確率数理工学9

⑪ 大数の法則と中心極限定理

- 大数の法則

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[\text{a.s.\ (強)}]{p\ (\text{弱})} \mu$$

（標本平均）　　　　　　（真の平均）

- 中心極限定理

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \rightsquigarrow \text{正規分布}$$

（注）それぞれどういう意味で「収束」しているかを明確に理解すること。
何となく「正規分布に収束する」という理解で止めない。

Thm（大数の弱法則）(☆)

$X_i\ (i=1,2,\ldots)$：互いに独立

$E[X_i]=\mu_i,\quad Var[X_i]=\sigma_i^2$ で、

$$\frac{\sum_{i=1}^{n}\sigma_i^2}{n^2} \longrightarrow 0 \quad かつ \quad \frac{\sum_{i=1}^{n}\mu_i}{n} \longrightarrow \mu \quad なら$$

← n^2 で割っている

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

である。

特に、X_i が i.i.d. で $E[X_i]=\mu$（有限）, $Var[X_i]<\infty$ なら

$$\overline{X}_n \xrightarrow{p} \mu$$

である。

証明

$\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$ とおくと、仮定より $\overline{\mu}_n \to \mu$ である。

$$P(|\overline{X}_n - \mu| \geqq \varepsilon) \leqq P\left(|\overline{X}_n - \overline{\mu}_n| \geqq \frac{\varepsilon}{2} \cup |\overline{\mu}_n - \mu| \geqq \frac{\varepsilon}{2}\right)$$

$$\leqq \underbrace{P\left(|\overline{X}_n - \overline{\mu}_n| \geqq \frac{\varepsilon}{2}\right)}_{①} + \underbrace{P\left(|\overline{\mu}_n - \mu| \geqq \frac{\varepsilon}{2}\right)}_{②}$$

- $\overline{\mu}_n \longrightarrow \mu$ より ② $\longrightarrow 0$
- 一方、

Markov

$$P\left(|\overline{X}_n - \overline{\mu}_n| \geqq \frac{\varepsilon}{2}\right) \leqq \frac{E[(\overline{X}_n - \overline{\mu}_n)^2]}{\frac{\varepsilon^2}{4}} \leqq \frac{E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right]}{\frac{\varepsilon^2}{4}}$$

$$= \frac{\frac{1}{n^2}\sum_{i=1}^{n} E[(X_i - \mu_i)^2]}{\frac{\varepsilon^2}{4}}$$

$$= \frac{\frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2}{\frac{\varepsilon^2}{4}} \longrightarrow 0 \quad (\because \text{仮定})$$

つまり、$\overline{X}_n \xrightarrow{P} \mu$ である。

//

確率収束は概収束に変えられる：大数の強法則

事象の列 $A_n \in \mathcal{F}$ $(n=1,2,\dots)$ があったとして，その上極限を

$$\limsup_{n\to\infty} A_n := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

とする。 $\omega \in \limsup A_n$ なら，任意の n に対し，ある $k \geq n$ が存在して，

$\omega \in A_k$ である．つまり，ω は無限個の A_n に含まれる．（逆も然り）

このことから

$$\limsup A_n = A_n \ \text{i.o.}$$

とも書く．(i.o. = infinitely often)

Thm (Borel - Cantelli の補題)

1. $\displaystyle\sum_{n=1}^{\infty} P(A_n) < \infty$ ならば

$$P(\limsup_{n\to\infty} A_n) = 0.$$

2. $(A_n)_{n=1}^{\infty}$ は独立で $\displaystyle\sum_{n=1}^{\infty} P(A_n) = \infty$ ならば

$$P(\limsup_{n\to\infty} A_n) = 1.$$

//

(証明) 1. 任意の N に対し，

$$P(\limsup_{n\to\infty} A_n) \leq P\left(\bigcup_{n=N}^{\infty} A_n\right)$$

$$\leq \sum_{n=N}^{\infty} P(A_n) \quad (\text{劣加法性})$$

である．今 $\displaystyle\sum_{n=1}^{\infty} P(A_n) < \infty$ より，$\displaystyle\lim_{N\to\infty} \sum_{n=N}^{\infty} P(A_n) = 0$ である．

よって，$N \to \infty$ とすることで，

$$P(\limsup_{n\to\infty} A_n) \leq \lim_{N\to\infty} \sum_{n=N}^{\infty} P(A_n) = 0$$

を得る．

2.

