Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

5 Chap.5. Consistency and Limiting Distributions

5.3 The Central Limit Theorem / 中心極限定理

これまでの章で見たように，$X_1, X_2, \ldots, X_n$ が $N(\mu, \sigma^2)$ からの無作為標本である時，確率変数

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

は全ての正整数 n に関して平均 0，分散 1 の正規分布に従う．

定理 5.3.1(中心極限定理)

$X_1, X_2, \ldots, X_n$ を，平均 μ，正の分散 σ^2 を持つ分布からの無作為標本とする．この時，確率変数

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

は平均 0，分散 1 の正規分布に従う確率変数に分布収束する．

証明 5.1. 証明のために，$-h < t < h$ にて $M(t) = E[e^{tX}]$ が存在することを仮定する．この積率母関数を，常に存在する特性関数に置き換えても同様の手順で証明が可能である．

仮定から，$-h < t < h$ に，関数

$$m(t) = E[e^{t(X-\mu)}] = e^{-t\mu}E[e^{tX}]$$

が存在する．$m(t)$ は $X - \mu$ の積率母関数であるため，$m(0) = 1, m'(0) = E[X - \mu] = 0, m''(0) = E[(X - \mu)^2] = \sigma^2$ である．テイラーの公式から，

$$\begin{aligned} m(t) &= m(0) + m'(0)(t - 0) + \frac{m''(\xi)t^2}{2} \\ &= 1 + \frac{m''(\xi)t^2}{2} \end{aligned}$$

を満たすような ξ が 0 と t の間に存在する．もし，$\sigma^2 t^2/2$ を挟み込むと，

$$m(t) = 1 + \frac{\sigma^2 t^2}{2} + \frac{[m''(\xi) - \sigma^2]t^2}{2} \tag{1}$$

となる．次に，

$$\begin{aligned}
M(t;n) &= E\left[\exp\left(t\frac{\sum_i X_i - n\mu}{\sigma\sqrt{n}}\right)\right] \\
&= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\right] E\left[\exp\left(t\frac{X_2 - \mu}{\sigma\sqrt{n}}\right)\right] \cdots E\left[\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right] \\
&= E\left[\exp\left(t\frac{X_1 - \mu}{\sigma\sqrt{n}}\right)\right] \cdots E\left[\exp\left(t\frac{X_n - \mu}{\sigma\sqrt{n}}\right)\right] \\
&= \left\{E\left[\exp\left(t\frac{X - \mu}{\sigma\sqrt{n}}\right)\right]\right\}^n \\
&= \left[m\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n,\ -h < \frac{t}{\sigma\sqrt{n}} < h
\end{aligned}$$

なる $M(t;n)$ を考える．*(1)* の t を $t/\sigma\sqrt{n}$ に置き換えると，

$$m\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{[m''(\xi_2) - \sigma^2]t^2}{2n\sigma^2},\ -h\sigma\sqrt{n} < t < h\sigma\sqrt{n}$$

を得る．ただしここで，$\xi_2 \in (0, t/\sigma\sqrt{n})$ である．以上をまとめると，$M(t;n)$ は

$$M(t;n) = \left\{1 + \frac{t^2}{2n} + \frac{[m''(\xi_2) - \sigma^2]t^2}{2n\sigma^2}\right\}^n$$

となる．$m''(t)$ は $t = 0$ で連続であり，$n \to \infty$ の時，$\xi \to 0$ となることから，

$$\lim_{n\to\infty} M(t;n) = e^{t^2/2}$$

を得る．これは，標準正規分布の積率母関数である．よって，確率変数 $Y_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ は，$n \to \infty$ の時，標準正規分布に従う．

¡note¿

例 5.1. *(平均の大標本における推測)* $X_1, X_2, \ldots, X_n$ を，平均 μ，正の分散 σ^2 を持つ分布からの無作為標本とする．この時，μ, σ 共に未知母数であるとする．$\bar{X}, S$ をそれぞれ，標本平均，標準偏差とする．第 *4* 章では，平均に関する信頼区間と検定を紹介したが，それらは，

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \xrightarrow{D} N(0,1) \tag{2}$$

であることに基づく．これを確認するために，左辺を以下のように改める．

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \left(\frac{\sigma}{S}\right) \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

