Markov Transitions and the Propagation of Chaos

Alex D. Gottlieb

Abstract

The propagation of chaos is a central concept of kinetic theory that serves to relate the equations of Boltzmann and Vlasov to the dynamics of many-particle systems. Propagation of chaos means that molecular chaos, i.e., the stochastic independence of two random particles in a many-particle system, persists in time, as the number of particles tends to infinity.

We establish a necessary and sufficient condition for a family of general n-particle Markov processes to propagate chaos. This condition is expressed in terms of the Markov transition functions associated to the n-particle processes, and it amounts to saying that chaos of random initial states propagates if it propagates for pure initial states.

Our proof of this result relies on the weak convergence approach to the study of chaos due to Sznitman and Tanaka. We assume that the space in which the particles live is homeomorphic to a complete and separable metric space so that we may invoke Prohorov's theorem in our proof.

We also show that, if the particles can be in only finitely many states, then molecular chaos implies that the specific entropies in the n-particle distributions converge to the entropy of the limiting single-particle distribution.

Contents

1 Introduction **2**

1.1 Overview 2

1.2 Definition of Propagation of Chaos and Statement of Main Result 4

2 Kinetic Theory and the Propagation of Chaos **9**

2.1 Dilute Gases 9

2.1.1 Boltzmann's equation 9

2.1.2 Particle systems for Boltzmann's equation 12

2.2 Plasmas and Stellar Systems 15

2.2.1 Vlasov's equation 16

2.2.2 McKean-Vlasov particle systems 17

3 Chaos and Weak Convergence **20**

3.1 Background 20

3.2 The Theorem of Sznitman and Tanaka 23

3.3 Chaos on Finite Sets 27

4 Propagation of Chaos **31**

4.1 Preliminaries 31

4.2 Lemmas 34

4.3 Theorems 37

5 Conclusion **42**

Chapter 1

Introduction

1.1 Overview

Kinetic theory is the analysis of nonequilibrium physical phenomena that emerge from the collective behavior of large numbers of particles. That analysis is accomplished by the techniques of probability theory; kinetic theory has an inherently statistical character. One of the notions of probability theory from which one can derive Boltzmann's equation and Vlasov's equation, two staples of kinetic theory, is the propagation of chaos.

The concept of propagation of chaos originated with Kac's Markovian models of gas dynamics [16]. Kac invented a class of interacting particle systems wherein particles collide at random with each other while the *density* of particles evolves deterministically in the limit of infinite particle number. A nonlinear evolution equation analogous to Boltzmann's equation governs the particle density. Grünbaum proved the propagation of chaos along Kac's lines for the spatially homogeneous Boltzmann equation given existence and smoothness assumptions on the Boltzmann semigroup [15]. The processes of Kac were further investigated with regard to their fluctuations about the deterministic infinite particle limit in [23, 39, 40].

McKean introduced propagation of chaos for interacting diffusions and analyzed what are now called McKean-Vlasov equations [21, 22]. Independently, Braun and Hepp [5] analyzed the propagation of chaos for Vlasov equations and proved a central limit theorem for the fluctuations. Analysis of the fluctuations and large deviations for McKean-Vlasov processes was carried out in [38, 35, 8]. Chorin [6] created a numerical method for the two dimensional Navier-Stokes equation by interleaving independent random walks into the discretized dynamics of interacting *vortex blobs*, smoothed and localized patches of vorticity that move without changing shape. Propagation of chaos has been studied in connection with this *random vortex method* by [20, 27, 25]. Other instances of the propagation of chaos have been studied in [26, 30, 31, 14]. A thorough analysis of the convergence of numerical schemes based on stochastic particle methods for McKean-Vlasov equations in one dimension is undertaken in [4, 37].

Finally, we refer the reader to the long, informative articles by Sznitman [36] and by Méléard [24] in Springer-Verlag's *Lecture Notes in Mathematics*.

The aforementioned authors are mostly concerned with proving that specific systems propagate chaos, rather than the propagation of chaos *per se*. The modest purpose of this dissertation is to clarify the definition of propagation of chaos in general, and not to prove that any particular system propagates chaos. The essential content of this dissertation is Definition 4.1 and Theorem 4.2 of Chapter 4.

This dissertation is organized as follows.

The rest of Chapter 1 is an informal summary of our point of view and contains a statement of our main theorem. We introduce general Markovian interacting particle systems and adopt a strong-sense definition of the propagation of chaos. We can then characterize the propagation of chaos in terms of the Markov transition functions that define the interacting particle systems.

Chapter 2 describes the most important instances of the propagation of chaos. The concept of propagation of chaos is most useful in (and was indeed motivated by) the kinetic theories of gases, plasmas, and stellar systems. Boltzmann's equation for dilute gases is discussed in Section 2.1 and Vlasov's equation for plasmas and stellar systems is discussed in Section 2.2.

Chapters 3 and 4 are meant to be self-contained, formal, and brisk. They contain the necessary background and the proofs of the theorems that flow from the point of view described in Section 1.2.

Chapter 3 discusses the theorem of Sznitman and Tanaka, which is our main technical tool. A detailed proof of this theorem is given in Section 3.2. Theorem 3.5 states that, if the underlying space is finite, p-chaos entails the convergence of specific entropy to the entropy of p.

Chapter 4 is dedicated to the proof of our main theorem and its corollaries. Although the theorems there are neither deep nor surprising, they should still be of interest to those who work with the propagation of chaos because they establish properties of the propagation of chaos which, though easy to take for granted, do require some proof. The proofs rely on the theorem of Sznitman and Tanaka and basic properties of convergence in law. Our main theorem requires the completeness of the basic space so that we may invoke Prohorov's theorem in its proof.

This work was supported by the U.S. Department of Energy at the Lawrence Berkeley National Laboratory, at the Mathematics Department, under Alexandre Chorin.

1.2 Definition of Propagation of Chaos and Statement of Main Result

Statistical mechanics and kinetic theory are probabilistic theories of many-body systems; their predictions are intended to be valid only when the number of particles is very large, typically as large as Avogadro's number. The equations of kinetic theory are obtained by studying the limiting behavior of n-particle systems as n tends to infinity. A key concept in such studies is the propagation of chaos.

The concept of propagation of chaos was motivated originally by the kinetic theories of gases and plasmas. Before we delve into these kinetic theories, we first set up, in this section, a very general framework for the study of interacting particle systems and the propagation of chaos.

Our particles shall live in a space S. For gases and plasmas S would be position-velocity space, a subset of $\mathbb{R}^6$, but for our general purposes S may be any separable metric space. The state of an n-particle system is a point in S^n, the n-fold Cartesian product of S with itself, also a metric space. Whether the dynamics of the n-particle system are deterministic or random, we desire that the future motion of a system of particles depend only on its current state, and not the entire history of the particles' motion. We stipulate that the n-particle dynamics are Markovian; the future depends on the past only through the present state. Markovian processes are defined by their transition functions, so our data includes one transition function $K_n(\mathbf{s}, B, t)$ for each n. That is, for each n, we are given the transition function

$$K_n : S^n \times \mathcal{B}_{S^n} \times [0, \infty) \longrightarrow [0, 1],$$

where $\mathcal{B}_{S^n}$ is the Borel σ-algebra on S^n. The transition functions K_n have the following interpretation. For $t \geq 0$, $\mathbf{s} \in S^n$, and $B \in \mathcal{B}_{S^n}$, the probability that the state at time t of an n-particle system belongs to B, given that the state was initially $\mathbf{s}$, is $K_n(\mathbf{s}, B, t)$. The Markov property implies that the transition functions satisfy the Chapman-Kolmogorov equations:

$$K_n(\mathbf{s}, B, t + t') = \int_{S^n} K(\mathbf{s}, d\mathbf{s}', t') K(\mathbf{s}', B, t)$$

for all $t, t' \geq 0, \mathbf{s} \in S^n$, and $B \in \mathcal{B}_{S^n}$.

We restrict our consideration to n-particle systems whose dynamics commute with permutations by imposing the following conditions on the transition functions. Let Π_n denote the set of permutations of $\{1, 2, \ldots, n\}$. If $\pi \in \Pi_n$ and $\mathbf{s} = (s_1, s_2, \ldots, s_n) \in S^n$, let

$$\pi \cdot \mathbf{s} = (s_{\pi(1)}, s_{\pi(2)}, \ldots, s_{\pi(n)}),$$

and, for $B \subset S^n$, let

$$\pi \cdot B = \{\pi \cdot \mathbf{s} : \mathbf{s} \in B\}.$$

We suppose that the transition functions satisfy

$$K_n(\mathbf{s}, B, t) = K_n(\pi \cdot \mathbf{s}, \pi \cdot B, t) \tag{1.1}$$

for all permutations π, points $\mathbf{s}$, Borel sets B, and times t.

Here, then, is the set-up. A separable metric space S is given, along with a sequence

$$\{K_n(\mathbf{s}, B, t)\}_{n=1}^{\infty}$$

of Markov transition functions that satisfy the permutation condition (1.1), K_n being a transition function on S^n. We have a Markovian dynamics of n particles, an n-particle system, for each n. Propagation of chaos, defined shortly, is an attribute of *families* of particle systems, indexed by n; a family of n-particle systems either does or does not propagate chaos.

In order to define the propagation of chaos we must first define the property of being chaotic — "chaos" for short. For each n, let ρ_n be a symmetric probability measure on S^n, i.e., a probability measure on S^n such that

$$\rho_n(\pi \cdot B) = \rho_n(B)$$

for all permutations π and all $B \in \mathcal{B}_{S^n}$. Let ρ be a probability measure on S.

Definition: *The sequence $\{\rho_n\}$ is ρ-**chaotic** if, for any natural number k and any bounded continuous functions $g_1(s), g_2(s), \ldots, g_k(s)$ on S,*

$$\lim_{n\to\infty} \int_{S^n} g_1(s_1)g_2(s_2)\cdots g_k(s_k)\rho_n(ds_1 ds_2 \cdots ds_n) = \prod_{i=1}^{k} \int_S g_i(s)\rho(ds).$$

In words, a sequence probability measures on the product spaces S^n is ρ-chaotic if, for fixed k, the joint probability measures for the first k coordinates tend to the product measure $\rho(ds_1)\rho(ds_2)\cdots\rho(ds_k) \equiv \rho^{\otimes k}$ on S^k. If the measures ρ_n are thought of as giving the joint distribution of n particles residing in the space S, then $\{\rho_n\}$ is ρ-chaotic if k particles out of n become more and more independent as n tends to infinity, and each particle's distribution tends to ρ. A sequence of symmetric probability measures on S^n is *chaotic* if it is ρ-chaotic for some probability measure ρ on S.

If a Markov process on S^n begins in a random state with distribution ρ_n, the distribution of the state after t seconds of Markovian random motion can be expressed in terms of the transition function K_n for the Markov process. The distribution at time t is the probability measure $U_t^n \rho_n$ defined by

$$U_t^n \rho_n(B) := \int_{S^n} K_n(\mathbf{s}, B, t)\rho_n(d\mathbf{s}) \tag{1.2}$$

for all $B \in \mathcal{B}_{S^n}$. If K_n satisfies the permutation condition (1.1) then $U_t^n \rho_n$ is symmetric whenever ρ_n is.

Definition: *A sequence*

$$\{K_n(\mathbf{s}, B, t)\}_{n=1}^{\infty}$$

whose n^{th} term is a Markov transition function on S^n that satisfies the permutation condition (1.1) **propagates chaos** *if, whenever $\{\rho_n\}$ is chaotic so is $\{U_t^n \rho_n\}$ for any $t \geq 0$, where U_t^n is as defined in (1.2).*

We sometimes say that a family of n-particle Markov *processes* propagates chaos when we really mean that the associated family of transition functions propagates chaos.

It follows from the definition of propagation of chaos that for each $t > 0$ there exists an operator U_t^∞ on probability measures such that $\{U_t^n \rho_n\}$ is $U_t^\infty \rho$-chaotic if $\{\rho_n\}$ is ρ-chaotic. This operator is typically nonlinear; even though U_t^∞ is derived from the linear operators U_t^n by taking a limit of sorts, it is not actually a limit of linear operators, and may be nonlinear. For families of interacting particle systems suited to the study of gases or plasmas, the semigroup $\{U_t^\infty\}_{t \geq 0}$ is the semigroup of solution operators for the Boltzmann or the Vlasov equation. (The existence of the operators U_t^∞ is part of our main theorem, stated shortly.)

We are adopting here a strong definition of the propagation of chaos. Other authors [24, p. 42][29, p. 98] have defined propagation of chaos in a weaker sense: a family of Markovian n-particle processes propagates chaos if $\{U_t^n \rho^{\otimes n}\}$ is chaotic for all $\rho \in \mathcal{P}(S)$ and $t > 0$, where $\rho^{\otimes n}$ is product measure on S^n. For these authors, only *purely* chaotic sequences of initial measures are required to "propagate" to chaotic sequences. This condition is strictly weaker than the one we adopt for our definition. For example, take $S = \{0, 1\}$ and let $\delta(x)$ or δ_x denote a point mass at x. Then, if

$$K_n(\mathbf{s}, \cdot, t) = \begin{cases} \delta_{(1,1,\ldots,1)} & \text{if } \mathbf{s} \neq (0, 0, \ldots, 0) \\ \delta_{(0,0,\ldots,0)} & \text{if } \mathbf{s} = (0, 0, \ldots, 0) \end{cases}$$

for all $t > 0$, the sequence $\{K_n\}$ propagates chaos in the weak sense, but not in the strong sense of our definition. Under these K_n's, the $\delta(0)$-chaotic sequence $\{\delta_{(0,0,\ldots,0)}\}$ is propagated to itself, while other $\delta(0)$-chaotic sequences are propagated to $\delta(1)$-chaotic sequences, and yet other $\delta(0)$-chaotic sequences are not propagated to chaotic sequences at all.

Our main result is a condition on the Markov transition functions for a family of n-particle processes that is necessary and sufficient for the propagation of chaos (in the strong sense). Before we state it we must recall the weak topology on probability measures and introduce some necessary notation.

