

## Theorems and Definitions

### Chapter 1

- Multistep Transition Probabilities
- Classification of States
- Stationary Distributions
- Detailed Balance Condition
  - Reversibility
  - Metropolis Hastings
- Limit Behavior
- Exit Times
- Infinite State Spaces

### Chapter 2

- Exponential Distribution
- Defining the Poisson Process
- Compound Poisson Processes
- Transformations
  - Superposition
  - Conditioning

### Chapter 4

- Continuous Time Markov Chains
  - Transition Probabilities
  - Limiting Behavior
  - Detailed Balance Condition
  - Exit Distributions
  - Exit Times

### Pinsky

- Limiting Distributions
- Classification of States
- Limit Theorem of MCs

# Theorems and Definitions

---

## Chapter 1

---

### Multistep Transition Probabilities

---

- Absorbing state:  $p(x, x) = 1$
- **D Theorem 1.1.** The m step transition probability  $P(X_{n+m} = j | X_n = i)$  is the mth power of the transition matrix p.
- Chapman-Kolmogorov Equation:  $p^{m+n}(i, j) = \sum_k p^m(i, k)p^n(k, j)$ 
  - Intuitively, to go from i to j in m + n steps, we have to go from i to some state k in m steps and then

from  $k$  to  $j$  in  $n$  steps. The Markov property implies that the two parts of our journey are independent.

## Classification of States

---

- First time to return to  $y$ :  $T_y = \min \{n \geq 1 : X_n = y\}$ 
  - Does not count time zero
- Probability of process returning to  $y$  starting at  $y$ :  $\rho_{yy} = P_y(T_y < \infty)$
- Stopping time:  $\{T = n\}$ 
  - The occurrence (or nonoccurrence) of the event “we stop at time  $n$ ,” can be determined by looking at the values of the process up to that time:  $X_0, \dots, X_n$
- **D Theorem 1.2** (Strong Markov Property). Suppose  $T$  is a stopping time. Given that  $T = n$  and  $X_t = y$ , any other information about  $X_0 \dots X_T$  is irrelevant for predicting the future, and  $X_{T+k}, k \geq 0$  behaves like the Markov chain with initial state  $y$ .
- Transient:  $\rho_{yy} < 1$ 
  - The probability of returning  $k$  times is  $\rho_{yy}^k \rightarrow 0$  as  $k \rightarrow \infty$ , eventually the Markov chain does not find its way back to  $y$ .
- Recurrent:  $\rho_{yy} = 1$ 
  - The probability of returning  $k$  times  $\rho_{yy}^k = 1$  so the chain returns to  $y$  infinitely many times.
  - An absorbing state is the strongest version of recurrence  $p(y, y) = 1$
- D Lemma 1.3. Suppose  $P_x(T_y \leq k) \geq \alpha > 0$  for all  $x$  in the state space  $S$ . Then  $P_x(T_y > nk) \leq (1 - \alpha)^n$
- D Definition 1.1. We say that  $x$  communicates with  $y$  and write  $x \rightarrow y$  if there is a positive probability of reaching  $y$  starting from  $x$ , that is, the probability  $\rho_{xy} = P_x(T_y < \infty) > 0$ 
  - In english, we can travel to  $y$  starting from  $x$  with positive probability
- D Lemma 1.4. If  $x \rightarrow y$  and  $y \rightarrow z$ , then  $x \rightarrow z$ . Communication is transitive.
- **D Theorem 1.5.** If  $\rho_{xy} > 0$  but  $\rho_{yx} < 1$  then  $x$  is transient
  - This allows us to identify all the transient states when the state space is finite.
- Lemma 1.6. If  $x$  is recurrent and  $\rho_{xy} > 0$  then  $\rho_{yx} = 1$ 
  - Directly follows from T1.5.
- Closed: A set  $A$  is closed if it is impossible to get out.  $p(i, j) = 0$  for  $i \in A, j \notin A$
- Irreducible: set  $B$  irreducible if whenever  $i, j \in B$   $i$  and  $j$  communicate.
  - In english, a set of states that all communicate with each other.
  - The whole MC is irreducible if all states communicate with each other.
- **D Theorem 1.7:** If  $C$  is a finite closed and irreducible set, then all states in  $C$  are recurrent.
- **D Theorem 1.8** (Decomposition Theorem): If the state space  $S$  is **finite**, then  $S$  can be written as a disjoint union  $T \cup R_1 \cup \dots \cup R_k$  where  $T$  is a set of transient states and the  $R_i, 1 \leq i \leq k$ , are closed irreducible sets of recurrent states.
  - For finite states, we can create groups based on which states communicate with each other.
- Lemma 1.9. If  $x$  is recurrent and  $x \rightarrow y$ , then  $y$  is recurrent.
- Lemma 1.10. In a finite closed set there has to be at least one recurrent state.

- Lemma 1.11 / 1.12.  $E_x N(y) = \rho_{xy} / (1 - \rho_{yy}) = \sum_{n=1}^{\infty} p^n(x, y)$ .
  - Expected number of visits to y starting from x
  - Recurrence defined by  $E_y N(y) = \infty$ , expected # of time periods that process is in state y is infinite.
- **D Theorem 1.13.** y is recurrent if and only if  $\sum_{n=1}^{\infty} p^n(y, y) = E_y N(y) = \infty$ 
  - Necessary and sufficient condition for recurrence.
  - In english, y is recurrent if the expected number of visits to y starting at y is infinite. Alternatively, y is recurrent if the infinite sum of transition probabilities from y to y is infinite.

