2

More uses of exchangeability: representations of complex random structures

David J. Aldous[a]

Abstract

We review old and new uses of exchangeability, emphasizing the general theme of exchangeable representations of complex random structures. Illustrations of this theme include processes of stochastic coalescence and fragmentation; continuum random trees; second-order limits of distances in random graphs; isometry classes of metric spaces with probability measures; limits of dense random graphs; and more sophisticated uses in finitary combinatorics.

AMS subject classification (MSC2010) 60G09, 60C05, 05C80

1 Introduction

Kingman's write-up [44] of his 1977 Wald Lectures drew attention to the subject of exchangeability, and further indication of the topics of interest around that time can be seen in the write-up [3] of my 1983 Saint-Flour lectures. As with any mathematical subject, one might expect some topics subsequently to wither, some to blossom and new topics to emerge unanticipated. This Festschrift paper aims, in informal lecture style,

(a) to recall the state of affairs 25 years ago (sections 2.1–2.3, 3.1);
(b) to briefly describe three directions of subsequent development that

[a] Department of Statistics, Evans Hall, University of California at Berkeley, Berkeley, CA 94720-3860; aldous@stat.berkeley.edu; `http://www.stat.berkeley.edu/users/aldous`. Research supported by N.S.F. Grant DMS-0704159.

have recently featured prominently in monographs [16, 43, 52] (sections 2.4, 3.1–3.2);

(c) to describe 3 recent rediscoveries, motivated by new theoretical topics outside mainstream mathematical probability, of the theory of representations of partially exchangeable arrays (sections 2.5, 5.1–5.2);

(d) to emphasize a general program that has interested me for 20 years. It doesn't have a standard name, but let me here call it *exchangeable representations of complex random structures* (section 4).

The survey focusses on mathematical probability; although the word *Bayesian* appears several times, I do not attempt to cover the vast territory of explicit or implicit uses of exchangeability in Bayesian statistics, except to mention here its use in *hierarchical models* [20, 36].

This article is very much a bird's-eye view. Of the monographs mentioned above, let me point out Pitman's *Combinatorial Stochastic Processes* [52], which packs an extraordinary number of detailed results into 200 pages of text and exercises. Exchangeability is a recurring theme in [52], which covers about half of the topics we shall mention (and much more, not related to exchangeability), and so [52] is a natural starting place for the reader wishing to get to grips with details.

2 Exchangeability

2.1 de Finetti's theorem

I use *exchangeability* to mean, roughly, 'applications of extensions of de Finetti's theorem'. Let me assume only that the reader is familiar with the definition of an exchangeable sequence of random variables

$$(Z_i, i \geq 1) \stackrel{d}{=} (Z_{\pi(i)}, i \geq 1) \text{ for each finite permutation } \pi$$

and with the common verbal statement of de Finetti's theorem:

> An infinite exchangeable sequence is distributed as a mixture of i.i.d. sequences.

Scanning the graduate-level textbooks on measure-theoretic probability on my bookshelves, the theorem makes no appearance in about half, a brief appearance in others [18, 31, 46] and only three [22, 35, 42] devote much more than a page to the topic.

A remarkable feature of de Finetti's theorem is that there are many ways to state essentially the same result, depending on the desired emphasis. This feature is best seen when you state results more in words rather than symbols, so that's what I shall do. Take a nice space S, and either don't worry what 'nice' means, or assume $S = \mathbb{R}$. Write $\mathcal{P}(S)$ for the space of probability measures on S. Write μ for a typical element of $\mathcal{P}(S)$ and write α for a typical *random* element of $\mathcal{P}(S)$, that is a typical *random measure*. When we define an infinite exchangeable sequence of S-valued random variables we are really defining an *exchangeable measure* (Θ, say) on $\mathcal{P}(S^\infty)$, where Θ is the distribution of the sequence.

Functional analysis viewpoint. The subset $\{\mu^\infty = \mu \times \mu \times \mu \times \ldots : \mu \in \mathcal{P}(S)\} \subset \mathcal{P}(S^\infty)$ is the set of extreme points of the convex set of exchangeable elements of $\mathcal{P}(S^\infty)$, and the identification

$$\Theta(\cdot) = \int_{\mathcal{P}(S)} \mu^\infty(\cdot)\ \Lambda(d\mu)$$

gives a bijection between probability measures Λ <u>on</u> $\mathcal{P}(S)$ (that is, $\Lambda \in \mathcal{P}(\mathcal{P}(S))$) and exchangeable measures Θ in $\mathcal{P}(S^\infty)$.

Probability viewpoint. Here are successively more explicit versions of the same idea. Let $(Z_i,\ 1 \le i < \infty)$ be exchangeable S-valued.

(a) Conditional on the tail (or invariant or exchangeable) σ-field of the sequence (Z_i), the random variables Z_i are i.i.d.

(b) There exists a random measure α such that, conditional on $\alpha = \mu$, the random variables Z_i are i.i.d.(μ).

(c) Giving $\mathcal{P}(S)$ the topology of weak convergence, the empirical measure $F_n = F_n(\omega, \cdot) = n^{-1} \sum_{i=1}^n 1_{(Z_i(\omega) \in \cdot)}$ converges a.s. to a limit random measure $\alpha(\omega, \cdot)$ satisfying (b).

Theoretical statistics viewpoint. In contexts where a frequentist would assume data are i.i.d.(μ) from an unknown distribution μ, a Bayesian would put a prior distribution Λ on possible μ; so de Finetti's theorem is saying that the Bayesian assumption is logically equivalent to the assumption that the data $(Z_i, i \ge 1)$ are exchangeable. Note a mathematical consequence. There is a posterior distribution $\Lambda_n(\omega, \cdot) \in \mathcal{P}(\mathcal{P}(S))$ for Λ given $(Z_1, \ldots, Z_n)$, and an extension of (c) above is

(d) $\Lambda_n(\omega, \cdot) \to \delta_{\alpha(\omega, \cdot)}$ a.s. in $\mathcal{P}(\mathcal{P}(S))$.

Such results are historically often used as a starting point for philosophical and mathematical discussion of consistency/inconsistency of frequentist and Bayesian methods, inspired of course by Bruno de Finetti himself.

But there's more! Later we encounter at least two further, somewhat different, viewpoints: explicit constructions (section 2.7), and our central theme of using exchangeability to describe complex structures (sections 3 and 4). This theme is related to the general features that

- exchangeable-like properties are preserved under weak convergence;
- parallel to representation theorems for infinite exchangeable-like structures, are convergence theorems giving necessary and sufficient condition for finite exchangeable-like structures to converge in distribution to an infinite such structure.

In the setting of de Finetti's theorem, the condition for finite exchangeable sequences $\mathbf{X}^{(n)} = (X_1^{(n)}, \ldots, X_n^{(n)})$ to converge in distribution to an infinite exchangeable sequence $\mathbf{X}$ is

$$\alpha_n \stackrel{d}{\to} \alpha \text{ on } \mathcal{P}(S)$$

where α is the 'directing' random measure for $\mathbf{X}$ in (b) above, and α_n is the empirical distribution of $(X_1^{(n)}, \ldots, X_n^{(n)})$. Note that when we talk of convergence in distribution to infinite sequences or arrays, we mean w.r.t. product topology, i.e. convergence of finite restrictions.

2.2 Exchangeability, 25 years ago

Here I list topics from the two old surveys [44, 3], for the purpose of saying a few words about those topics I will *not* mention further, while pointing to sections where other topics will be discussed further.

Classical topics not using de Finetti's theorem.

(a) Combinatorial aspects for classical stochastic processes, e.g. ballot theorems: [56].

(b) Weak convergence for 'sampling without replacement' processes (e.g. [17] Thm 24.1).

Variants of de Finetti's theorem. Several variants were already classical, for instance:

(c) Schoenberg's[1] theorem ([3](3.6)) for the special case of spherically symmetric sequences;
(d) the analogous representation ([3](3.9)) in the setting of two sequences $(X_i, 1 \le i < \infty;\ Y_j, 1 \le j < \infty)$ whose joint distribution is invariant under finite permutations of either;
(e) the *selection property* ([44] p.188), that the exchangeability hypothesis in de Finetti's theorem can be weakened to the assumption

$$(X_1, X_2, \ldots X_n) \stackrel{d}{=} (X_{k_1}, X_{k_2}, \ldots, X_{k_n}) \quad \text{for all } 1 \le k_1 < k_2 < \cdots < k_n.$$

Other variants had been developed in the 1970s, for instance:

(f) the analog for continuous-time processes with exchangeable *increments* [40];
(g) Kingman's paintbox theorem for exchangeable random partitions; see section 3.1.

