Discrete Markov Chains - Continuous Dynamics

June 18, 2015

Markov Processes in Continuous Space

Markov processes are not limited to the time-discrete and space-discrete case that we have considered up to this point. In fact, it is probably a lot more accurate to model real-world phenomena by random processes that are continuous in both space and time. Let us consider a stochastic process X_t for continuous times $t \geq 0$ that maps into a continuous state space Ω. Such a process satisfies the Markov property if for all $x_0, x_1, \ldots, x_n$ and $y \in \Omega$, $0 \leq t_0 < t_1 < \ldots < t_n$ and $\tau > 0$, we have

$$\mathbb{P}\left(X_{t_n+\tau} = y | X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\right) \quad = \quad \mathbb{P}\left(X_{t_n+\tau} = y | X_{t_n} = x_n\right) \tag{1}$$

As before, we assume that the transition probabilities on the right-hand side of Eq. (1) are time-homogeneous, i.e. they only depend on τ. This allows to define a transition probability distribution

$$\mathbb{P}(X_\tau \in C | X_0 = x),$$

for all $x \in \Omega$ and subsets $C \subset \Omega$. We also assume that these distributions have a smooth density function $p(x, y; \tau)$, s.t.

$$\mathbb{P}(X_\tau \in C | X_0 = x) \quad = \quad \int_C p(x, y; \tau)\, \mathrm{dy}.$$

This **transition density function** takes over the role of the transition matrix from the discrete case.

Examples

The simplest continuous Markov process is called Brownian motion

Definition 1. A one-dimensional stochastic process W_t defined for $t \geq 0$, is called a **Brownian motion** or a Wiener process if it satisfies:

1. $W_0 = 0$.

2. For any times $t_1, \ldots, t_k$, the increments $W_{t_k} - W_{t_{k-1}}, \ldots, W_{t_2} - W_{t_1}$ are independent random variables.

3. The increments $W_t - W_s$ are normally distributed according to $W_t - W_s \sim \mathcal{N}(0, \sqrt{t-s})$.

The transition density function of Brownian motion is just a Gaussian centred at x:

$$p(x, y; \tau) \quad = \quad \frac{1}{(2\pi\tau)^{1/2}} \exp(-\frac{1}{2\tau}(y-x)^2).$$

The intuition is that from the current position x of the process, the next position y at a short time step τ later is drawn from a Gaussian distribution with variance equal to the time step, centred at x. At least for a small time step τ, Brownian motion can indeed be sampled like this, to a good approximation:

1. Set $x_0 = 0$.

2. For $n = 1, \ldots$, draw x_n from a normal distribution with standard deviation $\tau^{1/2}$ and mean value x_{n-1}.

A more complicated example is given by **Brownian dynamics**. This process models the superposition of a deterministic motion with stochastic pertubations. Suppose that there is a potential energy function V and the change of a process X_t over an infinitesimal time Δt can be described by

$$\Delta X_t \quad = \quad -\Delta t \nabla V(X_t) + \sigma W_t.$$

The change of the process is determined by the negative gradient of the energy, driving the system towards the minima of V, plus a stochastic noise in the form of Brownian motion. One might be tempted to cast this into a differential equation via

$$\frac{dX_t}{dt} \quad = \quad -\nabla V(X_t) + \sigma dW_t.$$

It is unclear what the differential dW_t of Brownian motion should mean. Stochastic calculus can define this equation rigorously in an integral sense. For our purposes, it is sufficient to know that this equation defines a time- and state-continuous Markov process that can be approximated for small time steps τvia:

1. Choose some x_0.

2. For $n = 1, \ldots$, draw x_n from a normal distribution with standard deviation $\sigma\tau^{1/2}$ and mean value $x_{n-1} - \tau\nabla V(x_{n-1})$.

For more details, see e.g. the textbook [3].

Propagator

Many of the concepts we have derived for discrete dynamics can be transferred to the continuous case (see [4]).