## Stationary Distributions

---

- Define:  $q(i) = P(X_0 = i)$
- Stationary Distribution: If  $\pi p = \pi$ , then  $\pi$  is called a stationary distribution. If the distribution at time 0 is the same as the distribution at time 1, then by the Markov property it will be the distribution at all times
  - That after all the transfers are made, the amount of sand that ends up at each site is the same as the amount that starts there. Not necessarily from x to y and y to x - transfers around the whole system.
- Definition 1.2. A transition matrix p is said to be doubly stochastic if its COLUMNS sum to 1
- **D Theorem 1.14.** If p is a doubly stochastic transition probability for a Markov chain with N states, then the uniform distribution,  $\pi(x) = 1/N$  for all x, is a stationary distribution.

## Detailed Balance Condition

---

- $\pi(x)p(x, y) = \pi(y)p(y, x)$
- A stronger condition than the stationary distribution. In this case the detailed balance condition says that the amount of sand going from x to y in one step is exactly balanced by the amount going back from y to x.
- Birth and death chains are defined by the property that the state space is some sequence of integers and it is impossible to jump by more than one.
- Random walks on graphs: an adjacency matrix A.u; v/, which is 1 if there is an edge connecting u and v and 0 otherwise. Then  $p(u, v) = \frac{A(u, v)}{d(u)}$  and  $\pi(u)p(u, v) = cA(u, v) = cA(v, u) = \pi(v)p(v, u)$

## Reversibility

- **Theorem 1.15:** Fix n and let  $Y_m = X_{n-m}$  for  $0 \leq m \leq n$ . Then  $Y_m$  is a Markov chain with transition probability  $\hat{p}(i, j) = P(Y_{m+1} = j | Y_m = i) = \frac{\pi(j)p(j, i)}{\pi(i)}$ .
  - In english, if we watch the process  $X_m, 0 \leq m \leq n$  backwards, then it is a MC. For transition probability  $p(i, j)$  with stationary distribution  $\pi(i)$  and Let  $X_n$  be a realization of the Markov chain starting from the stationary distribution,  $P(X_0 = i) = \pi(i)$
  - When  $\pi$  satisfies DBC,  $\hat{p}(i, j) = \frac{\pi(j)p(j, i)}{\pi(i)} = p(i, j)$

## Metropolis Hastings

- MC  $q(x, y)$ , and a move is accepted with probability  $r(x, y) = \min \left\{ \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right\}$
- We get transition probability  $p(x, y) = q(x, y)r(x, y)$ . We run the chain for a long time until reaches equilibrium, then take samples at widely separated times. For an irreducible finite S chain, Theorem 1.22 guarantees  $\frac{1}{n} \sum_{m=1}^n f(X_m) \rightarrow \sum_x f(x)\pi(x)$

## Limit Behavior

- Period: The period of a state is the largest number that will divide all the  $n \geq 1$  for which  $p^n(x, x) > 0$
- Lemma 1.17. If  $\rho_{xy} > 0$  and  $\rho_{yx} > 0$ , then x and y have the same period - periodicity is a class property.
  - If you show aperiodicity for some states in an irreducible chain, then the whole chain is aperiodic.
- Lemma 1.18. If  $p(x, x) > 0$  then x has period 1.
  - Very useful in showing aperiodicity as a sufficient condition.
- Define
  - I: p is irreducible,
  - A: aperiodic (all states have period 1),
  - R: all states recurrent,
  - S: stationary distribution  $\pi$  exists
- **D Theorem 1.19 Convergence Theorem:**  $p^n(x, y) \rightarrow \pi(y)$  as  $n \rightarrow \infty$  for I, A, S
- **Theorem 1.20 (Asymptotic Frequency).** Suppose I and R. If  $N_n(y)$  be the number of visits to y up to time n, then  $\frac{N_n(y)}{n} \rightarrow \frac{1}{E_y T_y}$ 
  - The average probability of visiting y converges to the 1 over the expected min time to return to y starting from y.
  - If state space is infinite, we can have  $E_y T_y = \infty$  and the limit is 0
- **Theorem 1.21.** If I and S hold, then  $\pi(y) = 1/E_y T_y$  and hence the stationary distribution is unique.
- **Theorem 1.22.** Suppose I, S, and  $\sum_x |f(x)|\pi(x) < \infty$  then  $\frac{1}{n} \sum_{m=1}^n f(X_m) \rightarrow \sum_x f(x)\pi(x)$ 
  - Just extending 1.20
- **Theorem 1.23.** Suppose I, S.  $\frac{1}{n} \sum_{m=1}^n p^m(x, y) \rightarrow \pi(y)$ 
  - The average of the first n values of  $p^m(x, y)$  converges even in the periodic case
- Ergodic: State i is positive recurrent if i is recurrent and, starting in i, the expected return time to i is finite. Positive recurrent, aperiodic states are called ergodic.
- Limiting Probabilities: An irreducible, ergodic MC has limit independent of i  $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^n$ , where  $\pi_j$  is the solution to the system  $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$ ,  $j \geq 0$  s.t.  $\sum_{j=0}^{\infty} \pi_j = 1$ . Note  $\pi_j$  also the long run proportion of time that the process will be in state j

## Exit Times

- **Theorem 1.28:** Given a set F, let  $V_F = \min \{n \geq 0 : X_n \in F\}$ . Consider a Markov chain with state space S. Let A and B be subsets of S, so that  $C = S - (A \cup B)$  is finite. Suppose  $h(a) = 1$  for  $a \in A$ ,  $h(b) = 0$  for  $b \in B$  and that for x in C we have  $h(x) = \sum_y p(x, y)h(y)$ . If  $P_x(V_A \wedge V_B < \infty) > 0$  for all x in C, then  $h(x) = P_x(V_A < V_B)$

