3. The Voter Model

David Aldous

July 22, 2012

We now move on to the voter model, which (compared to the averaging model) 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.

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 $i \to j$ or $j \to i$.

Voter model. Initially each agent has a different "opinion" – agent i has opinion i. When i and j meet at time t with direction $i \to j$, 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 $\{\mathcal{V}_i(t), i \in \textbf{Agents}\}$ is a random partition of **Agents**. A natural quantity of interest is the **consensus time**

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

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 (cluster) $\mathcal{C}_i(t)$ of tokens. When i and j meet at time t with direction $i \to j$, 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 $\{\mathcal{C}_i(t), i \in \textbf{Agents}\}$ is a random partition of **Agents**. A natural quantity of interest is the **coalescence time**

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

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,

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

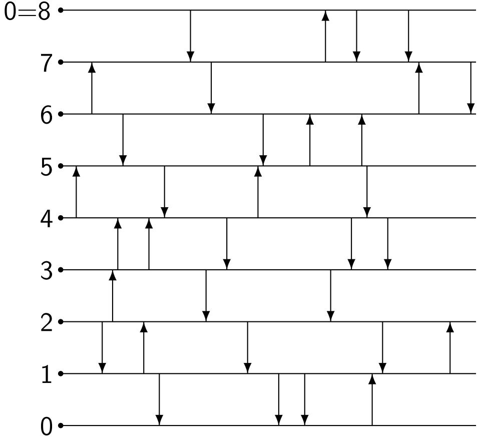
In particular $T^{\text{voter}} \stackrel{d}{=} T^{\text{coal}}$.

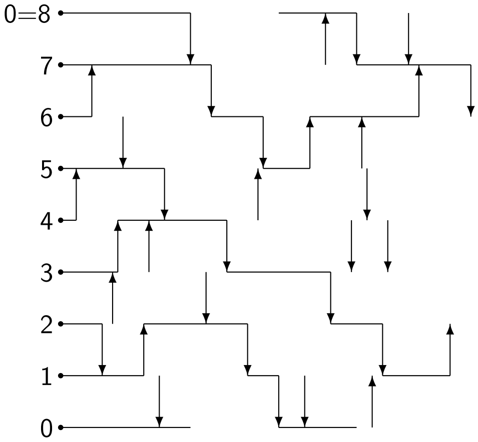
They are different as processes. For fixed i, note that $|\mathcal{V}_i(t)|$ can only change by ± 1, but $|\mathcal{C}_i(t)|$ jumps to and from 0.

In figures on next slides,
time "left-to-right" gives CMC,
time "right-to-left" with reversed arrows gives VM.

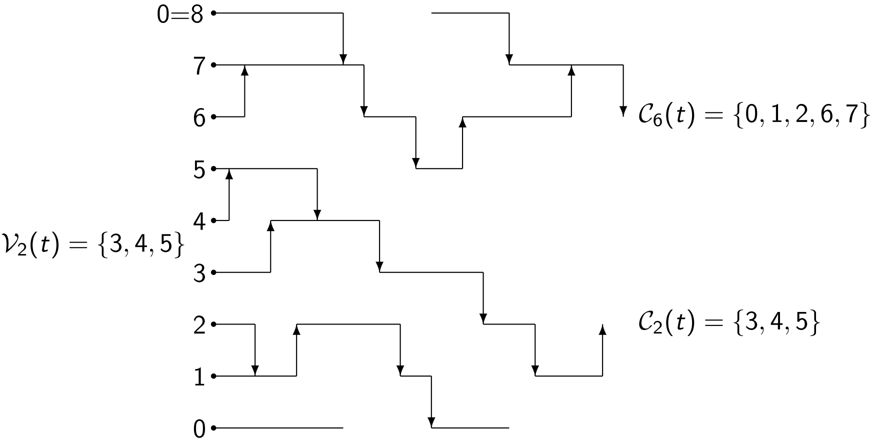
Note this depends on the symmetry assumption $\nu_{ij} = \nu_{ji}$ of the meeting process.

Schematic – the meeting model on the 8-cycle.





0=8
7
6
5
4
3
2
1
0


0=8
7
6
5
4
3
2
1
0
$\mathcal{C}_6(t) = \{0, 1, 2, 6, 7\}$
$\mathcal{V}_2(t) = \{3, 4, 5\}$
$\mathcal{C}_2(t) = \{3, 4, 5\}$

Literature on finite voter model has focussed on estimating $T^{\text{voter}} \stackrel{d}{=} T^{\text{coal}}$, 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^{voter} on the complete graph, both providing some bounds on other geometries.

