# Lecture notes of Stochastic Process
## lectured by prof. Hsueh-I Lu

pishen

AlgoLab, CSIE, NTU

April 16, 2012

LeoSW, windker

# Stochastic Process

### Definition

A Stochastic process is a set of random variables $\{X(t)|t \in T\}$ where T is a index ( time) set.
State Space: possible value of X(t) for each t, which is defined as subset of R.

## Markov Chain

### Definition

A Stochastic Process $\mathbb{X}$ with state space S is a Markov Chain if
$\exists 0 \le p_{ij} \le 1 \quad \forall i, j \in S$ such that

$$(a) \quad \sum_{j \in S} p_{ij} = 1 \quad \forall i \in S$$

$$(b) \quad P(X(t+1) = j | X(0) = i_0, X(1) = i_1, ..., X(t) = i) = p_{ij}$$
$$\forall t, i_0, i_1, ..., i_{t-1}$$

$\mathbb{P}$ denotes the matrix form of $p_{ij}$ with sum of any row is 1.

Lemma: $P(X(n) = j | X(0) = i) = \mathbb{P}^n[i, j]$

## Proof of lemma

We know statement is true for $(m + n) = 0$. For $(m + n) > 0$:

$$P(X(m + n) = j | X(0) = i) = \sum_{k \in S} P(X(m + n) = j \text{ and } X(m) = k | X(0) = i$$

$$= \sum_{k \in S} P(X(m + n) = j | X(m) = k \text{ and } X(0) = i$$

$$= \sum_{k \in S} P(X(m + n) = j | X(m) = k) \cdot P(X(m) =$$

$$= \sum_{k \in S} P^n[k, j] \cdot P^m[i, k]$$

$$= \sum_{k \in S} P^m[i, k] \cdot P^n[k, j]$$

$$= \mathbb{P}^n[i, j]$$

$=$: conditional on $X(m)$
$=$: definition of conditional probability
$=$: (see next page)

## Proof of lemma(cont)

$$P(X(m+n) = j | X(m) = k \text{ and } X(0) = i) = \sum_{r \in S} P(X(m+n) = j | X(m+n)$$

$$= \sum_{r \in S} P(X(m+n) = j | X(m+n)$$

$$= P(X(m+n) = j | X(m) = k)$$

$=:$ conditional on $X(m+n-1)$

$=:$ first part by definition of Markov chain and second part by inductive hypothesis

# Recurrent & transient

## Definition

The *recurrent probability* of state $i$ of Markov chain $\mathbb{X}$ is

$$f_i = P(\text{there exists an index } t \geq 1 \text{ with } X(t) = i | X(0) = i)$$

- State $i$ of $\mathbb{X}$ is *recurrent* if $f_i = 1$.
- State $i$ of $\mathbb{X}$ is *transient* if $f_i < 1$.

# Recurrent & transient (cont.)

- If state $i$ is recurrent, by the property of Markov chain, once it re-enter the state $i$, we can take it as starting from $X(0)$ again. Hence we know that it will keep re-entering the state $i$ again and again in the process.

- If state $i$ is transient, in each period it start going from $i$, it may have probability $1 - f_i$ that it won't come back anymore. Hence the probability that the process will be in state i for exactly $n$ periods equals $f_i^{n-1}(1 - f_i)$, $n \geq 1$, which is a geometric distribution.

## Recurrent & transient (cont.)

- From the preceding page, it follows that state $i$ is recurrent if and only if, starting in state $i$, the expected number of steps that the process is in state $i$ is infinite.
- We can also derive that, if the Markov chain has finite states, at least one state is recurrent.

## Expected number of visits

Let

$$I(n) = \begin{cases} 1 & \text{if } X(n) = i \\ 0 & \text{if } X(n) \neq i \end{cases}$$

we have $\sum_{n=0}^{\infty} I(n)$ represents the number of steps that the process is in state $i$, and

$$
\begin{aligned}
E\left[\sum_{n=0}^{\infty} I(n) | X(0) = i\right] &= \sum_{n=0}^{\infty} E[I(n)|X(0) = i] \\
&= \sum_{n=0}^{\infty} 1 \cdot P(X(n) = i | X(0) = i) \\
&= \sum_{n=0}^{\infty} P_{ii}^n
\end{aligned}
$$

We set $T = \sum_{n=0}^{\infty} I(n)$

## Lemma 1

From the above statements, we prove the following

### Lemma

*State $i$ is*

$$recurrent \iff \sum_{n=0}^{\infty} P_{ii}^n = \infty,$$

$$transient \iff \sum_{n=0}^{\infty} P_{ii}^n < \infty$$

## Proof of Lemma 1

$(\Rightarrow:)$
$(\Leftarrow:)$
Suppose state i is transient($f_i < 1$), consider $P(T = k) = f_i^{k-1} \cdot (1 - f_i)$.
Since T is a geometric distribution, we have

$$E[T] = \sum_{k=0}^{\infty} k \cdot f_i^{k-1} \cdot f_i$$
$$= \frac{1}{1 - f_i} < \infty$$

# Communicated states

### Definition

State $i$ and $j$ *communicate*, denoted $i \leftrightarrow j$, if there exist integers $m \geq 0$ and $n \geq 0$ such that

$$P_{ij}^m > 0 \text{ and } P_{ji}^n > 0$$

We say a Markov chain $\mathbb{X}$ is irreducible if $i \leftrightarrow j \quad \forall i, j \in S$

## Lemma 2

### Lemma

If $i \leftrightarrow j$, then the following statements hold.

- State $i$ is recurrent if and only if state $j$ is recurrent.
- State $i$ is transient if and only if state $j$ is transient.

Corollary: $\mathbb{X}$ is finite and irreducible $\implies$ all states are recurrent.

