(8.1) **Lemma.** *If there is a stationary distribution, then all states y that have $\pi(y) > 0$ are recurrent.*

Proof. (3.11) tells us that $E_x N(y) = \sum_{n=1}^{\infty} p^n(x,y)$, so

$$\sum_x \pi(x) E_x N(y) = \sum_x \pi(x) \sum_{n=1}^{\infty} p^n(x,y)$$

Interchanging the order of summation and using $\pi p^n = \pi$, the above

$$= \sum_{n=1}^{\infty} \sum_x \pi(x) p^n(x,y) = \sum_{n=1}^{\infty} \pi(y) = \infty$$

since $\pi(y) > 0$. Using (3.8) now gives $E_x N(y) = \rho_{xy}/(1-\rho_{yy})$, so

$$\infty = \sum_x \pi(x) \frac{\rho_{xy}}{1-\rho_{yy}} \le \frac{1}{1-\rho_{yy}}$$

the second inequality following from the facts that $\rho_{xy} \le 1$ and π is a probability measure. This shows that $\rho_{yy} = 1$, i.e., y is recurrent. □

With (8.1) in hand we are ready to tackle the proof of:

(4.5) **Convergence theorem.** *Suppose p is irreducible, aperiodic, and has stationary distribution π. Then as $n \to \infty$, $p^n(x,y) \to \pi(y)$.*

Proof. Let $S^2 = S \times S$. Define a transition probability $\bar{p}$ on $S \times S$ by

$$\bar{p}((x_1,y_1),(x_2,y_2)) = p(x_1,x_2)p(y_1,y_2)$$

In words, each coordinate moves independently. Our first step is to check that $\bar{p}$ is irreducible. This may seem like a silly thing to do first, but this is the only step that requires aperiodicity. Since p is irreducible, there are K, L, so that $p^K(x_1,x_2) > 0$ and $p^L(y_1,y_2) > 0$. Since x_2 and y_2 have period 1, it follows from (4.2) that if M is large, then $p^{L+M}(x_2,x_2) > 0$ and $p^{K+M}(y_2,y_2) > 0$, so

$$\bar{p}^{K+L+M}((x_1,y_1),(x_2,y_2)) > 0$$

Our second step is to observe that since the two coordinates are independent $\bar{\pi}(a,b) = \pi(a)\pi(b)$ defines a stationary distribution for $\bar{p}$, and (8.1) implies that all states are recurrent for $\bar{p}$. Let (X_n, Y_n) denote the chain on $S \times S$, and let T be the first time that the two coordinates are equal, i.e.,

$T = \min\{n \geq 0 : X_n = Y_n\}$. Let $V_{(x,x)} = \min\{n \geq 0 : X_n = Y_n = x\}$ be the time of the first visit to (x, x). Since $\bar{p}$ is irreducible and recurrent, $V_{(x,x)} < \infty$ with probability one. Since $T \leq V_{(x,x)}$ for our favorite x we must have $T < \infty$.

The third and somewhat magical step is to prove that on $\{T \leq n\}$, the two coordinates X_n and Y_n have the same distribution. By considering the time and place of the first intersection and then using the Markov property we have

$$\begin{aligned}
P(X_n = y, T \leq n) &= \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x, X_n = y) \\
&= \sum_{m=1}^{n} \sum_{x} P(T = m, X_m = x) P(X_n = y | X_m = x) \\
&= \sum_{m=1}^{n} \sum_{x} P(T = m, Y_m = x) P(Y_n = y | Y_m = x) \\
&= P(Y_n = y, T \leq n)
\end{aligned}$$

To finish up we observe that using the last equality we have

$$\begin{aligned}
P(X_n = y) &= P(X_n = y, T \leq n) + P(X_n = y, T > n) \\
&= P(Y_n = y, T \leq n) + P(X_n = y, T > n) \\
&\leq P(Y_n = y) + P(X_n = y, T > n)
\end{aligned}$$

and similarly $P(Y_n = y) \leq P(X_n = y) + P(Y_n = y, T > n)$. So

$$|P(X_n = y) - P(Y_n = y)| \leq P(X_n = y, T > n) + P(Y_n = y, T > n)$$

and summing over y gives

$$\sum_{y} |P(X_n = y) - P(Y_n = y)| \leq 2P(T > n)$$

If we let $X_0 = x$ and let Y_0 have the stationary distribution π, then Y_n has distribution π, and it follows that

$$\sum_{y} |p^n(x, y) - \pi(y)| \leq 2P(T > n) \to 0$$

proving the desired result. □

Next on our list is the equivalence of positive recurrence and the existence of a stationary distribution, (7.2), the first piece of which is:

(8.2) Theorem. *Let x be a positive recurrent state, let $T_x = \inf\{n \geq 1 : X_n = x\}$, and let*

$$\mu(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

Then $\pi(y) = \mu(y)/E_x T_x$ defines a stationary distribution.

To prepare for the proof of (6.5) note that $\mu(x) = 1$ so $\pi(x) = 1/E_x T_x$. Another useful bit of trivia that explains the norming constant is that the definition and (3.10) imply

$$\sum_{y \in S} \mu(y) = \sum_{n=0}^{\infty} P_x(T_x > n) = E_x T_x$$

Why is this true? This is called the "cycle trick." $\mu(y)$ is the expected number of visits to y in $\{0, \ldots, T_x - 1\}$. Multiplying by p moves us forward one unit in time so $\mu p(y)$ is the expected number of visits to y in $\{1, \ldots, T_x\}$. Since $X(T_x) = X_0 = x$ it follows that $\mu = \mu p$. Since π is just μ divided a constant to make the sum 1, π is a stationary distribution.

