3.4　中心極限定理

誤差の集積の分布はは正規分布に近づいていく

いよいよ最後の話題の中心極限定理に移る．最初に独立同分布な確率変数列に関する中心極限定理を少し強い条件の下で証明する．この形ででも十分実用に耐える．

定理 3.8 $X_1, X_2, \ldots$ は独立同分布の確率変数列で， $EX_1 = 0, \mathrm{Var}(X_1) = 1, E|X_1|^3 = m_3 < \infty$ ならば，$S_n = X_1 + \ldots + X_n$ として

$$Z_n = \frac{S_n}{\sqrt{n}}$$

の分布は標準正規分布に収束する．

証明　Z_n の特性関数 $\phi_{Z_n}(u) = E[e^{iuZ_n}]$ は　$\{X_n\}$ が独立同分布であることより

$$\begin{aligned}\phi_{Z_n}(u) = E\left[\exp\{\frac{iu}{\sqrt{n}}\sum_{k=1}^{n} X_k\}\right] &= \prod_{j=1}^{n} E\left[\exp\{\frac{iu}{\sqrt{n}} X_j\}\right] \\ &= \left(\phi_{X_1}(\frac{iu}{\sqrt{n}})\right)^n\end{aligned}$$

と表される．$e^{iz} = 1 + iz - \frac{z^2}{2} + R_n(z)$ と書くと，

$$\begin{aligned}e^{iz} - 1 &= \int_0^z ie^{ix}dx \\ e^{iz} - 1 - iz &= i\int_0^z (e^{ix} - 1)dx = -\int_0^z \int_0^x e^{iy}dydx \\ e^{iz} - 1 - iz + \frac{z^2}{2} &= -\int_0^z \int_0^x (e^{iy} - 1)dydx = -i\int_0^z \int_0^x \int_0^y e^{iw}dwdydx\end{aligned}$$

だから，

$$|R_n(z)| \le \frac{|z|^3}{6}$$

となるので，

$$\begin{aligned}\phi_{Z_n}(u) &= \left(E\big[1 + \frac{iu}{\sqrt{n}}X_1 - \frac{u^2}{2n}X_1^2 + R_n(\frac{iuX_1}{\sqrt{n}})\big]\right)^n \\ &= \left(1 - \frac{u^2}{2n} + ER_n(\frac{iuX_1}{\sqrt{n}})\right)^n\end{aligned}$$

ここで $n \to \infty$ のとき，

$$|ER_n(\frac{iuX_1}{\sqrt{n}})| \le \frac{u^3E|X^3|}{n^{3/2}} = o(\frac{1}{n})$$

なので，

$$\phi_{Z_n}(u) = \left(1 - \frac{u^2}{2n} + o(\frac{1}{n})\right)^n \to e^{-\frac{u^2}{2}}$$

となり， Z_n の分布は標準正規分布 $N(0,1)$ に近づく．

独立，同分布な確率変数列のときは $E|X_1|^3 < \infty$ の条件はなくても中心極限定理が成り立つことが知られている．同分布を仮定しない場合の最も一般的なリンデベルグの定理を紹介して，その応用としてこのことも説明しよう．

定理 3.9 $\{X_n\}$ を独立な確率変数列とする．これらが，以下の条件を満たすものとする．

(i) $EX_n = 0, \quad n = 1, 2, \ldots$

(ii) $\mathrm{Var}(X_n) = \sigma_n^2 < \infty \quad n = 1, 2, \ldots$

(iii) $B_n = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$ と書くとき，任意の $M > 0$ に対して

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{j=1}^{n} E\left(X_j^2;\ |X_j| > MB_n\right) = 0 \tag{3.15}$$

が成り立つならば，$Z_n = \frac{S_n}{B_n}$ の分布は標準正規分布に収束する．

証明は西尾真喜子「確率論」の第六章　§3, P.147–150 を参照．条件 (iii) の (3.15) はリンデベルグの条件としてよく知られている．

例 3.3 $\{X_n\}$ を独立，同分布の確率変数列として，

$$EX_1 = a, \quad \mathrm{Var}(X_1) = \sigma^2$$

のとき，

$$Z_n = \frac{X_1 + \ldots + X_n \quad - na}{\sigma\sqrt{n}}$$

の分布は標準正規分布に収束する．
実際，このこと確かめるには $Y_n = \frac{X_n - a}{\sigma}$ に対して Theorem 3.9 の条件を確

