Definitions and Theorems

Sunday, October 7, 2018 10:30 PM

Markov Chain

- Markov Property
  - $\mathbb{P}(X_{l+1} = x_{l+1} | X_0 = x_0, \dots, X_l = x_l) = \mathbb{P}(X_{l+1} = x_{l+1} | X_l = x_l)$
- Chapman-Kolmogorov Equation
  - $p^{m+n}(i,j) = \sum_{l \in S} p^m(i,l) p^n(l,j)$
- Stopping Time
  - $\{T = n\}$ can be expressed using the variables $X_0, X_1, \dots, X_n$
- Strong Markov Property
  - $\mathbb{P}(X_{T+1} = j | X_T = i, T = n) = p(i,j)$
- Return Time/Probability
  - $T_y = \min\{n \geq 1 | X_n = y\}$ is the time of first return
  - $T_y^k = \min\{n > T_y^{k-1} | X_n = y\}$ is the time of k-th return
  - $\rho_{xy}^k = \mathbb{P}_x(T_y^k < \infty)$ is the probability of reaching y from x for k times
- Number of Visits
  - $N(y)$: Number of visits to y after time 0
  - $N_n(y)$:Number of visits to y up to time n
- Initial Distribution
  - $\mathbb{P}_x(A) = (A | X_0 = x)$
  - $\mathbb{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mu(x_0) \prod_{l=0}^{n-1} p(x_l, x_{l+1})$
- Transient and Recurrent
  - y is transient $\Leftrightarrow \rho_{yy} = \mathbb{P}_y(T_y < \infty) < 1 \Leftrightarrow 1 - \rho_{yy} = \mathbb{P}_y(T_y = \infty) > 0$
  - y is recurrent $\Leftrightarrow \rho_{yy} = \mathbb{P}_y(T_y < \infty) = 1 \Leftrightarrow 1 - \rho_{yy} = \mathbb{P}_y(T_y = \infty) = 0$
- Communication: $x \Rightarrow y$ iff $p^n(x,y) > 0$ for some $n \geq 0$
- Closed (impossible to get out of): If $i \in C$ and $p(i,j) > 0$, then $j \in C$
- Irreducible (freely moved about): $i \Leftrightarrow j, \forall i, j \in C$
- $\text{Tail} - \text{Sum Formula: } \mathbb{E}N = \sum_{k=1}^{\infty} \mathbb{P}(N \geq k)$
- Theorems Related to Recurrence
  - $\mathbb{E}_x N(y) = \dfrac{\rho_{xy}}{1 - \rho_{yy}}$

- $\mathbb{E}_x N(y) = \sum_{n=1}^{\infty} p^n(x,y)$
- y is recurrent $\Leftrightarrow \sum_{n=1}^{\infty} p^n(y,y) = E_y N(y) = +\infty$
- If $x \Rightarrow y$ and $y \Rightarrow z$, then $x \Rightarrow z$
- If $x \Rightarrow y$ and $\rho_{yx} < 1$, then x is transient
- If x is recurrent and $x \Rightarrow y$, then $\rho_{yx} = 1$
- If x is recurrent and $x \Rightarrow y$, then y is recurrent
- In a finite closed set of states, there is at least one recurrent state
- Finite, Closed, Irreducible $\Rightarrow$ Recurrent
- $|S| < \infty \Rightarrow S = T \cup R_1 \cup \cdots \cup R_k$ for T, R_i disjoint, R_i irreducible

* Stationary Distribution/Measure
  - μ is a stationary measure $\Leftrightarrow \mu = \mu\mathcal{P} \Leftrightarrow \mu(j) = \sum_{i \in S} \mu(i) p(i,j)$
  - π is a stationary distribution $\Leftrightarrow \pi$ is a stationary measure and $\sum_{j \in S} \pi(j) = 1$
  - Normalize μ to get π: $\pi(k) = \frac{\mu(k)}{\sum_{l \in S} \mu(l)}$
* Positive vs Null Recurrent
  - x is positive recurrent if $\mathbb{E}_x T_x < \infty$
  - x is null recurrent if $\mathbb{E}_x T_x = \infty$
* Convergence Theorem
  - If a MC is irreducible, aperiodic, and π exists, then $\lim_{n \to \infty} p^n(x,y) = \pi(y)$
* Asymptotic Frequency
  - If a MC is irreducible and recurrent, then $\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y} \overset{\text{if exists}}{=} \pi(y)$
* Law of Large Numbers for MC
  - Suppose a MC is irreducible and π exists
  - If $\sum_{x \in S} |f(x)| \pi(x) < \infty$, then $\frac{1}{n} \sum_{l=1}^{n} f(X_l) \to \sum_{x \in S} f(x)\pi(x) = \mathbb{E}_\pi f(x_0)$
* Doubly Stochastic
  - A stochastic matrix is doubly stochastic if its column sum to 1 $i.e. \sum_{x \in S} p(x,y) = 1$
  - $\pi(x) = \frac{1}{N}, \forall x \in S$ is a stationary distribution $\Leftrightarrow$ the MC is doubly stochastic
* Detailed Balance
  - $\pi(x)p(x,y) = \pi(y)p(y,x), \forall x, y \in S$

- All distributions satisfying the detailed balance equations are stationary
- All random walks' graphs satisfy DBE's

- Exit Distribution
  - $\begin{cases} h(a) = 1, h(b) = 0 \\ h(x) = \sum_{y \in S} p(x,y)h(y), \forall x \in C \coloneqq S \setminus \{a,b\} \end{cases} \Rightarrow h(x) = \mathbb{P}_x(V_a < V_b)$
- Exit Time
  - Define $V_A \coloneqq \inf\{n \geq 0 | X_n \in A\}$ and $C \coloneqq S \setminus A$. Suppose $\mathbb{P}_x(V_A < \infty) > 0, \forall x \in C$
  - $\begin{cases} g(a) = 0, \forall a \in A \\ g(x) = 1 + \sum_{y \in C} g(y)p(x,y) \end{cases} \Rightarrow g(x) = \mathbb{E}_x[V_A]$

Poisson Process

- Exponential Distribution
  - $X \sim \text{Exp}(\lambda) \Leftrightarrow f_X(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \Leftrightarrow F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
  - $\mathbb{E}[X] = \frac{1}{\lambda}, \text{Var}[X] = \frac{1}{\lambda^2}$
  - $\mathbb{P}(X > s + t | X > s) = \mathbb{P}(X > t)$
- Gamma Distribution
  - $T \sim \text{Gamma}(n, \lambda) \Leftrightarrow T = \text{Sum of } n \text{ Exp}(\lambda) \Leftrightarrow f_T(t) = \begin{cases} \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{n-1}}{(n-1)!} & t \geq 0 \\ 0 & t < 0 \end{cases}$
  - $\mathbb{E}[T] = \frac{n}{\lambda}, \text{Var}[T] = \frac{n}{\lambda^2}$
- Poisson Distribution
  - $X \sim \text{Poisson}(\lambda) \Leftrightarrow p_X(n) = e^{-\lambda}\frac{\lambda^n}{n!} \Rightarrow \mathbb{E}[X] = \text{Var}[X] = \lambda$
- Poisson Process
  - Interarrival time: $\tau_k \overset{iid}{\sim} \text{Exp}(\lambda)$
  - Arrival time: $T_n = \tau_1 + \cdots + \tau_n \sim \text{Gamma}(n, \lambda)$
  - Number of arrivals up to time s: $N(s) \sim \text{Poisson}(\lambda s)$
- Equivalent Definition of Poisson Process
  - $N(0) = 0$ (with probability 1)
  - $N(t + s) - N(s) \sim \text{Poisson}(\lambda t)$
  - $N(t)$ has independent increments
- Compound Poisson Process
  - $S(t) = Y_1 + Y_2 + \cdots + Y_{N(t)} = \sum_{k=1}^{N(t)} Y_k$

- $S(t) = 0$ when $N(t) = 0$
- Mean and Variance of Random Sum
  - $E[S] = E[N]E[Y_1]$
  - $\text{Var}[S] = \mathbb{E}[N]\text{Var}[Y_1] + \text{Var}[N](\mathbb{E}[Y_1])^2$
- Mean and Variance of Compound Poisson Process
  - $\text{Var}(S) = \lambda\mathbb{E}[Y_1^2]$
  - $\mathbb{E}[S(t)] = \lambda t\mathbb{E}[Y_1]$
  - $\text{Var}[S(t)] = \lambda t\mathbb{E}[Y_1^2]$
- Thinning a Poisson Process
  - Define $N_j(t) = \sum_{k=1}^{N(t)} \mathbb{1}\{Y_k = j\}$ be the number of arrivales up to time t of type j
  - Then $N_1(t), N_2(t), \ldots$ are independent Poisson process with rate $\lambda_j = \lambda\mathbb{P}(Y_1 = j)$
- Superposition of Poisson Processes
  - Suppose $N_1(t), \ldots, N_k(t)$ are independent Poisson process with rates $\lambda_1, \ldots, \lambda_k$
  - Then $N(t) = N_1(t) + \cdots + N_k(t)$ is a Poisson process with rate $\lambda = \lambda_1 + \cdots + \lambda_k$
- Conditioning of Poisson Processes
  - $(T_1, \ldots, T_n | N(t) = n) \stackrel{D}{=} \left(U_{(1)}, \ldots, U_{(n)}\right)$
  - $f(t_1, \ldots, t_n) = \begin{cases} \frac{n!}{t^n} & 0 \le t_1 \le \cdots \le t_n \le t \\ 0 & \text{otherwise} \end{cases}$
- Binomial and Conditioning of Poisson Processes
  - $\mathbb{P}(N(s) = k | N(t) = n) = \binom{n}{k}\left(\frac{s}{t}\right)^k\left(1 - \frac{s}{t}\right)^{n-k}$ for $s < t$ and $0 \le k \le n$

Renewal Process

- Renewal process: Like a Poisson process, but waiting time t_k do not have to be $\text{Exp}(\lambda)$
- Arrival LLN: $\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$, where $\mu = \mathbb{E}[t_i]$
- Reward LLN
  - Let r_i = reward/cost of i-th renewal, and $R(t) = \sum_{i=1}^{N(t)} r_i$, then, $\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]}$
- Alternating LLN
  - Let $s_1, s_2, \ldots$ be the times in state 1, and $u_1, u_2, \ldots$ be times in state 2
  - Then the limiting fraction of time spent in state 1 is $\frac{\mathbb{E}[s_i]}{\mathbb{E}[s_i] + \mathbb{E}[u_i]}$
- Age and Residual Life
  - $A(t)$ = age = time since last renewal = $t - T_{N(t)}$
  - $Z(t)$ = residual life = time until next renewal = $T_{N(t)+1} - t$

- $\lim_{t\to\infty} \mathbb{P}(A(t) > x, Z(t) > y) = \frac{1}{\mathbb{E}[t_i]} \int_{x+y}^{\infty} \mathbb{P}(t_i > z) dz$
- Limiting PDF of $Z(t)$ is $g(z) = \frac{\mathbb{P}(t_i > z)}{\mathbb{E}[t_i]}$ for $z \geq 0$, and same for $A(t)$
- Limiting expected value of $A(t)$ and $Z(t)$ is $\frac{\mathbb{E}[t_i^2]}{2\mathbb{E}[t_i]}$
- If $t_k \sim f$ then the limiting joint PDF of $A(t)$ and $Z(t)$ is $\frac{f(a+z)}{\mathbb{E}[t_1]}$

Continuous Time Markov Processes

- Markov Property
  - For any time $0 \leq s_0 < \cdots < s_n < s$, and any states $j, i, i_n, \dots, i_0$, we have
  - $\mathbb{P}\left(X_{s+t} = j \middle| X_s = i, X_{s_n} = i_n, \dots, X_{s_0} = i_0\right) = \mathbb{P}(X_{s+t} = j | X_s = i) = \mathbb{P}(X_t = j | X_0 = i)$
- Chapman–Kolmogorov Equation
  - $p_{s+t}(i,j) = \sum_{k \in S} p_s(i,k) p_t(k,j)$
- Jump Rates: For any states $i \neq j, q_{ij} := \lim_{h\to 0} \frac{p_h(i,j)}{h}$
- Kolmogorov Equations
  - Define $\lambda_i = \sum_{k \neq i} q_{ik}$ to be the rate out of state i
  - Define $Q_{ij} = \begin{cases} q_{ij} & \text{if } i \neq j \\ -\lambda_i & \text{if } i = j \end{cases} \Leftrightarrow Q = \begin{bmatrix} -\lambda_1 & q(1,2) & q(1,3) & \cdots \\ q(2,1) & -\lambda_2 & q(2,3) & \cdots \\ q(3,1) & q(3,2) & -\lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$
  - Backward: $\frac{d}{dt}[p_t(i,j)] = \sum_{k \neq i} q(i,k) p_t(k,j) - \lambda_i p_t(i,j) \Leftrightarrow \frac{d}{dt}[p_t] = Q p_t$
  - Forward: $\frac{d}{dt}[p_t(i,j)] = \sum_{k \neq i} p_t(i,k) q(k,j) - p_t(i,j)\lambda_j \Leftrightarrow \frac{d}{dt}[p_t] = p_t Q$
- Stationary Distributions
  - $\mathbb{P}_\pi(X(t) = j) = \pi(j), \forall t > 0, j \in S \Leftrightarrow \pi p_t = \pi$
  - π is stationary if and only if $\pi Q = 0$
- Irreducibility
  - A CTMC $X(t)$ is irreducible if for any $i, j \in S$, there exists states $k_1, \dots, k_{n-1}$ s.t.
  - $q(i,k_1) q(k_1,k_2) \cdots q(k_{n-1},j) > 0$ *i.e.* "It is possible to go from i to j"
- Convergence Theorem
  - If $X(t)$ is a CTMC s.t. $X(t)$ is irreducible, and has a stationary distribution
  - Then, $\lim_{t\to\infty} p_t(i,j) = \pi(j), \forall i, j \in S$
- Detailed Balance
  - $\pi(i) q(i,j) = \pi(j) q(j,i), \forall j \neq i$

Review, Introduction to Stochastic Processes

Thursday, September 6, 2018 9:31 AM

Probability Space

- **Sample space**, Ω: set of all elementary outcomes in a random experiment
- **Events**, $\mathcal{F}$: set of subsets of the sample space
- **Probability measure** $\mathbb{P}$: function on the events that assigns probabilities to them
- $(\Omega, \mathcal{F}, \mathbb{P})$ form a probability space

Axioms of Probability Measure

1. For any event $A \in \mathcal{F}$, we must have $0 \leq \mathbb{P}(A) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. Countable additivity of $\mathbb{P}$

For disjoint events $A_1, A_2, A_3 \ldots, \mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} A_j$

Properties of Probability Measure

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Random Variables

- Definitions
  - A **random variable** X is a **function with domain $\boldsymbol{\Omega}$ and codomain $\boldsymbol{R}$**
  - A **discrete** RV is a RV where range is a **finite** set, or a **countably infinite** set
- Classic examples: Bernoulli, Binomial, Geometric

What is Stochastic Processes

- A **collection of random variables organized by an index set**
- More formally, $\{X(t) | t \in \mathcal{L}\}$ is a stochastic process, and $\mathcal{L}$ the index set
- We often classify and study the stochastic processes by properties of the index set

Common Choices for the Index Set

1. $\mathbb{Z}_{\geq 0} = \mathbb{N} = \{0,1,2,3 \ldots\}$
   - This gives us a sequence of RVs called **discrete time stochastic process**
   - Example: Pick a stock. Check its price each morning.
   - Usual notation: $X(t) = X_t$, often use n instead of t
2. $\mathbb{R}_{\geq 0} = [0, +\infty)$
   - This is called a **continuous time stochastic process**

- Example: Suppose you want to check the stock's price at **ANY** time.
- Notation: $X(t) = X_t$

3. $\mathcal{L}$ is a set of subsets of some larger universe U
    - Sometimes called a **point process**
    - Example
        - $U =$ All stocks on S&P500
        - $\mathcal{L} =$ Powerset of U (All subsets of U)
        - For all $A \in \mathcal{L}$, $X(A) =$ #Stocks in A that increase in value over 2018

State Space

- Definition
    - **The set of values of RVs can take** is called the **state space**, denoted by S
- Example
    - Suppose you are playing Monopoly
    - $X_n =$ Your position on Monopoly board after n rounds of play
    - This is a DTSP with $S = \{\text{All positions on the board}\}$

Basic Question for DTSPs

- What is the value of $\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)$ for any $x_0, x_1, \dots, x_n$?
- Idea: apply the chain rule / multiplication rule for conditional probability
- Conditional probability
    - $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(AB)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(AB) = \mathbb{P}(B)\mathbb{P}(A|B)$
- Generalized conditional probability
    - $\mathbb{P}(E_1E_2 \cdots E_n) = \mathbb{P}(E_1)\displaystyle\prod_{l=1}^{n-1}\mathbb{P}(E_{l+1}|E_1 \cdots E_l)$
- Formula for DTSPs in general
    - $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0)\displaystyle\prod_{l=0}^{n-1}\mathbb{P}(X_{l+1} = x_{l+1}|X_0 = x_0, \dots, X_l = x_l)$

Introduction to Markov Chain

Tuesday, September 11, 2018 9:21 AM

Markov Chain

- Markov assumption
    - Your next step **only depends on where you are**, not where you've been
- **Markov property**
    - $\mathbb{P}(\boldsymbol{X_{n+1}} = \boldsymbol{x_{n+1}} | \boldsymbol{X_0} = \boldsymbol{x_0}, \dots, \boldsymbol{X_n} = \boldsymbol{x_n}) = \mathbb{P}(\boldsymbol{X_{n+1}} = \boldsymbol{x_{n+1}} | \boldsymbol{X_n} = \boldsymbol{x_n}), \forall \boldsymbol{x_i}$
- Further assumption in this course: temporally homogeneous
    - $\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_{m+1} = j | X_m = i), \forall m, n$
- Transition probability
    - Since the subscript doesn't matter, we will use

      $$p(i, j) \coloneqq \mathbb{P}(X_{n+1} = j | X_n = i)$$

      to denote the transition probability from state i to state j
    - Therefore, for Markov chain
        - $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0) \prod_{l=0}^{n-1} p(x_l, x_{l+1})$

Initial distribution

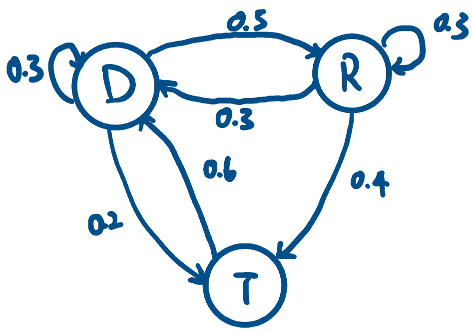
- If we know the **exact starting position** from the MC
    - Then $\mathbb{P}(X_0 = i) = 1$, for some $i \in S$
    - We may write $\mathbb{P}_i(X_n = j) \coloneqq \mathbb{P}(X_n = j | X_0 = i)$
- If the **starting position is random**
    - We need to assign an initial distribution/measure on S
    - Our usual notion for the initial distribution is μ
    - $\mu(i) \coloneqq \mathbb{P}(X_0 = i)$, where $\begin{cases} 0 \le \mu(i) \le 1 \\ \sum_{i \in S} \mu(i) = 1 \end{cases}$
    - We may write $\mathbb{P}_\mu(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \coloneqq \mu(x_0) \prod_{l=0}^{n-1} p(x_l, x_{l+1})$

Example: Highly Simplified Voter Model

- We randomly choose a US voter
- Start with 2012 ($n = 0$), then 2016 ($n = 1$), 2020 ($n = 2$), and so on
- In 2012, voters were split by D: 51%, R: 46%, T: 2%
- From one election to the next,
    - D votes D, R, T with probability 0.3, 0.5, 0.2

- R votes D, R, T with probability 0.3, 0.3, 0.4
- T votes D, R, T with probability 0.6, 0, 0.4

- What is the initial distribution for this model?
  - Coordinate form: $\mu(D) = 0.51, \mu(R) = 0.46, \mu(T) = 0.02$
  - Vector form: $\mu = [0.51 \quad 0.47 \quad 0.02]$
- How can we visualize this MC?



