9.20 Markov chains debug: markov-chains.tex

Suppose you observe where your friend John eats his lunch in a week. Let X_0 be the rv of what John eats at time $t = 0$ (say $t = 0$ means Sunday). Perhaps

$$\Pr[X_0 = b] = 0.3, \Pr[X_0 = m] = 0.1, \Pr[X_0 = w] = 0.6$$

The sample space is

$$S = \{b = \text{BURGER KING}, \;\; m = \text{MACDONALDS}, \;\; w = \text{WENDY'S}\}^7$$

For time $t = 1$, suppose

$$\Pr[X_1 = b] = 0.2, \Pr[X_1 = m] = 0.0, \Pr[X_1 = w] = 0.8$$

etc. You can think of John as going through a sequence of decisions: what to eat at time $t = 0$, what to eat at time $t = 1$, etc. The decision (of what to eat at time t) is not deterministic but depends on a probability. It is a stochastic process. Here, $I = \{0, 1, 2, ..., 6\}$ (where $0 = \text{SUNDAY}, ...$) One of the outcomes in S is

$$(m, b, b, m, w, b, m)$$

and

$$\begin{aligned}
X_0((m, b, b, m, w, b, m)) &= m \\
X_1((m, b, b, m, w, b, m)) &= b \\
X_2((m, b, b, m, w, b, m)) &= b
\end{aligned}$$

etc. Note that you don't expect John to eat according to a uniform probability distribution – that's probably the case for most people. For instance perhaps John's favorite restaurant about the three is Macdonalds. Perhaps John like Macdonalds twice as much as the rest.

But ... suppose it's not just that John like Macdonalds about twice as much as the rest, as in the number of times he visit Macdonalds is twice as many as another restaurant. It's John's habit that when he eats at Macdonalds on a day, there's a high probability that the next day he eats at Macdonalds again, but the chance of him eating at Macdonalds three days in a row is low. See the difference? In first case above, the probability

$$p((m, b, m, w, m, b, m))$$

is higher than

$$p((m, b, b, b, b, b, b))$$

and is about the same as

$$p((m, m, b, w, m, m, b))$$

However in the second case above, the probability of

$$p((m, b, m, w, m, b, m))$$

lower than

$$p((m, m, b, w, m, m, b))$$

In the second case, that X_1 depends on X_0 (more accurately $\Pr[X_1 = \bullet]$ depends on $\Pr[X_0 = \bullet]$), X_2 depends on X_1 and X_0, etc.

Let's formalize the above.

A **random/stochastic process** is a sequence of random variables random/stochastic process

$$X_i : S \to V$$

for $i \in I$. The S is a product of V, i.e., $S = V \times V \times ....$ The number of copies of V is the number of elements of I. In real world applications, usually the i runs over a set of values associated with time and I is usually discrete, and frequently I is infinite: $I = \{0, 1, 2, ...\}$.

(The word **stochastic** is Greek and means "guess". In math, stochastic means "based on randomness or probability".) stochastic

The above is not new. For instance you have already seen that a sequence of Bernoulli trials is a random/stochastic process where each X_i is independent of X_j for $i \neq j$. In this case $V = \{\text{SUCCESS}, \text{FAILURE}\} = \{1, 0\}$.

The definition of a random/stochastic process does not assume that X_1 depends on X_0. (Or rather does not assume $\Pr[X_1 = \bullet]$ depend on $\Pr[X_0 = \bullet]$.) For a sequence of Bernoulli trails, X_i and X_j (for $i \neq j$) are independent.

However, for this section, I'm actually interested in the situation where there are some dependencies between the X_i's. This is important. Why? Note that $X_0, X_1, ...$ (or rather $\Pr[X_0 = \bullet], \Pr[X_1 = \bullet], ...$) is coming from the same "system". In the above example, they come from "John", the system that stochastically decides on what to eat, so it's possible that X_1 (or rather the probability of $\Pr[X_1 = \bullet]$) depends on X_0 (or rather the probability of $\Pr[X_0 = \bullet]$). and when I look at $\Pr[X_1 = x_1]$, I might in fact be interested in

$$\Pr[X_1 = x_1 \mid X_0 = x_0]$$

Furthermore it's possible that X_2 depends on X_0, X_1, etc. Therefore in general we might be interested in

$$\Pr[X_i = x_i \mid X_0 = x_0,\ \ X_1 = x_1,\ \ \ldots,\ \ X_{i-1} = x_{i-1}]$$

This is very common. For instance if you think of words in an essay. Each word must have some dependencies on previous words. You cannot possible be interested in the 5th word alone – when looking at the 5th word, you are probably interested in the meaning of this 5th word in the context of the previous 4 words. A (reasonably well–written) piece of music is not made up of random notes only obeying probability of each note – a note must have some dependency of previous notes. Etc.

