Alma Mater Studiorum · Università di Bologna

DIPARTIMENTO DI MATEMATICA
Corso di Laurea (Triennale) in Matematica

Symmetric particle systems: from empirical measures to propagation of chaos

Tesi di Laurea in Probabilitá e statistica

Relatore:
Chiar.mo Prof.
Andrea Pascucci

Presentata da:
Alessio Santulli

Anno Accademico 2024-2025

Alla mia famiglia

Introduction

One of the most actively studied topics in many areas of applied mathematics is the theory of interacting particle systems. These statistical models are used to describe any collection of *particles*, from theoretical notions to real-world applications: such systems might represent a group of molecules in a gas, a flock of birds, a crowd of people.
In fact, the study of particle systems dates back more than a century in statistical mechanics, when physicists like Boltzmann and Maxwell were studying the behavior of particles in a gas, laying the foundations for the various thermodynamic laws that we know today. Boltzmann realized that it was impractical to describe the exact configuration of over trillions of molecules. Moreover, the molecules constantly collide into each other, a phenomenon known as *dependence* in probability and statistics, which makes predicting particle movements even harder: in any instant, a small change of the state of one molecule could result in a much bigger and non-negligible difference in the macroscopic scale. This is commonly known as the *butterfly effect* in chaos theory. Because of all this, a statistical description is preferred: mathematical tools like empirical measures and stochastic processes need to be used in order to reduce computational complexity.
For a simpler analysis, these particle systems are always assumed to be *symmetric*, evolving in a Polish state space: many results about convergence of measures may not apply otherwise. The main concept behind propagation of chaos is that for a large number of particles they behave almost independently from one another. This property carries on as time evolves, as discussed by Mark Kac in the 1950s [3], who was the one who defined chaos in a mathematical sense and came up with the expression *propagation of chaos*; Kac developed this idea during his studies about kinetic theory, but his publications are still extended today, including machine learning, biology and physics.

The goal of this thesis is to explore this asymptotic independence: for each individual element in the system, we will use the empirical measure to define an approximating distribution that approaches the true distribution in a rate of $O(1/N)$ for large Ns according to the total variation distance. This is a significant quantitative bound, as the total variation distance is the strongest and most intuitive amongst all metrics in the space of random measures. Furthermore, this bound is *uniform in time*, that is, it holds at every time instant.

Chapter 1 will introduce some key results about the total variation distance and probability theory that will be used later. Chapter 2 describes the empirical measures and their convergence: the aim is to show the differences between the independent and identically distributed case and the opposite. Finally, Chapter 3 will focus on propagation of chaos and the main setting and achievements of this report; Section 3.1 is a brief review of infinite particle systems and remarkable contributions by de Finetti.

Contents

Chapter 1

General results

1.1 Total variation distance

Definition 1.1 (Total variation distance)**.** Let (E, d_E) be a Polish space and μ,ν two probability measures in $P(E)$. The total variation distance between μ and ν is defined by

$$\|\mu - \nu\|_{TV} = 2 \inf_{X \sim \mu, Y \sim \nu} P(X \neq Y).$$

We say that μ converges to ν in total variation norm if

$$\|\mu - \nu\|_{TV} \overset{n\to\infty}{\longrightarrow} 0.$$

This is the strongest norm amongst the different notions of distance in the space of probability measures: intuitively, it shows the largest possible difference between the probabilities assigned by two measures to the same event.

Proposition 1.2. *If μ and ν are two probability measures on a measurable space $(E, \mathcal{F})$, then it holds:*

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

Proof. Since the total variation distance is defined with a coefficient 2 in front of the infimum, we will prove that

$$\inf_{X \sim \mu, Y \sim \nu} P(X \neq Y) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

We adapt the proof found in Chapter 3 of [5]. Let $X \sim \mu$ and $Y \sim \nu$. For $A \in \mathcal{F}$ we have

$$\begin{aligned}
|\mu(A)-\nu(A)| =&\Bigg|\bigg(P\Big((X\in A)\cap(X=Y)\Big)+P\Big((X\in A)\cap(X\neq Y)\Big)\bigg)-\\
&-\bigg(P\Big((Y\in A)\cap(X=Y)\Big)+P\Big((Y\in A)\cap(X\neq Y)\Big)\bigg)\Bigg| \leq\\
\leq&\Big|P\Big((X\in A)\cap(X=Y)\Big)-P\Big((Y\in A)\cap(X=Y)\Big)\Big|+\\
&\Big|P\Big((X\in A)\cap(X\neq Y)\Big)-P\Big((Y\in A)\cap(X\neq Y)\Big)\Big| =\\
=&0+P\bigg(\Big((X\in A)\setminus(Y\in A)\Big)\cap(X\neq Y)\bigg)\leq P(X\neq Y),
\end{aligned}$$

using the measure properties of monotonicity and additivity. We take the supremium over $A \in \mathcal{F}$ to obtain

$$\sup_{A\in\mathcal{F}}|\mu(A)-\nu(A)| \leq P(X\neq Y),$$

which is true for all $X \sim \mu$ and $Y \sim \nu$, meaning

$$\sup_{A\in\mathcal{F}}|\mu(A)-\nu(A)| \leq \inf_{X\sim\mu, Y\sim\nu} P(X\neq Y).$$

