

確率数理工学8

④ 確率不等式

Thm **重要**

(1) Markovの不等式

$$X: \text{r.v.}$$

$$X \geq 0 \quad (\text{a.s.}) \text{ とおす.}$$

$$\forall \varepsilon > 0 \text{ と}$$

$$P(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon}$$

(2) Chebyshevの不等式

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

Proof

$$(1) \quad E[X] = E[X \mathbb{1}\{X \geq \varepsilon\} + X \mathbb{1}\{X < \varepsilon\}] \\ \geq E[\varepsilon \mathbb{1}\{X \geq \varepsilon\}] + 0 \\ = \varepsilon \cdot E[\mathbb{1}\{X \geq \varepsilon\}] = \varepsilon P(X \geq \varepsilon) \quad //$$

$$(2) \quad P(|X - E[X]| \geq \varepsilon) = P(|X - E[X]|^2 \geq \varepsilon^2) \\ \leq \frac{E[|X - E[X]|^2]}{\varepsilon^2} = \frac{\text{Var}[X]}{\varepsilon^2} \\ \uparrow \\ \text{Markov} \quad //$$

例 $X = \frac{1}{n} \sum_{i=1}^n X_i$ ($X_i: \text{i.i.d.}, \text{Var}[X_i] = \sigma^2 < \infty$)

$$\Rightarrow \text{Var}[X] = \frac{1}{n} \sigma^2 \Rightarrow P(|X - E[X]| \geq \varepsilon) \leq \frac{\sigma^2}{n \varepsilon^2}$$

$$\Rightarrow \varepsilon = \frac{\sigma}{\sqrt{n}} \varepsilon' \text{ とおきなおせば: } P(|X - E[X]| \geq \frac{\sigma}{\sqrt{n}} \varepsilon') \leq \frac{1}{\varepsilon'^2}$$

$$\Rightarrow |X - E[X]| = O_p\left(\frac{\sigma}{\sqrt{n}}\right) : \text{重要}$$

Cor (Markovの不等式の一般化)

φ : 単調増加かつ $\varphi(x) > 0$ ($\forall x \in \mathbb{R}$), 可測

$$P(X \geq a) \leq \frac{E[\varphi(X)]}{\varphi(a)} \quad (\forall a \in \mathbb{R})$$

(\therefore) Markovの不等式より $P(X \geq a) = P(\varphi(X) \geq \varphi(a)) \leq \frac{E[\varphi(X)]}{\varphi(a)} //$

Cor $P(X \geq a) \leq \frac{E[e^{tx}]}{e^{ta}} \quad (\forall t \geq 0, a \in \mathbb{R})$

* モーメント母関数を用いて X の裾確率を評価できる。
→ 様々な確率集中不等式に用いられている。

Thm (Hoeffdingの不等式) \otimes

← 同一である必要はない

X_1, \dots, X_n : 独立 s.t. $a \leq X_i \leq b$ ($\forall i=1, 2, \dots, n$)

$E[X_i] = 0$ とする。

$\varepsilon > 0$ とする。

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i \geq \varepsilon\right) \leq \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right)$$

$$\left(P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i\right| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right)\right) \text{ とある}$$

Note

$$\varepsilon = \frac{b-a}{\sqrt{n}} \sqrt{\frac{1}{2}t} \quad \text{とおけば}$$

$$\left|\frac{1}{n} \sum_{i=1}^n X_i\right| \geq \frac{b-a}{\sqrt{n}} \sqrt{\frac{1}{2}t}$$

とある確率 $\leq 2e^{-t}$ 以下である。 $\Rightarrow \left|\frac{1}{n} \sum_{i=1}^n X_i\right| = O_p\left(\frac{1}{\sqrt{n}}\right)$

→ 大数の法則, 中心極限定理

\otimes 先の Chebyshevの不等式と比較。裾確率 \leq

$$O\left(\frac{1}{t^2}\right) \text{ とは } < O(e^{-t}) \text{ に } t \rightarrow \infty \text{ で } < \text{ なる。}$$

(2乗) (指数)

