第1章　確率・確率変数・期待値

1.1　確率

　不確実性や変動というものは多くの状況では避けることができないものである．確率論はこれらの概念を数学的に定式化する道具である．

1.1.1　試行・標本空間・事象

　実験とか観測とか調査とかを総称して**試行**という．試行を行ったときに起こり得る結果のすべての集まりを**標本空間**といい，Ω と記すことにする．標本空間 Ω の点を**標本点**とか**根元事象**という．Ω が有限または可算の場合を離散的という．

例 1.1 硬貨投げを考えよう．硬貨には表 (H) と裏 (T) があることから，試行のすべての集まりである標本空間は $\Omega = \{H, T\}$ となる．

　部分集合 $A \subset \Omega$ を**事象**といい，Ω を**全事象**，$\emptyset$ を**空事象**という．事象に対して，和，交わり (積)，直積，補集合，差などの演算を考える．

事象の記法　$\in$, $\notin$, $\emptyset$, $A \subset B$, $A \cap B$, $A \cup B$, A^c. ただし，A, B は事象とし，$A^c = \{\omega \in \Omega : \omega \notin A\}$, $A \cup B = \{\omega \in \Omega : \omega \in A \quad \text{または} \quad \omega \in B\}$ 等である． 和と交わりに対して

$$
\begin{aligned}
&A \cup A = A && A \cap A = A && \text{(ベキ等律)}\\
&A \cup B = B \cup A && A \cap B = B \cap A && \text{(交換律)}\\
&(A \cup B) \cup C = A \cup (B \cup C) && (A \cap B) \cap C = A \cap (B \cap C) && \text{(結合律)}
\end{aligned}
$$

が成り立つ．

命題 1.1 (事象の演算に関する基本的性質) A, B, C を事象とする．

(i)
$$
\begin{aligned}
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C)\\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C)
\end{aligned}
\qquad \text{(分配法則)}
$$

(ii)
$$
\begin{aligned}
(A \cup B)^c &= A^c \cap B^c\\
(A \cap B)^c &= A^c \cup B^c
\end{aligned}
\qquad \text{(ド・モルガンの法則)}
$$

証明　証明は略．(「集合・位相」(佐久間カ一浩著，共立) や「集合・位相入門」(松坂和夫著，岩波) 等を参照のこと)　□

　Γ を添え字の空間とし，$\gamma \in \Gamma$ の元によって添え字付けられる集合 A_γ の集まりを考え，これを

$$\{A_\gamma : \gamma \in \Gamma\}$$

と書く．任意の濃度を持つ添え字集合 Γ に対して

$$\begin{aligned}\cup_{\gamma\in\Gamma}A_\gamma &= \{\omega\in\Omega:\ \text{ある } \gamma\in\Gamma \text{ が存在して，} \omega\in A_\gamma\ \}\\ \cap_{\gamma\in\Gamma}A_\gamma &= \{\omega\in\Omega:\ \text{すべての } \gamma\in\Gamma \text{ に対して，} \omega\in A_\gamma\ \}\end{aligned}$$

とする．

例 1.2　たとえば，$\Omega=(0,1]$ とし，$A_i=[1/i,1]$, $i=1,2,\ldots$ とする．したがって，$\Gamma=\mathbb{N}$ (自然数) である．このとき，

$$\begin{aligned}\cup_{i=1}^\infty A_i &= \cup_{i=1}^n[1/i,1]\\ &= \{\omega\in(0,1]:\ \text{ある } i\in\mathbb{N} \text{ が存在して，} \omega\in[1/i,1]\ \}\\ &= \Omega\end{aligned}$$

定義 1.1　$\{A_\gamma:\ \gamma\in\Gamma\}$ が互いに排反 (互いに素) であるとは，すべての $\alpha,\beta\in\Gamma$ に対して，$\alpha\neq\beta$ ならば，$A_\alpha\cap A_\beta=\emptyset$ が成り立つことである．

命題 1.2 (和と交わりおよびド・モルガンの法則の一般化) $\{A_\gamma:\gamma\in\Gamma\}$ を集合族(1-1)とする．このとき，

(i) $\left(\cup_{\gamma\in\Gamma}A_\gamma\right)\cap B=\cup_{\gamma\in\Gamma}(A\cap B)$.

(ii) $\left(\cap_{\gamma\in\Gamma}A_\gamma\right)\cup B=\cap_{\gamma\in\Gamma}(A\cup B)$.

(iii) $\{\cup_{\gamma\in\Gamma}A\}^c=\cap_{\gamma\in\Gamma}A_\gamma^c$

(iv) $\{\cap_{\gamma\in\Gamma A}A\}^c=\cup_{\gamma\in\Gamma}A_\gamma^c$

証明　証明は略． □

定義 1.2　可算無限個の事象列 $\{A_n\}_{n=1}^\infty$ に対して，その**上極限事象**と**下極限事象**をそれぞれ

$$\begin{aligned}\overline{\lim}_{n\to\infty}A_n &= \cap_{m=1}^\infty\cup_{n=m}^\infty A_n\\ \underline{\lim}_{n\to\infty}A_n &= \cup_{m=1}^\infty\cap_{n=m}^\infty A_n\end{aligned}$$

上極限事象は $\{A_n\}$ のうちの無限個が起きるという事象を，下極限事象は $\{A_n\}$ のうちのある番号から先のすべてが起こるという事象を表す．すなわち，$\omega\in\overline{\lim}A_n$ とは，無限に多くの n に対して，$\omega\in A_n$ となることであり，$\omega\in\underline{\lim}A_n$ とは，ある番号から先のすべての n に対して $\omega\in A_n$ である．上極限と下極限が一致するとき，$\lim_{n\to\infty}A_n$ と書き，$\{A_n\}_{n=1}^\infty$ の極限事象という．

定理 1.1 つぎの関係式が成り立つ．

(i) $\liminf_{n\to\infty}A_n\subset\limsup_{n\to\infty}A_n$.

(ii) $\left(\liminf_{n\to\infty}A_n\right)^c=\limsup_{n\to\infty}A_n^c$.

(iii) $\left(\limsup_{n\to\infty}A_n\right)^c=\liminf_{n\to\infty}A_n^c$.

(iv) $A_1\subset A_2\subset\cdots$ ならば，$\lim_{n\to\infty}A_n=\cup_{n=1}^\infty A_n$.

(v) $A_1\supset A_2\supset\cdots$ ならば，$\lim_{n\to\infty}A_n=\cap_{n=1}^\infty A_n$.

