MATH 56A SPRING 2008
STOCHASTIC PROCESSES

KIYOSHI IGUSA

Contents

Date: March 16, 2008.

3. Continuous Markov Chains

"Continuous" means *continuous time*. We still have a countable (and thus discrete) set of states. I started by saying that continuous Markov chains are similar to Poisson processes. Later I discussed when they are transient, positive recurrent and null recurrent. There is also another interesting possibility. Continuous Markov chains can *explode*!

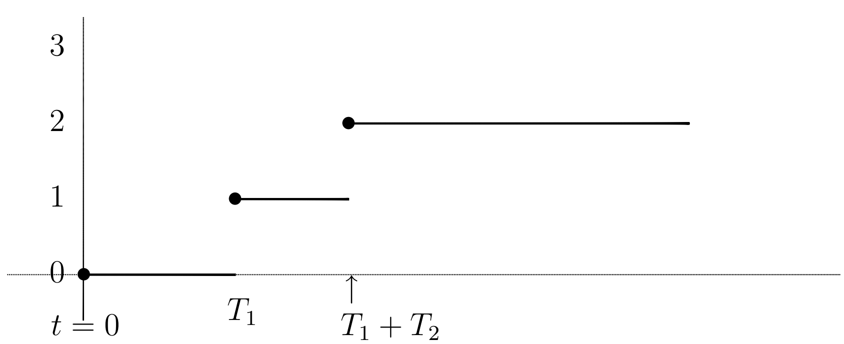
3.1. **Poisson process.** I started with the following example. Customers come into a store at a rate of one every $2\frac{1}{2}$ minutes. ($\lambda = \frac{2}{5}$ per minute.) Assume this is a *Poisson process*. This means:

(1) (independence) The number of occurrences of the event in one interval of time is independent of the number of occurrences in a different disjoint) time interval.
(2) (time homogeneous) The rate λ at which the event occurs is constant (independent of the time). For example, the rate is the same at midnight as it is at noon.
(3) Customers arrive one at a time. I.e., the event never occurs twice at exactly the same time. Even if you go with a friend, we assume that they have a high speed camera similar to those used on a race track to determine who enters the store first.

3.1.1. *conversion to Markov chain.* We convert this into a Markov chain by letting X_t be the number of occurrences of the Poisson event in the time interval $(0, t]$. This means that $X_0 = 0$. The state space is the set of nonnegative integers:

$$S = \{0, 1, 2, 3, \cdots\}$$

and the graph giving the state X_t at time t looks like this: In this



figure,

$$T_1 = \text{ 1st time that the event occurs}$$

$$T_2 = \text{ time between 1st and second occurrence of event}$$

Each time T_n is an exponential random variable.

3.1.2. *exponential variable.* Every Poisson process has an exponential variable associated with it. This is the time T between occurrences of the event. In the Markov chain T is how long you stay in the state you are in. I proved the following theorem in class:

Theorem 3.1.
$$\mathbb{P}(T_1 > t) = \mathbb{P}(X_t = 0) = e^{-\lambda t}.$$

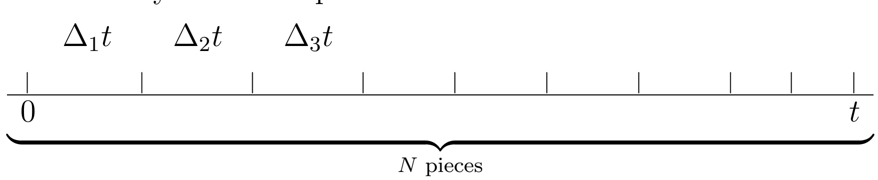
(And the same holds for $T_2, T_3, \cdots$.)

In order to prove this I needed the definition:

Definition 3.2. The *rate* λ of a Poisson process is defined to be the limit:
$$\lambda = \lim_{\Delta t \to 0} \frac{\mathbb{P}(\text{event occurs in } \Delta t)}{\Delta t}$$

This equation is often written using "little oh":
$$\mathbb{P}(\text{event occurs in } \Delta t) = \lambda \Delta t + o(\Delta t)$$
So, the probability that it doesn't occur is
$$\mathbb{P}(\text{event does not occur in } \Delta t) \approx 1 - \lambda \Delta t$$
The definition of the little oh is that this is a quantity that goes to zero faster than the named quantity:
$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$
If you compare this definition with the definition of the rate λ you will see that it says exactly the same thing.

