MAU34804

Lecture 15

2026-02-25

Markov matrices

For a Markov process let

$$p_j(t_i) = p(X(t_i) = x_j), \quad a_{j,k}^{(m)} = p(X(t_{i+m}) = x_j \,|\, X(t_i) = k), \quad a_{j,k} = a_{j,k}^{(1)}.$$

Then we have the matrix equations

$$\mathbf{p}(t_{i+m}) = A^m \mathbf{p}(t_i), \quad \mathbf{p}(t_{i+m}) = A^{(m)} \mathbf{p}(t_i), \quad A^{(m)} = A^m.$$

Only the last of these requires the second condition from the definition.
The k'th column of A is a set of conditional probabilities, all conditional on the same event $X(t_i) = k$ so the entries satisfy the usual conditions for probabilities of an exhaustive set of mutually exclusive events: they are non-negative and sum to 1.
In general square matrices A such that $a_{j,k} \geq 0$ for all j and k and such that $\sum_{j=1}^{n} a_{j,k} = 1$ are called *Markov matrices*.
Or roughly half the world calls them Markov matrices. For the other half the roles of rows and columns are reversed.

Stable probability distributions

The condition that $\sum_{j=1}^{n} a_{j,k} = 1$ is the same as the condition that $(1, 1, \ldots, 1)$ is an eigenvector of A^T with eigenvalue 1.

A^T and A have the same eigenvalues, so 1 must also be an eigenvalue of A.

A vector $\mathbf{p}$ with $p_k \geq 0$ for all k, $\sum_{k=1}^{n} p_k = 1$ and $A\mathbf{p} = \mathbf{p}$ is of interest because it represents a stable probability distribution for the Markov matrix with transition matrix A.

Such a vector is an eigenvector of A with eigenvalue 1.

We know there must be an eigenvector of eigenvalue 1 but we don't yet know that there's one satisfying the extra conditions $p_k \geq 0$ for all k and $\sum_{k=1}^{n} p_k = 1$.

Markov and Brouwer

A Markov matrix is a square matrix A satisfying $a_{j,k} \geq 0$ for all j and k and $\sum_{j=1}^{n} a_{j,k} = 1$.
If $p_k \geq 0$ for all k and $\sum_{k=1}^{n} p_k = 1$ then $\sum_{j=k}^{n} a_{j,k} p_k \geq 0$ for all j and

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_{j,k} p_k = \sum_{k=1}^{n} \sum_{j=1}^{n} a_{j,k} p_k = \sum_{k=1}^{n} p_k = 1.$$

In other words, multiplication from the left by A maps Δ to itself, where Δ is the standard simplex

$$\Delta = \{\mathbf{p} \in \mathbf{R}^n \colon p_k \geq 0, \sum_{k=1}^{n} p_k = 1\}.$$

Matrix multiplication is continuous so Brouwer's fixed point theorem tells us there is a fixed point, i.e. a $\mathbf{p} \in \Delta$ such that $A\mathbf{p} = \mathbf{p}$. We already knew that $A\mathbf{p} = \mathbf{p}$ had a non-zero solution but we didn't know it had a solution in Δ.

Problems, and solutions

Compared to the Banach fixed point theorem, Brouwer's theorem has two disadvantages.
- It's non-constructive, so we don't know how to find the fixed point.
- It only gives existence of a fixed point, not uniqueness.

The first problem isn't serious here. Brouwer doesn't tell us how to find the eigenvector(s), but linear algebra does.

The second problem is unavoidable without further hypotheses on A. $A = I$ is a perfectly good Markov matrix, although the Markov process it describes is a rather boring one, and every vector in Δ is an eigenvector with eigenvalue 1.

If we do make further hypotheses then we can get uniqueness though. The problem in our counterexample is that the probability of a transition is too low. If we assume that all possible transitions have positive probability then we can get uniqueness.

Markov and Banach

From now on consider Δ to have the metric d_1 rather than the Euclidean metric d_2. Multiplication by A is still continuous.

$$d_1(A\mathbf{p}, A\mathbf{q}) = \sum_{j=1}^{n} \left| \sum_{k=1}^{n} a_{j,k}(p_k - q_k) \right|.$$

Then

$$\begin{aligned} d_1(A\mathbf{p}, A\mathbf{q}) &\leq \sum_{j=1}^{n} \sum_{k=1}^{n} \left| a_{j,k}(p_k - q_k) \right| = \sum_{j=1}^{n} \sum_{k=1}^{n} a_{j,k} \left| p_k - q_k \right| \\ &= \sum_{k=1}^{n} \sum_{j=1}^{n} a_{j,k} \left| p_k - q_k \right| = \sum_{k=1}^{n} \left| p_k - q_k \right| = d_1(\mathbf{p}, \mathbf{q}). \end{aligned}$$

This is the $d_1(A\mathbf{p}, A\mathbf{q}) \leq c d_1(\mathbf{p}, \mathbf{q})$ we need for Banach, except with $c = 1$ rather than $c < 1$. Can we improve it?

