Markov Chain

Definition 44

The process $\{X_n\}$ is a *Markov chain* if it satisfies the Markov property:

$$\mathbb{P}(X_n = j | X_0 = x_0, \dots, X_{n-1} = i) = \mathbb{P}(X_n = j | X_{n-1} = i)$$

for all $i, j, x_0, \dots < x_{n-2} \in S$ and for all $n = 1,2,3, \dots$.

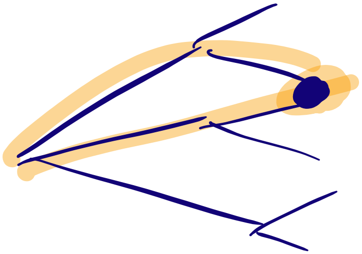
- The Markov property implies that:

$$\mathbb{P}\left(X_{n_k} = j \middle| X_{n_0} = x_0, \dots, X_{n_{k-1}} = i\right) = \mathbb{P}\left(X_{n_k} = j \middle| X_{n_{k-1}} = i\right)$$

for all $k, n, all\ n_0 \le n_1 \le \cdots \le n_{k-1} \le n_k$ and all $i, j, x_0, \dots, x_{n_{k-1}}$

- Also

$$\mathbb{P}(X_{n+m} = j | X_0 = x_0, \dots, X_m = i) = \mathbb{P}(X_{n+m} = j | X_m = i)$$



Homogeneous Chain

- The evolution of a markov chain is defined by its transition probability, defined by $\mathbb{P}(X_{n+1} = j | X_n = i)$ (where without loss of generality we may assume that S is an integer set.

Definition 45

- The chain $\{X_n\}$ is called *homogeneous* if its transition probabilities do not depend on the time, i.e.,

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

for all n, i, j. The *transition probability matrix* $\boldsymbol{P} = [p_{i,j}]$ *is the* $|S|$ *x* $|S|$ matrix of the transition probabilities, such that $p_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i)$

Transition Matrix

Theorem

The transition matrix **P** of a Markov chain is a *stochastic matrix*, that is, it has non-negative elements such that

$$\sum_{j \in S} p_{i,j} = 1$$

(sum of the elements on each row yields 1)

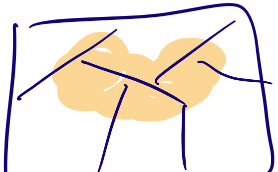
- In order to characterize the probability for n steps transitions, we introduce the n-step transition probability matrix with elements

$$p_{i,j}(m, m+n) = \mathbb{P}(X_{m+n} = j | X_m = i)$$

- By homogeneity, we have that $\boldsymbol{P}(m, m+1) = \boldsymbol{P}$.

- Furthermore, $\boldsymbol{P}(m, m+n) \triangleq \boldsymbol{P}^{(n)}$ does not depend on m.

$P(A|B_2) \cdot P(B_2)$



$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$

Transition Matrix

Theorem

$P(1, 3) = \sum P(1, 1+1) \cdot P(1+1, 1+2)$

$$p_{i,j}(m, m+n+r) = \sum_k p_{i,k}(m, m+n) p_{k,j}(m+n, m+n+r)$$

Therefore, $\boldsymbol{P}(m, m+n+r) = \boldsymbol{P}(m, m+n)\boldsymbol{P}(m+n, m+n+r)$. It follows that for homogeneous Markov chains, $\boldsymbol{P}(m, m+n) = P^n$, i.e., $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$

P^1 - trans. matrix 1-step

P^3 - — " — 3-steps ahead

$\boxed{P^3 = P \cdot P \cdot P}$

Initial State pmf

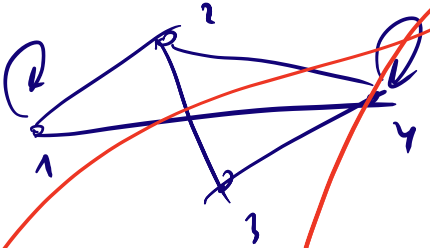
$u_n = u_0 \cdot P^n$

$S = \{1, 2, 3, 4\}$

- We let $\boldsymbol{u}(n)$ denote the pmf of X_n, that is, for each n we have that $\boldsymbol{u}(n)$ is a vector with $|S|$ non-negative components that sum to 1.