例 *5.1* に示すように，S は σ に確率収束するため，*5.2* 章の定理群を用いると，σ/S は 1 に確率収束する．(2) 式は中心極限定理，スラツキーの定理から導かれる．

例 5.2. 母比率の大標本における推測 $X_1, X_2, \ldots, X_n$ を，成功確率 p のベルヌーイ分布からの無作為標本とする．$\hat{p}$ を標本の成功したものの比率とすると，$\hat{p} = \bar{X}$ である．第 *4* 章では，大標本での比率の信頼区間と検定を紹介したが，これらは，

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{D} N(0,1) \tag{3}$$

に基づいている．これは，中心極限定理と，例 *5.1* と同様の方法で導くことができる．

例 5.3. *(カイ二乗検定のための大標本における推測)* $X_1, X_2, \ldots, X_n$ を，$Bin(1,p)$ からの無作為標本とし，$Y_n = X_1 + X_2 + \cdots + X_n$ とすると，$Y_n \sim Bin(n,p)$ である．また，中心極限定理から，

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0,1)$$

である．この時，定理 5.9 から，

$$\left(\frac{Y_n - np}{\sqrt{np(1-p)}}\right)^2 \xrightarrow{D} \chi^2(1)$$

を得る．

¡note¿

近似的に正規分布に従う統計量の関数について，興味を持つケースは頻繁にある．例えば，先ほどの例の Y_n を見てみると，Y_n は近似的に $N(np, np(1-p))$ に従う．Y_n の分散であるため，$np(1-p)$ は p の重要な関数である．ゆえに，もし，p が未知の場合，Y_n の分散を推定しようと考えるであろう．$E[Y_n/n] = p$(不偏推定量) であるため，$n(Y_n/n)(1 - Y_n)$ が分散の推定量として使ったり，その分布について知ろうとするだろう．このケースにおいて，この推定量は近似的に正

規分布に従うのだろうか．もしそうなら，この推定量の分布の平均，分散は何であろうか．これらの疑問を解決するために，Δ method が利用される．

以前に紹介した Δ method の使用例と同様に，標本平均の関数を考える．私たちは，$\bar{X}_n$ が母平均 μ に確率収束し，$\bar{X}_n$ が近似的に $N(\mu, \sigma/n)$ に従うことを知っているとする．我々の興味は $\bar{X}_n$ の関数，例えば $u(\bar{X}_n)$ にあるとしよう．ここで，$u(\cdot)$ は μ で微分可能で，$u'(\mu) \neq 0$ の関数である．定理 5.15(Δ method) から，$u(\bar{X}_n)$ は近似的に，$N(u(\mu)), [u'(\mu)^2\sigma^2/n]$ に従う．これは，少し変形すると，

$$\frac{u(\bar{X}_n) - u(\mu)}{\sqrt{[u'(\mu)^2]\sigma^2/n}} \xrightarrow{D} N(0,1)$$

となる．

¡note¿

5.4 Extensions to Multivariate Distributions / 多変量分布への拡張

この節では，確率変数ベクトルに対する漸近的性質を手短に紹介する．1 変数の場合に示された概念は，そのまま (直感にそぐわないような方法で) 多変数の場合に拡張される．ここでの紹介は簡潔であるため，興味のある読者は，Serfling(1980) を参照されたい．

以下の表記を利用する．$\boldsymbol{v} \in \mathbb{R}^p$ に対して，ユークリッドノルムは

$$\|\boldsymbol{v}\| = \sqrt{\sum_{i=1}^{p} v_i^2} \tag{4}$$

で定義される．このノルムは，以下の有用な 3 つの性質を満たす．

ユークリッドノルムの性質

1. $\forall \boldsymbol{v} \in \mathbb{R}^p$, $\|\boldsymbol{v}\| \geq \boldsymbol{0}$ であり，$\boldsymbol{v} = \boldsymbol{0}$ の時に限り，$\|\boldsymbol{v}\| = \boldsymbol{0}$ である．
2. $\forall \boldsymbol{v} \in \mathbb{R}^p$ と $a \in \mathbb{R}$ に対して，$\|a\boldsymbol{v}\| = |a|\|\boldsymbol{v}\|$ である
3. $\forall \boldsymbol{v},\ \boldsymbol{u} \in \mathbb{R}^p$ に対して，$\|\boldsymbol{v} + \boldsymbol{u}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{u}\|$ である．

