Invariance Principles for Non-uniform Random Mappings and Trees *

David Aldous and Jim Pitman

Technical Report No. 594

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

December 2001

Abstract

In the context of uniform random mappings of an n-element set to itself, Aldous and Pitman (1994) established a functional invariance principle, showing that many $n \to \infty$ limit distributions can be described as distributions of suitable functions of reflecting Brownian bridge. To study non-uniform cases, in this paper we formulate a *sampling invariance principle* in terms of iterates of a fixed number of random elements. We show that the sampling invariance principle implies many, but not all, of the distributional limits implied by the functional invariance principle. We give direct verifications of the sampling invariance principle in two successive generalizations of the uniform case, to p-mappings (where elements are mapped to i.i.d. non-uniform elements) and P-mappings (where elements are mapped according to a Markov matrix). We compare with parallel results in the simpler setting of random trees.

*Research supported in part by N.S.F. Grants DMS-9970901 and DMS-0071448

1 Introduction

A function M from the n-element set $[n] := \{1, 2, \dots, n\}$ to itself, in this context called a *mapping*, induces a digraph on vertex-set $[n]$ whose edges are $(i, M(i)), i \in [n]$. From a random function M we get a random digraph, and the subject of *random mappings* concerns exact and asymptotic properties of such random graphs, most commonly under the *uniform model* where M is uniform on all n^n mappings [10, 30, 34, 39], but also under various non-uniform models. Saying M is uniform is equivalent to saying that $M(1), M(2), \dots, M(n)$ are independent uniform on $[n]$, so a natural non-uniform model can be defined by requiring $M(1), M(2), \dots, M(n)$ to be independent with some general probability distribution p on $[n]$. Such *p-mappings* are the subject of both classical and current research, reviewed briefly in Section 1.1. A more general model is to take $M(1), \dots, M(n)$ independent on $[n]$ with probabilities $P(M(i) = j) = p_{ij}$ for some Markov matrix $P = (p_{ij})$; call this a *P-mapping.* One might guess that some $n \to \infty$ asymptotic results for uniform random mappings would extend to p-mappings or P-mappings, under appropriate conditions on the sequence $p^{(n)}$ or $P^{(n)}$, or to other models of non-uniform random mappings. The purpose of this paper is to set out a technical framework for studying such questions and specifying limit distributions.

Classical work on the uniform model focussed on specific statistics of mappings, such as component sizes and cycle lengths, which in the uniform case scale as order n and order $n^{1/2}$, and on joint distributions of such statistics [41]. As an extension, in the uniform model Aldous and Pitman [7] gave a Brownian bridge limit theorem which encompasses simultaneously limit distributions for many different statistics which scale as order n and order $n^{1/2}$. That *functional invariance principle*, based on coding mappings as walks, is reviewed in Section 4. A drawback is that the statement of the functional invariance principle is complicated and seemingly rather arbitrary.

In Section 3 we introduce a more direct method for studying random mappings, which we call the *sampling invariance principle.* This method is based on studying $n \to \infty$ asymptotics of the iterates of a fixed number of elements. Our central result is that a variety of asymptotic results extend from the uniform case to any model of non-uniform random mappings which satisfies the sampling invariance principle. Theorem 15 formulates this generally, and then we set down asymptotic distributional results more explicitly. Parallel results for random trees are simpler; we start by reviewing these in Section 2. The mathematical content of Theorem 15 (and the parallel Proposition 7 for trees) is that the sampling invariance principle can be reinterpreted as weak convergence of random functions, but using a weaker topology than is used in the functional invariance principle.

The essence of the sampling invariance principle is that, for fixed k, the union of the orbits of k elements converges (as $n \to \infty$, after rescaling) to a random "mapping with edge-lengths" $G[k]$ whose distribution has a simple explicit form. Such random graphs with edge-lengths, which also arise in the context of trees (Section 2.1), are perhaps of independent interest.

Our results are "abstract" in that instead of starting from a specific model, we are starting from the assumption that the model satisfies the sampling invariance principle, and exploring the non-obvious implications of that principle. For illustration we give two non-uniform random mapping models, which can be regarded as successive generalizations of the uniform model, and which can be shown directly to satisfy the sampling invariance principle. Theorem 25, proved in Section 6.1 as a simple consequence of Poisson approximation for the non-uniform birthday problem, establishes the sampling invariance principle for p-mappings under a natural condition. Then in Section 6.2 we indicate some conditions on P-mappings which imply the sampling invariance principle. It seems likely that other models of non-uniform random mappings could be shown to satisfy the sampling invariance principle, and this is a natural topic for future research.

1.1 p-mappings and p-trees

Though our main results are not tied to a particular model of random mappings, our approach was motivated by consideration of the p-mappings model. Older papers on that model focused on exact combinatorial properties related to Burtin's lemma [17, 32, 47] (see [45] for recent systematic discussion of combinatorial properties) and on asymptotics in the special case where all but one of the p-values are equal [12, 40, 48]. Asymptotics for general p were first considered explicitly only recently, when O'Cinneide and Pokrovskii [42] proved convergence of the rescaled height profile to an unspecified limit (our Corollary 19 reproves this and specifies the limit). However, asymptotics for the closely related p-trees model of random trees (Section 2.3(b)) have been extensively studied [18, 11, 9, 8]. It has recently become clear that an efficient way to study p-mapping asymptotics is to exploit the Joyal bijection between marked trees and mappings, which enables one to deduce asymptotics for p-mappings from already known asymptotics for p-trees. This approach, to be developed elsewhere [6], turns out to give stronger information about p-mappings than does the approach in this paper. But that method seems tied to the particular p-mappings model.

A final remark on our methodology. The sampling invariance principle for random trees (involving spanning subtrees of random vertices: Section 2.1) appears as part of the circle of ideas around the Brownian continuum random tree [4], but is somewhat overshadowed by the stronger and more succinct functional invariance principle for trees (Section 2.2). But in the context of mappings, studying the orbits of a fixed number of vertices (which is the essence of the sampling invariance principle) is very natural and often easy; and it is the statement of the functional invariance principle which is harder to interpret. That is why it seems worthwhile exploring the consequences of the sampling invariance principle.

2 Invariance principles for random trees

Consider the assertion

- For certain models of non-uniform random trees T on $[n]$, the $n \to \infty$ asymptotic distributions of many statistics should be the same as for the uniform model.

An *invariance principle* is a way of formalizing such an assertion. In this Section we give a slightly new perspective on some known results. We emphasize two apparently different ways of thinking about asymptotics (looking at spanning subtrees in Section 2.1; coding trees as walks in Section 2.2) and then describe carefully their relationship in Section 2.4. In Section 2.3 we recall hypotheses under which these invariance principles are true or conjectured.

2.1 Spanning subtrees and the sampling invariance principle

Consider a rooted tree T on vertex-set $[n]$. Take distinct vertices $\kappa_1, \ldots, \kappa_k$ and consider the spanning subtree on $\{\text{root}, \kappa_1, \ldots, \kappa_k\}$. Relabel vertex κ_i as i and unlabel other vertices, while still marking the root. Picture the resulting tree as on the left of figure 1. Take each edge to have length 1; then delete unlabeled degree-2 vertices to obtain a "tree with edge-lengths" where the edge-lengths are integers. Call this tree $\text{SPAN}(\kappa_1, \ldots, \kappa_k; T)$. For real $c > 0$ write $c \otimes \text{SPAN}(\kappa_1, \ldots, \kappa_k; T)$ for the tree obtained by multiplying all edge-lengths by c. For asymptotics we anticipate getting a "tree with edge-lengths" $\mathbf{t}$ as pictured on the right in figure 1. Such a tree $\mathbf{t}$ has the following properties.
(i) There is a degree-1 root, and k other leaves labeled by $[k]$.