Improvement

For any $\mathbf{p}, \mathbf{q} \in \Delta$ define

$$J_+ = \left\{ j \colon \sum_{k=1}^{n} a_{j,k}(p_k - q_k) > 0 \right\}, \quad J_- = \left\{ j \colon \sum_{k=1}^{n} a_{j,k}(p_k - q_k) < 0 \right\},$$

$$K_+ = \{ k \colon p_k - q_k > 0 \}, \quad K_- = \{ k \colon p_k - q_k < 0 \}.$$

$$\begin{aligned} d_1(A\mathbf{p}, A\mathbf{q}) &= \sum_{j=1}^{n} \left| \sum_{k=1}^{n} a_{j,k}(p_k - q_k) \right| = \sum_{j \in J_+} \sum_{k=1}^{n} a_{j,k}(p_k - q_k) - \sum_{j \in J_-} \sum_{k=1}^{n} a_{j,k}(p_k - q_k) \\ &= \sum_{k \in K_+} \left(\sum_{j \in J_+} a_{j,k} - \sum_{j \in J_-} a_{j,k} \right) |p_k - q_k| + \sum_{k \in K_-} \left(\sum_{j \in J_-} a_{j,k} - \sum_{j \in J_+} a_{j,k} \right) |p_k - q_k| \end{aligned}$$

Improvement, continued

$$d_1(A\mathbf{p}, A\mathbf{q}) = \sum_{k\in K_+}\left(\sum_{j\in J_+} a_{j,k} - \sum_{j\in J_-} a_{j,k}\right)|p_k - q_k| + \sum_{k\in K_-}\left(\sum_{j\in J_-} a_{j,k} - \sum_{j\in J_+} a_{j,k}\right)|p_k - q_k|$$

$$\leq \max_{1\leq k\leq n}\left|\sum_{j\in J_+} a_{j,k} - \sum_{j\in J_-} a_{j,k}\right|\left(\sum_{k\in K_+}|p_k - q_k| + \sum_{k\in K_-}|p_k - q_k|\right)$$

The parenthesised expression on the second line is just $d_1(\mathbf{p}, \mathbf{q})$.

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_{j,k}(p_k - q_k) = 0$$

and J_+ and J_- are the sets of j where the summand is positive or negative, respectively, so either all summands are 0, in which case $d_1(A\mathbf{p}, A\mathbf{q}) = 0$, or some are positive and some are negative.

Assume we're in the second case.

Improvement, conclusion

$$\sum_{j\in J_+} a_{j,k} - \sum_{j\in J_-} a_{j,k} = \sum_{j=1}^{n} a_{j,k} - \sum_{j\notin J_+} a_{j,k} - \sum_{j\in J_-} a_{j,k} = 1 - \sum_{j\notin J_+} a_{j,k} - \sum_{j\in J_-} a_{j,k} \leq 1 - 2\alpha,$$

where $\alpha = \min_{1\leq j,k\leq n} a_{j,k}$.
The same argument also works with the roles of J_+ and J_- reversed so

$$\left| \sum_{j\in J_+} a_{j,k} - \sum_{j\in J_-} a_{j,k} \right| \leq 1 - 2\alpha.$$

It follows that

$$d_1(A\mathbf{p}, A\mathbf{q}) \leq c d_1(\mathbf{p}, \mathbf{q}),$$

where $c = 1 - 2\alpha$.
The proof above works in the second case from the previous slide, but in the first case we had $d_1(A\mathbf{p}, A\mathbf{q}) = 0$, so the inequality also holds in that case.

Markov and Banach, again

If all entries of A are positive then $\alpha > 0$ so $c < 1$ and we can apply the Banach fixed point theorem to conclude that there is a *unique* $\mathbf{p} \in \Delta$ such that $A\mathbf{p} = \mathbf{p}$.

Of course not all the entries of the identity matrix are positive!

The Banach fixed point theorem also tells us that $\mathbf{p} = \lim_{i\to\infty} A^i\mathbf{p}_0$ for any $\mathbf{p}_0 \in \Delta$.

As a way to compute the fixed point this is probably less efficient than linear algebra but it still gives us useful information: No matter what the initial probability distribution is it will tend in the limit to the unique stable distribution.

Irreducibility

The example $A = I$ shows the need for some assumption on A beyond simply being Markov order to get a unique fixed point. Positivity of all the entries is sufficient condition, but unnecessarily strong, and fails in many interesting examples.

We say that A is *irreducible* if for every j and k there is some $m \geq 0$ such that the probability of a transition from the j'th state to the k'th state in m steps is positive.

It's not hard to see that if there is such an m then there is one with $m < n$.

Products of Markov matrices are Markov, as are averages of Markov matrices, so

$$B = \frac{1}{n} \sum_{i=0}^{n-1} A^i$$

is a Markov matrix.

If A is irreducible then all entries of B are positive, so we can apply the previous theorem to B to show that there is at most one fixed point in Δ.

All fixed points of A are fixed points of B, so there's at most one fixed point of A. The argument using Brouwer already showed there is at least one, so there is exactly one.