An introduction to mean field theory

DAVYDOV Michel

November 25th 2020

Object of study

Object of interest= evolution in time of a system of interacting particles.
Particles= very wide meaning (neurons, queues, players,...)

Microscopic approach

Describing the behavior of the system through a system of SDEs

Example

Let $N \geq 1$. We consider the stochastic processes $X^{k,N}$ on $\mathbb{R}$ for $1 \leq k \leq N$, verifying the SDEs

$$X^{k,N}(t) = X^{k,N}(0) + \omega^k(t) + \int_0^t \frac{1}{N} \sum_{j=1}^{N} b(X^{k,N}(s), X^{j,N}(s)) \, \mathrm{d}s \quad (1)$$

where ω^k are independent BM, and b is a globally Lipschitz function.

Example

Let $N \geq 1$. We consider the stochastic processes $X^{k,N}$ on $\mathbb{R}$ for $1 \leq k \leq N$, verifying the SDEs

$$X^{k,N}(t) = X^{k,N}(0) + \omega^k(t) + \int_0^t \frac{1}{N} \sum_{j=1}^{N} b(X^{k,N}(s), X^{j,N}(s))\, \mathrm{d}s$$

where ω^k are independent BM, and b is a globally Lipschitz function.

A few remarks

- Without the noise, just a set of ordinary ODEs.
- We have existence and trajectorial uniqueness of the solutions (stochastic Cauchy-Lipschitz theorem).
- Good modelization, but hard to compute.

Macroscopic approach

Describing the behavior of a simplified version of the initial system through a system of SDEs that are no longer interlocked and that describe the statistical distribution of the particles.

Example

We consider the stochastic processes $\overline{X}^k$ on $\mathbb{R}$ for $1 \leq k \leq N$, verifying the SDEs

$$\overline{X}^k(t) = \overline{X}^k(0) + \omega^k(t) + \int_0^t \int_{\mathbb{R}} b(\overline{X}^k(s), y)\mu_s^k(\mathrm{d}y)\,\mathrm{d}s, \qquad (2)$$

where ω^k are independent BM, b globally Lipschitz function and μ_s^k is the law of $\overline{X}^k(s)$.

A few remarks

- Infinite number of particles$->$ statistical approach (sometimes called thermodynamic limit)
- Heuristic: equation (2) is intuitively what we believe the limit of (1) verifies when N goes to infinity (law of large numbers)
- We obtain nonlinear SDEs, but they don't depend on each other anymore$->$ asymptotic independence
- Information lost compared to the initial model (correlations between particles for example, or finite size effects).

Putting the "mean-field" in "mean-field theory"

Note that we can rewrite (1) in terms of its empirical mean measure:

$$X^{k,N}(t) = X^{k,N}(0) + \omega^k(t) + \int_0^t \int_{\mathbb{R}} b(X^{k,N}(s), z) e_s^N(\mathrm{d}z)\,\mathrm{d}s, \quad (3)$$

where for all $r \geq 0$, $e_r^N = \frac{1}{N} \sum_{j=1}^N \delta_{X_r^{j,N}}$.

Remarks

- We are interested in the convergence in law of this empirical measure when N goes to infinity. If it converges, there is asymptotic independence and this is called propagation of chaos.
- We have exchangeability, that is, invariance by permutation of the law of $(X^{1,N}, ..., X^{N,N})$.

The results

- Well-posedness of the SDEs (2).
- Convergence in probability of the empirical mean measure on the space of probability measures on the space of trajectories.

Aim of the talk

- Discuss the general techniques used to prove the first type of results.
- Present the convergence of (1) to (2) on a simple toy model using coupling techniques.

Well-posedness of limit SDEs

Two main techniques: probabilistic and analytical (we won't discuss the second one here)

Probabilistic approach

- Define the "right" distance on the space of probability measures on the space of trajectories
- Introduce a function Φ on this space that associates to a measure the law of the solution to a linearized version of the SDEs (more detail on this later)
- Show that this function Φ has a unique fixed point.

Wasserstein distance for continuous trajectories

Let μ and ν be two probability measures on $\mathcal{C} = C([0,T],\mathbb{R}^N)$. We define the Wasserstein distance between them by

$$D_T(\mu,\nu) = \inf_{\Pi}\{\int_{\mathcal{C}\times\mathcal{C}} \sup_{s\in[0,T]} |\omega^1(s) - \omega^2(s)| \wedge 1 \, \mathrm{d}\Pi(\omega^1,\omega^2)\}, \qquad (4)$$

where Π probability measure on $\mathcal{C}\times\mathcal{C}$ such that its first marginal is μ and its second is ν (Π is called a coupling of μ and ν).

