z -Transforms for Stochastic Finite Automata

Jim Crutchfield

Physics Department
University of California
Berkeley, California 94720
chaos@gojira.berkeley.edu

A short review, in our notation, of methods for computing iterates of the transition matrix and state probabilities as a function of time for finitary stochastic machines.

1 Definitions

Consider a finitary stochastic machine $\mathrm{M} = (\mathbf{V}, \mathbf{E}, \mathcal{A}, \mathbf{T})$ with

1. $\mathbf{V} = \{v\}$ the set of states. Typically, we identify a particular state v_0 as the start state.
2. $\mathcal{A} = \{s\}$ is a set of symbols (i) emitted by the machine on taking an edge or (ii), in recognize-mode, that select a transition to be taken when reading a string composed of these symbols.
3. $\mathbf{E} = \left\{ e_{v \underset{s}{\rightarrow} v'} : v, v' \in \mathbf{V}, s \in \mathcal{A} \right\}$ the set of labeled edges or transitions between the states.
4. $\mathbf{T} = \left\{ T^{(s)} : s \in \mathcal{A} \right\}$ is a set of labeled, conditional transition matrices, where the probability $p_{vv's}$ of making a transition to state v' conditioned on being in state v and seeing symbol s is given by the matrix elements

$$p_{vv's} = \left(T^{(s)} \right)_{vv'} \tag{1}$$

The connection matrix is given by $T = \sum_{s \in \mathcal{A}} T^{(s)}$, with $p_{vv'} = (T)_{vv'}$. Note that it is a stochastic matrix

$$\sum_{v' \in \mathbf{V}} p_{vv'} = 1 \tag{2}$$

The state probability distribution at time t is denoted $p_{\mathbf{V}}(t) = \{p_{v_0}(t), p_{v_1}(t), \ldots\}$. It is normalized, $\sum_{v \in V} p_v(t) = 1$. Starting from an initial state distribution $p_{\mathbf{V}}(0)$, the temporal evolution is given by the connection matrix

$$p_{\mathbf{V}}(t+1) = p_{\mathbf{V}}(t) T \tag{3}$$

Note the left multiplication.

Iterating Eq. 3 over time gives the state distribution in terms of the initial distribution

$$p_{\mathbf{V}}(t) = p_{\mathbf{V}}(0) T^t \tag{4}$$

where T^t is the t-th iterate of the connection matrix.

Note that if we start the machine with the distribution $p_V(0) = (1, 0, 0, \ldots)$ (with an ϵ-machine this corresponds to the condition of total ignorance, i.e. beginning in the start state and with zero probabilistic information), then the first row of T^t is $p_{\mathbf{V}}(t)$.

Eq. 4 indicates that if we can compute T^t, then $p_{\mathbf{V}}(t)$ is given for all times. This is what the z-transform allows one to do; though there are a good number of other things one can do with it.

2 z-Transform

The z-transform of the state distribution is denoted

$$q_{\mathbf{V}}(z) = \mathcal{Z}(p_{\mathbf{V}}(t)) \tag{5}$$

and is defined by

$$\mathcal{Z}(p_{\mathbf{V}}(t)) = \sum_{t=0}^{\infty} p_{\mathbf{V}}(t) z^{-t} \tag{6}$$

The notation here assumes that the $\mathcal{Z}$ operator is applied component-wise to the vector $p_{\mathbf{V}}(t)$. The same will also be assumed to hold for matrices. Although time is discrete, the formal variable z is continuous and can be complex.

The inverse z-transform is denoted

$$p_{\mathbf{V}}(t) = \mathcal{Z}^{-1}(q_{\mathbf{V}}(z)) \tag{7}$$

and is given by

$$\mathcal{Z}^{-1}(q_{\mathbf{V}}(z)) = (2\pi i)^{-1} \int_{-\infty}^{\infty} dz q_{\mathbf{V}}(z) z^{t-1} \tag{8}$$

Computing the z-transform typically involves summing series; whereas, its inverse requires contour integration or, at least for those processes whose $q_{\mathbf{V}}(z)$ is a rational function, recalling Cauchy's theorem.

