# Probability and Stochastic Processes Course Notes

Chia-Hao Chang, chchangkh@utexas.edu

Fall 2018

## 1 Discrete Time Markov Chain

A stochastic process $\{X_t : t \in \mathcal{T}\}$ with state space $E$ is said to be a *discrete time* stochastic process if $\mathcal{T}$ is countable . In particular, $\mathcal{T} = \mathbb{N} \cup \{0\}$. i.e. $\{X_n : n \geq 0, n \in \mathbb{Z}\}$ is a discrete time stochastic process. In the following context, we assume the sate space $E$ is countable.

**Definition 1** (Discrete Time Markov Chain (DTMC))**.** A stochastic process $\{X_n : n \geq 0\}$ with sate space $E$ is said to be be a **discrete time Markov Chain** (DTMC) if for all integers $n \geq 0$ and all states $i_0, i_1, \cdots, i_{n-1}, i, j \in E$, we have

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \tag{1}$$

If, further, the right hand side does not depend on $n$, we say the Markov Chain is **homogeneous**.

Eqn.(1) is called the Markov property of a stochastic process.

**Definition 2** (Transition Probability Matrix (TPM))**.** Given a HMC, the associated **transition probability matrix** $\mathbf{P}$ is a (general) matrix with $[\mathbf{P}]_{ij} = p_{ij} = P(X_{n+1} = j | X_n = i)$.

Note that $\mathbf{P}$ might have infinite dimension when $|E| = \infty$. Note that $\mathbf{P}$ must satisfy

$$\sum_{j \in E} p_{ij} = 1, p_{ij} \geq 0 \tag{2}$$

for all states $i, j$. That is each row of $\mathbf{P}$ sums to 1. A matrix satisfying Eqn.(2) is said to be a **stochastic matrix**.

**Remark 1.** *Let $\{X_n : n \geq 0\}$ be a Markov chain, then*

$$\begin{aligned} &P(X_{n+1} = j_1, X_{n+2} = j_2, \cdots, X_{n+k} = j_k | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) \\ &= P(X_{n+1} = j_1, X_{n+2} = j_2, \cdots, X_{n+k} = j_k | X_n = i) \end{aligned}$$

*In particular, if $\{X_n : n \geq 0\}$ is a HMC, then*

$$P(X_{n+1} = j_1, X_{n+2} = j_2, \cdots, X_{n+k} = j_k | X_n = i) = p_{ij_1} p_{j_1 j_2} \cdots p_{j_{k-1} j_k}$$

Based on the remark, we have the following theorem

**Theorem 1** (Distribution of HMC)**.** *The distribution of a HMC at time $n$ is completely determined by the initial distribution $\pi^{(0)}$ and the transition probability matrix $\mathbf{P}$.*

**Theorem 2** (Chapman-Kolomogorov Theorem)**.** *Given any two states $i, j \in E$, the $n+m$ step transition probability $p_{ij}^{(n+m)}$ is given by*

$$p_{ij}^{(n+m)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}^{(m)}$$

In particular, the $n$ step transition probability matrix $\mathbf{P}^{(n)}$ is given by $\mathbf{P}^{(n)} = \mathbf{P}^n$ i.e. the $n$th power of transition probability matrix $\mathbf{P}$. Note using the Bayes Rule, we can in fact write the $n$ step transition probability as

$$p_{ij}^{(n)} = \sum_{i_1, i_2, \cdots, i_{n-1} \in E} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} j}$$

We close this subsection by noting the following property

$$\begin{aligned}
&P(X_{n+1} = j_1, X_{n+2} = j_2, \cdots, X_{n+k} = j_k, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0 | X_n = i) \\
&= P(X_{n+1} = j_1, X_{n+2} = j_2, \cdots, X_{n+k} = j_k | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) \\
&\qquad \times P(X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0 | X_n = i) \\
&= P(X_{n+1} = j_1, X_{n+2} = j_2, \cdots, X_{n+k} = j_k | X_n = i) P(X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0 | X_n = i)
\end{aligned}$$

In words: the future and the past at time $n$ are *conditionally independent* given the present state $X_n = i$.

