Continuous Time Markov Chains

Dan Crytser

March 11, 2011

1 Introduction

The theory of Markov chains of the discrete form $X_1, X_2, \ldots$ can be adapted to a continuous form $X_t, t \in [0, \infty)$. This requires use of a q-matrix of *transition probabilities*, which is also known as an infinitesimal generator for the Markov chain X_t. The q-matrix can be used to derive a related *jump matrix* which gives a discrete Markov chain called the jump chain J_n. The Markov chain chain X_t is given by running the chain J_n and then holding at each state for an amount of time given by independent Markov chains representing the hold times.

There is also a more abstract analytic approach following [And] some aspects of which we present here.

2 Definitions

A Markov chain X_t is a stochastic process with domain $[0, \infty \subset \mathbb{R}$ and countable state space E which satisfies the following Markov property: given a set of states $\{i_1, \ldots, i_k\} \subset E$ and a collection of times $\{t_1, \ldots, t_k\} \subset [0, \infty)$, we have

$$\mathbb{P}(X_{t_k} = i_k | X_{t_{k-1}} = i_{k-1}, \ldots, X_{t_1} = i_1) = \mathbb{P}(X_{t_k} = i_k | X_{t_{k-1}} = i_{k-1}).$$

In other words, the dependence of the current state on any set of previous states is only sensitive to the most recent of these prior states. Note that by evaluating $X(1), X(2), \ldots$ we obtain an ordinary discrete Markov chain (nothing special about the countable set $\{1, 2, \ldots\}$ here–we can generate a discrete Markov chain evaluating $\epsilon\mathbb{N}$ for any $\epsilon > 0$). Given states $i, j \in Y$ we define

$$P_{ij}(t, s) = \mathbb{P}(X_s = j | X_t = i).$$

The chain X_t is said to be *homogeneous* if for all states $i, j \in E$ and all $\epsilon > 0$ we have

$$P_{ij}(t + \epsilon, s + \epsilon) = P_{ij}(t, s).$$

In this case the $P_{ij}(t, s)$ depends only on the difference $t - s$, and we write

$$P_{ij}(t) = \mathbb{P}(X_t = j | X_0 = i).$$

All Markov chains in this paper are assumed to be homogeneous. The matrix valued function $P_{ij}(t)$ is called the *transition function* of the Markov chain X.

3 Properties of the Transition Function

There are a few properties transition function $P_{ij}(t)$ which are in some sense essential; that is, given a matrix valued function $Q_{ij}(t)$ with these properties, it is possible to create a Markov chain Z with $Q_{ij}(t)$ as its transition function. These are as follows:

1) $P_{ij}(t) \geq 0$ for all states i, j and

$$\sum_{j \in E} P_{ij}(t) = 1.$$

This rule states that after t time, the Markov chain is in some state of the state space E.

2) $P_{ij}(0) = \delta_{ij}$. Here the δ represents the Kronecker delta which is 1 when $i = j$ and 0 elsewhere. This says that if you start at state i, you cannot start also at some state j distinct from i.

3) The Chapman-Kolmogorov or semigroup property states that:

$$P_{ij}(s+t) = \sum_{k \in E} P_{ik}(s) P_{kj}(t).$$

The interpretation here is that to pass from state i to state j in $s+t$ time, you must first pass from state i to another state k in s time, and then pass from that state k to state j in k time. The algebraic term semigroup indicates that we can regard the pairs (i, j), (i, k), and (k, j) as differences in some semigroup and then we see some cancellation

$$i - j = (i - k) + (k - j).$$

Now we sketch the proof that any function P_{ij} which satisfies the three rules (1) - (3) must be a transition function. Note that as the function encodes first order or "rate"-type information, we'll need some initial condition. In this case, the initial condition is a vector $(p_i)_{i \in E}$ which describes the initial distribution of the states. Thus we have $p_i \geq 0$ and $\sum p_i = 1$.