- How should we organize the transition probability?
  - $\mathcal{P} = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.3 & 0.3 & 0.4 \\ 0.6 & 0 & 0.4 \end{bmatrix}$
  - $\mathcal{P}$ is called the **transition matrix** for the MC
  - Note: Rows sums to 1, columns do not have to sum to 1
- What is the probability that someone who votes R in 2012 votes T in 2016 and D in 2020?
  - $\mathbb{P}_R(X_1 = T, X_2 = D) = p(R,T) \cdot p(T,D) = 0.4 \times 0.6 = 0.24$
- What is the probability a 2012 R voter will vote D in 2020?
  - $\mathbb{P}_R(X_2 = D) = \sum_{s \in S} \mathbb{P}_R(X_1 = s, X_2 = D)$

$$= \mathbb{P}_R(X_1 = D, X_2 = D) + \mathbb{P}_R(X_1 = R, X_2 = D) + \mathbb{P}_R(X_1 = T, X_2 = D)$$
$$= p(R,D) \cdot p(D,D) + p(R,R) \cdot p(R,D) + p(R,T) \cdot p(T,D)$$
$$= 0.3 \times 0.3 + 0.3 \times 0.3 + 0.4 \times 0.6$$
$$= 0.42$$

Simple Random Walk, $\mathcal{P}^n$, Gambler's Ruin

Thursday, September 13, 2018 10:13 AM

Example: Simple Random Walk

- Let $\{Y_n\}_{n\geq 1}$ be iid with distribution $Y_n = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } 1-p=q \end{cases}$
- Let $\{X_n\}_{n\geq 0}$ be defined as $X_n = \begin{cases} 0 & \text{for } n=0 \\ \sum_{i=1}^n Y_i & \text{for } n\geq 1 \end{cases}$
- Question: Is $X_0, X_1, X_2, \ldots$ a Markov chain?
- We need to check whether the **Markov property** is satisfied
    - $\mathbb{P}(X_{n+1}=x_{n+1}|X_0=x_0,\ldots,X_n=x_n)=\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n)$
- Compute $\mathbb{P}(X_{n+1}=j|X_0=x_0,\ldots,X_n=i)$
    - $\mathbb{P}(X_{n+1}=j|X_0=x_0,\ldots,X_{n-1}=x_{n-1},X_n=i)$

$$= \frac{\mathbb{P}(X_0=x_0,\ldots,X_{n-1}=x_{n-1},X_n=i,X_{n+1}=j)}{\mathbb{P}(X_0=x_0,\ldots,X_{n-1}=x_{n-1},X_n=i)}, \text{by Bayes' law}$$

$$= \frac{\mathbb{P}(X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=i,X_{n+1}=j)}{\mathbb{P}(X_1=x_1,\ldots,X_{n-1}=x_{n-1},X_n=i)}, \text{since } X_0=0$$

$$= \frac{\mathbb{P}(Y_1=x_1,Y_2=x_2-x_1\ldots,Y_n=i-x_{n-1},Y_{n+1}=j-i)}{\mathbb{P}(Y_1=x_1,Y_2=x_2-x_1\ldots,Y_n=i-x_{n-1})}, \text{since } Y_{i+1}=X_{i+1}-X_i$$

$$= \frac{\mathbb{P}(Y_1=x_1)\mathbb{P}(Y_2=x_2-x_1)\cdots\mathbb{P}(Y_n=i-x_{n-1})\mathbb{P}(Y_{n+1}=j-i)}{\mathbb{P}(Y_1=x_1)\mathbb{P}(Y_2=x_2-x_1)\cdots\mathbb{P}(Y_n=i-x_{n-1})}$$

$$= \mathbb{P}(Y_{n+1}=j-i)$$

- Compute $\mathbb{P}(X_{n+1}=j|X_n=i)$
    - $\mathbb{P}(X_{n+1}=j|X_n=i)=\dfrac{\mathbb{P}(X_n=i,X_{n+1}=j)}{\mathbb{P}(X_n=i)}$

$$= \frac{\mathbb{P}(X_n=i,Y_{n+1}=j-i)}{\mathbb{P}(X_n=i)}, \text{since } X_{n+1}=X_n+Y_{n+1} \Leftrightarrow Y_{n+1}=X_{n+1}-X_n$$

$$= \frac{\mathbb{P}(X_n=i)\mathbb{P}(Y_{n+1}=j-i)}{\mathbb{P}(X_n=i)}, \text{since } X_n=Y_1+\cdots+Y_n \text{ is independent with } Y_{n+1}$$

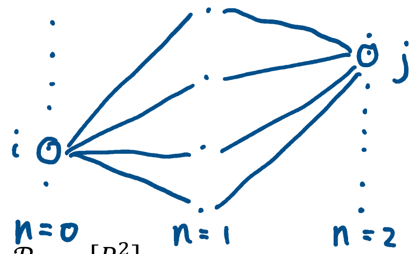
$$= \mathbb{P}(Y_{n+1}=j-i)$$

- Therefore $X_0, X_1, X_2, \ldots$ is a Markov chain

n-Step Transition Probabilities

- Motivation
    - Compute $\mathbb{P}(X_n=j|X_0=i)$, given the transition probabilities $p(l,k)$ for the MC
- Statement
    - Let $\mathcal{P}_{lk}=p(l,k)$ be the probability transition matrix, then $\mathbb{P}(\boldsymbol{X_n=j|X_0=i})=[\boldsymbol{\mathcal{P}^n}]_{\boldsymbol{ij}}$
- Proof
    - For $n=1$: True by definition of $\mathcal{P}$

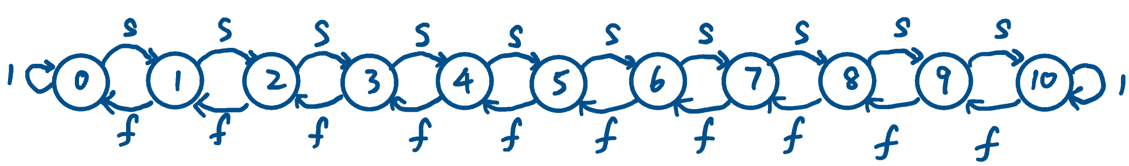
- For $n = 2$
  - $\mathbb{P}(X_2 = j | X_0 = i) = \sum_{l \in S} \mathbb{P}(X_2 = j, X_1 = l | X_0 = i)$
  - $= \sum_{l \in S} \frac{\mathbb{P}(X_2 = j, X_1 = l, X_0 = i)}{\mathbb{P}(X_0 = i)}$
  - $= \sum_{l \in S} \frac{\mu(i)p(i,l)p(l,j)}{\mu(i)} = \sum_{l \in S} p(i,l)p(l,j) = \sum_{l \in S} \mathcal{P}_{il}\mathcal{P}_{lj} = [P^2]_{ij}$



- The general case is proven via strong mathematical induction
- Corollary: **Chapman-Kolmogorov Equation**
  - $\boldsymbol{p^{m+n}(i,j) = \sum_{l \in S} p^m(i,l)p^n(l,j)}$
- Proof for Corollary
  - $p^{m+n}(i,j) = [\mathcal{P}^{m+n}]_{ij} = \sum_{l \in S} [\mathcal{P}^m]_{il}[\mathcal{P}^n]_{lj} = \sum_{l \in S} p^m(i,l)p^n(l,j)$

Example: Gambler's Ruin

- Background
  - You have \$7. You need \$10. There is a casino game where you either win or lose \$1.
  - The win probability is 0.45. You play the game until you have lost or met your goal.
- Model this problem with a Markov chain
  - $s = 0.45, f = 0.55, \mu(7) = 1$



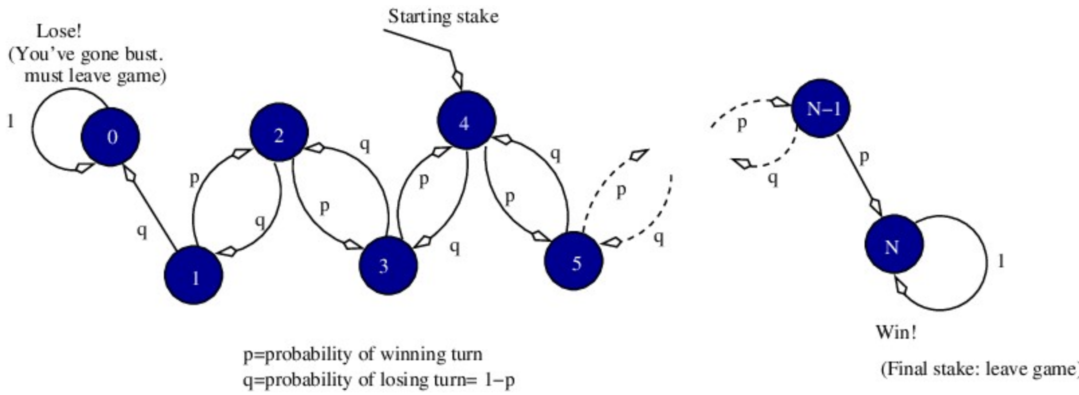
- Find the probability that you have met your goal by the 10th round
  - $\mathcal{P} = \begin{bmatrix} 1 & 0 & & & & & & & & & \\ f & 0 & s & & & & & & & & \\ & f & 0 & s & & & & & & & \\ & & f & 0 & s & & & & & & \\ & & & f & 0 & s & & & & & \\ & & & & f & 0 & s & & & & \\ & & & & & f & 0 & s & & & \\ & & & & & & f & 0 & s & & \\ & & & & & & & f & 0 & s & \\ & & & & & & & & f & 0 & s \\ & & & & & & & & & 0 & 1 \end{bmatrix}$
  - $\mathbb{P}(X_{10} = 10 | X_0 = 7) = p^{10}(7,10) = [\mathcal{P}^{10}]_{8,11} \approx 0.248$, note that the index starts with 1
- Find the probability you lost it all by round 10
  - $p^{10}(7,0) = [\mathcal{P}^{10}]_{8,1} \approx 0.042$

T, Strong Markov Property, T_y, ρ_{yy}, Recurrence

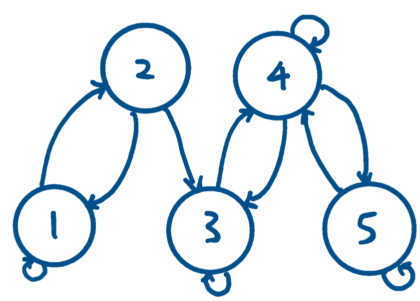
Tuesday, September 18, 2018 9:33 AM

Stopping Time

- Motivation
    - In the setting of Gambler's ruin



    - $T =$ The first time you have \$$N$
    - We can think of **stopping time** as a **criteria to quit** running the Markov chain
- Definition
    - Let T be a random variable taking values in $\{0,1,2,\dots,\infty\}$
    - T is a **stopping time** for a Markov chain $X_0, X_1, \dots$ if
        - The event $\{T = n\}$ can be **expressed using the variables $X_0, X_1, \dots, X_n$**
        - *i.e.* You can tell if you stop at time n based on the states of the MC through time n
- Example: Determine if the following RVs are stopping times



    - $T = \min\{n \geq 1 | X_n = 5\} =$ time of first visit to state 5
        - $\{T = n\} = \{X_n = 5, X_{n-1} \neq 5, \dots, X_1 \neq 5\}$
        - Therefore T is a stopping time
        - Note: We **do not include X_0**, since $n \geq 1$
    - $T = \max\{n \geq 1 | X_n = 2\} =$ time of final visit to state 2
        - $\{T = n\} \overset{a.s.}{=} \{X_n = 2, X_{n+1} = 3\}$
        - T is **not a stopping time**, since we need to know **$\{X_{n+1} = 3\}$ in the future**
    - $T =$ Time of the third visit to state 2
        - $\{T = n\} = \left\{X_n = 2, \left(\sum_{k=1}^{n-1} \mathbb{1}\{X_k = 2\}\right) = 2\right\}$, where $\mathbb{1}$ is a indicator function
        - Since $\{T = n\}$ could be expressed using $X_0, \dots, X_n$, it is a stopping time
    - $T =$ Time of final visit to state 2 after visiting state 5
        - $\{T = n\} = \emptyset$ for $n \neq 0$
        - So T is a stopping time for the MC

Strong Markov Property

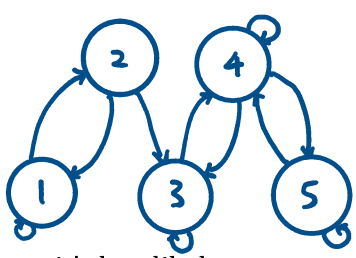
- Definition
  - Let T be a stopping time for the Markov chain $X_0, X_1, \dots$
  - Given that $T = n$ and $X_T = y$. Then
  - Any other information about $X_0, \dots, X_r$ is **irrelevant for future predictions**
  - And X_{T+k} $(k \neq 0)$ **behaves like a Markov chain** with initial state y
- Justification
  - Durret proves $\mathbb{P}(\boldsymbol{X_{T+1} = j | X_T = i, T = n}) = \boldsymbol{p(i, j)}$
- Why stopping times? Why no any random variables?
  - Suppose $T_y = \min\{n \geq 0 | X_{n+1} = y\}$
  - T_y is not a stopping time, since $\{T_y = n\} = \{X_{n+1} = y\}$
  - $\mathbb{P}\left(X_{T_y+1} = j \middle| X_{T_y} = i, T_y = n\right) = \begin{cases} 1 & \text{if } j = y \\ 0 & \text{if } j \neq y \end{cases}$

Return Time and Return Probability

- $\boldsymbol{T_y} = \mathbf{min}\{\boldsymbol{n \geq 1 | X_n = y}\}$ is called the **hitting time** of y or **time of first return** to y
- $\boldsymbol{\rho_{yy}} = \mathbb{P}_{\boldsymbol{y}}(\boldsymbol{T_y} < \infty)$ is called the **return probability**
- $\boldsymbol{T_y^k} = \mathbf{min}\{\boldsymbol{n \geq T_y^{k-1} | X_n = y}\}$ is called the **time of k-th return**
- $\boldsymbol{\rho_{yy}^k} = \mathbb{P}_{\boldsymbol{y}}(\boldsymbol{T_y^k} < \infty)$ is called the **k-th return probability**
  - Proof: Use strong Markov property and mathematical induction
- Note: k is label on T_y^k, but exponent on ρ_{yy}^k

Recurrent and Transient States

- Motivation



  - In the example above, it's less likely to return to state 1 and 2 as the time increase
  - While for state 3, 4 and 5, the chain returns to those states for infinitely many times
- Definition
  - If $\rho_{yy} < 1$, we say y is **transient** (not guaranteed to keep returning to y)
  - If $\rho_{yy} = 1$, we say y is **recurrent** (guaranteed to return to y forever)

Recurrence, Closed, Irreducible, Communication

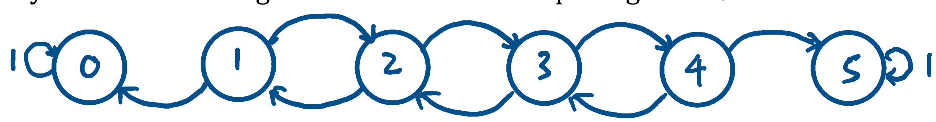
Thursday, September 20, 2018 9:31 AM

Introduction

- $\boldsymbol{T_y = \min\{n \geq 1 | X_n = y\}}$ is **the time of first return to $\boldsymbol{y}$**
- $\boldsymbol{\rho_{yy}} = \mathbb{P}_y(\boldsymbol{T_y} < \infty)$ is called the **return probability of y**
- It's easier to calculate the return probability rather than finding the PMF, $\mathbb{E}T_y$, etc
- But it's still difficult, so we try to **classify states categorically**
  - y is **transient** if $\rho_{yy} < 1$
  - y is **recurrent** if $\rho_{yy} = 1$
- It is possible to classify all states as transient or recurrent once at a time
- But we want to find a more efficient way to **classify the states in groups**

Example: Transient or Recurrent

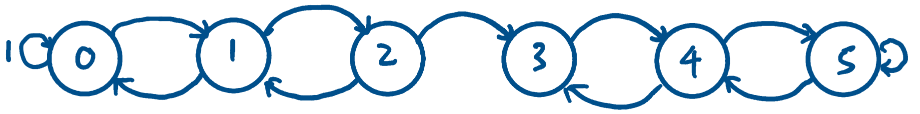
- Classify the states of the gambler's ruin MC for a prize goal of $5 as transient or recurrent



- Recurrent
  - $\rho_{00} = \mathbb{P}_0(T_0 < \infty) = p(0,0) = 1$
  - $\rho_{55} = \mathbb{P}_5(T_5 < \infty) = \rho(5,5) = 1$
  - So state 0 and state 5 is recurrent
- Transient
  - $\rho_{yy} < 1 \Leftrightarrow 1 - \rho_{yy} > 0 \Leftrightarrow \mathbb{P}_y(T_y = \infty) > 0$
  - $\mathbb{P}_2(T_2 = \infty) \geq \mathbb{P}_2(X_1 = 1, X_2 = 0) = p(2,1)p(1,0) > 0$
  - So state 2 is transient, similar for state 1, 3, and 4

Communication (Accessibility)

- Definition
  - We say that $\boldsymbol{x}$ **communicates with** $\boldsymbol{y}$ if $\boldsymbol{p^n(x,y) > 0}$ **for some** $\boldsymbol{n \geq 0}$, denoted by $x \Rightarrow y$
- Remark: Different from Textbook
  - Textbook uses $x \to y$ for communication
    - This single arrow is used in graphs to denote $p(x,y) > 0$
    - But since communication is more general than 1-step, we use double arrows
  - Textbook defines communication as $\mathbb{P}_x(T_y < \infty) = 1$
    - It is possible for $x \nRightarrow x$
    - But the usual convention is to ensure $x \Rightarrow x$, which is guaranteed for our definition
- Example



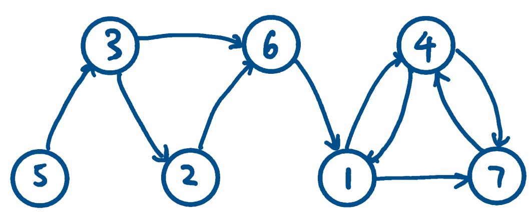
- Why $1 \Rightarrow 4$
  - $p^3(1,4) \geq p(1,2)p(2,3)p(3,4) > 0$
- Why $4 \nRightarrow 1$
  - Only $p(3,4), p(4,5) > 0$ for $p(4,j)$, so 4 cannot get to 3, 5 in one step
  - Thus $p(5,4)p(3,3)p(3,4) > 0$ are the only possible transitions from 3, 5
  - So for all $p^n(4,1) = 0$ *i.e.* $4 \nRightarrow 1$

Closed and Irreducible Sets

- A **closed set** of states is **impossible to get out of**
  - A set of states C is **closed** if the following condition is satisfied
  - If $i \in C$ and $p(i,j) > 0$, then $j \in C$
- A **irreducible set** of states can be **freely moved about** (you can go anywhere)
  - A set of states C is **irreducible** if $i \Leftrightarrow j, \forall i, j \in C$
- Example (in the graph above)
  - $\{1,2\}, \{3,4,5\}, \{4,5\}, \{2\}$ are irreducible sets
  - $\{3,4,5\}, \{1,2,3,4,5\}$ are closed sets

Decomposition of Finite State Space (Theorem 1.8)

- Statement
  - If the state space S is **finite**, then S can be written as a disjoint union
  - $\boldsymbol{T} \cup \boldsymbol{R_1} \cup \cdots \cup \boldsymbol{R_k}$ for $k \geq 1$ (at least one recurrent state), where
    - T is a set of transient states, and
    - R_i are closed irreducible sets of recurrent states.
- Example
  - Classify all states of the Markov chain with
  - $$\mathcal{P} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$



  - $T = \{2, 3, 5, 6\}$ is a set of transient states
  - $R_1 = \{1, 4, 7\}$ is a closed irreducible set of recurrent states

Number of Visits

- $N(y) =$ **Number of times** the Markov chain visit state y

Theorems Related to Recurrence

Monday, September 24, 2018 9:21 PM

Some Notation Reminders

- $T_y = \min\{n \geq 1 | X_n = y\}$
- $T_y^k = \min\{n > T_y^{k-1} | X_n = y\}$
- $N(y) =$ Number of times MC visits state y after time 0
- $\rho_{xy} = \mathbb{P}_x(T_y < \infty)$
- y is transient $\Leftrightarrow \rho_{yy} < 1$
- y is recurrent $\Leftrightarrow \rho_{yy} = 1$
- $x \Rightarrow y$ iff $p^n(x,y) > 0$ for some $n \geq 0$

Theorems Related to $N(y)$

- Lemma: **tail-sum formula**
  - If N is a RV taking values in $\{0,1,2,\ldots\}$, then $\mathbb{E}\boldsymbol{N} = \sum_{\boldsymbol{k}=\boldsymbol{1}}^{\infty} \mathbb{P}(\boldsymbol{N} \geq \boldsymbol{k})$
  - Define the indicator $\mathbb{1}_A = \begin{cases} 1 & A \text{ occurs} \\ 0 & A \text{ does not occur} \end{cases}$. Then
    - $N = \mathbb{1}_{\{N\geq 1\}} + \mathbb{1}_{\{N \geq 2\}} + \cdots = \sum_{k=1}^{\infty} \mathbb{1}_{\{N \geq k\}}$
  - Taking $\mathbb{E}$ on both side, we obtain
    - $\mathbb{E}N = \mathbb{E}\mathbb{1}_{\{N\geq 1\}} + \mathbb{E}\mathbb{1}_{\{N \geq 2\}} + \cdots = \mathbb{P}(N \geq 1) + \mathbb{P}(N \geq 2) + \cdots = \sum_{k=1}^{\infty} \mathbb{P}(N \geq k)$
- Lemma 1.11: $\mathbb{E}_{\boldsymbol{x}}\boldsymbol{N}(\boldsymbol{y}) = \dfrac{\boldsymbol{\rho_{xy}}}{\boldsymbol{1} - \boldsymbol{\rho_{yy}}}$
  - $\mathbb{E}_x N(y) = \sum_{k=1}^{\infty} \mathbb{P}_x(N(y) \geq k)$, by the tail-sum formula

$$= \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty) \text{, since } \{N(y) \geq k\} \text{ is the same as the } k\text{th return occurs}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}_x(T_y^k < \infty, T_y < \infty) \text{, since } \{T_y^k < \infty\} \text{ includes } \{T_y < \infty\}$$

$$= \sum_{k=1}^{\infty} \underbrace{\mathbb{P}_x(T_y^k < \infty | T_y < \infty)}_{\rho_{yy}^{k-1}} \underbrace{\mathbb{P}_x(T_y < \infty)}_{\rho_{xy}}$$

xy

yy

$$= \rho_{xy} \sum_{k=1}^{\infty} \rho_{yy}^{k-1} = \rho_{xy} \sum_{k=0}^{\infty} \rho_{yy}^{k} = \begin{cases} \dfrac{\rho_{xy}}{1-\rho_{yy}} & \text{if } \rho_{yy} < 1 \\ +\infty & \text{if } \rho_{yy} = 1 \end{cases}$$

- Lemma 1.12: $\mathbb{E}_x \boldsymbol{N(y)} = \sum_{n=1}^{\infty} \boldsymbol{p^n(x,y)}$
  - Use an indicator function to express $N(y)$: $N(y) = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = y\}}$
  - Then, $\mathbb{E}_x N(y) = \sum_{n=1}^{\infty} \mathbb{E}\mathbb{1}_{\{X_n=y\}} = \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = y) = \sum_{n=1}^{\infty} p^n(x,y)$
- Theorem 1.13: $\boldsymbol{y}$ **is recurrent** $\Leftrightarrow \sum_{n=1}^{\infty} \boldsymbol{p^n(y,y) = E_y N(y)} = +\infty$
  - y is recurrent $\Rightarrow \rho_{yy} = 1 \Rightarrow \mathbb{E}_y N(y) = \rho_{yy} \sum_{k=1}^{\infty} 1 = +\infty$
  - $\mathbb{E}_y N(y) = \sum_{k=0}^{\infty} \rho_{yy}^k = +\infty \Rightarrow \rho_{yy} = 1 \Rightarrow y$ is recurrent