$\mathbb{R}^p$ の標準基底を $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_p}$ で表すとする．すると，$\boldsymbol{v}' = (v_1, \ldots, v_p)$ は，

$$\boldsymbol{v} = \sum_{i=1}^{p} v_i \boldsymbol{e}_i$$

と書くことができる．確率変数ベクトルへの拡張は以下の補題が有用である．

補題 5.4.1

$\boldsymbol{v}' = (v_1, \ldots, v_p)$ を $\mathbb{R}^p$ のベクトルとする．この時，

$$|v_j| \le \|\boldsymbol{v}\| \le \sum_{i=1}^{p} |v_i|,\ for\ all\ j = 1, \ldots, p \tag{5}$$

が成り立つ．

証明 5.2. 全ての j に対して，

$$v_j^2 \le \sum_{i=1}^{p} v_i^2 = \|\boldsymbol{v}\|^2$$

であり，これルートをとると，$v_j \le \|\boldsymbol{v}\|$ を得る．さらに，

$$\|\boldsymbol{v}\| = \left\| \sum_{i=1}^{p} v_i \boldsymbol{e}_i \right\| \le \sum_{i=1}^{p} |v_i| \|\boldsymbol{e}_i\| = \sum_{i=1}^{p} |v_i|$$

であることから，補題の結果を得る． □

$\{\boldsymbol{X}_n\}$ で p 次元確率変数ベクトルの列を表すとする．絶対値は，$\mathbb{R}^1$ のユークリッドノルムであるので，確率変数ベクトルの確率収束の定義は以下のようになる．

定義 5.1. $\{\boldsymbol{X}_n\}$ で p 次元確率変数ベクトルの列とし，$\boldsymbol{X}$ を全ての要素が同一の標本空間で定義された確率変数ベクトルとする．$\{\boldsymbol{X}_n\}$ が $\boldsymbol{X}$ に確率収束するとは，任意の $\epsilon > 0$ に対して

$$\lim_{n\to\infty} P[\|\boldsymbol{X}_n - \boldsymbol{X}\| \ge \epsilon] = 0 \tag{6}$$

が成立することをいう．*1* 変数の場合と同じように，$\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$ と書く．

定理 5.4.1

$\{\boldsymbol{X}_n\}$ で p 次元確率変数ベクトルの列とし，$\boldsymbol{X}$ を全ての要素が同一の標本空間で定義された確率変数ベクトルとする．この時，

$$\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}\ if\ and\ only\ if\ X_{nj} \xrightarrow{P} X_j\ for\ all\ j = 1, \ldots, p$$

である．

証明 5.3. $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$ とする．全ての j について，(5) 式の最初の不等号から，ある $\epsilon > 0$ について，

$$\epsilon \leq |X_{nj} - X_j| \leq \|\boldsymbol{X}_n - \boldsymbol{X}\|$$

を得る．よって，

$$\overline{\lim_{n\to\infty}} P[|X_{nj} - X_j| \geq \epsilon] \leq \overline{\lim_{n\to\infty}} P[\|\boldsymbol{X}_n - \boldsymbol{X}\| \geq \epsilon] = 0$$

となる．これは定理の $\Rightarrow$ の結果である．

逆に，もし，全ての $j = 1, \ldots, p$ について，$X_{nj} \xrightarrow{P} X_j$ であるならば，(5) 式の二つ目の不等号から，任意の $\epsilon > 0$ について，

$$\epsilon \leq \|\boldsymbol{X}_n - \boldsymbol{X}\| \leq \sum_{j=1}^{p} |X_{nj} - X_j|$$

を得る．よって，

$$\begin{aligned}\overline{\lim_{n\to\infty}} P[\|\boldsymbol{X}_n - \boldsymbol{X}\| \geq \epsilon] &\leq \overline{\lim_{n\to\infty}} P[\sum_{j=1}^{p} |X_{nj} - X_j| \geq \epsilon] \\ &\leq \sum_{j=1}^{p} \overline{\lim_{n\to\infty}} P[|X_{nj} - X_j| \geq \epsilon/p] = 0\end{aligned}$$

