Exchangeability and Continuum Limits

David Aldous

June 24, 2013

This talk is a fast overview of several topics relating to exchangeability. I will give references to topics 1-4; topic 5 is work in progress.

1. Structure theory for exchangeable arrays
2. A general program for continuum limits of discrete random structures . . .
3. . . . illustrated by dense graph limits and by measured metric spaces
4. A conjectured compactification of some finite reversible Markov chains
5. The compulsive gambler and the metric coalescent

Structure theory for exchangeable arrays

Take independent Uniform$(0,1)$ random variables

$$U, \ (U_i, 1 \leq i < \infty), \ (U_{ij}, 1 \leq i < j < \infty)$$

and then set $U_{ji} = U_{ij}$ for $j > i$.
Take measurable $f : [0,1]^4 \to S$, symmetric in middle two arguments, define

$$X_{ij} = f(U, U_i, U_j, U_{ij}), \ i \neq j. \tag{1}$$

This gives an infinite random array which is symmetric ($X_{ij} = X_{ji}$) and has the **jointly exchangeable** property

$$(X_{ij}, i \neq j) \ \stackrel{d}{=} \ (X_{\pi(i)\pi(j)}, i \neq j) \text{ for each finite permutation } \pi \text{ of } \mathbb{N}.$$

Theorem (The representation theorem)

(a) Each infinite symmetric jointly exchangeable array of RVs is distributed as the array (1) for some f.
(b) (informally) The representing function f is unique up to replacing each U_i by $\phi(U_i)$ for measure-preserving $\phi : [0,1] \to [0,1]$, and similarly for U and U_{ij}.

This result became known around 1980. If forced to attach names, I attach Hoover-Aldous-Kallenberg. Olav's contributions have been

- Clear statement and proof of the uniqueness assertion (1989).
- Many variants (1988-95).
- Definitive monograph *Probabilistic symmetries and invariance principles* (2005).

Three remarks:

- The representation theorem is purely measure-theoretic; the range space S is a general Borel space. No topology involved.
- Concise recent treatment in Tim Austin's *Exchangeable Random Arrays*.
- Around 1980 others (Kingman; Diaconis-Freedman; Dawid; Lauritzen) were interested for various reasons (e.g. Bayesian statistics). The topics that follow in this talk were not anticipated then.

Continuum limits of discrete random structures

I will outline a "general program", where "general" $\neq$ "always works" but instead means "works in various settings that otherwise look different". For some different settings see my survey paper *More Uses of Exchangeability: Representations of complicated Random Structures* in the 2010 Kingman Festschrift.

Rather obvious idea:

> *One way of examining a complicated mathematical structure equipped with a probability measure is to sample IID random points and look at some form of induced substructure relating the random points*

which assumes we are **given** the complicated structure.

Less obvious idea:

> *We can often use exchangeability in the* **construction** *of complicated random structures as the $n \to \infty$ limits of random finite n-element structures $\mathcal{G}(n)$.*

What's the point? Use when there's no natural way to think of each $\mathcal{G}(n)$, as n varies, as taking values in the **same** space.

To expand the idea:

> *Within the n-element structure $\mathcal{G}(n)$ pick k IID random elements, look at an induced substructure on these k elements – call this $\mathcal{S}(n,k)$ – taking values in some space $S_{(k)}$ that depends on k but not n. Take a limit (in distribution) as $n \to \infty$ for fixed k, any necessary rescaling having been already done in the definition of $\mathcal{S}(n,k)$ – call this limit $\mathcal{S}_k$. Within the limit random structures $(\mathcal{S}_k, 2 \le k < \infty)$, the k elements are exchangeable, and the distributions are consistent as k increases and therefore can be used to* **define** *an infinite structure $\mathcal{S}_\infty$.*

Where one can implement this program, the random structure $\mathcal{S}_\infty$ will for many purposes serve as a $n \to \infty$ limit of the original n-element structures. Note that $\mathcal{S}_\infty$ makes sense as a rather abstract object, via the Kolmogorov extension theorem, but in concrete cases one tries

- to identify $\mathcal{S}_\infty$ with some more concrete construction
- to characterize all possible limits of a given class of finite structures.

