## Lecture 3: Discrete-Time Markov Chain - Part I

Instructor: Yen-Chi Chen

These notes are partially based on those of Mathias Drton.

### 3.1 Introduction

Before introducing Markov chain, we first talk about stochastic processes. A stochastic process is a family of RVs $X_{n}$ that is indexed by $n$, where $n \in \mathcal{T}$. Note that sometimes people write $X_{t}$ with $t \in \mathcal{T}$. The set $\mathcal{T}$ is called the index set. There are two common types of stochastic processes:

- $\mathcal{T}$ is discrete. For instance, $\mathcal{T}=\{0,1,2,3, \cdots\}$. Then $X_{n}$ is called a discrete-time stochastic process.
- $\mathcal{T}$ is continuous. For instance $\mathcal{T}=[0, \infty)$. Then $X_{n}$ is called a continuous-time stochastic process.

Each random variable $X_{n}$ can have a discrete, continuous, or mixed distribution. For example, in a queue $X_{n}$ could represent the time that the $n$-th customer waits after arrival before receiving service, with a distribution that has an atom at zero but is otherwise continuous. Usually, each $X_{n}$ takes its values in the same set, which is called the state space and denoted $S$. Therefore, $X_{n} \in S$.

We will focus on discrete time stochastic processes with a discrete state space in this course.
Let $\left\{X_{n}: n=0,1,2, \ldots\right\}$ be a discrete time stochastic process and the state space $S$ is discrete. The probability model of this process is determined by the joint distribution function:

$$
P\left(X_{0}=x_{0}, X_{1}=x_{1}, \cdots, X_{n}=x_{n}\right)
$$

for all $n=0,1, \ldots$ and $x_{0}, x_{1}, \ldots \in S$.
In general, this joint distribution function can be arbitrary so it is very complex. We need some addition modeling on the joint distribution function to make it simpler enough that we can analyze. One particular example is the IID assumptions $-X_{0}, X_{1}, \cdots$ are IID. In this case, the join distribution function can be factorized into products of identical functions.

However, IID assumptions are often too strong for some scenarios. For instance, when modeling the genetic drift, the IID assumption on each generation is not a good model. Here are two examples where the IID assumptions do not work.

Example 1: Excess number of heads over tails in tossing a coin. Assume we toss a coin $n$ times and let $X_{n}$ denote the excess number of heads over tails. Clearly, $X_{n} \in\{\cdots,-2,-1,0,1,2, \cdots\}$. Assume that each coin is tossed independently, then the process $\left\{X_{n}\right\}$ has a one-step memory such that

$$
P\left(x_{0}, \cdots, x_{n}\right)=P\left(x_{n} \mid x_{n-1}\right) \times P\left(x_{n-1} \mid x_{n-2}\right) \times P\left(x_{1} \mid x_{0}\right) \times P\left(x_{0}\right)
$$

where $P\left(x_{0}, \cdots, x_{n}\right)=P\left(X_{0}=x_{0}, X_{1}=x_{1}, \cdots, X_{n}=x_{n}\right)$. Clearly, $P\left(x_{0}, x_{1}\right) \neq P\left(x_{0}\right) P\left(x_{1}\right)$ so the process is not independent although $P\left(x_{0}, x_{2} \mid x_{1}\right)=P\left(x_{0} \mid x_{1}\right) P\left(x_{2} \mid x_{1}\right)$.

Example 2: Precipitation level. Let the precipitation level of day $n$ be $X_{n}$ where $X_{n}=0$ (dry) or 1 (wet). Assume that the precipitation level of a day given all the past history only depends on the precipitation level of the last two days. Namely,

$$
P\left(x_{n} \mid x_{n-1}, x_{n-2}, \cdots, x_{0}\right)=P\left(x_{n} \mid x_{n-1}, x_{n-2}\right)
$$

Then the joint distribution function is

$$
P\left(x_{0}, \cdots, x_{n}\right)=P\left(x_{n} \mid x_{n-1}, x_{n-2}\right) \times P\left(x_{n-1} \mid x_{n-2}, x_{n-3}\right) \times P\left(x_{2} \mid x_{1}, x_{0}\right) \times P\left(x_{1} \mid x_{0}\right) \times P\left(x_{0}\right)
$$

Again, it is clear that the RVs $X_{0}, X_{1}, \cdots, X_{n}$ are not IID. Although this model shows that the process has a two-step memory, we can reformulate it to a one-step memory process by defining a sequence of random vectors $Y_{0}, Y_{1}, \cdots, Y_{n}, \cdots$ such that $Y_{n}=\left(X_{n}, X_{n+1}\right)$. Then

$$
P\left(y_{n} \mid y_{n-1}, y_{n-2}, \cdots, y_{0}\right)=P\left(y_{n} \mid y_{n-1}\right)
$$

so the process $\left\{Y_{n}\right\}$ has a one-step memory.
The above two examples motivates us to study the process with a one-step memory. And such a stochastic process is known as the Markov chain .

### 3.2 Markov Chain

A discrete time stochastic process $\left\{X_{n}: n=0,1,2, \cdots\right\}$ is called a Markov chain if for every $x_{0}, x_{1}, \cdots, x_{n-2}, i, j \in$ $S$ and $n \geq 0$,

$$
P\left(X_{n}=i \mid X_{n-1}=j, \cdots, X_{0}=x_{0}\right)=P\left(X_{n}=i \mid X_{n-1}=j\right)
$$

whenever both sides are well-defined. The Markov chain has a one-step memory.
If the distribution function $P\left(X_{n}=x_{n} \mid X_{n-1}=x_{n-1}\right)=p_{i j}$ is independent of $n$, we called $\left\{X_{n}\right\}$ a homogeneous Markov chain. Otherwise we called it an inhomogeneous Markov chain. For a homogeneous Markov chain,

$$
\sum_{j \in S} p_{i j}=1, \quad p_{i j} \geq 0
$$

for every $i, j$. Note that sometimes people write $p_{i j}=p_{i \rightarrow j}$, where $p_{i \rightarrow j}$ stands for that the probability moving from state $i$ to state $j$.

Because $S$ is a discrete set, we often label it as $S=\{1,2,3, \cdots, s\}$ and the elements $\left\{p_{i j}: i, j=1, \cdots, s\right\}$ forms an $s \times s$ matrix $\mathbf{P}=\left\{p_{i j}\right\}$. $\mathbf{P}$ is called the transition (probability) matrix (t.p.m). The property of homogeneous Markov chain implies that

$$
\begin{equation*}
\mathbf{P} \geq 0, \quad \mathbf{P} \mathbf{1}_{s}=\mathbf{1}_{s} \tag{3.1}
\end{equation*}
$$

where $\mathbf{1}_{s}=(1,1,1,1, \cdots, 1)^{T}$ is the vector of 1 's. Note that any matrix satisfying equation (3.1) is called a stochastic matrix.

## Example 3: SIS (Susceptible-Infected-Susceptible) model.

Suppose we observe an individual over a sequence of days $n=1,2, \ldots$ and classify this individual each day as

$$
X_{n}= \begin{cases}I & \text { if infected } \\ S & \text { if susceptible }\end{cases}
$$

We would like to construct a stochastic model for the sequence $\left\{X_{n}: n=1,2, \ldots\right\}$. One possibility is to assume that the $X_{n}$ 's are independent with $P\left(X_{n}=I\right)=1-P\left(X_{n}=S\right)=\alpha$. However, this model is not very realistic since we know from experience that the individual is more likely to stay infected if he or she is already infected.

Since Markov chains are the simplest models that allow us to relax independence, we proceed by defining a transition probability matrix:

$$
\left.\mathbf{P}=\begin{array}{c}
\nearrow \\
I \\
S
\end{array} \quad \begin{array}{cc}
I & S \\
{[1-\alpha} & \alpha \\
\beta & 1-\beta
\end{array}\right]
$$

It can be helpful to visualize the transitions that are possible (have positive probability) by a transition diagram:


## Example 4: Example: Ehrenfest Model of Diffusion.

We start with $N$ particles in a closed box, divided into two compartments that are in contact with each other so that particles may move between compartments. At each time epoch, one particle is chosen uniformly at random and moved from its current compartment to the other compartment. Let $X_{n}$ be the number of particles in compartment 1 (say) at step $n$. This stochastic process is Markov by construction.


Transition probabilities of the Markov chain are:

$$
p_{i j}= \begin{cases}\frac{i}{N} . & \text { for } j=i-1 \\ 1-\frac{i}{N}, & \text { for } j=i+1 \\ 0, & \text { otherwise }\end{cases}
$$

The probability of transfer depends on the number of particles in each compartment. For $N=2$ we have states $0,1,2$ and t.p.m.

$$
\mathbf{P}=\left[\begin{array}{ccc}
0 & 1 & 0 \\
1 / 2 & 0 & 1 / 2 \\
0 & 1 & 0
\end{array}\right]
$$

and the transition diagrm


## Example 5: Snoqualmie Falls Precipitation.

There is a data on the precipitation (in inches), recorded by the UW Weather Service, at Sonqualmie Falls in the years 1948-1983. We examine the data for January only and consider dry $(=0)$ and wet $(=1)$ only. If we condition on the state on January 1st we obtain the frequencies of the four different transitions as:

|  | 0 | 1 | Total |
| :---: | :---: | :---: | :---: |
| 0 | 186 | 123 | 309 |
| 1 | 128 | 643 | 771 |
| Total | $(314)$ | $(766)$ | 1080 |

For example, there were 123 occasions on which a wet day followed a dry day.
From the table of frequencies we can compute the relative frequencies of transitions:

$$
\widehat{\mathbf{P}}=\left[\begin{array}{ll}
0.602 & 0.398 \\
0.166 & 0.834
\end{array}\right]
$$

This is an estimate (hence the hat!) of the t.p.m.