Theorems Related to Communication

- Lemma 1.9: If $\boldsymbol{x \Rightarrow y}$ and $\boldsymbol{y \Rightarrow z}$, then $\boldsymbol{x \Rightarrow z}$
  - $p^{n_1}(x,y) > 0$ and $p^{n_2}(y,z) > 0$ for some $n_1, n_2 \geq 0$
  - $p^{n_1+n_2}(x,z) \geq p^{n_1}(x,y) p^{n_2}(y,z) > 0$
  - Therefore $x \Rightarrow z$
- Theorem 1.5: If $\boldsymbol{x \Rightarrow y}$ and $\boldsymbol{\rho_{yx} < 1}$, then $\boldsymbol{x}$ **is transient**
  - Let $n \in \mathbb{N}$ s.t. $p^n(x,y) > 0$
  - $\mathbb{P}_x(T_x = \infty) \geq \mathbb{P}_x(T_x = \infty, X_n = y)$

$$= \underbrace{\mathbb{P}_x(T_x = \infty | X_n = y)}_{\mathbb{P}_y(T_x = \infty)} \underbrace{\mathbb{P}_x(X_n = y)}_{p^n(x,y)} = (1 - \rho_{yx}) p^n(x,y) > 0$$

  - So $\rho_{xx} = \mathbb{P}_x(T_x < \infty) = 1 - \mathbb{P}_x(T_x = \infty) < 1$
  - Therefore x is transient
- Lemma 1.6: If $\boldsymbol{x}$ **is recurrent** and $\boldsymbol{x \Rightarrow y}$, then $\boldsymbol{\rho_{yx} = 1}$
  - Use the contrapositive from the previous theorem
  - If x is recurrent, then $x \not\Rightarrow y$ or $\rho_{yx} = 1$
  - By assumption $x \Rightarrow y$, so $\rho_{yx} = 1$
- Lemma 1.9: If $\boldsymbol{x}$ **is recurrent** and $\boldsymbol{x \Rightarrow y}$, then $\boldsymbol{y}$ **is recurrent**
  - By the previous lemma, we have $y \Rightarrow x$
  - So there exists l, k s.t. $p^k(y,x) > 0$ and $p^l(x,y) > 0$
  - We want to show that $E_y N(y) = +\infty$

- $E_y N(y) = \sum_{n=1}^{\infty} p^n(y,y)$
- $\geq \sum_{n=1}^{\infty} p^{k+n+l}(y,y)$, the inequality holds since this is just one possible path
- $= \sum_{n=1}^{\infty} p^k(y,x) p^n(x,x) p^l(x,y)$, by Chapman– Kolmogorov equation
- $\geq p^l(x,y) p^k(y,x) \underbrace{\sum_{n=1}^{\infty} p^n(x,x)}_{\mathbb{E}_x N(x)}$, since only $p^n(x,x)$ depends on n
- $= p^l(x,y) p^k(y,x) \underbrace{\mathbb{E}_x N(x)}_{\infty} = +\infty$

  - Therefore y is recurrent

Finite, Closed ⇒ ∃ Recurrent State (Lemma 1.9)

- Statement
  - In a **finite closed** set of states, there is **at least one recurrent state**
- Proof
  - Let C be a closed finite set of states
  - Suppose that there is no recurrent state in C (*i.e.* $\mathbb{E}_x N(y) < \infty, \forall x, y \in C$)
  - Then, $\sum_{y \in C} \mathbb{E}_x N(y) = \sum_{y \in C} \sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty} \underbrace{\sum_{y \in C} p^n(x,y)}_{1} = +\infty$
  - This contradicts $\mathbb{E}_x N(y) < \infty$
  - So the assumption is wrong, there must be a recurrent state

Finite, Closed, Irreducible ⇒ Recurrent (Theorem 1.7)

- Statement
  - If C is a **finite closed** and **irreducible** set, then all states in C are **recurrent**
- Proof
  - By the previous lemma, there is at least one recurrent state x
  - Because C is irreducible, $x \Rightarrow y$ for all $y \in C$
  - So y is also recurrent by Lemma 1.9
  - Therefore all states in C are recurrent

Stationary Distribution/Measure, Renewal Chain

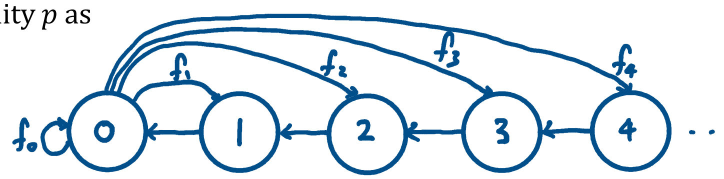
Thursday, September 27, 2018 9:32 AM

Stationary Distribution and Stationary Measure

- Motivation
  - Let $X_0, X_1, \dots$ be a Markov chain, and μ be its initial distribution
  - Then the distribution of X_i is
    - $\mathbb{P}_\mu(X_i = j) = \sum_{i \in S} \mu(i) p^n(i,j)\,, \forall j \in S$, or
    - $X_i \sim \mu \mathcal{P}^i$ (in matrix form)
  - What conditions must be satisfied so that $X_0, X_1, \dots$ **follow the same distribution**
- We say that $\mu: S \to \mathbb{R}_{\geq 0}$ is a **stationary/invariant measure** for a MC if
  - $\boldsymbol{\mu(j) = \sum_{i \in S} \mu(i) p(i,j)}$ (coordinate form), or
  - $\boldsymbol{\mu = \mu \mathcal{P}}$ (matrix form), or
  - μ is a left eigenvector of $\mathcal{P}$ with eigenvalue 1
- We say $\pi: S \to \mathbb{R}_{\geq 0}$ is a **stationary/invariant distribution** for a MC if
  - π is a stationary measure and $\boldsymbol{\sum_{j \in S} \pi(j) = 1}$
- How can we convert stationary measures into stationary distributions?
  - Given $\mu = [1,2,4,3]$, we can take $\pi = \frac{1}{\sum_{i \in S} \mu(i)} \mu$
  - But this may not work when $\sum_{i \in S} \mu(i)$ is not finite
- Example: Social Mobility (Example 1.18)
  - Given $\mathcal{P} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$
  - Find the stationary distribution for this MC
  - $[\pi_1, \pi_2, \pi_3] \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.4 & 0.4 \end{bmatrix} = [\pi_1, \pi_2, \pi_3]$
  - $\Rightarrow \begin{cases} 0.7\pi_1 + 0.3\pi_2 + 0.2\pi_3 = \pi_1 \\ 0.2\pi_1 + 0.5\pi_2 + 0.4\pi_3 = \pi_2 \\ 0.1\pi_1 + 0.2\pi_2 + 0.4\pi_3 = \pi_3 \end{cases} \Rightarrow \begin{cases} \pi_1 = 22/47 \\ \pi_2 = 16/47 \\ \pi_3 = 9/47 \end{cases}$
- How can we guarantee a stationary distribution exists
  - If a Markov chain is **irreducible** and **finite**, then
  - There is a **unique stationary distribution** π, and $\pi(j) > 0, \forall j \in S$
  - Proof: Linear algebra

Example: Renewal Chain (Countably Infinite State Space)

- $S = \mathbb{Z}_{\geq 0} = \{0,1,2,\dots\}$
- Let $\{f_k\}_{k\geq 0}$ be a distribution on S
- Define the transition probability p as
    - $p(0,k) = f_k$
    - $p(k,k-1) = 1$



- Let $f_k = \frac{6}{\pi^2} \cdot \frac{1}{(k+1)^2}$
- Obviously, 0 is recurrent $\Leftrightarrow \mathbb{P}_0(T_0 < \infty) = 1$
- What is $\mathbb{E}_0 T_0$?
    - $\mathbb{E}_0 T_0 = \sum_{k=1}^{\infty} k \mathbb{P}_0(T_0 = k) = \sum_{k=1}^{\infty} k f_{k-1} = \sum_{k=1}^{\infty} k \frac{6}{\pi^2} \cdot \frac{1}{k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} = +\infty$
- Find an invariant measure for this MC
    - Let μ be an invariant measure, then
    - $\mu(k) = \sum_{l=0}^{\infty} \mu(l) p(l,k)$

      $= \mu(0) \underbrace{p(0,k)}_{f_k} + \mu(k+1) \underbrace{p(k+1,k)}_{1}$, since we can only get k from 0 or $k+1$

      $= \mu(0) f_k + \mu(k+1)$
    - Thus, $\mu(k+1) = \mu(k) - \mu(0) f_k$
    - Solving the recursion, we have $\mu(k) = \mu(0)\left(1 - \sum_{l=0}^{k-1} f_l\right)$
    - Set $\mu(0) = 1$ (since we can freely scale the invariant measure by a positive number)
    - Then for $k \geq 1, \mu(k) = 1 - \sum_{l=0}^{k-1} f_l = \sum_{l=k}^{\infty} f_l = \sum_{l=k}^{\infty} \mathbb{P}_0(T_0 = l+1) = \mathbb{P}_0(T_0 \geq k+1)$
    - Note: $f = \mathbb{P}_0(T_0 = l+1)$ since we need 1 step to get l, and l steps to return to 0
- Can we make μ into a distribution?
    - $\sum_{k=0}^{\infty} \mu(k) = \sum_{k=0}^{\infty} \mathbb{P}_0(T_0 \geq k+1) \overset{\text{tail sum}}{=} \mathbb{E}_0 T_0 = +\infty$
    - So we cannot normalize μ into distribution
- Repeat this problem with $f_k = \frac{1}{2^{k+1}}$ (see next lecture)

Positive/Null Recurrent, Limit Behavior

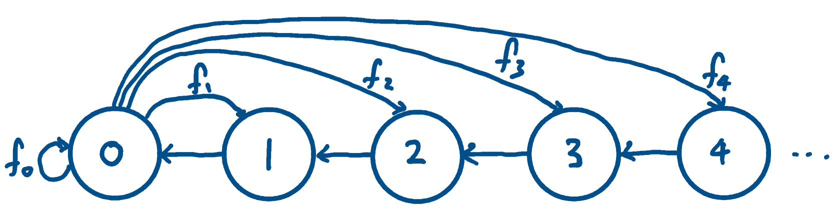
Tuesday, October 2, 2018 9:31 AM

Stationary Distribution and Stationary Measure

- Stationary measure
  - $\mu: S \to \mathbb{R}_{\geq 0}$ $s.t.$ $\boldsymbol{\mu(k)} = \sum_{l \in S} \boldsymbol{\mu(l) p(l, k)}$
- Stationary distribution
  - A stationary measure π with $\sum_{l \in S} \boldsymbol{\pi(l)} = \mathbf{1}$
  - Given $\sum_{l \in S} \mu(l) \neq \infty$, we can normalize μ by setting $\boldsymbol{\pi(k)} = \frac{\boldsymbol{\mu(k)}}{\sum_{l \in S} \boldsymbol{\mu(l)}}$
  - In finite case, we can solve for $\boldsymbol{\pi} = \boldsymbol{\pi \mathcal{P}}$ with $\sum_{l \in S} \pi(l) = 1$
- Motivation
  - If π is the **initial distribution**, then $X_0, X_1, \ldots$ all **have the same distribution**
  - $\mathbb{P}_{\boldsymbol{\pi}}(\boldsymbol{X_j} = \boldsymbol{x}) = \mathbb{P}_{\boldsymbol{\pi}}(\boldsymbol{X_k} = \boldsymbol{x}), \forall j, k \geq 0, \forall x \in S$

Example: Renewal Chain (Cont.)

- $S = \mathbb{Z}_{\geq 0} = \{0,1,2, \ldots\}$
- Let $\{f_k\}_{k \geq 0}$ be a distribution on S
- Define the transition probability p as
  - $p(0, k) = f_k$
  - $p(k, k-1) = 1$



- In the previous lecture, we set $f_k = \frac{6}{\pi^2} \cdot \frac{1}{(k+1)^2}$, and found
  - $\mathbb{E}_0 T_0 = +\infty$
  - $\mu(k) = \sum_{l=k}^{\infty} f_l = \mathbb{P}_0(T_0 \geq k+1)$
  - $\sum_{k=0}^{\infty} \mu(k) = +\infty \Rightarrow \pi$ does not exist

- If we set $f_k = \frac{1}{2^{k+1}}$, then
  - $\mathbb{E}_0 \mathrm{T}_0 = \sum_{k=1}^{\infty} k \mathbb{P}_0(T_0 = k) = \sum_{k=1}^{\infty} k f_{k-1} = \sum_{k=1}^{\infty} \frac{k}{2^k} \overset{\text{Geo}\left(\frac{1}{2}\right)}{=} 2$
  - $\mathbb{P}_0(T_0 = k) = f_{k-1} = \left(\frac{1}{2}\right)^k$, so $T_0 \sim \text{Geo}\left(\frac{1}{2}\right)$
  - $\sum_{l=0}^{\infty} \mu(l) = \sum_{l=0}^{\infty} \mathbb{P}_0(T_0 \geq k+1) = \sum_{l=1}^{\infty} \mathbb{P}_0(T_0 \geq k) \overset{\text{tail sum}}{=} \mathbb{E}_0 T_0 = 2$
  - $\pi(k) = \frac{\mu(k)}{2} = 2^{-k-1}$

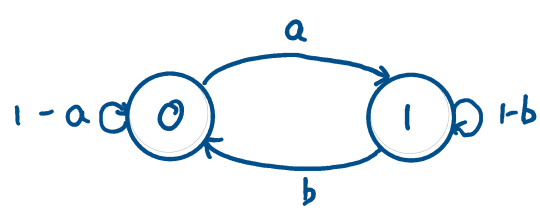
Positive vs Null Recurrent

- Motivation
  - In the previous example, even for recurrent states, it is possible to have $\mathbb{E}_x \boldsymbol{T}_x = \infty$
- Definition
  - Suppose x is recurrent, we say that
  - x is **positive recurrent** if $\mathbb{E}_x T_x < \infty$
  - x is **null recurrent** if $\mathbb{E}_x T_x = \infty$

Theorem Related to Recurrence and Stationary Measure/Distribution

- Suppose we have a MC with irreducible state space (finite or countably infinite)
- If all states are **recurrent**, then
  - The MC has a unique stationary measure μ up to multiplicative constants
  - $\boldsymbol{\mu(x) > 0}, \forall x \in S$
  - The stationary distribution $\pi(x) = \frac{1}{\mathbb{E}_x T_x}$ exists iff all states are **positive recurrent**
- Note: If $x \Leftrightarrow y$, then x and y are both **transient**, **positive recurrent**, or **null recurrent**

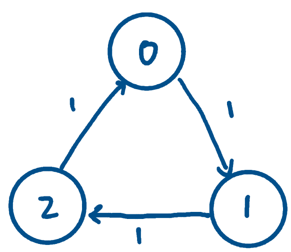
Example: Limit Behavior of Two State MC



- Find the n-step transitions
  - Compute $\mathbb{P}_0(X_n = 0)$
    - $\mathbb{P}_0(X_n = 0) = \mathbb{P}_0(X_{n-1} = 0)(1-a) + \mathbb{P}_0(X_{n-1} = 1)b$
    - Solving the recurrence, we have $\mathbb{P}_0(X_n = 0) = (1-a-b)\mathbb{P}_0(X_{n-1} = 0) + b$
  - Set $x_n = \mathbb{P}_0(X_n = 0)$. Then
    - $x_n = (1-a-b)x_{n-1} + b$

- $x_n - \frac{b}{a+b} = (1-a-b)\left(x_{n-1} - \frac{b}{a+b}\right)$
  - Set $y_n = x_n - \frac{b}{a+b}$. Then
    - $y_n = (1-a-b)y_{n-1}$
    - $y_n = (1-a-b)^n y_0$
  - Therefore $\mathbb{P}_0(X_n = 0) - \frac{b}{a+b} = (1-a-b)^n\left(\mathbb{P}_0(X_0 = 0) - \frac{b}{a+b}\right)$
  - $p^n(0,0) = (1-a-b)^n\left(1 - \frac{b}{a+b}\right) + \frac{b}{a+b} = \frac{b}{a+b} + (1-a-b)^n\frac{a}{a+b}$
  - $p^n(0,1) = 1 - p^n(0,0) = (1-(1-a-b)^n)\frac{a}{a+b}$
- Evaluate $\lim_{n\to\infty} p^n(x,y)$
  - $\lim_{n\to\infty} p^n(0,0) = \lim_{n\to\infty}\left(\frac{b}{a+b} + (1-a-b)^n\frac{a}{a+b}\right) = \frac{b}{a+b}$
  - $\lim_{n\to\infty} p^n(0,1) = \lim_{n\to\infty}\left((1-(1-a-b)^n)\frac{a}{a+b}\right) = \frac{a}{a+b}$
- Remark
  - $\boldsymbol{\pi}(\mathbf{0}) = \frac{b}{a+b} = \lim_{n\to\infty} \boldsymbol{p^n}(\mathbf{0},\mathbf{0})$
  - $\boldsymbol{\pi}(\mathbf{1}) = \frac{a}{a+b} = \lim_{n\to\infty} \boldsymbol{p^n}(\mathbf{0},\mathbf{1})$

Periodicity

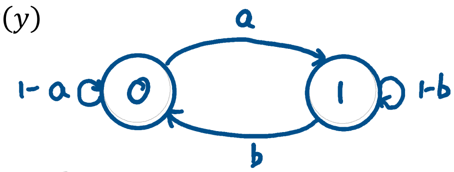


- For the MC on the right
  - $p(0,0) = 0$
  - $p^2(0,0) = 0$
  - $p^3(0,0) = 1$
- We observe a period of 3 for the n-th return probability of state 0
- We say a state is **aperiodic** if the state has a period of 1
- (The definition of periodicity will be formalized in the next lecture)

Periodicity, Limiting Behavior

Thursday, October 4, 2018 9:35 AM

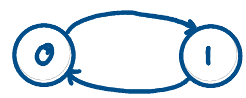
Example: Two State MC (Cont.)



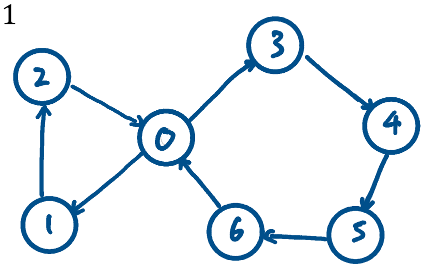
- For $0 < a, b < 1$, we showed that $\lim_{n\to\infty} p^n(x,y) = \pi(y)$
- This is very difficult to compute the limit explicitly
- We will prove theorem to show this often is true
- One minor issue that can prevent convergence is **periodicity**
- When $\alpha = \beta = 1$, $p(0,1) = 1; p^2(0,1) = 0; p^3(0,1) = 1, p^4(0,1) = 0, \cdots$

Periodicity

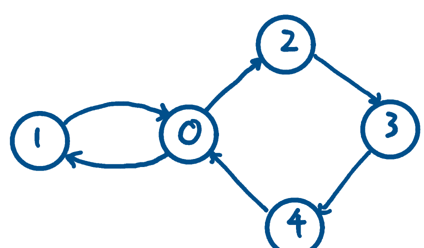
- Intuition
  - Period represents the **minimal length of gaps** between visits to that state
- Definition
  - The period of a state x is **gcd** $\underbrace{\{\boldsymbol{n \geq 1 | p^n(x,x) > 0}\}}_{\boldsymbol{I_x}}$
- Example 1: Two state chain with $a = b = 1$
  - $I_0 = \{n \geq 1 | p^n(0,0) > 0\} = \{2,4,6,8,\cdots\} \Rightarrow \gcd(I_0) = 2$
  - So state 0 has period 2, and same for state 1



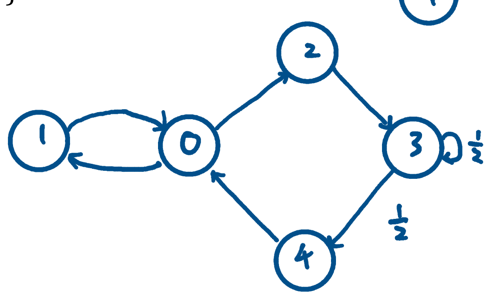
- Example 2: Find the period of 0
  - $p^3(0,0) > 0$ and $p^5(0,0) > 0$
  - So $p^{3k+5l} > 0$
  - $I_0 = \{3k + 5l | k, l \geq 0 \text{ not both equal to } 0\}$
  - $\gcd(I_0) = \gcd(3,5) = 1$
  - So 0 has period 1 (it is aperiodic)
  - $I_0 = \{3,5,6,8,9,10,11,12,\cdots\}$



- Example 3: Find the period of 0
  - $I_0 = \{2k + 4l | k > 0 \text{ or } l > 0\}$
    $= \{2(k + 2l) | k > 0 \text{ or } l > 0\}$
  - $\Rightarrow \gcd(I_0) = 2$
  - So 0 has period 2



- Example 4: Find the period of 0
  - $I_0 = \{2,4,5,6,7\cdots\}$
  - So 0 is aperiodic



Theorems Related to Periodicity

- Lemma 1.15: If $\boldsymbol{p^j(x,x) > 0}$ and $\boldsymbol{p^k(x,x) > 0}$, then $\boldsymbol{p^{j+k}(x,x) > 0}$

- Lemma 1.17: If $\boldsymbol{p(x,x) > 0}$, then x has **period 1** (is aperiodic)
- Lemma 1.16: If $\boldsymbol{x}$ **has period 1**, then $\exists \boldsymbol{n_0} \in \mathbb{N}$ **s.t.** $\boldsymbol{p^n(x,x) > 0}, \forall \boldsymbol{n \geq n_0}$
- Lemma 1.18: If $\boldsymbol{x \Leftrightarrow y}$, then x and y have the **same period**

