

## 2. The averaging process

David Aldous

June 19, 2012

Write-up published as “A Lecture on the Averaging Process” (with Dan Lanoue) in recent *Probability Surveys*. Intended as illustration of how one might teach this subject in a graduate course.

Background meeting model with rates  $(\nu_{ij})$ .

**Model: averaging process.**

Each agent initially has some amount of money;  
whenever agents meet they share their money equally.

$X_i(t)$  is the amount of money agent  $i$  has at time  $t$ .

Formally, the states are the real numbers  $\mathbb{R}$ ; initially  $X_i(0) = x_i$ , and the update rule, when agents  $i$  and  $j$  meet at  $t$ , is

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

Your immediate reaction to this model should be (cf. General Principle 1) “obviously the individual values  $X_i(t)$  converge to the average of initial values, so what is there to say?”.

**Exercise:** write a one-sentence outline proof that a post-first-year-grad student could easily turn into a complete proof.

Curiously, while this process has been used as an ingredient in more elaborate models, the only place it appears by itself is in some “gossip algorithms” literature which derives a version of the “global bound” later – see paper for citations.

We will show

- 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 the 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 1).
- Complementary to this global bound there is a “universal” (i.e. not depending on the meeting rates) bound for an appropriately defined *local* roughness of the time- $t$  configuration (Proposition 2).
- There is a duality relationship with coupled Markov chains (Lemma 3).
- Two open problems.

The analysis I will give parallels analysis of the well-known **voter model** – will compare and contrast later.

## 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

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

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

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:  $\sum_i X_i(t) = \sum_i x_i(0)$ .

## Relation with the associated Markov chain

Write  $\mathbf{1}_i$  for the initial configuration ( $1_{(j=i)}, j \in I$ ), that is agent  $i$  has unit money and other agents have none, and write  $p_{ij}(t)$  for the transition probabilities of the associated Markov chain.

### Lemma

*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

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

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

$$\mathbb{P}(U(t) = i \mid \mathbf{X}(t)) = X_i(t). \quad (2)$$

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

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



How one writes down a careful proof of (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  $j$ ,

$$\mathbb{P}(U(m) = i \mid \mathcal{F}_m) = \frac{1}{2}\mathbb{P}(U(m-1) = i \mid \mathcal{F}_{m-1}) + \frac{1}{2}\mathbb{P}(U(m-1) = j \mid \mathcal{F}_{m-1})$$

$$X_i(m) = \frac{1}{2}X_i(m-1) + \frac{1}{2}X_j(m-1)$$

and so inductively we have

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

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  $\mathbb{E}\mathbf{X}(t) \rightarrow \bar{\mathbf{x}}(0)$  as  $t \rightarrow \infty$ .

Amongst several ways one might proceed to argue that  $\mathbf{X}(t)$  itself converges to  $\bar{\mathbf{x}}(0)$ , the next leads to a natural explicit quantitative bound.

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  $L^2$  norm will be used in several different ways. For a possible time- $t$  configuration  $\mathbf{x}(t)$  of the averaging process, the quantity  $\|\mathbf{x}(t)\|_2$  is a number, and so the quantity  $\|\mathbf{X}(t)\|_2$  appearing in the proposition below is a random variable.

### Proposition (Global convergence theorem)

*From an initial configuration  $\mathbf{x}(0) = (x_i)$  with average zero, the time- $t$  configuration  $\mathbf{X}(t)$  of the averaging process satisfies*

$$\mathbb{E}\|\mathbf{X}(t)\|_2 \leq \|\mathbf{x}(0)\|_2 \exp(-\lambda t/4), \quad 0 \leq t < \infty \quad (3)$$

*where  $\lambda$  is the spectral gap of the associated MC.*

Before starting the proof let us 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 for the theory surrounding (4) and Lemma 2. The associated Markov chain, with generator  $\mathcal{N}$  at (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

$$\lambda = \inf_f \left\{ \frac{\mathcal{E}(f, f)}{\text{var}(f)} : \text{var}(f) \neq 0 \right\}. \quad (4)$$

Writing  $\pi$  for the uniform distribution on  $I$ , one can define a distance from uniformity for probability measures  $\rho$  to be the  $L^2$  norm of the function  $i \rightarrow \frac{\rho_i - \pi_i}{\pi_i}$ , and we write this distance in the equivalent form

$$d_2(\rho, \pi) = \left( -1 + n \sum_i \rho_i^2 \right)^{1/2}.$$

Recall result from Markov chain theory.

