15.1 Ergodic Theory

Ergodic theorems concern the limiting behavior of averages over time. We will show that asymptotically,

\[
\text{“time average } \approx \text{ space average”}
\]

Let \((X_t)_{t \geq 0} \sim \text{Markov}(\lambda, P)\) with countable space state \(I\). We will show that

\[
\frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \to \sum_{x \in I} f(x) \cdot \pi_x = \pi^\top f.
\]

**Theorem 15.1.** Define

\[
V_i(n) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = i\}}
\]

where \(i \in I\) and \(n \geq 1\). Assume that \(P\) is irreducible. Then

\[
\frac{V_i(n)}{n} \to \frac{1}{m_i} = \frac{1}{E_i(R_i)} \quad \text{as } n \to \infty
\]

almost surely.

**Corollary 15.2.** If \(P\) is transient or null recurrent, \(1/m_i = 0\).

**Theorem 15.3** (SLLN). If \(Y_0, Y_1, Y_2, \ldots\) are i.i.d. non-negative random variables with mean \(\mu \in (0, \infty]\), then

\[
\frac{1}{n} \sum_{k=0}^{n-1} Y_k \xrightarrow{a.s.} \pi \to \infty \mu.
\]

**Theorem 15.4.** Let \(P\) be irreducible, positive recurrent with invariant distribution \(\pi\). Take a
bounded function \( f \) on \( I \). Then

\[
\frac{1}{n} \sum_{i=0}^{n-1} f(X_i) \xrightarrow{\text{a.s.}} \lim_{n \to \infty} \sum_{x \in I} f(x) \cdot \pi_x.
\]

**Proof of Theorem 15.4.** First, we note that, \( \sum_{t=0}^{n-1} f(X_t) = \sum_{x \in I} f(x)V_x(n) \). We define,

\[
\text{Error}_n = \frac{1}{n} \sum_{t=0}^{n-1} f(X_t) - \sum_{x \in I} f(x)\pi_x = \sum_{x \in I} f(x) \left( \frac{1}{n} V_x(n) - \pi_x \right).
\]

Take \( \varepsilon > 0 \). Since \( \sum_{x \in I} \pi_x = 1 \), we can choose a finite subset \( F \subseteq I \), such that \( \sum_{x \notin F} \pi_x < \varepsilon \).

Then, using \( \sum_{x} V_x(n)/n = 1 \), we have

\[
|\text{Error}_n| \leq \sum_{x \in F} f(x) \left| \frac{1}{n} V_x(n) - \pi_x \right| + \sum_{x \notin F} f(x) \left| \frac{1}{n} V_x(n) - \pi_x \right| 
\]

\[
\leq M \sum_{x \in F} \left| \frac{1}{n} V_x(n) - \pi_x \right| + M \sum_{x \notin F} \left( \frac{1}{n} V_x(n) + \pi_x \right) 
\]

\[
\leq M \sum_{x \in F} \left| \frac{1}{n} V_x(n) - \pi_x \right| + M \sum_{x \notin F} \left( \frac{1}{n} V_x(n) - \pi_x \right) + 2M\varepsilon
\]

\[
\leq M \sum_{x \in F} \left| \frac{1}{n} V_x(n) - \pi_x \right| - M \sum_{x \notin F} \left( \frac{1}{n} V_x(n) - \pi_x \right) + 2M\varepsilon
\]

\[
\leq 2M \sum_{x \in F} \left| \frac{1}{n} V_x(n) - \pi_x \right| + 2M\varepsilon
\]

where \( M := \max_{x \in I} |f(x)| < \infty \). Choose \( N \), such that

\[
\sum_{x \in F} \left| \frac{1}{n} V_x(n) - \pi_x \right| < \varepsilon, \text{ for all } n \geq N.
\]

Then \( |\text{Error}_n| \leq 4M\varepsilon \), for all \( n \geq N \). Since \( \varepsilon \) is arbitrary, \( \text{Error}_n \to 0 \) almost surely as \( n \to \infty \). \( \blacksquare \)

**Proof of Theorem 15.1.** Assume that \( P \) is irreducible and recurrent. Let \( R_1^{(i)}, R_2^{(i)}, \ldots \) be the return times to \( i \). If the MC is transient, the proof is trivial. So assume that the MC is recurrent.

![Diagram](image)

Define,

\[
G_1^{(i)} = R_2^{(i)} - R_1^{(i)}
\]

\[
G_j^{(i)} = R_{j+1}^{(i)} - R_j^{(i)}, \forall j \geq 1.
\]
We know \((G_j^{(i)})_{j \geq 1}\) are i.i.d. non-negative random variables with mean \(m_i\), and
\[
\frac{G_1^{(i)} + G_2^{(i)} + \cdots + G_{V_i(n)-1}^{(i)}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{R_1^{(i)} + G_1^{(i)} + G_2^{(i)} + \cdots + G_{V_i(n)}^{(i)}}{V_i(n)}.
\]
For large \(n\)
\[
\frac{V_i(n) - 1}{V_i(n)} \cdot \frac{G_1^{(i)} + G_2^{(i)} + \cdots + G_{V_i(n)-1}^{(i)}}{V_i(n) - 1} \leq \frac{n}{V_i(n)} \leq \frac{R_1^{(i)} + G_1^{(i)} + G_2^{(i)} + \cdots + G_{V_i(n)}^{(i)}}{V_i(n)}.
\]
We know \(V_i(n) \to \infty\) as \(n \to \infty\). By SLLN,
\[
\frac{1}{n} \sum_{k=1}^{n} G_k^{(i)} \xrightarrow{a.s.}{n \to \infty} m_i.
\]
Thus, LHS and RHS \(\xrightarrow{a.s.} m_i\) as \(n \to \infty\). In particular,
\[
\frac{n}{V_i(n)} \xrightarrow{a.s.}{n \to \infty} m_i\quad\text{and}\quad \frac{V_i(n)}{n} \xrightarrow{a.s.}{n \to \infty} \frac{1}{m_i}.
\]