- **Density propagation:** Suppose that, at time $t = 0$, the system is distributed according to a distribution p_0. Then, the corresponding density at time $\tau > 0$ can be computed by

$$p_\tau(y) \quad = \quad \int_\Omega \mathrm{dx}\, p(x, y; \tau)\, p_0(x).$$

- **Stationary density:** If the process is sufficiently ergodic, which means in essence that the system cannot be decomposed into two dynamically independent components, and every state of the system will be visited infinitely often over an infinite run of the process, there is a unique probability distribution π which is invariant in time, i.e.

$$\pi(y) \quad = \quad \int_\Omega \mathrm{dx}\, p(x, y; \tau)\, \pi(x).$$

- **Detailed balance:** The detailed balance condition is written as

$$\pi(x)p(x, y; \tau) \quad = \quad \pi(y)p(y, x; \tau).$$

- **Propagator:** The time evolution defines a linear operator, called the propagator $\mathcal{P}(\tau)$:

$$\mathcal{P}(\tau)p(y) \quad = \quad \int_\Omega \mathrm{dx}\, p(x, y; \tau)\, p(x).$$

  The propagator satisfies the Chapman-Kolmogrov equation

$$\mathcal{P}(\tau_1 + \tau_2) \quad = \quad \mathcal{P}(\tau_1)\mathcal{P}(\tau_2).$$

  For detailed balance, the propagator is linear and self-adjoint w.r.t. the inner product

$$\langle f, g\rangle_{\pi^{-1}} \quad = \quad \int_\Omega \mathrm{dx}\, f(x)\, g(x)\, \pi^{-1}(x). \tag{2}$$

- **Eigenvalues:** Self-adjointness and compactness imply that the eigenvalues of the propagator are real-valued, discrete and bounded by the greatest eigenvalue $\lambda_1 = 1$. Due to the Markov property and the Chapman-Kolmogrov equation, the eigenvalues also satisfy

$$\begin{aligned} \lambda_m(\tau_1 + \tau_2) \quad &= \quad \lambda_m(\tau_1)\lambda_m(\tau_2) \\ \lambda_m(0) \quad &= \quad 1. \end{aligned}$$

  It follows that all eigenvalues are non-negative and can be written as exponential decays, with some rates $0 = \kappa_1 > \kappa_2 > \kappa_3 \ldots$:

$$\lambda_m(\tau) \quad = \quad e^{\kappa_m \tau}.$$

- **Spectral Decomposition:** To each eigenvalue, we can assign eigenfunctions $\pi = \phi_1, \phi_2, \ldots$. They are orthonormal w.r.t. the inner product Eq. (2), and the operator can be written in terms of the eigenfunctions via

$$\mathcal{P}(\tau)p \quad = \quad \sum_{m=1}^{\infty} e^{\kappa_m \tau} \langle p, \phi_m \rangle_{\pi^{-1}} \phi_m.$$

  This decomposition implies convergence of distributions towards the stationary density. Also, rates close to zero give rise to metastability of the system.

Discretization

How can we discretize the operator $\mathcal{P}(\tau)$? One approach is to compute an approximation of its eigenfunctions in terms of finitely many trial functions. For convenience, we will formulate all following results in terms of the **transfer operator**

$$\mathcal{T}(\tau)f(y) \quad = \quad \frac{1}{\pi(y)} \int_{\Omega} \mathrm{dx}\, p(x, y; \tau)\, \pi(x)\, f(x).$$

The transfer operator is equivalent to the propagator in the sense that it shares the eigenvalues with $\mathcal{P}(\tau)$, and its eigenfunctions ψ_m can be computed from the ϕ_m by $\psi_m = \pi^{-1}\phi_m$. The propagator eigenfunctions correspond to the left eigenvectors of a transition matrix, the others to the right eigenvectors. The functions ψ_m are orthonormal w.r.t. the inner product

$$\langle f, g \rangle_{\pi} \quad = \quad \int_{\Omega} \mathrm{dx}\, f(x)\, g(x)\, \pi(x).$$

Lemma 2. *For any normalized function $\hat{\phi}$ which is orthogonal to the first k eigenfunctions (possibly $k = 0$), we have:*

$$\langle \mathcal{T}(\tau)\hat{\psi}, \hat{\psi} \rangle_{\pi} \quad \leq \quad \lambda_{k+1}(\tau).$$

Equality holds exactly for the eigenfunction ψ_{k+1}.

