# Finite Markov Information-Exchange processes

David Aldous

February 9, 2011

**Model 2. Averaging model.** Here "information" is most naturally interpreted as money. When agents $i$ and $j$ meet, they split their combined money equally, so the values ($X_i(t)$ and $X_j(t)$) are replaced by the average $(X_i(t) + X_j(t))/2$.

The overall average is conserved, and rather obviously each agent's fortune $X_i(t)$ will converge to the overall average. It turns out to be easy to quantify this convergence: Proposition 1 below. We note first a simple relation with the associated Markov chain. Write $1_i$ for the initial configuration $X_j(0) = 1_{(i=j)}$ and $p_{ij}(t)$ for the transition probabilities for the Markov chain.

**Lemma**

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

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

So if $\mathbf{x}(0)$ is a probability distribution, then the means evolve as the distribution of the MC started with $\mathbf{x}(0)$ and slowed down by factor $1/2$.

**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 is as the MC 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

$$\mathbb{P}(U = i|\mathbf{X}) = X_i. \tag{1}$$

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

$$(\mathbf{x}, u) \to (\mathbf{x}^e, u): \ \text{rate } \nu_e, \quad \text{if } u \notin e$$

$$(\mathbf{x}, u) \to (\mathbf{x}^e, u): \ \text{rate } \nu_e/2, \quad \text{if } u \in e$$

$$(\mathbf{x}, u) \to (\mathbf{x}^e, u'): \ \text{rate } \nu_e/2, \quad \text{if } u \in e = (u, u').$$

This defines a process $(\mathbf{X}(t), U(t))$ consistent with the averaging process and satisfying (1). The latter implies $\mathbb{E}X_i(t) = \mathbb{P}(U(t) = i)$, and clearly $U(t)$ evolves as the MC at half speed.

The case of a general initial configuration follows by linearity. That is, for a given realization of the underlying meeting process, the map from $\mathbf{x}(0)$ to $\mathbf{X}(t)$ is linear.

**Lemma (repeat)**

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

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

So from MC theory the means $\mathbb{E}\mathbf{X}(t)$ converge to the limit constant $\overline{\mathbf{x}(0)}$ at asymptotic exponential rate given by half the spectral gap of the MC.
Next we study $\mathbf{X}(t)$ itself.

**Proposition (Global convergence in the averaging model)**

*From an initial configuration* $\mathbf{x} = (x_i)$ *with average zero and* $L^2$ *size* $||\mathbf{x}||_2 := \sqrt{n^{-1}\sum_i x_i^2}$, *the time-*$t$ *configuration* $\mathbf{X}(t)$ *satisfies*

$$\mathbb{E}||\mathbf{X}(t)||_2 \leq ||\mathbf{x}||_2 \exp(-\lambda t/4), \quad 0 \leq t < \infty \tag{2}$$

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

Theorems 4.1 and 5.1 of Shah (2008) give an analogous "gossip algorithm" result.

**Proof.** 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 $Q(t) := ||\mathbf{X}(t)||_2^2$,

$$\begin{aligned}
\mathbb{E}\left(\tfrac{dQ(t)}{dt}|_{t=0}|\mathbf{x}(0) = \mathbf{x}\right) &= -\sum_{\{i,j\}} \nu_{ij} \cdot n^{-1}(x_j - x_i)^2/2 \\
&= -\mathcal{E}(\mathbf{x}, \mathbf{x})/2 \\
&\leq -\lambda||\mathbf{x}||_2^2/2
\end{aligned}$$

The middle equality is just the definition of $\mathcal{E}$ and the final inequality is the extremal characterization

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

Writing $\mathcal{F}(t)$ for the filtration of the underlying meeting process we get the martingale-like inequality

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

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 $\surd$.