Taking the z-transform of Eq. 3 gives

$$\begin{aligned} \sum_{t=0}^{\infty} p_{\mathbf{V}}(t+1) z^{-t} &= \sum_{t=0}^{\infty} p_{\mathbf{V}}(t) T z^{-t} \\ \sum_{t=1}^{\infty} p_{\mathbf{V}}(t) z^{-(t-1)} &= q_{\mathbf{V}}(z) T \end{aligned} \tag{9}$$

$$\begin{aligned} z\left(\sum_{t=0}^{\infty} p_{\mathbf{V}}(t) z^{-t} - p_{\mathbf{V}}(0)\right) &= q_{\mathbf{V}}(z) T \\ z(q_{\mathbf{V}}(z) - p_{\mathbf{V}}(0)) &= q_{\mathbf{V}}(z) T \end{aligned} \tag{10}$$

Rearranging

$$\begin{aligned} z q_{\mathbf{V}}(z) - q_{\mathbf{V}}(z) T &= z p_{\mathbf{V}}(0) \\ q_{\mathbf{V}}(z)\big(I - z^{-1} T\big) &= p_{\mathbf{V}}(0) \end{aligned} \tag{11}$$

where I is the identity matrix. Thus, the z-transform of the state probability vector $p_{\mathbf{V}}(t)$ is given by

$$q_{\mathbf{V}}(z) = p_{\mathbf{V}}(0)\big(I - z^{-1} T\big)^{-1} \tag{12}$$

3 Response Matrix

Define the $\|\mathbf{V}\| \times \|\mathbf{V}\|$ matrix $\mathcal{T}(z) = \left(I - z^{-1}T\right)^{-1}$. Its inverse z-transform $\mathcal{Z}^{-1}$ is the so-called response matrix

$$R(t) = \mathcal{Z}^{-1}(\mathcal{T}(z)) \tag{13}$$

In terms of the response matrix the inverse z-transform of Eq. 12 is

$$p_{\mathbf{V}}(t) = \mathcal{Z}^{-1}(q_{\mathbf{V}}(z)) = p_{\mathbf{V}}(0)R(t) \tag{14}$$

Comparing this with Eq. 4, the response matrix is seen to be $R(t) = T^t$.

Thus, Eq. 13 gives the method for computing the powers of a stochastic matrix. Note that $(R(t))_{vv'}$ is the probability that at time t the system starting in state v at time $t = 0$ will go to state v' at time t.

The response matrix breaks into a transient T piece and a time-independent (asymptotic) piece A

$$R(t) = \mathrm{A} + \mathrm{T}(t) \tag{15}$$

For the probability distribution then we have

$$p_{\mathbf{V}}(t) = p_{\mathbf{V}}(0)\mathrm{A} + p_{\mathbf{V}}(0)\mathrm{T}(t) \tag{16}$$

There are several properties of the asymptotic matrix that depend on the structure of the connection matrix T

1. If T is nonrecurrent, i.e. all transient states, or is simply recurrent, i.e. a single strongly connected component, then the rows of A are identical and equal to $p_{\mathbf{V}}(\infty)$.
2. If T is multiple recurrent, i.e. with several strongly connected components, then the rows of A are not equal and $p_{\mathbf{V}}(\infty)$ depends on the initial distribution $p_{\mathbf{V}}(0)$. If one starts in state v with $p_{\mathbf{V}}(0) = (0, 0, \ldots, p_v = 1, \ldots, 0)$, however, the v-th row of A is $p_{\mathbf{V}}(\infty)$.

There are several notable properties of the transient matrix

1. $\mathrm{T}(t) \rightarrow 0, \text{ as } \quad t \rightarrow \infty$.
2. The row sums are zero: $\sum\limits_{v \in \mathbf{V}} (\mathrm{T}(t))_{vv'} = 0$. One thinks of each row as a perturbation of the associated $p_{\mathbf{V}}(\infty)$, which must normalized.
3. The v-th row of T is the set of transient components of $p_{\mathbf{V}}(t)$ starting in state v.

4 z-Transform Properties

There is a long list of properties that (i) are useful in computations and (ii) give some indication of what the z-transform is telling one. For example,

1. Relation to discrete-time Fourier transform $F(f) = \mathcal{F}(p_{\mathbf{V}}(t)) = q_{\mathbf{V}}\left(z = e^{2\pi i f}\right)$. Thus, restricting the z-transform to the unit circle gives the Fourier transform.

