は演習問題とする。 □

2.2 可測関数と確率変数

問 1.1 より、 \mathbf{R} 上の区間 $(a, b]$ 全体からなる集合を含む最小の σ 集合族が存在する。この集合族を \mathbf{R} の Borel 集合族といい、 $\mathcal{B}(\mathbf{R})$ と表す。Borel 集合族に属する集合を Borel 集合という。すべての区間は Borel 集合となることが定理 2.2 などと同様に証明できる。

補題 2.1 可測空間 (Ω, \mathfrak{B}) において、関数 $X : \Omega \rightarrow \mathbf{R}$ によって与えられる \mathbf{R} の部分集合族 $\{A \subset \mathbf{R}; X^{-1}(A) \in \mathfrak{B}\}$ は σ 集合族である。ただし、 $X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\}$ とする。

証明: $X^{-1}(\mathbf{R}) = \Omega$, $X^{-1}(A^c) = (X^{-1}(A))^c$, $X^{-1}(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} X^{-1}(A_n)$ となることを示せば、容易に証明できるので演習問題とする。 \square

定理 2.3 X が (Ω, \mathfrak{B}) 上の確率変数 \iff 任意の Borel 集合 $B \in \mathcal{B}(\mathbf{R})$ に対して $\{\omega; X(\omega) \in B\} \in \mathfrak{B}$.

証明: (\Leftarrow) $B = (-\infty, a]$ ととれば明らか。

(\Rightarrow) X を確率変数とし、 $\mathcal{A} = \{A \subset \mathbf{R}; X^{-1}(A) \in \mathfrak{B}\}$ とおく。定理 2.2 より $(a, b] \in \mathcal{A}$ であり、補題 2.1 より σ 集合族である。従って、 $\mathcal{B}(\mathbf{R})$ は区間 $(a, b]$ 全体からなる集合族を含む最小の σ 集合族であるから $\mathcal{A} \supset \mathcal{B}(\mathbf{R})$ となる。よって $B \in \mathcal{B}(\mathbf{R})$ なら $B \in \mathcal{A}$ となり、 $X^{-1}(B) \in \mathfrak{B}$ を得る。 \square

定義 2.1 可測空間 $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ において、関数 $f : \mathbf{R} \rightarrow \mathbf{R}$ が任意の $a \in \mathbf{R}$ に対して $\{x; f(x) \leq a\} \in \mathcal{B}(\mathbf{R})$ を満たすとき、 f は $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の可測関数であるという。

注意 2.1 f が $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上で可測であることは、 f が $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の確率変数であることは同値である。

定理 2.4 連続関数 $f : \mathbf{R} \rightarrow \mathbf{R}$ が $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の可測関数である。また、高々可算個の点を除いて連続な関数も $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の可測関数である。

証明: f は連続なので $\forall a \in \mathbf{R}$ に対して $\{x; f(x) < a\}$ は開集合となる。よって、 G が開集合なら $G \in \mathcal{B}(\mathbf{R})$ を示せばよい。これは、任意の $x \in G$ に対して $\delta_x > 0$ を $(x - \delta_x, x + \delta_x) \subset G$ ととれば、有理数の集合 Q は可算集合で、 $\bigcup_{x \in G \cap Q} (x - \delta_x, x + \delta_x) = G$ となるから従う。次に、 D_f は高々可算個の点からなる集合で、 f は $\mathbf{R} \setminus D_f$ で連続とする。 $\forall a \in \mathbf{R}$ に対して、 $C_{a,1} := \{x \in D_f; f(x) < a\}$ は高々可算なのでこれは Borel 集合。また、 $C_{a,2} := \{x \in \mathbf{R} \setminus D_f; f(x) < a\}$ は開集合 G を用いて $C_{2,a} = G \setminus D_f$ と表せる。実際、 $\forall x \in C_{a,2}$ に対して、 $\delta'_x > 0$ を $y \in (x - \delta'_x, x + \delta'_x) \setminus D_f$ ならば $|f(y) - f(x)| < f(a) - f(x)$ となるように選べば、このとき $f(y) < f(a)$ なので $(x - \delta'_x, x + \delta'_x) \setminus D_f \subset C_{a,2}$ 。よって、 $G = \bigcup_{x \in C_{a,2}} (x - \delta'_x, x + \delta'_x)$ とおくと、 G は開集合で、 $C_{a,2} = G \setminus D_f$ と表せるので $C_{a,2} \in \mathcal{B}(\mathbf{R})$ 。よって、 $\{x; f(x) < a\} = C_{a,1} \cup C_{a,2} \in \mathcal{B}(\mathbf{R})$ 。 \square

定理 2.5 X が (Ω, \mathfrak{B}) 上の確率変数、 $f : \mathbf{R} \rightarrow \mathbf{R}$ が $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の可測関数のとき、 $f(X(\omega))$ は (Ω, \mathfrak{B}) 上の確率変数である。

証明: $\forall a \in \mathbf{R}$ とし、 $B = \{x; f(x) \leq a\}$ とおくと、 f は可測だから $B \in \mathcal{B}(\mathbf{R})$ 。よって、定理 2.3 により $\{\omega; f(X(\omega)) \leq a\} = \{\omega; X(\omega) \in B\} \in \mathfrak{B}$ を得る。 \square

例 2.2 定理 2.4、定理 2.5 より、 X が確率変数であるとき $aX + b$ (a, b は定数), X^a , \sqrt{X} , $\log X$, e^X , $\cos X$ などは、すべて確率変数である。

2.3 分布関数

定義 2.2 確率空間 $(\Omega, \mathfrak{B}, P)$ 上の確率変数 X に対して、 $F(x) = P(X \leq x) = P(\{\omega; X(\omega) \leq x\})$ で定義される関数 $F : \mathbf{R} \rightarrow [0, 1]$ を確率変数 X の分布関数という。

$\{\omega; X(\omega) \leq a\}$ を単に $\{X \leq a\}$ と書くことが多い。以下このような略した書き方を用いる。また、確率変数 X を明示したいときは $F_X(x)$ と表す。

定理 2.6 $P(a < X \leq b) = F(b) - F(a)$, $a, b \in \mathbf{R}$ ($a < b$)

証明: $A = \{X \leq a\}$, $B = \{a < X \leq b\}$ とすると、 $A \cap B = \emptyset$, $A \cup B = \{X \leq b\}$ より、 $F(b) = P(A \cup B) = P(A) + P(B) = F(a) + P(a < X \leq b)$. \square

定理 2.7 $F(x)$ を確率変数 X の分布関数とする。このとき、次が成立する。

- (1) F は単調非減少である、即ち、 $x < y$ ならば $F(x) \leq F(y)$.
- (2) 右連続関数である、即ち、 $F(a+0) := \lim_{x \rightarrow a+0} F(x) = F(a)$ ($\forall a \in \mathbf{R}$).
- (3) $F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0$, $F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1$.

証明: (1) $x < y$ より $\{X \leq x\} \subset \{X \leq y\}$. 従って、定理 1.2 (3) により $F(x) = P(X \leq x) \leq P(X \leq y) = F(y)$ を得る。

(2) $\{a_n\}$ を単調減少で a に収束する任意の数値列とし、 $C_n = \{X \leq a_n\}$ とおく。このとき、 $\bigcap_{n=1}^{\infty} C_n = \{X \leq a\}$ となる。実際、「 \supset 」は $\forall n$ で $C_n \supset \{X \leq a\}$ だから明らか。 $\omega \in \{X > a\}$ とすると $X(\omega) > a$ で $a_n \rightarrow a+0$ ($n \rightarrow \infty$) より、ある N があって $n \geq N \implies a_n - a < X(\omega) - a$, これより $X(\omega) > a_n$, 即ち $\omega \notin C_N \supset \bigcap_{n=1}^{\infty} C_n$ となり「 \subset 」が従う。よって、 $\{C_n\}$ は単調減少であったらから、定理 1.4 (2) より $\lim_{n \rightarrow \infty} F(a_n) = \lim_{n \rightarrow \infty} P(C_n) = P(X \leq a) = F(a)$ となる。

(3) $\{x_n\}$ を単調減少で $-\infty$ に発散する任意の数値列とする。このとき $B_n = \{X \leq x_n\}$ とおくと、 $\{B_n\}$ は単調減少な事象の列で、また X は実数値なので、 $\bigcap_{n=1}^{\infty} \{X \leq x_n\} = \emptyset$. 従って、定理 1.4 (2) と命題 1.2 (1) により、 $\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(B_n) = P(\emptyset) = 0$ となり、前半の主張を得る。後半は、 $\{x_n\}$ を単調増加で ∞ に発散する任意の数値列とし、定理 1.4 (1) と $P(\Omega) = 1$ となることを用いれば、同様に証明できる (演習問題とする)。 \square

注意 2.2 関数 $F: \mathbf{R} \rightarrow \mathbf{R}$ が定理 2.7 の条件 (1), (2), (3) を満たせば、 $F(x)$ を分布関数としてもつ確率変数 X が存在することが知られている。

定義 2.3 確率空間 $(\Omega, \mathfrak{B}, P)$ 上の確率変数 X に対して、 $P(X \in E) = 1$ となる高々可算個の元からなる集合 $E \subset \mathbf{R}$ が存在するとき、 X を離散型確率変数という。

ここで、高々可算個の元からなる集合 E は Borel 集合である。実際、一点からなる集合 $\{x\}$ は $\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right]$ となるから Borel 集合である。よって、定義 1.1 (iii) により E は Borel 集合となる。

以下、確率変数 X がとりうるのは可算個の値 x_1, x_2, \dots であるとして説明する。(有限個の場合も同様である。)

このとき、 $p(x_i) = P(X = x_i)$ ($i \in \mathbf{N}$) とすると、このとき

$$p(x_i) > 0 \quad (i \in \mathbf{N}), \quad \sum_{i=1}^{\infty} p(x_i) = 1$$

が成立する。この p を確率変数 X の確率関数という。また、次のように表すとわかりやすい:

$$\left(\begin{array}{cccccc} x_1 & x_2 & \cdots & x_n & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{array} \right) \quad \text{または} \quad \begin{array}{c|cccccc} X & x_1 & x_2 & \cdots & x_n & \cdots \\ \hline P & p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{array}$$

これを確率変数 X の分布列という。この場合、分布関数は

$$F(x) = P(X \leq x) = \sum_{k: x_k \leq x} p(x_k)$$

与えられる。これは、不連続点だけで増加する関数で、 $p(x_i) = F(x_i) - F(x_i - 0)$ となる。離散型確率変数の分布関数を離散型分布関数という。

離散型分布の例

(a) 二項分布

歪んだコイン投げの試行のように、事象 A の起こる確率が p 、事象 A が起こらない確率が $q := 1 - p$ となる試行 (Bernoulli 試行という) を n 回行うとき、事象 A の出現回数を X とする。このとき、 X のとり得る値は $0, 1, \dots, n$ で

$$b(k; n, p) := P(X = k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n)$$

であり、この分布関数は $x < 0$ のとき $F(x) = 0$ 、 $x \geq 0$ のとき

$$F(x) = \sum_{k=0}^{[x]} \binom{n}{k} p^k q^{n-k}$$

となる。 $x \geq n$ のとき $F(x) = 1$ となる (二項定理による)。この分布を二項分布 $B(n, p)$ という。また、確率変数 X がこのような分布を持つとき X は $B(n, p)$ に従うといい $X \sim B(n, p)$ とかく。

(b) 負の二項分布

(a) と同様な事象 A の起こる確率が p となる Bernoulli 試行を、 A が n 回出現するまで反復するとき、余事象 A^c が出現する回数を Y とする。このとき、 Y のとり得る値は $0, 1, \dots$ で、 $Y = k$ となるとき、 $n + k$ 回の試行で事象 A は最後を除いて $n - 1$ 回、余事象 A^c は k 回出現するので、

$$P(Y = k) = \binom{n+k-1}{k} p^n q^k \quad (k = 0, 1, \dots) \quad (2.1)$$

となる。ここで、

$$\begin{aligned} \binom{n+k-1}{k} &= \frac{(n+k-1)(n+k-2)\cdots(n+1)n}{k!} \\ &= (-1)^k \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} = (-1)^k \binom{-n}{k} \end{aligned}$$

となるので、(2.1) は

$$P(Y = k) = \binom{-n}{k} p^n (-q)^k \quad (k = 0, 1, \dots)$$

と表される。これより、この分布を負の二項分布 $NB(n, p)$ という。(この分布は n が自然数でないときも定義できる。) ここで、 $(1+x)^\alpha$ の Taylor 展開式 $(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k$ ($-1 < x < 1$) により

$$\lim_{y \rightarrow \infty} F(y) = \lim_{y \rightarrow \infty} P(Y \leq y) = \sum_{k=0}^{\infty} P(Y = k) = 1$$

