CMOR 350 Notes

1 Probability Theory Review

Probability Model has 3 components:

- Sample Space(set of all possible outcomes)
- Events(any subset of the sample space) - any subset E of the sample space S
- Probability of events
- Probability($P(E)$) - number that satisfies:
  - $P(S) = 1$
  - $0 \leq P(E) \leq 1$
  - For any sequence of events $E_1, E_2, ...$ that are M.E. - $P(U_{n=1}^{\infty} E_n = \sum_{n=1}^{\infty} P(E_n)$
- Complement of Events(E^C) - all other events not E
- Intersection($E \cap F$) - event where E and F occurs
- Union($E \cup F$) - event where E or F occurs
- Mutually Exclusive($E \cap F =$) - cannot happen at same time

Probability Equations:

- $1 = P(S) = P(E \cup E^c) = P(E) + P(E^c) \implies P(E^c) = 1 - P(E)$
- $P(E \cup F) = P(E) + P(F) - P(EF)$
- When E, F are M.E. $\implies P(E \cup F) = P(E) + P(F) - P() = P(E) + P(F)$

Conditional Probability

- $P(E|F)$: conditional probability that E occurs given that F has occurred
  - If F, then for E to occur, both E and F must occur
  - Since we know that F has occurred - follows that F becomes new sample space
  - $P(E|F) = \frac{P(EF)}{P(F)}$

Independent Events:

- Events E, F are independent if $P(EF) = P(E)P(F), P(E|F) = P(E)$
- E and F are independent if knowledge that F occurred does not affect the probability that E occurs

Law of Total Probability:

- Let $E_1, E_2, E_3, ...$ be a set of mutually exclusive and exhaustive events. Then, for an event E:

$$P(E) = \sum_{n=1}^{\infty} P(E|E_n)P(E_n)$$

Bayes Formula:

- Suppose that $F_1, F_2, ...F_n$ are mutually exclusive events such that $U_{i=1}^{n} F_i = S$:

$$P(F_j|E) = \frac{P(EF_j)}{P(E)}$$
$$= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

2 Random Variables

- Function that assigns a real number to each outcome in the sample space
- Value of random variable is determined by the outcome of the experiment

Cumulative Distribution Function(CDF):

- F(.) of the random variable X is defined for any real number $b, -\infty < b < \infty$ by $F(b) = P\{X \leq b\}$
- F(B) - probability that random variable X takes on a value that is less than or equal to b
  - F(b) is non-decreasing
  - $lim_{b \to \infty} F(b) = F(\infty) = 1$
  - $lim_{b \to -\infty} F(b) = F(-\infty) = 0$
- $P\{a \leq X \leq b\} = F(b) - F(a)$ for all $a < b$
- Discrete random variables take on only a countable number of distinct values
- Continuous Random Variables take on a set of uncountable possible values

Probability Mass Function

- Suppose that X is a r.v. that takes on the values $x_1, x_2, ...$
- PMF $p(a)$ of X: $p(x_i) = P(X = x_i)$
- PMF $p(a)$ is strictly positive for at most a countable number of values so that $\sum_{i=1}^{\infty} p(x_i) = 1$

- Cumulative Distribution function F can be expressed as: $F(k) = P(X \leq k) = \sum_{x_i \leq k} p(x_i)$

Probability Density Function

- Nonnegative function $f(x)$ defined for all real $x \in (-\infty, \infty)$
- For any set B of real numbers:
$$P\{X \in B\} = \int_B f(x)dx$$
- $1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx$
- $P\{a \leq X \leq b\} = \int_a^b f(x)dx$
- CDF: $F(a) = P\{X \in (-\infty, a)\} = \int_{-\infty}^{a} f(x)dx$
- Density is the derivative of the cumulative distribution function:
$$\frac{d}{da}F(a) = f(a)$$

Expectation of RV - weighted average of all possible values:

- Discrete RV: $E[X] = \sum xp(x)$
- Continuous RV: $E[X] = \int_{-\infty}^{\infty} xf(x)dx$
- See expectations of well-known RVs
- Moment Generating Functions - expected value $E[X]$ is the mean / first moment of X. $E[X^n]$ is the n^{th} moment of X
- When a, b are constants: $E[aX + b] = aE[X] + b$
- If we calculate $f(t) = E(e^{tx})$, we get the moment generating function
  - $f'(t) = \frac{d}{dt}E[e^{tX}] = E[Xe^{tx}]$
  - $f'(0) = E[X]$
  - $f''(o) = E[X^2]$

Variance of RV:

- Expected square of deviation of X from its expected value:
$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$
- For variance of sum of independent RVs: - should add covariance functions if dependent
$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$
- If multiplied by a constant:
$$Var(cX) = c^2Var(X)$$

3 Conditional Probability and Expectation

Let X, Y be a bivariate RV

$$F(x,y) = P(X \leq x, Y \leq y), x, y \in (-\infty, \infty)$$

WTF Conditional Probability $P(X \leq x | Y \leq y)$

Conditional PMF / PDF:

- $$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x | Y = y\} \\ &= \frac{P(X = x; Y = y)}{P(Y = y)} \\ &= \frac{p(x,y)}{p_Y(y)} \end{aligned}$$

- $$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \tag{1}$$

Conditional Expectations:

- For discrete case:

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

- For continuous case

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

Computing Probabilities by Conditioning

- Let Y be a discrete random variable taking values $y_0, y_1, y_2, ...$ then, events $E = \{Y = y_i\}$ are mutually exclusive and exhaustive - by law of total probability:
  - $P(E) = \sum_{i=0}^{\infty} P(E|Y = y_i) P(Y = y_i)$
- When Y is a continuous random variable,
  - $P(E) = \int_{-\infty}^{\infty} P(E|Y = y) f(y) dy$

- Given that we have two independent continuous RVs with densities f_x, f_y, we know that $P\{X < Y\}$ can be described as

$$\int_y P(X < y) f_y(y) dy$$

- $E[X] = E[E[X|Y]]$

$$E[X] = \sum_y E[X|Y = y] P(Y = y) \text{ For Discrete}$$
$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_y(y) dy \text{ For Continuous}$$

Variances by Conditioning:

- $Var(X) = E[X^2] - (E[X]^2)$
- $E[X] = E[E[X|Y]])$
- $E[X^2] = E[E[X^2|Y]]$

4 Discrete-Time Markov Chains

Stochastic process - probabilistic model of a system that evolves randomly in time - looking at processes with discrete time and a discrete state space

Discrete-Time Stochastic Process:

- Consider a system that evolves randomly in time, observed at $n = 0, 1, 2, ..$
- X_n is the random state of the system at time n
- The sequence of random variables $\{X_0, X_1, X_2, ...\}$ is a discrete-time stochastic process, written as $\{X_n, N \geq 0\}$
- State space(S) - set of values that X_n can take, where $n \geq 0$

Discrete-Time Markov Chain(DTMC):

- Stochastic process $\{X_n, n \geq 0\}$ on state space S is a DTMC if, for all $i, j, i_{0,1}, ...i_{n-1}$ in S:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$

- Only most recent history is relevant - does not matter what system states are before period n - memoryless

- A DTMC is time-homogeneous(stationary) if for all $n = 0, 1, ...$

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) = p_{ij}$$

  - As long as the difference is one period - no matter what state - transition probabilities are the same
  - Can create a matrix of one-stop transition probabilities

    $$p_{ij} = P(X_{n+1} = j | X_n = i), \text{ p is an element of the transition matrix P}, 0 \leq p \leq 1$$

  - Probability that a process in state i goes to state j
  - Rows sum to 1 - stochastic matrix

General Random Walk:

- Condier $X_i, i \geq 1$ be iid with $P(X_i = j) = q_j, j = ..., -2, -1, 0, 1, 2, ...$
- Define $S_0 = 0$ and

$$S_n = \sum_{i=1}^{n} X_i$$

- $\{S_n, n \geq 0\}$ is called the general random walk and is a DTMC
- X is the amount moved, say on the real line moving in discrete locations for each step

$$P(S_{n+1} = j | S_n = 1)$$
$$= P(S_n + X_{n+1} | S_n = i)$$
$$= P(X_{n+1} = j - i) = q_{j-i}$$

  - The random walk $\{S_n, n \geq 0\}$ with $S_0 = 0$ and

    $$S_n = \sum_{i=1}^{n} X_i$$

    Is a simple random walk if for some $p, 0 < p < 1$:

    $$P(X_i = 1) = p, P(X_i = -1) = (1 - p)$$

  - In this situation, can only move up one or down one

- Can show as DTMC by
  - Writing X_{n+1} as a function of X_n
  - Showing that $P(X_n + 1 = j | X_n = i, X_{n+1} = i_{n+1} ... X_0 = i_0\} = P(X_{n+1} = j | X_n = i)$

5 Transient Behavior of DTMCs

- If I observe a system for 5 periods, what will be the state?
- Counting how many times a state is visited
- Transient Distribution:
  - Consider a stationary DTMC with tpm(transient probabability matrix) P on state space S
  - Initial distribution is given by $a = [a_i]$ where $a_i = P(X_0 = i)$ for $i \in S$
  - Need to predict state of system at time n
  - $P(X_n = j) =?$
    - $$\vec{a} = [a_1, a_2, ...a_N]$$
      $$a_i = P(X_0 = i)$$
      $$P(X_n = j) = \sum_{i \in S} P(X_n = j|X_0 = 1)P(X_0 = i) = \sum_{i \in S} P_{ij}^{(n)} a_i$$
      $$p_{ij}^{(n)} = P(X_n = j|X_0 = i)$$
      $$P^{(n)} = P^n$$
    - n step probability matrix is the same as the nth power of the one step probability matrix - Chapman-Kolmogoroc Equations

      $$\begin{aligned} P_{ij}^{(n+m)} = P(X_{n+m} = j|X_0 = i) &= \sum_{k \in S} P(X_{n+m} = j|X_n = k, X_0 = i)P(X_n = k|X_0 = i) \\ &= \sum_{k \in S} (P_{n+m} = j|X_n = k)P(X_{n=k}|X_0 = i) \\ &= \sum_{k \in S} P_{kj}^{(m)} P_{ik}^{(n)} = \sum_{k \in S} P_{ik}^{(n)} P_{kj}^{(m)} \end{aligned}$$
    - This is essentially one operation of matrix multiplication
    - $P^{(n+m)} = P^{(n)} P^{(m)}$
  - Finding probability of states after n states is:

    $$a^{(n)} = aP^{(n)}$$
- Occupancy Times - how to compute expected time that a DTMC spends in a particular state during a given time interval
  - $N_j(n)$ - number of times the DTMC visits state j over the time span $\{0, 1, ..., n\}$
  - Define:

    $$m_{ij}(n) = E(N_j(n)|X_0 = i)$$

- Where $m_{ij}(n)$ is the expected time spent(occupancy time) in state j up to time n starting from state i
- $M(n) = [m_{ij}(n)]$ is the occupancy time matrix

$$Z_N = \begin{cases} 1 \text{ if } X_n = j \\ 0 \text{ else} \end{cases}$$

$$N_j(n) = \sum_{r=0}^{n} Z_r$$

$$m_{ij}(n) = E[N_j(n)|X_0 = i] = E[\sum_{r=0}^{n} Z_r|X_0 = i]$$

$$= \sum_{r=0}^{n} E[Z_r|X_0 = i] = \sum_{r=0}^{n} 1 * P(Z_r = 1|X_0 = i)$$

$$= \sum_{r=0}^{n} P(X_r = 1|X_0 = i) = \sum_{r=0}^{n} P_{ij}^{(r)}$$

$$M(n) = [m_{ij}(n)] = \sum_{r=0}^{n} P^{(r)} = \sum_{r=0}^{n} P^r$$

- For a stationary DTMC on state space S with tpm P, the occupancy times matrix is given by

$$M(n) = \sum_{r=0}^{n} P^r$$

6 Limiting Behavior

- Markov chain which goes to infinity

$$\lim_{n\to\infty} P(X_n = j)$$

- Does the pmf of X_n approach to a limit?
- If the limiting distribution exists, is it unique? item How do we compute the limiting distribution?
- Computing the long-run proportion that the DTMC spends in state j

$$\lim_{n\to\infty} \frac{m_{ij}(n)}{n+1}$$

- For large n, if we get a matrix whose rows are identical, then we have a limiting distribution which is that row

- If we get a matrix which converges whose rows are not identical, then a limiting distribution exists but depends on the initial distribution
- If we dont get a convergent matrix then there is no limiting distribution
- If a limiting distribution exists, such that

$$\pi_j = \lim_{n\to\infty} P(X_n = j), j \in S$$

- Then it satisfies the following equations:

$$\pi_j = \sum_{i\in S} \pi_i p_{ij}, j \in S$$
$$\sum_{j\in S} \pi_j = 1$$

$$\lim_{n\to\infty} P(X_{n+1} = j) = \sum_{i\in S} P(X_{n+1} = j|X_n = i)P(X_n = i) = \lim_{n\to\infty} \sum_{i\in S} P_{ij}P(X_n = i)$$
$$\pi_j = \sum_{i\in S} P_{ij} \lim_{n\to\infty} P(X_n = i)$$

- Classification of States(CoS)
  - Given two stats i and j, a path from i to j is a sequence of transitions that begins at i and ends at j such that each transition in the sequence has a positive probability of occurring
  - State j is said to be accessible from state i if $p_{ij}^{(n)} > 0$ for some $n \geq 0$:
    * A path exists from i to j
    * It is possible for the system to eventually enter state j starting from state i
    * If state j is accessible from state i and state i is accessible from state j, then states i and j are said to communicate - $i \leftrightarrow j$
      · Any state communicates with itself 9 $P_{ii}^{(0)} = P(X_0 = i|X_0 = i) = 1$
      · If $i \leftrightarrow j$ then $j \leftrightarrow i$
      · If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$
    * Two states that communicate are said to be in the same class
    * Communication partitions the states into disjoint classes
    * A class may be a single state
    * If there is only one class(i.e. if all states communicate) then the Markov chain is said to be irreducible
    * If not irreducible then considered reducible
- For any state i, let f_{ii} denote the probability that, starting in state i, the process will ever reenter state i

  - State i is said to be transient if $f_{ii} < 1$ - positive probability that process will never return to transient state, state will be visited only a finite number. If N is the number of times state i is visited - geometric distribution - time until first failure

$$P(N = k) = f_{ii}^{k-1}(1 - f_{ii})$$
$$E[N] = \frac{1}{1 - f_{ii}}$$

    * Since we are starting at state i, we start count at 1
  - State i is said to be recurrent if $f_{ii} = 1$ - after leaving state, process will definitely return to this state again - state will be visited infinitely many times
    * Transient iff

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} < \infty$$

    * Recurrent iff

$$\sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty$$

  - Recurrence and transience are class properties - if state i is recurrent and state j communicates with the states then j is recurrent - same for transient

- State is absorbing if the process will never leave this state($p_{ii} = 1$)
- In a finite-state Markov chain, not all states can be transient
  - If every state is transient, then will eventually not come back to that state, will eventually run out of states
- Suppose state j is recurrent
  - Number of transitions until markov chain makes a transition into state j:

$$N_j = min\{n > 0 : X_n = j\}$$

  - First time markov chain enters into j - gives min
  - Expected Number of transitions that it takes the Markov chain when starting in state j to return to that state

$$m_j = E\{N_j | X_0 = j\}$$

  - State j is positive recurrent if $m_j < \infty$ and null recurrent is $m_j = \infty$

