平均をめぐって
-高校数学教育における平均の確率分布と信頼性-

杉山　高一*　，兵頭 昌†

統計学では母集団に正規分布や 2 項分布などの確率分布を想定する．これを母集団分布と呼ぶ．一般に，確率分布は未知の母数 (パラメータ) を含むから，母集団からの無作為標本に基づいて，未知の母数を推測することになる．これを統計的推測と呼ぶ．例えば，母集団に正規分布 $\mathcal{N}(\mu, \sigma^2)$ を仮定すると，正規分布は 2 つのパラメータ μ, σ^2 を含むから，これらのパラメータを標本に基づいて推測することにより母集団分布が規定され，母集団の性質・特徴を把握することが可能となる．本論文では, 特に高校の教科書で扱われる母集団分布がベルヌーイ分布である場合の平均の区間推定についての詳細を報告する．

Key words and phrases:　区間推定, 漸近正規性

1. 序

母数に関する推測は，無作為標本 $X_1, X_2, \cdots, X_n$ の関数を用いて行われる．これを統計量という．例えば,

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

$$S^2 = \frac{1}{n-1}\{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\}$$

は統計量であり，それぞれ標本平均，標本分散と呼ばれる．また，S を標本標準偏差という．標本平均と標本分散は，母集団分布の平均 (母平均) μ と分散 (母分散) σ^2 を推測するのに用いられる．母集団分布のパラメータを推測する際に，そのパラメータに対応する統計量の分布を知ることが必要になってくる．統計量の変動の様子を表す分布を標本分布と呼ぶ．$\mathrm{E}[X] = \mu, \mathrm{Var}[X] = \sigma^2$ である母集団分布より，無作為標本 $X_1, X_2, \cdots, X_n$ が得られたとする. このとき,

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[X_i\right] = \mu$$

となる．これを, 不偏性という．これは, 平均的に過大・過小の推定がないことを意味している．

また, 標本平均の分散は

$$\begin{aligned}\mathrm{Var}[\bar{X}] &= \mathrm{E}\left[\frac{1}{n^2}\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}\left[(X_i - \mu)^2\right] + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1}\mathrm{E}\left[(X_i - \mu)\right]\mathrm{E}\left[(X_j - \mu)\right] = \frac{\sigma^2}{n} \qquad (1)\end{aligned}$$

* 創価大学, 〒 192-8577, 住所：東京都八王子市丹木町 1-236, E-Mail:takakazusugiyam@yahoo.co.jp
† 東京理科大学理学部, 〒 162-8601, 東京都新宿区神楽坂 1-3, E-Mail:hyodoh_h@yahoo.co.jp

と計算できる．この式より, 標本の大きさ n が大きくなると $\bar{X}$ の分散は小さくなり，分布は μ のまわりに集中してくることがわかる．このことは，n が大きいほど μ のより精確な推測が可能なことを意味している．この意味は, 確率収束という概念と関係する．任意の $\varepsilon > 0$ に対して,

$$\Pr\{|\bar{X}-\mu| \geq \varepsilon\} \to 0 \ as \ n \to \infty \tag{2}$$

が成立するときに $\bar{X}$ は μ へ確率収束するといい $\bar{X} \xrightarrow{P} \mu$ と書く．この性質を有する推定量を, 一致推定量と呼ぶ．いま,

$$\begin{aligned}
\Pr\{|\bar{X}-\mu| \geq \varepsilon\} &= \mathrm{E}\left[I_{\{|\bar{X}-\mu|\geq\varepsilon\}}\right] \\
&\leq \mathrm{E}\left[I_{\{|\bar{X}-\mu|\geq\varepsilon\}}\frac{|\bar{X}-\mu|^2}{\varepsilon^2}\right] \\
&\leq \mathrm{E}\left[\frac{|\bar{X}-\mu|^2}{\varepsilon^2}\right] = \frac{\mathrm{Var}[\bar{X}]}{\varepsilon^2}
\end{aligned}$$

である．ただし, I は指示関数であり, $\{\}$ 内の命題が真であるときに 1 を返し, しからざれば 0 を返す．一方, (1) 式より $\lim_{n\to\infty}\mathrm{Var}[\bar{X}] = 0$ であるから, (2) 式が成立することがわかる．つまり, $\lim_{n\to\infty}\mathrm{Var}[\bar{X}] = 0$ は (2) 式が成立するための十分条件となっていることに気付く．