**Theory project:** Write $\mathrm{ent}(\mathbf{x}) = -\sum_i x_i \log x_i$ for the entropy of $\mathbf{x}$. If $\mathbf{x}(0)$ is a probability measure then from the global convergence result $\mathrm{ent}(\mathbf{X}(t))$ converges in distribution to $\mathrm{ent}(\pi) = \log n$. Can we get upper bounds on $\log n - \mathrm{ent}(\mathbf{X}(t))$ in terms of the log-Sobolev constant of the MC?

**Literature project;** what has been done on this model and variants? Certainly there is some literature on the "noise" variant below.

## Averaging process with noise

The 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 $(W_i(1), 1 \leq i \leq n)$. The effect of the conditioning is to make the variance of $W_i(1)$ become $(n-1)/n$ (instead of 1) and to introduce covariance $\mathbb{E}[W_i(1)W_j(1)] = -1/n$, $j \neq i$. Now take $\mathbf{W}(t)$ to be the $n$-dimensional Brownian motion associated with the time-1 distribution $\mathbf{W}(1) = (W_i(1), 1 \leq i \leq n)$.

For this process we can repeat the analysis in Proposition 1. The only change is that the noise makes an extra contribution to $\mathbb{E}(dQ(t)|\mathcal{F}(t))/dt$ of

$$n^{-1} \times \sigma^2(n-1)/n \times n = \sigma^2(n-1)/n.$$

So

$$\mathbb{E}(dQ(t)|\mathcal{F}(t))/dt \le -\lambda Q(t)/2 + \sigma^2(n-1)/n.$$

By soft arguments there must be a $t \to \infty$ limit distribution $\mathbf{X}(\infty)$ and $\mathbb{E}Q(t) \to \mathbb{E}Q(\infty)$, implying

$$0 \le -\lambda \mathbb{E}Q(\infty)/2 + \sigma^2(n-1)/n.$$

This rearranges to the following bound on the variability of the equilibrium distribution.

**Corollary (Equilibrium variability in averaging with noise)**

$$\mathbb{E}||\mathbf{X}(\infty)||_2^2 \le 2\tau_{rel}\, \sigma^2(n-1)/n.$$

We next reconsider the basic averaging model. Note (though we do not use this fact) that time-reversal gives the **self-duality** property

$$X(\mathbf{x}, t, j) \ \stackrel{d}{=} \ \sum_i X(1_j, t, i) x_i$$

where we write $X(\mathbf{x}, t, j)$ for the time-$t$ information of agent $j$ when starting from configuration $\mathbf{x}$.

Intuition suggests that, separate from the "global convergence" given by Proposition 1, there is some (often initially faster) "local smoothing" when the initial configuration is not locally smooth (e.g. if it were a realization of an IID process over agents).

In our setting the natural way to measure "local smoothness" of a function $f$ is by the Dirichlet form $\mathcal{E}(f,f)$, so we study the random process $\mathcal{E}(\mathbf{X}(t), \mathbf{X}(t))$, that is the Dirichlet form applied to a realization of the averaging process.

The technique we use is to extend the "random penny"/augmented process argument used in the opening lemma. It uses a particular coupling $(Z_1(t), Z_2(t))$ of two copies of the (half-speed) associated MC, defined as the following MC on product space.

Here $i, j, k$ denote distinct agents.

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

For comparison, for two independent chains the transitions $(i,j) \to (j,i)$ and $(i,i) \to (j,j)$ are impossible and in the other transitions above all the $1/4$ terms become $1/2$. (The coupling might be regarded as a weird mixture of discrete- and continuous-time effects).

Write $\mathbf{X}^a(t) = (\mathbf{X}^a_i(t))$ for the averaging process started from configuration $1_a$.

**Key identity;** for each choice of $a, b, i, j$, not requiring distinctness,

$$\mathbb{E}(X^a_i(t) X^b_j(t)) = \mathbb{P}(Z^{a,b}_1(t) = i, Z^{a,b}_2(t) = j)$$

where $(Z^{a,b}_1(t), Z^{a,b}_2(t))$ denotes the coupled process started from $(a, b)$.