- $\mathbb{X}$ is finite $\implies \exists i \in S$ is recurrent (proof later)
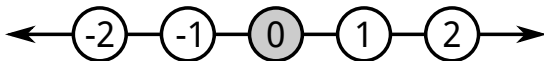- By Lemma 2, all states are recurrent

## Proof

Let $m$ and $n$ be nonnegative integers with $P_{ij}^m, P_{ji}^n > 0$. Suppose that state $j$ is recurrent, i.e., $\sum_{t=0}^{\infty} P_{jj}^t = \infty$. We have

$$\sum_{t=0}^{\infty} P_{ii}^t \geq \sum_{t=0}^{\infty} P_{ii}^{m+t+n}$$

$$\geq \sum_{t=0}^{\infty} P_{ij}^m \cdot P_{jj}^t \cdot P_{ji}^n$$

$$= P_{ij}^m \cdot P_{ji}^n \cdot \sum_{t=0}^{\infty} P_{jj}^t = \infty$$

Thus, state $i$ is also recurrent.

## Infinite drunken man problem



Let the state space consist of all integers. Let $X(0) = 0$ (i.e. at time 0 the drunken man is in state 0). The transition probabilities are such that

$$P_{i,(i+1)} = P_{i,(i-1)} = 0.5$$

holds for all states $i$ of $\mathbb{X}$.

# Gambler's ruin

1. Limiting probabilities
2. Stationary distribution
3. Long-run proportion
4. (Inverse of) Expected return time

# Limiting Probabilities

### Definition

Number $\pi_j$ is the *limiting probability* of $j$ if

$$\pi_j = \lim_{n \to \infty} P_{ij}^n$$

holds for all states $i \in S$ ($S \subseteq \mathbb{N}$ is the state space).

- $\pi_j$ is independent of $i$.

- $\lim_{n \to \infty} P^n = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix}$ , where $\pi = (\pi_1, \pi_2, \dots)$

## Stationary Probability Distribution

### Definition

Non-negative row vector $\pi = (\pi_1, \pi_2, \ldots)$ is a *stationary probability distribution* of $\mathbb{X}$ if $\pi \times P = \pi$ holds and $\sum_{i \in S} \pi_i = 1$

- $\pi$ is a normalized left eigenvector with eigenvalue $= 1$.
- If $X(0)$ has distribution $\pi$, then $X(t)$ has the same distribution $\pi$ for all $t \geq 1$. $\pi$ is also called as *steady-state distribution*.
- It doesn't mean that each $X(t)$ become independent. $\pi$ only means the distribution of $X(t)$ when the previous random variable's value is unknown.

# Theorem 1

### Theorem

Let $\mathbb{X}$ be an irreducible, aperiodic, positive recurrent Markov chain, then

- The limiting probability $\pi_j$ of each state $j$ exists.
- $\pi = (\pi_1, \pi_2, \ldots)$ is the unique stationary probability distribution.

- The proof will be stated at page 35.

# Expected return time

### Definition

The *expected return time* of state $i \in S$ is

$$\mu_i = \sum_{n \geq 1} n \cdot f_i^{(n)}$$

where

$$f_i^{(n)} = P(\min\{t : X(t) = i, t \geq 1\} = n | X(0) = i)$$

- $f_i = \sum_{n \geq 1} f_i^{(n)}$

# Positive recurrent & null recurrent

### Definition

State $i$ is *positive recurrent* if $\mu_i < \infty$

### Definition

State $i$ is *null recurrent* if $\mu_i = \infty$

- Both are recurrent states, and are *class properties*, which means that if state $i$ and $j$ communicate, they will share this property.
- If $\mathbb{X}$ is finite, then each recurrent state of $\mathbb{X}$ is positive recurrent. Proof stated at page 60.

## Example of null recurrent

### Example

For a Markov chain with $n$ states $(1, \ldots, n)$, if

$$P(X(t+1) = i + 1 | X(t) = i) = 1 - 1/n$$

and

$$P(X(t+1) = 1 | X(t) = i) = 1/n$$

According to geometric distribution (taking $p = 1/n$), the expectation value of "steps taken for state 1 to come back" will be $1/p = n$, hence $\lim_{n \to \infty} n = \infty$.

# Period of a chain

## Definition

The *period* of state $i$ is $d$ if $d$ is the largest integer such that

$$P_{ii}^n = 0$$

holds for all $n$ which is not divisible by $d$.

## Definition

If each state of $\mathbb{X}$ has period 1, then $\mathbb{X}$ is called *aperiodic*.

- If $P_{ii} > 0$ for all $i \in S$, then $\mathbb{X}$ is aperiodic.
- Period can be seen as the $\mathrm{gcd}$ of all $n$ that have $P_{ii}^n > 0$, note that $P_{ii}^{\mathrm{gcd}} > 0$ is not necessary.
- The period of drunken man problem is 2.

## Lemma 1

### Lemma

*If state $j$ is aperiodic and positive recurrent, then*

$$\pi_j \equiv \lim_{n \to \infty} P_{ij}^n$$

*exists and is positive for all states $i \in S$.*

- This can be proved by the Blackwell theorem in Renewal theory.
- It doesn't promise that each $\pi_j$ for different $i$ will be the same. But they will be the same if we add the irreducible property.

## Property of $\lim$

- The position of $\lim$ cannot be switched arbitrarily in an equation.

### Example

$$1 = \lim_{n \to \infty} \lim_{m \to \infty} \frac{m}{m+n} \neq \lim_{m \to \infty} \lim_{n \to \infty} \frac{m}{m+n} = 0$$

- $\lim$ would not influence the inequality.

### Example

If $f(n) \geq g(n)$, then $\lim_{n \to \infty} f(n) \geq \lim_{n \to \infty} g(n)$

- $\lim$ is linear operator under finite number of functions.

### Example

For $m < \infty$,

$$\sum_{i=1}^{m} \lim_{n\to\infty} f_i(n) = \lim_{n\to\infty} \sum_{i=1}^{m} f_i(n)$$

need an example of $m = \infty$

# Inequality 1

## Inequality

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

## Proof.

$$\lim_{m \to \infty} \sum_{j=1}^{m} \pi_j = \lim_{m \to \infty} \sum_{j=1}^{m} \lim_{n \to \infty} P_{ij}^n$$

$$= \lim_{m \to \infty} \lim_{n \to \infty} \sum_{j=1}^{m} P_{ij}^n$$

$$\leq \lim_{m \to \infty} \lim_{n \to \infty} \sum_{j \in S} P_{ij}^n = 1$$

$\square$

- The last equation works since $\sum_{j \in S} P_{ij}^n = 1$.