If X is a completely regular topological space (as normal topological spaces are), let $\mathcal{P}(X)$ denote the space of probability measures on X endowed with the weakest topology relative to which all the functions $I_g : \mathcal{P}(X) \longrightarrow \mathbb{R}$ are continuous, where

$$I_g(\mu) = \int_X g(x)\mu(ds)$$

and g ranges over the bounded and continuous real-valued functions on X. A sequence $\{\mu_n\}$ in $\mathcal{P}(X)$ converges to μ in this weak topology if

$$\int_X g(x)\mu_n(ds) \longrightarrow \int_X g(x)\mu(dx)$$

for all $g \in C_b(X)$, the space of bounded and continuous real-valued functions on X.

For ν a measure on S^n, let $\widetilde{\nu}$ denote its symmetrization: for all $B \in \mathcal{B}_{S^n}$,

$$\widetilde{\nu}(B) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \nu(\pi \cdot B).$$

For fixed $\mathbf{s} \in S^n$ and $t \geq 0$, denote by $\widetilde{K}_n(\mathbf{s}, \cdot, t)$ the symmetrization of the measure $K_n(\mathbf{s}, \cdot, t)$.

For

$$\mathbf{s} = (s_1, s_2, \ldots, s_n) \in S^n,$$

let $\varepsilon_n(\mathbf{s})$ denote the purely atomic probability measure

$$\varepsilon_n(\mathbf{s}) = \frac{1}{n} \sum_{i=1}^{n} \delta(s_i) \quad , \tag{1.3}$$

where $\delta(s)$ — Dirac's delta — denotes a point-mass at s. The function ε_n takes ordered n-tuples to purely atomic probability measures consisting of n point-masses of weight $\frac{1}{n}$ each.

We can now state our main theorem.

Main Theorem: *Let S be a complete, separable metric space. Let the Markov transitions K_n satisfy the permutation condition (1.1).*

Propagation of chaos by the sequence $\{K_n\}$ is equivalent to the following condition:

For every $t > 0$ there exists a continuous map

$$U_t^\infty : \mathcal{P}(S) \longrightarrow \mathcal{P}(S)$$

such that, if the sequence

$$\mathbf{s}_1 \in S, \quad \mathbf{s}_2 \in S^2, \quad \mathbf{s}_3 \in S^3, \ldots$$

is such that $\{\varepsilon_n(\mathbf{s}_n)\}_{n=1}^\infty$ converges to ρ in $\mathcal{P}(S)$, then the sequence of symmetric measures

$$\left\{\widetilde{K}_n(\mathbf{s}_n, \cdot, t)\right\}_{n=1}^\infty$$

is $U_t^\infty\rho$-chaotic.

The necessity of the condition of the preceding theorem is an easy consequence of the definition; its sufficiency is nontrivial. Our theorem shows that to prove propagation of chaos it is *sufficient* to verify that $\{U_t^n \rho_n\}$ is chaotic when the initial measures $\rho_n \in \mathcal{P}(S^n)$ are symmetric atomic measures of the form

$$\rho_n = \frac{1}{n!} \sum_{\pi \in \Pi_n} \delta(\pi \cdot \mathbf{s}_n); \quad \mathbf{s}_n \in S^n. \tag{1.4}$$

This sufficient condition can come in handy. In 1977, Braun and Hepp [5] proved the propagation of chaos for Vlasov's equation, provided the initial conditions are "pure initial states" of the form (1.4). Sznitman [35], in 1983, noted that Braun and Hepp require "purely atomic initial data" to propagate their chaos, implicitly suggesting that this restriction to special initial conditions weakens their result. Our theorem shows that it did indeed suffice for Braun and Hepp to verify propagation of chaos for purely atomic initial data.

The theorem is proved by expressing chaos in terms of weak convergence of probability measures in $\mathcal{P}(\mathcal{P}(S))$ and then applying Prohorov's theorem. Prohorov's theorem [1] states that a family $\mathcal{F}$ of probability measures on a complete and separable metric space is relatively compact if *and only if* $\mathcal{F}$ is tight. Our hypothesis that S is complete ensures that $\mathcal{P}(S)$ is also complete and enables us to apply Prohorov's theorem in $\mathcal{P}(\mathcal{P}(S))$. Chapter 4 is devoted to the proof of the theorem.

The study of chaos via weak convergence in the space $\mathcal{P}(\mathcal{P}(S))$ is due to Sznitman [34] and Tanaka [38]. They proved that a sequence of symmetric measures $\{\rho_n\}$ is ρ-chaotic if and only if the probability measures induced on $\mathcal{P}(S)$ by ε_n converge in $\mathcal{P}(\mathcal{P}(S))$ to $\delta(\rho)$, a point mass at $\rho \in \mathcal{P}(S)$. This device is essential to our approach and is discussed at length in Chapter 3.

Chapter 2

Kinetic Theory and the Propagation of Chaos

Boltzmann's equation for dilute gases and Vlasov's equation for plasmas govern the evolution, the change over time, of the density of particles in position-momentum space. The particle density changes due to interactions between the particles: binary collisions of molecules in a dilute gas or mutual electric forces acting between ions in a plasma. The rate of change of the particle density is determined by the particle density *itself* through the particle interactions. The evolution equations of Boltzmann and Vlasov are nonlinear because of the way the particle density affects its own evolution.

This chapter reviews the equations of Boltzmann and Vlasov for the sake of illuminating the meaning and physical relevance of the propagation of chaos. One may consult [33] for a more thorough treatment of kinetic theory.

Section 2.1 presents the theory of dilute gases from the point of view of the propagation of chaos. First, the classic derivation of Boltzmann's equation is repeated in 2.1.1. Then, in 2.1.2, two types of n-particle systems are introduced that satisfy Boltzmann's equation in the infinite particle limit.

Section 2.2 is about Vlasov's equation for plasmas and stellar systems. Vlasov's equation is introduced in 2.2.1 and rederived in terms of the propagation of chaos in 2.2.2.

2.1 Dilute Gases

2.1.1 Boltzmann's equation

In this section we summarize Boltzmann's derivation of his equation for a dilute gas. Our source is the first chapter of his *Lectures on Gas Theory* [3], written over a century ago.

Boltzmann modeled the molecules of the gas by hard spheres: balls of radius r that collide elastically according to simple mechanics. When a ball having velocity $\mathbf{v}$ collides with a ball having velocity $\mathbf{w}$, the collision instantaneously changes the velocity of the first ball from $\mathbf{v}$ to $\mathbf{v}'$ and the velocity of the second ball from $\mathbf{w}$ to $\mathbf{w}'$. Given the relative orientation of the balls at the time of impact, the, post-collisional or outgoing velocities are determined by the laws of conservation of energy and momentum. Suppose that, at the moment of impact, $\mathbf{l}$ is the unit vector parallel to the ray that originates at the center of the ball of velocity $\mathbf{v}$ and passes through the center of the ball of velocity $\mathbf{w}$. Such a collision, which we call a $(\mathbf{v}, \mathbf{w} : \mathbf{l})$ collision, changes the velocities of the balls to

$$\begin{aligned} \mathbf{v} \quad &\longrightarrow \quad \mathbf{v}' = \mathbf{v} + [(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l}]\mathbf{l} \\ \mathbf{w} \quad &\longrightarrow \quad \mathbf{w}' = \mathbf{w} - [(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l}]\mathbf{l}. \end{aligned} \tag{2.1}$$

A collision of type $(\mathbf{v}, \mathbf{w} : \mathbf{l})$ is only possible if $(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l} < 0$. Except during collisions, which have instantaneous duration, molecules (hard spheres in this model) travel inertially, with unchanging velocity. Let n denote the number of molecules in the gas, and let the number of molecules per unit volume of position-momentum space be given by the density

$$f(\mathbf{x}, \mathbf{v}, t) d\mathbf{x} d\mathbf{v},$$

so that the proportion of molecules which, at time t, are located in a region X of space and have velocities belonging to a set V of velocities is

$$\frac{1}{n} \int_V \int_X f(\mathbf{x}, \mathbf{v}, t) d\mathbf{x} d\mathbf{v}.$$

Boltzmann's equation tells how $f(\mathbf{x}, \mathbf{v}, t)$ changes due to the collisions detailed above.

The density $f(\mathbf{x}, \mathbf{v}, t) d\mathbf{x} d\mathbf{v}$ of molecules changes through the inertial motion of the molecules between collisions (called *free streaming*) and through collisions between molecules. Boltzmann's equation can be written

$$\frac{\partial}{\partial t} f(\mathbf{x}, \mathbf{v}, t) = -\mathbf{v} \cdot \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{v}, t) + Q[f(\mathbf{x}, \mathbf{v}, t)],$$

where $-\mathbf{v} \cdot \nabla_{\mathbf{x}} f$ gives the rate of change of f due to free streaming, and $Q[f]$, the collision operator applied to f, gives the rate of change of the density due to collisions.

Further assumptions are needed to determine $Q[f]$, the rate of change of $f(\mathbf{x}, \mathbf{v}, t)$ due to collisions. We know the effect of a $(\mathbf{v}, \mathbf{w} : \mathbf{l})$ collision, but we also need to know the rate at which those collisions are occurring. Boltzmann assumed that the rate at which $(\mathbf{v}, \mathbf{w} : \mathbf{l})$ collisions are happening at a point $\mathbf{x}$ of space is proportional to $r^2 \|(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l}\|$ and jointly proportional to the densities at $\mathbf{x}$ of molecules having velocities $\mathbf{v}$ and $\mathbf{w}$. These assumptions are the *Stosszahlansatz*, or collision-number-hypothesis: the rate of $(\mathbf{v}, \mathbf{w} : \mathbf{l})$ collisions at $\mathbf{x}$ is

$$r^2 \|(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l}\| f(\mathbf{x}, \mathbf{v}, t) f(\mathbf{x}, \mathbf{w}, t). \tag{2.2}$$

The rate of change of $f(\mathbf{x}, \mathbf{v}, t)$ due to collisions, $Q[f]$, equals the rate at which the molecules are receiving post-collisional velocities $\mathbf{v}$ less the rate at which molecules already having velocity $\mathbf{v}$ are colliding with other molecules and exchanging $\mathbf{v}$ for other velocities. The loss rate is easy to express, assuming the Stosszahlansatz:

$$L[f] := \frac{r^2}{2} \int_{\mathbb{R}^3} \int_{S_2} f(\mathbf{x}, \mathbf{v}, t) f(\mathbf{x}, \mathbf{w}, t) \|(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l}\| d\mathbf{l} d\mathbf{w}, \tag{2.3}$$

where S_2 is the unit sphere in $\mathbb{R}^3$ and $d\mathbf{l}$ indicates the normalized and uniform measure on the sphere S_2, is the number of molecules per unit volume at $\mathbf{x}$ of velocity $\mathbf{v}$ that will collide with other molecules between times t and $t + \Delta t$, divided by Δt.

There is a similar expression for the gain rate at which collisions are *resulting* in molecules having velocity $\mathbf{v}$. Observe that a binary collision can only produce a post-collisional, or outgoing, velocity $\mathbf{v}$ if the velocities before collision were $\mathbf{v} + ((\mathbf{w} - \mathbf{v}) \cdot \mathbf{l})\mathbf{l}$ and $\mathbf{w} - ((\mathbf{w} - \mathbf{v}) \cdot \mathbf{l})\mathbf{l}$ for some $\mathbf{w}$. Let

$$\begin{aligned} \mathbf{v}^* &= \mathbf{v} + ((\mathbf{w} - \mathbf{v}) \cdot \mathbf{l})\mathbf{l} \\ \mathbf{w}^* &= \mathbf{w} - ((\mathbf{w} - \mathbf{v}) \cdot \mathbf{l})\mathbf{l}. \end{aligned}$$

The number of molecules per unit volume that will end up having velocity $\mathbf{v}$ because of a collision that took place between times t and $t + \Delta t$, divided by Δt, equals

$$G[f] := \frac{r^2}{2} \int_{\mathbb{R}^3} \int_{S_2} f(\mathbf{x}, \mathbf{v}^*, t) f(\mathbf{x}, \mathbf{w}^*, t) \|(\mathbf{w} - \mathbf{v}) \cdot \mathbf{l}\| d\mathbf{l} d\mathbf{w}. \tag{2.4}$$

In fact, $\mathbf{v}^* = \mathbf{v}'$ and $\mathbf{w}^* = \mathbf{w}'$; if a $(\mathbf{v}, \mathbf{w} : \mathbf{l})$ collision changes $\mathbf{v}$ to $\mathbf{v}'$ and $\mathbf{w}$ to $\mathbf{w}'$, then a $(\mathbf{v}', \mathbf{w}' : \mathbf{l})$ collision changes $\mathbf{v}'$ to $\mathbf{v}$ and $\mathbf{w}'$ to $\mathbf{w}$. It is only a lucky accident that $\mathbf{v}^* = \mathbf{v}'$, so we emphasize, by introducing new notation, that $\mathbf{v}^*$ and $\mathbf{w}^*$ are supposed to be velocities for which a $(\mathbf{v}^*, \mathbf{w}^* : \mathbf{l})$ collision *results in* a velocity $\mathbf{v}$.

The net rate of change of $f(\mathbf{x}, \mathbf{v}, t)$ due to collisions equals the gain rate minus the loss rate: $Q[f] = G[f] - L[f]$. Boltzmann's equation is thus

$$\frac{\partial}{\partial t} f(\mathbf{x}, \mathbf{v}, t) + \mathbf{v} \cdot \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{v}, t) = G[f(\mathbf{x}, \mathbf{v}, t)] - L[f(\mathbf{x}, \mathbf{v}, t)], \tag{2.5}$$

where $G[f]$ and $L[f]$ are as defined in (2.4) and (2.3).

The existence of solutions of Boltzmann's equation (2.5) is difficult to prove. The state of the art is the global existence of mild solutions proved by Di Perna and Lions [9].

2.1.2 Particle systems for Boltzmann's equation

Kac [17], in his article *Foundations of Kinetic Theory* of 1954, propounds the relationship between Boltzmann's equation and certain n-particle Markovian jump processes. These n-particle systems are inherently stochastic; the collisions have random results and happen at random times. The dynamics are not the true dynamics of deterministically colliding molecules, rather, the stochastic motion of fictitious particles which obey the spatially homogeneous Boltzmann equation on the macroscopic level.