- **Theorem 1.29.** Let  $V_A = \inf \{n \geq 0 : X_n \in A\}$ . Suppose  $C = S - A$  is finite, and that  $P_x(V_A < \infty) > 0$  for any  $x$  in  $C$ . If  $g(a) = 0$  for all  $a$  in  $A$ , and for  $x$  in  $C$  we have  $g(x) = 1 + \sum_y p(x, y)g(y)$ , then  $g(x) = E_x(V_A)$ .
  - $g(x)$  is the expected time to some exit state starting in state  $x$ . For some exit state (absorbing)  $g(x) = 0$ , while for non absorbing states,  $g(x) = 1 + f(g(x))$ , ie. we add one since we have to take at least one step to exit, then have some system of equations dependent on other  $g(x)$  values.
  - This Theorem gives us a way to guess and verify the answer while ensuring that our solution is unique.

## Infinite State Spaces

- Positive recurrent:  $x$  is said to be positive recurrent if  $E_x T_x < \infty$
- Null recurrent: if a state is recurrent but not positive recurrent, meaning  $P_x(T_x < \infty) = 1$  and  $E_x T_x = \infty$ , it is null recurrent
  - In a reflecting random walk -  $p < 1/2$  0 is PR,  $p = 1/2$  0 is NR,  $p > 1/2$  0 is transient
  - In reflecting random walk, null recurrence thus represents the borderline between recurrence and transience. This is what we think in general when we hear the term.
- **Theorem 1.30:** For an irreducible chain the following are equivalent: (i) Some state is positive recurrent. (ii) There is a stationary distribution. (iii) All states are positive recurrent.
  - Think from Theorem 1.21,  $\pi(x) = \frac{1}{E_x T_x}$  then for null recurrent  $E_x T_x = \infty$ ,  $\pi(x) = 0$
- Lemma 1.31: The extinction probability  $\rho$  is the smallest solution of the equation  $\phi(x) = x$  with  $0 \leq x \leq 1$ .
  - A population in which each individual in the  $n$ th generation independently gives birth, producing  $k$  children (members of the  $n+1$  generation) with probability  $p_k$ .
  - Number of individuals at time  $n$  is  $X_n$ .  $\phi(\theta) = \sum_{k=0}^{\infty} p_k \theta^k$  is the generating function of the distribution  $p_k = P(Y_n = k)$ .
  - Let  $\rho$  be the probability that the process dies out starting from  $X_0 = 1$ , then  $\rho = \sum_{k=0}^{\infty} p_k \rho^k$ . It always has the trivial root  $\rho = 1$  since  $\phi(\rho) = \sum_{k=0}^{\infty} p_k \rho^k = 1$ .
  - Avg number of offspring of one individual:  $\mu = \sum_{k=0}^{\infty} k p_k$

## Chapter 2

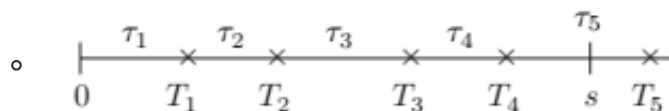
### Exponential Distribution

- Distribution function of  $T = \exp(\lambda)$   $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$ ,  $t \geq 0$ . Note  $\lambda$  is a rate (given a time, take reciprocal)
- Density function  $f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$
- Mean  $\frac{1}{\lambda}$ , Variance  $\frac{1}{\lambda^2}$
- Memoryless:  $P(T > t + s | T > t) = P(T > s)$ .
  - In words, "if we've been waiting for  $t$  units of time then the probability we must wait  $s$  more units of time is the same as if we haven't waited at all."

- Exponential Races: For  $S = \exp(\lambda)$ ,  $T = \exp(\mu)$ ,  $S \perp T$ . Then  $\min(S, T)$  has an exponential distribution with rate  $\lambda + \mu$ 
  - In order for the minimum of S and T to be larger than t, each of S and T must be larger than t
  - $P(\min(S, T) > t) = P(S > t, T > t) = P(S > t)P(T > t) = e^{-\lambda t} e^{-\mu t} = e^{-(\lambda + \mu)t}$
  - Who finishes first:  $P(S < T) = \int_0^\infty f_S(s)P(T > s)ds = \frac{\lambda}{\lambda + \mu}$
  - What is the time until both processes are done? Convert times to rates, The total service rate is  $\lambda + \mu$ , so the time until the first customer completes service is  $\exp(\lambda + \mu)$  with mean  $\frac{1}{\lambda + \mu}$ . Total waiting time is  $\frac{1}{\lambda + \mu} + P(S < T)\frac{1}{\lambda} + P(T < S)\frac{1}{\mu}$ . Another way to say this is  $E \max\{S, T\} = \frac{1}{\mu + \lambda} + \frac{\lambda}{\lambda + \mu} \cdot \frac{1}{\mu} + \frac{\mu}{\lambda + \mu} \cdot \frac{1}{\lambda}$ . See example 2.1, page 97.
- **Theorem 2.1:** Let  $T_i = \text{exponential}(\lambda_i)$ ,  $1 \leq i \leq n$  and  $V = \min(T_1, \dots, T_n)$ , I be the (random) index of the  $T_i$  that is smallest. Then
  1.  $P(V > t) = \exp(-(\lambda_1 + \dots + \lambda_n)t)$
  2.  $P(I = i) = \frac{\lambda_i}{\lambda_1 + \dots + \lambda_n}$
  3. I and V are independent
  - In words: the **time to first event** is exponential with rate  $\lambda_1 + \dots + \lambda_n$ , so the mean time to first event is  $\frac{1}{\lambda_1 + \dots + \lambda_n}$ .  $P(I = i)$  is the probability that a given part is the first to event. For time to second event / failure, see Example 2.2, Page 98.
- **Theorem 2.2:** Let  $\tau_1, \tau_2, \dots$  be independent  $\exp(\lambda)$ . The sum  $T_n = \tau_1 + \dots + \tau_n \sim \text{gamma}(n, \lambda)$ . That is, the density function  $T_n$  is given by  $f_{T_n}(t) = \lambda e^{-\lambda t} \cdot \frac{(\lambda t)^{n-1}}{(n-1)!}$  for  $t \geq 0$ , 0 otherwise.