証明　証明は略． □

1.1.2　確率の定義

定義 1.3　Ω の部分集合の集まり $\mathcal{F}$ は次の条件をみたすとき，Ω 上の**有限加法族**であるという．

(i)　$\Omega \in \mathcal{F}$

(ii)　$A \in \mathcal{F} \Longrightarrow A^c \in \mathcal{F}$

(iii)　$A_1, A_2, \ldots, A_k \in \mathcal{F} \Longrightarrow \cup_{i=1}^{k} A_i \in \mathcal{F}$

さらに，

(iii)'　$A_1, A_2, \ldots, A_k, \ldots \in \mathcal{F} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

をみたすとき，Ω 上の**完全加法族** (σ-加法族) という．

注意 1.1　$A_1, A_2, \ldots, A_k, \ldots \in \mathcal{F}$ ならば，$\cap_{i=1}^{\infty} A_i \in \mathcal{F}$ である．Ω 上の完全加法族で最も大きなものは Ω のすべての部分集合からなる族であり，最小の完全加法族は $\{\emptyset, \Omega\}$ である．

確率論において用いられる完全加法族の重要な例をいくつか考察する．まず，一般的な定義から始めよう．

定義 1.4　$\mathcal{F}$ は Ω の部分集合の任意の族とする．$\mathcal{F}$ を含む最小の完全加法族を $\sigma(\mathcal{F})$ と表す．言い換えれば，$\sigma(\mathcal{F})$ は $\mathcal{F}$ を含む完全加法族全体の交わりである．Ω のすべての部分集合全体は完全加法族であることから，少なくとも 1 つは $\mathcal{F}$ を含む完全加法族は存在することに注意しよう．

例 1.3　$\Omega = \mathbb{R}$ とし，次の部分集合族を考える．

(i)　$\mathcal{F}_1 = \{(a, b) : a, b \in \mathbb{R}, a < b\}$

(ii)　$\mathcal{F}_2 = \{[a, b) : a, b \in \mathbb{R}, a < b\}$

(iii)　$\mathcal{F}_3 = \{(a, b] : a, b \in \mathbb{R}, a < b\}$

(iv)　$\mathcal{F}_4 = \{[a, b] : a, b \in \mathbb{R}, a < b\}$

(v)　$\mathcal{F}_5 = \{(-\infty, a] : a \in \mathbb{R}\}$

(vi)　$\mathcal{F}_6 = \{(-\infty, a) : a \in \mathbb{R}\}$

(vii)　$\mathcal{F}_7$ は $\mathbb{R}$ の開集合の族

(viii)　$\mathcal{F}_8$ は $\mathbb{R}$ の閉集合の族

とする．このとき，

$$\sigma(\mathcal{F}_1) = \sigma(\mathcal{F}_2) = \sigma(\mathcal{F}_3) = \sigma(\mathcal{F}_4) = \sigma(\mathcal{F}_5) = \sigma(\mathcal{F}_6) = \sigma(\mathcal{F}_7) = \sigma(\mathcal{F}_8)$$

である．この完全加法族を $\mathcal{B}(\mathbb{R})$ と記し，$\mathbb{R}$ の**ボレル**集合族という．

定義 1.5　$\mathcal{F}$ を Ω 上の完全加法族としたとき，$(\Omega, \mathcal{F})$ のことを**可測空間**という．

定義 1.6　任意の可測空間 $(\Omega, \mathcal{F})$ に対して，確率とは $\mathcal{F}$ 上で定義された関数 P で以下の条件をみたすものである．

(i)　任意の $A \in \mathcal{F}$ について，$P(A) \geq 0$

(ii) $P(\Omega)=1$

(iii) $i=1,2,\ldots$ について $A_i\in\mathcal{F}$ かつ $A_i\cap A_j=\emptyset\,(i\neq j)$ ならば,

$$P(\cup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$$

$P(A)$ は事象 A の起きる確率とよばれる．$(\Omega,\mathcal{F},P)$ を**確率空間**とよぶ．

注意 1.2 Ω が離散のとき，$\mathcal{F}$ は Ω のすべての部分集合の集まりと通常する．また，$\Omega=\mathbb{R}$ のとき，$\mathbb{R}$ のすべての部分集合の集まりを採用するのでなく，ボレル集合族 $\mathcal{B}(\mathbb{R})$ を採用するのが標準的である．

命題 1.3 (確率の性質) (i) $P(\emptyset)=0$

(ii) 任意の $A\in\mathcal{F}$ に対し，$P(A)\leq 1$

(iii) 任意の $A\in\mathcal{F}$ に対し，$P(A^c)=1-P(A)$

(iv) 任意の $A,B\in\mathcal{F}$ に対し，$P(A\cap B^c)=P(A)-P(A\cap B)$

(v) $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

(vi) $A\subset B$ ならば，$P(A)\leq P(B)$

(vii) $i=1,2,\ldots$ に対し，$A_i\in\mathcal{F}$ ならば，

$$P(\cup_{i=1}^{\infty}A_i)\leq\sum_{i=1}^{\infty}P(A_i)$$

(viii) $A_i\subset A_{i+1}$ ならば，

$$P(\cup_{i=1}^{\infty}A_i)=\lim_{i\to\infty}P(A_i)$$

(ix) $A_i\supset A_{i+1}$ ならば，

$$P(\cap_{i=1}^{\infty}A_i)=\lim_{i\to\infty}P(A_i)$$

[(x) $A_i\supset A_{i+1}$ かつ $\cap_{i=1}^{\infty}A_i=\emptyset$ ならば，

$$P(\cap_{i=1}^{\infty}A_i)=\lim_{i\to\infty}P(A_i)=0$$

証明 (i)-(vi) は証明は略． □

1.2 条件つき確率と独立性

$(\Omega, \mathcal{F}, P)$ を確率空間とし，$B \in \mathcal{F}$ かつ $P(B) > 0$ とする．事象 B 上の条件つき確率とは，完全加法族 $\mathcal{F}$ 上で定義された確率であって，任意の $A \in \mathcal{F}$ について

$$P(A|\,B) = \frac{P(A \cap B)}{P(B)}$$

で定める．$P(A|\,B)$ を**事象 B が与えられたときの事象 A の条件つき確率**という．

命題 1.4 $P(\cdot\,|\,B)$ は確率測度である．すなわち，

(i) 任意の $A \in \mathcal{F}$ に対し，$P(A|\,B) \geq 0$

(ii) $P(\Omega|\,B) = 1$

(iii) $i = 1, 2, \ldots$ に対し，$A_i \in \mathcal{F}$ かつ $A_i \cap A_j = \emptyset\,(i \neq j)$ ならば，