### Lemma ( $L^2$ contraction lemma)

*The time- $t$  distributions  $\rho(t)$  of the associated Markov chain satisfy*

$$d_2(\rho(t), \pi) \leq e^{-\lambda t} d_2(\rho(0), \pi)$$

*where  $\lambda$  is the spectral gap of the associated MC.*

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

We don't actually use this lemma, but our global convergence theorem for the averaging process is clearly analogous.

**Notation for FMIE process dynamics.** We will write

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

to mean

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

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

$$\mathbb{E}(d\Phi(\mathbf{X}(t)) \mid \mathbf{X}(t) = \mathbf{x}) = \phi(\mathbf{x})dt \quad (5)$$

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

$$\phi(\mathbf{x}) = \sum_{\{i,j\}} \nu_{ij}(\Phi(\mathbf{x}^{ij}) - \Phi(\mathbf{x})) \quad (6)$$

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  $\Phi(\mathbf{x}^{ij}) - \Phi(\mathbf{x})$ .

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

$$\begin{aligned} \mathbb{E}(dQ(t) \mid \mathbf{X}(t) = \mathbf{x}) &= - \sum_{\{i,j\}} \nu_{ij} \cdot n^{-1} (x_j - x_i)^2 / 2 \, dt \\ &= -\mathcal{E}(\mathbf{x}, \mathbf{x}) / 2 \, dt \\ &\leq -\lambda \|\mathbf{x}\|_2^2 / 2 \, dt. \end{aligned} \tag{7}$$

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

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

So we have shown

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

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

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

Finally take the square root.



## 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 1. In this context we may regard the Dirichlet form

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

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  $\mathbb{E}\mathcal{E}(\mathbf{X}(t), \mathbf{X}(t))$  at finite times by giving an explicit bound for the integral over  $0 \leq t < \infty$ . Note that, from the fact that the spectral gap is strictly positive, we can see directly that  $\mathbb{E}\mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) \rightarrow 0$  exponentially fast as  $t \rightarrow \infty$ ; Proposition 2 is a complementary non-asymptotic result.

## Proposition

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  $\mathcal{N}$ . Can regard as instance of “General principle 3”.

**Proof.** By linearity we may assume  $\overline{\mathbf{x}(0)} = 0$ . As in the proof of Proposition 1 consider  $Q(t) := \|\mathbf{X}(t)\|_2^2$ . Using (7)

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

and hence

$$\mathbb{E} \int_0^\infty \mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) dt = 2(Q(0) - Q(\infty)) = 2\|\mathbf{x}(0)\|_2^2 \quad (8)$$

because  $Q(\infty) = 0$  by Proposition 1.

## General Principle 4: Duality

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 (recalled later) 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 (because of the continuous time setting) 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 partly replaced by synchronous ones.

Repeating the argument around (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 (The duality relation)

For each choice of  $a, b, i, j$ , not requiring distinctness, and for each  $t$ ,

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

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 (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, and for each  $t$ ,

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

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

**Open Problem.** 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, but the issue is to find an exact formula, or to somehow analyze asymptotics without an exact formula.

## Quantifying convergence via entropy

Parallel to Lemma 2 are quantifications of reversible Markov chain convergence in terms of the log-Sobolev constant of the chain, defined (cf. (4)) as

$$\alpha = \inf_f \left\{ \frac{\mathcal{E}(f, f)}{L(f)} : L(f) \neq 0 \right\}. \quad (9)$$

where

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

See Montenegro and Tetali (2006) for an overview, and Diaconis and Saloff-Coste (1996) 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 1 in which one quantifies closeness of  $\mathbf{X}(t)$  to uniformity via entropy, anticipating a bound in terms of 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. 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,

$$nL(\sqrt{\mathbf{x}}) = \log n - \text{Ent}(\mathbf{x}). \quad (10)$$

## Proposition

*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  $\alpha$  is the log-Sobolev constant of the associated Markov chain.*

The format closely parallels that of Proposition 1, though the proof is a little more intricate. See the paper for proof.

**Open Problem.** A standard test bench for Markov chain related problems is the discrete cube graph with vertex-set  $\{0, 1\}^d$  and rates  $\nu_{ij} = 1/d$  for adjacent vertices. In particular its log-Sobolev constant is known. Can you get stronger results for the averaging process on this cube than are implied by our general results?