となる．これは定理の $\Leftarrow$ の結果である． □

この結果から，多くの確率収束の定理は多変数の場合に拡張できる．
¡note¿

簡単な適用例として，標本平均と標本分散の多変数への拡張を考える．$\{\boldsymbol{X}_n\}$ を共通の平均ベクトル，分散共分散行列を持つ分布からの iid 標本とする．標本平均のベクトルを

$$\bar{\boldsymbol{X}}_n = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \tag{7}$$

で表すとする．大数の弱法則から，全ての j について $\bar{X}_j \xrightarrow{P} \mu_j$ である．よって定理 5.4.1 から

$$\bar{\boldsymbol{X}}_n \xrightarrow{P} \boldsymbol{\mu}$$

である．

次に標本分散について考える．まず，以下のように標本分散と，標本共分散を定義する．

$$S_{n,j}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2,\ for\ j = 1, \dots, p, \tag{8}$$

$$S_{n,jk}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ik} - \bar{X}_k)(),\ for\ j \neq k = 1, \dots, p, \tag{9}$$

4 次のモーメントが有限であると仮定するなら，大数の弱法則から，全ての標本分散と標本共分散は母分散，母共分散にそれぞれ確率収束することが示される (大数の強法則なら，より弱い仮定で収束が示される．具体的には 2 次のモーメントの存在の仮定だけで良い.)．もし，$\boldsymbol{S}$ を標本分散，標本共分散を要素として持つ $p \times p$ 行列とすると，$\boldsymbol{S} \xrightarrow{P} \boldsymbol{\Sigma}$ とかける．次に分布収束の定義を紹介する

定義 5.2. $\{\boldsymbol{X}_n\}$ をそれぞれの要素 $\boldsymbol{X_n}$が 分布関数 $F_n(\boldsymbol{x})$ を持つ p 次元確率変数ベクトルの列とし，$\boldsymbol{X}$ を分布関数 $F(\boldsymbol{x})$ を持つ確率変数ベクトルとする．この条件下で，$F(\boldsymbol{x})$ が連続である全ての点において，

$$\lim_{n\to\infty} F_n(\boldsymbol{x}) = F(\boldsymbol{x}) \tag{10}$$

が成立するとき，$\{\boldsymbol{X}_n\}$ は $\boldsymbol{X}$ に分布収束するといい，$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ で表す．

多変数の場合に，5.2 章の定理の拡張が考えられる．以下では，重要なものを 2 つ，証明なしで紹介する．(気が向いたら証明つけて更新します)

定理 5.4.2

$\{\boldsymbol{X}_n\}$ を $\boldsymbol{X}$ を分布収束する確率変数ベクトルの列とし，$g(\boldsymbol{x})$ を $\boldsymbol{X}$ の台で連続な関数とする．この時，

$$g(\boldsymbol{X}_n) \xrightarrow{D} g(\boldsymbol{X})$$

である．

分布収束することを定義を用いて示すことはしばしば困難である．以下の定理で示すように，1 変数の場合と同じく，分布収束は積率母関数の収束と等しい．

定理 5.4.3

$\{\boldsymbol{X}_n\}$ をそれぞれの要素 $\boldsymbol{X_n}$が 分布関数 $F_n(\boldsymbol{x})$ と積率母関数 $M_n(\boldsymbol{t})$ を持つ p 次元確率変数ベクトルの列とする．$\boldsymbol{X}$ を分布関数 $F(\boldsymbol{x})$ と積率母関数 $M(\boldsymbol{t})$ を持つ確率変数ベクトルとする．この時，$\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ は，任意の $\boldsymbol{t}$ *such that* $\|t\| < h$，ある $h > 0$ について，

$$\lim_{n\to\infty} M_n(\boldsymbol{t}) = M(\boldsymbol{t}) \tag{11}$$

が成り立つことと等しい．

これの証明は Tucker(1967) などに記されている．また，積率母関数ではなく，特性関数を用いるのが一般的である．

定理 5.4.4(多変数の中心極限定理)

$\{\boldsymbol{X}_n\}$ を共通の平均ベクトル $\boldsymbol{\mu}$，分散共分散行列 $\boldsymbol{\Sigma}$(半正定値) を持つ確率変数ベクトルの列とする．$\boldsymbol{0}$ の開近傍に共通の積率母関数 $M(\boldsymbol{t})$ が存在すると仮定する．また，

$$\boldsymbol{Y_n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}) = \sqrt{n}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$$