[give argument on board]

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

**Lemma (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 $\mathbf{x}(0)$.*

We will explore consequences in the special case where $\mathcal{N}$ is the rate matrix for RW on a $r$-regular graph. In this special case

$$\mathcal{E}(\mathbf{x},\mathbf{x}) = n^{-1} \sum_{e=(ij)} r^{-1}(x_i - x_j)^2$$

the sum over undirected edges; rearranging

$$n\mathcal{E}(\mathbf{x},\mathbf{x}) = \sum_i x_i^2 - 2r^{-1} \sum_{e=(ij)} x_i x_j .$$

So $n\, \mathbb{E}\mathcal{E}(\mathbf{X}(t),\mathbf{X}(t))$

$$\begin{aligned} &= \mathbb{E}\sum_i (X_i(t))^2 - 2r^{-1}\mathbb{E}\sum_{e=(ij)} X_i(t)X_j(t) \\ &= \mathbb{P}(Z_1(t) = Z_2(t)) - r^{-1}\mathbb{P}(Z_1(t) \sim Z_2(t)) \text{ by the cross-product lemma} \end{aligned}$$

where $i \sim j$ indicates the edge relationship. Now in this special case we can directly compute, from the dynamics of the coupling $(Z_1(t), Z_2(t))$,

$$\frac{d}{dt}\mathbb{P}(Z_1(t) = Z_2(t)) = -\tfrac{1}{2}\mathbb{P}(Z_1(t) = Z_2(t)) + \tfrac{1}{2r}\mathbb{P}(Z_1(t) \sim Z_2(t))$$

So we have proved

**Proposition**

*Take $\mathcal{N}$ to be the rate matrix for RW on a $r$-regular graph. For the averaging model $\mathbf{X}(t)$ started from a probability distribution $\mathbf{x}(0)$ over agents.*

$$\mathbb{E}\mathcal{E}(\mathbf{X}(t),\mathbf{X}(t)) = \frac{-2}{n}\frac{d}{dt}\mathbb{P}(Z_1(t) = Z_2(t))$$

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

Now $\mathbb{P}(Z_1(0) = Z_2(0)) = ||\mathbf{x}(0)||_2^2$ and the $t \to \infty$ limit distribution of $(Z_1(t), Z_2(t))$ is independent uniform, so integrating over $0 \le t < \infty$ gives

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

By scaling, the averaging process $\mathbf{Y}(t)$ from an initial configuration $\mathbf{y}$ with $\bar{\mathbf{y}} = 0$ satisfies (recall $\mathcal{N}$ is $r^{-1}\times$ the adjacency matrix of a $r$-regular graph; $\bar{\mathbf{y}} = 0$)

$$\mathbb{E} \int_0^\infty \mathcal{E}(\mathbf{Y}(t), \mathbf{Y}(t))\ dt = 2||\mathbf{y}||_2^2. \qquad (3)$$

This is a remarkable "universality" property of local smoothness in the averaging model.
[Discuss on board: IID starts]
**Theory project:** Is there an analogous result assuming only the the standardized setting ($\nu_i \equiv 1$)?
**Theory project:** Is there a universal bound on $\mathbb{E}\mathcal{E}(\mathbf{Y}(t), \mathbf{Y}(t))$ at each $t$?
**Theory project:** Give a sharp analysis of the behavior of the averaging process on the $n$-cycle/integers.
By analogy with theory surrounding Cheeger's inequality, it is possible that the 1-dimensional case is the "worst case" re orders of magnitude.

