

A lecture on the averaging process *

David Aldous[†]

Department of Statistics, 367 Evans Hall # 3860, U.C. Berkeley CA 94720
e-mail: aldous@stat.berkeley.edu url: www.stat.berkeley.edu

and

Daniel Lanoue

Department of Mathematics, 970 Evans Hall # 3840, U.C. Berkeley, CA 94720
e-mail: dlanoue@math.berkeley.edu

Abstract: To interpret interacting particle system style models as social dynamics, suppose each pair $\{i, j\}$ of individuals in a finite population meet at random times of arbitrary specified rates ν_{ij} , and update their states according to some specified rule. The *averaging process* has real-valued states and the rule: upon meeting the values $X_i(t-), X_j(t-)$ are replaced by $\frac{1}{2}(X_i(t-) + X_j(t-)), \frac{1}{2}(X_i(t-) + X_j(t-))$. It is curious this simple process has not been studied very systematically. We provide an expository account of basic facts and open problems.

AMS 2000 subject classifications: Primary 60K35; secondary 60K99.
Keywords and phrases: duality, interacting particle systems, rate of convergence, spectral gap, voter model.

1. Introduction

The models in the field known to mathematical probabilists as *interacting particle systems (IPS)* are often exemplified, under the continuing influence of Liggett [6], by the voter model, contact process, exclusion process, and Glauber dynamics for the Ising model, all on the infinite d -dimensional lattice, and a major motivation for the original development of the field was as rigorous study of phase transitions in the discipline of statistical physics. Models with similar mathematical structure have long been used as toy models in many other disciplines, in particular in *social dynamics*, as models for the spread of opinions, behaviors or knowledge between individuals in a society. In this context the nearest neighbor lattice model for contacts between individuals is hardly plausible, and because one has a finite number of individuals, the finite-time behavior is often more relevant than time-asymptotics. So the context is loosely analogous to the study of mixing times for finite Markov chains (Levin-Peres-Wilmer [5]). A general mathematical framework for this viewpoint, outlined below, was explored in a Spring 2011 course by the first author, available as informal beamer slides [2]. In this expository article (based on two lectures from that course) we treat a simple model where the “states” of an individual are continuous.

*This is an original expository article.

[†]Research supported by N.S.F Grant DMS-0704159

1.1. The general framework and the averaging process

Consider n agents (interpret as individuals) and a nonnegative array (ν_{ij}) , indexed by unordered pairs $\{i, j\}$, which is irreducible (i.e. the graph of edges corresponding to strictly positive entries is connected). Interpret ν_{ij} as the “strength of relationship” between agents i and j . Assume

- Each unordered pair i, j of agents with $\nu_{ij} > 0$ meets at the times of a rate- ν_{ij} Poisson process, independent for different pairs.

Call this collection of Poisson processes the *meeting process*. In the general framework, each agent i is in some state $X_i(t)$ at time t , and the state can only change when agent i meets some other agent j , at which time the two agents’ states are updated according to some rule (deterministic or random) which depends only on the pre-meeting states $X_i(t-), X_j(t-)$. In the averaging process the states are the real numbers \mathbb{R} and the rule is

$$(X_i(t+), X_j(t+)) = \left(\frac{1}{2}(X_i(t-) + X_j(t-)), \frac{1}{2}(X_i(t-) + X_j(t-))\right).$$

A natural interpretation of the state $X_i(t)$ is as the amount of money agent i has at time t ; the model says that when agents meet they share their money equally.

A mathematician’s reaction to this model might be “obviously the individual values $X_i(t)$ converge to the average of initial values, so what is there to say?”, and this reaction may explain the comparative lack of mathematical literature on the model. Here’s what we will say.

- If the initial configuration is a probability distribution (i.e. unit money split unevenly between individuals) then the vector of expectations in the averaging process evolves precisely as the probability distribution of an associated (continuous-time) Markov chain with that initial distribution (Lemma 1).
- There is an explicit bound on the closeness of the time- t configuration to the limit constant configuration (Proposition 2).
- Complementary to this global bound there is a “universal” (i.e. not depending on the meeting rates) bound on an appropriately defined *local* roughness of the time- t configuration (Proposition 4).
- There is a duality relationship with coupled Markov chains (section 2.4).