↳ 次ページへ

まず、

$$P\left(\left(\limsup_{n\to\infty} A_n\right)^c\right) = P\left(\bigcup_{k=1}^{\infty}\left(\bigcup_{n=k}^{\infty} A_n\right)^c\right)$$

$$\leq \sum_{k=1}^{\infty} P\left(\left(\bigcup_{n=k}^{\infty} A_n\right)^c\right) = \sum_{k=1}^{\infty} P\left(\bigcap_{n=k}^{\infty} A_n^c\right)$$

であるので、 $P\left(\bigcap_{n=k}^{\infty} A_n^c\right) = 0 \quad ({}^{\forall}k = 1, 2, \ldots)$ を示せば良い。

今任意の N に対し

$$P\left(\bigcap_{n=k}^{\infty} A_n^c\right) \leq P\left(\bigcap_{n=k}^{N} A_n^c\right) \overset{(A_n)_n\text{は独立}}{=} \prod_{n=k}^{N} P(A_n^c)$$

$$= \prod_{n=k}^{N} (1 - P(A_n)) \leq \prod_{n=k}^{N} e^{-P(A_n)}$$

$$= e^{-\sum_{n=k}^{N} P(A_n)}$$

である。 $\sum_{n=1}^{\infty} P(A_n) = \infty$ より ${}^{\forall}k$ で $\sum_{n=k}^{\infty} P(A_n) = \infty$ でもある。よって、

$N \to \infty$ とすることで、右辺 $\to 0$ である。　よって示された。 //

㊍ Thm（大数の強法則）

X_i $(i=1,2,\dots)$：独立

$E[X_i]=\mu$：有限

$Var[X_i]=\sigma^2<\infty$, $\quad \nu_4=E[|X_i-\mu|^4]<\infty$

ならば．

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

㊍概収束！

//

（証明）$^\forall\varepsilon>0$ に対し．

$$P\left(\limsup_{n\to\infty}|\bar{X}_n-\mu|\geqq\varepsilon\right)=0$$

を示す．これが示せれば，確率の連続性より

$$P\left(^\exists\lim_{n\to\infty}|\bar{X}_n-\mu|=0\right)=1-P\left(^\exists\varepsilon>0,\ \limsup_{n\to\infty}|\bar{X}_n-\mu|\geqq\varepsilon\right)$$
$$=1-P\left(\bigcup_{m=1}^{\infty}\left\{\limsup_{n\to\infty}|\bar{X}_n-\mu|\geqq\frac{1}{m}\right\}\right)$$
$$=1-\lim_{m\to\infty}P\left(\limsup_{n\to\infty}|\bar{X}_n-\mu|\geqq\frac{1}{m}\right)$$
$$=1-0=1$$

が示せる．

$A_n:=\{|\bar{X}_n-\mu|\geqq\varepsilon\}$ として，$(A_n)_{n=1}^{\infty}$ に Borel-Cantelli を適用する．

$$P(A_n)=P(|\bar{X}_n-\mu|\geqq\varepsilon)\leqq\frac{E[|\bar{X}_n-\mu|^4]}{\varepsilon^4}$$