I have shown all that's explicitly known about the averaging process itself, though more elaborate variant models have been studied.

We now move on to the voter model, which has a more substantial literature in the finite setting, so what's written here is far from complete. It would be a valuable project for someone to write a (50-page?) survey article.

## The voter model and coalescing MCs.

Here the update rule has a random (fair coin flip) component. Nicest to implement this within the meeting model via a “directed” convention: when agents  $i, j$  meet, choose a random direction and indicate it using an arrow  $i \rightarrow j$  or  $j \rightarrow i$ .

**Voter model.** Initially each agent has a different “opinion” – agent  $i$  has opinion  $i$ . When  $i$  and  $j$  meet at time  $t$  with direction  $i \rightarrow j$ , then agent  $j$  adopts the current opinion of agent  $i$ .

So we can study

$$\mathcal{V}_i(t) := \text{the set of } j \text{ who have opinion } i \text{ at time } t.$$

Note that  $\mathcal{V}_i(t)$  may be empty, or may be non-empty but not contain  $i$ . The number of different remaining opinions can only decrease with time.

**Minor comments.** (i) We can rephrase the rule as “agent  $i$  imposes his opinion on agent  $j$ ”.

(ii) The name is very badly chosen – people do not **vote** by changing their minds in any simple random way.

**Nuance.** In the classical, infinite lattice, setting one traditionally assumed only two different initial opinions. In our finite-agent case it seems more natural to take the initial opinions to be all different. Ultimate behavior is obvious (cf. General Principle 1): absorbed in one of the  $n$  “everyone has same opinion” configurations.

Note that one can treat the finite and infinite cases consistently by using IID  $U(0,1)$  opinion labels.

So  $\{\mathcal{V}_i(t), i \in \mathbf{Agents}\}$  is a random partition of **Agents**. A natural quantity of interest is the **consensus time**

$$T^{\text{voter}} := \min\{t : \mathcal{V}_i(t) = \mathbf{Agents} \text{ for some } i\}.$$

**Coalescing MC model.** Initially each agent has a token – agent  $i$  has token  $i$ . At time  $t$  each agent  $i$  has a (maybe empty) collection  $\mathcal{C}_i(t)$  of tokens. When  $i$  and  $j$  meet at time  $t$  with direction  $i \rightarrow j$ , then agent  $i$  gives his tokens to agent  $j$ ; that is,

$$\mathcal{C}_j(t+) = \mathcal{C}_j(t-) \cup \mathcal{C}_i(t-), \quad \mathcal{C}_i(t+) = \emptyset.$$

Now  $\{\mathcal{C}_i(t), i \in \mathbf{Agents}\}$  is a random partition of **Agents**. A natural quantity of interest is the **coalescence time**

$$T^{\text{coal}} := \min\{t : \mathcal{C}_i(t) = \mathbf{Agents} \text{ for some } i\}.$$

**Minor comments.** Regarding each non-empty cluster as a particle, each particle moves as the MC at half-speed (rates  $\nu_{ij}/2$ ), moving independently until two particles meet and thereby coalesce. Note this factor 1/2 in this section.

## The duality relationship.

For fixed  $t$ ,

$$\{\mathcal{V}_i(t), i \in \mathbf{Agents}\} \stackrel{d}{=} \{\mathcal{C}_i(t), i \in \mathbf{Agents}\}.$$

In particular  $\mathcal{T}^{\text{voter}} \stackrel{d}{=} \mathcal{T}^{\text{coal}}$ .

They are different as processes. For fixed  $i$ , note that  $|\mathcal{V}_i(t)|$  can only change by  $\pm 1$ , but  $|\mathcal{C}_i(t)|$  jumps to and from 0.

