Fluid limits for Markov chains I

James Norris

Durham Symposium on Markov Processes, Mixing Times and Cutoff 2017

Markov chains

Consider a discrete-time Markov chain $X = (X_n)_{n \geqslant 0}$ with state-space E and transition kernel p.

Thus, for $x \in E$ and $B \subseteq E$,

$$p(x, B) = \mathbb{P}(X_{n+1} \in B | X_n = x).$$

To be precise, our probability space $(\Omega, \mathcal{F}, \mathbb{P})$ will come with a filtration $(\mathcal{F}_n)_{n \geqslant 0}$, and the state-space has a σ-algebra $\mathcal{E}$, such that

- p is a measurable probability kernel on $(E, \mathcal{E})$
- X is $(\mathcal{F}_n)_{n \geqslant 0}$-adapted
- $\mathbb{P}(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B)$ a.s.

Markov chains

Sometimes instead X will be a continuous-time Markov chain with transition kernel q.

Write $q(x, B) = q(x)\pi(x, B)$ with $\pi(x, E) = 1$.

Then $X = (X_t)_{t \geqslant 0}$ is a continuous-time process, which runs through the values of a discrete-time Markov chain with transition kernel π, spending an independent exponential time of parameter $q(x)$ on each visit to x.

Precisely, in this case, there will be a continuous-time filtration $(\mathcal{F}_t)_{t \geqslant 0}$ and

- q is a finite measurable kernel on $(E, \mathcal{E})$
- X is cadlag and $(\mathcal{F}_t)_{t \geqslant 0}$-adapted
- $\mathbb{P}(J_1(t) > t + s \text{ and } X_{J_1(t)} \in B | \mathcal{F}_t) = e^{-q(X_t)s}\pi(X_t, B)$ a.s., where $J_1(t)$ is the time of the first jump by X after time t.

Fluid limits

When and how can we show that X is close to the solution $(x_t)_{t\geqslant 0}$ of a differential equation $\dot{x}_t = b(x_t)$?

In the discrete-time case, we will need to embed the discrete time-scale as $\varepsilon\mathbb{Z}^+$ with ε small.

We will work for now in continuous time, where this time-rescaling can be absorbed into the jump rate $q(x)$.

Fluid limits

Let $(\xi_t)_{t \geqslant 0}$ be a Markov chain in E with transition kernel q.

Choose a vector space V and a *coordinate map* $x : E \to V$.

Set $X_t = x(\xi_t)$ and consider the *drift* $\beta : E \to V$ given by

$$\beta(\xi) = \lim_{t \to 0} \mathbb{E}(X_t - X_0 | \xi_0 = \xi)/t = \int_E \{x(\eta) - x(\xi)\} q(\xi, d\eta).$$

Choose b so that $b(x(\xi)) = \beta(\xi)$ and solve $\dot{x}_t = b(x_t)$. Thus

$$b(x_0) = \lim_{t \to 0} (x_t - x_0)/t.$$

We may hope that $(X_t)_{t \geqslant 0}$ is close to $(x_t)_{t \geqslant 0}$.

Fluid limits

Let $(\xi_t)_{t\geqslant 0}$ be a Markov chain in E with transition kernel q.

Choose a vector space V and a *coordinate map* $x : E \to V$.

Set $X_t = x(\xi_t)$ and consider the *drift* $\beta : E \to V$ given by

$$\beta(\xi) = \lim_{t\to 0} \mathbb{E}(X_t - X_0|\xi_0 = \xi)/t = \int_E \{x(\eta) - x(\xi)\}q(\xi, d\eta).$$

Choose b so that $b(x(\xi)) = \beta(\xi)$ and solve $\dot{x}_t = b(x_t)$. Thus

$$b(x_0) = \lim_{t\to 0}(x_t - x_0)/t.$$

We may hope that $(X_t)_{t\geqslant 0}$ is close to $(x_t)_{t\geqslant 0}$.

Why try this?

Fluid limits

Let $(\xi_t)_{t\geqslant 0}$ be a Markov chain in E with transition kernel q.

Choose a vector space V and a *coordinate map* $x : E \to V$.