となることが証明できる。(演習問題とする。)

(c) 幾何分布

事象 A の起こる確率が p となる Bernoulli 試行の反復において、 A が初めて出現するまでの A^c の出現回数を X とすると、 X のとり得る値は $0, 1, 2, \dots$ であり、このときの分布は幾何分布といい、

$$G(k; p) := P(X = k) = q^k p \quad (k = 0, 1, 2, \dots)$$

である。容易にわかるように $P(X \geq k) = \sum_{l=k}^{\infty} P(X = l) = q^k$ となる。また、

$$P(X = k + s | X \geq k) = P(X = s) \quad (s, k = 0, 1, 2, \dots)$$

となる。実際、

$$(\text{左辺}) = \frac{P(\{X = k + s\} \cap \{X \geq k\})}{P(X \geq k)} = \frac{P(X = k + s)}{P(X \geq k)} = \frac{q^{k+s}p}{q^k} = q^s p = P(X = s)$$

となるからである。

(d) Poisson 分布

$\lambda > 0$ とする。確率変数 X が非負整数値で、その確率関数が

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots) \quad (2.2)$$

で与えられるとき、この確率変数 X は母数 λ の Poisson 分布 $Po(\lambda)$ に従うという。 $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ ($x \in \mathbf{R}$) に注意すれば、 $\sum_{k=0}^{\infty} P(X = k) = 1$ がわかる。Poisson 分布は次の命題で見ると、一定時間間隔の事故の件数などを表すと考えられる。

問 2.1 各 $n \in \mathbf{N}$ に対して、確率変数 X_n は二項分布 $B(n, p_n)$ に従うとする。ここで、 p_n は $0 < p_n < 1$ および $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ を満たすとする。このとき、 $\{X_n\}$ は Poisson 分布を近似している、即ち、次が成立する。

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots)$$

証明: $P(X = k) = 1 \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{(np_n)^k}{k!} \left\{1 - \frac{np_n}{n}\right\}^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \quad (n \rightarrow \infty) \quad \square$

定義 2.4 $F(x)$ を、確率空間 $(\Omega, \mathfrak{B}, P)$ 上の確率変数 X の分布関数とする。 $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 上の非負可測関数 $f(x)$ が存在して

$$F(x) = \int_{-\infty}^x f(t) dt \quad (\forall x \in \mathbf{R})$$

と表せるとき、 F を絶対連続な分布関数、 X を絶対連続型確率変数という。また、この $f(x)$ を X の密度関数 (density function) という。

注意 2.3 (1) X が絶対連続型なら $F(x)$ は連続である。

(2) $F(x)$ が x について連続であっても、密度関数 $f(x)$ が存在するとは限らない。(詳しくは関数解析学で Cantor 関数として勉強する。)

X が絶対連続型確率変数のとき、その密度関数 $f(x)$ は

$$f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

を満たす。また、 x が f の連続点であれば $\frac{dF}{dx}(x) = f(x)$ であり、 $P(X = a) = 0$ ($\forall a \in \mathbf{R}$) であり、

$$P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(t) dt \quad (a < b)$$

となる。

絶対連続型分布の例

(a) 一様分布

確率変数 X の密度関数 f あるいは分布関数 F が

$$f(x) = \begin{cases} \frac{1}{b-a} & (a < x < b) \\ 0 & (x \leq a \text{ または } b \leq x) \end{cases} \quad F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & (x \leq a) \\ \frac{x-a}{b-a} & (a < x \leq b) \\ 1 & (b < x) \end{cases}$$

で与えられるとき、確率変数 X は区間 (a, b) 上の一様分布 $U(a, b)$ に従うという。

(b) 指数分布 (教科書と異なる)

$\lambda > 0$ とする。確率変数 X の密度関数 $f(x)$ あるいは分布関数 $F(x)$ が次で与えられるとき、この分布を指数分布 $Ex(\lambda)$ という。

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad F(x) = \begin{cases} 1 - e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

指数分布は事故などの Poisson 事象が生起する時間間隔の分布として広く用いられている。

(c) 正規分布

$\mu \in \mathbf{R}, \sigma > 0$ とする。次の密度関数をもつ分布を正規分布 (normal distribution) という。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.3)$$

この正規分布を記号 $N(\mu, \sigma^2)$ で表す。このとき、分布関数は

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

となる。ここで、 $\int_{-\infty}^{\infty} f(t) dt = 1$ に注意する。これは、

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

である (これは微分積分学で習ったように、重積分に帰着し極座標を用いれば証明できる) から、 $z = \frac{x-\mu}{\sigma}$ とおくと、 $dx = \sigma dz$ で $\frac{x}{z} \Big|_{-\infty \rightarrow \infty}^{-\infty \rightarrow \infty}$ より、

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1.$$

正規分布 $N(\mu, \sigma^2)$ において、特に $\mu = 0, \sigma = 1$ のとき、この正規分布 $N(0, 1)$ を標準正規分布という。この密度関数と分布関数をそれぞれ $\phi(x), \Phi(x)$ とすると、

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$N(\mu, \sigma^2)$ に従う確率変数 X に対して、 $Z = (X - \mu)/\sigma$ とし、 $z = \frac{x-\mu}{\sigma}$ とおくと、変数変換 $s = \frac{t-\mu}{\sigma}$ を施すことで

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds = \Phi(z),$$

即ち、 $Z = (X - \mu)/\sigma$ は標準正規分布に従う。これより、 $P(a \leq X \leq b)$ などは、性質 $P(X \leq x) = \Phi(\frac{x-\mu}{\sigma})$ と標準正規分布表 (教科書 p.163) を利用して計算できる。 $\Phi(-z) = 1 - \Phi(z)$ ($z > 0$) にも注意する。

例えば、確率変数 X が $N(1, 4^2)$ に従うとき $P(0 \leq X \leq 5)$ を求めるには、 $Z = \frac{X-1}{4}$ とし、

$$\begin{aligned} P(0 \leq X \leq 5) &= P\left(\frac{0-1}{4} \leq Z \leq \frac{5-1}{4}\right) = P(-0.25 \leq Z \leq 1) = \Phi(1) - \Phi(-0.25) \\ &= 1 - (1 - \Phi(1)) - (1 - \Phi(0.25)) = 1 - 0.1587 - 0.4013 = 0.4400 \end{aligned}$$

となる。

正規分布は中心極限分布で現れる分布で、偏差値など日常幅広く活用されている。

2.4 多次元確率変数

教科書と番号が異なる。

定義 2.5 $X_1(\omega), \dots, X_n(\omega)$ を確率空間 $(\Omega, \mathfrak{B}, P)$ 上の確率変数とすると、

$$\mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)) \quad (2.4)$$

を n 次元確率変数または n 次元確率ベクトルという。ここに、 $\mathbf{X} = (X_1, \dots, X_n)$ とも略記される。

定義 2.6 任意の $(x_1, \dots, x_n) \in \mathbf{R}^n$ に関し、(2.4) の多次元確率変数 $\mathbf{X} = (X_1, \dots, X_n)$ の同時分布関数を次式で定義する。

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n).$$

注意 2.4 $P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$ は $P(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\} \cap \dots \cap \{X_n \leq x_n\})$ を表す。今後にもこのように “ \cap ” を略した記号を用いる。

n 次元 Borel 集合族を問 1.1 を用いて次のように定義する。

定義 2.7 \mathbf{R}^n のすべての开区間 $(a_1, b_1) \times (a_2, b_2) \times \dots \times (a_n, b_n)$, $a_i < b_i$, $i = 1, 2, \dots, n$, を含む最小の σ 集合族を、 n 次元 Borel 集合族といい、 $\mathcal{B}(\mathbf{R}^n)$ と表す。 $\mathcal{B}(\mathbf{R}^n)$ の元を Borel 集合という。また、関数 $\varphi: \mathbf{R}^n \rightarrow \mathbf{R}$ が任意の $B \in \mathcal{B}(\mathbf{R})$ に対して、集合族 $\varphi^{-1}(B) \in \mathcal{B}(\mathbf{R}^n)$ を満たすとき、 φ を $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ 上の可測関数、あるいは、Borel 可測であるという。

原理的には、任意の $B \in \mathcal{B}(\mathbf{R})$ に対して $P((X_1, \dots, X_n) \in B)$ の値は $F_{\mathbf{X}}(x_1, x_2, \dots, x_n)$ から定まる。例えば、2 次元の場合、 $B = (a_1, b_1] \times (a_2, b_2]$ の場合、 $B = \bigcap_{n=1}^{\infty} (a_1, b_1 + \frac{1}{n}) \times (a_2, b_2 + \frac{1}{n})$ より Borel 集合であることに注意する。このとき、

$$\begin{aligned} P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) &= P(a_1 < X_1 \leq b_1, X_2 \leq b_2) - P(a_1 < X_1 \leq b_1, X_2 \leq a_2) \\ &= P(X_1 \leq b_1, X_2 \leq b_2) - P(X_1 \leq a_1, X_2 \leq b_2) - \{P(X_1 \leq b_1, X_2 \leq a_2) - P(X_1 \leq a_1, X_2 \leq a_2)\} \\ &= F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2) \end{aligned}$$

を得る。

次の定理は定理 2.5 の拡張である。証明は省略する。

定理 2.8 X_1, \dots, X_n が (Ω, \mathfrak{B}) 上の確率変数、 $f: \mathbf{R}^n \rightarrow \mathbf{R}$ が $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$ 上の可測関数のとき、 $f(X_1(\omega), \dots, X_n(\omega))$ は (Ω, \mathfrak{B}) 上の確率変数である。

定理 2.7 と同様に次が成り立つ。証明は省略する。

命題 2.1 (1) 各変数 x_j ごとに分布関数 $F_{\mathbf{X}}(x_1, x_2, \dots, x_n)$ は非減少な右連続関数である。

$$(2) F_{\mathbf{X}}(x_1, \dots, -\infty, \dots, x_n) := \lim_{x_j \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_j, \dots, x_n) = 0.$$

$$\text{また、} F_{\mathbf{X}}(\infty, \dots, \infty) := \lim_{x_1, \dots, x_n \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_n) = 1.$$

多次元確率変数 $\mathbf{X} = (X_1, \dots, X_n)$ において、

$$P(X_j \leq x_j) = F_{\mathbf{X}}(\infty, \dots, \infty, x_j, \infty, \dots, \infty)$$

を X_j の 1 次元周辺分布という。同様に、多次元周辺分布も定義される。

一次元の場合同様に離散分布、絶対連続分布を以下のように定義する。

(1) 高々可算個の点からなる集合 $D \subset \mathbf{R}^n$ があって、 $P((X_1, \dots, X_n) \in D) = 1$ となる場合、この多次元確率変数を離散型という。

(2) R^n 上の可測関数 $f(x_1, \dots, x_n)$ が存在して

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \cdots \int_{-\infty}^{x_1} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \cdots dt_n$$

と表せるとき、 F を絶対連続な分布関数、 $\mathbf{X} = (X_1, \dots, X_n)$ を絶対連続型確率変数という。また、この $f(x_1, \dots, x_n)$ を \mathbf{X} の同時密度関数という。

絶対連続型確率変数 (X_1, \dots, X_n) に対し、 X_i の周辺分布は絶対連続型となり、その密度関数は他の座標をすべて R で積分してしまうことで得られる。実際、 $j = 1$ のとき証明すると

$$P(X_1 \leq x) = P(X \leq x_1, X_2 \in R, \dots, X_n \in R) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, t_2, \dots, t_n) dt_n \cdots dt_2 dt_1$$

より、 X_1 の周辺密度関数 f_{X_1} は

$$f_{X_1}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x, t_2, \dots, t_n) dt_n \cdots dt_2 \quad (2.5)$$

となるのがわかる。また、 $B \subset R^n$ が Borel 集合のとき、

$$P((X_1, \dots, X_n) \in B) = \iint \cdots \int_B f(t_1, t_2, \dots, t_n) dt_1 dt_2 \cdots dt_n \quad (2.6)$$

となるのが、(Lebesgue 積分論を用いて) 証明できる。

例 2.3 多次元確率変数とその分布を例示する。

(i) 多項分布

ある試行で対象とする事象 A_1, A_2, \dots, A_L がすべてで、互いに排反である、即ち、 $\Omega = \bigcup_{j=1}^L A_j$, $A_i \cap A_j = \emptyset$ ($i \neq j$) とする。また、1 回の試行で事象 A_j が出現する確率を p_j とし、この試行を n 回反復して各事象 A_j の出現回数を X_j ($j = 1, \dots, L$) とすれば、 (X_1, \dots, X_L) の同時分布は次式で示される。

$$P(X_1 = k_1, X_2 = k_2, \dots, X_L = k_L) = \frac{n!}{k_1! k_2! \cdots k_L!} p_1^{k_1} p_2^{k_2} \cdots p_L^{k_L}$$