とする．この時，$\boldsymbol{Y}_n$ は，$N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ に分布収束する．

証明 5.4. $\boldsymbol{t} \in \mathbb{R}^p$ を $\boldsymbol{0}$ の近傍に存在するベクトルとする．$\boldsymbol{Y}_n$ の積率母関数は，

$$\begin{aligned}
M_n(\boldsymbol{t}) &= E\left[\exp\left\{\boldsymbol{t}'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu})\right\}\right] \\
&= E\left[\exp\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{t}'(\boldsymbol{X}_i - \boldsymbol{\mu})\right\}\right] \\
&= E\left[\exp\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i\right\}\right] \ (W_i := \boldsymbol{t}'(\boldsymbol{X}_i - \boldsymbol{\mu}))\ (*)
\end{aligned}$$

とする．ここで，$W_1, \ldots, W_n$ は平均 0，分散 $Var(W_i) = \boldsymbol{t'\Sigma t}$ の独立同分布の確率変数である．よって，*1* 変数の場合の中心極限定理から，

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i \xrightarrow{D} N(0, \boldsymbol{t'\Sigma t}) \tag{12}$$

である．(12) は$\underline{確率変数\ 1/\sqrt{n}\sum_{i=1}^{n}W_i}$の積率母関数の $t = 1$ の場合である．よって，

$$M_n(\boldsymbol{t}) = E\left[\exp\left\{1 \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i\right\}\right] \to e^{1^2\boldsymbol{t'\Sigma t}/2} = e^{\boldsymbol{t'\Sigma t}/2}$$

を得る．$e^{\boldsymbol{t'\Sigma t}/2}$ は $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ の積率母関数であるため，定理の結果を得る． □

$\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ を平均 $\boldsymbol{\mu}$，分散共分散行列 $\boldsymbol{\Sigma}$ の分布からの無作為標本とする．$\bar{\boldsymbol{X}}_n$ を標本平均のベクトルとすると，中心極限定理から，

$$\bar{\boldsymbol{X}}_n \textit{ has an approximate } N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}) \tag{13}$$

が成立する．

定理 5.4.5

$\{\boldsymbol{X}_n\}$ を p 次元確率変数ベクトルの列とする．$\boldsymbol{X}_n \overset{D}{\to} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ とする．$\boldsymbol{A}$ を定数を要素ともつ $m \times p$ 行列とし，$\boldsymbol{b}$ を m 次元定数ベクトルとする．このとき，

$$\boldsymbol{A}\boldsymbol{X}_n + \boldsymbol{b} \overset{D}{\to} N(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$$

が成り立つ．

これの証明は，積率母関数を利用することで得られる．(余裕あるときに追加して更新します．すみません!)．以下の定理は，Δ method の拡張としてかなり便利である．証明は Serfling(1980) の第 3 章を参照 (ダジャレじゃないです)．

定理 5.4.6

$\{\boldsymbol{X}_n\}$ を p 次元確率変数ベクトルの列とする．以下を仮定する．

$$\sqrt{n}(\boldsymbol{X}_n - \mu_0) \overset{D}{\to} N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$$

さらに，$\boldsymbol{g}$ を $1 \leq k \leq p$ なる k について，$\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_k(\boldsymbol{x}))'$ とし，$k \times p$ 行列 $\boldsymbol{B}$，

$$\boldsymbol{B} = \left[\frac{\partial g_i}{\partial \mu_j}\right],\ i = 1, \ldots, k;\ j = 1, \ldots, p,$$

が連続で，$\boldsymbol{\mu}_0$ の近傍でも存在するとする．また，$\boldsymbol{\mu}_0$ での $\boldsymbol{B}$ を $\boldsymbol{B}_0$ とする．このとき，

$$\sqrt{n}(\boldsymbol{g}(\boldsymbol{X}_n) - \boldsymbol{g}(\boldsymbol{\mu}_0)) \overset{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{B_0}\boldsymbol{\Sigma}\boldsymbol{B_0'}) \tag{14}$$

が成り立つ．