The spatially homogeneous Boltzmann equation is the equation satisfied by a position-velocity density that does not depend on position: $f(v)dv$. So Kac imagines a gas of n particles on the line, particles whose positions are unimportant and are not given, but whose velocities

$$v_1,\ v_2, \dots,\ v_n; \qquad v_i \in \mathbb{R} \tag{2.6}$$

completely specify the state of the gas. Kac proposes a stochastic dynamics of these states driven by collisions between pairs of particles. Suppose that the state is initially given by the list (2.6). At a random time, a collision occurs. A collision changes the values of a random pair of the n velocities in the list, at random. Once the state of the gas has jumped to a new state due to a collision, another random time elapses, another collision occurs, and so forth. The random times are taken to be independent and to have exponential distributions with mean duration τ/n; the probability that a collision happens later than t seconds after the previous collision is $e^{-nt/\tau}$. Notice that the more particles there are, the faster collisions are occurring. Each collision only affects the velocities of two particles, the affected pair being selected at random from one of the $n(n-1)/2$ possible pairs of particles. Given that a pair of particles having velocities v and w collide, those two velocities change to another pair v' and w' satisfying the conservation of energy condition

$$(v')^2 + (w')^2 = v^2 + w^2,$$

but otherwise at random, so that (v', w') is randomly sampled from the uniform probability measure on the circle

$$\left\{(v', w') : (v')^2 + (w')^2 = v^2 + w^2\right\}.$$

Kac's n particle gas is thus a Markov jump process on $\mathbb{R}^n$, for each n.

In [16, 17], Kac proves that this family of n-particle gases propagates chaos. Indeed, the exact definition of chaos as the asymptotic independence of particles is due to Kac. The notion of chaos originates in Boltzmann [3], who derived his equation under a hypothesis of "molecular disorder (chaos)."

Kac proved that if the particles of each n-particle gas initially have independent and $f_0(v)dv$-distributed velocities, then at a later time t the velocities of a random pair become increasingly independent as $n \longrightarrow \infty$, even though the initial condition of *pure* independence or "molecular chaos" has been spoiled by collisions. The random velocity of a single particle at time t becomes increasingly $f(v,t)$-distributed as $n \longrightarrow \infty$, where $f(v,t)$ satisfies an analog of Boltzmann's equation, namely,

$$\begin{aligned} \frac{\partial}{\partial t} f(v,t) &= \frac{2}{\tau} \int_{\mathbb{R}} \int_0^{2\pi} f(v\cos\theta - w\sin\theta) f(v\sin\theta + w\cos\theta) d\theta dw - f(v) \\ f(v,0) &= f_0(v). \end{aligned}$$

Indeed, the sequence of n-particle joint distributions at time t is $f(v,t)dv$-chaotic.

Similar procedures yield particle systems for the spatially homogeneous Boltzmann equation [17]. The spatially homogeneous Boltzmann equation for hard spheres of radius r is

$$\begin{aligned} \frac{\partial}{\partial t} f(\mathbf{v},t) &= G[f(\mathbf{v},t)] - L[f(\mathbf{v},t)] \\ G[f] &= \frac{r^2}{2} \int_{\mathbb{R}^3} \int_{S_2} f(\mathbf{v}^*,t) f(\mathbf{w}^*,t) \|(\mathbf{w}-\mathbf{v}) \cdot \mathbf{l}\| d\mathbf{l} d\mathbf{w} \\ L[f] &= \frac{r^2}{2} \int_{\mathbb{R}^3} \int_{S_2} f(\mathbf{v},t) f(\mathbf{w},t) \|(\mathbf{w}-\mathbf{v}) \cdot \mathbf{l}\| d\mathbf{l} d\mathbf{w}, \end{aligned} \tag{2.7}$$

where $d\mathbf{l}$ is normalized surface area on the sphere S_2, and

$$\begin{aligned} \mathbf{v}^* &= \mathbf{v} + ((\mathbf{w}-\mathbf{v}) \cdot \mathbf{l})\mathbf{l} \\ \mathbf{w}^* &= \mathbf{w} - ((\mathbf{w}-\mathbf{v}) \cdot \mathbf{l})\mathbf{l}. \end{aligned} \tag{2.8}$$

One may devise several n-particle jump processes for the Boltzmann equation. Grünbaum [15] suggests one with a three-stage random mechanism for making jumps: given that the initial state of the gas or the state it has just jumped to is

$$(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n); \qquad \mathbf{v}_i \in \mathbb{R}^3,$$

1) Select two distinct particles at random (equiprobably), say the i^{th} and j^{th} particles where $i < j$.

2) If $\mathbf{v}_i = \mathbf{v}_j$ select another pair. Otherwise wait for an exponentially distributed random time of mean duration $\|\mathbf{v}_i - \mathbf{v}_j\|/(n-1)$.

3) Jump to $(\mathbf{v}_1, \ldots, \mathbf{v}^*{}_i, \ldots, \mathbf{v}^*{}_j, \ldots, \mathbf{v}_n)$ with probability proportional to $\frac{\|(\mathbf{v}_i - \mathbf{v}_j) \cdot \mathbf{l}\|}{\|\mathbf{v}_i - \mathbf{v}_j\|}$, where $\mathbf{v}^*{}_i, \mathbf{v}^*{}_j$, and $\mathbf{l}$ are as in (2.8).

Note that the jumps speed up as the number of particles increases so that the number of jumps per particle per unit time is roughly constant. Grünbaum proves that this family of n-particle processes propagates chaos and that the limit satisfies (2.7) under certain assumptions [15]. His proof relies on the theory of strongly continuous contraction semigroups.

Other jump processes similar to those of Kac have been treated by several authors. Uchiyama [40] proves propagation of chaos and a central limit theorem for families of Kac-type processes, on *countable* sets of velocities. Rezakhanlou and Tarver [31] prove an interesting propagation of chaos result for the discrete Boltzmann equation in one dimension. Their particles travel with constant velocities around a circle in between random collisions that become increasingly local as the number of particles increases. Graham and Méléard [24] prove the propagation of chaos for a variant of the Boltzmann equation with nonlocal collisions. Particles experience random Kac-type collisions, but do not need to be at the same spatial location in order to collide. Bird's numerical scheme for Boltzmann's equation [2] amounts to the simulation of one of the processes studied by Graham and Méléard.

The jump processes of Kac et alia are intrinsically stochastic, for collisions happen at random and have random results. On the other hand, the dynamics of real molecules are strictly deterministic, or are classically conceived to be such. Our idealized model for molecular dynamics, the hard sphere model, admits no randomness at all. When two particles collide, their outgoing velocities are determined by their incoming velocities and their attitude at collision. (It is true that the outcome of a simultaneous collision of three or more spheres may not be determined, but in a dilute enough hard sphere gas these collisions are so rare that they have negligible effect.) Since Boltzmann's equation is supposed to govern the macroscopic behavior of the density of a hard sphere gas, it ought to be derivable somehow from the deterministic dynamics of hard spheres. But alas, it would appear that the Boltzmann equation is not even *consistent* with molecular dynamics, much less derivable from it, for the molecular dynamics are reversible and Boltzmann's equation is irreversible. This apparent antinomy, known as Loschmidt's paradox, has been raising deep concerns about the validity of Boltzmann's equation for nearly as long as that equation has been known. It is therefore surprising and philosophically significant that (notwithstanding Loschmidt's paradox) Boltzmann dynamics can indeed be derived from molecular dynamics.

Grad [13] first advanced the idea that Boltzmann's equation may be derived in the *dilute limit*

$$nr^2 \longrightarrow \text{constant}$$

of hard sphere dynamics, and Lanford [18] succeeded in a rigorous derivation of Boltzmann's equation along the lines suggested by Grad. Lanford's theorem can be neatly expressed in terms of of chaos. This approach can be found in *The Mathematical Theory of Dilute Gases* by Cercignani, Illner, and Pulvirenti [7](pp. 90-93), who emphasize that the theorem of Lanford constitutes a *validation* of

Boltzmann's equation from the fundamental principles of molecular dynamics.

Let us describe Lanford's result.

Consider the deterministic dynamics of n hard spheres of radius $\frac{1}{\sqrt{n}}$. The phase space is formed by excising the points of $(\mathbb{R}^6)^n$ that represent configurations in which two or more spheres would overlap. The set of all initial configurations that lead to simultaneous collisions of three or more particles or to infinitely many collisions in finite time has measure zero and can be ignored. The trajectories through phase space are determined by the free motion of the spheres between collisions and the rule (2.1) for binary collisions. When a trajectory hits a boundary point of the phase space, i.e., when a collision occurs, the trajectory continues from the unique boundary point that the rule of elastic collision associates to it. This defines the deterministic dynamics of a dilute gas of n hard spheres of radius $\frac{1}{\sqrt{n}}$. Increasing n increases the number of particles but decreases the density, whence the term "dilute limit."

Lanford's theorem states (roughly) that there exists $\tau > 0$ on the order of the mean free time such that, if the initial n-particle densities are f_0-chaotic *in a very strong sense* , then the densities at a later time $t \leq \tau$ are f_t chaotic, where f_t is a mild solution of the Boltzmann equation with initial data f_0. The hypotheses on the initial data are that the k-marginals of the symmetric n-particle distributions are absolutely continuous with continuous densities, and those densities satisfy a growth bound depending on k and *converge uniformly on compact sets to* $f_0^{\otimes k}$ in the dilute limit $n \longrightarrow \infty$. This hypothesis on the initial data is stronger than mere chaos, and Lanford's theorem asserts that such strong initial chaos is propagated. The n-particle densities at a later time will be chaotic, says Lanford's theorem, but typically not chaotic in the same strong sense as were the initial densities. This "loss of convergence quality" is what permits the Boltzmann equation to be irreversible even though it is derived from reversible dynamics [7, p. 97].

Lanford's theorem says that chaos is propagated, but only if the initial densities converge uniformly on compact sets, et cetera. This kind of propagation of chaos differs from propagation of chaos as defined in this dissertation; it has to do with subtler properties of uniform and pointwise convergence of densities rather than simple weak convergence of distributions. We remark that the hard sphere gases of Grad and Lanford do *not* propagate chaos in our sense, nor do they satisfy the conclusions of our theorems about families of Markov processes that propagate chaos.

2.2 Plasmas and Stellar Systems

This section contains an account of the propagation of chaos for the Vlasov equation.

2.2.1 Vlasov's equation

Vlasov's equation [5, 33] is another important equation of kinetic theory. It governs the density in position-velocity space of particles that interact (without colliding) through long-range forces such as the electric forces between ions in a plasma or the gravitational attraction between stars in a galaxy.

Suppose, for simplicity, that all particles in the system are of the same species, each having mass m, and let $F(\mathbf{x})$ denote the force that a particle at the origin would exert on a particle at $\mathbf{x}$. For example, the force $F(\mathbf{x})$ is proportional to $\mathbf{x}/\|\mathbf{x}\|^3$ if the particles are electrons, and proportional to $-m^2\mathbf{x}/\|\mathbf{x}\|^3$ if the particles are stars. If $f(\mathbf{x},\mathbf{v},t)d\mathbf{x}d\mathbf{v}$ denotes the number of particles per unit volume near $(\mathbf{x},\mathbf{v})$ at time t, we find that the net force on a particle at $\mathbf{x}$ is

$$F_f(\mathbf{x}) := \int_{\mathbb{R}^3} F(\mathbf{x}-\mathbf{x}')f(\mathbf{x}',\mathbf{v}',t)d\mathbf{x}'d\mathbf{v}'. \tag{2.9}$$

The particle density $f(\mathbf{x},\mathbf{v},t)$ changes through the motion of particles subject to the force field $F_f(\mathbf{x})$.

Vlasov's equation for the density is

$$\frac{\partial}{\partial t}f(\mathbf{x},\mathbf{v},t) = -\mathbf{v}\cdot\nabla_{\mathbf{x}}f(\mathbf{x},\mathbf{v},t) - \frac{1}{m}F_f(\mathbf{x})\cdot\nabla_{\mathbf{v}}f(\mathbf{x},\mathbf{v},t), \tag{2.10}$$

where the net force field $F_f(\mathbf{x})$, defined in equation (2.9), depends on the particle density f itself. This is just an advection equation for the flow on $(\mathbf{x},\mathbf{v})$-space given by the time dependent flux $(\mathbf{v},F(\mathbf{x}))$, with the requirement that F equals F_f, i.e., the flux function at time t is determined through (2.9) by the solution at time t of the advection equation itself.

The preceding is a heuristic derivation of Vlasov's equation from the smoothed dynamics of a large but fixed number of particles. Vlasov's equation may be derived rigorously from the true dynamics of interacting particle systems, in the limit of infinite particle number. This rigorous derivation is the content of the theorems, stated in the next section, on the propagation of chaos for Vlasov and McKean-Vlasov equations.

Propagation of chaos clarifies the relationship between the Vlasov equation and the dynamics of gravitational systems and plasmas.

Imagine n particles of mass $\frac{1}{n}$ following the classical n-body evolution. As the number of particles tends to infinity and the initial distribution of particles approaches a distribution $f(\mathbf{x},\mathbf{v},0)d\mathbf{x}d\mathbf{v}$ of mass, Vlasov's equation is an increasingly correct description of the evolution of the mass density. The mass density follows equation (2.10) with m set to 1 and with $F(\mathbf{x})$ redefined as the gravitational force on a test particle of *unit* mass at $\mathbf{x}$ due to a particle of *unit* mass at the origin.

It is a little tougher to obtain a macroscopic equation for the density of charge in the limit of infinitely many ions. A k-fold increase of the number of electrons (say) in a plasma increases the forces by a factor of k^2 and the system becomes too energetic in the limit $n \longrightarrow \infty$. Vlasov dynamics can only result from proper scaling of mass and/or time. One possibility is to imagine n electrons of mass n each. In the limit $n \longrightarrow \infty$, the density of charge in position-velocity space satisfies equation (2.10), *mutatis mutandis*. An alternative scaling is found in [5]: Consider the dynamics of n ions of mass $\frac{1}{n}$ each. As n-tends to infinity, and time is slowed as $\frac{1}{n}$, one obtains a Vlasov equation for the density in position-*momentum* space.

2.2.2 McKean-Vlasov particle systems

We begin with a statement of the propagation of chaos for Vlasov's equation. It is assumed that the interparticle force is bounded and globally Lipschitz, an assumption that excludes the physical inverse-square forces of gravitational systems and plasmas. One way around this difficulty is to assume that the system is so dilute that particles never get too close to one another. The interparticle force could then be replaced with one without the singularity at zero distance that is still inversely proportional to the square of the distance between particles when that distance is not too small.