ただし、 $k_1 + k_2 + \cdots + k_L = n$ である。このような分布を多項分布という。

この多項分布に対し X_1 の 1 次元周辺分布が二項分布 $B(n, p_1)$ であることを示そう。簡単のため $L = 4$ として示す。一般の場合はそれを繰り返せばよい。 $X_1 = k_1, X_2 = k_2$ とすると $X_3 + X_4 = n_3 := n - k_1 - k_2$ であるから、

$$\begin{aligned} P(X_1 = k_1, X_2 = k_2) &= \sum_{k_3=0}^{n_3} P(X_1 = k_1, X_2 = k_2, X_3 = k_3, X_4 = n_3 - k_3) \\ &= \frac{n!}{k_1! k_2! n_3!} p_1^{k_1} p_2^{k_2} \sum_{k_3=0}^{n_3} \frac{n_3!}{k_3! (n_3 - k_3)!} p_3^{k_3} p_4^{n_3 - k_3} \\ &= \frac{n!}{k_1! k_2! n_3!} p_1^{k_1} p_2^{k_2} (p_3 + p_4)^{n_3} \end{aligned}$$

最後の等式は二項定理を用いた。特に $(X_1, X_2, X_3 + X_4)$ も多項分布となる。

$$\begin{aligned} P(X_1 = k_1) &= \sum_{k_2=0}^{n-k_1} P(X_1 = k_1, X_2 = k_2) \\ &= \frac{n!}{k_1! (n - k_1)!} p_1^{k_1} \sum_{k_2=0}^{n-k_1} \frac{(n - k_1)!}{k_2! n_3!} p_2^{k_2} (p_3 + p_4)^{n_3} \\ &= \frac{n!}{k_1! (n - k_1)!} p_1^{k_1} (p_2 + p_3 + p_4)^{n - k_1} = \binom{n}{k_1} p_1^{k_1} (1 - p_1)^{n - k_1} \end{aligned}$$

3 つ目の等号は $n_3 = n - k_1 - k_2$, 最後の等号は $p_1 + p_2 + p_3 + p_4 = 1$ を用いた。□

(ii) 多次元正規分布

$\mu = (\mu_1, \dots, \mu_n)$ と n 次正定値対称行列 $\Sigma = (\sigma_{ij})$ に対して、 n 次元確率変数 $X = (X_1, \dots, X_n)$ が次の密度関数をもつとき、この分布を n 次元正規分布 $N(\mu', \Sigma)$ という。

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} e^{-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)'}$$

ここに、 $x = (x_1, \dots, x_n)$ とし、 x' は x の転置 (この場合縦ベクトル) を表す。

これが R^n 上で積分すれば 1 となることは次のように示すことができる。

Σ は正定値対称行列なので、線形代数の対角化に関する定理から、直交行列 P と対角成分がすべて正の対角行列 $D = (d_{jj})$ がとれて、 $P'\Sigma P = D$ とできる。ここで、 $y = (y_1, \dots, y_n) = (x - \mu)P$ とすると、 $D^{-1} = (P'\Sigma P)^{-1} = P'\Sigma^{-1}P$ より

$$(x - \mu)\Sigma^{-1}(x - \mu)' = yP'\Sigma^{-1}Py' = yD^{-1}y' = \sum_{j=1}^n \frac{1}{d_{jj}} y_j^2$$

で、ヤコビアンは $\frac{\partial y}{\partial x} = \det P = \pm 1$, $\det \Sigma = \det D = \prod_{j=1}^n \lambda_{jj}$ なので、

$$\int \cdots \int_{R^n} f(x_1, \dots, x_n) dx_1 \cdots dx_n = \prod_{j=1}^n \int_{-\infty}^{\infty} \frac{1}{(2\pi d_{jj})^{1/2}} e^{-\frac{1}{2} \frac{y_j^2}{d_{jj}}} dy_j = 1$$

となる。 □

2.5 条件付き確率分布

この節では確率変数で条件付けをした分布について、離散型確率変数の場合と絶対連続型確率変数の場合に定義する。一般の確率変数について条件付き確率分布を定義するためには、Lebesgue 積分論の Radon-Nikodym の定理を必要とするのでこの授業では取り扱わない。

ここでは、簡単のため (X, Y) を 2 次元確率変数として、条件付き分布 $P(Y \in A | X = x)$ を定義する。

(i) X の周辺分布が離散型の場合

X のとり得る値が $\{x_1, x_2, \dots\}$ とすると、 $P(X = x_j) > 0 (\forall j)$ となることに注意する。この場合、

$$P(Y \in A | X = x_j) = \frac{P(X = x_j, Y \in A)}{P(X = x_j)}$$

と定め、これを $X = x_j$ の条件下での Y の条件付き確率分布という。ただし、 $P(X = x) = 0$ となる x では定義しないものとする。

(ii) X の周辺分布が絶対連続型の場合

この場合 $P(X = x) = 0 (\forall x)$ であるから工夫が必要となる。

X の周辺密度関数を $f_X(x)$ とする。(式 (2.5) のように定義される。) ここでは、簡単のため $f_X(x)$ の不連続点は高々可算個とし、 x は $f_X(x)$ の連続点で $f_X(x) > 0$ となるときに考える。 $P(Y < y | X = x)$ を次のように定義する。

$$P(Y \in A | X = x) = \lim_{\delta \rightarrow +0} P(Y \in A | x \leq X < x + \delta)$$

と定める。このとき、 $\xi_k = a + \frac{k}{n}(b - a)$ とすることで、

$$\int_a^b P(Y \in A | X = x) f_X(x) dx = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(Y \in A | \xi_{k-1} \leq X < \xi_k) \int_{\xi_{k-1}}^{\xi_k} f_X(x) dx = P(Y \in A, a \leq X < b)$$

となることに注意する。次に、 (X, Y) は絶対連続型で $f(x, y)$ をその同時密度関数とする。このとき、

$$P(Y < y | x \leq X < x + \delta) = \frac{P(Y < y, x \leq X < x + \delta)}{P(x \leq X < x + \delta)} = \frac{\int_x^{x+\delta} \int_{-\infty}^y f(s, t) dt ds}{\int_x^{x+\delta} f_X(s) ds}$$

ここで、分母分子を δ で割り $\delta \rightarrow +0$ とすることで、微分積分学の基本定理により

$$P(Y < y | X = x) = \frac{\int_{-\infty}^y f(x, t) dt}{f_X(x)}$$

を得る。これより、条件 $X = x$ の下での Y の分布は絶対連続型で、その密度関数 $f_{Y|X}(y|x)$ が

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} \quad (2.7)$$

となることがわかった。 $f_{Y|X}(y|x)$ を条件 $X = x$ のもとでの Y の密度関数という。□

例 2.4 (X, Y) を 2次元正規分布 $N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right)$ に従うとする。ただし、 Σ は正定値であったから、 $\det(\lambda I - \Sigma) = \lambda^2 - (\sigma_{11} + \sigma_{22})\lambda + \sigma_{11}\sigma_{22} - \sigma_{12}^2 = 0$ の根が 2 つとも正となるため、 $\sigma_{11}, \sigma_{22} > 0$, $\sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0$ を得る。これより、 $\sigma_1, \sigma_2 > 0$, $|\rho| < 1$ を用いて、 $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ と表せる。このとき、 $\det \Sigma = \sigma_1^2\sigma_2^2(1 - \rho^2)$ で $\Sigma^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{pmatrix}$ であるから、 (X, Y) の密度関数は

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left[-\frac{1}{2(1 - \rho^2)} \left\{ \frac{(x - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x - \mu_1)(y - \mu_2)}{\sigma_1\sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right\}\right] \quad (2.8)$$

となる。

次に、 $f_{Y|X}(y|x)$ を求め、 $X = x$ の条件下の Y の分布を求めよう。

まず、 X の周辺密度関数 $f_X(x)$ を求める。(2.8) の右辺の $\{\dots\}$ の中を y について平方完成すると

$$\frac{1}{\sigma_2^2} \left[(y - \mu_2) - \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1) \right]^2 + \frac{1 - \rho^2}{\sigma_1^2} (x - \mu_1)^2$$

となるから、 $t = \frac{1}{\sigma_2\sqrt{1 - \rho^2}} [y - \mu_2 - \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1)]$ とおくと、

$$f_X(x) = \int_{\mathbf{R}} f(x, y) dy = \frac{1}{2\pi\sigma_1} e^{-\frac{(x - \mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x - \mu_1)^2}{2\sigma_1^2}}$$

となる。(これより X の周辺分布は $N(\mu_1, \sigma_1^2)$ であることがわかった。) よって、

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1 - \rho^2}} \exp\left[-\frac{(y - \{\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1)\})^2}{2\sigma_2^2(1 - \rho^2)}\right]$$

を得る。これは、 $X = x$ の条件下の Y の分布が正規分布 $N(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$ となることを示している。□

例題 2.1 (X, Y) の同時密度関数を $f(x, y) = \begin{cases} Kxy & (0 \leq x, 0 \leq y, x^2 + y^2 \leq 1) \\ 0 & (\text{その他}) \end{cases}$ とする。

(1) 定数 K を求めよ。

(2) X の周辺密度関数 $f_X(x)$ を求めよ。

(3) $0 < x < 1$ のとき $f_{Y|X}(y|x)$ を求め、 $P(Y > \frac{1}{2} | X = \frac{1}{2})$ を計算せよ。

(4) $T = X + Y$ とする。 T の分布関数 $F_T(t)$ を求めよ。

解 (1) $1 = \iint_{\mathbf{R}^2} f(x, y) dx dy = \int_0^1 dx \int_0^{\sqrt{1-x^2}} Kxy dy = \dots = \frac{K}{8}$ より $K = 8$.

(2) $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ より、 $x \notin (0, 1)$ のとき明らかに $f_X(x) = 0$. $0 < x < 1$ のとき、

$$f_X(x) = \int_0^{\sqrt{1-x^2}} 8xy dy = 4x(1-x^2).$$

$$(3) f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \frac{2y}{1-x^2} & (0 \leq y \leq \sqrt{1-x^2}) \\ 0 & (\text{その他}) \end{cases}.$$

$$P\left(Y > \frac{1}{2} \mid X = \frac{1}{2}\right) = \int_{\frac{1}{2}}^{\infty} f_{Y|X}\left(y \mid \frac{1}{2}\right) dy = \int_{\frac{1}{2}}^{\frac{\sqrt{3}}{2}} \frac{2y}{1 - (\frac{1}{2})^2} dy = \frac{2}{3}.$$

(4) $t < 0$ のとき $F_T(t) = 0$, $\sqrt{2} \leq t$ のとき $F_T(t) = 1$ は明らか. $0 \leq t \leq 1$ のとき、

$$F_T(t) = P(X + Y \leq t) = \int_0^t dx \int_0^{1-t} 8xy dy = \dots = \frac{1}{3}t^3.$$

$1 \leq t < \sqrt{2}$ のとき、 $x + y = t$, $x^2 + y^2 = 1$ の交点の x 座標を α, β ($\alpha < \beta$) とすると、 $\alpha + \beta = t$, $\alpha\beta = \frac{t^2 - 1}{2}$. よって、 $(\beta - \alpha)^2 = 2 - t^2$ となることに注意する. よって、

$$\begin{aligned} 1 - F_T(t) &= P(X + Y > t) = \iint_{t-x \leq y \leq \sqrt{1-x^2}} 8xy dx dy = \int_{\alpha}^{\beta} dx \int_{t-x}^{\sqrt{1-x^2}} 8xy dy \\ &= \int_{\alpha}^{\beta} \{4x(1-x^2) - 4x(t-x)^2\} dx = \left[\frac{8}{3}tx^3 - 2x^4 - 2(t^2-1)x^2 \right]_{\alpha}^{\beta} \\ &= (\beta - \alpha) \left\{ \frac{8}{3}t((\alpha + \beta)^2 - \alpha\beta) - 2(\alpha + \beta)(\alpha^2 + \beta^2) - 2(t^2 - 1)(\alpha + \beta) \right\} \\ &= \sqrt{2-t^2} \left\{ \frac{4}{3}t(t^2 + 1) - 2t(t^2 - t^2 + 1) - 2(t^2 - 1)t \right\} = \dots = \frac{2}{3}t(2-t^2)^{3/2}. \end{aligned}$$

