Lecture Notes for EE306
Module II: Discrete Stochastic Processes

Department of Electrical and Electronics Engineering
Middle East Technical University

Elif Uysal

May 10, 2022

Chapter 1

Finite State Markov Chains

1.1 Introduction

Consider observing the value of a measured quantity at times $n = 1, 2, 3, \ldots$ such that X_n is the value measured at time n.

Ex: Starting $t = 0$, a temperature sensor reports a measurement every 10 minutes. Let X_n be the measurement reported at $t = 10n$ minutes.

Ex: An autonomous vehicle sends the GPS measurement of its location every second. X_n is the value of the location at the n^{th} second.

Ex: You check the price of Gold each day. X_n is the price on day n.

We want to model X_n as a random variable. But we do not want to stop at just modeling a single value measured at a single time n: we want to consider the evolution of the sequence of values measured at a sequence of times $n = 1, 2, \ldots$. In other words, we want a model for the whole *process* $\{X_1, X_2, X_3, \ldots\}$, so that we can make some computations regarding it, and make decisions such as:

- Given the vehicle's location 3 seconds ago, what are the chances it is dangerously close to another vehicle at the moment?
- The price of an ounce of Gold is 1963 USD today. Based on its price history over the last six months, is it a good time to sell?
- Looking at the most recent blood test of a patient, a doctor decides whether or not to change the dosage of their medication.

Such decisions occur all the time in daily life as well as in engineering problems. Especially with the proliferation of the Internet of Things (IoT) technology, autonomous driving, robotics, remote monitoring, social networking, and the fact that each of us run applications on our devices that monitor distant phenomena and make decisions constantly, such questions are perhaps more relevant than ever.

Definition 1 *A* ***discrete stochastic process*** *is a sequence of random variables, $\{X_n\}$, indexed by n. Typically, $n = 1, 2, \ldots$.*

Ex: The Bernoulli process with rate p is a sequence of IID Bernoulli random variables with parameter p:

$$X_i = \begin{cases} 1 & \text{, with probability p} \\ 0 & \text{, with probability } 1 - \text{p} \end{cases}$$

The Bernoulli process is memoryless: X_n does not depend on the *history* of the process, i.e. $\{X_i, i < n\}$, or the *future*, $\{X_i, i > n\}$.

In the rest, we will explore a type of stochastic process that has memory.

1.2 Introduction to Markov Chains

The Bernoulli process considered above was the simplest nontrivial case of stochastic process, with no memory. In many real-life and engineering problems this would be a poor model. Consider, for example, the location of a vehicle, X_t, at times $t = 1$ and $t = 2$. If the vehicle has finite speed, knowing X_1 will give us some information about X_2. Similarly, the price of gold is not determined independently each day! It either rises or falls from its value the previous day. The peak temperature today exhibits dependence on not only yesterday's peak temperature, but perhaps on the temperature of the previous day, through the weather system that has been in effect for several days. As in these examples, many discrete stochastic processes have *memory.*

In some cases, the value of the stochastic process X_{n+1} at time $n + 1$ depends on the past history $X_n, X_{n-1}, X_{n-2}, \ldots, X_0$, only through the present value, X_n. We call this a *Markov Chain* (MC).

Markov Chains are among the most interesting models in probability, built on the notion of conditional probability. Here, we briefly study the topic through a sequence of simple examples.

Ex: Suppose that if it rains today, then it will rain tomorrow with probability α, and if it does not rain today, then it will rain tomorrow with probability β, independently of the weather in previous days. Define a state space, S and a Markov Chain to model this situation.

Solution: Let $S = \{0, 1\}$. Suppose we say that the process is in state 1 when it rains and state 0 when it does not rain. Let $X_n \in S$ be the state at time n. $P(X_n = 0|X_{n-1} = 0, X_{n-2}, \ldots, X_1\}$ =P($\mathrm{X}_n = 0|X_{n-1} = 0) = 1 - \beta$, and $P(X_n = 0|X_{n-1} = 1, X_{n-2}, \ldots, X_1\}$ =P($\mathrm{X}_n = 0|X_{n-1} = 1) = 1 - \alpha$.

Draw a pictorial representation of this 2-state Markov Chain.

Specify values for α and β such that the resulting process is a Bernoulli process.

Ex: Burak is determined to take Calculus I until he passes the course. There is no eviction from the undergraduate program, which means he is offered an unlimited number of chances to take the course. However, Burak does not study at all, and does not learn the material. So, every semester that he takes Calculus, the probability that he passes is p, irrespective of how many times he took the course before. Let X_n be a random variable indicating whether Burak passed the course at the end of semester n or not. $X_n = 0$ if Burak fails at the end of the n^{th} semester, and $X_n = 1$ otherwise.

(a) Consider possible sample paths of the process and contrast with that of a Bernoulli process with rate p.

(b) Argue why this is NOT a Bernoulli process.

(c) Show that this process is Markov.

(d) Note that, the Bernoulli process is also a special case of a Markov Chain, but not all binary valued Markov Chains are Bernoulli processes.

Discussion: If Burak passes on the k^{th} attempt, we can set $X_{k+1} = X_{k+2} = ... = 1$. Imagine a *sequence* of X_ns that continue forever, even after Burak passes the course. Sample paths will look like

00001111111111111111111111...

That is, an infinite number of ones, possibly preceded by a number of zeroes.