## Defining the Poisson Process

- Definition: We say that X has a Poisson distribution with mean  $\lambda$ , or  $X = \text{Poisson}(\lambda)$ , if  $P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$  for  $n = 0, 1, 2, \dots$
- **Theorem 2.3:** For any  $k \geq 1$ ,  $EX(X - 1) \dots (X - k + 1) = \lambda^k$ , and hence  $\text{var}(X) = \lambda$
- **Theorem 2.4:** If  $X_i$  are independent  $\text{Pois}(\lambda_i)$  then  $X_1 + \dots + X_k = \text{Poisson}(\lambda_1 + \dots + \lambda_n)$
- Definition:  $\{N(s), s \geq 0\}$  is a Poisson process if
  1.  $N(0) = 0$
  2.  $N(t + s) - N(s) = \text{Poisson}(\lambda t)$
  3.  $N(t)$  has independent increments, ie. if  $t_0 < t_1 < \dots < t_n$ , then  $N(t_1) - N(t_0), \dots, N(t_n) - N(t_{n-1})$  are independent.
- **Theorem 2.5:** If n is large, the binomial(n,  $\lambda/n$ ) distribution is approximately  $\text{Pois}(\lambda)$ .
- Definition: Let  $\tau_1, \tau_2, \dots$  be independent  $\exp(\lambda)$  random variables. Let  $T_n = \tau_1 + \dots + \tau_n$  for  $n \geq 1$ ,  $T_0 = 0$  and define  $N(s) = \max\{n : T_n \leq s\}$ .
  - Think of  $\tau_n$  as times between arrivals of customers at the ATM, so  $T_n = \tau_1 + \dots + \tau_n$  is the arrival time of the nth customer and **N(s) is the number of arrivals by time s**. Below,  $N(s) = 4$  when  $T_4 \leq s < T_5$ , that is the 4th customer has arrived by time s but the 5th has not.



- Lemma 2.6:  $N(s)$  has a Poisson distribution with mean  $\lambda s$

- Lemma 2.7:  $N(t+s) - N(s), t \geq 0$  is a rate  $\lambda$  Poisson process and independent of  $N(r), 0 \leq r \leq s$ 
  - Suppose for concreteness picture) that by time  $s$  there have been four arrivals  $T_1, T_2, T_3, T_4$  that occurred at times  $t_1, t_2, t_3, t_4$ . We know that the waiting time for the fifth arrival must  $\tau_5 > s - \tau_4$ , but by the lack of memory property of the exponential distribution  $P(\tau_5 > s - t_4 + t | \tau_5 > s - t_4) = P(\tau_5 > t) = e^{-\lambda t}$ . This shows that the distribution of the first arrival after  $s$  is  $\exp(\lambda)$  and hence  $N(t+s) - N(s), t \geq 0$  is a PP.
- Lemma 2.8:  $N(t)$  has independent increments.
- **Theorem 2.9** (Poisson Approximation): Let  $X_{n,m}, 1 \leq m \leq n$  be independent random variables with  $P(X_m = 1) = p_m$  and  $P(X_m = 0) = 1 - p_m$ . Let  $S_n = X_1 + \dots + X_n, \lambda_n = ES_n = p_1 + \dots + p_n$  and  $Z_n = \text{Pois}(\lambda_n)$ . Then for any set  $A$   $|P(S_n \in A) - P(Z_n \in A)| \leq \sum_{m=1}^n p_m^2$ .
  - Why? For  $X, Y$  integer valued RVs, for a set  $A$   $|P(X \in A) - P(Y \in A)| \leq \frac{1}{2} \sum_n |P(X = n) - P(Y = n)|$ . The RHS is called the total variation distance between the distributions, denoted  $\|X - Y\|$
  - This is useful because it gives a bound on the difference between the distribution of  $S_n$  and the Poisson distribution with mean  $\lambda_n = ES_n$ . The approximation is good when  $\max_k p_k$  is small.
- Nonhomogeneous Poisson Processes: We say that  $\{N(s), s \geq 0\}$  is a PP with  $\lambda(r)$  if
  1.  $N(0) = 0$
  2.  $N(t)$  has independent increments
  3.  $N(t) - N(s)$  is Poisson with mean  $\int_s^t \lambda(r) dr$ 
    - In this case, the interarrival times are not exponential and they are not independent.

