Markov chain!

Peter Lukianchenko

11 September 2021

Syllabus

Module		1	
	Class	1	Intro + Statistics + Markov chain
		2	Markov Chain + Convergency
		3	Convergency + MGF
		4	Markov Chain
		5	Conditional Expectation
		6	Sigma Algebra
		7	Basics of Stochastics
Module		2	
	Class	1	Martingales
		2	Martingales
		3	Brownian Motion
		4	Ito Lemma
		5	Black Scholes Model
		6	Black Scholes Model
		7	BS Model + Coding
		8	BS Model + Coding
Module		3	
	Class	1	Course Overview and Recall all we did
		2	Stationary Time Series
		3	ETS
		4	ETS
		5	ARMA
		6	ARIMA + test
		7	Arch-Garch
		8	Point Estimators
		9	Fisher Information
		10	LR LM WALD
		11	Sufficient Estimator
Module		4	
	Class	1	UoL
		2	UoL
		3	UoL
		4	UoL

Reminder: Markov Property

- A sequence of RVs indexed by a variable $n \in \{0,1,2, \dots\}$ forms a discrete-time random process $\{X_n\} = \{X_n : n = 0,1,2, \dots\}$.
- The process if defined by the collection of all joint distributions of order m,

$$F_{X_{n_1}, \dots, X_{n_m}}(x_1, \dots, x_m)$$

for all instants $\{n_1, \dots, n_m\}$ and for all $m = 1, 2, 3, \dots$.

- If the process has the property that, given the present, the future is independent of the past, we say that the process satisfies the *Markov property*, or, it is a *Markov chain*.
- Furthermore, we are mainly interested in the case where X_t takes on values in some countable set S, called *state space*.

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 $\boldsymbol{P}(m, m+1) = \boldsymbol{P}$.

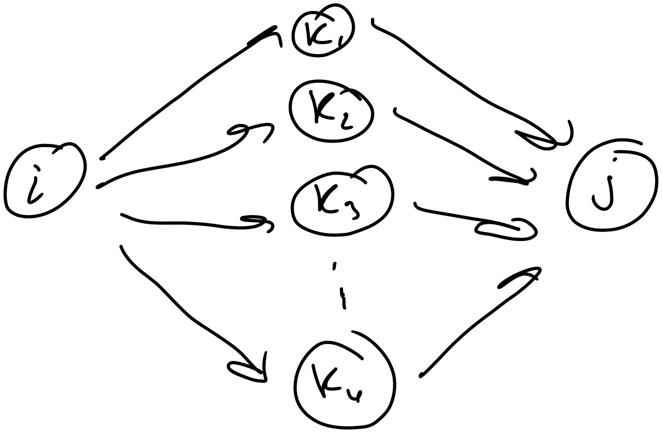
- Furthermore, $\boldsymbol{P}(m, m+n) \triangleq \boldsymbol{P}^{(n)}$ does not depend on m.

Transition Matrix

Theorem

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

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., $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$



Initial State pmf

- 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 $\boldsymbol{u}(0)$.

- We have a MC with transition matrix $\mathbf{P}$. Let $A \subset S$ denote a subset of states.
- For $i \notin A$, we are interested in the probability

$$\mathbb{P}(X_k \in A \text{ for some } k = 1, \ldots, m | X_0 = i) = \mathbb{P}\left(\cup_{k=1}^{m} \{X_k \in A\} | X_0 = i\right)$$

- Define the stopping time $N = \min\{n \geq 1 : X_n \in A\}$, and the new MC

$$W_n = \begin{cases} X_n & n < N \\ a & n \geq N \end{cases}$$

where a is a special symbol.

- $\{W_n\}$ has transition probability matrix

$$\begin{aligned} q_{i,j} &= p_{i,j} \quad \text{if } i \notin A,\ j \notin A \\ q_{i,a} &= \sum_{j \in A} p_{i,j} \quad \text{if } i \notin A \\ q_{a,a} &= 1 \end{aligned}$$

- Notice that the original MC enters a state in A by time m if and only if $W_m = a$, then

$$\mathbb{P}\left(\cup_{k=1}^{m}\{X_k \in A\} | X_0 = i\right) = \mathbb{P}(W_m = a | W_0 = i) = q_{i,a}(m) = [\mathbf{Q}^m]_{i,a}$$

- Now, we are interested in the probability of never entering the subset of states A in the times $0, \ldots, m$, i.e., for $i, j \notin A$ we wish to compute

$$\mathbb{P}\left(\{X_m = j\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i\right)$$