(ii) It is a binary tree, with unlabeled degree-3 branchpoints.
(iii) Each edge has a strictly positive real length.

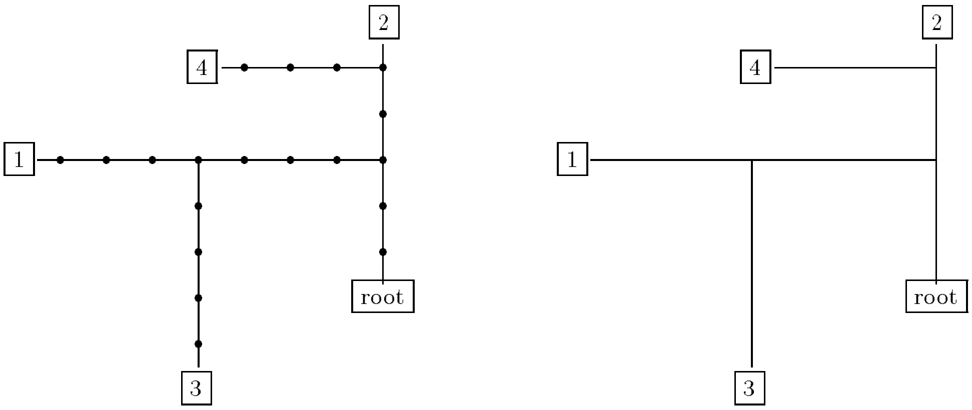


FIG. 1. *Instances of* $SPAN(\kappa_1, \ldots, \kappa_4; T)$ *and of a* $\mathbf{t} \in \mathbf{T}[4]$.

Write $\mathbf{T}[k]$ for the set of such graphs. A tree $\mathbf{t} \in \mathbf{T}[k]$ can be specified by its *shape* (pedantically, the shape is the equivalence class of isomorphic leaf-labeled rooted trees-without-edge-lengths) and by its $2k-1$ edge-lengths. Inductively, the number of shapes equals $(2k-3)!! := (2k-3)(2k-5)\cdots 3\cdot 1$ because the k'th leaf can be attached at $2k-3$ different places. (Make the convention $(-1)!! = 1$.) Following [4] Section 4.3, define a distribution for a random graph $T[k]$ with edge lengths $\mathcal{L}_1, \ldots, \mathcal{L}_{2k-1}$ as follows. For each of the $(2k-3)!!$ possible shapes t,

$$\begin{aligned} P(\mathrm{shape}(T[k]) = t, \mathcal{L}_1 \in d\ell_1, \quad \ldots \quad &, \mathcal{L}_{2k-1} \in d\ell_{2k-1}) \\ &= f_{2k-1}(\ell_1, \ldots, \ell_{2k-1})\ d\ell_1 \ldots d\ell_{2k-1} \quad (1) \end{aligned}$$

where

$$f_{2k-1}(\ell_1, \ldots, \ell_{2k-1}) := \left(\sum_{j=1}^{2k-1} \ell_j \right) \ \exp\left(-\tfrac{1}{2} \left(\sum_{j=1}^{2k-1} \ell_j \right)^2 \right), \qquad (2)$$

and where we adopt some arbitrary convention for ordering the edges associated with each possible shape. The convention does not matter because $f_{2k-1}(\cdot)$ is symmetric. Definition (1) - (2) implies that shape$(T[k])$ is uniform on the $(2k-3)!!$ possible shapes and that the edge-lengths are independent of the shape. Saying that (1) - (2) defines a *probability* density on $\mathbf{T}[k]$ is saying

$$(2k-3)!! \int \ldots \int f_{2k-1}(\ell_1, \ldots, \ell_{2k-1}) \; d\ell_1 \ldots d\ell_{2k-1} = 1;$$

check by rewriting the integral as $\frac{1}{(2k-2)!} \int_0^\infty s^{2k-2} \exp(-s^2/2) ds$. (There is an alternate interpretation of the distribution of $T[k]$ using a *line-breaking construction*: [4] Lemma 21.)

Note there is a natural notion of convergence in $\mathbf{T}[k]$: $\mathbf{t}^{(n)} \to \mathbf{t}$ if $\mathrm{shape}(\mathbf{t}^{(n)}) = \mathrm{shape}(\mathbf{t})$ ultimately and the edge-lengths converge. This convergence can be metrized. So convergence in distribution (written $\stackrel{d}{\to}$) of random trees with edge lengths means weak convergence with respect to the underlying metric.

By a *weighted tree* on vertices $[n]$ we mean a tree together with a probability distribution $q = (q_i)$ on the vertex-set $[n]$. Picture the q_i as deterministic "weights" on vertices, which for a random tree do not depend on the realization of the tree. We may consider the uniform distribution as a "default" choice of q, but allowing more generality may be useful. In the context of $n \to \infty$ asymptotics for such *random weighted trees* we always assume, without further mention, that the distributions $q = q^{(n)}$ satisfy

$$\max_i q_i \to 0 \text{ as } n \to \infty. \tag{3}$$

Definition 1 A model for random weighted trees (T, q) on $[n]$ satisfies the *sampling invariance principle with scaling constants* $c = c^{(n)}$ if, as $n \to \infty$,

$$c \otimes \mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T) \stackrel{d}{\to} T[k], \quad k \geq 1 \tag{4}$$

where $\{\kappa_1, \ldots, \kappa_k\}$ are independent of each other and of T with distribution $q^{(n)}$.

The word "sampling" is intended to convey the idea of "spanning tree on randomly sampled vertices". Note that for finite n the tree $\mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T)$ might not be in $\mathbf{T}[k]$, for instance if two of the κ_i coincide. However, one can make sense of convergence in distribution of random objects $\mathcal{X}_n$ even if the objects are well-defined only on events A_n with $P(A_n) \to 1$ (for instance by appending a "fictitious state"), and we adopt this view throughout the paper. Note that hypothesis (3) is exactly what is needed to ensure $\lim_n P(\kappa_1, \ldots, \kappa_k \text{ distinct }) = 1$.

Examples of models where Definition 1 is satisfied will be given shortly (section 2.3), after we recall an alternative notion of *invariance principle*.

2.2 Coding trees as walks: the functional invariance principle

In a rooted *ordered* tree, the children of each vertex are ordered as first, second, third .... For a rooted ordered tree on n vertices there is a *depth-first* relabeling of the vertices as $\{0, 1, \ldots, n-1\}$ defined as follows. Label the root as 0. Having labeled some vertex as i, give label $i+1$ to
(i) the first child of i, if any; else
(ii) the first not-yet-labeled child of parent(i), if any; else
(iii) the first not-yet-labeled child of parent(parent(i)); and so on.
After thus relabeling vertices in depth-first order, define a walk

$$w(i) = \text{height}(i), \ 0 \leq i \leq n-1$$

where the *height* of vertex i is its distance to the root. Note that the walk determines the tree, because vertex i is the next child of vertex $\max\{i' < i : w(i') = w(i) - 1\}$.

Given a weight function q on the vertices, we can define a rescaled walk $(\widetilde{w}(t), 0 \leq t \leq 1)$ by

$$\widetilde{w}(t) = w(i) \text{ for } q_0 + q_1 + \cdots + q_{i-1} \leq t < q_0 + q_1 + \cdots + q_i. \tag{5}$$

So $\widetilde{w}$ is an element of the usual space $D[0,1]$ of right-continuous functions with left limits [16]. The functional invariance principle (which we are about to define) relates to ordered trees, whereas the sampling invariance principle was defined for unordered trees. An unordered tree T can be made into an ordered tree by putting the children of each vertex into uniform random order. The resulting depth-first labeling of vertices of T by $\{0, 1, \ldots, n-1\}$ will be called the *randomized depth-first ordering* of T.