## Compound Poisson Processes

- We will embellish our Poisson process by associating an independent and identically distributed (i.i.d.) random variable  $Y_i$  with each arrival. By independent we mean that the  $Y_i$  are independent of each other and of the Poisson process of arrivals. The sum of  $Y$ 's up to time  $t$   $S(t) = Y_1 + \dots + Y_{N(t)}$  with  $S(t) = 0$  and  $N(t) = 0$
- **Theorem 2.10:** Let  $Y_1, Y_2, \dots$  be independent and identically distributed, let  $N$  be an independent nonnegative integer valued random variable, and let  $S = Y_1 + \dots + Y_N$  with  $S = 0$  when  $N = 0$ .
  1. If  $E|Y_i|, EN < \infty$  then  $ES = EN \cdot EY_i$
  2. If  $EY_i^2, EN^2 < \infty$  then  $\text{var}(S) = EN \text{var}(Y_i) + \text{var}(N)(EY_i)^2$
  3. If  $N$  is Poisson( $\lambda$ ) then  $\text{var}(S) = \lambda EY_i^2$ 
    - Suppose that the number of customers at a liquor store in a day has a Poisson distribution with mean 81 and that each customer spends an average of \$8 with a standard deviation of \$6. It follows from (i) in Theorem 2.10 that the mean revenue for the day is  $81 \times \$8 = \$648$ . The variance of total revenue by iii)  $81 \cdot \{(6)^2 + (8)^2\} = \$8100$

## Transformations

- **Theorem 2.11:** Let  $N_j(t)$  be the number of  $i \leq N(t)$  with  $Y_i = j$ .  $N_j(t)$  are independent rate  $\lambda P(Y_i = j)$  Poisson processes

- Eg,  $N_j(t)$  is the number of cars that have arrived by time  $t$  with exactly  $j$  people. There are two “surprises” here: the resulting processes are Poisson and they are independent. To drive the point home consider a Poisson process with rate 10 per hour, and then flip coins to determine whether the arriving customers are male or female. One might think that seeing 40 men arrive in one hour would be indicative of a large volume of business and hence a larger than normal number of women, but Theorem 2.11 tells us that the number of men and the number of women that arrive per hour are independent.
- **Theorem 2.12**(Poisson thinning): Suppose that in a Poisson process with rate  $\lambda$ , we keep a point that lands at  $s$  with probability  $p(s)$ . Then the result is a nonhomogeneous Poisson process with rate  $\lambda p(s)$ .
- **Theorem 2.13**: In the long run the number of calls in the system will be Poisson with mean  $\lambda \int_{r=0}^{\infty} (1 - G(r)) dr = \lambda \mu$ .
  - This applies to an M/G/inf Queue. While many people on the telephone show a lack of memory, there is no reason to suppose that the duration of a call has an exponential distribution, so we use a general distribution function  $G$  with  $G(0) = 0$  and mean  $\mu$ . Suppose that the system starts empty at time 0. The probability a call started at  $s$  has ended by time  $t$  is  $G(t-s)$ , so using Thm 2.12, the number of calls still in progress at time  $t$  is Poisson with mean  $\int_{s=0}^t \lambda(1 - G(t-s)) ds = \lambda \int_{r=0}^t (1 - G(r)) dr$ . Thm 2.13 takes the long run when  $t$  goes to infinity.
  - The mean number in the system is the rate at which calls enter times their average duration. In the argument above we supposed that the system starts empty, but the limiting result holds for any initial call number  $X_0$ .

## Superposition

- Taking one Poisson process and splitting it into two or more by using an i.i.d. sequence  $Y_i$  is called thinning. Going in the other direction and adding up a lot of independent processes is called superposition.
- **Theorem 2.14**: Suppose  $N_1(t), \dots, N_k(t)$  are independent Poisson processes with rates  $\lambda_1, \dots, \lambda_k$  then  $N_1(t) + \dots + N_k(t)$  is a Poisson process with rate  $\lambda_1 + \dots + \lambda_k$ .
  - Useful for computing the outcome of races between Poisson processes
  - Viewing the red and green Poisson processes as being constructed by starting with one rate  $\lambda + \mu$  Poisson process and flipping coins with probability  $p = \lambda/(\lambda + \mu)$  to decide the color

## Conditioning

- Let  $T_1, T_2, T_3, \dots$  be the arrival times of a Poisson process with rate  $\lambda$ , let  $U_1, U_2, \dots, U_n$  be independent and uniformly distributed on  $[0, t]$  and let  $V_1 < \dots < V_n$  be the  $U_i$  arranged in increasing order.
- **Theorem 2.15**: If we condition on  $N(t) = n$ , then the vector  $(T_1, T_2, \dots, T_n)$  has the same distribution as  $(V_1, V_2, \dots, V_n)$  and hence the set of arrival times  $\{T_1, T_2, \dots, T_n\}$  has the same distribution as  $\{U_1, U_2, \dots, U_n\}$ .
- **Theorem 2.16**: If  $s < t$  and  $0 \leq m \leq n$  then  $P(N(s) = m | N(t) = n) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}$ . That is the conditional distribution of  $N(s)$  given  $N(t) = n$  is  $\text{bin}(n, s/t)$ .