- For random walks, if we go forward n, then we must go backwards n

$$P_{00} = \binom{2n}{n}(\frac{1}{2})^{2n} = \frac{2n!}{n!n!}\frac{1}{2}^{2n} n! \approx n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$$

$$\sum_{n=0}^{\infty} P_{00}^{(n)} \approx \sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty$$

- Therefore all states in random walks must be recurrent - however expected time to visit state is infinity
- Period of state i is an integer $d(d \geq 1)$ such that $p_{ii}^{(n)} = 0$ for all values for n other than $d, 2d, 3d$, and d is the smallest integer with this property
  - Starting in i, it may be possible for process to enter state i only at times 2,4,6,8,... in which case state i has period 2
  - A state with period 1 is said to be aperiodic - needs a self-loop
  - Can be shown that periodicity is a class property
  - Positive recurrent, aperiodic states are ergodic
- Aperiodic, reducible, and positively recurrent markov chains guarantee a limiting distribution, following a unique solution of

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}, j \in S$$

$$\sum_{j \in S} \pi_j = 1$$

- If there is no solution of these equations, then markov chain is either transient or null recurrent and all $\pi_j = 0$

$$\pi_j = \lim_{n \to \infty} \frac{m_{ij}(n)}{n+1}$$

- If the DTMC is ergodic(irreducible, positive recurrent, and aperoidic):

$$lim_{n \to \infty} P(X_n = j)$$

- Exists and is independent of i. Furthermore,

$$\lim_{n \to \infty} P(X_n = j) = \pi_j > 0$$

- Long-run proportions are also called stationary probabilities
  - If initial state is chosen according to the probabilities π_j such that $P(X_0 = j) = \pi_j$, then $P(X_n = j) = \pi_j \forall n, j \in S$
  - If we start our initial states as long-run proportions, then all following states will be those proportions
- When reducible, long run proportions depend on initial state

7 Cost Models and First Passage Times

Cost Models

- Suppose system incurs a random cost of C(i) units very time it visits state i
- c(i) = E(C(i)): expected cost incurred at every visit to state i
- Expected total cost incurred up to time n starting from state i

$$g(i,n) = E[\sum_{r=0}^{n} C(X_r)|X_0 = i]$$
$$= \sum_{j\in S} m_{ij(n)c(j)}$$
$$= \sum_{r=0}^{n}\sum_{j\in S} E[c(X_r)|X_0 = i, X_r = j]P(X_r = j)$$
$$= \sum_{r=0}^{n}\sum_{j\in S} c(j)P_{ij}^{(r)} = \sum_{j\in S} c(j)[\sum_{r=0}^{n} P_{ij}^{(n)}]$$

- For an irreducible and positive recurrent DTMC, the expected long-run cost rate is

$$\sum_{j=1}^{N} \pi_j c(j)$$

- Probability that the DTMC makes a transition into state j for the first time in n periods starting from state i?
  - $f_{ij}^{(n)}$ - probability that the first passage time from state i to state j is equal to n

$$f_{ij}^{(1)} = p_{ij}$$
$$f_{ij}^{(2)} = \sum_{k\in S, k\neq j} p_{ij} f_{kj}^{(1)}$$
$$\vdots$$
$$f_{ij}^{(n)} = \sum_{k\in S, k\neq j} p_{ij} f_{kj}^{(n-1)}$$

- What is expected time that the DTMC makes a transition into state j for the first time starting from state i?

- μ_{ij} - expected first passage time from state i to state j

$$\begin{aligned}\mu_{ij} &= E[time\ to\ make\ the\ first\ transition\ into\ state\ j|X_0=i]\\ &= \sum_{k\in S} E[time\ to\ make\ the\ first\ transition\ into\ state\ j|X_0=i, X_1=k]p_{ik}\\ &= (1)p_{ij} + \sum_{k\neq j}(1+\mu_{kj})p_{ik}\\ &= 1 + \sum_{k\neq j} p_{ik}\mu_{kj}\end{aligned}$$

- Probability of Absorption
  - Suppose j is a recurrent state, $j \in R$ and i is a transient state, $i \in T$
  - Interested in computing probability f_{ij} of ever making a transition into state j given the process starts at $i \in T$

$$\begin{aligned}f_{ij} &= \sum_{k\in T} p_{ik}f_{kj} + \sum_{k\in R} p_{ik}f_{kj} + \sum_{k\notin R,T} p_{ik}f_{kj}\\ &= \sum_{k\in T} p_{ik}f_{kj} + \sum_{k\in R} p_{ik}\end{aligned}$$

8 Midterm Review

- 3-4 questions
- Condition Probabilities / Expectations - finding probabilities and expectations by conditioning(25%)
- DTMCs up to cost models(75%)
  - Define state - show it is a DTMC

$$X_{n+1} = X_n$$
$$P(X_{n+1}=j|X_n=i, X_{n-1}=i_{n-1}....,) = P(X_{n+1}=j|X_n=i)$$

  - Construct a P matrix
  - Can find probability of value at certain states by finding a "path" that gets there, e.g.:

$$P(X_4 \geq 2, X_5 \geq 2|X_0=0) = \sum_{j=2}^{4}\sum_{k=2}^{4} P^4_{0j}P^1_{jk}$$

9 Continuous-Time Markov Chains

- System changes states at arbitrary points in time, rather than at discrete intervals
- Evolution of process needs to be observed continuously
- Stochastic process $\{X(t), t \geq 0\}$ on state space S is a continuous-time Markov chain if, for all $i, j \in S$

$$P(X(s+t) = j|X(s) = i, X(u)0 \leq u \leq s) = P(X(s+t) = j|X(s) = j)$$

- A CTMC is said to be time-homogeneous(stationary) if

$$P(X(s+t) = j|X(s) = i) = P(X(t) = j|X(0) = i) = p_{ij}(t)$$

$$P(t) = [p_{ij}(t)] = \begin{pmatrix} p_{11}(t) & p_{12}(t) & p_{13}(t) & \dots \\ \vdots & \vdots & \vdots & \dots \\ p_{i1}(t) & p_{i2}(t) & p_{i3}(t) & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

- $P(t)$ has following properties:
  - $p_{ij}(t) \geq 0$
  - $\sum p_{ij}(t) = 1$
  - $p_{ij}(s+t) = \sum_{k \in S} p_{ik}(s)p_{kj}(t) = \sum_{k \in S} p_{ik}(t)p_{kj}(s)$

9.1 Example

Lifetime of machine is a r.v., denoted by T

$$X(t) = \begin{cases} 1 \text{ Working at time t} \\ 0 \text{ Otherwise} \end{cases}$$

Matrix is solely dependent on t, 1 until certain t is hit:

$$\begin{pmatrix} 1 & 0 \\ P(T \leq t) & P(T > t) \end{pmatrix}$$

$$P(X(T+S) = 1|X(S) = 1) = P(T > t+s|T > s) = \frac{1-(t+s)/a}{1-s/a}, T \sim U(0,a), P(T \leq t) = \frac{t}{a}$$

$$S = 0 : \frac{at-1}{a-1}$$

$$S = 1 : \frac{a-t}{a}$$

Since $S = 1 \neq S = 0$, not time-homogeneous

However, if $T \sim exp(\lambda)$

$$P(T \leq t) = 1 - e^{-\lambda t}, t \geq 0$$
$$P(T > t+s|T > s) = \frac{P(T > t+s, T > s}{P(T > S)} = \frac{P(T > t+s)}{P(T > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t)$$
$$P(t) = \begin{pmatrix} 1 & 0 \\ 1 - e^{-\lambda t} & e^{-\lambda t} \end{pmatrix}$$

What is the rate matrix? - if lifetime is exponential random with rate μ - machine changes state when it fails, transition from 1 to 0 - rate will be μ - $v_i = \mu, p_{10} = 1, q_{10} = \mu * 1$

$$S = \{0, 1\}$$
$$R = [[0, 0], [\mu, 0]]$$

If machine is repairded when failed - and repair time is exponentially distributed with rate λ - only difference is for $v_0 = \lambda, p_{01} = 1$, $q_{01} = \lambda$

Simpler Method for describing a CTMC:

- A CTMC is a stochastic process having the properties that each time it enters state i
  - The amount of time it spends in that state before making a transition into a different state is random variable which is exponentually distributed with mean $1/v_i$
  - When the process leaves state i it next enters state j with some probability, p_{ij}, the $p_{ij}s$ must satisfy $p_{ii} = 0$ and $\sum_j p_{ij} = 1 \forall i$
  - Starts in state i taking from Exp amount of time in that state, then moves to state j with probability p_{ij} then stays in for exp amount of time, etc...
  - Memoryless property of exponential distribution allows markovian property
  - If we observe at time n, would not know distribution of remaining time - would have different distribution - but since exponential is memoryless, remaining is also exponential with same rate
  - Self-loops(p_{ii}) is always 0
  - The amount of time spent in state i is called the sojourn time in state i, and time spent in each state is independent.