Proof. We want to calculate the probability $\mathbb{P}(T_1 > t)$. This is the probability that the event *does not occur* in the time interval $(0, t]$. To calculate this you break up this interval into little intervals:



If there are N equal pieces then each piece has length
$$\Delta t = \frac{t}{N}$$
The ith pieces is
$$\Delta_i t = \left(\frac{(i-1)t}{N}, \frac{it}{N} \right]$$

In order for the event not to occur during the entire interval, it needs to not occur in each of these little intervals $\Delta_i t$. Since these are independent:

$$\begin{aligned}\mathbb{P}(T_1 > t) &= \prod_{i=1}^{N} \mathbb{P}(\text{event does not occur in } \Delta_i t) \\ &\approx \prod_{N} (1 - \lambda \Delta t) = \prod_{N} \left(1 - \lambda \frac{t}{N}\right) \\ &= \left(1 - \frac{\lambda t}{N}\right)^{N}\end{aligned}$$

The exact value is given by taking a limit:

$$\mathbb{P}(T_1 > t) = \lim_{N \to \infty} \left(1 - \frac{\lambda t}{N}\right)^{N} = e^{-\lambda t}$$

□

From this equation we can find the cdf and pdf of $T = T_1$:

The *cumulative distribution function* (cdf) of a random variable T is

$$F_T(t) = \mathbb{P}(T \leq t).$$

For the exponential variable, it is:

$$F_T(t) = 1 - \mathbb{P}(T > t) = 1 - e^{-\lambda t} \text{ if } t \geq 0$$

(And $F_T(t) = 0$ if $t < 0$.)

The *probability density function* (pdf) is the derivative of the cumulative distribution function when the cdf is differentiable:

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{d}{dt}\left(1 - e^{-\lambda t}\right) = \lambda e^{-\lambda t}$$

3.2. **Definitions.** On the second day I gave more precise definitions:

A *continuous finite Markov chain* has set of states $S = \{1, 2, \cdots, n\}$ with transition probabilities given by an *infinitesimal generator* A. This is an $n \times n$ matrix $A = (\alpha(i,j))$ whose rows add to zero and whose negative entries are all along the diagonal. If the entries are written $a_{ij} = \alpha(i,j)$ the equations corresponding to these statements are:

(1) $a_{ij} \geq 0$ if $i \neq j$
(2) $a_{i1} + a_{i2} + \cdots + a_{in} = 0$

Example 3.3.

$$A = \begin{pmatrix} -3 & 1 & 2 \\ 0 & -2 & 2 \\ 3 & 1 & -4 \end{pmatrix}$$

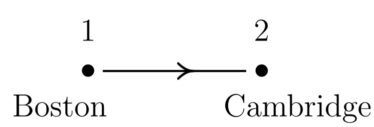
This is an infinitesimal generator. The negative numbers are all on the diagonal and the rows add up to zero. The numbers off the diagonal are positive or zero. Notice that the numbers can be greater than 1.

3.2.1. The numbers a_{ij} are the *infinitesimal transition probabilities*:

$$\boxed{a_{ij} = \alpha(i,j) = \lim_{t \to 0+} \frac{\mathbb{P}(X_t = j \mid X_0 = i)}{t}}$$

I used the following example to explain what this means.

Example 3.4. $\begin{pmatrix} -6 & 6 \\ 0 & 0 \end{pmatrix}$ This means: The transition from state 1 to



state 2 is a Poisson process with rate

$$\lambda = 6 \ /year$$

Question: If this represents people moving from Boston to Cambridge, how long does a Boston resident expect to remain in Boston?
Answer: 2 months.

The rate $\lambda = 6$ does not mean 6 people per year, it means 600% per year! This is the same as 50% per month. If half the people leave in one month then you might think the city will be empty in 2 months. But:

T = time until the 1st occurrence of the event (the jump $1 \to 2$).