→ X_i の有界性を用いてよりよい裾確率評価は可能である。

Proof

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i \geq \varepsilon\right) \leq \frac{E[e^{\sum_{i=1}^n X_i}]}{e^{n\varepsilon}} \quad (\varepsilon > 0)$$

$$= \frac{\prod_{i=1}^n E[e^{X_i}]}{e^{n\varepsilon}} \quad (\because \text{独立})$$

$E[e^{\sum_{i=1}^n X_i}]$ を評価する.

$$a \leq X_i \leq b \text{ かつ } X_i = db + (1-d)a \quad \left(d = \frac{X_i - a}{b - a} \text{ と書くと}\right)$$

$X \mapsto e^{\pi X}$ の凸性より

$$\boxed{0 \leq d \leq 1}$$

$$e^{\pi X_i} \leq d e^{\pi b} + (1-d) e^{\pi a}$$

である. 右辺の期待値をとると.

$$E[e^{\pi X_i}] \leq E[d] e^{\pi b} + E[1-d] e^{\pi a}$$

$$= -\frac{a}{b-a} e^{\pi b} + \frac{b}{b-a} e^{\pi a}$$

$$= p e^{\pi b} + (1-p) e^{\pi a}$$

$$\left(\text{ただし } p = -\frac{a}{b-a} \geq 0\right)$$

$$\psi(\pi) = \ln(p e^{\pi b} + (1-p) e^{\pi a}) \text{ とおくと}$$

これは $P(Z=b) = p, P(Z=a) = 1-p$ なる Z のモーメント母関数

$$\psi(0) = \ln(1) = 0$$

$$\psi'(0) = \frac{p b e^{\pi b} + (1-p) a e^{\pi a}}{p e^{\pi b} + (1-p) e^{\pi a}} \Big|_{\pi=0} = 0 \quad (\text{平均})$$

$$\psi''(\pi) = \frac{p b^2 e^{\pi b} + (1-p) a^2 e^{\pi a}}{p e^{\pi b} + (1-p) e^{\pi a}} - \frac{(p b e^{\pi b} + (1-p) a e^{\pi a})^2}{(p e^{\pi b} + (1-p) e^{\pi a})^2}$$

$$\Rightarrow p_{\pi} = \frac{p e^{\pi b}}{p e^{\pi b} + (1-p) e^{\pi a}} \text{ とおくと}$$

$Z_{\pi} \in \{b, a\}$ $P(Z_{\pi}=b) = p_{\pi}, P(Z_{\pi}=a) = 1-p_{\pi}$ なる r.v. とおくと.

$$\psi''(\pi) = b^2 p_{\pi} + a^2 (1-p_{\pi}) - (b p_{\pi} + a (1-p_{\pi}))^2$$

$$= \text{Var}[Z_{\pi}] = \text{Var}\left[Z_{\pi} - \frac{a+b}{2}\right] \leq E\left[\left(Z_{\pi} - \frac{a+b}{2}\right)^2\right] \leq \frac{(a-b)^2}{4}$$

まず.

$$\begin{aligned} \psi(t) &= \psi(0) + t \psi'(0) + \frac{1}{2} t^2 \psi''(\xi) \quad (0 \leq \xi \leq t) \\ &\leq 0 + 0 + \frac{1}{8} t^2 (a-b)^2 \end{aligned}$$

次に.

$$E[e^{tX_i}] \leq \exp\left(\frac{t^2}{8} (b-a)^2\right) \quad (t \geq 0)$$

よって

$$\frac{\prod_{i=1}^n E[e^{tX_i}]}{e^{tn\varepsilon}} \leq \exp\left(\frac{nt^2}{8} (b-a)^2 - tn\varepsilon\right)$$

$$\Rightarrow t = \frac{4\varepsilon}{(b-a)^2} \quad \text{と } t \wedge \psi(t) \text{ を}$$

$$\text{右辺} \leq \exp\left(\frac{2n\varepsilon^2}{(b-a)^2} - \frac{4n\varepsilon^2}{(b-a)^2}\right) = \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right)$$

$$\bullet \left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq \varepsilon \text{ のとき } P\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq \varepsilon\right) \leq P\left(\frac{1}{n} \sum_{i=1}^n X_i \geq \varepsilon\right) + P\left(\frac{1}{n} \sum_{i=1}^n (-X_i) \geq \varepsilon\right)$$

対称性より //

Note

• $a_i \leq X_i \leq b_i$ のように a, b が i に依存するとき.