Theorems Related to Limiting Behavior

- Convergence Theorem (Theorem 1.19)
  - Suppose a MC is **irreducible**, **aperiodic**, and has a stationary distribution π
  - Then $\boldsymbol{\lim_{n\to\infty} p^n(x,y) = \pi(y)}$
  - Note that the choice of x is arbitrary
- Asymptotic Frequency (Theorem 1.21)
  - Suppose a MC is **irreducible** and **recurrent**. Then
  - $\boldsymbol{\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y T_y}}$ where $N_n(y)$ is the number of visits to y up to time n
- Law of Large Numbers for MC (Theorem 1.23)
  - Suppose a MC is **irreducible** and has a stationary distribution π. Let $f: S \to \mathbb{R}$
  - If $\sum_{x\in S} |f(x)|\pi(x) < \infty$, then $\boldsymbol{\frac{1}{n}\sum_{l=1}^{n} f(X_l) \to \sum_{x\in S} f(x)\pi(x)} = \mathbb{E}_\pi f(x_0)$

Example 1.24: Inventory Chain

- A store may sell 0, 1, 2, 3 items with probabilities 0.3, 0.4, 0.2, 0.1
- Let X_n be number of units in store at end of the day
- We want to find the optimal inventory policy given the profit $g(X_n) = 12(3 - X_n) - 2X_n$
- We can compare average daily profit for restocking when $X_n = 0$ or 1 or 2
- If we restock when $X_n \leq 2$, then
  - $\mathcal{P} = \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \end{bmatrix} \Rightarrow \pi = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}^T$
- Average profit after n days is
  - $\frac{1}{n}\sum_{l=1}^{n} g(X_l) \overset{n\gg 1}{\approx} \sum_{s=0}^{3} g(s)\pi(s) = \sum_{s=0}^{3} [12(3-s) - 2s]\pi(s) = 9.40$
- Repeat for restocking when $X_n \leq 0$ and $X_n \leq 1$
- We will find out that it is optimal to restock when $X_n \leq 1$

Convergence Theorem

Tuesday, October 9, 2018 9:32 AM

Review: Markov Chain Convergence Theorem

- If a MC is **irreducible**, **aperiodic**, and has a **stationary distribution** $\boldsymbol{\pi}$
- Then $\lim_{n\to\infty} \boldsymbol{p^n(x,y) = \pi(y), \forall x, y \in S}$

Proof for Markov Chain Convergence Theorem

- Proof outline (using **coupling method**)
  - Consider two MCs with same transition probabilities, but different initial distributions
  - Let $x \in S$ be the **fixed initial state** for $X_0, X_1, \ldots$
  - Let π be the **initial distribution** for $Y_0, Y_1, \ldots$
  - We will show that $|\mathbb{P}_{\boldsymbol{x}}(\boldsymbol{X_n = y}) - \mathbb{P}_{\boldsymbol{\pi}}(\boldsymbol{Y_n = y})| \to \boldsymbol{0}$ as $n \to \infty$
  - Then $|p^n(x,y) - \pi(y)| \to 0$ as $n \to \infty$
- **Define a coupled MC**
  - Set $\bar{S} = S \times S$ as a new state space
  - Set $\bar{p}\big((x_1,y_1),(x_2,y_2)\big) = p(x_1,x_2)p(y_1,y_2)$
  - Use the initial distribution $\mu\big((x_0,y_0)\big) = \mathbb{1}_{\{x_0=x\}}\pi(y_0)$
  - We now have a single MC $(X_0,Y_0),(X_1,Y_1),\ldots$
- Show $\bar{\boldsymbol{p}}$ **is irreducible**
  - Let $(x_1,y_1),(x_2,y_2) \in \bar{S} = S \times S$ be arbitrary. We will show that $(x_1,y_1) \Rightarrow (x_2,y_2)$
  - Note that this is non-trivial, consider the product MC of $\mathcal{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
  - p is irreducible, so there exists k, l s.t.
    - $p^k(x_1,x_2) > 0$ and $p^l(y_1,y_2) > 0$
  - p is aperodic, so there exists n_x, n_y s.t.
    - $p^{n+l}(x_1,x_2) > 0$ and $p^{n+k}(x_1,x_2) > 0$ for $n > \max\{n_x,n_y\}$
  - Then $\bar{p}^{n+l+k}\big((x_1,y_1),(x_2,y_2)\big)$

    $= p^{n+l+k}(x_1,x_2)p^{n+l+k}(y_1,y_2)$

    $\geq \underbrace{p^k(x_1,x_2)}_{>0}\ \underbrace{p^{l+k}(x_1,x_2)}_{>0}\ \underbrace{p^l(y_1,y_2)}_{>0}\ \underbrace{p^{n+k}(y_1,y_2)}_{>0} > 0$ if $n > \max\{n_x,n_y\}$
  - Therefore $\bar{p}$ is irreducible
- Find **stationary distribution** for $\bar{p}$
  - Claim: $\bar{\pi}\big((x_0,y_0)\big) \coloneqq \pi(x_0)\pi(y_0)$ is a stationary distribution for $\bar{p}$
  - $\bar{\pi}\big((x_0,y_0)\big) = \sum_{(u,v)\in S\times S} \bar{p}\big((u,v),(x_0,y_0)\big)\bar{\pi}\big((x_0,y_0)\big)$

$$= \sum_{u\in S}\sum_{v\in S} p(u,x_0)p(v,y_0)\pi(x_0)\pi(y_0)$$

$$= \underbrace{\sum_{u\in S} p(u,x_0)\pi(x_0)}_{\pi(x_0)} \underbrace{\sum_{v\in S} p(v,y_0)\pi(y_0)}_{\pi(y_0)}$$

$$= \pi(x_0)\pi(y_0)$$

- $\sum_{(u,v)\in S\times S} \bar{\pi}\big((u,v)\big) = \sum_{u\in S}\sum_{v\in S}\pi(u)\pi(v) = \sum_{u\in S}\pi(u)\sum_{v\in S}\pi(v) = 1$
- Therefore $\bar{\pi}\big((x_0,y_0)\big)$ is a stationary distribution for $\bar{p}$

- Show that **X_n, Y_n must eventually meet**
  - Set $V_{(x,x)} \coloneqq \min\{n \geq 0 | X_n = Y_n = x\}$ and $T \coloneqq \min\{n \geq 0 | X_n = Y_n\}$
  - Since $\bar{p}$ is irrducible and has a stationary distribution, all states are recurrent
  - Thus, $\mathbb{P}_\mu\big(V_{(x,x)} < \infty\big) = 1 \Longrightarrow \mathbb{P}_\mu(T < \infty) = 1$, since $T \leq V_{(x,x)}$
- Show **X_n, Y_n have same distribution after meeting**
  - $\mathbb{P}_\mu(X_n = y, n \geq T) = \sum_{k=0}^{n}\sum_{z\in S}\mathbb{P}_\mu(X_k = z, T = k, X_n = y)$
  - $= \sum_{k=0}^{n}\sum_{z\in S}\mathbb{P}_\mu(X_n = y | X_k = z, T = k)\mathbb{P}_\mu(X_k = z, T = k)$
  - $= \sum_{k=0}^{n}\sum_{z\in S} p^{n-k}(z,y)\mathbb{P}_\mu(X_k = z, T = k)$, by strong Markov property
  - $= \sum_{k=0}^{n}\sum_{z\in S} p^{n-k}(z,y)\mathbb{P}_\mu(Y_k = z, T = k)$
  - $= \sum_{k=0}^{n}\sum_{z\in S}\mathbb{P}_\mu(Y_n = y | Y_k = z, T = k)\mathbb{P}_\mu(Y_k = z, T = k)$
  - $= \sum_{k=0}^{n}\sum_{z\in S}\mathbb{P}_\mu(Y_k = z, T = k, Y_n = y) = \mathbb{P}_\mu(Y_n = y, n \geq T)$
- Show $\big|\mathbb{P}_{\boldsymbol{\mu}}(\boldsymbol{X_n} = \boldsymbol{y}) - \mathbb{P}_{\boldsymbol{\mu}}(\boldsymbol{Y_n} = \boldsymbol{y})\big| \to \boldsymbol{0}$ as $n \to \infty$
  - $\big|\mathbb{P}_\mu(X_n = y) - \mathbb{P}_\mu(Y_n = y)\big| = \begin{vmatrix} \mathbb{P}_\mu(X_n = y, T > n) + \mathbb{P}_\mu(X_n = y, T \leq n) \\ -\mathbb{P}_\mu(Y_n = y, T > n) - \mathbb{P}_\mu(Y_n = y, T \leq n) \end{vmatrix}$

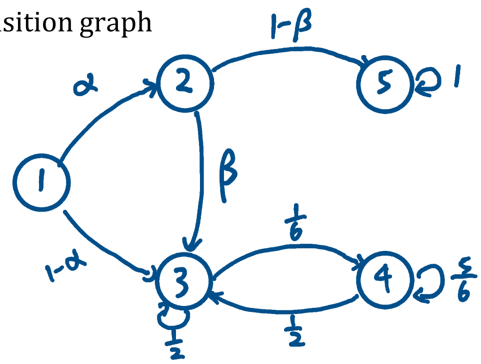
    $\leq \big|\mathbb{P}_\mu(X_n = y, T > n) - \mathbb{P}_\mu(Y_n = y, T > n)\big|$
  - $\sum_{y\in S}\big|\mathbb{P}_\mu(X_n = y) - \mathbb{P}_\mu(Y_n = y)\big| \leq \sum_{y\in S}\big|\mathbb{P}_\mu(X_n = y, T > n) - \mathbb{P}_\mu(Y_n = y, T > n)\big|$

    $\leq \sum_{y\in S}\mathbb{P}_\mu(X_n = y, T > n) + \sum_{y\in S}\mathbb{P}_\mu(Y_n = y, T > n)$

    $\leq 2\sum_{y\in S}\mathbb{P}_\mu(T > n) \to 0$ as $n \to \infty$, since T is finite

Example: Convergence Theorem

- Let MC X_1, X_2, X_3, X_4, X_5 be defined as
  - $\mathcal{P} = \begin{bmatrix} 0 & \alpha & 1-\alpha & 0 & 0 \\ 0 & 0 & \beta & 0 & 1-\beta \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/6 & 5/6 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ for $\alpha, \beta \in (0,1)$
- Draw the transition graph



- Classify states as transient or recurrent
  - $R_1 = \{3,4\}, R_2 = \{5\}$ are recurrent because they are closed, irreducible, finite
  - $T = \{1,2\}$ are transient
- Find the periods of recurrent states
  - $p(3,3), p(4,4), p(5,5) > 0$, so state 3, 4, 5 have period 1 (aperiodic)
- Find all stationary distributions
  - $\pi(1) = \pi(2) = 0$ because state 1, 2 are transient
  - The MC restricted to $R_1 = \{3,4\}$ has stationary distribution
    - $\pi^1 = \left[\frac{1/6}{1/6+1/2} \quad \frac{1/2}{1/6+1/2}\right] = \left[\frac{1}{4} \quad \frac{3}{4}\right]$
  - The MC restricted to $R_2 = \{5\}$ has stationary distribution
    - $\pi^2 = [1]$, since there is only one state
  - Therefore $\pi = \left[0 \quad 0 \quad s \cdot \frac{1}{4} \quad s \cdot \frac{3}{4} \quad (1-s) \cdot 1\right]$ for some constant $0 \le s \le 1$
- Compute $\lim_{n\to\infty} p^n(1,3)$
  - $\lim_{n\to\infty} p^n(1,3) = \lim_{n\to\infty} [p(1,3)p^{n-1}(3,3) + p(1,2)p(2,3)p^{n-2}(3,3)]$

$$= (1-\alpha) \lim_{n\to\infty} p^n(3,3) + \alpha\beta \lim_{n\to\infty} p^n(3,3)$$

$$= (1-\alpha+\alpha\beta) \lim_{n\to\infty} p^n(3,3) = (1-\alpha+\alpha\beta) \cdot \frac{1}{4}$$

Doubly Stochastic, Detailed Balance

Tuesday, October 16, 2018 9:33 AM

Doubly Stochastic Chains

- Stochastic matrix
  - The row of a MC's transition matrix **sums up to 1** $i.e. \sum_{y \in S} p(x, y) = 1$
  - Any matrix with **non-negative** values, and **row sum to 1** is called a **stochastic matrix**
  - Every stochastic matrix gives the transition probabilities for some MC
- Doubly stochastic
  - A stochastic matrix is **doubly stochastic** if its **column sum to 1** $i.e. \sum_{x \in S} p(x, y) = 1$
  - We say that a MC is doubly stochastic if its transition matrix is
- Stationary distribution of doubly stochastic MC
  - Statement
    - Suppose we have a finite state space MC, where $|S| = N$
    - $\boldsymbol{\pi(x)} = \frac{\mathbf{1}}{\boldsymbol{N}}, \forall x \in S$ is a **stationary distribution** $\Leftrightarrow$ the MC is **doubly stochastic**
  - $(\Longrightarrow)$ Assume π is a stationary distribution
    - $\pi(y) = \sum_{x \in S} \pi(x) p(x, y) \Leftrightarrow \frac{1}{N} = \frac{1}{N} \sum_{x \in S} p(x, y) \Leftrightarrow \sum_{x \in S} p(x, y) = 1$
    - So the MC is doubly stochastic
  - $(\Longleftarrow)$ Assume the MC is doubly stochastic
    - $\sum_{x \in S} \pi(x) p(x, y) = \frac{1}{N} \sum_{x \in S} p(x, y) = \frac{1}{N} = \pi(y), \forall y \in S$
    - Therefore $\pi(x) = \frac{1}{N}$ is a stationary distribution for this MC

Detailed Balance Condition

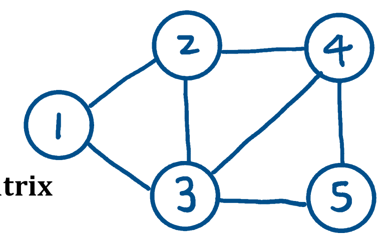
- Definition
  - We say a **distribution** satisfy the **detailed balance condition/equations** if
  - $\boldsymbol{\pi(x) p(x, y) = \pi(y) p(y, x), \forall x, y \in S}$
- Detailed balance condition and stationary distribution
  - Statement
    - All distributions satisfying the **detailed balance equations** are **stationary**
  - Proof
    - Suppose π satisfy the dtailed balance euqations *i.e.* $\pi(x) p(x, y) = \pi(y) p(y, x)$

- $\sum_{x \in S} \pi(x)p(x,y) = \sum_{x \in S} \pi(y)p(y,x) = \pi(y) \sum_{x \in S} p(y,x) = \pi(y) \Rightarrow \pi$ is stationary

- Example 1.29
  - $\mathcal{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.3 & 0.1 & 0.6 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$
  - Can $\mathcal{P}$ have a stationary distribution that satisfies DBE?
    - $\begin{cases} \pi(1)p(1,2) = \pi(2)p(2,1) \\ \pi(1)p(1,3) = \pi(3)p(3,1) \\ \pi(2)p(2,3) = \pi(3)p(3,2) \end{cases} \Rightarrow \begin{cases} 0.5 \cdot \pi(1) = 0.3 \cdot \pi(2) \\ 0 \cdot \pi(1) = 0.2 \cdot \pi(3) \\ 0.6 \cdot \pi(2) = 0.4 \cdot \pi(3) \end{cases} \Rightarrow \begin{cases} \pi(1) = 0 \\ \pi(2) = 0 \\ \pi(3) = 0 \end{cases}$
    - This is not a distribution, so none that satisfy DBE exists
  - Can it have any other stationary distributions?
    - Since $\mathcal{P}$ is doubly stochastic, so $\pi = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ is a stationary distribution
    - This is the only stationary distribution, as the MC is irreducible and finite

Random Markov on Graphs

- Undirected Graph
  - Undirected graph is a set of **vertices** and **edges**, $G = (V, E)$
  - $V = \{1,2,3,4,5\}$
  - $E = \{\{1,2\}, \{1,3\}, \{2,3\}, \{2,4\}, \{3,4\}, \{3,5\}, \{4,5\}\}$
  - $A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$ is called the **adjacency matrix**



  - The **neighbor** of a vertex are those vertices is share an edge with.
  - The **degree** of a vertex is the number of neighbors if has
- Random Walk on G
  - Set $S = V$. If in state V, you **choose a neighbor** of v **uniformaly** as the next state
  - Then $p(u,v) = \frac{A(u,v)}{\deg(u)}, \forall u, v \in V$
- Random walk and DBE
  - Statement
    - All **random walks' graphs satisfy DBE's**
  - Proof
    - $\pi(u)p(u,v) = \pi(v)p(v,u)$
    - $\Rightarrow \pi(u) \cdot \frac{A(u,v)}{\deg u} = \pi(v) \cdot \frac{A(u,v)}{\deg v}$
    - $\Rightarrow \frac{\pi(u)}{\deg u} = \frac{\pi(v)}{\deg v}$
    - If we set $\pi(x) = c \cdot \deg x, \forall x \in V$, then DBE is satisfied

- We just need to choose c so that π is a distribution
- $\sum_{v \in S} \pi(v) = \sum_{v \in S} c \cdot \deg x = 1 \Rightarrow c \coloneqq \frac{1}{\sum_{v \in S} \deg x}$
- Then $\pi(x) = \frac{\deg x}{\sum_{v \in S} \deg x} = \frac{\deg x}{2|E|}$

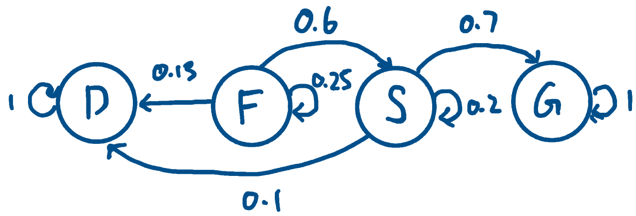
Reversibility

- Let $X_0, X_1, \ldots$ be a MC with transition probabilities p, stationary and initial distribution π
- Fix n and set $\boldsymbol{Y_m = X_{n-m}}, \forall m \in \{0,1,2, \ldots, n\}$ (*i.e.* $Y_0, \ldots, Y_n$ is a a time reversal for $X_0, \ldots, X_n$)
- Then Y_m is a MC with transition probability $\boldsymbol{\hat{p}(i,j) = \frac{\pi(j)p(j,i)}{p(i)}}$
- Moreover, if DBE's are satisfied, then $\hat{p} = p$
  - $\boldsymbol{\hat{p}(i,j) = \frac{\pi(j)p(j,i)}{\pi(i)} = \frac{\pi(i)p(i,j)}{\pi(i)} = p(i,j)}$

Exit Distributions

Thursday, October 18, 2018 9:31 AM

Exit Distribution Motivative Example: Community College



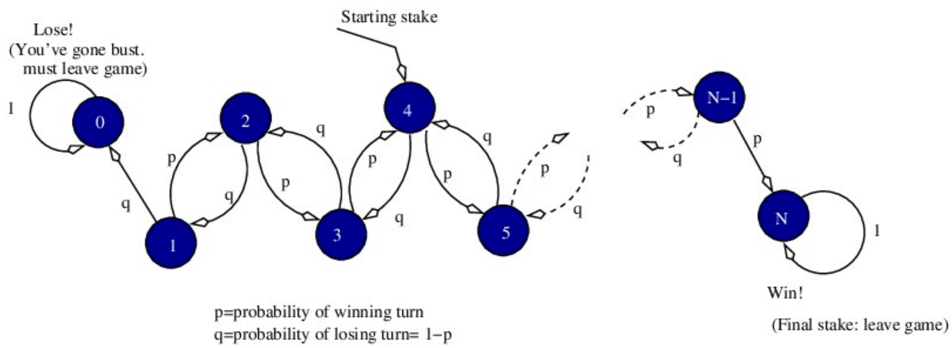
- Set $V_x \coloneqq \inf\{n \geq 0 | X_n = x\}$, and we want to compute $\mathbb{P}_F(V_G < V_D)$
- First step analysis: if $X_0 = F$, then $X_1 = D, F$, or S
- $$\begin{cases} \mathbb{P}_F(V_G < V_D) = \begin{pmatrix} \mathbb{P}_F(X_1 = D)\ \underbrace{\mathbb{P}_F(V_G < V_D | X_1 = D)}_{0} \\ +\mathbb{P}_F(X_1 = F)\ \underbrace{\mathbb{P}_F(V_G < V_D | X_1 = F)}_{\mathbb{P}_F(V_G < V_D)} \\ +\mathbb{P}_F(X_1 = S)\ \underbrace{\mathbb{P}_F(V_G < V_D | X_1 = S)}_{\mathbb{P}_S(V_G < V_D)} \end{pmatrix} \\ \mathbb{P}_S(V_G < V_D) = \begin{pmatrix} \mathbb{P}_S(X_1 = D)\ \underbrace{\mathbb{P}_S(V_G < V_D | X_1 = D)}_{0} \\ +\mathbb{P}_S(X_1 = F)\ \underbrace{\mathbb{P}_S(V_G < V_D | X_1 = F)}_{\mathbb{P}_F(V_G < V_D)} \\ +\mathbb{P}_S(X_1 = S)\ \underbrace{\mathbb{P}_S(V_G < V_D | X_1 = S)}_{1} \end{pmatrix} \end{cases}$$
- $$\begin{cases} \mathbb{P}_F(V_G < V_D) = 0.25 \cdot \mathbb{P}_F(V_G < V_D) + 0.6 \cdot \mathbb{P}_S(V_G < V_D) \\ \mathbb{P}_S(V_G < V_D) = 0.2 \cdot \mathbb{P}_F(V_G < V_D) + 0.7 \end{cases} \Rightarrow \begin{cases} \mathbb{P}_F(V_G < V_D) = 0.7 \\ \mathbb{P}_S(V_G < V_D) = 0.875 \end{cases}$$

Exit Distribution (Theorem 1.27)