# Inequality 2

### Inequality

For state $j \in S$, we have

$$\pi_j \geq \sum_{i \in S} \pi_i P_{ij}$$

### Proof.

For $m \geq 1$ and $n \geq 1$,

$$P_{ij}^{n+1} = \sum_{k \in S} P_{ik}^n P_{kj} \geq \sum_{k=1}^{m} P_{ik}^n P_{kj}$$

then

$$
\begin{aligned}
\pi_j &= \lim_{n \to \infty} P_{ij}^{n+1} \\
&\geq \lim_{n \to \infty} \sum_{k=1}^{m} P_{ik}^n P_{kj} = \sum_{k=1}^{m} \lim_{n \to \infty} P_{ik}^n P_{kj} = \sum_{k=1}^{m} \pi_k P_{kj}
\end{aligned}
$$

hence, we know

$$\lim_{m \to \infty} \sum_{k=1}^{m} \pi_k P_{kj} = \sum_{k \in S} \pi_k P_{kj} \leq \pi_j$$

$\square$

# Equality 1

## Equality

$$\pi_j = \sum_{i \in S} \pi_i P_{ij}$$

### Proof.

Assume that for some $j \in S$, $\pi_j > \sum_{i \in S} \pi_i P_{ij}$, then

$$\sum_{j \in S} \pi_j > \sum_{j \in S} \sum_{i \in S} \pi_i P_{ij}$$
$$= \sum_{i \in S} \pi_i \sum_{j \in S} P_{ij} = \sum_{i \in S} \pi_i$$

Since a value cannot be greater than itself, we got contradiction. $\square$

- $\sum$ should be represented by $\lim \sum$, the equation will still work.

- **Step 0**: existence of limiting probability.
- **Step 1**: existence of stationary probability distribution.
- **Step 2**: uniqueness.

# 0. Existence of limiting probability

## Proof.

By lemma 1, we know that there exists a $\pi_j$ for row $i$. Since the Markov chain is irreducible and all the states are positive recurrent, for any state $i'$ other than $i$, we know that $i'$ surely will visit $i$ in finite steps. Therefore, the $\pi_j$ value at row $i'$ will equal to the $\pi_j$ value at row $i$, which means that all the $\pi_j$ for column $j$ are the same, and is the limiting probability. $\square$

still not clear enough

# 1. Existence of stationary probability distribution

We want to prove that

## Target

*There's a vector $s = (s_1, s_2, \ldots)$ such that*

1. $\sum_{i \in S} s_i = 1$
2. $s \times P = s$

## Proof.

By lemma 1, we know that there exists a $\pi = (\pi_1, \pi_2, \ldots)$.
And by equality 1, we know that

$$(\pi_1, \pi_2, \ldots) \times P = (\pi_1, \pi_2, \ldots)$$

Hence $\pi$ can satisfy the 2nd part of our target.
Then, we take $k = \sum_{i \in S} \pi_i$. By inequality 1, we know that $k < \infty$, and can get

$$(\frac{\pi_1}{k}, \frac{\pi_2}{k}, \ldots) \times P = (\frac{\pi_1}{k}, \frac{\pi_2}{k}, \ldots)$$

where $\sum_{i \in S} \frac{\pi_i}{k} = 1$ also satisfy the 1st part of our target.
Therefore, this vector can be $s$, which means that it exists. $\qquad\square$

# 2. Uniqueness

## Target

*If $s = (s_1, s_2, \ldots)$ is a stationary distribution of $\mathbb{X}$, then $s = \pi$.*

- We'll prove this by inequality 3 & 4.

# Inequality 3

## Inequality

$$s_j \geq \pi_j, \forall j \in S$$

## Proof.

Let the distribution of $X(0)$ be $s$, by the property of stationary distribution, we have

$$s_j = P(X(n) = j) = \sum_{i \in S} P(X(n) = j | X(0) = i) P(X(0) = i)$$

$$= \sum_{i \in S} P_{ij}^n \cdot s_i$$

$$\geq \sum_{i=1}^{m} P_{ij}^n \cdot s_i$$

$$\Rightarrow s_j = \lim_{m \to \infty} \lim_{n \to \infty} s_j$$

$$\geq \lim_{m \to \infty} \lim_{n \to \infty} \sum_{i=1}^{m} P_{ij}^n \cdot s_i = \lim_{m \to \infty} \sum_{i=1}^{m} \pi_j \cdot s_i = \pi_j$$

$\square$

# Inequality 4

## Inequality

$$s_j \leq \pi_j, \forall j \in S$$

### Proof.

Similar in the proof above, $\forall m, n \geq 1$, we have

$$
\begin{aligned}
s_j &= \sum_{i \in S} P_{ij}^n \cdot s_i \\
&\leq \sum_{i=1}^{m} P_{ij}^n \cdot s_i + \sum_{i=m+1}^{\infty} s_i \\
\Rightarrow s_j &= \lim_{m \to \infty} \lim_{n \to \infty} s_j \\
&\leq \lim_{m \to \infty} \lim_{n \to \infty} \left( \sum_{i=1}^{m} P_{ij}^n \cdot s_i + \sum_{i=m+1}^{\infty} s_i \right) \\
&= \pi_j
\end{aligned}
$$

$\square$

## Example

$$P = \begin{pmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{pmatrix}, 0 < \alpha, \beta < 1$$

$$\pi = \left( \frac{\beta}{1 + \beta - \alpha}, \frac{1 - \alpha}{1 + \beta - \alpha} \right)$$

# Real world example: Hardy-Weinberg Law

## Example

There're two kinds of allele:

- dominant: **A**
- recessive: **a**

And three kinds of senotype with population proportion as follow:

- AA: $p$
- aa: $q$
- Aa: $r = 1 - (p + q)$

### Example (cont.)

$$
\begin{array}{c}
\phantom{P =}\quad AA \qquad\quad aa \qquad\quad Aa \\
P = \begin{array}{c} AA \\ aa \\ Aa \end{array}
\begin{pmatrix}
p + \frac{r}{2} & 0 & q + \frac{r}{2} \\
0 & q + \frac{r}{2} & p + \frac{r}{2} \\
\frac{p}{2} + \frac{r}{4} & \frac{p}{2} + \frac{r}{4} & \frac{p+q+r}{2}
\end{pmatrix}
\end{array}
$$

we get $\pi = (p, q, r)$ when

- $p = \left(p + \frac{r}{2}\right)^2$
- $q = \left(q + \frac{r}{2}\right)^2$
- $r = 2\left(p + \frac{r}{2}\right)\left(q + \frac{r}{2}\right)$

### Definition

We say that $r_j$ is the *long-run proportion* of state $j \in S$ if

$$r_j = \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} P_{ij}^t$$

holds for each state $i \in S$.