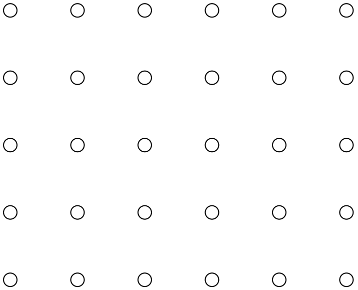
Follow-up idea:

> *This implicitly gives a topology on the space of all "complicated mathematical structures equipped with a probability measure" that we are studying.*

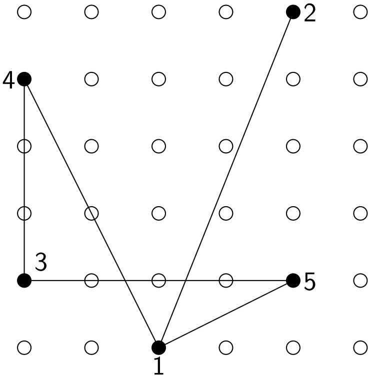
That is, a sequence of structures converges if, for each k, the induced substructures on k sampled elements converge.

Dense graph limits

Suppose that for each n there is a graph G_n on n vertices.
But we don't see the edges of G_n.
Instead we can sample k random vertices and see the induced subgraph on the sampled vertices.

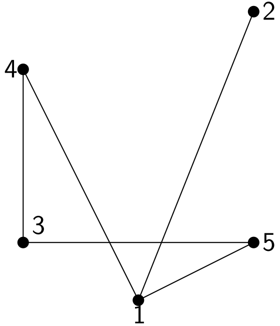


not square grid!



not square grid!

Induced subgraph $\mathcal{S}(n,k)$ on k of the n vertices of G_n.



One sense of "convergence" of graphs G_n is that for each fixed k the random subgraphs $\mathcal{S}(n,k)$ converge in distribution to some limit $\mathcal{S}(\infty,k)$.

This notion of "dense graph convergence" was introduced (in superficially different form) by Lovász - Szegedy (2006), and has attracted a large literature. In one sense it's atypical of the general methodology in that it applies to deterministic n-vertex graphs and gives a compactification; its use is in extremal graph theory, **not** in random graph theory.

Convergence of measured metric spaces

Consider the case where a "structure" is a *measured metric space (MMS)* written (S, d, μ); that is,

- (S, d) is a complete separable metric space;
- μ is a probability measure on S.

Take $(\xi_i, 1 \leq i < \infty)$ i.i.d. (μ) and consider the infinite random array

$$X_{ij} = d(\xi_i, \xi_j), \ 1 \leq i, j < \infty$$

Say (S, d, μ) and (S', d', μ') are *equivalent* is there is a measure-preserving isometry between them. Clearly, for equivalent MMSs we have

$$(X_{ij}, \ 1 \leq i, j < \infty) \ \stackrel{d}{=} \ (X'_{ij}, \ 1 \leq i, j < \infty)$$

and the uniqueness part of the representation theorem implies the converse: the distribution of the array $(X_{ij}, \ 1 \leq i, j < \infty)$ determines the MMS up to equivalence.

Consider now a sequence (S^n, d^n, μ^n) of MMSs. In terms of the arrays $\mathbf{X}^n$ defined above

$$X^n_{ij} = d^n(\xi^n_i, \xi^n_j), \ 1 \leq i, j < \infty$$

we can define a notion of convergence to a limit MMS

$$(S^n, d^n, \mu^n) \to (S^\infty, d^\infty, \mu^\infty) \text{ is defined to mean } \mathbf{X}^n \ \stackrel{d}{\to} \ \mathbf{X}^\infty.$$

It is not obvious that this is equivalent to convergence in some natural metric on the space of all (equivalence classes of) MMSs, but this is true, for the *Gromov-Prohorov metric*.

The equivalance is implicit in work of Gromov (with quite different motivations); a clearer treatment for probabilists is in Greven - Pfaffelhuber - Winter (2009). One "probabilistic" motivation involves continuum random trees and their generalizations. An n-vertex tree can be regarded as a metric space on n points by taking edge-lengths $= n^{-1/2}$; including the uniform distribution on the n vertices makes it a MMS. For various models of "uniform random n-vertex tree" $\mathcal{T}_n$ we have

$$\mathcal{T}_n \stackrel{d}{\to} \mathcal{T}_\infty \quad (2)$$

where the limit "Brownian CRT" can be constructed explictly from Brownian excursion, but is abstractly a random MMS.

Note the point: the **realizations** of $\mathcal{T}_n$ and $\mathcal{T}_\infty$ are MMSs, so to fit (2) into the usual theory of weak convergence we want a nice metric topology on the space of all MMSs.

A conjectured compactification of some finite reversible Markov chains

[There is a write-up on my Talks web page].