$$P(\cup_{i=1}^{\infty} A_i|\,B) = \sum_{i=1}^{\infty} P(A_i|\,B)$$

証明 条件付確率と確率の定義 1.3 よりわかる． □

Ω を互いに排反な事象 $A_1, A_2, \ldots, A_n$ に分割されるとする．すなわち，$A_i \cap A_j = \emptyset\,(i \neq j)$ かつ $\cup_{i=1}^{\infty} A_i = \Omega$ である．$i = 1, 2, \ldots, n$ に対し，$P(A_i) > 0$ のとき，任意の事象 B の確率は条件つき確率を用いてつぎのように表現できる．

$$P(B) = \sum_{i=1}^{\infty} P(A_i) P(B|\,A_i)$$

これを**全確率の法則**という．

事象 $A_1, A_2, \ldots, A_n$ は互いに排反で Ω を分割するとする．事象 B が起こる前と後のことを事前と事後とよぶことにする．事象 A_i に対して $P(A_i)$ は事象 B が起こる前に与えられる確率であるから**事前確率**といい，$P(A_i|\,B)$ は B が起こった後で与えられる確率であるから**事後確率**という．

定理 1.2 (ベイズの法則) 事後確率は事前確率を用いてつぎのように表現できる．

$$P(A_i|\,B) = \frac{P(A_i)P(B|\,A_i)}{\sum_{i=1}^{n} P(A_i)P(B|\,A_i)}$$

証明 全確率の法則を用いればよい． □

定義 1.7 2 つの事象 A, B が**確率的に独立**であるとは

$$P(A \cap B) = P(A)P(B)$$

が成り立つことをいう．独立でないとき，**従属**という．

注意 1.3 事象 A と B は独立で $P(A) > 0$，$P(B) > 0$ とする．このとき，

$$P(B|\,A) = P(B), \qquad P(A|\,B) = P(A)$$

である．

定義 1.8 事象 $A_1, A_2, \dots, A_n$ が独立であるとは，すべての $k \in \{1, 2, \dots, n\}$ と $i_1 < i_2 < \cdots < i_k, i_l \in \{1, 2, \dots, n\}, l = 1, 2, \dots, k,$ に対し

$$P(\cap_{l=1}^{k} A_{i_l}) = \Pi_{l=1}^{k} P(A_{i_l})$$

が成り立つことである．すべての $A_i, A_j, i, j = 1, 2, \dots, n, i \neq j$ に対し

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

が成り立つとき，**対独立**という．

注意 1.4 対独立であっても，独立でない場合がある．

定義 1.9 $\mathcal{F}_1$ と $\mathcal{F}_2$ を完全加法族 $\mathcal{F}$ の部分完全加法族とする．このとき， $\mathcal{F}_1$ と $\mathcal{F}_2$ が独立であるとは，すべての $A \in \mathcal{F}_1$ と $B \in \mathcal{F}_2$ に対し

$$P(A \cap B) = P(A)P(B)$$

が成り立つことをいう．

1.3 確率変数

$(\Omega, \mathcal{F}, P)$ を確率空間とし，X を Ω から $\mathbb{R}$ への写像とする．X によるボレル集合 $B \in \mathcal{B}(\mathbb{R})$ の**逆像**は Ω の部分集合で $X^{-1}(B)$ と記し

$$X^{-1}(B) = \{\omega \in \Omega \ X(\omega) \in B\}$$

で定義する．

定理 1.3 (逆像の性質) $B, B', \{B_\gamma : \gamma \in \Gamma\}$ はボレル集合とする．このとき

(i) $B \subset B'$ ならば，$X^{-1}(B) \subset X^{-1}(B')$

(ii) $X^{-1}(\cup_{\gamma\in\Gamma} B_\gamma) = \cup_{\gamma\in\Gamma} X^{-1}(B_\gamma)$
$X^{-1}(\cap_{\gamma\in\Gamma} B_\gamma) = \cap_{\gamma\in\Gamma} X^{-1}(B_\gamma)$

(iii) B と B' が互いに排反ならば，$X^{-1}(B)$ と $X^{-1}(B')$ も互いに排反である．

(iv) $X^{-1}(B^c) = \{X^{-1}(B)\}^c$

証明 集合・位相入門（松坂和夫，岩波）等を参照.） □

定義 1.10 写像 $X : \Omega \to \mathbb{R}$ が確率空間 $(\Omega, \mathcal{F}, P)$ 上の **(実) 確率変数**であるとは，任意の $B \in \mathcal{B}(\mathbb{R})$ に対して，$X^{-1}(B) \in \mathcal{F}$ をみたすときをいう．

注意 1.5 すべての $a \in \mathbb{R}$ に対して，$\{\omega \in \Omega : X(\omega) \le a\} \in \mathcal{F}$ をみたすことと同値である．また，すべての $a \in \mathbb{R}$ に対して，$\{X < a\} \in \mathcal{F}$ とも同値である．

確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数 X により $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度 P_X がつぎのように定まる：任意の $B \in \mathcal{B}(\mathbb{R})$ に対して

$$P_X(B) = P(X \in B) = P(\omega \in X^{-1}(B)) = P(X^{-1}(B))$$

P_X のことを X の分布という．

$X_1, X_2, \ldots, X_p$ が $(\Omega, \mathcal{F}, P)$ 上の確率変数のとき，$\boldsymbol{X} = (X_1, X_2, \ldots, X_p)'$ を p-次元確率変数という．

定理 1.4 (確率変数の性質) X，Y および $\{X_n\}_{n=1}^{\infty}$ は $(\Omega, \mathcal{F}, P)$ 上の確率変数 *(列)* とする．

(i) すべての $a, b \in \mathbb{R}$ に対し，$aX + bY$ も確率変数である．

(ii) $\max\{X, Y\}$ も $\min\{X, Y\}$ も確率変数である．

(iii) XY も確率変数である．

(iv) 各 $\omega \in \Omega$ に対し，$Y(\omega) \ne 0$ ならば，X/Y も確率変数である．

(v) $\sup_n X_n, \inf_n X_n, \limsup_{n\to\infty} X_n, \liminf_{n\to\infty} X_n$ も確率変数である．

.

証明 (i) を示すために，$X + Y$ と aX も確率変数であることを示す．各 t に対して，

$$\{X + Y < t\} = \cup_{r\in\mathbb{Q}}(\{X < r\} \cap \{Y < t - r\}) \in \mathcal{F}$$

となる．ただし，$\mathbb{Q}$ は有理数とする．つぎに，aX も確率変数であることを示す．$a > 0$ の場合，$\{aX \le t\} = \{X \le t/a\} \in \mathcal{F}$．$a < 0$ の場合，$\{aX \le t\} = \{X \ge t/a\} \in \mathcal{F}$.