## 1.1 Determine Markov Chain

**Theorem 3** (HMC Driven by White Noise)**.** *Let $\{Z_n : n \geq 1\}$ be an sequence of i.i.d. random variables with range $F$, $E$ a countable state space and $f : E \times F \to E$ be some function. Let $X_0$ be a random variable independent of $\{Z_n : n \geq 1\}$. The recurrence equation $X_{n+1} = f(X_n, Z_{n+1})$ defines an HMC.*

**Theorem 4** (Generalization of Theorem 3)**.** *Let things be the same as Theorem 3 except for the statistics of $X_0, Z_1, Z_2, \cdots$. Suppose $Z_{n+1}$ is conditionally independent of $Z_1, Z_2, \cdots, Z_n, X_0, X_1, X_2, \cdots, X_{n-1}$ given $X_n$. i.e.*

$$\begin{aligned}
&P(Z_{n+1} = k, Z_1 = i_1, \cdots, Z_n = i_n, X_0 = j_0, \cdots, X_{n-1} = j_{n-1} | X_n = i) \\
&= P(Z_{n+1} = k | X_n = i) P(Z_1 = i_1, \cdots, Z_n = i_n, X_0 = j_0, \cdots, X_{n-1} = j_{n-1} | X_n = i)
\end{aligned}$$

*or equivalently,*

$$\begin{aligned}
&P(Z_{n+1} = k | Z_1 = i_1, \cdots, Z_n = i_n, X_0 = j_0, \cdots, X_{n-1} = j_{n-1}, X_n = i) \\
&= P(Z_{n+1} = k | X_n = i)
\end{aligned}$$

*Then $\{X_n : n \geq 0\}$ defines an HMC. With $\mathbf{P}$ given by*

$$p_{ij} = P(f(i, Z_{n+1}) = j | X_n = i)$$

## 1.2 Topology of Markov Chain

**Definition 3** (Communication)**.** A state $j \in E$ is said to be **accessible** from state $i$ if there exists $n \geq 0$ s.t. $p_{ij}^{(n)} > 0$. If $j$ is accessible from $i$, we write $i \to j$. In particular, we have $p_{ii}^{(0)} = 1$ and $i \to i$ always holds. Two states $i$ and $j$ are said to **communicate** if $i \to j$ and $j \to i$ and is denoted by $i \leftrightarrow j$.

Note that $i \to j$ if and only if there exists at least one $m \geq 0$ and a path $(i, i_1, i_2, \cdots, i_{m-1}, j)$ s.t.

$$p_{ii_1} p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{m-2} i_{m-1}} p_{i_{m-1} j} > 0$$

**Remark 2** (Equivalence Relation)**.** *"$\leftrightarrow$" defines an equivalence relation on states $E$. Hence, $E/\leftrightarrow$ defines an equivalence class.*

**Definition 4** (Communication Class)**.** With Remark 2, the state space $E$ can be partitioned into $L$ equivalent classes, which are termed as **communication classes** of $E$.

**Definition 5** (Period)**.** The **period** $d_{ij}$ between states $i, j$ is defined as

$$d_{ij} = \gcd\{n : p_{ij}^{(n)} > 0\}$$

In particular, the **period** of a state $i \in E$ is defined as

$$d_i = \gcd\{n : p_{ii}^{(n)} > 0\}$$

**Theorem 5** (Period is Class Property)**.** *If $i \leftrightarrow j$, then $d_i = d_j$. That is, all the states in the same communication class have the same period.*

Further, if a the state space $E$ is finite with $|E| = m$, we only need to check up to the $m$th power of $\mathbf{P}^m$. i.e.

**Proposition 1.** *For any two states $i$ and $j$, $i \to j$ if and only if*

$$[I_m + \mathbf{P}^{(1)} + \mathbf{P}^{(3)} + \mathbf{P}^{(2)} + \cdots \mathbf{P}^{(m)}]_{ij} > 0$$

*i.e. at least one the $ij$-entry of $\mathbf{P}^n, 0 \leq n \leq m$, is nonzero.*

