Matrix exponentials

Markov chains

18. maj 2020

$$\mathrm{Mat}_{\varphi \text{ to } M} = \begin{array}{c c} & \begin{array}{ccccc} c_1 & c_2 & c_3 & c_4 & c_5 \end{array} \\ \begin{array}{c} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Introduction

In this brief note we will examine why matrix exponentials show up in the theory of continuous time Markov chains, and have a close enough look at the mathematical definition to get a better understanding of them, without getting bogged down in technical details.

We will see that we need to think of the exponential function $\exp(t) = e^t$ as a sort of infinite polynomial, rather than as e raised to the power of t, if we want to understand the matrix exponential. This is because we can plug a matrix A into the infinite polynomial and in that way define e^A, while "raising e to the power of A" makes no sense.

1 Why do matrix exponentials show up in continous time Markov chain theory?

The exponential function on the real line is the function that takes a number t and gives you e^t. You probably know its graph, and some of its most important properties:

- $e^{t+s} = e^t e^s$,
- $\frac{d}{dt}e^{qt} = qe^{qt}$.

The last property is the reason that the solution to the differential equation

$$p'(t) = qp(t) \tag{1}$$

with initial value $p(0) = 1$ is

$$p(t) = e^{qt}.$$

When working with continuous time Markov chains we usually know an $n \times n$ matrix Q of transition rates and would like to know the matrix function $P(t)$ of transition probabilities — that is $P(t)$ is the function that, for a given time t, gives the $n \times n$ matrix whose element on row i and coloumn j is the probability of being in state j of the Markov chain at time t, given that you are in state i at time 0.

The (Kolmogorov) backward differential equation states that P is the solution to the first order linear system of differential equations

$$P'(t) = QP(t) \tag{2}$$

with initial value

$$P(0) = I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

where I is the diagonal matrix with 1s in the diagonal and 0s everywhere else. The backward equations (2) looks a lot like a matrix version of the differential equation (1) and the initial value $P(0) = I$ is even much like the initial value $p(0) = 1$, since

I is the matrix version of 1. We know the solution to (1) as begin simply $p(t) = e^{qt}$ — wouldn't it be nice if we could somehow give meaning to e^{Qt} and hope that $P(t) = e^{Qt}$ solves (2)?

2 So, how to define e^{Qt}?

The problem of course is that it is not at all obvious what it means to raise e to the Qt'th power, since Qt is a matrix. We now move away from thinking of e^t as "e raised to the power of t" and toward thinking of it simply as a function $\exp(t) = e^t$ given by its so-called *Taylor series*.

All functions f that are infinitely often differentiable (which just means that you can keep differentiating it as many times as you want to) has a Taylor series. The Taylor series is the "infinite sum"

$$f(0) + f'(0)t + \frac{f''(0)t^2}{2} + \frac{f'''(0)t^3}{3!} + \frac{f^{(4)}(0)t^4}{4!} + \cdots$$

that is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)t^k}{k!},$$

where $f^{(k)}(0)$ means the k'th derivative of f evaluated in 0. You don't need to worry too much about what an infinite sum (or so-called *series*) is — all you need to know is that they sometime *converge* to a *limit* which is then said to be the sum of the series. As a simple example the series

$$\sum_{k=0}^{\infty} \frac{1}{2^k} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$$

turns out to converge to 2. Rigorously speaking, the statement

$$\sum_{k=0}^{\infty} \frac{1}{2^k} = 2,$$

means that you can get the *finite* sum

$$\sum_{k=0}^{N} \frac{1}{2^k} = 1 + \frac{1}{2} + \cdots + \frac{1}{2^N}$$

as close to 2 as you want to, by just using a large enough N. So $\sum_{k=0}^{\infty} \frac{1}{2^k}$ is the limit of $\sum_{k=0}^{N} \frac{1}{2^k}$ when N goes to infinity ($N \to \infty$). You can try adding some of the terms of the series on your calculator to convince yourself that this is actually the case.

