Markov chains via generating functions

Peter Doyle

Version 2A2 dated 27 June 2016

Abstract

We develop the theory of discrete-time finite-state Markov chains using generating functions. Discovering that this was anticipated by Feller took the wind out of our sails. But there is still much of interest here, including three Projects begging to be undertaken.

1 Introduction

A discrete-time finite-state Markov chain can be represented by an n-by-n square matrix P. Many questions about the behavior of the chain can be answered using the generating function for the powers of P.

2 The generating function for powers of a matrix

We begin by studying powers of an arbitrary square matrix, over any field. To emphasize this, let's call our matrix A instead of P.

We introduce the *generating function*

$$F = I + xA + x^2A^2 + \ldots = (I - xA)^{-1}.$$

F is a matrix of formal power series in x. These power series all represent rational functions of degree n:

$$F = \frac{1}{\det(I - xA)} \mathrm{Adj}(I - xA).$$

The numerators here are the entries of $\mathrm{Adj}(I - xA)$, the adjoint matrix (or transposed cofactor matrix, or whatever you call it) of $I - xA$. They are polynomials of degree $n - 1$. The common denominator $\det(I - xA)$ is a polynomial of degree n. We get it from the standard characteristic polynomial

$$\chi_A(x) = \det(xI - a) = x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} - \ldots + (-1)^n \sigma_n$$

by reversing the list of coefficients:

$$\det(I - xA) = x^n \chi_A(\frac{1}{x}) = 1 - \sigma_1 x + \sigma_2 x^2 - \ldots + (-1)^n \sigma_n x^n.$$

This means that the roots of $\det(I - xA)$ are the reciprocals of the roots of $\chi_A(x)$, which is to say that they are the reciprocals of the eigenvalues of A. But we hasten to say that throughout this discussion we will only use eigenvalues as a heuristic, never as a tool.

The relationship of the denominator of F to $\chi_A(x)$ arises because the generating function

$$F = (I - xA)^{-1} = \frac{1}{\det(I - xA)} \mathrm{Adj}(I - xA)$$

is a close relative of the *resolvent*

$$R = (xI - A)^{-1} = \frac{1}{\det(xI - A)} \mathrm{Adj}(xI - A),$$

whose denominator is

$$\chi_A(x) = \det(xI - a).$$

Specifically, we have

$$F = \frac{1}{x} R(\frac{1}{x})$$

and

$$R = \frac{1}{x} F(\frac{1}{x}).$$

3 The Cayley-Hamilton theorem

Rewrite the equation

$$I + xA + x^2A^2 + \ldots = \frac{1}{\det(I - xA)}\mathrm{Adj}(I - xA)$$

as

$$\det(I - xA)(I + xA + x^2A^2 + \ldots) = \mathrm{Adj}(I - xA).$$

On the left, we have a matrix of formal power series. On the right, we have a matrix of polynomials of degree $n - 1$. Look at the coefficient of x^n. On the right we get the 0 matrix. On the left we get

$$A^n - \sigma_1 A^{n-1} + \sigma_2 A^{n-2} - \ldots + (-1)^n \sigma_n I,$$

which is the result of substituting A into its characteristic polynomial χ_A. This proves the familiar Cayley-Hamilton theorem.

The Cayley-Hamilton Theorem.

$$\chi_A(A) = 0,$$

where 0 here represents the 0 matrix. ∎

Note. We can easily convert the algebraic proof we've just given of the Cayley-Hamilton theorem into a *bijective proof*, whereby the vanishing of $\chi_A(A)$ is demonstrated by explicit cancellation of terms. See section 6 below for an outline.

4 Recurrence relation

The sequence of matrices $I, A, A^2, \ldots$ satisfies the linear recurrence relation

$$A^N = \sigma_1 A^{n-1} - \sigma_2 A^{N-2} + \ldots + (-1)^n A^{N-n}$$

for all $N \geq n$. This means that the individual entries of $I, A, A^2, \ldots$ satisfy a common linear recurrence. (As we could have seen by noting that their generating functions are rational with a common denominator.) If we look at the ij entry of the sequence $I, A, A^2, \ldots, A^{2n-1}$, we'll be able to deduce the recurrence. If A is sufficiently general, there will be only one degree-n recurrence that will work, and this will tell us the characteristic polynomial of A.