- Brainstorming
  - Find $\mathbb{P}_x(V_a < V_b)$ for some $x, a, b \in S$
  - $\mathbb{P}_x(V_a < V_b) = \sum_{y \in S} \mathbb{P}_x(X_1 = y)\mathbb{P}_x(V_a < V_b | X_1 = y) = \sum_{y \in S} \overbrace{p(x,y)}^{\text{known}} \overbrace{\mathbb{P}_y(V_a < V_b)}^{\text{unknown}}$
  - So to find $\mathbb{P}_x(V_a < V_b)$, we need to find $\mathbb{P}_y(V_a < V_b), \forall y \in S$
- Observations (informal)
  - $\mathbb{P}_a(V_a < V_b) = 1$
  - $\mathbb{P}_b(V_a < V_b) = 0$
  - There are $|S|$ linear equations in $|S|$ variables
  - Define $h(x) \coloneqq \mathbb{P}_x(V_a < V_b)$, then we need to find $h: S \to \mathbb{R}$ that satisfies
    - $h(x) = \sum_{y \in S} p(x,y)h(y)$
    - $h(a) = 1, h(b) = 0$

- Theorem
  - Consider a MC with $|S| < \infty$
  - Let $a, b \in S$, and set $C \coloneqq S \setminus \{a, b\}$
  - Suppose $h: S \to \mathbb{R}$ satisfies
    - $h(a) = 1, h(b) = 0$
    - $h(x) = \sum_{y \in S} p(x, y) h(y), \forall x \in C$
  - If $\mathbb{P}_x(\min\{V_a, V_b\} < \infty) > 0, \forall x \in C$, then $h(x) = \mathbb{P}_x(V_a < V_b), \forall x \in S$

Exit Distribution Example: Gambler's Ruin



- Assume $p < \frac{1}{2}$, and we want to compute $\mathbb{P}_x(V_N < V_0)$
- Construct h
  - $h(0) = 0, h(N) = 1$
  - $h(x) = \sum_{y \in S} p(x, y) h(y) = p(x, x-1) h(x-1) + p(x, x+1) h(x+1)$
    $= p \cdot h(x-1) + q \cdot (x+1)$, for $x \in \{1, \dots, N-1\}$
  - $\Rightarrow \underbrace{p \cdot h(x) + q \cdot h(x)}_{h(x)} = p \cdot h(x-1) + q \cdot (x+1)$
  - $\Rightarrow p\big(h(x+1) - h(x)\big) = q\big(h(x) - h(x-1)\big)$
- Solve the recurrence equation
  - Set $u_x \coloneqq h(x+1) - h(x), \forall x \in \{1, \dots, N-1\}$
  - $u_x = \left(\frac{q}{p}\right) u_{x-1} \Longrightarrow u_x = \left(\frac{q}{p}\right)^x \mu_0$
  - $h(x) = \underbrace{h(x) - h(x-1)}_{u_{x-1}} + \underbrace{h(x-1) - h(x-2)}_{u_{x-2}} + \cdots - h(1) + \underbrace{h(1) - h(0)}_{u_0}$
    $= \sum_{l=0}^{x-1} u_l = u_0 \sum_{l=0}^{x-1} \left(\frac{q}{p}\right)^l = u_0 \frac{1 - (q/p)^x}{1 - q/p}$
  - $1 = h(N) = h(N) - h(0) = u_0 \frac{1 - (q/p)^N}{1 - q/p} \Longrightarrow u_0 = \frac{1 - q/p}{1 - (q/p)^N}$
- Therefore $h(x) = \frac{1 - (q/p)^x}{1 - (q/p)^N}$

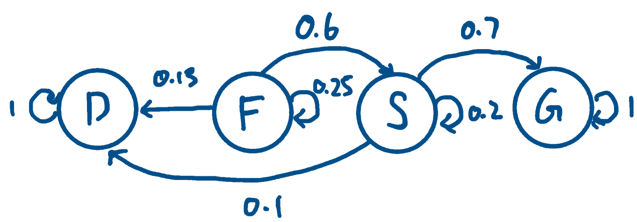
Exit Time

Tuesday, October 23, 2018 9:30 AM

Long Run Behavior of Markov Chains

- For **irreducible**, **aperiodic** MCs with $\boldsymbol{\pi}$, we have the **Convergence Theorem**
- If there are transient states in the MC, they will ultimately travel between recurrent states
- Two basic questions
  - **Which** closed set of recurrent states do you end up in? $\mathbb{P}_{\boldsymbol{x}}(\boldsymbol{V_a} < \boldsymbol{V_b})$
  - **How long** should we expect the MC to travel between transient states before ending up in a recurrent state? $\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{V_a}]$

Exit Time Motivating Example: Community College



- How long will the average student remain at this community college?
- Define $L = \{D, G\}$ and $V_L = \inf\{n \geq 0 | X_n \in L\}$. Then we need to find $\mathbb{E}_F[V_L]$
- $\mathbb{E}_F[V_L] = \sum_{l \in S} \underbrace{E_F[V_L | X_1 = l]}_{1+\mathbb{E}_l[V_L]} \underbrace{\mathbb{P}_F(X_1 = l)}_{p(F,l)}$, using firat step analysis

$$= \sum_{l \in S} (1 + \mathbb{E}_l[V_L])\, p(F, l), \text{ since we need 1 step to get from } F \text{ to } l$$

$$= 1 \cdot p(F, D) + (1 + \mathbb{E}_F[V_L])p(F, F) + (1 + \mathbb{E}_S[V_L])p(F, S)$$

$$= \underbrace{p(F, D) + p(F, F) + p(F, S)}_{1} + \mathbb{E}_F[V_L]p(F, F) + \mathbb{E}_S[V_L]p(F, S)$$

$$= 1 + \mathbb{E}_F[V_L]p(F, F) + \mathbb{E}_S[V_L]p(F, S)$$

$$= 1 + \mathbb{E}_F[V_L] \cdot 0.25 + \mathbb{E}_S[V_L] \cdot 0.6$$

- Similarly, we have $\mathbb{E}_S[V_L] = 1 + \mathbb{E}_S[V_L]p(S, S) = 1 + \mathbb{E}_S[V_L] \cdot 0.2$
- $\begin{cases} \mathbb{E}_F[V_L] = 1 + \mathbb{E}_F[V_L] \cdot 0.25 + \mathbb{E}_S[V_L] \cdot 0.6 \\ \mathbb{E}_S[V_L] = 1 + E_S[V_L] \cdot 0.2 \end{cases} \Rightarrow \begin{cases} \mathbb{E}_F[V_L] = 7/3 \\ \mathbb{E}_S[V_L] = 5/4 \end{cases}$

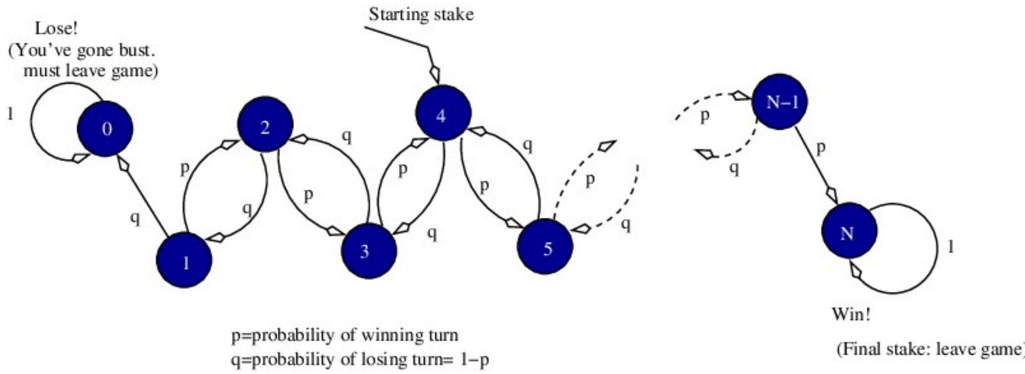
Exit Time (Theorem 1.28)

- Consider a MC with finite state space S
- Let $A \subseteq S$. Define $V_A \coloneqq \inf\{n \geq 0 | X_n \in A\}$ and $C \coloneqq S \setminus A$
- If $\mathbb{P}_x(V_A < \infty) > 0, \forall x \in C$, and $g: S \to \mathbb{R}$ satisfies
  - $\boldsymbol{g(a) = 0, \forall a \in A}$

- $g(x) = 1 + \sum_{y \in C} g(y)p(x,y)$

- Then $g(x) = \mathbb{E}_x[V_A]$ for all $x \in S$

Exit Time Example: Fair Gambler's Ruin



- Assume $p = q = 0.5$, how long should you expect to play the game?
- Set $A = \{0, N\}$, then we want to find $\mathbb{E}_x[V_A], \forall x \in \{1, \dots, N-1\}$
- Approach 1: **Use the theorem to verify/disprove a conjecture**
  - Claim: $\mathbb{E}_x[V_A] = x(N-x)$
  - Set $g(x) = x(N-x)$, then obviously $g(0) = g(N) = 0$
  - For $1 \le x \le N-1$
    - $1 + \sum_{y=1}^{N-1} g(y)p(x,y) = 1 + g(x-1)p(x,x-1) + g(x+1)p(x,x+1)$

$$= 1 + (x-1)\big(N-(x-1)\big) \cdot \frac{1}{2} + (x+1)\big(N-(x+1)\big) \cdot \frac{1}{2}$$

$$= Nx - x^2 = x(N-x) = g(x)$$

  - Therefore $g(x) = \mathbb{E}_x[T_A]$
- Approach 2: **Use the theorem to derive a solution**
  - By the theorem, we can define g as
    - $g(0) = g(N) = 0$
    - $g(x) = 1 + \frac{1}{2}g(x-1) + \frac{1}{2}g(x+1), \forall x \in \{1, \dots, N-1\}$
  - Solve as recurrence equations (or as a linear system)
  - $\big(g(x+1) - g(x)\big) = -2 + \big(g(x) - g(x-1)\big)$
  - Set $u_x = g(x+1) - g(x)$, then $u_x = -2 + u_{x-1} \Leftrightarrow u_x = u_0 - 2x, \forall x \in \{1, \dots, N-1\}$
  - $g(x) = g(x) - g(0) = \underbrace{g(x) - g(x-1)}_{u_{x-1}} + g(x+1) + \cdots + \underbrace{g(1) - g(0)}_{u_0}$

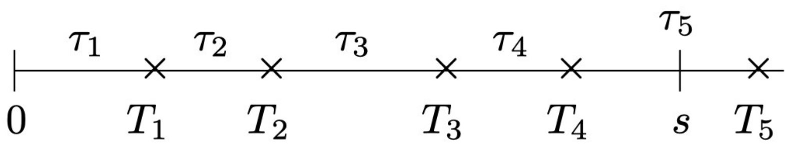
$$= \sum_{l=1}^{x} u_{x-1} = \sum_{l=1}^{x} \big(u_0 - 2(l-1)\big) = u_0 x - 2\frac{(x-1)x}{2} = u_0 x - (x-1)x$$

  - $g(N) = u_0 N - (N-1)N = 0 \Longrightarrow u_0 = N - 1$
  - Therefore $g(x) = (N-1)x - (x-1)x = x(N-x)$

Probability Review for Poisson Process

Thursday, October 25, 2018 9:32 AM

Renewal Process



- τ_k = **interarrival time**
- T_k = **arrival/renewal time**
- $N(s)$ = **number of renewals up to time s**

Definition of Poisson Process

- Let $\tau_1, \tau_2 \ldots \sim \text{Exp}(\lambda)$ be independent
- Set $T_0 = 0, T_k = T_{k-1} + \tau_k = \tau_1 + \cdots + \tau_k$
- Define $\boldsymbol{N(s) = \max\{n \geq 0 | T_n \leq s\}}$
- Then we call $\{N(s)\}$ a **Poisson process with rate λ**

Exponential Distribution

- Definition
  - We write that $\boldsymbol{X \sim \text{Exp}(\lambda)}$ for $\lambda > 0$ if
  - $\boldsymbol{f_X(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}}$, **or**
  - $\boldsymbol{F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}}$
- Survival function
  - $G(x) = \mathbb{P}(X > x) = 1 - F_X(x) = \begin{cases} e^{-\lambda x} & x \geq 0 \\ 1 & x < 0 \end{cases}$
- Expected value
  - $\mathbb{E}[X] = \int_0^\infty x f_X(x) dx = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$
- **$\text{Exp}(\lambda)$ is memoryless**
  - $\mathbb{P}(X > s + t | X > s) = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t)$

Gamma Distribution

- Definition
  - We say that $\boldsymbol{T \sim \text{Gamma}(n, \lambda)}$ if $\boldsymbol{f_T(t) = \begin{cases} \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{n-1}}{(n-1)!} & t \geq 0 \\ 0 & t < 0 \end{cases}}$

- Relation with exponential distribution
  - Let $\boldsymbol{\tau_1, \tau_2} \dots \sim \mathbf{Exp}(\boldsymbol{\lambda})$ be independent
  - Set $\boldsymbol{T_0 = 0, T_k = T_{k-1} + \tau_k = \tau_1 + \cdots + \tau_k}$**, then** $\boldsymbol{T_n} \sim \mathbf{Gamma}(\boldsymbol{n, \lambda})$
  - Proof by induction, the base case is trivial
  - For $n \geq 1$, $T_{n+1} = T_n + \tau_{n+1}$, where T_n and τ_{n+1} are independent
  - $f_{T_{n+1}}(t) = \left(f_{T_n} * f_{\tau_{n+1}}\right)(t) = \int_{-\infty}^{\infty} f_{T_n}(s) f_{\tau_{n+1}}(t-s)ds$

    $$= \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda(t-s)} ds = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^n}{n!} \text{ for } t \geq 0$$
  - So $T_{n+1} \sim \text{Gamma}(n+1, \lambda)$, which completes the proof

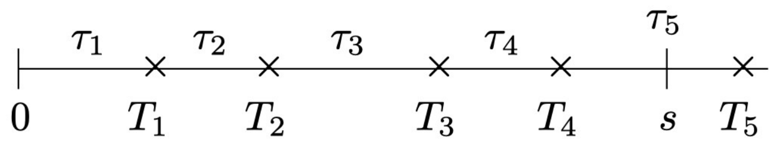
Poisson Distribution

- We say that $\boldsymbol{X} \sim \mathbf{Poisson}(\boldsymbol{\lambda})$ if $\boldsymbol{p_X(n) = e^{-\lambda} \frac{\lambda^n}{n!}}$ for $n = 0,1,2, \dots$
- $\mathbb{E}[X] = \sum_{n=1}^{\infty} n \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} \underbrace{\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}}_{e^\lambda} = \lambda \Rightarrow \mathbb{E}[\boldsymbol{X}] = \boldsymbol{\lambda}$
- $\mathbb{E}[X(X-1)] = \sum_{n=2}^{\infty} n(n-1) \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} = \lambda^2 \Rightarrow \mathbf{Var}[\boldsymbol{X}] = \boldsymbol{\lambda}$

Introduction to Poisson Process

Tuesday, October 30, 2018 9:31 AM

Poisson Process



- In the graph above,
  - τ_k = **interarrival time**
  - T_k = **arrival/renewal time**
  - $N(s)$ = **number of arrivals up to time** $\boldsymbol{s}$
- For Poisson process, we have
  - $\boldsymbol{\tau_k \overset{iid}{\sim} \mathrm{Exp}(\lambda)}$
  - $\boldsymbol{T_n = \tau_1 + \cdots + \tau_n \sim \mathrm{Gamma}(n, \lambda)}$
  - $\boldsymbol{N(s) \sim \mathrm{Poisson}(\lambda s)}$

Equivalent Definition of Poisson Process

- $\{N(s) | s \geq 0\}$ is a Poisson process with rate λ if and only if
  - $\boldsymbol{N(0) = 0}$ (with probability 1)
  - $\boldsymbol{N(t+s) - N(s) \sim \mathrm{Poisson}(\lambda t)}$
  - $N(t)$ has **independent increments**
- Independent increment
  - We say that $N(t)$ has **independent increments** if for any $t_0 < \cdots < t_n$, the random variables $N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1})$ are independent
  - The number of arrivals between any two intervals has no effect to each other
- Proof ($\Longleftarrow$)
  - $\mathbb{P}(N(s) = n) = \mathbb{P}(T_n \leq s, T_{n+1} > s) = \mathbb{P}(T_n \leq s, \tau_{n+1} > s - T_n)$

$$= \int_0^s \int_{s-t}^\infty f_{T_n, \tau_{n+1}}(t, u)\, du\, dt = \int_0^s f_{T_n}(t) \left(\int_{s-t}^\infty f_{\tau_{n+1}}(u)\, du \right) dt$$

$$= \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \left(\int_{s-t}^\infty \lambda e^{-\lambda u}\, du \right) dt = \int_0^s \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \left(e^{-\lambda(s-t)} \right) dt$$

$$= \frac{\lambda^n}{(n-1)!}\, e^{-\lambda s} \int_0^s t^{n-1}\, dt = \frac{\lambda^n}{(n-1)!}\, e^{-\lambda s} \left(\frac{s^n}{n} \right) = \frac{(\lambda s)^n}{n!}$$

Poisson Process Example: Arrival of Patients

- Patients arrive at a rate of 1 every 10 minutes (on average)
- This doctor does not see patient until at least 3 are waiting

- What is the expected waiting time until the first patient is seen
    - Let $\lambda = \frac{1 \text{ patient arrival}}{10 \text{ minutes}} = \frac{1}{10}$
    - $\mathbb{E}[T_3] = \mathbb{E}[\tau_1 + \tau_2 + \tau_3] = 3\mathbb{E}[\tau_1] = 3 \cdot \frac{1}{\lambda} = 30$
- What is the probability that no patient is seen in the first hour?
    - $\mathbb{P}(N(60) < 3) = \sum_{t=0}^{2} \mathbb{P}(N(60) = t) = \sum_{t=0}^{2} e^{-6} \cdot \frac{6^t}{t!} \approx 0.062$

Poisson Process Example: Arrival of Customers

- Suppose customers arrive at a rate of 5 per hour, following a Poisson process
- Your store is open from 9am to 6pm
- What is the probability that no customer arrives within 1 hour of opening?
    - $\mathbb{P}(N(1) = 0) = e^{-\lambda \cdot 1} \cdot \frac{(\lambda \cdot 1)^0}{0!} = e^{-5}$
- What is the probability that we have 2 customers from 9-10am, 3 customers from 10-10:30am and 5 customers from 2-3:30pm?
    - Use the notation $N(t_1, t_2] \coloneqq N(t_2) - N(t_1)$
    - $\mathbb{P}(N(0,1] = 2, N(1,1.5] = 3, N(5,6.5] = 5)$

$$= \mathbb{P}(N(0,1] = 2)\mathbb{P}(N(1,1.5] = 3)\mathbb{P}(N(5,6.5] = 5)$$

$$= \left(e^{-\lambda} \cdot \frac{\lambda^2}{2!}\right)\left(e^{-0.5\lambda} \cdot \frac{(0.5\lambda)^3}{3!}\right)\left(e^{-1.5\lambda} \cdot \frac{(1.5\lambda)^5}{5!}\right) \approx 0.00197$$

- What is the probability that we have 3 customers from 10-10:30am, given 12 customers from 10am-12pm?
    - $\mathbb{P}(N(1,1.5] = 3 | N(1,3] = 12)$

$$= \frac{\mathbb{P}(N(1,1.5] = 3, N(1,3] = 12)}{\mathbb{P}(N(1,3] = 12)} = \frac{\mathbb{P}(N(1,1.5] = 3, N(1.5,3] = 9)}{\mathbb{P}(N(1,3] = 12)}$$

$$= \frac{\left(e^{-5 \cdot 0.5} \frac{(5 \cdot 0.5)^3}{3!}\right)\left(e^{-5 \cdot 1.5} \frac{(5 \cdot 1.5)^9}{9!}\right)}{e^{-5 \cdot 2} \frac{(5 \cdot 2)^{12}}{12!}} = \binom{12}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^9$$

    - Note this is a binomial distribution

Inhomogeneous Poisson Process

- $\{N(s) | s \geq 0\}$ is an **inhomogeneous Poisson process with rate $\boldsymbol{\lambda(r)}$** if it satisfies
    - $N(0) = 0$ with probability 1
    - $N(t)$ has independent increment
    - **$\boldsymbol{N(t) - N(s)}$ is Poisson distributed with mean $\boldsymbol{\int_s^t \lambda(r) dr}$**

Compound Poisson Process

Thursday, November 1, 2018 5:30 PM

Variations on Poisson Process

- Inhomogeneous Poisson Process
- Compound Poisson Process
- Thinning a Poisson Process
- Superposition of Poisson Process
- Conditioning for Poisson Process

Compound Poisson Process

- Motivating example: Risk Theory
  - Suppose claims arrive as a Poisson process $N(t)$ with rate λ
  - How much money must the company pay out over time
  - Let Y_k be the amount of money company pays for k^{th} claim
  - Let $S(t)$ be the amount of money company paid out up to time t
  - Then $S(t) = Y_1 + Y_2 + \cdots + Y_{N(t)} = \sum_{k=1}^{N(t)} Y_k$
- Motivating example: Stock Prices
  - Suppose a stock price has changes occurs as a Poisson Process $N(t)$ with rate λ
  - Let Y_k be the k^{th} change in stock price
  - Let $S(t)$ be the total price change up to time t
  - Then $S(t) = \sum_{k=1}^{N(t)} Y_k$
- Definition
  - Let $\{N(t) | t \geq 0\}$ be a Poisson process with rate λ, and $Y_1, \ldots, Y_k$ be iid RVs
  - A **Compound Poisson Process** is defined by
    - $\boldsymbol{S(t) = Y_1 + Y_2 + \cdots + Y_{N(t)} = \sum_{k=1}^{N(t)} Y_k}$
    - $\boldsymbol{S(t) = 0 \text{ when } N(t) = 0}$
  - Note: $S(t)$ is a sum of random length

Random Sum (Theorem 2.10)

