Recap: Modes of convergence

- $X_n \xrightarrow{d} X$ iff

$$F_n(x) \to F(x), \quad \forall x \;\; \text{where } F \text{ is continuous}$$

- $X_n \xrightarrow{P} X$ iff

$$\lim_{n\to\infty} P\left[|X_n - X| > \epsilon\right] = 0, \; \forall \epsilon > 0$$

- $X_n \xrightarrow{r} X$ iff

$$E\left[|X_n - X|^r\right] \to 0 \qquad \text{as} \;\; n \to \infty$$

- $X_n \xrightarrow{a.s} X$ iff

$$P[X_n \to X] = 1 \;\; \text{or} \;\; P[\limsup |X_n - X| > \epsilon] = 0$$

Recap: Relations among different modes

- We have the following relations among different modes of convergence

$$X_n \xrightarrow{r} X \ \Rightarrow \ X_n \xrightarrow{P} X \ \Rightarrow \ X_n \xrightarrow{d} X$$

$$X_n \xrightarrow{a.s.} X \ \Rightarrow \ X_n \xrightarrow{P} X \ \Rightarrow \ X_n \xrightarrow{d} X$$

- All the implications are one-way and we have seen counter examples
- In general, almost sure convergence does not imply convergence in r^{th} mean and vice versa

Recap: Laws of large numbers

- Given X_i iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^{n} X_i$
- Weak law of large numbers: $\frac{S_n}{n} \xrightarrow{P} \mu$
- strong law of large numbers: $\frac{S_n}{n} \xrightarrow{a.s.} \mu$

Recap: Central Limit Theorem

- Given X_i iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^{n} X_i$
- Let $\tilde{S}_n = \frac{S_n - ES_n}{\sqrt{\text{var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$
- **(Lindberg-Levy) Central Limit Theorem**

$$\lim_{n\to\infty} P\left[\tilde{S}_n \le x\right] = \lim_{n\to\infty} P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt, \quad \forall x$$

Recap

- Take X_i iid, $EX_i = 0$, $\text{Var}(X_i) = 1$, $n = 1, 2, \cdots$
- $S_n = \sum_{i=1}^n X_i$
- Strong law of large numbers implies

$$\frac{S_n}{n} \stackrel{a.s.}{\to} 0$$

- Central Limit Theorem implies

$$\frac{S_n}{\sqrt{n}} \stackrel{d}{\to} \mathcal{N}(0, 1)$$

Recap: CLT

- X_i iid, $EX_i = \mu$, $\mathsf{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$
- Central Limit Theorem: $\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$
- It allows us to approximate distributions of sums of independent random variables

$$P[S_n \leq x] \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

- For example, binomial rv is well approximated by normal for large n

- CLT allows one to get rate of convergence of law of large numbers
- Let X_i iid, $EX_i = \mu$, Var$(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$
- By Law of large numbers, $\frac{S_n}{n} \to \mu$.
- Now, by CLT

$$
\begin{aligned}
P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] &= P\left[\left|\frac{S_n - n\mu}{\sigma\sqrt{n}}\right| > \frac{n\epsilon}{\sigma\sqrt{n}}\right] \\
&\approx 1 - \left(\Phi\left(\frac{n\epsilon}{\sigma\sqrt{n}}\right) - \Phi\left(-\frac{n\epsilon}{\sigma\sqrt{n}}\right)\right) \\
&= 2\left(1 - \Phi\left(\frac{n\epsilon}{\sigma\sqrt{n}}\right)\right)
\end{aligned}
$$

(because $\Phi(-x) = (1 - \Phi(x))$)

Confidence intervals

- Let X_i iid, $EX_i = \mu$, $\text{Var}(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$.
- Using CLT, we get

$$P\left[\left|\frac{S_n}{n} - \mu\right| > c\right] \approx 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right)$$

- If the RHS above is δ, then we can say that $\frac{S_n}{n} \in [\mu - c, \ \mu + c]$ with probability $(1 - \delta)$
- This interval is called the $100(1 - \delta)\%$ confidence interval.