Set $X_t = x(\xi_t)$ and consider the *drift* $\beta : E \to V$ given by

$$\beta(\xi) = \lim_{t\to 0} \mathbb{E}(X_t - X_0|\xi_0 = \xi)/t = \int_E \{x(\eta) - x(\xi)\}q(\xi, d\eta).$$

Choose b so that $b(x(\xi)) = \beta(\xi)$ and solve $\dot{x}_t = b(x_t)$. Thus

$$b(x_0) = \lim_{t\to 0}(x_t - x_0)/t.$$

We may hope that $(X_t)_{t\geqslant 0}$ is close to $(x_t)_{t\geqslant 0}$.

Why try this? Why might this work?

Why try this?

- Continuum limits
  - we justify the use of continuum equations to describe what begins as a particle model.
- Computation
  - in low dimension, numerical solution of the differential equation may be faster and more accurate,
  - in high dimension, numerical solution of the Markov chain may be faster and more accurate.
- Consistency
  - we identify effects which are independent of population size,
  - can then compute using N particles to predict outcomes for a population of size $N' \gg N$.

When will it work?

- Need jumps to be rapid and small
  - the *instantaneous variance* α should be small, where

$$\alpha(\xi) = \lim_{t \to 0} \mathbb{E}(|X_t - X_0|^2 | \xi_0 = \xi)/t = \int_E |x(\eta) - x(\xi)|^2 q(\xi, d\eta)$$

  - typically we are modelling averages over a large population, of size N say, so

$$\text{jump size} \sim O(1/N), \quad \text{rate} \sim N$$

giving

$$\beta(\xi) \sim O(1), \quad \alpha(\xi) \sim O(1/N).$$

When will it work?

- If the coordinate map $x : E \to V$ is not one-to-one, we may not be able to achieve exactly

$$b(x(\xi)) = \beta(\xi) = \int_E \{x(\eta) - x(\xi)\}q(\xi, d\eta).$$

- In the presence of fast variables in the Markov chain, b will oscillate wildly unless x is well chosen.
- The limit dynamics $(x_t)_{t\geqslant 0}$ may be infinite-dimensional
  - then care is needed to understand the linearized equation

$$\dot{y}_t = \nabla b(x_t)y_t.$$

- The limits $N \to \infty$ and $t \to \infty$ often do not commute
  - but I will explain an interpolation method which sometimes allows to exploit stability of the limit dynamics to obtain good long-time estimates.

Example: rock, paper, scissors

A population of N children mix rapidly in a playground.

At rate $1/N$, each of the $N(N-1)$ pairs of children meet and play rock, paper, scissors.

After each game, both children adopt the strategy of the winner.

The number of children playing each strategy thus evolves as a Markov chain $(R_t, P_t, S_t)_{t \geqslant 0}$ with non-zero transition rates such as

$$q(r, p, s; r+1, p, s-1) = \frac{rs}{N}.$$

The choice of coordinate map $x(\xi) = \xi/N$ then gives drift vector

$$\beta(\xi) = b(x(\xi)), \quad b(r, p, s) = (r(s-p), p(r-s), s(p-r)).$$

Exercise: show that $r_t p_t s_t =$ const. but $R_t P_t S_t \to 0$ a.s.

Example: join-the-shorter-queue with memory

Mitzenmacher, Prabakhar & Shah (2002), M. Luczak & N (2013)

Customers arrive at a system of N queues as a Poisson process of rate $N\lambda$, where $\lambda < 1$. The service requirement of each customer is exponentially distributed of mean 1. At all times, one of the queues, called the memory queue, is kept under observation.

On arrival, each customer selects a queue at random and compares its length with the memory queue. The customer joins the memory queue if that is shorter, and otherwise joins the selected queue. If the selected queue remains shorter than the memory queue after the customer has joined, then that becomes the memory queue.

This use of memory turns out to reduce markedly the numbers of longer queues, of length at least k say, from $N\lambda^k$ (without memory) to Na_k, where $a_k \leqslant e^{-e^{2k}}$ for k sufficiently large.

