Interacting particle systems as stochastic social dynamics

David Aldous

April 5, 2013

Analogy: **game theory** not about "games" (baseball, chess, ...) but about a particular setup (players choose actions separately, get payoffs) which is useful in other contexts (Google ads).

Analogously, my nominal topic is "flow of information through networks", but I'm going to specify a particular setup. Thousands of papers over the last ten years, in fields such as statistical physics; epidemic theory; broadcast algorithms on graphs; ad hoc networks; social learning theory, can be fitted into this setup. But it doesn't have a standard name – there exist names like "interacting particle systems" or "social dynamics" but these have rather fuzzy boundaries. The best name I can invent is *Finite Markov Information-Exchange* (FMIE) Processes.

A nice popular book on **game theory** (Len Fisher: *Rock, Paper, Scissors: Game Theory in Everyday Life*) illustrates the breadth of that subject by discussing 7 prototypical models with memorable names.

Prisoner's Dilemma; Tragedy of the Commons; Free Rider; Chicken; Volunteer's Dilemma; Battle of the Sexes; Stag Hunt.

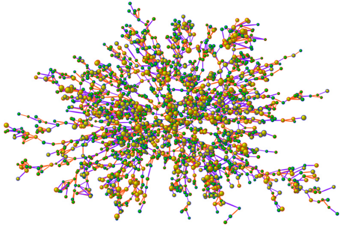
So let me describe the subject of FMIE processes via 8 prototypical and simple models with memorable names, invented for this talk because the standard names are uninformative.

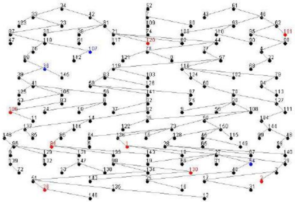
Hot Potato, Pandemic, Leveller, Pothead, Deference, Fashionista, Gordon Gekko, and Preserving Principia.

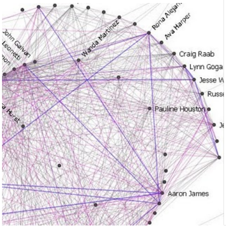
- On my web page are slides from a 2012 summer school lecture course, and a 30-page overview paper, which contains references.
- **Nothing is essentially new . . . . . .**
- Model at a high level of abstraction (= unreality!), not intended for real data.

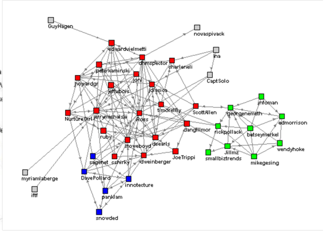
What (mathematically) is a social network?

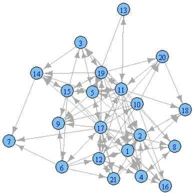
Usually formalized as a *graph*, whose vertices are individual people and where an edge indicates presence of a specified kind of relationship.











In many contexts it would be more natural to allow different strengths of relationship (close friends, friends, acquaintances) and formalize as a *weighted* graph. The interpretation of *weight* is context-dependent. In some contexts (scientific collaboration; corporate directorships) there is a natural quantitative measure, but not so in "friendship"-like contexts.

Our particular viewpoint is to identify "strength of relationship" with "**frequency of meeting**", where "meeting" carries the implication of "opportunity to exchange information".

Because we don't want to consider only social networks, we will use the neutral word **agents** for the n people/vertices. Write ν_{ij} for the weight on edge ij, the "strength of relationship" between agents i and j.

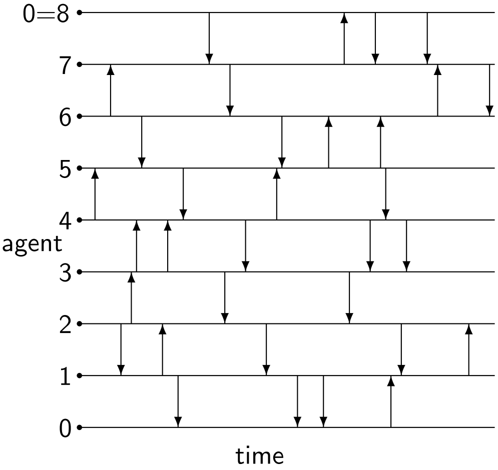
Here is the model for agents meeting (i.e. opportunities to exchange information).