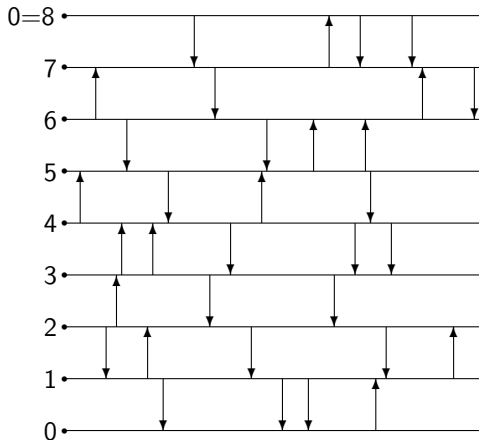
In figures on next slides,

time “left-to-right” gives CMC,

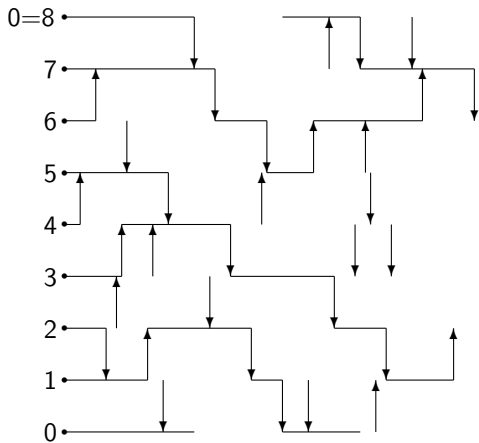
time “right-to-left” with reversed arrows gives VM.

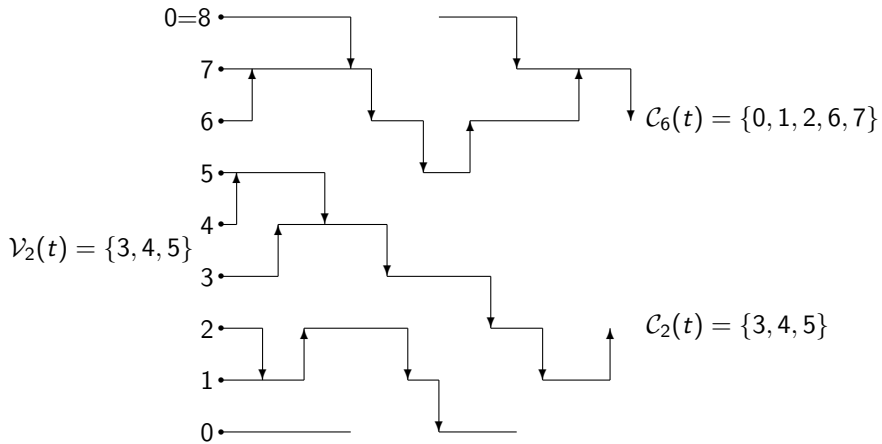
Note this depends on the symmetry assumption  $\nu_{ij} = \nu_{ji}$  of the meeting process.

# Schematic – the meeting model on the 8-cycle.









Literature on finite voter model has focussed on estimating  $\mathcal{T}^{\text{voter}} \stackrel{d}{=} \mathcal{T}^{\text{coal}}$ , and I will show some of this work.

But there are several other questions one can ask about the finite-time behavior . . . . .

## Voter model on the complete graph

There are two ways to analyze  $T_n^{\text{voter}}$  on the complete graph, both providing some bounds on other geometries.

Part of **Kingman's coalescent** is the continuous-time MC on states  $\{1, 2, 3, \dots\}$  with rates  $\lambda_{k,k-1} = \binom{k}{2}$ ,  $k \geq 2$ . For that chain

$$\mathbb{E}_m T_1^{\text{hit}} = \sum_{k=2}^m 1 / \binom{k}{2} = 2(1 - \frac{1}{m})$$

and in particular  $\lim_{m \rightarrow \infty} \mathbb{E}_m T_1^{\text{hit}} = 2$ .

In coalescing RW on the complete  $n$ -graph, the number of clusters evolves as the continuous-time MC on states  $\{1, 2, 3, \dots, n\}$  with rates  $\lambda_{k,k-1} = \frac{1}{n-1} \binom{k}{2}$ . So  $\mathbb{E} T_n^{\text{coal}} = (n-1) \times 2(1 - \frac{1}{n})$  and in particular

$$\mathbb{E} T_n^{\text{voter}} = \mathbb{E} T_n^{\text{coal}} \sim 2n. \quad (11)$$

