Discrete Markov Chains - Stationary Distributions

May 8, 2015

Definitions

Definition 1. A probability distribution $\boldsymbol{\pi} \in \mathbb{R}^n$ is a **stationary distribution** of $\mathbf{T}$ when

$$\boldsymbol{\pi}^T \mathbf{T} = \boldsymbol{\pi}^T \tag{1}$$

holds. Applying Eq. (1) m times leads to $\boldsymbol{\pi}^T \mathbf{T}^m = \boldsymbol{\pi}^T$. A stationary distribution is a left eigenvector of the transition matrix with eigenvalue one.
We investigate the conditions needed to ensure existence and uniqueness of stationary distributions. In this lecture, we are going to show that irreducibility is sufficient for the existence and uniqueness of a stationary distribution. Therefore, for the rest of this lecture, let $\mathbf{T}$ be the transition matrix of an irreducible Markov chain. We are going to show that:

Theorem 2. *Let* $\mathbf{T}$ *be the transition matrix of an irreducible Markov chain. Then there exists a unique stationary distribution* π *such that* $\pi^T \mathbf{T} = \pi$ *and* $\pi_i > 0$ *for all states* i*.*

Hitting Times and Existence

As a preparation, we need two technical lemmas and a definition:

Lemma 3. *Let* X *be a non-negative integer-valued random variable. Then*

$$\mathbb{E}(X) \;=\; \sum_{k=0}^{\infty} \mathbb{P}(X > k)$$

Proof. We have that

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{P}(X > k) &= \sum_{k=0}^{\infty} \sum_{l>k} \mathbb{P}(X = l) \\ &= \sum_{k=0}^{\infty} k\mathbb{P}(X = k). \end{aligned}$$

□

Definition 4. For any state i, we define the **first hitting time** of that state by

$$\tau_i \quad = \quad \min\{k \geq 1 : X_k = i\}.$$

We can now use Lemma 3 to show that the expected hitting time of any state is finite in an irreducible chain.

Lemma 5. *For any states i, j of an irreducible chain, we have that*

$$\mathbb{E}_i(\tau_j) \quad < \quad \infty.$$

Proof. The idea of the proof is to use irreducibility in order to show that the probability of the hitting time being large decays exponentially. Irreducibility implies that for any states i, j there is a positive integer r_{ij} and a positive number ϵ_{ij} s.t.

$$\mathbf{T}^{r_{ij}}(i,j) \quad = \quad \epsilon_{ij}.$$

We define $r := \max r_{ij}$ and $\epsilon := \min \epsilon_{ij}$. It follows that, whatever the value of X_k, the probability to hit state j between steps k and $k+r$ is at least ϵ (why?). Consequently, we have that

$$\begin{aligned}
\mathbb{P}_i(\tau_j > mr) \quad &\leq \quad \mathbb{P}_i(X_k \neq j,\, k = (m-1)r+1, \ldots, mr)\mathbb{P}_i(\tau_j > (m-1)r). \\
&\leq \quad (1-\epsilon)\mathbb{P}_i(\tau_j > (m-1)r).
\end{aligned}$$

Repeated application of this argument shows that

$$\mathbb{P}_x(\tau_j > mr) \quad \leq \quad (1-\epsilon)^m.$$

Now, we can use Lemma 3 to conclude:

$$\begin{aligned}
\mathbb{E}_i(\tau_j) \quad &= \quad \sum_{k=0}^{\infty} \mathbb{P}_i(\tau_j > k) \\
&\leq \quad r\sum_{m=0}^{\infty} \mathbb{P}_i(\tau_j > mr) \\
&\leq \quad r\sum_{m=0}^{\infty} (1-\epsilon)^m \\
&< \quad \infty.
\end{aligned}$$

□

Now, we can pick a candidate distribution for π and show that it is stationary. The intuition is that as soon as the chain returns to some state i for the first time, the chain basically restarts, as its memory is forgotten by the Markov property. Therefore, the expected relative number of visits to any state before returning to i should be the stationary distribution.

Definition 6. For a Markov chain starting in i, we define the following random variables:

$$R_i^j \quad = \quad \text{Number of visits to } j \text{before returning to } i$$

Lemma 7. *Fix a state i. Define the distribution*

$$\pi(j) \quad = \quad \frac{\mathbb{E}_i(R_i^j)}{\mathbb{E}_i(\tau_i)}. \tag{2}$$

Then π is a stationary distribution of the Markov chain $\mathbf{T}$.