2. Linearity: $\mathcal{Z}\big(ap_{\mathbf{V}}(t) + bp'_{\mathbf{V}}(t)\big) = aq_{\mathbf{V}}(z) + bq'_{\mathbf{V}}(z)$.
3. Time shift: $\mathcal{Z}(p_{\mathbf{V}}(t-k)) = z^{-k}q_{\mathbf{V}}(z)$.
4. Scaling: $a^{-t}p_{\mathbf{V}}(t) = q_{\mathbf{V}}(az), a > 0$.
5. Conjugation: $p_{\mathbf{V}}(t) = q^*_{\mathbf{V}}(z^*)$.
6. Time reversal: $p_{\mathbf{V}}(-t) = q_{\mathbf{V}}\big(\frac{1}{z}\big)$ and $p^*_{\mathbf{V}}(-t) = q^*_{\mathbf{V}}\big(\frac{1}{z^*}\big)$.
7. Convolution: $p_{\mathbf{V}}(t) \star g(t) = q_{\mathbf{V}}(z) \cdot G(z)$.

5 Examples

For a machine with a single recurrent state $p_V(t) = 1$ and the z-transform is

$$q_V(z) = \mathcal{Z}(1) = \sum_{t=0}^{\infty} z^{-t} = \frac{1}{1 - z^{-1}} \tag{17}$$

Going back the other way we ask for

$$\begin{aligned} \mathcal{Z}^{-1}\left(\frac{1}{1-z^{-1}}\right) &= (2\pi i)^{-1} \int_{-\infty}^{\infty} dz \frac{z^{t-1}}{1-z^{-1}} \\ &= (2\pi i)^{-1} \int_{-\infty}^{\infty} dz \frac{z^{t}}{z-1} \\ &= (2\pi i)^{-1} (2\pi i \cdot 1) = 1 \end{aligned} \tag{18}$$

The second line follows from the Residue theorem with the 1 in the second factor coming from the simple pole at $z = 1$, which has a residue of 1. The Residue theorem says that the value of the integral is

$$2\pi i \sum_{\{z_0 : \mathrm{poles}(f)\}} \mathrm{Res}(f(z), z_0), \tag{19}$$

where $\mathrm{Res}(f(z), z)$ is the residue of the integrand evaluated at z. Typically, with z-transforms the integrand looks like $f(z) = \frac{g(z)}{z-z_0}$ and, in this case, $\mathrm{Res}(f(z), z_0) = g(z_0)$.

Let's focus on a transient part of a machine: a state for which half its probability stays and half leaks away. Then we know that its probability decays exponentially $p_v(t) = 2^{-t}$. Then we have

$$q_V(z) = \mathcal{Z}\big(2^{-t}\big) = \sum_{t=0}^{\infty} 2^{-t}z^{-t} = \frac{1}{1-(2z)^{-1}} = \frac{2}{2 - z^{-1}} \tag{20}$$

(Try computing the inverse.)

For the third example, consider a bona fide machine with a transient state like that just considered and two simple recurrent components. The connection matrix is

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{21}$$

Then we have the matrix

$$\left(I - z^{-1}T\right) = \begin{pmatrix} 1 - \frac{1}{2}z^{-1} & -\frac{1}{4}z^{-1} & -\frac{1}{4}z^{-1} \\ 0 & 1 - z^{-1} & 0 \\ 0 & 0 & 1 - z^{-1} \end{pmatrix} \tag{22}$$

Its inverse is

$$\mathcal{T}(z) = \left(I - z^{-1}T\right)^{-1} = \begin{pmatrix} \frac{2}{2-z^{-1}} & \frac{z^{-1}}{2(1-z^{-1})(2-z^{-1})} & \frac{z^{-1}}{2(1-z^{-1})(2-z^{-1})} \\ 0 & \frac{1}{1-z^{-1}} & 0 \\ 0 & 0 & \frac{1}{1-z^{-1}} \end{pmatrix} \tag{23}$$

or

$$\mathcal{T}(z) = \frac{1}{1 - z^{-1}} \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{2}{2 - z^{-1}} \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{24}$$

The response function is then

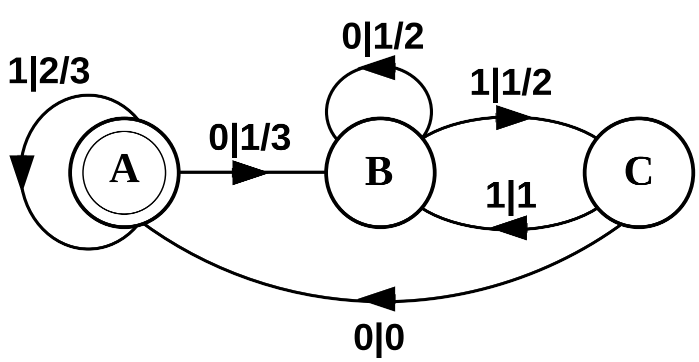
$$R(t) = \mathcal{Z}^{-1}(\mathcal{T}(z)) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 2^{-t} \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{25}$$

Note that the inverse transform used to get to this last line was already done in the first two examples.

