Basic Doeblin Theorem

Let \( P = ((p_{ij}))_{1 \leq i, j \leq N} \) be a transition probability matrix with \( \{1, \ldots, N\} \) as its state space, and use \( \|\rho\|_1 \) to denote \( \sum_{i=1}^{N} |\rho_i| \) for \( \rho = (\rho_1, \ldots, \rho_N) \).

**Lemma 1.** If \( p_{ij} \geq \frac{\epsilon}{N} \) for all \((i, j)\) and some \( \epsilon > 0 \), then \( \|\rho P^n\|_1 \leq (1-\epsilon)^n \|\rho\|_1 \) for all \( n \in \mathbb{N} \) and \( \rho \) satisfying \( \sum_{i=1}^{N} \rho_i = 0 \).

**Proof.** First note that \( \sum_{j=1}^{N}(\rho P)_j = \sum_{i=1}^{N} \rho_i \). Thus, it suffices to handle \( n = 1 \). Second, if \( \sum_{i=1}^{N} \rho_i = 0 \), then

\[
\|\mu P\|_1 = \sum_{j=1}^{N} \left| \sum_{i=1}^{N} \rho_i p_{ij} \right| = \sum_{j=1}^{N} \sum_{i=1}^{N} \rho_i (p_{ij} - \frac{\epsilon}{N}) \\
\leq \sum_{i=1}^{N} \sum_{j=1}^{N} |\rho_i| (p_{ij} - \frac{\epsilon}{N}) = (1-\epsilon)\|\rho\|_1.
\]

**Theorem (A) (Doeblin’s).** Under the conditions in Lemma 1, there exists exactly one probability vector \( \mu \) such that \( \mu P = \mu \). Moreover, for any probability vector \( \nu \) and all \( n \geq 1 \), \( \|\nu P^n - \mu\|_1 \leq 2(1-\epsilon)^n \).

**Proof.** Let \( \nu \) be a probability vector, and take \( \nu^{(n)} = \nu P^n \). Given \( n \geq 1 \), set \( \rho^{(n)} = \nu^{(n)} - \nu \), and observe that \( \sum_{i=1}^{N} \rho^{(n)}_i = 0 \) and \( \|\rho^{(n)}\|_1 \leq 2 \). Hence, since \( \nu^{(m+n)} - \nu^{(m)} = \rho^{(n)} P^m \), the lemma says that \( \|\nu^{(m+n)} - \nu^{(m)}\|_1 \leq 2(1-\epsilon)^m \) for all \( m, n \in \mathbb{N} \). Thus, by Cauchy’s convergence criterion, there exists a \( \mu \) to which \( \{\nu^{(n)}\}_n \) converges. Furthermore, because each \( \nu^{(n)} \) is a probability vector, so is \( \mu \). In addition,

\[
\mu P = \lim_{n \to \infty} \nu^{(n)} P = \lim_{n \to \infty} \nu P^{n+1} = \lim_{n \to \infty} \nu^{(n+1)} = \mu.
\]

That is, \( \mu P = \mu \). Finally, for any probability vector \( \nu \), \( \nu P^n - \mu = \nu P^n - \mu P^n = \rho P^n \), where \( \rho = \nu - \mu \), and so \( \|\nu P^n - \mu\|_1 \leq 2(1-\epsilon)^n \). In particular, if \( \nu P = \nu \), then \( \|\nu - \mu\|_1 = \|\nu P^n - \mu\|_1 \leq 2(1-\epsilon)^n \) for all \( n \geq 1 \), and so \( \|\nu - \mu\|_1 = 0 \).

**Corollary.** Suppose that, for some \( M \geq 1 \) and \( \epsilon > 0 \), \( (P^M)_{ij} \geq \frac{\epsilon}{N} \) for all \((i, j)\). Then again there is a unique probability vector \( \mu \) satisfying \( \mu P = \mu \). Moreover, for any probability vector \( \nu \), \( \|\nu P^n - \mu\|_1 \leq 2(1-\epsilon)\frac{n}{N} \).

**Proof.** By applying the preceding to \( P^M \), we know that there is a unique probability vector \( \mu \) such that \( \mu P^M = \mu \). Hence, since \( \mu P \) is a probability vector and \( (\mu P)P^M = (\mu P^M)P = \mu P \), it follows that \( \mu = \mu P \). In addition, again by the preceding, \( \|\nu (P^M) - \mu\|_1 \leq 2(1-\epsilon)^m \) for all probability vectors \( \nu \) and all \( m \geq 0 \). Hence, if \( n = mM + r \), where \( 0 \leq r < M \), then \( \|\nu P^n - \mu\|_1 = \|(\nu P^n)P^r\| - \|\mu\|_1 \leq 2(1-\epsilon)^m \).