Write $B^{\text{exc}} = (B^{\text{exc}}_t, 0 \leq t \leq 1)$ for *standard Brownian excursion* [13, 15]. Consider the following property for a model of random weighted trees (T, q) on $[n]$ and constants $c = c^{(n)}$.

Property 2 *The rescaled walk $\widetilde{W}(t)$ defined by (5) based on the randomized depth-first ordering satisfies*

$$(c\widetilde{W}(t), 0 \leq t \leq 1) \ \stackrel{d}{\to} \ (2B^{\text{exc}}_t, 0 \leq t \leq 1) \ \textit{as} \ n \to \infty$$

with respect to a specified metric on $D[0,1]$.

Definition 3 A model for random weighted trees (T, q) on $[n]$ satisfies the *functional invariance principle with scaling constants $c = c^{(n)}$* if Property 2 holds for the usual (Skorokhod J_1) metric on $D[0,1]$.

We shall shortly show (Proposition 7) that the functional invariance principle implies the sampling invariance principle. But before continuing the theoretical development to show this, let us discuss some specific models of random trees where these invariance principles are known or conjectured.

2.3 Models of non-uniform random trees

Model (a). The uniform random rooted tree on $[n]$ is a special case (where ξ has Poisson distribution) of the CBP(n) *model* (here CBP stands for *conditioned branching process*). This is the Galton-Watson branching process, with some offspring distribution ξ satisfying

$$E\xi = 1;\ 0 < \sigma^2 := \mathrm{var}\ \xi < \infty\ ,$$

conditioned to have total size equal to n.

Theorem 4 *The CBP(n) model (with uniform weights q) satisfies the functional invariance principle with scaling constants $\sigma n^{-1/2}$.*

This was proved in [4] with a slightly different walk coding, though it is easy to deduce the stated form. See [38] for a more direct proof and further references; and see [21, 22] for cases with infinite variance. Because of the special combinatorial structure of the CBP(n) model one can obtain more refined results, for instance the local limit theorem corresponding to Corollary 10 [20].

Model (b). Given a probability distribution p on $[n]$, a *p-tree* [18, 45] is a random rooted tree T on vertex-set $[n]$ whose distribution is defined by

$$P(T = \mathbf{t}) = \prod_{i \in [n]} p_i^{C_i \mathbf{t}} \tag{6}$$

where $C_i\mathbf{t}$ denotes the number of children of vertex i in $\mathbf{t}$. For a probability distribution p on $[n]$ write

$$c_p := \sqrt{\sum_i p_i^2}.$$

For a sequence $p^{(n)}$ of probability distributions on $[n]$, introduce the condition

$$\frac{\max_i p_i}{c_p} \to 0 \text{ as } n \to \infty. \tag{7}$$

For reasons to be touched upon in Section 7(b), one can view condition (7) for p-trees as closely analogous to Lindeberg's condition in the central limit theorem.

Theorem 5 *Under assumption (7), a sequence of p-trees with weight functions $q = p$ satisfies the sampling invariance principle with scaling constants c_p.*

This is obtained by combining Corollary 9 of Camarri and Pitman [18], who specified the limit in terms of a line-breaking construction, with Lemma 21 of Aldous [4], which obtains the formula stated for $T[k]$ from the line-breaking construction. One can deduce from [18] that (7) is also *necessary* for the sampling invariance principle. But assumption (7) is not enough to imply the functional invariance principle for p-trees (the example for p-mappings in Section 6.1 can be adapted to p-trees). We do not know the precise necessary and sufficient condition.

In the analogous result for random mappings (Theorem 25) the weight functions can be arbitrary. This is surely also true in Theorem 5, though we have not attempted a proof.

Model (c). Attach i.i.d. costs to the edges of the complete graph on $[n]$, and then let T be the minimum-cost spanning tree, rooted at 1 say. Frieze [25] studied the total cost of T, and Aldous [3] studied some asymptotic distributions associated with the tree T itself, but the following conjecture remains open.

Conjecture 6 *T satisfies the sampling invariance principle with uniform weights and with scaling constants $n^{-1/2}$.*

Model (d). Given an irreducible Markov transition matrix $P = (p_{ij})$ on $[n]$, one can define a *P-tree* via

$$P(T = \mathbf{t}) \propto \prod_{(i,j)\in\mathbf{t}} p_{ij}$$

where we are regarding $\mathbf{t}$ as a set of edges directed toward the root. Such trees arise as part of a circle of ideas around the *Markov chain tree theorem* [37]. It seems plausible that one can find conditions on P that imply invariance principles, but this setting has apparently not been studied.

2.4 Relating the sampling and functional invariance principles

The definitions of the sampling invariance principle and the functional invariance principle look quite different, but we now give a result revealing a close relationship. On the function space $D[0,1]$, the usual Skorokhod metric is stronger than the L_0 metric

$$||f_1 - f_2||_0 := \int_0^1 \min(1, |f_1(t) - f_2(t)|)\ dt.$$

It is easy to check there is yet another metric on $D[0,1]$, which we call the "$*$-metric", intermediate between the L_0 and the Skorokhod metrics, with the property

for $f_n, f \in D[0,1]$ with f continuous, $f_n \to^* f$ iff

$$\begin{aligned} ||f_n - f||_0 &\to 0 \text{ and} \\ \inf_{a \le t \le b} f_n(t) &\to \inf_{a \le t \le b} f(t) \text{ for all intervals } [a,b] \text{ of positive length.} \end{aligned} \quad (8)$$

Proposition 7 *A model for random weighted trees (T,q) on $[n]$ satisfies the sampling invariance principle with scaling constants $c = c^{(n)}$ if and only if Property 2 holds for the $*$-metric.*

Definition 3 and this alternate characterization of sampling invariance make it clear that the functional invariance principle does imply the sampling invariance principle. Proposition 7 is new, though the key underlying conceptual fact (Lemma 8) is known; let us explain this fact first.

Given $f : [0,1] \to [0,\infty)$ and distinct $\{u_1, u_2, \ldots, u_k\} \subset (0,1)$, one can define a tree-with-edge-lengths $\mathbf{t}(u_1, \ldots, u_k; f)$ by specifying
(i) leaf i is at height (distance from root) $f(u_i)$;
(ii) if $u_i < u_j$ then the branchpoint between the paths from the root to i and to j is at height $\min_{u_i \le t \le u_j} f(t)$.

We can make $\mathbf{t}(u_1, \ldots, u_k; f)$ into an ordered tree by giving the leaves i the order inherited from the natural order of the u_i in $(0,1)$.

Now let $U_1, \ldots, U_k$ be i.i.d. uniform $(0,1)$ random variables, independent of B^{exc}. Recall $T[k]$ from Section 2.1; we can make $T[k]$ into an ordered tree by putting, independently at each branchpoint, the two edges leading away from the root into random order.

Lemma 8 ([4] Corollary 22) $\mathbf{t}(U_1, \ldots, U_k; 2B^{\mathrm{exc}}) \stackrel{d}{=} T[k]$.