The following follow from the description of the problem:

$$P(X_{n+1}=0|X_n=i,X_{n-1},X_{n-2},\ldots,X_1) = P(X_{n+1}=0|X_n=i) = \begin{cases} 0, & \text{if } i=1 \\ 1-p, & \text{if } i=0 \end{cases}$$

and

$$P(X_{n+1}=1|X_n=i,X_{n-1},X_{n-2},\ldots,X_1) = P(X_{n+1}=1|X_n=i) = \begin{cases} 1, & \text{if } i=1 \\ p, & \text{if } i=0 \end{cases}$$

From the above, given the present value of the process, X_n, the future of the process, X_{n+1} depends only on X_n. That is, conditioned on the present, the future is independent of the past. In other words, the process is *Markov.*

1.2.1 Transition Probabilities

In the above example, the probability of going from *State 0* (*Did not pass the course yet*) to *State 1* (Done with Calculus) is

$$P(X_{n+1}=1|X_n=0)=p$$

The probability of going from *State 0* back to the same state (in other words, staying in that state) is:

$$P(X_{n+1}=0|X_n=0)=1-p$$

These are called *transition probabilities.* When in state 1, the probability of going to state 0 is zero, hence the only possibility is a *self-transition*, that is, staying in state 1.

The diagram in Fig.1.2.1, summarizes the MC visually. Go ahead and mark the transition probabilities on the diagram. Note that the transition probabilities out of every state add to 1.

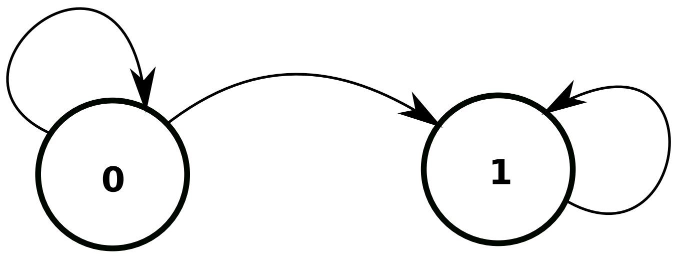

Figure 1.1: Markov Chain with one transient and one absorbing state.

In general, we define the transition probabilities[1] as:

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

For $i, j \in S$. Here $P_{ij} \geq 0$ is the probability that, when in state i, the process will next go to state j (j can be the same as i, in that case the process would simply be staying in i). As the process must make a transition into some state, we have, for all states i,:

$$\sum_{j \in S} P_{ij} = 1$$

The **transition probability matrix P** is defined as the matrix of values $\{P_{ij}\}$ for all $i, j \in S$. Due to the above, each row of **P** adds to 1.

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ . & . & . & . \\ . & . & . & . \\ P_{i0} & P_{i1} & P_{i2} & \dots . \\ . & . & . & \\ . & . & . & . \end{bmatrix}$$

Ex: For the rain/no rain example above,

$$\mathbf{P} = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}$$

[1]Our scope is limited to the case of a time-homogeneous MC.

Ex: (Random Walk) Consider a MC whose state space is the set of integers, such that, for some $a > 0$, the transitions are either to the right with probability a, or to the "left" with probability $1 - a$.

$$P_{i,i+1} = a, \quad P_{i,i-1} = 1 - a$$

for $i = 0, \pm 1, \pm 2, \ldots$ Note that the above is a MC with a countably infinite state-space.

Ex: Burak finally passed Calculus I, and he is now taking Calculus II. There is a 20 percent probability that he will pass Calculus II this semester. If not, he will either take it again next semester, or change his major, those decisions being equally likely. Every time Burak takes Calculus, the probability of passing is 20 percent.

(a) Draw the diagram of a MC for this problem, where the states are $\{1, 2, 3\}$, corresponding to *taking Calculus*, *passed Calculus*, *quit.* Mark the transition probabilities

(b) Let X_n be the state in Semester n, where $X_1 = 1$, $n = 1, 2, \ldots$. Write down the transition probability matrix $\mathbf{P}$ with entries $P_{ij} = P(X_{n+1} = j | X_n = i)$, for all $i, j \in \{1, 2, 3\}$.

Ex: (Ross, Example 4.4) Now, assume that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it has rained for the past two days, then it will rain tomorrow with probability 0.7; if it rained today but not yesterday, then it will rain tomorrow with probability 0.5; if it rained yesterday but not today, then it will rain tomorrow with probability 0.4; if it has not rained in the past two days, then it will rain tomorrow with probability 0.2. If we let the states be "rain, no rain" as before, is this a MC? If not, find a MC to represent this process.

Solution: We can form a MC with the following 4 states:

- state 0, if it rained both today and yesterday,
- state 1, if it rained today but not yesterday,
- state 2, if it rained yesterday but not today,
- state 3, if it did not rain either yesterday or today.

Please write down the transition probability matrix accordingly.

Ex: The 2 Umbrella Problem: I own two umbrellas. At time $t = 0$, one is at home, and the other is at the office. In the morning, if it is raining, I will take an umbrella with me as I walk from home to work. In the evening, as I walk back, I will again take an umbrella with me if it is raining. I never carry an umbrella if it is not raining. With this policy, I am interested in the steady-state probability of being caught in the rain with no umbrella. Construct a MC for this process, and specify $\mathbf{P}$.

1.2.2 n-step Transition Probabilities

We denote the probability of going from i to j in n steps as P_{ij}^n:

Definition 2

$$P_{ij}^n = P(X_{n+k} = j | X_k = i), \;\; n \geq 0, i, j \in S$$

Note that,

$$\begin{aligned} P_{ij}^{n+m} &= \sum_k P(X_{n+m+k} = j | X_{m+k} = l, X_k = i) P(X_{m+k} = l | X_k = i), \\ & n, m \geq 0, i, j, l \in S \end{aligned}$$

which implies that:

$$P_{ij}^{(n+m)} = \sum_k P_{il}^m P_{lj}^n \;\; n, m \geq 0, i, j, l \in S$$

In short,

$$P^{(n+m)} = P^{(n)} P^{(m)}$$

Ex: Burak is now taking EE202. His algorithm did not change, and is expressed by the 3-state MC represented below.