## Chapter 4

---



# Continuous Time Markov Chains

- Markov property:  $X_t, t \geq 0$  is an MC if for any  $0 \leq s_0 < s_1 < \dots < s_n < s$  and possible states  $i_0, \dots, i_n, j$ , we have  $P(X_{t+s} = j | X_s = i, X_{s_n} = i_n, \dots, X_{s_0} = i_0) = P(X_t = j | X_0 = i)$
- Transition probability: In continuous time there is no first time  $t > 0$  so we introduce for each  $t > 0$  a transition probability  $p_t(i, j) = P(X_t = j | X_0 = i)$ 
  - Eg.  $p_t(i, j) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} u^n(i, j)$
- **Theorem 4.1** (Chapman-Kolmogorov):  $\sum_k p_s(i, k) p_t(k, j) = p_{s+t}(i, j)$
- Jump rate: the transition probabilities  $p_t$  can be determined from their derivatives at 0:
 
$$q(i, j) = \lim_{h \rightarrow 0} \frac{p_h(i, j)}{h} \quad \text{for } j \neq i.$$
 If this limit exists (and it will in all the cases we consider) we will call  $q(i, j)$  the jump rate from  $i$  to  $j$ .
  - If we know the transition probability for  $t < t_0$  for any  $t_0 > 0$ , we know it for all  $t$
- Informal Construction: Let  $\lambda_i = \sum_{j \neq i} q(i, j)$  be the rate  $X_t$  leaves  $i$ . If  $X_t$  is in a state  $i$  with  $\lambda_i = 0$ , then  $X_t$  stays there forever and the construction is done. If  $\lambda_i > 0$ ,  $X_t$  stays at  $i$  for an exponentially distributed amount of time with rate  $\lambda_i$ , then goes to state  $j$  with probability  $r(i, j)$ .
  - $r(i, j)$  is the routing matrix, the probability the chain goes to  $j$  when it leave  $i$ .  $r(i, j) = q(i, j) / \lambda_i$

## Transition Probabilities

- Compute transition probability  $p$  from jump rates  $q$ :  $Q(i, j) = \begin{cases} q(i, j) & \text{if } j \neq i \\ -\lambda_i & \text{if } j = i \end{cases}$ 
  - For future computations note that the off-diagonal elements  $q(i, j)$ ,  $i \neq j$  are nonnegative, while the diagonal entry is a negative number chosen to make the row sum equal to 0.
  - If  $\lambda_i = 0$ ,  $X$  will never leave  $i$
- Kolmogorov Backward Equation:  $p'_t = Q p_t$ . Forward equation:  $p'_t = p_t Q$
- **Theorem 4.2** (Yule Process): The transition probabilities of the Yule process is given by
 
$$p_t(1, j) = e^{-\beta t} (1 - e^{-\beta t})^{j-1} \text{ for } j \geq 1, p_t(i, j) = \binom{j-1}{i-1} (e^{-\beta t})^i (1 - e^{-\beta t})^{j-i}$$
  - In this system each individual dies at rate  $\mu$  and gives birth to a new individual at rate  $\lambda$ .  $\mu = 0$  gives Yule process.
  - That is  $p_t(1, j)$  is a geom distribution with success probability  $e^{-\beta t}$  and hence mean  $e^{\beta t}$ . We get
 
$$P(\exp(-\beta t) Y_t > x) = P(Y_t > x e^{\beta t}) = (1 - 1/e^{\beta t})^{x e^{\beta t}} \rightarrow e^{-x} - e^{-\beta t} \text{ converges to a mean one exponential.}$$
- Lemma 4.3: For process  $Z(t)$  in which each individual gives birth at rate  $\lambda$  and dies at rate  $\mu$ . The transition rates are  $q(i, i+1) = \lambda i$ ,  $q(i, i-1) = \mu i$  else 0. It is enough to consider  $Z(0)=1$ ,
 
$$\frac{d}{dt} E Z(t) = (\lambda - \mu) E Z(t) \implies E_1 Z(t) = \exp((\lambda - \mu)t).$$
 For generating function  $F(x, t) = E x^{Z_0(t)}$ :
 
$$\partial F / \partial t = -(\lambda + \mu) F + \lambda F^2 + \mu = (1 - F)(\mu - \lambda F)$$

## Limiting Behavior

- Irreducible:  $X_t$  is irreducible if for any two states  $i$  and  $j$  it is possible to get from  $i$  to  $j$  in a finite number of jumps. To be precise, there is a sequence of states  $k_0 = i, k_1, \dots, k_n = j$  so that  $q(k_{m-1}, k_m) > 0$  for  $1 \leq m \leq n$ .

- Lemma 4.6: If  $X_t$  is irreducible, and  $t > 0$ , then  $p_t(i, j) > 0, \forall i, j$
- Stationary Distribution:  $\pi$  is SD if  $\pi p_t = \pi, \forall t > 0$ .
- Lemma 4.7:  $\pi$  is a SD if and only if  $\pi Q = 0$  (for  $Q(i, j) = \begin{cases} q(i, j) & \text{if } j \neq i \\ -\lambda_i & \text{if } j = i \end{cases}$  and  $\lambda_i = \sum_{j \neq i} q(i, j)$  is the total rate of transitions out of  $i$ ).
  - A test for stationarity in terms of the basic data used to describe the chain, the matrix of transition rates
- **Theorem 4.8:** If a continuous time Markov chain  $X_t$  is irreducible and has a stationary distribution  $\pi$ , then  $\lim_{t \rightarrow \infty} p_t(i, j) = \pi(j)$ 
  - Since lemma 4.6 implies that for any  $h > 0$ ,  $p_h$  is irreducible and aperiodic, so we can use Theorem 1.19 to get this result

## Detailed Balance Condition

- **Theorem 4.9:** If  $\pi(k)q(k, j) = \pi(j)q(j, k)$  for all  $j \neq k$  holds, then  $\pi$  is a stationary distribution.
  - The detailed balance condition implies that the flows of sand between each pair of sites are balanced, which then implies that the net amount of sand flowing into each vertex is 0, hence  $\pi Q = 0$
  - Easier to check than SD but does not always hold. Always holds for birth and death chains