Proof of Proposition 7. Now for a tree T and a weight function q, make the tree ordered as in Section 2.2, and write

$$\kappa(u) := \min\{i : q_0 + q_1 + \ldots + q_i \geq u\}. \tag{9}$$

So $\kappa(\cdot)$ depends on the tree via the ordering of vertices. Take $(U_1, \ldots, U_k)$ independent of the random trees T. So $\kappa(U_1), \ldots, \kappa(U_k)$ are independent random vertices chosen from distribution q. Therefore $\mathrm{SPAN}(\kappa_1, \ldots, \kappa_k; T) \stackrel{d}{=} \mathrm{SPAN}(\kappa(U_1), \ldots, \kappa(U_k); T)$. But from their definitions, $\mathrm{SPAN}(\kappa(U_1), \ldots, \kappa(U_k); T)$ and $\mathbf{t}(U_1, \ldots, U_k; \widetilde{W})$ are almost the same; the difference is that the heights of branchpoints in the latter are exactly 1 less than their heights in the former, and because $c \to 0$ this difference vanishes asymptotically. Thus by Lemma 8 the sampling invariance principle is equivalent to: for each k,

$$\mathbf{t}(U_1, \ldots, U_k; c\widetilde{W}) \stackrel{d}{\to} \mathbf{t}(U_1, \ldots, U_k; 2B^{\mathrm{exc}}). \tag{10}$$

Write $0 < V_1 < V_2 < \ldots < V_k < 1$ for the order statistics of $(U_1, \ldots, U_k)$. Consider the assertion

$$c\left(\widetilde{W}_n(V_1), \inf_{V_1 \leq t \leq V_2} \widetilde{W}_n(t), \widetilde{W}_n(V_2), \inf_{V_2 \leq t \leq V_3} \widetilde{W}_n(t), \ldots, \widetilde{W}_n(V_k)\right)$$

$$\stackrel{d}{\to} (2B^{\mathrm{exc}}(V_1), \inf_{V_1 \leq t \leq V_2} 2B^{\mathrm{exc}}(t), 2B^{\mathrm{exc}}(V_2), \inf_{V_2 \leq t \leq V_3} 2B^{\mathrm{exc}}(t), \ldots, 2B^{\mathrm{exc}}(V_k)). \tag{11}$$

Using the fact that (V_i) is independent of the permutation associating the (U_i) with the (V_i), and the random ordering of branches within trees, we see that (11) implies (10). Conversely, since in (10) the trees $\mathbf{t}(\cdot)$ are ordered, from a realization of $\mathbf{t}(\cdot)$ we can determine the permutation associating the (U_i) with the (V_i), and it follows that (10) implies (11).

We have now reformulated the sampling invariance principle as (11). Proposition 7 is a consequence of this reformulation, together with the following reformulation of $*$-convergence. (Note that in our setting the first and last components of the vectors below are automatically tending to zero).

Lemma 9 *Let $X_n(t)$ and $X(t)$ be processes in $D[0,1]$, and suppose $X(\cdot)$ has continuous paths. Then $X_n \stackrel{d}{\to} X$ with respect to the $*$-metric if and only if for each k*

$$\left(\inf_{0 \leq t \leq V_1} X_n(t), X_n(V_1), \inf_{V_1 \leq t \leq V_2} X_n(t), X_n(V_2), \inf_{V_2 \leq t \leq V_3} X_n(t), \ldots, X_n(V_k), \inf_{V_k \leq t \leq 1} X_n(t)\right)$$

$$\stackrel{d}{\to} \left(\inf_{0 \leq t \leq V_1} X(t), X(V_1), \inf_{V_1 \leq t \leq V_2} X(t), X(V_2), \inf_{V_2 \leq t \leq V_3} X(t), \ldots, X(V_k), \inf_{V_k \leq t \leq 1} X(t)\right). \tag{12}$$

Proof. The "only if" is immediate. For "if", the key fact (proved as Proposition 29 in the Appendix) is that convergence in distribution with respect to the L_0 metric is equivalent to

$$(X_n(V_1), X_n(V_2), \ldots, X_n(V_k)) \stackrel{d}{\to} (X(V_1), X(V_2), \ldots, X(V_k)), \quad k \geq 1. \tag{13}$$

So in particular, assumption (12) implies $X_n \stackrel{d}{\to} X$ in L_0. Now let us be more precise about notation and write $(V_{k,i}, 1 \leq i \leq k)$ for the order statistics of $(U_i, 1 \leq i \leq k)$. Fix $a < b$ and consider $i(k) \sim ak$ and $j(k) \sim bk$ as $k \to \infty$. For fixed k,

$$\inf_{V_{k,i(k)} \leq t \leq V_{k,j(k)}} X_n(t) \stackrel{d}{\to} \inf_{V_{k,i(k)} \leq t \leq V_{k,j(k)}} X(t)$$

and so by taking $k = k_n \to \infty$ sufficiently slowly

$$\inf_{V_{k_n,i(k_n)} \leq t \leq V_{k_n,j(k_n)}} X_n(t) \stackrel{d}{\to} \inf_{a \leq t \leq b} X(t),$$

using path-continuity of X. We could choose $i(k)$ such that $V_{k,i(k)} > a$ ultimately, or such that $V_{k,i(k)} < a$ ultimately, and analogously for $j(k)$; so

$$\inf_{a \leq t \leq b} X_n(t) \stackrel{d}{\to} \inf_{a \leq t \leq b} X(t). \tag{14}$$

Using the Skorokhod representation theorem ([24] Theorem 3.1.8) we may assume $X_n \stackrel{a.s.}{\to} X$ in L_0, and then convergence in (14) must be a.s. also. Since $a < b$ is an arbitrary interval, it follows that $X_n \stackrel{a.s.}{\to} X$ in the $*$-metric.

Remark. The construction of a tree-with-edge-lengths $\mathbf{t}(u_1, \ldots, u_k; 2B^{\text{exc}})$ with an arbitrary finite number k of leaves labeled by $\{u_1, \ldots, u_k\} \subset [0, 1]$ extends to the construction [4] of the Brownian *continuum random tree* (CRT) whose leaves are indexed by almost all $u \in [0, 1]$, but we do not emphasize that formalism in this paper.

2.5 Distributional limits implied by invariance principles

Here we briefly recall some instances of what can be deduced from each type of invariance. Write height(i, T) for the height of vertex i in tree T. In the context of an invariance principle, define the *rescaled cumulative height profile* by

$$H_n(u) := \sum_{i=1}^{n} q_i^{(n)} 1_{(\text{height}(i,T) \leq u/c_n)}.$$

Regard H_n as a random element of the space $\mathcal{D}$ of distribution functions, where $\mathcal{D}$ is equipped with the topology of "convergence in distribution". Note that the map $D[0,1] \to \mathcal{D}$ defined by

$$f \to H_f : \quad \text{where } H_f(u) := \int_0^1 1_{(f(t) \le u)} \, dt$$

is L_0-continuous at almost all realizations of B^{exc}. So we can use the continuous mapping theorem to get

Corollary 10 *Consider a model for random weighted trees (T, q) on $[n]$ which satisfies the sampling invariance principle with scaling constants c. Then*

$$H_n \stackrel{d}{\to} H$$

as random elements of $\mathcal{D}$, where

$$H(u) := \int_0^1 1_{(2B_t^{\text{exc}} \le u)} \, dt; \quad 0 \le u < \infty.$$

See [27, 31] for discussion of the explicit distribution of H.

The map $f \to \sup_{0 \le t \le 1} f(t)$ is continuous with respect to the Skorokhod topology but not with respect to the $*$-topology, so we cannot deduce the following from the *sampling* invariance principle.

Corollary 11 *Consider a model for random weighted trees (T, q) on $[n]$ which satisfies the functional invariance principle with scaling constants c. Then*

$$c \cdot \max_i \text{height}(i, T) \stackrel{d}{\to} \sup_{0 \le t \le 1} 2B_t^{\text{exc}}.$$

See [14] Section 4.1 for discussion of the explicit distribution of $\sup_t B_t^{\text{exc}}$.