- Let $Y_1, \ldots, Y_k$ be iid RVs, and N be an independent non-negative discrete RV
- Define $S = Y_1 + Y_2 + \cdots + Y_N$, and $S = 0$ if $N = 0$. Then
  - $\mathbb{E}[\boldsymbol{S}] = \mathbb{E}[\boldsymbol{N}]\boldsymbol{E}[\boldsymbol{Y_1}]$

- Note: $\mathbb{E}[S] = \mathbb{E}\left[\sum_{k=1}^{N} Y_k\right] \neq N\mathbb{E}[Y_1]$, since N is a random variable
- $\mathbb{E}[S|N = n] = \mathbb{E}\left[\sum_{k=1}^{n} Y_k \middle| N = n\right] = \sum_{k=1}^{n} \mathbb{E}[Y_k|N = n] = \sum_{k=1}^{n} \mathbb{E}[Y_k] = n\mathbb{E}[Y_1]$
- Therefore $\mathbb{E}[S|N] = N \cdot \mathbb{E}[Y_1]$
- $\mathbb{E}[S] = \mathbb{E}\big[\mathbb{E}[S|N]\big] = \mathbb{E}\big[\underbrace{N \cdot \mathbb{E}[Y_1]}_{N\times\text{constant}}\big] = \mathbb{E}[N] \cdot \mathbb{E}[Y_1]$

  - $\mathbf{Var}[\boldsymbol{S}] = \mathbb{E}[\boldsymbol{N}]\mathbf{Var}[\boldsymbol{Y_1}] + \mathbf{Var}[\boldsymbol{N}](\mathbb{E}[\boldsymbol{Y_1}])^2$
    - $\mathbb{E}[S^2|N = n] = \mathbb{E}[(Y_1 + \cdots + Y_n)^2]$

      $= \text{Var}[Y_1 + \cdots + Y_n] + (\mathbb{E}[Y_1 + \cdots + Y_n])^2$, since $E[X^2] = E[X]^2 + \text{Var}[X]$

      $= \text{Var}[Y_1] + \cdots + \text{Var}[Y_n] + (\mathbb{E}[Y_1] + \cdots + \mathbb{E}[Y_n])^2$, since $Y_1, \ldots, Y_n$ are iid

      $= n \cdot \text{Var}[Y_1] + n^2(\mathbb{E}[Y_1])^2$
    - Therefore $\mathbb{E}[S^2|N] = N \cdot \text{Var}[Y_1] + N^2(\mathbb{E}[Y_1])^2$
    - $\mathbb{E}[S^2] = \mathbb{E}\left[\mathbb{E}[S^2|N]\right]$

      $= \mathbb{E}[N \cdot \text{Var}[Y_1] + N^2(\mathbb{E}[Y_1])^2]$

      $= \mathbb{E}\big[\underbrace{N \cdot \text{Var}[Y_1]}_{N\times\text{constant}}\big] + \mathbb{E}\big[\underbrace{N^2(\mathbb{E}[Y_1])^2}_{N^2\times\text{constant}}\big]$

      $= \mathbb{E}[N] \cdot \text{Var}[Y_1] + \mathbb{E}[N^2](\mathbb{E}[Y_1])^2$
    - $\text{Var}[S] = \mathbb{E}[S^2] - (\mathbb{E}[S])^2$

      $= \left(\mathbb{E}[N] \cdot \text{Var}[Y_1] + \mathbb{E}[N^2](\mathbb{E}[Y_1])^2\right) - (\mathbb{E}[N] \cdot \mathbb{E}[Y_1])^2$

      $= \mathbb{E}[N] \cdot \text{Var}[Y_1] + \underbrace{\left(\mathbb{E}[N^2] - (\mathbb{E}[N])^2\right)}_{\text{Var}[N]}(\mathbb{E}[Y_1])^2$

      $= \mathbb{E}[N] \cdot \text{Var}[Y_1] + \text{Var}[N](\mathbb{E}[Y_1])^2$

- In particular, if $N \sim \text{Poisson}(\lambda)$, then
  - $\mathbf{Var}(\boldsymbol{S}) = \mathbb{E}[N] \cdot \text{Var}[Y_1] + \text{Var}[N](\mathbb{E}[Y_1])^2 = \lambda\text{Var}[Y_1] + \lambda(\mathbb{E}[Y_1])^2 = \boldsymbol{\lambda}\mathbb{E}[\boldsymbol{Y_1^2}]$
  - $\mathbb{E}[\boldsymbol{S}(\boldsymbol{t})] = \mathbb{E}[N(t)] \cdot \mathbb{E}[Y_1] = \boldsymbol{\lambda t}\mathbb{E}[\boldsymbol{Y_1}]$
  - $\mathbf{Var}[\boldsymbol{S}(\boldsymbol{t})] = \mathbb{E}[N(t)] \cdot \text{Var}[Y_1] + \text{Var}[N(t)](\mathbb{E}[Y_1])^2 = \lambda t\text{Var}[Y_1] + \lambda t(\mathbb{E}[Y_1])^2 = \boldsymbol{\lambda t}\mathbb{E}[\boldsymbol{Y_1^2}]$

Compound Poisson Process Example

- An insurance company pays claim at rate of 4 per week as a Poisson process
- The average payment for a claim is \$10,000. The standard deviation is \$6,000
- Find the mean and standard deviation of total payments for 4 weeks
- Given $E[Y_1] = 10000, \text{Var}[Y_1] = 6000^2 = 36000000, \lambda = 4$
- $\mathbb{E}[S(4)] = \lambda \cdot 4 \cdot \mathbb{E}[Y_1] = 4 \cdot 4 \cdot 10000 = 160000$
- $\text{Var}[S(4)] = \lambda \cdot 4 \cdot \mathbb{E}[Y_1^2] = \lambda \cdot 4 \cdot \left(\text{Var}[Y_1] + (\mathbb{E}[Y_1])^2\right)$

  $= 4 \cdot 4 \cdot \left(36000000 + (10000)^2\right) = 2.176 \times 10^9$
- $\text{SD}[S(4)] = \sqrt{\text{Var}[S(4)]} = 46647.6$

Thinning, Superposition, and Conditioning

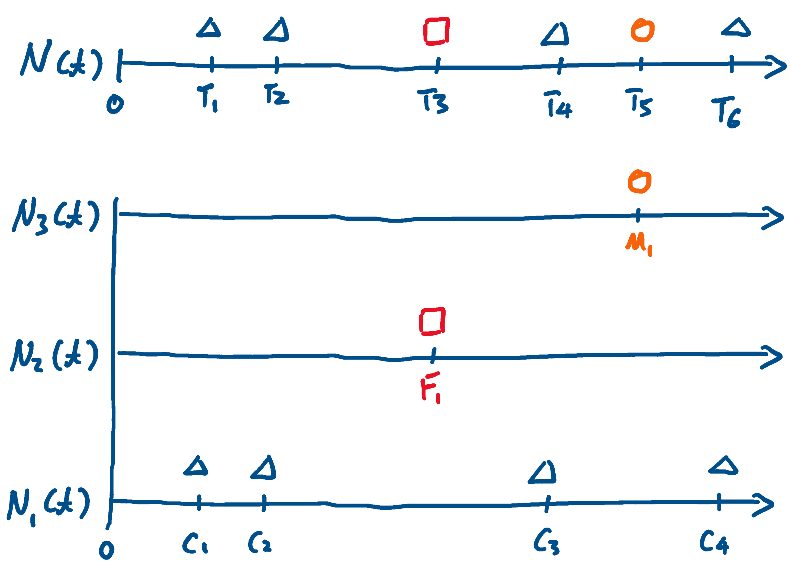
Tuesday, November 6, 2018 9:32 AM

General Idea for Thinning

- You have a Poisson process for arrivals, which are **filtered or categorized upon arrival**
- Not-so-surprising: The arrivals for a specific category form a **Poisson process**
- Surprising: The **process** for each category are **independent** of each other

Thinning Motivating Example: Highway Traffic

- Suppose vehicles pass a weigh station as a Poisson process with rate λ
- Let Y_k denote the type of the k^{th} vehicle that passes
- Assume that $\underbrace{\mathbb{P}(Y_k = 1) = 0.85}_{\text{Car}}, \underbrace{\mathbb{P}(Y_k = 2) = 0.10}_{\text{Truck}}, \underbrace{\mathbb{P}(Y_k = 3) = 0.05}_{\text{Motorcycle}}$



- General Idea: $N_1(t), N_2(t), N_3(t)$ will be **independent Poisson processes**

Thinning a Poisson Process (Theorem 2.11)

- Statement
  - Suppose $N(t)$ is a Poisson process with rate λ
  - Also, $Y_1, Y_2, \ldots$ are **iid** (and non-negative integer-valued) random variables
  - Define $N_j(t) = \sum_{k=1}^{N(t)} \mathbb{1}\{Y_k = j\}$ be the **number of arrivales up to time t of type j**
  - Then $N_1(t), N_2(t), \ldots$ are **independent Poisson process** with **rate $\lambda_j = \lambda \mathbb{P}(Y_1 = j)$**
- Proof (Binary Case)
  - Define $p = \mathbb{P}(Y_1 = 1)$ and $q = 1 - p = \mathbb{P}(Y_1 = 2)$
  - Claim: $N_1(t) \sim \text{Poisson}(p\lambda t)$ and $N_2(t) \sim \text{Poisson}(q\lambda t)$

- $\mathbb{P}(N_1(t)=j)=\sum_{n=j}^{\infty}\mathbb{P}(N_1(t)=j,N(t)=n)$

$$=\sum_{n=j}^{\infty}\underbrace{\mathbb{P}(N_1(t)=j|N(t)=n)}_{\sim\text{Binomial}(n,p)}\underbrace{\mathbb{P}(N(t)=n)}_{\sim\text{Poisson}(\lambda t)}$$

$$=\sum_{n=j}^{\infty}\binom{n}{j}p^j(1-p)^{n-j}e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

$$=e^{-\lambda t}\frac{p^j}{j!}\sum_{n=j}^{\infty}\frac{(\lambda t)^n(1-p)^{n-j}}{(n-j)!}$$

$$=e^{-\lambda t}\frac{p^j}{j!}\underbrace{\sum_{n=0}^{\infty}\frac{\big(\lambda t(1-p)\big)^n}{n!}}_{e^{(1-p)\lambda t}}(\lambda t)^j$$

$$=e^{-p\lambda t}\frac{(p\lambda t)^j}{j!}$$

- Therefore $N_1(t)\sim\text{Poisson}(p\lambda t)$, and similarly $N_2(t)\sim\text{Poisson}(q\lambda t)$

- Claim: $N_1(t)$ and $N_2(t)$ are independent
  - $\mathbb{P}(N_1(t)=j,N_2(t)=k)=\mathbb{P}(N_1(t)=j,N(t)=j+k)$

$$=\underbrace{\mathbb{P}(N_1(t)=j|N(t)=j+k)}_{\sim\text{Binomial}(j+k,p)}\underbrace{\mathbb{P}(N(t)=j+k)}_{\sim\text{Poisson}(\lambda t)}$$

$$=\binom{j+k}{j}p^jq^ke^{-\lambda t}\frac{(\lambda t)^{j+k}}{(j+k)!}$$

$$=\frac{(p\lambda t)^j}{j!}e^{-p\lambda t}\frac{(q\lambda t)^k}{k!}e^{-q\lambda t}$$

$$=\mathbb{P}(N_1(t)=j)\mathbb{P}(N_2(t)=k)$$

- Claim: $N_1(t)$ is a Poisson Process (same for $N_2(t)$)
  - Since $N_1(t)\le N(t)$, we have $\mathbb{P}(N_1(0)=0)=1$
  - In independence proof, we showed $N_1(t)\sim\text{Poisson}(p\lambda t)$
  - N_1 has independet increment
    - $N_1\big(t_j,t_{j+1}\big]=\sum_{k=N_1(t_j)+1}^{N_1(t_{j+1})}\mathbb{1}\{Y_k=1\}$
    - $N_1(t_1,t_2],\dots,N_1(t_{n-1},t_n]$ are sums independent random variables Y_k
    - Thus, the $N_1\big(t_j,t_{j+1}\big]$ will be independent for nonoverlapping intervals
  - Therefore $N_1(t)$ is a Poisson process with rate λ

Superposition of Poisson Processes (Theorem 2.13)

- Suppose $N_1(t),\dots,N_k(t)$ are independent Poisson process with rates $\lambda_1,\dots,\lambda_k$
- Then $\boldsymbol{N(t)=N_1(t)+\cdots+N_k(t)}$ **is a Poisson process** with rate $\boldsymbol{\lambda=\lambda_1+\cdots+\lambda_k}$
- The proof is like thinning theorem proof, but a little easier. Proceed by mathematical induction

Order Statistics

- Definition
  - Let $X_1, \dots, X_n$ be iid random variables
  - Define $X_{(i)}$ be the k-th smallest element in $\{X_1, \dots, X_n\}$
    - $X_{(1)} = \min\{X_1, \dots, X_n\}$
    - $X_{(2)} = \min\left(\{X_1, \dots, X_n\} \setminus \{X_{(1)}\}\right)$
    - $\vdots$
    - $X_{(n)} = \max\{X_1, \dots, X_n\}$
  - Then $X_{(1)}, \dots, X_{(n)}$ are the **order statistics** for $X_1, \dots, X_n$
- Fact
  - If $U_1, \dots, U_n \overset{iid}{\sim} \text{Unif}[0, t]$, then the joint PDF for $U_{(1)}, \dots, U_{(n)}$ is
  - $f(u_1, \dots, u_n) = \begin{cases} \frac{n!}{t^n} & 0 \le u_1 \le \cdots \le u_n \le t \\ 0 & \text{otherwise} \end{cases}$

Conditioning of Poisson Processes (Theorem 2.14)

- For a Poisson process, the **conditional distribution of arrival times** satisfies
  - $(\boldsymbol{T_1, \dots, T_n | N(t) = n}) \overset{\boldsymbol{D}}{=} \left(\boldsymbol{U_{(1)}, \dots, U_{(n)}}\right)$
- Specifically, the joint PDF given $N(t) = n$ is
  - $f(t_1, \dots, t_n) = \begin{cases} \frac{n!}{t^n} & 0 \le t_1 \le \cdots \le t_n \le t \\ 0 & \text{otherwise} \end{cases}$

Binomial and Conditioning of Poisson Processes (Theorem 2.15)

- Statement
  - Suppose $s < t$ and $0 \le k \le n$. Then
  - $\mathbb{P}(\boldsymbol{N(s) = k | N(t) = n}) = \boldsymbol{\binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}}$
  - In other words, $(N(s)|N(t) = n) \sim \text{Binomial}(n, s/t)$
- Proof (using order statistics)
- Proof (proceed directly from definition of condition probability)

Poisson Process Comprehensive Problems

Thursday, November 8, 2018 9:39 AM

Exercise 2.47

- Problem setup
  - $N_1(t) \coloneqq$ number of trucks that have passed up to time t
  - $N_2(t) \coloneqq$ number of cars that have passed up to time t
  - N_1 and N_2 are Poisson process with rate 40 and 100 respectively
  - 1/8 of trucks and 1/10 of cars go to Bojangle's
  - $B_1(t) \coloneqq$ number of trucks that have gone to Bojangle's up to time t
  - $B_2(t) \coloneqq$ number of cars that have gone to Bojangle's up to time t
  - Then B_1 and B_2 are Poisson process with rate 5 and 10 respectively
- Find the probability that exactly 6 trucks arrive at Bojangle's between noon and 1PM
  - $\mathbb{P}(B_1(1) = 6) = e^{-5}\dfrac{5^6}{6!}$
- Given that there were 6 truck arrivals at Bojangle's between noon and 1PM, what is the probability that exactly two arrived between 12:20 and 12:40?
  - $\mathbb{P}\left(B_1\left(\frac{1}{3},\frac{2}{3}\right] = 2 \middle| B_1(1) = 6\right) = \binom{6}{2}\left(\frac{2/3 - 1/3}{1}\right)^2\left(1 - \frac{2/3 - 1/3}{1}\right)^4 = \binom{6}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4$
- Suppose that trucks always have 1 passenger; 30% of the cars have 1 passenger, 50% have 2, and 20% have 4. Find the μ and σ^2 of the number of customers arrive at Bojangle's in one hour.
  - Define
    - $S_1(t) \coloneqq$ number of customers that arrive in trucks up to time t
    - $S_2(t) \coloneqq$ number of customers that arrive in cars up to time t
    - $Y_{1,k} \coloneqq$ number of passengers in k^{th} truck to arrive at Bojangle's
    - $Y_{2,k} \coloneqq$ number of passengers in k^{th} cars to arrive at Bojangle's
    - $S_l(t) \coloneqq \displaystyle\sum_{k=1}^{B_l(t)} Y_{l,k}$
    - $S(t) \coloneqq S_1(t) + S_2(t)$ to be total customers up to time t
  - Compute $\mathbb{E}[S(1)] = \mathbb{E}[S_1(1)] + \mathbb{E}[S_2(1)]$
    - $\mathbb{E}[S_1(1)] = \mathbb{E}[B_1(1)]\mathbb{E}[Y_{1,1}] = (5 \cdot 1) \cdot 1 = 5$
    - $\mathbb{E}[S_2(1)] = \mathbb{E}[B_2(1)]\mathbb{E}[Y_{2,1}] = (10 \cdot 1) \cdot (1 \times 0.3 + 2 \times 0.5 + 4 \times 0.2) = 21$
    - $\Rightarrow \mathbb{E}[S(1)] = \mathbb{E}[S_1(1)] + \mathbb{E}[S_2(1)] = 26$
  - Compute $\operatorname{Var}[S(1)] = \operatorname{Var}[S_1(1)] + \operatorname{Var}[S_2(1)]$ (by independence)
    - $\operatorname{Var}[S_1(1)] = 5\mathbb{E}[Y_{1,1}^2] = 5$

        - $\text{Var}[S_2(1)] = 10\mathbb{E}[Y_{2,1}^2] = 10(1^2 \times 0.3 + 2^2 \times 0.5 + 4^2 \times 0.2) = 55$
        - $\Rightarrow \text{Var}[S(1)] = \text{Var}[S_1(1)] + \text{Var}[S_2(1)] = 60$

Exercise 2.27

- Problem setup
    - The next bus arrival time is uniformly distributed over the next hour
    - Cars pass at a rate of 6 per hour (following a Poisson process)
    - 1/3 of car will pick up a hitchhiker
- Define
    - T_B ≔ time bus arrives, then $T_B \sim \text{Unif}[0,1]$
    - $N(t)$ ≔ the number of car passed up to time t, then $N(t)$ is a Poisson process with $\lambda = 6$
    - $H(t)$ ≔ the number of car pick up a hitchhiker up to time t, then $H(t)$ is a P.P. with $\lambda = 2$
    - T_1 ≔ arrival time for first car that will pick up a hitchhiker, then $T_1 \sim \text{Exp}(2)$
- What is the probability someone takes the bus rather than hitchhikes?
    - $\mathbb{P}(T_B < T_1) = \int_0^1 \int_y^\infty f_{T_1}(x) f_{T_B}(y)\, dx\, dy = \int_0^1 \int_y^\infty 2e^{-2x}\, dx\, dy = \frac{1}{2}\left(1 - \frac{1}{e^2}\right)$

Exercise 2.50

- Problem setup
    - $N(t)$ ≔ number of typos author has made **in the first *t* pages**
    - $N_f(t)$ ≔ number of typos found in the first t pages
    - Then $N(t), N_f(t)$ are Poisson processes with rate λ and 0.9λ respectively
    - X ≔ number of typos found in full manuscript, then $X = N_f(200)$
- Compute the expected number of typos
    - $\mathbb{E}[X] = \mathbb{E}[N_f(200)] = 200 \cdot 0.9\lambda = 180\lambda$
- Estimate λ if the total number of typos is 108
    - $180\lambda \approx 108 \Rightarrow \hat{\lambda} = \frac{108}{180} = 0.6$

More Exercises on Poisson Process

Tuesday, November 13, 2018 9:36 AM

Exercise 2.45

- Problem setup
  - Signals are sent as a Poisson process with rate λ
  - Each signal reaches its target with probability p and fails with probability $q = 1 - p$
  - $N_1(t) :=$ # successful transimissions up to time t
  - $N_2(t) :=$ # failed transimissions upto time t
- Find the distribution of $\left(N_1(t), N_2(t)\right)$
  - This is asking for the joint PMF of $N_1(t), N_2(t)$
  - $N_1(t)$ and $N_1(t)$ are thinned versions of the general singal process
  - So $N_1(t)$ and $N_2(t)$ are Poisson proecss with rates $p\lambda$ and $(1-p)\lambda$, respectively
  - Additionally, $\boldsymbol{N_1(t)}$ **and** $\boldsymbol{N_2(t)}$ **are independent**
  - $\mathbb{P}(N_1(t) = j, N_2(t) = k) = \mathbb{P}(N_1(t) = j)\mathbb{P}(N_2(t) = k)$

$$= \left[e^{-p\lambda t}\frac{(p\lambda t)^j}{j!}\right]\left[e^{-(1-p)\lambda t}\frac{\left((1-p)\lambda t\right)^k}{k!}\right] = e^{-\lambda t}\frac{(p\lambda t)^j\left((1-p)\lambda t\right)^k}{j!\,k!}$$

- $L :=$ # signals lost before the first success. Find the distribution of L
  - We can compute $\mathbb{P}(L \geq k)$, then $\mathbb{P}(L = k) = \mathbb{P}(L \geq k) - \mathbb{P}(L \geq k+1)$
  - $F_k :=$ time of k^{th} failed signal, $S_k :=$ time of k^{th} successful signal
  - $\mathbb{P}(L \geq k) = \mathbb{P}(F_k < S_1) = \int_0^\infty f_{F_k}(t)\int_t^\infty f_{S_1}(s)ds\,dt$

$$= \int_0^\infty q\lambda e^{-q\lambda t}\frac{(q\lambda t)^{k-1}}{(k-1)!}\underbrace{\left(\int_t^\infty p\lambda e^{-p\lambda s}\,ds\right)}_{e^{-p\lambda s}}dt = \int_0^\infty q\lambda e^{-\lambda t}\frac{(q\lambda t)^{k-1}}{(k-1)!}dt$$

$$= q^k\int_0^\infty \underbrace{\lambda e^{-\lambda t}\frac{(\lambda t)^{k-1}}{(k-1)!}}_{\text{Gamma Dist.}}dt = q^k$$