Let $F : \mathbb{R}^6 \longrightarrow \mathbb{R}^6$ be bounded and Lipschitz. For each n, define a deterministic n-particle process in $\mathbb{R}^6$ by the following system of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{d}{dt}\mathbf{x}_i^n(t) &= \mathbf{v}_i^n(t) \\ \frac{d}{dt}\mathbf{v}_i^n(t) &= \frac{1}{n}\sum_{j=1}^{n} F(\mathbf{x}_i^n - \mathbf{x}_j^n) \end{aligned} \tag{2.11}$$

for $i = 1, 2, \ldots, n$. Braun and Hepp [5] prove that if the initial conditions

$$\mathbf{x}_1^n(0), \mathbf{v}_1^n(0), \mathbf{x}_2^n(0), \mathbf{v}_2^n(0), \ldots, \mathbf{x}_n^n(0), \mathbf{v}_n^n(0)$$

are such that

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{(\mathbf{x}_i^n(0), \mathbf{v}_i^n(0))} \longrightarrow \mu_0 \in \mathcal{P}(\mathbb{R}^6),$$

then, for each $t > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{(\mathbf{x}_i^n(t), \mathbf{v}_i^n(t))} \longrightarrow \mu_t,$$

where $\mu_t \in \mathcal{P}(\mathbb{R}^6)$ is the weak solution at time t of the Vlasov equation

$$\begin{aligned}
\frac{\partial}{\partial t} f(\mathbf{x}, \mathbf{v}, t) &= -\mathbf{v} \cdot \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{v}, t) - F_f(\mathbf{x}) \cdot \nabla_{\mathbf{v}} f(\mathbf{x}, \mathbf{v}, t) \\
F_f(\mathbf{x}) &= \int_{\mathbb{R}^3} F(\mathbf{x} - \mathbf{x}') f(\mathbf{x}', \mathbf{v}', t) d\mathbf{x}' d\mathbf{v}' \\
\mu_0 &= f(\mathbf{x}, \mathbf{v}, 0) d\mathbf{x} d\mathbf{v}.
\end{aligned} \tag{2.12}$$

Thanks to our Corollary 4.2, this theorem of Braun and Hepp implies that the family of n-particle processes (2.11) propagates chaos. The fact that the result of Braun and Hepp implies the propagation of chaos is also noted in [29, p. 99].

The deterministic particle systems (2.11) may be generalized to interacting *diffusions*. A diffusion is a Markov process with continuous trajectories, like the solution of a stochastic differential equation. McKean [22] initiated the study of propagation of chaos for diffusions and what is now called the McKean-Vlasov equation.

Let $\mathbf{v} : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ be bounded and globally Lipschitz. For each n, consider the system of n stochastic differential equations (SDEs)

$$dX_i^n = \left\{ \frac{1}{n} \sum_{j=1}^n \mathbf{v}(X_i^n, X_j^n) \right\} dt + \left\{ \frac{1}{n} \sum_{j=1}^n \sigma(X_i^n, X_j^n) \right\} dW_i, \tag{2.13}$$

for random vectors $X_1^n, X_2^n, \ldots, X_n^n$ in $\mathbb{R}^d$. The Wiener processes

$$W_1, \quad W_2, \quad W_3, \quad \ldots$$

are taken to be independent of one another and of the random initial conditions

$$X_1^n(0), \quad X_2^n(0), \quad \ldots, \quad X_n^n(0).$$

Each system of SDEs has a unique solution and defines a Markov transition function

$$K_n(\mathbf{x}, d\mathbf{y}, t) : (\mathbb{R}^d)^n \times \mathcal{B}_{(\mathbb{R}^d)^n} \times [0, \infty) \longrightarrow [0, 1]$$

by

$$\int_{\mathbb{R}^d} \phi(\mathbf{y}) K_n(\mathbf{x}, d\mathbf{y}, t) := \mathbb{E}^{\mathbf{x}}[\phi(X_1^n(t), \ldots, X_n^n(t))].$$

In other words, for fixed $t \geq 0$ and $\mathbf{x} \in (\mathbb{R}^d)^n$, $K_n(\mathbf{x}, \cdot, t)$ is the distribution of the position at time t of random trajectory

$$\mathbf{X}^n \equiv (X_1^n(t), X_2^n(t), \ldots, X_n^n(t)) \in \Big[C([0, \infty), \mathbb{R}^d) \Big]^n$$

that started at $\mathbf{X}^n(0) = \mathbf{x}$.

McKean [21, 22] proves that if the particles are initially stochastically independent but with a common distribution μ_0, then the sequence of n-particle joint distributions at time t is μ_t-chaotic, μ_t being the (weak) solution at time t of the nonlinear McKean-Vlasov equation

$$\begin{aligned}
\frac{\partial}{\partial t} f_t(\mathbf{x}) &= -\nabla \cdot [V_f(\mathbf{x}) f_t(\mathbf{x})] + \frac{1}{2} \Delta [D_f(\mathbf{x}) f_t(\mathbf{x})] \\
V_f(\mathbf{x}) &= \int_{\mathbb{R}^d} \mathbf{v}(\mathbf{x}, \mathbf{x}') f_t(\mathbf{x}') d\mathbf{x}' \\
D_f(\mathbf{x}) &= \left(\int_{\mathbb{R}^d} \sigma(\mathbf{x}, \mathbf{x}') f_t(\mathbf{x}') d\mathbf{x}' \right)^2 \\
f_0(\mathbf{x}) d\mathbf{x} &= \mu_0.
\end{aligned} \tag{2.14}$$

McKean's result includes that of Braun and Hepp: when $\sigma \equiv 0$ there is no diffusion and the system of SDEs (2.13) becomes a Vlasov system of ODEs like (2.11). Braun and Hepp seem unaware, in their paper of 1977, of McKean's important work of 1966. They use a different method to prove the propagation of chaos for Vlasov systems. Though they only treat the deterministic (Vlasov) case, their method can be generalized to prove that interacting (McKean-Vlasov) diffusions also propagate chaos.

McKean really proves much more than the propagation of chaos. Suppose the initial positions $X_1^n(0), \ldots, X_n^n(0)$ for the n-particle systems are taken to be the first n terms of a sequence $Z_1, Z_2, Z_3, \ldots$ of independent and μ_0 distributed random variables. McKean proves that $X_i^n(t)$, the random position of the i^{th} particle at time t, converges in mean square to $X_i^\infty(t)$ as n tends to infinity. The X_i^∞ are independent and identically distributed. $X_1^\infty(t)$ is sometimes called the *nonlinear process* and it satisfies the SDE

$$\begin{aligned}
dX &= \left\{ \int_{\mathbb{R}^d} \mathbf{v}(X, y) \mu_t(dy) \right\} dt + \left\{ \int_{\mathbb{R}^d} \sigma(X, y) \mu_t(dy) \right\} dW_1 \\
\mu_t &= \mathrm{Law}(X(t))
\end{aligned}$$

with $X(0) = Z_1$.

Chapter 3

Chaos and Weak Convergence

Chaos of a sequence of symmetric measures is equivalent to weak convergence of certain probability measures. This observation, due to Sznitman and Tanaka, is the subject of this chapter. First, in Section 3.1, the theory of weak convergence of probability measures is reviewed. The theorem of Sznitman and Tanaka is proved in Section 3.2. We examine this equivalence in the simplest context of finite probability spaces in Section 3.3.

This chapter ends with Theorem 3.5: on a finite space, chaos implies convergence of specific entropy to the entropy of the single-particle distribution.

3.1 Background

Let X be a set and $\mathcal{F}$ a class of subsets of X that contains the empty set and is closed under complementation and countable unions. $(X, \mathcal{F})$ is called a *measurable space*, and the sets in $\mathcal{F}$ are called *measurable*. A *probability measure* or *law* on $(X, \mathcal{F})$ is a countably additive, nonnegative function

$$P : \mathcal{F} \longrightarrow [0, 1]$$

satisfying $P(X) = 1$. The measure $P(F)$ of a set $F \in \mathcal{F}$ is the *probability of* F. Countable additivity requires the probability of a union of a sequence of disjoint measurable sets to equal the sum of their probabilities. The simplest probability measure is a *point mass* at a point $x \in X$, denoted $\delta(x)$ or δ_x; $\delta_x(F)$ equals one if $x \in F$, otherwise it equals zero.

Let $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ be measurable spaces, and let $h : X \longrightarrow Y$ be *measurable*, i.e., $h^{-1}(G) \in \mathcal{F}$ whenever $G \in \mathcal{G}$, where $h^{-1}(G)$ is the inverse image of G under h. Any probability measure P on X induces a probability measure $P \circ h^{-1}$ on Y via h. The probability measure *induced by* h is defined for $G \in \mathcal{G}$ by

$$\left(P \circ h^{-1}\right)(G) := P(h^{-1}(G)).$$

This definition implies that for any integrable function ϕ on $(Y, \mathcal{G}, P \circ h^{-1})$,

$$\int_Y \phi(y) P \circ h^{-1}(dy) = \int_X \phi(h(x)) P(dx).$$

Now let $(X, \mathcal{T})$ be a Hausdorff topological space with topology $\mathcal{T}$. The *Borel σ-algebra*, $\mathcal{B}$, is the smallest σ-algebra containing $\mathcal{T}$. The Borel algebra is thus the smallest σ-algebra with respect to which any function continuous on $(X, \mathcal{T})$ is measurable. The set of probability measures on $(X, \mathcal{B})$ is denoted $\mathcal{P}(X)$.

We often call probability measures simply "laws."

Let $C_b(X)$ denote the continuous and bounded real-valued functions on $(X, \mathcal{T})$. The set of laws $\mathcal{P}(X)$ is endowed with the weakest topology rendering continuous the maps

$$P \in \mathcal{P}(X) \longmapsto \int_X g(x) P(dx) \in \mathbb{R},$$

for all $g \in C_b(X)$. This is known as the weak topology on $\mathcal{P}(X)$. A net of laws $\{P_\beta\}$ in $\mathcal{P}(X)$ converges to P in the weak topology if and only if the nets $\{\int g P_\beta\}$ converge to $\int g P$ for all $g \in C_b(X)$.

We consider exclusively the case that X is homeomorphic to a separable metric space (S, d_S), so that we may use certain results of the theory of weak convergence. The theory of weak convergence of laws is customarily expounded for laws on separable metric spaces, and especially *complete* and separable metric spaces, because of the influence of Prohorov's original study [28] of 1956. Around the same time, Le Cam [19] developed the theory of of weak convergence of laws on completely regular topological spaces.

For separable metric spaces (S, d_S), the weak topology on $\mathcal{P}(S)$ is metrizable. Two metrics on $\mathcal{P}(S)$ that generate the weak topology are the Lévy-Prohorov distance LP and Dudley's distance BL^*. The Dudley distance between two laws $\mu, \nu \in \mathcal{P}(S)$ is

$$BL^*(\mu, \nu) := \sup_{g \in BL_1} \left\{ \left| \int_S g(s)\mu(ds) - \int_S g(s)\nu(ds) \right| \right\},$$

where g ranges over the class BL_1 of bounded Lipschitz functions from S to $\mathbb{R}$ defined as

$$BL_1 := \left\{ g(s) : \sup_{s \in S} \{|g(s)|\} + \sup_{s \neq t \in S} \{|g(s) - g(t)|/d_S(s,t)\} \leq 1 \right\}.$$

The Lévy-Prohorov distance between μ and ν is

$$LP(\mu, \nu) := \inf \left\{ \delta > 0 : \nu(B) \leq \mu(B^{+\delta}) + \delta \text{ for all closed sets } B \right\},$$

where $B^{+\delta}$ is the set of all points in S that are within δ of some point of B:

$$B^{+\delta} := \{s \in S : d_S(s, B) < \delta\}.$$

These metrics are discussed in Chapter 11 of the textbook *Real Analysis and Probability*, by Dudley [10].

The general theory of weak convergence in law on Polish spaces is due to Prohorov. (A topological space is *Polish* if it is homeomorphic to a complete, separable metric space.) Prohorov's theorem characterizes compact sets in $\mathcal{P}(X)$ when X is Polish, much as the Arzelà-Ascoli theorem characterizes compactness in the space of continuous functions on a compact Hausdorff space. The Arzelà-Ascoli theorem states that a family of functions on a compact space is relatively compact (has compact closure) in the topology of uniform convergence if and only if the family is equicontinuous and bounded. Prohorov's theorem states that a family of laws on a Polish space is relatively compact if and only if it is *tight*.

Tightness is a simple condition:

Definition 3.1 *Let $\Sigma \subset \mathcal{P}(X)$ be a family of laws on a topological space.*

The family Σ is **tight** *if for each $\epsilon > 0$ there exists a compact $K_\epsilon \subset X$ such that*

$$\sup_{\mu \in \Sigma} \{\mu(X \setminus K_\epsilon)\} < \epsilon.$$

Tightness implies relative compactness, and the conditions are equivalent in separable, topologically complete spaces:

Theorem 3.1 (Prohorov) *Suppose $(X, \mathcal{T})$ is homeomorphic to a separable metric space. Then, if $\Sigma \subset \mathcal{P}(X)$ is tight, its closure is compact in $\mathcal{P}(X)$.*

If $(X, \mathcal{T})$ is Polish (homeomorphic to a complete, separable metric space) then $\Sigma \subset \mathcal{P}(X)$ is tight if and only if its closure is compact in $\mathcal{P}(X)$.

It follows from Prohorov's theorem that $\mathcal{P}(X)$ is Polish if X is Polish [10]. This is important to us since the proof of our main theorem requires the application of Prohorov's theorem to $\mathcal{P}(X)$.

Convergence of Probability Measures by Patrick Billingsley [1] is a charming classic monograph on the theory of weak convergence of laws and its applications. Unfortunately, this text is missing some essential material, especially the metric approach to weak convergence. It is well complemented by the material in [10].

3.2 The Theorem of Sznitman and Tanaka

Let (S, d_S) be a separable metric space with Borel algebra $\mathcal{B}_S$. Let S^n denote the n-fold product of S with itself;

$$S^n := \{(s_1, s_2, \ldots, s_n) : s_i \in S \quad \text{for} \quad i = 1, 2, \ldots, n\}.$$

S^n is itself metrizable in a variety of equivalent ways that all generate the same topology and the same Borel algebra $\mathcal{B}_{S^n}$.