To visualize the distinction between the sampling and functional invariance principles, consider trees $T^{(n)}$ satisfying the functional invariance principle with respect to uniform weights. Take $m = o(n)$ and make trees $\hat{T}^{(m+n)}$ by linking an arbitrary m-vertex tree $\tilde{T}^{(m)}$ to the root of $T^{(n)}$. Then $\hat{T}^{(m+n)}$ will still satisfy the *sampling* invariance principle, but in general not the functional invariance principle. One can easily make examples where the convergence of maximum heights assertion in Corollary 11 fails for $\hat{T}^{(m+n)}$. In the random walk coding, the point is that the walk excursion for $\tilde{T}^{(m)}$ is vanishingly short and so is not noticed by L_0 convergence. At the technical level, if one knows that the weak invariance principle holds, then to prove the strong invariance principle one needs only to prove tightness of the rescaled walks in the Skorokhod metric. It would be interesting to find useful sufficient conditions for tightness in our random tree context.

3 Formulating the sampling invariance principle for random mappings

In a sense, the rest of the paper is devoted to describing the "random mappings" analogs of the results for random trees in Section 2. The reader may wish first to look at Section 6 for specific models of random mappings.

For a mapping M and an element $i \in [n]$, consider the iterates $M^0(i) = i, M^1(i) = M(i), M^{j+1}(i) = M(M^j(i)),\ j \geq 1$. There is a smallest number $L_1 + L_2$ such that

$$M^{L_1+L_2}(i) = M^{L_1}(i) \text{ for some } 0 \leq L_1 < L_1 + L_2$$

and the associated L_1 is unique. As shown in the left of figure 2, we picture the set of interates of i as a line of length L_1 attached to a cycle of length $L_2 - L_1$.

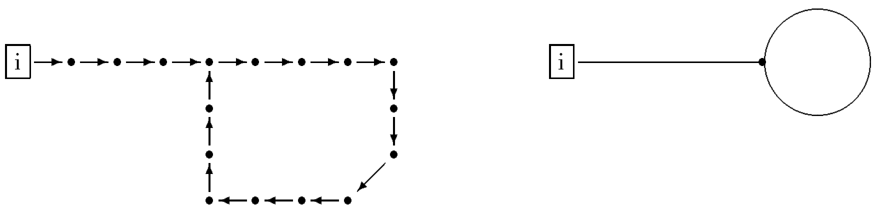


FIG. 2.

For the *uniform* random mapping M, the elementary "birthday problem" argument shows

$$P(L_1 = x, L_2 = y) = (n)_{x+y}/n^{x+y+1}, \quad x = 0, 1, 2, \ldots;\ y = 1, 2, \ldots$$

where $(n)_k := n(n-1)(n-2)\cdots(n-k+1)$. In the $n \to \infty$ limit,

$$n^{-1/2}(L_1, L_2) \ \xrightarrow{d}\ (\mathcal{L}_1, \mathcal{L}_2), \text{ where } f_{\mathcal{L}_1,\mathcal{L}_2}(\ell_1, \ell_2) = \exp(-(\ell_1+\ell_2)^2/2). \quad (15)$$

We picture the limit as a "mapping with edge-lengths", as in right of figure 2: there are two edges, a line of length $\mathcal{L}_1$ and a loop of length $\mathcal{L}_2$.

Instead of starting with a single vertex, one can fix k distinct vertices $\kappa_1, \ldots, \kappa_k$ and consider the graph of all iterates $(M^j(\kappa_i), j \geq 0, 1 \leq i \leq k)$. Relabel vertex κ_i as i and unlabel other vertices. Picture the resulting graph as on the top of figure 3, which shows an example where no κ_j falls in the orbit of any $\kappa_i,\ i < j$. Take each edge to have length 1; then delete unlabeled degree-2 vertices to obtain a "mapping with edge-lengths" where the edge-lengths are integers. Call this graph ORBITS$(\kappa_1, \ldots, \kappa_k; M)$. For asymptotics, after suitably scaling edge-lengths we anticipate getting a "mapping with edge-lengths" $\mathbf{g}$ as pictured on the bottom in figure 3.

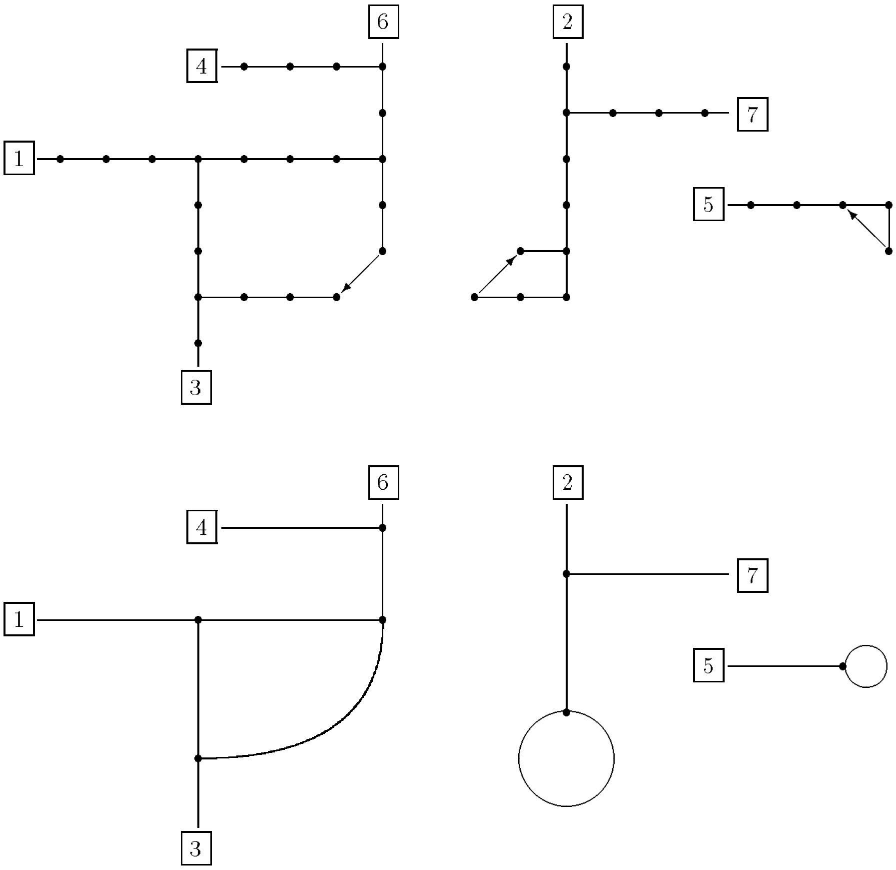


FIG. 3. *Instances of ORBITS*$(\kappa_1, \ldots, \kappa_7; M)$ *and of a* $\mathbf{g} \in \mathbf{G}[7]$.

Such a graph $\mathbf{g}$ has the following properties.
(i) Each component consists of some number $\gamma \geq 1$ of trees attached to a directed cycle consisting of γ edges.
(ii) Each tree is an unordered binary tree, with labeled leaves. The degree-3 branchpoints within trees or where trees meet cycles are unlabeled.
(iii) The set of leaf-labels is $[k]$.
(iv) Each edge has a strictly positive real length.

Write $\mathbf{G}[k]$ for the set of such graphs, which one could call *mappings with edge lengths.* A graph $\mathbf{g} \in \mathbf{G}[k]$ can be specified by its *shape* (pedantically, the shape is the equivalence class of isomorphic leaf-labeled graphs-without-edge-lengths) and by its edge-lengths. Inductively, the number of edges equals $2k$ (adding a new leaf creates two extra edges) and the number of shapes equals $(2k-1)!! := (2k-1)(2k-3)(2k-5)\cdots 3\cdot 1$ because the k'th leaf can be attached at $2k-1$ different places (the $2k-2$ existing edges, or a new component). This closely parallels the discussion of tree-with-edge-lengths in Section 2.1. Analogous to the definition (1) of $T[k]$, we now define a distribution for a random mapping with edge lengths, $G[k]$, as follows.