Finite versions. The general forms of de Finetti's theorem and some classical variants can be proved by comparing sampling with and without replacement. This method [24] also yields finite-n variants.

Mathematical population genetics, the coalescent and the Poisson–Dirichlet distribution. Exchangeability is involved in this large circle of ideas, developed in part by Kingman in the 1970s, which continues to prove fruitful in many ways. For the population genetics aspects of Kingman's work see the article by Ewens and Watterson [33] in this volume; also the *Kingman coalescent* fits into the more general *stochastic coalescent* material in section 3.2.

The subsequence principle. The idea emerged in the 1970s that, from any tight sequence of random variables, one can extract a subsequence which is close to exchangeable, close enough that one can prove analogs of classical limit theorems (CLT and LIL, for instance) for the subsequence. General versions of this principle were established in [1, 15], which pretty much killed the topic.

[1] Persi Diaconis observes that the result is hard to deduce from Schoenberg [54] and should really be attributed to Freedman [34].

Sufficient statistics and mixtures of Markov chains. One can often make a Bayesian interpretation of 'sufficient statistic' in terms of some context-dependent invariance property [25]. Somewhat similarly, one can characterize mixtures of Markov chains via the property that transition counts are sufficient statistics [23].

2.3 Partially exchangeable arrays

The topic, emerging around 1980, of *partially exchangeable arrays*, plays a role in what follows and so requires more attention. Take a measurable function $f : [0,1]^2 \to S$ which is symmetric, in the sense $f(x,y) \equiv f(y,x)$. Take $(U_i, i \geq 1)$ i.i.d. Uniform$(0,1)$ and consider the array

$$X_{\{i,j\}} := f(U_i, U_j) \tag{2.1}$$

indexed by the set $\mathbb{N}_{(2)}$ of unordered pairs $\{i,j\}$. The exchangeability property of (U_i) implies what we shall call the *partially exchangeable* property for the array:

$$(X_{\{i,j\}},\ \{i,j\} \in \mathbb{N}_{(2)}) \stackrel{d}{=} (X_{\{\pi(i),\pi(j)\}},\ \{i,j\} \in \mathbb{N}_{(2)}) \quad \text{for each finite permutation } \pi. \tag{2.2}$$

Note this is a weaker property than the 'fully exchangeable' property for the countable collection $(X_{\{i,j\}},\ \{i,j\} \in \mathbb{N}_{(2)})$, because the permutations of $\mathbb{N}_{(2)}$ which are of the particular form $\{i,j\} \to \{\pi(i), \pi(j)\}$ for a finite permutation π of $\mathbb{N}$ are only a subset of all permutations of $\mathbb{N}_{(2)}$.

Aside from construction (2.1), how else can one produce an array with this partially exchangeable property? Well, an array with i.i.d. entries has the property, and so does the trivial case where all entries are the same r.v. After a moment's thought we realize we can combine these ideas as follows.

Take a function $f : [0,1]^4 \to S$ such that $f(u, u_1, u_2, u_{12})$ is symmetric in (u_1, u_2), and then define

$$X_{\{i,j\}} := f(U, U_i, U_j, U_{\{i,j\}}) \tag{2.3}$$

where all the r.v.s in the families $U, (U_i, i \in \mathbb{N}), (U_{\{i,j\}},\ \{i,j\} \in \mathbb{N}_{(2)})$ are i.i.d. Uniform$(0,1)$. Then the resulting array $\mathbf{X} = (X_{\{i,j\}})$ is partially exchangeable.

Finding oneself unable to devise any other constructions, it becomes natural to conjecture that every partially exchangeable array has a representation (in distribution) of form (2.3). This was proved by Hoover

[39] and (in the parallel non-symmetric setting) by Aldous [2], the latter proof having been substantially simplified due to a personal communication from Kingman.

Constructions of partially exchangeable arrays appear in Bayesian statistical modeling; see e.g. the family of copulae introduced in [19] in the context of a semi-parametric model for Value at Risk.

2.4 Fast forward

Such *partially exchangeable representation theorems* were the state of the art in the 1984 survey [3]. They were subsequently extended systematically by Kallenberg, both for arrays and analogs such as exchangeable-increments continuous-parameter processes, and for the *rotatable matrices* to be mentioned in section 2.6, during the late 1980s and early 1990s. The whole topic of representation theorems is the subject of Chapters 7–9 of Kallenberg's 2005 monograph [43]. Not only does this monograph provide a canonical reference to the theorems, but also its introduction provides an excellent summary of the topic.

In the particular setting above we have

Theorem 2.1 (Partially Exchangeable Representation Theorem) *An array* $\mathbf{X}$ *which is partially exchangeable, in the sense* (*2.2*)*, has a representation in the form* (*2.3*)*.*

This is one of the family of results described carefully in Chapter 7 of [43]. There are analogous results for higher-parameter arrays (X_{ijk}), and for arrays in which the 'joint exchangeability' assumption (2.2) is replaced by a 'separate exchangeability' assumption for non-symmetric arrays $(X_{i,j}, 1 \le i, j < \infty)$:

$$(X_{i,j}, 1 \le i, j < \infty) \stackrel{d}{=} (X_{\pi_1(i),\pi_2(j)}, 1 \le i, j < \infty)$$
$$\text{for finite permutations } \pi_1, \pi_2.$$

One aspect of this theory is surprisingly subtle, and that is the issue of *uniqueness* of representing functions f. In representation (2.3), if we take Lebesgue-measure-preserving maps ϕ_0, ϕ_1, ϕ_2 from $[0,1]$ to $[0,1]$, then the arrays $\mathbf{X}$ and $\mathbf{X}^*$ obtained from f and from $f^*(u, u_1, u_2, u_{12}) := f(\phi_0(u), \phi_1(u_1), \phi_1(u_2), \phi_2(u_{12}))$ must have the same distribution. But this is not the only way to make arrays have the same distribution: there are other ways to construct measure-preserving transformations of $[0,1]^4$, and (because measure-preserving transformations are not invertible in general) one needs to insert randomization variables. (I thank

the referee for correcting a blunder in my first draft, and for the comment "this may be a major reason why non-standard analysis is effective here".) For an explicit statement of the uniqueness result in two dimensions see [41] and for higher dimensions see [43].

Relative to the Big Picture of Mathematics, this theory of partial exchangeability was perhaps regarded during 1980–2005 as a rather small niche inside mathematical probability—and ignored outside mathematical probability. So it is ironic that around 2004–8 it was rediscovered in at least three different contexts outside mainstream mathematical probability. Let me say one such context right now and the others later (sections 5.1 and 5.2).

2.5 Isometry classes of metric spaces with probability measures

The definition of *isometry* between two metric spaces (S_1, d_1) and (S_2, d_2) contains an 'if there exists ...' expression. Asking for a *characterization* of metric spaces up to isometry is asking for a scheme that associates some notion of 'label' to each metric space in such a way that two metric spaces are isometric if and only if they have the same label. I am not an expert on this topic, but I believe there is no known such characterization.

But suppose instead we consider 'metric spaces with probability measure', (S_1, d_1, μ_1) and (S_2, d_2, μ_2), and require the isometry to map μ_1 to μ_2. It turns out there is now a conceptually simple characterization. Given (S, d, μ), take i.i.d.(μ) random elements $(\xi_i, 1 \le i < \infty)$ of S, form the array (of form (2.1))

$$X_{\{i,j\}} = d(\xi_i, \xi_j);\ \{i,j\} \in \mathbb{N}_{(2)} \tag{2.4}$$

and let Ψ be the distribution of the infinite random array. It is obvious that, for two isometric 'metric spaces with probability measure', we get the same Ψ, and the converse is a simple albeit technical consequence of the uniqueness part of Theorem 2.1, implying:

$$\begin{gathered}\text{'metric spaces with probability measure' are characterized} \\ \text{up to isometry by the distribution } \Psi. \end{gathered} \tag{2.5}$$

This result was given by Vershik [57], as one rediscovery of part of the general theory of partial exchangeability.