次に, 標本平均の分布について議論する．一般に, 標本平均の正確な分布は, 母集団分布に依存する．実は, 標本の大きさが十分大きい場合, $\bar{X}$ の μ 周りの分布は, 適当な条件の下で正規分布に近づく．この事実を $\bar{X}$ の漸近正規性という．また, $\bar{X}$ の漸近正規性は, $X_1, X_2, \cdots, X_n$ が独立, μ と σ が有限であれば成立し, X_i の分布に因らないということが大きい点である．本講演では, 母集団分布が正規分布の場合を議論したが, ここでは母集団分布がベルヌーイ分布である場合も $\bar{X}$ の漸近正規性が成立することを示し, その事実を利用した近似区間推定法を与える．尚, 高校の教科書においてこの近似区間推定法が掲載されていることもあるが, ここではどのような意味の近似なのかを明らかにする．2 節において母集団分布にベルヌーイ分布を想定した場合の標本平均の漸近正規性を示す．さらに, その結果を利用した比率の近似信頼区間を導出する．もちろんこれらの結果に新規性はないが, ここでは, 高校の教科書で紹介される近似区間推定がどのように導き出されるかを概説することに重点を置く．

2. 母集団分布がベルヌーイ分布の場合の標本平均とその漸近的性質

母集団の中で, ある性質をもつものの比率 p の区間推定を考える．区間推定の場合, 事前に適当な $\alpha \in (0,1)$ を決めておき,

$$\Pr\{L \leq p \leq U\} = 1-\alpha$$

を満たすような $X_1, \ldots, X_n$ の関数 L と U を与える事を目標とする．このようにして得られる区間 $[L, U]$ を, $100(1-\alpha)$ の信頼区間といい, 信頼の精度 $100(1-\alpha)$ を信頼度という．高校の教科書に掲載されている信頼度 $100(1-\alpha)$ の近似信頼区間は

$$\left[\bar{x} - z_{\alpha/2}\frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sqrt{\bar{x}(1-\bar{x})}}{\sqrt{n}}\right]$$

である．ただし, z_α は上側 $100\alpha\%$ 点である．例えば, 失業率調査において，労働力人口の 10000 人を調査したところ 500 人が失業していた．母集団失業率 p の信頼係数 95% の近似信頼区間は,

$$0.05 - 1.96\frac{\sqrt{0.05\times 0.95}}{100} < p < 0.05 + 1.96\frac{\sqrt{0.05\times 0.95}}{100} = [0.0457, 0.0543]$$

と算出される．

ところで, 近似といったがどのような意味を持つ近似であるかについて考えてみる．結論を先に述べると, 被覆確率

$$\Pr\left\{\bar{X} - z_{\alpha/2}\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \ \le\ p\ \le\ \bar{X} + z_{\alpha/2}\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}\right\}$$

が $n\to\infty$ とすれば信頼度 $1-\alpha$ へ収束するという意味を持つ近似である．この事実を示すには, $\bar{X}$ をスチューデント化した統計量の極限分布が正規分布であることを示す必要がある．本論文では, 数理統計学における幾つかの有用な定理を紹介しつつこの事実を導く．

ある性質を持てば 1, もたなければ 0 の値をとるような確率変数を X と定義すると，母集団における X の確率分布は以下で与えられるベルヌーイ分布と呼ばれる確率分布になる．

$$\Pr\{X=1\}=p,\ \Pr\{X=0\}=1-p$$

ここで, ベルヌーイ分布に従う確率変数は, 2 値 (例えば, 0 または,1) を取り, 期待値は $\mathrm{E}[X]=p$, 分散は $\mathrm{Var}[X]=p(1-p)$ となる．このとき, p の推定量として, $\bar{X}$ が用いられる．$\bar{X}$ は, 不偏性・一致性に加え以下の定理で与えられる漸近正規性を有する．この定理は, 2 項分布の正規分布への収束をいっておりド・モアブル-ラプラスの定理と呼ばれる．オリジナルの証明は, 局所極限定理と積分型極限定理に分けられる．前者から後者は比較的簡単に導くことができるが, 前者の証明にはスターリングの公式のを用いて 2 項分布の漸近性についての相当に詳しい評価を行う必要がある．一方で, ド・モアブル-ラプラスの定理は, 中心極限定理の特別な場合にあたる．最初に中心極限定理を数学的に証明に示したのは, Aleksandr Lyapunov (1857 年 6 月 6 日-1918 年 11 月 3 日) であると言われている．本論文では, リャプノフ型中心極限定理を利用した証明方法を紹介する．