The state probability vector for starting in the first state is $p_{\mathbf{V}}(t) = \left(2^{-t}, 2^{-1}\left(1 - 2^{-t}\right), 2^{-1}\left(1 - 2^{-t}\right)\right)$. If the machine started in state 2, the probabilities would be $p_{\mathbf{V}}(t) = (0, 1, 0)$; in state 3, $p_{\mathbf{V}}(t) = (0, 0, 1)$. The latter two distributions are also the asymptotic ones for starting in those states. The asymptotic state probabilities for starting in the first state, though, are given by the first row of the time independent matrix$p_{\mathbf{V}}(\infty) = \left(0, \frac{1}{2}, \frac{1}{2}\right)$.

A final and more "realistic" example: See "Knowledge and Meaning ... Chaos and Complexity", from which the following example is taken, for a discussion of the use of computing the state probabilities as a function of time: how an observer comes to know of the state of a process and what meaning an observer ascribes to a given measurement.

Figure 1 The even system generates sequences $\{\ldots 01^{2n}0 \ldots : n = 0, 1, 2, \ldots\}$ of 1s of even length, i.e. even parity. There are three states $\mathbf{V} = \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$. The state A with the inscribed circle is the start state v_0. The edges are labeled $s|p$ where $s \in \mathbf{A}$ is a measurement symbol and $p \in [0, 1]$ is a conditional transition probability.



Consider the even system: the three state stochastic automaton shown in Figure 1. Here we consider only the recurrent state probabilities. They are updated via the stochastic connection matrix

$$\vec{p}_{\mathbf{V}}(t+1) = \vec{p}_{\mathbf{V}}(t)\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix} \tag{26}$$

where $\vec{p}_{\mathbf{V}}(t) = (p_{\mathrm{B}}(t), p_{\mathrm{C}}(t))$ and the initial distribution is $\vec{p}_{\mathbf{V}}(0) = (1,0)$. Using the z-transform, we first find that

$$I - z^{-1}T = \begin{pmatrix} 1 - \frac{1}{2}z^{-1} & -\frac{1}{2}z^{-1} \\ -z^{-1} & 1 \end{pmatrix} \tag{27}$$

We then determine this matrix's inverse, which is

$$\mathcal{T}(z) = \left(I - z^{-1}T\right) = \begin{pmatrix} \frac{2z^2}{(2z+1)(z-1)} & \frac{z}{(2z+1)(z-1)} \\ \frac{2z}{(2z+1)(z-1)} & \frac{2z^2-z}{(2z+1)(z-1)} \end{pmatrix} \tag{28}$$

The response matrix is

$$R(t) = \begin{pmatrix} \mathcal{Z}^{-1}(\mathcal{T}_{00}) & \mathcal{Z}^{-1}(\mathcal{T}_{01}) \\ \mathcal{Z}^{-1}(\mathcal{T}_{10}) & \mathcal{Z}^{-1}(\mathcal{T}_{11}) \end{pmatrix} \tag{29}$$

Applying the Residue Theorem to determine the integral for each component, we find

$$R(t) = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} + \frac{1}{3}\left(-\frac{1}{2}\right)^t \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix} \tag{30}$$

If the initial distribution is $p_{\mathbf{V}}(0) = (1,0)$, then the result then gives the time-dependent state probabilities as

$$\begin{aligned} p_{\mathrm{B}}(t) &= \frac{2}{3} + \frac{1}{3}\left(-\frac{1}{2}\right)^t \qquad t = 0,1,2,\ldots \\ p_{\mathrm{C}}(t) &= \frac{1}{3} - \frac{1}{3}\left(-\frac{1}{2}\right)^t \qquad t = 0,1,2,\ldots \end{aligned} \tag{31}$$

At infinite time $p_{\mathbf{V}}(\infty) = \left(\frac{2}{3}, \frac{1}{3}\right)$.