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i \geq \varepsilon\right) \leq \exp\left(-\frac{2n\varepsilon^2}{\frac{1}{n} \sum_{i=1}^n (b_i - a_i)^2}\right)$$

と主張できる.

• 証明は)

$$E[e^{tX_i}] \leq e^{\frac{t^2}{2} \sigma^2} \quad (b_i): \text{ sub-Gaussian}$$

← Gauss の条件が緩くても成り立つ

よって X_i が有界で $\sigma < 2\varepsilon$

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i \geq \varepsilon\right) \leq \exp\left(-\frac{n\varepsilon^2}{2\sigma^2}\right)$$

が成り立つ。

Thm (Bernstein の不等式)

X_1, \dots, X_n : 独立 (同-分布限定なし)

$$E[X_i] = 0, \text{Var}[X_i] \leq \sigma^2, |X_i| \leq M \text{ (a.s.)}$$

なら,

$$P\left(\frac{1}{n} \sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{nt^2}{2\sigma^2 + \frac{2}{3}Mt}\right) \quad (t \geq 0) //$$

\Rightarrow 分散の情報を利用して.

分散が小さい場合は、よりよいバウンドを得る.

* Hoeffding, Bernstein および他の拡張は統計学・学習理論で頻りに用いられる.

参考 Thm (Gaussian 集中不等式)

X_1, \dots, X_n は独立同一に標準正統分布に従うとする.

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ は L -Lipschitz 連続とする.

$$P\left(f(X_1, \dots, X_n) - E[f(X_1, \dots, X_n)] \geq t\right) \leq e^{-\frac{t^2}{2L^2}} \quad (t > 0) //$$

* 確率率は次元 n に依存しない!

ある有限な関数クラス G に対し.

$$f(x_1, \dots, x_n) := \sup_{g \in G} \frac{1}{n} \sum_{i=1}^n x_i g(z_i) \quad (z_i \text{ は固定})$$

これを用いる. \rightarrow ノンパラ回归, 学習理論で現れる (Gaussian width)

Thm (Talagrand の集中不等式)

$G: \|\cdot\|_\infty$ -norm に閉じた有限な関数クラス

$\forall g \in G$ 2: $E[g(X)] = 0, \text{Var}[g] \leq \sigma^2, \|g\|_\infty \leq M$ なら.

$$P\left(\sup_{g \in G} \frac{1}{n} \sum_{i=1}^n g(X_i) \geq 2 E\left[\sup_{g \in G} \frac{1}{n} \sum_{i=1}^n g(X_i)\right] + \sqrt{\frac{2t\sigma^2}{n}} + \frac{2M}{n}t\right) \leq e^{-t} //$$

\rightarrow 2 次元頻出の不等式

④ 確率変数数列の収束 ★重要

(Ω, \mathcal{F}, P) : prob. sp.

X_1, X_2, \dots : r.v. の列 on (Ω, \mathcal{F}, P)

X : 別の r.v.

F_n : X_n の分布関数

F : X の分布関数

Def (確率変数数列の収束) ★重要

1. 概然収束 (convergence almost surely)

$$X_n \xrightarrow{\text{a.s.}} X \stackrel{\text{def}}{\iff} P\left(\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1$$

2. 確率収束 (convergence in probability)

$$X_n \xrightarrow{P} X \stackrel{\text{def}}{\iff} \forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

3. p -次平均収束, L^p -収束 (conv. in p -th degree mean, conv. in L^p)

$$X_n \xrightarrow{L^p} X \stackrel{\text{def}}{\iff} \|X_n - X\|_p \rightarrow 0 \quad (p \geq 1)$$