定理 1 ($\bar{X}$ の漸近正規性) $p\in(0,1)$ を固定し, $n\to\infty$ の下で,

$$Y_n = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する．

(証明) リャプノフ型中心極限定理 (付録の定理 A.4 を参照) を利用する．$q\equiv 1-p$ とすると,

$$\begin{aligned} B_n &= \left(\sum_{i=1}^n \mathrm{E}[|X_i-p|^4]\right)^{1/4} = n^{1/4}p^{1/4}q^{1/4}\{(q-p)^2+pq\}^{1/4}, \\ C_n &= \left(\sum_{i=1}^n \mathrm{E}[(X_i-p)^2]\right)^{1/2} = n^{1/2}p^{1/2}q^{1/2} \end{aligned}$$

であるので, $B_n/C_n = n^{-1/4}(pq)^{-1/4}\{(q-p)^2+pq\}^{1/4}$ となる. したがって, $\lim_{n\to\infty}(B_n/C_n) = 0$ なので, リャプノフ型中心極限定理より

$$Y_n = \frac{\sum_{i=1}^{n}(X_i - p)}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する. □

この漸近正規性に基づき, 標準正規分布の上側 $100(\alpha/2)\%$ 点 $z_{\alpha/2}$ を近似として用いる近似信頼区間

$$\left[\bar{X} - z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \ \le\ p \ \le\ \bar{X} + z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

を考えることができる. 定理1で与えれらる漸近正規性より, この信頼区間の被覆確率

$$\Pr\left\{\bar{X} - z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \ \le\ p \ \le\ \bar{X} + z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right\}$$

は $p \in (0,1)$ を固定し $n \to \infty$ の下で, 信頼度 $1-\alpha$ へ収束する. つまり, この近似信頼区間は, 標本の大きさが大きければ大きいほど良い近似となっている.

しかしながら, $p(1-p)$ は神でない限り知ることが出来ない. よって, 手元の標本から $p(1-p)$ を見積もる必要がある. 推定量として,

$$\bar{X}(1-\bar{X})$$

を考える. 高校の教科書では n が大きいとき $\bar{X}(1-\bar{X})$ は粗 $p(1-p)$ とみなせるという理由で置き換えを正当化しているが, 厳密ではない. そもそも $\bar{X}(1-\bar{X})$ は確率変数である. 厳密に近似を正当化するには, 定理1における確率変数の分母の $p(1-p)$ をこの推定量で置き換えた統計量

$$Z_n = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}}$$

の極限分布を考える必要がある. この統計量を本論文では, スチューデント化統計量と呼ぶ. スチューデント化統計量の極限分布を論じる前に幾つか準備を行う. $\bar{X}(1-\bar{X})$ の推定量としての性質を見る必要がある. 以下の補題は $\bar{X}(1-\bar{X})$ の一致性を与えている.

補題 1 $p \in (0,1)$ を固定し, $n \to \infty$ の下で,

$$\bar{X}(1-\bar{X}) \xrightarrow{P} p(1-p)$$

が成立する.

(証明) $\bar{X}(1-\bar{X})$ の $p(1-p)$ 周りの2次モーメントが0へ収束することが言えれば十分である. $n\bar{X}$ が2項分布に従うことを利用しモーメント評価を行うと,

$$\begin{aligned}
&\mathrm{E}\left[\frac{|\bar{X}(1-\bar{X}) - p(1-p)|^2}{\varepsilon^2}\right] \\
&= \frac{(1-p)p\left[n^2(1-2p)^2 + n\{11(1-p)p-2\} - 6(1-p)p + 1\right]}{\varepsilon^2 n^3} \\
&\to 0 \ as \ n \to \infty
\end{aligned}$$

をえる．よって, 補題が示された． □

補題 1 と定理 A.3(連続写像定理) より,

$$\sqrt{\frac{\bar{X}(1-\bar{X})}{p(1-p)}} \xrightarrow{P} 1 \ as \ n \to \infty \tag{3}$$

をえる．定理 1[$\bar{X}$ の漸近正規性] より,

$$\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \ as \ n \to \infty \tag{4}$$

