Continuous Time Markov Processes

E.H.F.

2004

Updated September 2009. Thanks for Tomi Silander for finding a few mistakes in the original draft.

1 Introduction

We discuss continuous time Markov processes as both a method for sampling an equilibrium distribution and simulating a dynamical system. We begin with a brief review of an exponentially distributed random variable. Then we define a Markov jump process and derive the Chapman Kolmolgorov and Master equation. The limiting properties of these systems are similar to those of a Markov chain, and these results are stated. The relationship between a Markov chain and jump process is then developed, and finally we use examples in statistical mechanics to compare and contrast these two methods. This article provides the mathematical foundation for the often used continuous time Monte Carlo simulation (see *Monte Carlo Methods in Statistical Physics* by Newman and Barkema).

2 Exponential Distribution

We define a random variable X on $[0, \infty)$ that represents the time interval to the transition of a process. For example, X could represent the lifetime of a light bulb, and the transition is the burning out of the light bulb. We assume this random variable has the memoryless property,

$$P(X > t + s | X > s) = P(X > t) \tag{1}$$

which implies that if the transition has not occurred at time s, then its probability of the transition at time $t+s$ is the same as the probability of the transition at time t. Starting the process at time s given that it had not made a transition by time s is equivalent to starting the process at time zero. The process doesn't "remember" that it has not made a transition yet. For any random variable,

$$P(X > t+s|X > s)P(X > s) = P(X > t+s, X > s) = P(X > t+s) \quad (2)$$

where the second equality holds since the object must be alive at time s if it's alive at time $t+s$. Then the memoryless property in Eqn. 1 implies that

$$P(X > t)P(X > s) = P(X > t+s) \quad (3)$$

If we define $G(t) \equiv P(X > t)$, then $G(t+s) = G(t)G(s)$. The exponential function $G(t) = e^{-\gamma t}$ satisfies this relationship, so then $P(X > t) = e^{-\gamma t}$ for $\gamma > 0$, the distribution function is

$$F_\gamma(t) \equiv P(X \leq t) = 1 - e^{-\gamma t} \quad (4)$$

and the probability density is $f(t) = \gamma e^{-\gamma t}$. This defines X as a random variable with an exponential distribution with parameter γ, and we write $X \sim \exp(\gamma)$. If t is time, then the parameter γ has units of inverse time, and one can think of this as the transition rate. A straightforward integration by parts shows that the expected value of $X \sim \exp(\gamma)$ is γ^{-1}.

2.1 Discrete version of this process

Let's consider a discrete version of this random variable with the memoryless property. We define Y as a random variable on the positive integers. Then $Y = n$ means that the transition occurs at the n step of the process. If we associate a time interval Δt with a step in the process, then $Y = n$ can be associated with the time $n\Delta t$. If the process has a probability p of making a transition at each integer time, then

$$P(Y = n) = (1-p)^{n-1}p \quad (5)$$

since there is no transition at the previous $n-1$ time steps. A random variable with this density has a geometric distribution, and we write $Y \sim \text{geo}(p)$. This discrete random variable has the memoryless property because the transition

can occur at any step with equal probability. The expected value of this random variable is

$$\begin{aligned} EY &= \sum_{n=1}^{\infty} n(1-p)^{n-1}p \\ &= p\sum_{n=1}^{\infty}\left(-\frac{d}{dp}(1-p)^n\right) = -p\frac{d}{dp}\left(\sum_{n=1}^{\infty}(1-p)^n\right) = -p\frac{d}{dp}\frac{1-p}{p} = \frac{1}{p} \end{aligned}$$

so p is a transition rate, or the probability of a transition per time Δt.

For the probability density of this distribution, we can write

$$P(Y=n) = pe^{(n-1)\ln(1-p)} \tag{6}$$

so the time dependence is exponential with rate $[\ln(1-p)]^{-1}$. For small values of p, $\ln(1-p) \approx -p$, so then probability density looks like

$$pe^{-pn} \tag{7}$$

for large n where p is the rate of the process. This implies that there is a relationship between this discrete random variable $Y \sim \text{geo}(p)$ and the exponentially distributed random variable $X \sim \exp(\gamma)$. Specifically, let an integer n correspond to the time $t = n$. Then Y is a good approximation for X when $\gamma \ll 1$ and $n \gg 1$. The exponentially distributed random variable is the continuous analog of a geometrically distributed random variable, and $X \sim \exp(\gamma)$ is a continuous time process that has the same transition probability at every time.

