第2章　独立確率変数列の極限定理

2.1　独立性

たとえば赤いさいころと白いさいころを2個同時に投げたとき，赤いさいころの出す目の数と白いさいころの出す目の数の間には何の関係もないものと思われる．このことを数学的にはどのように表すのかを考える．このときに対応する確率空間は二つのさいころの出目を並べた

$$\begin{aligned} \Omega &= \{(i,j); 1 \le i,j \le 6\}, \quad \mathcal{F} = \ \Omega \text{ の部分集合全体} \\ P(i,j) &= \frac{1}{36} \end{aligned}$$

となる．赤いさいころの出目を X，白いさいころの出目を Y と書くとき，上の確率は

$$P(i,j) = P(X=i, Y=j) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(X=i)P(Y=j)$$

と書けることに注意すると，$A, B \subset \{1,2,\ldots,6\}$ のとき，

$$\begin{aligned} P(X \in A,\ Y \in B) &= \sum_{i \in A} \sum_{j \in B} P(X=i, Y=j) \\ &= \sum_{i \in A} \sum_{j \in B} P(X=i)P(Y=j) \\ &= \left(\sum_{i \in A} P(X=i) \right) \left(\sum_{j \in B} P(Y=j) \right) \\ &= P(X \in A)P(Y \in B) \end{aligned} \tag{2.1}$$

と，確率の積になる．数学的にはこの式が独立性の定義になる．

定義 2.1 確率空間 $(\Omega, \mathcal{F}, P)$ において，二つの事象 $A, B \in \mathcal{F}$ が独立であるとは，

$$P(A \cap B) = P(A)P(B) \tag{2.2}$$

となるときに言う．さらに，n 個の事象 $A_1, A_2, \ldots, A_n \in \mathcal{F}$ が 独立であるとは，この中から任意に選んだ $A_{k_1}, \ldots, A_{k_m}$ （ただし， $1 \le m \le n$ ）について，

$$P(A_{k_1} \cap \ldots \cap A_{k_m}) = \prod_{i=1}^{m} P(A_{k_i}) \tag{2.3}$$

となるときに言う．最後に，無限個の事象族 $A_\lambda \in \mathcal{F}$, $\lambda \in \Lambda$ が独立であるとは，この 任意の有限部分族 $A_{\lambda_1}, \ldots, A_{\lambda_n}$ が独立なときに言う．

二つの集合族が独立であるというのも考えることがある．

定義 2.2 確率空間 $(\Omega, \mathcal{F}, P)$ において，二つの集合族 $\mathcal{G}, \mathcal{H}$ が 独立であるとは，それぞれの任意有限個の部分族 $\{A_k\}_{k=1}^n \subset \mathcal{G}$, $\{B_\ell\}_{\ell=1}^m \subset \mathcal{H}$ について，

$$P(A_1 \cap \ldots \cap A_n \cap B_1 \cap \ldots \cap B_m) = P(A_1 \cap \ldots \cap A_n)P(B_1 \cap \ldots \cap B_m)$$

が成り立つときに言う．

確率変数の独立性については次の様になる．

定義 2.3 二つの確率変数 X, Y が独立とは，任意のボレル集合 A, B に対して

$$P(X \in A,\ Y \in B) = P(X \in A)P(Y \in B)$$

となるときに言う．同様に n 個の確率変数 $X_1, \ldots, X_n$ が 独立であるとは，任意の $A_1, \ldots, A_n \in \mathcal{F}$ に対して

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{j=1}^{n} P(X_j \in A_j)$$

となるときに言う．
無限個の確率変数 $\{X_\lambda; \lambda \in \Lambda\}$ が独立とはこの中の任意有限個の確率変数の組 $X_{\lambda_1}, \ldots, X_{\lambda_n}$ が独立なときに言う．

定理 2.1 確率変数列 $X_1, \ldots, X_n$ が独立であることと次の条件が成立することは同値である．任意の有界なボレル可測関数 $f_1, \ldots, f_n$ に対して

$$E[\prod_{j=1}^{n} f(X_j)] = \prod_{j=1}^{n} E[f(X_j)] \tag{2.4}$$

が成り立つことである．

証明　（十分性）任意のボレル集合 $B_1, \ldots, B_n$ に対して，$f_j(\omega) = 1_{A_j}(\omega)$ とおくと，これらは有界なボレル関数だから (2.4) から

$$E[\prod_{j=1}^{n} 1_{A_j}(X_j)] = \prod_{j=1}^{n} E[1_{A_j}(X_j)] = \prod_{j=1}^{n} P(X_j \in A_j) \tag{2.5}$$

一方，

$$\prod_{j=1}^{n} 1_{A_j}(X_j(\omega)) = 1 \Leftrightarrow X_j(\omega) \in A_j \ \forall j$$

だから

$$E[\prod_{j=1}^{n} 1_{A_j}(X_j)] = P(\{X_1 \in A_1\} \cap \ldots \cap \{X_n \in A_n\})$$