2.6 Rotatable arrays and random matrices with symmetry properties

In Theorem 2.1 we described $\mathbf{X} = (X_{ij})$ as an *array* instead of a *matrix*, partly because of the extension to higher-dimensional parametrizations and partly because we never engage matrix multiplication. Now regarding $\mathbf{X}$ as a matrix, one can impose stronger 'matrix-theoretic' assumptions and ask for characterizations of the random matrices satisfying such assumptions. One basic case, *rotatable matrices*, is where the $n \times n$ restrictions are invariant in distribution under the orthogonal group, and the characterization goes back to [2]. Two other cases (I thank the referee for suggesting (ii) and the subsequent remark) are

(i) non-negative definite jointly exchangeable arrays: [30, 51];
(ii) rotatable *Hermitian* matrices [50], motivated indirectly by problems in quantum mechanics and thereby related to the huge literature on semicircular laws.

Returning to the basic case of rotatable matrices, for the higher-dimensional analogs the basic representations are naturally stated in terms of multiple Wiener–Itô integrals, which form the fundamental examples of rotatable random functionals. Such multiple Wiener–Itô integrals are also a basic tool in [49], a subject with important applications to analysis.

2.7 Revisiting de Finetti's theorem

Returning to a previous comment, the theory of partially exchangeable representation theorems reminds us that one can take a similar view of de Finetti's theorem itself, to add to the list in section 2.1.

Construction viewpoint. Given a measurable function $f : [0,1]^2 \to S$ and i.i.d. Uniform$(0,1)$ random variables $(U; U_i, i \geq 1)$, the process $(Z_i, i \geq 1)$ defined by $Z_i = f(U, U_i)$ is exchangeable, and every exchangeable process arises (in distribution) in this way from some f.

3 Using exchangeability to describe complex structures

Here is my attempt at articulating the first part of the central theme of this paper.

One way of examining a complex mathematical structure is to sample i.i.d. random points and look at some form of induced substructure relating the random points.

The idea being that the i.i.d. sampling induces some kind of 'exchangeability' on the distribution of the substructure, when the substructure is regarded as an object in its own right.

The 'isometry' result (2.5) nicely fits this theme—the substructure is simply the induced metric on the sampled points. The rest of the present paper seeks to illustrate that this, admittedly very vague, way of looking at structures can indeed be useful, conceptually and/or technically. Let us mention here two prototypical examples (which will reappear later) of what a 'substructure' might be. Given k vertices $v(1), \ldots, v(k)$ in a graph, one can immediately see two different ways to define an induced substructure.

(i) The induced subgraph on vertices $1, \ldots, k$: there is an edge (i, j) iff the original graph has an edge $(v(i), v(j))$.
(ii) The distance matrix: $d(i, j)$ is the number of edges in the shortest path from $v(i)$ to $v(j)$.

But before considering graph theoretic examples, let us explain with hindsight how Kingman's work on exchangeable random partitions fits this theme.

3.1 Exchangeable random partitions and Kingman's paintbox theorem

The material here is covered in detail in Pitman [52] Chapters 2–4.

Given a discrete sub-probability distribution, one can write the probabilities in decreasing order as $p_1 \geq p_2 \geq \ldots > 0$ and then write $p_{(\infty)} := 1 - \sum_j p_j \geq 0$ to define a probability distribution $\mathbf{p}$. Imagine objects $1, 2, 3, \ldots$ each independently being colored, assigned color j with probability p_j or assigned with probability $p_{(\infty)}$ a unique color (different from that assigned to any other object). Then consider the resulting 'same color' equivalence classes as a random partition of $\mathbb{N}$. So a realization of this process might be

$$\{1, 5, 6, 9, 13, \ldots\}, \ \{2, 3, 8, 11, 15, \ldots\}, \ \{4\}, \ \{7, 23, \ldots\}, \ \ldots\ldots; \quad (3.1)$$

sets in the partition are either infinite or singletons. This *paintbox*($\mathbf{p}$)

distribution on partitions is exchangeable in the natural sense. *Kingman's paintbox theorem*, an analog of de Finetti's theorem, states that every exchangeable random partition of $\mathbb{N}$ is distributed as a mixture over $\mathbf{p}$ of paintbox($\mathbf{p}$) distributions.

We mentioned in section 2.1 as a general feature that, associated with a representation theorem like this, there will be a convergence theorem. Here are two slightly different ways of looking at the convergence theorem in the present setting. Suppose that for each n we are given an arbitrary random (or non-random) partition $\mathcal{G}(n)$ of $\{1, 2, \ldots, n\}$. For each $k < n$ sample without replacement k times from $\{1, 2, \ldots, n\}$ to get $U_n(1)$, $\ldots$, $U_n(k)$, consider the induced partition on the sampled elements $U_n(1)$, $\ldots$, $U_n(k)$, and relabel these elements as 1, $\ldots$, k to get a random partition $\mathcal{S}(n, k)$ of $\{1, 2, \ldots, k\}$. This random partition $\mathcal{S}(n, k)$ is clearly exchangeable. If there is a limit

$$\mathcal{S}(n, k) \stackrel{d}{\to} \mathcal{S}_k \text{ as } n \to \infty \tag{3.2}$$

(the set of all possible partitions of $\{1, 2, \ldots, k\}$ is finite, so there is nothing technically sophisticated here) then the limit $\mathcal{S}_k$ is exchangeable; and if a limit (3.2) exists for all k then the family $(\mathcal{S}_k, 1 \le k < \infty)$ specifies the distribution of an exchangeable random partition of $\mathbb{N}$, to which Kingman's paintbox theorem can be applied.

The specific phrasing above was chosen to fit a general framework in section 4.1 later, but here is a more natural phrasing. For any random partition of $\{1, \ldots, n\}$ write $\mathbf{F}^{(n)} = (F_1^{(n)}, F_2^{(n)}, \ldots)$ for the *ranked empirical frequencies*, the numbers $n^{-1} \times$(sizes of sets in partition) in decreasing order. For a paintbox($\mathbf{p}$) distribution the SLLN implies $\mathbf{F}^{(n)} \to \mathbf{p}$ a.s., and so Kingman's paintbox theorem implies that for any infinite exchangeable random partition Π, the limit $\mathbf{F}^{(n)} \to \mathbf{F}$ exists a.s. and is the 'directing random measure' (conditional on $\mathbf{F} = \mathbf{p}$ the distribution of Π is paintbox($\mathbf{p}$)). Now suppose for each n we have an *exchangeable* random partition $\Pi^{(n)}$ of $\{1, 2, \ldots, n\}$ and write $\mathbf{F}^{(n)}$ for its ranked empirical frequencies. The *convergence theorem* states that the sequence $\Pi^{(n)}$ converges in distribution (meaning its restriction to $\{1, \ldots, k\}$ converges, for each k) to some limit Π, which is necessarily some infinite exchangeable random partition with some directing random measure $\mathbf{F}$, if and only if $\mathbf{F}^{(n)} = (F_j^{(n)},\ 1 \le j < \infty) \stackrel{d}{\to} \mathbf{F} = (F_j,\ 1 \le j < \infty)$.

A final important idea is *size-biased order*. In the context of exchangeable random partitions this just means writing the components in a realization, as at (3.1), starting with the component containing element 1,

then the component containing the least element not in the first component, and so on. In the infinite case, the frequencies $\mathbf{F}^* = (F_1^*, F_2^*, \ldots)$ of the components in size-biased order are just a random permutation of the frequencies $\mathbf{F}$ given by Kingman's paintbox theorem. In the paintbox($\mathbf{p}$) case, replacing non-random $\mathbf{F} = \mathbf{p}$ by random $\mathbf{F}^*$ is perhaps merely complicating matters, but in the general case of random $\mathbf{F}$ it is often more natural to work with the size-biased order than with the ranked order. For instance, the size-biased order codes information such as

$$\mathbb{E}(F_1^*)^m = \mathbb{E}\sum_{i\geq 1} F_i^{m+1}.$$

I am highlighting these 'structural' results as part of my overall theme, but in many ways the concrete examples are more interesting. The one-parameter Poisson–Dirichlet(θ) family was already recognized 25 years ago as a mathematically canonical family of measures arising in several different contexts: the Ewens sampling formula in neutral population genetics, the 'Chinese restaurant process' construction, a construction via subordinators, the size-biased order of asymptotic frequencies is the GEM distribution; and special cases arise as limits of component sizes in random permutations and in random mappings. Subsequently the two-parameter Poisson–Dirichlet(α, θ) distribution introduced by Pitman–Yor [53] was shown to possess many analogous properties. The paper [37] in this volume gives the flavor of current work in this direction.