Rate Matrix

- For any pair of states i, j let $q_{ij} = vp_{ij}$ - the rate of going into state j once system is in state i

- q_{ij} is the rate, when i nstate i at which the process makes a transition into state j

- q_{ij} are called instantenous transition rates

- $$R_{ij} = \begin{cases} q_{ij}, i \neq j \\ 0, i = j \end{cases}$$

Generator Matrix

- In first row, sum up all rates and put negative in the $1, 1$ spot so that sum to 0- work through all rows
- Makes easier to find limiting probabilities

Properties of Exponential Random Variables

- $X_i \sim Exp(\lambda_i), i = 1, 2, ..., k$ and $X_1, X_2, ..., X_k$ are independent
- $X = min\{X_1, X_2, ..., X_k\}$
- Useful to find the next event type - time where CTMC makes next transition
- $X \sim Exp(\sum_{i=1}^{k} \lambda_i)$

$$\begin{aligned} P(X > t) &= P(min\{X_1, X_2, ..., X_k\} > t) \\ &= P(X_1 > t, X_2 > t, ...X_k > t) \\ &= P(X_1 > t)P(X_2 > t)...P(X_k > t) \\ &= e^{-\lambda_1 t}e^{-\lambda_2 t}...e^{-\lambda_k t} \\ &= e^{(\sum_{i=1}^{k} \lambda_i)t} \end{aligned}$$

- Define $Z = i$ if $X_i = minX_1, X_2, ..., X_k$. That is, the minimum of k exponential random variables is the ith one - finding $P(Z = i)$ - want to know that the probability that the ith event occurs before the others so that CTMC will make transition

$$\begin{gathered} P(Z = i) = P(X_i = min\{X_1, X_2, ..., X_k\}) \\ = \int_0^\infty P(X_i = min\{X_1, X_2, ..., X_k\}|X_1 = x)\lambda_i e^{-\lambda_i x}dx \\ = \int_0^\infty (e^{-\lambda_2 x}e^{-\lambda_3 x}...e^{-\lambda_k x})\lambda_1 e^{-\lambda_i x}dx \\ = \lambda_i \int_0^\infty e^{-\sum_{i=1}^{k} \lambda_i}dx \\ = \lambda_i \frac{1}{\sum_{i=1}^{k} \lambda_i} \int_0^\infty (\sum_{i=1}^{k} \lambda_i)e^{(-\lambda_{i=1}^{k} \lambda_i)x}dx \\ = \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i} \end{gathered}$$

- Sum of identical exponential random variables, $X_1+X_2+...X_k \sim Erlang(k, \lambda)$ - special case of Gamma

9.2 Example

You arrive at a post office. There are two clerks, both busy with a customer. Amount of time a clerk spends with each customer is $Exp(\lambda)$. What is the probability that you are the last one to leave among the three people?

- Essentially want to find probability that new persons total time is less than the that the person(currently with the clerk who does not leave first)'s remaining time
- Say S is service time of new person, R is remaining time for the person who is still in service(does not leave first) - want to find probability that $P(R < S)$
- $S \sim Exp(\lambda)$, $R \sim Exp(\lambda)$ - due to memoryless property of exponential
- $P(R < S) = P(min(R, S) = R) = \frac{\lambda}{2\lambda} = 1/2$

9.3 Example

- You arrive at a post office having two clerks at a moment when both are busy but there is no one else waiting in line. You will enter service when either clerk becomes free.
- Service times for clerk i are exponential with rate $\lambda_i, i = 1, 2$
- Define T as the amount of time that you spend in the post office. What is $E[T]$?
- T = waiting time(W) + service time(S)
- Say R_i = remaining service time of server i

$$E[T] = E[W] + E[S]$$
$$W = min(R_1, R_2) \sim Exp(\lambda_1 + \lambda_2)$$
$$E[W] = \frac{1}{\lambda_1 + \lambda_2}$$
$$E[S] = E[S|R_1 < R_2]P(R_1 > R_2) + E[S|R_1 > R_2]P(R_2 > R_1)$$
$$\frac{1}{\lambda_1}\frac{\lambda_1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_2}\frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{2}{\lambda_1 + \lambda_2}$$

9.4 Example

- Consider a workshop with two machines, each with its own repair person. Machine lifetime is $Exp(\mu)$ and repair time is $Exp(\lambda)$. Define X(t) as the number of machines operating at time t. Is X(t) a CTMC?
- X(t) - number of working machines at t
- $S = \{0, 1, 2\}$
- Want to find rate matrix

- State 0 - Remaining Repair Time for Machine i- R_i
  - Next event will occur at $min(R_1, R_2)$ - whenever one repair completes, system goes to 1, cannot go to 2 as $R_1 \neq R_2$
  - $min(R_1, R_2) \sim Exp(2\lambda)$ - time spent in state 0, 2λ is rate out of state 0
  - $v_0 = 2\lambda$, always goes to 1 so $p_{01} = 1$, $r_{01} = 2\lambda$
- State 1 - Remaining lifetime of working machine is $L \sim Exp(\mu)$, R is remaining repair time of down machine $R \sim Exp(\lambda)$
  - Transition time is $min(L, R)$ - whatever happens first - amount of time spent in state 1 - $v_1 = \lambda + \mu$
  - Moves to state 0 when $p_{10} = P(L < R) = \frac{\mu}{\mu+\lambda}$
  - Moves to state 1 when $p_{12} = P(L > R) = \frac{\lambda}{\mu+\lambda}$
  - $r_{10} = (\lambda + \mu)\frac{\mu}{\mu+\lambda} = \mu$
  - $r_{12} = \lambda$
- State 2 - $min(L_1, L_2) \sim Exp(2\mu)$
  - $p_{21} = 1$, $v_2 = 2\mu$, $r_{21} = 2\mu$

$$Rate = \begin{pmatrix} 0 & 2\lambda & 0 \\ \mu & 0 & \lambda \\ 0 & 2\mu & 0 \end{pmatrix}$$

- WHat is the next event that triggers a state transition?