Markov の不等式

であるが．

（1次，3次の cross-term は 0）

$$E[|\bar{X}_n-\mu|^4]=\frac{1}{n^4}\sum_{i=1}^{n}E[(X_i-\mu)^4]+\underbrace{\frac{\binom{4}{2}}{n^4}\sum_{i<j}E[(X_i-\mu)^2(X_j-\mu)^2]}_{\boxed{=\frac{6}{n^4}\frac{n(n-1)}{2}\sigma^4=\frac{3(n-1)}{n^3}\sigma^4}}$$

$$=\frac{\nu_4}{n^3}+\frac{3(n-1)}{n^3}\sigma^4\leqq\frac{K}{n^2}\quad(K:=\nu_4+3\sigma^4<\infty)$$

よって．$\displaystyle\sum_{n=1}^{\infty}P(A_n)\leqq\frac{K}{\varepsilon^4}\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty$ なので．Borel-Cantelli より

$P(\limsup_{n\to\infty}A_n)=0$．つまり $P(\limsup_{n\to\infty}|\bar{X}_n-\mu|\geqq\varepsilon)=0$ //

チェックせよ．

㊟ 実は X_n が i.i.d. なら分散やモーメントの条件を外せる。
(独立同一)

Thm
X_i : i.i.d. , $E[X_i]=\mu$ (有限) とする。 ← 分散の仮定は不要

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

//

(証明はWeb上の補足資料を参照)

• 分布の極限

Thm (Levyの連続性定理) ㊟

X_n : r.v.

ϕ_n : X_n の特性関数 ($\phi_n(t)=E[e^{itX_n}]$)

(i) X : r.v.

ϕ : X の特性関数

$$X_n \rightsquigarrow X \iff \phi_n(t) \to \phi(t) \quad ({}^\forall t \in \mathbb{R})$$

(各点収束)

(ii) ある $\phi(t)$ が存在して、$\phi_n(t) \to \phi(t)$ (${}^\forall t \in \mathbb{R}$) かつ $\phi(t)$ が $t=0$ で連続なら、ϕ を特性関数として持つ r.v. X が存在して、

$$X_n \rightsquigarrow X$$

が成り立つ。 //

※ 多変量でも同様に、$\phi(t) := E[e^{it^T X}]$ ($t \in \mathbb{R}^d$) として

「$X_n \rightsquigarrow X \iff \phi_n(t) \to \phi(t) \quad ({}^\forall t \in \mathbb{R}^d)$」

が成り立つ。

Cor (Cramer-Wold device)

$$X_n \rightsquigarrow X \iff t^T X_n \rightsquigarrow t^T X \quad ({}^\forall t \in \mathbb{R}^d)$$

↑ d次元 r.v.

Thm (中心極限定理)　Central Limit Theorem (CLT)

X_n : i.i.d.

$E[X_n]=\mu$: 有限

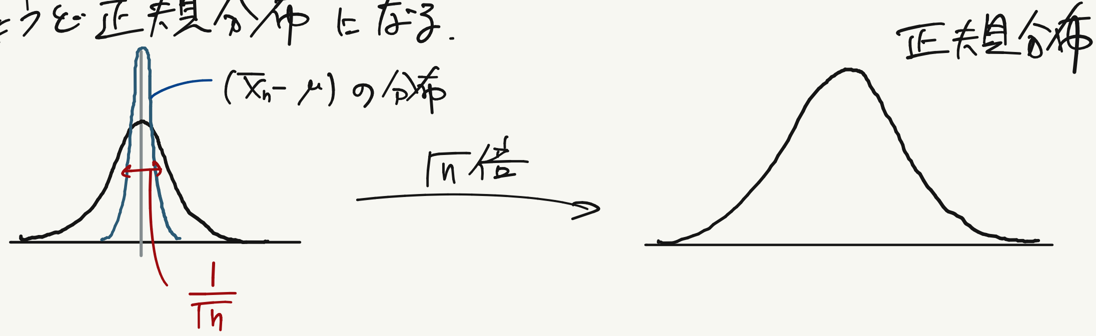
$Var[X_n]=\sigma^2 \underline{\underline{< \infty}}$, $\sigma^2 \neq 0$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ とすると、