$$P\left[\left|\frac{S_n}{n}-\mu\right|>c\right]=2\left(1-\Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right)$$

- Suppose $c=\frac{1.96\sigma}{\sqrt{n}}$
- Then

$$P\left[\left|\frac{S_n}{n}-\mu\right|>\frac{1.96\sigma}{\sqrt{n}}\right]=2\left(1-\Phi\left(1.96\right)\right)=0.05$$

- Denoting $\bar{X}=\frac{S_n}{n}$, the 95% confidence interval is $\left[\bar{X}-\frac{1.96\sigma}{\sqrt{n}},\ \bar{X}+\frac{1.96\sigma}{\sqrt{n}}\right]$
- One generally uses an estimate for σ obtained from X_i
- In analyzing any experimental data the confidence intervals or the variance term is important

central limit theorem

- CLT essentially states that sum of many independent random variables behaves like a Gaussian random variable
- It is very useful in many statistics applications.
- We stated CLT for iid random variables.
- While independence is important, all rv need not have the same distribution.
- Essentially, the variances should not die out.

- We have been considering sequences: $X_n, \ n = 1, 2, \cdots$
- We have so far considered only the asymptotic properties or limits of such sequences.
- Any such sequence is an example of what is called a random process or stochastic process
- Given n rv, they are completely characterized by their joint distribution.
- How doe we specify or characterize an infinite collection of random variables?
- We need the joint distribution of every finite subcollection of them.

Markov Chains

- Let $X_n, \; n = 0, 1, \cdots$ be a sequence of discrete random variables taking values in S.
  Note that S would be countable
- We say it is a Markov chain if $\forall n, x_i$

$$P[X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1} \cdots X_0 = x_0] = P[X_{n+1} = x_{n+1} | X_n = x_n]$$

- We can write it as

$$f_{X_{n+1}|X_n X_{n-1} \cdots X_1} \; = \; f_{X_{n+1}|X_n}$$

- Conditioned on X_n, X_{n+1} is independent of $X_{n-1}, X_{n-2}, \cdots$
- We think of X_n as state at n
- For a Markov chain, given the current state, the future evolution is independent of the history of how you reached the current state

Example

- Let X_i be iid discrete rv taking integer values.
- Let $Y_0 = 0$ and $Y_n = \sum_{i=1}^{n} X_i$
- $Y_n, \ n = 0, 1, \cdots$ is a Markov chain with state space as integers
- Note that $Y_{n+1} = Y_n + X_{n+1}$ and X_{n+1} is independent of $Y_0, \cdots, Y_n$.

$$P[Y_{n+1} = y | Y_n = x, Y_{n-1}, \cdots] = P[X_{n+1} = y - x]$$

- Thus, Y_{n+1} is conditionally independent of $Y_{n-1}, \cdots$ conditioned on Y_n
- Sum of iid random variables is a Markov chain

- In this example, we can think of X_n as the number of people or things arriving at a facility in the n^{th} time interval.
- Then Y_n would be total arrivals till end of n^{th} time interval.
- Number of packets coming into a network switch, number people joining the queue in a bank, can be modeled as Markov chains.
  In such applications, it is called a queing chain.
- Markov chain is a useful model for many dynamic systems or processes

- The Markov property is: given current state, the future evolution is independent of the history of how we came to current state.
- It essentially means the current state contains all needed information about history
- We are considering the case where states as well as time are discrete.
- It can be more general

Transition Probabilities

- Let $\{X_n,\ n = 0, 1, \cdots\}$ be a Markov Chain with (countable) state space S

$$Pr[X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} \cdots X_0] = Pr[X_{n+1} = x_{n+1} | X_n = x_n], \forall x$$

(**Notice the big change of notation**)

- Define function $P : S \times S \to [0,\ 1]$ by

$$P(x, y) = Pr[X_{n+1} = y | X_n = x]$$

- P is called the state transition probability function. It satisfies
  - $P(x, y) \geq 0,\ \forall x, y \in S$
  - $\sum_{y \in S} P(x, y) = 1, \forall x \in S$
- If S is finite then P can be represented as a matrix