If X_i depends only on X_{i-1} and not X_j for $0 \le j < i-1$ then we say the random/stochastic process X_i $(i \in I)$ is a **Markov process**. This means that Markov process

$$\Pr[X_i = x_i \mid X_0 = x_0,\ \ X_1 = x_1,\ \ \ldots,\ \ X_{i-1} = x_{i-1}] = \Pr[X_i = x_i \mid X_{i-1} = x_{i-1}]$$

for $i \in I$.

More generally, if X_i only depends on X_{i-1} then we say that $X_i(i \in I)$ form an order 1 Markov process. If X_i only depends on X_{i-1} and X_{i-2}, we have an order 2 Markov process. Etc. I will only talk about order 1 Markov processes in this set of notes. So "Markov process" means "order 1 Markov process".

We say that X_i $(i \in I)$ is a **Markov chain** if V is discrete. Markov chain

Furthermore, X_i $(i \in I)$ is a **discrete-time Markov chain** is V is discrete and I is discrete (remember you want to think of I as time, but mathematically it does not matter whether I is a set of times or not). One more: X_i $(i \in I)$ is a **continuous-time Markov chain** if V is discrete and I is continuous. discrete-time Markov chain continuous-time Markov chain

I will restrict to the case where I is frequently $\{0, 1, 2, ...\}$ and V is a finite set. So for me, from now on, when I say "Markov chain" I mean discrete-time Markov chain where V is finite.

The set V in the context of Markov processes is frequently called the **state space**. Each value of V is called a **state**. This is very similar to automata theory and theory of computation (see CISS362). state space

Example 9.20.1. Suppose in Columbia, whether it rains depends on whether it rained or not the previous day:

- If it rained today, then there's a 80% chance that it will rain tomorrow.
- If it did not rain today, then there's a 60% chance that it will not rain tomorrow.

In this case the states are $V = \{\text{RAIN}, \text{NO-RAIN}\}$.

	RAIN	NO-RAIN
RAIN	0.8	0.2
NO-RAIN	0.4	0.6

I'm going to rewrite the above as a matrix. Let me relabel the states RAIN, NO-RAIN as $0, 1$, of course using a rv X (say). Then the above table becomes

	0	1
0	0.8	0.2
1	0.4	0.6

which can be written as a matrix:

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

probability transition matrix

P is called the **probability transition matrix** of this Markov chain. $P_{00} = 0.8$ is the probability that it will rain tomorrow if it rains today, i.e., the probability that if the system is at state 0, it will transition to state 0. $P_{01} = 0.2$ is the probability that it will not rain tomorrow if it rains today, i.e., the probability that if the system is at state 0, it will transition to state 1. Etc. In general P_{rc} is the probability that the system at state r will transition to state c:

$$r \xrightarrow{P_{rc}} c$$

You'll see that writing the probabilities as the above matrix will help (because this allows me to use matrix operations).

In general, if a Markov chain has n states, the probability transition matrix is an $n \times n$ matrix where P_{rc} is the probability the system will transition from state r to the next state of state c. Furthermore for each $0 \le r < n$,

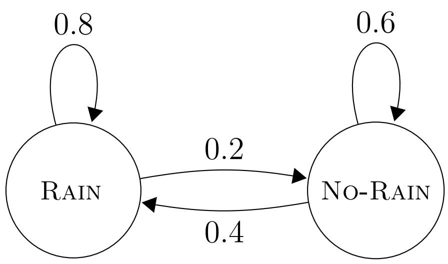
$$\sum_{c=0}^{n-1} P_{rc} = 1$$

Of course the states if not already written using $0, 1, 2, ..., n-1$ can be relabeled as such using a rv X.