To handle general irreducible \( P \)'s, we introduce matrices \( A_n \equiv \frac{1}{N} \sum_{m=0}^{n-1} P^m \) for \( n \geq 1 \). Observe that, for each \( n \geq 1 \), \( A_n \) is again a transition probability matrix. In addition, note that all the matrices \( A_n \) and \( P^m \) commute with one another. Finally, because the state space is finite, it is clear that \( P \) is irreducible if and only if there exists a \( M > 1 \) and an \( \epsilon > 0 \) such that \( A_M \geq \frac{\epsilon}{N} \) for all \((i, j)\).

**Lemma 2.** For any probability vector \( \nu \), \( \|\nu A_n A_m - \nu A_m\|_1 \leq \frac{m-1}{n} \) for all \( m, n \geq 1 \).

**Proof.** First observe that

\[
\|\nu A_n A_m - \nu A_m\|_1 = \frac{1}{m} \left\| \sum_{k=0}^{m-1} (\nu P^k A_n - \nu A_n) \right\|_1 \\
\leq \frac{1}{m} \left\| \nu P^k A_n - \nu A_n \right\|_1.
\]
Second, for each $k \geq 0$,

$$\nu P^k A_n - \nu A_n = \frac{1}{n} \sum_{\ell=0}^{n-1} (\nu P^{k+\ell} - \nu P^n) = \frac{1}{n} \sum_{\ell=k}^{k+n-1} \nu P^\ell - \frac{1}{n} \sum_{\ell=0}^{n-1} \nu P^\ell,$$

from which it is easy to see first that $\|\nu P^k A_n - \nu A_n\|_1 \leq \frac{2k}{n}$ and then that

$$\|\nu A_n A_m - \nu A_m\|_1 \leq \frac{2}{mn} \sum_{k=0}^{m-1} k = \frac{m-1}{n}.$$ \hfill \Box

**Theorem (B).** Assume that $(A_M)_{ij} \geq \frac{\epsilon}{N}$ for some $M \geq 1$ and $\epsilon > 0$ and all $(i,j)$. Then there is exactly one probability vector $\mu$ satisfying $\mu P = \mu$. In fact, for any probability vector $\nu$ and $n \geq 1$,

$$\|\nu A_n - \mu\|_1 \leq M - \frac{1}{n}.$$ 

**Proof.** First, note that, by Doeblin’s Theorem, there is precisely one $\mu$ such that $\mu A_M = \mu$. In particular, because $\mu PA_M = \mu A_M P = \mu P$, this shows that $\mu = \mu P$. On the other, if $\nu P = \nu$, then $\nu = \nu A_M$, and so, $\nu = \mu$. In other words, $\mu$ is the one and only $\nu$ satisfying $\nu P = \nu$.

To finish the proof, note that, by Lemmas 1 & 2,

$$\|\nu A_n - \mu\|_1 \leq \|\nu A_n - \nu A_M A_n\|_1 + \|\nu A_M A_n - \mu A_M\|_1 \leq \frac{M-1}{n} + (1-\epsilon)\|\nu A_n - \mu\|_1.$$ \hfill \Box

**Remark:** It is important to realize this last theorem is as well as one can do when $P$ is irreducible but does not satisfy the conditions of the Corollary to Doeblin’s Theorem. To see this, consider the case when $N = 2$ and

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

It is then easy to check that $P^n$ is either the identity $I$ or $P$ depending on whether $n$ is even or odd. Hence $P^n$ isn’t converging to anything as $n \to \infty$. On the other hand, $A_n$ equals

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \text{ or } \begin{bmatrix} \frac{1}{2} - \frac{1}{n} & \frac{1}{2} + \frac{1}{n} \\ \frac{1}{2} + \frac{1}{n} & \frac{1}{2} - \frac{1}{n} \end{bmatrix}$$

depending on whether $n$ is even or odd. In particular,

$$\|(1,0)A_n - \left(\frac{1}{2}, \frac{1}{2}\right)\|_1 = \frac{1}{n} \text{ when } n \text{ is odd.}$$

**Interpretation of $\mu$**

Throughout this section, we will be assuming that $\{1, \ldots, N\}$ is the state space and that, for some $M \geq 1$ and $\epsilon > 0$,

$$(A_M)_{ij} \geq \frac{\epsilon}{N} \text{ for all } (i,j).$$

In addition, we will be using $\rho_j = \inf\{m \geq 1 : X_m = j\} (\equiv \infty \text{ if } X_m \neq j \text{ for any } m \geq 1)$ to denote the first time of return to the state $j$. Our goal is to prove that $\mu_j = \frac{1}{R_j}$, where $R_j \equiv \mathbb{E}[\rho_j|X_0 = j]$. 

