

Notes for Math 450

Continuous-time Markov chains and Stochastic Simulation

Renato Feres

These notes are intended to serve as a guide to chapter 2 of Norris's textbook. We also list a few programs for use in the simulation assignments. As always, we fix the probability space (Ω, \mathcal{F}, P) . All random variables should be regarded as \mathcal{F} -measurable functions on Ω . Let S be a countable (or finite) state set, typically a subset of \mathbb{Z} . A *continuous-time random process* $(X_t)_{t \geq 0}$ is a family of random variables $X_t : \Omega \rightarrow S$ parametrized by $t \geq 0$.

1 Q -matrices - Text sec 2.1

The basic data specifying a continuous-time Markov chain is contained in a matrix $Q = (q_{ij})$, $i, j \in S$, which we will sometimes refer to as the *infinitesimal generator*, or as in Norris's textbook, the *Q -matrix* of the process, where S is the state set. This is defined by the following properties:

1. $q_{ii} \leq 0$ for all $i \in S$;
2. $q_{ij} \geq 0$ for all $i, j \in S$ such that $i \neq j$;
3. $\sum_{j \in S} q_{ij} = 0$ for all $i \in S$.

The motivation for introducing Q -matrices is based on the following observation (see theorems 2.1.1 and 2.1.2 in Norris's text), which applies to a finite state set S : the matrices $P(t) = e^{tQ}$, for $t \geq 0$, defined by the convergent matrix-valued Taylor series

$$P(t) = I + \frac{tQ}{1!} + \frac{t^2Q^2}{2!} + \frac{t^3Q^3}{3!} + \dots$$

constitute a family of stochastic matrices. $P(t) = (p_{ij}(t))$ will be seen to be the transition probability matrix at time t for the Markov chain (X_t) associated to Q . The chain (X_t) will be defined later not directly in terms of the transition probabilities but from two discrete-time processes (the holding times and jump chains) associated to Q , to be defined later. Only after that will we derive the interpretation

$$p_{ij}(t) = P_i(X_t = j) = P(X_t = j | X_0 = i).$$

In fact, we will prove later that a continuous-time Markov chain $(X_t)_{t \geq 0}$ derived from a Q -matrix satisfies:

$$P(X_{t_{n+1}} = j | X_{t_n} = i, X_{t_{n-1}} = i_{n-1}, \dots, X_{t_0} = i_0) = p_{ij}(t_{n+1} - t_n)$$

for all times $0 \leq t_0 \leq t_1 \leq \dots \leq t_{n+1}$ and all states $j, i_0, \dots, i_{n-1}, i$. We take this property for granted for the time being and examine a few examples.

2 A few examples

The information contained in a Q -matrix is conveniently encoded in a transition diagram, where the label attached to the edge connecting state i to state j is the entry q_{ij} . We disregard all self-loops.

Example 1. We begin by examining the process defined by the diagram of figure 1.

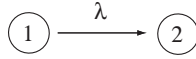


Figure 1: After an exponential random time with parameter λ , the process switches from state 1 to state 2, and then remains at 2.

The Q -matrix associated to this diagram has entries $-q_{11} = q_{12} = \lambda$ and $q_{21} = q_{22} = 0$. To obtain the stochastic matrix $P(t)$ we use Theorem 2.1.1, which shows that $P(t)$ satisfies the equation $P'(t) = QP(t)$ with initial condition $P(0) = I$. It is immediate that the entry $p_{11}(t)$ of $P(t)$ must satisfy the differential equation $y' = -\lambda y$ with the initial condition $y(0) = 1$. This gives $y(t) = e^{-\lambda t}$, so that

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

The key remark is that the time of transition from 1 to 2 has exponential distribution with parameter λ . (See Lecture Notes 3, section 5.) We will recall later some of the salient features of exponentially distributed random variables.

Example 2. The next example refers to the diagram of figure 2.