The probability transition matrix can be described using a finite state diagram (see CISS362) where the transitions are probabilities. Furthermore there is no initial state and there are no accept states. For instance for the above rain example

	RAIN	NO-RAIN
RAIN	0.8	0.2
NO-RAIN	0.4	0.6

I can draw



Exercise 9.20.1. Continuing the above example:

debug: exercises/disc-prob-54/question.tex

(a) If it rained today, what is the probability that it will rain in two days? (Hint: What are all the relevant paths of state transitions?)
(b) What has the above computation got to do with the probability transition matrix?

(Go to solution, page 8187) □

Let $P^{(k)}_{rc}$ be the probability of transitioning from state r to state c in k transitions. Clearly

$$P^{(1)}_{rc} = P_{rc}$$

and

$$P^{(2)}_{rc} = (P^2)_{rc}$$

where $P^2 = P \cdot P$ where $\cdot$ being the matrix product. (Note: Read the above carefully. That's $(P^2)_{rc}$ and not $(P_{rc})^2$)!) It's reasonable to assume that

$$P^{(3)}_{rc} = (P^3)_{rc}$$

In particular

$$P^{(3)}_{00} = (P^3)_{00}$$

Exercise 9.20.2. Continuing with the rain example:

debug: exercises/disc-prob-58/question.tex

(a) Compute $P_{00}^{(3)}$.
(b) Compute P^3 where P is the probability transition matrix of our rain example.
(c) If $(P^3)_{00}$ the same as $P_{00}^{(3)}$?

(Go to solution, page 8188) □

Exercise 9.20.3. Prove that if P is the probability transition matrix of a Markov chain with n states, then probability of transition from state r to state c in k transitions is

debug: exercises/disc-prob-55/question.tex

$$P_{rc}^k$$

i.e., the (r, c)–entry of the matrix P^k. (Go to solution, page 8189) □

Back to our probability transition matrix of rain in Columbia:

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}$$

I have calculated the probability of rain two days from a day that rains. Now suppose my friend John is visiting Columbia for two days. He gave me 10 possible days he will be visiting Columbia. Out of these 10 days, according to the forecast, 4 of the days are rainy days and 6 are not. John is interested in the probability of rain two days after his arrival. How would we compute that? Clearly the probability that it will rain is

$$0.4P_{00}^2 + 0.6P_{10}^2$$

right? Note that this is the $(0, 0)$–entry of

$$\begin{bmatrix} 0.4 & 0.6 \end{bmatrix} P^2 = \begin{bmatrix} 0.4P_{00}^2 + 0.6P_{10}^2 \\ 0.4P_{01}^2 + 0.6P_{11}^2 \end{bmatrix}$$

The matrix P^2 is

$$P^2 = \begin{bmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{bmatrix}$$

and

$$\begin{bmatrix} 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{bmatrix} = \begin{bmatrix} 0.56 & 0.44 \end{bmatrix}$$

The probability that it will rain on the day John leaves Columbia is 0.56

(equivalently it will not rain with probability of 0.44).

Note that in general if a random day is chosen with probability p of rain, then to compute the probability of rain after k transitions, we look at

$$\begin{bmatrix} p & q \end{bmatrix} P^k$$

where $q = 1 - p$.

Furthermore note that if John is guaranteed to arrive on a day that is raining, then the initial vector is

$$\begin{bmatrix} 1 & 0 \end{bmatrix}$$

and the probability of rain after two days is the $(0, 0)$–entry of

$$\begin{bmatrix} 1 & 0 \end{bmatrix} P^2$$

You can see in general that if P is a probability transition matrix of n states, then given a probability row vector v

$$vP^k$$

gives you the probability vector after k transitions.

(Note: In the study of Markov chains, it's common to use the symbol π for this probability vector.)

Note that besides being a vector, the entries of v are ≥ 0 and sum up to 1:

- $v_i \geq 0$ for $0 \leq i < n$
- $\sum_{i=0}^{n-1} v_i = 1$

The probability vector v is the initial probabilistic state. In the situation where you know exactly what is the beginning state, say you begin at state i ($0 \leq i < n$), then the probability vector is

$$\begin{bmatrix} 0 & ... & 1 & ... & 0 \end{bmatrix}$$

where all entries are 0s except that there is a 1 at column i.