T is an exponential variable with rate $\lambda = 6$. This means

$$\mathbb{P}(T > t) = e^{-\lambda t} = e^{-6t}.$$

$$e^{-\lambda t} = \underbrace{1 - \lambda t}_{1st\ order} + \frac{\lambda^2 t^2}{2!} + \frac{\lambda^3 t^3}{3!} + \cdots$$

- When t is very small, t^2, t^3, etc are negligible and

$$e^{-\lambda t} \approx 1 - \lambda t$$

$$\mathbb{P}(T > t) \approx 1 - \lambda t$$

with error $o(t)$.

$$\frac{\mathbb{P}(T \leq t)}{t} \approx \frac{\lambda t}{t} \approx \lambda$$

with error $o(t)/t$. Since $o(t)/t \to 0$ as $t \to 0$, we get:

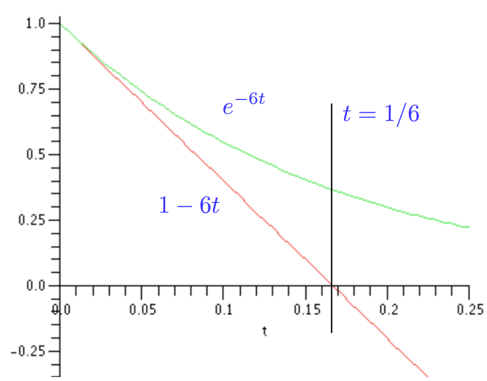
$$\boxed{\lim_{t \to 0+} \frac{\mathbb{P}(T \leq t)}{t} = \lambda}$$

- $t = 1/6$ (2 months):

$$\mathbb{P}(T > t) = e^{-6t} = e^{-6/6} = e^{-1} \approx 0.368$$

This is the probability that you don't jump to state 2. I.e., if people are leaving Boston at the rate of 50% per month then in 2 months 63.2% will have left and 36.8% will remain. The number 0 is just the first order approximation:

$$e^{-\lambda t} \approx 1 - \lambda t = 1 - 6(\tfrac{1}{6}) = 0.$$



3.2.2. *computation of expected value.* The definition of expected value is

$$\mathbb{E}(T) = \int t\, f_T(t)\, dt$$

Since we know that

$$pdf = f_T(t) = \lambda e^{-\lambda t}$$

for $t \geq 0$ and $f_T(t) = 0$ for $t < 0$, we can compute the integral:

$$\mathbb{E}(T) = \int_0^\infty t\lambda e^{-\lambda t} dt = -te^{\lambda t} - \frac{1}{\lambda} e^{-\lambda t}|_0^\infty$$

$$= 0 - (-0 - 1/\lambda) = \boxed{\frac{1}{\lambda}}$$

which in this case is 1/6 year = 2 months.

3.3. **probability transition matrix.**

Definition 3.5. For any $t \geq 0$ we can ask which state the system will be in at time t given that it starts in state i:

$$p_t(i,j) := \mathbb{P}(X_t = j \mid X_0 = i)$$

These numbers give the *probability transition matrix*

$$P_t := (p_t(i,j))$$

3.3.1. *relation to infinitesimal generator.* This matrix is related to the infinitesimal generator A by the following formula:

$$\boxed{P_t = e^{tA}}$$

which I proved later. First I explained how to compute this (in the case when the eigenvalues of A are distinct).

Suppose that A has the right eigenvectors $V_1, V_2, \cdots, V_n$(these are column vectors) with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. If we put the vectors V_j side by side to make a square matrix

$$Q = (V_1, V_2, \cdots, V_n)$$

Then the eigenvalue equation $AV_j = \lambda_j V_j$ becomes the matrix equation:

$$AQ = QD, \qquad D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

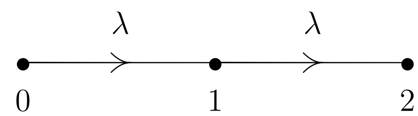
This means

$$A = QDQ^{-1}$$

Theorem 3.6.

$$P_t = e^{tQAQ^{-1}} = Qe^{tD}Q^{-1} = Q \begin{pmatrix} e^{t\lambda_1} & & & 0 \\ & e^{t\lambda_2} & & \\ & & \ddots & \\ 0 & & & e^{t\lambda_n} \end{pmatrix} Q^{-1}$$