To an expert in IPS these four observations will be more or less obvious, and three are at least implicit in the literature, so we are not claiming any essential novelty. Instead, our purpose is to suggest using this model as an expository introduction to the topic of these social dynamics models for an audience not previously familiar with IPS but familiar with the theory of finite-state Markov chains (which will always mean *continuous time* chains). This “suggested course material” occupies section 2. In a course one could then continue to the voter model (which has rather analogous structure) and comment on similarities and differences in mathematical structure and behavior for models based on other rules. We include some such “comments for students” here.

As a side benefit of the averaging process being simple but little-studied and analogous to the well-studied voter model, one can easily suggest small research projects for students to work on. Some projects are given as open problems in section 3, and the solution to one (obtaining a bound on entropy distance from the limit) is written out in section 3.4.

1.2. Related literature

The techniques we use have been known in IPS for a very long time, so it is curious that the only literature we know that deals explicitly with models like the averaging process is comparatively recent. Two such lines of research are mentioned below, and are suggested further reading for students. Because more complicated IPS models have been studied for a very long time, we strongly suspect that the results here are at least implicit in older work. But the authors do not claim authoritative knowledge of IPS.

Shah [8] provides a survey of *gossip algorithms*. His Theorems 4.1 and 5.1 are stated very differently, but upon inspection the central idea of the proof is the discrete-time analog of our Proposition 2, proved in the same way.

Acemoglu et al [1] treat a model where some agents have fixed (quantitative) opinions and the other agents update according to an averaging process. They appeal to the duality argument in our section 2.4, as well as a “coupling from the past” argument, to study the asymptotic behavior.

2. Basic properties of the averaging process

Write $I = \{i, j \dots\}$ for the set of agents and $n \geq 2$ for the number of agents. Recall that the array of non-negative meeting rates $\nu_{\{i,j\}}$ for unordered pairs $\{i, j\}$ is assumed to be irreducible. We can rewrite the array as the symmetric matrix $\mathcal{N} = (\nu_{ij})$ in which

$$\nu_{ij} = \nu_{\{i,j\}}, j \neq i; \quad \nu_{ii} = -1. \quad (2.1)$$

Then \mathcal{N} is the generator of the Markov chain with transition rates ν_{ij} ; call this the *associated Markov chain*. The chain is reversible with uniform stationary distribution.

Comment for students. The associated Markov chain is also relevant to the analysis of various social dynamics models other than the averaging process.

Throughout, we write $\mathbf{X}(t) = (X_i(t), i \in I)$ for the averaging process run from some non-random initial configuration $\mathbf{x}(0)$. Of course the sum is conserved: $\sum_i X_i(t) = \sum_i x_i(0)$.

2.1. Relation with the associated Markov chain

We note first a simple relation with the associated Markov chain. Write $\mathbf{1}_i$ for the initial configuration $(1_{(j=i)}, j \in I)$ – agent i has unit money and other agents

have none – and write $p_{ij}(t)$ for the transition probabilities for the associated Markov chain.

Lemma 1. *For the averaging process with initial configuration $\mathbf{1}_i$ we have $\mathbb{E}X_j(t) = p_{ij}(t/2)$. More generally, from any deterministic initial configuration $\mathbf{x}(0)$, the expectations $\mathbf{x}(t) := \mathbb{E}\mathbf{X}(t)$ evolve exactly as the dynamical system*

$$\frac{d}{dt}\mathbf{x}(t) = \frac{1}{2}\mathbf{x}(t)\mathcal{N}.$$

The time- t distribution $\mathbf{p}(t)$ of the associated Markov chain evolves as $\frac{d}{dt}\mathbf{p}(t) = \mathbf{p}(t)\mathcal{N}$. So if $\mathbf{x}(0)$ is a probability distribution over agents, then the *expectation* of the averaging process evolves as the distribution of the associated Markov chain started with distribution $\mathbf{x}(0)$ and slowed down by factor $1/2$. But keep in mind that the averaging process has more structure than this associated chain.

Proof. The key point is that we can rephrase the dynamics of the averaging process as

when two agents meet, each gives half their money to the other.