Now partitions are rather simple structures, and the paintbox theorem (which can be derived from de Finetti's theorem) isn't so convincing as an illustration of the theme 'using exchangeability to describe complex structures'. The theme becomes more visible when we consider the more complex setting of partitions evolving over time, and this setting arises naturally in the following context.

3.2 Stochastic coalescence and fragmentation

The topic of this section is treated in detail in Bertoin [16], the third monograph in which exchangeability has recently featured prominently. The topic concerns models in which, at each time, unit mass is split into clusters of masses $\{x_j\}$. One studies models of dynamics under which clusters split (*stochastic fragmentation*) or merge (*stochastic coalescence* or *coagulation*[2]) according to some random rules.

[2] The word *coagulation*, introduced in German in [55], sounds strange to the native English speaker to whom it suggests blood clotting; *coalescence* seems a more apposite English word.

Conceptually, states are unordered collections $\{x_j\}$ of positive numbers with $\sum_j x_j = 1$. What is a good mathematical representation of such states? The first representation one might devise is as vectors in decreasing order, say $\mathbf{x}^{\downarrow} = (x_1, x_2, \ldots)$. But this representation has two related unsatisfactory features; fragmenting one cluster forces one to relabel the others; and given the realizations at two times, one can't identify a particular cluster at the later time as a fragment of a particular cluster at the earlier time. These difficulties go away if we think instead in terms of sampling 'atoms' and tracking which cluster they are in. A uniform random atom will be in a mass-x cluster of a configuration $\mathbf{x}^{\downarrow}$ with probability x; sampling atoms $i = 1, 2, 3, \ldots$ and taking the partition of $\{1, 2, 3, \ldots\}$ into 'atoms of the same cluster' components gives an exchangeable random partition Π with paintbox($\mathbf{x}^{\downarrow}$) distribution.

Thus instead of representing a process as $(\mathbf{X}^{\downarrow}(t), 0 \le t < \infty)$ we can represent it as a partition-valued process $(\Pi(t), 0 \le t < \infty)$ which tracks the positions of (i.e. the clusters containing) particular atoms. For fixed t, both $\Pi(t)$ and $\mathbf{X}^{\downarrow}(t)$ give the same information about the cluster masses—and note that clusters in $\Pi(t)$ automatically appear in size-biased order. But as processes in t, $(\Pi(t), 0 \le t < \infty)$ gives more information than $(\mathbf{X}^{\downarrow}(t), 0 \le t < \infty)$, and in particular avoids the unsatisfactory features mentioned above.

Now in one sense this is merely a technical device, but I find it does give some helpful insights.

The basic general stochastic models. In the basic model of *stochastic fragmentation*, different clusters evolve independently, a mass-x cluster splitting at some stochastic rate λ_x into clusters whose relative masses $(x_j/x,\ j \ge 1)$ follow some probability distribution $\mu_x(\cdot)$. (So the model neglects detailed 3-dimensional geometry; the shape of a cluster is assumed not to affect its propensity to split, and different clusters do not interact). Especially tractable is the *self-similar* case where $\mu_x = \mu_1$ and $\lambda_x = x^{\alpha}$ for some *scaling exponent* α. Such processes are closely related to classical topics in theoretical and applied probability—the log-masses form a continuous time branching random walk, and the mass of the cluster containing a sample atom forms a continuous-time Markov process on state space $(0, 1]$.

The basic model for *stochastic coalescence* is to have n particles, initially in single particle clusters of masses $1/n$, and let clusters merge according to a kernel $\kappa(x, x')$ indicating the rate (probability per unit time) at which a typical pair of clusters of masses x and x' may merge.

For fixed n this is just a finite-state continuous-time Markov chain, but it is natural to study $n \to \infty$ limits, and there are two different regimes. On the time-scale where typical clusters contain $O(1)$ particles, i.e. have mass $O(1/n)$, there is an intuitively natural *hydrodyamical limit* (law of large numbers), that is differential equations for the relative proportions $y_i(t)$ of i-particle clusters in the $n \to \infty$ limit. This *Smoluchowski coagulation equation* has a long history in several areas of science such as physical chemistry, as indicated in the survey [6]. Recent theoretical work has made rigorous the connection between the stochastic and deterministic models, and part of this is described in [16] Chapter 5. A different limit regime concerns the time-scale when the largest clusters contain order n particles, i.e. have mass of order 1. In this limit we have real-valued cluster sizes evolving over time $(-\infty, \infty)$ and 'starting with dust' at time $-\infty$, that is with the largest cluster mass $\to 0$ as $t \to -\infty$ (just as, in the basic fragmentation model, the largest cluster mass $\to 0$ as $t \to +\infty$) and $\to 1$ as $t \to +\infty$. (So these models incidentally provide novel examples within the classical topic of *entrance boundaries* for Markov processes).

Finally recall *Kingman's coalescent*, as a model of genealogical lines of descent within neutral population genetics, which (with its many subsequent variations) has become a recognized topic within mathematical population genetics—see e.g. Wakeley [58]. Although the background story is different, it can mathematically be identified with the constant rate ($\kappa(x, x') = 1$) stochastic coalescent in the present context.

Discussion and special cases. There are three settings above (fragmentation; discrete-particle coalescence; continuous-mass coalescence) which one might formalize differently, but the advantage of the 'exchangeable random partition' set-up is that each can be represented as a partition-valued process $(\Pi(t))$. Intuitively, coalescence and fragmentation are time-reversals of each other, and it is noteworthy that

(i) there are several fascinating examples of special models where a precise duality relation exists and is useful (see e.g. section 4.4 (iv));
(ii) but there seems to be no *general* precise duality relationship within the usual stochastic models.

In the general models the processes $(\Pi(t))$ are all Markov, as processes on partitions of $\mathbb{N}$. One can consider their restrictions $(\Pi_k(t))$ to partitions of $\{1, \ldots, k\}$, i.e. consider masses of the components containing k

sampled atoms. In general $(\Pi_k(t))$ will not be Markov, but it is Markov in special cases, which are therefore particularly tractable. One such case ([16] section 3.1) is *homogeneous fragmentation*, where each cluster has the same splitting rate ($\lambda_x \equiv \lambda_1$). Another such case ([16] Chapter 4) is the elegant general theory of *exchangeable coalescents*, which eliminates the 'only binary merging' aspect of Kingman's coalescent, and is interpretable as $n \to \infty$ limit genealogies of more general models in population genetics.

4 Construction of, and convergence to, infinite random combinatorial objects

4.1 A general program

de Finetti's theorem refers specifically to *infinite* sequences. Of course we can always try to view an infinite object as a limit of finite objects, and in the '25 years ago' surveys such convergence ideas were explicit in the context of weak convergence for 'sampling without replacement' processes, in finite versions such as [24], and in some other contexts, such as Kingman's theory of exchangeable random partitions. I previously stated the first part of our central theme as

> One way of examining a complex mathematical structure is to sample i.i.d. random points and look at some form of induced substructure relating the random points

which assumes we are *given* the complex structure. But now the second and more substantial part of the theme is that we can often use exchangeability in the *construction* of complex random structures as the $n \to \infty$ limits of random finite n-element structures $\mathcal{G}(n)$.

> Within the n-element structure $\mathcal{G}(n)$ pick k random elements, look at the induced substructure on these k elements—call this $\mathcal{S}(n,k)$. 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 it with some more concrete construction.

4.2 First examples

To invent a name for the program above, let's call it *exchangeable representations of complex random structures*. Let me first mention three examples.