Lemma

$\boldsymbol{u}(m+n) = \boldsymbol{u}(m)\mathbf{P}^n$, and hence $\boldsymbol{u}(n) = \boldsymbol{u}(0)\mathbf{P}^n$. This describes the pmf of X_n in terms of the initial state pmf $\boldsymbol{u}(0)$.



$1 - 4 - 3 - 4 - 2 - 1 - 3$

$u_0 = \begin{bmatrix} 10\% \\ 20\% \\ 30\% \\ 40\% \end{bmatrix}$

$u_1 = u_0 \cdot P$

$u_n = u_0 \cdot P^n$

$\boxed{A\vec{v} = \lambda \cdot \vec{v}}$

1

$1 \cdot u = u \cdot P$

Example

Let $S = \{1, 2, 3, 4, 5,6\}$ and consider the transition matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Stationary Distribution

Definition

The vector π is called a *stationary distribution* of the chain if it has entries $\{\pi_j\colon\ j \in S\}$ such that:

a) $\pi_j \geq 0$ for all j, and $\sum_{j\in S} \pi_j = 1$.

b) it satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, that is, $\pi_j = \sum_i \pi_i p_{i,j}$ for all $j \in S$.

- This is called "stationary distribution" since if X_0 is distributed with $\boldsymbol{u}(0) = \boldsymbol{\pi}$, then all X_n will have the same distribution, in fact

$$\mathbf{u}(n) = \mathbf{u}(0)\mathbf{P}^n = \boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi}\mathbf{P}\mathbf{P}^{n-1} = \boldsymbol{\pi}\mathbf{P}^{n-1} = \cdots = \boldsymbol{\pi}$$

- Given the classification of chains and the decomposition theorem, we shall assume that the chain is *irreducible*, that is, its state space is formed by a single equivalence class of intercommunicating (persistent) states C or by the class of transient states T.

Stationary Distribution

Theorem

A irreducible chain has a stationary distribution π if and only if all states are *non-null persistent*. In this case, π is unique and satisfies $\pi_j = \frac{1}{\mu_j}$,

where is the *mean recurrence time* of state j.

Let $S = \{1, 2\}$ and consider the transition matrix

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

Convergence

Convergence

Definition *Let $\{x_n, n \geq 1\}$ be a real-valued sequence, i.e., a map from $\mathbb{N}$ to $\mathbb{R}$. We say that the sequence $\{x_n\}$ converges to some $x \in \mathbb{R}$ if there exists an $n_0 \in \mathbb{N}$ such that for all $\epsilon > 0$,*

$$|x_n - x| < \epsilon, \forall\, n \geq n_0.$$

We say that the sequence $\{x_n\}$ converges to $+\infty$ if for any $M > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $x_n > M$.
We say that the sequence $\{x_n\}$ converges to $-\infty$ if for any $M > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $x_n < -M$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real-valued random variables defined on this probability space.

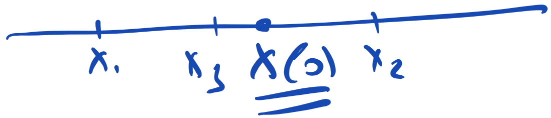
Convergence

Definition **[Definition 0 (Point-wise convergence or sure convergence)]**
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ *is said to converge point-wise or surely to* X *if*

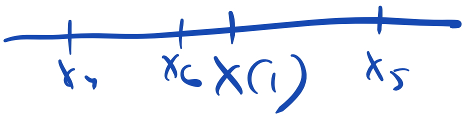
$$X_n(\omega) \to \underline{X(\omega)}, \quad \forall\, \omega \in \Omega.$$