The second way is to consider the variant of the voter model with only 2 opinions, and to study the number  $X(t)$  of agents with the first opinion. On the complete  $n$ -graph,  $X(t)$  evolves as the continuous-time MC on states  $\{0, 1, 2, \dots, n\}$  with rates

$$\lambda_{k,k+1} = \lambda_{k,k-1} = \frac{k(n-k)}{2(n-1)}.$$

This process arises in classical applied probability (e.g. as the Moran model in population genetics). We want to study

$$T_{0,n}^{\text{hit}} := \min\{t : X(t) = 0 \text{ or } n\}.$$

By general birth-and-death formulas, or by comparison with simple RW,

$$\mathbb{E}_k T_{0,n}^{\text{hit}} = \frac{2(n-1)}{n} (k(h_{n-1} - h_{k+1}) + (n-k)(h_{n-1} - h_{n-k+1}))$$

where  $h_m := \sum_{i=1}^m 1/i$ . This is maximized by  $k = \lfloor n/2 \rfloor$ , and

$$\max_k \mathbb{E}_k T_{0,n}^{\text{hit}} \sim (2 \log 2) n.$$

Now we can couple the true voter model ( $n$  different initial opinions) with the variant with only 2 opinions, initially held by  $k$  and  $n - k$  agents. (Just randomly assign these two opinions, initially). From this coupling we see

$$\mathbb{P}_k(T_{0,n}^{\text{hit}} > t) \leq \mathbb{P}(T_n^{\text{voter}} > t)$$

$$\mathbb{P}_k(T_{0,n}^{\text{hit}} > t) \geq \frac{2k(n-k-1)}{n(n-1)} \mathbb{P}(T_n^{\text{voter}} > t)$$

In particular, the latter with  $k = \lfloor n/2 \rfloor$  implies

$$\mathbb{E} T_n^{\text{voter}} \leq (4 \log 2 + o(1)) n.$$

This is weaker than the correct asymptotics (11).

## Voter model on general geometry

Suppose the flow rates satisfy, for some constant  $\kappa$ ,

$$\nu(A, A^c) := \sum_{i \in A, j \in A^c} n^{-1} \nu_{ij} \geq \kappa \frac{|A|(n - |A|)}{n(n - 1)}.$$

On the complete graph this holds with  $\kappa = 1$ . We can repeat the analysis above – the process  $X(t)$  now moves at least  $\kappa$  times as fast as on the complete graph, and so

$$\mathbb{E} T_n^{\text{voter}} \leq (4 \log 2 + o(1)) n / \kappa.$$

This illustrates another general principle.

## General Principle 5: Bottleneck statistics give crude general bounds

For a geometry with given rate matrix  $\mathcal{N} = (\nu_{ij})$ , the quantity

$$\nu(A, A^c) = \sum_{i \in A, j \in A^c} n^{-1} \nu_{ij}$$

has the interpretation, in terms of the associated continuous-time Markov chain  $Z(t)$  **at stationarity**, as “flow rate” from  $A$  to  $A^c$

$$\mathbb{P}(Z(0) \in A, Z(dt) \in A^c) = \nu(A, A^c) dt.$$

So if for some  $m$  the quantity

$$\phi(m) = \min\{\nu(A, A^c) : |A| = m\}, \quad 1 \leq m \leq n - 1$$

is small, it indicates a possible “bottleneck” subset of size  $m$ .

For many FMIE models, one can obtain upper bounds (on the expected time until something desirable happens) in terms of the parameters  $(\phi(m), 1 \leq m \leq n/2)$ . Such bounds are always worth noting, though

- $\phi(m)$  is not readily computable, or simulate-able
- The bounds are often rather crude for a specific geometry



More elegant to combine the family  $(\phi(m), 1 \leq m \leq n/2)$  into a single parameter, but the appropriate way to do this is (FMIE) model-dependent. In the voter model case above we used the parameter

$$\kappa := \min_A \frac{n(n-1)\nu(A, A^c)}{|A|(n-|A|)} = n(n-1) \min_m \frac{\phi(m)}{m(n-m)}.$$

Quantities like  $\kappa$  are descriptive statistics of a weighted graph. In literature you see the phrase “isoperimetric inequalities” which refers to bounds for particular weighted graph. In our setting – bounding behavior of a particular FMIE process in terms of the geometry – “bottleneck statistics” seems a better name.

## Coalescing MC on general geometry

Issues clearly related to study of the *meeting time*  $T_{ij}^{\text{meet}}$  of two independent copies of the MC, a topic that arises in other contexts. Under enough symmetry (e.g. continuous-time RW on the discrete torus) the relative displacement between the two copies evolves as the same RW run at twice the speed, and study of  $T_{ij}^{\text{meet}}$  reduces to study of  $T_k^{\text{hit}}$ .