- It represents the average appearance times of state $j$ in the whole process.
- We will show that (in theorem 3) if $\mathbb{X}$ is irreducible, then the long-run proportion of all states exist.

# Theorem 2

### Theorem (type 1)

*If $r_j$ exists for each $j \in S$ and $\sum_{j \in S} r_j > 0$, then $r = (r_1, r_2, \ldots)$ is the unique stationary distribution of $\mathbb{X}$.*

or

### Theorem (type 2)

*If $r_j$ exists for each $j \in S$ and **a stationary distribution exists**, then $r = (r_1, r_2, \ldots)$ is the unique stationary distribution of $\mathbb{X}$.*

# Proof

**Existence of stationary distribution in type 1:**
Let

$$R = \begin{pmatrix} r \\ r \\ \vdots \end{pmatrix} = \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} P^t$$

then

$$\begin{aligned} R \times P &= \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} P^{t+1} \\ &= \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} P^t + \lim_{n \to \infty} \frac{1}{n}(P^{n+1} - P) \\ &= R \end{aligned}$$

As stated later, $\sum_{j \in S} r_j \le 1$, hence by normalizing $r$, we prove that stationary distribution exist.

- $(\lim f(n)) \cdot g(n) = \lim f(n) \cdot g(n)$?

## Proof (cont.)

**Uniqueness:**

Let $\pi$ be an arbitrary stationary distribution, then

$$
\begin{aligned}
r &= \pi \times R \\
&= \pi \times \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} P^t \\
&= \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} \pi \times P^t \\
&= \lim_{n \to \infty} \frac{1}{n} \sum_{1 \le t \le n} \pi \\
&= \pi
\end{aligned}
$$

## Proof (cont.)

**Prove that $\sum_{j \in S} r_j \leq 1$:**

$$
\begin{aligned}
\sum_{j \in S} r_j &= \lim_{m \to \infty} \sum_{j=1}^{m} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} P_{ij}^t \\
&= \lim_{m \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{m} P_{ij}^t \\
&\leq \lim_{m \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{j \in S} P_{ij}^t \\
&= \lim_{m \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} 1 = 1
\end{aligned}
$$

## Example 1

On a highway, if we know the probability that

- A truck is followed by a truck: $1/4$
- A truck is followed by a car: $3/4$
- A car is followed by a truck: $1/5$
- A car is followed by a car: $4/5$

We can construct a matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} T & C \end{array} \\
\begin{array}{c} T \\ C \end{array} & \begin{pmatrix} 1/4 & 3/4 \\ 1/5 & 4/5 \end{pmatrix}
\end{array}
$$

and get the portion of trucks and cars on the whole highway as the eigenvector $(4/19, 15/19)$ (we will know that long-run proportion exists by Theorem 3).

## Example 2

For a system which has several good and bad states, we have a matrix $P$:

$$
\begin{array}{c}
\phantom{g_1} \\
g_1 \\
g_2 \\
\vdots \\
b_1 \\
b_2 \\
\vdots
\end{array}
\begin{array}{cccccc}
g_1 & g_2 & \cdots & b_1 & b_2 & \cdots
\end{array}
\left(
\begin{array}{cccccc}
\phantom{x} & \phantom{x} & \phantom{x} & \phantom{x} & \phantom{x} & \phantom{x} \\
\\
\\
\\
\\
\\
\end{array}
\right)
$$

Example 2 (cont.)

**Q1:** Breakdown rate (breakdown times / total time)
The long-run frequency of going to a bad state from a good state is

$$\sum_{i \in g} \sum_{j \in b} r_i P_{ij}$$

Example 2 (cont.)

**Q2:** The expected time $\mu_G$ (resp. $\mu_B$) of staying in good (resp. bad) states once we reach a good (resp. bad) state?

**Ans:**

For each $t = 1, 2, \ldots$, let $G_t$ (resp. $B_t$) be the length of the $t$-th good (resp. bad) phase of consecutive good (resp. bad) states. By the strong law of large numbers,

$$P\left(\lim_{t \to \infty} \frac{G_1 + B_1 + G_2 + B_2 + \cdots + G_t + B_t}{t} = \mu_G + \mu_B\right) = 1$$

Since the reciprocal of above is the breakdown rate, we get equation (1):

$$P\left(\sum_{i \in G}\sum_{j \in B} \pi_i P_{ij} = \frac{1}{\mu_G + \mu_B}\right) = 1$$

Example 2 (cont.)

Also, with probability 1, we get equation (2):

$$P\left(\sum_{i\in G} r_i = \lim_{t\to\infty} \frac{G_1 + G_2 + \cdots + G_t}{G_1 + B_1 + \cdots + G_t + B_t} = \frac{\mu_G}{\mu_G + \mu_B}\right) = 1$$

Then, by $(2)/(1)$, we get that

$$P\left(\mu_G = \frac{\sum_{i\in G} r_i}{\sum_{i\in G}\sum_{j\in B} r_i P_{ij}}\right) = 1$$

- $\lim \frac{f(n)}{g(n)} = \frac{\lim f(n)}{\lim g(n)}$?