9.5 Example

Consider 4 machines with 2 repair persons. The machines are identical and the lifetimes are independent $Exp(\mu)$ random variables. When machines fail, they are served in the order of failure by 2 repair persons. Each failed machine needs one repair person, and the repair times are $Exp(\lambda)$ random variables. A repaired machine behaves like a brand new machine

- Define $X(t)$ as number of machines operating at time t
- What is the rate matrix?
- $S = \{0, 1, 2, 3, 4\}$
- State 0
  - 2 Machines being repaired - one machine will be repaired first
  - Only transition from 0 to 1 - rate is $min(R_1, R_2) = 2\lambda$
  - State 1 - one machine can fail or one machine can be fixed

  - Rate time will be $min(L, R_1, R_2)$, L is lifetime of working machine and R is remaining time to fix for a down machine
  - Transition from 1 to 0 happens when machine breaks before - rate is μ
  - Transition from 1 to 2 happens when fix happens before break - 2λ
- State 2 - 2 up 2 down - $min(L_1, L_2, R_1, R_2) \sim Exp(2\lambda + 2\mu)$
  - When going from 2 to 3, fix before break, rate is 2λ
  - Going from 2 to 1 - break before fix - rate is 2μ
- State 3 $min(L_1, L_2, L_3, R)$ - moves to state 4 when $R > L_1, L_2, L_3$, rate is μ - moves to state 2 when $min(L_i) < R$- 3μ
- Similar for state 4 - $min(L_1, L_2, L_3, L_4)$, only decreases when L breaks - rate is 4μ

10 Poisson Process

- A stochastic process $\{N(t), t \leq 0\}$is said to be a counting process if N(t) represents the total number of events that occur by time t
- $N(t) \geq 0$ and integer valued
- if $s < t$, then $N(s) \leq N(t)$
- For $s < t, N(t) - N(s)$ - number of events that occur in (s,t]

Poisson Process

- $X_1, X_2, X_3, , ...$ - sequence of independent $Exp(\lambda)$ random variables representing inter-event times
  - $S_0 = 0$
  - $S_n = X_1, X_2, ... + X_n$
  - $N(t) = max(n \geq 0 : S_n \leq t)$
- $N(t)$ is special counting process and is called the Poisson process, $PP(\lambda)$
- Since S_n is Erlang Distributed

$$P(N(t) = k) = P(N(t) \geq k) - P(N(t) \geq k + 1)$$
$$\{N(t) \geq k\} \iff \{S_k \leq t\}$$
$$P(S_k \leq t) = [1 - \sum_{r=0}^{k-1} \frac{e^{-\lambda t}(\lambda t)^r]}{r\ !}$$
$$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

- If I have a poisson counting process - observing some events and interevent times at exponential randoms - $P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$, $E[N(t)] = \lambda t$
- Properties:
  - Independent Increments - number of events that occur at disjoint intervals are independent $N(10)$ is independent of $N(15) - N(10)$
  - Stationary Increments - distribution of the number of events that occur in any interval of time depends only on the interval length - $N(15) - N(10)$ and $N(5)$ have the same distribution
  - When h is small, $P(N(h) = 1) \approx \lambda h, P(N(h) \geq 2) \approx 0$

10.1 Example

- Consider the arrivals to ER of a hospital during a 24-hour interval 6am to 6am. Assume arrivals follow a Poisson process with rate $\lambda = 5$ patients per hour
  - What is the probability of seeing 57 patients by 9pm if the ER has had 52 patients by 8pm?
    - $$N(t) \sim PP(\lambda)$$ $$P(N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$
    - Know that $T_i \sim Exp(\lambda)$, where T_i is the interarrival time
    - Since starting at 6 am(14 hours before 8pm) and independent increments, essentially finding $P(N(15) = 57|N(14) = 52) = P(N(15) - N(14) = 5|N(14) = 52) = P(N(1) = 5) = \frac{e^{-5(1)}(5(1))^5}{5!}$
  - On average, how much time is there between two arrivals?
    - $$T_i \sim Exp(\lambda)$$ $$E[T_i] = \frac{1}{\lambda} = \frac{1}{5}$$
  - At 8am, what is the expected number of people that arrive during the next 2 hours?
    - $E[N(4) - N(2)] = E[N(2)] = \lambda t = 5 \cdot 2 = 10$
  - What is the expected time until the 8th person arrives?
    - $8/\lambda$ hours, $8 \cdot E[T_i]$
    - Distributed as sum of exponential(Erlang distribution)

- Of all patients arriving to ER, 63% are treated and released. Rest is admitted to the hospital. What can we say about the admissions?
    * $\lambda = 5$ patients per hour
    * 63% leave, 36% admitted, since poisson process, if process is split up, branches are also poisson process weighted by probabilitiy of branch
    * If leave - $PP(\lambda_0 = \lambda p)$
    * If admitted - $PP(\lambda_1 = \lambda(1-p))$
    * When many poisson processes sum to one process(superposition), then also results in a poisson process

11 Limiting Distribution

- $p_j = \lim_{t\to\infty} P(X(t) = j)$
- limit exists if all states of Markov chain communicate
- Markov chain is positive recurrent
- If limit exists, $p = [p_1, p_2, ..., p_i, ...]$ is the limiting distribution
- An irreducible positive recurrent $CTMC\{X(t), t \geq 0\}$ with rate matrix R has a unique limiting distribution. It is given by the solution to

$$\sum_{i\in S} p_i = 1$$
$$p_j v_j = \sum_{i\in S} p_i q_{ij} j \in S$$

- Rate out of state j = rate into state j

11.1 Example

Machine reliability, machine is up for $Exp(\mu)$ and down for $Exp(\lambda)$

$$R = \begin{bmatrix} 0 & \lambda \\ \mu & 0 \end{bmatrix}$$

- $\lambda p_0 = \mu p_1$ - state 0
- $\mu p_1 = \lambda p_0$ - state 1
- $p_0 + p_1 = 1$
- $p_0 = \frac{\lambda}{\lambda+\mu}$
- $p_1 = \frac{\mu}{\lambda+\mu}$

11.2 Example

Consider a service system with 3 servers. Service times are exponentially distributed with rate $R_i \sim \mu$. Customers arrive to this system according to a $PP(\lambda)$

- $S = \{0, 1, 2, ...\}$
- Any increase in state number - rate of λ - extra customer
- Any decrease in state number - customer is served and leaves, when in state 1 - just μ, when in state 2 - $min(R_1, R_2) - 2\mu$, when in state 3 or above - $min(R_1, R_2, R_3) - 3\mu$
- State 0 $\lambda p_0 = \mu p_1$
- State 1 $(\lambda + \mu)p_1 = \lambda p_0 + 2\mu p_2$
- State 2-∞ $(\lambda + 2\mu)p_2 = \lambda p_1 + 3\mu p_3$
- Essentially an infinite birth and death process

For a finite birth and death process, where rate out of state 1 $= \lambda_1$ and rate into state 1 $= \mu_0$, finding limiting distribution

- State 0 - $\lambda_0 p_0 = \mu_1 p_1$ - $p_1 = \frac{\lambda_0}{\mu_1} p_1$
- State 1 - $(\lambda_1 + \mu_1)p_1 = \lambda_0 p_0 + \mu_2 p_2$ - $p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2}$
- State i - $(\lambda_i + \mu_i)p_i = \lambda_{i-1} p_{i-1} + \mu_{i+1} p_{i+1}$ - $p_i = \frac{\lambda_0 \lambda_1 ... \lambda_{i-1}}{\mu_1 \mu_2 ... \mu_i} p_0$, $z_i = \frac{\lambda_0 \lambda_1 ... \lambda_{i-1}}{\mu_1 \mu_2 ... \mu_i}$

$$p_i = z_i p_0$$

$$p_0 + \sum_{i=1}^{k} p_i = 1$$

$$p_0 + p_0 \sum_{i=1}^{k} z_i = 1$$

$$p_0 = \frac{1}{1 + \sum_{i=1}^{k} z_i}$$

$$p_i = z_9 \frac{1}{1 + \sum_{i=1}^{k}}$$

$$p_i = \frac{z_i}{1 + \sum_{i=1}^{k} z_i}$$

- State k(upper bound) - $\lambda_{k-1} p_{k-1} = \mu_k p_k$
- To solve easily, write everything in terms of some reference state(i.e. p_0)
- As $k \to \infty$(infinite states), $p_i \to \infty$ if $\sum_{i=0}^{\infty} z_i$ diverges, if converges then $p_i > 0$ and a limiting distribution exists

11.3 Example

A manufacturing facility consists of a single machine that can be turned on and off. If machine is on, it produces items according to a PP(λ). Demand for items arrive according to a PP(). The machine is controlled as follows:

- If the number of items in stock reaches a maximum number of 4 (storage capacity), the machine is turned off.
- It is turned on again when the number of items in stock decreases to 2.
- How can we model this system as a CTMC?
- Suppose that demands occur at the rate of 5 per hour, and the average time to manufacture one item is 10 minutes. What is the limiting distribution?
  - $\lambda = 1/10$ - manufacturing rate
  - $\mu = 5/60 = 1/12$

Describe $X(t), \{X(t), t \geq 0\}$ as the number of items in storage, where $S = \{[0,1](1), [1,1](2), [2,1](3), [3,0](4$ need to know both state of machine and storage as when there are 3 items, could be on or off depending on context

- From state $[0,1]$ will move to $[1,1]$ with rate λ - only creates new
- From state $[1,1]$ can either move to $[2,1]$ with rate λ(new item) or to $[0,1]$ with rate μ(customer picks up item)
- From state $[2,1]$ can go back to $[1,1]$ or goes to $[3,1]$ with new customer
- From $[3,1]$, goes up to $[4,0]$(storage capacity) or down to $[2,1]$
- From $[4,0]$ can only go back to $[3,0]$
- $[3,0]$ can only go back to $[2,1]$

$$Q = \begin{bmatrix} -\lambda & \lambda & & & & \\ \mu & -(\mu+\lambda) & \lambda & & & \\ & \mu & -(\mu+\lambda) & \lambda & & \\ & & \mu & -(\mu+\lambda) & \lambda & \\ & & & \mu & -\mu & \lambda \\ & & & \mu & & -\mu \end{bmatrix}$$

$$R = \begin{bmatrix} & \lambda & & & & \\ \mu & & \lambda & & & \\ & \mu & & \lambda & & \\ & & \mu & & \lambda & \\ & & & \mu & & \lambda \\ & & & \mu & & \end{bmatrix}$$

For balance equations:

$$
\begin{gathered}
(1) : \lambda p_1 = \mu p_2 \\
(2) : (\lambda + \mu) p_2 = \lambda p_1 + \mu p_3 \\
(3) : (\lambda + \mu) p_3 = \lambda p_2 + \mu p_4 + \mu p_6 \\
(4) : (\lambda + \mu) p_4 = \lambda p_3 \\
(5) : \mu p_5 = \lambda p_4 \\
(6) : \mu p_6 = \mu p_5 \\
p_1 + p_2 + ... p_6 = 1
\end{gathered}
$$

Finding $min(Exp(\mu), Exp(\lambda)) \sim Exp(\mu + \lambda)$

Can also solve $pQ = 0$ to calculate limiting distribution, where Q is the generator matrix

12 Transient Analysis

Compute $P(t)$ - probabilities of CTMC

- Use uniformization - creates equivalence system which allows time spent in each state to have same rate μ
- Consider a Markov chain s.t. time spent in every state is $Exp(v)$ where $v_i = \sum_{j \in S} q_{ij}$, $p_{ij} = \frac{q_{ij}}{v_i}$
- Define $N(t)$ as the number of transitions by time t - then $N(t)$ is $PP(v)$

12.1 Example

Given $R = \begin{bmatrix} 0 & 2 & 3 & 0 \\ 4 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 0 & 3 & 0 \end{bmatrix}$

$$
\begin{gathered}
v_1 = 5 \\
v_2 = 6 \\
v_3 = 4 \\
v_4 = 4
\end{gathered}
$$

Define $v \geq max_{i \in S}\{v_1\}$, say $v = 6$, define matrix $\hat{P} = \begin{cases} 1 - \frac{v_1}{v}, i = j \\ \frac{p_{ij} v_i}{v} = \frac{q_{ij}}{v}, i \neq j \end{cases}$ - describes probability that state will transfer to a certain state after one transition happens - can also have fictitious transitions where state ”transitions” to itself

$$\hat{P} = \begin{bmatrix} \frac{1}{6} & \frac{2}{6} & \frac{3}{6} & 0 \\ \frac{4}{6} & 0 & \frac{2}{6} & 0 \\ 0 & \frac{2}{6} & \frac{2}{6} & \frac{2}{6} \\ \frac{1}{6} & 0 & \frac{3}{6} & \frac{2}{6} \end{bmatrix}$$

$$X(t) = \hat{X}_{N(t)}$$

$$p_{ij}(t) = P(X(t) = j|X(0) = i) = \sum_{k=0}^{M} P(X(t) = j)|X(0) = i, N(t) = k)P(N(t) = k|X(0) = i) =$$

$$\sum_{k=0}^{\infty} P(X_{N(t)} = j|\hat{X}_0 - i, N(t) \hat{=} k)P(N(t) = k|\hat{X}_0 = i)$$

$$= \sum_{k=0}^{\infty} P(\hat{X}_k = j|\hat{X}_0 = i) = \sum_{k=0}^{\infty} P(\hat{X}_k = j|\hat{X}_0 = i)\frac{e^{-vt}(vt)^k}{k!} =$$

$$\sum_{n=0}^{\infty} \hat{P_{ij}}^{(n)} e^{-vt}\frac{(vt)^n}{n!}$$

$$P(t) = \sum_{k=0}^{\infty} e^{-vt}\frac{(vt)^k}{k!}\hat{P}^k \approx \sum_{k=0}^{M} e^{-vt}\frac{(vt)^k}{k!}\hat{P}^k, M = max(vt + 5\sqrt{vt}, 20)$$

12.2 Example

Find $P(5)$

$$R = \begin{bmatrix} 0 & 2 & 3 & 0 \\ 4 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 0 & 3 & 0 \end{bmatrix}$$

$$\hat{v}t = 6(5) = 30$$

$$P(5) = \sum_{k=0}^{M} \frac{\hat{P}^k e^{-30}(30)^k}{k!}$$

$$M = max(30 + 5\sqrt{30}, 20) = 58$$

$$P(5) = \begin{bmatrix} 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \end{bmatrix}$$

13 Occupancy Times

- The expected time that a CTMC spends in a given state during a specified time interval is called the occupancy time

- Let $m_{ij}(T)$ be occupancy time in state j during the interval $[0, T]$, starting in state i at time zero

$$Y_j(t) = \begin{cases} 1, X(t) = j \\ 0, else \end{cases}$$
$$m_{ij}(T) = E(\int_0^T Y_j(t)dt | X0 = i) = \int_0^T p_{ij}(t)dt$$

Using uniformization:

$$p_{ij}(t) = \sum_{k=0}^{\infty} e^{-vt} \frac{(vt)^k}{k!} [\hat{P^k}]_{ij}$$
$$m_{ij}(T) = \int_0^T p_{ij}(t)dt = \sum_{k=0}^{\infty} \frac{1}{v} P(Y > k)[\hat{P^k}]_{ij}, Y \sim Pois(vT)$$

Occupancy matrix is given by

$$M(t) = \sum_{k=0}^{\infty} \frac{1}{v} P(Y > k) \hat{P^k}$$

If CTMC incurs an expected cost of $c(j)$ every time state j is visited, total cost incurred up to time T is - if initial state is i

$$\sum_{j \in S} c(j) m_{ij}(T)$$

13.1 Example

The machine is up for Exp(μ) and down for Exp (λ).Suppose the system operates 24 hours a day. How do we compute the expected amount of time during which the machine is on during the next 24 hours. Assume the stock level is 4 at the beginning. How to compute expected amount of time during next 24 hours?