- The above condition is equivalent to $W_m = j$ conditionally on $W_0 = i$, hence

$$\mathbb{P}\left(\{X_m = j\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i\right) = q_{i,j}(m)$$

- Next, we wish to consider the first entrance probability, for $i \notin A$ and $j \in A$

$$
\begin{aligned}
&\mathbb{P}\left(\{X_m = j\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i\right) = \\
&= \sum_{r \notin A} \mathbb{P}\left(\{X_m = j\} \cap \{X_{m-1} = r\} \cap_{k=1}^{m-2} \{X_k \notin A\} | X_0 = i\right) \\
&= \sum_{r \notin A} \mathbb{P}\left(X_m = j | X_{m-1} = r\right) \mathbb{P}\left(\{X_{m-1} = r\} \cap_{k=1}^{m-2} \{X_k \notin A\} | X_0 = i\right) \\
&= \sum_{r \notin A} p_{r,j} q_{i,r}(m-1)
\end{aligned}
$$

- When $i \in A$ and $j \notin A$, we can determine

$$
\mathbb{P}\left(\{X_m = j\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i\right)
$$

- By conditioning on the first transition:

$$\mathbb{P}\left(\{X_m = j\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i\right) =$$

$$= \sum_{r \notin A} \mathbb{P}\left(\{X_m = j\} \cap_{k=2}^{m-1} \{X_k \notin A\} \cap \{X_1 = r\} | X_0 = i\right)$$

$$= \sum_{r \notin A} \mathbb{P}\left(\{X_m = j\} \cap_{k=2}^{m-1} \{X_k \notin A\} | X_1 = r\right) \mathbb{P}(X_1 = r | X_0 = i)$$

$$= \sum_{r \notin A} q_{r,j}(m-1) p_{i,r}$$

- Finally, we can calculate the conditional probability of $X_m = j$ given that the chains starts at $X_0 = i$ and has not entered the subset A in any time till time m. For $i, j \notin A$, we have

$$\mathbb{P}(X_m = j | \{X_0 = i\} \cap_{k=1}^{m} \{X_k \notin A\}) =$$

$$= \frac{\mathbb{P}(\{X_m = j\} \cap_{k=1}^{m} \{X_k \notin A\} | X_0 = i)}{\mathbb{P}(\cap_{k=1}^{m} \{X_k \notin A\} | X_0 = i)}$$

$$= \frac{\mathbb{P}(\{X_m = j\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i)}{\sum_{r \notin A} \mathbb{P}(\{X_m = r\} \cap_{k=1}^{m-1} \{X_k \notin A\} | X_0 = i)}$$

$$= \frac{q_{i,j}(m)}{\sum_{r \notin A} q_{i,r}(m)}$$

Persistent and Transient States

- A state i $\in S$ is called *persistent* (or recurrent) if

$$\mathbb{P}(X_n = i \; for \; some \; n \; \geq \; 1 | X_0 = i) \; = 1$$

- Otherwise, if the above probability is strictly less than 1, the state is called *transient*.

- We are interested in the *first passage* probability

$$f_{i,j}(n) = \mathbb{P}(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j | X_0 = i)$$

- We define $f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}(n)$. **Note**: state j is persistent if and only if $f_{j,j} = 1$.

Persistent and Transient States

- We define the generating functions

$$P_{i,j}(s) = \sum_{n=0}^{\infty} p_{i,j}(n)s^n, \qquad F_{i,j}(s) = \sum_{n=0}^{\infty} f_{i,j}(n)s^n$$

with the conditions: $p_{i,j}(0) = \delta_{i,j}$, $f_{i,j}(0) = 0$ for all $i, j \in S$.

- We shall assume $|s| < 1$, since in this way the generating functions converge, and occasionally evaluate limits for $s \uparrow 1$ using Abel's theorem.

- Notice that $f_{i,j} = F_{i,j}(1)$.

Theorem 48

a) $P_{i,i}(s) = 1 + P_{i,i}(s)F_{i,i}(s)$,

b) $P_{i,j}(s) = F_{i,j}(s)\, P_{j,j}(s)$, for $i \neq j$.

Persistent and Transient States

Corollary

a) State j is persistent if $\sum_n p_{j,j}(n) = \infty$. If this holds, then $\sum_n p_{i,j}(n) = \infty$ for all states i such that $f_{i,j} > 0$.

b) State j is transient if $\sum_n p_{j,j}(n) < \infty$. If this holds then $\sum_n p_{i,j}(n) < \infty$ for all i.

c) If j is transient, then $p_{i,j}(n) \to 0$ as $n \to \infty$.