(a) Show that $P(X_{n+2} = j | X_n = i)$ is given by $\sum_{k=1}^{3} p_{ik} p_{kj}$. Note that this is the product of the i^{th} row and the j^{th} column of $\mathbf{P}$. In other words, the probability of going from state i to state j in two steps is given by the ij^{th} entry of the $\mathbf{P}^2$ matrix. Compute and write down the matrix $\mathbf{P}^2$.

(b) Argue similarly that the n step transition probabilities are given by entries of the matrix $\mathbf{P}^n$.

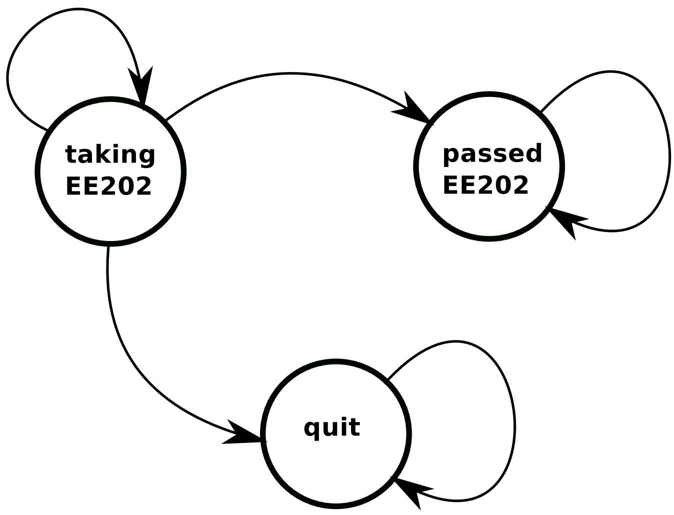

Figure 1.2: Markov Chain with two absorbing states.

(c) Notice that the second and third rows of P^n will always be fixed, while the first row changes with n. Show that the first entry of the first row, P^n_{11} is equal to 0.4^n. Similarly, let the second entry be $P^n_{12} = a_n$. Observe that $a_n = 0.4a_{n-1} + 0.2$. Let $P^n_{13} = b_n$. Observe that $b_n = 0.4b_{n-1} + 0.4$.

(d) Based on the observations just made, what is the matrix P^n converging to, as n grows? Interpret the results: what is the probability that Burak will eventually pass the course? What is the probability that he will change his major?

The above example is a Markov Chain with two *absorbing states*: states 2 and 3. If we start in State 1, we will eventually leave this state and get absorbed in either state 2 or 3.

Now, we will compute the probability that Burak eventually passes EE202, through an alternative approach:

(a) Let a be the probability that the eventual state is passing EE202, given that we start at time $n = 1$ by taking EE202: that is, $a = P(\lim_{n\to\infty} X_n = 2|X_1 = 1)$, and b be the probability that the eventual state is quitting: that is, $b = P(\lim_{n\to\infty} X_n = 3|X_1 = 1)$. Argue that $b = 1 - a$.
(Hint: We have three types of sequences possible: those that hit 2 and get stuck there, those that hit 3 and get stuck there, and those that always stay at 1. The probability of the last kind of sequence is zero. To see this, consider the first column of the matrix $\mathbf{P}^n$ computed above, and see that it converges to the all-zero vector.)

(b) Show that we can write the following equation for a:

$$a = 1 \times p_{12} + 0 \times p_{13} + a \times p_{11}$$

(Hint: Just like we found two step transition probabilities above, we can condition on the next move we make starting from state 1, and apply the law of total probability. $P(\lim_{n\to\infty} X_n = 2|X_1 = 1) = \sum_{k=1}^{3} P(\lim_{n\to\infty} X_n = 2|X_1 = 1, X_2 = k)P(X_2 = k|X_1 = 1)$)

(c) Solve the above equation, determine the values of a and b.

Ex: Recall the 2 Umbrella Problem, where I started on the morning of day 1 with one umbrella at each location. Suppose it independently rains each morning and evening with probability 0.5. Compute the probability that, on the way back home on day 2, I am left in the rain with no umbrella.

Ex: There is an insect walking (from what I gather, randomly) on my windowsill. Out of boredom, and in order to observe the movements of the insect better, I marked 4 equally spaced regions on the windowsill. At location 1 at the left hand corner, there is a spider web. If the insect gets to that location, it will be captured. Upon careful observation, I determined that the transition probabilities between locations are as in the matrix **P** given below:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

(a) Given that the insect starts at location 2, what is the probability, that after 2 steps, it is at location 4?

(b) Given that the insect starts at location 2, what is the probability, that after infinitely many steps (in practice, after a sufficiently long time), it is observed at location 4?

Ex: Random Walk with Two Thresholds: Consider a random walk on the set of integers: The probability of moving right or left are each equal to 1/3. We stay in place with probability 1/3. Given $X_0 = 0$, compute the probability of the walk reaching the value 2 before ever hitting -2.

Ex: Random Walk on a Finite Set: Now, truncate the random walk in the previous example at -2 and 2, such that when we reach these endpoints, the probability of staying in place is 2/3. Given $X_0 = 0$, compute the expected number of transitions until the walk hits 2.

1.2.3 Classification of States

Definition 3 *In a MC, a state j is said to be **accessible** from state i if there is a value for $n \geq 0$ s.t. $P_{ij}^{(n)} > 0$.*

A straightforward corollary of this definition is the following:

Theorem 1 *If j is NOT **accessible** from state i, then given that the process starts in i, the probability that the process will ever enter j is zero.*

Proof.

$$\begin{aligned} P(\text{ever enter } j|\text{start in } i) &= \lim_{n\to\infty} P(\bigcup_{k=1}^{n} X_k = j|X_0 = i) \\ &\leq \lim_{n\to\infty} \sum_{k=1}^{n} P(X_k = j|X_0 = i) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} P_{ij}^{(k)} \\ &= 0 \end{aligned}$$

Definition 4 *If state j is accessible from state i and state i is accessible from state j, then i and j are said to **communicate**, denoted $i \leftrightarrow j$.*

Show that, if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.

