Lecture-08: Convergence to mean-field model

1 Classical approach

Definition 1.1. Let $\mathcal{X} \triangleq \mathcal{Z}^N$ for countable $\mathcal{Z}$. Consider an N interacting particle CTMC $X^N : \Omega \to \mathcal{X}^{\mathbb{R}_+}$ with associated empirical distribution $A^N : \Omega \to \mathcal{M}(\mathcal{Z})^{\mathbb{R}_+}$ such that $(A^N : N \in \mathbb{N})$ is a density-dependent family of CTMCs. We let denote $A^N(\infty)$ the stationary empirical distribution of N particle CTMC, and $a : \mathbb{R}_+ \to \mathcal{M}(\mathcal{Z})$ denote the solution of the associated mean-field model and $a^* \in \mathcal{M}(\mathcal{Z})$ denote its equilibrium point.

Classical approach to show convergence in distribution of $\lim_{N\to\infty} A^N(\infty) \to a^*$ involves the following three steps.

1. Show the convergence of CTMCs to the trajectory of the mean-field model for any finite time interval $[0,t]$. That is, we show
$$\lim_{N\to\infty} \sup \left\{ d(A^N(s), a(s)) : s \in [0,t] \right\} = 0,$$
where $d : \mathcal{M}(\mathcal{Z}) \times \mathcal{M}(\mathcal{Z}) \to \mathbb{R}_+$ is some measure of distance. This can be proved using different techniques including Kurtz's theorem, propagation of chaos, or the convergence of the transition semigroup of CTMCs.
2. Show the asymptotic stability of the mean-field model. That is, we show $\lim_{t\to\infty} a(t) = a^*$. Lyapunov theorem or LaSalle invariance principle can often be used for proving the stability. This implies that
$$\lim_{t\to\infty} \lim_{N\to\infty} A^N(t) = \lim_{t\to\infty} a(t) = a^*$$
3. Show the exchange of limits. That is, $\lim_{N\to\infty} \lim_{t\to\infty} A^N(t) = \lim_{t\to\infty} \lim_{N\to\infty} A^N(t)$. This shows the convergence of the stationary distribution, i.e. $\lim_{N\to\infty} \lim_{t\to\infty} A^N(t) = a^*$.

2 Ordinary differential equations

Definition 2.1. For any d-dimensional Euclidean space $\mathbb{R}^d$, we define d-norm for any $x \in \mathbb{R}^d$ as
$$\|x\|_d \triangleq \Big(\sum_{i=1}^d |x_i|^d \Big)^{\frac{1}{d}}.$$

Definition 2.2. The space of continuous functions from interval $[0,T]$ to $\mathbb{R}^d$ is denoted by $C([0,T],\mathbb{R}^d)$. We can define a sup norm on this space for each $f \in C([0,T],\mathbb{R}^d)$ and a norm $\|\|$ on $\mathbb{R}^d$ as
$$\|f\| \triangleq \sup_{t\in[0,T]} \|f(t)\|. \tag{1}$$
We can define a metric on $C([0,T],\mathbb{R}^d)$ as $d_T(f,g) \triangleq \|f-g\|$.

Theorem 2.3. *The normed vector space $C([0,T],\mathbb{R}^d)$ with the norm defined in (1) is complete, and hence a Banach space.*

Definition 2.4. The space of continuous functions from interval $\mathbb{R}_+ \to \mathbb{R}^d$ is denoted by $C(\mathbb{R}_+,\mathbb{R}^d)$. For any functions $f,g \in C(\mathbb{R}_+,\mathbb{R}^d)$, we can define corresponding projections on $C([0,T],\mathbb{R}^d)$ as $f\mathbb{1}_{[0,T]}, g\mathbb{1}_{[0,T]}$, and denote a metric on $C(\mathbb{R}_+,\mathbb{R}^d)$ as $d_T(f,g) \triangleq \left\|(f-g)\mathbb{1}_{[0,T]}\right\|$. We define another metric as
$$d(f,g) \triangleq \sum_{T\in\mathbb{N}} 2^{-T}(d_T(f,g) \wedge 1). \tag{2}$$