以上をまとめて、

$$F_T(t) = \begin{cases} 0 & (t < 0) \\ \frac{1}{3}t^3 & (0 \leq t < 1) \\ 1 - \frac{2}{3}t(2-t^2)^{3/2} & (1 \leq t < \sqrt{2}) \\ 1 & (\sqrt{2} \leq t) \end{cases} \quad \square$$

2.6 確率変数の独立性

定義 2.8 (1) 確率変数の列 X_1, X_2, \dots, X_n が独立であるとは、任意の $x_1, x_2, \dots, x_n \in \mathbf{R}$ に対して

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1)P(X_2 \leq x_2) \cdots P(X_n \leq x_n) \quad (2.9)$$

となるときにいう。

(2) 無限個の確率変数の族 $\{X_\lambda\}$ が独立であるとは、その任意の有限部分列 $X_{\lambda_1}, \dots, X_{\lambda_q}$ が独立であるときにいう。

確率変数の列 X_1, X_2, \dots, X_n が独立であることと、任意の Borel 集合 $B_1, \dots, B_n \subset \mathbf{R}$ に対して

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = P(X_1 \in B_1)P(X_2 \in B_2) \cdots P(X_n \in B_n) \quad (2.10)$$

となるのが同値である。(Lebesgue 積分論により容易に証明できる。) 特に、離散型確率変数であれば

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1)P(X_2 = x_2) \cdots P(X_n = x_n) \quad \forall x_1, x_2, \dots, x_n \in \mathbf{R}$$

と同値となる。絶対連続型確率変数の場合は次の定理で見るように同時密度関数がそれぞれの一次元周辺密度関数の積に分解できればよい。

定理 2.9 絶対連続型確率変数の列 X_1, \dots, X_n が独立であることのための必要十分条件は、 (X_1, \dots, X_n) の同時密度関数を $f(x_1, \dots, x_n)$ 、 X_j の周辺密度関数を $f_{X_j}(x)$ ($j = 1, \dots, n$) とするとき

$$f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n), \quad \forall x_1, x_2, \dots, x_n \in \mathbf{R} \quad (2.11)$$

が成立することである。

証明: 絶対連続型であるから (2.9) を密度関数で表すと

$$\int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, t_2, \dots, t_n) dt_1 \cdots dt_n = \left(\int_{-\infty}^{x_1} f(t) dt \right) \times \cdots \times \left(\int_{-\infty}^{x_n} f(t) dt \right). \quad (2.12)$$

これを x_1, \dots, x_n でそれぞれ 1 回ずつ偏微分すれば (2.11) を得る。逆に、(2.11) を各 x_j に関して $(-\infty, x_j]$ の範囲で定積分すれば (2.12) を得る。□

例 2.5 $X = (X_1, \dots, X_n)$ を n 次元正規分布 $N(\boldsymbol{\mu}', \Sigma)$ に従うとする。 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ 、 $\Sigma = (\sigma_{ij})$ とかく。このとき、 X_1, \dots, X_n が独立であるための必要十分条件が、 Σ が対角行列であること、即ち、 $i \neq j$ となる任意の i, j で $\sigma_{ij} = 0$ となることである。

証明: (必要性) Σ が対角行列のとき Σ^{-1} も対角行列でその jj 成分は $1/\sigma_{jj}$ であるから、明らかに

$$\frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\sigma_{jj}}} e^{-\frac{(x_j-\mu_j)^2}{2\sigma_{jj}}}. \quad (2.13)$$

が成立する。よって、定理 2.9 より独立であることが従う。

(十分性) 共分散の節で証明する (定理 4.12 とその系)。□

例 2.6 X_1, \dots, X_n を独立で同分布に従う (*i.i.d.* と書く) 確率変数とし、 $F(x)$ で X_1 の分布関数を表す。このとき、次が成立する。

- (1) $F_{(n)}(x)$ を $X_{(n)} := \max\{X_1, \dots, X_n\}$ の分布関数とすると、 $F_{(n)}(x) = F(x)^n$ となる。
- (2) $F_{(1)}(x)$ を $X_{(1)} := \min\{X_1, \dots, X_n\}$ の分布関数とすると、 $1 - F_{(1)}(x) = (1 - F(x))^n$ となる。

証明: (1) $F_{X_{(n)}}(x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{j=1}^n P(X_j \leq x) = F(x)^n$.

(2) $1 - F_{X_{(1)}}(x) = P(X_{(1)} > x) = P(X_1 > x, \dots, X_n > x) = \prod_{j=1}^n P(X_j > x) = (1 - F(x))^n$. □

例 2.7 X_1, X_2 が独立で、各 X_i が Poisson 分布 $Po(\lambda_i)$ ($\lambda_i > 0$) に従うとき、 $Y_1 = X_1 + X_2$ と $Y_2 = X_2$ の同時分布を求めよ。また、 Y_1 の周辺分布を求めよ。

解: (Y_1, Y_2) のとりうる値は、 $(y_1, y_2) \in \mathbf{Z}^2 : y_1 \geq y_2 \geq 0$ である。このとき、同時分布は

$$P(Y_1 = y_1, Y_2 = y_2) = P(X_1 = y_1 - y_2, X_2 = y_2) = \frac{\lambda_1^{y_1-y_2} \lambda_2^{y_2}}{(y_1 - y_2)! y_2!} e^{-\lambda_1 - \lambda_2}$$

となる。また、 Y_1 の周辺分布は $y_1 \in \{0\} \cup \mathbf{N}$ に対して

$$P(Y_1 = y_1) = \sum_{y_2=0}^{y_1} \frac{\lambda_1^{y_1-y_2} \lambda_2^{y_2}}{(y_1 - y_2)! y_2!} e^{-\lambda_1 - \lambda_2} = \frac{1}{y_1!} \sum_{y_2=0}^{y_1} \binom{y_1}{y_2} \lambda_1^{y_1-y_2} \lambda_2^{y_2} e^{-\lambda_1 - \lambda_2} = \frac{(\lambda_1 + \lambda_2)^{y_1}}{y_1!} e^{-(\lambda_1 + \lambda_2)}$$

となり、これは Y_1 が Poisson 分布 $Po(\lambda_1 + \lambda_2)$ に従うことを示している。□

3 確率変数の変換

3.1 はじめに

3.2 離散型確率変数の変換

各自読んでおくように伝える。(省略するが試験には出すかもしれない。)

3.3 絶対連続型確率変数の変換

定理 3.1 n 次元確率変数 (X_1, \dots, X_n) の密度関数を $f_X(x_1, \dots, x_n)$ とする。 $y_i = u_i(x_1, \dots, x_n)$ ($1 \leq i \leq n$) が $f_X(x_1, \dots, x_n) > 0$ なる領域で次の仮定 (1)–(3) を満たす R^n から R^n への 1 対 1 変換とする。

- (1) 変換 $y_i = u_i(x_1, \dots, x_n)$ ($1 \leq i \leq n$) およびその逆変換 $x_i = v_i(y_1, \dots, y_n)$ ($1 \leq i \leq n$) はともに連続である。
- (2) 偏導関数 $\frac{\partial x_i}{\partial y_j}$ ($1 \leq i, j \leq n$) が存在して連続である。
- (3) ヤコビアン $\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}$ がゼロにならない。

このとき、 $Y_i = u_i(X_1, \dots, X_n)$ ($1 \leq i \leq n$) で確率ベクトル (X_1, \dots, X_n) を (Y_1, \dots, Y_n) に変換するとき、 (Y_1, \dots, Y_n) の分布は絶対連続型でその密度関数 $f_Y(y_1, \dots, y_n)$ は次式で与えられる。

$$f_Y(y_1, \dots, y_n) = f_X(v_1(y_1, \dots, y_n), \dots, v_n(y_1, \dots, y_n)) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|.$$

証明: $(a_1, \dots, a_n) \in R^n$ に対して $E = \{(x_1, \dots, x_n); u_i(x_1, \dots, x_n) \leq a_i, 1 \leq i \leq n\}$ とおくと、この変換で E は $E' = \{(y_1, \dots, y_n); y_i \leq a_i, 1 \leq i \leq n\}$ に対応するから、多変数関数の積分の変数変換の公式により

$$\begin{aligned} P(Y_1 \leq a_1, \dots, Y_n \leq a_n) &= P(u_1(X_1, \dots, X_n) \leq a_1, \dots, u_n(X_1, \dots, X_n) \leq a_n) \\ &= P((X_1, \dots, X_n) \in E) = \int \cdots \int_E f_X(x_1, \dots, x_n) dx_1 \cdots dx_n = \int \cdots \int_{E'} f_Y(y_1, \dots, y_n) dy_1 \cdots dy_n \\ &= \int_{-\infty}^{a_n} \cdots \int_{-\infty}^{a_1} f_Y(y_1, \dots, y_n) dy_1 \cdots dy_n \quad \square \end{aligned}$$

定理 3.2 (X, Y) の密度関数を $f(x, y)$ とする。このとき、 $Z = X + Y$ とすると Z の密度関数 $f_Z(z)$ は

$$f_Z(z) = \int_{-\infty}^{\infty} f(t, z-t) dt$$

となる。

証明: $Z = X + Y, T = X$ という変換を考えると、 $X = T, Y = Z - T$ よりヤコビアンは

$$\frac{\partial(x, y)}{\partial(t, z)} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

よって、定理 3.1 より (T, Z) の密度関数は $f_{(T,Z)}(t, z) = f(t, z-t) | -1 | = f(t, z-t)$ となるので、 $f_Z(z)$ は Z の周辺密度関数となるから、 t で積分して与式で与えられることがわかる。 \square

定理 3.3 確率変数 X_1, \dots, X_n は独立で、各 X_i は正規分布 $N(\mu_i, \sigma_i^2)$ に従うとする。このとき、 $Y = X_1 + \dots + X_n$ は正規分布 $N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$ に従う。

証明: $n = 2$ の場合に証明する。一般の場合は、 $X_1 + \dots + X_k$ と X_{k+1} が独立になることに注意して、これを繰り返し用いればよい。 X_i の密度関数は $f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}}$ であり、 (X_1, X_2) の同時密度関数は $f_{X_1}(x_1)f_{X_2}(x_2)$ であった。したがって、 $Y = X_1 + X_2$ の密度関数は定理 3.2 より、

$$f_Y(y) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{(t - \mu_1)^2}{2\sigma_1^2} - \frac{(y - t - \mu_2)^2}{2\sigma_2^2}\right] dt.$$

ここで、 $u = t - \mu_1$ とおき、 \exp の中を u について平方完成すると

$$-\frac{u^2}{2\sigma_1^2} - \frac{(y - u - \mu_1 - \mu_2)^2}{2\sigma_2^2} = -\frac{(\sigma_1^2 + \sigma_2^2)}{2\sigma_1^2\sigma_2^2} \left[x - \frac{\sigma_1^2(y - \mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2} \right]^2 - \frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}$$

となり、 $s = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2} \left[x - \frac{\sigma_1^2(y - \mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2} \right]$ とおき、 $\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du = \sqrt{2\pi}$ となることを用いると、

$$f_Y(y) = \frac{1}{\sqrt{2\pi}(\sigma_1^2 + \sigma_2^2)} \exp\left[-\frac{(y - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right]. \quad \square$$

定義 3.1 $\alpha > 0, \beta > 0$ とする。確率変数 X が密度関数

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

をもつとき、 X はガンマ分布 $\Gamma(\alpha, \beta)$ に従うという。ここで、 $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$ はガンマ関数を表す。特に、ガンマ分布 $\Gamma(\frac{\nu}{2}, 2)$ に従うとき、自由度 ν のカイ二乗分布 χ_ν^2 に従うという。

補題 3.1 (ベータ関数 $B(p, q)$ 、ガンマ関数 $\Gamma(p)$ に関する公式の復習) 以下が成り立つ。

(1) $\Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$.

