Section 8: Non-stationary Transition Probabilities

Contents

8.1 Computing Multi-step Transition Probabilities in Discrete Time

In modeling the dynamics of an S-valued Markov chain $X = (X_n : n \geq 0)$ with non-stationary transition probabilities, we need to specify the sequence of (one-step) transition matrices $(P(n) : n \geq 1)$. In this case,

$$P(X_{n+1} = y | X_j : 0 \leq j \leq n) = P(n+1,\, X_n,\, y),$$

where $P(n+1,\, x,\, y)$ is the $(x,\, y)$'th entry of $P(n+1)$. Such models arise in settings in which one needs to incorporate explicit time-of -day, day-of-week, or seasonality effects.

Let P_n be the matrix in which the (x, y)'th entry is given by

$$P_n(x, y) = P(X_n = y | X_0 = x).$$

Note that

$$\begin{aligned} P_n(x, y) &= \sum_{x_1, x_2, \ldots, x_{n-1}} P(1, x, x_1)(2, x_1, x_2) \cdots P(n, x_{n-1}, y) \\ &= (P(1)P(2) \cdots P(n))(x, y). \end{aligned}$$

proving the following result.

Proposition 8.1.1 The n step transition matrix P_n is given by

$$P_n = P(1)P(2) \cdots P(n).$$

As pointed out earlier, it is preferable (from a computational complexity viewpoint) to structure one's computation so that the calculations involve matrix/vector products rather than matrix/matrix products.

Suppose, in particular, that we wish to compute $\mu_n = (\mu_n(y) : y \in S)$, where μ_n is a row vector with y'th entry given by

$$\mu_n(y) = P(X_n = y).$$

In addition to the modeler needing to specify $(P(n) : n \geq 1)$, one now also needs to provide the *initial distribution* $\mu = (\mu(x) : x \in S)$ in which

$$\mu(x) = P(X_0 = x).$$

Proposition 8.1.2 The sequence $\mu_n : n \geq 1)$ can be computed recursively via

$$\mu_n = \mu_{n-1} P(n)$$

subject to $\mu_0 = \mu$.

Note that the distribution of the chain at time n can be recursively computed from that at time $n-1$ (i.e. a *forwards recursion*).

Consider next the probability of computing the expected reward $\mathrm{E}[f(X_n)|X_j = x]$, where $f : S \to \mathbb{R}_+$ is a non-negative function. Put

$$u^*(j, x) = \mathrm{E}[f(X_n)|X_{n-j} = x],$$

and note that

$$\begin{aligned} u^*(j+1, x) &= \mathrm{E}[f(X_n)|X_{n-j-1} = x] \\ &= \sum_y \mathrm{E}[f(X_n) I(X_{n-j} = y)|X_{n-j-1} = x] \\ &= \sum_y P(n-j, x, y) u^*(j, y) \end{aligned}$$

so that if we let $u^*(j)$ be the (column) vector in which the x'th entry is $u^*(j, x)$, we can write the above in matrix/vector form as

$$u^*(j+1) = P(n-j) u^*(j)$$

for $0 \leq j < n$. This yields our next proposition.

Proposition 8.1.3 The sequence $(u^*(j) : 0 \leq j \leq n)$ satisfies the recursion

$$u^*(j+1) = P(n-j) u^*(j)$$

for $0 \leq j < n$ subject to $u^*(0) = f$.

In this case, the recursion computes $(\mathrm{E}[f(X_n)|X_{n-j-1} = x] : x \in S)$ from $(\mathrm{E}[f(X_n)|X_{n-j} = x] : x \in S)$, so that the expectation starting from time $n-j-1$ is computed from that starting at time $n-j$ (i.e. a *backwards recursion*).

Remark 8.1.1 If X has stationary transition probabilities, then

$$\mathrm{E}_x f(X_j) = \mathrm{E}[f(X_n)|X_{n-j} = x]$$

so that

$$\mathrm{E}_x f(X_{j+1}) = \sum_y P(x, y) \mathrm{E}_y f(X_j).$$

Exercise 8.1.1 For $C^c \subset S$, let $T = \inf\{n \geq 0 : X_n \in C^c\}$. Consider

$$\mathrm{E}_x \sum_{j=0}^{T-1} f(X_j)$$

for a given reward function $f : S \to \mathbb{R}_+$. Discuss how to efficiently compute this via matrix/vector operations, when the initial distribution μ is given.

Exercise 8.1.2 Suppose that $X = (X_n : n \geq 0)$ is a Markov chain with non-stationary transition probabilities in which X_i takes values in S_i for $i \geq 0$. (In other words, the state space at time i can depend on i. This permits one, for example, to use a richer state space at some times of day than at other times.) How do the previous propositions generalize to this setting?

8.2 Asymptotic Loss of Memory

Systems with non-stationary transition probabilities are systems for which the notion of equilibrium and steady-state typically fail to make sense. So, for such models, computing steady-state/equilibrium/staionary distributions is not a meaningful element of any stochastic analysis of the system.

However, many models with non-stationary transition probabilities exhibit an *asymptotic loss of memory*, by which we mean that for each $x, y \in S$,

$$P(X_{n+k} = z | X_k = x) - P(X_{n+k} = z | X_k = y) \to 0$$

as $n \to \infty$. The relation asserts that the system at time $n + k$ has a distribution that is essentially independent of the system state at time k when n is large. This can be useful, for example, in running simulations to compute probabilities and expectations at a given time m, with m (very) large. One can initiate such a simulation at time $m - n$ with an arbitrary initialization, provided that n is selected large enough that the system has lost its memory of the initial state by time m. (Consider, for example, simulating morning rush hour traffic at 7AM. How far back does one need to start the simulation to get a good sample of 7AM's typical traffic?)