For each possible shape g,

$$P(\text{shape}(G[k]) = g, \mathcal{L}_1 \in d\ell_1, \ldots, \mathcal{L}_{2k} \in d\ell_{2k}) = f^*_{2k}(\ell_1, \ldots, \ell_{2k})\ d\ell_1 \ldots d\ell_{2k} \tag{16}$$

where

$$f^*_{2k}(\ell_1, \ldots, \ell_{2k}) := \exp\left(-\tfrac{1}{2}\left(\sum_{j=1}^{2k}\ell_j\right)^2\right). \tag{17}$$

Compare with (2) and note the missing prefactor in (17); a calculation at (19) later will illuminate the connection between f^*_{2k} and f_{2k-1}. As before, definition (16) implies that shape$(G[k])$ is uniform on the $(2k-1)!!$ possible shapes and that edge-lengths are independent of shape.

We can introduce *weighted* mappings in the same way as weighted trees: there is a probability distribution (weight function) $q = q^{(n)}$ on $[n]$. We can talk about convergence in $\mathbf{G}[k]$ as we did in $\mathbf{T}[k]$: $\mathbf{g}^{(n)} \to \mathbf{g}$ if shape$(\mathbf{g}^{(n)}) =$ shape$(\mathbf{g})$ ultimately and the edge-lengths converge. We can now copy the format of Definition 1.

Definition 12 *A model of random weighted mappings (M, q) on $[n]$ satisfies the sampling invariance principle with scaling constants $c = c^{(n)}$ if, as $n \to \infty$,*

$$c \otimes \text{ORBITS}(\kappa_1, \ldots, \kappa_k; M) \ \stackrel{d}{\to}\ G[k], \quad k \geq 1 \tag{18}$$

where $\kappa_1, \ldots, \kappa_k$ are independent of each other and of M with distribution $q^{(n)}$.

As noted below Definition 1, our standing hypothesis (3) on the weights q implies $\lim_n P(\kappa_1, \ldots, \kappa_k \text{ distinct }) = 1$ and we only require the left side to take values in $\mathbf{G}[k]$ on events of probability $\to 1$ as $n \to \infty$.

3.1 A construction of $G[k]$

Motivation for studying the random mapping with edge-lengths $G[k]$ with distribution (17) comes from its appearance as the limit (18) for the uniform random mapping (this is elementary, or a special case of Theorem 25 below). Now we already mentioned that the analogous random tree with edge-lengths $T[k]$ has a direct line-breaking construction [4, 9]. We will show below that there is a simple way to construct $G[k]$ from $T[k+1]$. This construction, illustrated in figure 4, is a "graphs with edge lengths" analog of Joyal's bijection exploited in [6], though we will not elaborate on the analogy here.

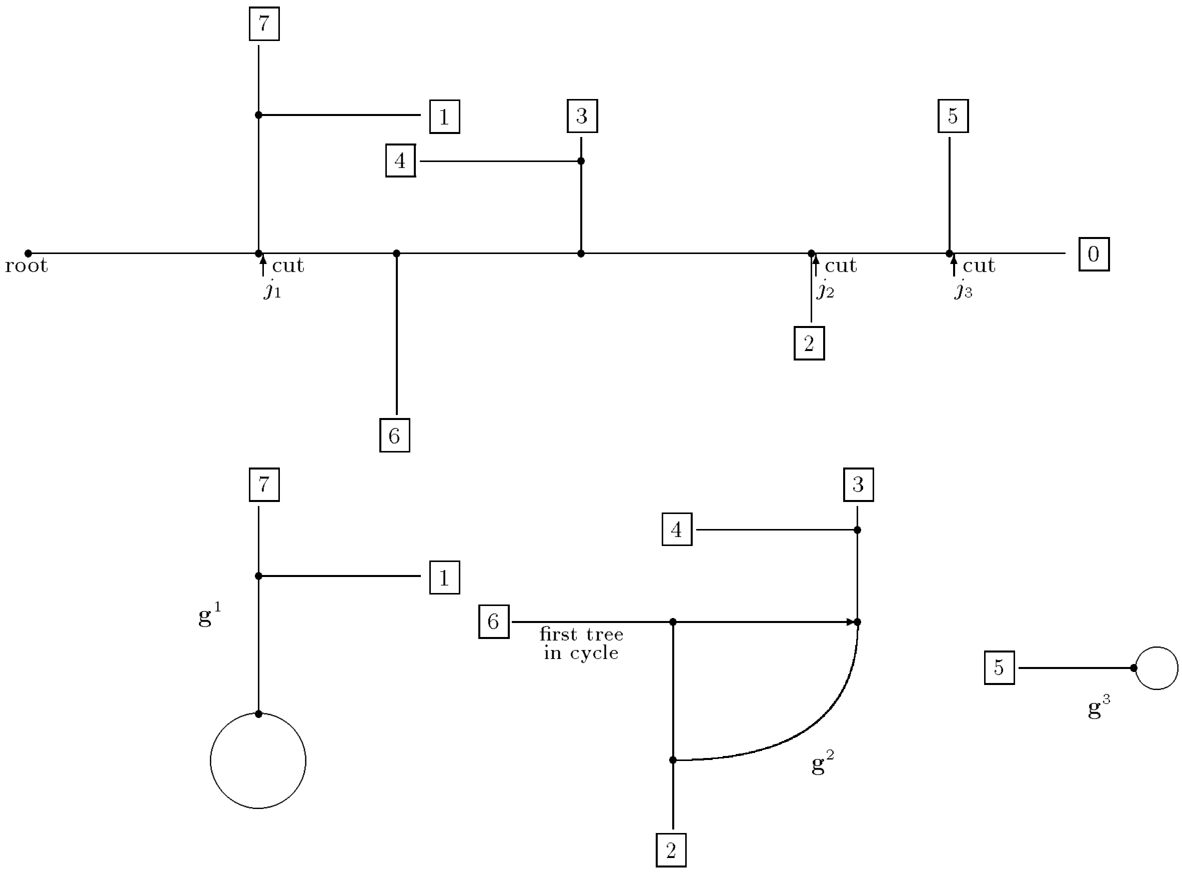


FIG. 4. *The map $J : \mathbf{T}[0,k] \to \mathbf{G}[k]$ which takes the distribution of $T[k+1]$ to the distribution of $G[k]$.*

Start with a tree with edge lengths $\mathbf{t} = \mathbf{t}^1$ with leaves labeled $0, 1, \ldots, k$.

In the figure, leaf i is denoted $\boxed{\mathrm{i}}$. Call the path from the root to leaf 0 the *spine*. Let j_1 be the junction where the path from leaf 1 to the root meets the spine. At j_1 cut away the edge leading toward leaf 0. Make the line from j_1 to root into a cycle by identifying the root with j_1, to form a component $\mathbf{g}^1$. In the remaining part of the original tree, make a new root at the endpoint previously at j_1 to define a new tree with edge lengths, say $\mathbf{t}^2$. Repeat recursively, letting j_2 be the junction where the path in $\mathbf{t}^2$ from the lowest-numbered leaf (except 0) meets the spine, to construct another component $\mathbf{g}^2$ with a cycle defined by identifying its root with j_2. Continue until the remaining tree $\mathbf{t}^{j+1}$ consists only of leaf 0 and a single edge; discarding that tree, the components $\mathbf{g}^1, \ldots, \mathbf{g}^j$ form a mapping with edge lengths. Cycles are directed according to the direction along the original spine from the root to leaf 0.