(1) $\Gamma(p+1) = p\Gamma(p)$ ($p > 0$) 特に、自然数 n に対して $\Gamma(n) = (n-1)!$ となる。

(3) $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ ($p, q > 0$). ただし、 $B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1} dx$ ($p, q > 0$) はベータ関数。

証明 (1) $\Gamma(1) = \int_0^{\infty} e^{-x} dx = [-e^{-x}]_0^{\infty} = 1,$

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} x^{-\frac{1}{2}} e^{-x} dx = \int_0^{\infty} \left(\frac{t^2}{2}\right)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2} t dt = \sqrt{2} \int_0^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2} \frac{\sqrt{2\pi}}{2} = \sqrt{\pi}. \quad (x = \frac{1}{2}t - 2)$$

(2) $\Gamma(p+1) = \int_0^{\infty} x^p (-e^{-x})' dx = [-x^{p+1} e^{-x}]_0^{\infty} + \int_0^{\infty} px^{p-1} (-e^{-x})' dx = p\Gamma(p).$

(3) まず、 $D; x > 0, y > 0$ とすると、 $\Gamma(p)\Gamma(q) = \iint_D e^{-x-y} x^{p-1} y^{q-1} dx dy$ となる。ここで、 $n \geq 2$ に対して $K_n: \frac{1}{n} \leq x+y \leq n, \frac{1}{n-1}x \leq y \leq (n-1)x$ とすると、 $\{K_n\}$ は D に収束する増大列である。このとき、 $x = uv, y = u(1-v)$ と変換すると $u = x+y, v = \frac{x}{x+y}$ より、 K_n は $F_n: \frac{1}{n} \leq u \leq n, \frac{1}{n} \leq v \leq 1 - \frac{1}{n}$ に対応し、ヤコビアンは $\frac{\partial(x, y)}{\partial(u, v)} = -u$ となる。よって、

$$\begin{aligned} \iint_{K_n} e^{-x-y} x^{p-1} y^{q-1} dx dy &= \iint_{F_n} e^{-uv-u(1-v)} (uv)^{p-1} \{u(1-v)\}^{q-1} | -u | dudv \\ &= \int_{\frac{1}{n}}^n u^{p+q-1} e^{-u} du \int_{\frac{1}{n}}^{1-\frac{1}{n}} v^{p-1} (1-v)^{q-1} dv \end{aligned}$$

となり、 $n \rightarrow \infty$ として $\Gamma(p)\Gamma(q) = \Gamma(p+q)B(p, q)$ を得る。 \square

定理 3.4 確率変数 X_1, \dots, X_n は独立で、各 X_i はガンマ分布 $\Gamma(\alpha_i, \beta)$ に従うとする。このとき、 $Y = X_1 + \dots + X_n$ はガンマ分布 $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ に従う。

証明: $n = 2$ の場合に証明する。一般の場合は、 $X_1 + \dots + X_k$ と X_{k+1} が独立になることに注意してこれを繰り返し用いればよい。 X_i の密度関数を f_{X_i} とすると、 $Y = X_1 + X_2$ の密度関数は定理 3.2 より、

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(t)f_{X_2}(y-t) dt = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \int_0^{\max\{y,0\}} t^{\alpha_1-1}(y-t)^{\alpha_2-1} e^{-\frac{t+y-t}{\beta}} dt \quad (3.1)$$

$y \leq 0$ のとき $f_Y(y) = 0$ は明らか。 $y > 0$ のとき、 $t = ys$ と置換することで、

$$\int_0^y t^{\alpha_1-1}(y-t)^{\alpha_2-1} dt = y^{\alpha_1+\alpha_2-1} \int_0^1 s^{\alpha_1-1}(1-s)^{\alpha_2-1} ds = y^{\alpha_1+\alpha_2-1} B(\alpha_1, \alpha_2)$$

となり、補題 3.1(3) を用いて (3.1) に代入して、

$$f_Y(y) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y^{\alpha_1+\alpha_2-1} \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} e^{-\frac{y}{\beta}} = \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y^{\alpha_1+\alpha_2-1} e^{-\frac{y}{\beta}}. \quad \square$$

定理 3.5 確率変数 X が標準正規分布 $N(0, 1)$ に従うとすると、 X^2 は自由度 1 のカイ二乗分布 χ_1^2 に従う。特に、確率変数 X_1, \dots, X_n は独立で、各 X_i は標準正規分布 $N(0, 1)$ に従うとする。このとき、 $Y = X_1^2 + \dots + X_n^2$ は自由度 n のカイ二乗分布 χ_n^2 に従う。

証明: X が $N(0, 1)$ に従うとすると、 $a \leq 0$ のとき $P(X^2 \leq a) = 0$ で、 $a > 0$ のとき、

$$P(X^2 \leq a) = P(-\sqrt{a} \leq X \leq \sqrt{a}) = \int_{-\sqrt{a}}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2 \int_0^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

ここで、 $y = x^2$ とすると、

$$2 \int_0^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_0^a \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} y^{-\frac{1}{2}} dy$$

ここで、 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ なので、最初の主張は従う。後者の主張は、 X_1, \dots, X_n は独立であれば、 X_1^2, \dots, X_n^2 も独立であるので、定理 3.4 からの帰結となる。 \square

4 期待値

4.1 Lebesgue 積分

Lebesgue 積分論に基づいて \mathbf{R}^n 上の Borel 可測関数 $\varphi(x)$ と n 次元確率変数 $\mathbf{X} = (X_1, \dots, X_n)$ に対して、 $\int_{\Omega} \varphi(\mathbf{X}(\omega)) dP(\omega)$ の定義の概略を述べる。

1st step $\varphi(x)$ が単関数のとき、即ち、 $\varphi(x) = \sum_{k=1}^N a_k 1_{A_k}(x)$ と表せるとき。ここで、 $a_1, \dots, a_N \in \mathbf{R}$ 、 $A_1, \dots, A_N \in \mathcal{B}(\mathbf{R}^n)$ とし、 $1_A(x)$ は A の定義関数を表す。

$$1_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

このとき、次のように定義する。(本来は well-defined かどうか調べなければならないが省略する。)

$$\int_{\Omega} \varphi(\mathbf{X}(\omega)) dP(\omega) = \sum_{k=1}^N a_k P(\mathbf{X} \in A_k). \quad (4.1)$$

2nd step $\varphi(x) \geq 0$ のとき。各 $q \in \mathbf{N}$ に対して、

$$A_{q,i} = \left\{ \mathbf{x} \in \mathbf{R}^n; \frac{i}{2^q} \leq \varphi(\mathbf{x}) < \frac{i+1}{2^q} \right\} \quad (i = 0, 1, \dots, q2^q - 1), \quad A_{q,q2^q} = \left\{ \mathbf{x} \in \mathbf{R}^n; \varphi(\mathbf{x}) \geq q \right\}$$

とし、 $\varphi_q(\mathbf{x}) = \sum_{i=0}^{q^{2^q}} \frac{i}{2^q} 1_{A_{q,i}}(\mathbf{x})$ と定める。このとき、 $\varphi_q(\mathbf{x})$ は単関数で (4.1) により $\int_{\Omega} \varphi_q(\mathbf{X}(\omega)) dP(\omega)$ は定義されるが、 $A_{q,i} = A_{q+1,2i} \cup A_{q+2,2i+1}$, $A_{q+1,2i} \cap A_{q+2,2i+1} = \emptyset$ ($i = 0, 1, \dots, q^{2^q} - 1$) となるので、

$$\frac{i}{2^q} P(\mathbf{X} \in A_{q,i}) \leq \frac{2i}{2^{q+1}} P(\mathbf{X} \in A_{q+1,2i}) + \frac{2i+1}{2^{q+1}} P(\mathbf{X} \in A_{q+1,2i+1})$$

となり、 $\int_{\Omega} \varphi_q(\mathbf{X}(\omega)) dP(\omega) \leq \int_{\Omega} \varphi_{q+1}(\mathbf{X}(\omega)) dP(\omega)$ を得る。よって、 $\{\int_{\Omega} \varphi_q(\mathbf{X}(\omega)) dP(\omega)\}$ は単調増加列なので、無限大かもしれないがその $q \rightarrow \infty$ とした極限は定義される。これを用いて、次のように定める。

$$\int_{\Omega} \varphi(\mathbf{X}(\omega)) dP(\omega) = \lim_{q \rightarrow \infty} \int_{\Omega} \varphi_q(\mathbf{X}(\omega)) dP(\omega). \quad (4.2)$$

3rd step 一般の $\varphi(\mathbf{x})$ について。

$$\varphi^+(\mathbf{x}) = \max\{\varphi(\mathbf{x}), 0\}, \quad \varphi^-(\mathbf{x}) = \max\{-\varphi(\mathbf{x}), 0\} \quad (4.3)$$

とおく。このとき、 $\varphi(\mathbf{x}) = \varphi^+(\mathbf{x}) - \varphi^-(\mathbf{x})$, $|\varphi(\mathbf{x})| = \varphi^+(\mathbf{x}) + \varphi^-(\mathbf{x})$ となることに注意する。これを用いて、 $\int_{\Omega} |\varphi(\mathbf{X}(\omega))| dP(\omega) < \infty$ のとき (この判定は 2nd step より可能)、各 $* = +, -$ に対して 2nd step の定義と $0 \leq \varphi^*(\mathbf{x}) \leq |\varphi(\mathbf{x})|$ より $0 \leq \int_{\Omega} \varphi^*(\mathbf{X}(\omega)) dP(\omega) \leq \int_{\Omega} |\varphi(\mathbf{X}(\omega))| dP(\omega) < \infty$ となることに注意して

$$\int_{\Omega} \varphi(\mathbf{X}(\omega)) dP(\omega) = \int_{\Omega} \varphi^+(\mathbf{X}(\omega)) dP(\omega) - \int_{\Omega} \varphi^-(\mathbf{X}(\omega)) dP(\omega) \quad (4.4)$$

と定める。

注意 4.1 上記の定義は、 $\mu(B) = P(\mathbf{X} \in B)$ ($B \in \mathcal{B}(\mathbf{R}^n)$) によって定義される確率測度 μ (\mathbf{X} の分布という) に対して、 $\int_{\mathbf{R}^n} \varphi(\mathbf{x}) d\mu(\mathbf{x})$ を定義したことに相当する。特に、 $n = 1$ のときこの積分を確率変数 X の分布関数 $F(x)$ を用いて、 $\int_{-\infty}^{\infty} \varphi(x) dF(x)$ と表すことがある。

以下の Lebesgue 積分論の定理 4.1, 4.2 が成立することに注意する。(証明は関数解析学で勉強のこと。)

定理 4.1 (1) 積分は線形性や単調性を有する。(期待値の形で定理 4.3 で述べる。)

(2) $\varphi(\mathbf{x}) \geq 0$, $\int_{\Omega} \varphi(\mathbf{X}(\omega)) dP(\omega) = 0 \implies P(\varphi(\mathbf{X}) = 0) = 1$.

注意 4.2 $P(\varphi(\mathbf{X}) = 0) = 1$ のとき、ほとんどいたるところ (*almost everywhere*) $\varphi(\mathbf{X}) = 0$ といい、「 $\varphi(\mathbf{X}) = 0$ a.e.」と表す。(ほとんど確かに (*almost surely*) といい、「 $\varphi(\mathbf{X}) = 0$ a.s.」と表すこともある。)

定理 4.2 (1) (単調収束定理) (R, \mathfrak{A}, μ) を測度空間とする。可測関数の列 $\{f_q(x)\}$ が非負値で単調増加であるとする、 $0 \leq f_1(x) \leq \dots \leq f_q(x) \leq \dots$ a.e. このとき、次が成立する。

$$\lim_{q \rightarrow \infty} \int_R f_q(x) d\mu(x) = \int_R \left(\lim_{q \rightarrow \infty} f_q(x) \right) d\mu(x).$$

(2) (Lebesgue の収束定理) (R, \mathfrak{A}, μ) を測度空間とする。可測関数の列 $\{f_q(x)\}$ に対して、ある可測関数 $M(x)$ で $\int_R M(x) d\mu(x) < \infty$ を満たすもの (可積分という) が存在して、 $|f_q(x)| \leq M(x)$ a.e. ($q = 1, 2, \dots$) を満たすとするとする。このとき、 $f(x) = \lim_{q \rightarrow \infty} f_q(x)$ a.e. が存在すれば、 $f(x)$ も可積分であり、次が成立する。

$$\lim_{q \rightarrow \infty} \int_R f_q(x) d\mu(x) = \int_R f(x) d\mu(x).$$

4.2 期待値の定義

§4.1 で準備した積分を用いて、 $\varphi \geq 0$ あるいは $\int_{\Omega} |\varphi(\mathbf{X}(\omega))| dP(\omega) < \infty$ のとき、期待値 $E[\varphi(\mathbf{X})]$ を

$$E[\varphi(\mathbf{X})] = \int_{\Omega} \varphi(\mathbf{X}(\omega)) dP(\omega)$$

で定義する。次の定理が成立する。証明は省略する。

定理 4.3 $\varphi, \psi : \mathbf{R}^n \rightarrow \mathbf{R}$ を Borel 可測関数、 a, b, c を定数とする。このとき、次が成立する。

- (1) $\varphi(\mathbf{X}) = c$ (定数) a.e. のとき、 $E[\varphi(\mathbf{X})] = c$.
- (2) $\varphi(\mathbf{X}) \leq \psi(\mathbf{X})$ a.e. のとき、 $E[\varphi(\mathbf{X})] \leq E[\psi(\mathbf{X})]$.
- (3) $|E[\varphi(\mathbf{X})]| \leq E[|\varphi(\mathbf{X})|]$.
- (4) $E[|\varphi(\mathbf{X})|] < \infty, E[|\psi(\mathbf{X})|] < \infty$ のとき、 $E[a\varphi(\mathbf{X}) + b\psi(\mathbf{X})] = aE[\varphi(\mathbf{X})] + bE[\psi(\mathbf{X})]$.