First consider the general reversible case. In terms of the associated MC define a parameter

$$\tau^* := \max_{i,j} \mathbb{E}_i T_j^{\text{hit}}.$$

The following result was conjectured long ago but only recently proved. Note that on the complete graph the mean coalescence time is asymptotically  $2\times$  the mean meeting time.

### Theorem (Oliveira 2010)

*There exist numerical constants  $C_1, C_2 < \infty$  such that, for any finite irreducible reversible MC,  $\max_{i,j} \mathbb{E} T_{ij}^{\text{meet}} \leq C_1 \tau^*$  and  $\mathbb{E} T^{\text{coal}} \leq C_2 \tau^*$ .*

Proof is intricate.

To seek “ $1 \pm o(1)$ ” limits, let us work in the meeting model setting (stationary distribution is uniform) and write  $\tau_{\text{meet}}$  for mean meeting time from independent uniform starts. In a sequence of chains with  $n \rightarrow \infty$ , impose a condition such as the following. For each  $\varepsilon > 0$

$$n^{-2} |\{(i, j) : \mathbb{E} T_{ij}^{\text{meet}} \notin (1 \pm \varepsilon)\tau_{\text{meet}}\}| \rightarrow 0. \quad (12)$$

**Open problem.** Assuming (12), under what further conditions can we prove  $\mathbb{E} T^{\text{coal}} \sim 2\tau_{\text{meet}}$ ?

This project splits into two parts.

**Part 1.** For fixed  $m$ , show that the mean time for  $m$  initially independent uniform walkers to coalesce should be  $\sim 2(1 - \frac{1}{m})\tau_{\text{meet}}$ .

**Part 2.** Show that for  $m(n) \rightarrow \infty$  slowly, the time for the initial  $n$  walkers to coalesce into  $m(n)$  clusters is  $o(\tau_{\text{meet}})$ .

Part 1 is essentially a consequence of known results, as follows.

From old results on mixing times (RWG section 4.3), a condition like (12) is enough to show that  $\tau_{\text{mix}} = o(\tau_{\text{meet}})$ . So – as a prototype use of  $\tau_{\text{mix}}$  – by considering time intervals of length  $\tau$ , for  $\tau_{\text{mix}} \ll \tau \ll \tau_{\text{meet}}$ , the events “a particular pair of walker meets in the next  $\tau$ -interval” are approximately independent. This makes the “number of clusters” process behave as the Kingman coalescent.

**Note.** That is the hack proof. Alternatively, the explicit bound involving  $\tau_{\text{rel}}$  on exponential approximation for hitting time distributions from stationarity is applicable to the meeting time of two walkers, so a more elegant way would be to find an extension of that result applicable to the case involving  $m$  walkers.

Part 2 needs some different idea/assumptions to control short-time behavior.

**(restate) Open problem.** Assuming (12), under what further conditions can we prove  $\mathbb{E}T^{\text{coal}} \sim 2\tau_{\text{meet}}$ ?

What is known rigorously?

Cox (1989) proves this for the torus  $[0, m-1]^d$  in dimension  $d \geq 2$ . Here  $\tau_{\text{meet}} = \tau_{\text{hit}} \sim m^d R_d$  for  $d \geq 3$ .

Cooper-Frieze-Radzik (2009) prove Part 1 for the random  $r$ -regular graph, where  $\tau_{\text{meet}} \sim \tau_{\text{hit}} \sim \frac{r-1}{r-2}n$ .

Cooper-Elsässer-Ono-Radzik (2012) prove (essentially)

$$\mathbb{E}T^{\text{coal}} = O(n/\lambda)$$

where  $\lambda$  is the spectral gap of the associated MC. But this bound is of correct order only for expander-like graphs.

**(repeat earlier slide)**

Literature on finite voter model has focussed on estimating  $\mathcal{T}^{\text{voter}} \stackrel{d}{=} \mathcal{T}^{\text{coal}}$ , and I have shown some of this work.

But there are several other questions one can ask about the finite-time behavior. Let's recall what we studied for the averaging process.

### (repeat earlier slide: averaging process)

- 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 the associated (continuous-time) Markov chain with that initial distribution (Lemma 1).
- There is a duality relationship with coupled Markov chains (Lemma 3).
- There is an explicit bound on the closeness of the time- $t$  configuration to the limit constant configuration (Proposition 1).
- Complementary to this global bound there is a “universal” (i.e. not depending on the meeting rates) bound for an appropriately defined *local* roughness of the time- $t$  configuration (Proposition 2).
- The entropy bound.