Sometimes the Taylor series of a function converges to the function itself, so that for all numbers t we have

$$f(t) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)t^k}{k!}.$$

This is in fact the case with the exponential function, that is

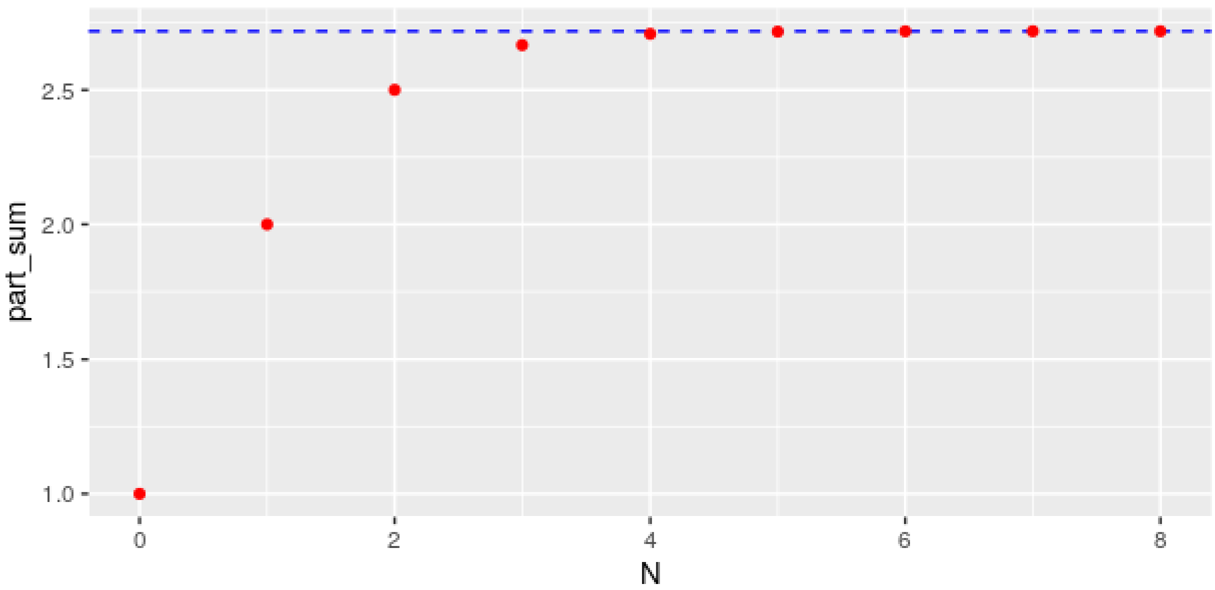
$$\exp(t) = \sum_{k=0}^{\infty} \frac{\exp^{(k)}(0)t^k}{k!}.$$

Since exp differentiated is still just exp (that is $\frac{d}{dt}e^t = e^t$) and $\exp(0) = 1$ we have $\exp^{(k)}(0) = 1$ for all k (you keep getting exp all the k times as you calculate the derivatives) the Taylor series is

$$e^t = \sum_{k=0}^{\infty} \frac{t^n}{n!} = 1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \cdots$$

In fact this is how the exponential function is usually defined in rigorous mathematics. To visualize the concept of convergence, we can check the above convergence for $t = 1$ in R. The statement then is that

$$\begin{aligned} e = e^1 &= 1 + 1 + \frac{1^2}{2} + \frac{1^3}{3!} + \cdots \\ &= 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \cdots \end{aligned}$$



Figur 1: Convergence of the Taylor series for exp in $t = 1$. The points show the sum of the first N terms in the Taylor series, while the line shows $e \approx 2.718$.

It is a mathematical result that you can differentiate term by term in some series (including the exponential Taylor series) which gives us the proof that $\frac{d}{dt}e^t = e^t$:

$$\begin{aligned}
\frac{d}{dt}e^t &= \frac{d}{dt}\left(1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots\right) \\
&= \frac{d}{dt}1 + \frac{d}{dt}t + \frac{d}{dt}\frac{t^2}{2} + \frac{d}{dt}\frac{t^3}{3!} + \frac{d}{dt}\frac{t^4}{4!} + \cdots \\
&= 0 + 1 + \frac{2t}{2} + \frac{3t^2}{3!} + \frac{4t^3}{4!} + \cdots \\
&= 1 + t + \frac{t^2}{2} + \frac{t^3}{3!} + \cdots \\
&= e^t.
\end{aligned}$$

We are finally ready to define the exponential of a matrix. For an $n \times n$ matrix A, the matrix exponential $\exp(A)$ or e^A is defined as plugging in A in the Taylor series for the usual exponential function of real numbers:

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots .$$

See the next section of this note for examples of the definition in use — this will help with the understanding. If this was to be done properly we would have to prove that this always converges, but we will contend ourselves with mentioning that this can be proven. This definition means that our guess for a solution of the backwards differential equations is

$$P(t) = e^{Qt} = I + Qt + \frac{(Qt)^2}{2} + \frac{(Qt)^3}{3!} \cdots = I + Qt + \frac{Q^2t^2}{2} + \frac{Q^3t^3}{3!} + \cdots$$