To see this, note that if there exist a $m$ step path $(i, i_1, i_2, \cdots, i_m, j)$ from $i$ to $j$. Then there exists at least two states, say $i_\alpha$ and $i_\beta$ s.t. $i_\alpha = i_\beta$. Consider the the states right before and after them. i.e. $(i_{\alpha-1}, i_\alpha, i_{\alpha+1})$ and $(i_{\beta-1}, i_\beta, i_{\beta+1})$. We have a nonzero probability from $i_{\alpha-1}$ to $i_\alpha$ and $i_\beta$ to $i_{\beta+1}$, which means there exists a reduced chain $(i, i_1, i_2, \cdots, i_{\alpha-1}, i_\alpha, i_{\beta+1}, \cdots, i_m, j)$ which has at most $m$ steps of transition. If the two $i_\alpha$ and $i_\beta$ are consecutive, then we can naively remove one of them since $p_{ii}^{(1)} = 1$. This completes the proof.

**Definition 6** (Closed Set)**.** A subset $S_0$ of $E$ is said to be **closed** if for all states $i \in S_0$, $\sum_{j \in S_0} p_{ij} = 1$. That is, any state in $S_0$ can go to states in $S_0$ only.

**Definition 7** (Absorbing State)**.** A state $i$ is said to be an **absorbing state** if $\{i\}$ is a closed set. Equivalently, a state $i$ is **absorbing** if $p_{ii} = 1$.

It is clear that if there is an absorbing state $i \in E$, then the Markov chain can not be irreducible. Unless the state space is trivial, i.e., $E = \{i\}$.

## 1.3 Regeneration

**Definition 8** (Filtration)**.** A **filtration** of $\sigma$-algebra $\mathcal{F}$ is a sequence of sub-$\sigma$-algebras $\{\mathcal{G}_n : n \geq 0\}$ s.t. for all $n \geq 0$, $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$.

**Definition 9** (Stopping Times)**.** A **stopping time** $\tau$ with respect to a filtration $\{\mathcal{G}_n : n \geq 0\}$ is a random variable taking values in $\mathbb{N} \cup \{0\}$ s.t. for all $m \geq 0$, $\{\tau = m\} \in \mathcal{G}_m$.

In particular, if a **stopping time** $\tau$ with respect to a random process $\{X_n : n \geq 0\}$ is a random variable $\tau$ s.t. $\{\tau = m\} \in \sigma\{X_0, X_1, \cdots, X_m\}$, where $\sigma\{X_0, X_1, \cdots, X_m\}$ is the $\sigma$-algebra generated by the random variables $\{X_0, X_1, \cdots, X_m\}$. In other words, for any stopping time $\tau$, the event $\{\tau = m\} = g_m(X_0, X_1, \cdots, X_m)$. i.e. $\tau = m$ is expressible in terms of the random variables before time $m$, namely $X_0, X_1, X_2, \cdots, X_m$.

**Example 1** (Return Time)**.** *The* **return time** $T_i$ *of a state $i$ is the first time that the HMC visits state $i$ after $X_0$. i.e.*

$$T_i := \inf\{n \geq 1 : X_n = i\}$$

**Example 2** (Visit Time/hitting time)**.** *The* **visit time***(or* **hitting time***)* $S_i$ *is defined as the first time that the HMC visits state $i$, including $X_0$ i.e.*

$$S_i := \inf\{n \geq 0 : X_n = i\}$$

*More generally speaking, let $A \in E$ be a subset of states in $E$. The* **hitting time** *of $A$ is defined as*

$$S_A := \inf\{n \geq 0 : X_n \in A\}$$

The above definition of $S_A$ is particularly useful in the calculation of the Gambler's Ruin problem.

**Example 3** (Successive Return Times)**.** *Let $\tau_1 = T_i$, $\tau_2$, $\tau_3$, $\cdots$ be the* **successive return times** *to state $i$. i.e. $\tau_k$ is the time index of the $k$th visit to state $i$. Then, $\{\tau_k : k \geq 1\}$ are stopping times.*

In particular, we have

**Proposition 2.** *Let $\tau_k$ be the $k$th successive return time of state $i$ of the HMC $\{X_n : n \geq 0\}$, then*

$$\{\tau_k = m\} = \left\{ \sum_{j=1}^{m} \mathbb{1}\{X_j = i\} = k, X_m = i \right\} \tag{3}$$