Ex: Consider the Markov Chain models for the 2 Umbrella Problem, the 202 Problem, the Spider-and-Insect Problem, and the Random Walk. Which states communicate with each other?

Definition 5 *A state i is called recurrent if, starting in i, the probability of ever returning to i is one. Otherwise, state i is transient.*

Properties:

- Let $A(i)$ be the set of states accessible from i. State i is recurrent only if i is accessible from any $j \in A(i)$.
- If i is recurrent, the process will reenter state i *infinitely often.*
- If state i is transient, the total number of times that i is visited, until it is left for good, is a Geometric random variable (and thus it is finite with probability 1) (Why).
- If i is recurrent, then all $j \in A(i)$ are recurrent. $\{i\} \cup A(i)$ forms a **recurrent class**.
- In a Markov Chain with a finite set of states, not all states can be transient.
- The set of states can be partitioned as $S = T \bigcup C_1 \bigcup C_2 \ldots$ where T is the set of transient states, and $C_i's$ are recurrent classes.

- If a finite-state MC is irreducible, it consists of a single recurrent class.

Ex: In the MC examples considered above, list transient states and recurrent classes.

Ex: Consider the random walk on the set of integers, with probability q of going to the right, and $1 - q$ of going to the left. Note that all states are either recurrent, or all states are transient (Why?). (For those who are interested: What is the condition for all states to be transient? There is an argument that uses the solution of the "Cliff hanger" problem).

Definition 6 *Let* $d_i = gcd\{n,\ s.t.\ P_{ii}^{(n)} > 0\}$. *If* $d_i > 1$, *state* i *is periodic with period* d_i.

Corollary 1 *All states in a recurrent class have the same period.*

Ex: Provide examples of Markov Chains with periods 2 and 3.

Argue that, if $P_{ii} > 0$ for and $i \in S$, the MC cannot be periodic.

Ex: Consider the Markov Chain given by the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

Study P^n for several values of n. What do you observe? Do you think the entries of P^n will converge as n increases?

Ex: Now consider the Markov Chain given by the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix}$$

Study P^n for several values of n. Do you think the entries of P^n will converge as n increases? What does each column converge to?

Ex: Now consider the Markov Chain belonging to the "Taking 202" problem. Do you think the entries of P^n will converge as n increases? What does each column converge to?

1.2.4 Steady-state probabilities

We observe from the three examples above, that when all recurrent classes in the MC are aperiodic, then as $n \to \infty$ the entries of the n-step transition probability matrix converge:

$$P_{ij}^{(n)} \to \pi_{ij}$$

Moreover, if the MC is has a single recurrent class (and possibly some transient states), these limiting values do not depend on the initial state:

$$\pi_{ij} = \pi_j$$

A Markov Chain which has a single recurrent class, that is also aperiodic, is called *ergodic*.

Ex: Consider two LEDs controlled by two switches in the following way: Switch 1 toggles the state of one of the LEDs, chosen equally likely at random (For example, if both LEDs are OFF (00), one of them will turn ON (01 or 10) when Switch 1 is flipped; from 10 or 01, if Switch 1 is flipped, they may go to 11 or 00.) Switch 2 turns both LEDs OFF.

Each minute, someone comes and flips one of the switches at random (each switch is chosen with equal probability.)

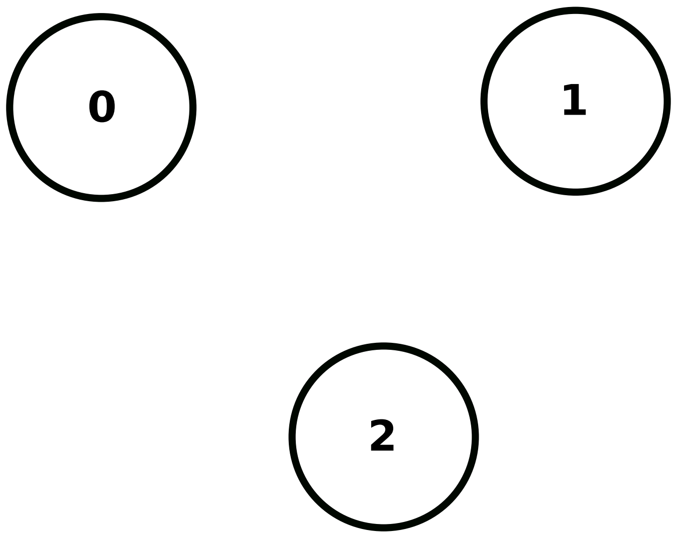

Figure 1.3: Ergodic Markov Chain for the switch flipping example: Mark the transition probabilities according to the probabilities described in the problem statement. All states are *recurrent*.

(a) Let the state of the system be the number of LEDs that are ON. Draw the diagram of a Markov Chain that models this process, and mark the transition probabilities.

(b) The P^n matrix converges to a matrix π in this problem. The entries of π depend only on the column number, not the row. That is, the probability of being in state j after a long amount of time is a constant number, that does not depend on the starting state i. In other words, the ij^{th} entry depends only on i, $\pi_{ij} = \pi_j$. (Note that $\pi_{ij} = lim_{n\to\infty} P(X_n = j|X_1 = i)$.) To find these limiting probabilities, we will apply the following logic: First, apply the Total Probability Theorem:

$$P(X_n = j) = \sum_{k=0}^{2} P(X_n = j|X_{n-1} = k)P(X_{n-1} = k)$$

If n is very large (large enough that steady state has been reached), than the probability of being in state k at time n should be the same as the probability of being in state k at time $n-1$. So, replace $P(X_{n-1} = k)$ by π_k and $P(X_n = k)$ also by π_k, for $k = 1, 2, 3$. We get a set of equations for the π_js:

$$\pi_j = \sum_{k=0}^{2} \pi_k p_{kj}$$

$j = 1, 2, 3$. In addition to these three equations (only two of which are linearly independent) we have another equation:

$$\sum_{j=0}^{2} \pi_j = 1$$

Solve this system of equations to find the steady-state probabilities.