4. 法則収束 (convergence in law, distribution weak convergence)

$$X_n \xrightarrow{w} X \stackrel{\text{def}}{\iff} \forall x \text{ の } F(x) \text{ の } \underline{\text{連続点}} \text{ } x \text{ において}$$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

※ L^p -空間は L^p -収束に関して 完備 (任意の $Cauchy$ -列が L^p -空間内で収束先を持つ)
 \implies L^p -空間は Banach 空間

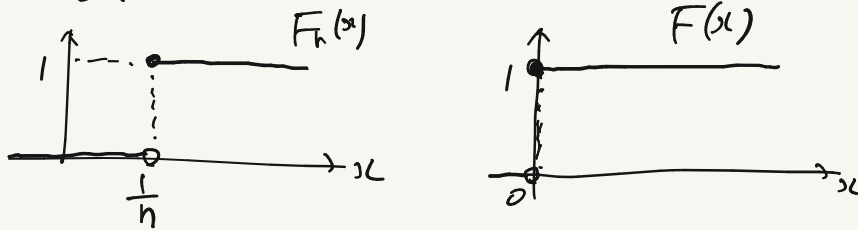
• 弱収束について.

- Fの連続点 についてのみ定義 (2つの理由):

$$X_n \in P(X_n = \frac{1}{n}) = 1 \text{ なる r.v.}$$

$$X \in P(X=0) = 1 \text{ なる r.v.}$$

とすると.



$$\frac{1}{n} \rightarrow 0 \text{ かつ } X_n \xrightarrow{d} X \text{ と思いたう.}$$

かつ $\lim_{n \rightarrow \infty} F_n(0) \neq F(0)$ かつ 不連続点 $x=0$ の時

収束が成り立たない. \rightarrow 不連続点を除外した方が都合よい.

弱収束 とは言う.

別定義 (Portmanteauの定理)

$$\textcircled{\otimes} \bullet X_n \xrightarrow{d} X \iff \forall f: \text{有界連続} \text{ に対し. } E[f(X_n)] \rightarrow E[f(X)]$$

• 任意の開集合 A に対し.

$$\limsup P(X_n \in A) \leq P(X \in A)$$

• 任意の開集合 B に対し.

$$\liminf P(X_n \in B) \geq P(X \in B)$$

(c.f. C_0 : 有界連続関数の集合. C_0 上の任意の有界線形写像 Φ はある正則 Borel 測度 μ を用いて.

$$\Phi(f) = \int f d\mu$$

と書ける.

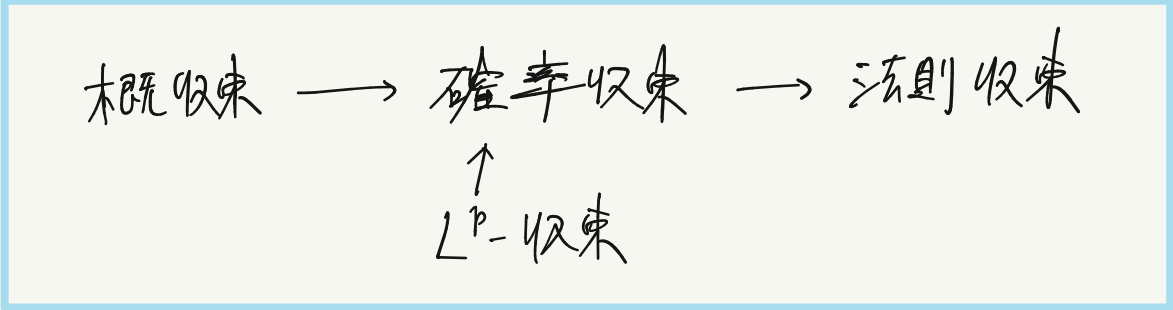
(Rieszの表現定理)

Thm (収束間の関係性) ~~⑧~~

$$(1) X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X$$

$$(2) X_n \xrightarrow{p} X \Rightarrow X_n \rightsquigarrow X$$

$$(3) X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{p} X$$



Proof

(1) $X_n \xrightarrow{a.s.} X$ かつ

$$P\left(\bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \left\{ |X_n - X| \leq \frac{1}{m} \right\}\right) = 1$$