(ii) を示すためには，

$$\{\max\{X, Y\} \leq t\} = \{X \leq t\} \cup \{Y \leq t\}$$

と

$$\{\min\{X, Y\} \leq t\} = \{X \leq t\} \cap \{Y \leq t\}$$

に注意すればよい．
(iii) を示すためには，

$$\{X^2 \leq t\} = \{-\sqrt{t} \leq X \leq \sqrt{t}\} = \{X \leq \sqrt{t}\} \setminus \{X \leq -\sqrt{t}\}$$

から X^2 は確率変数になることに注意する．さらに，

$$XY = \frac{1}{2}\{(X+Y)^2 - (X-Y)^2\}$$

から証明される．
(iv) を示すためには，

$$\{\sup_n X_n \leq t\} = \cap_n \{X_n \leq t\}, \qquad \{\inf_n X_n \geq t\} = \cap_n \{X_n \geq t\}$$

から $\sup_n X_n$ と $\inf_n X_n$ は確率変数であることがわかる．さらに，

$$\limsup_n X_n = \inf_n \sup_{k \geq n} X_k, \qquad \liminf_n X_n = \sup_n \inf_{k \geq n} X_k$$

から (iv) は示された． □

定義 1.11 $(\Omega, \mathcal{F}, P)$ 上の確率変数 X の**分布関数**を

$$F_X(x) = P((-\infty, x]) = P(X \leq x) = P(\omega \in \Omega :\ X(\omega) \leq x)$$

で定義する．ただし，x は任意に実数である．

定理 1.5 (分布関数の性質) (i) *(単調性)* $x, y \in \mathbb{R}$ とする．このとき，$x < y$ ならば，$F_X(x) \leq F_X(y)$

(ii) *(右連続性)* $\lim_{y \to x+0} F_X(y) = F_X(x)$

(iii) 任意の $x \in \mathbb{R}$ に対して，$0 \leq F_X(x) \leq 1$ で $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$, $F_X(\infty) = \lim_{x \to \infty} F_X(x) = 1$ である．

証明 □

確率変数 X の分布関数が $F_X(\cdot)$ である場合「確率変数 X は分布 $F(\cdot)$ に従う」といい，“$X \sim F_X$ ” と書くことにする．

注意 1.6 分布関数 $F(\cdot)$ が与えれたとき，$(\mathbb{R}, \mathcal{B}(\mathbb{R})$ 上の確率測度 μ で

$$F(x) = \mu((-\infty, x]), \qquad x \in \mathbb{R}$$

となるものが一意的に存在することが知られている．

X と Y を確率変数としたとき，X と Y の**分布が等しい**とは，任意の $A \in \mathcal{B}(\mathbb{R})$ に対し

$$P_X(A) = P_Y(A)$$

が成り立つ[1-2]ときをいう．

X と Y の分布が等しいために必要十分条件は，すべての $x \in \mathbb{R}$ に対し，$F_X(x) = F_Y(x)$ が成り立つことである．

1.4 確率密度関数と確率関数

定義 1.12 $F_X(\cdot)$ を確率変数 X の分布関数とする．X が**連続型確率変数**であるとは，F_X が $\mathbb{R}$ 上の連続関数のときをいう．また，X が**離散型確率変数**であるとは，F_X が $\mathbb{R}$ 上の階段関数のときをいう．

定義 1.13 X を離散型確率変数とする．X の**確率関数**を

$$f_X(x) = P(X = x), \qquad \text{すべての } x$$

で定める．

注意 1.7 確率変数 X が離散型のとき，$\{x \in \mathbb{R} : f_X(x) > 0\}$ は高々可算個であることを示すことができる．また，$f_X(x) > 0$ なる点に対し，

$$f_X(x) = P(X \le x) - P(X < x) = F_X(x) - \lim_{y \to x-0} F_X(y)$$

となる．$S = \{x \in \mathbb{R} : f_X(x) > 0\}$ としたとき，

$$p_i = f_X(x_i), \qquad x_i \in S, \qquad i = 1, 2, \ldots$$

を離散型確率変数 X の分布とよぶことにする．p_i と $x_i, i = 1, 2, \ldots,$ を表にまとめたものを**確率分布表**という．

命題 1.5 (確率関数の性質) (i) $0 \le f_X(x) \le 1$

(ii) $\sum_{x \in S} f_X(x) = 1$

証明 定義よりわかる． □

定義 1.14 X を連続型確率変数とし，$F_X(\cdot)$ をその分布関数とする．$\mathbb{R}$ 上の関数 $f_X(\cdot)$ で任意の $x \in \mathbb{R}$ に対して

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy$$

をみたすものが存在するとき，$f_X(\cdot)$ を X の**確率密度関数**という．

X は密度 $f_X(\cdot)$ を持つとか分布 $f_X(\cdot)$ に従うという．さらに，$F_X(\cdot)$ が $\mathbb{R}$ 上で微分可能ならば

$$f_X(x) = \frac{d}{dx} F_X(x)$$

となる．

注意 1.8 X が連続型確率変数ならば，任意の $x \in \mathbb{R}$ において，$P(X = x) = 0$ となる．すなわち，$P(\omega \in \Omega : X(\omega) = 0) = 0$ である．なぜならば，任意の $\epsilon > 0$ に対して

$$\begin{aligned} P(X = x) &\le P(x - \epsilon < X \le x) \\ &= P(X \le x) - P(X \le x - \epsilon) \\ &= F_X(x) - F_X(x - \epsilon) \end{aligned}$$

となり，$\epsilon \downarrow 0$ とすれば，$F_X(\cdot)$ の連続性より

$$P(X = x) \le \lim_{\epsilon \downarrow 0} \{F_X(x) - F_X(x - \epsilon)\} = 0$$

となる．

命題 1.6 (確率密度関数の性質) (i) $f_X(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

(iii) 任意のボレル集合 $B \in \mathcal{B}(\mathbb{R})$ に対して

$$P(X \in B) = \int_B f_X(x)\,dx$$

証明 定義よりわかる. □

1.5 確率変数の期待値

定義 1.15 確率変数 X は確率関数または密度関数 $f_X(\cdot)$ を持つとき，X の期待値を

$$\mathbb{E}[X] = \begin{cases} \sum_x x f_X(x), & \textit{(離散型)} \\ \int_{-\infty}^{\infty} x f_X(x)\,dx, & \textit{(連続型)} \end{cases}$$

で定義する．ただし，離散型の場合は $\sum_x |x| f_X(x) < \infty$ のとき，連続型の場合は $\int_{-\infty}^{\infty} |x| f_X(x)\,dx < \infty$ のとき，X の期待値を定義することにする．期待値が定義されるとき，X の期待値が存在するという．