**Theorem (C) (Mean Ergodic).** If \( \mu \) is the distribution of \( X_0 \), then

\[
\max_j \mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=0}^{n-1} 1_{(j)}(X_m) - \mu_j \right)^2 \right] \leq \frac{2M}{n\epsilon}.
\]

**Proof.** Set \( f = (1_{(j)}(1) - \mu_j, \ldots, 1_{(j)}(N) - \mu_j) \). Then

\[
\frac{1}{n} \sum_{m=0}^{n-1} 1_{(j)}(X_m) - \mu_j = \frac{1}{n} \sum_{m=0}^{n-1} f_{X_m},
\]
and so

\[
\mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=0}^{n-1} 1_{(j)}(X_m) - \mu_j \right)^2 \right] = \frac{2}{n^2} \sum_{k=0}^{n-1} \sum_{\ell=0}^{n-k-1} \mathbb{E}[f_{X_k} f_{X_{k+\ell}}] - \frac{1}{n^2} \sum_{m=0}^{n-1} \mathbb{E}[f_\mu^2]
\]

\[
\leq \frac{2}{n^2} \sum_{k=0}^{n-1} \mathbb{E} \left[ f_{X_k} \sum_{\ell=0}^{n-k-1} (P^\ell f)_{X_k} \right] = \frac{2}{n^2} \sum_{k=0}^{n-1} (n-k) \mathbb{E}[f_{X_k} (A_{n-k} f)_{X_k}].
\]

But, by the Corollary to Theorem (A),

\[
|\langle A_n f \rangle_i| = |\langle A_n \rangle_i - \mu_j| \leq \frac{M}{m\epsilon},
\]
and so the preceding leads to

\[
\mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=0}^{n-1} 1_{(j)}(X_m) - \mu_j \right)^2 \right] \leq \frac{2M}{n\epsilon} \quad \square.
\]

**Lemma 3.** For every initial distribution and \( \alpha < - \log \left(1 - \frac{\epsilon}{N} \right) \),

\[
\mathbb{E} \left[ e^{\alpha u_{ij}} \right] \leq 2 + (e^\alpha - 1) \left( 1 - e^\alpha \left(1 - \frac{\epsilon}{N} \right) \right)^{-1}.
\]

In particular, all moments of \( \rho_j \) are finite.

**Proof.** Set \( u_{ij}(m) = \mathbb{P}(\rho_j > mM \mid X_0 = i) \). Then \( u_{ij}(0) = 1 \) and

\[
u_{ij}(m + 1) = \sum_{k=1}^{N} \mathbb{P}(\rho_j > (m+1)M \& X_{m+M} = k \mid X_0 = i)
\]

\[
= \sum_{k=1}^{N} \mathbb{P}(\rho_j > M \mid X_0 = k) \mathbb{P}(\rho_j > mM \& X_{m+M} = k \mid X_0 = i)
\]

\[
\leq \max_k u_{kj}(1) \sum_{k=1}^{N} \mathbb{P}(\rho_j > mM \& X_{m+M} = k \mid X_0 = i) \leq \left( \max_k u_{kj} \right) u_{ij}(m).
\]

Hence, if \( u \equiv \max_k u_{kj}(1) \), then \( u_{ij}(m) \leq u^m \) for all \( m \geq 0 \) and \((i, j)\). Finally,

\[
1 - u_{ij}(1) = \mathbb{P}(\rho_j \leq M \mid X_0 = i) \geq \max_{1 \leq m \leq M} (P^m)_{ij} \geq (A_M)_{ij} \geq \frac{\epsilon}{N},
\]
and so \( u \leq 1 - \frac{\epsilon}{N} \). Equivalently, \( \mathbb{P}(\frac{\rho_j}{M} > m) \leq (1 - \frac{\epsilon}{N})^m \), from which the asserted estimates are easy. \( \square \)