The *marginal* of a law $\rho_n \in (S^n)$ on the first k-coordinates ($k \le n$) is the law $\rho_n^{(k)} \in \mathcal{P}(S^k)$ induced by the projection

$$(s_1, s_2, \ldots, s_n) \longmapsto (s_1, s_2, \ldots, s_k).$$

Equivalently,

$$\rho_n^{(k)}(B_1, B_2, \ldots, B_k) = \rho_n(B_1, B_2, \ldots, B_k, S, S, \ldots, S),$$

for all $B_1, B_2, \ldots, B_k \in \mathcal{B}_S$. If $\rho \in \mathcal{P}(S)$, the *product law* $\rho^{\otimes n} \in \mathcal{P}(S^n)$ is the law $\rho(ds_1)\rho(ds_2)\cdots\rho(ds_n)$. Note that $(\rho^{\otimes n})^{(k)} = \rho^{\otimes k}$.

Let Π_n denote the set of permutations of $\{1, 2, \ldots, n\}$. The permutations Π_n act on S^n by permuting coordinates: the map $\pi\cdot : S^n \longrightarrow S^n$ is

$$\pi \cdot (s_1, s_2, \ldots, s_n) := (s_{\pi(1)}, s_{\pi(2)}, \ldots, s_{\pi(n)}).$$

If E is any subset of S^n, define

$$\pi \cdot E = \{\pi \cdot \mathbf{s} : \mathbf{s} \in E\}.$$

A law ρ on S^n is *symmetric* if $\rho(\pi \cdot B) = \rho(B)$ for all $\pi \in \Pi_n$ and all $B \in \mathcal{B}_{S^n}$. Products $\rho^{\otimes n}$ are symmetric, for example. The *symmetrization* $\widetilde{\rho}$ of a law $\rho \in \mathcal{P}(S^n)$ is the symmetric law such that

$$\widetilde{\rho}(B) := \frac{1}{n!} \sum_{\pi \in \Pi_n} \rho(\pi \cdot B),$$

for all $B \in \mathcal{B}_{S^n}$.

Definition 3.2 (Kac, 1954) *Let (S, d_S) be a separable metric space. Let ρ be a law on S, and for $n = 1, 2, \ldots$, let ρ_n be a symmetric law on S^n.*

*The sequence $\{\rho_n\}$ is ρ-**chaotic** if, for each natural number k and each choice*

$$\phi_1(s), \quad \phi_2(s), \quad \ldots, \quad \phi_k(s)$$

of k bounded and continuous functions on S,

$$\lim_{n\to\infty} \int_{S^n} \phi_1(s_1)\phi_2(s_2)\cdots\phi_k(s_k)\rho_n(ds_1 ds_2 \ldots ds_n) = \prod_{i=1}^{k} \int_S \phi_i(s)\rho(ds). \tag{3.1}$$

In case S is Polish, condition 3.1 implies the weak convergence of the marginals to products $\rho^{\otimes k}$, because the class of functions of the form

$$\phi_1(x_1)\phi_2(x_2)\cdots\phi_k(x_k); \qquad \phi_1,\ldots,\phi_k \in C_b(S) \tag{3.2}$$

is a *convergence determining class* for $\mathcal{P}(S^k)$[12]. Condition (3.1) shows that the sequence of the marginals $\rho_n^{(k)}$ converges to $\rho^{\otimes k}$ weakly *for functions of the form* 3.2 , hence it converges weakly. Thus, *if S is Polish*, a sequence $\{\rho_n\}$ of symmetric laws on S^n is ρ-chaotic if and only if

$$\lim_{n\to\infty} \rho_n^{(k)} = \rho^{\otimes k},$$

for any natural number k.

It turns out, however, that S does not need to be Polish. It will be seen from the proof of the next theorem that condition (3.1) implies the convergence of the marginals $\rho_n^{(k)}$ *even if S is not Polish*, but only separable.

The following theorem of Sznitman and Tanaka states that a sequence of symmetric laws is chaotic if and only if the induced sequence of laws of the random empirical measures converges to a point mass. Let

$$\varepsilon_n((s_1,s_2,\ldots,s_n)) := \frac{1}{n}\sum_{i=1}^{n}\delta(s_i) \tag{3.3}$$

define a map from S^n to $\mathcal{P}(S)$. These maps are measurable for each n, and $\varepsilon_n(\pi\cdot\mathbf{s}) = \varepsilon_n(\mathbf{s})$ for all $\mathbf{s}\in S^n, \pi\in\Pi_n$.

Theorem 3.2 (Sznitman, Tanaka) *$\{\rho_n\}$ is ρ-chaotic if and only if*

$$\rho_n\circ\varepsilon_n^{-1} \longrightarrow \delta(\rho) \tag{3.4}$$

in $\mathcal{P}(\mathcal{P}(S))$.

Proof:

Suppose $\{\rho_n\}$ is ρ-chaotic.

A sequence of laws $\{\mu_n\}$ on a completely regular topological space X converges to $\delta(x)\in\mathcal{P}(X)$ if and only if for each neighborhood N of x

$$\lim_{n\longrightarrow\infty} \mu_n(X\setminus N) = 0 \tag{3.5}$$

for each neighborhood N of x. Therefore, to prove the convergence of $\rho_n\circ\varepsilon_n^{-1}$ to $\delta(\rho)$ in $\mathcal{P}(\mathcal{P}(S))$ it suffices to verify (3.5) on a subbase of neighborhoods of $\rho\in\mathcal{P}(S)$. The class of sets of the form

$$\left\{\nu\in\mathcal{P}(S) : \left|\int_S g(s)\nu(ds) - \int_S g(s)\rho(ds)\right| < \epsilon\right\}; \quad \epsilon>0, g\in C_b(S)$$

is a neighborhood subbase at $\delta(\rho)$, so it suffices to show that

$$\rho_n \circ \varepsilon_n^{-1} \left(\left\{ \nu : \left| \int_S g(s)\nu(ds) - \int_S g(s)\rho(ds) \right| \geq \epsilon \right\} \right) \longrightarrow 0. \tag{3.6}$$

Writing $\int_S g(s)\nu(ds)$ as $< g, \nu >$, we calculate

$$\begin{aligned}
& \int_{S^n} |< g, \varepsilon_n(\mathbf{s}) > - < g, \rho >|^2 \, \rho_n(d\mathbf{s}) \\
= & \int_{S^n} \left(\frac{1}{n} \sum_{i=1}^{n} g(s_i) - < g, \rho > \right)^2 \rho_n(ds_1 ds_2 \cdots ds_n) \\
= & \frac{1}{n^2} \sum_{i,j=1}^{n} \int_{S^n} (g(s_i) - < g, \rho >)(g(s_j) - < g, \rho >) \, \rho_n(d\mathbf{s}) \\
= & \frac{1}{n} \int_S (g(s) - < g, \rho >)^2 \, \rho_n^{(1)}(ds) \\
& + \frac{n-1}{n} \int_{S \times S} (g(s_1) - < g, \rho >)(g(s_2) - < g, \rho >) \, \rho_n^{(2)}(ds_1 ds_2),
\end{aligned}$$

the last equality by the symmetry of ρ_n. Thus condition (3.1) for $k = 1, 2$ implies that

$$\int_{S^n} |< g, \varepsilon_n(\mathbf{s}) > - < g, \rho >|^2 \, \rho_n(d\mathbf{s}) \longrightarrow 0,$$

and hence that (3.6) holds. Condition (3.1) thus implies condition (3.4).

Now suppose that $\rho_n \circ \varepsilon_n^{-1}$ tends to $\delta(\rho)$.

For natural numbers $k \leq n$, let $\mathcal{J}_{n:k}$ and $\mathcal{I}_{n:k}$ denote respectively the set of all maps and the set of injections from $\{1, 2, \ldots, k\}$ into $\{1, 2, \ldots, n\}$. Define the map $\varepsilon_{n:k}$ from S^n to $\mathcal{P}(S^k)$ by

$$\varepsilon_{n:k}((s_1, s_2, \ldots, s_n)) := \frac{(n-k)!}{n!} \sum_{i \in \mathcal{I}_{n:k}} \delta_{(s_{i(1)}, \ldots, s_{i(k)})}. \tag{3.7}$$

$\varepsilon_{n:k}(\mathbf{s})$ is the empirical measure of k-tuples of coordinates of $\mathbf{s}$, sampled without replacement. Define also

$$\vartheta_{n:k}(\mathbf{s}) := \varepsilon_n(\mathbf{s})^{\otimes k} = \frac{1}{n^k} \sum_{j \in \mathcal{J}_{n:k}} \delta_{(s_{j(1)}, \ldots, s_{j(k)})} \ , \tag{3.8}$$

the empirical measure of *all* k-tuples from $\mathbf{s}$. When $n >> k$, these two empirical measures are close in total variation (TV) and *a fortiori* in Dudley's distance on $\mathcal{P}(S)$:

$$\begin{aligned}
& BL^*\left(\varepsilon_{n:k}(\mathbf{s}), \varepsilon_n(\mathbf{s})^{\otimes k}\right) \\
\leq\ & \left\| \frac{(n-k)!}{n!} \sum_{i \in \mathcal{I}_{n:k}} \delta_{(s_{i(1)}, \ldots, s_{i(k)})} - \frac{1}{n^k} \sum_{j \in \mathcal{J}_{n:k}} \delta_{(s_{j(1)}, \ldots, s_{j(k)})} \right\|_{TV} \\
\leq\ & 2\left(1 - \frac{n!}{n^k (n-k)!}\right) .
\end{aligned}$$

Since this bound is uniform in $\mathbf{s}$, it follows that $\rho_n \circ \varepsilon_{n:k}^{-1}$ is near $\rho_n \circ \vartheta_{n:k}^{-1}$ in $\mathcal{P}(\mathcal{P}(S^k))$. In fact, both the Lévy-Prohorov and the Dudley distances between the two laws are bounded above:

$$\begin{aligned}
BL^*\left(\rho_n \circ \varepsilon_{n:k}^{-1}, \rho_n \circ \vartheta_{n:k}^{-1}\right) &\leq 2\left(1 - \frac{n!}{n^k (n-k)!}\right) \\
\text{and} \qquad LP\left(\rho_n \circ \varepsilon_{n:k}^{-1}, \rho_n \circ \vartheta_{n:k}^{-1}\right) &\leq 2\left(1 - \frac{n!}{n^k (n-k)!}\right) .
\end{aligned} \tag{3.9}$$

Condition (3.4) and definition (3.8) imply that $\rho_n \circ \vartheta_{n:k}^{-1}$ converges to $\delta(\rho^{\otimes k})$ in $\mathcal{P}(\mathcal{P}(S^k))$. By (3.9), $\rho_n \circ \varepsilon_{n:k}^{-1}$ converges to $\delta(\rho^{\otimes k})$ as well.

Now, if $\phi \in C_b(S^k)$,

$$\begin{aligned}
& \lim_{n\to\infty} \int_{S^n} \phi(s_1, s_2, \ldots, s_k) \rho_n(d\mathbf{s}) \\
=\ & \lim_{n\to\infty} \int_{S^n} \left\{ \frac{(n-k)!}{n!} \sum_{i \in \mathcal{I}_{n:k}} \phi(s_{i(1)}, \ldots, s_{i(k)}) \right\} \rho_n(d\mathbf{s}) \\
=\ & \lim_{n\to\infty} \int_{S^n} < \phi, \varepsilon_{n:k}(\mathbf{s}) > \rho_n(d\mathbf{s}) \\
=\ & \lim_{n\to\infty} \int_{\mathcal{P}(S^k)} < \phi, \mu > \rho_n \circ \varepsilon_{n:k}^{-1}(d\mu) \\
=\ & \lim_{n\to\infty} \int_{\mathcal{P}(S^k)} < \phi, \mu > \rho_n \circ \vartheta_{n:k}^{-1}(d\mu) \\
=\ & \int_{\mathcal{P}(S^k)} < \phi, \mu > \delta(\rho^{\otimes k})(d\mu) \\
=\ & \int_{S^k} \phi(s_1, s_2, \ldots, s_k) \rho(ds_1) \cdots \rho(ds_k).
\end{aligned}$$

Thus, condition (3.4) implies (3.1). □

The preceding arguments have actually proved the following stronger version of Theorem 3.2.

Theorem 3.3 *Let S be a separable metric space and for each n let ρ_n be a symmetric law on S^n.*

The following are equivalent:

Kac's condition for $k=2$*: For all $\phi_1, \phi_2 \in C_b(S)$,*

$$\lim_{n\to\infty} \int_{S^n} \phi_1(s_1)\phi_2(s_2)\rho_n(d\mathbf{s}) = \int_S \phi_1(s)\rho(ds) \int_S \phi_2(s)\rho(ds); \tag{3.10}$$

Condition of Sznitman and Tanaka*: For all natural numbers k, the laws $\rho_n \circ \varepsilon_{n:k}^{-1}$ converge to $\delta(\rho^{\otimes k})$ in $\mathcal{P}(\mathcal{P}(S))$ as n tends to infinity, where $\varepsilon_{n:k}$ is the empirical measure defined in (3.7) ;*

Weak convergence of marginals*: For all k, the marginals $\rho_n^{(k)}$ converge weakly to $\rho^{\otimes k}$ as n tends to infinity.*

3.3 Chaos on Finite Sets

Throughout this section, let $S = \{s_1, s_2, \ldots, s_k\}$ be a finite set.

For each natural number n, let

$$\rho_n(x_1, x_2, \ldots, x_n)$$

be a symmetric law on $S^n = S \times S \times \cdots \times S$. Because of its symmetry, ρ_n is entirely determined by the probability function

$$P_n(j_1, j_2, \ldots, j_k); \quad \sum_{i=1}^{k} j_i = n \tag{3.11}$$

that gives the probability there are j_1 coordinates equal to s_1, j_2 coordinates equal to s_2, and so on. The probability of $(x_1, x_2, \ldots, x_n) \in S^n$ is

$$\rho_n(x_1, x_2, \ldots, x_n) = P_n(j_1, j_2, \ldots, j_k) / \frac{n!}{j_1! \cdots j_k!} \quad , \tag{3.12}$$

where $j_i(x_1, x_2, \ldots, x_n)$ is the number of coordinates of $(x_1, x_2, \ldots, x_n)$ that equal s_i.