Now we need to construct a couple (X, Y) of random variables for which the equality is true. We define $p := 1-\sup_{A\in\mathcal{F}}|\mu(A)-\nu(A)| \in [0,1]$ and study each of the following cases:

1. If $p = 0$, then $\sup_{A\in\mathcal{F}}|\mu(A)-\nu(A)| = 1$ and the supports of μ and ν are disjoints. This gives $\sum_{x\in E}\mu(x)\nu(x) = 0$. We can then choose $X \sim \mu$ and $Y \sim \nu$ such that X and Y are two independent random variables. Therefore
$$P(X=Y) = \sum_{x\in E}\mu(x)\nu(x) = 0$$

2. If $p = 1$, then $\sup_{A\in\mathcal{F}}|\mu(A)-\nu(A)| = 0$ and $\mu = \nu$. We take $X \sim \mu$ and $Y = X$.

3. If $0 < p < 1$, then we define the following measure: $\mu(x)\wedge\nu(x) := min(\mu(x),\nu(x))$. Let $U \sim \frac{1}{p}(\mu\wedge\nu), V \sim \frac{1}{1-p}(\mu-(\mu\wedge\nu))$ and $W \sim \frac{1}{1-p}(\nu-(\mu\wedge\nu))$. We notice that $p = \sum_{x\in E}(\mu(x)\wedge\nu(x))$, because
$$\sum_{x\in E}(\mu(x)\wedge\nu(x)) = \frac{1}{2}\sum_{x\in E}(\mu(x)+\nu(x)-|\mu(x)-\nu(x)|) =$$

$$= 1 - \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)| = p.$$

Now let B be a Bernoulli random variable independent of U,V,W, with $P(B = 1) := p$. We can finally define:

$$(X, Y) = \begin{cases} (U, U) & \text{if } B = 1 \\ (V, W) & \text{if } B = 0 \end{cases}$$

This way we have $X \sim \mu$ and $Y \sim \nu$, and P(V=W)=0 due to the definition of the laws of V and W, thus

$$P(X = Y) = P(U = U)P(B = 1) + P(V = W)P(B = 0) = p$$

$\square$

Remark 1.3. This characterization makes it easier to prove that the total variation distance satisfies the properties of a distance: indeed, we have $\|\mu - \nu\|_{TV} \geq 0$ and

$$\|\mu - \nu\|_{TV} = 0 \quad \Leftrightarrow \quad \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)| = 0 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \mu(A) - \nu(A) = 0 \quad \forall A \in \mathcal{F} \quad \Leftrightarrow \quad \mu = \nu.$$

Moreover, using the triangle inequality, it holds that $\|\mu - \nu\|_{TV} \leq \|\mu - \tau\|_{TV} + \|\tau - \nu\|_{TV}$, because

$$\|\mu - \nu\|_{TV} = \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)| = \sup_{A\in\mathcal{F}} |\mu(A) - \tau(A) + \tau(A) - \nu(A)| \leq$$

$$\leq \sup_{A\in\mathcal{F}} |\mu(A) - \tau(A)| + \sup_{A\in\mathcal{F}} |\tau(A) - \nu(A)| = \|\mu - \tau\|_{TV} + \|\tau - \nu\|_{TV}$$

Proposition 1.4. *The total variation distance satisfies:*

$$\|\mu - \nu\|_{TV} = \sup_{\|\varphi\|_\infty \leq 1} \left| \int_E \varphi(x)\mu(\mathrm{d}x) - \int_E \varphi(x)\nu(\mathrm{d}x) \right|$$

This is a dual representation of the total variation distance, often used in analysis: we take the supremum of the expectation of a measurable bounded function φ with respect to the measure $\mu - \nu$ over all test functions φ.

Proof. Let $\tau := \mu - \nu$ (see [7]). Using the Hahn-Jordan decomposition, we define two sets $E^+, E^- \in \mathcal{F}$ such that $E^+ \cap E^- = \emptyset$ and $E^+ \cup E^- = E$. We also define

$$\tau^+(A) := \tau(A \cap E^+) \quad \text{and} \quad \tau^-(A) := \tau(A \cap E^-)$$

which are two non-negative measures that satisfy $\tau = \tau^+ + \tau^-$.
Given a test function φ such that $\|\varphi\|_\infty \leq 1$,

$$\int_E \varphi \,\mathrm{d}\mu - \int_E \varphi \,\mathrm{d}\nu = \int_E \varphi \,\mathrm{d}\tau = \int_{E^+} \varphi \,\mathrm{d}\tau^+ - \int_{E^-} \varphi \,\mathrm{d}\tau^- \leq$$
$$\leq \int_{E^+} 1 \,\mathrm{d}\tau^+ - \int_{E^-} (-1) \,\mathrm{d}\tau^- = \tau^+(E) + \tau^-(E) = 2\tau^+(E)$$

where the latter equality holds because $0 = \mu(E) - \nu(E) = \tau(E) = \tau^+(E) - \tau^-(E)$. Likewise, $\int_E \varphi \,\mathrm{d}\nu - \int_E \varphi \,\mathrm{d}\mu \leq 2\tau^+(E)$. Moreover, we have an equality if $\varphi = \mathbb{1}_{E^+} - \mathbb{1}_{E^-}$, thus

$$\sup_{\|\varphi\|_\infty \leq 1} \left| \int_E \varphi \,\mathrm{d}\mu - \int_E \varphi \,\mathrm{d}\nu \right| = 2\tau^+(E).$$