Properties

- D_T is a distance on $\mathcal{P}(\mathcal{C})$.
- $(\mathcal{P}(\mathcal{C}), D_T)$ is a complete metric space
- D_T is nondecreasing in T.

The linearized SDE

Let $\Phi : \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C})$ such that for all m, $\Phi(m)$ is the law of the solution of the SDE on $\mathbb{R}^N$:

$$Z(t) = \overline{X}(0) + B(t) + \int_0^t \int_{\mathbb{R}^N} b(Z(s), y) m(\mathrm{d}y)\,\mathrm{d}s,$$

The original SDE for reference

$$\overline{X}(t) = \overline{X}(0) + B(t) + \int_0^t \int_{\mathbb{R}^N} b(\overline{X}(s), y) \mu_s(\mathrm{d}y)\,\mathrm{d}s,$$

with μ_s is the law of $\overline{X}(s)$.

Main result

Φ has a unique fixed point.

Idea of proof

- Show that $D_t(\Phi(\mu), \Phi(\nu)) \leq C_t \int_0^t D_s(\mu, \nu)\, \mathrm{d}s$. (uses Gronwall's lemma)
- Show a similar bound for iterations of Φ.
- Conclude that $(\Phi^n(\mu))_n$ is a Cauchy sequence, from which it follows that its limit is a fixed point of Φ.

Consequence

- Existence and uniqueness (in law and trajectorial) of the solution to (2)

A toy model for proving convergence

To simplify calculations, let us consider the initial model with the function $b(x, y) = x - y$. We then have the following SDEs:

$$X^{k,N}(t) = X^{k,N}(0) + \omega^k(t) + \int_0^t (X^{k,N}(s) - \frac{1}{N}\sum_{j=1}^{N} X^{j,N}(s))\,\mathrm{d}s \quad (5)$$

where ω^k are independent BM.

Remarks

- The SDE is now linear
- Disregarding the noise, there is exponential convergence to the mean position of the particles.

The new intuitive limit process

The limit process now verifies the following SDE:

$$\overline{X}^k(t) = \overline{X}^k(0) + \omega^k(t) + \int_0^t (\overline{X}^k(s) - \mathrm{E}[\overline{X}^k(s)])\,\mathrm{d}s, \qquad (6)$$

where ω^k are independent BM.

Remarks

- Still nonlinearity at the limit (of McKean-Vlasov type)
- There is invariance in law for all k.
- Given μ_0 the law of $\overline{X}^k(0)$, if μ_0 has a finite first moment, (6) becomes equivalent to a linear SDE with an explicit solution (Ornstein-Uhlenbeck process)

The coupling

We consider the probability space $\Omega = (\mathbb{R} \times \mathcal{C}(\mathbb{R}^+, \mathbb{R}))^{\mathbb{N}}$ endowed with $(\mu_0 \otimes W)^{\mathbb{N}}$ with the real coordinates being the i.i.d. initial conditions (denoted Y^k hereafter) and the trajectory coordinates being the i.i.d. Brownian motions (denoted B_t^k). We construct on Ω the following processes:

- the processes $(X_t^{k,N})$ verifiying

$$X^{k,N}(t) = Y^k(0) + B^k(t) + \int_0^t (X^{k,N}(s) - \frac{1}{N}\sum_{j=1}^{N} X^{j,N}(s))\,\mathrm{d}s \quad (7)$$

- the processes $(\overline{X}_t^k)$ verifiying

$$\overline{X}^k(t) = Y^k(0) + B^k(t) + \int_0^t (\overline{X}^k(s) - \mathrm{E}[\overline{X}^k(s)])\,\mathrm{d}s \quad (8)$$

The convergence theorem

Suppose that μ_0 has a finite second moment ν_0.
For all $1 \leq k \leq N$, for all finite $T > 0$,

$$\sqrt{N}\,\mathrm{E}[\sup_{t\in[0,T]} |X^{k,N}(t) - \overline{X}^k(t)|] \leq (\nu_0 + \frac{1}{2})Te^{2T}. \tag{9}$$