一般に，$\mathbb{R}$ 上のボレル可測関数(1-3) $g(\cdot)$ に対し，$g(X)$ の期待値を

$$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x) f_X(x), & \text{(離散型)} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\,dx, & \text{(連続型)} \end{cases}$$

で定める．$g(X)$ の期待値の存在は X の期待値の存在と同様に定める．

例 1.4 X の確率分布は

X の取る値	0	1	合計
確率	1/2	1/2	1

とする．このとき，

$$\mathbb{E}[X] = 0 \times f_X(0) + 1 \times f_X(1) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2}$$

となる．

例 1.5 X は確率密度関数

$$f_X(x) = \begin{cases} 1, & \textit{(}0 \le x \le 1\textit{)} \\ 0, & \textit{(その他)} \end{cases}$$

を持つとする．このとき，

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x \cdot 1\,dx = \frac{1}{2}$$

例 1.6 X は確率密度関数

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

を持つとする．このとき，X の期待値は存在しない．なぜならば，$M \ge 1$ に対して

$$\mathbb{E}[|X|] \ge \int_1^M x \frac{1}{\pi(1+x^2)}\,dx \ge \int_1^M \frac{x}{\pi(2x^2)}\,dx = \frac{1}{2\pi} \int_1^M \frac{dx}{x}$$

よって

$$\mathbb{E}[|X|] \ge \lim_{M\to\infty} \frac{1}{2\pi} \int_1^M \frac{dx}{x} = \infty$$

よりわかる．

定理 1.6 (期待値の性質) (i) 定数 c に対して，$\mathbb{E}[c] = c$

(ii) ふたつのボレル可測関数 $h(\cdot)$ と $g(\cdot)$ および定数 a, b に対して

$$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)]$$

(iii) $h(x) \geq 0$ ならば，$\mathbb{E}[h(X)] \geq 0$

(iv) $|\mathbb{E}[h(X)]| \leq \mathbb{E}[|h(X)|]$

(v) X が非負値確率変数(1-4)のとき，$\mathbb{E}[X] = 0$ ならば，$P(X = 0) = 1$ である．

ただし，上記において，いずれの期待値も存在するものと仮定する．

証明 積分の性質からわかる． □

1.6 積率と積率母関数

定義 1.16 正の各整数 n に対して，$\mathbb{E}[|X|^n] < \infty$ のとき，

$$\mu_n' = \mathbb{E}[X^n]$$

を X の n 次の**積率**という．さらに，

$$\mu_n = \mathbb{E}[(X-\mu)^n]$$

を**中心まわりの n 次の積率**という．ただし，$\mu = \mathbb{E}[X]$ である．

定義 1.17 X の中心まわりの 2 次の積率を**分散**といい，$\mathbb{VAR}[X]$ と書く．すなわち，$\mathbb{VAR}[X] = \mathbb{E}[(X-\mathbb{E}[X])^2]$ である．

注意 1.9 $\mathbb{E}[X^2] < \infty$ ならば，$\mathbb{E}[|X|] < \infty$ であることに注意せよ．これは Hölder の不等式からわかる．以下では直接的に確認する．簡単なために，X を連続型確率変数とし，確率密度関数 $f_X(x)$ を持つとして議論を進める．

$$\mathbb{E}[|X|] = \int_{-\infty}^{\infty} |x| f_X(x)\,dx = \int_{-\infty}^{-1} |x| f_X(x)\,dx + \int_{-1}^{1} |x| f_X(x)\,dx + \int_{\infty}^{1} |x| f_X(x)\,dx$$

となる．$x \le -1$ または $x \ge 1$ のときは，$|x| f_X(x) \le x^2 f_X(x)$ となり，$-1 \le x \le 1$ のときは $|x| f_X(x) \le f_X(x)$ となることに注意すれば，

$$\mathbb{E}[|X|] \le \int_{-\infty}^{-1} x^2 f_X(x)\,dx + \int_{-1}^{1} f_X(x)\,dx + \int_{\infty}^{1} x^2 f_X(x)\,dx \le \int_{-\infty}^{\infty} x^2 f_X(x)\,dx + \int_{-\infty}^{\infty} f_X(x)\,dx = \mathbb{E}[X^2] + 1 < \infty$$

がわかる．

定理 1.7 (分散の性質) (i) $\mathbb{VAR}[a+bX] = b^2\mathbb{VAR}[X]$．ただし，$a, b$ は定数である．

(ii) $\mathbb{VAR}[X] = \mathbb{E}[X^2] - \{\mathbb{E}[X]\}^2$

(iii) $\mathbb{VAR}[X] = 0 \Longrightarrow P(X = \mathbb{E}[X]) = 1$

証明 □

定義 1.18 ある正の数 t_0 が存在して，すべての $t \in (-t_0, t_0)$ に対し，e^{tX} の期待値が存在するならば，X の**積率母関数**を

$$M_X(t) = \mathbb{E}[e^{tX}], \qquad t \in (-t_0, t_0)$$

で定義する．このような t_0 が存在しないとき，X の積率母関数は存在しないという．

定理 1.8 (積率母関数の性質) $M_X(t)$ を X の積率母関数とする．このとき，正の整数 n に対して

$$\mathbb{E}[X^n] = \left.\frac{d^n}{dt^n} M_X(t)\right|_{t=0}$$

となる．

証明 X が連続型の場合のみを示す．微分記号と積分記号の交換が可能であると仮定すれば，

$$\begin{aligned}\frac{d}{dt}M_X(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \\ &= \int_{-\infty}^{\infty}\left(\frac{d}{dt}e^{tx}\right) f_X(x)\,dx \\ &= \int_{-\infty}^{\infty}\left(xe^{tx} f_X(x)\right)\,dx = \mathbb{E}[Xe^{tX}]\end{aligned}$$

したがって

$$\left.\frac{d}{dt}M_X(t)\right|_{t=0} = \mathbb{E}[Xe^{tX}]\big|_{t=0} = \mathbb{E}[X]$$

□

例 1.7 X の確率分布は

X の取る値	0	1	合計
確率	1/2	1/2	1

とする．このとき，$t \in \mathbb{R}$ に対し

$$M_X(t) = \mathbb{E}[e^{tX}] = 1 \times f_X(0) + e^t \times f_X(1) = \frac{1}{2}(1+e^t)$$

となる．よって

$$\mathbb{E}[X] = \left.\frac{d}{dt}M_X(t)\right|_{t=0} = \left.\frac{1}{2}e^t\right|_{t=0} = \frac{1}{2}$$