Proof. We note first of all that $\sum_j \pi(j) = \mathbb{E}_i(\tau_i)$, hence π is finite by Lemma 5 and the denominator in Eq. (2) is just the normalization. We will show that

$$\tilde{\pi}(j) \quad = \quad \mathbb{E}_i(R_i^j)$$

is a stationary vector of the Markov chain. The first step is to arrive at a more useful expression for $\tilde{\pi}$: On the set of trajectories generated by $\mathbf{T}$, define random varibles

$$I_j^m \quad = \quad \begin{cases} 1 & \text{if } X_m = j \text{ and } \tau_i > m \\ 0 & \text{otherwise.} \end{cases}$$

We conclude that $R_i^j = \sum_{m=0}^{\infty} I_j^m$ and

$$\begin{aligned} \tilde{\pi}(j) \quad &= \quad \sum_{m=0}^{\infty} \mathbb{E}_i(I_j^m) \\ &= \quad \sum_{m=0}^{\infty} \mathbb{P}_i(X_m = j,\, \tau_i > m). \end{aligned}$$

Now, we can multiply our candidate distribution to the left of $\mathbf{T}$:

$$\begin{aligned} \sum_j \tilde{\pi}(j)\mathbf{T}_{jk} \quad &= \quad \sum_j \sum_{m=0}^{\infty} \mathbb{P}_i(X_m = j,\, \tau_i > m)\mathbf{T}_{jk} \\ &= \quad \sum_j \sum_{m=0}^{\infty} \mathbb{P}_i(X_m = j,\, X_{m+1} = k,\, \tau_i \geq m+1) \\ &= \quad \sum_{m=0}^{\infty} \mathbb{P}_i(X_{m+1} = k,\, \tau_i \geq m+1). \\ &= \quad \sum_{m=1}^{\infty} \mathbb{P}_i(X_m = k,\, \tau_i \geq m). \end{aligned}$$

The second line is not straightforward. It is a consequence of the Markov property (in which way?). Finally, the last expression is indeed equal to $\tilde{\pi}(k)$:

$$\begin{aligned}
\sum_{m=1}^{\infty} \mathbb{P}(X_m = k,\, \tau_i \geq m) &= \tilde{\pi}(k) - \mathbb{P}_i(X_0 = k,\, \tau_i > 0) + \sum_{m=1}^{\infty} \mathbb{P}_i(X_m = k,\, \tau_i = m) \\
&= \tilde{\pi}(k) - \mathbb{P}_i(X_0 = k) + \mathbb{P}_i(X_{\tau_i} = k) \\
&= \tilde{\pi}(k).
\end{aligned}$$

The last equality is due to the cancellation of the two extra terms. If $k \neq i$, both terms vanish, and if $k = i$, both terms equal one. □

Uniqueness

Lemma 8. *The space of right eigenvectors of* $\mathbf{T}$ *with eigenvalue one is one-dimensional and it is spanned by the constant vector* $\mathbf{v}$, $v_i \equiv 1$.

Proof. By a direct calculation, we see that $\mathbf{v}$ is a right eigenvector of $\mathbf{T}$ with eigenvalue 1:

$$\begin{aligned}
\sum_{j=1}^{n} \mathbf{T}_{ij} v_j &= \sum_{j=1}^{n} \mathbf{T}_{ij} \\
&= 1 \\
&= v_i.
\end{aligned}$$

Next, suppose there is non-constant right eigenvector $\mathbf{w}$ with eigenvalue 1. It follows that there is one i s.t. $w_i = M$ is maximal among all w_j. Now, suppose that there is one k such that $\mathbf{T}_{ik} = \epsilon > 0$ and $w_k < M$. The eigenvalue equation then implies that:

$$\begin{aligned}
M &= v_i \\
&= \mathbf{T}_{ik} v_k + \sum_{j \neq k} \mathbf{T}_{ij} v_j \\
&< \epsilon M + (1 - \epsilon) M \\
&= M.
\end{aligned}$$

This argument implies that $\mathbf{w}$is constant on all states that are accessible from state i. By irreducibility, $\mathbf{w}$ is constant on all states. □

Lemma 9. *The stationary vector* π *of an irreducible Markov chain* $\mathbf{T}$ *is unique.*

Proof. The multiplicity of right and left eigenspaces is always identical, hence, the space of left eigenvectors with eigenvalue one is one-dimensional by Lemma 8. Clearly, there can only be one left eigenvector such that its elements sum up to one. □

Reversible Transition Matrices

From now on, we will be interested in reversible transition matrices:

Definition 10. A transition matrix $\mathbf{T}$ with stationary distribution π is called **reversible** with respect to π if it satisifes the **detailed balance** condition

$$\begin{aligned} \pi_i \mathbf{T}_{ij} &= \pi_j \mathbf{T}_{ji}, \forall i, j. \\ \mathbf{\Pi T} &= \mathbf{T}^T \mathbf{\Pi}, \end{aligned} \tag{3}$$

where $\mathbf{\Pi} = \mathrm{diag}(\pi)$.Reversibility states that the joint probability of seeing the system first in state i and then in state j one step later is the same as vice versa. We will mostly focus our attention on reversible transition matrices, as in the biological application, Eq. (3) generally holds even on the microscopic (continuous) level, and should therefore be transferred to the discretized level.