Fluid limits for Markov chains III

James Norris

Research supported by EPSRC grant EP/I03372X/1

Durham Symposium on Markov Processes, Mixing Times and Cutoff 2017

Convergence of empirical distributions in Kantorovich metric

Lemma (eg. Fournier & Guillin (2015))

For all $d \geqslant 3$, there is a constant $C(d) < \infty$ with the following property. Let μ be a probability measure supported in $B_0 = (-1, 1]^d$ and let μ_N be the empirical distribution of a sample of size N from μ. Then

$$\mathbb{E}(W_1(\mu_N, \mu)) \leqslant C(d) N^{-1/d}.$$

Here, W_1 is the Wasserstein–Kantorovich metric, given by

$$W_1(\mu, \nu) = \sup_{f \in \mathcal{F}} \langle f, \mu - \nu \rangle, \quad \mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)$$

where $\mathcal{F}$ is the set of all Lipschitz functions on B_0 of Lipschitz constant 1.

Proof

For all $\ell \geqslant 0$, there is a set $\mathcal{P}_\ell$ of $2^{\ell d}$ translates of $(-2^{-\ell}, 2^{-\ell}]^d$ which cover B_0. Fix $L \geqslant 0$. Given $f \in \mathcal{F}$, write

$$f = \sum_{\ell=0}^{L-1} \sum_{B \in \mathcal{P}_\ell} a_B 1_B + g$$

where $a_{B_0} = \langle f \rangle_{B_0}$ and, for $\ell \geqslant 1$ and $B \in \mathcal{P}_\ell$,

$$a_B = \langle f \rangle_B - \langle f \rangle_{\pi(B)}.$$

Here $\langle f \rangle_B$ is the average of f over B, and $\pi(B)$ is the unique element of $\mathcal{P}_{\ell-1}$ containing B. It suffices to consider the case $f(0) = 1$. Then, since $f \in \mathcal{F}$, for some constant $c_d < \infty$,

$$|a_B| \leqslant 2^{-\ell} c_d, \quad |g(v)| \leqslant 2^{-L} c_d.$$

Now, by Cauchy–Schwarz,

$$\langle f, \mu_N - \mu \rangle = \sum_{\ell=0}^{L-1} \sum_{B \in \mathcal{P}_\ell} a_B(\mu_N(B) - \mu(B)) - \langle g, \mu_N - \mu \rangle$$

$$\leqslant c_d \sum_{\ell=0}^{L-1} 2^{(d-2)\ell/2} \left(\sum_{B \in \mathcal{P}_\ell} (\mu_N(B) - \mu(B))^2 \right)^{1/2} + 2c_d 2^{-L}$$

The RHS does not depend on f, so is an upper bound for $W_1(\mu_N, \mu)$. Note that $\mathrm{var}(\mu_N(B)) \leqslant \mu(B)/N$. Take expectations and use Cauchy–Schwarz again to obtain

$$\mathbb{E}(W_1(\mu_N, \mu)) \leqslant c_d \sum_{\ell=0}^{L-1} 2^{(d-2)\ell/2} N^{-1/2} + 2c_d 2^{-L}.$$

Optimize at $L = \lceil d^{-1} \log_2 N \rceil$ for the claimed estimate. □

Recap: Kac's model for a dilute gas

We denote by $\mathcal{S}$ the *Boltzmann sphere* of velocity distributions, which is 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 a convenient state-space for dynamics which preserve the total momentum and kinetic energy, where we chose the reference frame to make the average momentum zero and chose the units to normalize the total kinetic energy.

Recap: Kac's model for a dilute gas

We denote by $\mathcal{S}$ the *Boltzmann sphere* of velocity distributions, which is 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 a convenient state-space for dynamics which preserve the total momentum and kinetic energy, where we chose the reference frame to make the average momentum zero and chose the units to normalize the total kinetic energy.

We denote by $\mathcal{S}_N$ 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 a dilute gas

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_*'.

- This is one of a class of random processes proposed by Kac in 1954 as models for the evolution by collisions of particle velocities in a spatially homogeneous dilute gas.

- This is one of a class of random processes proposed by Kac in 1954 as models for the evolution by collisions of particle velocities in a spatially homogeneous dilute gas.

- The collision rate is proportional to relative speed $|v - v_*|$ as might be expected physically.