Example 3.7. (I got stuck on this example in class but I will finish it here.) Take the example of the customers coming into the store at a rate of λ. Recall that X_t is the number of customers who visit the store in the time interval $(0,t]$. I took a finite approximation of this



with only three states $S = \{0, 1, 2\}$. So,

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & 0 \end{pmatrix}$$

This matrix is not diagonalizable! But I can use a sneaky trick to find $P_t = e^{tA}$. By 0.3.1, P_t is the unique solution of the differential equation

$$\boxed{\frac{d}{dt} P_t = A P_t}$$

with initial condition $P_0 = I_3$ (the 3×3 identity matrix). Since A is upper triangular, we know that P_t has the form

$$P_t = \begin{pmatrix} e^{-\lambda t} & x(t) & y(t) \\ 0 & e^{-\lambda t} & z(t) \\ 0 & 0 & 1 \end{pmatrix}$$

The equation $P_t' = AP_t$ means that the derivative of $x(t)$ is the $(1, 2)$ entry of AP_t:

$$x'(t) = -\lambda x(t) + \lambda e^{-\lambda t}$$

and $P_0 = I_3$ means that $x(0) = 0$. So,

$$x(t) = \lambda t e^{-\lambda t}$$

Since the rows of P_t add up to 1 we get:

$$P_t = \begin{pmatrix} e^{-\lambda t} & \lambda t e^{-\lambda t} & 1 - (1 + \lambda t) e^{-\lambda t} \\ 0 & e^{-\lambda t} & 1 - e^{-\lambda t} \\ 0 & 0 & e^0 \end{pmatrix}$$

The theorem assumes the eigenvalues are distinct. The differential equation in the box always holds and has a unique solution.

Proof. I gave the proof that $e^{tA} = e^{QtDQ^{-1}} = Qe^{tD}Q^{-1}$. First, recall the definition:

$$e^{tA} := I + tA + \frac{t^2 A^2}{2} + \frac{t^3 A^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}$$

But

$$A^n = (QDQ^{-1})^n = QD\underbrace{Q^{-1}Q}_{I}D\underbrace{Q^{-1}Q}_{I}D\cdots$$
$$= QD^nQ^{-1}$$

So,

$$e^{tA} = \sum \frac{Qt^nD^nQ^{-1}}{n!} = Q\left(\sum \frac{t^nD^n}{n!}\right)Q^{-1}$$
$$P_t = e^{tA} = Qe^{tD}Q^{-1}$$

□

3.3.2. *finding the invariant distribution.* π. This is the probability vector so that

$$\pi P_t = \pi \quad \text{for all } t \geq 0$$

Substituting

$$P_t = e^{tA} = I + tA + \frac{t^2A^2}{2} + \cdots$$

this equation becomes:

$$\pi e^{tA} = \pi + \underbrace{t\pi A + t^2\frac{\pi A^2}{2} + \cdots}_{0}$$
$$= \pi$$
$$\Leftrightarrow \pi A = 0$$

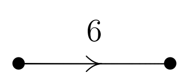
Proof: If we take t very small then $t^2, t^3, \cdots$ are negligible and we have

$$\pi e^{tA} \approx \pi(I + tA) = \pi + t\pi A$$

which is equal to π if and only if $\pi A = 0$.

So, the invariant distributions are the left null vectors of A which have nonnegative coordinates adding up to 1.

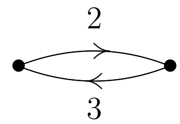
Example 3.8. Take the really simple case



$$A = \begin{pmatrix} -6 & 6 \\ 0 & 0 \end{pmatrix}$$

This has rank 1. So it only has one null vector:

$$\pi = (0, 1)$$



Example 3.9. This is an important case:

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

This also has rank 1 and

$$\pi = \left(\frac{3}{5}, \frac{2}{5} \right)$$

This is the solution of the equation

$$2\pi(1) = 3\pi(2)$$

The ratio of $\pi(1) : \pi(2)$ must be $3 : 2$ and you need to divide by 5 to make it a probability vector.