### 3.3 Property of Markov chain

Suppose we observe a finite realization of the discrete Markov chain and want to compute the probability of this random event:

$$
\begin{aligned}
& P\left(X_{n}=i_{n}, X_{n-1}=i_{n-1}, \cdots, X_{1}=i_{1}, X_{0}=i_{0}\right) \\
& =P\left(X_{n}=i_{n} \mid X_{n-1}=i_{n-1}, \cdots, X_{0}=i_{0}\right) P\left(X_{n-1}=i_{n-1}, \cdots, X_{0}=i_{0}\right) \\
& =p_{i_{n-1}, i_{n}} \times P\left(X_{n-1}=i_{n-1} \mid X_{n-2}=i_{n-2}, \ldots, X_{0}=i_{0}\right) \times P\left(X_{n-2}=i_{n-2}, \ldots, X_{0}=i_{0}\right) \\
& =\cdots \\
& =p_{0 i_{0}} p_{i_{0}, i_{1}} p_{i_{1}, i_{2}} \cdots p_{i_{n-2}, i_{n-1}} p_{i_{n-1}, i_{n}}
\end{aligned}
$$

where $p_{0}=\left(p_{01}, p_{02}, \ldots\right)^{T}$ is the distribution of $X_{0}$, called the initial distribution of $\left\{X_{n}\right\}$. Thus, every Markov chain is fully specified by its transition probability matrix $\mathbf{P}$ and initial distribution $p_{0}$.
The Markov chain has a powerful property called the Markov property - the distribution of $X_{m+n}$ given a set of previous states depends only on the latest available state. Assume that we observe a Markov chain from $n=0,1, \cdots, n$ and we are analyzing the distribution of $X_{m+n}$. Then

$$
\begin{equation*}
P\left(X_{m+n}=j \mid X_{n}=i, X_{n-1}=i_{n-1}, \ldots, X_{0}=i_{0}\right)=P\left(X_{m+n}=j \mid X_{n}=i\right) \tag{3.2}
\end{equation*}
$$

To give an intuition about how we obtain the Markov property, consider a simple case where $n=1$ and $m=2$.

$$
\begin{aligned}
P\left(X_{3}=i_{3} \mid X_{1}=i_{1}, X_{0}=i_{0}\right) & =\sum_{i_{2}} P\left(X_{3}=i_{3}, X_{2}=i_{2} \mid X_{1}=i_{1}, X_{0}=i_{0}\right) \\
& =\sum_{i_{2}} P\left(X_{3}=i_{3} \mid X_{2}=i_{2}, X_{1}=i_{1}, X_{0}=i_{0}\right) P\left(X_{2}=i_{2} \mid X_{1}=i_{1}, X_{0}=i_{0}\right) \\
& \stackrel{!}{=} \sum_{i_{2}} P\left(X_{3}=i_{3} \mid X_{2}=i_{2}, X_{1}=i_{1}\right) P\left(X_{2}=i_{2} \mid X_{1}=i_{1}\right) \\
& =\sum_{i_{2}} P\left(X_{3}=i_{3}, X_{2}=i_{2} \mid X_{1}=i_{1}\right) \\
& =P\left(X_{3}=i_{3} \mid X_{1}=i_{1}\right)
\end{aligned}
$$

To argue the equality ' $\stackrel{!}{=}$, observe that

$$
P\left(X_{3}=i_{3} \mid X_{2}=i_{2}, X_{1}=i_{1}, X_{0}=i_{0}\right)=P\left(X_{3}=i_{3} \mid X_{2}=i_{2}\right)
$$

But we also have that

$$
\begin{aligned}
P\left(X_{3}=i_{3} \mid X_{2}=i_{1}, X_{1}=i_{1}\right) & =\sum_{i_{0}} P\left(X_{3}=i_{3}, X_{0}=i_{0} \mid X_{2}=i_{2}, X_{1}=i_{1}\right) \\
& =\sum_{i_{0}} P\left(X_{3}=i_{3} \mid X_{2}=i_{2}, X_{1}=i_{1}, X_{0}=i_{0}\right) P\left(X_{0}=i_{0} \mid X_{2}=i_{2}, X_{1}=i_{1}\right) \\
& =P\left(X_{3}=i_{3} \mid X_{2}=i_{2}\right) \sum_{i_{0}} P\left(X_{0}=i_{0} \mid X_{2}=i_{2}, X_{1}=i_{1}\right) \\
& =P\left(X_{3}=i_{3} \mid X_{2}=i_{2}\right)
\end{aligned}
$$

We can represent the Markov chain using a simple graphical model:


The claim of the Markov property is now obvious from the theorem on conditional independence and graphical factorization. Indeed, the latest available state serves as a separating set.

Using the graph representation, we obtain an interesting property about a Markov chain: the past and the future are independent given the present.

To see this, again we consider a simple case where $n=2$ and we have $X_{0}, X_{1}, X_{2}$. Here $X_{0}$ denotes the past, $X_{1}$ denotes the present, and $X_{2}$ denotes the future. Then

$$
\begin{aligned}
P\left(X_{0}=i_{0}, X_{2}=i_{2} \mid X_{1}=i_{1}\right) & =\frac{P\left(X_{0}=i_{0}, X_{1}=i_{1}, X_{2}=i_{2}\right)}{P\left(X_{1}=i_{1}\right)} \\
& =\frac{P\left(X_{2}=i_{2} \mid X_{1}=i_{1}, X_{0}=i_{0}\right) P\left(X_{1}=i_{1}, X_{0}=i_{0}\right)}{P\left(X_{1}=i_{1}\right)} \\
& =P\left(X_{2}=i_{2} \mid X_{1}=i_{1}\right) \frac{P\left(X_{1}=i_{1}, X_{0}=i_{0}\right)}{P\left(X_{1}=i_{1}\right)} \\
& =P\left(X_{2}=i_{2} \mid X_{1}=i_{1}\right) P\left(X_{0}=i_{0} \mid X_{1}=i_{1}\right)
\end{aligned}
$$

for any $i_{0}, i_{1}, i_{2} \in S$. Namely, $X_{0}$ and $X_{2}$ are conditional independent given $X_{1}$.

## 3.4 n-step Transition Probability and Chapman-Kolmogorov Equation

For a Markov chain, we define the $n$-step transition probability as

$$
p_{i j}^{(n)}=P\left(X_{n}=j \mid X_{0}=i\right)
$$

The $n$-step transition probability is time invariant.
Lemma 3.1 Let $\left\{X_{n}\right\}$ be a homogeneous Markov chain and let $p_{i j}^{(n)}$ be the $n$-step transition probability. Then for any $k=0,1,2, \cdots$,

$$
P\left(X_{n+k}=j \mid X_{k}=i\right)=p_{i j}^{(n)} .
$$

Proof:

$$
\begin{aligned}
P\left(X_{n+k}=j \mid X_{k}=i\right) & =\sum_{i_{k+1}, i_{k+2}, \cdots, i_{n+k-1}} P\left(X_{n+k}=j \mid X_{n+k-1}=i_{n+k-1}\right) \times \cdots \times P\left(X_{k+1}=i_{k+1} \mid X_{k}=i\right) \\
& =\sum_{i_{k+1}, i_{k+2}, \cdots, i_{n+k-1}} P\left(X_{n}=j \mid X_{n-1}=i_{n+k-1}\right) \times \cdots \times P\left(X_{1}=i_{k+1} \mid X_{0}=i\right) \\
& =P\left(X_{n}=j \mid X_{0}=i\right)=p_{i j}^{(n)} .
\end{aligned}
$$

The $n$-step transition probabilities are related to each other via the famous Chapman-Kolmogorov Equation.

Lemma 3.2 Let $\left\{X_{n}\right\}$ be a homogeneous Markov chain and let $p_{i j}^{(n)}$ be the $n$-step transition probability. Then for any $n, m=0,1,2, \cdots$

$$
\begin{equation*}
p_{i j}^{(n+m)}=\sum_{k \in S} p_{i k}^{(n)} p_{k j}^{(m)} . \tag{3.3}
\end{equation*}
$$

## Proof:

$$
\begin{aligned}
p_{i j}^{(n+m)} & =P\left(X_{n+m}=j \mid X_{0}=i\right) \\
& =\sum_{k \in S} P\left(X_{n+m}=j, X_{m}=k \mid X_{0}=i\right) \\
& =\sum_{k \in S} P\left(X_{n+m}=j \mid X_{m}=k, X_{0}=i\right) P\left(X_{m}=k \mid X_{0}=i\right) \\
& =\sum_{k \in S} P\left(X_{n+m}=j \mid X_{m}=k\right) P\left(X_{m}=k \mid X_{0}=i\right) \quad \text { (Markov property) } \\
& =\sum_{k \in S} P\left(X_{n}=j \mid X_{0}=k\right) P\left(X_{m}=k \mid X_{0}=i\right) \quad \text { (time-invariant property) } \\
& =\sum_{k \in S} p_{i k}^{(n)} p_{k j}^{(m)}
\end{aligned}
$$

The Chapman-Kolmogorov Equation (equation (3.3)) also implies

$$
\begin{aligned}
\text { Forward equation : } p_{i j}^{(n+1)} & =\sum_{k} p_{i k}^{(n)} p_{k j}, \text { for } n=1,2, \ldots \text { and } \\
\text { Backward equation : } p_{i j}^{(n+1)} & =\sum_{k} p_{i k} p_{k j}^{(n)}, \text { for } n=1,2, \ldots
\end{aligned}
$$

The forward equation singles out the final step and has the initial state $i$ fixed. The equation is most useful when interest centers on the $p_{i j}^{(n)}$,s for a particular $i$ but all values of $j$. Conversely, the backward equation singles out the change from the initial state $i$ and has the final state $j$ fixed. This equation is useful when interest is in the $p_{i j}^{(n)}$ 's for a particular $j$ but all values of $i$. The backward equation can be interesting, in particular, when there is an absorbing state $j$ from which there is no escape $\left(p_{j j}=1\right)$.