$$\sqrt{n}(\bar{X}_n-\mu) \rightsquigarrow N(0,\sigma^2)$$

Note

$\bar{X}_n-\mu \longrightarrow 0$ (a.s) であるが、$\sqrt{n}$倍膨らませることで、ちょうど正規分布になる。



(証明)　$\sqrt{n}(\bar{X}_n-\mu)$の特性関数

$$\Phi_n(t) = E\left[e^{it\sqrt{n}(\bar{X}_n-\mu)}\right]$$

$$= E\left[e^{\sum_{j=1}^{n}\frac{it}{\sqrt{n}}(X_j-\mu)}\right]$$

$$= \prod_{j=1}^{n} \underbrace{E\left[e^{\frac{it}{\sqrt{n}}(X_j-\mu)}\right]}_{\phi\left(\frac{t}{\sqrt{n}}\right)}$$

$\phi\left(\frac{t}{\sqrt{n}}\right)$ とおく。 ← $X_j-\mu$の特性関数

$$\phi\left(\frac{t}{\sqrt{n}}\right) = \phi(0) + \frac{t}{\sqrt{n}}\phi'(0) + \frac{1}{2}\phi''(0)\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)$$ ← 分散が有限ならば ϕは連続2回微分可能

$$= 1 + \underbrace{0}_{\because\ E[X_j-\mu]=0} - \frac{1}{2}\sigma^2\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)$$

より、

$$\Phi_n(t) = \left(1-\frac{1}{2}\sigma^2\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{-\frac{1}{2}\sigma^2 t^2}$$

↓ $N(0,\sigma^2)$の特性関数

あとはLevyの連続性定理より、CLTが従う。

※ CLT は $\dfrac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \rightsquigarrow N(0,1)$ も言える.

Ex. (Bernoulli 試行の CLT)

$P(X_j=1)=\theta,\quad P(X_j=0)=1-\theta \qquad (0<\theta<1)$

$E[X_j]=\theta,\quad Var[X_j]=\theta(1-\theta)$ なので.

$$\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sqrt{\theta(1-\theta)}} \rightsquigarrow N(0,1)$$

//

ある選挙に2人の候補者 A, B がいる.

$X_i=1$ (有権者 i が A に投票), $X_i=0$ (i が B に投票) とする.

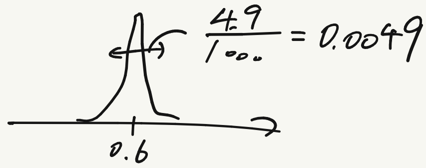
A を支持する人の割合 60%, B を支持する人の割合 40% とする.

X_i は $\theta=0.6$ の Bernoulli 分布に従う.

今 $n=1$万人 の開票後, $\bar{X}_n$ はおよそ

$$N\left(\theta, \frac{\theta(1-\theta)}{n}\right) = N\left(0.6, \frac{0.4\times 0.6}{10^4}\right) = N(0.6,\ 24\times 10^{-6})$$

に従う



Thm (Poisson の小数の法則)

$Y_n \sim B(n, \theta_n)$ (n 回のコイン投げ)

$n\theta_n \longrightarrow \lambda$ ($n\to\infty$) とする. このとき,

$Y_n \rightsquigarrow Po(\lambda)$

Proof Y_n の特性関数 $= \phi_n(t) = (\theta_n e^{it} + (1-\theta_n))^n$

$$= (1-\theta_n(1-e^{it}))^n$$

$$= \left\{1-\frac{\lambda}{n}[(1-e^{it})+o(1)]\right\}^n$$

$$\longrightarrow \exp(-\lambda(1-e^{it}))$$: $Po(\lambda)$ の特性関数

//

例 不良品が出る確率が低い製品を大量に生産すると. その中に入っている不良品の個数は大体ポアソン分布.