We can do better if we look at the traces, which satisfy the same linear recurrence. That's because the traces of $I, A, A^2, \ldots$ are determined entirely by the characteristic polynomial, while individual entries are not. The generating function for the sequence of traces is

$$\mathrm{Tr}(F) = n + x\mathrm{Tr}(A) + x^2\mathrm{Tr}(A^2) + \ldots.$$

Write this as

$$\mathrm{Tr}(F) = n + s_1 x + s_2 x^2 + \ldots.$$

The coefficients $s_1 = \sigma_1, s_2, \ldots$ represent the power sums of the eigenvalues of A, and we know very well that the initial subsequence $n, s_1, s_2, \ldots, s_n$ determines $\sigma_1, \sigma_2, \ldots, \sigma_n$ and vice versa. But there is no need to appeal to authority here, or ever to mention eigenvalues.

$$\mathrm{Tr}(F) = \frac{1}{\det(I - xA)}\mathrm{Tr}(\mathrm{Adj}(I - xA)).$$

$$\begin{aligned}
\mathrm{Tr}(\mathrm{Adj}(I - xA)) &= \mathrm{Tr}_{n-1}(I - xA) \\
&= n - (n-1)\sigma_1 x + (n-2)\sigma_2 x^2 + \ldots + (-1)^{n-1}\sigma_{n-1}x^{n-1} \\
&= (n - x\frac{d}{dx})\det(I - xA).
\end{aligned}$$

$$\mathrm{Tr}(F) = n - x\frac{d}{dx}\log\det(I - xA).$$

$$\frac{d}{dx}\log\det(I - xA) = \frac{1}{x}(\mathrm{Tr}(F) - n) = s_1 x + s_2 x^2 + \ldots.$$

$$\det(I - xA) = \exp(\int_0^x \frac{1}{x}(F - n)dx) = \exp(s_1 x + \frac{s_2}{2}x^2 + \frac{s_3}{3}x^3 + \ldots).$$

Thus

$$1 - \sigma_1 x + \sigma_2 x^2 - \ldots + (-1)^n \sigma_n x^n = \exp(s_1 x + \frac{s_2}{2}x^2 + \frac{s_3}{3}x^3 + \ldots)$$

Truncating this identity of formal power series expresses the sequence $\sigma_1, \ldots, \sigma_n$ in terms of $s_1, \ldots, s_n$.

Project 1 *This identity is begging for a bijective proof, or at least a combinatorial proof: It will likely be necessary to split terms and then recombine them.*

5 The generating function of a Markov chain

Let P be an n-by-n matrix of non-negative real numbers whose rows sum to 1. Such a matrix is called a *stochastic matrix*. It represents the transition mechanism for a Markov chain, with P_{ij} being the probability of moving from state i to state j. The generating function of P is

$$F = (I - xP)^{-1}.$$

The key to understanding the long-run behavior of the chain is to look at the Laurent expansion of the generating function $F = (I - xP)^{-1}$ at $x = 1$.

Since P is stochastic, 1 is an eigenvalue of P:

$$P1 = 1.$$

This means that the generating function F will have a simple pole at $x = 1$. Of course F is a matrix, so what we mean here is that each entry of the matrix has (at most) a simple pole at $x = 1$. We denote by

$$P^\infty = (1 - x)F(x)|_{x=0}$$

the residue of the pole at $x = 1$ (or rather, its negative). The notation P^∞ suggests the limit of $I, P, P^2, \ldots$, which is justified because when we plug in $x = 1$ after simplifying $(x-1)F(x)$ we're effectively taking the limit at $x = 1$ of

$$(1 - x)F(x) = I + x(P - I) + x^2(P^2 - P) + \ldots.$$

This corresponds to rewriting the sequence $I, P, P^2, \ldots$ as the telescoping series

$$I + (P - I) + (P^2 - P) + \ldots$$

and then using Abel summation. This sum will be genuinely convergent if the chain is *eventually aperiodic*, meaning that it is aperiodic when restricted to any terminal strongly connected component of the state transition graph. Such chains include (strictly) aperiodic chains and absorbing chains. But the great thing is that P^∞ works as we expect for the limit of $I, P, P^2, \ldots$ even when this limit doesn't exist in the naive sense.