1. Our discussion (section 3.1) of exchangeable random partitions fits the program but is atypically simple, in that the limit $\mathcal{S}_\infty$ is visibly the same kind of object (an exchangeable random partition) as is the finite object $\mathcal{G}(n)$. But when we moved on to coalescence and fragmentation *processes* in section 3.2, our 'exchangeability' viewpoint prompts consideration of the limit process as the partition-valued process $(\Pi(t))$, which is rather different from the finite-state Markov processes arising in coalescence for finite n.

2. A conceptually similar example arises in the technically more sophisticated setting of measure-valued diffusions (μ_t). In such processes the states are probability measures μ on some type-space, representing a 'continuous' infinite population. But one can alternately represent μ via an infinite i.i.d. sequence of samples from μ, and thereby represent the state more directly as a discrete countable infinite population $(Z(i), i \geq 1)$ and the process as a particle process $(Z_t(i), i \geq 1)$. This viewpoint was emphasized in the *look-down construction* of Kurtz–Donnelly [28, 29].

3. As a complement to the characterization (2.5) of metric spaces with a probability measure (p.m.), we can define a notion of *convergence* of such objects, say of finite metric spaces with p.m.s (S_n, d_n, μ_n) to a continuous limit $(S_\infty, d_\infty, \mu_\infty)$. The definition is simply that we have weak convergence

$$\mathbf{X}^n \stackrel{d}{\to} \mathbf{X}^\infty$$

of the induced random arrays defined as at (2.4):

$$X^n_{\{i,j\}} = d_n(\xi^n_i, \xi^n_j);\ \{i,j\} \in \mathbb{N}_{(2)}$$

for i.i.d. $(\xi^n_i, i \geq 1)$ with distribution μ_n. This definition provides an intriguing complement to the more familiar notion of Gromov–Hausdorff distance between compact metric spaces.

Let us move on to the fundamental setting where the program gives a substantial payoff, the formalization of *continuum random trees* as rescaled limits of finite random trees.

4.3 Continuum random trees

The material here is from the old survey [4].

Probabilists are familiar with the notion that rescaled random walk converges in distribution to Brownian motion. Now in the most basic case—simple symmetric RW of length n—we are studying the uniform distribution on a combinatorial set, the set of all 2^n simple walks of length n. So what happens if we study instead the uniform distribution on some other combinatorial set? Let us consider the set of n-vertex trees. More precisely, consider either the set of rooted labeled trees (Cayley's formula says there are n^{n-1} such trees), or the set of rooted ordered trees (counted by the Catalan number $\frac{1}{n}\binom{2n-2}{n-1}$), and write $\mathcal{T}_n$ for the uniform random tree.

Trees fit nicely into the 'substructure' framework. Vertices $v(1), \ldots, v(k)$ of a tree define a spanning (sub)tree. Take each maximal path $(w_0, w_1, \ldots, w_\ell)$ in the spanning tree whose intermediate vertices have degree 2, and contract to a single edge of length ℓ. Applying this to k independent uniform random vertices from a n-vertex model $\mathcal{T}_n$, then rescaling edge-lengths by the factor $n^{-1/2}$, gives a tree we'll call $\mathcal{S}(n, k)$. We visualize such trees as in Figure 4.1, vertex $v(i)$ having been relabeled as i.

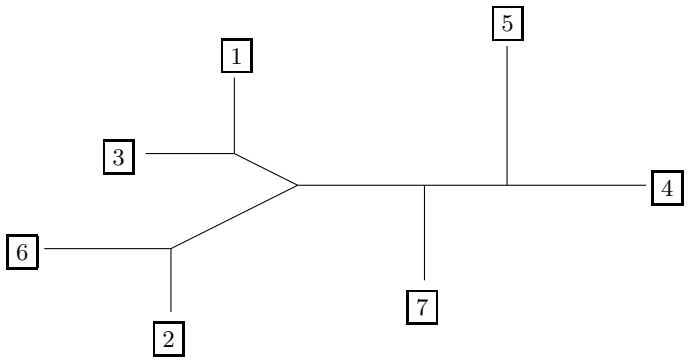

Figure 4.1 A leaf-labeled tree with edge-lengths.

In the two models for $\mathcal{T}_n$ mentioned above, one can do explicit calculations of the distribution of $\mathcal{S}(n,k)$, and use these to show that in distribution there is an $n \to \infty$ limit $\mathcal{S}_k$ which (up to a model-dependent scaling constant we'll ignore in this informal exposition) is the following distribution.

(i) The state space is the space of trees with k leaves labeled 1, 2, ..., k and with unlabeled degree-3 internal vertices, and where the $2k-3$ edge-lengths are positive real numbers.

(ii) For each possible topological shape, the chance that the tree has that particular shape and that the vector of edge-lengths $(L_1, \ldots, L_{2k-3})$ is in $([l_i, l_i + dl_i], 1 \le i \le 2k-3)$ equals $s \exp(-s^2/2) dl_1 \ldots dl_{2k-3}$, where $s = \sum_i l_i$.

One can check from the explicit formula what must be true from the general program, that for fixed k the distribution is exchangeable (in labels $1, \ldots, k$), and the distributions are consistent as k increases (that is, the subtree of $\mathcal{S}_{k+1}$ spanned by leaves $1, \ldots, k$ is distributed as $\mathcal{S}_k$).

So some object $\mathcal{S}_\infty$ exists, abstractly—but what is it, more concretely? A *real tree* is a metric space with the 'tree' property that between any two points v and w there is a unique path. This implicitly specifies a length measure λ such that the metric distance $d(v,w)$ equals the length measure of the set of points on the path from v to w. When a real tree is equipped with a mass measure μ of total mass 1, representing a method for picking a vertex at random, I call it a *continuum tree*. We will consider random continuum trees—which I call continuum random trees or CRTs because it sounds better!—and the Portmanteau Theorem below envisages realizations of $\mathcal{S}_\infty$ as being equipped with a mass measure.

Returning to the n-vertex random tree models $\mathcal{T}_n$, by assigning 'mass' $1/n$ to each vertex we obtain the analogous 'mass measure' on the vertices, used for randomly sampling vertices. The next result combines existence, construction and convergence theorems. The careful reader will notice that some details in the statements have been omitted.

The Portmanteau Theorem [4, 5]

1. **Law of spanning subtrees.** There exists a particular *Brownian CRT* which agrees with $\mathcal{S}_\infty$ in the following sense. Take a realization of the Brownian CRT, then pick k i.i.d. vertices from the mass measure, and consider the spanning subtree on these k vertices. The unconditional law of this subtree is the law in (ii) above.

2. **Construction from Brownian excursion.** Consider an excursion-type function $f : [0,1] \to [0,\infty)$ with $f(0) = f(1) = 0$ and $f(x) > 0$ elsewhere. Use f to define a continuum tree as follows. Define a pseudo-metric on $[0,1]$ by: $d(x,y) = f(x) + f(y) - 2\min(f(u) : x \le u \le y)$, $x \le y$. The continuum tree is the associated metric space, and the mass measure is the image of Lebesgue measure on $[0,1]$. Using this construction with standard Brownian excursion (scaled by a factor 2) gives the Brownian CRT.
3. **Line-breaking construction.** Cut the line $[0,\infty)$ into finite segments at the points of a non-homogeneous Poisson process of intensity $\lambda(x) = x$. Build a tree by inductively attaching a segment $[x_i, x_{i+1}]$ to a uniform random point of the tree built from the earlier segments. The tree built from the first $k-1$ segments has the law (ii) above. The metric space closure of the tree built from the whole half-line is the Brownian CRT, where the mass measure is the a.s. weak limit of the empirical law of the first k cut-points.
4. **Weak limit of conditioned critical Galton–Watson branching processes and of uniform random trees.** Take a critical Galton–Watson branching process where the offspring law has finite non-zero variance, and condition on total population until extinction being n. This gives a random tree. Rescale edge-lengths to have length $n^{-1/2}$. Put mass $1/n$ on each vertex. In a certain sense that can be formalized, the $n \to \infty$ weak limit of these random trees is the Brownian CRT (up to a scaling factor). This result includes as special cases the two combinatorial models $\mathcal{T}_n$ described above.

