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**

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

- 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 is also important to get information on rate of convergence of law of large numbers.

Recap: Markov Chain

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

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

- We can write it as

$$f_{X_{n+1}|X_n, \cdots X_0}(x_{n+1}|x_n, \cdots, x_0) = f_{X_{n+1}|X_n}(x_{n+1}|x_n), \ \forall x_i$$

- For a Markov chain, given the current state, the future evolution is independent of the history of how you reached the current state

Recap: Transition Probabilities

- Let $\{X_n,\ n = 0, 1, \cdots\}$ be a Markov Chain with (countable) state space S
- Transition probability function is $P : S \times S \rightarrow [0,\ 1]$

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

  The chain is said to be homogeneous when this is not a function of time.
- 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

Recap: Initial State Probabilities

- Let $\{X_n\}$ be a Markov Chain with state space S
- Initial state probabilities $\pi_0 : S \rightarrow [0,\ 1]$

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

It satisfies

  - $\pi_0(x) \geq 0,\ \forall x \in S$
  - $\sum_{x \in S} \pi_0(x) = 1$

- The P and π_0 determine all joint distributions

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

- 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 P, and π_0, determine joint distribution of $X_0, \cdots, X_m$ for any m
- 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 $X_0, \cdots, X_m$
- This shows that the transition probabilities, P, and initial state probabilities, π_0, completely specify the chain.

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

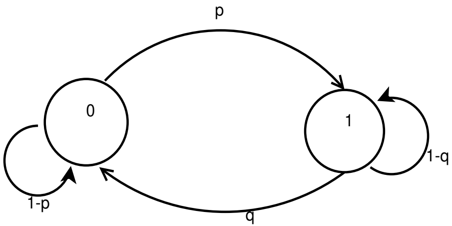
$$Pr[X_0 = x_0, \cdots X_n = x_n] = \pi_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n)$$

- 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

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

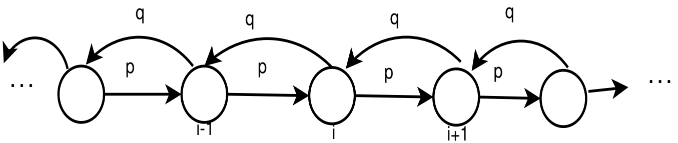
An example

- 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

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

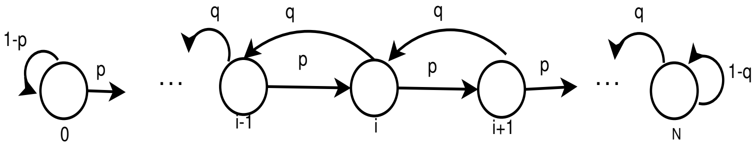
Birth-Death chain

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



- In general, birth-death chains may have self-loops on states
- Random walk: $X_i \in \{-1, +1\}$, iid, $S_n = \sum_{i=1}^{n} X_i$
- We can have 'reflecting boundary' at 0
- Queuing chains can also be birth-death chains

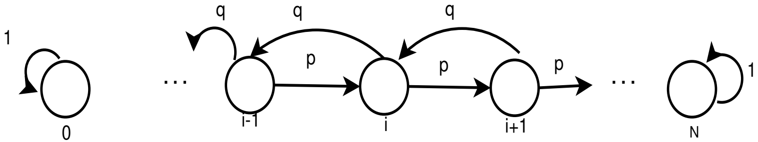
- We can have birth-death chains with finite state space also



- 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

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

- 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

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

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

- Now we get

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

Chapman-Kolmogorov Equations

- Define: $P^n(x, y) = Pr[X_n = y | x_0 = x]$
- These are called n-step transition probabilities.
- From what we showed, n-step transition probabilities satisfy

$$P^{m+n}(x, y) = \sum_z P^m(x, z) P^n(z, y)$$

- These are known as Chapman-Kolmogorov equations
- This relationship is intuitively clear

- Specifically, using Chapman-Kolmogorov equations,

$$P^2(x, y) = \sum_z P(x, z)P(z, y)$$

- For a finite chain, P is a matrix
- Thus $P^2(x, y)$ is the $(x, y)^{th}$ element of the matrix, $P \times P$
- That is why we use P^n for n-step transition probabilities

- Define: $\pi_n(x) = Pr[X_n = x]$.
- Then we get

$$\begin{aligned}\pi_n(y) &= \sum_x Pr[X_n = y | X_0 = x]\ Pr[X_0 = x] \\ &= \sum_x \pi_0(x) P^n(x, y)\end{aligned}$$

- In particular

$$\begin{aligned}\pi_{n+1}(y) &= \sum_x Pr[X_{n+1} = y | X_n = x]\ Pr[X_n = x] \\ &= \sum_x \pi_n(x) P(x, y)\end{aligned}$$