# Theorem 3

## Theorem

*If $\mathbb{X}$ is irreducible, then the long-run proportion $r_i$ exists with probability 1, moreover,*

1. *If state $i$ is positive recurrent (i.e. $0 < \mu_i < \infty$), then $P(r_i = \frac{1}{\mu_i}) = 1$.*
2. *If state $i$ is null recurrent (i.e. $\mu_i = \infty$) or transient, then $P(r_i = 0) = 1$.*

**Part 1:**

Suppose $X(0) = i$, $T_k$ is the number of steps required for the $k$-th $i$ goes to $(k+1)$-st $i$, then by the strong law of large number,

$$P\left(\lim_{k\to\infty} \frac{T_1 + T_2 + \cdots + T_k}{k} = \mu_i\right) = 1$$

$$\Rightarrow P\left(r_i = \lim_{k\to\infty} \frac{k}{T_1 + T_2 + \cdots + T_k} = \frac{1}{\mu_i}\right) = 1$$

- $\lim(A/B) = \frac{1}{\lim(B/A)}$?

## Proof (cont.)

**Part 2:**

1. If $i$ is transient, $i$ will only appear finite times in the long-run, hence

$$r_i = \frac{finite}{\infty} = 0$$

2. If $i$ is null recurrent, $\mu_i$ is $\infty$, then

$$P\left(\lim_{k \to \infty} \frac{T_1 + T_2 + \cdots + T_k}{k} = \infty\right) = 1$$

$$P\left(r_i = \lim_{k \to \infty} \frac{k}{T_1 + T_2 + \cdots + T_k} = 0\right) = 1$$

(The first equation is not promised by the strong law of large number. But if it's not $\infty$, we can say that $\mu_i$ is not $\infty$, which is a contradiction.)

# Example 1

## Example (type 1)

If $\mathbb{X}$ is **irreducible** and finite, then $\mathbb{X}$ has no null recurrent states.

## Example (type 2)

If $\mathbb{X}$ is finite, then $\mathbb{X}$ has no null recurrent states.

- Finite irreducible imply positive recurrent.

## Proof

- **Type 1:**
  If there's a state which is null recurrent, by irreducible, all the states
  will be null recurrent. Then, all states have $P(r_i = 0) = 1$. By
  changing the proof in page 51 into finite states version, we know that
  $\sum r_i = 1$. So it's impossible for finite $r_i$, which are all close to 0, to
  sum up to 1.

- **Type 2:**
  If it's not irreducible, the finite set of communicated null recurrent
  states still form an irreducible and finite Markov chain, which can fit
  the requirement of type 1.

# Example 2

## Example

In the drunken man problem with infinite states, no state will be positive recurrent.

- Infinite drunken man imply no positive recurrent. Note that it doesn't mean all infinite irreducible Markov chain has no positive recurrent state.

## Proof

If all the states are positive recurrent, then by theorem 3, we know that all the $r_i > 0$ and is a finite value. Since each state of drunken man problem has the same structure, all the $r_i$ has same value. We then set $r = \epsilon \cdot \min(r_1, r_2, \ldots)$ $(0 < \epsilon < 1)$ such that $r_i > r > 0, \forall i$. And get

$$\sum_{i \in S} r_i > \sum_{i \in S} r = \infty > 1$$

which is contradiction to page 51.

### Example

There's a hotel, with $N$ representing the number of newly occupied rooms each day ($N$ is a poisson distribution with parameter $\lambda$). And the number of consecutive check-in days of each room is a geometric distribution with probability $p$ ($p$ is the probability of check-out). $X(t)$ is the number of occupied rooms in day $t$.

## Q1: $P_{ij} = ?$

We set $R_i$ as a binomial distribution with parameter $(i, 1-p)$, which represents the number of rooms which will remain occupied in the next day, then

$$
\begin{aligned}
P_{ij} &= P(R_i + N = j) \\
&= \sum_{k \geq 0} P(R_i + N = j | R_i = k) P(R_i = k) \\
&= \sum_{k \geq 0} P(N = j - k) P(R_i = k) \\
&= \sum_{0 \leq k \leq \min(i,j)} \frac{e^{-\lambda} \cdot \lambda^{j-k}}{(j-k)!} \binom{i}{k} (1-p)^k p^{1-k}
\end{aligned}
$$

We guess (by a dream?) there's a stationary distribution which is a poisson distribution with parameter $\lambda_0$. Setting $X(0)$ with this distribution. And let $R$ as the number of rooms in $X(0)$ which remain check-in in the next day ($R$ is a poisson distribution with parameter $\lambda_0(1-p)$). $X(1)$ will have distribution $R + N$, which is a poisson distribution with parameter $\lambda_0(1-p) + \lambda$. Then since $X(0)$ is a stationary distribution, it will have the same distribution with $X(1)$, which means that $\lambda_0 = \lambda_0(1-p) + \lambda$, and we get $\lambda_0 = \lambda/p$.

After getting $r_i$, we get that with probability 1,

$$\mu_i = \frac{1}{P(X(0) = i)} = \frac{i!}{e^{-\lambda/p} \cdot (\lambda/p)^i}$$

<span style="color:red">need clean up</span>

# Corollary of theorem 2 & 3

## Corollary

If $\mathbb{X}$ is irreducible, then

$\mathbb{X}$ is positive recurrent $\iff$ $\mathbb{X}$ admits a stationary distribution.

## Moving to transient states

For transient states $i$ and $j$, we define the following:

**1** Expected steps in a transient state:

### Definition

$E$ is a matrix where $E_{ij}$ is the expected number of steps with $X(t) = j$ when $X(0) = i$.