(c) Suppose, after the switches have been flipped randomly as described above many times, I come in at a certain time. What is the probability that I see both LEDs ON? What is the long term fraction of time that both LEDs are ON?

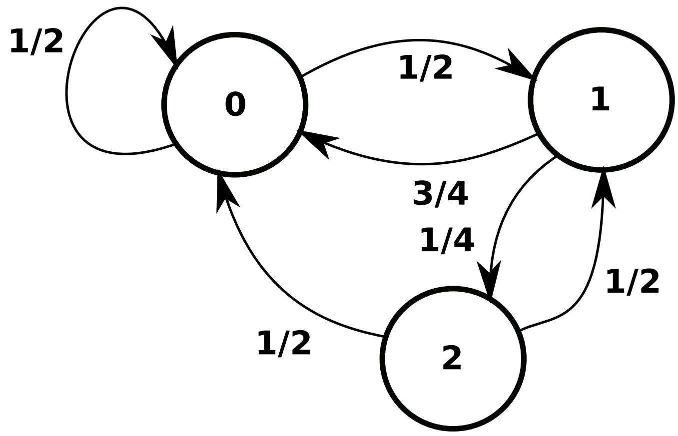

Figure 1.4: Ergodic Markov Chain for the switch flipping example with the transition probabilities marked. All states are *recurrent*.

Let us summarize these observations in a theorem that we will state without proof.

Theorem 2 *Consider a Markov Chain with $k < \infty$ states with a single recurrent class, which is aperiodic. Then, there is a set of values $0 \leq \pi_j \leq 1$, such that:*

[a]For each j,

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j, \forall i$$

The π_j are the unique solution satisfying $\sum_{m=1}^{k} \pi_m = 1$ to the set of equations:

$$\pi_j = \sum_{m=1}^{k} \pi_m P_{mj}, j = 1, 2, \ldots, k \tag{1.1}$$

We can write the RHS in the above as a matrix multiplication:

$$\boldsymbol{\pi} = \boldsymbol{\pi} P$$

where $\boldsymbol{\pi}$ is the probability vector $[\pi_1 \pi_2 \ldots \pi_k]$.

Remarks:

- The set of **global balance equations** given in (1.1) contains only $m-1$ linearly independent equations (one equation will be redundant.) Therefore, (1.1) has an infinite number of solutions. We are looking for the unique solution that is a probability vector (i.e. satisfying the normalization $\sum_{m=1}^{k} \pi_m = 1$.)

- . We have

$$\pi_j = 0, \text{ for all transient states } j \pi_j > 0, \text{ for all recurrent states } j$$

- The π_j are called **stationary probabilities** because if the process started with these probabilities:

$$\mathbf{P}(X_o = j) = \pi_j, \forall j$$

  Then, by the Chapman-Kolmogorov eqns,

$$\mathbf{P}(X_1 = j) = \sum_m \mathbf{P}(X_o = m) P_{mj} = \sum_m \pi_m P_{mj} = \pi_j, \forall j$$

  Hence at any future time the distribution of the state is the same.

Long Term Frequency of Occurrence

Theorem 2 says that in an ergodic MC, the limiting values $\{\pi_j\}$ are stationary probabilities of observing each state j. It can also be shown that in such an ergodic process, with probability 1, the value π_j is also equal to the long term fraction of time that state j is visited in a sample path.

Here is a sketch of an argument for this: Suppose, every time we visit state j, we get a unit reward. Given that we start in state i, consider the long term average rate of reward collected:

$$\lim_{n \to \infty} \frac{R_{ij}(n)}{n}$$

Every time we receive an award, the process renews itself. It can be then shown (using Law of Large Numbers arguments) that the above fraction (number of renewals per time) converges to the its expectation, the expected reward per time in almost all sample paths. (We have to leave the proof of this intuitive result outside the scope in this course.) Note that this expectation is the expectation of an indicator random variable, which takes the value 1 whenever we are in state i. In other words, it is equal to the steady-state probability of a visit to state i:

$$\lim_{n\to\infty} \frac{R_{ij}(n)}{n} = \pi_j \text{ with probability 1} \tag{1.2}$$

Ex: In the 2 Umbrella Problem, if it rains independently with probability p each morning and evening, how often does one get caught in the rain with no umbrella?

How about the long term frequency of transitions from j to k?

Balance equations:

Ex: Compute the steady-state distribution in the Rain/No Rain problem where $P_{12} = a,\ P_{21} = b$.

Note that the probabilities of Rain and No Rain are in proportion to a and b. Is this a coincidence?

Birth-death Chains:

Ex: Random Walk with Reflecting Barriers:

Ex: Geo/Geo/1 Queue:

Mean First Passage and Recurrence Times

Imagine getting a unit reward whenever we visit state i. Every time state i is visited, the process is renewed. Let us name the duration (that is, the number of transitions) between any two consecutive reward instants an inter-reward epoch. As the process is renewed after each reward, the durations of these epochs are independent. By the same argument, they each have the same distribution. In other words, they are IID. Then, the average of epoch durations satisfies the Strong Law of Large Numbers. This means that the long term average of the time between rewards is equal to its expectation, with probability 1. This expected time is the Expected Recurrence Time for state i. With this argument, we obtain:

$$lim_{n\to\infty}\frac{R_{ij}(n)}{n} = \frac{1}{M_i}, \text{ with probability 1.}$$

Combining this with (1.2), we obtain:

$$\pi_i = \frac{1}{M_i}$$

In the rest, we will explore how to compute the mean recurrence time of a recurrent state.