The Q -matrix for this example is

$$Q = \begin{pmatrix} -\mu & \lambda_1 & \dots & \lambda_N \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

where $\mu = \lambda_1 + \dots + \lambda_N$. Once again, we use the equation $P'(t) = QP(t)$ with initial condition $P(0) = I$ to obtain the transition probabilities. It is clear from

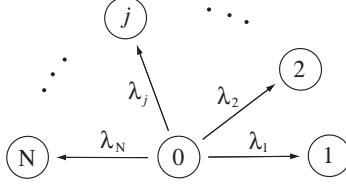


Figure 2: After an exponential random time with parameter $\lambda = \lambda_1 + \dots + \lambda_N$, the process switches from state 0 to one of the states $1, \dots, N$, and then remains there.

the equation

$$p'_{ij}(t) = \sum_{k=0}^N q_{ik} p_{kj}(t)$$

that, whenever the present state is $i \neq 0$, the transition to $j \neq i$ has probability $p_{ij}(t) = 0$. For $i = 0$ we have $p'_{00}(t) = -\mu p_{00}(t)$, $p_{00}(0) = 1$, and for $j \neq 0$, $p'_{0j}(t) = -\mu p_{0j}(t) + \lambda_j$, $p_{0j}(0) = 0$. The solution is easily seen to be

$$p_{0j}(t) = \begin{cases} e^{-\mu t} & \text{for } j = 0 \\ \frac{\lambda_j}{\mu} (1 - e^{-\mu t}) & \text{for } j \neq 0. \end{cases}$$

The solution can be interpreted as follows: assuming that the process is initially at state zero, the transition to a state $j \neq 0$ happens at an exponentially distributed random time with parameter $\mu = \lambda_1 + \dots + \lambda_N$. At that jump time, the new state j is chosen with probability

$$g_j = \frac{\lambda_j}{\lambda_1 + \dots + \lambda_N}.$$

Example 3. We study now the process defined by the diagram of figure 3. The transition probabilities $p_{ij}(t)$ can be obtained as in the previous examples.

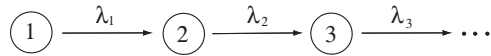


Figure 3: A continuous-time birth process.

The Q -matrix for this example has entries

$$q_{ij} = \begin{cases} -\lambda_i & \text{if } j = i \\ \lambda_i & \text{if } j = i + 1 \\ 0 & \text{if } j \neq i, i + 1. \end{cases}$$

It can be shown in this case that at each state i the process waits a random time, exponentially distributed with parameter λ_i , then jumps to the next state $i + 1$. The mean holding (waiting) time at state i is $1/\lambda_i$. (See properties of exponential distribution in Lecture Notes 3.) Denote by S_n the holding time before the n -th transition (to state n). Note that the expected value of the sum $\zeta = S_1 + S_2 + \dots$ is finite if

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} < \infty.$$

In this case, J must be finite with probability 1. (If a random variable assumes the value ∞ with positive probability, its expected value is infinite. This is clear since the weighted average of a set of numbers that includes ∞ with positive weight is necessarily equal to ∞ .) The random variable ζ is called the first *explosion time* of the process. If the process has finite explosion time, it will run through an infinite number of transitions in finite time. We will have more to say about this phenomenon later.

We consider in more detail the special case of the last example having constant $\lambda_i = \lambda$. The transition probabilities in this case can be calculated (see example 2.1.4 in text) to be

$$p_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}.$$

In particular, the transition from $i = 0$ to j in time t has the Poisson distribution with parameter λt . (See Lecture Notes 3.) Therefore, the process $(X_t)_{t \geq 0}$ of example 3 for constant λ , and starting at $X_0 = 0$, has the following characterization: for each t , X_t is a Poisson random variable with parameter λt .