次に、 \mathbf{X} が離散型、絶対連続型のときの計算方法に関する定理を与える。 $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}$ を Borel 可測関数とする。

定理 4.4 \mathbf{X} が離散型確率変数のとき、そのとりうる値を x_1, x_2, \dots とすれば、

$$E[\varphi(\mathbf{X})] = \sum_{k=1}^{\infty} \varphi(x_k) P(\mathbf{X} = x_k). \quad (4.5)$$

証明: $\varphi \geq 0$ とする。§4.1 の 2nd step のように $\varphi_q(x)$ を定義すると、

$$\begin{aligned} E[\varphi_q(\mathbf{X})] &= \sum_{i=0}^{q2^q} \frac{i}{2^q} P(\mathbf{X} \in A_{q,i}) = \sum_{i=0}^{q2^q} \frac{i}{2^q} \sum_{k:\mathbf{x}_k \in A_{q,i}} P(\mathbf{X} = \mathbf{x}_k) \\ &= \sum_{i=0}^{q2^q} \sum_{k:\mathbf{x}_k \in A_{q,i}} \varphi_q(\mathbf{x}_k) P(\mathbf{X} = \mathbf{x}_k) = \sum_{k=1}^{\infty} \varphi_q(\mathbf{x}_k) P(\mathbf{X} = \mathbf{x}_k) \end{aligned}$$

ここで、 $\{\varphi_q\}$ は非負値単調増加列なので、 $\lim_{q \rightarrow \infty} \varphi_q(x) = \varphi(x)$ に注意して単調収束定理 (個数測度として) を用いれば (4.5) を得る。一般の場合は、(4.4) として定義するのだから、 $\varphi \geq 0$ の場合より成立する。 \square

定理 4.5 \mathbf{X} が絶対連続型確率変数のとき、その密度関数を $f(x_1, \dots, x_p)$ とすれば、

$$E[\varphi(\mathbf{X})] = \int \cdots \int_{\mathbf{R}^p} \varphi(x_1, \dots, x_p) f(x_1, \dots, x_p) dx_1 \cdots dx_p. \quad (4.6)$$

証明: 記号を簡単にするため 1 次元確率変数として記す。定理 4.4 の証明と同様に $\varphi \geq 0$ とし、§4.1 の 2nd step の $\varphi_q(x)$ を用いると、

$$\begin{aligned} E[\varphi_q(X)] &= \sum_{i=0}^{q2^q} \frac{i}{2^q} P(X \in A_{q,i}) = \sum_{i=0}^{q2^q} \frac{i}{2^q} \int_{A_{q,i}} f(x) dx \\ &= \sum_{i=0}^{q2^q} \int_{A_{q,i}} \varphi_q(x) f(x) dx = \int_{\mathbf{R}} \varphi_q(x) f(x) dx \end{aligned}$$

ここで、 $\{\varphi_q(x) f(x)\}$ は非負値単調増加列なので、単調収束定理を用いて (4.6) を得る。一般の φ に対しては、(i) と同様に (4.4) として定義するのだから明らか。 \square

確率変数の期待値と独立性の関係で、次は重要である。

定理 4.6 確率変数 X_1, \dots, X_n は独立であるとし、 $\varphi_1, \dots, \varphi_n$ は Borel 可測であるとする。このとき、もし、任意の j で $E[|\varphi_j(X_j)|] < \infty$ であれば、 $E[|\varphi_1(X_1)\varphi_2(X_2)\cdots\varphi_n(X_n)|] < \infty$ であり、次が成立する。

$$E[\varphi_1(X_1)\varphi_2(X_2)\cdots\varphi_n(X_n)] = E[\varphi_1(X_1)]E[\varphi_2(X_2)]\cdots E[\varphi_n(X_n)]. \quad (4.7)$$

証明: 簡単のため $n = 2$ の場合に証明する。 $X_1 = X, X_2 = Y, \varphi_1 = \varphi, \varphi_2 = \psi$ と記す。

1st step $\varphi(x), \psi(y)$ が単関数 $\varphi(x) = \sum_{i=1}^M a_i 1_{A_i}(x), \psi(y) = \sum_{j=1}^N b_j 1_{B_j}(y)$ のとき。 $1_A(x)1_B(y)$ も $x \in A$ かつ $y \in B$ のときのみ 1 になる定義関数だから、 $\varphi(x)\psi(y) = \sum_{i=1}^M \sum_{j=1}^N a_i b_j 1_{A_i}(x)1_{B_j}(y)$ も単関数で、独立性の定義 2.8 とそのあとの注意により、

$$\begin{aligned} E[\varphi(X)\psi(Y)] &= \sum_{i=1}^M \sum_{j=1}^N a_i b_j P(X \in A_i, Y \in B_j) = \sum_{i=1}^M \sum_{j=1}^N a_i b_j P(X \in A_i)P(Y \in B_j) \\ &= \sum_{i=1}^M a_i P(X \in A_i) \sum_{j=1}^N b_j P(Y \in B_j) = E[\varphi(X)]E[\psi(Y)] \end{aligned}$$

2nd step $\varphi \geq 0, \psi \geq 0$ のとき。 §4.1 の 2nd step と同様に、各 $n \in \mathbb{N}$ に対して単関数 φ_q, ψ_q を定義する。このとき、1st step により

$$E[\varphi_q(X)\psi_q(Y)] = E[\varphi_q(X)]E[\psi_q(Y)]$$

となる。ここで、単調収束定理 (定理 4.2 (1)) により $q \rightarrow \infty$ とすることで (4.7) を得る。特に、 $E[|\varphi(X)|] < \infty$ かつ $E[|\psi(Y)|] < \infty$ であれば、 $E[|\varphi(X)\psi(Y)|] = E[|\varphi(X)|]E[|\psi(Y)|] < \infty$ となることもわかった。

3rd step 一般の φ, ψ の場合。 $(\varphi\psi)(x, y) = \varphi(x)\psi(y)$ と書くものとする。このとき、(4.3) のように、 $\varphi^+, \varphi^-, \psi^+, \psi^-, (\varphi\psi)^+, (\varphi\psi)^-$ を定める。このとき、

$$(\varphi\psi)^+(x, y) = \varphi^+(x)\psi^+(y) + \varphi^-(x)\psi^-(y), \quad (\varphi\psi)^-(x, y) = \varphi^+(x)\psi^-(y) + \varphi^-(x)\psi^+(y)$$

に注意する。これより、

$$\begin{aligned} E[\varphi(X)]E[\psi(Y)] &= (E[\varphi^+(X)] - E[\varphi^-(X)])(E[\psi^+(Y)] - E[\psi^-(Y)]) \\ &= (E[\varphi^+(X)]E[\psi^+(Y)] + E[\varphi^-(X)]E[\psi^-(Y)]) - (E[\varphi^+(X)]E[\psi^-(Y)] + E[\varphi^-(X)]E[\psi^+(Y)]) \\ &= (E[\varphi^+(X)\psi^+(Y)] + E[\varphi^-(X)\psi^-(Y)]) - (E[\varphi^+(X)\psi^-(Y)] + E[\varphi^-(X)\psi^+(Y)]) \\ &= E[(\varphi\psi)^+(X, Y)] - E[(\varphi\psi)^-(X, Y)] = E[\varphi(X)\psi(Y)] \quad \square \end{aligned}$$

4.3 積率 (モーメント) ・ 分散

定義 4.1 X を確率変数、 $r > 0$ とする。

(1) $E[X^r]$ を r 次の積率 (モーメント) という。

特に、 $r = 1$ のとき、 $E[X]$ を X の期待値もしくは平均という。

(2) $\mu = E[X]$ とするとき、 $E[(X - \mu)^r]$ を期待値のまわりの r 次の積率という。

特に、 $r = 2$ のとき、 $E[(X - \mu)^2]$ を X の分散といい $V(X)$ と表す。また、 $\sigma(X) = V(X)^{1/2}$ を X の標準偏差という。

(3) $m_X(t) = E[e^{tX}], t \in \mathbb{R}$, を X の積率母関数という。 $\psi_X(t) = \log E[e^{tX}], t \in \mathbb{R}$, を X のキュムラント母関数という。

ただし、上記はそれぞれの期待値が定義される場合にのみ考えるものとする。

定理 4.7 X の r 次の積率が存在するとき、 $s (\leq r)$ 次の積率も存在する。

証明: $|X|^s \leq \max\{1, |X|^r\} \leq 1 + |X|^r$ であるから、 $E[|X|^s] \leq 1 + E[|X|^r] < \infty$ となり存在する。 \square

この定理により、 $E[X^2] < \infty$ のとき、平均 $E[X]$ も存在する。また、このとき $(X - c)^2 = X^2 - 2cX + c^2$ となることから、分散 $V(X)$ が存在する。

定理 4.8 $E[X^2] < \infty$ とする。このとき次が成り立つ。

- (1) $V(X) = E[X^2] - E[X]^2$.
- (2) $V(aX + b) = a^2V(X)$ (a, b は定数).
- (3) $V(X) = 0$ ならば $X = \mu$ a.e., ただし $\mu = E[X]$.

証明: (1), (2) は演習問題とする。(3) は定理 4.1 (2), 注意 4.2 から自明。 \square

例 4.1 重要な分布の期待値と分散を例示する。

[A] 離散型確率変数の例

(1) 二項分布 $B(n, p)$

$P(X = k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \dots, n$, ($q = 1 - p$) より

$$\begin{aligned} E[X] &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1} q^{n-1-(k-1)} \\ &= np(p+q)^{n-1} = np. \end{aligned}$$

また、

$$\begin{aligned} E[X(X-1)] &= \sum_{k=0}^n k(k-1) \frac{n!}{k!(n-k)!} p^k q^{n-k} = n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-2-(k-2))!} p^{k-2} q^{n-2-(k-2)} \\ &= n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2. \end{aligned}$$

となるから、 $E[X^2] = E[X(X-1) + X] = E[X(X-1)] + E[X] = n(n-1)p^2 + np$. よって、 $V(X) = E[X^2] - E[X]^2 = np(1-p)$. \square

(2) Poisson 分布 $Po(\lambda)$

$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \dots$ より

$$E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

また、二項分布と同様にして、

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2.$$

となる。よって、 $V(X) = E[X(X-1)] + E[X] - E[X]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

(3) 幾何分布

$P(X = k) = q^k p$, $k = 0, 1, \dots$, ($q = 1 - p$). まず、級数 $\sum_{k=0}^{\infty} q^k$ の収束半径は 1 であったから、 $|q| < 1$ のときこの級数は何度でも微分でき、

$$\begin{aligned} \sum_{k=1}^{\infty} k q^{k-1} &= \frac{d}{dq} \left(\sum_{k=0}^{\infty} q^k \right) = \frac{d}{dq} \left(\frac{1}{1-q} \right) = \frac{1}{(1-q)^2} \\ \sum_{k=2}^{\infty} k(k-1) q^{k-2} &= \frac{d}{dq} \left(\sum_{k=1}^{\infty} k q^{k-1} \right) = \frac{d}{dq} \left(\frac{1}{(1-q)^2} \right) = \frac{2}{(1-q)^3} \end{aligned}$$

よって、

$$E[X] = \sum_{k=1}^{\infty} kq^k p = qp \sum_{k=1}^{\infty} kq^{k-1} = \frac{pq}{(1-q)^2} = \frac{q}{p},$$

$$V(X) = E[X(X-1)] + E[X] - E[X]^2 = \sum_{k=2}^{\infty} k(k-1)q^k p + E[X] - E[X]^2$$

$$= \frac{q^2 p}{(1-q)^3} + \frac{q}{p} - \left(\frac{q}{p}\right)^2 = \frac{q}{p}.$$