In informal language, this implies that the motion of a random penny - which at a meeting of its owner agent is given to the other agent with probability $1/2$ - is as the associated Markov chain at half speed, that is with transition rates $\nu_{ij}/2$.

To say this in symbols, we augment a random partition $\mathbf{X} = (X_i)$ of unit money over agents i by also recording the position U of the “random penny”, required to satisfy

$$\mathbb{P}(U = i|\mathbf{X}) = X_i.$$

Given a configuration \mathbf{x} and an edge e , write \mathbf{x}^e for the configuration of the averaging process after a meeting of the agents comprising edge e . So we can define the augmented averaging process to have transitions

$$\begin{aligned} (\mathbf{x}, u) &\rightarrow (\mathbf{x}^e, u) : \text{rate } \nu_e, & \text{if } u \notin e \\ (\mathbf{x}, u) &\rightarrow (\mathbf{x}^e, u) : \text{rate } \nu_e/2, & \text{if } u \in e \\ (\mathbf{x}, u) &\rightarrow (\mathbf{x}^e, u') : \text{rate } \nu_e/2, & \text{if } u \in e = (u, u'). \end{aligned}$$

This defines a process $(\mathbf{X}(t), U(t))$ consistent with the averaging process and (intuitively at least – see below) satisfying

$$\mathbb{P}(U(t) = i|\mathbf{X}(t)) = X_i(t). \tag{2.2}$$

The latter implies $\mathbb{E}X_i(t) = \mathbb{P}(U(t) = i)$, and clearly $U(t)$ evolves as the associated Markov chain slowed down by factor $1/2$. This establishes the first assertion of the lemma. The case of a general initial configuration follows via the following *linearity property* of the averaging process. Writing $\mathbf{X}(\mathbf{y}, t)$ for the averaging process with initial configuration \mathbf{y} , one can couple these processes as \mathbf{y} varies by using the same realization of the underlying meeting process. Then clearly

$$\mathbf{y} \rightarrow \mathbf{X}(\mathbf{y}, t) \text{ is linear.}$$

How one writes down a careful proof of (2.2) depends on one's taste for details. We can explicitly construct $U(t)$ in terms of “keep or give” events at each meeting, and pass to the embedded jump chain of the meeting process, in which time m is the time of the m 'th meeting and \mathcal{F}_m its natural filtration. Then on the event that the m 'th meeting involves i and j ,

$$\begin{aligned}\mathbb{P}(U(m) = i | \mathcal{F}_m) &= \frac{1}{2}\mathbb{P}(U(m-1) = i | \mathcal{F}_{m-1}) + \frac{1}{2}\mathbb{P}(U(m-1) = j | \mathcal{F}_{m-1}) \\ X_i(m) &= \frac{1}{2}X_i(m-1) + \frac{1}{2}X_j(m-1)\end{aligned}$$

and so inductively we have

$$\mathbb{P}(U(m) = i | \mathcal{F}_m) = X_i(m)$$

as required. □

For a configuration \mathbf{x} , write $\bar{\mathbf{x}}$ for the “equalized” configuration in which each agent has the average $n^{-1} \sum_i x_i$. Lemma 1, and convergence in distribution of the associated Markov chain to its (uniform) stationary distribution, immediately imply $\mathbb{E}\mathbf{X}(t) \rightarrow \bar{\mathbf{x}}$ as $t \rightarrow \infty$. Amongst several ways one might proceed to argue that $\mathbf{X}(t)$ itself converges to $\bar{\mathbf{x}}$, the next leads to a natural explicit quantitative bound.

2.2. The global convergence theorem

Proposition 2 (The global convergence theorem). *From an initial configuration $\mathbf{x}(0) = (x_i)$ with average zero and standard deviation $\|\mathbf{x}(0)\|_2 := \sqrt{n^{-1} \sum_i x_i^2}$, the time- t configuration $\mathbf{X}(t)$ of the averaging process satisfies*

$$\mathbb{E}\|\mathbf{X}(t)\|_2 \leq \|\mathbf{x}(0)\|_2 \exp(-\lambda t/4), \quad 0 \leq t < \infty \quad (2.3)$$

where λ is the spectral gap of the associated MC.