- This is one of a class of random processes proposed by Kac in 1954 as models for the evolution by collisions of particle velocities in a spatially homogeneous dilute gas.

- The collision rate is proportional to relative speed $|v - v_*|$ as might be expected physically.

- Note that $v' + v'_* = v + v_*$ and write $v' - v'_* = \sigma|v - v_*|$.

- This is one of a class of random processes proposed by Kac in 1954 as models for the evolution by collisions of particle velocities in a spatially homogeneous dilute gas.

- The collision rate is proportional to relative speed $|v - v_*|$ as might be expected physically.

- Note that $v' + v'_* = v + v_*$ and write $v' - v'_* = \sigma|v - v_*|$.

- The transition $(v, v_*) \to (v', v'_*)$ models an elastic collision

$$|v|^2 + |v_*|^2 = \frac{|v + v_*|^2 + |v - v_*|^2}{2} = |v'|^2 + |v'_*|^2.$$

- This is one of a class of random processes proposed by Kac in 1954 as models for the evolution by collisions of particle velocities in a spatially homogeneous dilute gas.
- The collision rate is proportional to relative speed $|v - v_*|$ as might be expected physically.
- Note that $v' + v'_* = v + v_*$ and write $v' - v'_* = \sigma|v - v_*|$.
- The transition $(v, v_*) \to (v', v'_*)$ models an elastic collision

$$|v|^2 + |v_*|^2 = \frac{|v + v_*|^2 + |v - v_*|^2}{2} = |v'|^2 + |v'_*|^2.$$

- The simple rule to choose the direction of $v' - v'_*$ uniformly at random, corresponds (in 3 dimensions) to a model for collisions between *spherical* particles – by an elementary geometric calculation.

- Kac's purpose was to shed light on the Boltzmann equation which, at least formally, should govern the behaviour of his process in the limit $N \to \infty$.

- Kac's purpose was to shed light on the Boltzmann equation which, at least formally, should govern the behaviour of his process in the limit $N \to \infty$.
- McKean (1966) and Tanaka (1978, 1983) proved results on other versions of Kac's model.

- Kac's purpose was to shed light on the Boltzmann equation which, at least formally, should govern the behaviour of his process in the limit $N \to \infty$.

- McKean (1966) and Tanaka (1978, 1983) proved results on other versions of Kac's model.

- Sznitman (1984) proved weak convergence in probability for hard spheres to solutions of Boltzmann's equation – formulated as convergence in distribution and asymptotic independence of particles.

- Kac's purpose was to shed light on the Boltzmann equation which, at least formally, should govern the behaviour of his process in the limit $N \to \infty$.

- McKean (1966) and Tanaka (1978, 1983) proved results on other versions of Kac's model.

- Sznitman (1984) proved weak convergence in probability for hard spheres to solutions of Boltzmann's equation – formulated as convergence in distribution and asymptotic independence of particles.

- Mischler and Mouhot (2013) have established quantitative versions of Sznitman's result (and much more) with good long-time properties.

- Kac's purpose was to shed light on the Boltzmann equation which, at least formally, should govern the behaviour of his process in the limit $N \to \infty$.

- McKean (1966) and Tanaka (1978, 1983) proved results on other versions of Kac's model.

- Sznitman (1984) proved weak convergence in probability for hard spheres to solutions of Boltzmann's equation – formulated as convergence in distribution and asymptotic independence of particles.

- Mischler and Mouhot (2013) have established quantitative versions of Sznitman's result (and much more) with good long-time properties.

- I will describe a new approach to the question of convergence, following the approach of Markov chain fluid limits, which leads to an explicit pathwise estimate in Wasserstein distance.

We seek a fluid limit with coordinate map $x(\mu) = \mu$. 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 the spatially homogeneous Boltzmann equation

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

with the same initial data.