Lem（Slutskyの補題）

$X_n \rightsquigarrow X$ かつ $Y_n \rightsquigarrow c$（定数）のとき（X_n, Y_nは独立とは限らない）

(1) $X_n + Y_n \rightsquigarrow X + c$

(2) $X_n Y_n \rightsquigarrow cX$ //

（証明は略）

Thm（分散が未知の場合のCLT）

$$\hat{\sigma}_n^2 \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \quad \text{（標本分散）}$$

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\hat{\sigma}_n} \rightsquigarrow N(0,1)$$

（信頼区間の計算に使える）

Proof 大数の法則より $\hat{\sigma}_n^2 \longrightarrow \mathrm{Var}[X_1]$ (a.s.) なので

Slutskyの補題とCLTより従う //

―Delta法

Thm（Delta法）

$(X_n)_n$: i.i.d.

$\sqrt{n}(\overline{X}_n - \mu) \rightsquigarrow N(0, \sigma^2)$ を仮定

f: 1回微分可能かつ，$f'(\mu) \neq 0$ とおくと．

$$\frac{\sqrt{n}\left(f(\overline{X}_n) - f(\mu)\right)}{|f'(\mu)|\sigma} \rightsquigarrow N(0,1)$$

$X_n = o_p(a_n) \iff \frac{X_n}{a_n} \xrightarrow{P} 0$

スモールオーダー

（略証）
$$f(\overline{X}_n) - f(\mu) = f(\mu) + (\overline{X}_n - \mu)f'(\mu) + o_p\left(\frac{1}{\sqrt{n}}\right) - f(\mu)$$
$$= (\overline{X}_n - \mu)f'(\mu) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

よって．
$$\frac{\sqrt{n}\left(f(\overline{X}_n) - f(\mu)\right)}{|f'(\mu)|\sigma} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \cdot \underbrace{\frac{f'(\mu)}{|f'(\mu)|}}_{=\pm 1} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

$\rightsquigarrow N(0,1)$ (by Slutsky) //

<u>Thm</u> (大変量のCLT)

X_i : $\mathbb{R}^d$-値 r.v. (i.i.d.)

$E[X_i] = \mu \in \mathbb{R}^d$

$E[(X_i-\mu)(X_i-\mu)^T] = \Sigma \in \mathbb{R}^{d\times d}$

($\Sigma \succ 0$ とする)

$\Longrightarrow \quad \sqrt{n}(\overline{X}_n - \mu) \rightsquigarrow N(0, \Sigma)$

(多変量正規分布)

//

さらに、$f: \mathbb{R}^d \to \mathbb{R}$ が微分可能で

$$\nabla f(\mu) = \left[\frac{\partial f}{\partial x_1}(\mu), \cdots, \frac{\partial f}{\partial x_d}(\mu)\right]^T (\in \mathbb{R}^d) \neq 0$$

ならば、

$$\sqrt{n}(f(\overline{X}_n) - f(\mu)) \rightsquigarrow N(0, \nabla f(\mu)^T \Sigma \nabla f(\mu))$$

//

<u>Ex.</u>

$$\begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix}, \cdots, \underset{\displaystyle X_n}{\overset{}{}}\begin{pmatrix} X_{n1} \\ X_{n2} \end{pmatrix} : \text{i.i.d.}$$

$$\mu = E[X_i] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \in \mathbb{R}^2, \quad \Sigma = Cov(X_i) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

CLTより、$\sqrt{n}(\overline{X}_n - \mu) \rightsquigarrow N(0, \Sigma)$ である。

$f(x, z) = xz$ とすると、$\nabla f(x, z) = \begin{pmatrix} z \\ x \end{pmatrix}$ より、

$$\nabla f(\mu)^T \Sigma \nabla f(\mu) = \mu_2^2 \sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2 \sigma_{22}$$

であるので、

$$\sqrt{n}(\overline{X}_{1,n}\overline{X}_{2,n} - \mu_1\mu_2) \rightsquigarrow N(0, \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22})$$

//