If we collect the $n$-step transition probabilities into the matrix $\mathbf{P}^{(n)}=\left\{p_{i j}^{(n)}\right\}$, then Kolmogorov's forward and backward equations can be rewritten in matrix form as

$$
\mathbf{P}^{(n+1)}=\mathbf{P}^{(n)} \mathbf{P}=\mathbf{P} \mathbf{P}^{(n)},
$$

where $\mathbf{P}^{(1)}=\mathbf{P}$. Therefore, $\mathbf{P}^{(n)}=\mathbf{P}^{n}$.
This matrix form also implies a cool property about the marginal distribution of $X_{n}$. Assume that $X_{0}$ has a distribution $p_{0}=\left(p_{01}, p_{02}, \cdots, p_{0 s}\right)^{T}$. Let $p_{n}=\left(p_{n 1}, \cdots, p_{n s}\right)^{T}$ be the marginal distribution of $X_{n}$, i.e., $p_{n j}=P\left(X_{n}=j\right)$. Then

$$
\begin{aligned}
p_{n j}=P\left(X_{n}=j\right) & =\sum_{i} P\left(X_{n}=j, X_{0}=i\right) \\
& =\sum_{i} P\left(X_{n}=j \mid X_{0}=i\right) P\left(X_{0}=i\right) \\
& =\sum_{i} p_{0 i} p_{i j}^{(n)}
\end{aligned}
$$

Using the matrix form, we obtain

$$
p_{n}^{T}=p_{0}^{T} \mathbf{P}^{n}
$$

## Example 3: SIS model (revisited).

Recalled that SIS model has a transition probability

$$
\left.\mathbf{P}=\begin{array}{c}
\nearrow \\
0 \\
1
\end{array} \begin{array}{cc}
0 & 1 \\
{\left[\begin{array}{c}
1-\alpha \\
\beta
\end{array}\right.} & 1-\beta
\end{array}\right]
$$

Note that we use $\{0,1\}$ to denote the state $I$ and $S$ in the SIS model.
Assume that the initial distribution $p_{0}=(1-\alpha, \alpha)$, i.e., $P\left(X_{0}=0\right)=1-\alpha$. Moreover, assume that $\beta=1-\alpha$ so the distribution of $X_{1}$ will be

$$
\begin{aligned}
p_{1}^{T} & =p_{0}^{T} \mathbf{P}=(1-\alpha, \alpha)\left[\begin{array}{ll}
1-\alpha & \alpha \\
1-\alpha & \alpha
\end{array}\right] \\
& =\left[(1-\alpha)^{2}+\alpha(1-\alpha), \alpha(1-\alpha)+\alpha^{2}\right]=(1-\alpha, \alpha)=p_{0}^{T}
\end{aligned}
$$

What will be the distribution of $X_{n}$ ? Using the matrix form, we know that

$$
p_{n}^{T}=p_{0}^{T} \mathbf{P}^{n}=p_{1}^{T} \mathbf{P}^{n-1}=p_{0}^{T} \mathbf{P}^{n-1}=\cdots=p_{0}^{T}
$$

Therefore, $P\left(X_{n}=0\right)=1-\alpha$ and $P\left(X_{n}=1\right)=\alpha$ for all $n=1,2,3, \cdots$.
A more interesting fact is the joint distribution of $X_{0}, X_{1}, \cdots, X_{n}$ :

$$
\begin{aligned}
P\left(X_{0}=i_{0}, X_{1}=i_{1}, \cdots X_{n}=i_{n}\right) & =p_{0 i_{0}} p_{i_{0} i_{1}} p_{i_{1} i_{2}} \cdots p_{i_{n-1} i_{n}} \\
& =\alpha^{i_{0}}(1-\alpha)^{1-i_{0}} \alpha^{i_{1}}(1-\alpha)^{1-i_{\alpha}^{i_{0}}(1-\alpha)^{1-i_{0}}} \cdots \alpha^{i_{n}}(1-\alpha)^{1-i_{n}}
\end{aligned}
$$

which is the joint PMF of IID random Bernoulli variables with parameter $\alpha$. Therefore, under this special case, the Markov chain reduces to IID Bernoulli RVs.

Note that in general, when the rows of t.p.m are the same, the corresponding Markov chain is a sequence of IID RVs whose distribution is given by the first/any row of the t.p.m.

Example 5: Snoqualmie Falls Precipitation (revisited).
In the Snoqualmie Falls Precipitation problem, we have a t.p.m

$$
\widehat{\mathbf{P}}=\left[\begin{array}{ll}
0.602 & 0.398 \\
0.166 & 0.834
\end{array}\right]
$$

If we consider the 2 -step transition probability,

$$
\widehat{\mathbf{P}}^{2}=\left[\begin{array}{ll}
0.428 & 0.572 \\
0.238 & 0.762
\end{array}\right]
$$

When we consider $n$-step transition probability with $n$ large (in this case, $n \geq 10$ ), it turns out that the $n$-step transition probability matrix becomes

$$
\widehat{\mathbf{P}}^{n}=\left[\begin{array}{ll}
0.294 & 0.706 \\
0.294 & 0.706
\end{array}\right]
$$

This implies that the initial distribution is uninformative - whether it is dry or wet on Jan 18th tells us little about Jan 27th.

Recall that in the previous example, we saw that when the rows of a t.p.m. are the same, the corresponding random variables are IID. Therefore, if we consider another sequence of RVs $\left\{Y_{0}(n)=X_{k}, Y_{1}(n)=\right.$ $\left.X_{k+n}, Y_{2}(n)=X_{k+2 n}, \cdots\right\}$, then $Y_{0}(n), Y_{1}(n), Y_{2}(n), \cdots$ are IID when $n \rightarrow \infty$.

After seeing this example, one may conjecture that if the $\operatorname{limit} \mathbf{P}_{\infty}=\lim _{n \rightarrow \infty} \mathbf{P}^{n}$ will have equal rows. However, this is not always true. A counterexample is

$$
\mathbf{P}=\left[\begin{array}{ll}
0 & 1 \\
1 & 0
\end{array}\right]
$$

### 3.5 First Step Analysis of Markov Chain

First-step analysis is a general strategy for solving many Markov chain problems by conditioning on the first step of the Markov chain. We demonstrate this technique on a simple example - the Gambler's ruin problem.

Gambler's ruin problem: Two players bet one dollar in each round. Player 1 wins with probability $\alpha$ and loses with probability $\beta=1-\alpha$. We assume that player 1 starts with $a$ dollars and player 2 starts with $b$ dollars. Let $X_{n}$ be the fortune of player 1 after $n$ rounds. $X_{n}$ can take values from 0 to $a+b$ :

$$
p_{i j}=P\left(X_{n+1}=j \mid X_{n}=i\right)= \begin{cases}\alpha & \text { if } j=i+1 \\ \beta & \text { if } j=i-1 \\ 0 & \text { otherwise }\end{cases}
$$

Let $T$ be the time of rounds when one of the players loses all his/her money. Because of the randomness of this model, $T$ is also a random variable. We are interested in the probability that player 1 wins the game, which occurs when $X_{T}=a+b$. Apparently, this probability depends on the initial amount of money that player 1 has so we will denote it as

$$
u(a)=P\left(X_{T}=a+b \mid X_{0}=a\right)
$$

Note that $u(0)=0$ and $u(a+b)=1$. We may view $u(a)$ as the probability that the chain is absorbed into the state $a+b$ at the hitting time $T$ when the chain starts at $X_{0}=a$.

First step analysis proceeds as follows:

$$
\begin{aligned}
u(a) & =P\left(X_{T}=a+b \mid X_{0}=a\right) \\
& =\sum_{j=1}^{a+b} P\left(X_{T}=a+b, X_{1}=j \mid X_{0}=a\right) \\
= & \sum_{j=1}^{a+b} P\left(X_{T}=a+b \mid X_{1}=j, X_{0}=a\right) P\left(X_{1}=j \mid X_{0}=a\right) \\
= & \sum_{j=1}^{a+b} P\left(X_{T}=a+b \mid X_{1}=j\right) P\left(X_{1}=j \mid X_{0}=a\right) \quad \text { (Markov property) } \\
= & P\left(X_{T}=a+b \mid X_{1}=a+1\right) P\left(X_{1}=a+1 \mid X_{0}=a\right)+ \\
& P\left(X_{T}=a+b \mid X_{1}=a-1\right) P\left(X_{1}=a-1 \mid X_{0}=a\right) \\
= & u(a+1) \alpha+u(a-1) \beta
\end{aligned}
$$

Therefore, we have $u(a)=u(a+1) \alpha+u(a-1) \beta$ with two boundary conditions: $u(0)=0, u(a+b)=1$. Because $\alpha+\beta=1$,

$$
(\alpha+\beta) u(a)=u(a)=u(a+1) \alpha+u(a-1) \beta
$$

which implies

$$
\alpha(u(a+1)-u(a))=\beta(u(a)-u(a-1))
$$

Define $v(a)=u(a)-u(a-1)$. Then the above leads to

$$
\alpha v(a+1)=\beta v(a) \Rightarrow v(a+1)=\frac{\beta}{\alpha} v(a) .
$$

By telescoping,

$$
v(a+1)=\frac{\beta}{\alpha} v(a)=\left(\frac{\beta}{\alpha}\right)^{2} v(a-1)=\cdots=\left(\frac{\beta}{\alpha}\right)^{a} v(1) .
$$

Using the boundary condition $u(0)=0$,

$$
\begin{aligned}
u(a) & =u(a)-u(0) \\
& =\sum_{j=1}^{a}[u(j)-u(j-1)] \\
& =\sum_{j=1}^{a} v(j) \\
& =v(1) \sum_{j=0}^{a-1}\left(\frac{\beta}{\alpha}\right)^{j} \\
& = \begin{cases}v(1) \times a & \text { if } \alpha=\beta \\
v(1) \times\left(\frac{1-\left(\frac{\beta}{\alpha}\right)^{a}}{1-\left(\frac{\beta}{\alpha}\right)}\right) & \text { if } \alpha \neq \beta\end{cases}
\end{aligned}
$$

To find out $v(1)$, we use the other boundary condition

$$
\begin{aligned}
1 & =u(a+b)=u(a+b)-u(0) \\
& =\sum_{j=1}^{a+b}[u(j)-u(j-1)] \\
& =\sum_{j=1}^{a+b} v(j) \\
& = \begin{cases}v(1) \times(a+b) & \text { if } \alpha=\beta \\
v(1) \times\left(\frac{1-\left(\frac{\beta}{\alpha}\right)^{a+b}}{1-\left(\frac{\beta}{\alpha}\right)}\right) & \text { if } \alpha \neq \beta\end{cases}
\end{aligned}
$$

Therefore,

$$
v(1)= \begin{cases}\frac{1}{a+b} & \text { if } \alpha=\beta \\ \frac{1-\frac{\beta}{\alpha}}{1-\left(\frac{\beta}{\alpha}\right)^{a+b}} & \text { if } \alpha \neq \beta\end{cases}
$$

and

$$
u(a)= \begin{cases}\frac{a}{a+b} & \text { if } \alpha=\beta \\ \frac{1-\left(\frac{\beta}{\alpha}\right)^{a}}{1-\left(\frac{\beta}{\alpha}\right)^{a+b}} & \text { if } \alpha \neq \beta\end{cases}
$$

### 3.6 Classification of States

We now turn to a classification of the states of a Markov chain that is crucial to understanding the behavior of Markov chains.