となり，$X_1, \ldots, X_n$ は独立．
（必要性）$X_1, \ldots, X_n$ が独立とする．このとき定義から任意のボレル集合 $A_1, \ldots, A_n$ に対して (2.5) が成り立つ．これから各 f_j が階段関数のとき，つまり

$$f_j(\omega) = \sum_{k=1}^{N_j} c_k^{(j)} 1_{A_k^{(j)}}(\omega)$$

の形のときにも (2.4) は次のようにして成り立つ．

$$
\begin{aligned}
E[\prod_{j=1}^{n} f_j(X)] &= \sum_{k_1=1}^{N_1}\cdots\sum_{k_n=1}^{N_n} E[\prod_{j=1}^{n} c_k^{(j)} 1_{A_{k_j}}^{(j)}(X_j)] \\
&= \sum_{k_1=1}^{N_1}\cdots\sum_{k_n=1}^{N_n} \prod_{j=1}^{n} E[c_k^{(j)} 1_{A_{k_j}}^{(j)}(X_j)] \\
&= \prod_{j=1}^{n}\left(E[\sum_{k_j=1}^{N_j} c_{k_j}^{(j)} 1_{A_{k_j}}^{(j)}]\right) \\
&= \prod_{j=1}^{n} E[f_j(X_j)]
\end{aligned}
$$

各 f_j が非負のときは $f_j^{(\nu)} \nearrow f_j$ となる非負の階段関数を取ることにより単調収束定理から

$$
\begin{aligned}
E[\prod_{j=1}^{n} f_j(X_j)] &= \lim_{\nu\to\infty} E[\prod_{j=1}^{n} f_j^{(\nu)}(X_j)] \\
&= \lim_{\nu\to\infty} \prod_{j=1}^{n} E[f_j^{(\nu)}(X_j)] \\
&= \prod_{j=1}^{n} E[f_j(X_j)]
\end{aligned}
$$

最後に，一般の有界なボレル可測関数 f_j については，$f_j = f_j^+ - f_j^-$ と書くことで，

$$
\begin{aligned}
E[\prod_{j=1}^{n} f_j(X_J)] &= \sum_{\varepsilon_1=+1,-1}\cdots\sum_{\varepsilon_n=+1,-1} \varepsilon_1\cdots\varepsilon_n E[\prod_{j=1}^{n} f_j^{\varepsilon_j}(X_j)] \\
&= \sum_{\varepsilon_1=+1,-1}\cdots\sum_{\varepsilon_n=+1,-1} \varepsilon_1\cdots\varepsilon_n \prod_{j=1^n} E[f_j^{\varepsilon_j}(X_j)] \\
&= \prod_{j=1}^{n} E[f_j(X_J)]
\end{aligned}
$$

2.1.1 大数の弱法則 (The Weak Laws of Large Numbers)

「そもそも，確率とは何か」という質問に対してなんと答えたらいいのであろうか? **確率** の直観的意味は，相対的な頻度であるといえるであろう．たとえば，さいころの目の出方は理想的にはどの目の出方も同等なはずであるが，現実的には少し偏りがあるとしても不思議ではない．この偏りはどのようにして調べたらよいかというと，何度も何度も繰り返して投げて実験を繰り返して，それぞれの目が n 回の試行のうち何%出たかを調べるのが普通である．つまり，われわれは**相対的な頻度はその目の出る確率に近づいていく**と考えている．これは経験則といっても良い．

これから述べていく大数の弱法則，後の節で述べる大数の強法則はこの経験則を数学的な枠組みの中で証明したもので，**実用に耐えうる**定理として有名な定理である．いつものように確率空間は $(\Omega, \mathcal{F}, P)$ で表すことにする．

定義 2.4 $X_1, X_2, \ldots$ を確率変数の列とする．このとき，$\{X_n\}$ が確率変数 X に **確率収束する**とは任意の $\varepsilon > 0$ に対して，

$$\lim_{n\to\infty} P\big[|X_n - X| > \varepsilon\big] = 0 \tag{2.6}$$

となるときに言う．

定理 2.2 $\{X_n\}_{n=1}^{\infty}$ を独立で，分布がすべて同じ確率変数の列とする．今，$EX_1 = m, \mathrm{Var}(X_1) = \sigma^2 < \infty$ ならば，

$$S_n = X_1 + \ldots + X_n$$

とおくとき，$\frac{1}{n}S_n$ は期待値 m に確率収束する．つまり，任意の $\varepsilon > 0$ に対して

$$\lim_{n\to\infty} P\big(|\frac{1}{n}S_n - m| > \varepsilon\big) = 0$$

証明 チェビシェフの不等式から[5]

$$P(|S_n - nm| > n\varepsilon) \leq \frac{E\big[|S_n - nm|^2\big]}{n^2\varepsilon^2} = \frac{\mathrm{Var}(S_n)}{n^2\varepsilon^2} \tag{2.7}$$

$\{X_n\}$ が独立なので，問題 4.2 により，

$$\mathrm{Var}(S_n) = \sum_{j=1}^{n} \mathrm{Var}(X_j) = n\sigma^2$$

これを (2.7) に代入して，

$$P(|\frac{1}{n}S_n - m| > \varepsilon) = P(|S_n - nm| > n\varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