Intuition for the proof

Exponential bound: use Gronwall's lemma. Here is a simple version of it (that can be generalized to a much less stringent setting):
Let u and b be a nonnegative continuous functions on $\mathbb{R}^+$, let a be a nonnegative constant (or nonnegative continuous function). If u verifies the following integral inequality for all $t \in \mathbb{R}^+$:

$$u(t) \leq a + \int_0^t b(s)u(s)\,\mathrm{d}s,$$

then

$$u(t) \leq a e^{\int_0^t b(s)\,\mathrm{d}s}.$$

The proof

By the coupling, we have

$$|X^{k,N}(t) - \overline{X}^k(t)| \leq \int_0^t |X^{k,N}(s) - \overline{X}^k(s)| + |\frac{1}{n}\sum_{j=1}^n (X^{j,N}(s) - \overline{X}^j(s))| + |\frac{1}{n}\sum_{j=1}^n \overline{X}^j(s) - \mathrm{E}[\overline{X}^1(s))]| \,\mathrm{d}s$$

Let $\delta(t) = \mathrm{E}[\sup_{t\in[0,T]} |X^{k,N}(t) - \overline{X}^k(t)|]$. Then taking the expectation in the equation above, we have:

$$\delta(t) \leq 2\int_0^t \delta(s)\,\mathrm{d}s + \int_0^t \mathrm{E}[|\frac{1}{n}\sum_{j=1}^n \overline{X}^j(s) - \mathrm{E}[\overline{X}^1(s))]|]\,\mathrm{d}s$$

The proof, continued

Let $D_N(t) = \mathrm{E}[|\frac{1}{n}\sum_{j=1}^{n}\overline{X}^j(t) - \mathrm{E}[\overline{X}^1(t))]|$. It is the first moment mean of an empirical mean of centered independent r.v.s. By the Cauchy-Schwarz inequality, we have:

$$D_N(t) \leq \frac{1}{\sqrt{n}}\mathrm{var}(\overline{X}^1(t)).$$

Recall that $\overline{X}^1(t)$ is the solution of an SDE with an explicit solution (Ornstein-Uhlenbeck process). Taking the variance in that explicit solution, we show that

$$\mathrm{var}(\overline{X}^1(t)) \leq \nu_0 + \frac{1}{2}.$$

The proof, end

Finally, we have

$$\delta(t) \leq 2\int_0^t \delta(s)\,\mathrm{d}s + \frac{1}{\sqrt{n}}(\nu_0 + \frac{1}{2})t.$$

Applying Gronwall's lemma, the result follows.

Consequence for the convergence

Note that (9) gives a control of the Wasserstein distance between the laws of $(X_t^{k,N})$ and $(\overline{X}_t^k)$. Therefore, for all $1 \leq k \leq N$, for all finite $t \geq 0$, $\mathcal{L}(X_t^{k,N}) \to \mathcal{L}(\overline{X}_t^k)$ when $N \to \infty$.

Remark: stronger result

It is actually possible to prove the convergence of the empirical mean:

$$\frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{j,N}} \to \mathcal{L}(\overline{X}_t^k)$$

when $N \to \infty$.

Propagation of chaos

The previous result is usually called *propagation of chaos* because it can be shown that in a certain sense, the asymptotic independence of the N-particle system is equivalent to the weak convergence of the empirical mean.

Spatial generalizations

- It is possible to introduce a spatial component to the interactions by taking them of the form

$$\frac{1}{N}\sum_{j=1}^{N} K(y_i, y_j) b(X^{k,N}(s), X^{j,N}(s))$$

  with y_i the spatial localization of the i-th particle and K the kernel of spatial dependence. In this case, you have to consider both the convergence of the processes $X^{k,N}(s)$ to some limit process and the convergence of $\frac{1}{n}\sum_{i=1}^{N}\delta_{y_i}$ to a certain limit distribution.
- Another way to preserve the geometry of the N-particle system is to consider *replica mean-field* limits, looking at the limit of M replicas of the initial particle system with interactions uniformly randomly routed in between the replicas when M goes to infinity.

Julien Chevallier. "Approximation par champ-moyen : le couplage à la Sznitman pour les nuls". Jan. 2017. URL: https://hal.archives-ouvertes.fr/hal-01433292.

Alain-Sol Sznitman. "Topics in propagation of chaos". In: *Ecole d'Ete de Probabilites de Saint-Flour XIX, vol.1464* (1989), pp. 165–251.