An equivalence relation " $\sim$ " is a binary relation between elements of a set satisfying

1. Reflexivity: $i \sim i$ for all $i$
2. Symmetry: $i \sim j \Rightarrow j \sim i$
3. Transitivity: $i \sim j, j \sim k \Rightarrow i \sim k$.

For a set $\mathcal{S}$ and $a \in \mathcal{S},\{s \in \mathcal{S}: s \sim a\}$ is called an equivalence class. Equivalence relations will allow us to split Markov chain state spaces into equivalence classes.

State $j$ is accessible from state $i(i \rightarrow j)$ if there exists $m \geq 0$ such that $p_{i j}^{(m)}>0$. We say that $i$ communicates with $j(i \leftrightarrow j)$ if $j$ is accessible from $i$ and $i$ is accessible from $j$. A set of states $\mathcal{C}$ is a communicating class if every pair of states in $\mathcal{C}$ communicates with each other, and no state in $\mathcal{C}$ communicates with any state not in $\mathcal{C}$.

Proposition 3.3 Communication of states is an equivalence relation.

Proof: Reflexivity and symmetry are clear. To prove transitivity, let $i \leftrightarrow j$ and $j \leftrightarrow k$. We then want to show that $i \leftrightarrow k$.

Note that $i \rightarrow j$ if and only if the transition diagram contains a path from $i$ to $j$. So there is a path from $i$ to $j$ and from $j$ to $k$. Concatenate the two to obtain a path from $i$ to $k$, which testifies to the fact that $i \rightarrow k$.

Analogously, we have $k \rightarrow i$ and conclude that $i \leftrightarrow k$.

A set of states $\mathcal{C}$ is closed if $\sum_{j \in \mathcal{C}} p_{i j}=1$ for all $i \in \mathcal{C}$.
Example 6. Consider a Markov chain with the following transition diagram:


Then

- Communication classes: $\{1,2,3\},\{4,5\}$, and $\{6\}$.
- Closed sets: $\{6\},\{4,5,6\},\{1,2,3,4,5,6\}$.

Note that a single state (node in a transition diagram) may belong to multiple closed set. However, a single state can only belong to a communication class.

A Markov chain $\left\{X_{n}\right\}$ is called irreducible if it has only one communication class, i.e., for all $i$ and $j$, $i \leftrightarrow j$. For state $i, d_{i}=\operatorname{gcd}\left\{n \geq 1: p_{i i}^{(n)}>0\right\}$ is called its period, where gcd $=$ greatest common divisor and $d_{i}=+\infty$ if $p_{i i}^{(n)}=0$ for all $n \geq 1$.
Example 7. Consider the example with state space $S=\{0,1,2, \ldots\}$ and $X_{n}$ such that

$$
P\left(X_{n+1}=i \mid X_{n}=0\right)= \begin{cases}p & \text { if } i=1 \\ 1-p & \text { if } i=0 \\ 0 & \text { otherwise }\end{cases}
$$

and for $j \neq 0$

$$
P\left(X_{n+1}=i \mid X_{n}=j\right)= \begin{cases}p & \text { if } i=j+1 \\ 1-p & \text { if } i=j-1 \\ 0 & \text { otherwise }\end{cases}
$$

Then, $d_{2}=\operatorname{gcd}\{2,4,5,6, \ldots\}=1$ though 1 is not in the list (think about why $p_{22}^{(5)}>0$ ).
Example 8: Simple (1-D) Random Walk on the Integers. Consider another example with state space $\mathbb{Z}$. Let $X_{n}$ be the position at time $n$. Then

$$
P\left(X_{n+1}=i-1 \mid X_{n}=i\right)=q \text { and } P\left(X_{n+1}=i+1 \mid X_{n}=i\right)=p
$$

with $p=1-q$. Suppose we start at 0 , then it is clear that we cannot return to 0 after an odd number of steps, so $p_{00}^{(2 n+1)}=0$ for all $n$, i.e.

$$
d_{0}=\operatorname{gcd}\left\{n \geq 1: p_{00}^{(n)}>0\right\}=\operatorname{gcd}\{2,4,6, \ldots\}=2
$$

Proposition 3.4 Period is a communication class property. Namely, $i \leftrightarrow j \Rightarrow d_{i}=d_{j}$.

Proof: $i \leftrightarrow j \Rightarrow$ there exists $n_{1}, n_{2}$ such that $p_{i j}^{\left(n_{1}\right)}>0$ and $p_{j i}^{\left(n_{2}\right)}>0$. Then, by Chapman-Kolmogorov:

$$
p_{i i}^{\left(n_{1}+n_{2}\right)}=\sum_{k} p_{i k}^{\left(n_{1}\right)} p_{k i}^{\left(n_{2}\right)} \geq p_{i j}^{\left(n_{1}\right)} p_{j i}^{\left(n_{2}\right)}>0
$$

Consequently we know that $d_{i} \mid n_{1}+n_{2}$. For example, suppose $n_{1}=3$ and $n_{2}=5$, then $n_{1}+n_{2}=3+5=8$. Therefore we know that $d_{i} \leq 8$, i.e. we could have $d_{i}=8,4,2,1$; we know we could return after 8 time steps, but it could be less.

Note: $a \mid b$ means $a$ divides $b$, i.e. there is an integer $c$ s.t. $b=a c$.
Now, take any $n$ such that $p_{j j}^{(n)}>0$. Then

$$
\begin{aligned}
p_{i i}^{\left(n_{1}+n_{2}+n\right)} & =\sum_{k} p_{i k}^{\left(n+n_{1}\right)} p_{k i}^{\left(n_{2}\right)} \geq p_{i j}^{\left(n+n_{1}\right)} p_{j i}^{\left(n_{2}\right)} \\
& =\left[\sum_{k} p_{i k}^{\left(n_{1}\right)} p_{k j}^{(n)}\right] p_{j i}^{\left(n_{2}\right)} \geq p_{i j}^{\left(n_{1}\right)} p_{j j}^{(n)} p_{j i}^{\left(n_{2}\right)}>0 .
\end{aligned}
$$

Hence, $d_{i} \mid n_{1}+n_{2}+n$.
Together, $n_{1}+n_{2}=c_{1} d_{i}$ and $n_{1}+n_{2}+n=c_{2} d_{i}$ imply that $n=\left(c_{2}-c_{1}\right) d_{i}$ and as a result, $d_{i} \mid n$ for all $n$ such that $p_{j j}^{(n)}>0$.

Since $d_{i}$ is a divisor of the set $\left\{n: p_{j j}^{(n)}>0\right\}$ and $d_{j}$ is the greatest common divisor of the same set (by definition of period), $d_{i} \leq d_{j}$.

By symmetry, $d_{j} \leq d_{i} \Rightarrow d_{i}=d_{j}$.

In a communication class, all states have the same period. Since all states communicate in an irreducible Markov chains, it makes sense to define the period of such a Markov chain. If $d_{i}=1$, state $i$ is called aperiodic. An irreducible Markov chain with period 1 is also called aperiodic.

Theorem 3.5 (Lattice Theorem (Brémaud p.75)) Suppose $d$ is the period of an irreducible Markov chain. Then for all states $i, j$ there exists $m \in\{0, \ldots, d-1\}$ and $k \geq 0$ such that

$$
p_{i j}^{(m+n d)}>0, \quad \forall n \geq k
$$

Theorem 3.6 (Cyclic Classes) For any irreducible Markov chain one can find a unique partition of $S$ into $d$ classes $C_{0}, C_{1}, \ldots, C_{d-1}$ such that for all $k$, and for $i \in C_{k}$,

$$
\sum_{j \in C_{k+1}} p_{i j}=1
$$

where, by convention $C_{d}=C_{0}$ and where $d$ is maximal (that is, there is no other such partition $C_{0}^{\prime}, C_{1}^{\prime}, \ldots, C_{d^{\prime}-1}^{\prime}$ with $\left.d^{\prime}>d\right)$.

Proof: Fix a state $i$ and classify states $j$ by the value of $m$ in Lattice Theorem.
The number $d \geq 1$ is the period of the chain. The classes $C_{0}, C_{1}, \ldots, C_{d-1}$ are called the cyclic classes.
Example 8: Simple (1-D) Random Walk on the Integers (revisited). Random walk on $S=\mathbb{Z}=$ $C_{0}+C_{1}$ where $C_{0}$ and $C_{1}$ are the sets of even and odd integers.

### 3.7 Strong Markov Property

The Markov property states that the random variable at time $n+m$ conditional on its behavior at time $n$ is independent of the at time prior to $n$. However, what if the time $n$ is random?

Say we are interested in the behavior of $X_{T+m}$ given $X_{T}$, where $T$ is the first time that the Markov chain hits the state 0 . Do we still have the Markov property?