- Each pair i, j of agents with $\nu_{ij} > 0$ meets at random times, more precisely at the times of a rate-ν_{ij} Poisson process.

Call this the **meeting model**. It is parametrized by the symmetric matrix $\mathcal{N} = (\nu_{ij})$ without diagonal entries.

Regard a meeting model as a "geometric substructure". One could use any geometry, but most existing literature uses variants of 4 basic geometries for which explicit calculations are comparatively easy.

Schematic – the meeting model on the 8-cycle.



The 4 popular basic geometries.

Most analytic work implicitly takes $\mathcal{N}$ as the (normalized) adjacency matrix of an unweighted graph, such as the following,

Complete graph or **mean-field**.

$$\nu_{ij} = 1/(n-1), \ j \neq i.$$

d-**dimensional grid (discrete torus)** $\mathbb{Z}_m^d; \quad n = m^d$.

$$\nu_{ij} = 1/(2d) \text{ for } i \sim j.$$

Small worlds. The grid with extra long edges, e.g. chosen at random with chance $\propto (\text{length})^{-\alpha}$.

Random graph with prescribed degree distribution. A popular way to make a random graph model to "fit" observed data is to take the observed degree distribution (d_i) and then define a model interpretable as "an n-vertex graph whose edges are random subject to having degree distribution (d_i)". This produces a locally tree-like network – unrealistic but analytically helpful.

In this talk we'll assume as a default **normalized rates**

$$\nu_i := \sum_j \nu_{ij} = 1 \text{ for all } i.$$

A natural "geometric" model is to visualize agents having positions in 2-dimensional space, and take ν_{ij} as a decreasing function of Euclidean distance. This model (different from "small wolds") is curiously little-studied, perhaps because hard to study analytically.

What is a FMIE process?

Such a process has two levels.

1. Start with a **meeting model** as above, specified by the symmetric matrix $\mathcal{N} = (\nu_{ij})$ without diagonal entries.

2. Each agent i has some "information" (or "state") $X_i(t)$ at time t. When two agents i, j meet at time t, they update their information according to some **update rule** (deterministic or random). That is, the updated information $X_i(t+), X_j(t+)$ depends only on the pre-meeting information $X_i(t-), X_j(t-)$ and (perhaps) added randomness.

The update rule is chosen based on the real-world phenomenon we are studying. A particular FMIE **model** is just a particular update rule. The general math issue is to study how the behavior of any particular model depends on the "geometry" in the meeting model.

Can't expect any substantial "general theorem" but there are five useful "**general principles**" we'll mention later.

Two models seem basic, both conceptually and mathematically.

Model: Hot Potato.

`There is one token. When the agent` i `holding the token meets another agent` j`, the token is passed to` j`.`

The natural aspect to study is $Z(t) =$ the agent holding the token at time t. This $Z(t)$ is the continuous-time Markov chain with transition rates (ν_{ij}).

As we shall see, for some FMIE models the interesting aspects of their behavior can be related fairly directly to behavior of this **associated Markov chain**, while for others any relation is not so visible.

I'll try to give one result for each model, so here is an (undergraduate homework exericise) result for Hot Potato. For the geometry take the $n = m \times m$ discrete torus. Take two adjacent agents. Starting from the first, what is the mean time for the Potato to reach the second?

Answer: $n - 1$.

Take two adjacent agents on $\mathbb{Z}_m^2$. Starting from the first, what is the mean time for the Potato to reach the second?

Answer: $n - 1$. Because
(i) Just assuming normalized rates, the symmetry $\nu_{ij} = \nu_{ji}$ implies mean return time to any agent $= n$, regardless of geometry.
(ii) Takes mean time one to leave initial agent; by symmetry of the particular graph it doesn't matter which neighbor is first visited.

Model: Pandemic.

`Initially one agent is infected. Whenever an infected agent meets another agent, the other agent becomes infected.`

Pandemic has been studied in many specific geometries, but (in contrast to the Markov chain model) there are no general theorems. I will give one specific result and one general conjecture.

The "deterministic, continuous" analog of our "stochastic, discrete" model of an epidemic is the **logistic equation**

$$F'(t) = F(t)(1 - F(t))$$

for the proportion $F(t)$ of a population infected at time t. A solution is a shift of the basic solution

$$F(t) = \frac{e^t}{1 + e^t}, \quad -\infty < t < \infty. \qquad \textbf{logistic function}$$

Distinguish initial phase when the proportion infected is $o(1)$, followed by the pandemic phase. Write $X_n(t)$ for the proportion infected. On the **complete** n**-vertex graph** geometry,

(a) During the pandemic phase, $X_n(t)$ behaves as $F(t)$ to first order.
(b) The time until a proportion q is infected is

$$\log n + F^{-1}(q) + G_n \pm o(1),$$

where G_n is a random time-shift ("founder effect").