Abusing notation slightly, write $\mathbf{T}[0, k]$ for the set of trees with edge lengths with leaf-set $\{0, 1, \ldots, k\}$.

Lemma 13 *The map $J : \mathbf{T}[0, k] \to \mathbf{G}[k]$ described above carries the probability density f_{2k+1} defined at (2) to the density f_{2k}^* defined at (17).*

Proof. It is straightforward to check the map is a bijection between the two sets of *shapes*, each having cardinality $(2k-1)!!$. Checking the assertion about densities reduces to checking the consequence of integrating out the contribution from the length ℓ_0 of the discarded edge incident to 0, that is to checking

$$f_{2k}^*(\ell_1, \ldots, \ell_{2k}) = \int_0^\infty (\sum_{j=0}^{2k} \ell_j) \ \exp\left(-\tfrac{1}{2}\left(\sum_{j=0}^{2k} \ell_j\right)^2\right) \ d\ell_0. \tag{19}$$

But writing $s = \sum_{j=1}^{2k} \ell_j$, this is just the integration by parts formula

$$\int_0^\infty (\ell_0 + s) \exp(-\tfrac{1}{2}(\ell_0 + s)^2) \ d\ell_0 = \exp(-\tfrac{1}{2} s^2).$$

4 Brownian bridge and random mapping-walks

Here we recall from [7] how to code a mapping as a walk, and a version of the functional invariance principle for the uniform model of random mappings.

4.1 Coding mappings as walks plus marks

Recall from Section 2.2 the coding of a rooted ordered tree as a walk. The corresponding coding for mappings is more intricate. Here's the key conceptual idea from [7], with a different coding.

> To a mapping M on $[n]$ one can associate a *walk* and *marks*. The walk and marks determine the mapping, up to vertex-labels. Given a mapping, the definition of the walk and marks involves three levels of choices of orderings.
> (a) The trees in the mapping's digraph are naturally "unordered trees"; we need to make them ordered trees.
> (b) The components of the digraph are unordered; we need to impose an order.
> (c) The trees attached to a cycle in a component are (only) cyclically ordered; we need to specify a "first" tree.
> The mapping and these choices determine the walk. Some choices will make the probabilistic structure more tractable.

Here are the details, deferring order choices until later. A *walk* is a sequence $0 = w(0), w(1), w(2), \ldots, w(n-1)$ satisfying

$$w(i) \in \{0, 1, 2, 3, \ldots\}, \quad w(i+1) \leq w(i) + 1.$$

So maybe $w(i+1) < w(i) - 1$. The *marks* are integers $0 = d(0) < d(1) < \ldots \leq n-1$, and for each marked integer $d(j)$ we require $w(d(j)) = 0$.

Using a mapping and order choices to define a walk and marks. An order on components and an order of trees within each component specify an order on all the trees, so breaking cyclic edges makes a forest (consisting of trees, whose roots are the original cyclic vertices) whose trees are in specified order. To each tree is associated a walk, as in Section 2.2. Concatenating the walks $\tilde{w}$ for each tree, in the specified order of trees, gives the walk w for the mapping M, and defines a relabeling of the vertices as $\{0, 1, \ldots, n-1\}$. Then mark each i for which, after this relabeling, vertex i is in a different component of the mapping than all vertices $i' < i$. So the marked vertices are the first vertex of each component.

Using the walk and marks to define a mapping. We will label vertices as $\{0, 1, \ldots, n-1\}$. The components of the digraph of M will be $\{0, 1, \ldots, d(1) - 1\}$, $\{d(1), d(1) + 1, \ldots, d(2) - 1\}$, ... The value $w(i)$

will be the height of vertex i above the cycles of M. So the cyclic vertices i will be exactly the vertices with $w(i) = 0$. If $i < j$ are the positions of successive visits of $w(\cdot)$ to 0, then $(\tilde{w}(0), \tilde{w}(1), \ldots, \tilde{w}(j-i-1)) = (w(i), w(i+1), \ldots, w(j-1))$ is a walk which determines a tree, and the order of the excursions within each component determines the cyclic order in which trees are attached to the cycle.

Together with the walk we define

$$l(j) := \sum_{i=0}^{i} 1_{(w(i)=0)}. \tag{20}$$

So $l(j)$ is the number of vertices $i \leq j$ which are cyclic.

We now specify our order choices, for coding a mapping as a walk. Other choices are possible, as discussed in [1]. The choices will use external randomization.

(a) Within each tree, children of each vertex are put in uniform random order.

(b) Order components in q-biased random order. That is, choose random $v_1 \in [n]$ according to $q^{(n)}$, and let C_1 be the component containing vertex v_1. Then choose random $v_2 \in [n] \setminus C_1$ with probabilities proportional to $q^{(n)}$, and let C_2 be the component containing vertex v_2; and so on.

(c) Within component C_j, put trees in cyclic order such that the tree containing vertex v_j is last.

Summary. Given a deterministic or random mapping on $[n]$, the construction above leads to a random walk $(w(i), 0 \leq i \leq n-1)$ and random marks $0 = d(0) < d(1) < \ldots \leq n-1$, and a relabeling of vertices by $\{0, 1, \ldots, n-1\}$. From the walk and marks we can reconstruct the mapping digraph, up to permutation of vertices. Figure 5 later illustrates such a walk and marks (more precisely, their rescaling described below). Call $(w(i), 0 \leq i \leq n-1)$ the *mapping-walk*.

Finally, for a component C write $q(C) := \sum_{i \in C} q_i$ for the weight of C, and write $\mathrm{cycle}(C)$ for the cycle length in component C. We will give results for the joint distributions

$$\left((q(C_j), c \cdot \mathrm{cycle}(C_j)), j \geq 1\right) \tag{21}$$

where as above the components (C_j) are in q-biased random order.

4.2 Rescaling mapping-walks

To discuss weak convergence as $n \to \infty$ of mapping-walks $(w(i), 0 \le i \le n) = (w^{(n)}(i), 0 \le i \le n)$ associated with random mappings $M = M^{(n)}$, we need to introduce rescalings. Given weights $q(i) = q^{(n)}(i)$, set

$$\tau(i) := q(0) + q(1) + \cdots + q(i-1)$$

and define the rescaled walk $(\bar{w}(t), 0 \le t \le 1)$ by

$$\bar{w}(t) := w(i) \text{ for } \tau(i) \le t < \tau(i+1).$$

Define a rescaled "cyclic vertex counting" process $(\bar{l}(t), 0 \le t \le 1)$ by

$$\bar{l}(t) := l(i) \text{ for } \tau(i) \le t < \tau(i+1).$$

And rescale the marks $0 = d(0) < d(1) < \ldots$ by defining

$$\bar{d}(j) := \tau(d(j)),\ j = 0, 1, 2, \ldots$$

so that $0 = \bar{d}(0) < \bar{d}(1) < \ldots < 1$.

4.3 The Brownian bridge limit

Write $B^{|\mathrm{br}|} = (B_t^{|\mathrm{br}|}, 0 \le t \le 1)$ for *reflecting Brownian bridge*, obtained from standard Brownian motion (B_t) via

$$B_t^{|\mathrm{br}|} := |B_t - tB_1|.$$

It turns out that the role of Brownian excursion for random trees is played by reflecting Brownian bridge for random mappings. To code the component structure of the mappings, we require some external randomization provided by an infinite i.i.d. uniform $(0,1)$ sequence $(\widetilde{U}_i)$. Define $D_0 = 0$,

$$D_1 := \inf\{t \ge \widetilde{U}_1 : B_t^{|\mathrm{br}|} = 0\}$$

and then for $r = 1, 2, \ldots$ define

$$D_{r+1} := \inf\{t \ge D_r + \widetilde{U}_{r+1}(1 - D_r) : B_t^{|\mathrm{br}|} = 0\}. \tag{22}$$

Write $(L(t), 0 \le t \le 1)$ for *local time at* 0 for $B^{|\mathrm{br}|}$, normalized so that $P(L(1) > s) = \exp(-s^2/2),\ s > 0$ [36].