Let $N(i)$ denote the number of times a chain visit state i given that its starting point is i. It is intuitively clear that

$$\mathbb{P}(N(i) = \infty) = \begin{cases} 1 & i \text{ is persistent} \\ 0 & i \text{ is transient} \end{cases}$$

- Let $T_j = \min\{n : X_n = j\}$ denote the time to the first visit o state j. This is a random variable, with the convention that $T_j = \infty$ is the visit never occurs.
- We have that $\mathbb{P}(T_j = \infty | X_0 = j) > 0$ if and only if j is transient an in this case $E[T_j | X_0 = j] = \infty$.

Mean Recurrence Time, Null State

- We define the *mean recurrence time* μ_j of a state j as

$$\mu_j = \mathbb{E}[T_j|X_0 = j] = \begin{cases} \sum_n n f_{j,j}(n) & j \text{ is persistent} \\ \infty & j \text{ is transient} \end{cases}$$

- **Note:** μ_j may be infinity even though the state is persistent!

Definition 47

A persistent state j is called *null* if $\mu_j = \infty$ and *non-null* or *positive* if $\mu_j < \infty$.

Theorem 49

A persistent state i is null if and only if $p_{i,i}(n) \rightarrow 0$ as $n \rightarrow \infty$. If this holds, then $p_{j,i}(n) > 0$ for all j.

Communicating and Intercommunicating States

- We say that state i *communicates* with state j (written $i \to j$) if $p_{i,j}(n) > 0$ for some $n \geq 0$. In this case, state j can be reached from state i with positive probability.
- We say that state i and j *intercommunicate* if $i \to j$ and $j \to i$. In this case, we write $i \leftrightarrow j$.
- Clearly, $i \leftrightarrow i$, since $p_{i,i}(0) = 1$.
- Also, $i \to j$ if and only if $f_{i,j} > 0$.
- Intercommunication $\leftrightarrow$ induces an equivalence relation on the state space.
- In fact, if: $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$. Hence, S can be partitioned into equivalence classes of mutually intercommunicating states.

Communicating and Intercommunicating States

Theorem 50

If $i \leftrightarrow j$ then:

a) i and j have the same period;

b) i is transient if and only if j is transient;

c) i is null persistent if and only if j is null persistent;

Decomposition theorem

- In light of Decomposition theorem (slide 22) we can identify the following behaviors:

- If the chain starts in some state $i \in C_r$, then it stays in C_r forever.

- If the chain starts in T, then either it stays in T forever or eventually will end up into some C_r.

- If S is finite (finite-state Markov chain), then the first possibility cannot occur

Lemma 34

If S is finite, then at least one state is persistent and all persistent states are non-null.

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 50

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 52

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

Cauchy's

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 $\epsilon > 0$,*

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

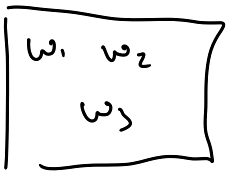
We say that the sequence $\{x_n\}$ converges to $+\infty$ if for any $M > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $x_n > M$.
We say that the sequence $\{x_n\}$ converges to $-\infty$ if for any $M > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,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

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*

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



$X_n(\omega_1) \to X(\omega_1)$ by C's def
$X_n(\omega_2) \to X(\omega_2)$ by C's def
$X_n(\omega_3) \to X(\omega_3)$ by C's def

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*

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

$$P\left(|X_n - X| < \varepsilon \mid \omega \right) \to 1$$

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

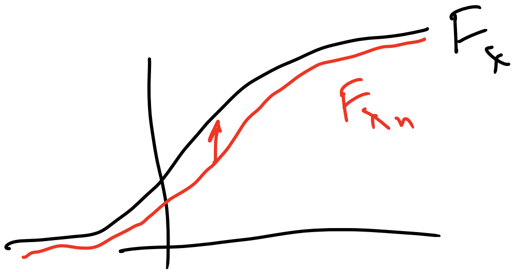
Definition **[Definition 3 (convergence in r^{th} mean)]**
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ *is said to converge in* r^{th} *mean to* X *if*

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

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

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

$[0;1]$ – sample space

$$X_1 = \begin{cases} 1 & \text{if } \omega \in [0;1] \\ 0, & \text{oth} \end{cases} \qquad X_2 = \begin{cases} 2 & \omega \in [0;\frac{1}{2}] \\ 0, & \dots \end{cases} \qquad X_3 = \begin{cases} 3 & \omega \in [0;\frac{1}{3}] \\ 0, & \dots \end{cases}$$

$X_1 = 1$

Convergence

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

$\rightarrow 0$

$P(X_n \to 0) \to 1$

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

Do not have a conv. in mean squared

$\lim_{n\to\infty} E[(X_n - X)^2] \to 0$

Consider the following two expressions:

m.s.