Machie is on when system is in states $1, 2, 3, 4$ - need to compute

$$m_{51}(24) + m_{52}(24) + m_{53}(24) + m_{54}(24)$$
$$Q = \begin{bmatrix} -6 & 6 & 0 & 0 & 0 & 0 \\ 5 & -11 & 6 & 0 & 0 & 0 \\ 0 & 5 & -11 & 6 & 0 & 0 \\ 0 & 0 & 5 & -11 & 6 & 0 \\ 0 & 0 & 0 & 0 & -5 & 5 \\ 0 & 0 & 5 & 0 & 0 & -5 \end{bmatrix}, v = 11$$
$$\hat{P} = \begin{bmatrix} \frac{5}{11} & \frac{6}{11} & 0 & 0 & 0 & 0 \\ \frac{5}{11} & 0 & \frac{6}{11} & 0 & 0 & 0 \\ 0 & \frac{5}{11} & 0 & \frac{6}{11} & 0 & 0 \\ 0 & 0 & \frac{5}{11} & 0 & \frac{6}{11} & 0 \\ 0 & 0 & 0 & 0 & \frac{6}{11} & \frac{5}{11} \\ 0 & 0 & \frac{5}{11} & 0 & 0 & \frac{6}{11} \end{bmatrix}$$
$$M(24) = \sum_{k=0}^{\infty} \frac{1}{v} P(Y > k) \hat{P}^k \approx \sum_{k=0}^{345} \frac{1}{11} P(Y > k) \hat{P}^k$$
$$Y \sim Pois(vT)$$

First Passage Times - want to find first passage time to state j from state i - define $T_j = min(t \geq 0 : X(t) = j)$ as the first passage time into state j -

$$\mu_{ij} = E(T_j | X_0 = i)$$
$$= \sum_{k \in S} E(T_j | X_0 = i, X(Y) = k) P(X(Y) = k | X(0) = i)$$
$$\sum_{k \neq j} (\frac{1}{v_i} + \mu_{kj}) \frac{q_{ik}}{v_i} + \frac{1}{v_i} \frac{q_{ij}}{v_i}$$
$$= \frac{1}{v_i} + \sum_{k \neq j} \frac{q_{ik}}{v_i} \mu_{kj}$$

The first passage time to state j from state i can be calculated by solving the following set of linear equations

$$v_i \mu_{ij} = 1 + \sum_{k \neq j} q_{ik} \mu_{kj}$$

For example, R, find expected amount of time CTMC takes to reach state 4 from state 1:

$$R = \begin{bmatrix} 0 & 2 & 3 & 0 \\ 4 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 0 & 3 & 0 \end{bmatrix}$$

$$v_1 = 5, v_2 = 6, v_3 = 4, v_4 = 4$$
$$5\mu_{13} = 1 + 2\mu_{24} + 3\mu_{34}$$
$$6\mu_{24} = 1 + 4\mu_{14} + 2\mu_{34}$$
$$4\mu_{34} = 1 + 2\mu_{24}$$
$$\mu_{14} = 1.27$$

14 Queueing Models

- Single Stage System - jobs go into one stage of a server
- Serial System - jobs go into many stages
- Network of Queues - many stages with different types of jobs where jobs have different routing of stages

Basic Questions -

- How many customers are in queue
- How long does a customer spend in a queue
- How many customers are lost due to capacity limitations
- How busy are the servers

Measures -

- L - expected number of customers in the system
- L_Q - expected number of customers waiting in queue
- W - expected amount of time a customer spends in the system
- W_Q - expected amount of time a customer spends waiting in queue

Little's Law -

- If λ is average arrival rate of customers to the system

$$L = \lambda W$$
$$L_Q = \lambda W_Q$$

Kendall-Lee Notation for Queueing Systems:

- 1/2/3/4/5/6
- 1 - distribution of customer inter-arrival times
- 2- distribution of service times
  - M - Markovian(Exponential) Distribution
  - G - General Distribution
  - D - Deterministic
  - E_k - Erlang distribution
- 3 - number of servers
- 4 - maximum number of customers allowed in system(default inf)
- Service discipline(default is FCFS)
- Size of population from which customers are drawn(default inf)
- e.g. a markov chain process with inter-arrival and service times both markovian with one server is M/M/1

M/M/1 Queue

- Customers arrive one at a time
- At most one customer is in service at any time
- Customers are processed on a FCFS basis
- A_1 is arrival of 1st job, A_2 is arrival of 2nd, etc.
- D_1 is departure time of 1st job, D_2 is departure time of 2nd, etc.
- What is L(average number of customers in system)? - $L = \sum_{i=0} ip_i$, where $p_i = \frac{z_i}{\sum_{j=0}^{\infty} z_i}$
- Utilization - fraction of time server is busy
- For $z_i = (\lambda/\mu)^i = u^i$, not stable if $u \geq 1$
- if $u < 1$ - $\sum_{i=0}^{\infty} u^i = \frac{1}{1-u}$
- $p_0 = 1 - u$
- $p_i = (1-u)u^i$
- Expected Number in System $L = \sum_{i=0}^{\infty} ip_i = \frac{\lambda}{\mu-\lambda} = \frac{u}{1-u}$
- $L = \lambda W, W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda}$

- $W = W_Q + \frac{1}{\mu}, W_Q = W - \frac{1}{\mu}$
- $L_Q = \lambda W_Q$
- As utilitzation increases - expected number of customers balloons up

M/M/1/K Queue

- K - maximum number of customers that system can accommodate
- Essentially birth and death system with an upper limit
- p_0 never diverges as we take a finite sum $\sum_{i=0}^{k} u_i$, if $u = 1$, $p_i = \frac{1}{k+1}$
- $u \neq 1$ - $\sum_{i=0}^{k} u^i = \frac{1-u^{k+1}}{1-u}$
- $p_0 = \frac{1-u}{1-u^{k+1}}$
- $p_i = \frac{u^i(1-u)}{1-u^{k+1}}$
- Effective Arrival Rate - arrival rate of customers who actually go into system - $\lambda_e = \lambda(1 - p_k)$

M/M/c Queue

- c - number of servers in station where servers work in parallel
- Inter-arrival times $\sim Exp(\lambda)$
- Service times at each server $\sim Exp(\mu)$
- Birth and death process, $\lambda_1 = \lambda$. $\mu_i = min(c, i)\mu$

$$\rho_i = \begin{cases} \frac{1}{i!}(\frac{\lambda}{\mu})^i, 0 \leq i \leq c-1 \\ \frac{c^c}{c!}\mu^i, i \geq c \end{cases}$$

$$p_0 = [\sum_{i=0}^{c-1} \frac{1}{i!}(\frac{\lambda}{\mu})^i + \frac{c^c}{c!}\frac{u^c}{1-u}]^{-1}$$

$$u = \frac{\lambda}{c\mu}$$

- Probability that an arriving customer has to wait?

$$\sum_{i=c}^{\infty} p_i = \frac{p_c}{1-u}$$

- Expected number of busy servers:

$$B = \sum_{i=0}^{\infty} min(i, c)p_i = \frac{\lambda}{\mu}$$

- Limiting distribution only exists if $\lambda \leq c\mu$ - number of customers in system in other case is infinite
- Utilization

$$u = \frac{\lambda}{c\mu} < 1$$

- Average Number in the system

$$L = \sum_{i=0}^{\infty} ip_i = \frac{\lambda}{\mu} + p_c \frac{u}{(1-u)^2}$$

- Average wait time in the system

$$W = \frac{1}{\mu} + \frac{1}{c\mu}\frac{p_c}{(1-u)^2}$$

- Throughput $= \lambda$
- Probability of all servers busy

$$p_c = \frac{1}{\left[\frac{c!}{(cu)^c}\sum_{i=0}^{c-1}\frac{(cu)^i}{i!} + \frac{1}{1-u}\right]}$$

14.1 Example

- Suppose we would like to design a service system where customers do not wait more than 5 minutes on average.
- We estimate the average service time to be 3.8 minutes and the arrival rate to be 47.2 per hour.
- Assuming arrivals follow a Poisson process and service times are exponential, how many servers are required?
- Essentially an M/M/c queue where we want to find c where $W_Q \leq 5$

$$\lambda = 47.2$$
$$\frac{1}{\mu} = 3.8, \mu = 15.79$$
$$u = \frac{47.2}{15.79c} < 1$$
$$W_Q = W - \frac{1}{\mu}$$
$$= \frac{1}{c\mu}\frac{p_c}{(1-u)^2}$$
$$W_Q(c=3) = 5.9hrs$$
$$W_Q(c=4) = 0.0316hrs$$