Ex: Consider the Rain/No Rain example, with $P_{11} = 0.8$, and $P_{22} = 0.4$. Compute the mean number of dry days between any two rainy days.

Ex: Consider the given Markov Chain containing a single recurrent class, and some transient states.

(a) Compute the mean first passage time to state 1, starting at any given state i at time $n = 0$. (Hint: Turn state 1 into an absorbing state, and write a set of equations for the expected time for absorption, starting from each state other than 1)

(b) Compute the mean recurrence time for state 1. (Hint: Use your answers to the previous problem. Why do you not need to include any of the transient states in the computation?)

Chapter 2

The Poisson Process

In this chapter, we will briefly review the Exponential distribution, and next, we will define and explore Counting Processes. We will then introduce the Poisson Process.

2.1 The Exponential Distribution

An exponential r.v. has the following PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0, & \text{o.w.} \end{cases},$$

where λ is a positive parameter.

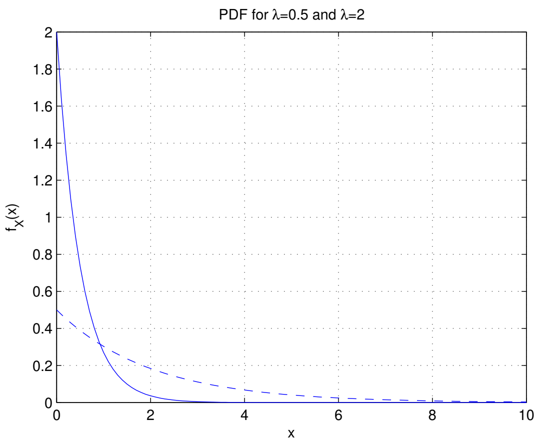

Figure 2.1: Exponential PDF

Memorylessness: Show that the exponential random variable is memoryless. That is, let X be an exponential with parameter λ. Given $X > t$, find $\mathrm{P}(X > t + x)$.)

Ex: Suppose you are waiting at the bus stop to get onto the next bus. Would you rather the bus inter-arrival times be exponential with mean 10 minutes, or exactly 10 minutes?

Racing Exponentials: Show that the minimum of k independent exponential random variables is an exponential random variable, with parameter ....

Ex: Suppose you arrive at a bank that has two tellers, and suppose that both tellers are busy when you arrive, but there is no-one else in queue. Somebody else arrives right after you, while you are waiting. If service times for teller i are exponential with rate λ_i, i = 1,2, find the expected time you will spend at the bank, under each of the following strategies:

(a) You pick a teller at random and wait until the customer being served by that teller leaves.

(b) You will queue up in front of both tellers, and join the first teller that gets freed up.

2.2 Counting Processes

Consider arrivals (of busses, customers, photons, e-mails, etc) occurring at random points in time. A Counting Process is a model for such an arrival process, that counts the number of arrival events by time t.

A counting process $\{N(t), t \geq 0\}$ is an integer valued random process, such that $N(0) = 0$, $N(\tau) \geq N(t)$ for all $\tau \geq t$.

Ex: Sketch a sample path of a counting process and notice that it looks like a staircase.

Let $X_i, i \geq 1$ be the times between arrival events. We will refer to these as "inter-arrival times". We can alternatively define the counting process $N(t)$ by specifying these inter-arrival epochs.

We can also express the process in terms of the times of the arrivals, S_1, S_2, $\ldots$ such that $S_i \leq S_{i+1}$. Let $S_0 = 0$ for completeness.

Ex: Mark the S_i's in the sketch you made above. Also mark the inter-arrival times $X_{i+1} = S_{i+1} - S_i$. If these are continuous random variables, notice that each jump in the staircase has height 1, with probability 1.

Definition 7 *Let $\{S_n, n \geq 1\}$ be the arrival times in the counting process. Note that $S_n = \sum_{i=1}^{n} X_i$.*

Definition 8 *A counting process is a sequence of non-negative random variables $\{X_i, i \geq 1\}$ where $X_i \geq i \; \forall \; i$.*

Hence, the process is alternatively characterized by $\{N(t), t \geq 0\}$, or $\{S_n, n \geq 1\}$, or

$\{X_i, i \geq 1\}$. We make the following observations which will help in solving problems using counting process models.

Properties

- $\{S_n \leq t\} \Rightarrow \{N(t) \geq n\}$, and $\{N(t) \geq n\} \Rightarrow \{S_n \leq t\}$
  - Show that this is true.

- $\{S_n > t\} \Rightarrow \{N(t) < n\}$, and $\{N(t) < n\} \Rightarrow \{S_n > t\}$
  - This is a corollary of the above, because

$$\{S_n > t\}^c \equiv \{S_n \leq t\} \Leftrightarrow \{N(t) \geq n\} \equiv \{N(t) < n\}^c$$

- $\{S_n \geq t\} \Rightarrow \{N(t) \leq n\}$, but $\{N(t) \leq n\}$ does not imply $\{S_n \geq t\}$
  - Counter example:

- $\{S_n < t\} \Rightarrow \{N(t) \geq n\}$, but $\{N(t) \geq n\}$ does not imply $\{S_n < t\}$
  - Note that $\{N(t) \geq n\} = \{S_n < t\} \bigcup \{S_n = t\}$

2.3 The Poisson Process

Definition 1 of the Poisson Process: *A counting process is called a Poisson Process if the times between arrivals are IID, Exponential random variables.*

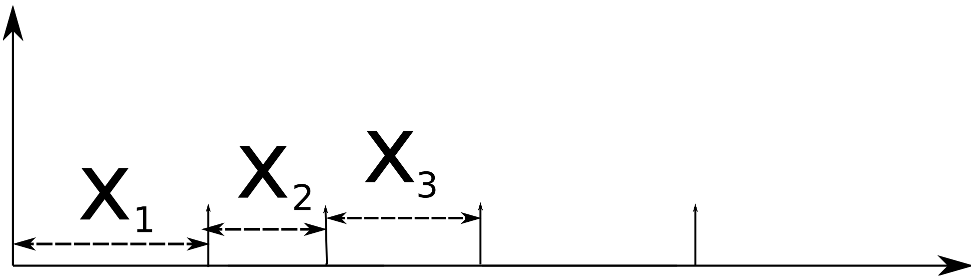

Figure 2.2: Inter-arrival times X_1, X_2, X_3 in a Poisson process

In other words, in a Poisson process of rate λ, the inter-arrival times $\{X_i\}$, $i \geq 1$ are independent and Exponential with rate λ. Note that the mean time between two arrivals is $\frac{1}{\lambda}$.