For an exponentially distributed random variable with arbitrary rate γ, we let $p = \gamma/N$ for some integer N so that $p \ll 1$. By dividing the rate by this integer, we have multiplied the characteristic time of the discrete process by the same integer; it takes many more steps for the density of the discrete process to decay. Then if we associate $n = 1$ with $t = 1/N$ and $n = N$ with $t = 1$, and the random variable $Y \sim \text{geo}(\gamma/N)$ is a good approximation to $X \sim \exp(\gamma)$. For certain parameters, a discrete process can be a good approximation to a continuous time process. There will be a similar relationship between a Markov chain and Markov jump process.

3 Markov Jump Processes

3.1 Definitions

In the following, we use Γ to denote the set of all states of a system and $x, y \in \Gamma$ to denotes states of the system. A jump process is a random variable $X(t)$ parameterized by time $t \in [0, \infty)$. This random variable starts in an initial state x_0 at time $t = 0$ and stays in this state until some time t_1 when it makes a transitions to a different state x_1. It stays in this state until a later time $t_2 > t_1$ at which is jumps to different state x_2. Then if $t_1, t_2, \ldots$ are the set of jump times, then $X(t) = x_0$ for $t \in [0, t_1)$, $X(t) = x_1$ for $t \in [t_1, t_2)$, and so on. This defines a Markov jump process $X(t)$. We assume that $\lim_{n\to\infty} t_n = \infty$, so the jump process $X(t)$ is defined for all non-negative values of t.

The jump process $X(t)$ changes states by two procedures. We define $r(x, y)$ as the transition probability from the state y to state x such that $r(x, x) = 0$ for all $x \in \Gamma$ and

$$\sum_x r(x, y) = 1. \tag{8}$$

If the jump process is in state y, then it makes a transition to a new state x according to $r(x, y)$. Once the process is in a state x, the time that it stays in this state is a random variable governed by the distribution function $F_x(t)$. Each state x can have a different distribution $F_x(t)$. The time spent in state x and the choice of the next state y are independent random variables, so then

$$P(\tau \leq t, X(\tau) = y, X(0) = x) = r(y, x)F_x(t). \tag{9}$$

We assume that the transition probability $r(y, x)$ is independent of time and the distribution functions $F_x(t)$ for all $x \in \Gamma$ remain the same throughout the jump process.

We define $p(y, x; t)$ as the conditional probability that the jump process is in state y at time t given that it was in state x at time 0. This conditional probability plays the same role in a jump process as the transition probability $p(y, x)$ played in Markov chains.

3.2 Markov Property

Given times $0 < t_1 < t_2 < \ldots < t_n < s$ and $t > 0$, the Markov property for a jump process is

$$P(X(t+s) = y | X(s) = x, X(t_n) = x_n, \ldots, X(t_1) = x_1) = p(y, x; t) \quad (10)$$

or that the conditional probability of state y at a time t given the states of the jump process at $n+1$ previous times only depends on the state x at the most recent time s and the time interval t to the present time $t+s$. Given a set of previous states at earlier times, the Markov jump process 'forgets' all but the state at the most recent time. The jump process starts all over again at this most recent time s. It is possible to prove that a jump process is a Markov jump process if and only if $F_x(t)$ is the exponential distribution for all $x \in \Gamma$, or that

$$F_x(t) = 1 - e^{-\gamma_x t} \quad (11)$$

where γ_x is the rate of the exponential distribution for state x. The probability density for this distribution is $f_x(t) = \gamma_x e^{-\gamma_x t}$. Intuitively, this makes sense since the Markov property is similar to the memoryless property of the exponential distribution. The form of Eqn. 10 is similar to Eqn. 1 in that a conditional probability at an arbitrary time is related to a condition probability at time zero. With the Markov property, the jump process is "doing the same thing" at every time. One can see how this might be a useful model for systems at equilibrium.

3.3 Chapman-Kolmolgorov and Master Equations

With the Markov property, joint probabilities can be written as

$$P(X(t+s) = x, X(s) = y, X(0) = z) = p(x, y; t)p(y, z; s)P(X(0) = z) \quad (12)$$

for $t, s > 0$ with the obvious generalization to higher order joint probabilities. We can derive a useful equation for the conditional probabilities $p(x, y; t)$ by summing the previous equation over y:

$$P(X(t+s) = x, X(0) = z) = \sum_y p(x, y; t)p(y, z; s)P(X(0) = z). \quad (13)$$

Since $p(x, z; t+s) \equiv P(X(t+s) = x, X(0) = z)/P(X(0) = z)$ we obtain

$$p(x, z; t+s) = \sum_y p(x, y; t)p(y, z; s). \quad (14)$$

This is the Chapman-Kolmolgorov equation, and a Markov jump process must satisfy this equation. This equation is analogous to the Chapman-Kolmolgorov equations in polymer physics used to derive diffusion equations for single chain partition functions.