Given P and k, it would be nice if we can compute

$$vP^k$$

quickly. It would be better if we can compute P^k quickly given any $k \geq 0$. Of course ... eigenvalues come to the rescue!

Let's find a nonzero probability vector v such that

$$vP = \lambda v$$

This is not the usual eigenvector condition we are used to. Let me take the transpose:

$$\begin{aligned} (vP)^t &= (\lambda v)^t \\ \therefore \quad P^t v^t &= \lambda v^t \\ \therefore \quad (P^t - \lambda I_2)v^t &= 0 \\ \therefore \quad \det(P^t - \lambda I_2) &= 0 \end{aligned}$$

Note that v^t is a column vector. Of course

$$P^t = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}^t = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 1/5 & 3/5 \end{bmatrix}$$

Therefore from

$$\det\left(P^t - \lambda I_2\right) = 0$$

I get

$$\begin{aligned} \det \begin{bmatrix} 4/5 - \lambda & 2/5 \\ 1/5 & 3/5 - \lambda \end{bmatrix} &= 0 \\ \therefore \quad (4/5 - \lambda)(3/5 - \lambda) - 2/5 \times 1/5 &= 0 \\ \therefore \quad \lambda^2 - (7/5)\lambda + 2/5 &= 0 \\ \therefore \quad \lambda &= \frac{7/5 \pm \sqrt{49/25 - 4(2/5)}}{2} \\ &= \frac{7/5 \pm \sqrt{9/25}}{2} \\ &= \frac{7/5 \pm 3/5}{2} \\ &= 1, 2/5 \end{aligned}$$

The eigenvalues are $\lambda_0 = 1$ and $\lambda_1 = 2/5$. (When the eigenvalues are all real, it's a convention to list them in descending order.) Since the eigenvalues are distinct, their eigenvectors are linearly independent eigenvectors and the matrix P^t is diagonalizable.

For $\lambda_0 = 1$, if $v_0 = [x_0, y_0]$ is the eigenvector, then

$$\begin{aligned}
\begin{bmatrix} 4/5-1 & 2/5 \\ 1/5 & 3/5-1 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} &= 0 \\
\therefore \quad \begin{bmatrix} -1/5 & 2/5 \\ 1/5 & -2/5 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} &= 0 \\
\therefore \quad \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} &= 0 \\
\therefore \quad \begin{bmatrix} -1x_0 & 2y_0 \\ x_0 & -2y_0 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\therefore \quad -x_0 + 2y_0 &= 0 \\
\therefore \quad y_0 &= x_0/2
\end{aligned}$$

Hence $v_0 = [x_0, x_0/2]$. Choosing $x_0 = 2$, $v_0 = [2, 1]$. (The eigenspace of $\lambda_0 = 1$ is $\mathbb{R}[2, 1]$.)

For $\lambda_1 = 2/5$, if $v_1 = [x_1, y_1]$ is an eigenvector of λ_1, then

$$\begin{aligned}
\begin{bmatrix} 4/5-2/5 & 2/5 \\ 1/5 & 3/5-2/5 \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} &= 0 \\
\therefore \quad \begin{bmatrix} 2/5 & 2/5 \\ 1/5 & 1/5 \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} &= 0 \\
\therefore \quad \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} &= 0 \\
\therefore \quad \begin{bmatrix} 2x_1 + 2y_1 \\ x_1 + y_1 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\therefore \quad x_1 + y_1 &= 0 \\
\therefore \quad y_1 &= -x_1
\end{aligned}$$

Hence $v_1 = [x_1, -x_1]$. Choosing $x_1 = 1$, $v_1 = [1, -1]$. (The eigenspace of $\lambda_1 = 2/5$ is $\mathbb{R}[1, -1]$.)

Note: For both cases, to find the relationship between x_0, y_0 or x_1, y_1, a more systematic method, especially for a larger matrix size, is to use row reduction.