In rather vague words, the conjecture is

> *Given a sequence of n-state reversible chains which does not have the L^2 cutoff property, there is a subsequence in which, after relabeling states, the transition densities converge to those of some limit general-state-space reversible Markov process.*

We emphasize that the *n*-state chains are arbitrary in the sense that we do not assume any connection between the chains as *n* varies. The conjectured behavior is a compactness assertion, in the spirit of the Lovász - Szegedy work on dense graph limits.

For simplicity we work with uniform stationary distributions, but we anticipate that the general reversible case will be similar.

Consider an n-state irreducible continuous-time Markov chain with symmetric transition rate matrix. Write $p(i,j;t)$ for the transition probabilities $\mathbb{P}(X(t) = j | X(0) = i)$. Consider the function

$$G(t) := \sum_i p(i,i;t).$$

The basic convergence theorem implies $G(t) \to 1$ as $t \to \infty$, and the spectral representation gives the more detailed result

$$G(t) = 1 + \sum_{u=2}^{n} e^{-\lambda_u t} \qquad (3)$$

where $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n$ are the eigenvalues associated with the transition rate matrix. One can regard the time τ at which $G(\tau) = 3/2$ as one of many possible notions of "mixing time". Rescaling time by this τ, we can standardize according to the convention

$$G(1) = 3/2.$$

The notion of L^2 cutoff studied in detail in (Chen - Saloff-Coste 2010) is, in our context, the property that a sequence of time-standardized chains has

$$G^{(n)}(t) \to \infty,\ t < 1; \quad G^{(n)}(t) \to 1,\ t > 1. \tag{4}$$

Now imagine a continuous-space analog. That is, a probability measure π on a space S and an S-valued Markov process $X^*(t)$ such that for $t > 0$ there exist transition densities

$$p^*(x, \cdot; t) = \text{density of } \mathbb{P}(X^*(t) \in \cdot | X^*(0) = x) \text{ w.r.t. } \pi$$

which are symmetric: $p^*(x, y; t) = p^*(y, x; t)$. The analog of $G(t)$ is

$$G^*(t) := \int_S p^*(x, x; t)\ \pi(dx).$$

Assume $G^*(t) < \infty$ for $t > 0$. Then we expect the analog of (3)

$$G^*(t) = 1 + \sum_{u=2}^{\infty} e^{-\lambda_u^* t} \tag{5}$$

where $0 = \lambda_1^* < \lambda_2^* \leq \lambda_3^* \leq \ldots$ are the eigenvalues associated with the appropriate generator. And we can standardize to make $G^*(1) = 3/2$.

Now consider a sequence of chains with $n \to \infty$, where n is the number of states, but without assuming any relation between the chains as n varies, except for the "no L^2 cutoff" assumption

$$\sup_n G^{(n)}(t) < \infty \ \forall 0 < t < 1. \tag{6}$$

As a standard analytic fact, because each $G^{(n)}$ is of form (3) there is a subsequence in which $G^{(n)}(\cdot) \to G^*(\cdot)$ for some limit function

$$G^*(t) = 1 + \sum_{u=2}^{\infty} e^{-\lambda_u^* t}$$

of form (5). This starts to hint at what is going on; the conjectured limit continuous-space process will be one with this function $G^*(\cdot)$.

Isomorphic processes

Consider a Markov process on a measurable space S with stationary distribution π, which we will view naively as a family of symmetric densities $p^*(\cdot, \cdot; t)$ satisfying the Chapman-Kolmogorov relations, with a UTC (unspecified technical condition) on the $t \downarrow 0$ behavior. Analogous to our treatment of MMSs, define an infinite partially exchangeable random array (whose entries are functions of t) by

$$\text{take i.i.d.}(\pi) \text{ random elements } (\xi_i, 1 \le i < \infty) \text{ of } S$$

$$\text{set } X^*_{ij} = p^*(\xi_i, \xi_j; t), \ \ i, j \ge 1. \tag{7}$$

This array has some distribution Ψ. There is a natural notion of "isomorphism" between two stationary Markov processes X^1 and X^2 on different spaces S_1 and S_2: processes are isomorphic if there exists a bijection $\phi : S_1 \to S_2$ that preserves joint distributions

$$(\phi(X^1_0), \phi(X^1_t)) \ \stackrel{d}{=} \ (X^2_0, X^2_t)$$

and hence preserves transition densities. And as before it is obvious that, for two isomorphic processes, we get the same Ψ.