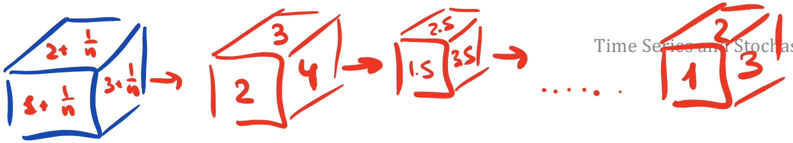
$\omega = 0, 1$

$X_n(0): \; X_1 \; X_2 \; X_3$



$X_n(1): \; X_4, \; X_5, \; X_6$





Convergence

Definition **Definition 1 (Almost sure convergence or convergence with probability 1)]**
A sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ *is said to converge almost surely or with probability 1 (denoted by a.s. or w.p. 1) to* X *if*

$$\mathbb{P}\left(\{\omega | X_n(\omega) \to X(\omega)\}\right) = 1.$$

Convergence

Definition **[Definition 2 (convergence in probability)]**
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ *is said to converge in probability (denoted by i.p.) to* X *if*

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0, \quad \forall\, \epsilon > 0.$$

Convergence

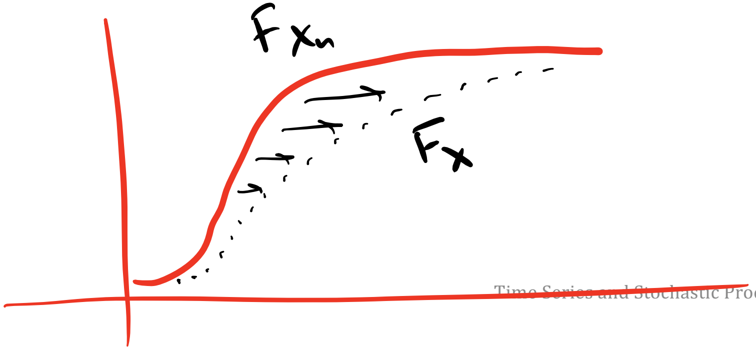
Definition **[Definition 3 (convergence in r^{th} mean)]**
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ *is said to converge in* r^{th} *mean to* X *if*

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

Convergence

Definition **[Definition 4 (convergence in distribution or weak convergence)]**
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in distribution to X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \forall\, x \in \mathbb{R} \text{ where } F_X(\cdot) \text{ is continuous.}$$



Convergence

(1) *Point-wise Convergence:* $X_n \xrightarrow{\text{p.w.}} X$.

(2) *Almost sure Convergence:* $X_n \xrightarrow{\text{a.s}} X$ or $X_n \xrightarrow{\text{w.p.}} {}^1 X$.

(3) *Convergence in probability:* $X_n \xrightarrow{\text{i.p.}} X$.

(4) *Convergence in r^{th} mean:* $X_n \xrightarrow{r} X$. When $r = 2$, $X_n \xrightarrow{\text{m.s.}} X$.

(5) *Convergence in Distribution:* $X_n \xrightarrow{\text{D}} X$.

Convergence

Example: Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and a sequence of random variables $\{X_n, n \geq 1\}$ defined by

$$X_n(\omega) = \begin{cases} n, & \text{if } \omega \in \left[0, \frac{1}{n}\right], \\ 0, & \text{otherwise.} \end{cases}$$

Convergence

$$X_n = \begin{cases} n, & \text{with probability } \frac{1}{n}, \\ 0, & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

Clearly, when $\omega \neq 0$, $\lim_{n\to\infty} X_n(\omega) = 0$ but it diverges for $\omega = 0$. This suggests that the limiting random variable must be the constant random variable 0. Hence, except at $\omega = 0$, the sequence of random variables converges to the constant random variable 0. Therefore, this sequence does not converge surely, but converges almost surely.

Convergence

For some $\epsilon > 0$, consider

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\left(|X_n| > \epsilon\right) &= \lim_{n\to\infty} \mathbb{P}\left(X_n = n\right), \\ &= \lim_{n\to\infty} \left(\frac{1}{n}\right), \\ &= 0. \end{aligned}$$

Hence, the sequence converges in probability.