Some random time does not have the Markov property. Recall that the Markov property states that for any $m<n<k, X_{m} \perp X_{k} \mid X_{n}$. Let $\left\{X_{n}\right\}$ be a Markov chain with state space $S=\{1,2,3\}$ and consider a random time

$$
T=\inf \left\{n \geq 1:\left(X_{n-1}, X_{n}, X_{n+1}\right)=(2,1,3) \text { or }(3,1,2)\right\}
$$

Then the probability

$$
P\left(X_{k}=3 \mid X_{n}=1, X_{m}=2\right)=P\left(X_{T+1}=3 \mid X_{T}=1, X_{T-1}=2\right)=1 \neq P\left(X_{k}=3 \mid X_{n}=1\right)
$$

because when $X_{m}=X_{n-1}=3, P\left(X_{k}=2 \mid X_{n}=1, X_{m}=3\right)=P\left(X_{T+1}=2 \mid X_{T}=1, X_{T-1}=3\right)=1$ so $P\left(X_{k}=3 \mid X_{n}=1, X_{m}=3\right)=0$. Thus, the conditional probability of $X_{k}$ given $X_{n}$ depends on $X_{m}$, which is a violation of Markov property.

Therefore, it is crucial to identify a class of random time such that the Markov property holds. It turns out that there is a simple class of random times that has the Markov property. This class is called the stopping time.

A random variable $\tau \in\{1,2,3, \cdots\} \cup\{\infty\}$ is called a stopping time if the event $\{\tau=m\}$ can be expressed in terms of $X_{0}, X_{1}, \cdots, X_{m}$. Intuitively, a stopping time is a random time such that we can observe it when the time arrives.

## Examples 9: Stopping times.

- Return time. Let $T_{i}=\inf \left\{n \geq 1: X_{n}=i\right\}$ is a stopping time because $\left\{T_{i}=m\right\}=\left\{X_{1} \neq\right.$ $\left.i, \cdots, X_{m-1} \neq i, X_{m}=i\right\} . T_{i}$ is interpreted as the first time the chain returns to state $i$.
- Successive Returns. Let $\tau_{k}$ be the time of the $k$-th return to state $i$ (note that $\tau_{1}=T_{i}$ ). Then $\tau_{k}$ is a stopping time because

$$
\left\{\tau_{k}=m\right\}=\left\{\sum_{n=1}^{m} I\left(X_{n}=i\right)=k, X_{m}=i\right\} .
$$

- Counterexample - non-stopping time: Let $\tau=\inf \left\{n \geq 1: X_{n+1}=i\right\}$ is not a stopping time because when the time arrives at $m,\{\tau=m\}=\left\{X_{1} \neq i, \cdots, X_{m} \neq i, X_{m+1}=i\right\}$ depends on $X_{m+1}$.

Stopping time is a very important class of random variable in statistics. Many statistical procedure involves a stopping time. For instance, if we are performing a sequence of experiments and we will stop when we observe certain behavior such as a high signal or enough anomaly. Then the time (of related to the number of sample) is a stopping time. If we want to use data from this sequence of experiments, then we need to use theorems of stopping time (such as optional sampling theorem).

Theorem 3.7 (Strong Markov Property) Let $\left\{X_{n}\right\}$ be a homogeneous Markov chain with a transition probability matrix $\mathbf{P}=\left\{p_{i j}\right\}$ and let $\tau$ be a stopping time with respect to $\left\{X_{n}\right\}$. Then for any integer $k$,

$$
P\left(X_{\tau+k}=j \mid X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)=P\left(X_{k}=j \mid X_{0}=i\right)=p_{i j}^{(k)}
$$

and

$$
P\left(X_{\tau+k}=j \mid X_{\tau}=i\right)=P\left(X_{k}=j \mid X_{0}=i\right)=p_{i j}^{(k)} .
$$

Proof: We first prove the first equality.

$$
\begin{align*}
P\left(X_{\tau+k}=j \mid X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right) & =\frac{P\left(X_{\tau+k}=j, X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)}{P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)} \\
& =\frac{\sum_{r=1}^{\infty} P\left(X_{\tau+k}=j, X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right)}{P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)} \tag{3.4}
\end{align*}
$$

Now, because $\tau$ is a stopping time, the event $\{\tau=r\}$ can be expressed as $X_{0}, \cdots, X_{r}$ so the Markov property implies

$$
P\left(X_{\tau+k}=j \mid X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right)=P\left(X_{\tau+k}=j, X_{\tau}=i\right)=p_{i j}^{(k)}
$$

Therefore, equation (3.4) becomes

$$
\begin{aligned}
P\left(X_{\tau+k}=j \mid X_{\tau}=i, 0\right. & \left.\leq \ell<\tau, X_{\ell}=i_{\ell}\right) \\
& =\frac{\sum_{r=1}^{\infty} P\left(X_{\tau+k}=j, X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right)}{P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)} \\
& =\frac{\sum_{r=1}^{\infty} P\left(X_{\tau+k}=j \mid X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right) P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right)}{P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)} \\
& =\frac{\sum_{r=1}^{\infty} p_{i j}^{(k)} P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right)}{P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)} \\
& =p_{i j}^{(k)} \frac{\sum_{r=1}^{\infty} P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}, \tau=r\right)}{P\left(X_{\tau}=i, 0 \leq \ell<\tau, X_{\ell}=i_{\ell}\right)} \\
& =p_{i j}^{(k)} .
\end{aligned}
$$

The second equality follows simply from the first equality:

$$
\begin{aligned}
P\left(X_{\tau+k}=j \mid X_{\tau}=i\right) & =\frac{\sum_{r=1}^{\infty} P\left(X_{\tau+k}=j \mid X_{r}=i, \tau=r\right) P\left(X_{r}=i, \tau=r\right)}{P\left(X_{\tau}=i\right)} \\
& =p_{i j}^{(k)} \frac{\sum_{r=1}^{\infty} P\left(X_{r}=i, \tau=r\right)}{P\left(X_{\tau}=i\right)} \\
& =p_{i j}^{(k)} .
\end{aligned}
$$

### 3.8 Stationary distribution

It is often of great interest to study the limiting behavior of a Markov chain $X_{n}$ when $n \rightarrow \infty$. Here, for simplicity, we assume that our Markov chain is homogeneous. A property of limiting behavior is that $X_{n}$ and $X_{n+1}$ should have the same distribution when $n$ is large. So we are interested in understanding if a Markov chain will eventually converge to a 'stable' distribution (formally, we will call it a stationary distribution). In particular, we would like to know given a Markov chain,

- does this chain has a stationary distribution?
- if so, what is the stationary distribution?
- and does this stationary distribution unique?

It turns out that to answer these questions, we will use concepts related to return time. Thus, we start with understanding properties about return time.

### 3.8.1 Return Times

Let $N_{i}=\sum_{i=1}^{\infty} I\left(X_{n}=i\right)$ denotes the number of visits of $\left\{X_{n}\right\}$ to state $i$ not counting the initial state. We also define the following notations:

$$
P\left(\cdot \mid X_{0}=i\right)=P_{i}(\cdot), \quad \mathbb{E}\left(\cdot \mid X_{0}=i\right)=\mathbb{E}_{i}(\cdot)
$$

Note that the quantity $N_{i}$ may equal to $\infty$. It is a finite number with a non-zero probability if there are some states such that when the chain enters one of them, the chain never go back to state $i$. Later we will describe this phenomena using the concept of transient states and recurrent states.

Example 6 (revisited). Consider a Markov chain with the following transition diagram:


As can be seen easily, when the Markov chain enters states $\{4,5,6\}$, it never comes back to any of $\{1,2,3\}$. Thus, $N_{1}$ takes a non-trivial probability to be a finite number.

Let $T_{i}=\inf \left\{n \geq 1: X_{n}=i\right\}$ be the return time. Then the following events can be defined using either $T_{i}$ or $N_{i}$ :

$$
\left\{T_{i}=\infty\right\}=\left\{N_{i}=0\right\}, \quad\left\{T_{i}<\infty\right\}=\left\{N_{i}>0\right\}
$$

These are useful later.
We then define $f_{j i}=P_{j}\left(T_{i}<\infty\right)=P_{j}\left(N_{i}>0\right)$ to be the probability of reaching state $i$ in a finite number of time when the chain starts at state $j$. Note that because $P_{j}\left(T_{i}=\infty\right)+P_{j}\left(T_{i}<\infty\right)=1$, we have $f_{i i}=P_{i}\left(T_{i}<\infty\right)$ and $P_{i}\left(T_{i}=\infty\right)=1-f_{i i}$.

## Proposition 3.8

$$
P_{j}\left(N_{i}=r\right)= \begin{cases}f_{j i} f_{i i}^{r-1}\left(1-f_{i i}\right) & \text { if } r \geq 1 \\ 1-f_{j i} & \text { if } r=0\end{cases}
$$

Proof: The case $r=0$ is very simple because $\left\{N_{i}=0\right\}=\left\{T_{i}=\infty\right\}$. Thus, $P_{j}\left(N_{i}=0\right)=P_{j}\left(T_{i}=\infty\right)=$ $1-P_{j}\left(T_{i}<\infty\right)=1-f_{j i}$.

For the rest of cases, we will do a proof by induction. Before doing that, we first investigate the case $P_{j}\left(N_{i}=r\right)$ for $r>0$. Let $\tau_{r}$ be the $r$-th return time. Note that the event $\left\{X_{\tau_{r}}=i\right\}=\left\{N_{i} \geq r\right\}$.

Then

$$
\begin{aligned}
P_{j}\left(N_{i}=r\right) & =P_{j}\left(N_{i}=r, X_{\tau_{r}}=i\right) \\
& =P_{j}\left(N_{i}=r \mid X_{\tau_{r}}=i\right) P_{j}\left(X_{\tau_{r}}=i\right) \\
& =P_{j}\left(\sum_{t=\tau_{r}+1}^{\infty} I\left(X_{t}=i\right)=0 \mid X_{\tau_{r}}=i\right) P_{j}\left(X_{\tau_{r}}=i\right) \\
& =P_{i}\left(\sum_{t=1}^{\infty} I\left(X_{t}=i\right)=0 \mid X_{0}=i\right) P_{j}\left(X_{\tau_{r}}=i\right) \quad \text { (Strong Markov property) } \\
& =P_{i}\left(N_{i}=0\right) P_{j}\left(N_{i} \geq r\right) \\
& =P_{i}\left(T_{i}=\infty\right) P_{j}\left(N_{i} \geq r\right)
\end{aligned}
$$

Therefore, we conclude

$$
P_{j}\left(N_{i}=r\right)=P_{i}\left(T_{i}=\infty\right) P_{j}\left(N_{i} \geq r\right)=\left(1-f_{i i}\right) P_{j}\left(N_{i} \geq r\right)
$$

To start with the proof by induction, consider $r=1 . P_{j}\left(N_{i} \geq 1\right)=1-P_{j}\left(N_{i}=0\right)=f_{j i}$ so $P_{j}\left(N_{i}=1\right)=$ $\left(1-f_{i i}\right) f_{j i}$, which agrees with what we need for $r=1$.