となる．X の分散は

$$\mathbb{VAR}[X] = \mathbb{E}[(X-\mathbb{E}[X])^2] = (0-\frac{1}{2})^2 \times f_X(0) + (1-\frac{1}{2})^2 \times f_X(1) = \frac{1}{4}$$

となる．一方，

$$\mathbb{E}[X^2] = \frac{d^2}{dt^2}M_X(t) = \left.\frac{1}{2}e^t\right|_{t=0} = \frac{1}{2}$$

より

$$\mathbb{VAR}[X] = \mathbb{E}[X^2] - \{\mathbb{E}[X]\}^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

となる．

命題 1.7 確率変数 X と Y の分布関数を $F_X(\cdot)$ と $F_Y(\cdot)$ とする．ある正の整数 t_0 が存在して，すべての $t \in (-t_0, t_0)$ に対し，X と Y の積率母関数 $M_X(t)$ と $M_Y(t)$ が存在して，$M_X(t) = M_Y(t)$ ならば，すべての $x \in \mathbb{R}$ に対して $F_X(x) = F_Y(x)$ である．

証明 証明は略. □

1.7 確率変数の変換

確率変数 X は分布関数 $F_X(\cdot)$ を持つとする．$g(\cdot)$ をボレル可側関数としたとき，$Y = g(X)$ も確率変数になり，Y の分布 P_Y はつぎのように定まる：任意のボレル集合 $A \in \mathcal{B}(\mathbb{R})$ に対して

$$P(Y \in A) = P(g(X) \in A)$$

また，Y の分布関数は

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

となる．P_Y は P_X と g に依存するので，$F_Y(\cdot)$ は $F_X(\cdot)$ と g に依存するであろう．さらに，存在するならば，$F_Y(\cdot)$ の確率密度関数 $f_Y(\cdot)$ も求めることができれば便利である．

以下の例では，$F_X(\cdot)$ は微分可能とする．

例 1.8 X の線形変換を考える．すなわち，$Y = a + bX$ である．ここで，a, b は定数で $b \neq 0$ とする．$b > 0$ のとき

$$F_Y(y) = P(Y \le y) = P(a + bX \le y) = P\left(X \le \frac{y-a}{b}\right) = F_X\left(\frac{y-a}{b}\right)$$

また，$b < 0$ のとき

$$F(y) = P(Y \le y) = P\left(X \ge \frac{y-a}{b}\right) = 1 - P\left(X \le \frac{y-a}{b}\right) = 1 - F_X\left(\frac{y-a}{b}\right)$$

となる．したがって

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{|b|}f_X\left(\frac{y-a}{b}\right)$$

となる．線形関すは狭義単調関数である．

例 1.9 X を正値確率変数とする．すなわち，$P(X \le 0) = 0$ である．$Y = \log X$ とおく．このとき，$y \in \mathbb{R}$ に対して

$$F_Y(y) = P(Y \le y) = P(X \le e^y) = F_X(e^y)$$

となる．したがって

$$f_Y(y) = \frac{d}{dy}F_Y(y) = e^y f_X(e^y)$$

となる．$g(x) = \log x$ も $\mathbb{R}^+$ 上では狭義単調関数である．

例 1.10 $Y = X^2$ とおく．これは $\mathbb{R}$ 上の単調関数ではない．しかし，$\mathbb{R}^+$ 上と $\mathbb{R}^-$ 上では単調である．$y \ge 0$ に対して，

$$\begin{aligned}
F_Y(y) &= P(Y \le y) = P(\omega \in \Omega :\, (X(\omega)^2) \le y) \\
&= P(\omega \in \Omega :\, (X(\omega)^2) \le y,\, X(\omega) > 0) + P(\omega \in \Omega :\, (X(\omega)^2) \le y,\, X(\omega) < 0) \\
&\quad + P(\omega \in \Omega :\, X(\omega) = 0) \\
&= P(0 < X < \sqrt{y}) + P_X(-\sqrt{y} \le X < 0) + P(X = 0) \\
&= P(-\sqrt{y} \le X \le \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}$$

となる．したがって，

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_X(-\sqrt{y})$$

となる．$y<0$ に対しては $F_Y(y)=0$ となるので，$f_Y(y)=0$ となることは容易にわかる．よって

$$f_Y(y)=\begin{cases}\frac{1}{2\sqrt{y}}f_X(\sqrt{y})+\frac{1}{2\sqrt{y}}f_X(-\sqrt{y}), & (y>0),\\ 0, & (\text{その他})\end{cases}$$

となる．

これらの例を踏まえて，一般の $Y=g(X)$ の場合を考える．そのために

$$\mathcal{X}=\{x:\, f_X(x)>0\},\qquad \mathcal{Y}=\{y:\, y=g(x), x\in\mathcal{X}\}$$

とおく．$\mathcal{X}$ と $\mathcal{Y}$ を X と Y の台とよぶ．たとえば，$f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ で $Y=X^2$ のとき，$\mathcal{X}=\mathbb{R}$ ならば，$\mathcal{Y}=[0,\infty)$ となる．

いま，関数 $g:\mathcal{X}\to\mathcal{Y}$ を考える．もし，g が一対一上への写像ならば，逆写像 $g^{-1}:\mathcal{Y}\to\mathcal{X}$ は定義できる．しかし，一般には，逆像で定義する．すなわち，

$$g^{-1}(y)=\{x\in\mathcal{X}:\, g(x)=y\}$$

さらに，ボレル集合 A に対し

$$g^{-1}(A)=\{x\in\mathcal{X}:\, g(x)\in A\}$$

とする．すると

$$P(Y\in A)=P(\omega\in\Omega:\, g(X(\omega))\in A)=P(X\in g^{-1}(A))$$

となる．しかし，Y の確率密度関数を求めるためには，追加の条件が必要となる．

命題 1.8 X を連続型確率変数とし，f_X をその確率密度関数とし，$\mathcal{X}$ 上で連続とする．
(i) 逆関数 $g^{-1}(\cdot)$ は $\mathcal{Y}$ 上で定義され，連続微分可能とする．このとき

$$f_Y(y)=\begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, & (y\in\mathcal{Y}),\\ 0, & (\text{その他})\end{cases}$$