- The state transition probability function is given by

$$P(x, y) = Pr[X_{n+1} = y | X_n = x]$$

- In general, this can depend on n though our notation does not show it
- If the value of that probability does not depend on n then the chain is called homogeneous
- For a homogeneous chain we have

$$Pr[X_{n+1} = y | X_n = x] = Pr[X_1 = y | X_0 = x], \ \forall n$$

- In this course we will consider only homogeneous chains

Initial State Probabilities

- Let $\{X_n\}$ be a Markov Chain with state space S
- Define function $\pi_0 : S \to [0,\ 1]$ by

$$\pi_0(x) = Pr[X_0 = x]$$

- It is the pmf of the rv X_0
- Hence it satisfies
  - $\pi_0(x) \geq 0,\ \forall x \in S$
  - $\sum_{x \in S} \pi_0(x) = 1$
- From now on, without loss of generality, we take $S = \{0, 1, \cdots\}$

- Let X_n be a (homogeneous) Markov chain
- Then we have

$$\begin{aligned} Pr[X_0 = x_0, X_1 = x_1] &= Pr[X_1 = x_1 | X_0 = x_0] \; Pr[X_0 = x_0], \forall x_0, x_1 \\ &= P(x_0, x_1)\pi_0(x_0) = \pi_0(x_0)P(x_0, x_1) \end{aligned}$$

- Now we can extend this as

$$\begin{aligned} Pr[X_0 = x_0, X_1 = x_1, X_2 = x_2] &= Pr[X_2 = x_2 | X_1 = x_1, X_0 = x_0] \\ &\qquad Pr[X_0 = x_0, X_1 = x_1] \\ &= Pr[X_2 = x_2 | X_1 = x_1] \\ &\qquad Pr[X_0 = x_0, X_1 = x_1] \\ &= P(x_1, x_2) \; P(x_0, x_1)\pi_0(x_0) \\ &= \pi_0(x_0)P(x_0, x_1)P(x_1, x_2) \end{aligned}$$

- This calculation is easily generalized to any number of time steps

$$
\begin{aligned}
Pr[X_0 = x_0, \cdots X_n = x_n] &= Pr[X_n = x_n | X_{n-1} = x_{n-1}, \cdots X_0 = x_0] \\
&\qquad Pr[X_{n-1} = x_{n-1}, \cdots X_0 = x_0] \\
&= Pr[X_n = x_n | X_{n-1} = x_{n-1}] \\
&\qquad Pr[X_{n-1} = x_{n-1}, \cdots X_0 = x_0] \\
&= P(x_{n-1}, x_n) Pr[X_{n-1} = x_{n-1}, \cdots X_0 = x_0] \\
&= P(x_{n-1}, x_n) Pr[X_{n-1} = x_{n-1} | X_{n-2} = x_{n-2}] \\
&\qquad Pr[X_{n-2} = x_{n-2}, \cdots X_0 = x_0] \\
&\vdots \\
&= \pi_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n)
\end{aligned}
$$

- We showed

$$Pr[X_0 = x_0, \cdots X_n = x_n] = \pi_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n)$$

- This shows that the transition probabilities, P, and initial state probabilities, π_0, completely specify the chain.
- They give us the joint distribution of any finite subcollection of the rv's
- Suppose you want joint distribution of $X_{i_1}, \cdots X_{i_k}$
- Let $m = \max(i_1, \cdots, i_k)$
- We know how to get joint distribution of $X_0, \cdots, X_m$.
- The joint distribution of $X_{i_1}, \cdots X_{i_k}$ is now calculated as a marginal distribution from the joint distribution of $X_0, \cdots, X_m$

Example: 2-state chain

- Let $S = \{0, 1\}$.
- We can write the transition probabilities as a matrix

$$P = \left[\begin{array}{cc} P(0,0) & P(0,1) \\ P(1,0) & P(1,1) \end{array}\right] = \left[\begin{array}{cc} 1-p & p \\ q & 1-q \end{array}\right]$$