Assume that it works for $r \leq k$. Now we show that it works for $r=k+1$.

$$
\begin{align*}
P_{j}\left(N_{i} \geq k+1\right) & =1-P_{j}\left(N_{i} \leq k\right) \\
& =1-\sum_{r=1}^{k} f_{j i} f_{i i}^{r-1}\left(1-f_{i i}\right)-\left(1-f_{j i}\right) \\
& =1-f_{j i}\left(1-f_{i i}\right) \sum_{r=1}^{k} f_{i i}^{r-1}-\left(1-f_{j i}\right)  \tag{3.5}\\
& =1-f_{j i}\left(1-f_{i i}^{k}\right)-\left(1-f_{j i}\right) \\
& =f_{j i} f_{i i}^{k} .
\end{align*}
$$

Thus,

$$
P_{j}\left(N_{i}=k+1\right)=P_{j}\left(N_{i} \geq k+1\right)-P_{j}\left(N_{i} \geq k+2\right)=f_{j i} f_{i i}^{k}-f_{j i} f_{i i}^{k+1}=f_{j i} f_{i i}^{k}\left(1-f_{i i}\right)
$$

which is the desired quantity.

The above formula also gives an interesting result on the case of 'starting from state $i$, returning to state $i$ ' when we set $j=i$ :

$$
P_{i}\left(N_{i}=r\right)=f_{i i}\left(1-f_{i i}\right), \quad P_{i}\left(N_{r}>r\right)=f_{i i}^{r+1}
$$

where $f_{i i}=P_{i}\left(T_{i}<\infty\right)$.
We have seen many situations that $T_{i}$ and $N_{i}$ are closely related. Here is another result about their relationship.

## Corollary 3.9

$$
P_{i}\left(N_{i}=\infty\right)=1 \Leftrightarrow P_{i}\left(T_{i}<\infty\right)=1
$$

and

$$
P_{i}\left(T_{i}<\infty\right)<1 \Leftrightarrow P_{i}\left(N_{i}=\infty\right)=0 \Leftrightarrow \mathbb{E}_{i}\left(N_{i}\right)<\infty .
$$

Proof: Homework.
Corollary 3.9 links the finiteness of $T_{i}$ and $N_{i}$ and also relates it to the expectation. With the following formula of expectation, Corollary 3.9 will be very useful:

$$
\begin{equation*}
\mathbb{E}(X)=\sum_{t=1}^{\infty} P(X \geq t) \tag{3.6}
\end{equation*}
$$

when $X$ is a random variable taking integer values.

### 3.8.2 Recurrence and Transience

Based on the return time property, we classify a state $i$ as

$$
\begin{cases}\text { recurrent/persistent, } & \text { if } P_{i}\left(T_{i}<\infty\right)=f_{i i}=1 \\ \text { transient, } & \text { otherwise. }\end{cases}
$$

Furthermore, a recurrent state is called


Note that: either $P_{i}\left(N_{i}=\infty\right)=0$ or $P_{i}\left(N_{i}=\infty\right)=1$, with nothing in between (if $f_{i i}<1$, then $P_{i}\left(N_{i}=\right.$ $\infty)=0$; if $f_{i i}=1$, then $P_{i}\left(N_{i}=\infty\right)=1$ ). This, together with Corollary 3.9, implies that $\mathbb{E}_{i}\left(N_{i}\right)=\infty \Longleftrightarrow$ $P_{i}\left(N_{i}=\infty\right)=1$.

Note that:

$$
f_{i i}=P_{i}\left(T_{i}<\infty\right)=1 \Longleftrightarrow P_{i}\left(N_{i}=\infty\right)=1
$$

In other words, if a Markov chain returns to state $i$ in finite time, then the chain visits this state infinitely often.

Proposition 3.10 State $i$ is recurrent $\Longleftrightarrow \sum_{n=1}^{\infty} p_{i i}^{(n)}=\infty$.

Proof: State $i$ is recurrent $\Longleftrightarrow P_{i}\left(T_{i}<\infty\right)=f_{i i}=1 \Longleftrightarrow P_{i}\left(N_{i}=\infty\right)=1$ by Corollary 3.9.
It is easy to see that $P_{i}\left(N_{i}=\infty\right)=1 \Longleftrightarrow \mathbb{E}_{i}\left(N_{i}\right)=\infty$.
Using equation (3.6), $\mathbb{E}_{i}\left(N_{i}\right)=\sum_{n=1}^{\infty} p_{i i}^{(n)}$ and the result follows.

Proposition 3.11 Recurrence is a communication class property, i.e. if $i \leftrightarrow j$ and $i$ is recurrent, then $j$ is recurrent.

Proof: Homework.

Example: Gambler's Ruin. Recall that in Gambler's ruin, 0 and $a+b$ states are absorbing. herefore, $\sum_{n=1}^{\infty} p_{00}^{(n)}=\sum_{n=1}^{\infty} p_{a+b, a+b}^{(n)}=\sum_{n=1}^{\infty} 1=\infty$. Hence, 0 and $a+b$ are recurrent states. Once they are reached we stay there forever. Consider state 1 :

$$
P_{1}\left(T_{1}<\infty\right)=1-P_{1}\left(T_{1}=\infty\right) \leq 1-q<1 \text { if } q \in(0,1)
$$

Therefore, by definition, 1 is a transient state. Since states $\{1, \ldots, a+b-1\}$ form a communication class, all states in this class are also transient. These states are transient because they occur a finite number of times before absorption into states 0 or $a+b$.

Example 8: 1-D Random Walk (revisited). Let $X_{n}$ be a random walk on the set of all integers $\mathbb{Z}$ such that

$$
p_{i j}= \begin{cases}p & \text { if } j=i+1 \\ q:=1-p & \text { if } j=i-1\end{cases}
$$

Let's study recurrence of state 0 . We know that $p_{00}^{(2 n+1)}=0$ for all $n \geq 0$ and that, conditional on $X_{0}=0$, $X_{2 n}={ }_{d} \xi_{1}+\cdots+\xi_{2 n}$, where $\xi_{1}, \ldots, \xi_{n}$ are i.i.d. with $P\left(\xi_{i}=1\right)=1-P(\xi=-1)=p$. Hence,

$$
p_{00}^{(2 n)}=P\left(X_{2 n}=0 \mid X_{0}=0\right)=\binom{2 n}{n} p^{n} q^{n}
$$

Recall that Stirlings formula says that $n!\sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2 \pi}$, meaning that

$$
\lim _{n \rightarrow \infty} \frac{n!}{n^{n+\frac{1}{2}} e^{-n} \sqrt{2 \pi}}=1
$$

Therefore,

$$
\begin{aligned}
p_{00}^{(2 n)} & =\frac{(2 n)!}{n!n!} p^{n} q^{n} \\
& \sim \frac{(2 n)^{2 n+\frac{1}{2}} e^{-2 n} \sqrt{2 \pi}}{n^{2 n+1} e^{-2 n} 2 \pi}(p q)^{n} \\
& =\frac{2^{2 n+\frac{1}{2}} n^{2 n+\frac{1}{2}}}{n^{2 n+1} 2^{\frac{1}{2}} \sqrt{\pi}}(p q)^{n}=\frac{(p q)^{n} 2^{2 n}}{\sqrt{\pi n}}=\frac{(4 p q)^{n}}{\sqrt{\pi n}} .
\end{aligned}
$$

We deduce that

$$
\sum_{n=1}^{\infty} p_{00}^{(n)}=\sum_{n=1}^{\infty} p_{00}^{(2 n)}=\infty \quad \Leftrightarrow \quad 4 p q \geq 1 \quad \Leftrightarrow \quad p=q=\frac{1}{2}
$$

(Ratio Test: Let $\sum_{n=1}^{\infty} a_{n}$ be a series which satisfies $\lim _{n \rightarrow \infty}\left|\frac{a_{n+1}}{a_{n}}\right|=k$. If $k>1$ the series diverges, if $k<1$ the series converges.) Conclusion: Only the symmetric random walk is recurrent on $\mathbb{Z}$. Interestingly, the symmetric random walk on $\mathbb{Z}^{2}$ is also recurrent, but it is transient on $\mathbb{Z}^{n}$ for $n \geq 3$, See Brémaud (1999, p. 98).

### 3.8.3 Invariant Measures

With the knowledge about recurrence, we are able to talk about the invariant measures and stationary distribution of a stochastic matrix.

A vector $x \neq 0$ is called an invariant measure of a stochastic matrix $\mathbf{P}$ if

- $x_{i} \geq 0$ for each $i$, and
- $x^{T} \mathbf{P}=x^{T}$, i.e., $x_{i}=\sum_{j} x_{j} p_{j i}$ for each $i$.

A probability vector $\pi$ on a Markov chain state space is called a stationary distribution of a stochastic $\operatorname{matrix} \mathbf{P}$ if $\pi^{T} \mathbf{P}=\pi^{T}$, i.e., $\pi_{i}=\sum_{j} \pi_{j} p_{j i}$ for each $i$.

The equation $x^{T} \mathbf{P}=x^{T}$ or $\pi^{T} \mathbf{P}=\pi^{T}$ is also called the global balance equaitons - the probability flow in equals the flow out. Note that for an invariance measure $x$ such that $c=\sum_{i} x_{i}<\infty, c^{-1} x$ is a stationary distribution. But it may happen that $c=\infty$ for some invariant measure so one cannot always normalize it.