3 Jump times and holding times - Text sec. 2.2

Since the set of states is discrete and the time parameter is continuous, it is clearly not possible for the sample paths $X_t(\omega)$ to be continuous functions of t . At random times $J_0 = 0, J_1, J_2, \dots$, called the *jump times* (or transition times) the process will chance to a new state, and the sequence of states constitute a discrete-time process Y_0, Y_1, Y_2, \dots .

It is convenient to assume that sample paths are *right-continuous*. This means that for all $\omega \in \Omega$, there is a positive ϵ such that $X_s(\omega) = X_t(\omega)$ for s, t such that $t \leq s \leq t + \epsilon$. In particular, $X_t = Y_n$ for $J_n \leq t < J_{n+1}$.

More formally, we define the jump times of the process $(X_t)_{t \geq 0}$ inductively as follows: $J_0 = 0$ and, having obtained J_n we define J_{n+1} as

$$J_{n+1} = \inf\{t \geq J_n | X_t \neq X_{J_n}\}.$$

The *infimum*, or *inf*, of a set A of real numbers is the unique number a (not necessarily in A) such that every element of A is greater than or equal to a (i.e., a is a lower bound for A) any no other lower bound for A is greater than a .

Thus J_{n+1} is the least random time greater than J_n at which the process takes a new value $X_t \neq X_{J_n}$. In the definition of J_{n+1} the infimum is evaluated for each sample path. A more explicit statement is that for each $\omega \in \Omega$, $J_{n+1}(\omega)$ is the infimum of the set of times $t \geq t_n = J_n(\omega)$ such that $X_t(\omega)$ is different from $X_{t_n}(\omega)$. It could happen that the process gets stuck at an absorbing state and no further transitions occur. In this case $J_{n+1}(\omega) = \infty$. In this case we define $X_{J_n} = X_\infty$ (the final value of X_t). If all J_n are finite, the final value of the process is not defined.

We also define the *holding times* S_n , $n = 1, 2, \dots$, as the random variables

$$S_n = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise.} \end{cases}$$

The right-continuity condition implies that the holding times S_n are positive for all n , that is, there cannot be two state transitions happening at the same time. It is, nevertheless, possible in principle for a sequence of jump times to accumulate at a finite time. In other words, the random variable

$$\zeta = \sum_{n=1}^{\infty} S_n$$

may be finite. The random variable ζ is called the first *explosion time*. As we saw in the birth process in the previous section, it is possible that the holding times of a sequence of state transitions become shorter and shorter, so that the chain undergoes an infinite number of transitions in a finite amount of time. This is called an *explosion*. We will describe later simple conditions for the process to be non-explosive.

The analysis of a continuous-time Markov chain $(X_t)_{t \geq 0}$ can be approached by studying the two associated processes: the holding times S_n and the *jump process* Y_n , $n = 0, 1, 2, \dots$. This is explained in the next section.

4 The jump matrix Π - Text sec. 2.6

For a given Q -matrix $Q = (q_{ij})$ we associate a stochastic matrix $\Pi = (\pi_{ij})$, called the *jump matrix*, as follows. Write $q_i = -q_{ii}$ for all $i \in S$. Note that q_i is non-negative and $q_i = \sum_{j \neq i} q_{ij}$. Now define Π as follows: for each $i \in S$, if $q_i > 0$ set the diagonal entry of the i -th row of Π to zero and the other entries to $\pi_{ij} = q_{ij}/q_i$. If $q_i = 0$, set $\pi_{ii} = 1$ and the other entries in row i to 0. In other words, define:

$$\pi_{ij} = \begin{cases} q_{ij}/q_i & \text{if } q_i \neq 0 \text{ and } j \neq i \\ 0 & \text{if } q_i \neq 0 \text{ and } j = i \\ 0 & \text{if } q_i = 0 \text{ and } j \neq i \\ 1 & \text{if } q_i = 0 \text{ and } j = i. \end{cases}$$

As an example, consider the process specified by the Q -matrix:

$$Q = \begin{pmatrix} -2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 3 \\ 2 & 0 & 0 & -4 & 2 & 0 \\ 0 & 3 & 1 & 0 & -5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The associated transition diagram is given in figure 4.