We can differentiate this term by term (this would technically also require further justification) to get that

$$\begin{aligned} P'(t) &= \frac{d}{dt}I + \frac{d}{dt}Qt + \frac{d}{dt}\frac{Q^2t^2}{2} + \frac{d}{dt}\frac{Q^3t^3}{3!} + \cdots \\ &= 0 + Q + \frac{2Q^2t}{2} + \frac{3Q^3t^2}{3!} + \cdots \\ &= QI + QQt + Q\frac{Q^2t^2}{2} + \cdots \\ &= Q(I + Qt + \frac{Q^2t^2}{2} + \cdots) \\ &= Qe^{tQ}. \end{aligned}$$

We even see that (here 0 denotes the 0 matrix with all zero entries):

$$P(0) = e^{Q0} = e^0 = I + 0 + \frac{0^2}{2} + \cdots = I,$$

so $P(t) = e^{Qt}$ is in fact a solution to the backwards differential equation. A mathematical result on uniqueness of solutions to certain types of differential equations gives us that $P(t) = e^{Qt}$ is in fact *the* solution to the backwards differential equation.

3 Three examples of calculating matrix exponentials by hand

Matrix exponentials will usually be found using software such as R, but in simple cases they can be calculated using pencil and paper. We will examine two examples to illustrate the definition.

Example 1. Consider the 2×2 matrix

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}.$$

The reader can convince herself that

$$A^k = \begin{pmatrix} 3^k & 0 \\ 0 & 2^k \end{pmatrix}.$$

Hence, using the definition of a matrix exponential, we get that

$$\begin{aligned} e^A &= I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \cdots \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} \frac{3^2}{2} & 0 \\ 0 & \frac{2^2}{2} \end{pmatrix} + \begin{pmatrix} \frac{3^3}{3!} & 0 \\ 0 & \frac{2^3}{3!} \end{pmatrix} + \cdots \\ &= \begin{pmatrix} 1 + 3 + \frac{3^2}{2} + \frac{3^3}{3!} + \cdots & 0 \\ 0 & 1 + 2 + \frac{2^2}{2} + \frac{2^3}{3!} + \cdots \end{pmatrix} \\ &= \begin{pmatrix} e^3 & 0 \\ 0 & e^2 \end{pmatrix}. \end{aligned}$$

In general, the matrix exponential of a diagonal matrix is just the diagonal matrix with the usual exponential function applied on the diagonal elements. As a small exercise, the reader can try finding the matrix exponential of

$$\begin{pmatrix} 7 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}.$$

Example 2. Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Since $A^2 = 0$ (the zero matrix), and thus $A^k = 0$ for $k \geq 2$, we have

$$e^A = I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \cdots = I + A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Example 3. Consider the matrix

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix},$$

which is the transition matrix of the two state Markov chain with states i_1 and i_2 with transition rate 1 from i_1 to i_2 and also rate 1 from i_2 to i_1. To find $P(t)$ we wish to calculate e^{Qt}. It can be checked that with

$$B = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$$

we have

$$Qt = B^{-1} \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} tB$$

and hence

$$\begin{aligned} (Qt)^k &= \left(B^{-1} \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} tB \right)^k \\ &= B^{-1} \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} tBB^{-1} \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} tB \cdots B^{-1} \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} tB \\ &= B^{-1} \begin{pmatrix} (-2)^k & 0 \\ 0 & 0 \end{pmatrix} t^k B \end{aligned}$$

since the Bs and B^{-1}s in the middle cancel out.

Using the definition of matrix exponentials we now get

$$\begin{aligned}
e^{Qt} &= I + Qt + \frac{(Qt)^2}{2} + \cdots \\
&= I + B^{-1}\begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} tB + B^{-1}\frac{1}{2}\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} t^2 B + \cdots \\
&= B^{-1}\left(I + \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} t + \frac{1}{2}\begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix}^2 t^2 + \cdots \right) B \\
&= B^{-1}\begin{pmatrix} e^{-2t} & 0 \\ 0 & 1 \end{pmatrix} B \\
&= \begin{pmatrix} \frac{1}{2} + \frac{1}{2}e^{-2t} & \frac{1}{2} - \frac{1}{2}e^{-2t} \\ \frac{1}{2} - \frac{1}{2}e^{-2t} & \frac{1}{2} + \frac{1}{2}e^{-2t} \end{pmatrix}.
\end{aligned}$$

Therefore

$$P(t) = e^{Qt} = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}e^{-2t} & \frac{1}{2} - \frac{1}{2}e^{-2t} \\ \frac{1}{2} - \frac{1}{2}e^{-2t} & \frac{1}{2} + \frac{1}{2}e^{-2t} \end{pmatrix}.$$