4.4 Complements to the continuum random tree

More recent surveys by Le Gall [45] and by Evans [32] show different directions of development of the preceding material over the last 15 years. For instance

(i) the *Brownian snake* [45], which combines the genealogical structure of random real trees with independent spatial motions.
(ii) Diffusions on real trees: [32] Chapter 7.
(iii) *Continuum-tree valued diffusions.* There are several natural ways to define Markov chains on the space of n-vertex trees such that the stationary distribution is uniform. Since the $n \to \infty$ rescaled limit of the stationary distribution is the Brownian CRT, it is natural to conjecture that the entire rescaled process can be made to converge

to some continuum-tree valued diffusion whose stationary distribution is the Brownian CRT. But this forces us to engage a question that was deliberately avoided in the previous section: what exactly is the space of all continuum trees, and when should we consider two such trees to be the same? This issue is discussed carefully in [32], based on the notion of the Gromov–Hausdorff space of all compact spaces. Two specific continuum-tree valued diffusions are then studied in Chapters 5 and 9 of [32].

(iv) Perhaps closer to our 'exchangeability' focus, a surprising aspect of CRTs is their application to stochastic coalescence. For $0 < \lambda < \infty$ split the Brownian CRT into components at the points of a Poisson process of rate λ along the skeleton of the tree. This gives a vector $Y(\lambda) = (Y_1(\lambda), Y_2(\lambda), \ldots)$ of masses of the components, which as λ increases specifies a fragmentation process. Reversing the direction of time by setting $\lambda = e^{-t}$ provides a construction of the (standard) additive coalescent [8], that is the stochastic coalescent (section 3.2) with kernel $\kappa(x, y) = x + y$ 'started from dust'. This result is non-intuitive, and notable as one of a handful of precise instances of the conceptual duality between stochastic coalescence and fragmentation. Also surprisingly, there are different ways that the additive coalescent can be 'started from dust', and these can also be constructed via fragmentation of certain inhomogeneous CRTs [10]. This new family of CRTs satisfies analogs of the Portmanteau Theorem, and in particular there is an explicit analog of the formula (ii) in section 4.3 for the distribution of the subtree $\mathcal{S}_k$ spanned by k random vertices [9]. This older work is complemented by much current work, the flavor of which can be seen in [38].

(v) A function $F : \{1, \ldots, n\} \to \{1, \ldots, n\}$ defines a directed graph with edges $(i, F(i))$, and the topic *random mappings* studies the graph derived from a random function F. One can repeat the section 4.1 general program in this context. Any k sampled vertices define an induced substructure, the subgraph of edges $i \to F(i) \to F(F(i)) \to \ldots$ reachable from some one of the sampled vertices. Analogously to Figure 4.1, contract paths between sampled vertices/junctions to single edges, to obtain (in the $n \to \infty$ limit) a graph $\mathcal{S}_k$ with edge-lengths, illustrated in Figure 4.2. The theory of $n \to \infty$ limits of random mappings turns out to be closely related to that of random trees; the approach based on studying the consistent family $(\mathcal{S}_k, k \geq 1)$ was developed in Aldous–Pitman [11].

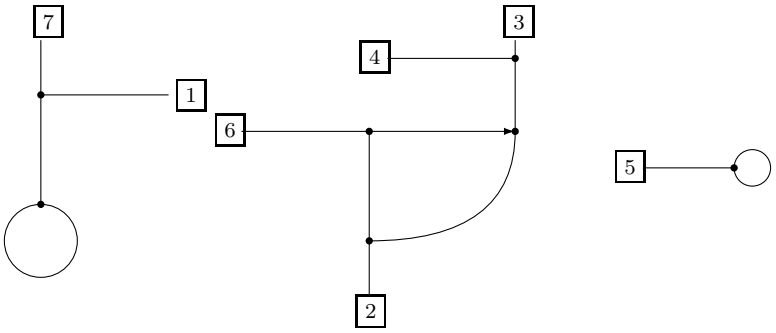

Figure 4.2 The substructure $\mathcal{S}_k$ of a random mapping.

4.5 Second-order distance structure in random networks

In the context (section 4.3) of continuum random trees the substructure was distances between sampled points. At first sight one might hope that in many models of size-n random networks one could repeat that analysis and find an interesting limit structure. But the particular feature of the models in section 4.3 is 'first order randomness' of the distance D_n between two random vertices; D_n/ED_n has a non-constant limit distribution, leading to the randomness in the limit structure. Other models tend to fall into one of two categories. For geometric networks (vertices having positions in $\mathbb{R}^2$; route-lengths as Euclidean length) the route-length tends to grow as constant c times Euclidean distance, so any limit structure reflects only the randomness of sampled vertex positions and the constant c, not any more interesting properties of the network. In non-geometric (e.g. Erdős–Rényi random graph) models, D_n tends to be first-order constant. So counter-intuitively, we don't know any other first-order random limit structures outside the 'somewhat tree-like' context.

Understanding *second*-order behavior in spatial models is very challenging—for instance, the second order behavior of first passage percolation times remains a longstanding open problem. But one can get second order results in simple 'random graph' type models, and here is the basic example (mentioned in Aldous and Bhamidi [7] as provable by the methods of that paper). The probability model used for a random n-vertex

network $\mathcal{G}_n$ starts with the complete graph and assigns independent Exponential(rate $1/n$) random lengths $L_{ij} = L_{ji} = L_e$ to the $\binom{n}{2}$ edges $e = (i,j)$. In this model $ED_n = \log n + O(1)$ and $\mathrm{var}(D_n) = O(1)$, and there is second-order behavior—a non-constant limit distribution for $D_n - \log n$.

Now fix $k \geq 3$ and write $D_n(i,j) \stackrel{d}{=} D_n$ for the distance between vertices i and j. We expect a joint limit

$$(D_n(1,2) - \log n, \ldots, D_n(1,k) - \log n) \stackrel{d}{\to} (D(1,2), \ldots, D(1,k)) \quad (4.1)$$

and it turns out the limit distribution is

$$(D(1,2), \ldots, D(1,k)) \stackrel{d}{=} (\xi_1 + \eta_{12}, \ldots, \xi_1 + \eta_{1k})$$

where ξ_1 has the double exponential distribution

$$\mathbb{P}(\xi \leq x) = \exp(-e^{-x}), \ -\infty < x < \infty,$$

the η_{1j} have logistic distribution

$$\mathbb{P}(\eta \leq x) = \tfrac{e^x}{1+e^x}, \ -\infty < x < \infty$$

and (here and below) the r.v.s in the limits are independent. Now we can go one step further: we expect a joint limit for the array

$$(D_n(i,j) - \log n, 1 \leq i < j \leq k) \stackrel{d}{\to} (D(i,j), 1 \leq i < j \leq k)$$

and it turns out that the joint distribution of the limit is

$$(D(i,j), 1 \leq i < j \leq k) \stackrel{d}{=} (\xi_i + \xi_j - \xi_{ij}, 1 \leq i < j \leq k)$$

where the limit r.v.s all have the double exponential distribution. Of course the limit here must fit into the format of the partially exchangeable representation theorem (Theorem 2.1), and it is pleasant to see an explicit function f.

5 Limits of finite deterministic structures

Though we typically envisage limiting random structures arising as limits of finite *random* structures, it also makes sense to consider limits of finite *deterministic* structures. Let me start with a trivial example. Suppose that for each n we have a sequence $b_{n,1}, \ldots, b_{n,n}$ of n bits (binary digits), and write p_n for the proportion of 1s. For each k and n, sample

k random bits from $b_{n,1}, \ldots, b_{n,n}$ and call the samples $X_{n,1}, \ldots, X_{n,k}$. Then, rather obviously, the property

$$(X_{n,1}, \ldots, X_{n,k}) \stackrel{d}{\to} (k \text{ independent Bernoulli}(p)) \quad \text{as } n \to \infty; \text{ for each } k$$

is equivalent to the property $p_n \to p$.

Now in one sense this illustrates a big limitation to the whole program —sampling a substructure might lose most of the interesting information in the original structure! But a parallel procedure in the deterministic graph setting (next section) does get more interesting results, and more sophisticated uses are mentioned in section 5.2.

5.1 Limits of dense graphs

Suppose that for each n we have a graph G_n on n vertices. Write p_n for the proportion of edges, relative to the total number $\binom{n}{2}$ of possible edges. We envisage the case $p_n \to p \in (0,1)$.