The following example shows that, unlike $\mathbb{E}||\mathbf{X}(t)||^2$,

$$\mathbb{E}\ \mathcal{E}(\mathbf{X}(t), \mathbf{X}(t)) \text{ is not necessarily decreasing in } t.$$

Take states $1, 2, 3, 4$ with $\nu_{12} = \nu_{34} = 1$ and $\nu_{23} = \delta$, for small $\delta$. Take initial configuration $\mathbf{x}(0) = (-1, -1, 1, 1)$, so $\mathcal{E}(\mathbf{x}(0), \mathbf{x}(0)) = \delta$. Now at the first time $T$ that the averaging process changes state, it changes to to $(-1, 0, 0, 1)$, and then $\mathcal{E}(\mathbf{X}(T), \mathbf{X}(T)) = 1/2$.

Speculation [**theory project**] – can we use the averaging process to estimate mixing times for the associated MC?

Consider the following type of algorithm. At time 0 pick 25 random agents $a$ ; from each, start the averaging process with one unit at $a$. Now each agent $i$ tracks the 25 processes $X_i^a(t),\ a = 1, \ldots, 25$. The agent continues until the first time $t$ that at least 20 of the 25 processes have $X_i^a(t) \geq \frac{4}{5n}$, then outputs that time as $T^{\text{out}}$.

Can we relate $T^{\text{out}}$ to some theoretical notion of mixing time?
[board: discuss type of algorithm]

For the usual reason (some subset of $n/100$ agents may be almost disconnected) we cannot hope to estimate $\tau_{\text{mix}}$ or $\tau_{\text{rel}}$. Instead, for the MC $Z(t)$ consider a mixing time of the form

$$\tau = \min\{t : |\mathbb{P}_a(Z(t) = i) - 1/n| \text{ is small for most pairs } (a, i)\}.$$

We will outline a possible argument that $T^{\text{out}} = O(\tau)$ – the other direction would be more interesting.

Suppose we can show that the coupling doesn't make much difference, specifically that for $t$ of order $\tau$, for most pairs $(a, i)$

$$\mathbb{P}(Z_1^{a,a}(t) = i, Z_2^{a,a}(t) = i) \approx [P_a(Z(t) = i)]^2$$

with a bound on the error. Applying the key identity, the error in the displayed approximation represents the variance of $X_i^a(t)$, so proving the error is small would prove

$$X_i^a(t) \text{ is approximately } 1/n \text{ for most pairs } (a, i).$$

So $T^{\text{out}}$ will be $O(\tau)$ for most $i$.

## The interchange [exclusion] process

The **interchange process** is the FMIE process where there are $n$ distinguishable tokens, one at each agent. When two agents meet, they exchange tokens.

This is a finite-site variant of the **exclusion process** which has been much studied as an infinite site IPS. To get the exclusion process from the interchange process, declare some of the tokens to be invisible, and declare the visible tokens to be indistinguishable particles. Then the visible particles evolve according to the exclusion process rules:

A visible particle at $i$ attempts to jump to $j$ at rate $\nu_{ij}$; the jump is aborted if $j$ is occupied by a visible particle.

These processes don't seem to naturally fit our "information-exchange" theme, so we don't emphasize them, but they may be useful for comparison purposes. Note that a single token in the interchange process moves as the associated MC. In particular, here is a recent hard result (**talk project?**).

**Theorem (Caputo-Liggett-Richthammer (2009))**

*The relaxation time $\tau_{rel}$ of the interchange process equals the relaxation time $\tau_{rel}$ of the associated MC.*

This result is loosely related to analysis of the previous "algorithmic" scheme, because it suggests that the relaxation time for the coupled process (in the key identity) should be $\tau_{\mathrm{rel}}^{MC}$. If true we could deduce $T^{\mathrm{out}} = O(\tau_{\mathrm{rel}} \log n)$. [discuss on board].

The theorem above suggests studying the (variation distance) mixing time $\tau_{\mathrm{mix}}$ for the interchange process. Recent hard results of Oliveira (2010) give bounds of the form

$$\tau_{\mathrm{mix}}^{IP} = O(\tau_{\mathrm{mix}}^{MC} \log n).$$

(**talk project?**).