4.5 大数の法則と中心極限定理

定理 4.15 (大数の (弱) 法則) $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[|X_1|] < \infty$ する．$n \to \infty$ のとき，

$$\bar{X}_n = \frac{1}{n}\sum_{n=1}^{n} X_i \xrightarrow{P} \mu \tag{4.18}$$

が成立する．ただし，$\mathbb{E}[X_1] = \mu$ である．

注意 4.4 $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[X_1^2] < \infty$ する．いま，$\mathbb{E}[X_1] = \mu$, $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ と書くことにする．このとき，任意の正数 ϵ に対して，

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = \frac{1}{\epsilon^2}\mathbb{E}[(\bar{X}_n - \mu)^2] = \frac{1}{\epsilon^2}\mathbb{VAR}[\bar{X}_n] = \frac{1}{\epsilon}\frac{1}{n}\mathbb{VAR}[X_1] \to 0 \qquad (n \to \infty)$$

を得る．ただし，上の不等号はチェビシェフの不等式を利用した．$\mathbb{E}[X_1^2] < \infty$ を仮定すれば，チェビチェフの不等式より (4.18) はわかるが，大数の弱法則は $\mathbb{E}[|X_1|] < \infty$ ならば，(4.18) が成立することを主張している．

例 4.8 $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[X_1] = \mu$, $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ とする．このとき，

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \xrightarrow{P} \sigma^2$$

を示そう．まず，

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2\right)$$

に注意する．$\mathbb{E}[|X_1|] < \infty$ から大数の弱法則を用いれば，$\bar{X}_n \xrightarrow{P} \mu$ になる．さらに，定理 4.9 から $\bar{X}_n^2 \xrightarrow{P} \mu^2$ を得る．また，$\mathbb{E}[X_1^2] = \mathbb{VAR}[X_1] + \{\mathbb{E}[X_1]\}^2 < \infty$ から大数の弱法則を用いれば，

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mu\mathbb{E}[X_1^2]$$

となる．最後に，系 4.3 から $n \to \infty$ のとき，

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \xrightarrow{P} \mathbb{E}[X_1^2] - \mu^2 = \mathbb{VAR}[X_1]$$

を得る．

定理 4.16 (中心極限定理) $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[X_1] = \mu$, $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ とし，

$$Z_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

とする．$n \to \infty$ のとき，

$$Z_n \xrightarrow{d} N(0,\,1)$$

が成立する．すなわち，任意の実数 x に対し，

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$$

が成立する．

例 4.9 $X_1, X_2, \ldots, X_n$ は独立同一に母数 p, $0 < p < 1$ のベルヌーイ分布に従うとする．すると $T_n = \sum_{i=1}^n X_i$ は母数 n と p の二項分布に従う．したがって，$t_1 < t_2$ に対して，

$$\mathbb{P}(t_1 \le T_n \le t_2) = \sum_{x=t_1}^{t_2} f_{T_n}(x)$$

となる．ただし，

$$f_{T_n}(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,\,1,\,2,\ldots,\,n$$

である．この確率の近似値を中心極限定理を利用して求めよう．$\mathbb{E}[X_1] = p$, $\mathbb{VAR}[X_1] = p(1-p)$ と中心極限定理から

$$\mathbb{P}\left(\frac{\sqrt{n}((1/n)T_n - p)}{\sqrt{p(1-p)}} \le x\right) = \mathbb{P}\left(\frac{T_n - np}{np(1-p)} \le x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$$

となる．これより

$$\begin{aligned}
\mathbb{P}(t_1 \le T_n \le t_2) &= \mathbb{P}(t_1 - (1/2) < T_n \le t_2 + (1/2)) \\
&= \mathbb{P}(T_n \le t_2 + (1/2)) - \mathbb{P}(T_n \le t_1 - (1/2)) \\
&= \mathbb{P}\left(\frac{T_n - np}{\sqrt{np(1-p)}} \le \frac{t_2 - np + (1/2)}{\sqrt{np(1-p)}}\right) - \mathbb{P}\left(\frac{T_n - np)}{\sqrt{p(1-p)}} \le \frac{t_2 - np - (1/2)}{\sqrt{np(1-p)}}\right) \\
&\approx \Phi\left(\frac{t_2 - np + (1/2)}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{t_1 - np - (1/2)}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

となる．ただし，

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

とした．

いま，$n = 25$ と $p = 0.2$ と $t_1 = 3$, $t_2 = 5$ とし，$\mathbb{P}(3 \le T_{25} \le 5)$ の近似値を求めよう：

$$\begin{aligned}
\mathbb{P}(3 \le T_{25} \le 5) &= \mathbb{P}(2.5 < T_{25} \le 5.5) \\
&\approx \Phi\left(\frac{5.5 - 5}{2}\right) - \Phi\left(\frac{2.5 - 5}{2}\right) \\
&= \Phi(0.25) - \Phi(-1.25) \fallingdotseq 0.599 - 0.106 \\
&= 0.493
\end{aligned}$$

となる．