Exercise 9.20.4. Here's a shortcut. First, here are some standard facts about matrices. Some easy to prove – try to prove them. For the rest, verify using some examples. After that look for a book on linear algebra and study the proofs for the statements below. debug: exercises/disc-prob-57/question.tex

Theorem 9.20.1. *Let M and N be a square matrix of the same size.*

(a) $\det(M) = \det(M^t)$
(b) $\operatorname{tr}(M) = \operatorname{tr}(M^t)$
(c) *M and M^t have the same eigenvalues.*
(d) *The sum of eigenvalues of M is the trace of M.*
(e) *If M and N are similar, then M and N have the same determinant, trace, and eigenvalues.*

(M and N are similar matrices if there is a change of basis matrix P such that $PMP^{-1} = N$.)

Proposition 9.20.1. *Let P be a probability transition matrix.*

(a) *1 is an eigenvalue of P.*
(b) *If P has size 2×2, then the eigenvalues of P^t are $1, P_{00} + P_{11} - 1$.*

(Go to solution, page 8190) □

I want to diagonalize P^t. I have

$$P^t = QDQ^{-1}$$

where

$$D = \begin{bmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{bmatrix}$$

is a diagonal matrix and where the columns of Q are the eigenvectors v_0, v_1 (as column vectors), i.e.,

$$Q = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$$

The inverse of Q is

$$Q^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}^{-1} = \frac{1}{3} \begin{bmatrix} -1 & -1 \\ -1 & 2 \end{bmatrix}$$

You should verify that

$$QDQ^{-1} = P^t$$

Recall that we want to compute powers of P^t. I have

$$\left(P^t\right)^n = QD^nQ^{-1} = Q\begin{bmatrix}\lambda_0^n & 0 \\ 0 & \lambda_1^n\end{bmatrix}Q^{-1}$$

For instance if I'm interested in the probabilisties after 10 transitions, I compute $(P^t)^{10}$:

$$\begin{aligned}\left(P^t\right)^{10} &= QD^{10}Q^{-1} \\ &= \begin{bmatrix}2 & 1 \\ 1 & -1\end{bmatrix}\cdot\begin{bmatrix}1 & 0 \\ 0 & (2/5)^{10}\end{bmatrix}\cdot\frac{1}{3}\begin{bmatrix}-1 & -1 \\ -1 & 2\end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix}2 & (2/5)^{10} \\ 1 & -(2/5)^{10}\end{bmatrix}\begin{bmatrix}1 & 1 \\ 1 & -2\end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix}2+(2/5)^{10} & 2-2(2/5)^{10} \\ 1-(2/5)^{10} & 1+2(2/5)^{10}\end{bmatrix}\end{aligned}$$

Therefore 10 days from a raining day, the probabilities are

$$\begin{aligned}\left(P^t\right)^{10}\begin{bmatrix}1 \\ 0\end{bmatrix} &= \frac{1}{3}\begin{bmatrix}2+(2/5)^{10} & 2-2(2/5)^{10} \\ 1-(2/5)^{10} & 1+2(2/5)^{10}\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix}2+(2/5)^{10} \\ 1-(2/5)^{10}\end{bmatrix} \\ &= \begin{bmatrix}0.6667... \\ 0.3332...\end{bmatrix}\end{aligned}$$

i.e., there's a 0.6667... (i.e. 67%) chance of rain 10 days after a rainy day. (And of course a 0.3332... chance of no rain.)

(In the above, I used column vector v^t for probabilities because it's more common to write an eigenvector equation as

$$Mv = \lambda v$$

where M is a square matrix. If you want to stick to row vectors, i.e., use

$$vP$$

instead of

$$P^tv^t$$

the eigenvector equation is

$$\begin{aligned} vP &= \lambda v \\ \therefore \quad v(P - \lambda I_2) &= 0 \\ \therefore \quad \det(P - \lambda I_2) &= 0 \end{aligned}$$

etc. Of course you will arrive at the same result. Note that if M is a square matrix, $\det M = \det M^t$.)

Hmmm the probability of rain (in the above) seems to be 2/3.

Write a program to compute and print the probabilities k days after a rainy day. Is the probability of rain k days after a raining day approaching 2/3 as k grows?