Convergence

Consider the following two expressions:

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left[|X_n|^2\right] &= \lim_{n\to\infty} \left(n^2 \times \frac{1}{n} + 0\right), \\
&= \infty.
\end{aligned}
$$

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{E}\left[|X_n|\right] &= \lim_{n\to\infty} \left(n \times \frac{1}{n} + 0\right), \\
&= 1.
\end{aligned}
$$

X, (o) : 2 2 2 2 2 10 2 2 2 2 2 10 2 2 2 2 2

Convergence

$$
\begin{tikzcd}
\text{p.w.} \arrow[r, Rightarrow] & \text{a.s.} \arrow[dr, Rightarrow] & & \\
& & \text{i.p.} \arrow[r, Rightarrow] & \text{D} \\
r^{th} \text{ mean } (r \geq 1) \arrow[urr, Rightarrow] \arrow[rrr, Rightarrow] & & & s^{th} \text{ mean } (r > s \geq 1)
\end{tikzcd}
$$

Convergence

Theorem $\quad X_n \xrightarrow{r} X \implies X_n \xrightarrow{\text{i.p.}} X, \quad \forall\, r \geq 1.$

Proof: Consider the quantity $\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right)$. Applying Markov's inequaliy, we get

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) &\leq \lim_{n\to\infty} \frac{\mathbb{E}\left[|X_n - X|^r\right]}{\epsilon^r}, \ \forall \epsilon > 0, \\ &\overset{(a)}{=} 0, \end{aligned}$$

where (a) follows since $X_n \xrightarrow{r} X$. Hence proved.

$$P(|A| > \varepsilon) \leq \frac{E(A^r)}{\varepsilon^r}$$

Convergence

Theorem $\quad X_n \xrightarrow{\text{i.p.}} X \implies X_n \xrightarrow{\text{D}} X.$

Proof: Fix an $\epsilon > 0$.

$$\begin{aligned} F_{X_n}(x) &= \mathbb{P}(X_n \le x), \\ &= \mathbb{P}(X_n \le x, X \le x+\epsilon) + \mathbb{P}(X_n \le x, X > x+\epsilon), \\ &\le F_X(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon). \end{aligned}$$

0

Similarly,

$$\begin{aligned} F_X(x-\epsilon) &= \mathbb{P}(X \le x-\epsilon), \\ &= \mathbb{P}(X \le x-\epsilon, X_n \le x) + \mathbb{P}(X \le x-\epsilon, X_n > x), \\ &\le F_{X_n}(x) + \mathbb{P}(|X_n - X| > \epsilon). \end{aligned}$$

Thus,

$$F_X(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \le F_{X_n}(x) \le F_X(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon).$$

As $n \to \infty$, since $X_n \xrightarrow{\text{i.p.}} X$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$. Therefore,

$$F_X(x-\epsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x+\epsilon), \ \forall \epsilon > 0.$$

If F is continuous at x, then $F_X(x-\epsilon) \uparrow F_X(x)$ and $F_X(x+\epsilon) \downarrow F_X(x)$ as $\epsilon \downarrow 0$. Hence proved.

Convergence

$\checkmark$ $E((X_n - X)^R) \to 0$ $E((X_n - X)^S) \to 0$?

Theorem $X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X$, *if* $r > s \geq 1$.

$$(\mathbb{E}[|X_n - X|^s])^{1/s} \leq (\mathbb{E}[|X_n - X|^r])^{1/r},$$

$$(E(A^S))^{1/S} \leq (E(A^R))^{1/R}$$

$S < R$

Convergence

Theorem $\quad X_n \xrightarrow{\text{i.p.}} X \;\not\Rightarrow\; X_n \xrightarrow{\text{r}} X$ *in general.*

Proof: Proof by counter-example:
Let X_n be an independent sequence of random variables defined as

$$X_n = \begin{cases} n^3, & \text{w.p. } \frac{1}{n^2}, \\ 0, & \text{w.p. } 1\text{-}\frac{1}{n^2}. \end{cases}$$

Then, $\mathbb{P}(|X_n| > \epsilon) = \frac{1}{n^2}$ for large enough n, and hence $X_n \xrightarrow{\text{i.p.}} 0$. On the other hand, $\mathbb{E}[|X_n|] = n$, which diverges to infinity as n grows unbounded. ∎

Convergence

Theorem $X_n \xrightarrow{\text{D}} X \;\not\Rightarrow\; X_n \xrightarrow{\text{i.p.}} X$ *in general.*