In most cases, it is easier to work with a differential equation that is equivalent to the Chapman-Kolmolgorov equation. To derive such an equation, consider the conditional probability $p(x, y; \Delta t)$ for small Δt. We choose Δt small enough so that the Markov jump process can make at most one transition in the time interval Δt. If τ is the time the jump process stays in state y, then the probability that the jump process moves from y to x in this time interval Δt is

$$P(\tau \leq \Delta t) r(x, y) = F_y(\Delta t) r(x, y) = \Big(1 - e^{-\gamma_y \Delta t}\Big) r(x, y) \approx \gamma_y r(x, y) \Delta t \quad (15)$$

where in the last step we consider small Δt. We define $w(x, y) \equiv \gamma_y r(x, y)$ as the transition rate from y to x, or the probability of a transition from y to x in a time interval Δt. Then $w(x, y)\Delta t$ is the probability that the process moves from y to x in the time interval Δt. The conditional probability is

$$p(x, y; \Delta t) \approx \left(1 - \sum_{x'} w(x', y) \Delta t\right) \delta(x, y) + w(x, y) \Delta t + O(\Delta t^2) \quad (16)$$

for small Δt. The term in the parentheses is the probability of staying in the state y after the time interval, and there is an unrestricted summation over x' since $w(x, x) = 0$. While we have not justified this rigorously, it should be intuitively clear. Inserting this expression into Eqn. 14 gives

$$p(x, z; s + \Delta t) = \sum_y \left[\left(1 - \sum_{x'} w(x', y) \Delta t\right) \delta(x, y) + w(x, y) \Delta t\right] p(y, z; s) \quad (17)$$

and performing the summation over y in the first two terms and rearranging gives

$$\frac{1}{\Delta t}\Big(p(x, z; s + \Delta t) - p(x, z; s)\Big) = -\sum_{x'} w(x', x) p(x, z; s) + \sum_y w(x, y) p(y, z; s). \quad (18)$$

Taking the limit as $\Delta t \to 0$ gives

$$\frac{d}{ds} p(x, z; s) = \sum_y \Big(w(x, y) p(y, z; s) - w(y, x) p(x, z; s)\Big) \quad (19)$$

which is the Master equation. The first term on right of the equation is gain in probability from transitions into state x while the second term is the loss in probability from transitions from state x to other states. The rate of change in the conditional probability depends on the balance of these two terms. In the mathematics literature, this equation is the *forwards equation*.[1] If we define

$$\ell^T(x,y) \equiv w(x,y) - \delta(x,y)\sum_z w(z,y) \tag{20}$$

when we can express the Master equation as

$$\frac{d}{ds}p(x,z;s) = \sum_y \ell^T(x,y)p(y,z;s) \tag{21}$$

In the mathematics literature, $\ell^T(x,y)$ for all $x,y \in \Gamma$ are called the infinitesimal parameters of the process.

3.4 Properties

A stationary distribution for a Markov jump is a distribution $b^*(x)$ such that for all states x and times t

$$\sum_y p(x,y;t)b^*(y) = b^*(x). \tag{22}$$

If we take the time derivative of this equation and carelessly exchange the summation and integration, then

$$\sum_y \frac{d}{dt}p(x,y;t)b^*(y) = \sum_y \ell^T(x,y)b^*(y) = 0 \tag{23}$$

so the stationary distribution must be in the null space of $\ell^T(x,y)$. We expect that there are similar properties in the $t \to \infty$ limit for Markov jump processes as in the $n \to \infty$ limit for Markov chains. A Markov jump process is *irreducible* if for all $x,y \in \Gamma$ the probability to start in state x and reach state y in finite time is nonzero. A Markov jump process is *positive recurrent* if for all $x \in \Gamma$ the mean return time is finite. It is possible to prove that there exists a stationary distribution for irreducible, positive

[1] There is a backwards equation where the transition rates "act" on the earlier state y in $p(x,y;t)$.

recurrent Markov jump processes. There is no such thing as an aperiodic Markov jump process, since the process exists for all positive time and not just integer times. It is possible to prove that

$$\lim_{t\to\infty} p(x,y;t) = b^*(x) \tag{24}$$

so a Markov jump process goes to its stationary distribution in the limit of long times. This result is independent of the initial distribution.