Furthermore, for any $A \in \mathcal{F}$ we have

$$\mu(A) - \nu(A) = \tau(A) = \tau^+(A) - \tau^-(A) \leq \tau^+(A) \leq \tau^+(E) \quad \text{and} \quad \text{similarly,}$$
$$\nu(A) - \mu(A) = -\tau(A) = \tau^-(A) - \tau^+(A) \leq \tau^-(A) \leq \tau^-(E) = \tau^+(E).$$

This way $|\mu(A) - \nu(A)| \leq \tau^+(E)$, and we can achieve the equality if $A = E^+$, hence

$$\sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)| = \tau^+(E) = \frac{1}{2} \sup_{\|\varphi\|_\infty \leq 1} \left| \int_E \varphi \,\mathrm{d}\mu - \int_E \varphi \,\mathrm{d}\nu \right|$$

□

1.2 Convergence of probability measures

Understanding the different types of convergence of probability measures is a fundamental concept in probability theory and statistics [4]: a central notion when studying sequences of measures is to determine how they behave as they approach the limit measure. These tools will be necessary when working with probability measures in the following sections.

Remark 1.5. Throughout this review, the expression $\langle \mu, \varphi \rangle$ will be used as an alternative notation for the integral of a test function φ over a measure μ.

Definition 1.6 (Weak convergence)**.** A sequence of probability measures $(\mu_N)_N$ converges weakly (or in distribution) towards μ when

$$\forall \varphi \in C_b(E), \quad \langle \mu_N, \varphi \rangle \overset{N\to\infty}{\longrightarrow} \langle \mu, \varphi \rangle .$$

Since $(\mu_N)_N, \mu \in P(E)$, the weak convergence in $P(E)$ is defined as the related weak-$\star$ convergence in $C_b(E)^\star$, given that $P(E)$ is a subset of $C_b(E)^\star$.

Remark 1.7. In probability theory, this means that

$$\lim_{N\to\infty} \mathbb{E}[\varphi(\mu_N)] = \mathbb{E}[\varphi(\mu)]$$

for all $\varphi \in C_b(E)$

We also recall other types of convergence for a more complete picture of the topic.

Definition 1.8. We consider a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$

1. We say X_n converges in probability towards X if

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \epsilon\right) = 0 \quad \forall \epsilon > 0.$$

2. We say X_n converges in L^p towards X for $p \geq 1$ if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = 0.$$

3. We say X_n converges almost surely towards X if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

for almost every $\omega \in \Omega$.

Among all types of convergence, almost sure convergence is the strongest : it requires that $X_n(\omega)$ converges to a limit for almost every outcome ω. The following result is an important milestone in probability theory and statistics that we will use in the following sections.

Theorem 1.9 (Strong Law of Large Numbers). *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of independent and identically distributed random variables. We consider*

$$M_n := \frac{1}{n}\sum_{i=1}^{n} X_i,$$

which is the sample mean of the first n *random variables of the sequence. Then it holds:*

$$M_n \overset{n\to\infty}{\to} \mathbb{E}[X_1] \quad \text{almost surely.}$$

Chapter 2

Empirical measures

This section will introduce a fundamental object in probability and statistics. The empirical measure addresses the problem of approximating an unknown probability measure through *empirical samples*. Empirical measures is greatly used in many fields such as econometrics and finance. The approximation improves as the sample size increases, which motivates the need of a mathematical analysis in order to understand the different properties of the true measure.
The ideal case occurs when the samples are independent and identically distributed, which allows us to apply some relevant results in Section 2.1. However, in many real-world situations, this assumption does not apply, since there are often some dependencies, like in the study of finite particle systems; Section 2.2 is dedicated to a non-i.i.d setting as we try to preserve some of the theoretical results using the concept of *exchangeability*.

2.1 I.i.d. setting

Definition 2.1 (Empirical measure)**.** Let $(X_1, X_2, \dots, X_n)$ be a collection of independent and identically distributed random variables with values in the state space E. The empirical measure associated to this sequence is defined as:

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

For all $n \in \mathbb{N}$ the empirical measure is a discrete probability measure which puts $\frac{1}{n}$ mass on each sample. If μ is the true probability measure, Portmanteau's theorem grants weak convergence of $(\mu_n)_{n\in\mathbb{N}}$ towards μ, meaning $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$ for all bounded continuous functions f (see Appendix A). However, convergence in total variation often fails as seen in the example below.

Example 2.2. We choose $X_1, X_2, \ldots, X_n \sim U[0,1]$, independent random variables with the uniform distribution on [0,1]. The empirical measure will be

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}.$$

We have $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ weakly, where μ is the Lebesgue measure on [0,1]. Indeed, using the Strong Law of Large Numbers, we get

$$\int f\mathrm{d}\mu_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \overset{a.s.}{\to} \mathbb{E}[f(X_1)] = \int_0^1 f(x)\mathrm{d}x = \int f\mathrm{d}\mu$$

for all $f \in C_b([0,1])$. Now we define $A_n = \{X_1, X_2, \ldots, X_n\} \subseteq [0,1]$: this is a countable set for all $n \in \mathbb{N}$, meaning $\mu(A_n) = 0$, thus

$$\|\mu_n - \mu\|_{TV} = \sup_{A\subseteq[0,1]} |\mu_n(A) - \mu(A)| \geq |\mu_n(A_n) - \mu(A_n)| = |1-0| = 1.$$

So taking the limit $n \to \infty$, $\|\mu_n - \mu\|_{TV}$ cannot converge to zero.