14.2 Example

- Consider an M/M/c queue. Suppose each server costs $\$C_s$ per hour and the cost of waiting is quantified as $\$C_w$ per customer per hour. What is the optimal number of servers to employ to minimize the long-run average cost per hour?
  - Hourly Employee Cost: $C_s c$
  - Cost of waiting: $C_w L$
  - Want to minimize $C_s c + C_w L$
  - Find minimum using a grid-search, derivative wrt c is very tedious

$G/G/\infty$ Queue

- Customer never has to wait for service to begin
- Customer's entire stay in the system may be thought of as his or her service time

$$W_Q = 0$$
$$W = E(S)$$
$$L = \lambda E(S)$$

Multiple Servers, non-identical service rates

- Job inter-arrival times are $Exp(\lambda)$
- Service times at server 1 are $Exp(\mu_1)$
- Service times at server 2 are $Exp(\mu_2)$
- K - maximum number of jobs that system can accommodate

Assuming that $\mu_1 > \mu_2$, if both machines idle, job is assigned to faster machine

Can be modeled as a CTMC - $S = \{0, 1, 1', 2, 3, 4\}$, where 1 describes the state where only server 1 is working and $1'$ describes the state where only server 2 is working - can calculate p_i from there

$$\lambda = 3$$
$$\mu_1 = 3$$
$$\mu_2 = 2$$
$$K = 4$$

- Expected number of customers in system

$$L = 0p_0 + 1(p_1 + p_{1'}) + 2p_2 + 3p_3 + 4p_4$$

- Expected number in queue

$$L_Q = 0p_0 + 0(p_1 + p_{1'}) + 0p_2 + 1p_3 + 2p_4$$

- Throughput - since upper bound on capacity

$$TH = \lambda(1 - p_4) = \lambda_e$$
$$TH = \mu_1 p_1 + \mu_2 p_{1'} + (\mu_1 + \mu_2)(p_2 + p_3 + p + 4)$$

- Expected number of busy machines

$$E(B) = 1(p_1 + p_{1'}) + 2(p_2 + p_3 + p_4)$$

- Server utilization

$$u = \frac{E(B)}{u}$$

$u_1 = p_1 + p_2 + p_3 + p_4 -$ Utility of machine 1 - What percent of time is machine working?

- Expected waiting time in system(Cycle)

$$L = \lambda W$$
$$W = L/\lambda_e$$
$$W_Q = \frac{W_Q}{\lambda_e}$$

- Expected service time

$$W - W_Q = \frac{L - L_Q}{\lambda_e}$$

15 Exponential Queues in Series

- In many situations, customer's service is not complete until customer has been served in multiple number of stages
- If exponential, can decompose into n single stage queues
- Probability distribution of the number of customers at station j is the same as an $M/M/c_j$ system with arrival rate λ_j and service rate μ_j, does not hold if not exponential
- If $\exists j, c_j\mu_j \leq \lambda_j$, then no steady state distribution of customers exists
- In finding L, add up expected number at each station
- To find W, apply $L = \lambda W$, $\lambda = \sum r_i$

16 Markov Decision Processes

Sequential Decision Making

- Set of decision epochs(when do decisions get made?)
- Set of system states
- Set of avaliable actions
- Set of state and action dependent immediate rewards or costs

16.1 Example

Want to find the shortest path between two nodes in a graph

- Greedy solution by pickign shortest at each stage - not necessarily optimal
- Can enumerate or use Djikstra's - not efficient
- Work backwards - define $f(i)$ as the minimum distance required to reach J from $i, i \in \{A, B, ..., I\}$
- $f(i)$ =min cost from i to J - "cost to go function"

MDP Notation

- At beginning of each stage/ period, process is observed to be in some state s in S, where S is the state space
- The total number of periods is N, N is the planning horizon
- At each period, an action a in $A(s)$ must be selected, $A(s)$ is the set of feasible actions in state s
- Process evolves(probabilistically) to another state at the beginning of the next period - next state depends on the history only through the current state and action - Markovian
- Probability of making a transition from state s to j given action a is p^a_{sj}
- Immediate(expected) return resulting from action a in state s is $r(s, a)$
- $v_t(x)$ - maximum total reward when there are t periods to go until the end of the planning horizon starting in state x

$$v_t(s) = max_{a \in A(s)}\{r(s, a) + \sum_{j \in S} p^a_{sj} v_{t-1}(j)\}, t \in \{1, 2, 3, ..., N\}$$

16.2 Example

- Suppose we have a machine that is either running or is broken down. If it runs throughout one week, it makes a profit of \$100.
- When the machine is broken down at the start of the week, it may either be repaired at a cost of \$40, in which case it will fail during the week with a probability of 0.4, or it may be replaced at a cost of \$150 by a new machine that is guaranteed to run through its first week of operation.
- If it is running at the start of the week and we perform preventive maintenance, the probability that it will fail during the week is 0.4. If we do not perform such maintenance, the probability of failure is 0.7. However, maintenance will cost \$20.
- Formulate this problem as an MDP for a planning horizon of N weeks with the objective of maximizing expected profit. If it fails during the week, profit is zero.
- S = $\{0, 1\}$
- Action Space A(s)

$$A(0) = \{Repair(R), Replace(P)\}$$
$$A(1) = \{Maintenance(M), Do\ Nothing(N)$$

Transition Probabilities

$$P_{00}^{R} = 0.4, P_{01}^{R} = 0.6, P_{00}^{P} = 0, P_{01}^{P} = 1, P_{10}^{M} = 0.4, P_{11}^{M} = 0.6, P_{10}^{N} = 0.7, P_{11}^{N} = 0.3$$

One-stage expected rewards, $r(s, q)$:

$$r(0, R) = 100(0.6) - 40 - 20$$
$$r(0, P) = 100 - 150 = -50$$
$$r(1, M) = -20 + 0.6(100) = 40$$
$$r(1, N) = 0.3(100) = 30$$

Value Functions:

$v_n(5)$ = max total expected profit starting in states and there are n periods to go until the end of the planning horizon

$$v_n(0) = max\{20 + 0.4v_{n-1}(0) + 0.6v_{n-1}(1), -50 + v_{n-1}(1)$$
$$v_n(1) = max\{40 + 0.4v_{n-1}(0) + 0.6v_{n-1}(1), 30 + 0.7v_{n-1}(0) + 0.3v_{n-1}(1)$$

16.3 Example

- Consider single flight leg with capacity C
- Requests originate from m fare classes, where class i pays fare r_i and $r_1 > r_2 > ...r_m$
- Each passenger requests a single seat upon arrival
- Reservations are accepted during a horizon of fixed length T. The booking horizon $[0, T]$ is divided into N subintervals in such a way that, during each subinterval, the probability of two or more requests is 0. These subintervals, or periods, are numbered in reverse order, with period N corresponding to the start of the booking horizon, and period 0 denoting the scheduled departure time
- p_{in} - probability that a class i customer arrives in period n
- p_{0n} - probability of no arrivals during period n, and

$$\sum_{i=0}^{m} p_{in} = 1$$

- The objective is to maximize the total expected revenue for the booking horizon by deciding whether to accept or reject a customer request upon arrival

MDP Model:

- x - number of reservatiosn accepted
- $u_n(x)$ - maximum expected revenue from periods $n, n-1, ..., 0$ when x reservations have been accepted, $x = 0, ..., C$

$$n \in 1, ..., N$$

$$u_n(x) = \begin{cases} \sum_{i=1}^{m} p_{in} max\{r_i + u_{n-1}(x+1), u_{n-1}(x)\} + p_{0n}u_{n-1}(x), x = 0, 1, ...C-1 \\ u_{n-1}(x), x = C \end{cases}$$

$$u_0(x) = 0, x \in 0, 1, ..., C$$

Test:

- Absorption Probabilities
- DTMC
- CTMC
- Queueing
- MDPs