Let Δ_{k-1} denote the unit simplex in $\mathbb{R}^k$:

$$\Delta_{k-1} = \left\{ (q_1, q_2, \ldots, q_k) : \sum_{i=1}^{k} q_i = 1, q_i \geq 0 \right\}.$$

Given ρ_n, define a law μ_n on Δ_{k-1} by

$$\mu_n := \sum_{\mathbf{j}} P_n(\mathbf{j}) \delta(\mathbf{j}/n) \quad , \tag{3.13}$$

where $\mathbf{j}$ ranges over k-tuples of nonnegative integers that sum to n.

Finally, let $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ be a point of Δ_{k-1}, and let p denote the law on S given by $p(s_i) = p_i$. With these definitions and notations, we can formulate simpler versions of Definition 3.2 and Theorem 3.2 for finite probability spaces:

Definition 3.3 (Chaos for Finite State Spaces) *The sequence* $\{\rho_n\}$ *is* p-**chaotic** *if for each natural number* m *and each* $(z_1, z_2, \ldots, z_m)$ *in* S^m*,*

$$\lim_{n\to\infty} \sum_{x_1,\ldots,x_{n-m}\in S} \rho_n(z_1, z_2, \ldots, z_m, x_1, x_2, \ldots, x_{n-m}) = \prod_{i=1}^{m} p(z_i).$$

Theorem 3.4 *The sequence* $\{\rho_n\}$ *is p-chaotic if and only if* μ_n *converges weakly to* $\delta(\mathbf{p})$*.*

Equivalently, $\{\rho_n\}$ *is p-chaotic if and only if*

$$\lim_{n\to\infty} \sum_{(j_1,\ldots,j_k)} P_n(j_1, \ldots, j_k) F\left(\frac{j_1}{n}, \frac{j_2}{n}, \ldots, \frac{j_k}{n}\right) = F(\mathbf{p}), \tag{3.14}$$

for every continuous function F *on the simplex* Δ_{k-1}*.*

Proof:

This is a special case of Theorem 3.2. □

Formula (3.14) will be used to prove that, on finite probability spaces, chaos implies convergence of specific entropy. We are borrowing the expression "specific entropy" from statistical mechanics, where it refers to entropy per particle.

For laws π and μ on a measurable space $(X, \mathcal{F})$, the entropy of μ relative to π is defined to be

$$H_\pi(\mu) := -\int_X \left[\frac{d\mu}{d\pi}\right] \log\left[\frac{d\mu}{d\pi}\right] d\pi$$

if μ is absolutely continuous relative to π with density $\left[\frac{d\mu}{d\pi}\right]$, and to equal $-\infty$ otherwise.

Relative entropy is nonpositive, but might equal $-\infty$. $H_\pi(\mu)$ achieves its maximum of 0 only when $\mu = \pi$. If X is a Polish space, $H_\pi(\mu)$ is a upper semicontinuous function of μ relative to the weak topology on $\mathcal{P}(X)$. The entropy of a joint law is less than or equal to the sum of the entropies of its marginals, with equality only if the joint law is a product measure. That is, if $\mu \in \mathcal{P}(X \times X)$ with marginals $\mu_1, \mu_2 \in \mathcal{P}(X)$, then

$$H_{\pi\otimes\pi}(\mu) \leq H_\pi(\mu_1) + H_\pi(\mu_2), \tag{3.15}$$

for any *reference law* $\pi \in \mathcal{P}(X)$. The reader is referred to [11, pp. 32-40] for properties of the relative entropy.

Now, if π is a reference law and $\{\rho_n\}$ is a p-chaotic sequence of laws on a general separable metric space (where chaos has been defined), the subadditivity (3.15) of entropy guarantees that

$$\limsup_{n\to\infty} \frac{1}{n} H_{\pi^{\otimes n}}(\rho_n) \leq H_\pi(p).$$

The left hand side of this inequality is what we are calling the specific entropy. In case $\{\rho_n\}$ is *purely* chaotic, i.e., in case $\rho_n = p^{\otimes n}$ for all n, the specific entropy always equals the entropy of p. At the other extreme, when the symmetric laws of a p-chaotic sequence $\{\rho_n\}$ are not absolutely continuous relative to the laws $p^{\otimes n}$, the above inequality is strict, for then

$$\lim_{n\to\infty} \frac{1}{n} H_{p^{\otimes n}}(\rho_n) = -\infty < 0 = H_p(p).$$

However, *if the space is finite*, one can prove that the specific entropy of a chaotic sequence does converge:

Theorem 3.5 (Specific Entropy Converges) *Let $S = \{s_1, s_2, \ldots, s_k\}$ be a finite set, and for each n let $\rho_n \in \mathcal{P}(S^n)$ be a symmetric law.*

If the sequence $\{\rho_n\}$ is p-chaotic, then

$$\lim_{n\to\infty} \left(-\frac{1}{n} \sum_{\mathbf{x}\in S^n} \rho_n(\mathbf{x}) \log \rho_n(\mathbf{x}) \right) = -\sum_{i=1}^{k} p_i \log p_i.$$

Proof:

By the relationship (3.12) between ρ_n and P_n,

$$-\frac{1}{n} \sum_{\mathbf{x}\in S^n} \rho_n(\mathbf{x}) \log \rho_n(\mathbf{x}) = -\frac{1}{n} \sum_{(j_1,\ldots,j_k)} P_n(j_1,\ldots,j_k) \log \frac{P_n(j_1,\ldots,j_k)}{\frac{n!}{j_1!\cdots j_k!}}.$$

This equals

$$-\frac{1}{n} \sum_{\mathbf{j}} P_n(\mathbf{j}) \log P_n(\mathbf{j}) + \frac{1}{n} \sum_{\mathbf{j}} P_n(\mathbf{j}) \log \left(\frac{n!}{j_1! \cdots j_k!} \right), \tag{3.16}$$

abbreviating $(j_1, j_2, \ldots, j_k)$ by $\mathbf{j}$. The first addend in (3.16) is $O\left(\frac{\log n}{n}\right)$, since it equals an n^{th} part of the entropy of a probability function $P_n(\mathbf{j})$ on fewer than n^k points, which entropy cannot exceed $\log(n^k) = k \log n$.

Using Stirling's approximation

$$\log j! = j \log j - j + \epsilon_j$$

where $0 < \epsilon_j < 1 + \log j$, and the fact that $n = \sum_{i=1}^{k} j_i$, one finds that

$$\begin{aligned}
\log\left(\frac{n!}{j_1!\cdots j_k!}\right) &= n\log n - n + \epsilon_n - \sum_{i=1}^{k}(j_i\log j_i - j_i) - \sum_{i=1}^{k}\epsilon_{j_i} \\
&= n\log n - \sum_{i=1}^{k} j_i\log j_i + O\left(\log n\right) \\
&= -\sum_{i=1}^{k} j_i\log\frac{j_i}{n} + O\left(\log n\right).
\end{aligned}$$

Substituting this into the second term of (3.16) shows that

$$-\frac{1}{n}\sum_{\mathbf{x}\in S^n}\rho_n(\mathbf{x})\log\rho_n(\mathbf{x}) = -\sum_{\mathbf{j}} P_n(\mathbf{j})\sum_{i=1}^{k}\left(\frac{j_i}{n}\right)\log\left(\frac{j_i}{n}\right) + O\left(\frac{\log n}{n}\right). \tag{3.17}$$

Since ρ_n is ρ-chaotic, formula (3.14) of Theorem 3.4 tells us that

$$\lim_{n\to\infty}\left\{-\sum_{\mathbf{j}} P_n(\mathbf{j})\sum_{i=1}^{k}\left(\frac{j_i}{n}\right)\log\left(\frac{j_i}{n}\right)\right\} = -\sum_{i=1}^{k} p_i\log p_i. \tag{3.18}$$

By (3.17) and (3.18) the specific entropy converges to the entropy of p:

$$\lim_{n\to\infty}\left(-\frac{1}{n}\sum_{\mathbf{x}\in S^n}\rho_n(\mathbf{x})\log\rho_n(\mathbf{x})\right) = -\sum_{i=1}^{k} p_i\log p_i. \qquad \blacksquare$$

Chapter 4

Propagation of Chaos

This brief chapter is devoted to the proof of the main theorem stated in Section 1.2. Definitions are given and the approach is outlined in Section 4.1. Lemmas are proved in Section 4.2 that expedite the proofs of the theorems of Section 4.3.

4.1 Preliminaries

Let $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ be two measurable spaces. A Markov transition function $K(x, E)$ on $X \times \mathcal{G}$ is a function that satisfies the following two conditions:

(1) $K(x, \cdot)$ is a probability measure on $(Y, \mathcal{G})$ for each $x \in X$, and

(2) $K(\cdot, E)$ is a measurable function on $(X, \mathcal{F})$ for each $E \in \mathcal{G}$.

Whenever X and Y are measurable spaces and there is no confusion about what their σ-algebras are supposed to be, we usually speak of Markov transitions from X to Y rather than transition *functions*. In particular, if S and T are metric spaces, a Markov transition from S to T is a transition function on $S \times \mathcal{B}_T$.

A Markov *process* on a state space $(X, \mathcal{F})$ determines a family, indexed by time, of Markov transitions from X to itself: $\{K(x, E, t)\}_{t \geq 0}$. The transitions satisfy — in addition to (1) and (2) above — the Chapman-Kolmogorov equations

$$K(x, E, s+t) = \int_X K(x, dy, s) K(y, E, t); \quad s, t \geq 0, x \in X, E \in \mathcal{F}.$$

Let (S, d_S) and (T, d_T) be separable metric spaces. For each n, let K_n be a Markov transition from S^n to T^n. We assume that the Markov transition function K_n is symmetric in the sense that, if π is a permutation in Π_n and A is a Borel subset of T^n,

$$K_n(\pi \cdot \mathbf{s}, \pi \cdot A) = K_n(\mathbf{s}, A). \tag{4.1}$$

Definition 4.1 (Propagation of Chaos) *Let $\{K_n\}$ be as above.*

The sequence $\{K_n\}$ **propagates chaos** *if, whenever $\{\rho_n\}$ is a ρ-chaotic sequence of measures on S^n, the measures*

$$\int_{S_n} K_n(\mathbf{s}, \cdot)\rho_n(d\mathbf{s})$$

on T^n are τ-chaotic for some $\tau \in \mathcal{P}(T)$.

When we say that a family of n-particle Markov *processes* on a state space S **propagates chaos** we mean that, for each fixed time $t > 0$, the family of associated n-particle transition functions $\{K_n(\mathbf{s}, E, t)\}$ propagates chaos.

Most Markov processes of interest are characterized by their laws on nice *path spaces*, such as $C([0,\infty), S)$ the space of continuous paths in S, or the space $D([0,\infty), S)$ of right continuous paths in S having left limits. For such processes, the function that maps a state $s \in S$ to the law of the process started at s defines a Markov transition from S to the entire path space. Now, if a sequence of transitions $K_n(s, \cdot)$ from S^n to the path spaces $C([0,\infty), S^n)$ or $D([0,\infty), S^n)$ propagates chaos, then, *a fortiori*, it propagates chaotic sequences of initial laws to chaotic sequences of laws on S^n at any (fixed) later time. We have defined the propagation of chaos for sequences of Markov transitions from S^n to a (possibly) different space T^n, instead of simply from S^n to itself, with the case where T is path space especially in mind. This way, our ensuing study will pertain even to those families of processes that propagate the chaos of initial laws to the chaos of laws on the whole path space.

We are going to prove that a sequence of Markov transitions $\{K_n\}_{n=1}^{\infty}$ propagates chaos if and only if

$$\left\{\widetilde{K}_n(\mathbf{s}_n, \cdot)\right\}_{n=1}^{\infty}$$

is chaotic whenever $\mathbf{s}_n \in S^n$ satisfy $\varepsilon_n(\mathbf{s}_n) \longrightarrow p$ in $\mathcal{P}(S)$. We employ the weak convergence characterization of chaos of Sznitman and Tanaka and we assume that S is Polish. To study the *propagation* of chaos, we project the transitions K_n from S^n to T^n onto transitions from $\varepsilon_n(S^n)$ to $\mathcal{P}(T)$, and then apply Theorem 3.2, which projects chaotic sequences of symmetric laws on the spaces S^n onto convergent sequences of laws on $\mathcal{P}(S)$.

From now on, the notation ε_n is used both for the map from S^n to n-point empirical measures on S and for the same kind of map on T^n.

Markov transitions K_n from S^n to T^n induce Markov transition functions H_n from $\varepsilon_n(S^n)$ to $\varepsilon_n(T^n)$. The induced transition function can be defined in terms of a Markov transition J_n from $\varepsilon_n(S^n)$ to S^n

which acts as a kind of inverse of ε_n. For fixed $\zeta \in \varepsilon_n(S^n)$, let $J_n(\zeta, \cdot)$ denote the atomic probability measure on S^n that allots equal probability to each of the points in $\varepsilon^{-1}(\{\zeta\})$, a set containing at most $n!$ points. Putting it another way, $J_n(\zeta, E)$ equals the proportion of points $\mathbf{s} \in S^n$ such that $\varepsilon_n(\mathbf{s}) = \zeta$ that lie in $E \subset S^n$. A Markov transition K_n from S^n to T^n induces a Markov transition H_n from $\varepsilon_n(S^n)$ to $\varepsilon_n(T^n)$ defined by

$$H_n(\zeta, G) := \int_{\mathbf{s} \in S^n} J_n(\zeta, d\mathbf{s}) K_n\left(\mathbf{s}, \varepsilon_n^{-1}(G)\right) \tag{4.2}$$

for $\zeta \in \varepsilon_n(S^n)$ and G a measurable subset of $\varepsilon_n(T^n)$. Note that if $\mathbf{s} \in S^n$,

$$H_n(\varepsilon_n(\mathbf{s}), \cdot) = K_n(\mathbf{s}, \cdot) \circ \varepsilon_n^{-1}, \tag{4.3}$$

where the maps ε_n written on the left and right hand sides are, respectively, the maps from S^n and T^n to empirical measures in $\mathcal{P}(S)$ and $\mathcal{P}(T)$.

Theorem 3.2 shows that propagation of chaos by a sequence K_n is equivalent to the following condition on the induced transitions H_n.