- Now we can calculate the joint distribution, e.g., of X_1, X_2 as

$$\begin{aligned} Pr[X_1 = 0, X_2 = 1] &= \sum_{x=0}^{1} Pr[X_0 = x, X_1 = 0, X_2 = 1] \\ &= \sum_{x=0}^{1} \pi_0(x) P(x,0) P(0,1) \\ &= \pi_0(0)(1-p)p + \pi_0(1)qp \end{aligned}$$

- We can similarly calculate probabilities of any events involving these random variables

$$
\begin{aligned}
Pr[X_2 \neq X_0] &= Pr[X_2 = 0, X_0 = 1] + Pr[X_2 = 1, X_0 = 0] \\
&= \sum_{x=0}^{1} (\pi_0(1) P(1, x) P(x, 0) + \pi_0(0) P(0, x) P(x, 1))
\end{aligned}
$$

- We have the formula

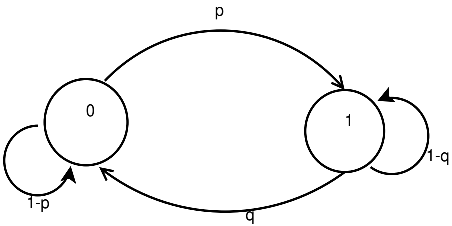
$$
Pr[X_0 = x_0, \cdots X_n = x_n] = \pi_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n)
$$

- This can easily be seen through a graphical notation.

- Consider the 2-state chain with $S = \{0, 1\}$ and

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

- We can represent the chain through a graph as shown below



- The nodes represent states. The edges show possible transitions and the probabilities. We can easily see

$$Pr[X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 0] = \pi_0(0)p(1-q)q$$

- The two state chain can model the 'working/down' state of a machine
- We can ask: what is the probability the machine would be working for the next two days given it is working today

$$
\begin{aligned}
Pr[X_1 = 1, X_2 = 1 | X_0 = 1] &= \frac{Pr[X_0 = 1, X_1 = 1, X_2 = 1]}{Pr[X_0 = 1]} \\
&= \frac{\pi(1) P(1,1) P(1,1)}{\pi(1)}
\end{aligned}
$$

- We may want to know about $\lim_{n \to \infty} Pr[X_n = 1]$

Another example

- A man has 4 umbrellas. carries them from home to office and back when needed. Probability of rain in the morning and evening is same, namely, p.
- What should be the state?

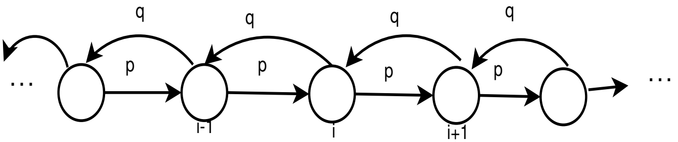
Another example

- A man has 4 umbrellas. carries them from home to office and back when needed. Probability of rain in the morning and evening is same, namely, p.
- What should be the state?
- $S = \{0, 1, \cdots, 5\}$. The transition probabilities are

$$P = \left[\begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1-p & p \\ 2 & 0 & 0 & 1-p & p & 0 \\ 3 & 0 & 1-p & p & 0 & 0 \\ 4 & 1-p & p & 0 & 0 & 0 \end{array}\right]$$

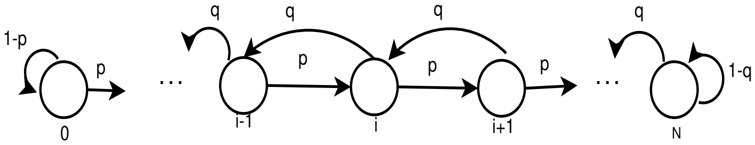
Birth-Death chain

- The following Markov chain is known as a birth-death chain



- e.g., Random walk: $X_i \in \{-1, +1\}$, iid, $S_n = \sum_{i=1}^{n} X_i$
- The queing chain would be a birth-death chain if at most one person can join and one person can leave the queue in any time step.
- In general, the transition probabilities can be different for different states.
- Birth-death chains can also have self-loops on states