をえる．ここで,

$$Z_n = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}} \Big/ \sqrt{\frac{\bar{X}(1-\bar{X})}{p(1-p)}}$$

であることに注意すると, (3), (4) と定理 A.1(スラツキーの定理) より次の結果を得る．

定理 2 $p \in (0,1)$ を固定し, $n \to \infty$ の下で,

$$Z_n = \frac{\sqrt{n}(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する．

注意 1 高校の教科書に掲載されている近似信頼区間は, 定理で与えられる極限分布である正規分布の上側 α パーセント点を利用した近似であることが判る．

また, 定理 2 より次の系をえる．

系 1 $p \in (0,1)$ を固定し, $n \to \infty$ の下で,

$$\Pr\left\{\bar{X} - z_{\alpha/2}\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \ \le \ p \ \le \ \bar{X} + z_{\alpha/2}\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}\right\} \to 1-\alpha$$

が成立する．

この系から, 冒頭で述べた「$n \to \infty$ とすると被覆確率が信頼度 $1-\alpha$ へ収束する」という事実が示されている．つまり, この近似はサンプルサイズ n が十分大きい場合に精度が良い近似となる．

付録

ここでは, 中心極限定理の証明のために幾つかの定義及び重要な結果を紹介する．

まずはじめに, 分布収束及び関連する有用な結果を幾つか紹介する．

定義 1 (分布収束) $\{X_1, X_2, \ldots\}$ を確率変数列, X を確率変数, $F_X(x)$ を X の分布関数とする．このとき, $F_X(x)$ の任意の連続点において

$$\lim_{n\to\infty} \Pr\{X_n \le x\} = F_X(x)$$

であることを X_n が X へ分布収束するといい,

$$X_n \xrightarrow{\mathcal{D}} X$$

と表す．

以下の２つの定理は, 分布収束や確率収束に絡んだ便利な結果である.

定理 A. 1 (スラツキーの定理) 実定数 $c \neq 0$, 確率変数 X, および２つの確率変数の列 $\{X_1, X_2, \ldots\}$, $\{Y_1, Y_2, \ldots\}$ について,

$$X_n \xrightarrow{\mathcal{D}} X,\ Y_n \xrightarrow{P} c \quad \Rightarrow \quad \frac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \frac{X}{c}$$

が成立する.

(証明) 一般性を失うことなく $c > 0$ とし, X および $Z_n \equiv \frac{X_n}{Y_n}$ の分布関数をそれぞれ $F(x)$, $H_n(x)$ とする. 定理の条件の下で,

$$\lim_{n\to\infty} H_n(x) = F(cx) \tag{5}$$

を示せばよい. $0 < {}^\forall\varepsilon < c$ に対して,

$$\begin{aligned}
H_n(x) &= \Pr\left\{\frac{X_n}{Y_n} \le x\right\} \\
&= \Pr\left\{\frac{X_n}{Y_n} \le x,\ |Y_n - c| \le \varepsilon\right\} + \Pr\left\{\frac{X_n}{Y_n} \le x,\ |Y_n - c| > \varepsilon\right\} \\
&= \Pr\{X_n \le xY_n,\ c-\varepsilon \le Y_n \le c+\varepsilon\} + \Pr\left\{\frac{X_n}{Y_n} \le x,\ |Y_n - c| > \varepsilon\right\} \qquad (6)\\
&\le \begin{cases} \Pr\{X_n \le x(c+\varepsilon)\} + \Pr\{|Y_n - c| > \varepsilon\} & \text{if } x \ge 0 \\ \Pr\{X_n \le x(c-\varepsilon)\} + \Pr\{|Y_n - c| > \varepsilon\} & \text{if } x < 0 \end{cases} \qquad (7)
\end{aligned}$$

となる. よって,

$$\limsup_{n\to\infty} H_n(x) \quad \le \quad F(cx) \tag{8}$$

を得る. 一方 (6) の表現を再評価することで, (7) とは逆の不等式

$$\begin{aligned}
H_n(x) &= \Pr\{X_n \le xY_n,\ c-\varepsilon \le Y_n \le c+\varepsilon\} + \Pr\left\{\frac{X_n}{Y_n} \le x,\ |Y_n - c| > \varepsilon\right\} \\
&\ge \Pr\{X_n \le xY_n,\ c-\varepsilon \le Y_n \le c+\varepsilon\} \\
&\ge \begin{cases} \Pr\{X_n \le x(c-\varepsilon)\} - \Pr\{|Y_n - c| > \varepsilon\} & \text{if } x \ge 0 \\ \Pr\{X_n \le x(c+\varepsilon)\} - \Pr\{|Y_n - c| > \varepsilon\} & \text{if } x < 0 \end{cases}
\end{aligned}$$

