Recap

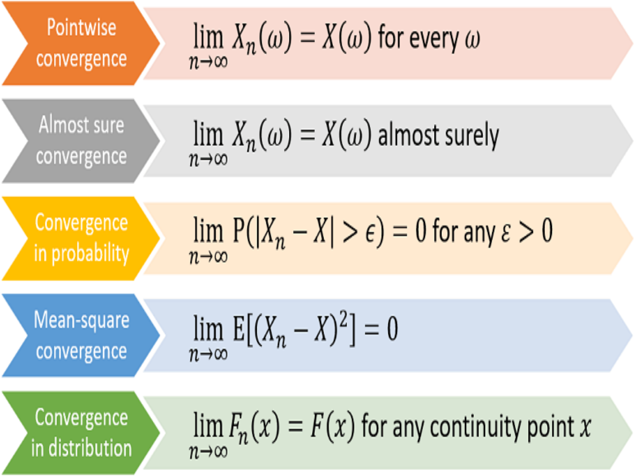


Image from statlect.com

Towards CLT

- Recall $\hat{\mu}_n = \frac{S_n}{n}$ where $S_n = \sum_{i=1}^{n} X_i$
- $\{X_i\}$ is i.i.d. with mean μ and variance σ^2.
- $E[\hat{\mu}_n] = \mu$ and $var(\hat{\mu}_n) = \frac{\sigma^2}{n}$
- Now consider $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. (centering and scaling). What is the mean and variance of Y_n?
- $E[Y_n] = 0$ and $Var(Y_n) = 1$. What is $F_{Y_n}(\cdot)$?
- What is $\lim_{n \to \infty} F_{Y_n}(\cdot)$? ANS: $F_{N(0,1)}(\cdot)$
- In other words, Y_n converges to $Y = N(0,1)$ in distribution.

CLT

Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables each with mean μ and variance $0 < \sigma^2 < \infty$.Denote $\hat{\mu}_n = \frac{\sum_{i=1}^n X_i}{n}$ and $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Then Y_n converges to $N(0,1)$ in distribution.

- X_i could be ANY discrete or continuous r.v. with finite mean and variance.
- What is the consequence when $E[X_i] = 0$ and $Var(X_i) = 1$.
- In this case, $Y_n = \frac{S_n}{\sqrt{n}}$ and it converges in distribution to $N(0,1)$.
- $\frac{S_n}{n}$ converges almost surely to 0 but $\frac{S_n}{\sqrt{n}}$ converges to a random variable $\mathcal{N}(0,1)$.

CLT

Let $\{X_n, n \geq 0\}$ denote a sequence of i.i.d random variables each with mean μ and variance $0 < \sigma^2 < \infty$.Denote $\hat{\mu}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ and $Y_n = \frac{\hat{\mu}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Then Y_n converges to $N(0,1)$ in distribution.

- CLT given a way to find approximate disribution of $\hat{\mu}_n$.
- Note that for large enough n, we can use the approximation that $Y_n \sim \mathcal{N}(0,1)$.
- Since Gaussianity is preserved under affine transformation, $\hat{\mu}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Example from probabilitycourse.com

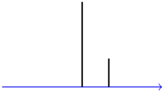
Assumptions:

- $X_1, X_2 \ldots$ are iid Bernoulli(p).
- $Z_n = \dfrac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}}$.

We choose $p = \frac{1}{3}$.

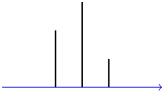
$$Z_1 = \frac{X_1 - p}{\sqrt{p(1-p)}}$$

PMF of Z_1



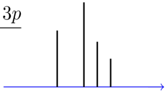
$$Z_2 = \frac{X_1 + X_2 - 2p}{\sqrt{2p(1-p)}}$$

PMF of Z_2



$$Z_3 = \frac{X_1 + X_2 + X_3 - 3p}{\sqrt{3p(1-p)}}$$

PMF of Z_3



$$Z_{30} = \frac{\sum_{i=1}^{30} X_i - 30p}{\sqrt{30p(1-p)}}$$

PMF of Z_{30}

Markov Chains

Introduction to Stochastic processes

- Stochastic process $\{X(t), t \in T\}$ is a collection of random variables defined such that for every $t \in T$ we have $X(t) : \Omega \rightarrow \mathcal{S}$.
- These random variables could be dependent and need not have identical distribution.
- T is the parameter space (often resembles time) and $\mathcal{S}$ is the state space.
- When T is countable, we have a discrete time process.
- If T is a subset of real line, we have a continuous time process.
- State space could be integers or real numbers

