Finite Markov Information-Exchange processes

David Aldous

March 30, 2011

Note that previous FMIE models were "non-equilibrium". We digress to a quite different FMIE model designed to have an equilibrium.

I can remember Bertrand Russell telling me of a horrible dream. He was in the top floor of the University Library, about A.D. 2100. A library assistant was going round the shelves carrying an enormous bucket, taking down books, glancing at them, restoring them to the shelves or dumping them into the bucket. At last he came to three large volumes which Russell could recognize as the last surviving copy of *Principia Mathematica*. He took down one of the volumes, turned over a few pages, seemed puzzled for a moment by the curious symbolism, closed the volume, balanced it in his hand and hesitated ....
(G. H. Hardy, *A Mathematician's Apology*)

Want to store information (informally, a *book*) somewhere in a data-storage network, for a LONG time, longer than the reliable lifetime of an individual node. Need more than one copy of each book. Cost of storage/communication of *title* of book is negligible. Set time unit so that cost of storage of book for one time unit = cost of communicating the book.

Goal: A distributed algorithm which maintains a small number of copies of each book in an unreliable network over times much longer than lifetimes of individual vertices. The algorithm doesn't know the current number of copies.

In our FMIE setting, with (large) n vertices, set μ ($= 10$, say) for desired average number of copies, then set $p = \mu/n$. We want to define a particle process (particle = copy of book) such that, in the "reliable network" setting,

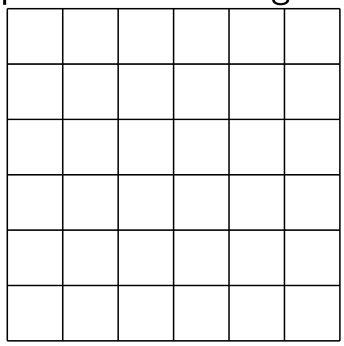
$$\text{the stationary distribution is independent Bernoulli}(p) \text{ conditioned on non-empty.} \tag{1}$$

The evolution rule is simple. Use the directed meeting model. At a directed meeting $(i \to j)$,

- if i has a particle then j "resets" to have a particle with chance p and no particle with chance $1 - p$;
- if i has no particle then j does not change state.

Note (1) holds by checking the reversible equilibrium criterion.

What happens on a low-degree graph?



First-order effect: Isolated particles do RW at rate $p/2$.
Second-order effect: A particle splits into two non-adjacent particles at rate $O(p^2)$. Two particles which become adjacent have chance $O(1)$ to merge.
Math Insight: Could directly define a process of particles doing RW, splitting, coalescing – but wouldn't know its stationary distribution. This **constrained Ising** model has these qualitative properties and a simple stationary distribution.

Continuing in the "reliable" setting, what can one say about mixing/relaxation times? Take μ fixed and $n \to \infty$. There are two kinds of obstacles ......

From motion of single particle, get lower bound $\Omega(p^{-1} \times$ mixing/relaxation time of MC).

From the mean-field birth-coalescence chain, that is the birth-death chain with

$$q_{i,i+1} = \mu i; \quad q_{i+1,i} = (i+1)i$$

and Poisson(μ) stationary distribution. This has finite mixing/relaxation time, so we get lower bound $\Omega(p^{-2})$. This bound is typically larger.

xxx studied on infinite lattice xxx refs but not xxx finite.
xxx theory project to prove anything.

xxx unreliable network.
xxx have implicitly assumed the typical meeting rate $\nu = \nu_i$ is $\Theta(1)$. In order to restore lost copies we want

$$p^2 \nu \gg 1/L$$

for $L :=$ reliable lifetime. This becomes

$$\nu \gg n^2/L.$$

Key point is that this should work on a dynamic (changing) graph – can have much more drastic changes than independent single-site failures.
Challenge to prove anything like this!

Two variants of the averaging model

Model 2 (the averaging model), with state interpreted as opinion, is a rather extreme instance of a "consensus-seeking" model – extreme because agents ultimately retain no memory of their original opinion. One can invent variants where some aspects of opinion are fixed, and we will look at two such models.

Subjective assessment of an average. Consider a model of how humans assess the average opinion. Presumably people are in fact biased towards the opinions of those who are close in the underlying geometry. Here is a model for such subjective assessment.