(4) 負の二項分布 $NB(\alpha, p)$ ($\alpha > 0, 0 < p < 1$)

$$P(X = n) = \binom{\alpha + n - 1}{n} p^\alpha (1-p)^n, \quad n = 0, 1, 2, \dots$$

$$\text{まず、} \binom{\alpha + n - 1}{n} = \frac{(\alpha + n - 1)(\alpha + n - 2) \cdots \alpha}{n!} = (-1)^n \frac{(-\alpha)(-\alpha - 1) \cdots (-\alpha - n + 1)}{n!} = (-1)^n \binom{-\alpha}{n}$$

と $(1+x)^{-\alpha} = \sum_{n=0}^{\infty} \binom{-\alpha}{n} x^n$ ($|x| < 1$) より、 $\sum_{n=0}^{\infty} P(X = n) = 1$ となることに注意する。(1)-(3) と同様に、

$$E[X] = \sum_{n=1}^{\infty} n \binom{\alpha + n - 1}{n} p^\alpha (1-p)^n = \sum_{n=1}^{\infty} (-1)^n n \binom{-\alpha}{n} p^\alpha (1-p)^n = p^\alpha \sum_{n=1}^{\infty} n \binom{-\alpha}{n} (p-1)^n$$

$$= p^\alpha (-\alpha)(p-1) \sum_{n=1}^{\infty} \binom{-\alpha-1}{n-1} (p-1)^{n-1} = p^\alpha (-\alpha)(p-1)(1+p-1)^{-\alpha-1} = \frac{\alpha(1-p)}{p}.$$

同様に、 $n(n-1) \binom{-\alpha}{n} = (-\alpha)(-\alpha-1) \binom{-\alpha-2}{n-2}$ より $E[X(X-1)]$ が求まり、 $V(X) = \frac{\alpha(1-p)}{p^2}$ となる。

(5) 期待値が存在しない例

とり得る値を $x_k = (-1)^{k-1} k$, $k = 1, 2, \dots$, その確率を $P(X = (-1)^{k-1} k) = \frac{1}{k(k+1)}$, $k = 1, 2, \dots$ とすると、

$$\sum_{k=1}^{\infty} P(X = (-1)^{k-1} k) = \sum_{k=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) = 1$$

となり、 X は確かに確率変数となるが、

$$E[|X|] = \sum_{k=1}^{\infty} k \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty$$

となり期待値は存在しない。この場合、

$$\sum_{k=1}^{\infty} (-1)^{k-1} k \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k+1} = 1 - \sum_{l=1}^{\infty} \frac{(-1)^{l-1}}{l} = 1 - \log 2$$

と条件収束し、一見期待値が存在しそうだが、この場合は期待値が存在しないとする。

[B] 絶対連続型確率変数の例

(1) 正規分布 $N(\mu, \sigma^2)$

まず、 $\int_0^{\infty} ze^{-z^2/2} dz = \int_0^{\infty} (-e^{-z^2/2})' dz = 1$ に注意する。これより、

$$E[|X|] = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |\sigma z + \mu| e^{-\frac{z^2}{2}} dz$$

$$\leq \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z| e^{-\frac{z^2}{2}} dz + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \frac{2\sigma}{\sqrt{2\pi}} + |\mu| < \infty.$$

ただし、 $z = \frac{x-\mu}{\sigma}$ と変換した。よって、期待値 $E[X]$ は存在して、

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-\frac{z^2}{2}} dz \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \mu. \end{aligned}$$

同様に、

$$\begin{aligned} V(X) &= E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z (-e^{-\frac{z^2}{2}})' dz = \frac{\sigma^2}{\sqrt{2\pi}} [-ze^{-\frac{z^2}{2}}]_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sigma^2. \end{aligned}$$

(2) 一様分布 $U(a, b)$

$$E[X] = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

$$V(X) = E[X^2] - E[X]^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}.$$

(3) 指数分布 $Ex(\lambda)$

$$E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx = [-xe^{-\lambda x}]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}.$$

$$V(X) = E[X^2] - E[X]^2 = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

(4) Cauchy 分布

X が Cauchy 分布に従うとは、その密度関数が $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ ($-\infty < x < \infty$) ととなるときにいう。実際、次に注意する。

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} [\arctan x]_{-\infty}^{\infty} = \frac{1}{\pi} \left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right) \right] = 1$$

このとき、

$$\int_{-\infty}^{\infty} |x| \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx = \frac{2}{\pi} \left[\frac{1}{2} \log(1+x^2) \right]_0^{\infty} = \infty$$

となるので、Cauchy 分布の平均は存在しない。

4.4 共分散と相関係数

以下、 $L^2(\Omega)$ で、確率変数 X で $E[X^2] < \infty$ を満たすもの全体を表すとする。

定理 4.9 (Cauchy-Schwarz の不等式) $X, Y \in L^2(\Omega)$ のとき、 $E[XY]$ は定義され、

$$|E[XY]|^2 \leq E[X^2]E[Y^2] \quad (4.8)$$

が成立する。等号成立のための必要十分条件は、 $(a, b) \neq (0, 0)$ なる定数 a, b が存在して $aX + bY = 0$ a.e. となることである。

証明: $E[X^2] = 0$, 即ち、 $X = 0$ a.e. のとき、(4.8) は $0 = 0$ として明らかに成立し、また、 $a = 1, b = 0$ として $aX + bY = 0$ a.e. も成立する。 $E[X^2] > 0$ のとき、

$$0 \leq E[(tX + Y)^2] = t^2 E[X^2] + 2tE[XY] + E[Y^2] = E[X^2] \left\{ t + \frac{E[XY]}{E[X^2]} \right\}^2 - \frac{(E[XY])^2}{E[X^2]} + E[Y^2] \quad (4.9)$$

が任意の t に対して成立するので、(4.8) は成立する。一方、(4.8) で等号が成立するとき、(4.9) で $t = -E[XY]/E[X^2]$ とれば $E[(tX+Y)^2] = 0$ となる。よって、 $tX+Y = 0$ a.e. を得る。逆に、 $aX+bY = 0$ a.e. とすると、 $P(X \neq 0) > 0$ より $b \neq 0$ となることに注意する。このとき、 $Y = aX/b$ a.e. より

$$|E[XY]|^2 = \left(\frac{a}{b}\right)^2 E[X^2]^2, \quad E[X^2]E[Y^2] = E[X^2]\left(\frac{a}{b}\right)^2 E[X^2]$$

となるので、(4.8) の等号が成立する。 \square

系 4.1 $X, Y \in L^2(\Omega)$ のとき $E[(X - \mu_1)(X - \mu_2)]$ は定義され、

$$|E[(X - \mu_1)(X - \mu_2)]|^2 \leq V(X)V(Y) \quad (4.10)$$

が成立する。ただし、 $E[X] = \mu_1$, $E[Y] = \mu_2$ とした。(4.10) で等号成立のための必要十分条件は、 $(a, b) \neq (0, 0)$ なる定数 a, b が存在して $a(X - \mu_1) + b(Y - \mu_2) = 0$ a.e. となることである。

証明: 定理 4.9 の X, Y を $|X - \mu_1|, |Y - \mu_2|$ に置き換えると、

$$E[|(X - \mu_1)(Y - \mu_2)|^2] \leq E[(X - \mu_1)^2]E[(Y - \mu_2)^2] < \infty$$

より、 $E[(X - \mu_1)(Y - \mu_2)]$ は定義される。他は、定理 4.9 の X, Y を $X - \mu_1, Y - \mu_2$ に置き換えればよい。 \square

定義 4.2 $X, Y \in L^2(\Omega)$ なる確率変数 X, Y の共分散 $\text{Cov}(X, Y)$ 、相関係数 $\rho(X, Y)$ を次で定義する。

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - \mu_1)(Y - \mu_2)] = E[XY] - \mu_1 E[X] - \mu_2 E[Y] + \mu_1 \mu_2 = E[XY] - \mu_1 \mu_2, \\ \rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} \end{aligned}$$

ただし、 $\mu_1 = E[X], \mu_2 = E[Y]$ とし、相関係数は $V(X)V(Y) > 0$ のときのみ定義されるものとする。

定理 4.10 相関係数 $\rho(X, Y)$ について次が成立する。

- (i) $-1 \leq \rho(X, Y) \leq 1$.
- (ii) $\rho(X, Y) = 1$ のとき、 $\frac{X - \mu_1}{\sqrt{V(X)}} = \frac{Y - \mu_2}{\sqrt{V(Y)}}$ a.e.
- (iii) $\rho(X, Y) = -1$ のとき、 $\frac{X - \mu_1}{\sqrt{V(X)}} = -\frac{Y - \mu_2}{\sqrt{V(Y)}}$ a.e.

証明: (i) 定義と系 4.1 より明らか。

(ii), (iii) 仮定と系 4.1 により $(a, b) \neq (0, 0)$ があって $a(X - \mu_1) + b(Y - \mu_2) = 0$ a.e. となる。従って、 $a^2(X - \mu_1)^2 = b^2(Y - \mu_2)^2$ の期待値をとって、 $a^2V(X) = b^2V(Y)$ 。ここで、 $V(X)V(Y) > 0$ より a, b はともに 0 ではないことがわかる。(一方が 0 なら双方とも 0 となるため。) よって、 $Y - \mu_2 = -a(X - \mu_1)/b$ a.e. より、

$$V(Y) = \left(\frac{a}{b}\right)^2 V(X), \quad \rho(X, Y) = \frac{E\left[-\frac{a}{b}(X - \mu_1)^2\right]}{V(X)^{1/2}\left|\frac{a}{b}\right|V(X)^{1/2}} = -\frac{\frac{a}{b}}{\left|\frac{a}{b}\right|}.$$

よって、 $\rho(X, Y) = 1$ なら、 $\frac{a}{b} < 0$ で $\frac{a}{b} = -\frac{\sqrt{V(Y)}}{\sqrt{V(X)}}$, $\rho(X, Y) = -1$ なら、 $\frac{a}{b} > 0$ で $\frac{a}{b} = -\frac{\sqrt{V(Y)}}{\sqrt{V(X)}}$ となり主張を得る。 \square

$|\rho(X, Y)| = 1$ の X と Y は完全相関であるという。 $\rho(X, Y) > 0$ のとき正の相関、 $\rho(X, Y) < 0$ のとき負の相関という。また、 $\rho(X, Y) = 0$ のとき X と Y は無相関という。

例題 4.1 同時密度関数が $f(x, y) = \begin{cases} 24xy & (0 \leq x, 0 \leq y, x + y \leq 1) \\ 0 & (\text{その他}) \end{cases}$ である確率変数 X, Y について、その共分散 $\text{Cov}(X, Y)$ と相関係数 $\rho(X, Y)$ を求める。

解 $f_X(x), f_Y(y)$ を X, Y の周辺密度関数とする。 $0 \leq x \leq 1$ のとき、

$$f_X(x) = \int_0^{1-x} 24xy \, dy = 12x(1-x)^2$$

より、 $f_X(x) = \begin{cases} 12x(1-x)^2 & (0 \leq x \leq 1) \\ 0 & (\text{その他}) \end{cases}$. 全く同様に、 $f_Y(y) = \begin{cases} 12y(1-y)^2 & (0 \leq y \leq 1) \\ 0 & (\text{その他}) \end{cases}$ である。よって、補題 3.1 を用いて計算すると、

$$E[X] = \int_0^1 x \cdot 12x(1-x)^2 \, dx = 12 \int_0^1 x^2(1-x)^2 \, dx = 12 \text{B}(3, 3) = 12 \frac{\Gamma(3)\Gamma(3)}{\Gamma(6)} = 12 \frac{2 \cdot 2}{5!} = \frac{2}{5} (= E[Y])$$

$$E[X^2] = \int_0^1 x^2 \cdot 12x(1-x)^2 \, dx = 12 \int_0^1 x^3(1-x)^2 \, dx = 12 \text{B}(4, 3) = 12 \frac{\Gamma(4)\Gamma(3)}{\Gamma(7)} = 12 \frac{3! \cdot 2!}{6!} = \frac{1}{5} \text{ より}$$

$$V(X) = E[X^2] - E[X]^2 = \frac{1}{5} - \left(\frac{2}{5}\right)^2 = \frac{1}{25} (= V(Y))$$

一方 $D : 0 \leq y \leq 1-x, 0 \leq x \leq 1$ とすると、

$$\begin{aligned} E[XY] &= \iint_D xy \cdot 24xy \, dx dy = \int_0^1 dx \int_0^{1-x} 24x^2y^2 \, dy = \int_0^1 [8x^2y^3]_{y=0}^{y=1-x} \, dx \\ &= \int_0^1 8x^2(1-x)^3 \, dx = 8 \text{B}(3, 4) = \dots = \frac{2}{15} \text{ より} \end{aligned}$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{2}{15} - \left(\frac{2}{5}\right)^2 = -\frac{2}{75}$$

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{-2/75}{1/25} = -\frac{2}{3}. \quad \square$$

定理 4.11 確率変数 X, Y について、次の (i)–(iv) は同値である。

(i) $\text{Cov}(X, Y) = 0$.