Let Ω denote the set of all functions from $[0, \infty)$ to E, the state space. We have for every $t \in [0, \infty)$ the evaluation function which carries $\omega \in \Omega$ to $\omega(t) \in E$. If $A \subset \Omega$ is of the form

$$A = \{\omega \in \Omega | \omega(t_1) \in E_1, \ldots, \omega(t_n) \in E_n\}$$

where $0 \leq t_1 < t_2 < \ldots < t_n$, and $E_1, \ldots E_n \subset E$, then we call A a finite dimensional rectangle with respect to the times $\{t_i\}$ and the sets $\{E_i\}$. If A is a finite dimensional rectangle with respect to a family of singletons, i.e. $E_k = \{i_k\}$, then we say that A is narrow, and define

$$P_i(A) = \prod_{m=1}^{n} P_{i_{m-1}, i_m}(t_m - t_{m-1}),$$

where $t_0 = 0$ and $i_0 = i$. Basically $P_i(A)$ determines the probability that the chain starts at the state i and then follows through the path dictated by the finite rectangle A. Note that if A is any finite dimensional rectangle, then it can be decomposed as the union of a countable set of narrow finite dimensional rectangles, because a finite product of countable sets is countable.

Thus if A is decomposed into narrow rectangles $A_1, \ldots$, we can define

$$P_i(A) = \sum_{k=1}^{\infty} P_i(A_k).$$

Let $\mathcal{F}_0$ denote the class of all finite unions of disjoint finite rectangles. Then $\mathcal{F}_0$ is an algebra of subsets, because taking a union of disjoint finite rectangles yields another union of disjoint finite rectangles. The closure under complementation is slightly trickier–basically to guarantee that you are not within the narrow rectangle $\{(t_k, i_k)|k = 1, \ldots, n\}$, you have to have one $j \in [n]$ such that $X(t_j) \neq i_j$, and this set can be decomposed via the inclusion exclusion principle into a finite union of disjoint sets.

Now we note that we have all the ingredients needed to construct a measure via Caratheodory's theorem (see Folland for details). Caratheodory's theorem gives us a σ-algebra $\mathcal{F}$ which contains $\mathcal{F}_0$. As specified in the construction, we see that for the 1-dimensional rectangle $\{(0, i), (t, j)\}$ we have probability $P_{ij}(t)$. Now given an arbitrary element F of $\mathcal{F}$, we set

$$\mathbb{P}(F) \sum_{i \in E} p_i P_i(F)$$

where the numbers p_i are the initial probability distribution. Note that intuitively this says that the likelihood that F "happens" is the sum of the likelihoods that the chain starts in some state i and manages to run through F, as i ranges over E. This gives us a probability space. The Markov chain is then the family of evaluation functions $X(t)$ which take a function $\omega \in \Omega$ and return $\omega(t)$.

Note that the Markov chain property is satisfied, because the probability that $X(t)$ is i is simply the measure of some finite rectangles. These conditional probability $\mathbb{P}(X(t_k) = i_k|X(t_{k-1}) = i_{k-1}, \ldots, X(t_1) = i_1)$ can be directly computed and the terms involving $t_{k-2}, \ldots, t_1$ all vanish on division.

This completes the connection between transition functions and continuous time Markov chains. Thus we will sometimes switch between the two terms.

4 The q-matrix

The definition above is very abstract and non-computational, not unlike the definition of a topological manifold. However, just as in topology we can specify a manifold more concretely by giving the transition functions, we can specify a continuous time Markov chain by determining certain infinitesimal data.

We say that a transition function $P_{ij}(t)$ is standard if $\lim_{t\to 0} P_{ii}(t) = 1$. This seems fairly mild in practice–it indicates only that we can guarantee with arbitrary probability that the chain will stay put if we use a small enough time window. For our purposes that matrices $P_{ij}(t)$ will all be stochastic, which is to say that their row sums will all equal 1.

Then we have a proposition (Proposition 2.2 in Anderson p.9) which states that the functions $P_{ii}(t)$ are differentiable at 0 (although possibly with infinite derivative). We say that $P_{ii}(t) := -q_i$. Then the state i is stable is $q_i < +\infty$ and instantaneous elsewise. The number q_i represents the speed with which the chain to move away from state i, given that it started there. Thus, an instantaneous state is one from which the chain desires to move immediately. An absorbing state is a state i such that $q_i = 0$, i.e. one in which the chain displays no desire to move away from the state i.

Another theorem of Anderson (Prop 2.4) states that if i is a stable state then $P'_{ij}(t)$ exists and is continuous on $[0, \infty)$. The matrix $Q = [q_{ij}] = [P'_{ij}(0)]$ is called the q-matrix or the infinitesimal generator of the Markov chain.