**2** Probability of reaching a transient state:

### Definition

$F$ is a matrix where

$$F_{ij} = P(X(t) = j \text{ for some } t \geq 1 | X(0) = i)$$

# Computing $E$ & $F$

### Theorem

*For a Markov chain $\mathbb{X}$ consisting finite transient states,*

$$E = (I - T)^{-1}$$

*where $I$ is an identity matrix, $T$ is the induced matrix of $P$ by all the transient states in $P$. Moreover,*

$$F_{ij} = \frac{E_{ij} - \delta_{ij}}{E_{jj}} \text{ ,where } \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

## Proof

Conditioned on $X(1)$, we have

$$E_{ij} = \underbrace{\delta_{ij}}_{\text{step}=0} + \underbrace{\sum_k P_{ik} \cdot E_{kj}}_{\text{step} \geq 1} = \sum_k T_{ik} \cdot E_{kj}$$

The 2nd equation works since the process will not go back to transient state once it enter a recurrent state. Then, we have

$$I \times E = E = I + T \times E$$
$$\implies (I - T) \times E = I$$
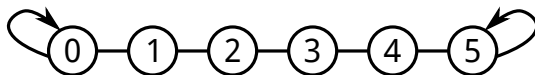$$\implies E = (I - T)^{-1}$$

## Proof (cont.)

Conditioned on whether or not $X(t) = j$ holds for some $t \geq 1$, we have

$$E_{ij} = \underbrace{\delta_{ij}}_{\text{step}=0} + \underbrace{F_{ij} \cdot E_{jj}}_{\text{steps} \geq \text{ the first } j}$$

therefore,

$$F_{ij} = \frac{E_{ij} - \delta_{ij}}{E_{jj}}$$

## Example: Gambler's ruin



$$T = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \begin{pmatrix} 0 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 \end{pmatrix} \end{array} \quad E = \begin{pmatrix} 1.6 & 1.2 & 0.8 & 0.4 \\ 1.2 & 2.4 & 1.6 & 0.8 \\ 0.8 & 1.6 & 2.4 & 1.2 \\ 0.4 & 0.8 & 1.2 & 1.6 \end{pmatrix}$$

$$F = \begin{pmatrix} 0.375 & 0.5 & 1/3 & 0.25 \\ 0.75 & 1.75/3 & 1.9/3 & 0.5 \\ 0.5 & 1.9/3 & 1.75/3 & 0.75 \\ 0.25 & 1/3 & 0.5 & 0.375 \end{pmatrix}$$

## Branching process

In the beginning, there're $X(0)$ life forms, each life form has probability $p_i$ of becoming $i$ life forms in the next step.

- state 0 is recurrent (absorbing).
- if $p_0 > 0$, all other states $(1, 2, \ldots)$ are transient since
  $P(X(t+1) = 0 | X(t) = i) = p_0{}^i > 0$

We'll show that

$$E[X(n)] = \mu^n \cdot X(0)$$

where

$$\mu = \sum_{j \geq 1} j \cdot p_j = E[Z_k]$$

and $Z_k$ is the number of offspring of the $k$-th life form, all $Z_k$ are i.i.d.

## Proof

$$
\begin{aligned}
E[X(n)] &= E[E[X(n)|X(n-1)]] \\
&= E\left[ E\left[ \sum_{k=1}^{X(n-1)} Z_k | X(n-1) \right] \right] \\
&= E[X(n-1) \cdot \mu] \\
&= \mu \cdot E[X(n-1)] \\
&= \mu^n \cdot X(0)
\end{aligned}
$$

# Probability of extinction

## Definition

$e_i$ is the probability of extinction when $X(0) = i$.

**Case 1:** $\mu < 1$

$$
\begin{aligned}
1 - e_i &= \lim_{n \to \infty} P(X(n) \geq 1 | X(0) = i) \\
&= \lim_{n \to \infty} \sum_{j \geq 1} P(X(n) = j | X(0) = i) \\
&\leq \lim_{n \to \infty} \sum_{j \geq 1} j \cdot P(X(n) = j | X(0) = i) \\
&= \lim_{n \to \infty} E[X(n) | X(0) = i] \\
&= \lim_{n \to \infty} \mu^n \cdot i = 0
\end{aligned}
$$

**Case 2:** $\mu \geq 1$

$$e_2 = e_1{}^2, \quad e_3 = e_2 \cdot e_1, \quad \ldots$$

$$\begin{aligned}
e_1 &= P(\text{extinct}|X(0) = 1) \\
&= \sum_{j \geq 0} P(\text{extinct}|X(1) = j) \cdot P_{1j} \\
&= \sum_{j \geq 0} e_j \cdot p_j \\
&= \sum_{j \geq 0} e_1{}^j \cdot p_j
\end{aligned}$$

We then solve the above equation to get $e_1$.

## Example

$$p_0 = p_1 = 0.25, \quad p_2 = 0.5$$
$$\implies \mu = 1 \cdot 0.25 + 2 \cdot 0.5 > 1$$
$$\implies e_1 = e_1{}^0 \cdot 0.25 + e_1{}^1 \cdot 0.25 + e_1{}^2 \cdot 0.5$$
$$\implies e_1 = \{1/2, \ 1\}$$

Since $\mu > 1$, we know $\lim_{n \to \infty} E[X(n)] = \infty$.
But if $e_1 = 1$, we have $\lim_{n \to \infty} P(X(n) = 0) = 1$, which would not make $\lim_{n \to \infty} E[X(n)] = \infty$, hence $e_1 \neq 1$.

# Reversed Markov chain

## Definition

Let $\mathbb{X}$ (resp. $\mathbb{Y}$) be a Markov chain with matrix $P$ (resp. $Q$).
We say that $\mathbb{Y}$ is the *reversed chain* of $\mathbb{X}$ if there exists a stationary distribution $\pi$ of $\mathbb{X}$ such that

$$\pi_i \cdot Q_{ij} = \pi_j \cdot P_{ji}$$

holds for all states $i, \ j \in S$.

# Observation 1

### Observation

*The reversed sequence $\mathbb{Y}$ of $\mathbb{X}$ is a Markov chain.*

## Proof

$$P(Y(n) = i_0 | Y(n-1) = i_1, Y(n-2) = i_2, \ldots, Y(n-k) = i_k)$$
$$= P(X(n) = i_0 | X(n+1) = i_1, X(n+2) = i_2, \ldots, X(n+k) = i_k)$$
$$= \frac{P(X(n) = i_0, X(n+1) = i_1, \ldots, X(n+k) = i_k)}{P(X(n+1) = i_1, \ldots, X(n+k) = i_k)}$$
$$= \frac{P(X(n) = i_0) \cdot P(X(n+1) = i_1 | X(n) = i_0) \cdot P_{i_1 i_2} \cdots P_{i_{k-1} i_k}}{P(X(n+1) = i_1) \cdot P_{i_1 i_2} \cdots P_{i_{k-1} i_k}}$$
$$= \frac{P(X(n) = i_0, X(n+1) = i_1)}{P(X(n+1) = i_1)}$$
$$= P(X(n) = i_0 | X(n+1) = i_1)$$
$$= P(Y(n) = i_0 | Y(n-1) = i_1)$$

## Observation 2

### Observation

If $\mathbb{Y}$ is the reversed sequence of Markov chain $\mathbb{X}$ and $\pi$ is a stationary distribution of $\mathbb{X}$, then

$$\pi_i \cdot Q_{ij} = \pi_j \cdot P_{ji}$$

holds for all $i, j \in S$, where $Q$ is the transition matrix of $\mathbb{Y}$.