## Other aspects of finite-time behavior (voter model)

1. Recall our “geometry-invariant” theme (General Principle 2). Here an invariant property is

expected total number of opinion changes =  $n(n - 1)$ .

2. If the proportions of agents with the various opinions are written as  $\mathbf{x} = (x_i)$ , the statistic  $q := \sum_i x_i^2$  is one measure of concentration - diversity of opinion. So study  $Q(t) := \sum_i (n^{-1} |\mathcal{V}_i(t)|)^2$ . Duality implies

$$\mathbb{E}Q(t) = \mathbb{P}(T^{\text{meet}} \leq t)$$

where  $T^{\text{meet}}$  is meeting time for independent MCs with uniform starts. Can study in special geometries.

3. A corresponding “local” measure of concentration - diversity is the probability that agents  $(I, J)$  chosen with probability  $\propto \nu_{ij}$  (“neighbors”) have same opinion at  $t$ . (“Diffusive clustering”: Cox (1986))



4. The statistic  $q := \sum_i x_i^2$  emphasizes large clusters (large time); the statistic  $\text{ent}(\mathbf{x}) = -\sum_i x_i \log x_i$  emphasizes small clusters (small time). So one could consider

$$\mathbb{E}(t) := -\sum_i (n^{-1}|\mathcal{V}_i(t)|) \log(n^{-1}|\mathcal{V}_i(t)|)$$

Apparently not studied – involves similar short-time issues as in the  $\mathbb{E} T^{\text{coal}} \sim 2\tau_{\text{meet}}$ ? question.

## General Principle 6: Approximating finite graphs by infinite graphs

For two of the standard geometries, there are local limits as  $n \rightarrow \infty$ .

- For the torus  $\mathbb{Z}_m^d$ , the  $m \rightarrow \infty$  limit is the infinite lattice  $\mathbb{Z}^d$ .
- For the “random graphs with prescribed degree distribution” model, (xxx not yet introduced?) the limit is the associated Galton-Watson tree.

There is also a more elaborate random network model (Aldous 2004) designed to have a more “interesting” local weak limit for which one can do some explicit calculations – it’s an **Open Topic** to use this as a testbed geometry for studying FMIE processes.

So one can attempt to relate the behavior of a FMIE process on such a finite geometry to its behavior on the infinite geometry. This is simplest for the “epidemic” (FPP) type models later, but also can be used for MC-related models, starting from the following

**Local transience principle.** For a large finite-state MC whose behavior near a state  $i$  can be approximated by a transient infinite-state chain, we have

$$\mathbb{E}_\pi T_i^{\text{hit}} \approx R_i / \pi_i$$

where  $R_i$  is defined in terms of the approximating infinite-state chain as  $\int_0^\infty p_{ii}(t) dt = \frac{1}{\nu_i q_i}$ , where  $q_i$  is the chance the infinite-state chain started at  $i$  will never return to  $i$ .

The approximation comes from the finite-state mean hitting time formula via a “interchange of limits” procedure which requires ad hoc justification.

Conceptual point here: local transience corresponds roughly to voter model consensus time being  $\Theta(n)$ .

In the case of simple RW on the  $d \geq 3$ -dimensional torus  $\mathbb{Z}_m^d$ , so  $n = m^d$ , this identifies the limit constant in  $\mathbb{E}_\pi T_i^{\text{hit}} \sim R_d n$  as  $R_d = 1/q_d$  where  $q_d$  is the chance that RW on the infinite lattice  $\mathbb{Z}^d$  never returns to the origin.

In the “random graphs with prescribed degree distribution” model, this argument (and transience of RW on the infinite Galton-Watson limit tree) shows that  $\mathbb{E}_\pi T_i^{\text{hit}} = \Theta(n)$ .

## A final thought

For Markov chains, mixing times and hitting times seem “basic” objects of study in both theory and applications. These objects may look quite different, but Aldous (1981) shows some connections, and in particular an improvement by Peres-Sousi (2012) shows that (variation distance) mixing time for a reversible chain agrees (up to constants) with  $\max_{A:\pi(A)\geq 1/4} \max_i \mathbb{E}_i T_A^{\text{hit}}$ .

We have seen that the behaviors of the Averaging Process and the Voter Model are closely related to the mixing and hitting behavior of the associated MC. Is there any direct connection between properties of these two FMIE processes? Does the natural coupling tell you anything?