## Exit Distributions

- Definition: Jump chain with transition probabilities  $r(i, j) = \frac{q(i, j)}{\lambda_i}$  where  $\lambda_i = \sum_{j \neq i} q(i, j)$ 
  - Let  $V_k = \min \{t \geq 0 : X_t = k\}$  be the time of the first visit to  $x$
  - Let  $T_k = \min \{t \geq 0 : X_t = k \text{ and } X_s \neq k \text{ for some } s < t\}$  be the time of first return to  $x$ . Note if  $X_0 = k$  then the chain stays at  $k$  for an amount of time that is exponential with rate  $\lambda_k$
- Definition: Branching process, has jump rates  $q(i, i+1) = \lambda i$  and  $q(i, i-1) = \mu i$ . 0 is an absorbing state.
  - For  $i \geq 0$  we have  $r(i, i+1) = \frac{\lambda}{\lambda+\mu}$   $r(i, i-1) = \frac{\mu}{\lambda+\mu}$
  - Absorption certain for  $\lambda \leq \mu$ . For  $\lambda > \mu$ , probability of avoiding extinction is  $P_1(T_0 = \infty) = 1 - \frac{\mu}{\lambda}$
  - Alternatively, can say  $\rho = P_1(T_0 < \infty)$  and for chain to leave 0,  $\rho = \frac{\mu}{\lambda+\mu} \cdot 1 + \frac{\lambda}{\lambda+\mu} \cdot \rho^2$ . Solve for  $\rho$  st  $\mu/\lambda < 1$ .
- Definition: Let  $R$  be the portion of the  $Q$  matrix with  $i, j \in C$ . Then by  $\sum_j Q(i, j)h(j) = 0$  for  $i \in C$  we have  $-\sum_{j \in C} R(i, j)h(j) = v(i)$  for  $v(i) = \sum_{j \in A} Q(i, j)$ ,  $i \in C$ . Finally then  $h(i) = (-R)^{-1}v$ , where  $(-R)^{-1}(i, j)$  is the expected amount of time spent in state  $j$  when we start in  $i$ .

## Exit Times

- First step in continuous time: For  $C = S - A$  finite,  $P_i(V_A < \infty) > 0$  for each  $i \in C$ ,  $g(a) = 0$  for  $a \in A$ , then for  $i \in C$ :  $g(i) = \frac{1}{\lambda_i} + \sum_{j \neq i} \frac{q(i, j)}{\lambda_i} g(j)$ .
  - By Theorem 1.29:  $g(i) = E_i V_A$
  - Alternatively, for  $R$  the part of  $Q$  with  $i, j \in C$ , and  $g(a) = 0$  for  $a \in A$ , then  $g = (-R)^{-1}\mathbf{1}$ . Since  $(-R)^{-1}(i, j)$  is the expected time spent at  $j$  when we start at  $i$ , when we multiply by  $\mathbf{1}$  we sum over

all  $j$  and get the exit time.

- Exit Times and SD:
  - $\pi(0) = \frac{1/\lambda_0}{1/\lambda_0 + E_1 T_0}$
  - $E_1 T_0 = \frac{1 - \pi(0)}{\lambda \pi(0)}$

## Pinsky

---

### Limiting Distributions

---

- **Regular:** On a finite number of states labeled  $0, 1, \dots, N$ ,  $P$  has the property that when raised to some power  $k$ , the matrix  $P^k$  has all of its elements strictly positive.
  - Not all Markov chains are regular. The identity matrix remains in its initial state - it has a limiting distribution but it clearly depends on its initial state
- Limiting Probability Distribution:  
 $\pi = (\pi_0, \pi_1, \dots, \pi_N)$ , where  $\pi_j > 0$  for  $j = 0, 1, \dots, N$  and  $\sum_j \pi_j = 1$ 
  - Primary interpretation:  $\pi_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)} = \lim_{n \rightarrow \infty} \Pr\{X_n = j | X_0 = i\}$ . In words, after the process has been in operation for a long duration, the probability of finding the process in state  $j$  is  $\pi_j$ , irrespective of the starting state.
  - Secondary interpretation:  $\pi_j$  gives the long run mean fraction of time that the process  $\{X_n\}$  is in state  $j$ .
- For a regular transition probability matrix,  $\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi_j > 0$  for  $j = 0, 1, \dots, N$ 
  - In the long run the probability of finding the Markov chain in state  $j$  is approximately  $\pi_j$  no matter in which state the chain began at time 0.
- Sufficient conditions for regularity:
  - For every pair of states  $i, j$  there is a path  $k_1, \dots, k_r$  for which  $P_{ik_1} P_{k_1 k_2} \dots P_{k_r j} > 0$  - ie. you can reach any state through transitions of others
  - There is at least one state  $i$  for which  $P_{ii} > 0$  - diagonals have at least one positive entry
- Theorem 4.1: Let  $P$  be a regular transition probability matrix on the states  $0, 1, \dots, N$ . Then the limiting distribution  $\pi = (\pi_0, \pi_1, \dots, \pi_N)$  is the unique nonnegative solution of the equations  
$$\pi_j = \sum_{k=0}^N \pi_k P_{kj}, \quad j = 0, 1, \dots, N \text{ and } \sum_{k=0}^N \pi_k = 1$$
  - In english limiting distribution sets up a system of equations to be solved

### Classification of States

---

- Accessible:  $j$  is said to be accessible from state  $i$  if  $P_{ij}^{(n)} > 0$  for some integer  $n \geq 0$ ; i.e., state  $j$  is accessible from state  $i$  if there is positive probability that state  $j$  can be reached starting from state  $i$  in some finite number of transitions.
- Communicate: When two state  $i$  and  $j$  are each accessible to each other.
- Equivalence classes: The states in an equivalence class are those that communicate with each other.
  - It may be possible starting in one class to enter some other class with positive probability; if so, however, it is clearly not possible to return to the initial class, or else the two classes would

together form a single class.