Recall some background facts about reversible chains, here specialized to the case of uniform stationary distribution (that is, $\nu_{ij} = \nu_{ji}$) and in the continuous-time setting. See Chapter 3 of Aldous-Fill [3] for the theory surrounding (2.4) and Lemma 3.

A function $f: I \rightarrow \mathbb{R}$ has (with respect to the uniform distribution) average \bar{f} , variance $\text{var } f$ and L^2 norm $\|f\|_2$ defined by

$$\begin{aligned}\bar{f} &:= n^{-1} \sum_i f_i \\ \|f\|_2^2 &:= n^{-1} \sum_i f_i^2 \\ \text{var } f &:= \|f\|_2^2 - (\bar{f})^2.\end{aligned}$$

The associated Markov chain, with generator \mathcal{N} at (2.1), has *Dirichlet form*

$$\mathcal{E}(f, f) := \frac{1}{2}n^{-1} \sum_i \sum_{j \neq i} (f_i - f_j)^2 \nu_{ij} = n^{-1} \sum_{\{i,j\}} (f_i - f_j)^2 \nu_{ij}$$

where $\sum_{\{i,j\}}$ indicates summation over *unordered* pairs. The spectral gap of the chain, defined as the gap between eigenvalue 0 and the second eigenvalue of \mathcal{N} , is characterized as

$$\lambda = \inf_f \left\{ \frac{\mathcal{E}(f, f)}{\text{var}(f)} : \text{var}(f) \neq 0 \right\}. \quad (2.4)$$

Writing π for the uniform distribution on I , there is an L^2 distance from uniformity for probability measures ρ , defined as the L^2 norm of the function $(\rho_i - \pi_i)/\pi_i$ which we write as

$$\|\rho - \pi\|_{2(m)}^2 = -1 + n \sum_i \rho_i^2$$

the subscript (m) as a reminder the arguments here are probability measures, not functions.

Lemma 3 (L^2 contraction lemma). *The time- t distributions $\rho(t)$ of the associated Markov chain satisfy*

$$\|\rho(t) - \pi\|_{2(m)} \leq e^{-\lambda t} \|\rho(0) - \pi\|_{2(m)}.$$

This is optimal, in the sense that the rate of convergence really is $\Theta(e^{-\lambda t})$.

A few words about notation for process dynamics. We will write

$$\mathbb{E}(dZ(t)|\mathcal{F}(t)) = [\leq] Y(t)dt$$

to mean

$$Z(t) - Z(0) - \int_0^t Y(s)ds \text{ is a martingale [supermartingale],}$$

because we find the former “differential” notation much more intuitive than the integral notation. In the context of a social dynamics process we typically want to choose a functional Φ and study the process $\Phi(\mathbf{X}(t))$, and we write

$$\mathbb{E}(d\Phi(\mathbf{X}(t))|\mathbf{X}(t) = \mathbf{x}) = \phi(\mathbf{x})dt \quad (2.5)$$

so that $\mathbb{E}(d\Phi(\mathbf{X}(t))|\mathcal{F}(t)) = \phi(\mathbf{X}(t))dt$. We can immediately write down the expression for ϕ in terms of Φ and the dynamics of the particular process; for the averaging process,

$$\phi(\mathbf{x}) = \sum_{\{i,j\}} \nu_{ij} (\Phi(\mathbf{x}^{ij}) - \Phi(\mathbf{x})) \quad (2.6)$$

where \mathbf{x}^{ij} is the configuration obtained from \mathbf{x} after agents i and j meet and average. This is just saying that agents i, j meet during $[t, t + dt]$ with chance $\nu_{ij} dt$ and such a meeting changes $\Phi(\mathbf{X}(t))$ by the amount $\Phi(\mathbf{x}^{ij}) - \Phi(\mathbf{x})$.

Comment for students. In proofs we refer to (2.5, 2.6) as “the dynamics of the averaging process”. Everybody actually does the calculations this way, though some authors manage to disguise it in their writing.