\Leftarrow 任意の m に対し、ある N が存在し、任意の $n \geq N$ について

$$|X_n - X| \leq \frac{1}{m} \text{ が成立する事象}$$

と書ける。よって、 $\forall \varepsilon > 0$ に対し、十分大なる $m_0 \in \mathbb{Z}$ として

$$\frac{1}{m_0} \leq \varepsilon \text{ と } \mathbb{Z} \ni \mathbb{Z}$$

$$1 = P\left(\bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \{ \dots \}\right)$$

$$\leq P\left(\bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \left\{ |X_n - X| \leq \frac{1}{m_0} \right\}\right)$$

$$= \lim_{N \rightarrow \infty} P\left(\bigcap_{n \geq N} \left\{ |X_n - X| \leq \frac{1}{m_0} \right\}\right)$$

$$\leq \lim_{N \rightarrow \infty} P\left(\left\{ |X_N - X| \leq \frac{1}{m_0} \right\}\right)$$

$$\leq \lim_{N \rightarrow \infty} P\left(\left\{ |X_N - X| \leq \varepsilon \right\}\right)$$

つまり、 $X_n \xrightarrow{p} X$ である。

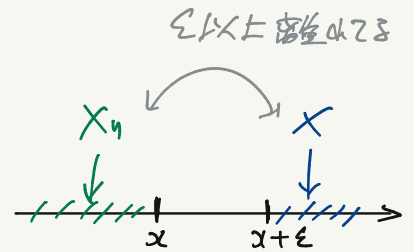
(2) $x \in F$ の連続点とする.

$$F_n(x) = P(X_n \leq x)$$

$$= P(\{X_n \leq x\} \cap \{x \leq x + \varepsilon\})$$

$$+ P(\{X_n \leq x\} \cap \{x > x + \varepsilon\})$$

$$\leq \underbrace{P(X \leq x + \varepsilon)}_{F(x + \varepsilon)} + \underbrace{P(|X_n - x| \geq \varepsilon)}_{\downarrow \lim_{n \rightarrow \infty} 0 \quad (\because X_n \xrightarrow{P} x)}$$



$$F(x - \varepsilon) = P(X \leq x - \varepsilon)$$

$$= P(\{X \leq x - \varepsilon\} \cap \{X_n \leq x\})$$

$$+ P(\{X \leq x - \varepsilon\} \cap \{X_n > x\})$$

$$\leq \underbrace{P(X_n \leq x)}_{F_n(x)} + \underbrace{P(|X - X_n| \geq \varepsilon)}_{\downarrow \lim_{n \rightarrow \infty} 0}$$

$$\text{以上より} \quad F(x - \varepsilon) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F(x + \varepsilon)$$

F は連続な点 x : $\varepsilon \rightarrow 0$ とする $\varepsilon > 0$.

$$F(x) = \lim_{n \rightarrow \infty} F_n(x)$$

を得る.

(3) Markov の不等式 (*)

$$\forall \varepsilon > 0: P(|X_n - x| \geq \varepsilon) \leq \frac{E[|X_n - x|^p]}{\varepsilon^p} \rightarrow 0$$

$$(\because X_n \xrightarrow{L^p} x)$$

(注) 弱収束にあつては X_n と X は同じ prob. sp. 上に u なる点を持つ. 確率収束と概収束は同じ prob. sp. 上に u なる点を持つ.

反例

(1) の逆は成り立たない例

(各 $n \in \mathbb{Z}^+$ に対して X_n と X は十分近い)。
 $(X_n(\omega))_{n=1}^{\infty}$ と $n \in \mathbb{Z}^+$ の列を考慮して $X(\omega)$ は収束しない
 状況を考慮してほしい。

$\Omega = [0, 1)$, P は Ω 上の一様分布

$n = 1, 2, 2^2, 2^3, \dots$ とする ($n = 2^k$)

$$X_{n,i} \stackrel{\text{def}}{=} \begin{cases} 1 & (\frac{i-1}{n} \leq \omega < \frac{i}{n}) \\ 0 & \text{otherwise} \end{cases} \quad (i = 1, 2, \dots, n)$$

$k=0$: $X_1 = X_{1,1} = 1$ (a.s.)