4 Relationship between Markov chains and jump processes

In this section, we show that for a Markov chain with a transition probability $p(y,x)$ we can construct a corresponding Markov jump process. First, consider the probability that a Markov chain remains in state x for n steps. If $p(x,x) = 0$, then the only possibility is $n = 0$; it must always make a transition and never stays in the same state. Otherwise, the probability is

$$p(x,x)^n = e^{n\ln p(x,x)} = e^{n\ln(1-p_g(x))} \tag{25}$$

where $p_g(x) \equiv \sum_{y(\neq x)} p(y,x)$ is the probability of leaving the state x. For $p_g(x) \ll 1$, we obtain that $\ln(1-p_g(x)) \approx -p_g(x)$ so then this probability is exponentially distributed with rate $p_g(x)$. Note that we always have the condition $p_g(x) \leq 1$.

This suggests that we define a Markov jump process by

$$r(y,x) \equiv \frac{p(y,x)}{p_g(x)} \tag{26}$$

for $x \neq y$ so that this quantity is normalized and $\gamma_x \equiv \gamma p_g(x)$ where γ^{-1} has units of time. This defines the time scale of the jump process, and the choice $\gamma = 1$ makes the n step of the Markov chain equivalent to $t = n$ in the Markov jump process. Since $p_g(x) \leq 1$, then a random variable $X \sim \exp(\gamma_x)$ has expectation $EX \geq \gamma_x^{-1}$. For the transition probability, we obtain $w(y,x) = \gamma p(y,x)$ so then

$$\ell^T(x,y) = \gamma\Big(1-\delta(x,y)\Big)p(x,y) - \gamma\delta(x,y)\sum_{z(\neq x)} p(z,y). \tag{27}$$

Now the gain and loss of probability in the Master equation is related to the transition probabilities $p(y, x)$ for $x \neq y$ of the Markov chain. We can rewrite this as

$$\ell^T(x, y) = \gamma p(x, y) - \gamma\delta(x, y)\sum_z p(z, y) = \gamma p(x, y) - \gamma\delta(x, y) \tag{28}$$

from which it is straightforward to see that a stationary distribution $b^*(y)$ that satisfies $\sum_y \ell^T(x, y)b^*(y) = 0$ must also satisfy $\sum_y p(x, y)b^*(y) = b^*(x)$. A stationary distribution for the Markov jump process we have defined is the same as the stationary distribution for the original Markov chain. So then a Markov jump process can also be used to sample equilibrium distributions. It is particularly useful as $p(x, x)$ approaches 1, or when the rejection rate in the Metropolis algorithm is large; a Monte Carlo simulation will be stuck in certain states often. We will give examples in the next section.

Moreover, Markov jump process is often the definition of a dynamical model for a statistical mechanical system at equilibrium. To construct the model, one defines transition rates $w(y, x)$ that satisfy the detailed balance and assumes the dynamics of the system is a Markov process governed by the Master equation. A Markov chain is an approximation to the Markov jump process that is the actual dynamics of the system, much like the geometrically distributed discrete random variable is an approximation to the continuous time exponentially distributed random variable.

5 Examples

5.1 Simple Lattice Gas

Consider a lattice gas with excluded volume interactions. A state of the system is given by

$$x = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\} \tag{29}$$

where $\mathbf{r}_i$ is the position of the i particle on a simple d dimensional lattice. Particles can't occupy the same site; otherwise, there are no interactions between the particles. The equilibrium statistics of this excluded volume lattice gas are trivial, as only states in which all particles occupy different sites are allowed, and all allowed states have the same probability. We construct a Markov chain with the transition probability

$$p(x, y) = t(x, y)\min\left(1, \frac{b^*(x)}{b^*(y)}\right) \tag{30}$$

where $t(x, y) = (2dN)^{-1}$ for a pair of allowed states x and y that differ by the position of one particle by one lattice space. This is a normalized, symmetric trial probability $t(x, y)$. With the simple interactions of this model, the min function is 1 if state x has nonzero probability and 0 otherwise. An equivalent Markov jump process is obtained by setting

$$w(x, y) = \gamma 2dNp(x, y) \tag{31}$$

where γ^{-1} is the basic time unit. We choose a factor of $2dN$ so that γ is the rate of one particle to move in one direction. Our interest in constructing a Markov process for this model is to study dynamics, since it is possible to sample the equilibrium distribution by other, more simple means. Still, it is a useful example to compare a Markov chain with a Markov jump process simulation.