Ex: I am waiting for the bus, and bus arrivals are known to be a Poisson process at rate 1 bus per 10 minutes. Starting at time $t = 0$, what is the expected time until the arrival of the third bus?

The memorylessness property of the exponential distribution carries on to the Poisson process, which means it starts fresh at any point in time. We make this precise in the following exercise.

Ex: Distribution of residual time: Following an arbitrary amount of time $t > 0$ after an arrival in the Poisson process, let R be the duration until the next arrival. This is called the "residual time". Show that R has the same distribution as a regular inter-arrival time. This is referred to as the "fresh-start" property of the Poisson process.

Ex: Given that I arrive at the bus stop at $t = 19$ and learn that I have missed the second bus by two minutes, how much do I expect to wait?

As a consequence of the fresh-start property, one can easily see that the numbers of arrivals in any two disjoint intervals are independent. We say that the Poisson process has the "independent increments" property.

Definition 9 *A counting process is said to have "independent increments" if, for any* $t_1 < t_2 \leq t_3 < t_4$, $N(t_2) - N(t_1)$ *is independent of* $N(t_4) - N(t_3)$.

Another useful property of the Poisson process (that follows from the fresh-start property) is the following:

Definition 10 *A counting process is said to have "stationary increments" if the number of arrivals in a time interval depends only on the length of the interval (and not on where the interval is). That is, for any* $\tau > 0$, *the distribution of* $N(t + \tau) - N(t)$ *does not depend on the value of* t, *and is the same as the distribution of* $N(\tau) - N(0) = N(\tau)$.

Ex: Compute $P(N(t) < 1)$. Compute $P(N(t + \delta) - N(t) < 1)$. Use this to compute the probability that there is at least one arrival in an interval of size δ. How does this probability scale with δ, as $\delta \to 0$?

Definition 11 *A function $f(\cdot)$ is said to be $o(\delta)$ if it decays faster than linearly as its argument goes to zero, that is,*

$$\lim_{\delta \to 0} \frac{f(\delta)}{\delta} = 0$$

Ex: Which of the following functions are $o(\delta)$?

(a) $f(x) = x$

(b) $f(x) = x^2$

(c) $f(x) = x^3 + cx$, where c is a real valued constant

(d) $f(x) = ah(x) + bg(x)$, where $h(\cdot)$ and $g(\cdot)$ are both $o(\delta)$, a and b are constants.

We can use the observations above to make an alternative definition for the Poisson process. We will show later that this definition is equivalent to Definition 1 of the Poisson process.

Definition 2 of the Poisson Process: *A counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with rate $\lambda > 0$ if $\{N(0) = 0\}$, if the process has stationary and independent increments, and also satisfies the following:*

(i) $P(N(\delta) = 1) = \lambda\delta + o(\delta)$, and

(ii) $P(N(\delta) \geq 2) = o(\delta)$.

We can use the above to discover that $N(t)$ is Poisson distributed with rate λt. We can do this through writing the distribution of the number of arrivals in a large interval, and using the Poisson approximation for the Binomial, or through computing the MGF as we will do below. First, let's recall the Poisson distribution:

Definition 12 *A random variable Y is said to have the Poisson Distribution with mean a, if*

$$\mathbf{P}(M = k) = \frac{a^k e^{-a}}{k!}, \; k = 0, 1, \ldots$$

Show that the Moment Generating function of a Poisson random variable Y, with mean a, is given by

$$M_Y(s) = e^{a(e^{-s}-1)}$$

Ex: Compute the MGF of the number of arrivals, $N(t)$, by time t, in the Poisson process whose rate is λ.

We have found that $E[\exp(-sN(t)] = e^{\lambda t(e^{-s}-1)}$. This is the MGF of a Poisson random variable with mean λt. Recalling that a MGF fully determines the corresponding distribution, we conclude that in a Poisson process with rate λ, the number of arrivals by time t is Poisson with mean λt. This explains why λ is the "rate" of the process. Now, by the stationary increments property, we conclude that the number of arrivals in any time interval of size t is Poisson with mean λt. This brings us to the following alternative definition for the Poisson process.

Definition 3 of the Poisson Process: *A counting process* $\{N(t), t \geq 0\}$ *is said to a Poisson process with rate* $\lambda > 0$ *if* $\{N(0) = 0\}$*, if the process has stationary and independent increments, and, for any* $s, t \geq 0$

$$\mathbf{P}(N(t+s) - N(s) = k) = e^{\lambda t}\frac{(\lambda t)^k}{k!},\ k = 0, 1, 2, \ldots$$

Ex: Show that Definition 3 implies Definition 2. As Def. 3 was already derived from Def.2, we conclude that the two definitions are equivalent.

Ex: Show that Definition 3 implies Definition 1: Let X_1 be the time until the first arrival in the process in Definition 3. Compute the CDF of X_1. How about the CDF of $X_i, i > 1$?

Ex: I get email according to a Poisson process at rate $\lambda = 1.5$ arrivals per minute. If I check my email every hour, what is the expected number of new messages I find in my inbox when I check my email? What is the probability that I find no messages? One message? Repeat for an e-mail checking period of two hours. (Note that we find very small values, because we are looking for the probability of getting exactly 0 or 1 messages, in a relatively long time period. Repeat the computations for a time period of 1 and 2 minutes.)