Proof of Proposition 2. A configuration \mathbf{x} changes when some pair $\{x_i, x_j\}$ is replaced by the pair $\{(x_i + x_j)/2, (x_i + x_j)/2\}$, which preserves the average and reduces $\|\mathbf{x}\|_2^2$ by exactly $n^{-1}(x_j - x_i)^2/2$. So, writing $Q(t) := \|\mathbf{X}(t)\|_2^2$,

$$\begin{aligned} \mathbb{E}(dQ(t)|\mathbf{X}(t) = \mathbf{x}) &= - \sum_{\{i,j\}} \nu_{ij} \cdot n^{-1}(x_j - x_i)^2/2 dt \\ &= -\mathcal{E}(\mathbf{x}, \mathbf{x})/2 dt \\ &\leq -\lambda\|\mathbf{x}\|_2^2/2 dt. \end{aligned} \tag{2.7}$$

The first inequality is by the dynamics of the averaging process, the middle equality is just the definition of \mathcal{E} for the averaging process, and the final inequality is the extremal characterization

$$\lambda = \inf\{\mathcal{E}(g, g)/\|g\|_2^2 : \bar{g} = 0\}.$$

So we have shown

$$\mathbb{E}(dQ(t)|\mathcal{F}(t)) \leq -\lambda Q(t) dt/2.$$

The rest is routine. Take expectation:

$$\frac{d}{dt} \mathbb{E}Q(t) \leq -\lambda \mathbb{E}Q(t)/2$$

and then solve to get

$$\mathbb{E}Q(t) \leq \mathbb{E}Q(0) \exp(-\lambda t/2)$$

in other words

$$\mathbb{E}\|\mathbf{X}(t)\|_2^2 \leq \|\mathbf{x}\|_2^2 \exp(-\lambda t/2), \quad 0 \leq t < \infty.$$

Finally take the square root.

2.3. A local smoothness property

Thinking heuristically of the agents who agent i most frequently meets as the “local” agents for i , it is natural to guess that the configuration of the averaging process might become “locally smooth” faster than the “global smoothness” rate implied by Proposition 2. In this context we may regard the Dirichlet form $\mathcal{E}(f, f)$ as measuring the “local smoothness”, more accurately the local roughness, of a function f , relative to the local structure of the particular meeting process. The next result implicitly bounds $\mathbb{E}\mathcal{E}(\mathbf{X}(t), \mathbf{X}(t))$.

Proposition 4. For the averaging process with arbitrary initial configuration $\mathbf{x}(0)$,

$$\mathbb{E} \int_0^\infty \mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) dt = 2 \operatorname{var} \mathbf{x}(0).$$

This looks slightly magical because the bound does not depend on the particular rate matrix \mathcal{N} , but of course the definition of \mathcal{E} involves \mathcal{N} .

Proof. By linearity we may assume $\overline{\mathbf{x}(0)} = 0$. As in the proof of Proposition 2 consider $Q(t) := \|\mathbf{X}(t)\|_2^2$. Using (2.7)

$$\frac{d}{dt} \mathbb{E} Q(t) = \mathbb{E} \mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) / 2$$

and hence

$$\mathbb{E} \int_0^\infty \mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) dt = 2(Q(0) - Q(\infty)) = 2\|\mathbf{x}(0)\|_2^2 \quad (2.8)$$

because $Q(\infty) = 0$ by Proposition 2. □

2.4. Duality with coupled Markov chains

Comment for students. Notions of *duality* are one of the interesting and useful tools in classical IPS, and equally so in the social dynamics models we are studying. The duality between the voter model and coalescing chains is the simplest and most striking example. The relationship we develop here for the averaging model is less simple but perhaps more representative of the general style of duality relationships.

The technique we use is to extend the “random penny”/augmented process argument used in Lemma 1. Now there are two pennies, and at any meeting there are independent decisions to hold or pass each penny. The positions $(Z_1(t), Z_2(t))$ of the two pennies behave as the following MC on product space, which is a particular coupling of two copies of the (half-speed) associated MC. Here i, j, k denote distinct agents.