Proof: Proof by counter-example:
Let X be a Bernoulli random variable with parameter 0.5, and define a sequence such that $X_i = X\ \forall\ i$. Let $Y = 1 - X$. Clearly, $X_i \xrightarrow{\text{D}} Y$. But, $|X_i - Y| = 1$, $\forall\ i$. Hence, X_i does not converge to Y in probability. ■

Convergence

Theorem $X_n \xrightarrow{\text{i.p.}} X \not\Rightarrow X_n \xrightarrow{\text{a.s.}} X$ *in general.*

Moment generating Function

$X \xrightarrow{\infty} 0$

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} 1, & \text{w.p. } \frac{1}{n}, \\ 0, & \text{w.p. } 1\text{-}\frac{1}{n}. \end{cases}$$

$$\to X_n = \begin{cases} 1, & 1/2 \\ 0, & 1/2 \end{cases}$$

$\lim_{n\to\infty} \mathbb{P}(|X_n| > \epsilon) = \lim_{n\to\infty} \mathbb{P}(X_n = 1) = \lim_{n\to\infty} \frac{1}{n} = 0$. So, $X_n \xrightarrow{\text{i.p.}} 0$.

Let A_n be the event that $\{X_n = 1\}$. Then, A_n's are independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. By Borel-Cantelli Lemma 2, w.p. 1 infinitely many A_n's will occur, *i.e.*, $\{X_n = 1\}$ i.o.. So, X_n does not converge to 0 almost surely. ■

Convergence

Theorem $X_n \xrightarrow{s} X \not\Rightarrow X_n \xrightarrow{r} X$ *if* $r > s \geq 1$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} n, & \text{w.p. } \frac{1}{n^{\frac{r+s}{2}}}, \\ 0, & \text{w.p. } 1\text{-} \frac{1}{n^{\frac{r+s}{2}}}. \end{cases}$$

Hence, $\mathbb{E}[|X_n^s|] = n^{\frac{s-r}{2}} \to 0$. But, $\mathbb{E}[|X_n^r|] = n^{\frac{r-s}{2}} \to \infty$.

Convergence

Theorem $X_n \xrightarrow{\text{m.s.}} X \;\not\Rightarrow\; X_n \xrightarrow{\text{a.s.}} X$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} 1, & \text{w.p. } \frac{1}{n}, \\ 0, & \text{w.p. } 1\text{-}\frac{1}{n}. \end{cases}$$

$\mathbb{E}[X_n^2]=\frac{1}{n}$. So, $X_n \xrightarrow{\text{m.s.}} 0$. X_n does not converge to 0 almost surely.

Convergence

Theorem $\quad X_n \xrightarrow{\text{a.s.}} X \;\not\Rightarrow\; X_n \xrightarrow{\text{m.s.}} X$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent of random variables defined as

$$X_n(\omega) = \begin{cases} n, & \omega \in (0, \frac{1}{n}), \\ 0, & \text{otherwise.} \end{cases}$$

We know that X_n converges to 0 almost surely. $\mathbb{E}[X_n^2]=n \longrightarrow \infty$. So, X_n does not converge to 0 in the mean-squared sense. ■

Before proving the implication $X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\text{i.p.}} X$, we derive a sufficient condition followed by a necessary and sufficient condition for almost sure convergence.

Convergence

Theorem 28.20 [Skorokhod's Representation Theorem]
Let $\{X_n, n \geq 1\}$ and X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that X_n converges to X in distribution. Then, there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, and random variables $\{Y_n, n \geq 1\}$ and Y on $(\Omega', \mathcal{F}', \mathbb{P}')$ such that,

a) $\{Y_n, n \geq 1\}$ *and* Y *have the same distributions as* $\{X_n, n \geq 1\}$ *and* X *respectively.*

b) $Y_n \stackrel{a.s.}{\to} Y$ *as* $n \to \infty$.

Convergence

Theorem 28.21 [Continuous Mapping Theorem]
If $X_n \xrightarrow{D} X$, and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, then $g(X_n) \xrightarrow{D} g(X)$.