We recall that the *cumulative distribution function* (CDF) of a distribution μ is a real valued function defined as $F_\mu(x) = \mu((-\infty, x])$; in our case, the CDF of the empirical measure μ_n is

$$F_n(x) = \mu_n((-\infty, x]) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,x]}(X_i)$$

which is called the *empirical cumulative distribution function* (ECDF). As n goes to infinity, F_n converges uniformly to F_μ almost surely: this is stated by Glivenko-Cantelli's theorem, the idea of the proof uses the strong law of large numbers, as

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,x]}(X_i) \xrightarrow{a.s.} \mathbb{E}[\mathbb{1}_{(X_1\leq x)}] = P(X_1 \leq x) = F_\mu$$

2.2 Non i.i.d setting

Since we need to link empirical measures to propagation of chaos, we also have to consider non independent frameworks: this is because in a finite particle system all the elements collide into each other (the particles are said to interact), meaning that the particle distributions *depend* on each other. As a result, every property in the last section usually fail: the only assumption we need to make is some type of symmetry.

Definition 2.3 (Exchangeability)**.** A family of random variables $(X^i)_{i\in I}$ is *exchangeable* when the law of $(X^i)_{i\in I}$ is invariant under every permutation of a finite number of indexes $i \in I$.

This means that the joint distribution of the sequence $(X_1, X_2, \ldots, X_n)$ equals the joint distribution of $(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)})$ for any $n \in \mathbb{N}$ and any permutation $\sigma \in \mathcal{S}_n$. With this hypothesis, the empirical measure does converge weakly to a (random) non fixed measure, which can depend on different factors.

Example 2.4 (Pólya urn)**.** Suppose we have an urn with a black ball and a white ball. At each iteration we draw a ball from the urn uniformly, we reinsert it and add another ball of the same color. Let $X_n = \{b, w\}$ indicate the color drawn at the n-th iteration.
The variables $X_1, X_2, \ldots, X_n$ are not i.i.d because every draw depends on the previous ones, however they are *exchangeable* because the probability of extracting a certain color sequence does not depend on the order but only on the number of balls in the urn. Let's define the empirical measure:

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}.$$

It evaluates the frequency of both outcomes in the first n draws, so $\mu_n(b) = \frac{\text{Number of blacks drawn in } n \text{ iterations}}{n}$. As n goes to infinity, this ratio converges to a fixed value $\alpha := \lim_{n\to\infty} \mu_n(b)$, which is drawn uniformly from [0,1]: consequently, we get

$$\mu_n \overset{n\to\infty}{\longrightarrow} \mu_\alpha := \alpha\delta_b + (1-\alpha)\delta_w,$$

which is a Bernoulli random variable with a non fixed parameter α. Intuitively, if the first ball drawn is black, the second iteration will likely result in a black draw, meaning α will likely be greater than 0.5.

Example 2.5. Let $X_n \sim Be_{p_n}$ where $p_n = \frac{1+sin(log(n))}{2}$; these are neither identically distributed nor exchangeable. The empirical measure is characterized by:

$$\mu_n(1) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We want to prove that the empirical measure does not converge weakly. Let's assume it does converge to a random measure μ: by choosing $\varphi = id$ in Remark 1.7 we would get $\mathbb{E}[\mu_n(1)] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[\mu(1)]$, but this cannot be verified by any measure μ because

$$\mathbb{E}[\mu_n(1)] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(X_i = 1)] = \frac{1}{n}\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \frac{1 + sin(log(i))}{2n}$$

is a real-valued sequence that does not converge as $n \to \infty$ due to the oscillations of the sine function.

These examples illustrate that exchangeability can still lead to weak convergence of empirical measures, as opposed to Example 2.5.

Chapter 3

Propagation of chaos

So far we have illustrated the main tools about empirical measures and their general properties. The final step of this thesis is to use these concepts in order to establish the framework necessary for studying the propagation of chaos.
Let us recall the setting and all the key assumptions:

- E is the state space, assumed to be Polish (i.e. separable and completely metrizable);

- We consider a system of N particles in E, described by:

$$\mathcal{X}_I^N \equiv (\mathcal{X}_t^N)_{t\in I} \equiv (X_t^1, \cdots, X_t^N)_{t\in I};$$

- Each element X_t^i is a stochastic process in the space E: it is a function that keeps track of the position of the i-th particle at time $t \in I \equiv [0, T]$;

- We also assume that $(X_t^1, \ldots, X_t^N)$ are exchangeable; this means that at any time t, $X_t^1, \ldots, X_t^N$ are exchangeable according to Definition 2.3. Since $\mathcal{X}_I^N$ is a finite system, the N particles, as well as the related stochastic processes, are *not* independent: the particles interact with each other and their trajectory is affected by collisions. Because of this, for a large number of particles, it becomes impractical to work with a microscopic description, as we need to keep track of every single position. Therefore, a statistical analysis is required so that we can approximate the particles behavior without storing too much information;

- We define $f_t^N \in P(E^N)$ as the true joint distribution of the N-particle system: it simply indicates the configuration of the system at time t all at once;

- For bigger values of N, it is best to use the empirical measure:

$$\mu_{\mathcal{X}_t^N} := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^i} \in P(E).$$

  This type of approach is commonly used in various optimization problems: we choose to study a one dimensional object, which belongs to $P(E)$, instead of $f_t^N \in P(E^N)$.