Model: Subjective assessment of an average. Each agent i has a fixed real-valued opinion $y(i)$, centered so that $\bar{y} := n^{-1} \sum_i y_i = 0$, and a subjective assessment of other's opinions, a random process $X_i(t)$. We suppose that when an agent j communicates with another agent, the impression that the other agent gets of j's opinion is a linear combination of j's own opinion and j's view of the average opinion:

$$\lambda y(j) + (1 - \lambda) X_j(t-),$$

where either $0 < \lambda < 1$ or $\lambda > 1$.
[discuss on board: compromising or polarizing]

Now when agent i meets agent j, agent i updates his opinion of the consensus as an average of his previous opinion and the opinion received from j:

$$X_i(t+) = \mu X_i(t-) + (1 - \mu)[\lambda y(j) + (1 - \lambda) X_j(t-)]$$

where here $0 < \mu < 1$. (note that j also updates).

This completes the model description. We consider the case $0 < \lambda < 1$, where it turns out that $X_i(t) \overset{d}{\to} X_i(\infty)$ as $t \to \infty$ and we study the limit distribution. Write $W(i, t), t = 0, 1, 2, 3$ for the discrete-time jump chain associated with the meeting rate matrix. Then

Lemma

$\mathbb{E} X_i(\infty) = \mathbb{E} y(W(i, G_\lambda))$ *where* G_λ *has Geometric(λ) distribution.*

Note that the value of μ does not matter.

[argument on board]

Theory project: similar formula for $\mathrm{var} X_i(\infty)$?

How can we extract some insight from the identity in the lemma? Assume normalized meeting rates, so that the dominant eigenvalue $1 - \lambda_2$ for the jump chain matches the eigenvalue λ_2 of the continuous-time chain.

Digression: In the context of real-valued MC-like FMIE processes, it may be helpful to consider an initial configuration $y(i)$ of the form

$$y(i) = \sigma_{\text{glo}}\alpha(i) + \sigma_{\text{loc}}z(i) \tag{2}$$

where α is the normalized eigenvector associated with the eigenvalue λ_2 of the associated continuous-time MC, and $z = z(i)$ is a typical realization of IID, mean zero and variance one, RVs. This form constitutes a "mixture of opposites", because we may interpret σ_{glo} and σ_{loc} as indicating the amounts of "purely global" and "purely local" variation. See (3) below.

In the current model, let us measure the extent of influence of the geometry on the subjective assessments of the average via the ratio

$$\rho := \mathbb{E}[y(U)X_U(\infty)]/\mathbb{E}[y^2(U)]$$

where U is a uniform random agent. This is somewhat reasonable, because the ratio would be zero if $X_U(\infty)$ were the correct average (zero), and would be one if $X_U(\infty)$ were replaced by the agent's own opinion $y(U)$.

To study ρ when y is of the form (2), note

$$\mathbb{E}\alpha^2(U) = 1, \ \mathbb{E}z^2(U) \approx 1, \ \mathbb{E}[\alpha(U)z(U)] \approx 0$$

so

$$\mathbb{E}[y^2(U)] \approx \sigma_{\text{glo}}^2 + \sigma_{\text{loc}}^2. \tag{3}$$

From the eigenvector interpretation of α

$$\mathbb{E}[\alpha(U)\alpha(W(U,g))] = (1-\lambda_2)^g.$$

We can calculate [board]

$$\mathbb{E}[y(U)X_U(\infty)] \approx \sigma_{\text{glo}}^2\mathbb{E}(1-\lambda_2)^{G(\lambda)} + \sigma_{\text{loc}}^2\mathbb{E}[z(U)\ z(W(U,G(\lambda))\ ]$$

$$\approx \sigma_{\text{glo}}^2\frac{\lambda(1-\lambda_2)}{1-(1-\lambda)(1-\lambda_2)} + \sigma_{\text{loc}}^2\mathbb{P}(W(U),G(\lambda) = U).$$

In terms of the relaxation time,

$$\mathbb{E}[y(U)X_U(\infty)] \approx \sigma^2_{\text{glo}} \frac{\lambda(\tau_{\text{rel}}-1)}{\lambda(\tau_{\text{rel}}-1)\ +1} + \sigma^2_{\text{loc}} \mathbb{P}(W(U), G(\lambda) = U).$$

We interpret this as saying that the global and local structure of the geometry affect this particular FMIE model via two different weighting factors. For the effect of the global structure to be small we need $\lambda = o(1/\tau_{\text{rel}})$. For the effect of the local structure to be small we just need, in the case where there is a $n \to \infty$ limit matching infinite geometry, $\lambda = o(1)$.