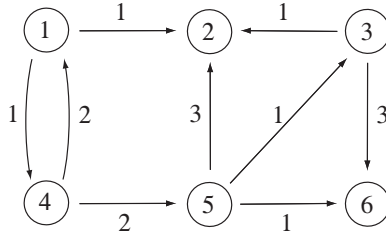


Figure 4: Transition diagram for a continuous-time Markov chain. The corresponding Π -matrix is given in the text.

The Π -matrix for this process is:

$$\Pi = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{3}{5} & \frac{1}{5} & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We can now present the general description of a continuous-time Markov chain as consisting of two independent discrete-time processes: the holding times S_n and the jump process Y_n associated to the Π -matrix. We call this the *hold-and-jump process*. It won't be immediately apparent why the hold-and-jump process has the transition probabilities matrix $P(t) = e^{tQ}$. We will return to this point a little later and show the different but equivalent ways in which the continuous-time Markov chain can be represented.

Here is the main definition. First suppose that the process is non-explosive. For a given Q -matrix Q and initial probability distribution λ on S , let $(Y_n)_{n \geq 0}$ be a discrete-time Markov chain $\text{Markov}(\lambda, \Pi)$, where Π is the Π -matrix associated to Q . Having the Y_n , let now S_1, S_2, \dots be a sequence of independent exponential random variables with parameters q_{Y_0}, q_{Y_1}, \dots , respectively. For each $n \geq 1$, define the n -th jump time by $J_n = S_1 + \dots + S_n$, and $J_0 = 0$. Finally, define a *hold-and-jump process* with initial probability distribution λ and

generator matrix Q , written $\text{Markov}(\lambda, Q)$, as the (right-continuous) process $(X_t)_{t \geq 0}$ given by

$$X_t = Y_n \quad \text{if} \quad J_n \leq t < J_{n+1}.$$

This definition can be modified to include explosive chains. This amounts to adding to S an extra state, denoted ∞ , which is attained after explosion, and defining $X_t = \infty$ if t does not lie in any of the intervals $[J_n, J_{n+1})$.

Representing a continuous-time Markov chain as a hold-and-jump process is particularly useful as it suggests a method of stochastic simulation. We pursue this in the next section.

5 Simulation of the hold-and-jump process

We begin by restating the description of the hold-and-jump process, making use of some of the basic properties of exponential random variables. See Lecture Notes 3 for a discussion of exponential random variables and their simulation. In particular, we should keep in mind the following: (i) if T is an exponential random variable of parameter 1, then αT is an exponential random variable of parameter $1/\alpha$; and (ii) if $M^{(j)}$ are independent exponential random variables of parameters λ_j , then $M = \inf_j M^{(j)}$ is an exponential random variable of parameter $\lambda = \sum_j \lambda_j$. Here, M can be interpreted as the time of the first occurring event among events with exponential times $M^{(j)}$.

To obtain a sample chain of holding times (S_n) and states (Y_n) , we do the following: First choose $X_0 = Y_0$ with probability distribution λ . Then choose an array $(T_n^{(j)} : n \geq 1, j \in S)$ of independent exponential random variables of parameter 1. Inductively, for $n = 0, 1, 2, \dots$, if $Y_n = i$ set:

$$\begin{aligned} S_{n+1}^{(j)} &= T_{n+1}^{(j)} / q_{ij} \quad \text{for } j \neq i, \\ S_{n+1} &= \inf_{j \neq i} S_{n+1}^{(j)} \end{aligned}$$

Now choose the new state Y_{n+1} according to the rule:

$$Y_{n+1} = \begin{cases} j & \text{if } S_{n+1}^{(j)} = S_{n+1} < \infty; \\ i & \text{if } S_{n+1} = \infty. \end{cases}$$

Conditional on $Y_n = i$, the $S_{n+1}^{(j)}$ thus obtained are independent exponential random variables of parameter q_{ij} , for all $j \neq i$, and S_{n+1} is exponential of parameter q_i ; Y_{n+1} has distribution $(\pi_{ij} : j \in S)$, and S_{n+1} and Y_{n+1} are independent and (conditional on $Y_n = i$) independent of Y_0, Y_1, \dots, Y_n and S_1, S_2, \dots, S_n .