Now as before we describe how to take a potential q-matrix and construct a Markov chain out of it.

A q-matrix $Q = [q_{ij}]$ is a real matrix (possibly countable dimensional) which satisfies

$$q_{ii} \leq 0$$

$$q_{ij} \geq 0, j \neq i$$

$$\sum_j q_{ij} = 0.$$

Let $q_i = \sum_{j \neq i} q_{ij}$. As before we want this to represent the total tendency of the chain to leap out of the state i.

Now we define a "jump matrix" which will allow us to discretize the problem. If $q_i = 0$, then let $J_{ij} = \delta_{ij}$. If $q_i \neq 0$, then let $J_{ij} = q_{ij}/q_i$. Now we can describe a Markov process as follows: take the jump matrix J and use it run a discrete -time Markov chain X^J. This gives us a sequence $i_0, i_1, \ldots$ of states visited by X^J. Between jumps, say at state i_k, we insert a holding-time given by an exponential distribution of parameter q_{i_k}. The memoryless property of the exponential insures that this is in fact a Markov process.

5 Kolmogorov Backward and Forward Equations

This section presents more analysis of the transition function via the q-matrix. The Kolmogorov backward equation takes the form

$$P_{ij}(t) = \delta_{ij} e^{-q_i t} + \int_0^t e^{-q_i s} \sum_{k \neq i} q_{ik} P_{kj}(t-s) ds$$

where $t \geq 0$ and i, j are simply states of the Markov chain. A proposition in Anderson states that if P is the transition matrix of a continuous time Markov chain, with affiliated q-matrix $Q = [q_{ij}]$, then P satisfies the Kolmogorov equation. Initially he proves this with analysis, but there is an appealing "literal" interpretation.

We are trying to measure the likelihood that the chain starts in i and ends in state j at time t. The question is boring when $q_i = 0$, for if $j \neq i$ then both sides of the equation are equal to 0, and if $j = i$ then both sides of the equation are equal to 1. Therefore assume that $q_i > 0$. If $i \neq j$ then at some point $s < t$ we must have the chain leaving, after living in state i with parameter q_i. This accounts for the density term $q_i e^{-q_i s}$. After leaving the state i there is an immediate transition (via the jump matrix q_{ik}/qi)) to a new state $k \neq i$. After the chain hits state k at time s, there is $t - s$ time remaining for the chain to get back, and the likelihood of this occurring is $P_{kj}(t - s)$. In the case where $i = j$, we can either pursue this jaunt through the states, or possibly stay in the state i for the entire interval $[0, t]$ with probability given by the holding time $e^{-q_i t}$.

There are an analogous set of equations known as the Kolmogorov forward equations which describe the evolution of "nice" Markov chains in which there are only finitely many jumps over a finite time interval. These take the form:

$$P_{ij}(t) = \delta_{ij} e^{-q_i t} + \int_0^t e^{-q_i s} \sum_{k \neq j} P_{ik}(t - s) q_{kj} ds.$$

These also have a nice literal interpretation. In order for the chain to land at j, it must reach some penultimate state other than j, say k. The $q_{kj}ds$ represents the likelihood of jumping from state k to state j duing a tiny window at time $t - s$. The quantity $P_{ik}(t - s)$ represents the likelihood of landing in the eligible state $k \neq j$ after time $t - s$ has elapsed. The quantity $e^{-q_j s}$ represents the likelihood that after the chain makes the jump from k to j, with s time left to spare, that it stays put and doesn't leave state j. The $\delta_{ij} e^{-q_i t}$ term, as before, represents the probability that the chain never moves.

6 Bibliography

Aldous, David. Reversible Markov Chains and Random Walks on Graphs. 1999. http://www.stat.berkel dous/RWG/book.html

Anderson, William J. *Continuous Time Markov Chains.* 1991. Springer Verlag. New York.

Feres, Renato. Notes for Math 450: Continuous Time Markov Chain and Stochastic Simulation. 2007. http://www.math.wustl.edu/ feres/Math450syll.html

Folland, Gerald. Real Analysis: Modern Techniques and Their Applications. 1999. Wiley. New York.