Of course this is all rather obvious from the model definition – it's just nice we can do a calculation.

Here is an alternative model whose analysis is quite similar. Details have been written out in Acemoglu et al (2010), in somewhat more generality than below.

In Model 2 (Averaging model) and the model above, when i and j meet they both update. But the behavior of the "asymmetric" version where only one agent updates is qualitatively similar. So consider the directed meeting model (there is a meeting $i \to j$ at rate $\nu_{ij}/2$) and the associated asymmetric averaging model:

at a meeting $i \to j$ at time t, set

$$X_j(t+) = (1-\theta)X_j(t-) + \theta X_i(t-).$$

Note that $\theta = 1$ is essentially the voter model; we take $0 < \theta < 1$. Now suppose some specified subset $\mathcal{S}$ of agents are "stubborn" and never update their opinions. Call this

Model: Asymmetric averaging model with some fixed opinions.

Why is there a limit distribution? Can apply a standard method, **iterated random functions**.

[board]

In the FMIE setting, we imagine the meeting model run over time $(-\infty, 0]$. Suppose we can construct a time-0 random configuration $\mathbf{X}^*$ such that, for any configuration $\mathbf{x}(t_0)$ at time $t_0 < 0$, the resulting time-0 configuration $\mathbf{X}(\mathbf{x}(t_0))$ satisfies

$$\mathbf{X}(\mathbf{x}(t_0)) \to \mathbf{X}^* \text{ a.s. as } t_0 \to -\infty.$$

Then $\mathbf{X}^*$ has the stationary distribution, and from any time-0 configuration the forwards process satisfies

$$\mathbf{X}(t) \ \stackrel{d}{\to}\ \mathbf{X}^* \text{ as } t \to \infty.$$

How does this work in the present model?

Condition on the meeting process over time $(-\infty, 0]$. We want to construct $X_i(0)$. Run t backards from 0; at each meeting we "accept" with chance θ. The first accepted meeting is at some time $-T_1$ and with some agent J_1. If J_1 is not a stubborn agent then continue; run t backwards from $-T_1$ until the first accepted meeting, with some agent J_2. And so on.

As in the previous model, $J_0 = i, J_1, J_2$ is the associated jump chain. Eventually this chain hits some stubborn agent J^*. This J^* depends on the history of the meeting process ($\mathcal{H}$, say) and the extra randomness of accepting/rejecting. From the model description we can get the time-0 value as the expectation over the latter

$$X_i(0) = \mathbb{E}(x_{J^*} | \mathcal{H})$$

Underlying this is the continuous-time MC $Q(t)$ which follows only the "accept" meetings. It is distributed as the usual associated MC, slowed down by the factor θ.

As in Model 2, one can study variances of the stationary distribution $\mathbf{X}$ via a specific coupling of two versions of this process. (what we wrote in Model 2 was the $\theta = 1/2$ case). In terms of the coupling $(Q^i(t), Q^j(t))$ started at (i, j) and stopped at the first hitting times $T^i_{\mathcal{S}}, T^j_{\mathcal{S}}$ on $\mathcal{S}$,

$$\mathbb{E}[X_i X_j] = \mathbb{E}[x(Q^i(T^i_{\mathcal{S}}))x(Q^j(T^j_{\mathcal{S}}))]. \quad (4)$$

This is Theorem 3 of Acemoglu et al (2010). That paper gives calculations for some standard geometries; also a general result (Theorem 4) formalizing the following intuition.

Write $\pi(\mathcal{S}) = |\mathcal{S}|/n$ for the proportion of stubborn agents, and write τ for a suitable mixing time (of the usual associated MC). If $\tau\pi(\mathcal{S}) \ll 1$ then the MC gets close to the uniform distribution before hitting $\mathcal{S}$. So each stationary value X_i will likely be close to the expectation, for the chain started from stationarity, of the opinion of the first stubborn agent met:

$$X_i \approx \mathbb{E}_\pi x(Q(T_{\mathcal{S}})).$$

The proof uses standard MC tools, such as variation distance mixing times and the approximate exponential distribution for stationary-start hitting times.