Theorem (The randomly-shifted logistic limit)

For Pandemic on the complete n-vertex graph, there exist random G_n such that

$$\sup_t |X_n(t) - F(t - \log n - G_n)| \to 0 \text{ in probability}$$

where F is the logistic function and $G_n \xrightarrow{d} G$ with Gumbel distribution $\mathbb{P}(G \le x) = \exp(-e^{-x})$.

Pandemic can be viewed as a "dynamical" version of **first passage percolation**. Assign to edges (a, b) random lengths with Exponential (rate ν_{ab}) distribution and consider

$$T_{ij} = \text{length of shortest path } \pi_{ij} \text{ between } i \text{ and } j.$$

Then T_{ij} is the time for Pandemic started at i to reach j.

First passage percolation (with general IID distribution of edge-lengths) on the lattice $\mathbb{Z}^d$ has been well-studied. The **shape theorem** gives the first order behavior of the infected region in Pandemic: linear growth of a deterministic shape. Rigorous understanding of second order behavior is a famous hard problem.

The essence of the shape theorem is that T_{ij} is close (first-order) to its expectation. Here is a conjecture for arbitrary geometries.

$\xi_{ab} =$ length of edge (a, b) has Exponential (rate ν_{ab}) distribution

$T_{ij} =$ length of shortest path π_{ij} between i and j.

Conjecture

With arbitrary rates (ν_{ij}), if (in a sequence of geometries)

$$\frac{\max\{\xi_{ab} : \ (a, b) \text{ edge in } \pi_{ij}\}}{\mathbb{E}\, T_{ij}} \to_p 0 \tag{1}$$

then

$$\frac{T_{ij}}{\mathbb{E}\, T_{ij}} \to_p 1$$

Easy to show (1) is necessary.

Model: Leveller.

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 obviously each agent's fortune $X_i(t)$ will converge to the overall average. Note 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 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 $1/2$.

It turns out to be easy to quantify global convergence to the average.

Proposition (Global convergence in Leveller)

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 λ *is the spectral gap of the associated MC.*

Results like this have appeared in several contexts, e.g. gossip algorithms. Here is a more subtle result. Suppose normalized meeting rates. Because an agent interacts with nearby agents, guess that some sort of "local averaging" occurs independent of the geometry.

For a "test function" $g : \textbf{Agents} \to \mathbb{R}$ write

$$\begin{aligned}\bar{g} &= n^{-1} \sum_i g_i \\ ||g||_2^2 &= n^{-1} \sum_i \pi_i g_i^2 \\ \mathcal{E}(g,g) &= n^{-1} \tfrac{1}{2} \sum_i \sum_{j \neq i} \nu_{ij} (g_j - g_i)^2 \quad \text{(the Dirichlet form).}\end{aligned}$$

When $\bar{g} = 0$ then $||g||_2$ measures "global" variability of g whereas $\mathcal{E}(g,g)$ measures "local" variability relative to the underlying geometry.

Proposition (Local smoothness in Leveller)

For normalized meeting rates associated with a r-regular graph; and initial $\bar{\mathbf{x}} = 0$,

$$\mathbb{E} \int_0^\infty \mathcal{E}(\mathbf{X}(t), \mathbf{X}(t))\, dt = 2||\mathbf{x}||_2^2. \quad (3)$$

Model: Pothead.

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.

Officially called the **voter model** (VM). Very well studied. View as "paradigm example" of a FMIE; can be used to illustrate all 5 of the "general principles".

We 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.

General principle 1. If an agent has only a finite number of states, the the total number of configurations is finite, so elementary Markov chain theory tells us qualitative asymptotics.

Here "all agents have opinion i" are the absorbing configurations – the process must eventually be absorbed in one. A natural quantity of interest is the **consensus time**

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

General principle 2. Time-reversal duality.

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 $i \to j$, 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.

The duality relationship.