となる. よって,

$$\liminf_{n\to\infty} H_n(x) \quad \ge \quad F(cx) \tag{9}$$

をえる. (8) と (9) より (5) が示された. □

定理 A. 2 (連続写像定理) 確率変数の列を $\{X_1, X_2, \ldots\}$ とし, I を区間とし $c \in I$ とする. I で連続な関数を $g(x)$ とする. このとき,

$$X_n \xrightarrow{P} c \ \Rightarrow\ g(X_n) \xrightarrow{P} g(c)$$

である.

(証明) $[c-\eta,\ c+\eta] \subseteq I$ となるような $\eta > 0$ をとる. 任意の $\varepsilon > 0$ に対して,

$$
\begin{aligned}
&\Pr\{|g(X_n) - g(c)| > \varepsilon\} \\
&= \Pr\{|g(X_n) - g(c)| > \varepsilon,\ X_n \in [c-\eta, c+\eta]\} + \Pr\{|g(X_n) - g(c)| > \varepsilon,\ X_n \notin [c-\eta, c+\eta]\} \\
&\leq \Pr\{|g(X_n) - g(c)| > \varepsilon,\ X_n \in [c-\eta, c+\eta]\} + \Pr\{|X_n - c| > \eta\}
\end{aligned}
$$

となる. 仮定より, $\Pr\{|X_n - c| > \eta\} \to 0$ なので,

$$\Pr\{|g(X_n) - g(c)| > \varepsilon,\ X_n \in [c-\eta, c+\eta]\} \to 0$$

が示せれば十分である. 簡単のため,

$$\Pr\{|g(X_n) - g(c)| > \varepsilon,\ X_n \in [c-\eta, c+\eta]\},\ \Pr\{|g(X_n) - g(c)| \leq \varepsilon,\ X_n \in [c-\eta, c+\eta]\}$$

を, それぞれ, 単に

$$\Pr\{|g(X_n) - g(c)| > \varepsilon\},\ \Pr\{|g(X_n) - g(c)| \leq \varepsilon\}$$

と書く. g は連続なので,

$$^{\forall}\varepsilon > 0, ^{\exists}\delta > 0\ s.t.\ |X_n - c| \leq \delta \Rightarrow |g(X_n) - g(c)| \leq \varepsilon$$

なので, $\{X_n \| X_n - c| \leq \delta\} \subseteq \{X_n \| g(X_n) - g(c)| \leq \varepsilon\}$. すなわち,

$$
\begin{aligned}
&\Pr\{|X_n - c| \leq \delta\} \leq \Pr\{|g(X_n) - g(c)| \leq \varepsilon\} \\
\Rightarrow\ &\Pr\{|X_n - c| > \delta\} \geq \Pr\{|g(X_n) - g(c)| > \varepsilon\}
\end{aligned}
$$

となる. 一方, $\Pr\{|X_n - c| > \delta\} \to 0$ であったので, はさみうちの原理より, 所望の結果を得る. □

次に, 特性関数と関連するトピックを幾つか紹介する. 特性関数は積率を次々と生み出す機能を持つ関数である. そのために積率の母関数とも呼ばれる. 積率母関数は実数領域で定義される関数を指し示すことが多いが, 実数領域ではコーシー分布のように積率母関数が定義できない分布があるため, ここでは複素数領域に踏み込んだ特性関数で議論する.

定義 2 (特性関数) 確率変数 X の特性関数とは,

$$\varphi(t) = \mathrm{E}[\mathrm{e}^{itX}],\ t \in \mathbb{R}$$

によって定義される実数変数による複素数値をとる関数をいう.

特性関数が確率変数の積率を次々生みだすことは, 特性関数が確率変数の分布に深く関係していることを予想させる. 実際に, ある分布の極限分布が知りたければ, その特性関数の極限を見てやればよい. 本論文ではこの事実を認めたうえで, Lindeberg-Feller の中心極限定理の証明を与えた.