What if I start with a non-rainy day and ask what is the probability vector after 10 days? That should be

$$\left(P^t\right)^{10} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

I get

$$\begin{aligned} \left(P^t\right)^{10} \begin{bmatrix} 0 \\ 1 \end{bmatrix} &= \frac{1}{3} \begin{bmatrix} 2 + (2/5)^{10} & 2 - 2(2/5)^{10} \\ 1 - (2/5)^{10} + 1 & 1 + 2(2/5)^{10} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2 - 2(2/5)^{10} \\ 1 + 2(2/5)^{10} \end{bmatrix} \\ &= \begin{bmatrix} 0.6665... \\ 0.3334... \end{bmatrix} \end{aligned}$$

Wait ... what? How come it seems that regardless of whether I begin with a rainy day or a non-rainy day, after 10 days the chance of rain is almost 2/3??? In fact could it be that no matter what probability vector v I begin with, after many days the probability vector is going to be close to $[2/3, 1/3]$? In other words, using Calculus notation, is it true that

$$\lim_{k \to \infty} vP^k = [2/3, 1/3]?$$

or equivalently

$$\lim_{k \to \infty} (P^t)^k v^t = \begin{bmatrix} 2/3 & 1/3 \end{bmatrix}^t$$

This is helpful (if true). Anytime John comes and visit Columbia and stays for a "long time" (let's just say more than 10 days), if he asks for the chance of rain on his departure day, I'll just say it's about 2/3 chance of rain. No need to re-calculate based on condition of date of arrival!

By the way did you notice that the vector

$$\lim_{k\to\infty} (P^t)^k v^t = \begin{bmatrix} 2/3 & 1/3 \end{bmatrix}$$

looks familiar? The eigenvector of $\lambda_0 = 1$ is

$$[2 \quad 1]$$

If I scale this eigenvector into a probability vector, I get

$$[2/3 \quad 1/3]$$

(Remember that an eigenvector for an eigenvalue is not unique since any scalar multiple of the eigenvector is also an eigenvector of that eigenvalue.) See that?

One can ask, if given any probabilistic transition matrix P and a probability vector v, does vP^k converge, i.e., gets closer and closer to some fixed vector?

Now this cannot be true for all P ...

Exercise 9.20.5. For $n = 2$, find a probabilistic transition matrix P and vector of probabilities v such that vP^n does not converge, i.e., does not get closer and closer to a fixed vector of probabilities. (Go to solution, page 8191) □

debug: exercises/disc-prob-56/question.tex

Let's analyze what's happening. Going back to this computation:

$$\begin{aligned}
\left(P^t\right)^{10} &= QD^{10}Q^{-1} \\
&= \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} (2/5)^{10} & 0 \\ 0 & 1 \end{bmatrix} \cdot \frac{1}{3} \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{3} \begin{bmatrix} (2/5)^{10} & 2 \\ -(2/5)^{10} & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{3} \begin{bmatrix} (2/5)^{10} + 2 & -2(2/5)^{10} + 2 \\ -(2/5)^{10} + 1 & 2(2/5)^{10} + 1 \end{bmatrix}
\end{aligned}$$

It's easy to see that

$$\left(P^t\right)^k = \frac{1}{3}\begin{bmatrix} (2/5)^k + 2 & -2(2/5)^k + 2 \\ -(2/5)^k + 1 & 2(2/5)^k + 1 \end{bmatrix}$$

Therefore

$$\lim_{k\to\infty} \left(P^t\right)^k = \frac{1}{3}\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$$

Therefore

$$\lim_{k\to\infty} \left(P^t\right)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$$

and

$$\lim_{k\to\infty} \left(P^t\right)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$$

In fact

$$\lim_{k\to\infty} \left(P^t\right)^k \begin{bmatrix} p \\ 1-p \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$$

for any $0 \le p \le 1$.

In general, if

$$\lim_{k\to\infty} \left(P^t\right)^k$$

converges, the limit is the **steady state** or **equilibrium** of the Markov chain (or of P). steady state equilibrium

A probability vector v is **steady state vector** if steady state vector

$$\lim_{k\to\infty} \left(P^t\right)^k v = v$$

Questions:

1. Given P, does P have a steady state?
2. If P has a steady state, is there a probability vector v such that $\lim_{k\to\infty} vP^k$ converges, i.e., does P have a steady state vector?