かめればよい．(i), (ii) は明らかに成り立っているので，(iii) つまりリンデベルグの条件を調べればよい．$B_n = \sigma\sqrt{n}$ なので

$$\frac{1}{n\sigma^2}\sum_{j=1}^{n} E(Y_j^2;\ |Y_j| > M\sigma\sqrt{n}) = \frac{1}{\sigma}E(Y_1^2; |Y_1| > M\sigma\sqrt{n})$$

右辺は X_1 の（従って Y_1 の）2 乗可積分性から $n \to \infty$ のとき 0 に収束する．

例 3.4 リャプノフの定理

$\{X_n\}$ が独立で $EX_n = 0$, $\mathrm{Var}(X_n) = \sigma_n^2 < \infty$, $n = 1, 2, \ldots$ とする．ある $\delta > 0$ が存在して，

$$c_n^{2+\delta} = E(|X_n|^{2+\delta}) < \infty \quad n = 1, 2, \ldots$$

とする．$B_n, C_n > 0$ を

$$B_n^2 = \sigma_1^2 + \ldots + \sigma_n^2, \quad C_n^{2+\delta} = c_1^{2+\delta} + \ldots + c_n^{2+\delta}$$

と定めるとき，$\dfrac{C_n}{B_n} \to 0 \quad (n \to \infty)$ ならば，

$$Z_n = \frac{S_n}{B_n}$$

の分布は標準正規分布に収束する．

これもリンデベルグの条件を確かめればよい．Hölder の不等式と *Chebyshev* の不等式により

$$\begin{aligned}
\frac{1}{B_n^2}\sum_{j=1}^{n} E\{X_j^2;\ |X_j| > MB_n\} &\le \frac{1}{B_n^2}\sum_{j=1}^{n} E\{|X_j^{2+\delta}|\}^{2/(2+\delta)}\{P(|X_j| > MB_n)\}^{\delta/(2+\delta)} \\
&\le \frac{1}{B_n^2}\sum_{j=1}^{n} c_j^2\left[E\left(\frac{|X_j|^{2+\delta}}{(MB_n)^{2+\delta}}\right)\right]^{\delta/(2+\delta)} \\
&= \sum_{j=1}^{n}\frac{1}{M^\delta}\frac{c_j^{2+\delta}}{B_n^{2+\delta}} \\
&= \frac{1}{M^\delta}\left(\frac{C_n}{B_n}\right)^{2+\delta} \to 0
\end{aligned}$$

例 3.5 Y_n が 2 項分布 $B_{n,p}$ に従うとき,

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

の分布は標準正規分布に近づく.

これは 0 または 1 をとる確率変数 X_n が同じ分布

$$P(X_n = 1) = p, \quad P(X_n = 0) = 1 - p$$

をとるとき $S_n = X_1 + \ldots + X_n$ の分布が Y_n と同じ $B(n, p)$ に従うことをまず確認する.

$$\varphi_{S_n}(t) = E(e^{itS_n}) = E(e^{it\sum_{j=1}^n X_j}) = (pe^{it} + 1 - p)^n = (1 + p(e^{it} - 1))^n$$

だから,たしかに S_n の分布は $B(n, p)$. 期待値を 0, 分散を 1 にするため $\tilde{X}_n = \frac{X_n - p}{\sqrt{p(1-p)}}$ とおくと,$\tilde{X}_n$ は明らかに定理 3.8 の条件を満たしているので $\tilde{S}_n = \sum_{j=1}^n \tilde{X}_j = \frac{\sum_{j=1}^n X_j - np}{\sqrt{p(1-p)}}$ に対して

$$\frac{\tilde{S}_n}{\sqrt{n}} = \frac{S_n - np}{\sqrt{np(1-p)}}$$

の分布は標準正規分布に近づく.