·····A.3·····

大数の法則と中心極限定理

ここでは，大数の法則と中心極限定理を証明なしで述べる．さらに，対称分布からのランダム標本の基づくメディアンは分布の中心に確率収束し，漸近正規性をもつことをこの二つの定理を用いて示す．

定理 A.11　$X_1, X_2, \ldots, X_n$ は独立に同一の分布[1]に従うとする．また，

$$\mathbb{E}|X_1| < \infty, \qquad \mathbb{E}[X_1] = \xi$$

であるとする．このとき，

$$\bar{X}_n \xrightarrow{a.s.} \xi$$

となる．

証明　略．

定理 A.12　$\{X_n\}_{n\geq 1}$ は独立同一の分布に従う確率変数の列とし，

$$\mathbb{E}(X_1) = \xi, \qquad \mathbb{VAR}(X_1) = \sigma^2$$

を満足するものとする．このとき，

$$\frac{\sqrt{n}(\bar{X}_n - \xi)}{\sigma} \xrightarrow{L} N(0,1)$$

が成立する．

証明　概略のみを示す．$Y \sim N(0,1)$ とすれば，

$$\mathbb{E}[e^{\sqrt{-1}tY}] = \exp(-\frac{t^2}{2})$$

である．また，

$$\begin{aligned}\mathbb{E}\left[\exp\left(\sqrt{-1}t\frac{\sqrt{n}(\bar{X}_n - \xi)}{\sigma}\right)\right] &= \Pi_{i=1}^n \mathbb{E}\left[\exp\left\{\sqrt{-1}\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_i - \xi}{\sigma}\right)\right\}\right] \\ &= \left[\mathbb{E}\exp\left\{\sqrt{-1}\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_1 - \xi}{\sigma}\right)\right\}\right]^n\end{aligned}$$

[1] 同一ではなくい独立な確率変数列にたいしては，2 次の積率が有限であれば，同様な結果が成立することが知られている．

となる．また，

$$\mathbb{E}\left[\exp\left\{\sqrt{-1}\left(\frac{t}{\sqrt{n}}\right)\left(\frac{X_1-\xi}{\sigma}\right)\right\}\right]=1-\frac{t^2}{2n}+o(\frac{t^2}{n})$$

より

$$\mathbb{E}\left[\exp\left(\sqrt{-1}t\frac{\sqrt{n}(\bar{X}_n-\xi)}{\sigma}\right)\right]=\left[1-\frac{t^2}{2n}+o(\frac{t^2}{n})\right]^n\to e^{-\frac{t^2}{2}}$$

となり，定理 1.15 からわかる．□

定理 A.13　$X_{n1}, X_{n2}, \dots, X_{nn}$ は独立同一分布に従い，各 $X_{ni}, i=1, 2, \dots, n,$ は平均 ξ_n と分散 σ_n^2 なる分布関数 $F_n(x)$ をもち，さらに 3 次の積率も有限とする．このとき，

$$\frac{\mathbb{E}_n\{|X_{n1}-\xi_n|^3\}}{\sigma_n^3}=o(\sqrt{n}) \tag{1.3}$$

ならば，

$$\frac{\sqrt{n}(\bar{X}_n-\xi_n)}{\sigma_n}\xrightarrow{d}N(0,\,1)$$

となる．特に，(1.3) が有限ならば，上の結果は成立する．ただし，$\mathbb{E}_n$ は F_n に関する期待値で，$\bar{X}_n=(1/n)(X_{n1}+X_{n2}+\cdots+X_{nn})$ である．

証明　これは Berry - Essenn の定理

$$\mathbb{P}\left[\frac{\sqrt{n}(\bar{X}_n-\xi_n)}{\sigma_n}\le\right]-\Phi(x)\le\frac{C}{\sqrt{n}}\frac{\mathbb{E}_n\{|X_{n1}-\xi_n|^3\}}{\sigma_n^3}$$

よりわかる．ただし，$\Phi(x)$ は標準正規分布の分布関数である．□

例 A.4　$X_1, X_2, \dots, X_n$ を $F_\theta(x)=P_\theta(X_1\le x)=F(x-\theta)\,(x\in\mathbb{R})$ からのランダム標本とする．ただし，θ は未知の定数とし，$F(x)$ は狭義単調増加で連続とし，$F(0)=1/2$ とする．$X_1, X_2, \dots, X_n$ に基づいて，メディアン θ の推定を考える．