$$\begin{aligned} (i, j) \rightarrow (i, k) & : \text{rate } \frac{1}{2} \nu_{jk} \\ (i, j) \rightarrow (k, j) & : \text{rate } \frac{1}{2} \nu_{ik} \\ (i, j) \rightarrow (i, i) & : \text{rate } \frac{1}{4} \nu_{ij} \\ (i, j) \rightarrow (j, j) & : \text{rate } \frac{1}{4} \nu_{ij} \\ (i, j) \rightarrow (j, i) & : \text{rate } \frac{1}{4} \nu_{ij} \\ (i, i) \rightarrow (i, j) & : \text{rate } \frac{1}{4} \nu_{ij} \\ (i, i) \rightarrow (j, i) & : \text{rate } \frac{1}{4} \nu_{ij} \\ (i, i) \rightarrow (j, j) & : \text{rate } \frac{1}{4} \nu_{ij}. \end{aligned}$$

For comparison, for two independent chains the transitions $(i, j) \rightarrow (j, i)$ and $(i, i) \rightarrow (j, j)$ are impossible and in the other transitions above all the $1/4$ terms

become 1/2. Intuitively, in the coupling the pennies move independently except for moves involving an edge between them, in which case the asynchronous dynamics are replaced by synchronous dynamics.

Repeating the argument around (2.2) – an exercise for the dedicated student – gives the following result. Write $\mathbf{X}^a(t) = (\mathbf{X}_i^a(t))$ for the averaging process started from configuration $\mathbf{1}_a$.

Lemma 5 (The duality relation). *For each choice of a, b, i, j , not requiring distinctness,*

$$\mathbb{E}(X_i^a(t)X_j^b(t)) = \mathbb{P}(Z_1^{a,b}(t) = i, Z_2^{a,b}(t) = j)$$

where $(Z_1^{a,b}(t), Z_2^{a,b}(t))$ denotes the coupled process started from (a, b) .

By linearity the duality relation implies the following – apply $\sum_a \sum_b x_a(0)x_b(0)$ to both sides.

Corollary 6 (Cross-products in the averaging model). *For the averaging model $\mathbf{X}(t)$ started from a configuration $\mathbf{x}(0)$ which is a probability distribution over agents,*

$$\mathbb{E}(X_i(t)X_j(t)) = \mathbb{P}(Z_1(t) = i, Z_2(t) = j)$$

where $(Z_1(t), Z_2(t))$ denotes the coupled process started from random agents $(Z_1(0), Z_2(0))$ chosen independently from $\mathbf{x}(0)$.

3. Projects and open problems

The material in section 2 is our suggestion for what to say in a lecture course. It provides an introduction to some general techniques, and one can then move on to more interesting models. In teaching advanced graduate courses it is often hard to find “projects” – from homework problems to open problems whose solution might be worth a research paper – to engage students, but here it is quite easy to invent projects.

3.1. Foundational issues

There are several “foundational” issues, or details of rigor, we have skipped over, and to students who are sufficiently detail-oriented to notice we suggest they investigate the issue themselves.

- (i) Give a formal construction of the averaging process consistent with the process dynamics described in section 2.2.
- (ii) Relate this notation of duality to the abstract setup in Liggett [6] section 2.3.
- (iii) Give a careful proof of Lemma 5.

3.2. Homework-style problems

- (i) From the definition of the averaging process, give a convincing one sentence outline argument for $X_i(t) \rightarrow \bar{\mathbf{x}}(0)$ a.s. for each i .
- (ii) *Averaging process with noise.* This variant model can be described as

$$dX_i(t) = \sigma dW_i(t) + (\text{dynamics of averaging model})$$

where the “noise” processes $W_i(t)$ are defined as follows. First take n independent standard Normals conditioned on their sum equalling zero – call them $(W_i(1), 1 \leq i \leq n)$. Now take $\mathbf{W}(t)$ to be the n -dimensional Brownian motion associated with the time-1 distribution $\mathbf{W}(1) = (W_i(1), 1 \leq i \leq n)$. By modifying the proof of Proposition 2, show this process has a limit distribution $\mathbf{X}(\infty)$ such that

$$\mathbb{E} \|\mathbf{X}(\infty)\|_2^2 \leq \frac{2\sigma^2(n-1)}{\lambda n}.$$

- (iii) Give an example to show that $\mathbb{E}\mathcal{E}(\mathbf{X}(t), \mathbf{X}(t))$ is not necessarily a decreasing function of t .

Answers to (ii,iii) are outlined in [2].