For each n let $(U_{n,i}, i \geq 1)$ be i.i.d. uniform on $1, \ldots, n$. Consider the infinite $\{0,1\}$-valued matrix $\mathbf{X}^n$:

$$X^n_{i,j} = 1((U_{n,i}, U_{n,j}) \text{ is an edge of } G_n).$$

When $n \gg k^2$ the k sampled vertices $(U_{n,1}, \ldots, U_{n,k})$ of G_n will be distinct and the $k \times k$ restriction of $\mathbf{X}^n$ is the incidence matrix of the induced subgraph $\mathcal{S}(n,k)$ on these k vertices. Suppose there is a limit random matrix $\mathbf{X}$:

$$\mathbf{X}^n \stackrel{d}{\to} \mathbf{X} \text{ as } n \to \infty \tag{5.1}$$

in the usual product topology, that is

$$(X^n_{i,j},\ 1 \leq i,j \leq k) \stackrel{d}{\to} (X_{i,j},\ 1 \leq i,j \leq k) \text{ for each } k.$$

(Note that by compactness there is always a *subsequence* in which such convergence holds.) Now each $\mathbf{X}^n$ has the partially exchangeable property (2.2), and the limit $\mathbf{X}$ inherits this property, so we can apply the representation theorem (Theorem 2.1) to describe the possible limits. In the $\{0,1\}$-valued case we can simplify the representation. First consider a function of form (2.3) but not depending on the first coordinate—that is, a function $f(u_i, u_j, u_{\{i,j\}})$. Write

$$q(u_i, u_j) = \mathbb{P}(f(u_i, u_j, u_{\{i,j\}}) = 1).$$

The distribution of a $\{0,1\}$-valued partially exchangeable array of the

special form $f(U_i, U_j, U_{\{i,j\}})$ is determined by the symmetric function $q(\cdot,\cdot)$, and so for the general form (2.3) the distribution is specified by a probability distribution over such symmetric functions.

This all fits our section 4.1 general program. From an arbitrary sequence of finite deterministic graphs we can (via passing to a subsequence if necessary) extract a 'limit infinite random graph' $\mathcal{S}_\infty$ on vertices 1, 2, ..., defined by its incidence matrix $\mathbf{X}$ in the limit (5.1), and we can characterize the possible limits. But what is a more concrete interpretation of the relation between $\mathcal{S}_\infty$ and the finite graphs (G_n)? To a probabilist the verbal expression of (5.1)

> the restriction $\mathcal{S}_k$ of $\mathcal{S}_\infty$ to vertices 1, ..., k is distributed as the $n \to \infty$ limit of the induced subgraph of G_n on k random vertices

is clear enough, but here is a translation into more graph-theoretic language, following [26]. For finite graphs F, G write $t(F,G)$ for the proportion of all mappings from vertices of F to vertices of G that are graph homomorphisms, i.e. map adjacent vertices to adjacent vertices. Suppose F has k vertices, and we label them arbitrarily as 1, ..., k. Take the subgraph $G[k]$ of G on k randomly sampled vertices, labeled 1, ..., k, and note that whether we sample with or without replacement makes no difference to $n \to \infty$ limits. Then $t(F,G)$ is the probability that F is a subgraph of $G[k]$. Now write $t_=(F,G)$ for the probability that $F = G[k]$. For fixed k, a standard inclusion-exclusion argument shows that, for a sequence (G_n), the existence of either family of limits

$$\lim_n t(F, G_n) \text{ exists, for each graph } F \text{ on vertices } \{1,\ldots,k\}, \quad (5.2)$$

$$\lim_n t_=(F, G_n) \text{ exists, for each graph } F \text{ on vertices } \{1,\ldots,k\}, \quad (5.3)$$

implies existence of the other family of limits.

In our program, the notion of $\mathcal{S}_\infty$ being the limit of G_n was defined by (5.1), which is equivalent to requiring existence of limits (5.3) for each k, in which case the limits are just $\mathbb{E}t_=(F, \mathcal{S}_k)$. And as indicated above, the partially exchangeable representation theorem (Theorem 2.1) characterizes the possible limit structures $\mathcal{S}_\infty$. A recent line of work in graph theory, initiated by Lovász and Szegedy [48], started by defining convergence in the equivalent way via (5.2) and obtained the same characterization. This is the second recent rediscovery of special cases of partially exchangeable representation theory. Diaconis and Janson [26] give a very clear and detailed account of the relation between the two settings, and Diaconis–Holmes–Janson [27] work through to an expli-

cit description of the possible limits for a particular subclass of graphs called *threshold graphs*. Of course the line of work started in [48] has been developed further to produce new and interesting results in graph theory—see e.g. [21].

5.2 Further uses in finitary combinatorics

The remarkable recent survey by Austin [12] gives a more sophisticated treatment of the theory of representations of jointly exchangeable arrays, with the goal ([12] section 4) of clarifying connections between that theory and topics involving limits in finitary combinatorics, such as those in our previous section. I don't understand this material well enough to do more than copy a few phrases, as follows. Section 4.1 of [12] gives a general discussion of 'extraction of limit objects', somewhat parallel to our section 4.1, but with more detailed discussion of different possible precise mathematical structures. The paper continues, describing connections with the 'hypergraph regularity lemmas' featuring in combinatorial proofs of Szemerédi's Theorem, and with the structure theory within ergodic theory that Furstenberg developed for his proof of Szemerédi's Theorem. A subsequent technical paper Austin–Tao [13] applies such methods to the topic of hereditary properties of graphs or hypergraphs being testable with one-sided error; informally, this means that if a graph or hypergraph satisfies that property 'locally' with sufficiently high probability, then it can be modified into a graph or hypergraph which satisfies that property 'globally'.

6 Miscellaneous comments

1. To get an idea of the breadth of the topic, *Mathematical Reviews* created an 'exchangeability' classification 60G09 in 1984, which has attracted around 300 items; *Google Scholar* finds around 350 citations of the survey [3]; and the overlap is only around 50%. The topics in this paper, centered around structure theory—theory and applications of extensions of de Finetti's theorem—are in fact only a rather small part of this whole. In particular the 'exchangeable pairs' idea central to Stein's method [14] is really a completely distinct field.

2. Our central theme involved exchangeability, but one can perhaps view it as part of a broader theme:

a mathematical object equipped with a probability measure is sometimes a richer and more natural structure than the object by itself.

For instance, elementary discussion of fractals like the Sierpiński gasket view the object as a set in $\mathbb{R}^2$, but it comes equipped with its natural 'uniform probability distribution' which enables richer questions—the measure of small balls around a typical point, for example. Weierstrass's construction of a continuous nowhere differentiable function seems at first sight artificial—where would such things arise naturally?—but then the fact that the Brownian motion process puts a probability measure on such functions indicates one place where they do arise naturally. Analogously the notion of *real tree* (section 4.3) may seem at first sight artificial—how might such objects arise naturally?—but then realizing they arise as limits of random finite trees indicates one place where they do arise naturally. Of course the underlying structure 'a space with a metric and a measure' arises in many contexts, for example (under the name *metric measure space*) in the context of differential geometry questions [47].

Acknowledgments I thank Persi Diaconis, Jim Pitman and an anonymous referee for very helpful comments.

References

[1] Aldous, D. J. 1977. Limit theorems for subsequences of arbitrarily-dependent sequences of random variables. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **40**(1), 59–82.

[2] Aldous, D. J. 1981. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, **11**, 581–598.

[3] Aldous, D. J. 1985. Exchangeability and related topics. Pages 1–198 of: Hennequin, P. L. (ed), *École d'Été de Saint-Flour XIII—1983*. Lecture Notes in Math., vol. 1117. Berlin: Springer-Verlag.

[4] Aldous, D. J. 1991. The continuum random tree II: an overview. Pages 23–70 of: Barlow, M. T., and Bingham, N. H. (eds), *Stochastic Analysis*. Cambridge: Cambridge Univ. Press.

[5] Aldous, D. J. 1993. The continuum random tree III. *Ann. Probab.*, **21**, 248–289.

[6] Aldous, D. J. 1999. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli*, **5**, 3–48.