Some **theory projects.**
1. Cleaner proofs?
2. Adapt to previous model.
3. Is there an analog of the "universality property" of local smoothness in the averaging model?

Model 4 (ordered consensus-seeking model) envisages agents as "slaves to authority". Here is a conceptually opposite "slaves to fashion" model, whose analysis is surprisingly similar.

Model: Fashionista.
Take a general meeting model. At the times of a rate-λ Poisson process, a new fashion originates with a uniform random agent, and is time-stamped. When two agents meet, they each adopt the latest (most recent time-stamp) fashion.

There is a stationary distribution, for the random partition of agents into "same fashion", to which one can include the "time before present" that the fashion started. Existence of the stationary distribution follows easily from the "FMIE version of iterated random functions" method used in the previous model.

Let us repeat the analysis of Model 4 in tbe complete graph geometry.

Model 4. Ordered consensus-seeking model; complete graph geometry,
(i) The agents are labelled 1 through n. Agent i initially has opinion i.
(ii) When two agents meet, they adopt the same opinion, the smaller of the two labels.

Here is some heuristic analysis of $(X_1^n(t), \ldots, X_k^n(t))$, where $X_k^n(t)$ is the proportion of the population with opinion k at time t.

The first component $X_1^n(t)$ evolves as the epidemic process, which during the pandemic phase follows a deterministic function $H_1(t)$ satisfying the logistic equation $H_1' = H_1(1 - H_1)$ whose general solution is $H_1(t) = F(t + c_1)$ for the logistic function F. We can rephrase the randomly-shifted logistic limit theorem to say

$$(X_1^n(\log n + s), \ -\infty < s < \infty) \to (F(C_1 + s), \ -\infty < s < \infty)$$

where $C_1 = \log \xi_1 \overset{d}{=} -G$ for Exponential(1) ξ_1 and Gumbel G.

Use the same "martingale" argument to see that, durng the pandemic phase, the random process $(X_1^n(\cdot), X_2^n(\cdot))$ will follow a determinstic function $(H_1(\cdot), H_2(\cdot))$ that satisfies the DEs

$$\begin{aligned} H_1' &= H_1(1 - H_1) \\ H_2' &= H_2(1 - H_1 - H_2) - H_1 H_2. \end{aligned}$$

We can solve these by simply observing that $H_1 + H_2$ must satisfy the logistic equation; so the general solution is

$$(H_1(t), H_2(t)) = (F(t + c_1), F(t + c_2) - F(t + c_1))$$

for $c_2 > c_1$. So we expect limit behavior of the form

$$((X_1^n(\log n+s), X_2^n(\log n+s), \ldots, X_k^n(\log n+s)), \ -\infty < s < \infty) \to \quad (5)$$

$$((F(C_1+s), F(C_2+s)-F(C_1+s), , \ldots, F(C_k+s)-F(C_{k-1}+s)), \ -\infty < s < \infty)$$

for some random $C_1 < C_2 < \ldots < C_k$.

We can determine the C_j by considering large negative s. From the Yule process approximation to the initial phase we have

$$X_j^n(\log n + s) \sim e^s \xi_j; \quad \text{for IID Exponential(1)}(\xi_j)$$

But since $F(s) \sim e^s$ we have

$$F(C_j + s) - F(C_{j-1} + s) \sim e^s(e^{C_j} - e^{C_{j-1}}).$$

This gives the equations

$$\begin{aligned} e^{C_j} - e^{C_{j-1}} &= \xi_j \quad j \geq 2 \\ e^{C_1} &= \xi_1. \end{aligned}$$

which have solution

$$C_j = \log(\xi_1 + \ldots + \xi_j), \ j \geq 1. \quad (6)$$

Fashionista model; complete graph geometry.

We start with some notational issues. Consider a configuration of points on $(0,1)$ with only finitely many points in any interval $[\varepsilon, 1-\varepsilon]$. We want to think of the space of such configurations in the usual "point process" way. For notational simplicity we write a configuration as $\mathbf{s} = (s_i, -\infty < i < \infty)$ but only the ordering of the i is relevant; $\mathbf{s}$ is the same as $(s_{i+1}, -\infty < i < \infty)$.