3.3. Open problems

- (i) Can you obtain any improvement on Proposition 4? In particular, assuming a lower bound on meeting rates:

$$\sum_j \nu_{ij} \geq 1 \text{ for all } i$$

is it true that

$$\mathbb{E}\mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) \leq \phi(t) \text{var}(\mathbf{x}(0))$$

for some universal function $\phi(t) \downarrow 0$ as $t \uparrow \infty$?

- (ii) One can define the averaging process on the integers – that is, $\nu_{i,i+1} = 1, -\infty < i < \infty$ – started from the configuration with unit total mass, all at the origin. By Lemma 1 we have

$$\mathbb{E}X_j(t) = p_j(t)$$

where the right side is the time $-t$ distribution of a continuous-time simple symmetric random walk, which of course we understand very well. But what can you say about the second-order behavior of this averaging process? That is, how does $\text{var}(X_j(t))$ behave and what is the distributional limit of $(X_j(t) - p_j(t))/\sqrt{\text{var}(X_j(t))}$? Note that duality gives an expression for the variance in terms of the coupled random walks – the issue is how to analyze its asymptotics explicitly.

- (iii) The discrete cube $\{0, 1\}^d$ graph is a standard test bench for Markov chain related problems, and in particular its log-Sobolev constant is known [4]. Can you get stronger results for the averaging process on this cube than are implied by our general results?

3.4. Quantifying convergence via entropy

Parallel to Lemma 3 are quantifications of reversible Markov chain convergence in terms of the log-Sobolev constant of the chain, defined (cf. (2.4)) as

$$\alpha = \inf_f \left\{ \frac{\mathcal{E}(f, f)}{L(f)} : L(f) \neq 0 \right\}. \quad (3.1)$$

where

$$L(f) = n^{-1} \sum_i f_i^2 \log(f_i^2 / \|f\|_2^2).$$

See Montenegro and Tetali [7] for an overview, and Diaconis and Saloff-Coste [4] for more details of the theory, which we do not need here. One problem posed in the Spring 2011 course was to seek a parallel of Proposition 2 in which one quantifies closeness of $\mathbf{X}(t)$ to uniformity via entropy, using the log-Sobolev constant of the associated Markov chain in place of the spectral gap. Here is one solution to that problem.

For a configuration \mathbf{x} which is a probability distribution write

$$\text{Ent}(\mathbf{x}) := - \sum_i x_i \log x_i$$

for the entropy of the configuration (note [4] writes “entropy” to mean relative entropy w.r.t. the stationary measure). Consider the averaging process where the initial configuration is a probability distribution. By concavity of the function $-x \log x$ it is clear that in the averaging process $\text{Ent}(\mathbf{X}(t))$ can only increase, and hence $\text{Ent}(\mathbf{X}(t)) \uparrow \log n$ a.s. (recall $\log n$ is the entropy of the uniform distribution). So we want to bound $\mathbb{E}(\log n - \text{Ent}(\mathbf{X}(t)))$. For this purpose note that, for a configuration \mathbf{x} which is a probability distribution,

$$nL(\sqrt{\mathbf{x}}) = \log n - \text{Ent}(\mathbf{x}). \quad (3.2)$$

Proposition 7. *For the averaging process whose initial configuration is a probability distribution $\mathbf{x}(0)$,*

$$\mathbb{E}(\log n - \text{Ent}(\mathbf{X}(t))) \leq (\log n - \text{Ent}(\mathbf{x}(0))) \exp(-\alpha t/2)$$

where α is the log-Sobolev constant of the associated Markov chain.

Proof. For $0 < t < 1$ define

$$\begin{aligned} a(t) &:= \frac{1}{2} - \sqrt{t(1-t)} \geq 0 \\ b(t) &:= \log 2 + t \log t + (1-t) \log(1-t) \geq 0. \end{aligned}$$

The key inequality is

$$a(t) \leq \frac{1}{2}(2t-1)^2 \leq b(t). \quad (3.3)$$

For the left inequality note that for $|z| \leq 1$ we have $\sqrt{1-z^2} \geq 1-z^2$ and so

$$a(t) = \frac{1}{2}(1 - \sqrt{1 - (2t-1)^2}) \leq \frac{1}{2}(2t-1)^2.$$

For the right inequality first consider, for $-1 \leq z \leq 1$,

$$g(z) := (1+z) \log(1+z) + (1-z) \log(1-z) - z^2.$$

Then $g(0) = 0$, $g(-z) = g(z)$ and $g''(z) = \frac{2}{1-z^2} - 2 \geq 0$, implying $g(z) \geq 0$ throughout $-1 \leq z \leq 1$. So

$$\frac{1}{2}g(2t-1) \geq 0$$

and upon expansion this is the right inequality of (3.3).