  - $\mathbb{P}(L = k) = \mathbb{P}(L \geq k) - \mathbb{P}(L \geq k+1) = q^k(1-q) = (1-p)^k p$
  - So $L \sim \text{Geometric}(p)$
  - Note: $\{\boldsymbol{L = k}\} = \{$**First** $\boldsymbol{k}$ **transimissions fail**, $\boldsymbol{k+1}$ **transimission succeeds**$\}$

Examples of Conditional Poisson Process

- $N(t)$ is a Poisson process with rate λ
- Recall that the PDF of $(T_1, \dots, T_n | N(t) = n)$ is $f(t_1, \dots, t_n) = \begin{cases} \frac{n!}{t^n} & 0 \leq t_1 \leq \cdots \leq t_n \leq t \\ 0 & \text{otherwise} \end{cases}$
- Compute $\mathbb{E}[T_1 | N(1) = 2]$

- $\mathbb{E}[T_1|N(1)=2] = \int_0^1 \int_0^{t_2} t_1 \cdot \frac{2!}{1^2} dt_1\, dt_2 = \int_0^1 t_2^2\, dt_2 = \frac{1}{3}$

- Compute $\mathbb{E}[T_1 T_2|N(1)=2]$
  - $\mathbb{E}[T_1 T_2|N(1)=2] = \int_0^1 \int_0^{t_2} t_1 t_2 \cdot \frac{2!}{1^2} dt_1\, dt_2 = \int_0^1 t_2^3\, dt_2 = \frac{1}{4}$
- Compute $\mathbb{E}[T_2|N(4)=3]$
  - $\mathbb{E}[T_2|N(4)=3] = \int_0^4 \int_0^{t_3} \int_0^{t_2} t_2 \cdot \frac{3!}{4^3} dt_1 dt_2 dt_3 = \int_0^4 \int_0^{t_3} t_2^2 \cdot \frac{3!}{4^3} dt_2 dt_3 = \int_0^4 \frac{2!}{4^3} t_3^3\, dt_3 = 2$
- Compute $\mathbb{E}[T_1|N(1)=n]$
  - Let $U_1, \dots, U_n \overset{iid}{\sim} \text{Unif}([0,1])$ and define $T = \min\{U_1, \dots, U_n\}$, then $E[T_1|N(1)=n] = \mathbb{E}[T]$
  - $F_T(t) = 1 - \mathbb{P}(T > t) = 1 - \mathbb{P}(U_1 > t, \dots, U_n > t) = 1 - (1-t)^n \Rightarrow f_T(t) = n(1-t)^{n-1}$
  - $E[T_1|N(1)=n] = \mathbb{E}[T] = \int_0^1 tn(1-t)^{n-1} dt = \frac{1}{n+1}$
  - Alternatively, $\mathbb{E}[T_1|N(1)=n] = \int_0^1 \int_0^{t_n} \cdots \int_0^{t_3} \int_0^{t_2} t_1 \frac{n!}{1^n} dt_1 dt_2 \cdots dt_{n-1} dt_n$

$$= \int_0^1 \int_0^{t_n} \cdots \int_0^{t_3} \frac{n!}{2!} t_2^2 dt_2 \cdots dt_{n-1} dt_n$$

$$= \int_0^1 \int_0^{t_n} \cdots \int_0^{t_4} \frac{n!}{3!} t_3^3 dt_3 \cdots dt_{n-1} dt_n = \cdots$$

$$= \int_0^1 \frac{n!}{n!} t_n^t\, dt_n = \left[\frac{1}{n+1} t_n^{n+1}\right]_{t_n=0}^{t_n=1} = \frac{1}{n+1}$$

Introduction to Renewal Process

Tuesday, November 20, 2018 9:30 AM

Renewal Process

- Renewal process is more general than Poisson process
- The structure is the same as a Poisson process, but we **do not assume $\boldsymbol{\tau_i \sim \mathrm{Exp}(\lambda)}$**
- We use the notation $\boldsymbol{t_1, t_2, \ldots \overset{iid}{\sim} F}$ where F is a CDF for a non-negative distribution
- With very few assumptions, it is difficult to say much in general

Arrival Law of Large Numbers

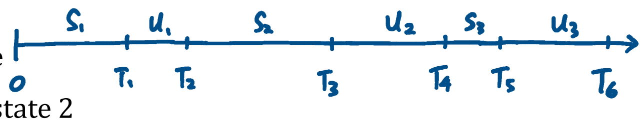
- Statement
  - Let $\mu = \mathbb{E}[t_i]$ be the mean interarrival
  - **If $\mathbb{P}(t_i > 0) > 0$ then $\dfrac{N(t)}{t} \to \dfrac{1}{\mu}$ as $t \to \infty$**
- Recall Strong Law of Large Numbers
  - If $X_1, X_2, \ldots \overset{iid}{\sim} F$ with $\mathbb{E}[X_1] = \mu_F$, then $\dfrac{X_1 + X_2 + \cdots + X_n}{n} \to \mu_F$ as $n \to \infty$
- Proof
  - Using the strong law of large numbers $\lim_{t\to\infty} \dfrac{T_{N(t)}}{N(t)} = \lim_{t\to\infty} \dfrac{t_1 + \cdots + t_{N(t)}}{N(t)} \to \mu$
  - Also, we know that $T_{N(t)} \le t < T_{N(t)+1}$
  - Therefore, $\underbrace{\dfrac{\mathrm{T}_{N(t)}}{N(t)}}_{\to\mu} \le \dfrac{t}{N(t)} < \dfrac{T_{N(t)+1}}{N(t)} = \underbrace{\dfrac{T_{N(t)+1}}{N(t)+1}}_{\to\mu} \cdot \underbrace{\dfrac{N(t)+1}{N(t)}}_{\to 1}$
  - As $t \to \infty, \mu \le \lim_{t\to\infty} \dfrac{t}{N(t)} \le \mu \cdot 1 = \mu$
  - Therefore $\lim_{t\to\infty} \dfrac{N(t)}{t} = \dfrac{1}{\mu}$

Renewal Reward Process

- Idea
  - With each arrival, there is an **associated reward (or cost)**
- Notation
  - $r_k =$ value/cost of k^{th} arrival
  - $N(t) =$ number of arrivals up to time t
  - $\boldsymbol{R(t) = \sum_{k=1}^{N(t)} r_k =}$ **cumulative reward up to time $\boldsymbol{t}$**
- Key assumptions
  - $(r_1, t_1), (r_2, t_2), \ldots$ is an iid sequence of rewards and waiting times

- Reward/Cumulative Law of Large Number: $\boldsymbol{\frac{R(t)}{t} \to \frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]}}$ **as** $\boldsymbol{t \to \infty}$
  - $\frac{R(t)}{N(t)} = \frac{1}{N(t)} \sum_{k=1}^{N(t)} r_k \to \mathbb{E}[r_i]$ as $t \to \infty$ by law of large numbers
  - $\frac{R(t)}{t} = \frac{R(t)}{N(t)} \cdot \frac{N(t)}{t} = \mathbb{E}[r_i] \cdot \frac{1}{\mathbb{E}[T_i]} = \frac{\mathbb{E}[r_i]}{\mathbb{E}[T_i]}$ as $t \to \infty$ by arrival LLN

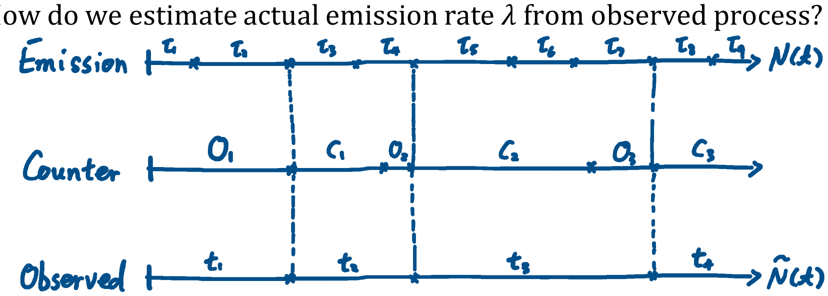
Alternating Renewal Process



- For the graph on the right, we have
  - s_1 time in state 1, u_1 time in state 2
  - s_2 time in state 3, u_2 time in state 4, and so on.
  - $s_1, s_2, \ldots \overset{iid}{\sim} F$ and $u_1, u_2, \ldots \overset{iid}{\sim} G$
  - $s_1, u_1, s_2, u_2, \ldots$ are independent
- Alternating renewal LLN
  - The **long-run fraction of time spent in state 1 is** $\boldsymbol{\frac{\mu_F}{\mu_F + \mu_G}}$
  - Reframe as a renweal reward process with $t_k = s_k + u_k$ and $r_k = s_k$
  - Then $R(t) = \sum_{k=1}^{N(t)} r_k = \sum_{k=1}^{N(t)} s_k =$ total time spent in state 1 up to time t
  - Therefore, $\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]} = \frac{\mu_F}{\mu_F + \mu_G}$

Application: Geiger Counter

- Problem background
  - Radioactive particles are emitted as a Poisson process with unknown rate λ
  - Geiger counter locks for a random amount of time when a particle registers
  - Then it opens and waits for next particle
- Two processes: particle emission and particle observation
- How do we estimate actual emission rate λ from observed process?



  - $O_k \overset{iid}{\sim} \text{Exp}(\lambda)$ and independent of $C_1, C_2, \ldots$
  - Set $\gamma_t = \frac{\tilde{N}(t)}{t}$, then for large values t, we can use arrival LLN
  - $\gamma_t \approx \frac{1}{\mathbb{E}[t_k]} = \frac{1}{\mathbb{E}[C_k] + \mathbb{E}[O_{k+1}]} = \frac{1}{\mathbb{E}[C_k] + \lambda^{-1}} \Rightarrow \boldsymbol{\hat{\lambda} = \frac{\gamma_t}{1 - \gamma_t \mathbb{E}[C_1]}}$

Renewal Process, Age and Residual Life

Tuesday, November 27, 2018 9:32 AM

Review: LLN for Renewal Process

- Renewal process: Like a Poisson process, but **waiting time t_k do not have to be Exp(λ)**
- Arrival LLN: $\lim\limits_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$, where $\mu = \mathbb{E}[t_i]$
- Reward LLN
  - Let r_i = reward/cost of i-th renewal, and $R(t) = \sum_{i=1}^{N(t)} r_i$, then, $\lim\limits_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]}$
- Alternating LLN
  - Let $s_1, s_2, \ldots$ be the times in state 1, and $u_1, u_2, \ldots$ be times in state 2
  - Then the limiting fraction of time spent in state 1 is $\frac{\mathbb{E}[s_i]}{\mathbb{E}[s_i] + \mathbb{E}[u_i]}$

Exercise 3.2: Alternating Renewal Process

- Let $J_1, J_2, \ldots$ be the length of jobs, and $S_1, S_2, \ldots$ be the time she spends between jobs
- Given that $\mathbb{E}[J_k] = 11$ and $S_k \sim \text{Exp}[1/3]$, what fraction of Monica's life will be work?
- This is an **alternating renewal process** where state 1 is "Monica is employed"
- By the Alternating LLN, Monica will work $\frac{\mathbb{E}[J_k]}{\mathbb{E}[J_k] + \mathbb{E}[S_k]} = \frac{11}{11+3} = \frac{11}{14}$ of the time

Exercise 3.4: Renewal Reward Process

- Taxi customers arrive to the stand independently, with interarrival times $t_k \sim F$
- The amount each customer pays r_k follows a distribution G
- What is the long-run amount of money per unit time that taxis at the stand collect
- Let $R(t) = \sum_{k=1}^{N(t)} r_k$ = total fares collected up to time t, then we want to find $\lim\limits_{t\to\infty} \frac{R(t)}{t}$
- By the **Renewal Reward LLN**, $\lim\limits_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]} = \frac{\mu_G}{\mu_F}$

Example 3.4: Renewal Reward Process

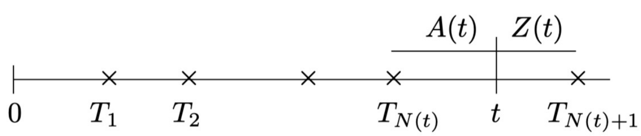
- The lifetime of a car follows some continuous distribution with density function h
- Mr. Brown's policy:
  - If the car breaks, buy a new one for \$A, and repair for \$B
  - If the car survives to time T, buy a new one for \$A
- What is the long-run average cost per unit time of this policy?
- This is a **renewal process where the renewal is buying a new car**

th

- Let t_i be time between car purchases and r_i be cost of buying i^{th} car
- Then by the **reward LLN**, the **long-run cost per unit time is** $\frac{\mathbb{E}[\boldsymbol{r_i}]}{\mathbb{E}[\boldsymbol{t_i}]}$
- Let $s_i \sim h$ be the lifetime of i^{th} car, then $\boldsymbol{t_i = \min\{s_i, T\}}$
- $\mathbb{E}[t_i] = \mathbb{E}[\min\{s_i, T\}] = \int_0^\infty \min\{s, T\}\, h(s)\, ds = \int_0^T s \cdot h(s) ds + T \int_T^\infty h(s) ds$
- $\mathbb{E}[r_i] = (A + B)\mathbb{P}(s_i < T) + A \cdot \mathbb{P}(s_i \geq T) = A + B \cdot \mathbb{P}(s_i < T) = A + B \int_0^T h(s) ds$
- Therefore, $\frac{\mathbb{E}[r_i]}{\mathbb{E}[t_i]} = \frac{A + B \int_0^T h(s) ds}{\int_0^T s \cdot h(s) ds + T \int_T^\infty h(s) ds}$
- Challenging follow-up: use this solution to choose optimal value of replacement time T

Age and Residual Life

- Introduction



  - $\boldsymbol{A(t)}$ **= age = time since last renewal =** $\boldsymbol{t - T_{N(t)}}$
  - $\boldsymbol{Z(t)}$ **= residual life = time until next renewal =** $\boldsymbol{T_{N(t)+1} - t}$
- What is the limiting distribution for $A(t)$ and $Z(t)$?
  - Consider a renewal process with continuous waiting times between renewals

1. Let $x, y \geq 0$ be fixed values

   Let $\boldsymbol{R(t)}$ **be the total time up to** $\boldsymbol{t}$ **for which age** $> \boldsymbol{x}$ **and residual life** $> \boldsymbol{y}$, then

   $$\lim_{t\to\infty} \mathbb{P}(A(t) > x, Z(t) > y) = \lim_{t\to\infty} \frac{R(t)}{t}$$

   $$= \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}\{A(s) > x, Z(s) > y\} ds = \frac{1}{\mathbb{E}[t_i]}\int_{x+y}^\infty \mathbb{P}(t_i > z) dz$$

2. Thus, $\lim_{t\to\infty} \mathbb{P}(Z(t) > y) = \frac{1}{\mathbb{E}[t_i]}\int_y^\infty \mathbb{P}(t_i > z) dz$

   So the **limiting PDF of** $\boldsymbol{Z(t)}$ **is** $\boldsymbol{g(z)} = \frac{\mathbb{P}(\boldsymbol{t_i > z})}{\mathbb{E}[\boldsymbol{t_i}]}$ for $z \geq 0$, and same for $A(t)$

3. The **limiting expected value of** $\boldsymbol{A(t)}$ **and** $\boldsymbol{Z(t)}$ **is** $\frac{\mathbb{E}[\boldsymbol{t_i^2}]}{2\mathbb{E}[\boldsymbol{t_i}]}$

4. If $t_k \sim f$ then the **limiting joint PDF of** $\boldsymbol{A(t)}$ **and** $\boldsymbol{Z(t)}$ **is** $\frac{\boldsymbol{f(a+z)}}{\mathbb{E}[\boldsymbol{t_i}]}$ for $a, z \geq 0$

- Example
  - Given $t_i \sim \text{Gamma}(2, \lambda)$, what is limiting density for $A(t)$?
  - $g(z) = \frac{\mathbb{P}(t_1 > z)}{\mathbb{E}[t_1]} = \frac{1}{2/\lambda}\int_z^\infty \lambda e^{-\lambda t}\frac{(\lambda t)^{2-1}}{(2-1)!} dt = \frac{\lambda}{2} e^{-\lambda z}(\lambda z + 1)$ for $z \geq 0$

Continuous Time Markov Processes

Thursday, November 29, 2018 9:34 AM

Continuous Time Markov Processes

- We say that X_t with $t > 0$ is a continuous time Markov process if
- For any time $0 \le s_0 < \cdots < s_n < s$, and any states $j, i, i_n, \dots, i_0$, we have
- $\mathbb{P}\left(\boldsymbol{X_{s+t} = j \middle| X_s = i, X_{s_n} = i_n, \dots, X_{s_0} = i_0}\right) = \mathbb{P}(\boldsymbol{X_{s+t} = j | X_s = i}) = \mathbb{P}(\boldsymbol{X_t = j | X_0 = i})$
- The equation above is called the **(continuous) Markov property**
- We denote the **transition probability** $\mathbb{P}(\boldsymbol{X_t = j | X_0 = i})$ **by** $\boldsymbol{p_t(i, j)}$

Poisson Process is Markovian

- Change $N(0)$ to be some starting number of points. Then
- $\mathbb{P}(\boldsymbol{N(s+t) = j | N(s) = i, N(s_n) = i_n, \dots, N(s_0) = i_0})$

$$= \frac{\mathbb{P}(N(s+t) = j, N(s) = i, N(s_n) = i_n, \dots, N(s_0) = i_0)}{\mathbb{P}(N(s) = i, N(s_n) = i_n, \dots, N(s_0) = i_0)}$$

$$= \frac{\mathbb{P}(N(s_0) = i_0, N(s_0, s_1] = i_1 - i_0, \dots, N(s_n, s] = i - i_n, N(s, s+t] = j - i)}{\mathbb{P}(N(s_0) = i_0, N(s_0, s_1] = i_1 - i_0, \dots, N(s_n, s] = i - i_n)}$$

$$= \mathbb{P}(N(s, s+t] = j - i) \cdot \frac{\mathbb{P}(N(s) = i)}{\mathbb{P}(N(s) = i)}$$

$$= \frac{\mathbb{P}(N(s, s+t] = j - i, N(s) = i)}{\mathbb{P}(N(s) = i)}$$

$$= \frac{\mathbb{P}(N(s+t) = j, N(s) = i)}{\mathbb{P}(N(s) = i)}$$

$$= \mathbb{P}(\boldsymbol{N(s+t) = j | N(s) = i})$$

Construction from a Discrete Time Markov Chain

- Procedure
  - Suppose $Y_0, Y_1, \dots$ is a DTMC with transition probability $u(i, j)$
  - Let $N(t)$ be a Poisson process with rate λ
  - Then $\boldsymbol{X_t = Y_{N(t)}}$ **is a continuous time Markov chain**
- Intuition
  - Transitions occur at random times according to the Poisson process
- Significance
  - This gives one general procedure for constructing continuous time Markov chain

Chapman–Kolmogorov Equation

- Equation
  - $p_{s+t}(i, j) = \sum_{k \in S} p_s(i, k) p_t(k, j)$

- Proof
  - $p_{s+t}(i,j) = \mathbb{P}(X_{s+t} = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_{s+t} = j, X_s = k | X_0 = i)$

    $= \sum_{k \in S} \underbrace{\mathbb{P}(X_{s+t} = j | X_0 = i)}_{p_t(k,j)} \underbrace{\mathbb{P}(X_s = k | X_0 = i)}_{p_s(i,k)} = \sum_{k \in S} p_s(i,k) p_t(k,j)$
- Importance
  - Suppose we know $p_t(i,j)$ for all $t \in [0, t_0)$
  - Then for all $s \in [t_0, 2t_0)$, we have $\boldsymbol{p_s(i,j) = p_{s/2+s/2}(i,j) = \sum_{k \in S} p_{s/2}(i,k) p_{s/2}(k,j)}$
  - Thus **for arbitrarily small $\boldsymbol{t_0}$, we can always find $\boldsymbol{p_s(i,j)}$** for all $s \geq t_0$

Jump Rates

- Definition
  - For any states $i \neq j$, the jump rate from i to j is defined as $q_{ij} \coloneqq \lim_{h \to 0} \frac{p_h(i,j)}{h}$
- Example of CTMCs constructed from DTMC
  - $q_{ij} = \lim_{h \to 0} \frac{p_h(i,j)}{h} = \lim_{h \to 0} \left[\frac{1}{h} \sum_{n=0}^{\infty} e^{-\lambda h} \frac{(\lambda h)^n}{n!} u^n(i,j) \right]$

    $= \lim_{h \to 0} \left[\lambda e^{-\lambda h} u(i,j) + \sum_{n=1}^{\infty} \frac{\lambda^n h^{n-1}}{n!} u^n(i,j) \right] = \lambda u(i,j)$
  - Note that the jump rate q_{ij} is the **rate for a thinned Poisson process**

Construction From Jump Rates

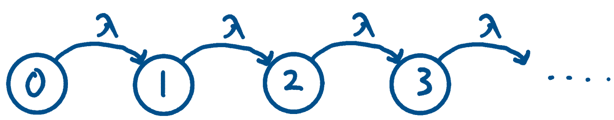
- Procedure
  - Suppose we know $q(i,j)$ for all states $i \neq j$
  - Define $\boldsymbol{\lambda(i) = \sum_{j \neq i} q(i,j)}$ to be the **rate at which the MC leaves $\boldsymbol{i}$**
  - Define $\boldsymbol{r(i,j) = \frac{q(i,j)}{\lambda_i}}$ **with $\boldsymbol{r(i,i) = 0}$** to be the **transition probability** from i to j
  - Let **$\boldsymbol{Y_0, Y_1}, \ldots$ be a DTMC** with transition matrix $r(i,j)$, and $\boldsymbol{\tau_0, \tau_1, \ldots \overset{iid}{\sim} \mathrm{Exp}(1)}$
  - Define $t_i = \frac{\tau_i}{\lambda(Y_{i-1})} \sim \mathrm{Exp}\big(\lambda(Y_{i-1})\big)$, and $T_i = \sum_{n=0}^{i} t_n$, for $i \geq 0$
  - Set $X_t = Y_{i-1}$ for $T_{i-1} \leq t < T_i$, then X_t is a CTMC
- Caveat
  - $\lim_{n \to \infty} T_n = T_\infty$ could be finite, then X_t is only defined for $0 \leq t < T_\infty$
  - One fix is to **set $\boldsymbol{X_t = \Delta}$ (cemetery state) for $\boldsymbol{t \geq T_\infty}$**