Proposition 4.1 *The sequence of Markov transitions $\{K_n\}_{n=1}^{\infty}$ propagates chaos if and only if, whenever $\{\mu_n \in \mathcal{P}(\varepsilon_n(S^n))\}_{n=1}^{\infty}$ converges in $\mathcal{P}(\mathcal{P}(S))$ to $\delta(p)$, the sequence*

$$\left\{ \int_{\varepsilon_n(S^n)} H_n(\zeta, \cdot) \mu_n(d\zeta) \right\}_{n=1}^{\infty} \tag{4.4}$$

converges in $\mathcal{P}(\mathcal{P}(T))$ to $\delta(q)$, for some $q \in \mathcal{P}(T)$.

Proof:

$\{\mu_n \in \mathcal{P}(\varepsilon_n(S^n))\}$ converges to $\delta(p)$ if and only if

$$\left\{ \int_{\varepsilon_n(S^n)} J_n(\zeta, \cdot) \mu_n(d\zeta) \right\}_{n=1}^{\infty}$$

is chaotic. Therefore, the sequence of transitions $\{K_n\}$ propagates chaos if and only if

$$\left\{ \int_{S^n} K_n(\mathbf{s}, \cdot) \int_{\varepsilon_n(S^n)} J_n(\zeta, d\mathbf{s}) \mu_n(d\zeta) \right\}_{n=1}^{\infty} \tag{4.5}$$

is chaotic whenever $\mu_n \longrightarrow \delta(p)$.

Now, using definition (4.2) of the transitions H_n, we find

$$\begin{aligned}
&\left(\int_{S^n} K_n(\mathbf{s},\cdot)\int_{\varepsilon_n(S^n)} J_n(\zeta, d\mathbf{s})\mu_n(d\zeta)\right)\circ\varepsilon_n^{-1} \\
= &\int_{\varepsilon_n(S^n)}\left(\int_{\mathbf{s}\in S^n} J_n(\zeta, d\mathbf{s})K_n(\mathbf{s},\cdot)\circ\varepsilon_n^{-1}\right)\mu_n(d\zeta) \\
= &\int_{\varepsilon_n(S^n)} H_n(\zeta,\cdot)\mu_n(d\zeta).
\end{aligned}$$

Thus, by Theorem 3.2, the sequence (4.5) is chaotic if and only if the sequence (4.4) converges to a point mass in $\mathcal{P}(\mathcal{P}(T))$. Therefore, $\{K_n\}$ propagates chaos if and only if (4.4) converges to $\delta(q)$, for some $q \in \mathcal{P}(T)$. □

Proposition 4.1 implies that if a sequence of Markov transitions $\{K_n\}$ propagates chaos, then the sequence $\{H_n(\zeta_n,\cdot)\}$ converges to a point mass whenever $\{\zeta_n \in \varepsilon_n(S^n)\}$ converges in $\mathcal{P}(S)$. That this condition implies propagation of chaos (and is not just a necessary condition) is equivalent to our main theorem. To prove the sufficiency of the condition, we use Lemma 4.3 of the next section.

The lemmas of Section 4.2 are presented in a general context. In Section 4.3 we apply these lemmas to propagation of chaos. In that context the induced transitions H_n, thought of as functions from $\varepsilon_n(S^n)$ to $\mathcal{P}(\mathcal{P}(T))$, behave like the maps f_n of the lemmas.

4.2 Lemmas

Let (X, d_X) be a metric space, and $D_1 \subset D_2 \subset \cdots$ an increasing chain of Borel subsets of X whose union is dense in X. For each natural number n, let f_n be a measurable real-valued function on D_n.

Consider the following four conditions on the sequence $\{f_n\}_{n=1}^{\infty}$. They are listed in order of decreasing strength.

[A] Whenever $\{\mu_n\}$ is a weakly convergent sequence of probability measures on X with μ_n supported on D_n, then the sequence

$$\left\{\int_X f_n(x)\mu_n(dx)\right\}_{n=1}^{\infty}$$

of real numbers converges as well.

[B] Whenever $\{\mu_n\}$ is a sequence of probability measures on X that converges weakly to $\delta(x)$ for some $x \in X$, and μ_n is supported on D_n, then the sequence $\left\{\int_X f_n(x)\mu_n(dx)\right\}$ also converges.

[C] Whenever $\{d_n\}$ is a convergent sequence of points in X, with $d_n \in D_n$, then $\{f_n(d_n)\}$ also converges.

[D] For any compact $K \subset X$, and for any $\epsilon > 0$, there exists a natural number N such that, whenever $m \geq n \geq N$ and $d \in D_n \cap K$, then

$$|f_m(d) - f_n(d)| < \epsilon.$$

Lemma 4.1 *[A] $\Rightarrow$ [B] $\Rightarrow$ [C] $\Rightarrow$ [D].*

Proof:

Clearly [A] $\Rightarrow$ [B]. Setting $\mu_n = \delta(d_n)$ in [B] shows that [B] $\Rightarrow$ [C].

To show that [C] $\Rightarrow$ [D], suppose that [C] holds but that [D] fails to hold for some compact $K \subset X$ and some $\epsilon > 0$. Then there exists an $\epsilon > 0$, two increasing sequences of natural numbers $\{n(k)\}$ and $\{m(k)\}$ with

$$n(k+1) > m(k) > n(k)$$

for all k, and a sequence of points $d_k \in D_{n(k)} \cap K$, such that

$$\left| f_{m(k)}(d_k) - f_{n(k)}(d_k) \right| \geq \epsilon. \tag{4.6}$$

Since K is compact, there exists an increasing sequence of natural numbers $\{k(j)\}$ such that $\{d_k(j)\}_{j=1}^{\infty}$ converges. Now define the convergent sequence $\{e_i \in D_{n(k(i))}\}_{i=1}^{\infty}$ by $e_i = d_k(j)$ when $n(k(j)) \leq i < n(k(j+1))$. By [C], the sequence $\{f_i(e_i)\}$ converges. But $\{f_i(e_i)\}$ *does not converge* along the subsequence indexed by

$$n(k(1)), \quad m(k(1)), \quad n(k(2)), \quad m(k(2)), \ldots$$

because of (4.6) and the fact that

$$e_{n(k(j))} = e_{m(k(j))} = d_{k(j)}.$$

This contradiction shows that [C] must imply [D]. $\square$

Lemma 4.2 *If condition [C] holds, then whenever $\{d_n \in D_n\}$ converges to $x \in X$, the limit of $\{f_n(d_n)\}$ depends only on x. The function of x which may thus be defined as*

$$f(x) := \lim_{n \to \infty} f_n(d_n)$$

when $d_n \longrightarrow x$, is continuous.

Proof:

Assume that condition [C] holds.

Suppose $\{d_n \in D_n\}$ and $\{e_n \in D_n\}$ are two sequences that both converge to $x \in X$. Then the sequence $d_1, e_2, d_3, e_4, d_5, e_6, \ldots$ also converges to x, and its n^{th} term is a member of D_n. By condition [C], the sequence $f_1(d_1), f_2(e_2), f_3(d_3), \ldots$ converges. This shows that $\lim f_n(d_n) = \lim f_n(e_n)$.

Suppose $x_k \longrightarrow x$ in X. Given $\epsilon > 0$, it is possible to find an increasing sequence $\{n(k)\}$ of natural numbers and a sequence of points $\{e_{n(k)} \in D_{n(k)}\}$ such that $d_X(e_{n(k)}, x_k) < \frac{1}{k}$ while $|f_{n(k)}(e_{n(k)}) - f(x_k)| < \epsilon$. Then $\{e_{n(k)}\}$ converges to x just as $\{x_k\}$ does, so $\lim_{k\to\infty} f_{n(k)}(e_{n(k)}) = f(x)$. Now

$$|f(x_k) - f(x)| \leq |f(x_k) - f_{n(k)}(e_{n(k)})| + |f_{n(k)}(e_{n(k)}) - f(x)|.$$

Since the last term tends to zero,

$$\limsup_{k\to\infty} |f(x_k) - f(x)| \leq \epsilon.$$

Since ϵ may be arbitrarily small, $f(x_k) \longrightarrow f(x)$, which shows that f is continuous. $\square$

Lemma 4.3 *If (X, d_X) is a complete and separable metric space, and the functions f_n are bounded uniformly in n, then conditions [A], [B], and [C] are all equivalent.*

Proof:

It remains to show that [C] $\Rightarrow$ [A] when (X, d_X) is complete and separable, and $\sup_{d\in D_n}\{|f_n(d)|\} \leq B$ for all n.

Suppose that $\{\mu_n\}$ converges to $\mu \in \mathcal{P}(X)$, where $\mu_n(X \setminus D_n) = 0$. Since X is complete and separable, Prohorov's theorem implies that $\{\mu_n\}$ is tight. Thus, given $\epsilon > 0$, there exists a compact $K_\epsilon \subset X$ such that $\mu_n(X \setminus K_\epsilon) < \epsilon$ for all n. With $f : X \longrightarrow \mathbb{R}$ as defined in Lemma 4.2,

$$\begin{aligned}\left|\int_X f_n(x)\mu_n(dx) - \int_X f(x)\mu(dx)\right| \leq& \\ \left|\int_{K_\epsilon} f_n(x) - f(x)\mu_n(dx)\right| &+ 2B\epsilon \\ &+ \left|\int_X f(x)\mu_n(dx) - \int_X f(x)\mu(dx)\right|.\end{aligned}$$

Condition [C] implies condition [D], a sort of uniform convergence on compact sets that entails that

$$\lim_{n\to\infty} \int_{K_\epsilon} |f_n(x) - f(x)|\, \mu_n(dx) = 0.$$

Therefore,

$$\limsup_{n\to\infty} \left|\int_X f_n(x)\mu_n(dx) - \int_X f(x)\mu(dx)\right| \leq 2B\epsilon.$$

Since ϵ is arbitrarily small, it follows that

$$\int_X f_n(x)\mu_n(dx) \longrightarrow \int_X f(x)\mu(dx). \qquad \blacksquare$$

4.3 Theorems

Let S and T be separable metric spaces. For each natural number n, let K_n be a Markov transition from S^n to T^n that satisfies the permutation condition (4.1). Let H_n be the transition from $\varepsilon_n(S^n)$ to $\mathcal{P}(T)$ that is induced by K_n, as defined in (4.2).

Theorem 4.1 *If a sequence of Markov transitions $\{K_n\}$ propagates chaos, then there exists a continuous function*

$$F : \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$$

such that, if $\varepsilon_n(\mathbf{s}_n) \longrightarrow p$ in $\mathcal{P}(S)$ with $\mathbf{s}_n \in S^n$, then

$$\left\{\widetilde{K}_n(\mathbf{s}_n, \cdot)\right\}_{n=1}^{\infty}$$

is $F(p)$-chaotic.

Proof:

The arguments of Lemma 4.1 and Lemma 4.2 will be adapted to prove this.

Take $\mathcal{P}(S)$ with one of the metrics for the weak topology to be the metric space (X, d_X) of those lemmas, and take $\varepsilon_n(S^n)$ to be D_n. For each bounded and continuous function $\phi \in C_b(\mathcal{P}(T))$ define the functions $\widehat{\phi}_n : \varepsilon_n(S^n) \longrightarrow \mathbb{R}$ by

$$\widehat{\phi}_n(\zeta) := \int_{\mathcal{P}(T)} \phi(\eta) H_n(\zeta, d\eta), \tag{4.7}$$

where H_n is as defined in (4.2). These functions ϕ_n will play the role of the functions f_n of the lemmas.

By hypothesis, $\{K_n\}$ propagates chaos. Proposition 4.1 therefore implies that whenever $\{\mu_n \in \mathcal{P}(\varepsilon_n(S^n))\}_{n=1}^{\infty}$ converges in $\mathcal{P}(\mathcal{P}(S))$ to $\delta(p)$, then

$$\int_{\varepsilon_n(S^n)} \widehat{\phi}_n(\zeta)\mu_n(d\zeta) \longrightarrow \phi(q) \tag{4.8}$$

for some $q \in \mathcal{P}(T)$. In fact, Proposition 4.1 implies that q does not depend on our choice of ϕ: the same q works for all ϕ in (4.8).

Condition (4.8) resembles condition [B] of Lemma 4.1. Lemma 4.1 and Lemma 4.2 can now be applied to show that there exists a continuous function $G_\phi(p)$, depending on ϕ, such that if $\{\mathbf{s}_n \in S^n\}$ is a sequence satisfying $\varepsilon_n(\mathbf{s}_n) \longrightarrow p$ in $\mathcal{P}(S)$, then

$$\widehat{\phi}_n(\varepsilon_n(\mathbf{s}_n)) \longrightarrow G_\phi(p).$$

By (4.8), $G_\phi(p) = \phi(q)$ for some $q \in \mathcal{P}(T)$ *that does not depend on* ϕ. The only way that all the G_ϕ's can have this form and yet all be continuous is for the dependence of q on p to be continuous: there must be a continuous F from $\mathcal{P}(S)$ to $\mathcal{P}(T)$ such that $G_\phi(p) = \phi(F(p))$ for all $\phi \in C_b(\mathcal{P}(T))$.

Thus, there exists a continuous function F from $\mathcal{P}(S)$ to $\mathcal{P}(T)$ such that

$$[\varepsilon_n(\mathbf{s}_n) \longrightarrow p] \implies \left[\widehat{\phi}_n(\varepsilon_n(\mathbf{s}_n)) \longrightarrow \phi(F(p))\right]$$

for all $\phi \in C_b(\mathcal{P}(T))$. This fact, and the definitions (4.2) and (4.7) of H_n and $\widehat{\phi}$, imply that

$$[\varepsilon_n(\mathbf{s}_n) \longrightarrow p] \implies \left[\widetilde{K}_n(\mathbf{s}_n, \cdot) \circ \varepsilon_n^{-1} \longrightarrow \delta(F(p))\right].$$

Finally, by Theorem 3.2, we have that

$$[\varepsilon_n(\mathbf{s}_n) \longrightarrow p] \implies \left\{\widetilde{K}_n(\mathbf{s}_n, \cdot)\right\}_{n=1}^{\infty} \text{ is } F(p)\text{−chaotic.} \qquad \blacksquare$$

When (S, d_S) is complete and separable, the necessary condition of Theorem 4.1 is also *sufficient*.