Waiting Times: Let $S_n = \sum_{i=1}^{n} X_i$, the time of the n^{th} arrival in the Poisson process with rate λ, $n \geq 1$. Show that S_n obeys the Erlang distribution of order n, with parameter λ:

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{(n-1)}}{(n-1)!}$$

(Hint: Note that $\{S_n \leq t\} \equiv \{N(t) \geq n\}$.)

Ex: Suppose, starting at time $t = 8:25$, students start entering Building A according to a Poisson process at a rate of 1 student per minute.

(a) Find the distribution of the time elapsed up to and including the arrival of the 10^{th} student.

(b) Let T be the elapsed time between the second and fourth arrivals. What is the probability that $T > 3$ minutes?

(c) Write an expression to compute the third moment of T.

Ex: An alternative way to obtain the density of the n^{th} inter-arrival time is the following: Write the probability $\mathbf{P}(t < S_n < t + \delta)$, divide by δ and take the limit $\delta \to 0$.

The time-reversed process is also Poisson: We can show that the reverse residual time distribution is the same as the inter-arrival time distribution.

Ex: In the bus problem, what is the expected number of people on the bus that I get on? (Hint: Consider the people that arrived in the two minutes before I arrived, as well as the people that arrive while I am waiting.)

The random incidence "paradox": Note from the above example that when I arrive at random, the inter-arrival time I sample is expected to have twice the duration of the average inter-arrival duration. Consider the following: if bus arrivals are a Poisson process, in order to understand whether buses are too crowded or not, should planners interview people, or bus drivers?

Ex: 3 buses departed from METU to Kizilay this morning. The first bus had only 10 passengers on board. The second bus had 70, and the third had 40 passengers. EGO, the company who runs the buses, wanted to find out whether the buses are too crowded or not, on average. They decided to sample several passengers at random, and ask them their estimate of how many people there were on the bus. What is the expected result? Compare this with the average number of customers on the three buses.

The example above is to illustrate the random incidence "paradox". In the Poisson process, when a point in time is selected at random, larger intervals are more likely to get selected.

2.3.1 Splitting and Merging Poisson Processes

Ex: Show that, when we send each arrival of a Poisson process at rate λ to a process A with probability p, and process B with probability $1 - p$, the resulting processes A and B are Poisson with rates $p\lambda$ and $(1 - p)\lambda$. (Hint: (i) Express the transform of the interarrival time as a geometric sum of Exponentials, and use Definition 1 of the Poisson process, or, (ii) use the Baby Bernoulli definition of the Poisson process.)

Note: It can be shown that the processes resulting from randomly splitting each arrival in a Poisson process as above are independent of each other.

Ex: Show that when we merge two INDEPENDENT Poisson processes at rates λ_a and λ_b, we get a Poisson process at rate $\lambda_a + \lambda_b$. (Hint: Consider racing Exponentials.)

Extending the above by induction, we find that when we merge independent Poisson processes of rates $\lambda_i, i = 1, 2, \ldots, k$, we obtain a Poisson process of rate $\sum_{i=1}^{k} \lambda_i$.

2.3.2 Examples related to the Poisson Process

Ex: Starting at $t = 0$, buses depart from the METU campus according to a Poisson process at rate 0.2 buses per minute. Each bus is independently assigned to go to Kizilay, or to Ulus, with probabilities $5/6$ and $1/6$, respectively. Starting at $t = 0$, people arrive at the bus stop according to a Poisson process of rate 1.8 people per minute, independently of the bus arrival process. People get on the first bus that comes (they do not distinguish between the destinations.)

(a) I arrive at the bus stop and a bus just left as I was arriving. How long do I expect to wait?

(b) I arrive at random. How much do I expect to wait? What is the probability that the first bus I see is to Ulus?

(c) Compute the probability that exactly 5 buses depart in the first hour. Compute the probability that at least 5 buses depart in a one-hour interval.

(d) What is the probability that among the first 5 bus departures, 3 go to Ulus? Also compute the probability that the fifth bus departure is the 3rd bus departure to Ulus.

(e) Find the PMF and the expectation of N, the number of people getting on a bus.

(f) I show up at the bus stop at a random time. What is the expected number of people on my bus?

(g) I arrive at the bus stop and learn that within the half hour previous to my arrival, exactly 6 buses departed. What is the probability that all 6 actually departed in the 10 minutes prior to my arrival? (The interested student can show a generalization of this: Given the number of arrivals occurring in a given interval, the arrival times are uniformly distributed in the interval. We will use this result in the analysis of "shot noise" in the next lecture.)

Ex: "The Random Telegraph Signal" Consider a process $X(t), t \geq 0$, that takes values ± 1 and changes polarity at time instants that arrive according to a Poisson process. Show that, if $X(0)$ takes the values 1 or -1 equiprobably, then $X(t) = 1$ with probability 1/2 at an t. Compute the covariance of $X(t_1)$ and $X(t_2)$ for any t_1 and t_2.

Ex: "Shot Noise" Consider an impulse train $Z(t) = \sum_{i=1}^{\infty} \delta(t - S_i)$, where $\{S_i, i \geq 1\}$ are the arrival times in a Poisson process of rate λ. Let's pass $Z(t)$ through an LTI system with impulse response $h(t)$. The resulting filtered process, $X(t) = \sum_{i=1}^{\infty} h(t - S_i)$ is known as "shot noise". For example, suppose $h(t)$ is the current pulse resulting from a photon hitting a detector. Then, $X(t)$ would be value of the current at time t. Compute $E[X(t)]$.

This completes our introduction to discrete stochastic processes. The interested student can follow EE531, available on METU OpenCourseWare:
https://ocw.metu.edu.tr/course/view.php?id=323