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

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

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

In coalescing RW on the complete n-graph, the number of clusters evolves as the continuous-time MC on states $\{1, 2, 3, \ldots, n\}$ 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

$$\mathbb{E}\, T_n^{\text{voter}} = \mathbb{E}\, T_n^{\text{coal}} \sim 2n. \tag{1}$$

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 $\{0, 1, 2, \ldots, n\}$ with rates

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

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

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

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

$$\mathbb{E}_k T^{\mathrm{hit}}_{0,n} = \tfrac{2(n-1)}{n} \; \left(k(h_{n-1} - h_{k+1}) + (n-k)(h_{n-1} - h_{n-k+1})\right)$$

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

$$\max_k \mathbb{E}_k T^{\mathrm{hit}}_{0,n} \sim (2 \log 2) \; n.$$

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

$$\mathbb{P}_k(T^{\text{hit}}_{0,n} > t) \leq \mathbb{P}(T^{\text{voter}}_n > t)$$

$$\mathbb{P}_k(T^{\text{hit}}_{0,n} > t) \geq \tfrac{2k(n-k-1)}{n(n-1)}\mathbb{P}(T^{\text{voter}}_n > t)$$

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

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

This is weaker than the correct asymptotics (1).

Voter model on general geometry

Suppose the flow rates satisfy, for some constant κ,

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

On the complete graph this holds with $\kappa = 1$. We can repeat the analysis above – the process $X(t)$ now moves at least κ 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 4: 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

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

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 $(\phi(m), 1 \le m \le n/2)$. 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 κ 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.

Digression: regularity conditions for finite irreducible MCs

	Degrees of freedom
general	$n(n-1)$
reversible, π general	$\frac{1}{2}n(n-1)+(n-1)$
reversible, π uniform	$\frac{1}{2}n(n-1)$
vertex-transitive	$n-1$

A cute fact about general MCs is

$$\sum_j \pi_j \mathbb{E}_i T_j^{\text{hit}} := \tau_0$$

does not depend on i, and is the natural "ave hitting time" statistic for a MC.

Coalescing MC on general geometry

Issues clearly related to study of the *meeting time* T_{ij}^{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}^{meet} reduces to study of T_k^{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. Proof is intricate.

Theorem (Oliveira 2010)

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

In our meeting model setting of uniform stationary distribution this typically gives us the correct order of magnitude of $\mathbb{E} T^{\text{coal}} = \mathbb{E} T^{\text{voter}}$.

To seek "$1 \pm o(1)$" limits, that is to generalize the Kngman coalescent result, write τ_{meet} for mean meeting time from independent uniform starts. In a sequence of chains with $n \to \infty$, impose a condition such as the following. For each $\varepsilon > 0$

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

Open problem. Assuming (2), 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 $m(n) \to \infty$ 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 (2) is enough to show that the (variation distance) mixing time $\tau_{\mathrm{mix}} = o(\tau_{\mathrm{meet}})$. So – as a prototype use of τ_{mix} – by considering time intervals of length τ, for $\tau_{\mathrm{mix}} \ll \tau \ll \tau_{\mathrm{meet}}$, the events "a particular pair of walker meets in the next τ-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 τ_{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 (2), 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-Elsässer-Ono-Radzik (2012) prove (essentially)

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

where λ 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 $T^{\text{voter}} \stackrel{d}{=} T^{\text{coal}}$, 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
- There is a duality relationship with coupled Markov chains
- There is an explicit bound on the closeness of the time-t configuration to the limit constant configuration
- 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
- The entropy bound.

Other aspects of finite-time behavior (voter model)

1. The voter model has a simple "geometry-invariant" property: expected total number of opinion changes $= n(n-1)$.

This holds because "number of agents with opinion i" is a martingale.

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^{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)). Can again study by duality.

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

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

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

General Principle 5: Approximating finite graphs by infinite graphs

For two of the standard geometries, there are local limits as $n \to \infty$.
• For the torus $\mathbb{Z}_m^d$, the $m \to \infty$ limit is the infinite lattice $\mathbb{Z}^d$.
• For the "random graphs with prescribed degree distribution" model, 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 be 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 $\Theta(n)$ (as seen in the $\mathbb{Z}^d$ case).

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

We have seen that the behaviors of the Averaging Process/ Voter Model are closely related to the mixing/hitting behavior of the associated MC. For Markov chains, mixing times and hitting times are “basic” objects of study in both theory and applications. They are not directly related to each other, but Aldous (1981) showed some indirect 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 $\max_{A:\pi(A)\geq 1/4} \max_i \mathbb{E}_i T_A^{\text{hit}}$.

Is there any indirect connections between properties of the Averaging Process/Voter Model? Does the natural coupling tell you anything?

Bathtub problem. Invent models which combine Averaging Process/Voter Model. (I will mention two in lecture 5).