## 1.4 Strong Markov Property and Regenerative Cycle

**Theorem 6** (Strong Markov Property)**.** *Let $\{X_n : n \geq 0\}$ be a HMC with countable state space $E$ and transition probability matrix $\mathbf{P}$. Let $\tau$ be a stopping time with respect to the HMC. Then for any state $i \in E$, the following hold:*

*(a) The process before $\tau$ and the process after $\tau$ are independent.*

*(b) The process after $\tau$ is an HMC with transition matrix $\mathbf{P}$.*

Let $N_i$ be the number of visit to state $i$ strictly after time 0, that is,

$$N_i := \#\text{ of visits to } i = \sum_{n=1}^{\infty} \mathbb{1}\{X_n = i\}$$

In fact, the distribution of $N_i$ can be found in closed form.

**Theorem 7** (Visit to a state)**.** *The distribution of $N_i$ given $X_0 = j$ is given by*

$$P_j(N_i = r) = P(N_i = r|X_0 = j) = \begin{cases} f_{ji}f_{ii}^{r-1}(1 - f_{ii}), & r \geq 1 \\ 1 - f_{ji}, & r = 0 \end{cases}$$

*where $f_{ji}$ is given by*

$$f_{ji} = P_j(T_i < \infty) = \sum_{n=1}^{\infty} P_j(T_i = n)$$

The next theorem is one of the most important theorem in the application of strong Markov Property

**Theorem 8** (Regenerative Cycle Theorem)**.** *Let $\{X_n : n \geq 0\}$ be an HMC with initial state 0 which is almost surely visited infinitely often. Denoting by $\tau_0 = 0, \tau_1 = T_0, \tau_2, \tau_3, \cdots$, the successive return times to state 0, then the trajectories*

$$\{X_{\tau_k}, X_{\tau_k+1}, \cdots, X_{\tau_{k+1}-1}\}, \quad k \geq 1 \tag{4}$$

*are independently and identically distributed (i.i.d.)*

The trajectories are *i.i.d.* in the sense that the "length" of the vector, i.e. $\tau_{k+1} - \tau_k$ are *i.i.d.* so are the values of the entries $X_{\tau_k}, X_{\tau_k+1}, \cdots, X_{\tau_{k+1}-1}$.

### 1.5 Recurrence and Transience

Before we begin our discussion, let's first define some notations commonly used in this section. For any state $i \in E$,

$$\begin{aligned} f_{ii} &= P(\{X_n : n \geq 1\} \text{ ever reenter } i|X_0 = i) \\ &= P(\bigcup_{n=1}^{\infty} \{X_n = i\}|X_0 = i) \end{aligned}$$

Intuitively, $f_i$ is if we start at $i$, the probability of visiting state $i$ *at least once* in the entire future.

**Definition 10** (Recurrence and Transience)**.** A state $i \in E$ is said to be **recurrent** if $f_{ii} = 1$, and **transient** if $f_{ii} < 1$.

One quantity of particular interest is the *expected number of visits* to a state $i$, given that we are currently at state $i$. Let us now start computing this quantity. The first way utilizes the property of *regenerative cycles* supplement here. After each visit to state $i$ at some time $n$, the probability of visiting state $i$ in the future is $f_i$. And between each visit, the processes are *i.i.d.*. Hence,

$$P(\text{exactly } n \text{ visits to } i|X_0 = i) = \underbrace{f_i f_i \cdots f_i}_{n \text{ times}}(1 - f_i) = f_i^n(1 - f_i), \quad n = 0, 1, 2, \cdots$$

This is a *geometric random variable* with parameter $f_i$. The expected value is thus given by

$$\mathbb{E}[\text{number of visits to } i|X_0 = i] = \sum_{n=0}^{\infty} n \cdot P(\text{exactly } n \text{ visits to } i|X_0 = i) = \frac{f_i}{1 - f_i} \tag{5}$$

Let's take a deeper look at Eqn.(5). When $i$ is recurrent, the expected number of visit is $\infty$, which means state $i$ will be visited **infinitely often**. Another way to compute the expectation is by the indicator trick

$$N_i := \# \text{ of visits to } i = \sum_{n=1}^{\infty} \mathbb{1}\{X_n = i\}$$