定理 A. 3 (Lindeberg-Feller の中心極限定理) $X_1, X_2, \cdots, X_n$ を分布関数 G_i に従う独立な確率変数列とする. いま, 各 i に対して, $\mathrm{E}[X_i] = \mu_i$, $\mathrm{E}[(X_i - \mu_i)^2] = \sigma_i \neq 0$ が存在するとする. さらに,$C_n = \left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}$ とするとき任意の $\varepsilon > 0$ に対して

$$(L).\ \lim_{n\to\infty} \frac{1}{C_n^2} \sum_{i=1}^{n} \mathrm{E}\left[|X_i - \mu_i|^2\ I_{\{|X_i - \mu_i| > \varepsilon C_n\}}\right] = 0$$

が成り立つとする．このとき,

$$Y_n = \frac{\sum_{i=1}^n (X_i - \mu_i)}{C_n}$$

の分布関数は$\mathcal{N}(0,1)$の分布関数へ収束する．

(証明) $\varphi(t) \to \mathrm{e}^{-t^2/2}\ as\ n \to \infty$ を示せば十分である．$Z_i = X_i - \mu_i\ (i = 1, 2, \cdots, n)$ とおく．いま

$$\begin{aligned} \varphi(t) &= \mathrm{E}[\mathrm{e}^{itY_n}],\ t \in \mathbb{R} \\ &= \mathrm{E}[\mathrm{e}^{it\sum_{i=1}^n Z_i/C_n}] = \prod_{i=1}^n \mathrm{E}[\mathrm{e}^{itZ_i/C_n}] \end{aligned}$$

である．そこで,

$$\left|\prod_{i=1}^n \mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - e^{-t^2/2}\right| = \left|\prod_{i=1}^n \mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - \prod_{i=1}^n \mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}\right|$$

の評価を考える．

評価の際には, 次の公式を用いる．a_i,b_i を複素数とすると（実数でもよい）, $n \in \mathbb{N}$ に対して,

$$\prod_{i=1}^n a_i - \prod_{i=1}^n b_i = \sum_{i=1}^n \prod_{j=0}^{i-1} a_j (a_i - b_i) \prod_{j=i+1}^{n+1} b_j$$

が成立する．ただし, $a_0 = 1, b_{n+1} = 1$.（n に関する帰納法で示せる.）

この公式を用いることで,

$$\begin{aligned} \left|\prod_{i=1}^n \mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - e^{-t^2/2}\right| &= \left|\prod_{i=1}^n \mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - \prod_{i=1}^n \mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}\right| \\ &\leq \sum_{i=1}^n \left|\prod_{j=1}^{i-1} \mathrm{E}[\mathrm{e}^{itZ_j/C_n}]\right| \left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - \mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}\right| \\ &\quad \times \left|\prod_{j=i+1}^n \mathrm{e}^{-t^2\mathrm{E}[Z_j^2]/(2C_n^2)}\right| \\ &\leq \sum_{i=1}^n \left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - \mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}\right| \end{aligned}$$

を得る．よって, リンデンベルグ条件が与えられたとき,

$$\sum_{i=1}^n \left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}] - \mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}\right| \to 0\ as\ n \to \infty$$

がいえればよい．リンデンベルグ条件は, 以下のように書き換えられる．
任意の $\varepsilon > 0,\ \eta > 0$ に対して, $n > n_0$ ならば

$$\frac{1}{C_n^2} \sum_{i=1}^n \mathrm{E}[(X_i - \mu_i)^2 I_{\{|X_i - \mu_i| > \varepsilon C_n\}}] < \eta$$

となるような n_0 が存在する. リンデンベルグ条件が与えられたとき, 任意の正の実数 η_0 に対して,

$$\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]]\right|$$

を満足する n_0 が存在することを示す. 準備として, e^{ix} の原点周りのテイラーの公式を 2 通り考えると,

$$\begin{aligned}
\left|\mathrm{e}^{ix}-(1+ix+\frac{(ix)^2}{2})\right| &\le x^2,\\
\left|\mathrm{e}^{ix}-(1+ix+\frac{(ix)^2}{2})\right| &\le \frac{1}{6}|x|^3
\end{aligned}$$