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

$$X_n^2 = \begin{cases} n^2 & , \frac{1}{n} \\ 0^2 & ; 1 - \frac{1}{n} \end{cases}$$

r=1

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n|\right] = \lim_{n\to\infty}\left(n \times \frac{1}{n} + 0\right); \\ = 1 \neq 0$$

$E[X_n^2] = n^2 \cdot \frac{1}{n} + 0^2 \cdot (1 - \frac{1}{n})$

Convergence

$$
\begin{tikzcd}
\text{p.w.} \arrow[r, Rightarrow] & \text{a.s.} \arrow[dr, Rightarrow, "4"] & & \\
& & \text{i.p.} \arrow[r, Rightarrow, "2"] & \text{D} \\
& r^{th} \text{ mean } (r \geq 1) \arrow[ur, Rightarrow, "1"] \arrow[rr, Rightarrow, "3"'] & & s^{th} \text{ mean } (r > s \geq 1)
\end{tikzcd}
$$

$r>s$ r.m. + $\Rightarrow$ s. m +

$$\lim_{n\to\infty} E[(X_n - X)^7] \to 0 \Rightarrow \lim_{n\to\infty} E[(X_n - X)^4] \to 0$$

Convergence

Marvov ineq-ty $P(X > a) < \frac{E[X^r]}{a^r}$

1

$E[(X_n - X)^r] \xrightarrow{n\to\infty} 0$

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

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.

Convergence

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

Proof: Fix an $\epsilon > 0$.

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

$\to 0$

$F_{X_n} \to F_X$

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

$\lim E[\ldots]^{\frac{1}{s}} \leq \lim E[\ldots]^{\frac{1}{r}} = 0$ $f(x)$ convex fn's

$$\boxed{f(x^s)^{1/s} \leq f(x^r)^{\frac{1}{r}} \quad r > s \geq 1}$$

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

Jensen's Inequality

$$\underbrace{\left(\mathbb{E}[|X_n - X|^s]\right)^{1/s}}_{\downarrow\, 0} \leq \underbrace{\left(\mathbb{E}[|X_n - X|^r]\right)^{1/r}}_{\downarrow\, 0},$$

Squeeze Th'm

Convergence

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

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

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

$E[X_n^r] = n^{3r-2} \rightarrow \infty$

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

$$X = \begin{cases} 1 & , \ 1/2 \\ 0 & , \ 1/2 \end{cases} \rightarrow \quad Y = 1 - X = \begin{cases} 0 & , \ 1/2 \\ 1 & , \ 1/2 \end{cases}$$

$$Y - X = \begin{cases} 1 \\ -1 \end{cases}$$

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

$y = 1 - X$

Convergence

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

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

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

$P(X_n = 0) \to 1$

$\lim_{n\to\infty} \mathbb{P}(|X_n| > \epsilon) = \lim_{n\to\infty} \mathbb{P}(X_n = 1) = \lim_{n\to\infty} \frac{1}{n} = 0$. 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. ■

$P(X_n = 1) = \frac{1}{n} > 0$

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

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

$\boxed{\mathbb{E}[X_n^2]=\frac{1}{n}. \text{ So, } X_n \xrightarrow{\text{m.s.}} 0.}$ 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) $Y_n \overset{a.s.}{\to} Y$ *as* $n \to \infty$.

Convergence

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

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, $Y_n \stackrel{a.s.}{\to} Y$. Further, from continuity of g,
$\{\omega \in \Omega' \mid g(Y_n(\omega)) \to g(Y(\omega))\} \supseteq \{\omega \in \Omega' \mid Y_n(\omega) \to Y(\omega)\}$,
$\Rightarrow \mathbb{P}(\{\omega \in \Omega' \mid g(Y_n(\omega)) \to g(Y(\omega))\}) \geq \mathbb{P}(\{\omega \in \Omega' \mid Y_n(\omega) \to Y(\omega)\})$,
$\Rightarrow \mathbb{P}(\{\omega \in \Omega' \mid g(Y_n(\omega)) \to g(Y(\omega))\}) \geq 1$,
$\Rightarrow g(Y_n) \stackrel{a.s.}{\to} g(Y)$,
$\Rightarrow g(Y_n) \stackrel{D}{\to} g(Y)$.
This completes the proof since, $g(Y_n)$ has the same distribution as $g(X_n)$, and $g(Y)$ has the same distribution as $g(X)$. $\blacksquare$

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

Thank you for your attention!

See next week!