Example 9: Two-State Markov Chain. Consider a Markov chain with two states and a transition probability matrix

$$
\mathbf{P}=\left[\begin{array}{cc}
1-p & p \\
q & 1-q
\end{array}\right] \quad 0<p<1, \quad 0<q<1
$$

The global balance equations:

$$
\begin{aligned}
& {\left[\pi_{0}, \pi_{1}\right]\left[\begin{array}{cc}
1-p & p \\
q & 1-q
\end{array}\right]=\left[\pi_{0}, \pi_{1}\right] \text { or }} \\
& \left\{\begin{array}{ll}
(1-p) \pi_{0}+q \pi_{1} & =\pi_{0} \\
p \pi_{0}+(1-q) \pi_{1} & =\pi_{1}
\end{array} \quad \Rightarrow \quad p \pi_{0}=q \pi_{1} \quad \Rightarrow \quad \pi_{0}=\frac{q}{p} \pi_{1} .\right.
\end{aligned}
$$

Using that $\pi_{0}+\pi_{1}=1$, we obtain

$$
\frac{q}{p} \pi_{1}+\pi_{1}=1 \quad \Rightarrow \quad \pi_{1}=\frac{p}{p+q}
$$

and deduce that the global balance equations have the unique solution

$$
\pi^{T}=\left[\frac{q}{p+q}, \frac{p}{p+q}\right]
$$

which is the stationary distribution.
Example: Gambler's Ruin (simple version). Let the total fortune of both players be $a+b=4$. Then

$$
\mathbf{P}=\left[\begin{array}{lllll}
1 & 0 & 0 & 0 & 0 \\
q & 0 & p & 0 & 0 \\
0 & q & 0 & p & 0 \\
0 & 0 & q & 0 & p \\
0 & 0 & 0 & 0 & 1
\end{array}\right]
$$

By inspection, vectors $\pi_{\alpha}^{T}=[\alpha, 0,0,0,1-\alpha]$ satisfy global balance equations: $\pi_{\alpha}^{T} \mathbf{P}=\pi_{\alpha}^{T}$ for any $\alpha \in(0,1)$. So the Gambler's ruin chain has an uncountable number of stationary distributions.

Here, we see the case where a Markov chain may have infinite number of stationary distribution. And in some cases it may not even have a stationary distribution! So returning to our original questions, we would like to know (i) when will a Markov chain has a stationary distribution? and (ii) how to find a stationary distribution? and (ii) when the stationary distribution will be unique?

The following proposition partially answer the first question. Note that a Markov chain is recurrent if all its states are recurrent.

Proposition 3.12 Let $\left\{X_{n}\right\}$ be an irreducible, recurrent, homogeneous Markov chain with transition probability matrix $\mathbf{P}$. For each $i \in S$ define

$$
y_{i}=\mathbb{E}_{0}\left[\sum_{n=1}^{\infty} I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right]
$$

where 0 is an arbitrary reference state and $T_{0}=\inf \left\{n \geq 1: X_{n}=0\right\}$ is the first return time to 0 . Then $y_{i} \in(0, \infty)$ for all $i \in S$, and $\mathbf{y}^{T}=\left[y_{0}, y_{1}, \ldots\right]$ is an invariant measure of $\mathbf{P}$.

Note: For $i \neq 0, y_{i}$ is the expected number of visits to state $i$ before returning to 0 .
Before starting the proof, we note the following three properties.
(P1) When $i=0$,

$$
y_{0}=\mathbb{E}_{0}\left[\sum_{n=1}^{\infty} I\left(X_{n}=0\right) I\left(n \leq T_{0}\right)\right]=1
$$

because for $n \geq 1, X_{n}=0$ if and only if $n=T_{0}$.
(P2)

$$
\begin{aligned}
\sum_{i \in S} y_{i} & =\sum_{i \in S} \mathbb{E}_{0}\left[\sum_{n=1}^{\infty} I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right] \\
& =\mathbb{E}_{0}\left[\sum_{n=1}^{\infty} \sum_{i \in S} I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right] \\
& =\mathbb{E}_{0}\left[\sum_{n=1}^{\infty} I\left(n \leq T_{0}\right)\right] \\
& =\mathbb{E}_{0}\left(T_{0}\right) .
\end{aligned}
$$

(P3) For any $i \in S$, we define

$$
q_{0 i}^{(n)} \equiv \mathbb{E}_{0}\left(I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right)=P_{0}\left(X_{1} \neq 0, X_{2} \neq 0, \cdots, X_{n-1} \neq 0, X_{n}=i\right)
$$

to be the probability of visiting state $i$ at time point $n$ before returning to state 0 . Thus,

$$
y_{i}=\sum_{n=1}^{\infty} \mathbb{E}_{0}\left(I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right)=\sum_{n=1}^{\infty} q_{0 i}^{(n)}
$$

and $q_{0 i}^{(1)}=\mathbb{E}_{0}\left(I\left(X_{1}=i\right) I\left(1 \leq T_{0}\right)\right)=p_{0 i}$.

Proof: This proof consists of two parts. In the first part, we prove that each $y_{i}$ satisfies $y_{i}=\sum_{j \in S} y_{j} p_{j i}$. In the second part, we will show that $0<y_{i}<\infty$ for every $i \in S$.

Part 1. To show that $y_{i}=\sum_{j \in S} y_{j} p_{j i}$, we analyze $q_{0 i}^{(n)}$ defined in property (P3):

$$
\begin{aligned}
q_{0 i}^{(n)} & =P_{0}\left(X_{1} \neq 0, X_{2} \neq 0, \cdots, X_{n-1} \neq 0, X_{n}=i\right) \\
& =\sum_{j \neq 0} P_{0}\left(X_{1} \neq 0, X_{2} \neq 0, \cdots, X_{n-1}=j, X_{n}=i\right) \\
& =\sum_{j \neq 0} P_{0}\left(X_{n}=i \mid X_{1} \neq 0, X_{2} \neq 0, \cdots, X_{n-1}=j\right) \underbrace{P_{0}\left(X_{1} \neq 0, X_{2} \neq 0, \cdots, X_{n-1}=j\right)}_{=q_{0 j}^{(n-1)}} \\
& =\sum_{j \neq 0} P\left(X_{n}=i \mid X_{n-1}=j\right) q_{0 j}^{(n-1)} \quad \text { (Markov property) } \\
& =\sum_{j \neq 0} q_{0 j}^{(n-1)} p_{j i} .
\end{aligned}
$$

Thus,

$$
\begin{aligned}
y_{i} & =\sum_{n=1}^{\infty} q_{0 i}^{(n)} \\
& =p_{0 i}+\sum_{n=2}^{\infty} q_{0 i}^{(n)} \\
& =p_{0 i}+\sum_{n=2}^{\infty} \sum_{j \neq 0} q_{0 j}^{(n-1)} p_{j i} \\
& =p_{0 i}+\sum_{n=1}^{\infty} \sum_{j \neq 0} q_{0 j}^{(n)} p_{j i} \\
& =p_{0 i}+\sum_{j \neq 0} \underbrace{\left(\sum_{n=1}^{\infty} q_{0 j}^{(n)}\right)}_{=y_{j}} p_{j i} \\
& =\underbrace{y_{0}}_{=1} p_{0 i}+\sum_{j \neq 0} y_{j} p_{j i} \\
& =\sum_{j \in S}^{\sum_{j}} y_{j} p_{j i} .
\end{aligned}
$$

Part 2. Now we show that $0<y_{i}<\infty$. First note that $y_{0}=1$ so we only need to focus on cases $y \neq 0$.
Because the Markov chain is irreducible, for each state $i$ there exists a number $n(i) \geq 1$ such that $p_{0 i}^{(n(i))}>0$. Then using the fact that $y^{T}=y^{T} \mathbf{P}$ implying $y^{T}=y^{T} \mathbf{P}^{(n(i))}$,

$$
y_{i}=\sum_{j \in S} y_{j} p_{j i}^{(n(i))}=\underbrace{y_{0} p_{0 i}^{(n(i))}}_{>0}+\sum_{j \neq 0} y_{j} p_{j i}^{(n(i))}>0
$$

To show that $y_{i}<\infty$, we prove by contradiction. Assume that $y_{i}=\infty$. Because the Markov chain is irreducible, there exists a constant $k(i)$ such that $p_{i 0}^{(k(i))}>0$. Then

$$
y_{0}=\sum_{j \in S} y_{j} p_{j 0}^{(k(i))}=\underbrace{y_{i} p_{i 0}^{(k(i))}}_{=\infty}+\sum_{j \neq i} y_{j} p_{j 0}^{(k(i))}=\infty,
$$

a contradiction. Thus, $y_{i}<\infty$.

Proposition 3.13 The invariant measure of an irreducible and recurrent chain is unique up to a multiplicative factor.

Proof: See Brémaud (1999, p. 102).

Proposition 3.14 An irreducible, recurrent and homogeneous Markov chain is positive recurrent $\Leftrightarrow$ all of its invariant measures y satisfy $\sum_{i \in S} y_{i}<\infty$.

Proof: By Proposition 3.12, there is an invariant measure $y$ with

$$
y_{i}=\mathbb{E}_{0}\left[I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right] .
$$

Moreover, by Proposition 3.13, this invariant measure is unique up to a multiplicative factor. So what remains to prove is to show that $\sum_{i \in S} y_{i}<\infty$.

Using property (P2),

$$
\sum_{i \in S} y_{i}=\mathbb{E}_{0}\left(T_{0}\right)
$$

Therefore, positive recurrent $\Longleftrightarrow \mathbb{E}_{0}\left(T_{0}\right)<\infty \Longleftrightarrow \sum_{i \in S} y_{i}<\infty$.