Proof. Expand $\hat{\psi}$ in terms of the orthonormal basis ψ_m:

$$
\begin{aligned}
\langle \mathcal{T}(\tau)\hat{\psi}, \hat{\psi}\rangle_\pi &= \sum_{i,j=k+1}^{\infty} c_i c_j \langle \mathcal{T}(\tau)\psi_i, \psi_j\rangle_\pi (\text{Orthonormality to first k eigenfunctions}) \\
&= \sum_{i,j=k+1}^{\infty} c_i c_j \lambda_i(\tau)\langle \psi_i, \psi_j\rangle_\pi (\text{Eigenfunctions}) \\
&= \sum_{i=k+1}^{\infty} c_i^2 \lambda_i(\tau) (\text{Orthonormality of eigenfunctions}) \\
&\leq \lambda_{k+1}(\tau) \sum_{i=k+1}^{\infty} c_i^2 (\text{Ordering of eigenfunctions}) \\
&= \lambda_{k+1}(\tau). (\text{Normalization})
\end{aligned}
$$

□

This result gives rise to the following idea: Given a set of N trial functions $\chi_1, \ldots, \chi_N$, let us compute an approximation of the first M eigenfunctions $\psi_1, \ldots, \psi_m$ by linear combinations of the trial functions. This leads to the
Optimization Problem: *Find M linear combinations*

$$
\hat{\psi}_m = \sum_{i=1}^{N} V_{im}\chi_i,
$$

such that

$$
\sum_{m=1}^{M} \langle \mathcal{T}(\tau)\hat{\psi}_m, \hat{\psi}_m\rangle_\pi \rightarrow \max \tag{3}
$$

$$
\langle \hat{\psi}_m, \hat{\psi}_{m'}\rangle_\pi = \delta_{m,m'}. \tag{4}
$$

Theorem 3. *The optimization problem Eqs. (3) and (4) is solved by the first M eigenvectors of the generalized eigenvalue problem (see [1, 2]):*

$$
\begin{aligned}
\mathbf{C}^\tau \mathbf{V} &= \mathbf{C}^0 \mathbf{V} \hat{\mathbf{\Lambda}}. \qquad (5) \\
\mathbf{C}^\tau_{ij} &= \langle \mathcal{T}(\tau)\chi_i, \chi_j\rangle_\pi \\
\mathbf{C}^0_{ij} &= \langle \chi_i, \chi_j\rangle_\pi.
\end{aligned}
$$

Finally, we check that a discrete Markov model arises naturally if we choose a specific trial space:

Lemma 4. *For a partition of the state space Ω into disjoint sets S_i, $i = 1, \ldots, N$, choose the basis functions as indicator functions of the sets: $\chi_i = \mathbf{1}_{S_i}$. Then, the optimzation problem Eq. (5) becomes a Markov model eigenvalue problem.*

Proof. We evaluate the matrix entries:

$$
\begin{aligned}
\mathbf{C}^{\tau}_{ij} &= \langle \mathcal{T}(\tau)\chi_i, \chi_j \rangle_\pi \\
&= \int_{S_i}\int_{S_j} \pi(x)\, p(x,y;\tau)\,\mathrm{dx}\,\mathrm{dy} \\
&= \mathbb{P}(X_\tau \in S_j,\, X_0 \in S_i) \\
\mathbf{C}^{0}_{ij} &= \delta_{ij} \int_{S_i} \pi(x)\,\mathrm{dx} \\
&= \delta_{ij}\mathbb{P}(X \in S_i).
\end{aligned}
$$

Thus, the left-hand side matrix contains the unconditional jump probabilities between the sets S_i, S_j, while the right-hand side matrix is diagonal and contains the stationary probabilities of the sets. Upon multiplication by the inverse of $\mathbf{C}^0$, we obtain the matrix of transition probabilities between the sets. □

References

[1] F. Noé and F. Nüske. A variational approach to modeling slow processes in stochastic dynamical systems. *SIAM Multiscale Model. Simul.*, 11:635–655, 2013.

[2] F. Nüske, B. Keller, G. Pérez-Hernández, A. S. J. S. Mey, and F. Noé. Variational approach to molecular kinetics. *J. Chem. Theory Comput.*, 10:1739–1752, 2014.

[3] Bernt Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag New York, Inc., 1992.

[4] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.*, 134:174105, 2011.