Looking more closely at $F(x)$ near $x = 1$, we write

$$F(x) = \frac{1}{1 - x}P^\infty + N + O(x - 1).$$

The $O(x-1)$ here means an identity up to order $x-1$ between formal power series in $x-1$. The matrix

$$\begin{aligned} N &= (1-x)F(x)|_{x=1} \\ &= ((I-P^\infty)+x(P-P^\infty)+x^2(P^2-P^\infty)+\ldots))|_{x=1} \\ &= (I-(P-P^\infty))^{-1}-P^\infty \end{aligned}$$

is the *Green's function*. So here we have another Abel summation. If P is eventually aperiodic, then the sum will actually converge to N, and we'll have

$$N = (I-P^\infty)+(P-P^\infty)+(P^2-P^\infty)+\ldots,$$

which is how we want to think about the Green's function even if the sum doesn't converge without help from Abel.

In the case of an ergodic chain, we can think of N as telling the excess expected visits to j for a chain starting from i compared to a chain starting from equilibrium. The same goes for a chain with some transient states but only one ergodic component. When there is more than one ergodic component, we weight these excess visits by the probability of landing in the various ergodic components.

6 Bijective proof of the Cayley-Hamilton theorem

As always,

$$F = (I-xA)^{-1}.$$

The trick here is to find a bijective proof of the matrix identity

$$F\Delta = \mathrm{Adj}(I-xA)$$

where $\Delta = \det(I-xA)I$, i.e, the diagonal matrix with all diagonal entries equal to $\det(I-xA)$. Looking at the coefficient of x^n will give a proof of Cayley-Hamilton.

Take the obvious (partial) bijection showing that

$$\Delta = (I-xA)\mathrm{Adj}(I-xA)$$

(row expansion of the determinant, plus vanishing of the determinant when two rows are equal). Multiplying on the left by F promotes this bijection in a trivial way to a bijection showing that

$$F\Delta = F(I - xA)\mathrm{Adj}(I - xA).$$

Similarly, take the obvious bijection showing

$$F(I - xA) = I$$

(telescoping sum) and multiply on the right by $\mathrm{Adj}(I - xA)$ to get a bijection showing that

$$F(I - xA)\mathrm{Adj}(I - xA) = \mathrm{Adj}(I - xA).$$

Now we have bijections showing that

$$F\Delta = F(I - xA)\mathrm{Adj}(I - xA) = \mathrm{Adj}(I - xA).$$

By 'subtraction', or 'cobordism', or 'the involution principle', or whatever you want to call it, these combine to give a bijective proof that

$$F\Delta = \mathrm{Adj}(I - xA).$$

Looking at the coefficient of x^n gives a bijective proof of Cayley-Hamilton. The bijection that emerges turns out to be (modulo trivial differences) the bijection found by Howard Straubing [1] and described by Doron Zeilberger in his paper on a combinatorial approach to matrix algebra [2].

Project 2 *I would love to see someone write out the details of this. Even better would be to have a computer system that would implement automatically the transition from algebraic proof to bijective proof. This program would have many applications to problems in combinatorial matrix algebra, for example in extending Sally Picciotto's work on the UCSD code for trees.*

References

[1] H. Straubing. A combinatorial proof of the Cayley-Hamilton theorem. *Discrete Mathematics*, 43:273–279, 1983.

[2] Doron Zeilberger. A combinatorial approach to matrix algebra. *Discrete Mathematics*, 56:61–72, 1985.