Examples of Stochastic Processes

- Sequence $\{X_i\}$ of i.i.d random variables.
- General random walk: If $X_1, X_2, \dots$ is a sequence i.i.d of random variables, then $S_n = \sum_{i=1}^{n} X_i$ is a random walk.
- Weiner process: $\{X(t), t \geq 0\}$ is a Weiner process if
  1. for every $t > 0$, $X(t) \sim \mathcal{N}(0, t)$.
  2. Often called as Brownian Motion as it was used by Robert Brown to describe motion of particle suspended in liquid.
  3. It is a scaling limit of a random walk (zoomed out BM).
  4. Trajectories are continuous but not differntiable (Financial modeling)
  5. Limit of Functional CLT (CLT for Stochastic processes)
- $\{X_n, n \geq 0\}$ is a martingale if $E[X_{n+1}|X_1, \dots, X_n)] = X_n$. (Applications in Finance, Optimal Stopping, pricing)
- $\{X(t), t \geq 0\}$ is a Markov process if for $t_1 < t_2 < \dots t_n < t$ $P(X(t) = j | X(t_1) = x_1, \dots, X(t_n) = i) = P(X(t) = j | X(t_n) = i)$

Discrete time Markov Chains (DTMC)

- A stochastic process $\{X_n, n \in \mathbb{Z}_+\}$ is a discrete time Markov chain if for any $n_1 < n_2 < \ldots < n_k < n$,

$$P(X_n = j | X_{n_1} = x_1, .., X_{n_k} = i) = P(X_n = j | X_{n_k} = i)$$

- This is called as the Markov property.
- $P(next\ state | past\ states,\ present\ state) = P(next\ state |\ present\ state)$
- Why Chain? You can view the successive random variables as a chain of states being visited in a sequence and where the next state visited depends only on the current state.
- We will throughout assume that the state space $\mathcal{S}$ is countable.

Recap

- Stochastic process $\{X(t), t \in T\}$ is a collection of random variables defined such that for every $t \in T$ we have a random variable $X(t)$ taking values in state space $\mathcal{S}$.
- A stochastic process that satisfies the Markov property is a Markov chain.
- $P(X_n = j | X_{n_1} = x_1, .., X_{n_k} = i) = P(X_n = j | X_{n_k} = i)$
- Markov chain is a stochastic process where the next state of the process depends on the present state butnot on previous states.

Running example: Coin with memory!

- In a Markovian coin with memory, the outcome of the next toss depends on the current toss.
- $X_n = 1$ for heads and $X_n = -1$ otherwise. $\mathcal{S} = \{+1, -1\}$.
- Sticky coin : $P(X_{n+1} = 1 | X_n = 1) = 0.9$ and $P(X_{n+1} = -1 | X_n = -1) = 0.8$ for all n.
- Flippy Coin: $P(X_{n+1} = 1 | X_n = 1) = 0.1$ while $P(X_{n+1} = -1 | X_n = -1) = 0.3$ for all n.
- This can be represented by a transition diagram (see board)
- The transition probability matrix P for the two cases is $P_s = \begin{bmatrix} 0.9 & .1 \\ 0.2 & 0.8 \end{bmatrix}$ and $P_f = \begin{bmatrix} 0.1 & 0.9 \\ 0.7 & 0.3 \end{bmatrix}$
- The row corresponds to present state and the column corresponds to next state.

Running example: Dice with memory!

- In a markovian dice with memory, the outcome of the next roll depends on the current roll.
- $X_n = i$ for $i \in \mathcal{S}$ where $\mathcal{S} = \{1, \ldots, 6\}$.
- Example transition probability matrix

$$P = \begin{bmatrix} 0.9 & .1 & 0 & 0 & 0 & 0 \\ 0 & .9 & .1 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.9 & 0.1 \\ 0.1 & 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}$$

- State transition diagram on board
- Consider $S_n = \sum_{i=1}^{n} X_i$ and $\hat{\mu}_n = \frac{S_n}{n}$. What is $\lim_{n\to\infty} \hat{\mu}_n$?
- Cannot invoke SLLN as $\{X_i\}$ are not i.i.d.
- We will see later SLLN for Markov chains!

Finite dimensional distributions

- Consider a Markov dice with transition probability P.
- What is $P(X_0 = 4, X_1 = 5, X_2 = 6)$?
- $= P(X_2 = 6|X_1 = 5, X_0 = 4)P(X_1 = 5|X_0 = 4)P(X_0 = 4)$
- $= p_{65}p_{54}P(X_0 = 4)$.
- What is $P(X_0 = 4)$?
- This probability of starting in a particular state is called initial distribution of the markov chain.

Finite dimensional distributions

- Consider a DTMC $\{X_n, n \geq 0\}$ with transition matrix P.
- We assume M states and X_0 denotes the initial state.
- You can start in any starting state or may pick your starting state randomly.
- Let $\bar{\mu} = (\mu_1, \ldots, \mu_M)$ denote the initial distribution, i.e., $P(X_0 = x_0) = \mu_{x_0}$.
- How does one obtain the finite dimensional distribution $P(X_0 = x_0, X_1 = x_1, , X_2 = x_2)$?
- $P(X_0 = x_0, X_1 = x_1, , X_2 = x_2) = p_{x_1,x_2} p_{x_0,x_1} \mu_{x_0}$.
- In general, $P(X_0 = x_0, X_1 = x_1, \ldots X_k = x_k) = p_{x_{k-1},x_k} \times \ldots \times p_{x_0,x_1} \mu_{x_0}$

Chapman Kolmogorov Equations for DTMC

- Consider a Markov coin and its transition probability matrix $P = \begin{bmatrix} p_{1,1} & p_{1,-1} \\ p_{-1,1} & p_{-1,-1} \end{bmatrix}$.