Before turning this into an algorithm for simulating the Markov chain, we briefly mention another useful interpretation of the process which has some intuitive appeal. We imagine the transitions diagram of the process as depicting a system of chambers (vertices) with a gate for each directed edge from chamber i to chamber j . The gates are directed, so that gate (i, j) controls the transit from

chamber i to j , while flow from j to i is controlled by gate (j, i) . Each gate opens independently of all others at random times for very brief moments, and whenever it does, anyone waiting to pass will immediately take the opportunity to do so. Over a period $[0, t]$, the gate (i, j) will open at times according to a Poisson process with parameter q_{ij} . In other words, The number of times, $N_{ij}(t)$, that the gate opens during $[0, t]$ is a Poisson random variable with parameter $q_{ij}t$, and these events are distributed over $[0, t]$ uniformly. Now, someone moving in this maze, presently waiting in chamber i , will move next to the chamber whose gate from i opens first. The chain then corresponds to the sequence of chambers the person visits and the waiting times in each of them.

We now describe an algorithm that implements the hold-and-jump chain. The following assumes a finite state space S :

1. Initialize the process at $t = 0$ with initial state i drawn from the distribution λ ;
2. Call the current state i ; simulate the time of the next event, t' , as an $\text{Exp}(q_i)$ random variable;
3. Set the new value of t as $t \leftarrow t + t'$;
4. Simulate the new state j : if $q_i = 0$, set $j = i$ and stop. If $q_i \neq 0$, simulate a discrete random variable with probability distribution given by the i -th row of the Π -matrix, i.e., q_{ij}/q_i , $j \neq i$;
5. If t is less than a pre-assigned maximum time T_{\max} , return to step 2.

The following program implements this algorithm in Matlab.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [t y]=ctmc(n,pi,Q)
%Obtain a sample path with n events for a
%continuous-times Markov chain with initial
%distribution pi and generator matrix Q.
%The output consists of two row vectors:
%the event times t and the vector of states y.
%Vectors t and y may be shorter than n if
%an absorbing state is found before event n.
%Uses samplefromp(pi,n).
t=[0];
y=[samplefromp(pi,1)]; %initial state
for k=1:n-1
    i=y(k);
    q=-Q(i,i);
    if q==0
        break
    else
        s=-log(rand)/(-Q(i,i)); %exponential holding time
```



```

        t=[t t(k)+s];
        p=Q(i,:);
        p(i)=0;
        p=p/sum(p);
        y=[y samplefromp(p,1)];
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

As an example to illustrate the use of the program, consider the birth-and-death chain given by the diagram of figure 5.

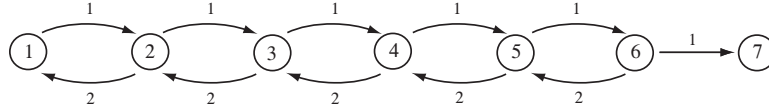


Figure 5: Diagram of a birth-and-death chain with an absorbing state.

The first 50 events of a sample path of the chain of figure 5 are shown in figure 6.