For fixed t,

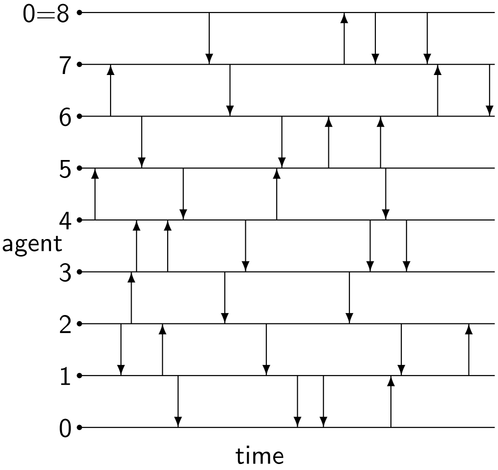
$$\{\mathcal{V}_i(t), i \in \textbf{Agents}\} \stackrel{d}{=} \{\mathcal{C}_i(t), i \in \textbf{Agents}\}.$$

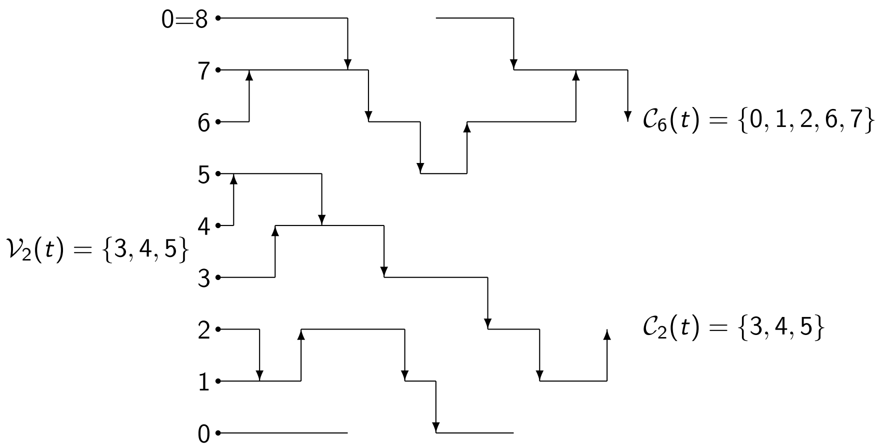
In particular $T^{\text{voter}} \stackrel{d}{=} T^{\text{coal}}$.
They are different as processes. For fixed i, note that $|\mathcal{V}_i(t)|$ can only change by ± 1, but $|\mathcal{C}_i(t)|$ jumps to and from 0.

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

Note the time-reversal argument 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
$\mathcal{C}_6(t) = \{0, 1, 2, 6, 7\}$
$\mathcal{V}_2(t) = \{3, 4, 5\}$
$\mathcal{C}_2(t) = \{3, 4, 5\}$

Random walk (RW) on $\mathbb{Z}^d$ is a classic topic in mathematical probability. Can analyze CRW model to deduce, on $\mathbb{Z}_m^d$ as $m \to \infty$ in fixed $d \geq 3$

$$\mathbb{E}\, T^{\text{voter}} = \mathbb{E}\, T^{\text{coal}} \sim c_d m^d = c_d n.$$

Very easy to show directly in the CRW model on complete graph that

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

There is a different analysis of VM on **complete graph**, by first considering only two initial opinions. The process

$$N(t) = \text{number with first opinion}$$

evolves as the continuous-time MC on states $\{0, 1, 2, \ldots, n\}$ with rates

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

Leads to an upper bound on complete graph

$$\mathbb{E}\, T^{\text{voter}} \leq (4 \log 2) n.$$

Moral of general principle 2: Sometimes the dual process is easier to analyze.

General principle 3. One can often get (maybe crude) bounds on the behavior of a given model on a general geometry in terms of **bottleneck statistics** for the rates (ν_{ij}).

Define κ as the largest constant such that

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

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

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

General principle 4. For many simple models there is some specific aspect which is "invariant" in the sense of depending only on n, not on the geometry.

