Stationary distribution via expected visits during an excursion

Consider an irreducible, recurrent Markov chain $\{X_n : n = 0, 1, \ldots\}$ on a countable state space $\mathcal{S}$ with transition probabilities $P(i, j)$. For an arbitrarily chosen (but fixed) state i, define $T_i := \inf\{n \in \mathbb{N} : X_n = i\}$ and

<1>

$$\begin{aligned}
\lambda_j &:= \mathbb{E}_i\{\text{ number of visits to state } j \text{ up to time } T_i\ \} \\
&= \mathbb{E}_i\left(\sum\nolimits_{n\in\mathbb{N}} 1\{X_n = j, n \le T_i\}\right) \\
&= \sum\nolimits_{n\in\mathbb{N}} \mathbb{P}_i\{X_n = j, n \le T_i\} \qquad \text{taking expectation term-by-term} \\
&= P(i, j) + \sum\nolimits_{m\ge 2} \mathbb{P}_i\{X_m = j, m \le T_i\}
\end{aligned}$$

Notice that $\lambda_i = 1$, because the first visit to i ends the excursion.

Over many (independent) excursions, we expect the chain to average λ_j visits to state j. On the average, a fraction $P(j, k)$ of those visits should be followed immediately by a visit to state k. (Does this heuristic work when j equals i? Maybe we need to consider the state for X_1 to make it work.) Summing over all possible j that might be visited right before a visit to state k, we could then hope that

<2>

$$\lambda_k = \sum\nolimits_{j\in\mathcal{S}} \lambda_j P(j, k) \qquad \text{for every } k.$$

To make the argument rigorous, we would need to show that everything works according to averages.

Alternatively, first invoke the Markov property,

$$\mathbb{P}_i\{X_1 = j_1, X_2 = j_2, \ldots, X_n = j, X_{n+1} = k\} = \mathbb{P}_i\{X_1 = j_1, X_2 = j_2, \ldots, X_n = j\}P(j.k),$$

then sum over all $j_1, \ldots, j_{n-1}$ not equal to i to get

<3>

$$\mathbb{P}_i\{T_i \ge n, X_n = j, X_{n+1} = k\} = \mathbb{P}_i\{T_i \ge n, X_n = j\}P(j, k).$$

Sum both sides of the last equality over n and j, using the form <1> for λ_j.

$$\sum\nolimits_{n\in\mathbb{N}} \sum\nolimits_{j\in\mathcal{S}} \mathbb{P}_i\{T_i \ge n, X_n = j, X_{n+1} = k\} = \sum\nolimits_{j\in\mathcal{S}} \lambda_j P(j, k).$$

On the left-hand side, split the event $\{T_i \ge n\}$ into a union of two disjoint events, $\{T_i = n\} \cup \{T_i \ge n+1\}$, thereby breaking the sum into

<4>

$$\sum\nolimits_{n\in\mathbb{N}} \sum\nolimits_{j\in\mathcal{S}} \mathbb{P}_i\{T_i = n, X_n = j, X_{n+1} = k\} + \sum\nolimits_{n\in\mathbb{N}} \sum\nolimits_{j\in\mathcal{S}} \mathbb{P}_i\{T_i \ge n+1, X_n = j, X_{n+1} = k\}$$

The sum over j collapses the second double sum to

$$\sum\nolimits_{n\in\mathbb{N}} \mathbb{P}_i\{T_i \ge n+1, X_{n+1} = k\} = \sum\nolimits_{m\ge 2} \mathbb{P}_i\{T_i \ge m, X_m = k\}$$

If the final sum were to start at $m = 1$ we woud have the expression for λ_k. We lack only the summand $\mathbb{P}_i\{T_i \ge 1, X_1 = k\} = P(i, k)$.

The collapse of the first double sum in <4> is even more dramatic, because only the summands with j equal to i survive: if $T_i = n$ then we must have $X_n = i$. The double sum reduces to

$$\sum\nolimits_{n\in\mathbb{N}} \mathbb{P}_i\{T_i = n, X_n = i, X_{n+1} = k\} = \sum\nolimits_{n\in\mathbb{N}} \mathbb{P}_i\{T_i = n, X_n = i\}P(i, k)$$

the final factorization coming from <3>. (Why can we replace $T_i \ge n$ by $T_i = n$ when j equals i?) With the factorization achieved, we can discard the redundant $X_n = i$ from the event $\{T_i = n, X_n = i\}$. The sum over n corresponds to a union over disjoint events $\{T_i = n\}$; the probabilities $\mathbb{P}_i\{T_i = n, X_n = i\}$ sum to 1, leaving us the $P(i, k)$ needed to reduce <4> to λ_k.

If $\sum_j \lambda_j < \infty$, the standardized values $\pi_k = \lambda_k / \sum_j \lambda_j < \infty$ sum to 1. Equality <2> then gives us the equations that identify $\{\pi_k : k \in \mathcal{S}\}$ as a stationary probability distribution for the chain. From <1>,

$$\sum\nolimits_{j\in\mathcal{S}} \lambda_j = \sum\nolimits_{n\in\mathbb{N}} \sum\nolimits_{j\in\mathcal{S}} \mathbb{P}_i\{X_n = j, n \le T_i\} = \sum\nolimits_{n\in\mathbb{N}} \mathbb{P}_i\{n \le T_i\} = \mathbb{E}_i \sum\nolimits_{n\in\mathbb{N}} 1\{n \le T_i\} = \mathbb{E}_i T_i.$$

That is, if the chain is ***positive recurrent*** (meaning that $\mathbb{E}_i T_i < \infty$) it has a stationary probability distribution.