[7] Aldous, D. J., and Bhamidi, S. 2007. *Edge Flows in the Complete Random-Lengths Network*. `http://arxiv.org/abs/0708.0555`.

[8] Aldous, D. J., and Pitman, J. 1998. The standard additive coalescent. *Ann. Probab.*, **26**, 1703–1726.

[9] Aldous, D. J., and Pitman, J. 1999. A family of random trees with random edge lengths. *Random Structures & Algorithms*, **15**, 176–195.

[10] Aldous, D. J., and Pitman, J. 2000. Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Theory Related Fields*, **118**, 455–482.

[11] Aldous, D. J., and Pitman, J. 2002. Invariance principles for non-uniform random mappings and trees. Pages 113–147 of: Malyshev, V. A., and Vershik, A. M. (eds), *Asymptotic Combinatorics with Applications to Mathematical Physics*. Dordrecht: Kluwer. Available at `http://www.stat.berkeley.edu/tech-reports/594.ps.Z`.

[12] Austin, T. 2008. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.*, **5**, 80–145.

[13] Austin, T., and Tao, T. 2008. *On the Testability and Repair of Hereditary Hypergraph Properties*. `http://arxiv.org/abs/0801.2179`.

[14] Barbour, A. D., and Chen, L. H. Y. (eds). 2005. *Stein's Method and Applications*. Lecture Notes Series, vol. 5. Singapore: Singapore Univ. Press, for Institute for Mathematical Sciences, National University of Singapore.

[15] Berkes, I., and Péter, E. 1986. Exchangeable random variables and the subsequence principle. *Probab. Theory Related Fields*, **73**(3), 395–413.

[16] Bertoin, J. 2006. *Random Fragmentation and Coagulation Processes*. Cambridge Stud. Adv. Math., vol. 102. Cambridge: Cambridge Univ. Press.

[17] Billingsley, P. 1968. *Convergence of Probability Measures*. New York: John Wiley & Sons.

[18] Billingsley, P. 1995. *Probability and Measure*. Third edn. Wiley Ser. Probab. Math. Stat. New York: John Wiley & Sons.

[19] Bingham, N. H., Kiesel, R., and Schmidt, R. 2003. A semi-parametric approach to risk management. *Quant. Finance*, **3**(6), 426–441.

[20] Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent Dirichlet allocation. *J. Machine Learning Res.*, **3**, 993–1022.

[21] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., and Vesztergombi, K. 2008. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, **219**(6), 1801–1851.

[22] Chow, Y. S., and Teicher, H. 1997. *Probability Theory: Independence, Interchangeability, Martingales*. Third edn. Springer Texts Statist. New York: Springer-Verlag.

[23] Diaconis, P., and Freedman, D. A. 1980a. de Finetti's theorem for Markov chains. *Ann. Probab.*, **8**(1), 115–130.

[24] Diaconis, P., and Freedman, D. A. 1980b. Finite exchangeable sequences. *Ann. Probab.*, **8**(4), 745–764.

[25] Diaconis, P., and Freedman, D. A. 1984. Partial exchangeability and sufficiency. Pages 205–236 of: *Statistics: Applications and New Directions (Calcutta, 1981)*. Calcutta: Indian Statist. Inst.

[26] Diaconis, P., and Janson, S. 2008. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, **28**(1), 33–61.

[27] Diaconis, P., Holmes, S., and Janson, S. 2009. *Threshold Graph Limits and Random Threshold Graphs*. `http://arxiv.org/abs/0908.2448`.

[28] Donnelly, P., and Kurtz, T. G. 1996. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, **24**(2), 698–742.

[29] Donnelly, P., and Kurtz, T. G. 1999. Particle representations for measure-valued population models. *Ann. Probab.*, **27**(1), 166–205.

[30] Dovbysh, L. N., and Sudakov, V. N. 1982. Gram-de Finetti matrices. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, **119**, 77–86, 238, 244–245. Problems of the Theory of Probability Distribution, VII.

[31] Durrett, R. 2005. *Probability: Theory and Examples*. Third edn. Statistics/Probability Series. Pacific Grove, CA: Brooks/Cole.

[32] Evans, S. N. 2008. *Probability and Real Trees*. Lecture Notes in Math., vol. 1920. Berlin: Springer-Verlag.

[33] Ewens, W. J., and Watterson, G. A. 2009. Kingman and mathematical population genetics. In: Bingham, N. H., and Goldie, C. M. (eds), *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press.

[34] Freedman, D. A. 1962. Invariants under mixing which generalize de Finetti's theorem. *Ann. Math. Statist*, **33**, 916–923.

[35] Fristedt, B., and Gray, L. 1997. *A Modern Approach to Probability Theory*. Probab. Appl. Ser. Boston, MA: Birkhäuser.

[36] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2004. *Bayesian Data Analysis*. Second edn. Texts Statist. Sci. Ser. Boca Raton, FL: Chapman & Hall/CRC.

[37] Gnedin, A. V., Haulk, C., and Pitman, J. 2009. Characterizations of exchangeable partitions and random discrete distributions by deletion properties. In: Bingham, N. H., and Goldie, C. M. (eds), *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*. London Math. Soc. Lecture Note Ser. Cambridge: Cambridge Univ. Press.

[38] Haas, B., Pitman, J., and Winkel, M. 2009. Spinal partitions and invariance under re-rooting of continuum random trees. *Ann. Probab.*, **37**(4), 1381–1411.

[39] Hoover, D. N. 1979. *Relations on Probability Spaces and Arrays of Random Variables*. Preprint, Institute of Advanced Studies, Princeton.

[40] Kallenberg, O. 1973. Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **27**, 23–36.

[41] Kallenberg, O. 1989. On the representation theorem for exchangeable arrays. *J. Multivariate Anal.*, **30**(1), 137–154.

[42] Kallenberg, O. 1997. *Foundations of Modern Probability*. Probab. Appl. (N.Y.). New York: Springer-Verlag.
[43] Kallenberg, O. 2005. *Probabilistic Symmetries and Invariance Principles*. Probab. Appl. (N.Y.). New York: Springer-Verlag.
[44] Kingman, J. F. C. 1978. Uses of exchangeability. *Ann. Probab.*, **6**(2), 183–197.
[45] Le Gall, J.-F. 2005. Random trees and applications. *Probab. Surv.*, **2**, 245–311 (electronic).
[46] Loève, M. 1978. *Probability Theory II*. Fourth edn. Grad. Texts in Math., vol. 46. New York: Springer-Verlag.
[47] Lott, J. 2007. Optimal transport and Ricci curvature for metric-measure spaces. Pages 229–257 of: *Surveys in Differential Geometry*. Surv. Differ. Geom., vol. 11. Somerville, MA: Int. Press.
[48] Lovász, L., and Szegedy, B. 2006. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, **96**(6), 933–957.
[49] Nualart, D. 2006. *The Malliavin Calculus and Related Topics*. Second edn. Probab. Appl. Ser. Berlin: Springer-Verlag.
[50] Olshanski, G., and Vershik, A. 1996. Ergodic unitarily invariant measures on the space of infinite Hermitian matrices. Pages 137–175 of: *Contemporary Mathematical Physics*. Amer. Math. Soc. Transl. Ser. 2, vol. 175. Providence, RI: Amer. Math. Soc.
[51] Panchenko, D. 2009. *On the Dovbysh-Sudakov Representation Result*. `http://front.math.ucdavis.edu/0905.1524`.
[52] Pitman, J. 2006. *Combinatorial Stochastic Processes*. Lecture Notes in Math., vol. 1875. Berlin: Springer-Verlag.
[53] Pitman, J., and Yor, M. 1997. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**(2), 855–900.
[54] Schoenberg, I. J. 1938. Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.*, **44**(3), 522–536.
[55] Smoluchowski, M. von. 1916. Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen. *Physik. Z.*, **17**, 557–585.
[56] Takács, L. 1967. *Combinatorial Methods in the Theory of Stochastic Processes*. New York: John Wiley & Sons.
[57] Vershik, A. M. 2004. Random metric spaces and universality. *Uspekhi Mat. Nauk*, **59**(2(356)), 65–104.
[58] Wakeley, J. 2008. *Coalescent Theory: An Introduction*. Greenwood Village, CO: Roberts & Co.