となるので,

$$\left|\mathrm{e}^{ix}-\left(1+it\frac{Z_i}{C_n}+(it)^2\frac{Z_i^2}{2C_n^2}\right)\right|\le t^2\min\left\{|x|^2,\frac{1}{6}|t||x|^3\right\}$$

をえる. さらに,

$$\begin{aligned}
&\left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]-\left(1+(it)^2\frac{\sigma_i^2}{2C_n^2}\right)\right|\\
&\le \mathrm{E}\left[\left|\mathrm{e}^{itZ_i/C_n}-\left(1+it\frac{Z_i}{C_n}+(it)^2\frac{Z_i^2}{2C_n^2}\right)\right|\right]\le t^2\mathrm{E}\left[\min\left\{\frac{|Z_i|^2}{C_n^2},\frac{|t|}{6}\frac{|Z_i|^3}{C_n^3}\right\}\right]\\
&= t^2\mathrm{E}\left[\min\left\{\frac{|Z_i|^2}{C_n^2},\frac{|t|}{6}\frac{|Z_i|^3}{C_n^3}\right\}I_{\{|Z_i|<C_n\varepsilon\}}\right]+t^2\mathrm{E}\left[\min\left\{\frac{|Z_i|^2}{C_n^2},\frac{|t|}{6}\frac{|Z_i|^3}{C_n^3}\right\}I_{\{|Z_i|\ge C_n\varepsilon\}}\right]\\
&\le t^2\mathrm{E}\left[\frac{|t|}{6}\frac{|Z_i|^3}{C_n^3}I_{\{|Z_i|<C_n\varepsilon\}}\right]+t^2\mathrm{E}\left[\frac{|Z_i|^2}{C_n^2}I_{\{|Z_i|\ge C_n\varepsilon\}}\right]\\
&< \frac{|t|^3\varepsilon}{6C_n^2}\mathrm{E}[|Z_i|^2]+\frac{t^2}{C_n^2}\mathrm{E}\left[|Z_i|^2I_{\{|Z_i|\ge C_n\varepsilon\}}\right]
\end{aligned}$$

をえる. リンデンベルグ条件より, 任意の正の実数 η と ϵ に対して,

$$\begin{aligned}
&\sum_{i=1}^{n}\left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]-\left(1+(it)^2\frac{\sigma_i^2}{2C_n^2}\right)\right|\\
&\le \frac{|t|^3\varepsilon}{6C_n^2}\sum_{i=1}^{n}\mathrm{E}[|Z_i|^2]+\frac{t^2}{C_n^2}\sum_{i=1}^{n}\mathrm{E}\left[|Z_i|^2I_{\{|Z_i|\ge C_n\varepsilon\}}\right]\\
&< \frac{|t|^3\varepsilon}{6}+t^2\eta
\end{aligned}$$

をえる. ゆえに, 任意の正の実数 η と ϵ に対して,

$$\sum_{i=1}^{n}\left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\right|<\frac{|t|^3\varepsilon}{6}+t^2\eta\ (n\ge n_0) \tag{10}$$

を満たす n_0 が存在することが示された.

一方で, テイラー展開の剰余項を考慮すると,

$$\begin{aligned}
&\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\right| \\
&<\frac{t^4}{4}\sum_{i=1}^{n}\left(\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)^2 \le \frac{t^4}{4}\max_{i=1,\ldots,n}\left(\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\left(\sum_{i=1}^{n}\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right) \\
&=\frac{t^4}{16}\max_{i=1,\ldots,n}\left(\frac{\mathrm{E}[Z_i^2]}{C_n^2}\right)
\end{aligned}$$

を得る. リンデンベルグ条件より, $i=1,\ldots,n$ について, 一様に

$$\begin{aligned}
\mathrm{E}\left[\frac{|Z_i|^2}{C_n^2}\right] &= \mathrm{E}\left[\frac{|Z_i|^2}{C_n^2}I_{\{|Z_i|<C_n\varepsilon\}}\right]+\mathrm{E}\left[\frac{|Z_i|^2}{C_n^2}I_{\{|Z_i|\ge C_n\varepsilon\}}\right] \\
&\le \mathrm{E}\left[\frac{|Z_i|^2}{C_n^2}I_{\{|Z_i|<C_n\varepsilon\}}\right]+\sum_{i=1}^{n}\mathrm{E}\left[\frac{|Z_i|^2}{C_n^2}I_{\{|Z_i|\ge C_n\varepsilon\}}\right] \\
&< \varepsilon^2+\eta \ (n\ge n_0)
\end{aligned}$$