To see why positive recurrent is important, consider the following example about a $1-D$ random walk on all integers $\mathbb{Z}$ with $p \neq q$ is transient and recurrent if $p=q=0.5$. This Markov chain has an invariant measure $\mathbf{y}^{T}=[1,1, \ldots]$ for any $p$ and $q$ since

$$
\mathbf{P}=\left[\begin{array}{ccccccc}
\ddots & \ddots & \ddots & \cdots & \cdots & \cdots & \cdots \\
\cdots & q & 0 & p & \cdots & \cdots & \cdots \\
\cdots & \cdots & q & 0 & p & \cdots & \cdots \\
\cdots & \cdots & \cdots & q & 0 & p & \cdots \\
\cdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots
\end{array}\right]
$$

Since this measure is not normalizable (the state space is $\mathbb{Z}$ ), the 1-D random walk can not be positive recurrent. Thus, we see that an irreducible homogeneous Markov chain can have an invariant measure and still be transient or null recurrent.

Lemma 3.15 Let $\left\{X_{n}\right\}$ be a homogeneous Markov chain with state space $S$ and $n$-step transition probability matrix $\mathbf{P}^{n}=\left\{p_{i j}^{(n)}\right\}$. If $i \in S$ is a transient state, then $\lim _{n \rightarrow \infty} p_{j i}^{(n)}=0$ for all $j \in S$.

Proof: This proof relies a trick - if $\sum_{n=1}^{\infty} p_{j i}^{(n)}<\infty$, then $\lim _{n \rightarrow \infty} p_{j i}^{(n)}=0$. Thus, we only need to show that $\sum_{n=1}^{\infty} p_{j i}^{(n)}<\infty$ when $i \in S$ is a transient state.
By definition,

$$
\sum_{n=1}^{\infty} p_{j i}^{(n)}=\sum_{n=1}^{\infty} P_{j}\left(X_{n}=i\right)=\sum_{n=1}^{\infty} \mathbb{E}_{j}\left(I\left(X_{n}=i\right)\right)=\mathbb{E}_{j}\left(\sum_{n=1}^{\infty} I\left(X_{n}=i\right)\right)=\mathbb{E}_{j}\left(N_{i}\right)
$$

So we can switch our goal to $\mathbb{E}_{j}\left(N_{i}\right)$.
Because $N_{i}$ is a RV taking integer values, we can rewrite its expectation as

$$
\mathbb{E}_{j}\left(N_{i}\right)=\sum_{k=1}^{\infty} P_{j}\left(N_{i} \geq k\right)=\sum_{k=0}^{\infty} P_{j}\left(N_{i} \geq k+1\right)
$$

So we only need to compute each of $P_{j}\left(N_{i}>k\right)$. Now, recalled from the proof of Proposition 3.8 (equation (3.5)),

$$
P_{j}\left(N_{i} \geq k+1\right)=f_{j i} f_{i i}^{k}
$$

We obtain

$$
\mathbb{E}_{j}\left(N_{i}\right)=\sum_{k=0}^{\infty} P_{j}\left(N_{i} \geq k+1\right)=f_{j i} \sum_{k=0}^{\infty} f_{i i}^{k}
$$

Because the state $i$ is transient, $f_{i i}<1$ so the above summation becomes

$$
\mathbb{E}_{j}\left(N_{i}\right)=f_{j i} \sum_{k=0}^{\infty} f_{i i}^{k}=\frac{f_{j i}}{1-f_{i i}}<\infty
$$

which is the desired result.

Finally, we obtain the criterion for stationary distribution.

Theorem 3.16 (Stationary Distribution Criterion) An irreducible homogeneous Markov chain is positive recurrent if and only if it has a stationary distribution. Moreover, if the stationary distribution $\pi^{T}=\left[\pi_{1}, \pi_{2}, \ldots\right]$ exists, it is unique and $\pi_{i}>0$ for all $i \in S$.

## Proof: $\Rightarrow$ :

By Propositions 3.12 and 3.14, the vector $y$ defined in Proposition 3.12 is an invariant measure with $\sum_{i \in S} y_{i}<$ $\infty$. Thus, the probability vector $\pi=y / \sum_{i \in S} y_{i}$ is the stationary distribution.

The uniqueness follows from Proposition 3.13.
$\Leftarrow$ :
To prove this direction, we use proof by contradiction. Because recurrence is a communication class property (Proposition 3.11) and the Markov chain is irreducible, the fact that a state $i$ is transient implies every state is transient. Let $\pi$ be a stationary distribution and we assume that the Markov chain is transient.
By Lemma $3.15, \lim _{n \rightarrow \infty} p_{j i}^{(n)}=0$ for any state $j \in S$. Since $\pi$ is a stationary distribution, $\pi^{T}=\pi^{T} \mathbf{P}^{n}$.
Using the dominated convergence theorem (we can exchange summation and limit) ${ }^{1}$,

$$
\pi_{i}=\lim _{n \rightarrow \infty} \pi_{i}=\lim _{n \rightarrow \infty} \sum_{j \in S} \pi_{j} p_{j i}^{(n)}=\sum_{j \in S} \pi_{j} \lim _{n \rightarrow \infty} p_{j i}^{(n)}=\sum_{j \in S} \pi_{j} \times 0=0
$$

for every state $i \in S$.
Then we conclude $\sum_{i \in S} \pi_{i}=0 \neq 1$, a contradiction to the definition of stationary distribution. Thus, the Markov chain is recurrent then by Proposition 3.14, the Markov chain is positive recurrent.

[^0]In the above case, we are working on a state space $S$ that may possibly contain infinite number of states. In many realistic scenarios the number of states is finite. Does the finiteness of state number gives us any benefits? The answer is yes - and it gives us a huge benefit.

Theorem 3.17 An irreducible homogeneous Markov chain on a finite state space is positive recurrent. Therefore, it always has a stationary distribution.

## Proof:

We first prove that the chain is recurrent. We proceed by proof by contradiction. Assume that the chain is transient. In the proof of Lemma 3.15, we have shown that if a state $i$ is transient, then

$$
\sum_{n=1}^{\infty} p_{j i}^{(n)}=\mathbb{E}_{j}\left(N_{i}\right)=f_{j i} \sum_{k=0}^{\infty} f_{i i}^{k}=\frac{f_{j i}}{1-f_{i i}}<\infty
$$

Because the number of state space is finite,

$$
\sum_{i \in S} \sum_{n=1}^{\infty} p_{j i}^{(n)}<\infty
$$

However, if we exchange the summations,

$$
\sum_{i \in S} \sum_{n=1}^{\infty} p_{j i}^{(n)}=\sum_{n=1}^{\infty} \sum_{i=1}^{s} p_{j i}^{(n)}=\sum_{n=1}^{\infty} 1=\infty
$$

which is a contradiction. So we conclude that the chain is recurrent.
To see if the chain is positive recurrent, note that Proposition 3.12 shows that there exists an invariant measure $y$. Because the number of state space is finite, $\sum_{i \in S} y_{i}<\infty$ so by Proposition 3.14, the chain is positive recurrent.

Finally, we end this lecture on the relation between the return time and the stationary distribution.

Theorem 3.18 Let $\left\{X_{n}\right\}$ be an irreducible homogeneous positive recurrent Markov chain. Then

$$
\pi_{i}=\frac{1}{\mathbb{E}_{i}\left(T_{i}\right)}
$$

where $\pi=\left(\pi_{1}, \cdots, \pi_{s}\right)$ is the stationary distribution of $\left\{X_{n}\right\}$ and $T_{i}=\inf \left\{n \geq 1: X_{n}=i\right\}$ is the return time to state $i$.

Proof: Define a vector $y$ such that $y_{i}=\mathbb{E}_{0}\left(\sum_{n=1}^{\infty} I\left(X_{n}=i\right) I\left(n \leq T_{0}\right)\right)$. We already know that such a vector describes an invariant measure and $\pi_{i}=\frac{y_{i} n=1}{\sum_{j \in S} y_{j}}$.
Now we consider the case $i=0$. Then $y_{0}=\mathbb{E}_{0}\left(\sum_{n=1}^{\infty} I\left(X_{n}=0\right) I\left(n \leq T_{0}\right)\right)=1$ by property (P1). Moreover, $\sum_{i \in S} y_{i}=\mathbb{E}_{0}\left(T_{0}\right)$ due to property (P2). Thus, $\pi_{0}=\frac{y_{0}}{\sum_{i \in S} y_{i}}=\frac{1}{\mathbb{E}_{0}\left(T_{0}\right)}$.

Because state 0 is just a reference state, we can apply the same argument to any other state. Thus, we conclude that $\pi_{i}=\frac{1}{\mathbb{E}_{i}\left(T_{i}\right)}$ for each $i \in S$.

Here is a short summary about what we have learned so far:

1. Irreducibility + recurrence $\Rightarrow$ There exists an invariant measure that is unique up to a proportionality constant.
2. Irreducibility + positive recurrence $\Leftrightarrow$ Irreducibility + there exists a stationary distribution $\pi$ and it is unique. Moreover, when $\pi$ exists, $\pi_{i}>0$ and $\pi_{i}=1 / \mathbb{E}_{i}\left[T_{i}\right]$.
3. Irreducibility + finite state-space $\Rightarrow$ Irreducibility + positive recurrence.

Here is a summary about the classification of states. Recall $f_{i i}=P_{i}\left(T_{i}<\infty\right)$ is the probability of return to $i$ given we start at $i$ and $\mathbb{E}_{i}\left(T_{i}\right)$ is the expected return time. State $i$ is called:

1. Recurrent if $f_{i i}=1$.
2. Transient if $f_{i i}<1$.
3. Positive Recurrent if $f_{i i}=1$ and $\mathbb{E}_{i}\left(T_{i}\right)<\infty$.
4. Null Recurrent if $f_{i i}=1$ and $\mathbb{E}_{i}\left(T_{i}\right)=\infty$.
5. Periodic with period $d_{i}$ if $p_{i i}^{(n)}=0$ for all $n$ not divisible by $d_{i}$, and $d_{i}(>1)$ is the greatest such integer.
6. Aperiodic if $d_{i}=1$.
7. Ergodic if 3. and 6. apply.
8. Absorbing if $p_{i i}=1$.

[^0]:    ${ }^{1}$ This works because $\pi_{j} p_{j i}^{(n)} \leq \pi_{j}$ for every $n$ and $\sum_{j \in S} \pi_{j}=1$.