$k=1$: $X_2 = X_{2,1} = \mathbb{1}_{[0, \frac{1}{2})}$

$X_3 = X_{2,2} = \mathbb{1}_{[\frac{1}{2}, 1)}$

$k=2$: $X_4 = X_{4,1} = \mathbb{1}_{[0, \frac{1}{4})}$

$X_5 = X_{4,2} = \mathbb{1}_{[\frac{1}{4}, \frac{1}{2})}$

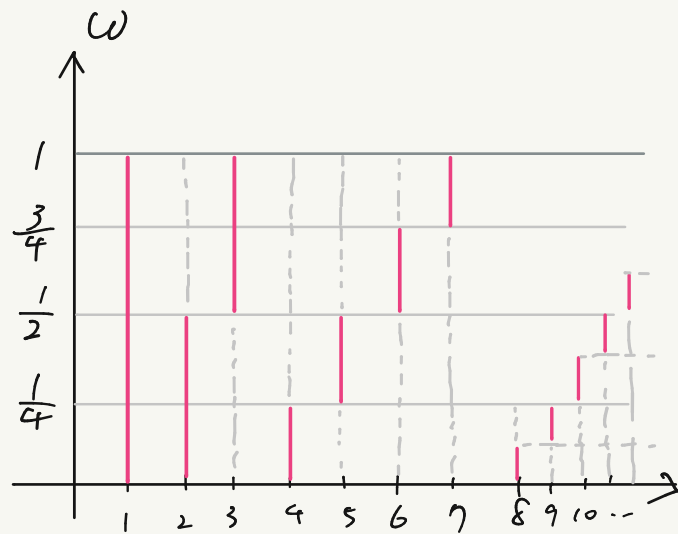
$X_6 = X_{4,3} = \mathbb{1}_{[\frac{1}{2}, \frac{3}{4})}$

$X_7 = X_{4,4} = \mathbb{1}_{[\frac{3}{4}, 1)}$

$k=3$: $X_8 = X_{8,1} = \mathbb{1}_{[0, \frac{1}{8})}$

\vdots

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & (\omega \in A) \\ 0 & (\omega \notin A) \end{cases}$$



— 2. 1, 4, 8, 16, 32, 64, 128, ...

とすると $P(|X_m| \geq \epsilon) \equiv \frac{1}{2^k}$ (if $2^k \leq m < 2^{k+1}$) ($0 < \forall \epsilon < 1$)

すなわち $X_m \xrightarrow{P} 0$

しかし $\forall \omega \in [0, 1)$ (a.s.) $\limsup_{n \rightarrow \infty} X_n(\omega) = 1$,

$\liminf_{n \rightarrow \infty} X_n(\omega) = 0$ (f.o.v.)

a.s. 2 収束しない (すなわち $X_n \xrightarrow{a.s.} X = 0$ は成り立たない) //

(2) の逆例: 成り立たない例]

$\Omega = \{0, 1\} \subset \mathbb{R}$. $P(\omega=0) = P(\omega=1) = \frac{1}{2}$ とする.

$$X_n(\omega) = \begin{cases} 1 & (\omega=0) \\ 0 & (\omega=1) \end{cases}$$

$$X(\omega) = \begin{cases} 0 & (\omega=0) \\ 1 & (\omega=1) \end{cases}$$

しかし、当然 $|X_n - X| = 1$ (a.s.) となる: $X_n \xrightarrow{p} X$ ではない.

しかし、 $F_n(x) = \begin{cases} 0 & (x < 0) \\ \frac{1}{2} & (0 \leq x < 1) \\ 1 & (1 \leq x) \end{cases} = F(x)$ である.

$F_n(x) = F(x)$ ($\forall x \in \mathbb{R}$) となる: 特異的. $X_n \xrightarrow{d} X$ である.

\uparrow $\lim F_n(x) = F(x)$ である. //

演習問題は近日別途公開