## Proof

Let $\mathbb{X}$ and $\mathbb{Y}$ have distribution $\pi$

$$\begin{aligned}
\pi_i \cdot Q_{ij} &= P(Y(n-1) = i) \cdot P(Y(n) = j \,|\, Y(n-1) = i) \\
&= P(Y(n-1) = i,\, Y(n) = j) \\
&= P(Y(n-1) = i \,|\, Y(n) = j) \cdot P(Y(n) = j) \\
&= P(X(n+1) = i \,|\, X(n) = j) \cdot P(X(n) = j) = \pi_j \cdot P_{ji}
\end{aligned}$$

# Observation 3

## Observation

*Let $P$ (resp. $Q$) be the transition matrix of $\mathbb{X}$ (resp. $\mathbb{Y}$), if vector $\pi$ satisfy the following*

- $\sum_{i \in S} \pi_i = 1$
- $\pi_i \geq 0 \quad \forall i \in S$
- $\pi_i \cdot Q_{ij} = \pi_j \cdot P_{ji} \quad \forall i, j \in S$

*then $\mathbb{Y}$ is the reversed sequence of $\mathbb{X}$.*

- The long-run proportion of $i \to j$ in the sequence of $\mathbb{Y}$ is equal to the long-run proportion of $j \to i$ in the sequence of $\mathbb{X}$.
- Reversed Markov chain is the reversed sequence.

## Proof

From the third property, we have

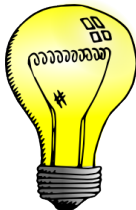$$\sum_{j \in S} \pi_i \cdot Q_{ij} = \pi_i = \sum_{j \in S} \pi_j \cdot P_{ji} \quad \forall i \in S$$

From the 2nd equation, we know that $\pi \times P = \pi$, hence $\pi$ is a stationary distribution of $\mathbb{X}$.

Then by observation 2, we know that for any $\pi$, there's a reversed sequence $\mathbb{Y}'$, whose transition matrix $Q'$ satisfy

$$\pi_i \cdot Q'_{ij} = \pi_j \cdot P_{ji} \quad \forall i, j \in S$$

hence $\mathbb{Y} = \mathbb{Y}'$, which is a reversed sequence of $\mathbb{X}$.

There's a room which need to be lighted by one bulb, when the bulb in use fails, it will be replaced by a new one on next day.

- $X(n) = i$ if the bulb in use on day $n$ is in its $i$th day of use.
- $L$ is a random variable representing the lifetime of a bulb.

We want to know the stationary probability $\pi_i$ of state $i$.

$\mathbb{X}$ is a irreducible, positive recurrent, aperiodic Markov chain which has the sequence like this:

$$1, 2, 3, 1, 2, 3, 4, 5, 1, 1, 2, 1, 2, 3, 4, \ldots$$

We know that

$$P_{i1} = P(\text{buld, on its } i\text{th day of use, fails}) = \frac{P(L = i)}{P(L \geq i)} = 1 - P_{i(i+1)}$$

And the expected return time of state $1$ is $E[L]$, which means that the long-run proportion of state $1$ is $1/E[L]$ by page 57.

## Example: Bulb's life (cont.)

Take $\mathbb{Y}$ (with matrix $Q$) as the reversed chain of $\mathbb{X}$, we know that for all $i \in S$,

- $Q_{(i+1)i} = 1$
- $Q_{1i} = P(L = i)$
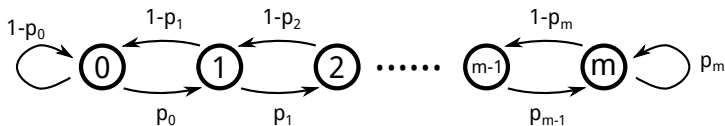- $\pi_1 \cdot Q_{1i} = \pi_i \cdot P_{i1}$

Hence,

$$\pi_i = \frac{\pi_1 \cdot Q_{1i}}{P_{i1}} = \frac{P(L = i) \cdot P(L \geq i)}{E[L] \cdot P(L = i)} = \frac{P(L \geq i)}{E[L]}$$

# Time-reversible

## Definition

$\mathbb{X}$ is *time-reversible* if $\mathbb{X}$ is the reversed chain of $\mathbb{X}$.

## Example: Reversed drunken man



- $0 < p_0 \leq 1$
- $0 \leq p_m < 1$
- $0 < p_i < 1 \quad \forall i = 1, \ldots, m-1$

The long-run proportion of transition $i \to i+1$ and $i+1 \to i$ are the same, since one must go back to $i$ from $i+1$ in order to go to $i+1$ from $i$. Hence the drunken man problem is time-reversible.