Proof: By Skorokhod's Representation Theorem, there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, and $\{Y_n, n \geq 1\}$, Y on $(\Omega', \mathcal{F}', \mathbb{P}')$ such that, $Y_n \xrightarrow{a.s.} Y$. Further, from continuity of g,
$\{\omega \in \Omega' \mid g(Y_n(\omega)) \to g(Y(\omega))\} \supseteq \{\omega \in \Omega' \mid Y_n(\omega) \to Y(\omega)\}$,
$\Rightarrow \mathbb{P}(\{\omega \in \Omega' \mid g(Y_n(\omega)) \to g(Y(\omega))\}) \geq \mathbb{P}(\{\omega \in \Omega' \mid Y_n(\omega) \to Y(\omega)\})$,
$\Rightarrow \mathbb{P}(\{\omega \in \Omega' \mid g(Y_n(\omega)) \to g(Y(\omega))\}) \geq 1$,
$\Rightarrow g(Y_n) \xrightarrow{a.s.} g(Y)$,
$\Rightarrow g(Y_n) \xrightarrow{D} g(Y)$.
This completes the proof since, $g(Y_n)$ has the same distribution as $g(X_n)$, and $g(Y)$ has the same distribution as $g(X)$. ■

Convergence

Theorem 28.23 *If* $X_n \xrightarrow{D} X$*, then* $C_{X_n}(t) \longrightarrow C_X(t)$*,* $\forall t$.

Proof: If $X_n \xrightarrow{D} X$, from Skorokhod's Representation Theorem, there exist random variables $\{Y_n\}$ and Y such that $Y_n \xrightarrow{a.s.} Y$.
So,

$$\cos(Y_n t) \longrightarrow \cos(Yt), \ \cos(X_n t) \longrightarrow \cos(Xt), \ \forall t.$$

As $\cos(\cdot)$ and $\sin(\cdot)$ are bounded functions,

$$\mathbb{E}[\cos(Y_n t)] + i\mathbb{E}[\sin(Y_n t)] \longrightarrow \mathbb{E}[\cos(Yt)] + i\mathbb{E}[\sin(Yt)], \ \forall t.$$

$$\Rightarrow C_{Y_n}(t) \longrightarrow C_Y(t), \ \forall t.$$

We get,

$$C_{X_n}(t) \longrightarrow C_X(t), \ \forall t,$$

since distributions of $\{X_n\}$ and X are same as those of $\{Y_n\}$ and Y respectively, from Skorokhod's Representation Theorem. ■

Convergence

Example 1: Let the random variable U be uniformly distributed on $[0, 1]$. Consider the sequence defined as:

$$X(n) = \frac{(-1)^n U}{n}.$$

1. *Almost sure convergence:* Suppose

   $$U = a.$$

   The sequence becomes

   $$\begin{aligned} X_1 &= -a, \\ X_2 &= \frac{a}{2}, \\ X_3 &= -\frac{a}{3}, \\ X_4 &= \frac{a}{4}, \\ &\vdots \end{aligned}$$

   In fact, for any $a \in [0, 1]$

   $$\lim_{n \to \infty} X_n = 0,$$

   therefore, $X_n \xrightarrow{a.s.} 0$.

Convergence

Convergence in mean square sense:

In order to answer this question, we need to prove that

$$\lim_{n\to\infty} E\left[|X_n - 0|^2\right] = 0.$$

We know that,

$$\begin{aligned}
\lim_{n\to\infty} E\left[|X_n - 0|^2\right] &= \lim_{n\to\infty} E\left[X_n^2\right], \\
&= \lim_{n\to\infty} E\left[\frac{U^2}{n^2}\right], \\
&= \lim_{n\to\infty} \frac{1}{n^2} E\left[U^2\right], \\
&= \lim_{n\to\infty} \frac{1}{n^2} \int_0^1 u^2 du, \\
&= \lim_{n\to\infty} \frac{1}{n^2} \left.\frac{u^3}{3}\right]_0^1, \\
&= \lim_{n\to\infty} \frac{1}{3n^2}, \\
&= 0.
\end{aligned}$$

Hence, $X_n \xrightarrow{m.s.} 0$.