いま，

$$m_n=\begin{cases}\frac{n+1}{2}, & (n\text{ が奇数のとき}) \\ \frac{n}{2}, & (n\text{ が偶数のとき})\end{cases}$$

とおく．順序統計量を $X_{(1)}\le\cdots\le X_{(n)}$ としたとき，

$$Z_n=X_{(m_n)}$$

を**標本メディアン**という．これを用いて，メディアン θ を推定することを考える．標本メディアン Z_n の漸近的な性質を調べてみよ．

まず，$Z_n\xrightarrow{P}\theta$ を示そう．任意の $\epsilon>0$ に対して，標本メディアンの定義から

$$P(Z_n>\theta+\epsilon)=P\left(\frac{1}{n}\sum_{i=1}^n I\{X_i>\theta+\epsilon\}\ge\frac{n-m_n+1}{n}\right)$$

$$P(Z_n<\theta-\epsilon)=P\left(\frac{1}{n}\sum_{i=1}^n I\{X_i>\theta-\epsilon\}\le\frac{n-m_n}{n}\right)$$

となる．いま，

$$2\delta = \min\left\{F_\theta(\theta+\epsilon) - \frac{1}{2}, \frac{1}{2} - F_\theta(\theta-\epsilon)\right\}$$

とおけば，$F_\theta(\theta) = 1/2$ に注意すれば，$\delta > 0$ となることがわかる．

$$A_n = \left\{\left|\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} - (1 - F_\theta(\theta+\epsilon))\right| \le \delta\right\}$$

$$B_n = \left\{\left|\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta-\epsilon\} - (1 - F_\theta(\theta-\epsilon))\right| \le \delta\right\}$$

とおけば，大数の法則から $n \to \infty$ のとき

$$P(A_n) \to 1, \qquad P(B_n) \to 1 \tag{1.4}$$

となる．また，$\frac{1}{n} < \delta$ ならば，A_n が起これば，

$$\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} \le \delta + 1 - F_\theta(\theta+\epsilon) = \delta + \frac{1}{2} + \frac{1}{2} - F_\theta(\theta+\delta) \le \frac{1}{2} - 2\delta + \delta$$

$$\le \delta + \frac{1}{2} - 2\delta = \frac{1}{2} - \delta < \frac{n - m_n + 1}{n}$$

となることが $2\delta \le F_\theta(\theta+\epsilon) - (1/2)$ に注意すればわかる．これより，事象

$$\left\{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} \ge \frac{n - m_n + 1}{n}\right\}$$

と A_n は排反になることがわかる．したがって，

$$C_n = A_n \cap \left\{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta+\epsilon\} \ge \frac{n - m_n + 1}{2}\right\} = \emptyset$$

となることと (1.4) から，$n \to \infty$ のとき，

$$P(Z_n > \theta+\epsilon) \le P(C_n) + P(A_n^c) \to 0$$

となる．同様に，$\frac{1}{n} < \delta$ ならば，B_n が起これば，

$$\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta-\epsilon\} \ge -\delta + 1 - F_\theta(\theta-\epsilon) = -\delta + \frac{1}{2} + \frac{1}{2} - F(\theta-\delta) \ge -\delta + \frac{1}{2} + 2\delta$$

$$= \frac{1}{2} + \delta > \frac{n - m_n}{n}$$

となることが $2\delta \le (1/2) - F_\theta(\theta-\epsilon)$ に注意すればわかる．これより，事象

$$\left\{\frac{1}{n}\sum_{i=1}^{n} I\{X_i > \theta-\epsilon\} \le \frac{n - m_n}{n}\right\}$$

と B_n は排反になることがわかる．したがって，

$$D_n = B_n \cap \left\{ \frac{1}{n} \sum_{i=1}^{n} I\{X_i > \theta - \epsilon\} \leq \frac{n - m_n}{n} \right\} = \emptyset$$

となることがわかる．さらに，(1.4) から，$n \to \infty$ のとき，

$$P(Z_n < \theta - \epsilon) \leq P(D_n) + P(B_n^c) \to 0$$

となる．これらから，$n \to \infty$ のとき

$$P(|Z_n - \theta| > \epsilon) \leq P(Z_n > \theta + \epsilon) + P(Z_n < \theta - \epsilon) \to 0$$