## Example: Reversed drunken man (cont.)

$$\pi_0 \cdot p_0 = \pi_1 \cdot (1 - p_1)$$
$$\pi_1 \cdot p_1 = \pi_2 \cdot (1 - p_2)$$
$$\vdots$$
$$\pi_{m-1} \cdot p_{m-1} = \pi_m \cdot (1 - p_m)$$

Thus,

$$\pi_1 = \pi_0 \cdot p_0 / (1 - p_1)$$
$$\pi_2 = \pi_1 \cdot p_1 / (1 - p_2)$$
$$\vdots$$
$$\pi_m = \pi_{m-1} \cdot p_{m-1} / (1 - p_m)$$

## Example: Reversed drunken man (cont.)

$$\pi_i = \underbrace{\frac{\prod_{j=0}^{i-1} p_j}{\prod_{j=1}^{i}(1 - p_j)}}_{q_i} \cdot \pi_0 \quad \forall i = 1, \ldots m$$

$$\Longrightarrow \pi_0 + \sum_{i=1}^{m} \pi_i = 1 = \pi_0 + \sum_{i=1}^{m} q_i \cdot \pi_0$$

$$\Longrightarrow \pi_0 = \frac{1}{1 + \sum_{i=1}^{m} q_i}$$

$$\Longrightarrow \pi_k = \frac{q_k}{1 + \sum_{i=1}^{m} q_i} \quad \forall k = 0, 1, \ldots m$$

## Example: Two bukkits of balls

There're two bukkits contain total $m$ balls.
In each step, we randomly choose one ball and put it in another bukkit.
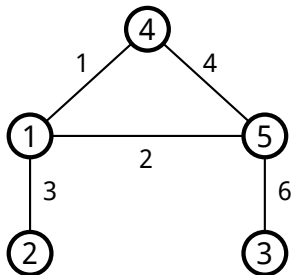Let $X(n)$ represent the number of balls in the first bukkit, it's the Markov chain of previous example with

$$p_0 = 1, \ p_m = 0, \ p_i = \frac{m-i}{m} \quad \forall i = 1, \ldots, m-1$$

We can get that

$$q_i = \frac{\prod_{j=0}^{i-1} \frac{m-j}{m}}{\prod_{j=1}^{i} \frac{j}{m}} = \frac{\prod_{j=0}^{i-1} m-j}{\prod_{j=1}^{i} j} = \binom{m}{i} \quad \forall i = 1, \ldots m$$

$$\implies \pi_0 = \frac{1}{1 + \sum_{i=1}^{m} \binom{m}{i}} = \frac{1}{2^m} \implies \pi_k = \frac{\binom{m}{k}}{2^m} \quad \forall k = 0, 1, \ldots m$$

# Example: A random walk



$$P_{ij} = \frac{w(i,j)}{\sum_k w(i,k)}$$

where $w(a, b)$ is the weight of edge $(a, b)$.
To make it as a time-reversible chain, we let

$$\pi_i = \frac{\sum_k w(i,k)}{\sum_\ell \sum_k w(\ell, k)}$$

We can see that

$$\pi_i \cdot P_{ij} = \pi_j \cdot P_{ji}$$

## Hastings-Metropolis sampling algorithm

Design an irreducible Markov chain $\mathbb{X}$ such that the unique stationary
distribution of $\mathbb{X}$ is the distribution of random variable $Y$.
Since the long-run proportion of state $i$ is $P(Y = i)$,

$$\lim_{n \to \infty} \frac{X(1) + X(2) + \ldots + X(n)}{n} = \sum_{i \in S} i \cdot P(Y = i) = E[Y] = \mu$$

While computing $\mu$ by the law of large number is difficult (hard to sample
on $Y$), we use this alternative method to compute $\mu$ by generating a
sequence of $\mathbb{X}$, which is sometime easier.

## Hastings-Metropolis sampling algorithm (cont.)

There's a random variable $Y$ such that

$$P(Y = i) = \frac{b_i}{C}$$

for some unknown (or intractable) $C = \sum_{i \in S} b_i$.
We then design a Markov chain $\mathbb{X}$ that

- $P_{ii} = Q_{ii} + \sum_{k \in S, k \neq i} Q_{ik} \cdot (1 - q_{ik})$
- $P_{ij} = Q_{ij} \cdot q_{ij} \quad \forall j \neq i$

where

- $Q$ is the transition matrix of an arbitrary irreducible Markov chain which has the same state space as $\mathbb{X}$.
- $q$ is a matrix to be determined later.

# Hastings-Metropolis sampling algorithm (cont.)

For $n = 0, 1, \ldots,$

1. If $X(n) = i$, set $Z$ such that $P(Z = j) = Q_{ij} \quad \forall j \in S$.
2. If $Z = j$, set $X(n+1)$ such that
   - $P(X(n+1) = j) = q_{ij}$
   - $P(X(n+1) = i) = 1 - q_{ij}$

One can see that this satisfies the requirement on previous page.

## Hastings-Metropolis sampling algorithm (cont.)

Then, we let

$$q_{ij} = \min\left(\frac{b_j \cdot Q_{ji}}{b_i \cdot Q_{ij}}, 1\right)$$

$$\implies b_i \cdot Q_{ij} \cdot q_{ij} = b_j \cdot Q_{ji} \cdot q_{ji}$$

$$\implies \frac{b_i}{C} \cdot P_{ij} = \frac{b_j}{C} \cdot P_{ji}$$

By observation 3 on page 83, we know that $(b_1/C, b_2/C, \ldots)$ is the stationary distribution of $\mathbb{X}$.

## Example: Space of permutations

### Example

Let $S$ consist of all the permutations $(x_1, x_2, \ldots, x_n)$ of $\{1, 2, \ldots, n\}$ that

$$\sum_{i=k}^{n} k \cdot x_k \geq \frac{n^3}{4}$$

- This is same as $Y$ in page 95 with $C = |S|$ and $b_i = 1 \ \forall i$.
- $S$ is hard to compute.
- We need to design a matrix $Q$ such that when given a permutation $x$, it's efficient to compute the value of $Q_{xy} \ \forall y \in S$.

## Example: Space of permutations (cont.)

We let

$$Q_{xy} = \frac{1}{N(x)} \quad \text{, if } y \text{ can be obtained from } x \text{ by one swap}$$

where $N(x)$ is the number of permutations that can be obtained from $x$ by one swap. For example:

$$\underbrace{(1,2,3,4,5)}_{y} \leftrightarrow \underbrace{(1,3,2,4,5)}_{x} \leftrightarrow \underbrace{(1,3,4,2,5)}_{y}$$

This chain is irreducible since each $x \in S$ can go to $(x_1, x_2, \ldots, x_n)$, where $x_1 \leq x_2 \leq \ldots \leq x_n$, by several swaps.
Also, given a $x$, finding all the obtainable $y$ can be done efficiently.