Boltzmann's equation

Recall, for $\mu \in \mathcal{S}$, we define a signed measure $b(\mu)$ on $\mathbb{R}^3$ by

$$\langle f, b(\mu) \rangle$$
$$= \int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \{f(v') + f(v'_*) - f(v) - f(v_*)\} |v - v_*| \mu(dv) \mu(dv_*) d\sigma.$$

A process $(\mu_t)_{t \geqslant 0}$ in $\mathcal{S}$ is a (measure) solution to the spatially homogeneous Boltzmann equation if, for all bounded measurable functions f of compact support in $\mathbb{R}^3$ and all $t \geqslant 0$,

$$\langle f, \mu_t \rangle = \langle f, \mu_0 \rangle + \int_0^t \langle f, b(\mu_s) \rangle ds.$$

Boltzmann's equation

Recall, for $\mu \in \mathcal{S}$, we define a signed measure $b(\mu)$ on $\mathbb{R}^3$ by

$$\langle f, b(\mu) \rangle$$
$$= \int_{\mathbb{R}^3 \times \mathbb{R}^3 \times S^2} \{ f(v') + f(v'_*) - f(v) - f(v_*) \} |v - v_*| \mu(dv) \mu(dv_*) d\sigma.$$

A process $(\mu_t)_{t \geqslant 0}$ in $\mathcal{S}$ is a (measure) solution to the spatially homogeneous Boltzmann equation if, for all bounded measurable functions f of compact support in $\mathbb{R}^3$ and all $t \geqslant 0$,

$$\langle f, \mu_t \rangle = \langle f, \mu_0 \rangle + \int_0^t \langle f, b(\mu_s) \rangle ds.$$

Lu and Mouhot (2012) have shown that, for all $\mu_0 \in \mathcal{S}$ there is a unique solution $(\mu_t)_{t \geqslant 0}$ starting from μ_0.

Weighted Wasserstein distance

For functions f on $\mathbb{R}^3$ we will write $\|f\|$ for the smallest constant such that, for all v, v',

$$|\hat{f}(v)| \leqslant \|f\|, \quad |\hat{f}(v) - \hat{f}(v')| \leqslant \|f\| |v - v'|.$$

where $\hat{f}(v) = f(v)/(1 + |v|^2)$.

Weighted Wasserstein distance

For functions f on $\mathbb{R}^3$ we will write $\|f\|$ for the smallest constant such that, for all v, v',

$$|\hat{f}(v)| \leqslant \|f\|, \quad |\hat{f}(v) - \hat{f}(v')| \leqslant \|f\||v - v'|.$$

where $\hat{f}(v) = f(v)/(1 + |v|^2)$.

We will use on $\mathcal{S}$ the distance function

$$W(\mu, \nu) = \sup_{\|f\|=1} \langle f, \mu - \nu \rangle.$$

This is a type of weighted Wasserstein-1 distance well adapted to the Boltzmann sphere.

Theorem (N. (2016))

For all

$$\varepsilon > 0, \quad \lambda < \infty, \quad p > 8, \quad T < \infty$$

there is a constant $C < \infty$ *with the following property.*

Let $(\mu_t^N)_{t \geqslant 0}$ *be a Kac process in* $\mathcal{S}_N$ *and let* $(\mu_t)_{t \geqslant 0}$ *be a solution to the spatially homogeneous Boltzmann equation.*

Assume that

$$\langle |v|^p, \mu_0 \rangle \leqslant \lambda, \quad \langle |v|^p, \mu_0^N \rangle \leqslant \lambda.$$

Then, with probability exceeding $1 - \varepsilon$, *for all* $t \in [0, T]$, *we have*

$$W(\mu_t^N, \mu_t) \leqslant C(W(\mu_0^N, \mu_0) + N^{-1/3}).$$

Theorem (N. (2016))

For all

$$\varepsilon > 0, \quad \lambda < \infty, \quad p > 8, \quad T < \infty$$

there is a constant $C < \infty$ *with the following property.*

Let $(\mu_t^N)_{t \geqslant 0}$ *be a Kac process in* $\mathcal{S}_N$ *and let* $(\mu_t)_{t \geqslant 0}$ *be a solution to the spatially homogeneous Boltzmann equation.*

Assume that

$$\langle |v|^p, \mu_0 \rangle \leqslant \lambda, \quad \langle |v|^p, \mu_0^N \rangle \leqslant \lambda.$$

Then, with probability exceeding $1 - \varepsilon$*, for all* $t \in [0, T]$*, we have*

$$W(\mu_t^N, \mu_t) \leqslant C(W(\mu_0^N, \mu_0) + N^{-1/3}).$$

A similar estimate holds for all $p > 2$ but we have to replace the optimal $1/3$ by some $\alpha(p) > 0$. For $\tau > 0$, a similar estimate holds without moment restriction, and with power $1/3$, for $t \in [\tau, T]$ if we replace $W(\mu_0^N, \mu_0)$ by $W(\mu_\tau^N, \mu_\tau)$.