Hence, the expected number of visits to $i$ is simply

$$\mathbb{E}[\sum_{n=1}^{\infty} \mathbb{1}\{X_n = i\}|X_0 = i] = \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{1}\{X_n = i\}|X_0 = i] = \sum_{n=1}^{\infty} P(X_n = i|X_0 = i) = \sum_{n=1}^{\infty} P_{ii}^{(n)}$$

Thus, we have the following theorem

**Theorem 9** (Characterization of recurrence via $n$ step transition probabilities). *A state $i$ is recurrent if and only if $\sum_{i=1}^{\infty} P_{ii}^{(n)} = \infty$.*

In fact, this is a more general result of *potential matrix*

**Definition 11** (Potential Matrix). The **potential matrix G** associated with the transition probability matrix **P** is given by

$$\mathbf{G} = \sum_{n=0}^{\infty} \mathbf{P}^n$$

In particular, note that the expected number of visit to state $j$, given that we are at state $i$ is

$$\mathbb{E}[N_j|X_0 = i] = \mathbb{E}[\sum_{n=0}^{\infty} \mathbb{1}\{X_n = j\}|X_0 = i] = \sum_{n=0}^{\infty} p_{ij}^n = [\mathbf{G}]_{ij}$$

Starting from $i$, we will meet $j$ infinitely often if and only if $\sum_{n=0}^{\infty} p_{ij}^n$ diverges.

**Theorem 10** (Recurrence is a class property). *If $i \leftrightarrow j$, then $i$ and $j$ are either both recurrent or both transient.*

**Definition 12** (Invariant Measure). A nontrivial vector $x$ is said to be an invariant measure of the stochastic matrix $\mathbf{P}$ if for all $i \in E$,

$$x_i = \sum_{j \in E} x_j p_{ji}$$

and

$$0 \leq x_i < \infty$$

**Theorem 11** (Canonical Invariant Measure). *Let $\mathbf{P}$ be the transition matrix of the irreducible, recurrent HMC $\{X_n : n \geq 0\}$. Let $0$ be an arbitrary state and let $T_0$ be the return time of state $0$. Define for all $i \in E$,*

$$x_i = \mathbb{E}_0 \left[ \sum_{n \geq 1} \mathbb{1}\{X_n = i\} \mathbb{1}\{n \leq T_0\} \right] = \mathbb{E}_0 \left[ \sum_{n=1}^{T_0} \mathbb{1}\{X_n = i\} \right]$$

*Then, for all $i \in E$*

$$x_i \in (0, \infty)$$

*and $x = (x_i : i \in E)$ is an invariant measure of $\mathbf{P}$. $x$ is called the* **canonical invariant measure** *of $\mathbf{P}$.*

**Theorem 12** (Uniqueness of Invariant Measure). *The invariant measure of an irreducible recurrent stochastic matrix is unique up to a scaling factor.*

**Theorem 13** (Positive and Null Recurrent Criterion). *An irreducible recurrent Markov chain is positive recurrent* **if and only if** *its invariant measure $x$ is summable. i.e.*

$$\sum_{i \in E} x_i < \infty$$

*i.e. an irreducible recurrent Markov Chain is positive recurrent if and only if an stationary distribution $\pi$ exists.*

**Remark 3** (Invariant Measure Does not imply Recurrent). *An irreducible Markov Chain may possess an invariant measure $x$ but still being transient. For example, the 1-D random walk with $p \neq 1/2$.*

**Theorem 14** (Fluid's Equation). *Let $\{X_n : n \geq 0\}$ be an HMC with state space $E$ (either finite or infinite) and a transition probability matrix $\mathbf{P} = [p_{ij}]$. $(S, \bar{S})$ is a cut on $E$ and $x$ is an invariant measure (assumed to exist), then*

$$\sum_{j \in \bar{S}} \sum_{i \in S} x_i p_{ij} = \sum_{i \in S} \sum_{j \in \bar{S}} x_j p_{ji} \tag{6}$$

*In particular, when a stationary distribution exists, Eqn.(6) reads*

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

We provide another criterion to determine whether a Markov chain is positive recurrent or null recurrent.

**Theorem 15** (Foster's Theorem). *Let $\{X_n \coloneqq n \geq 0\}$ be an irreducible, aperiodic Markov chain. Then it is positive recurrent if and only if there exists a function $V(\cdot) \coloneqq E \to \mathbb{R}$ s.t.*

*(a)* $V(i) \geq 0, \quad \forall i \in E$.