While the theory of propagation of chaos involves analyzing how stochastic processes behave over time, it is helpful to first consider a static framework, as the dynamic theory heavily relies on static results; it also enables us to establish important approximation results (like the one in Theorem 3.2) without dealing with time complexity. For this reason, from now on we will drop the t subscript and focus on the behavior of the particle system $\mathcal{X}^N$ at a fixed time in I.
If we consider the following bijective map

$$\begin{aligned} \Phi: \; E^N &\longrightarrow P(E) \\ \bar{x}^N &\longmapsto \mu_{\bar{x}^N} \end{aligned}$$

we see that the law of the empirical measure is given by $\mu_{\mathcal{X}^N} \sim f^N \circ \Phi^{-1} =: F^N$. More precisely, there is a one-to-one mapping between the quotient $E^N/\mathcal{S}_n$ and $\hat{P}_N(E)$, where $\hat{P}_N(E)$ indicates the space of empirical measures of size N on E.

Lemma 3.1. *$\forall \varphi \in C_b(E)$, it holds that:*

$$\mathbb{E}\left[\langle \mu_{\mathcal{X}^N}, \varphi\rangle\right] = \left\langle f^1, \varphi \right\rangle$$

Proof. By simply applying exchangeability and linearity of the expected value, we get:

$$\mathbb{E}\left[\langle \mu_{\mathcal{X}^N}, \varphi\rangle\right] = \mathbb{E}\left[\int_E \varphi(x)\mathrm{d}\mu_{\mathcal{X}^N}(x)\right] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(X^i)\right] =$$

$$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\varphi(X^i)\right] = \frac{1}{N}\sum_{i=1}^{N}\int_E \varphi(x)\mathrm{d}f^i(x) = \frac{1}{N}N\int_E \varphi(x)\mathrm{d}f^1(x) = \left\langle f^1, \varphi\right\rangle$$

□

This relation allows to derive the first marginal of f^N from the expectation of the empirical measure; thanks to the exchangeability we can actually reconstruct the law of every particle. The k-marginals also match the natural definition of marginal from the joint distribution for each $1 \leq k \leq N$:

$$\left\langle f^k, \varphi_k\right\rangle := \left\langle f^N, \varphi_k \otimes 1^{\otimes(N-k)}\right\rangle$$

.

Theorem 3.2 (Approximation rate of marginals)**.** *For $1 \leq k \leq N$, let the moment measure $\widetilde{F}^k \in P(E^k)$ be defined by:*

$$\left\langle \widetilde{F}^k, \varphi_k\right\rangle = \int_{P(E)} \left\langle \nu^{\otimes k}, \varphi_k\right\rangle F^N(\mathrm{d}\nu), \qquad \forall \varphi_k \in C_b(E^k).$$

Then it holds that

$$\left\|f^k - \widetilde{F}^k\right\|_{TV} \leq \frac{2k(k-1)}{N}$$

Proof. Given that F^N is the law of the random measure $\mu_{\mathcal{X}^N}$ we get $\left\langle \widetilde{F}^k, \varphi_k\right\rangle = \left\langle \mathbb{E}\left[\mu_{\mathcal{X}^N}^{\otimes k}\right], \varphi_k\right\rangle$ for all $\varphi_k \in C_b(E^k)$; so for any test function φ_k, we have:

$$\left\langle \widetilde{F}^k, \varphi_k\right\rangle = \left\langle \mathbb{E}\left[\mu_{\mathcal{X}^N}^{\otimes k}\right], \varphi_k\right\rangle =$$

$$= \int_{P(E)} \left\langle \nu^{\otimes k}, \varphi_k\right\rangle f^N(\Phi^{-1}(\mathrm{d}\nu)) \overset{\nu=\Phi(\bar{x}^N)}{=} \int_{E^N} \left\langle \mu_{\bar{x}^N}^{\otimes k}, \varphi_k\right\rangle f^N(\mathrm{d}\bar{x}^N).$$

Using the symmetry of f^k and the definition of marginal we also get:

$$\left\langle f^k, \varphi_k\right\rangle = \int_{E^N} \frac{1}{N!}\sum_{\sigma\in\mathcal{S}_N} \varphi_k(x^{\sigma(1)}, \ldots, x^{\sigma(k)}) f^N(\mathrm{d}\bar{x}^N)$$

and we combine the two:

$$\left|\left\langle f^k - \widetilde{F}^k, \varphi_k\right\rangle\right| = \left|\left\langle f^k, \varphi_k\right\rangle - \left\langle \widetilde{F}^k, \varphi_k\right\rangle\right| =$$

$$= \left|\int_{E^N}\left(\frac{1}{N!}\sum_{\sigma\in\mathcal{S}_N}\varphi_k(x^{\sigma(1)}, \ldots, x^{\sigma(k)}) - \left\langle \mu_{\bar{x}^N}^{\otimes k}, \varphi_k\right\rangle\right) f^N(\mathrm{d}\bar{x}^N)\right| \leq$$

$$\leq \sup_{\bar{x}^N\in E^N}\left|\frac{1}{N!}\sum_{\sigma\in\mathcal{S}_N}\varphi_k(x^{\sigma(1)}, \ldots, x^{\sigma(k)}) - \left\langle \mu_{\bar{x}^N}^{\otimes k}, \varphi_k\right\rangle\right|.$$