(ii) X と Y は無相関である。

(iii) $E[XY] = E[X]E[Y]$.

(iv) $V(X + Y) = V(X) + V(Y)$.

証明: (i) \iff (ii): 定義より明らか。

(i) \iff (iii): $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$ よりわかる。

(i) \iff (iv): $V(X + Y) = E[(X - E[X] + Y - E[Y])^2] = V(X) + V(Y) + 2\text{Cov}(X, Y)$ より従う。 \square

注意 4.3 X, Y が独立であれば、無相関である。実際、 X, Y が独立であれば、定理 4.6 より (iii) が成立する。よって、定理 4.11 より無相関となる。

逆は成立しない。例えば、 $0 < q < 1$ とし、 $P((X, Y) = (0, 1)) = q, P((X, Y) = (1, 0)) = P((X, Y) = (-1, 0)) = (1 - q)/2$ となる確率変数 X, Y を考えると、

$$P(X = 0) = P(Y = 1) = P((X, Y) = (0, 1)) = q$$

より $P(X = 0, Y = 1) = P((X, Y) = (0, 1)) \neq P(X = 0)P(Y = 1)$ となり X と Y は独立ではない。

一方、 $P(XY = 0) = P((X, Y) = (0, 1)) + P((X, Y) = (1, 0)) + P((X, Y) = (-1, 0)) = 1$ より $E[XY] = 0, P(X = 1) = P((X, Y) = (1, 0)) = (1 - q)/2, P(X = -1) = P((X, Y) = (-1, 0)) = (1 - q)/2$ より

$E[X] = 1 \cdot P(X = 1) + (-1) \cdot P(X = -1) + 0 \cdot P(X = 0) = 0$ となり、 $E[XY] = E[X]E[Y]$ が成立するので $\text{Cov}(X, Y) = \rho(X, Y) = 0$.

ただし、 (X, Y) が 2 次元正規分布であれば X, Y が独立であることと、無相関であることが同値となる (cf. 例 2.5, 定理 4.12, 系 4.2)。

例 4.2 X_1, X_2, \dots, X_n は独立で各 $k = 1, \dots, n$ に対して $E[X_k] = \mu, V(X_k) = \sigma^2$ とする。このとき、 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ (標本平均), $U^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$ (不偏分散) とおくと、 $E[\bar{X}] = \mu, E[U^2] = \sigma^2$ となる。

証明 $E[\bar{X}] = \frac{1}{n} \sum_{k=1}^n E[X_k] = \mu$. $Y_k = X_k - \mu$ とおくと、 $X_1 - \bar{X} = \frac{n-1}{n} Y_1 - \frac{1}{n} \sum_{k=2}^n Y_k$. 今、 $E[Y_k^2] = V(X_k) = \sigma^2, E[Y_k Y_l] = \text{Cov}(X_k, X_l) = 0$ ($k \neq l$) だから、

$$E[(X_1 - \bar{X})^2] = \frac{(n-1)^2}{n^2} E[Y_1^2] + \frac{1}{n^2} \sum_{k=2}^n E[Y_k^2] = \frac{n-1}{n} \sigma^2.$$

同様に、各 k について $E[(X_k - \bar{X})^2] = \frac{n-1}{n} \sigma^2$ となるので、 $E[U^2] = \frac{1}{n-1} \sum_{k=1}^n \frac{n-1}{n} \sigma^2 = \sigma^2$. \square

定義 4.3 p 次元確率変数 $\mathbf{X} = (X_1, \dots, X_p)$ について、ベクトル $E[\mathbf{X}] = (E[X_1], \dots, E[X_p])$ を \mathbf{X} の平均ベクトルという。また、 $\text{Cov}(X_i, X_j)$ をその (i, j) 成分とする $p \times p$ 対称行列を、 \mathbf{X} の共分散行列といい、 $V(\mathbf{X})$ で表す。

定理 4.12 $\mathbf{X} = (X_1, \dots, X_p)$ を p 次元正規分布 $N(\boldsymbol{\mu}', \boldsymbol{\Sigma})$ に従うとする。 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p) \in \mathbf{R}^p, \boldsymbol{\Sigma} = (\sigma_{ij})$ は正定値対称行列であった。このとき、 $E[\mathbf{X}] = \boldsymbol{\mu}, V(\mathbf{X}) = \boldsymbol{\Sigma}$ となる。

証明: $E[X_m] = \mu_m, 1 \leq m \leq p, \text{Cov}(X_m, X_n) = \sigma_{mn}, 1 \leq m, n \leq p$ を示せばよい。まず、 $\lambda > 0$ に対して

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda)^{1/2}} e^{-\frac{1}{2} \frac{y^2}{\lambda}} dy = 1, \quad \int_{-\infty}^{\infty} y \frac{1}{(2\pi\lambda)^{1/2}} e^{-\frac{1}{2} \frac{y^2}{\lambda}} dy = 0, \quad \int_{-\infty}^{\infty} y^2 \frac{1}{(2\pi\lambda)^{1/2}} e^{-\frac{1}{2} \frac{y^2}{\lambda}} dy = \lambda \quad (4.11)$$

に注意する。例 2.3 (ii) の記号を用いる。 $\mathbf{X} = (X_1, \dots, X_p)$ の密度関数は

$$f(x_1, \dots, x_p) = \frac{1}{(2\pi)^{p/2} (\det \boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})'}, \quad \mathbf{x} = (x_1, \dots, x_p).$$

$\boldsymbol{\Sigma}$ は正定値対称行列なので、線形代数の対角化に関する定理から、直交行列 $P = (p_{ij})$ と対角成分がすべて正の対角行列 $D = (\lambda_{ij})$ がとれて、 $P' \boldsymbol{\Sigma} P = D$ とできる。ここで、 $\mathbf{y} = (y_1, \dots, y_p) = (\mathbf{x} - \boldsymbol{\mu}) P$ とすると、

$$(\mathbf{x} - \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})' = \mathbf{y} P' \boldsymbol{\Sigma}^{-1} P \mathbf{y}' = \mathbf{y} D^{-1} \mathbf{y}' = \sum_{j=1}^p \frac{1}{\lambda_{jj}} y_j^2, \quad x_m - \mu_m = \sum_{k=1}^p p_{mk} y_k$$

で、 $|\frac{\partial \mathbf{y}}{\partial \mathbf{x}}| = |\det P| = 1, \det \boldsymbol{\Sigma} = \det D = \prod_{j=1}^p \lambda_{jj}$. よって、

$$\begin{aligned} E[X_m - \mu_m] &= \int \cdots \int_{\mathbf{R}^p} (x_m - \mu_m) f(x_1, \dots, x_p) dx_1 \cdots dx_p \\ &= \sum_{k=1}^p p_{mk} \int \cdots \int_{\mathbf{R}^p} y_k \left[\prod_{j=1}^p \frac{1}{(2\pi\lambda_{jj})^{1/2}} e^{-\frac{1}{2} \frac{y_j^2}{\lambda_{jj}}} \right] dy_1 \cdots dy_p \\ &= \sum_{k=1}^p p_{mk} \int_{-\infty}^{\infty} y_k \frac{1}{(2\pi\lambda_{kk})^{1/2}} e^{-\frac{1}{2} \frac{y_k^2}{\lambda_{kk}}} dy_k \prod_{j \neq k} \int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda_{jj})^{1/2}} e^{-\frac{1}{2} \frac{y_j^2}{\lambda_{jj}}} dy_j. \end{aligned}$$

よって、(4.11) により $E[X_m - \mu_m] = 0$, すなわち、 $E[X_m] = \mu_m$. 次に、

$$\begin{aligned} \text{Cov}(X_m, X_n) &= \int \cdots \int_{\mathbf{R}^p} (x_m - \mu_m)(x_n - \mu_n) f(x_1, \dots, x_p) dx_1 \cdots dx_p \\ &= \sum_{k=1}^p \sum_{l=1}^p p_{mk} p_{nl} \int \cdots \int_{\mathbf{R}^p} y_k y_l \left[\prod_{j=1}^p \frac{1}{(2\pi\lambda_{jj})^{1/2}} e^{-\frac{1}{2} \frac{y_j^2}{\lambda_{jj}}} \right] dy_1 \cdots dy_p. \end{aligned}$$

ここで、(4.11) であるから、

$$\int \cdots \int_{\mathbf{R}^p} y_k y_l \left[\prod_{j=1}^p \frac{1}{(2\pi\lambda_{jj})^{1/2}} e^{-\frac{1}{2} \frac{y_j^2}{\lambda_{jj}}} \right] dy_1 \cdots dy_p = \begin{cases} \lambda_{kk} & (k = l) \\ 0 & (k \neq l) \end{cases}$$

よって、 $\text{Cov}(X_m, X_n) = \sum_{k=1}^p p_{mk} p_{nk} \lambda_{kk}$. 一方、 PD の (i, j) 成分は $\sum_{k=1}^p p_{ik} \lambda_{kj} = p_{ij} \lambda_{jj}$, PDP' の (m, n) 成分は $\sum_{k=1}^p p_{mk} \lambda_{kk} p_{nk}$. ここで、 $\Sigma = PDP'$ であったから、

$$\text{Cov}(X_m, X_n) = \sum_{k=1}^p p_{mk} p_{nk} \lambda_{kk} = \sigma_{mn}. \quad \square$$

系 4.2 $X = (X_1, \dots, X_p)'$ を p 次元正規分布 $N(\mu, \Sigma)$ に従うとする。このとき、 X_1, \dots, X_p が独立であるための必要十分条件は Σ が対角行列であることである。

証明: 必要性は例 2.5 で示した。十分性について、 $\Sigma = (\sigma_{ij})$ とかく。 $i \neq j$ ならば、 X_i, X_j は独立なので定理 4.11 により $\text{Cov}(X_i, X_j) = 0$. よって、定理 4.12 により $\sigma_{ij} = \text{Cov}(X_i, X_j) = 0$ となる。 \square

例 4.3 2次元確率変数 (X, Y) の同時密度関数が $f(x, y) = \frac{1}{\pi e} \exp\left\{-\frac{1}{2}Q(x, y)\right\}$, $(x, y) \in \mathbf{R}^2$, であるとす。ただし、

$$Q(x, y) = x^2 + 2xy + 5y^2 - 2x + 2y$$

とする。このとき、 $V(X)$ と $\text{Cov}(X, Y)$ を求めよ。

解: $Q(x, y) = (x + y - 1)^2 - (y - 1)^2 + 5y^2 + 2y = (x + y - 1)^2 + (2y + 1)^2 - 2$ に注意して、 $S = X + Y - 1$, $T = 2Y + 1$ とおく。このとき、 (S, T) の同時密度関数 $f_{(S, T)}(s, t)$ を求めると、 $s = x + y - 1, t = 2y + 1$ より $y = \frac{1}{2}(t - 1), x = s - \frac{1}{2}t + \frac{3}{2}$ に注意して、

$$f_{(S, T)}(s, t) = f\left(s - \frac{1}{2}t + \frac{3}{2}, \frac{1}{2}(t - 1)\right) \left| \frac{\partial(x, y)}{\partial(s, t)} \right| = \frac{1}{\pi e} \exp\left\{-\frac{1}{2}(s^2 + t^2 - 2)\right\} \frac{1}{2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}.$$

これは、 S, T は独立でともに標準正規分布 $N(0, 1)$ に従うことを示している。よって、

$$E[S] = E[T] = 0, \quad E[S^2] = V(S) = E[T^2] = V(T) = 1, \quad E[ST] = \text{Cov}(S, T) = 0$$

となる。これより、 $Y = \frac{1}{2}(T - 1), X = S - \frac{1}{2}T + \frac{3}{2}$ より、

$$\begin{aligned} E[Y] &= \frac{1}{2}(E[T] - 1) = -\frac{1}{2}, & E[X] &= E[S] - \frac{1}{2}E[T] + \frac{3}{2} = \frac{3}{2}, \\ V(X) &= E[(X - E[X])^2] = E\left[\left(S - \frac{1}{2}T + \frac{3}{2} - \frac{3}{2}\right)^2\right] = E[S^2] - 2 \cdot \frac{1}{2}E[ST] + \frac{1}{4}E[T^2] = \frac{5}{4}, \\ \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] = E\left[\left(S - \frac{1}{2}T\right)\frac{1}{2}T\right] = \frac{1}{2}E[ST] - \frac{1}{4}E[T^2] = -\frac{1}{4}. \quad \square \end{aligned}$$