- Given $X_0 = 1$, what is $P(X_2 = 1)$?

$$\begin{aligned} P(X_2 = 1|X_0 = 1) &= P(X_2 = 1|X_1 = 1, X_0 = 1)P(X_1 = 1|X_0 = 1) \\ &+ P(X_2 = 1|X_1 = -1, X_0 = 1)P(X_1 = -1|X_0 = 1) \\ &= p_{1,1}^2 + p_{-1,1}p_{1,-1} \end{aligned}$$

- Here the first inequality follow from the fact that

$$P(C|A) = P(C|BA)P(B|A) + P(C|B^cA)P(B^c|A)$$ HW: Verify

- Similarly, $P(X_2 = -1|X_0 = 1), P(X_2 = 1|X_0 = -1), P(X_2 = -1|X_0 = -1)$ can be obtained and these are elements of a two-step transition matrix $P^{(2)}$.

Chapman Kolmogorov Equations for DTMC

- The two step transition probability matrix $P^{(2)}$ is given by
$$P^{(2)} = \begin{bmatrix} p_{1,1}^2 + p_{1,-1}p_{-1,1} & p_{1,1}p_{1,-1} + p_{1,-1}p_{-1,-1} \\ p_{-1,1}p_{1,1} + p_{-1,-1}p_{-1,1} & p_{-1,1}p_{1,-1} + p_{-1,-1}^2 \end{bmatrix}.$$
- This implies that $P^{(2)} = P \times P = P^2$.
- In general, $P^{(n)} = P^n$.
- Chapman-Kolmogorov equations are a further generalization of this.
$$\boxed{P^{(n+l)} = P^{(n)}P^{(l)}}$$
- We wont see the proof of this.

Classification of states

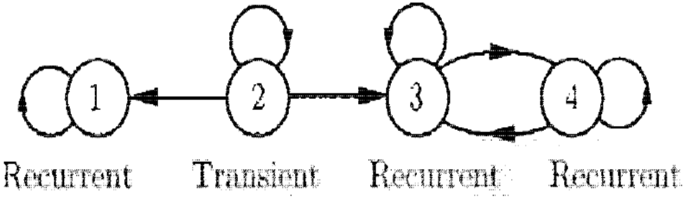
- Consider a Markov process with state space $\mathcal{S}$
- We say that j is accessible from i if $p_{ij}^n > 0$ for some n.
- This is denoted by $i \rightarrow j$.
- if $i \rightarrow j$ and $j \rightarrow i$ then we say that i and j communicate. This is denoted by $i \leftrightarrow j$.

  > A chain is said to be irreducible if $i \leftrightarrow j$ for all $i, j \in \mathcal{S}$.

- Are the examples of Markovian coin and dice we have considered till now irreducible? check!

Recurrent and Transient states

- We say that a state i is recurrent if $F_{ii} = P($ ever returning to i having started in i $) = 1$.
- F_{ii} is not easy to calculate. (Not part of this course)
- If a state is not recurrent, it is transient.
- For a transient state i, $F_{ii} < 1$.
- If $i \leftrightarrow j$ and i is recurrent, then j is recurrent.



Limiting probabilities

- $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.6 & 0.4 \\ 0.6 & 0.4 & 0 \end{bmatrix}$ $P^5 = \begin{bmatrix} .06 & .3 & .64 \\ .18 & .38 & .44 \\ .38 & .44 & .18 \end{bmatrix}$ $P^{30} = \begin{bmatrix} .23 & .385 & .385 \\ .23 & .385 & .385 \\ .23 & .385 & .385 \end{bmatrix}$
- $P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ $\lim_{n\to\infty} P^n = \begin{bmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{bmatrix}$
- What is the interpretation of $\lim_{n\to\infty} p_{ij}^{(n)} = [\lim_{n\to\infty} P^n]_{ij}$?
- $\pi_j = \lim_{n\to\infty} p_{ij}^{(n)}$ denotes the probability of being in state j at time n when starting in state i.
- For an M state DTMC, $\pi = (\pi_1, \ldots, \pi_M)$ denotes the limiting or stationary distribution.
- How do we obtain the stationary distribution π?

Stationary distribution

The **stationary distribution** can be obtained as a solution to the equation $\pi = \pi P$.

- Obtain stationary distribution for the Markov Chain with transition probability $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0.6 & 0.4 \\ 0.6 & 0.4 & 0 \end{bmatrix}$
- The limiting distribution need not exist for some Markov chains, but the stationary distribution exists. For example for $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
- How to tackle such cases? We will see it (among other thing) in CS3.307.