So for $x_1, x_2 > 0$ we have from (3.3) that $(x_1 + x_2)b(\frac{x_1}{x_1+x_2}) \geq (x_1 + x_2)a(\frac{x_1}{x_1+x_2})$ and this inequality rearranges to

$$x_1 \log x_1 + x_2 \log x_2 - (x_1 + x_2) \log\left(\frac{x_1+x_2}{2}\right) \geq \left(\frac{x_1+x_2}{2} - \sqrt{x_1 x_2}\right). \quad (3.4)$$

For the rest of the proof we can closely follow the format of the proof of Proposition 2. A configuration \mathbf{x} changes when some pair $\{x_i, x_j\}$ is replaced by the pair $\{(x_i + x_j)/2, (x_i + x_j)/2\}$, which increases the entropy of the configuration by $-(x_i + x_j) \log(x_i + x_j)/2 + x_i \log x_i + x_j \log x_j$. So the dynamics of the averaging process give the first equality below.

$$\begin{aligned} \mathbb{E}(d\text{Ent}(\mathbf{X}(t)) | \mathbf{X}(t) = \mathbf{x}) &= \sum_{\{i,j\}} \nu_{ij} \cdot \left(-(x_i + x_j) \log(x_i + x_j)/2 + x_i \log x_i + x_j \log x_j \right) dt \\ &\geq \sum_{\{i,j\}} \nu_{ij} \cdot \left(\frac{x_i+x_j}{2} - \sqrt{x_i x_j} \right) dt \text{ by (3.4)} \\ &= \sum_{\{i,j\}} \nu_{ij} \cdot \frac{(\sqrt{x_i} - \sqrt{x_j})^2}{2} dt \\ &= \frac{n}{2} \mathcal{E}(\sqrt{\mathbf{x}}, \sqrt{\mathbf{x}}) dt \quad (\text{definition of } \mathcal{E}) \\ &\geq \frac{\alpha n}{2} L(\sqrt{\mathbf{x}}) dt \quad (\text{definition of } \alpha) \\ &= \frac{\alpha}{2} (\log n - \text{Ent}(\mathbf{x})) dt \text{ by (3.2)}. \end{aligned}$$

So $D(t) := \log n - \text{Ent}(\mathbf{X}(t))$ satisfies

$$\mathbb{E}(dD(t) | \mathcal{F}(t)) \leq \frac{-\alpha}{2} D(t) dt.$$

The rest of the proof exactly follows Proposition 2. □

References

- [1] ACEMOGLU, D., COMO, G., FAGNINI, F. AND OZDAGLAR, A. (2011). Opinion fluctuations and disagreement in social networks. <http://arxiv.org/abs/1009.2653>
- [2] ALDOUS, D. (2011). *Finite Markov Information-Exchange Processes*. Lecture Notes from Spring 2011. <http://www.stat.berkeley.edu/~aldous/260-FMIE/Lectures/index.html>.

- [3] ALDOUS, D. AND FILL, J. *Reversible Markov Chains and Random Walks on Graphs*. <http://www.stat.berkeley.edu/~aldous/RWG/book.html>
- [4] DIACONIS, P. AND SALOFF-COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112
- [5] LEVIN, D. A., PERES, Y. AND WILMER, E. L. (2009), *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937
- [6] LIGGETT, T. M. (1985). *Interacting Particle Systems*. Springer-Verlag, New York. MR776231
- [7] MONTENEGRO, R. AND TETALI, P. (2006). Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* **1** 1–121. MR2341319
- [8] SHAH, D. (2008). Gossip algorithms. *Foundations and Trends in Networking* **3** 1-125.