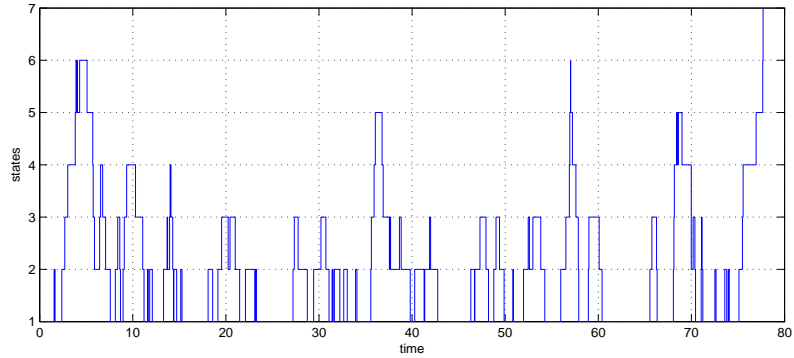


Figure 6: A sample path for the continuous-time Markov chain of figure 5. The initial state is 1. The chains stopped at after 157 events, at the absorbing state 7.

To obtain the sample path of figure 6, we used the following commands:

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Q=zeros(7,7);
Q(1,[1 2])=[-1 1];

```

```

Q(2,[1 2 3])=[2 -3 1];
Q(3,[2 3 4])=[2 -3 1];
Q(4,[3 4 5])=[2 -3 1];
Q(5,[4 5 6])=[2 -3 1];
Q(6,[5 6 7])=[2 -3 1];
pi=[1 0 0 0 0 0 0];
[t y]=ctmc(50,p,Q);
stairs(t,y)
grid
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

It is worth to mention the following modification of the algorithm given above for simulating a continuous-time Markov chain. It is based on the property that if we have k independent exponential random variables T_1, \dots, T_k of parameters $\lambda_1, \dots, \lambda_k$, respectively, then $T = \min_i \{T_i : i = 1, \dots, k\}$ is also an exponential random variable of parameter $\lambda_1 + \dots + \lambda_k$. Although it may not be immediately obvious, the following algorithm produces another version of exactly the same process as the first algorithm:

1. Initialize the process at $t = 0$ with initial state i drawn from the distribution λ ;
2. Call the current state i . For each potential next state l ($k \neq i$), simulate a time t_l with the exponential distribution of parameter q_{il} . Let j be the state for which t_j is minimum among the t_l ;
3. Set the new value of t as $t \leftarrow t + t_j$;
4. Let the new state be j ;
5. If t is less than a pre-assigned maximum time T_{\max} , return to step 2.

Although equivalent, this algorithm is less efficient since it requires the simulation of more random variables.

6 Remarks about explosion times - Text sec. 2.7

Recall that it is possible for a process to go through an infinite sequence of transitions in a finite amount of time, a phenomenon called *explosion*. If the holding times are S_1, S_2, \dots , then the explosion time is defined by

$$\zeta = S_1 + S_2 + \dots$$

The process is called *explosive* if ζ is there is a positive probability that ζ is finite.

Explosive processes can be extended beyond the first explosion time ζ . Section 2.9 has more on the topic. We will not discuss this issue in any depth, but only mention that any of the following conditions is enough to ensure that the process is non-explosive (see Theorem 2.7.1):

1. The state set S is finite;
2. There exists a finite number M such that $q_i \leq M$ for all $i \in S$;
3. The initial state, $X_0 = i$, is recurrent for the jump chain having transition probabilities matrix Π .

7 Kolmogorov's equations - Text sec. 2.8, 2.4

We have at this point two different description of the continuous-time Markov chain: first as a hold-and-jump process, specified by a discrete time Markov chain with transition probabilities given by a Π -matrix, and a sequence of random times S_n defining the waiting times between two transition events. Second, we have the process with a transition probabilities matrix $P(t)$ and generator Q so that $P(t) = e^{tQ}$. We need to show that these are different aspects of the same process. We will follow the textbook from this point on.

References

- [Nor] J.R. Norris. *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge U. Press, 1997.
- [Wil] Darren J. Wilkinson. *Stochastic Modelling for Systems Biology*, Chapman & Hall/CRC, 2006.