$n \to \infty$ のとき，右辺は 0 に収束する．

定理 2.3 確率変数列 $X_1, \ldots, X_n$ が独立ですべて同じ分布を持ち，可積分ならば，$EX_1 = m$ として，$\frac{S_n}{n}$ は m に確率収束する．

証明 $C > 0$ に対して X_n^C を

$$X_n^C(\omega) := \begin{cases} X_n(\omega), & |X_n(\omega)| \leq C \text{ のとき} \\ C, & \text{その他} \end{cases}$$

と定義する．X_n^C は有界なので，２乗可積分で，その期待値と分散が存在．同分布なので，その値は n によらない．平均を m^C，分散を $(\sigma^C)^2$ と書く．$|X_1^C(\omega)| \leq |X_1(\omega)|$ で $C \to \infty$ のとき $X_1^C(\omega) \to X_1(\omega)$ なので，ルベーグの優収束定理により

$$\lim_{C\to\infty} m^C = \lim_{C\to\infty} E[X_1^C] = E[X_1] = m \text{ および } \lim_{C\to\infty} E[|X_1^C - X_1|] = 0$$

いま，任意の $\varepsilon > 0$ に対して，C が十分大きいとき $|m - m^C| < \varepsilon$ となり，$S_n^C = X_1^C + \ldots + X_n^C$ と書くと，

$$|\frac{S_n}{n} - m| \leq |\frac{S_n}{n} - \frac{S_n^C}{n}| + |\frac{S_n^C}{n} - m^C| + |m - m^C|$$

[5] 定理 1.10 で f として $f(x) = x^2 (x \geq 0 \text{ のとき}), = 0 (x < 0 \text{ のとき})$ とおき，確率変数 X として $X = |S_n - nm|$ をとる，このとき $f(X) = f(|S_n - nm|) = |S_n - nm|^2$ となる．

なので，

$$P\big(|\frac{S_n}{n}-m|>\varepsilon\big) \le P\big(|\frac{S_n-S_n^C}{n}|>\frac{\varepsilon-|m-m^C|}{2}\big) + P\big(|\frac{S_n^C}{n}-m^C|>\frac{\varepsilon-|m-m^C|}{2}\big)$$

右辺第２項は 定理 2.2 から $n\to\infty$ のとき 0 に収束する．一方，チェビシェフの不等式から[6]

$$\begin{aligned} \text{右辺第１項} &\le \frac{2E|S_n-S_n^C|}{n(\varepsilon-|m-m^C|)} \\ &\le \sum_{j=1}^{n}\frac{2E|X_j-X_j^C|}{n(\varepsilon-|m-m^C|)} \\ &= \frac{2E|X_1-X_1^C|}{\varepsilon-|m-m^C|} \end{aligned}$$

以上より，

$$0\le \limsup_{n\to\infty} P\big(|\frac{S_n}{n}-m|>\varepsilon\big) \le \frac{2E|X_1-X_1^C|}{\varepsilon-|m-m^C|} \tag{2.8}$$

(2.8) 左辺は C に無関係で $C\to\infty$ のとき (2.8) 右辺は 0 に収束する．

練習問題 2.1　X,Y が確率変数であるとき，組 X,Y の**同時分布** $\mu_{X,Y}$ とは $\mathbf{R}^2$ 上の確率で，任意のボレル集合 A,B に対して

$$\mu_{X,Y}(A\times B)=P(X\in A, Y\in B)$$

によって定義される．ただし，$A\times B$ は

$$A\times B=\{(x,y)\in\mathbf{R}^2; x\in A, y\in B\}$$

で定義される直積図形である．このとき，X, Y が独立ならば，

$$\mu_{X,Y}(A\times B)=\mu_X(A)\mu_Y(B)$$

が成り立つことを示せ．ただし，μ_X,μ_Y はそれぞれ X,Y の分布である．

練習問題 2.2　X,Y が独立で，X^2,Y^2 が可積分のとき，

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$$

がなりたつことを証明せよ．ついでに

$$\mathrm{Var}(X)=E(X^2)-(EX)^2$$

が成り立つことも証明せよ．

練習問題 2.3　大数の法則は同分布でなくても成り立つ．$\{X_n\}$ を独立で期待値が 0 の確率変数列とする．さらにある定数 $C>$ に対して

$$\mathrm{Var}(X_n)=\sigma_n^2\le C \qquad n=1,2,\dots \tag{2.9}$$

とする．

(i)　チェビシェフの不等式を使って

$$P\big(|\frac{S_n}{n}|>\varepsilon\big)\le\frac{C}{n\varepsilon^2}$$

を示せ．

(ii)　このとき大数の弱法則が成立することを証明せよ．

[6] 今度は 定理 1.10 の f として $f(x)=\max\{0,x\}$ をとる．確率変数 X としては $|S_n-S_n^C|$ をとる．このとき，$f(X)=f(|S_n-S_n^C|)=|S_n-S_n^C|$ である．