We rewrite the first part as

$$\frac{1}{N!}\sum_{\sigma\in\mathcal{S}_N}\varphi_k(x^{\sigma(1)},\dots,x^{\sigma(k)}) = \frac{1}{A_{k,N}}\sum_{\substack{i_1,\dots,i_k\\ \text{pairwise distinct}}}\varphi_k(x^{i_1},\dots,x^{i_k}),$$

where $A_{k,N} := N!/(N-k)!$ represents the number of pairwise distinct tuples $(i_1,\dots,i_k)$ of integers between 1 and N, whereas for the second part we have:

$$\begin{aligned}\left\langle \mu_{\bar{x}^N}^{\otimes k},\varphi_k\right\rangle &= \frac{1}{N^k}\sum_{i_1,\dots,i_k}\varphi_k(x^{i_1},\dots,x^{i_k}) = \\ &= \frac{1}{N^k}\sum_{\substack{i_1,\dots,i_k\\ \text{pairwise distinct}}}\varphi_k(x^{i_1},\dots,x^{i_k}) + R_{k,N},\end{aligned}$$

where $R_{k,N}$ keeps track of the non pairwise distinct tuples (i.e. there is at least one repetition). If we go back to the expression inside the supremum we get:

$$\begin{aligned}&\left|\frac{1}{N!}\sum_{\sigma\in\mathcal{S}_N}\varphi_k(x^{\sigma(1)},\dots,x^{\sigma(k)}) - \left\langle \mu_{\bar{x}^N}^{\otimes k},\varphi_k\right\rangle\right| = \\ &= \left|\frac{N^k - A_{k,N}}{A_{k,N}N^k}\sum_{\substack{i_1,\dots,i_k\\ \text{pairwise distinct}}}\varphi_k(x^{i_1},\dots,x^{i_k}) - R_{k,N}\right| \le \\ &\le \left|\frac{N^k - A_{k,N}}{A_{k,N}N^k}\left\|\varphi_k\right\|_\infty A_{k,N} + \left\|\varphi_k\right\|_\infty\left(1-\frac{A_{k,N}}{N^k}\right)\right| = 2\left\|\varphi_k\right\|_\infty\left(1-\frac{A_{k,N}}{N^k}\right).\end{aligned}$$

Lastly, we obtain that

$$1-\frac{A_{k,N}}{N^k} = 1-\prod_{i=0}^{k-1}\left(1-\frac{i}{N}\right) \le 1-\left(1-\frac{k-1}{N}\right)^k;$$

using the formula $(1-x)^k \ge 1-kx$ for $|x|<1$ from the binomial series we get

$$1-\left(1-\frac{k-1}{N}\right)^k \le \frac{k(k-1)}{N}$$

and we conclude by using Proposition 1.4. □

This theorem is a major result in the theory of propagation of chaos: it provides a quantitative bound on the total variation norm between the k-marginal f^k and the k-th moment measure $\widetilde{F}^k$: their difference with respect

to the strongest norm is controlled by a fixed constant which goes to zero as $N \to \infty$, meaning that as the number of particles increases the k-th moment measure characterizes the marginal f^k.
From the probabilistic point of view, we can interpret the moment $\widetilde{F}^k$ this way: we sample a random measure ν with distribution F^N (which again is the law of the empirical measure $\mu_{\mathcal{X}_t^N}$), then we sample k *independent* values from ν, thus forming the product measure $\nu^{\otimes k}$, and we take the average of $\nu^{\otimes k}$ integrating over F^N.
We started our analysis with a system of interacting particles, however this result shows that, as N increases, the laws of each particle (the k-marginals) asymptotically become i.i.d: independent because of the definition of $\widetilde{F}^k$ and identically distributed due to the convergence of the empirical measure to a random measure (Section 2.2). Natural extensions of these ideas can be found in [3] using the notion of chaoticity developed by Kac: specifically, given $f \in P(E)$, a sequence f^N of symmetric probability measures on E^N is said to be f-chaotic when for any $k \in \mathbb{N}$ and any function $\varphi_k \in C_b(E^N)$,

$$\lim_{N\to\infty} \left\langle f^N, \varphi_k \otimes \mathbb{1}^{\otimes(N-k)} \right\rangle = \left\langle f^{\otimes k}, \varphi_k \right\rangle .$$

This means that the k-marginal satisfies $f^k \to f^{\otimes k}$ weakly.