Exercise below.

if I keep the eigenvalues and two eigenvectors around, say

$$v_0 = [a, b], \quad v_1, = [c, d]$$

Let the inverse matrix of

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

be

$$\begin{bmatrix} a' & c' \\ b' & d' \end{bmatrix}$$

i.e., $a' = d/\Delta$, $b' = -b/\Delta$, $c' = -c/\Delta$, $d' = a'/\Delta$ where $\Delta = ad - bc$. I see

$$\begin{aligned} \left(P^t\right)^k &= QD^kQ^{-1} \\ &= \begin{bmatrix} a & c \\ b & d \end{bmatrix} \cdot \lim_{k\to\infty} \begin{bmatrix} \lambda_0^k & 0 \\ 0 & \lambda_1^k \end{bmatrix} \cdot \begin{bmatrix} a' & c' \\ b' & d' \end{bmatrix} \end{aligned}$$

Assuming $\lambda_0 = 1$ and $0 \leq \lambda_1 < 1$,

$$\begin{aligned} \lim_{k\to\infty} \left(P^t\right)^k &= \begin{bmatrix} a & c \\ b & d \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} a' & c' \\ b' & d' \end{bmatrix} \\ &= \begin{bmatrix} a & 0 \\ b & 0 \end{bmatrix} \cdot \begin{bmatrix} a' & c' \\ b' & d' \end{bmatrix} \\ &= \begin{bmatrix} aa' & ac' \\ ba' & bc' \end{bmatrix} \end{aligned}$$

Now if I apply this matrix to a vector of probabilities $v = [p \;\; q]$,

$$\begin{aligned} \lim_{k\to\infty} \left(P^t\right)^k \begin{bmatrix} p \\ q \end{bmatrix} &= \begin{bmatrix} aa' & ac' \\ ba' & bc' \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} \\ &= \begin{bmatrix} a(pa' + qc') \\ b(pa' + qc') \end{bmatrix} \end{aligned}$$

From $q = 1 - p$ and $a' = d/\Delta, c' = -c/\Delta$:

$$\begin{aligned} \lim_{k\to\infty} \left(P^t\right)^k \begin{bmatrix} p \\ q \end{bmatrix} &= \begin{bmatrix} a(pa/\Delta - (1-p)c/\Delta) \\ b(pa/\Delta - (1-p)c/\Delta) \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} p(-cb) + (1-p)(ca) \\ p(-db) + (1-p)(da) \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} ac + pc(-b-a) \\ ad + pd(-b-a) \end{bmatrix} \end{aligned}$$

Note that if $a + b = 0$ we have

$$\lim_{k\to\infty} \left(P^t\right)^k v^t = \frac{a}{\Delta} \begin{bmatrix} c \\ d \end{bmatrix}$$

where $[c\ d]$ is the eigenvector of $\lambda_0 = 1$.

For our case, $a = 1, b = -1$ and therefore $a + b = 0$. Substituting $c = 2, d = 1$ into the above,

$$\lim_{k\to\infty} \left(P^t\right)^k v^t = \frac{1}{\Delta} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The probability of rain is $2/3$ as $k \to \infty$ regardless of whether the initial day is a rainy day or not.

Exercise 9.20.6. Convert this to a sequence of v and c where v=vowel and c=consonant: debug: exercises/disc-prob-65/question.tex

> A man entered who could hardly have been less than six feet six inches in height, with the chest and limbs of a Hercules. His dress was rich with a richness which would, in England, be looked upon as akin to bad taste. Heavy bands of astrakhan were slashed across the sleeves and fronts of his double-breasted coat, while the deep blue cloak which was thrown over his shoulders was lined with flame-coloured silk and secured at the neck with a brooch which consisted of a single flaming beryl. Boots which extended halfway up his calves, and which were trimmed at the tops with rich brown fur, completed the impression of barbaric opulence which was suggested by his whole appearance. He carried a broad-brimmed hat in his hand, while he wore across the upper part of his face, extending down past the cheekbones, a black vizard mask, which he had apparently adjusted that very moment, for his hand was still raised to it as he entered. From the lower part of the face he appeared to be a man of strong character, with a thick, hanging lip, and a long, straight chin suggestive of resolution pushed to the length of obstinacy.

Create a Markov chain with state v and c. Analyze this Markov chain for steady state. (Go to solution, page 8191) □

Exercise 9.20.7. Use a huge book (say from gutenberg ... even better use as many books as you can) and generate a markov chain of words and puncuation debug: exercises/disc-prob-66/question.tex