となるので,

$$\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\right|<\frac{t^4}{16}(\varepsilon^2+\eta)$$

が主張できる. ゆえに, 任意の正の実数 η と ϵ に対して,

$$\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\right|<\frac{t^4}{16}(\varepsilon^2+\eta)\ (n\ge n_0) \tag{11}$$

を満足する n_0 が存在することが示された. 三角不等式, (10) と (11) より,

$$\begin{aligned}
&\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]]\right| \\
&=\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)+\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)-\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]\right| \\
&\le\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\right|+\sum_{i=1}^{n}\left|\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]-\left(1-t^2\frac{\mathrm{E}[Z_i^2]}{2C_n^2}\right)\right| \\
&<\frac{t^4}{16}(\varepsilon^2+\eta)+\frac{|t|^3\varepsilon}{6}+t^2\eta
\end{aligned}$$

をえる. 以上から, 任意の正の実数 η と ϵ に対して,

$$\sum_{i=1}^{n}\left|\mathrm{e}^{-t^2\mathrm{E}[Z_i^2]/(2C_n^2)}-\mathrm{E}[\mathrm{e}^{itZ_i/C_n}]\right|<\frac{t^4}{16}(\varepsilon^2+\eta)+\frac{|t|^3\varepsilon}{6}+t^2\eta\ (n\ge n_0)$$

を満足する n_0 が存在することが示された. すなわち,

$$\varphi(t)\to e^{-t^2/2}\ as\ n\to\infty$$

が示された. □

次に, リャプノフ型中心極限定理を紹介する. リャプノフ条件は, リンデンベルグ条件よりも強い条件であるが, モーメントに関する条件であるため扱い易いことが多い. 本論文では, この定理を利用して漸近正規性を導いている.

定理 A. 4 (リャプノフ型中心極限定理) $X_1, X_2, \cdots, X_n$ を独立な確率変数列とする．いま，各 i に対して, $\mathrm{E}[X_i] = \mu_i$, $\mathrm{E}[(X_i - \mu_i)^2] = \sigma_i \neq 0$, $\mathrm{E}[|X_i - \mu_i|^k] = \beta_i$(ある $k \in \mathbb{N},\ k \geq 3$) が存在するとする．さらに,

$$B_n = \left(\sum_{i=1}^{n} \beta_i \right)^{1/k}, \ C_n = \left(\sum_{i=1}^{n} \sigma_i^2 \right)^{1/2}$$

に対して, $\lim_{n\to\infty}(B_n/C_n) = 0$ ならば,

$$Y_n = \frac{\sum_{i=1}^{n}(X_i - \mu_i)}{C_n}$$

の分布関数は $\mathcal{N}(0,1)$ の分布関数へ収束する．

(証明) $|X_i - \mu_i| > \varepsilon C_n$ より, $\frac{|X_i - \mu_i|}{\varepsilon C_n} > 1$ なので,

$$\begin{aligned}
&\frac{1}{C_n^2} \sum_{i=1}^{n} \mathrm{E}[|X_i - \mu_i|^2 I_{\{|X_i - \mu_i| > \varepsilon C_n\}}] \\
&\leq \frac{1}{\varepsilon^{\delta} C_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}[|X_i - \mu_i|^{2+\delta} I_{\{|X_i - \mu_i| > \varepsilon C_n\}}] \\
&\leq \frac{1}{\varepsilon^{\delta} C_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}[|X_i - \mu_i|^{2+\delta}] \\
&= \frac{1}{\varepsilon^{\delta} C_n^{k}} \sum_{i=1}^{n} \mathrm{E}[|X_i - \mu_i|^{k}] = \frac{1}{\varepsilon^{\delta}} \left(\frac{B_n}{C_n} \right)^{k}
\end{aligned}$$

となる．ただし, $k = \delta + 2,\ \delta \in \mathbb{N}$ である．もし, リャプノフ条件 $\lim_{n\to\infty}(B_n/C_n) = 0$ が成立すれば, リンデンベルグ条件が成立するので, 定理 A.3 から漸近正規性が示される． □

謝辞

本稿執筆に際し, 清水良一先生 (統計数理研究所元所長，同名誉教授) から多大なるご支援と，貴重なご助言を賜りました．ここに記して, 感謝の意を表します.