となる．
(ii) $\{\mathcal{X}_i\}_{i=0}^k$ を $\mathcal{X}$ の分割(1-5)とし，$P(X\in\mathcal{X}_0)=0$ とし，$f_X(\cdot)$ は各 $\mathcal{X}_i, i=1,2,\ldots,k,$ 上で連続とする．さらに，各 $\mathcal{X}_i, i=1,2,\ldots,k,$ 上で定義された関数 $g_i(\cdot)$ が存在して，$x\in\mathcal{X}_i$ に対し，$g(x)=g_i(x)$ が成立し，$g_i(x)$ は $\mathcal{X}_i$ 上で狭義単調関数とし，$g_i^{-1}(\cdot)$ は $\mathcal{Y}\setminus g(\mathcal{X}_0)$ 上で定義され，連続微分可能とする．このとき，

$$f_Y(y)=\begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y))\left|\frac{d}{dy}g_i^{-1}(y)\right|, & (y\in\mathcal{Y}\setminus g(\mathcal{X}_0)),\\ 0, & (\text{その他})\end{cases}$$

となる．

証明 (i) g が狭義単調増加のとき，$y\in\mathcal{Y}$ に対し，

$$F_Y(y)=P(Y\le y)=P(X\le g^{-1}(y))=F_X(g^{-1}(y))$$

となる．よって

$$f_Y(y)=\frac{d}{dy}F_Y(y)=f_X(g^{-1}(x))\frac{d}{dy}g^{-1}(y)$$

となる．また，g が狭義単調減少のとき，

$$F_Y(y)=P(Y\le y)=P(X\ge g^{-1}(y))=1-F_X(g^{-1}(y))$$

となる．よって，

$$f_Y(y) = \frac{d}{dy}F_Y(y) = f_X(g^{-1}(x))\left(-\frac{d}{dy}g^{-1}(y)\right)$$

となる．g が単調増加のときは $(d/dy)g^{-1}(y) > 0$，g が単調減少のときは $(d/dy)g^{-1}(y) < 0$ に注意すればよい．

(ii) $y \in \mathcal{Y}$ に対し，

$$F_Y(y) = P(Y \le y) = \sum_{i=1}^{k} P(g(X) \le y, X \in \mathcal{X}_i) = \sum_{i=1}^{k} P(X \in g_i^{-1}((-\infty, y])$$

となる．最右辺の各項に対して，(i) の議論を適用すれば，(ii) は証明される． □

例 1.11 確率変数 X は確率密度関数

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad -\infty < x < \infty$$

を持つとする．$Y = X^2$ としたとき，Y の確率密度関数を求めよう．関数 $g(x) = x^2$ は $(-\infty, 0)$ 上と $(0, \infty)$ 上のそれぞれで単調関数である．$\mathcal{Y} = (0, \infty)$ として，命題 1.8 を利用するために，

$$\begin{aligned}
\mathcal{X}_0 &= \{0\}, \\
\mathcal{X}_1 &= (-\infty, 0), \qquad g_1(x) = x^2, \qquad g_1^{-1}(y) = -\sqrt{y} \\
\mathcal{X}_2 &= (0, \infty), \qquad g_2(x) = x^2, \qquad g_2^{-1}(y) = \sqrt{y}
\end{aligned}$$

とおく．Y の確率密度関数は

$$\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi}}e^{-(-\sqrt{y})^2/2}\left|-\frac{1}{2\sqrt{y}}\right| + \frac{1}{\sqrt{2\pi}}e^{-(\sqrt{y})^2/2}\left|\frac{1}{2\sqrt{y}}\right| \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-y/2}, \qquad 0 < y < \infty
\end{aligned}$$

となる．

例 1.12 確率変数 X は連続で分布関数 $F_X(\cdot)$ を持つとし，$Y = F_X(X)$ とおく．このとき，Y は確率密度関数

$$f_Y(y) = \begin{cases} 1 & (0 < y < 1), \\ 0 & (\text{その他}) \end{cases} \tag{1.1}$$

を持つ．

$F_X(x)$ は x に関して狭義単調増加として証明(1-6)をする．$0 \le F_X(x) \le 1$ であり，$P(F_X(X) < 0) = 0$ と $P(F_X(X) > 1) = 0$ となる確率はゼロなので，$y \le 0$ のときは $P(Y \le y) = 0$ である．$0 < y < 1$ に対して，

$$\begin{aligned}
P(Y \le y) &= P(F_X(X) \le y) \\
&= P(F_X^{-1}(F_X(X)) \le F_X^{-1}(y)) \\
&= P(X \le F_X^{-1}(y)) \\
&= F_X(F_X^{-1}(y)) \\
&= y
\end{aligned}$$

となる．$y \ge 1$ に対しては $P(Y \le y) = 1$ となる．したがって，

$$F_Y(y) = \begin{cases} 0, & (y \le 0), \\ y, & (0 < y < 1), \\ 1, & (y \ge 1) \end{cases}$$

となる．$0 < y < 1$ 上で $F_Y(y)$ は微分可能で導関数は 1 となる．また，$y = 0$ における $F_Y(y)$ の導関数を左微分で，$y = 1$ における $F_Y(y)$ の導関数を右微分で定めると Y の確率密度関数は (1.1) となることがわかる．

1.8　確率変数の不等式 1

定理 1.9 (マルコフの不等式)　非負値確率変数 X が有限の期待値 $\mathbb{E}[X] < \infty$ をもつとき，任意の正数 $a > 0$ に対して，

$$P\{X \geq a\} \leq \frac{1}{a}\mathbb{E}[X]$$

が成立する．

証明　　X が連続型確率の場合を示す．X の確率密度関数を $f_X(x)$ とする．

$$\mathbb{E}[X] = \int_0^\infty x f_X(x)\,dx \geq \int_a^\infty x f_X(x)\,dx \geq a\int_a^\infty f_X(x)\,dx = a\,\mathbb{P}(X \geq a)$$

よりわかる．
証明

定理 1.10 (チェビシェフの不等式)　確率変数 X が平均 μ，分散 σ^2 をもつとき，任意の正数 $a > 0$ に対して，

$$\mathbb{P}\{|X-\mu| \geq a\} \leq \frac{\sigma^2}{a^2}$$

が成り立つ．

証明　Markov の不等式から

$$\mathbb{P}\{|X-\mu| \geq a\} = \mathbb{P}\{|X-\mu|^2 \geq a^2\} \leq \frac{\mathbb{E}[(X-\mu)^2]}{a^2}$$

よりわかる．□

例 1.13　　X は平均 μ，分散 $\sigma^2\,(\sigma > 0)$ の分布に従うとする．このとき，任意の正の実数 t に対して，

$$\mathbb{P}\{|X-\mu| \geq t\sigma\} \leq \frac{1}{(t\sigma)^2}\mathbb{E}[(X-\mu)^2] = \frac{1}{t^2}$$