となり，$Z_n \xrightarrow{P} \theta$ がわかる．

つぎに，$f = dF/dx$ とし，P_0 を $\theta = 0$ のときに対応する確率測度とする．すなわち，$P(X - \theta \leq x) = P_0(X \leq x)$ である．$f(0) > 0$ の仮定もとで，Z_n の極限分布を求める：

$$P[\sqrt{n}(Z_n - \theta) \leq t] = P_0[\sqrt{n} Z_n \leq t] = P_0 \left[X_{(m_n)} \leq \frac{t}{\sqrt{n}} \right]$$

S_n を $t/\sqrt{n}$ を越える $X_i (i = 1, 2, \ldots, n)$ の個数とすれば，

$$X_{(m_n)} \leq \frac{t}{\sqrt{n}} \Leftrightarrow S_n \leq n - m_n$$

となる．S_n は二項分布 $Bi(n, p_n)$ に従う．ただし，$p_n = 1 - F(t/\sqrt{n})$ である．したがって，

$$P_0 \left[S_n \leq n - m_n \right] = P_0 \left[\frac{S_n - np_n}{\sqrt{np_n(1 - p_n)}} \leq \frac{(n - m_n) - np_n}{\sqrt{np_n(1 - p_n)}} \right]$$

となる．$p_n \to 1 - F(0) = 1/2 (n \to \infty)$ より，

$$\begin{aligned} \frac{(n - m_n) - np_n}{\sqrt{np_n(1 - p_n)}} &= \frac{\sqrt{n}\left(\frac{1}{2} - p_n\right) + \frac{(n/2) - m_n}{\sqrt{n}}}{\sqrt{p_n(1 - p_n)}} \\ &\sim 2\sqrt{n}\left(\frac{1}{2} - p_n\right) \\ &= 2t \frac{F(t/\sqrt{n}) - F(0)}{t/\sqrt{n}} \to 2tf(0) \end{aligned}$$

となることに注意する．いま，$Y_{ni} = I\{X_i \geq t/\sqrt{n}\}$, $i = 1, 2, \ldots, n$, とおけば，$Y_{n1}, Y_{n2}, \ldots, Y_{nn}$ は独立同一分布に従い，$P_0(Y_{ni} = 1) = p_n$ かつ $P_0(Y_{ni} = 0) = 1 - p_n$ となる．また，$S_n = \sum_{i=1}^{n} Y_{ni}$ となる．定理 $A.13$ を用いて

$$\frac{S_n - np_n}{\sqrt{np_n(1 - p_n)}} \xrightarrow{d} N(0, 1)$$

を示す．そのために，(1.3) を確認する：$n \to \infty$ のとき，$p_n \to (1/2)$ になることに注意しれば，

$$\mathbb{E}|Y_1 - p_n|^3 = p_n(1 - p_n)^3 + (1 - p_n)p_n^3 = p_n(1 - p_n)\{p_n^2 + (1 - p_n)^2\} \to \frac{1}{8}$$

と $\sigma_n = \sqrt{p_n(1-p_n)} \to (1/2)$ となるので，

$$\frac{\mathbb{E}|Y_1 - p_n|^3}{\sigma_n^3} \to 1$$

がわかる．よって，(1.3) を確認は確認された．したがって，

$$P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \le 2tf(0)\right] \to \Phi(2f(0)t)$$

となる．ただし，$\Phi(\cdot)$ は標準正規分布の分布関数である．さらに，Slutsky の補題（定理 $A.9$ ）を用いれば

$$\begin{aligned}
&P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \le \frac{(n-m_n) - np_n}{\sqrt{np_n(1-p_n)}}\right] \\
&= P_0\left[\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} + \left\{(2f(0)t) - \frac{(n-m_n) - np_n}{\sqrt{np_n(1-p_n)}}\right\} \le 2f(0)t\right] \\
&\to \Phi(2f(0)t)
\end{aligned}$$

となることがわかる．したがって

$$P_\theta[\sqrt{n}(Z_n - \theta)] \to \Phi(2tf(0))$$

となる．したがって，

$$\sqrt{n}(Z_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{4f^2(0)}\right)$$

をえる．