*(b)* $\sum_{j \in E} p_{ij} V(j) = \mathbb{E}[V(X_{t+1}) | X_t = i] \leq V(i) - 1$

*(c)* $\sum_{j \in E} p_{0j} V(j) = \mathbb{E}[V(X_{t+1}) | X_t = 0] \leq \infty$

## 1.6 First-Step Analysis

Of particular interest is the exact value of $\mathbb{E}_i[T_i]$, which is the mean return time to state $i$, given the initial state is $X_0 = i$ since it determines whether the Markov chain at hand is positive recurrent or not. More generally speaking, when we would like to find the values of $\mathbb{E}[\tau | X_0 = i]$, where $\tau$ is a stopping time of the Markov chain, instead of finding the exact probability distribution of $\tau = \ell$ for all possible $\ell$'s, we usually apply the trick called **first step analysis**. Before we start, let's look at the following lemma, which is a even more general result than the first step analysis.

**Lemma 1** (Partition on Expectation Values). *Let $(\Omega, \mathcal{F}, P)$ be the underlying probability triplet and $\{C_j : j \in J\}$ be a partition of $\Omega$. Given any function $f(X) = f(X_0, X_1, \cdots)$ and $A \in \mathcal{F}$ with $P(A) \neq 0$, we have*

$$\mathbb{E}[f(X) | A] = \sum_{j \in J} \mathbb{E}[f X | A, C_j] P(C_j | A)$$

One might think of Lemma 1 as a variation of the *law of total probability*. Some commonly used partitions include $\{X_n = m : m \in E\}$ or $\{\tau = \ell : \ell \geq 1\}$, where $\tau$ is a stopping time.

**Proposition 3** (First Step Analysis). *Let $\tau \coloneqq \inf\{n \geq 1\colon X_n = 0\}$ is the return time of state 0 and $\mathbf{P} = [p_{ij}]$ be the transition probability matrix of $\{X_n : n \geq 0\}$. Then, we have*

$$\mathbb{E}[\tau|X_0 = i] = 1 + \sum_{j\neq 0} \mathbb{E}[\tau|X_0 = j]p_{ij} \tag{7}$$

Depending on the definition of the return time $\tau$, the form of Eqn.(7) may be different.

## 1.7 Ergodic Theorem

In this section, we are going to explore the relation between the *probabilistic average* and the *temporal average*(or ensemble average).

**Proposition 4.** *Let $\{X_n\colon n \geq 0\}$ be an irreducible, recurrent HMC and let $x$ be the canonical invariant measure associated with state $0 \in E$, i.e.,*

$$x_i = \mathbb{E}\left[\sum_{n\geq 1} \mathbb{1}\{X_n = i\}\mathbb{1}\{n \leq T_0\}\right]$$

*where $T_0 = \inf\{n \geq 1 : X_n = 0\}$ is the return time to 0. Define $\nu(n) \coloneqq \sum_{k=1}^{n} \mathbb{1}\{X_k = 0\}$. Let $f : E \to \mathbb{R}$ s.t.*

$$\sum_{i\in E} |f(i)|x_i < \infty$$

*Then for any initial distribution $\mu$, $P_\mu - a.s.$,*

$$\lim_{N\uparrow\infty} \frac{1}{\nu(N)} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)x_i$$

**Theorem 16** (Ergodic Theorem). *Let $\{X_n\colon\ n \geq 0\}$ be an irreducible, positive recurrent Markov chain with stationary distribution $\pi = (\pi_i\colon i \in E)$. Let $f\colon E \to \mathbb{R}$ be s.t. $\sum_{i\in E} |f(i)|\pi_i < \infty$. Then for any initial distribution $\mu$, $P_\mu - a.s.$,*

$$\lim_{N\uparrow\infty} \frac{1}{N} \sum_{k=1}^{N} f(X_k) = \sum_{i\in E} f(i)\pi_i$$

Note that on the left hand side, it is the *empirical/temporal* average. That is, for each realization $\omega \in \Omega$, the empirical mean *almost surely* converges to the *statistical* average, which is a *deterministic* value.