Proof. To formalize this intuition, let $\bar{p}_n(x, y) = P_x(X_n = y, T_x > n)$ and interchange sums to get

$$\sum_y \mu(y) p(y, z) = \sum_{n=0}^{\infty} \sum_y \bar{p}_n(x, y) p(y, z)$$

Case 1. Consider the generic case first: $z \neq x$.

$$\begin{aligned}\sum_y \bar{p}_n(x, y) p(y, z) &= \sum_y P_x(X_n = y, T_x > n, X_{n+1} = z) \\ &= P_x(T_x > n + 1, X_{n+1} = z) = \bar{p}_{n+1}(x, z)\end{aligned}$$

Here the second equality holds since the chain must be somewhere at time n, and the third is just the definition of $\bar{p}_{n+1}$. Summing from $n = 0$ to ∞, we have

$$\sum_{n=0}^{\infty} \sum_y \bar{p}_n(x, y) p(y, z) = \sum_{n=0}^{\infty} \bar{p}_{n+1}(x, z) = \mu(z)$$

since $\bar{p}_0(x, z) = 0$.

Case 2. Now suppose that $z = x$. Reasoning as above we have

$$\sum_y \bar{p}_n(x, y) p(y, x) = \sum_y P_x(X_n = y, T_x > n, X_{n+1} = x) = P_x(T_x = n + 1)$$

Summing from $n = 0$ to ∞ we have

$$\sum_{n=0}^{\infty} \sum_{y} \bar{p}_n(x, y)p(y, x) = \sum_{n=0}^{\infty} P_x(T_x = n + 1) = 1 = \mu(x)$$

since $P_x(T = 0) = 0$. □

With (8.2) established we can now easily prove:

(4.7) **Theorem.** *If the state space S is finite then there is at least one stationary distribution.*

Proof. By (3.5) we can restrict our attention to a closed irreducible subset of S, and hence suppose without loss of generality that the chain is irreducible. Let $y \in S$. In view of (8.2) it is enough to prove that y is positive recurrent, i.e., $E_y T_y < \infty$. To do this we note that irreducibility implies that for each x there is a $k(x)$ so that $P_x(T_y \le k(x)) > 0$. Since S is finite, $K = \max\{k(x) : x \in S\} < \infty$, and there is an $\alpha > 0$ so that $P_x(T_y \le K) \ge \alpha$. The pedestrian lemma (3.3) now implies that $P_x(T_y > nK) \le (1 - \alpha)^n$, so $E_x T_y < \infty$ for all $x \in S$ and in particular $E_y T_y < \infty$. □

To prepare for the second piece of (7.2) we now prove:

(8.3) **Theorem.** *Suppose p is irreducible. Then for any $x \in S$, as $n \to \infty$*

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

Proof. Consider the first the case in which y is transient. (3.10) implies that $EN_n(y) < \infty$ so $N_n(y) < \infty$ and hence $N_n(y)/n \to 0$ as $n \to \infty$. On the other hand transience implies $P_y(T_y = \infty) > 0$, so $E_y T_y = \infty$ and $1/E_y T_y = 0$.

Turning to the recurrent case, suppose that we start at y. Let $R(k) = \min\{n \ge 1 : N_n(y) = k\}$ be the time of the kth return to y. Let $R(0) = 0$ and for $k \ge 1$ let $t_k = R(k) - R(k-1)$. Since we have assumed $X_0 = y$, the times between returns, $t_1, t_2, \ldots$ are independent and identically distributed so the strong law of large numbers for nonnegative random variables implies that

$$R(k)/k \to E_y T_y \le \infty$$

From the definition of $R(k)$ it follows that $R(N_n(y)) \le n < R(N_n(y) + 1)$. Dividing everything by $N_n(y)$ and then multiplying and dividing on the end by $N_n(y) + 1$, we have

$$\frac{R(N_n(y))}{N_n(y)} \le \frac{n}{N_n(y)} < \frac{R(N_n(y)+1)}{N_n(y)+1} \cdot \frac{N_n(y)+1}{N_n(y)}$$

Letting $n \to \infty$, we have $n/N_n(y)$ trapped between two things that converge to $E_y T_y$, so

$$\frac{n}{N_n(y)} \to E_y T_y$$

To generalize now to $x \neq y$, observe that the strong Markov property implies that conditional on $\{T_y < \infty\}$, $t_2, t_3, \dots$ are independent and identically distributed and have $P_x(t_k = n) = P_y(T_y = n)$ so

$$R(k)/k = t_1/k + (t_2 + \cdots + t_k)/k \to 0 + E_y T_y$$

and we have the conclusion in general. □

From (8.3) we can easily get:

(6.5) **Theorem.** *If p is an irreducible transition probability and has stationary distribution π, then*

$$\pi(y) = 1/E_y T_y$$

Why is this true? From (8.3) it follows that

$$\frac{N_n(y)}{n} \to \frac{1}{E_y T_y}$$

Taking expected value and using the fact that $N_n(y) \leq n$, it can be shown that this implies

$$\frac{EN_n(y)}{n} \to \frac{1}{E_y T_y}$$

By the reasoning that led to (3.11), we have $E_x N_n(y) = \sum_{m=1}^{n} p^m(x,y)$. The convergence theorem implies $p^n(x,y) \to \pi(y)$, so we have

$$\frac{E_x N_n(y)}{n} \to \pi(y)$$

Comparing the last two results gives the desired conclusion. □

We are now ready to put the pieces together.

(7.2) **Theorem.** *For an irreducible chain the following are equivalent:*

(i) Some x is positive recurrent.

(ii) There is a stationary distribution.

(iii) All states are positive recurrent.