- Period: We define the period of state  $i$ , written  $d(i)$ , to be the greatest common divisor (g.c.d.) of all integers  $n \geq 1$  for which  $P_{ii}^{(n)} > 0$ 
  - If  $P_{ii} > 0$  for some single state  $i$ , then that state now has period 1 and is called aperiodic
  - Can look at  $ii$  entry in  $P$  at increasing powers of  $P$  to see period
  - Property 1: If  $i$  communicates with  $j$ , the  $d(i) = d(j)$  - period is constant in each class
  - Property 2: State  $i$  with period  $d(i)$ , then an integer  $N$  exists such that  $P_{ii}^{(nd(i))} > 0$  - a return to state  $i$  can occur at all large multiples of the period  $d(i)$
  - Property 3: If  $P_{ji}^{(m)} > 0$ , then  $P_{ji}^{(m+nd(i))} > 0$  for all  $n$  sufficiently large.
- Return at  $n$ th transition:  $P_{ii}^{(n)} = \sum_{k=0}^n f_{ii}^{(k)} P_{ii}^{(n-k)}$ ,  $n \geq 1$ 
  - When process starts at  $i$ , the probability it returns to state  $i$  at some time is  $f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)} = \lim_{N \rightarrow \infty} \sum_{n=0}^N f_{ii}^{(n)}$
- Recurrence:  $f_{ii} = 1$ , otherwise transient.
  - A state  $i$  is recurrent if and only if, after the process starts from state  $i$ , the probability of its returning to state  $i$  after some finite length of time is one.
- Theorem 4.2: A state  $i$  is recurrent if and only if  $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$ 
  - Conversely, suppose  $\sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty$ . Then,  $M$  is a random variable whose mean is finite, and thus,  $M$  must be finite. That is, starting from state  $i$ , the process returns to state  $i$  only a finite number of times.

## Limit Theorem of MCs

- For recurrent state  $i$ ,  $f_{ii}^{(n)}$  is the probability distribution of the first return time,  $R_i = \min \{n \geq 1; X_n = i\}$ .  $f_{ii} = 1$  and  $R_i$  is also finite.
- Mean duration between visits to state  $i$  is  $m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$ 
  - After starting in  $i$ , then, on the average, the process is in state  $i$  once every  $m_i$  units of time
- Theorem 4.3 (Basic Limit Theorem): Consider a **recurrent irreducible aperiodic** Markov chain. Let  $P_{ii}^{(n)}$  be the probability of entering state  $i$  at the  $n$ th transition given  $X_0 = i$  (initial state is  $i$ ). Let  $f_{ii}^{(n)}$  be the probability of first returning to state  $i$  at the  $n$ th transition for  $n = 0, 1, \dots$  where  $f_{ii}^{(0)} = 0$  then  $\lim_{n \rightarrow \infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^{(n)}} = \frac{1}{m_i}$ 
  - Under the same conditions as in (a),  $\lim_{n \rightarrow \infty} P_{ji}^{(n)} = \lim_{n \rightarrow \infty} P_{ii}^{(n)}$  for all states  $j$ .
  - The limit theorem applies verbatim to any aperiodic recurrent class.
- Positive Recurrent:  $\lim_{n \rightarrow \infty} P_{ii}^{(n)} > 0$  for one  $i$  in an aperiodic recurrent class, then  $\pi_j > 0$  for all  $j$  in the class of  $i$ . The class is then positive recurrent or strongly ergodic.
  - State  $i$  is pos. recurrent if  $m_i = E[R_i | X_0 = i] < \infty$
- Null Recurrent: If each  $\pi_i = 0$  and class is recurrent, then null recurrent or weakly ergodic.
  - State  $i$  is null recurrent if  $m_i = \infty$
  - Follows from  $\lim_{n \rightarrow \infty} P_{ii}^{(n)} = \pi_i = 1/m_i$
- Theorem 4.4: In a **positive recurrent aperiodic class** with states  $j = 0, 1, 2, \dots$ ,  $\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$ ,  $\sum_{i=0}^{\infty} \pi_i = 1$  and the  $\pi$ 's are uniquely determined by the set of equations  $\pi_i \geq 0$ ,  $\sum_{i=0}^{\infty} \pi_i = 1$  and  $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$  for  $j = 0, 1, \dots$

- These  $\pi'_i$ s are called a stationary probability distribution of the MC. The term “stationary” derives from the property that a Markov chain started according to a stationary distribution will follow this distribution at all points of time.
- When initial state is selected according to the stationary distribution, then
 
$$\Pr\{X_n = i, X_{n+1} = j\} = \pi_i P_{ij}$$
- Periodic Case: If  $i$  is a member of a recurrent periodic irreducible Markov chain with period  $d$ , one can show that  $P_{ii}^m = 0$  if  $m$  is not a multiple of  $d$  and that for multiple of  $d$   $\lim_{n \rightarrow \infty} P_{ii}^{nd} = \frac{d}{m_i}$
- A **unique stationary distribution**  $\pi = (\pi_0, \pi_1, \dots)$  exists for a **positive recurrent periodic irreducible** Markov chain, and the mean fraction of time in state  $i$  converges to  $\pi_i$  as the number of stages  $n$  grows to infinity. The convergence of does not require the chain to start in state  $i$ .