Example: join-the-shorter-queue with memory

To see this is true with high probability as $N \to \infty$, we can do a fluid limit for the Markov chain $\xi_t = (Z_t, Y_t)$, where

$$\begin{aligned} Z_t &= (Z_t^k : k \in \mathbb{N}), \\ Z_t^k &= \text{ proportion of queues of length } \geqslant k, \\ Y_t &= \text{ length of memory queue.} \end{aligned}$$

Try as coordinate map $x(z, y) = z$. Then (exercise)

$$\beta_k(z, y) = \lambda z_{k-1} 1_{\{y \geqslant k-1\}} - \lambda z_k 1_{\{y \geqslant k\}} - (z_k - z_{k+1}).$$

This looks unpromising because of the strong dependence of $\beta(z, y)$ on the uncontrolled fast variable y.

I will explain a general procedure to deal with this problem.

Example: Kac's model for a dilute gas

Write $\mathcal{S}$ for the set of probability measures μ on $\mathbb{R}^3$ such that

$$\langle v, \mu \rangle = \int_{\mathbb{R}^3} v \mu(dv) = 0, \quad \langle |v|^2, \mu \rangle = \int_{\mathbb{R}^3} |v|^2 \mu(dv) = 1.$$

This is the *Boltzmann sphere* of velocity distributions, where the reference frame is chosen to make the average momentum zero and units are chosen to normalize the total kinetic energy.

Example: Kac's model for a dilute gas

Write $\mathcal{S}$ for the set of probability measures μ on $\mathbb{R}^3$ such that

$$\langle v, \mu \rangle = \int_{\mathbb{R}^3} v\mu(dv) = 0, \quad \langle |v|^2, \mu \rangle = \int_{\mathbb{R}^3} |v|^2 \mu(dv) = 1.$$

This is the *Boltzmann sphere* of velocity distributions, where the reference frame is chosen to make the average momentum zero and units are chosen to normalize the total kinetic energy.

Write $\mathcal{S}_N$ for the subset of $\mathcal{S}$ consisting of N-particle normalized empirical measures

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{v_i}.$$

Example: Kac's model for a dilute gas

Write $\mathcal{S}$ for the set of probability measures μ on $\mathbb{R}^3$ such that

$$\langle v, \mu \rangle = \int_{\mathbb{R}^3} v\mu(dv) = 0, \quad \langle |v|^2, \mu \rangle = \int_{\mathbb{R}^3} |v|^2 \mu(dv) = 1.$$

Write $\mathcal{S}_N$ for the subset of $\mathcal{S}$ consisting of N-particle normalized empirical measures

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{v_i}.$$

Kac's model for particle velocities in a dilute gas is the Markov chain $(\mu_t^N)_{t \geqslant 0}$ in $\mathcal{S}_N$ with the following transition rule: for every pair of velocities v, v_*

- at rate $|v - v_*|/N$, draw a sphere with poles v, v_*
- choose randomly a new axis, with poles v', v'_* say
- replace v, v_* by v', v'_*.

Example: Kac's model for a dilute gas

For the fluid limit, we will take $x(\mu) = \mu$. Then $\beta = b$, where

$$\langle f, b(\mu)\rangle = \lim_{t\to 0} \mathbb{E}(\langle f, \mu_t^N - \mu_0^N\rangle | \mu_0^N = \mu)/t$$
$$= \int_{\mathbb{R}^3\times\mathbb{R}^3\times S^2} \{f(v') + f(v'_*) - f(v) - f(v_*)\}|v - v_*|\mu(dv)\mu(dv_*)d\sigma.$$

Here $d\sigma$ is the uniform distribution on S^2 and

$$v' + v'_* = v + v_*, \quad v' - v'_* = \sigma|v - v_*|.$$

For large N, it is natural to guess that μ_t^N is close to the solution of

$$\dot{\mu}_t = b(\mu_t)$$

with the same initial data. This is the *spatially homogeneous Boltzmann equation*.

Next lecture

- Martingales
- Martingale inequalities
- Gronwall
- Localization
- Long-time estimates for stable flows
- Averaging over fast variables