となる．したがって，$t = 2$ の場合

$$\mathbb{P}\{|X-\mu| \geq 2\sigma\} \leq \frac{1}{2^2} = 0.25$$

となる．

例 1.14　　Z は標準正規分布に従うとする．このとき，任意の正の実数 t に対して，

$$\mathbb{P}\{|Z| \geq t\} \leq \sqrt{\frac{2}{\pi}}\frac{e^{-t^2/2}}{t} \tag{1.2}$$

となる．
　$t = 2$ のとき，Chebychev の不等式から

$$\mathbb{P}\{|Z| \geq 2\} \leq \frac{1}{2^2} = 0.25$$

となる．しかし，(1.2) から

$$\mathbb{P}\{|Z| \geq 2\} \leq \sqrt{\frac{2}{\pi}}\frac{e^{-2}}{2} = 0.054$$

となる．また，

$$\mathbb{P}\{|Z| \geq 3\} \leq \sqrt{\frac{2}{\pi}}\frac{e^{-4.5}}{3} = 0.00295$$

となる．

1.2) は以下からわかる．

$$\begin{aligned}\mathbb{P}\{Z \geq t\} &= \frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-z^2/2}\,dz \\ &\leq \frac{1}{\sqrt{2\pi}}\int_t^\infty \frac{z}{t}e^{-z^2/2}\,dz \\ &= \frac{1}{\sqrt{2\pi}}\left[-\frac{1}{t}e^{-z^2/2}\right]_t^\infty = \frac{1}{\sqrt{2\pi}}\frac{e^{-t^2/2}}{t}\end{aligned}$$

と

$$\mathbb{P}\{|Z| \geq t\} = \mathbb{P}\{\{Z \geq t\} \cup \{Z \leq -t\}\} = \mathbb{P}\{Z \geq t\} + \mathbb{P}\{Z \leq -t\} = 2\,\mathbb{P}\{Z \geq t\}$$

からわかる．

ある区間 $I = (a,\, b)$ の上の実数値連続関数 $h(x)$ が凸 (convex) であるとは，任意の $c \in (1,\, 0)$ と任意の $x_1,\, x_2 \in I$ に対して，

$$h(cx_1 + (1-c)x_2) \leq ch(x_1) + (1-c)h(x_2)$$

が成り立つことである．もし， $x_1 \neq x_2$ に対して「$\leq$ 」のかわりに「$<$ 」が常に成立するならば，狭義の凸関数という．

定理 1.11 (イェンセンの不等式) X を確率変数とし，$h(x)$ を X の値域を含む区間上で凸な関数とする．X と $h(X)$ の期待値が有限のとき，

$$\mathbb{E}[h(X)] \geq h(\mathbb{E}[X])$$

が成立する．もし，$h(x)$ が狭義の凸関数のとき，等号が成り立つのは 1 点分布の時にかぎる．

証明 任意の固定した x_0 とある線形関数 $g(x) = ax + b$ が存在して，$h(x_0) = g(x_0)$ とすべての x に対して $h(x) \geq g(x)$ が成立する．$x_0 = \mathbb{E}[X]$ として，上のことを利用すれば，

$$\begin{aligned}\mathbb{E}[h(X)] &\geq \mathbb{E}[g(X)] \\ &= \mathbb{E}[aX + b] \\ &= a\mathbb{E}[X] + b \\ &= g(\mathbb{E}[X]) = h(\mathbb{E}[X])\end{aligned}$$

□

例 1.15 X を $\mathbb{E}[X^2] < \infty$ なる確率変数とする．Jensen の不等式を利用するために，まず，$\mathbb{E}[|X|] < \infty$ を確認する．たとえば，X が確率密度関数 $f_X(x)$ を持てば

$$\begin{aligned}\mathbb{E}[|X|] &= \int_{-\infty}^\infty |x|f_X(x)\,dx \\ &= \int_{-\infty}^{-1} |x|f_X(x)\,dx\,dx + \int_{-1}^{1} |x|f_X(x)\,dx\,dx + \int_{1}^{\infty} |x|f_X(x)\,dx\,dx \\ &\leq \int_{-\infty}^{-1} x^2f_X(x)\,dx\,dx + \int_{-1}^{1} f_X(x)\,dx\,dx + \int_{1}^{\infty} x^2f_X(x)\,dx \\ &\leq \int_{-\infty}^{\infty} |x|^2f_X(x)\,dx + \int_{-\infty}^{\infty} f_X(x)\,d = \mathbb{E}[X^2] + 1 < \infty\end{aligned}$$

となる．X が離散型確率変数のときも同様にすれば，$\mathbb{E}[|X|] \le \mathbb{E}[X^2]+1$ であることがわかる．さらに，$\mathbb{E}[|X|]$ の評価が Jensen の不等式から得られる．$h(x)=x^2$ とすれば，$h(x)$ は狭義の凸関数なので，Jensen の不等式より

$$\{\mathbb{E}[X]\}^2 = h(\mathbb{E}[X]) \le \mathbb{E}[h(X)] = \mathbb{E}[X^2]$$

となる．したがって，

$$|\mathbb{E}[X]| \le \sqrt{\mathbb{E}[X^2]}$$

をえる．

例 1.16 $a_1, a_2, \ldots, a_n$ は正の数とし，

$$\begin{aligned} m_A &= \frac{1}{n}(a_1+a_2+\cdots+a_n) \qquad (\text{算術平均}), \\ m_G &= \{a_1a_2\cdots a_n\}^{1/n} \qquad (\text{幾何平均}), \\ m_H &= \frac{1}{\frac{1}{n}\left(\frac{1}{a_1}+\frac{1}{a_2}+\cdots+\frac{1}{a_n}\right)} \qquad (\text{調和平均}) \end{aligned}$$

とする．このとき，

$$m_H \le m_G \le m_A$$

が成立する．

Jensen の不等式を用いて示す：そのために，X を確率変数として

$$\mathbb{P}\{X=a_i\} = \frac{1}{n}, \qquad i=1,2,\ldots,n$$

とする．$-\log x$ は凸関数なので，$\mathbb{E}[\log X] \le \log(\mathbb{E}[X])$ が成立する．したがって，

$$\log m_G = \frac{1}{n}\sum_{i=1}^{n}\log a_i = \mathbb{E}[\log X] \le \log(\mathbb{E}[X]) = \log\left(\frac{1}{n}\sum_{i=1}^{n}a_i\right) = \log m_A$$

より $m_G \le m_A$ がわかる．また，

$$\log\frac{1}{m_H} = \log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i}\right) = \log\left(\mathbb{E}\left[\frac{1}{X}\right]\right) \ge \mathbb{E}\left[\log\left(\frac{1}{X}\right)\right] = -\mathbb{E}[\log X] = -\log m_G$$

から $m_G \ge m_H$ がわかる．