Theorem 4.2 (Main Theorem) *Suppose (S, d_S) is a complete, separable metric space. Then $\{K_n\}$ propagates chaos if and only if there exists a continuous function*

$$F : \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$$

such that, whenever $\varepsilon_n(\mathbf{s}_n) \longrightarrow p$ in $\mathcal{P}(S)$ with $\mathbf{s}_n \in S^n$, then

$$\left\{\widetilde{K}_n(\mathbf{s}_n, \cdot)\right\}_{n=1}^{\infty}$$

is $F(p)$-chaotic.

Proof:

We have just demonstrated that the condition is necessary (Theorem 4.1). Next we demonstrate its sufficiency:

Suppose $P_n \in \mathcal{P}(S^n)$ is p-chaotic. Let $\mu_n = P_n \circ \varepsilon_n^{-1}$. Then $\mu_n \longrightarrow \delta(p)$ in $\mathcal{P}(\mathcal{P}(S))$ by Theorem 3.2. Our goal is to prove that

$$\int_{\mathcal{P}(S)} H_n(\zeta, d\eta)\mu_n(d\zeta) \longrightarrow \delta(F(p)),$$

where H_n is as defined in (4.2). This is enough, by Proposition 4.1, to demonstrate that chaos propagates.

By hypothesis, if $\{\mathbf{s}_n \in S^n\}$ is such that $\varepsilon_n(\mathbf{s}_n)$ converges to p then $\{\widetilde{K}_n(\mathbf{s}_n,\cdot)\}$ is $F(p)$-chaotic. By Theorem 3.2 and the fact that

$$\widetilde{K}_n(\mathbf{s}_n,\cdot)\circ\varepsilon_n^{-1} = H_n(\varepsilon_n(\mathbf{s}_n),\cdot),$$

the hypothesis is equivalent to the statement that, if $p_n \in \varepsilon_n(S^n)$ for each n, then

$$[p_n \longrightarrow p] \implies [H_n(p_n,\cdot) \longrightarrow \delta(F(p))]\,. \tag{4.9}$$

Let $\phi \in C_b(\mathcal{P}(T))$ be a bounded and continuous function on $\mathcal{P}(T)$, and define functions $\widehat{\phi}_n : \varepsilon_n(S^n) \longrightarrow \mathbb{R}$ by

$$\widehat{\phi}_n(\zeta) := \int_{\mathcal{P}(T)} \phi(\eta) H_n(\zeta, d\eta). \tag{4.10}$$

These functions are uniformly bounded in n since ϕ is bounded.

The hypothesis (4.9) and equation (4.10) imply that

$$\lim_{n\to\infty} \widehat{\phi}_n(p_n) = \phi(F(p)) \tag{4.11}$$

when $p_n \longrightarrow p$ with $p_n \in \varepsilon_n(S^n)$. We are assuming S is complete and separable, therefore so is $\mathcal{P}(S)$ [1].

We may now apply Lemma 4.3 with

$$X = \mathcal{P}(S), \quad D_n = \varepsilon_n(S^n), \quad f_n = \widehat{\phi}_n,$$

and conclude that

$$\lim_{n\to\infty} \int_{\mathcal{P}(S)} \widehat{\phi}_n(\zeta)\mu_n(d\zeta) = \phi(F(p)) \tag{4.12}$$

for any sequence $\{\mu_n\}$ that converges to $\delta(p)$ in $\mathcal{P}(\mathcal{P}(S))$.

By equations (4.12) and (4.10),

$$\begin{aligned}
\phi(F(p)) &= \lim_{n\to\infty} \int_{\mathcal{P}(S)} \widehat{\phi}_n(\zeta)\mu_n(d\zeta) \\
&= \lim_{n\to\infty} \int_{\mathcal{P}(S)} \int_{\mathcal{P}(T)} \phi(\eta) H_n(\zeta, d\eta)\mu_n(d\zeta) \\
&= \lim_{n\to\infty} \int_{\mathcal{P}(T)} \phi(\eta) \int_{\mathcal{P}(S)} H_n(\zeta, d\eta)\mu_n(d\zeta),
\end{aligned}$$

for all $\phi \in C_b(\mathcal{P}(T))$. This implies that

$$\int_{\mathcal{P}(S)} H_n(\zeta, d\eta)\mu_n(d\zeta) \longrightarrow \delta(F(p))$$

in $\mathcal{P}(\mathcal{P}(S))$, completing the proof. $\square$

Theorem 4.1 states that the *limit-law map* $F : \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$ must be continuous. If S is complete and separable then, conversely, any continuous map F is a possible limit-law map. This fact is a corollary of Theorem 4.2:

Corollary 4.1 *Suppose (S, d_S) is a complete, separable metric space. Then, for any continuous F : $\mathcal{P}(S) \longrightarrow \mathcal{P}(T)$, there exists a sequence of Markov transitions $\{K_n\}_{n=1}^{\infty}$ that propagates chaos, and for which*

$$\{K_n(\mathbf{s}_n, \cdot)\}_{n=1}^{\infty}$$

is $F(p)$-chaotic whenever $\varepsilon_n(\mathbf{s}_n) \longrightarrow p$ in $\mathcal{P}(S)$.

Proof:

Let $F : \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$ be continuous. For each n and each $\mathbf{s} \in S_n$, let $K_n(\mathbf{s}, \cdot)$ be the n-fold product measure

$$K_n(\mathbf{s}, \cdot) := F(\varepsilon_n(\mathbf{s})) \otimes F(\varepsilon_n(\mathbf{s})) \otimes \cdots \otimes F(e_n(\mathbf{s})). \tag{4.13}$$

Suppose the points $\mathbf{s}_n \in S^n$ are such that $\varepsilon_n(\mathbf{s}_n)$ converges to p as n tends to infinity. Since F is continuous , $F(\varepsilon_n(\mathbf{s}_n))$ converges to $F(p)$ as well, so it is clear from Definition 3.2 that the sequence of symmetric measures $\{K_n(\mathbf{s}_n, \cdot)\}$ is $F(p)$-chaotic. By Theorem 4.2, $\{K_n\}$ propagates chaos. □

The Markov transition functions $\{K_n\}$ may well be *deterministic*, that is, the n-particle dynamics may simply be given by a point-transformation from S^n to T^n. These point-transformations are measurable maps from S^n to T^n that commute with permutations of coordinates.

Let $f_n : S^n \longrightarrow T^n$ be a measurable map that commutes with permutations of n-coordinates, i.e., such that

$$f_n(s_{\pi(1)}, s_{\pi(2)}, \ldots, s_{\pi(n)}) = \pi \cdot f_n(s_1, s_2, \ldots, s_n) \tag{4.14}$$

for each point $\mathbf{s} \in S^n$ and each permutation π of the symbols $1, 2, \ldots, n$. Given f_n, define the Markov transition K_n from S^n to T^n by

$$K_n(\mathbf{s}, E) = \mathbf{1}_E(f_n(\mathbf{s}))$$

when $\mathbf{s} \in S^n$ and $E \in \mathcal{B}_{T^n}$. Say that $\{f_n\}_{n=1}^{\infty}$ *propagates chaos* if the sequence of deterministic transition functions $\{K_n\}$ propagates chaos.

The following is an immediate corollary of Theorem 4.2:

Corollary 4.2 (Deterministic Case) *Let S be a Polish space, and for each n let f_n be a measurable map from S^n to T^n that commutes with permutations as in (4.14).*

$\{f_n\}$ propagates chaos if and only if there exists a continuous function

$$F : \mathcal{P}(S) \longrightarrow \mathcal{P}(T)$$

such that $\varepsilon_n(f_n(\mathbf{s}_n)) \longrightarrow F(p)$ *in* $\mathcal{P}(T)$ *whenever* $\varepsilon_n(\mathbf{s}_n) \longrightarrow p$ *in* $\mathcal{P}(S)$.

Chapter 5

Conclusion

We have studied the propagation of chaos by families of Markov processes, having adopted a simple definition of propagation of chaos, namely, that the processes propagate *all* chaotic sequences of initial laws to chaotic sequences. Authors who wish to prove that certain families of processes propagate chaos often show only that sequences of initial laws of the form $\rho^{\otimes n}$ are propagated to chaotic sequences, that is, they show that pure chaos is propagated to chaos. We have remarked that this does not imply unqualified propagation of chaos.

Propagation of chaos, in its unqualified sense, entails the continuity of the limit dynamics. Families of Markov processes on Polish spaces propagate chaos if and only if the associated Markov transition functions satisfy the condition of Theorem 4.2.

Our definition of propagation of chaos may be too simplistic to cover some situations of interest. For instance, the subtle propagation of chaos phenomenon that is operative in Lanford's validation of Boltzmann's equation — where the chaos of the initial laws propagates if those laws have densities that converge uniformly — is not subject to our treatment here.

Further foundational research on the propagation of chaos phenomenon of Lanford's theorem is called for. Is the phenomenon endemic to the Boltzmann-Grad limit, or is it, like the propagation of chaos that is the subject of our theorems, a more general probabilistic phenomenon that should appear in other parts of kinetic theory?

Bibliography

[1] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, New York, 1968.

[2] G. A. Bird. *Molecular Gas Dynamics.* Clarendon Press, Oxford, 1976.

[3] L. Boltzmann. *Lectures on Gas Theory.* Dover Publications, New York, 1995.

[4] M. Bossy and D. Talay. A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Mathematics of Computation* 66 (217): 157-192, 1997.

[5] W. Braun and K. Hepp. The Vlasov dynamics and its fluctuations in the $\frac{1}{n}$ limit of interacting classical particles. *Communications in Mathematical Physics* 56: 101-113, 1977.

[6] A. J. Chorin. Numerical study of slightly viscous flow. *Journal of Fluid Mechanics* 57: 785-796, 1973.

[7] C. Cercignani, R. Illner, and M. Pulvirenti. *The Mathematical Theory of Dilute Gases.* Springer-Verlag, New York, 1994.

[8] D. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics* 20: 247-308, 1987.

[9] R. Di Perna and P.L. Lions. On the Cauchy problem for the Boltzmann equation. *Annals of Mathematics* 130: 321-366, 1989.

[10] R. M. Dudley. *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Pacific Grove, California, 1989.

[11] P. Dupuis and R. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* John Wiley & Sons, New York, 1997.

[12] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence.* John Wiley & Sons, New York, 1986.

[13] H. Grad. On the kinetic theory of rarefied gases. *Communications in Pure and Applied Mathematics* 2: 331-407, 1949.

[14] C. Graham. McKean-Vlasov Ito-Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stochastic Processes and their Applications* 40: 69-82, 1992.

[15] F. A. Grünbaum. Propagation of chaos for the Boltzmann equation. *Archive for Rational Mechanics and Analysis* 42: 323-345, 1971.

[16] M. Kac. *Probability and Related Topics in Physical Sciences.* American Mathematical Society, Providence, Rhode Island, 1976.

[17] M. Kac. Foundations of kinetic theory. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol III.* University of California Press, Berkeley, California, 1956.

[18] O.E. Lanford III. The evolution of large classical systems. *Lecture Notes in Physics*, 35: 1-111. Springer-Verlag, Berlin, 1975.

[19] L. Le Cam. *Convergence in Distribution of Stochastic Processes.* University of California Publications in Statistics. 2 (11): 207-236, 1957.

[20] C. Marchioro and M. Pulvirenti. Hydrodynamics in two dimensions and vortex theory. *Communications in Mathematical Physics* 84: 483-503, 1982.

[21] H. P. McKean, Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Science* 56: 1907-1911, 1966.

[22] H. P. McKean, Jr. Propagation of chaos for a class of nonlinear parabolic equations. *Lecture Series in Differential Equations* 7: 41-57. Catholic University, Washington, D.C., 1967.

[23] H. P. McKean, Jr. Fluctuations in the kinetic theory of gases. *Communications in Pure and Applied Mathematics* 28: 435-455, 1975.

[24] S. Méléard. Asymptotic behavior of some interacting particle systems; McKean-Vlasov and Boltzmann models. *Lecture Notes in Mathematics*, 1627. Springer-Verlag, Berlin, 1995.

[25] S. Méléard. A probabilistic proof of the vortex method for the 2D Navier-Stokes equations. Preprint.

[26] K. Oelschläger. A law of large numbers for moderately interacting diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 69: 279-322, 1985.

[27] H. Osada. Propagation of chaos for the two dimensional Navier-Stokes equation. *Probabilistic Methods in Mathematical Physics.* Academic Press, Boston, 1987.

[28] Yu. V. Prohorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability and it Applications* 1: 157-214, 1956.

[29] M. Pulvirenti. Kinetic limits of stochastic particle systems. *Lecture Notes in Mathematics*, 1627. Springer-Verlag, Berlin, 1995.

[30] F. Rezakhanlou. Kinetic limits for a class of interacting particle systems. *Probability Theory and Related Fields* 104: 97-146, 1996.

[31] F. Rezakhanlou and J. Tarver. Boltzmann-Grad limit for a particle system in continuum. *Annales de l'Institut Henri Poincaré Probab. Stat.* 33 (6): 753-796, 1997.

[32] T. Shiga and H. Tanaka. Central limit theorem for a Markovian system of particles in mean field interaction. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 69: 439-459, 1985.

[33] H. Spohn. *Large Scale Dynamics of Interacting Particles.* Springer-Verlag, Berlin, 1991.

[34] A. Sznitman. Équations de type de Boltzmann, spatialement homogènes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 66: 559-592, 1984.

[35] A. Sznitman. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of Functional Analysis* 56: 311-336, 1984.

[36] A. Sznitman. Topics in propagation of chaos. *Lecture Notes in Mathematics*, 1464. Springer-Verlag, Berlin, 1991.

[37] D. Talay. Probabilistic numerical methods for partial differential equations: elements of analysis. *Lecture Notes in Mathematics*, 1627. Springer-Verlag, Berlin, 1995.

[38] H. Tanaka. Limit theorems for certain diffusion processes with interaction. *Taniguchi Symposium on Stochastic Analysis* (pp. 469-488). Katata, Kyoto, 1982.

[39] K. Uchiyama. Fluctuations of Markovian systems in Kac's caricature of a Maxwellian gas. *Journal of the Mathematical Society of Japan* 35 (3): 477-499, 1983.

[40] K. Uchiyama. Fluctuations in a Markovian system of pairwise interacting particles. *Probability Theory and Related Fields* 79: 289-302, 1988.