Theorem 2.5. *The metric defined in (2) metrizes the topology on vector space $C(\mathbb{R}_+,\mathbb{R}^d)$ that renders projections to finite time as continuous functions.*

3 Convergence in probability

For a density-dependent family of empirical distribution CTMCs $(A^N : \Omega \to \mathcal{M}(\mathcal{Z})^{\mathbb{R}_+} : N \in \mathbb{N})$, the sample path of distribution process is $A^N(t)$. The solution $a : \mathbb{R}_+ \to \mathcal{M}(\mathcal{Z})$ to the corresponding McKean-Vlasov equation is deterministic. For a fix T, one can view sample path $(A^N(t) : t \in [0,T])$ and $(a(t) : t \in [0,T])$ as elements of space $C([0,T], \mathcal{M}(\mathcal{Z}))$.

Remark 1. The sequence of random paths $((A^N(t) : t \in [0,T]) : N \in \mathbb{N})$ converge in probability to $(a(t) : t \in [0,T])$, if for every $\epsilon > 0$, we have

$$\lim_{N\to\infty} P\left\{d_T(A^N, a) > \epsilon\right\} = \lim_{N\to\infty} P\left\{\sup_{t\in[0,T]} \left\|A^N(t) - a(t)\right\| > \epsilon\right\} = 0.$$

4 Total variation distance

Consider state space $\mathcal{X} \triangleq \mathcal{Z}^N$ for N particle system, where each particle $n \in [N]$ has a countable state space $\mathcal{Z}$. The closeness of two distributions on $\mathcal{Z}$ can be measured by the following distance metric.

Definition 4.1. The **total variation distance** between two probability distributions $\mu, \nu \in \mathcal{M}(\mathcal{X})$ is defined by

$$d_{\text{TV}}(\mu,\nu) \triangleq \|\mu - \nu\|_{\text{TV}} \triangleq \max\left\{|\mu(A) - \nu(A)| : A \subseteq \mathcal{X}\right\}.$$

Remark 2. This definition is probabilistic in the sense that the distance between μ and ν is the maximum difference between the probabilities assigned to a single event by the two distributions.

Proposition 4.2. *Let $\mu, \nu \in \mathcal{M}(\mathcal{X})$, and $B \triangleq \{x \in \mathcal{X} : \mu(x) \geqslant \nu(x)\}$, then*

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x) - \nu(x)| = \sum_{x\in B}[\mu(x) - \nu(x)].$$

Proof. Consider an event $A \subseteq \mathcal{X}$. Since $\mu(x) - \nu(x) < 0$ for any $x \in A \cap B^c$, we have

$$\mu(A) - \nu(A) = \sum_{x\in A}(\mu-\nu)(x) \leqslant \sum_{x\in A\cap B}(\mu-\nu)(x) \leqslant \sum_{x\in B}(\mu-\nu)(x) = \mu(B) - \nu(B).$$

Similarly, we observe that $(\nu - \mu)(x) > 0$ for all $x \in A \setminus B$, and hence

$$\nu(A) - \mu(A) = \sum_{x\in A}(\nu-\mu)(x) \leqslant \sum_{x\in A\cap B^c}(\nu-\mu)(x) \leqslant \sum_{x\in B^c}(\nu-\mu)(x) = \nu(B^c) - \mu(B^c) = \mu(B) - \nu(B).$$

Since $A \subseteq \mathcal{X}$ was arbitrary event, it follows that $\sup_{A\subseteq\mathcal{X}} |\mu(A) - \nu(A)| \leqslant \mu(B) - \nu(B)$, and the equality is achieved for $A = B$ and $A = B^c$. Thus, we get that

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2}[\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x) - \nu(x)|.$$

□

Proposition 4.3. *For probability distributions $\mu, \nu \in \mathcal{M}(\mathcal{X})$, total variation distance $\|\mu - \nu\|_{\text{TV}}$ is a metric.*

Proof. We verify that the total variation distance satisfies the following three necessary properties for it to be distance metric. Let $\mu, \nu, \pi \in \mathcal{M}(\mathcal{X})$.