Conjecture

If two symmetric Markov processes (on different spaces) have the same Ψ *then they are isomorphic.*

Convergence of processes

Now do exactly the same array construction for chains on finite sets S^n:

- take i.i.d. uniform random elements $(\xi_i, 1 \le i < \infty)$ of S^n
- set $X^n_{ij} = p_n(\xi_i, \xi_j; t),\ i, j \ge 1$.

Time-standardize, and recall that the "no L^2 cutoff" assumption lets us assume

$$G^n(t) \to G^*(t) \text{ as } n \to \infty \tag{8}$$

for some limit function with $1 < G^*(t) < \infty$ for $0 < t < \infty$. This is just saying that $\mathbb{E}X^n_{11} \to G^*(\cdot)$. Now an easy argument gives $\mathbb{E}X_{12} \le \mathbb{E}X_{11}$, and so we can take a subsequence in which

$$(X^n_{ij}, i, j \ge 1) \to_d (X^*_{ij}, i, j \ge 1) \text{ as } n \to \infty \tag{9}$$

(in the usual sense of convergence of finite sub-arrays) for some limit random function-valued array.

Conjecture

For any array $(X^*_{ij}, i, j \ge 1)$ that arises as a limit (9) from finite chains, there exists a general-space chain with some transition densities p^* such that the representation (7) holds.

The compulsive gambler process

There is a general setup: "interacting particle systems reinterpreted as stochastic social dynamics".

- n agents; each in some state $\in S$
- each pair of agents (i,j) meet at times of a Poisson process of given rate ν_{ij}
- enter such a meeting in states $(X_i(t-), X_j(t-))$, leave in states $(X_i(t+), X_j(t+))$ given by some deterministic or random rule $F : S \times S \to S \times S$.

As a simple but less-familar example, in the **averaging process** (Aldous-Lanoue 2012) we take $S = \mathbb{R}$ as money and when agents meet they share their money equally.

The "process" is the rule; and we seek to study how (non-asymptotic) behavior depends on the finite meeting rates (ν_{ij}). Analogous to study of mixing times for finite Markov chains.

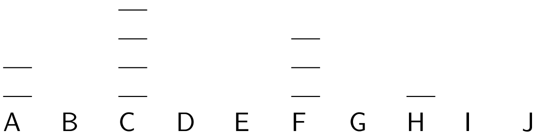
In the **Compulsive Gambler** process, agents initially have a 100-krone note each. When two agents with non-zero money meet, they instantly play a fair game in which one wins the other's money.

Interesting as methodology: there are 4 techniques which are useful for studying this process. (Work in progress, with grad student Dan Lanoue).

1. Martingales.
2. Comparison with the Kingman Coalescent chain (which is the mean-field model $\nu_{ij} \equiv 1$) for number of agents with non-zero money.

3. Imagine the initial currency notes have IID random serial numbers. The Compulsive Gambler process is the same (unconditionally, if you don't see the serial numbers) as the process in which the winner of each bet is determined as the possessor of the lowest serial number note.

4. If we run the process as above (determined by serial numbers) and see how much money each agent has at time t



then the allocation of the ordered serial numbers $(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$ is uniform random.

This is an exchangeability property, reminiscent of the theory of exchangeable coalescents (Bertoin et al).

Here is one of several directions of our "work in progress".

Place the n agents at positions $s_1, \ldots, s_n$ in a metric space (S, d). Define meeting rates as a decreasing function of distance

$$\nu_{ij} = \phi(d(s_i, s_j)).$$

Rescale so there unit money in total; initially each agent has $1/n$ money. Can now regard state space of CG process as the space $\mathcal{P}(S)$ of probability measures on S. Given a continuous distribution $\mu \in \mathcal{P}(S)$, choose $(s_i, 1 \leq i < \infty)$ such that

$$\mu^{(n)} := \text{ empirical dist.}(s_1, \ldots, s_n) \quad \to \mu.$$

Natural to guess that the CG processes started from $\mu^{(n)}$ converge to some limit process, which at times $t > 0$ has locally finite support but which converges to μ as $t \downarrow 0$. We call this the *metric coalescent*.

Key observation: one of our previous techniques implies a construction of the finite-agent CG process (based on uniform random order) which extends to a construction of the metric coalescent (based on IID samples from μ).

3. Imagine the initial currency notes have IID random serial numbers. The Compulsive Gambler process is the same (unconditionally, if you don't see the serial numbers) as the process in which the winner of each bet is determined as the possessor of the lowest serial number note.

The detailed argument involves rather subtle exchangeability properties of the construction.