3.1 Infinite particle systems

As anticipated in the previous section we will look at the highly studied framework involving the limit $N \to \infty$. We will consider an infinite sequence of exchangeable random variables $\mathcal{X} \equiv (X_1, X_2, \ldots)$: like in the finite version in Definition 2.3, the law of $(X^i)_{i\in I}$ is invariant under every permutation of an infinite number of indexes $i \in I$. Likewise, the set $(X_1, X_2, \ldots)$ is described by the infinite dimensional symmetric measure f^∞: for each $k \in \mathbb{N}$ we define the k-marginal as the joint law of $(X_1, X_2, \ldots, X_k)$. Exchangeability is crucial here so that the following compatibility relation is satisfied:

$$\forall \varphi_j \in C_b(E^j), \quad \left\langle f^k, \varphi_j \otimes \mathbb{1}^{\otimes(k-j)} \right\rangle = \left\langle f^j, \varphi_j \right\rangle$$

for $1 \le j \le k$, which basically says that the j-marginal of f^k is f^j.
We also consider the infinite sequence of empirical measures: for $N \in \mathbb{N}$,

$$\mu_{\mathcal{X}^N} = \sum_{i=1}^{N} \delta_{X_i}$$

Definition 3.3 (Moment measure)**.** For $k \in \mathbb{N}$ the k-moment measure of $\pi \in P(P(E))$ is defined by:

$$\pi^k := \int_{P(E)} \nu^{\otimes k} \pi(\mathrm{d}\nu) \in P(E^k).$$

Similar to Theorem 3.2 this is like saying $\langle \pi^k, \varphi_k \rangle = \mathbb{E}_{\nu \sim \pi} \left[\langle \nu^{\otimes k}, \varphi_k \rangle \right]$ for $\varphi_k \in C_b(E^k)$; in addition, the moments $(\pi^k)_k$ satisfy the compatibility property.

Definition 3.4 (Convergence determining subsets)**.** A subset $\mathcal{F} \subset C_b(E)$ is called convergence determining if, for any sequence $(\mu_N)_N \in P(E)$ and $\mu \in P(E)$, the condition

$$\forall \varphi \in \mathcal{F}, \quad \langle \mu_N, \varphi \rangle \overset{N\to\infty}{\longrightarrow} \langle \mu, \varphi \rangle$$

implies that $\mu_N \to \mu$ weakly.

Basically, whenever the set of test functions $C_b(E)$ is too large it is sufficient to only check test functions in a convergence determining subset. We use the moment measures to prove the following characterization.

Proposition 3.5. *A sequence $(\pi_N)_N \in P(P(E))$ of random measures converges weakly towards $\pi \in P(P(E))$ if and only if*

$$\pi_N^k \overset{N\to\infty}{\longrightarrow} \pi^k$$

for all $k \geq 1$.

Proof. The direct implication follows easily from the continuity of the maps $\pi \to \pi^k$, as we obtain:

$$\langle \pi_N^k, \varphi \rangle = \int_{P(E)} \langle \nu^{\otimes k}, \varphi \rangle \, \pi_N(\mathrm{d}\nu) \overset{N\to\infty}{\longrightarrow} \int_{P(E)} \langle \nu^{\otimes k}, \varphi \rangle \, \pi(\mathrm{d}\nu) = \langle \pi^k, \varphi \rangle$$

for $\varphi \in C_b(E)$ and for any $k \geq 1$.
Conversely, we consider the functions:

$$\begin{array}{rccc} R_\varphi : & P(E) & \longrightarrow & \mathbb{R} \\ & \mu & \longmapsto & \langle \mu^{\otimes k}, \varphi \rangle \end{array}$$

which are called *monomial functions*. The weak convergence of $(\pi_N)_N$ towards π implies that $\langle \pi_N, R_\varphi \rangle$ converges towards $\langle \pi, R_\varphi \rangle$ for all monomial functions. The conclusion follows from [1], Lemma 3.9, by proving that the subset generated by these functions is convergence determining. □

We now look at an important result due to de Finetti: the setting of the classical de Finetti's theorem is with Bernoulli random variables [2] but it can be extended to generic E-valued random variables for some Polish space E.

Theorem 3.6 (De Finetti)**.** *Let E be a locally compact Polish space. Consider an infinite sequence $(f^N)_N$ of symmetric probability measures on E^N which satisfy the usual compatibility relation. Then there exists a (unique) measure $\pi \in P(P(E))$ such that:*

$$f^N = \pi^N := \int_{P(E)} \nu^{\otimes N} \pi(\mathrm{d}\nu)$$

The opposite of de Finetti's theorem is presented in the following representation theorem: we are able to link empirical measures to infinite particle systems.

Theorem 3.7 (De Finetti representation theorem)**.** *Let $\pi \in P(P(E))$. Then there exists an infinite sequence of E-valued exchangeable random variables $(X_i)_{i\in\mathbb{N}}$ such that, for any $k \geq 1$, the joint law of $(X_1, \dots, X_k)$ is $\pi^{\otimes k}$. Moreover, the limit*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} \delta_{X_i} \in P(E)$$

exists almost surely and it is π-distributed, or in other words the law of the limit of the empirical measure is π.

The first theorem states that an infinite exchangeable particle system is always associated to a unique element in $P(P(E))$, whereas the second one says that a random measure can always be represented by an infinite exchangeable particle system.