We proved that,

\[ \text{"time average } \approx \text{ space average" \& "any positive recurrent, irreducible MC is ergodic"} \]

### 15.2 Metropolis-Hasting Algorithm

Suppose that \(I\) is finite and \(\pi\) is a distribution on \(I\). We want to generate a sample from \(\pi\) and for some function \(f\), approximate
\[
\pi^f = \sum_{x \in I} \pi_x f(x).
\]
Given \(I\) finite and distribution \(\pi\) on \(I\), suppose we know a transition matrix \(P\) on \(I \times I\), which is easy to sample from. Moreover, suppose that \(P\) is irreducible and aperiodic. Under \(P\), the Markov chain moves \(x \to y\) with probability \(p_{xy}\).

Now, we choose \(a_{xy} \in (0, 1]\) and modify the Markov chain as follows, starting at \(x\)

- accept \(x \to y\) with probability \(a_{xy}\) for \(y \in I\),
- otherwise, stay at \(x\).

The new Markov chain has transition matrix \(\hat{P} = ((\hat{p}_{xy}))\) given by
\[
\hat{p}_{xy} = \begin{cases} a_{xy}p_{xy} & \text{if } y \neq x \\ 1 - \sum_{z \neq x} a_{xz}p_{xz} & \text{if } y = x. \end{cases}
\]
Suppose we want to have Detailed Balance for \((\pi, \hat{P})\), i.e.,
\[
\pi_x \hat{p}_{xy} = \pi_y \hat{p}_{yx} \text{ or } \pi_x a_{xy} p_{xy} = \pi_y a_{yx} p_{yx} \text{ for all } x \neq y.
\]
Note that \(\pi_x a_{xy} p_{xy} = \pi_y a_{yx} p_{yx}\) implies that the common value must be smaller than
\[
\leq \pi_x P_{xy} \land \pi_y P_{yx}.
\]
Moreover, equality holds if
\[
a_{xy} = \frac{1}{\pi_x P_{xy}} \cdot (\pi_x P_{xy} \land \pi_y P_{yx}) = 1 \land \frac{\pi_y P_{yx}}{\pi_x P_{xy}}, \text{ for all } y \neq x.
\]
Thus, \(\pi\) is an invariant distribution for the transition matrix
\[
\hat{p}_{xy} = \begin{cases} 
(1 \land \frac{\pi_y P_{yx}}{\pi_x P_{xy}}) \cdot p_{xy} & \text{if } y \neq x \\
1 - \sum_{z \neq x} (1 \land \frac{\pi_z P_{xz}}{\pi_x P_{xy}}) \cdot p_{xz} & \text{if } y = x.
\end{cases}
\]
If \(P\) is symmetric, i.e., \(p_{xy} = p_{yx}\) for all \(x, y\), then \(a_{xy} = 1 \land (\pi_y / \pi_x)\), for all \(y \neq x\).
Now, suppose that \(\pi_x \propto w_x\) for some function \(w\), i.e., \(\pi_x = \frac{w_x}{\sum_{y \in I} w_y}\), then \(\pi_y / \pi_x = w_y / w_x\) and we do not need to compute the complicated term \(\sum_{y \in I} w_y\).

**Example 15.5.** Let \(G = (V, E)\) be a connected finite graph, which may not be regular. Our goal is to generate a sample uniformly at random from \(V\), i.e., \(\pi_x = 1 / |V|\). It is easy to run the random walk on \(G\) with transition matrix \(P\) given by \(p_{xy} = 1/d_x\) for \(x \sim y\) and invariant distribution \(\frac{d_x}{|E|}, x \in V\). We take
\[
a_{xy} = 1 \land (d_x / d_y)
\]
to get the new Markov chain with a transition matrix
\[
\hat{p}_{xy} = \begin{cases} 
\frac{d_x}{d_y} \land \frac{1}{d_y} & \text{if } y \sim x \\
1 - \sum_{z \sim x} (\frac{d_x}{d_z} \land \frac{1}{d_z}) & \text{if } y = x.
\end{cases}
\]
with invariant distribution \(\pi\).

**Example 15.6.** Suppose we want to estimate the value \(\max_{x \in [0,1]} f(x)\). First we discretize \([0,1]\) by \(I\) and try to find the set \(A^* = \{x \mid f(x) = \max_{y \in I} f(y)\}\). For a large number \(\lambda \gg 1\), take \(\pi_x \propto \lambda^{f(x)}\) for \(x \in I\). Then, as \(\lambda \to \infty\), we have
\[
\pi_x \to \frac{1}{|A^*|} \cdot 1_{\{x \in A^*\}}.
\]
In Simulated Annealing, we start with a simple random walk on \(I\) with transition matrix \(P\) and for \(\lambda \gg 1\), use Metropolis-Hasting algorithm to generate a sample \(X_n\) from \(\pi_x \propto \lambda^{f(x)}\). Take \(\lambda \to \infty\), in an appropriate way so that \(X_n \in A^*\) as \(n \to \infty\).