M/M/s Queue, Kolmogorov Equations

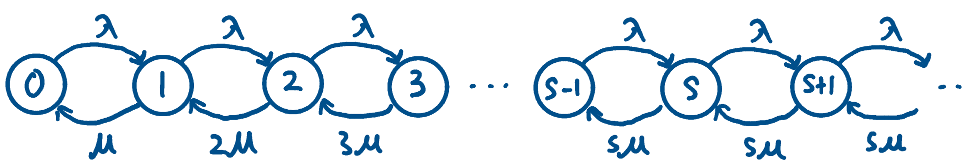
Tuesday, December 4, 2018 11:20 AM

CTMCs Constructed from Jump Rates

- Poisson process
  - Waiting time between customers is an Exp(λ) random variable
  - As a CTMC, the state space is $S = \{0,1,2,\ldots\}$



  - The jump rates are $\begin{cases} q(n, n+1) = \lambda & \forall n \in S \\ q(i,j) = 0 & j \neq i+1 \end{cases}$
- M/M/s Queue
  - A line of customers is being helped by $\boldsymbol{s}$ **servers**
  - Customers **arrive** as a **Poisson process with rate $\boldsymbol{\lambda}$**
  - Each server requires an **Exp($\boldsymbol{\mu}$) of time to serve** their customer
  - $X(t) :=$ #Customers in system (being served and in line) at time t



  - The **jump rates are** $\begin{cases} \boldsymbol{q(n, n+1) = \lambda} & \boldsymbol{n \geq 0} \\ \boldsymbol{q(n, n-1) = n\mu} & \boldsymbol{1 \leq n < s} \\ \boldsymbol{q(n, n-1) = s\mu} & \boldsymbol{n \geq s} \end{cases}$

Kolmogorov Equations

- Motivation
  - How do we get $p_t(i,j)$ from the transition rates $q(i,j)$
- Kolmogorov equations (coordinate form)
  - Define $\boldsymbol{\lambda_i} = \sum_{k \neq i} \boldsymbol{q_{ik}}$ **to be the rate out of state $\boldsymbol{i}$**
  - **Backward**: $\frac{d}{dt}[p_t(i,j)] = \sum_{k \neq i} q(i,k)p_t(k,j) - \lambda_i p_t(i,j)$
  - **Forward**: $\frac{d}{dt}[p_t(i,j)] = \sum_{k \neq i} p_t(i,k)q(k,j) - p_t(i,j)\lambda_j$
- Kolmogorov equations (matrix form)
  - Define the transition rate matrix (or jump rate matrix) Q as

- $\boldsymbol{Q_{ij} = \begin{cases} q_{ij} & \text{if } i \neq j \\ -\lambda_i & \text{if } i = j \end{cases} \Leftrightarrow Q = \begin{bmatrix} -\lambda_1 & q(1,2) & q(1,3) & \cdots \\ q(2,1) & -\lambda_2 & q(2,3) & \cdots \\ q(3,1) & q(3,2) & -\lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}$
  - Then we have $\begin{cases} \textbf{Backward: } \frac{d}{dt}[p_t] = Qp_t \\ \textbf{Forward: } \frac{d}{dt}[p_t] = p_tQ \end{cases}$
- Why we need Kolmogorov equations
  - Given the transition rates $q(i,j)$, we can find $p_t(i,j)$ by solving the ODEs
- Is matrix or coordinate form better?
  - Matrix form is nice for general proofs and theory
  - Coordinate form is nice for specific examples, especially when most $q(i,j) = 0$

Solving Forward Kolmogorov Equations

- Claim: $\boldsymbol{e^{tQ}}$ **solves Forward Kolmogorov equation**
  - $\frac{de^{tQ}}{dt} = \frac{d}{dt}\left[\sum_{n=0}^{\infty} \frac{(tQ)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{d}{dt}\left[\frac{(tQ)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{t^{n-1}Q^n}{(n-1)!} = Q\sum_{n=1}^{\infty} \frac{(tQ)^{n-1}}{(n-1)!} = Qe^{tQ}$
- The initial condition is $p_0 = I$, because
  - $p_0(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$
  - $e^{0Q} = \sum_{n=0}^{\infty} \frac{(0Q)^n}{n!} = \frac{(0Q)^0}{0!} = I$
- Why not always use $p_t = e^{tQ}$ for all CTMSs?
  - **Matrix exponentials** are **hard to compute**, especially for infinite state space

Derivation of Forward Kolmogorov Equations

- $\frac{d}{dt}[p_t(i,j)] = \lim_{h\to 0} \frac{p_{t+h}(i,j) - p_t(i,j)}{h}$

$$= \lim_{h\to 0} \frac{1}{h}\left[\sum_{k\in S} p_t(i,k)p_h(k,j) - p_t(i,j)\right]$$

$$= \lim_{h\to 0} \frac{1}{h}\left[\sum_{k\neq j} p_t(i,k)p_h(k,j) + p_t(i,j)p_h(j,j) - p_t(i,j)\right]$$

$$= \lim_{h\to 0} \frac{1}{h}\left[\sum_{k\neq j} p_t(i,k)p_h(k,j) - p_t(i,j)\big(1 - p_h(j,j)\big)\right]$$

$$= \lim_{h\to 0} \frac{1}{h}\left[\sum_{k\neq j} p_t(i,k)p_h(k,j) - p_t(i,j)\sum_{k\neq j} p_h(j,k)\right]$$

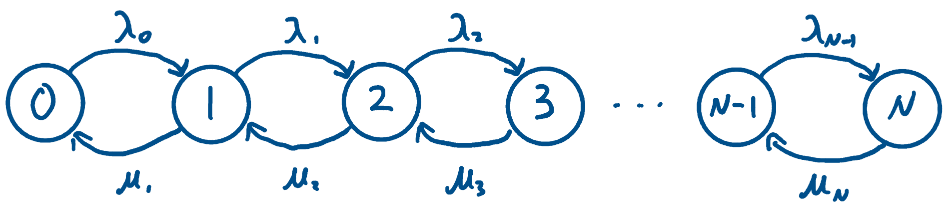
$$= \lim_{h\to 0} \frac{1}{h}\left[\sum_{k\neq j} p_t(i,k)p_h(k,j)\right] - p_t(i,j)\lim_{h\to 0} \frac{1}{h}\left[\sum_{k\neq j} p_h(j,k)\right]$$

$$= \sum_{k \neq j} p_t(i,k) \lim_{h \to 0} \frac{p_h(k,j)}{h} - p_t(i,j) \sum_{k \neq j} q(j,k)$$

$$= \sum_{k \neq j} p_t(i,k) q(k,j) - p_t(i,j) \lambda_j$$

Example: Birth and Death Processes

- The state space is $S = \{0,1,2,\dots,N\}$
- Only nonzero rates are $\begin{cases} q(n,n+1) = \lambda_n \\ q(n,n-1) = \mu_n \end{cases}$
- Note the conflict in notation. Usually $\lambda_n = \sum_{k \neq n} q_{nk} = q_{nn}$



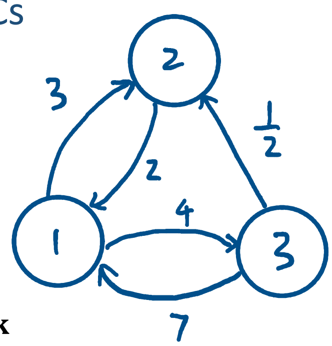
- Kolmogorov equations
  - $p_t'(i,j) = p_t(i,j-1)\lambda_{j-1} + p_t(i,j+1)\mu_{j+1} - p_t(i,j)(\lambda_j + \mu_j), \forall j = 1,\dots,N-1$
  - $p_t'(i,0) = p_t(i,1)\mu_1 - p_t(i,0)\lambda_0$
  - $p_t'(i,N) = p_t(i,N-1)\lambda_{N-1} - p_t(i,N)\mu_N$

Properties of CTMC

Thursday, December 6, 2018 9:37 AM

Intuitive View of CTMCs

- Transition graph



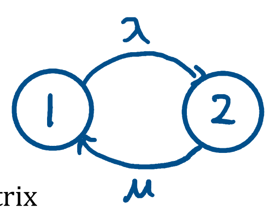
- **Exponential alarm clock**
    - An alarm clock that **goes off after a random Exp amount of time**
- Explanation on transition graph
    - Each **edge** in the graph represents an **exponential clock** with the edge weight as rate
    - When you **land in a new state**, the **clocks on the out edges begin**
    - Then your CTMC **takes the path of the clock that goes off first**

Foundational Work

- Make this informal description formal
- Show it possesses the Markov property
- Use Kolmogorov equations to determine $p_t(i,j)$ for a MC defined by jump rates

Two States Chains

- Transition graph



- Transition rate matrix
    - $Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$
- Backward equation
    - $\frac{d}{dt}[p_t] = Qp_t \Leftrightarrow \begin{bmatrix} p_t'(1,1) & p_t'(1,2) \\ p_t'(2,1) & p_t'(2,2) \end{bmatrix} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} \begin{bmatrix} p_t(1,1) & p_t(1,2) \\ p_t(2,1) & p_t(2,2) \end{bmatrix}$
    - Since $\begin{cases} p_t(1,2) = 1 - p_t(1,1) \\ p_t(2,2) = 1 - p_t(2,1) \end{cases}$, we only need to find $p_t(1,1), p_t(2,2)$
    - $\begin{cases} p_t'(1,1) = -\lambda p_t(1,1) + \lambda p_t(2,1) \\ p_t'(2,1) = \quad \mu p_t(1,1) - \mu p_t(2,1) \end{cases} \Rightarrow \underbrace{p_t'(1,1) - p_t'(2,1)}_{g'(t)} = -(\lambda + \mu) \underbrace{\left(p_t(1,1) - p_t(2,1)\right)}_{g(t)}$
    - Solving the equation above ,we have $g(t) = Ce^{-(\lambda+\mu)t}$, where $C = 1$
    - Thus, $p_t(1,1) - p_t(2,1) = e^{-(\lambda+\mu)t}$

- $\begin{cases} p_t'(1,1) = -\lambda e^{-(\lambda+\mu)t} \\ p_t'(2,1) = \ \mu e^{-(\lambda+\mu)t} \end{cases} \Rightarrow \begin{cases} p_t(1,1) = \dfrac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} + \dfrac{\mu}{\lambda+\mu} \\ p_t(2,1) = -\dfrac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} + \dfrac{\mu}{\lambda+\mu} \end{cases}$

Stationary Distributions

- Recall from DTMC
  - Coordinate form: $\mathbb{P}_\pi(X_n = j) = \pi(j), \forall n \geq 0, j \in S$
  - Matrix form: $\pi \mathcal{P}^n = \pi, \forall n \geq 0 \Leftrightarrow \pi \mathcal{P} = \pi$
- Continuous time
  - Coordinate form: $\mathbb{P}_{\boldsymbol{\pi}}(\boldsymbol{X(t) = j}) = \boldsymbol{\pi(j)}, \forall \boldsymbol{t > 0}, \boldsymbol{j \in S}$
  - Matrix form: $\boldsymbol{\pi p_t = \pi}$
- Claim: **$\boldsymbol{\pi}$ is stationary if and only if $\boldsymbol{\pi Q = 0}$**
  - Assume $\pi Q = 0$, we want to show that $\pi p_t = \pi$
  - $\pi p_t = \pi e^{tQ} = \pi \sum_{n=0}^{\infty} \frac{(tQ)^n}{n!} = \pi + \pi \sum_{n=1}^{\infty} \frac{t^n}{n!} Q^n = \pi + 0 = \pi$

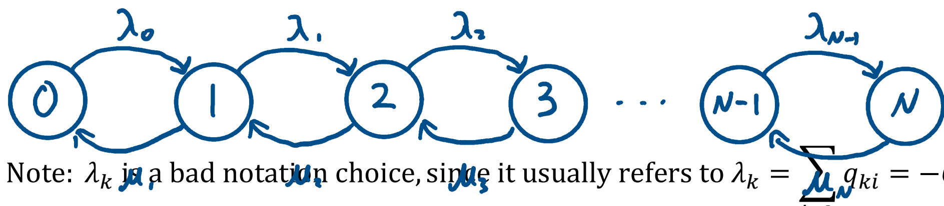
Convergence Theorem

- Irreducibility
  - A CTMC $X(t)$ is **irreducible** if **for any $\boldsymbol{i, j \in S}$**, there **exists states $\boldsymbol{k_1, \ldots, k_{n-1}}$** s.t.
  - $\boldsymbol{q(i, k_1) q(k_1, k_2) \cdots q(k_{n-1}, j) > 0}$ *i.e.* "It is possible to go from i to j"
- Fact about periodicity
  - If $X(t)$ is **irreducible**, then $\boldsymbol{p_t(i,j) > 0}$, for all $t > 0$ and $i, j \in S$
- Convergence theorem
  - If $X(t)$ is a CTMC s.t. $X(t)$ is **irreducible**, and **has a stationary distribution**
  - Then, $\boldsymbol{\lim_{t \to \infty} p_t(i,j) = \pi(j), \forall i, j \in S}$
- Proof
  - $p_h(i,j) > 0$ for all $h > 0$ and $i, j \in S$
  - p_h is a stochastic matrix that is irreducible, aperaodic, and has stationary distribution π
  - By Discrete Time Convergence Theorem, $\lim_{n \to \infty} p_{nh}(i,j) = \pi(j)$
  - Since this is true for all $h > 0$, we have $\lim_{t \to \infty} p_t(i,j) = \pi(j)$

Detailed Balance

- Definition
  - We say π satisfies the **detailed balance equations** if
  - $\boldsymbol{\pi(i) q(i,j) = \pi(j) q(j,i), \forall j \neq i}$
- Fact
  - Any distribution satisfying the detailed balance equations is a stationary distribution
- Example: Birth and Death Process

- $S = \{0,1,2,\dots,N\}$ with $N = \infty$ as a possible choice



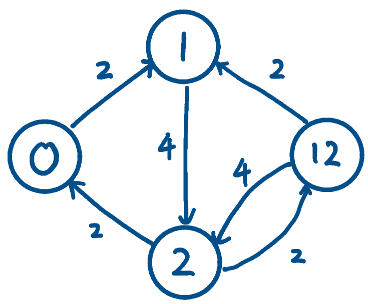
- Note: λ_k is a bad notation choice, since it usually refers to $\lambda_k = \sum_{i\neq k} q_{ki} = -Q_{kk}$
- Exercise: Show that Birth and Death processes satisfy the detailed balanced equations
- The transition rates for this Markov chain is $\begin{cases} q(n,n+1) = \lambda_n & \forall n \in \{0,\dots,N-1\} \\ q(n,n-1) = \mu_n & \forall n \in \{1,\dots,N\} \\ q(i,j) = 0 & \text{otherwise} \end{cases}$
- Let π be a distribution that satisfies the detailed balance equation. Then
- For $j \neq i+1$ or $i-1$
  - $\pi(i) \cdot 0 = \pi(j) \cdot 0$, which is automatically satisfied
- For $i \in \{0,\dots,N-1\}$
  - $\pi(i)q(i,i+1) = \pi(i+1)q(i+1,i)$
  - $\pi(i)\lambda_i = \pi(i+1)\mu_{i+1}$
  - $\pi(i+1) = \dfrac{\lambda_i}{\mu_{i+1}}\pi(i) = \dfrac{\lambda_i\lambda_{i-1}\cdots\lambda_1\lambda_0}{\mu_{i+1}\mu_i\cdots\mu_2\mu_1}\pi(0)$

CTMC Exercises

Tuesday, December 11, 2018 9:36 AM

Exercise 4.8(a)

- Two station queueing network
- Arrivals only occur to first station at rate 2
- Arriving customer at first station leaves if server is busy
- First server works at rate 4, second server works at rate 2
- When a customer is done as station 1, they go to station 2 immediately
- If station 2 already has a customer, the customer from station 1 leaves
- Model this as a CTMC with $S = \{0,1,2,12\}$
- Find the proportion of customers that enter the system
- An arriving customer enters the system if station 1 is open
- This only happens when the system is in state 0 or 2, so **we want $\boldsymbol{\pi(0) + \pi(2)}$**
- The jump rate matrix is
  - $Q = \begin{matrix} & 0 & 1 & 2 & 12 \\ 0 & -2 & 2 & 0 & 0 \\ 1 & 0 & -4 & 4 & 0 \\ 2 & 2 & 0 & -4 & 2 \\ 12 & 0 & 2 & 4 & -6 \end{matrix}$

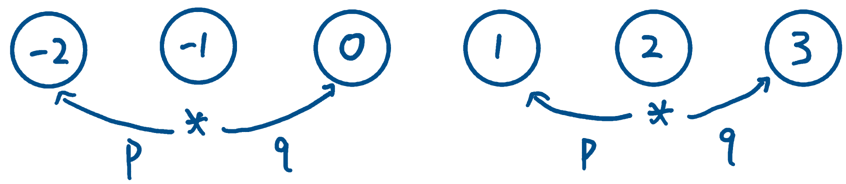


- **Detailed balance does not work**
  - $\pi(0)q(0,1) = \pi(1)q(1,0)$
  - $2\pi(0) = \pi(1) \cdot 0 = 0$
  - Thus, $\pi = [0 \quad 0 \quad 0 \quad 0]$ is the only solution satisifies DB
- Solving $\pi Q = 0$ with $\pi(0) + \pi(1) + \pi(2) + \pi(12) = 1$, we have
  - $\begin{cases} -2\pi(0) + 2\pi(2) = 0 \\ 2\pi(0) - 4\pi(1) + 2\pi(12) = 0 \\ 4\pi(1) - 4\pi(2) + 4\pi(12) = 0 \\ 2\pi(2) - 6\pi(12) = 0 \\ \pi(0) + \pi(1) + \pi(2) + \pi(12) = 1 \end{cases} \Rightarrow \pi = \begin{bmatrix} \frac{1}{3} & \frac{2}{9} & \frac{1}{3} & \frac{1}{9} \end{bmatrix}$

Exercise 4.13

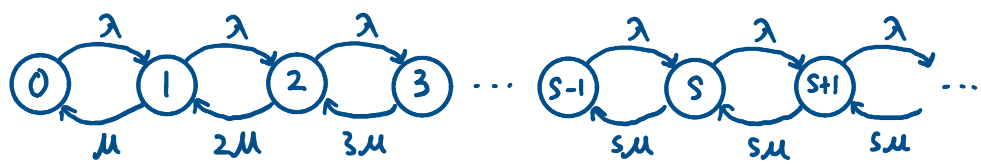
- 15 lily pads and 6 frogs
- Each frog gets the urge to jump to a new pad at rate 1
- When they jump, they choose 1 of 9 available pads uniformly at random
- Find the **stationary distribution** for the set of occupied lily pads
- Define $L = \{1,2,\dots,15\}$ and $S = \{s \subseteq L \mid |s| = 6\}$
- Then the only non-zero transition rates are
  - $q(\{a,b,c,d,e,f\},\{g,b,c,d,e,f\}) = \frac{1}{9}$ for any distinct $a,b,c,d,e,f,g \in L$

- To find π, use the **detailed balance equation**
  - $\pi(\{a, \dots, f\})q(\{a, \dots, f\}, \{g, b, \dots, f\}) = \pi(\{g, b, \dots, f\})q(\{g, b, \dots, f\}, \{a, \dots, f\})$
  - $\pi(\{a, \dots, f\}) \cdot \frac{1}{9} = \pi(\{g, b, \dots, f\}) \cdot \frac{1}{9}$
  - $\pi(\{a, \dots, f\}) = \pi(\{g, b, \dots, f\})$
- Therefore all the rates must be equal $\Rightarrow \pi(s) = \frac{1}{|S|} = \binom{15}{6}^{-1}$
- Asymmetric Simple Exclusion Process (with $p \neq q$)



Stationary Distribution of M/M/s Queue

- Find constraints on λ, μ so that a stationary distribution exists for the M/M/s



- The jump rates are $\begin{cases} q(n, n+1) = \lambda & n \geq 0 \\ q(n, n-1) = n\mu & 1 \leq n < s \\ q(n, n-1) = s\mu & n \geq s \end{cases}$
- Use the formula for birth and death process
  - $\boldsymbol{\pi(n) = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \pi(0)} = \begin{cases} \frac{\lambda^n}{n!\,\mu^n}\pi(0) & 1 \leq n < s \\ \frac{\lambda^n}{s!\,s^{n-s}\mu^n}\pi(0) & n \geq s \end{cases}$
- In order for π to be a distribution, we need $\sum_{n=0}^{\infty} \pi(n) < \infty$
  - $\sum_{n=0}^{\infty} \pi(n) = \sum_{n=0}^{s-1} \pi(n) + \sum_{n=s}^{\infty} \pi(n) = \underbrace{\pi(0) \sum_{n=0}^{s-1} \frac{\lambda^n}{n!\,\mu^n}}_{<\infty} + \frac{\pi(0)}{s!} \frac{\lambda^s}{\mu^s} \sum_{n=0}^{\infty} \left(\frac{\lambda}{s\mu}\right)^n < \infty$
  - We want $\sum_{n=0}^{\infty} \left(\frac{\lambda}{s\mu}\right)^n < \infty \Rightarrow \frac{\lambda}{s\mu} < 1 \Leftrightarrow \boldsymbol{\lambda < s\mu}$