Consider the Metropolis algorithm that constructs a Markov chain:

- pick a particle at random where each particle has an equal probability
- pick a direction at random where each direction has an equal probability
- if there is a different particle at the new position of the chosen particle, reject the move; otherwise, accept the move

The algorithm is simple, and this is the main advantage of Metropolis type algorithms. The Markov chain also gives dynamical information. The main disadvantage is that the rejection rate becomes large as the fraction of sites occupied by particles increases.

The definition of a Markov jump process is a prescription for performing a simulation of a jump process. If we are in an initial state x, pick a time interval to the next step from an exponential distribution with rate γ_x, pick a new state from the transition probability $r(y, x)$, and update both the time and the new state. The iteration of these steps simulates the Markov jump process. The complicated part of the algorithm is determining γ_x for each state x. This can be done constructing a data structure that contains information about the allowed transitions for each particle; each direction in which each particle can move contributes a factor of γ to the total rate γ_x. We will call this data structure the transition structure. Now suppose we have an initial state of the system. First, we initialize the transition structure. Then the Markov jump process can be simulated by the following algorithm:

- pick a time interval from an exponential distribution with rate γ_x
- pick any one of the allowed transitions with equal probability
- update the time of the system by the chosen time interval and change to state y
- update transition structure to reflect the allowed transitions of state y

This algorithm is more difficult to implement because of this last step. This can take quite a bit of computation. The benefit is that we have a much more efficient algorithm for systems with a large fraction of particles. In this situation, γ_x is on average small, so then the time intervals to the next step are on average large.

5.2 Lattice Gas with Interactions

Now consider a lattice gas with attractive interactions in addition to the excluded volume interaction. We have the same allowed states as previously, but now each is weighted by Boltzmann factor $e^{-\mathcal{H}(x)}$ where the Hamiltonian is

$$\mathcal{H}(x) = -\frac{\epsilon}{2}\sum_i N_i(x) \tag{32}$$

where $\epsilon > 0$ and $N_i(x)$ is the number of particles that occupy the neighboring sites of particle i in state x. For each pair of particles that occupy neighboring sites, we obtain an energy decrease of $-\epsilon$. This model has non-trivial equilibrium statistics, and we are interested in sampling this equilibrium distribution with a Markov jump process.

To specify the Markov jump process, we must define the transition rates $w(y, x)$. Just as in the excluded volume lattice gas, we allow transitions between states that differ by the position of one particle by one lattice space, so then we have the same trial probability $t(y, x)$. To construct the transition rates, consider the corresponding transition probability for a Markov chain which must satisfy the detailed balance condition:

$$\frac{p(y,x)}{p(x,y)} = \frac{b^*(y)}{b^*(x)} = e^{-[\mathcal{H}(y)-\mathcal{H}(x)]} \tag{33}$$

For any pair of states x and y such that $t(y, x) > 0$,

$$\frac{p(y,x)}{p(x,y)} = e^{+\epsilon N_i(y)-\epsilon N_i(x)} \tag{34}$$

where i is the particle whose position differs. This suggests that we choose

$$p(y,x) \approx e^{-\epsilon N_i(x)} \tag{35}$$

so then our Markov jump process is defined by the transition rates

$$w(y,x) = \gamma[2dN]t(y,x)e^{-\epsilon N_i(x)} \tag{36}$$

The advantage of this choice is that the rates only depend on the initial state x. This formula also makes physical sense, since there is a lower rate to make transitions out of low energy states, or states in which $N_i(x)$ is large.

Our algorithm to simulate the Markov jump process for this system is the following:

- in any state x, pick a random time step according to an exponential distribution with rate $\gamma_x = \sum_y w(y,x)$.
- pick a new state y with the probability

$$\frac{t(y,x)e^{-\epsilon N_i(x)}}{\sum_y t(y,x)e^{-\epsilon N_i(x)}} \tag{37}$$

- update the new time and new state, and update the transition structure

For this system, the transition structure is more complicated, since we must store a real number as the rate for each transition. This Markov jump process also specifies dynamics for a lattice gas with attractive interactions.