Positivity, i.e. $\|\mu - \nu\|_{\text{TV}} \geqslant 0$ and equals zero iff $\mu = \nu$. Note from Proposition 4.2 that $\|\mu - \nu\|_{\text{TV}} = \sum_{x\in B}[\mu(x) - \nu(x)]$, where $B = \{x \in \mathcal{X} : \mu(x) \geqslant \nu(x)\}$, hence $\|\mu - \nu\|_{\text{TV}} \geqslant 0$. The set B can not be empty as μ and ν are probability distributions and one can not dominate the other for all $x \in \mathcal{X}$. Similarly $\mu(x) = \nu(x)$ for all $x \in B$ iff $B = \mathcal{X}$. Thus, $\|\mu - \nu\|_{\text{TV}} = 0$ iff $\mu = \nu$.

Symmetry, i.e. $\|\mu - \nu\|_{\text{TV}} = \|\nu - \mu\|_{\text{TV}}$. To verify symmetry, we note that

$$\|\mu - \nu\|_{\text{TV}} = \mu(B) - \nu(B) = \nu(B^c) - \mu(B^c) = \|\nu - \mu\|_{\text{TV}}.$$

Triangle inequality, i.e. $\|\mu-\nu\|_{\text{TV}} \leqslant \|\mu-\pi\|_{\text{TV}} + \|\pi-\nu\|_{\text{TV}}$**.** This follows from the triangle inequality for absolute value function and noting that

$$\|\mu-\nu\|_{\text{TV}} = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x)-\nu(x)| \leqslant \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x)-\pi(x)| + \frac{1}{2}\sum_{x\in\mathcal{X}} |\pi(x)-\nu(x)| = \|\mu-\pi\|_{\text{TV}} + \|\pi-\nu\|_{\text{TV}}$$

□

Proposition 4.4. *Let* $\mu,\nu \in \mathcal{M}(\mathcal{X})$ *and* $F \triangleq \left\{f \in \mathbb{R}^{\mathcal{X}} : \max_{x\in\mathcal{X}} |f(x)| \leqslant 1\right\}$ *be a set of observables, then*

$$\|\mu-\nu\|_{\text{TV}} = \frac{1}{2}\sup\left\{\sum_{x\in\mathcal{X}} f(x)\mu(x) - \sum_{x\in\mathcal{X}} f(x)\nu(x) : f \in F\right\}.$$

Proof. Since $\max_x |f(x)| \leqslant 1$, it follows that

$$\frac{1}{2}\left|\sum_x f(x)(\mu(x)-\nu(x))\right| \leqslant \frac{1}{2}\sum_x |\mu(x)-\nu(x)| = \|\mu-\nu\|_{\text{TV}}.$$

For the reverse inequality, we define $f^*(x) \triangleq \mathbb{1}_{\{x\in B\}} - \mathbb{1}_{\{x\notin B\}}$ in terms of set $B \triangleq \{x \in \mathcal{X} : \mu(x) \geqslant \nu(x)\}$. It is clear that $\max_x |f(x)| = 1$, and we have

$$\frac{1}{2}\left|\sum_x f^*(x)(\mu(x)-\nu(x))\right| = \frac{1}{2}\sum_x |\mu(x)-\nu(x)| = \|\mu-\nu\|_{\text{TV}}.$$

□

4.1 Coupling and total variation distance

Definition 4.5. A **coupling** of two probability distributions $\mu,\nu \in \mathcal{M}(\mathcal{X})$ is a pair of random variables $(X,Y) : \Omega \to \mathcal{X}^2$ defined on a single probability space $(\Omega,\mathcal{F},P)$ such that the marginal distribution of X is μ and the marginal distribution of Y is ν. That is, a coupling (X,Y) satisfies $P\{X=x\} = \mu(x)$ and $P\{Y=y\} = \nu(y)$ for all $x,y \in \mathcal{X}$.