Ideas from the proof

Recall that, for bounded functions f of compact support,

$$\langle f, \mu_t \rangle = \langle f, \mu_0 \rangle + \int_0^t \langle f, b(\mu_s) \rangle ds$$

while

$$\langle f, \mu_t^N \rangle = \langle f, \mu_0^N \rangle + M_t^f + \int_0^t \langle f, b(\mu_s^N) \rangle ds.$$

Ideas from the proof

Recall that, for bounded functions f of compact support,

$$\langle f, \mu_t \rangle = \langle f, \mu_0 \rangle + \int_0^t \langle f, b(\mu_s) \rangle ds$$

while

$$\langle f, \mu_t^N \rangle = \langle f, \mu_0^N \rangle + M_t^f + \int_0^t \langle f, b(\mu_s^N) \rangle ds.$$

Subtract to obtain a linear equation for $\mu_t^N - \mu_t$

$$\langle f, \mu_t^N - \mu_t \rangle = \langle f, \mu_0^N - \mu_0 \rangle + M_t^f + 2 \int_0^t \langle f, b(\rho_s, \mu_s^N - \mu_s) \rangle ds.$$

where $\rho_t = (\mu_t^N + \mu_t)/2$ and we have written b also for the bilinear form associated to the quadratic form b.

We can write

$$M_t^f = \int_{(0,t]\times\mathbb{R}^3} f(v) M(ds, dv)$$

for a certain Poisson-type martingale measure M.

Then, we will show, for $s \in [0, t]$, there is a way to propagate $f_t = f$ linearly back from t to s to obtain f_s so that

$$\langle f_t, \mu_t^N - \mu_t \rangle = \langle f_0, \mu_0^N - \mu_0 \rangle + \int_{(0,t]\times\mathbb{R}^3} f_s(v) M(ds, dv).$$

It remains to estimate the right-hand side uniformly over $\|f\| = 1$.

Linearized Kac process – propagation of errors

We set up an auxiliary branching process of positive and negative particles in $\mathbb{R}^3$ which provides a stochastic realization of the linearized Boltzmann equation, linearized around $(\rho_t)_{t \geqslant 0}$.

Linearized Kac process – propagation of errors

We set up an auxiliary branching process of positive and negative particles in $\mathbb{R}^3$ which provides a stochastic realization of the linearized Boltzmann equation, linearized around $(\rho_t)_{t \geqslant 0}$.

This can be thought of as the difference of two Kac processes, maximally coupled, which differ by a single particle of a given velocity v, introduced at a given time s.

Linearized Kac process – propagation of errors

We set up an auxiliary branching process of positive and negative particles in $\mathbb{R}^3$ which provides a stochastic realization of the linearized Boltzmann equation, linearized around $(\rho_t)_{t\geqslant 0}$.

This can be thought of as the difference of two Kac processes, maximally coupled, which differ by a single particle of a given velocity v, introduced at a given time s.

The branching rule is that each positive particle v, at rate

$$2|v - v_*|\rho_t(dv_*)d\sigma dt,$$

dies and is replaced by two positive particles $v' = v'(v, v_*, \sigma)$ and $v'_* = v'_*(v, v_*, \sigma)$ and one negative particle v_*, and a similar rule holds for negative particles.

The branching rule is that each positive particle v, at rate

$$2|v - v_*|\rho_t(dv_*)d\sigma dt,$$

dies and is replaced by two positive particles $v' = v'(v, v_*, \sigma)$ and $v'_* = v'_*(v, v_*, \sigma)$ and one negative particle v_*, and a similar rule holds for negative particles.

Write $\Lambda_t^\pm$ for the un-normalized empirical measures of $\pm$ particles at time t. Fix $t \geqslant 0$ and a function f_t on $\mathbb{R}^3$. Define for $s \in [0, t]$

$$f_s(v) = E_{(s,v)}\langle f_t, \Lambda_t^+ - \Lambda_t^- \rangle.$$

The branching rule is that each positive particle v, at rate

$$2|v - v_*|\rho_t(dv_*)d\sigma dt,$$

dies and is replaced by two positive particles $v' = v'(v, v_*, \sigma)$ and $v'_* = v'_*(v, v_*, \sigma)$ and one negative particle v_*, and a similar rule holds for negative particles.