Set \( \rho_j^{(0)} = 0 \) and, for \( n \geq 1 \), \( \rho_j^{(n)} = \inf \{ m \mid \rho_j^{(n-1)} : X_m = j \} \). That is, for \( n \geq 1 \), \( \rho_j^{(n)} \) is the time of the \( n \)th return to \( j \).
Lemma 4. As \( n \to \infty \),
\[
\max_j \mathbb{E} \left[ \left( \frac{\rho_j(n)}{n} - R_j \right)^2 \mid X_0 = j \right] \to 0,
\]
and so, for each \( \delta > 0 \),
\[
\lim_{n \to \infty} \max_j \mathbb{P} \left( \left| \frac{\rho_j(n)}{n} - R_j \right| \geq \delta \mid X_0 = j \right) = 0.
\]

Proof. Set \( \tau_n(j) = \rho_j(n) - \rho_j(n-1) \) for \( n \geq 1 \). Then, because
\[
\mathbb{P}(\tau_{n+1}(j) = m_{n+1} \mid \tau_1(j) = m_1, \ldots, \tau_n(j) = m_n)
= \mathbb{P}(\tau_{n+1}(j) = m_{n+1} \mid \rho_j(n) = m_1 + \cdots + m_n) = \mathbb{P}(\rho_j = m_{n+1} \mid X_0 = j).
\]
This proves that, conditioned on \( X_0 = j \), \( \{\tau_n(j) : n \geq 1\} \) is a sequence of mutually independent random variables which have the same distribution as \( \rho_j \). Hence, just as in the proof of the weak law of large numbers,
\[
\mathbb{E} \left[ \left( \frac{\rho_j(n)}{n} - R_j \right)^2 \mid X_0 = j \right] = \mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=1}^{n} (\tau_m(j) - R_j) \right)^2 \mid X_0 = j \right] = \frac{1}{n} \text{var}(\rho_j \mid X_0 = j)
\]
and
\[
\mathbb{P} \left( \left| \frac{\rho_j(n)}{n} - R_j \right| \geq \delta \mid X_0 = j \right) \leq \frac{\text{var}(\rho_j \mid X_0 = j)}{n\delta^2}.
\]

Theorem (D). For each \( j \), \( \mu_j = \pi_j = \frac{1}{R_j} \).

Proof. First note that, because \( \mu_j = (\mu A_M)_j \), \( \mu_j \geq \frac{2}{N} \) for all \( j \). Hence, from by Theorem (C),
\[
\frac{2M}{nc} \geq \sum_j \mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{(j)}(X_m) - \mu_j \right)^2 \mid X_0 = j \right] \mu_j
\geq \frac{c}{N} \max_j \mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{(j)}(X_m) - \mu_j \right)^2 \mid X_0 = j \right],
\]
and so
\[
(*) \quad \lim_{n \to \infty} \max_j \mathbb{E} \left[ \left( \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{(j)}(X_m) - \mu_j \right)^2 \mid X_0 = j \right] = 0.
\]
Next, use \([x]\) is the integral part (i.e., the largest integer dominated by \( x \in \mathbb{R} \)), and observe that, when \( X_0 = j \),
\[
\left| \frac{1}{[nR_j]} \sum_{m=0}^{[nR_j]-1} \mathbf{1}_{(j)}(X_m) - \frac{1}{R_j} \right| = \left| \frac{1}{[nR_j]} \sum_{m=0}^{[nR_j]-1} \mathbf{1}_{(j)}(X_m) - \frac{n}{[nR_j]} \right| + \frac{1}{n}
= \frac{1}{[nR_j]} \left| \sum_{m=0}^{[nR_j]-1} \mathbf{1}_{(j)}(X_m) - \rho_j^{(n)} \right| + \frac{1}{n}
\leq \frac{[nR_j] - \rho_j^{(n)}}{[nR_j]} + \frac{1}{n} \leq \left| R_j - \frac{\rho_j^{(n)}}{n} \right| + \frac{2}{n}.
where. Thus, by Lemma 4,

\[
\lim_{n \to \infty} E \left[ \left| \frac{1}{n R_j} \sum_{m=0}^{[n R_j] - 1} 1_\{j\}(X_m) - \frac{1}{R_j} \right|^2 \bigg| X_0 = j \right] = 0.
\]

At the same time, by (*),

\[
\lim_{n \to \infty} E \left[ \left( \frac{1}{n R_j} \sum_{m=0}^{[n R_j] - 1} 1_\{j\}(X_m) - \mu_j \right)^2 \bigg| X_0 = j \right] = 0,
\]

and so we are done. \(\square\)