Remark 3. A coupling always exists since any two distributions μ and ν, can always have an independent coupling.

Definition 4.6. For distributions $\mu,\nu \in \mathcal{M}(\mathcal{X})$, the coupling (X,Y) is **optimal** if $\|\mu-\nu\|_{\text{TV}} = P\{X \neq Y\}$.

Remark 4. When the two distributions are not identical, it will not be possible for the random variables to always have the same value. Total variation distance between μ and ν determines how close can a coupling get to having X and Y identical.

Proposition 4.7. *Let* $\mu,\nu \in \mathcal{M}(\mathcal{X})$*, then* $\|\mu-\nu\|_{\text{TV}} = \inf\{P\{X\neq Y\} : (X,Y) \text{ a coupling of distributions } (\mu,\nu)\}$.

Proof. For any coupling (X,Y) of the distributions μ,ν and any event $A \subseteq \mathcal{X}$, we have

$$\mu(A)-\nu(A) = P\{X\in A\} - P\{Y\in A\} \leqslant P\{X\in A, Y\notin A\} \leqslant P\{X\neq Y\}.$$

Therefore, it follows that $\|\mu-\nu\|_{\text{TV}} \leqslant P\{X\neq Y\}$ for all couplings (X,Y) of distributions μ,ν.

Next we find a coupling (X,Y) for which $\|\mu-\nu\|_{\text{TV}} = P\{X\neq Y\}$. In terms of the set $B = \{x\in\mathcal{X} : \mu(x) \geqslant \nu(x)\}$, we can write

$$p \triangleq \sum_{x\in\mathcal{X}} \mu(x)\wedge\nu(x) = \mu(B^c) + \nu(B) = 1 - (\mu(B)-\nu(B)) = 1 - \|\mu-\nu\|_{\text{TV}}.$$

By the definition of p, we have $\gamma_3 \triangleq \frac{\mu\wedge\nu}{p} \in \mathcal{M}(\mathcal{X})$. Using the definition of B, we also define the following two distributions $\gamma_1,\gamma_2 \in \mathcal{M}(\mathcal{X})$ as

$$\gamma_1 \triangleq \frac{\mu-\nu}{\|\mu-\nu\|_{\text{TV}}}\mathbb{1}_B, \qquad \gamma_2 \triangleq \frac{\nu-\mu}{\|\mu-\nu\|_{\text{TV}}}\mathbb{1}_{B^c}.$$

We define a binary random variable $\xi : \Omega \to \{0,1\}$ such that $\mathbb{E}\xi = p$, and the conditional distribution of (X,Y) given ξ such that

$$P(\{(X,Y) = (x,y)\} \mid \xi) = \gamma_3(x)\mathbb{1}_{\{x=y\}}\xi + (1-\xi)\gamma_1(x)\gamma_2(y)\mathbb{1}_{\{x\neq y\}}.$$

Since $\gamma_1, \gamma_2, \gamma_3$ are distributions, it follows that $P\{(X,Y) = (x,y)\} = p\gamma_3(x)\mathbb{1}_{\{x=y\}} + (1-p)\gamma_1(x)\gamma_2(y)\mathbb{1}_{\{x\neq y\}}$ is a joint distribution function. From the definition of the set B, we observe that

$$P\{X = x\} = p\gamma_3(x) + (1-p)\gamma_1(x) = \mu(x) \wedge \nu(x) + (\mu(x) - \nu(x))\mathbb{1}_{\{x\in B\}} = \mu(x)$$
$$P\{Y = y\} = p\gamma_3(y) + (1-p)\gamma_2(y) = \mu(y) \wedge \nu(y) + (\nu(y) - \mu(y))\mathbb{1}_{\{y\notin B\}} = \nu(y).$$

That is, (X,Y) is a coupling of the distributions μ, ν and $P\{X \neq Y\} = 1 - p = \|\mu - \nu\|_{\mathrm{TV}}$. □