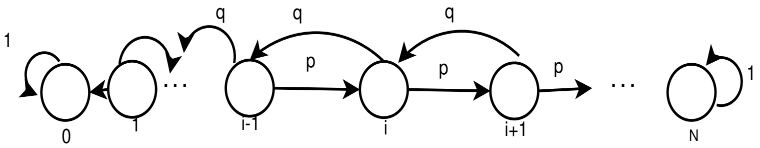
- We can have birth-death chains with finite state space also



- We can say it has ‘reflecting boundaries’
- This chain keeps visiting all the states again and again.

Gambler's Ruin chain

- The following chain is called Gambler's ruin chain



- Here, the chain is ultimately absorbed either in 0 or in N
- Here state can be the current funds that the gambler has

- The transition probabilities we defined earlier are also called one step transition probabilities

$$P(x, y) = Pr[X_{n+1} = y | X_n = x] = Pr[X_1 = y | X_0 = x]$$

- We can define transition probabilities for multiple steps, that is, $Pr[X_n = y | X_0 = x]$
- We first look at one consequence of markov property
- The Markov property implies that it is the most recent past that matters. For example

$$Pr[X_{n+m} = y | X_n = x, X_0] = Pr[X_{n+m} = y | X_n = x]$$

- We consider a simple case

$$
\begin{aligned}
Pr[X_3 = y | X_1 = x, X_0 = z] &= \frac{Pr[X_3 = y, X_1 = x, X_0 = z]}{Pr[X_1 = x, X_0 = z]} \\
&= \frac{\sum_w \pi_0(z) P(z,x) P(x,w) P(w,y)}{\pi_0(z) P(z,x)} \\
&= \sum_w P(x,w) P(w,y)
\end{aligned}
$$

- We also have

$$
\begin{aligned}
Pr[X_3 = y | X_1 = x] &= Pr[X_2 = y | X_0 = x] \\
&= \frac{\sum_w \pi_0(x) P(x,w) P(w,y)}{\pi_0(x)} \\
&= \sum_w P(x,w) P(w,y)
\end{aligned}
$$

- Thus we get

$$
Pr[X_3 = y | X_1 = x, X_0 = z] = Pr[X_3 = y | X_1 = x]
$$

- Using similar algebra, we can show that

$$\begin{aligned} Pr[X_{m+n} = y | X_m = x, X_{m-1} \cdots] &= Pr[X_{m+n} = y | X_m = x] \\ &= Pr[X_n = y | X_0 = x] \end{aligned}$$

- Or, in general,

$$f_{X_{m+n}|X_m,\cdots,X_0}(y|x,\cdots) = f_{X_{m+n}|X_m}(y|x)$$

- Using the same algebra, we can show

$$\begin{aligned} \mathrm{Pr}[X_{m+n} = y | X_m = x,\ X_{m-k} \in A_k,\ k = 1, \cdots, m] &= \\ Pr[X_{m+n} = y | X_m = x] \end{aligned}$$

$$\begin{aligned} &Pr[X_{m+n+r} \in B_r,\ r = 0, \cdots, s \mid X_m = x,\ X_{m-k} \in A_k,\ k = 1, \cdots, m] \\ &\quad = Pr[X_{m+n+r} \in B_r,\ r = 0, \cdots, s \mid X_m = x] \end{aligned}$$

- Now we get

$$
\begin{aligned}
Pr[X_{m+n} = y | X_0 = x] &= \sum_z Pr[X_{m+n} = y, X_m = z | X_0 = x] \\
&= \sum_z Pr[X_{m+n} = y | X_m = z, X_0 = x] \; Pr[X_m = z | X_0 = x] \\
&= \sum_z Pr[X_{m+n} = y | X_m = z] \; Pr[X_m = z | X_0 = x] \\
&= \sum_z Pr[X_n = y | X_0 = z] \; Pr[X_m = z | X_0 = x]
\end{aligned}
$$