Write $\Lambda_t^{\pm}$ for the un-normalized empirical measures of $\pm$ particles at time t. Fix $t \geqslant 0$ and a function f_t on $\mathbb{R}^3$. Define for $s \in [0, t]$

$$f_s(v) = E_{(s,v)}\langle f_t, \Lambda_t^+ - \Lambda_t^- \rangle.$$

Then we can show

$$\langle f_t, \mu_t^N - \mu_t \rangle = \langle f_0, \mu_0^N - \mu_0 \rangle + \int_{(0,t]\times\mathbb{R}^3} f_s(v) M(ds, dv).$$

Lemma

For all functions f_t on $\mathbb{R}^3$, the function

$$f_s(v) = E_{(s,v)}\langle f_t, \Lambda_t^+ - \Lambda_t^- \rangle$$

satisfies, for all $s, s' \leqslant t$ and all $v \in \mathbb{R}^3$,

$$\|f_s\| \leqslant C(T)\|f_t\|, \quad |f_s(v) - f_{s'}(v)| \leqslant C(T)(1 + |v|^3)|s - s'|\|f_t\|.$$

Here

$$C(T) = 6(T+1)e^{4Tm_3(T)}, \quad m_3(T) = \sup_{t \leqslant T}\langle 1 + |v|^3, \mu_t + \mu_t^N \rangle.$$

Lemma

For all functions f_t on $\mathbb{R}^3$, the function

$$f_s(v) = E_{(s,v)}\langle f_t, \Lambda_t^+ - \Lambda_t^- \rangle$$

satisfies, for all $s, s' \leqslant t$ and all $v \in \mathbb{R}^3$,

$$\|f_s\| \leqslant C(T)\|f_t\|, \quad |f_s(v) - f_{s'}(v)| \leqslant C(T)(1+|v|^3)|s-s'|\|f_t\|.$$

Here

$$C(T) = 6(T+1)e^{4Tm_3(T)}, \quad m_3(T) = \sup_{t \leqslant T}\langle 1+|v|^3, \mu_t + \mu_t^N \rangle.$$

This allows us to estimate the first term when $\|f_t\| \leqslant 1$

$$\langle f_0, \mu_0^N - \mu_0 \rangle \leqslant C(T)W(\mu_0^N, \mu_0), \quad t \leqslant T.$$

Then we show for the second term an estimate valid with high probability of the form

$$\int_{(0,t]\times\mathbb{R}^3} f_s(v)M(ds,dv) \leqslant CN^{-1/3}, \quad t \leqslant T.$$

Then we show for the second term an estimate valid with high probability of the form

$$\int_{(0,t]\times\mathbb{R}^3} f_s(v)M(ds,dv) \leqslant CN^{-1/3}, \quad t \leqslant T.$$

The integrand is given by

$$f_s(v) = E_{(s,v)}\langle f_t, \Lambda_t^+ - \Lambda_t^-\rangle$$

so depends implicitly on t and is not adapted in the filtration of M.

The estimate is uniform in $t \leqslant T$ and in $\|f_t\| \leqslant 1$.

Obtained by multiple use of the hierarchical type of estimation used for sample means – compare

$$\langle f, \mu_N - \mu\rangle \quad \text{and} \quad \int_{(0,t]\times\mathbb{R}^3} f_s(v)M(ds,dv).$$

Thus we can go from

$$\langle f_t, \mu_t^N - \mu_t \rangle = \langle f_0, \mu_0^N - \mu_0 \rangle + \int_{(0,t]\times\mathbb{R}^3} f_s(v) M(ds, dv)$$

to

$$W(\mu_t^N, \mu_t) \leqslant C(W(\mu_0^N, \mu_0) + N^{-1/3})$$

with high probability.

Future directions

- Laws of large numbers for function-valued and measure-valued Markov processes
  - Generic discrete-to-continuous problem away from criticality, away from equilibrium
  - First level and prerequisite for analysis of fluctuations
  - Limit dynamics $\dot{x}_t = b(x_t)$ as PDE or more general evolution equation