Hitting times

- Let y be a state.
- We define hitting time for y as the random variable

$$T_y = \min\{n > 0 \ : \ X_n = y\}$$

- T_y is the first time that the chain is in state y (after $t = 0$ when the chain is initiated).
- It is easy to see that $Pr[T_y = 1 | X_0 = x] = P(x, y)$.
- We often need conditional probability conditioned on the initial state.
- Notation: $P_z(A) = Pr[A | X_0 = z]$
- We write the above as $P_x(T_y = 1) = P(x, y)$

$$T_y = \min\{n > 0 \ : \ X_n = y\}$$

- We can now get

$$\begin{aligned}
P_x(T_y = 2) &= \sum_{z \neq y} P(x,z)\ P(z,y) = \sum_{z \neq y} P(x,z)\ P_z(T_y = 1) \\
P_x(T_y = m) &= Pr[T_y = m | X_0 = x] \\
&= \sum_{z \neq y} Pr[T_y = m | X_1 = z, X_0 = x]\ Pr[X_1 = z | X_0 = x] \\
&= \sum_{z \neq y} P(x,z) Pr[T_y = m | X_1 = z] \\
&= \sum_{z \neq y} P(x,z) P_z(T_y = m-1)
\end{aligned}$$

- Similarly we can get:

$$P^n(x,y) = \sum_{m=1}^{n} P_x(T_y = m) P^{n-m}(y,y)$$

- We can derive this as

$$\begin{aligned} P^n(x,y) &= Pr[X_n = y | X_0 = x] \\ &= \sum_{m=1}^{n} Pr[T_y = m, X_n = y | X_0 = x] \end{aligned}$$

$$= \sum_{m=1}^{n} Pr[X_n = y | T_y = m, X_0 = x] \; Pr[T_y = m | X_0 = x]$$

$$\begin{aligned} &= \sum_{m=1}^{n} Pr[X_n = y | X_m = y] \; Pr[T_y = m | X_0 = x] \\ &= \sum_{m=1}^{n} P^{n-m}(y,y) \; P_x(T_y = m) \end{aligned}$$

transient and recurrent states

- Define $\rho_{xy} = P_x(T_y < \infty)$.
- It is the probability that starting in x you will visit y
- Note that

$$\rho_{xy} = \lim_{n \to \infty} P_x(T_y < n) = \sum_{n=1}^{\infty} P_x(T_y = n)$$

Definition: A state y is called transient if $\rho_{yy} < 1$; it is called recurrent if $\rho_{yy} = 1$.

- Intuitively, all transient states would be visited only finitely many times while recurrent states are visited infinitely often.
- For any state y define

$$I_y(X_n) = \begin{cases} 1 & \text{if } X_n = y \\ 0 & \text{otherwise} \end{cases}$$

- Now, the total number of visits to y is given by

$$N_y = \sum_{n=1}^{\infty} I_y(X_n)$$

- We can get distribution of N_y as

$$\begin{aligned}
P_x(N_y \geq 1) &= P_x(T_y < \infty) = \rho_{xy} \\
P_x(N_y \geq 2) &= \sum_m P_x(T_y = m) P_y(T_y < \infty) \\
&= \rho_{yy} \sum_m P_x(T_y = m) = \rho_{yy}\ \rho_{xy} \\
P_x(N_y \geq m) &= \rho_{yy}^{m-1}\ \rho_{xy} \\
P_x(N_y = m) &= P_x(N_y \geq m) - P_x(N_y \geq m+1) \\
&= \rho_{yy}^{m-1}\ \rho_{xy} - \rho_{yy}^{m}\ \rho_{xy} = \rho_{xy}\ \rho_{yy}^{m-1}(1 - \rho_{yy}) \\
P_x(N_y = 0) &= 1 - P_x(N_y \geq 1) = 1 - \rho_{xy}
\end{aligned}$$

- Notation: $E_x[Z] = E[Z|X_0 = x]$
- Define

$$\begin{aligned} G(x,y) &\triangleq E_x[N_y] \\ &= E_x\left[\sum_{n=1}^{\infty} I_y(X_n)\right] \\ &= \sum_{n=1}^{\infty} E_x\left[I_y(X_n)\right] \\ &= \sum_{n=1}^{\infty} P^n(x,y) \end{aligned}$$

- $G(x,y)$ is the expected number of visits to y for a chain that is started in x.

Theorem:

(i). Let y be transient. Then

$$P_x(N_y < \infty) = 1, \ \forall x \ \text{ and } \ G(x,y) = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty, \ \forall x$$

(ii) Let y be recurrent. Then

$$P_y[N_y = \infty] = 1, \ \text{ and } \ G(y,y) = E_y[N_y] = \infty$$

$$P_x[N_y = \infty] = \rho_{xy}, \ \text{ and } \ G(x,y) = \begin{cases} 0 & \text{if } \rho_{xy} = 0 \\ \infty & \text{if } \rho_{xy} > 0 \end{cases}$$