Example 3.8. Let $E = \{0, 1\}$, which means that $P(E)$ contains all and only probability measures of the form $\tau_p = p\delta_1 + (1-p)\delta_0$, hence $P(E) \equiv [0, 1]$. We fix $\pi = \text{Uniform}(0, 1)$: after sampling p from the uniform on $[0, 1]$ we define an infinite sequence of identically distributed random variables $X_i \sim Be_p$ which are *conditionally* independent, that is they are independent given the value p sampled from π. The sequence $(X_i)_{i\in\mathbb{N}}$ is exchangeable because it is infinite and τ_p is discrete, hence we only consider the number of positive or negative outcomes and not their order. Furthermore we can compute the joint law of $(X_1, \ldots, X_k)$ as:

$$P(X_1 = x_1, \ldots, X_k = x_k) = \int_0^1 \left(\prod_{i=1}^{k} p^{x_i}(1-p)^{1-x_i} \right) \pi(\mathrm{d}p) = \pi^{\otimes k}.$$

Lastly about the empirical measure,

$$\mu_k = \frac{1}{k} \sum_{i=1}^{k} \delta_{X_i};$$

as usual, this represents the number of successes in k trials. Taking the limit as k tends to infinity, we see that

$$\mu_k(1) = \frac{1}{k} \sum_{i=1}^{k} X_i \xrightarrow{i.i.d} p \quad \text{almost surely.}$$

So the limit $\lim_{k\to\infty} \mu_k$ exist almost surely and is equal to τ_p; since p was sampled form π, this limit is decided from the distribution π.

Appendix A

Portmanteau's theorem

A good characterization of the weak convergence can be expressed in Portmanteau's theorem [6].

Theorem A.1. *Let $(P_n)_{n\in\mathbb{N}}, P$ be probability measures. The following conditions are equivalent:*

i) P_n *converges weakly towards* P*;*

ii) $\limsup_{n\to\infty} P_n(F) \le P(F)$ *for all closed set* F*;*

iii) $\liminf_{n\to\infty} P_n(A) \ge P(A)$ *for all open set* A*;*

iv) $\lim_{n\to\infty} P_n(B) = P(B)$ *for all* P*-continuity sets* B*, that is a set* B *which satisfies* $P(\partial B) = 0$*.*

Proof. $i) \Rightarrow ii)$. If d is the distance defined on the space E, we set $d(x, F) := \min_{z\in F} d(x, z)$. For $\epsilon > 0$, let us define $f(x) := (1 - d(x, F)/\epsilon)^+$; this is a bounded and (uniformly) continuous function on E because $0 \le f(x) \le 1$ for all $x \in E$ and $|f(x) - f(y)| \le d(x, y)/\epsilon$. By definition, we have that $\forall x \in F, f(x) = 1$ and $\forall x \in F^\epsilon, f(x) = 0$, where $F^\epsilon := \{x \in E | d(x, F) < \epsilon\}$. Therefore for all $x \in E$,

$$\mathbb{1}_F(x) \le f(x) \le \mathbb{1}_{F^\epsilon}(x).$$

If we apply the expected value over P_n and P, we obtain

$$P_n(F) \le \int f \mathrm{d}P_n \le P_n(F^\epsilon) \quad \text{and}$$
$$P(F) \le \int f \mathrm{d}P \le P(F^\epsilon).$$

Using these inequalities, we get

$$\limsup_{n\to\infty} P_n(F) \le \limsup_{n\to\infty} \int f \mathrm{d}P_n = \int f \mathrm{d}P \le P(F^\epsilon)$$

and we conclude letting $\epsilon \to 0$.
ii) $\Leftrightarrow$ *iii)*. Trivial by taking complements.
ii) & *iii)* $\Rightarrow$ *iv)*. For all sets B we have

$$\begin{aligned} P(\overline{B}) \ge \limsup_{n\to\infty} P_n(\overline{B}) &\ge \limsup_{n\to\infty} P_n(B) \ge \\ &\ge \liminf_{n\to\infty} P_n(B) \ge \liminf_{n\to\infty} P_n(intB) \ge P(intB). \end{aligned}$$

If B is a P-continuity set then $P(\overline{B}) = P(B) = P(intB)$ and *iv)* follows.
iv) $\Rightarrow$ *i)*. Given a test function $f \in C_b(E)$, we may assume without loss of generality that $0 \le f \le 1$ by linearity. $U_t := \{x \in E | f(x) > t\}$ is an event, so we can write

$$\int f \mathrm{d}P_n = \int_0^\infty P_n(U_t)\mathrm{d}t = \int_0^1 P_n(U_t)\mathrm{d}t$$

and likewise for P. Since f is continuous, we have $\partial U_t \subseteq \{x \in E | f(x) = t\}$: this means that U_t is a P-continuity set except for countably many t. We conclude using *iv)* and the bounded convergence theorem:

$$\int f \mathrm{d}P_n = \int_0^1 P_n(U_t)\mathrm{d}t \overset{n\to\infty}{\longrightarrow} \int_0^1 P(U_t)\mathrm{d}t = \int f \mathrm{d}P.$$

$\square$

Bibliography

[1] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. i. models and methods. arXiv preprint arXiv:2203.00446, May 2022.

[2] P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, 8:745–764, 1980.

[3] M. Kac. *Foundations of Kinetic Theory*. University of California Press, Berkeley and Los Angeles, California, 1956.

[4] A. Pascucci. Teoria della probabilitá. Lecture notes, 2021.

[5] D. Pollard. Asymptopia, chapter 3, Feb 2005. Version 15feb05, Yale University.

[6] F. Simenhaus. Limit theorems. Lecture notes for Convergence 2024, Université Paris-Dauphine, 2019.

[7] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.