In the fashionista model write $\mathbf{X}^n(t) = (X_i^n(t), -\infty < i < \infty)$ for the proportions of agents adopting the different fashions i at time t, where the fashions i are ordered from "most recent" to "least recent", that is $i+1$ originated before i. We study $\mathbf{X}^n$ by studying the cumulative process $\mathbf{S}^n$, that is $S_i^n(t) = \sum_{j \le i} X_j^n(t)$, and regarding a realization of $\mathbf{S}^n(t)$ as a point configuration, as above.

For the stationary version of the fashionista model, we will argue that as $n \to \infty$ the process $\mathbf{S}^n$ converges (no rescaling is involved) to a limit process $\mathbf{S}$ with the following general structure.

Take the points $(C_i, \ -\infty < i < \infty)$ of a stationary process on $(-\infty, \infty)$. Write F for the logistic function. Define, for $-\infty < t < \infty$,

$$S_i(t) = F(C_i + t). \quad (7)$$

The resulting process $\mathbf{S}(t) = (S_i(t), -\infty < i < \infty)$ is stationary (as t varies)

We repeat the general method of analysis used for Model 4: first argue that the limit process must be of the form $\mathbf{X}(t)$ above for some (C_i); then determine the distribution of (C_i) by considering the initial "pandemic" stage..

We study the stationary fashionista process over time $-\infty < t < \infty$. Consider some $-t_0 = -\log n \pm O(1)$ and some $-t_1 = -O(1)$. Over time $[-t_1, t_1]$ write $Y^n_{-t_0}(t)$ for the proportion of agents at time t adopting some fashion introduced during $[-t_0, -t_1]$. The process $t \to Y^n_{-t_0}(t)$ evolves essentially as the epidemic process over time $[-t_1, t_1]$; whenever an agent in this group meets another agent, the other agent adopts one of the group's fashions. The $O(1)$ new fashions introduced over $[-t_1, t_1]$ only attract $O(1)$ agents and so make negligible contribution.

This argument shows that a process $t \to S_i^n(t)$ must (to first-order) follow the logistic curve as its values increase over the range $[\varepsilon, 1-\varepsilon]$. So the limit process must be of the form (7):

$$S_i(t) = F(C_i + t).$$

Because the limit process (as a limit of stationary processes) must be stationary, the (C_i) must form a stationary point process.

Now consider the fashions at time $t = 0$ adopted by small but non-negligible proportions of the population. More precisely, consider fashions originating during the time interval $[-\log n + t_n, -\log n + 2t_n]$, where $t_n \to \infty$ slowly. For a fashion originating at time $-\log n + \eta$, the time-0 set of adopting agents will be a subset of the corresponding epidemic process, which we know has proportional size $\xi \exp(-\eta) = \exp(-\eta + \log \xi)$ where ξ has Exponential(1) distribution.

The times $-\log n + \eta_j$ of origination of different fashions form by assumption a rate-λ Poisson process, and after we impose IID shifts $\log \xi_j$ we note (as an elementary property of Poisson processes) that the shifted points $-\log n + \eta_j + \log \xi_j$ still form a rate-λ Poisson process, say γ_j, on $(-\infty, \infty)$. So the sizes of small recent fashion groups (that is letting $j \to -\infty$), for which overlap between fashions becomes negligible, are approximately $\exp(\gamma_j)$. Summing over $j \leq i$ gives

$$\sum_{j \leq i} \exp(\gamma_j) \approx F(C_i) \approx \exp(C_i)$$

and we end up with the representation

$$C_i = \log \left(\sum_{j \leq i} \exp(\gamma_j) \right) = \gamma_i + \log \left(\sum_{k \geq 1} \exp(\gamma_{i-k} - \gamma_i) \right). \qquad (8)$$

Digression: Consider any FMIE process where
(i) states are qualitative (categories) rather than quantitative (numerical)
(ii) there is a stationary distribution.

Two natural summary statistics of the stationary distribution of such a process are
(a) $s := \mathbb{E} \sum_i (|C_i|/n)^2$, where C_i is the set of category-i agents.
(b) $\mu := (1/n) \times$ (mean meeting rate between agents of different categories).

[discuss s on board]

In our (complete graph) fashionista model we can just write down expressions for s or μ (which are closely related, here) in the $n \to \infty$ limit.

[project: do these simplify?]