Already noted for Hot Potato and for Leveller. For Pothead,

$$\text{mean number opinion changes per agent} = n - 1.$$

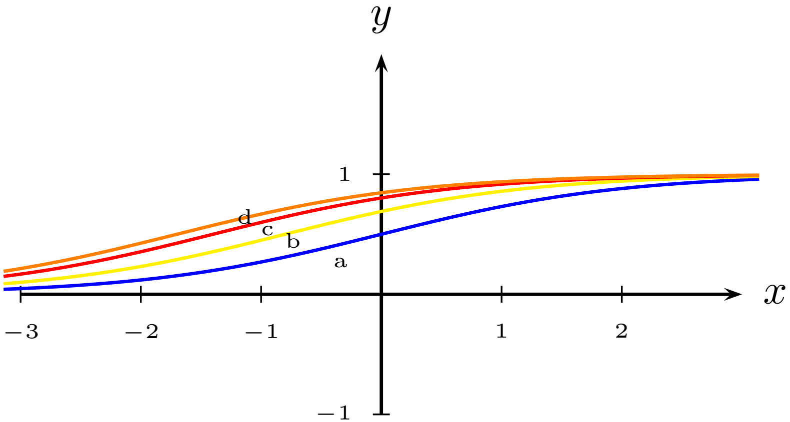
Model: Deference

(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.

Clearly opinion 1 spreads as Pandemic, so the "ultimate": behavior of Deference is not a new question. A challenging open problem is what one can deduce about the geometry (meeting process) from the short term behavior of Deference.

Easy to give analysis in **complete graph** model, as a consequence of the "randomly-shifted logisic" result for Pandemic. Study $(X_1^n(t), \ldots, X_k^n(t))$, where $X_k^n(t)$ is the proportion of the population with opinion k at time t.

Key insight: opinions 1 and 2 combined behave as one infection in Pandemic, hence as a random time-shift of the logistic curve F.



y
x
1
−1
−3
−2
−1
1
2
a
b
c
d

So we expect $n \to \infty$ limit behavior of the form

$$((X_1^n(\log n+s), X_2^n(\log n+s), \dots, X_k^n(\log n+s)), \ -\infty < s < \infty) \to \quad (4)$$

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

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

We can determine the C_j by the fact that in the initial phase the different opinions spread independently. It turns out

$$C_j = \log(\xi_1 + \dots + \xi_j), \ j \geq 1 \quad (5)$$

where $(\xi_i, i \geq 1)$ are IID Exponential(1).

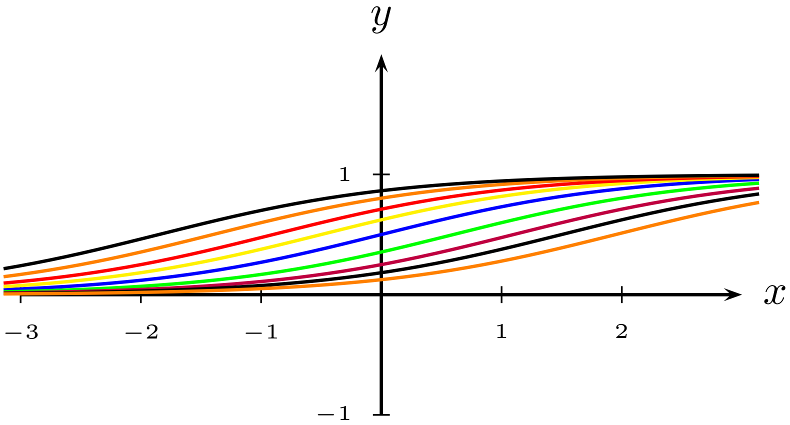
The Deference model envisages agents as "slaves to authority". Here is a conceptually opposite "slaves to fashion" model, whose analysis is mathematically surprisingly similar.

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 must be some equilibrium distribution, for the random partition of agents into "same fashion".

For the **complete graph** geometry, we can copy the analysis of Deference. Combining all the fashions apparing after a given time, these behave (essentially) as one infection in Pandemic (over the pandemic window), hence as a random time-shift of the logistic curve F. So when we study the vector $(X_k^n(t), -\infty < k < \infty)$ of proportions of agents adopting different fashions k, we expect $n \to \infty$ limit behavior of the form



y
x
1
−1
−3
−2
−1
1
2

$$(X_k^n(\log n + s),\ -\infty < k < \infty) \to \tag{6}$$
$$(F(C_k + s) - F(C_{k-1} + s),\ -\infty < k < \infty)$$

where $(C_k,\ -\infty < k < \infty)$ are the points of some stationary process on $(-\infty, \infty)$.

Knowing this form for the $n \to \infty$ asymptotics, we can again determine the distribution of (C_i) by considering the initial stage of spread of a new fashion. It turns out that

$$C_i = \log\left(\sum_{j \le i} \exp(\gamma_j)\right) = \gamma_i + \log\left(\sum_{k \ge 1} \exp(\gamma_{i-k} - \gamma_i)\right). \tag{7}$$

where η_j are the times of a rate-λ Poisson process.

The FMIE models I've shown were chosen as representative of the "mathematical fundamental" ones, but hundreds of others have been studied, and it's easy to invent your own model (my student Dan Lanoue is studying the **iPod model**). Here's another direction.

Game-theoretic aspects of FMIE processes

Our FMIE setup rests upon a **given** matrix (ν_{ij}) of meeting rates. We can add an extra layer to the model by taking as basic a given matrix (c_{ij}) of meeting **costs**. This means that for i and j to meet at rate ν_{ij} incurs a cost of $c_{ij}\nu_{ij}$ per unit time. Now we can allow agents to **choose** meeting rates, either
[reciprocal] i and j agree on a rate ν_{ij} and share the cost
[unilateral] i can choose a "directed" rate ν_{ij} but pays all the cost.

One can now consider models of the following kind. Information is spread at meetings, and there are benefits associated with receiving information. Agents seek to maximize their payoff = benefit - cost.

Our setup is rather different from what you see in a Game Theory course.

- $n \to \infty$ agents; rules are symmetric.
- allowed strategies parametrized by real θ.
- Distinguish one agent **ego**.
- $\mathrm{payoff}(\phi, \theta)$ is payoff to **ego** when **ego** chooses ϕ and all other agents choose θ.
- payoff is "per unit time" in ongoing process.

The Nash equilibrium value θ^{Nash} is the value of θ for which **ego** cannot do better by choosing a different value of ϕ, and hence is the solution of

$$\frac{d}{d\phi} \mathrm{payoff}(\phi, \theta) \Big|_{\phi=\theta} = 0. \tag{8}$$

So we don't use any Game Theory – we just need a formula for $\mathrm{payoff}(\phi, \theta)$.

Model: Gordon Gecko game

The model's key feature is **rank based rewards** – toy model for gossip or insider trading.

- New items of information arrive at times of a rate-1 Poisson process; each item comes to one random agent.

Information spreads between agents in ways to be described later [there are many variants], which involve communication costs paid by the *receiver* of information, but the common assumption is

- The j'th person to learn an item of information gets reward $R(\frac{j}{n})$.

Here $R(u),\ 0 < u \leq 1$ is a decreasing function with

$$R(1) = 0; \quad 0 < \bar{R} := \int_0^1 R(u)du < \infty.$$

Note the total reward from each item is $\sum_{j=1}^{n} R(\frac{j}{n}) \sim n\bar{R}$. That is, the average reward per agent per unit time is $\bar{R}$.

Because average reward per unit time does not depend on the agents' strategy, the "social optimum" protocol is for agents to communicate slowly, giving payoff arbitrarily close to $\bar{R}$. But if agents behave selfishly then one agent may gain an advantage by paying to obtain information more quickly, and so we seek to study Nash equilibria for selfish agents.

Instead of taking the geometry as the complete graph or discrete torus $\mathbb{Z}_m^2$, let's jump to the more interesting "Ma Bell" geometry. That is

The $m \times m$ torus with short and long range interactions

Geometry model. The agents are at the vertices of the $m \times m$ torus. Each agent i may, at any time, call any of the 4 neighboring agents j (at cost 1), or call any other agent j at cost $c_m \geq 1$, and learn all items that j knows.

Poisson strategy. An agent's strategy is described by a pair of numbers $(\theta_{\text{near}}, \theta_{\text{far}}) = \theta$:

at rate θ_{near} the agent calls a random neighbor
at rate θ_{far} the agent calls a random non-neighbor.

This model obviously interpolates between the complete graph model ($c_m = 1$) and the nearest-neighbor model ($c_m = \infty$). It turns out the interesting case is

$$1 \ll c_m \ll m^2.$$

We have to analyze Pandemic on this geometry, to get a formula for $\mathrm{payoff}(\phi, \theta)$; then doing the calculus it turns out

$$\theta^{\text{Nash}}_{\text{near}} \text{ is order } c_m^{-1/2} \text{ and } \theta^{\text{Nash}}_{\text{far}} \text{ is order } c_m^{-2}.$$

In particular the Nash cost $\asymp c_m^{-1/2}$ and the Nash equilibrium is efficient.