Aldous and Pitman proved a variety of results for the uniform model of random mappings, including the following.

Theorem 14 ([7]) *For the uniform random mapping M on $[n]$, with uniform weights, as $n \to \infty$*

$$n^{-1/2}\bar{w}^{(n)} \xrightarrow{d} 2B^{|\mathrm{br}|} \tag{23}$$

$$n^{-1/2}\bar{l}^{(n)} \xrightarrow{d} L \tag{24}$$

in the sense of weak convergence of processes in the usual Skorokhod metric on $D[0,1]$. Moreover

$$\left(\left(q(C_j^{(n)}), n^{-1/2} \cdot \mathrm{cycle}(C_j^{(n)})\right), j \geq 1\right) \xrightarrow{d} \left((D_j - D_{j-1}, L(D_j) - L(D_{j-1})), j \geq 1\right) \tag{25}$$

The proof of Theorem 14 in [7] used a different, but asymptotically equivalent, way of coding trees as walks. Also, in section 4.1 we used a specific choice of ordering of components to define the mapping-walk. It turns out there is an alternate choice based on size-biasing of *cycles* which also leads to $B^{|\mathrm{br}|}$ asymptotics. These random mapping considerations lead to two different recursive decompositions of $B^{|\mathrm{br}|}$, whose structure is explored in [1], including explicit descriptions of the distribution of the right side of (25).

Note that the total cycle length is $\bar{l}^{(n)}(1) = \sum_j \mathrm{cycle}(C_j^{(n)})$. The asymptotic result

$$n^{-1/2} \cdot \sum_j \mathrm{cycle}(C_j^{(n)}) \xrightarrow{d} L(1) \tag{26}$$

follows from (24) but cannot be deduced directly from (25).

5 Functional implications of the sampling invariance principle for random mappings

We finally arrive at the central point of the paper. Though the assertion of convergence to $B^{|\mathrm{br}|}$ in Theorem 14 is elegant, the choice of extra properties (24,25) seem somewhat arbitrary, as do the details of the mapping-walk definition. In contrast, the sampling invariance principle is a "natural" assertion without any arbitrary choices. One anticipates, by analogy with Proposition 7 for random trees, that the sampling invariance principle itself should imply results in the general format of Theorem 14. Theorem 15 and Proposition 20 provide such results.

Theorem 15 *Suppose a model of random weighted mappings* (M, q) *on* $[n]$ *satisfies the sampling invariance principle with scaling constants* $c = c^{(n)}$*. Then as* $n \to \infty$

$$((q(C_j), c \cdot \text{cycle}(C_j)), j \geq 1) \stackrel{d}{\to} ((D_j - D_{j-1}, L(D_j) - L(D_{j-1})), j \geq 1) \tag{27}$$

and

$$c\bar{w} \stackrel{d}{\to} 2B^{|\text{br}|} \tag{28}$$

in the sense of weak convergence of processes in the $*$*-metric defined in Section 2.4.*

Remarks. Here (27) just repeats (25). The conclusion of Theorem 15 is weaker than Theorem 14 in two ways. First, in (28) we have a weaker topology on function space than in (23) – in the trees setting, this was the *only* difference. Second, we do not have convergence (26) of total cycle length (nor the refinement (24)). See Section 5.1 for further discussion of asymptotic distributions.

The proof of Theorem 15, which occupies the rest of this Section, follows the general outline of the proof of Proposition 7, with extra complications dealing with component ordering conventions. We will need an analog of Lemma 8, stated as Lemma 16 below, which links sampling invariance to Brownian bridge. We then proceed to the analog of (11), stated as Lemma 17. First we need to specify an operation $\mathbf{g}(\cdot)$ (analogous to $\mathbf{t}(\cdot)$ in Section 2.4) which describes how graphs-with-edge-lengths are obtained from functions and marks.

Construction (illustrated in figure 5). Given:

• a function $f : [0, 1] \to [0, \infty)$;

• a non-decreasing function $l : [0, 1] \to [0, \infty)$ with $l(0) = 0$ and whose points of increase are contained in $\{t : f(t) = 0\}$;

• marks $0 = d(0) < d(1) < \ldots \leq 1$, such that each $f(d(i)) = 0$;

• distinct $u_1, \ldots, u_k \subseteq (0, 1)$

we construct a graph-with-edge-lengths $\mathbf{g}(u_1, \ldots, u_k; f, l, d)$ as follows.

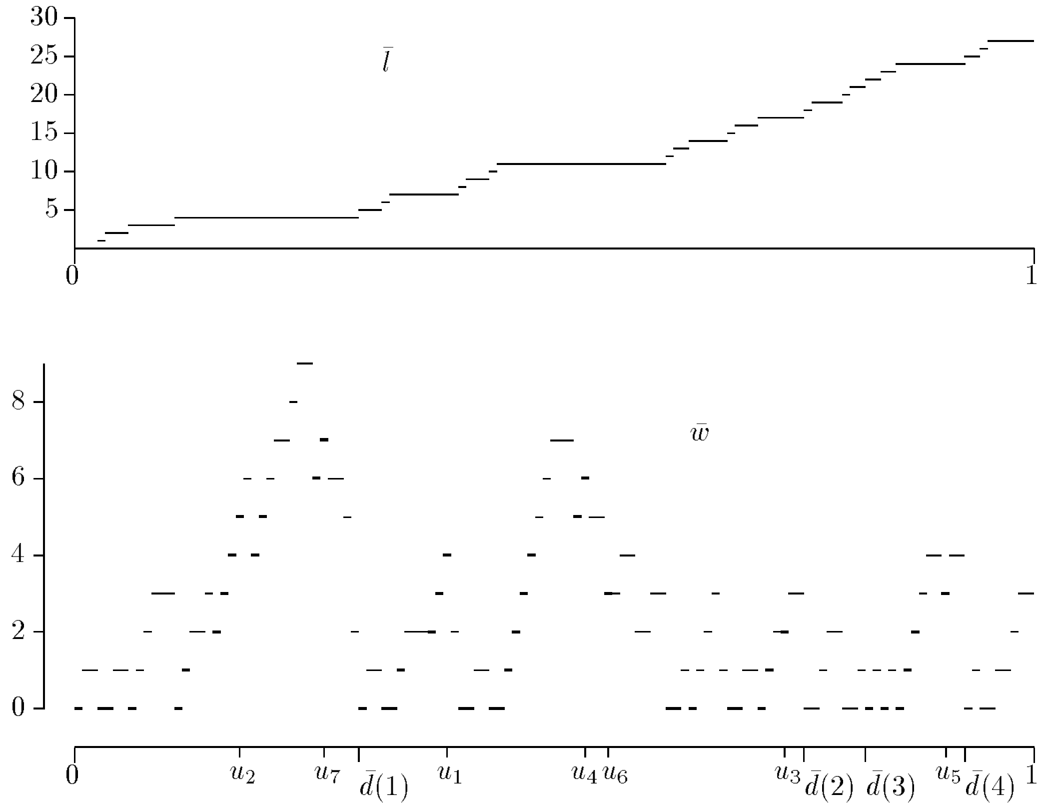
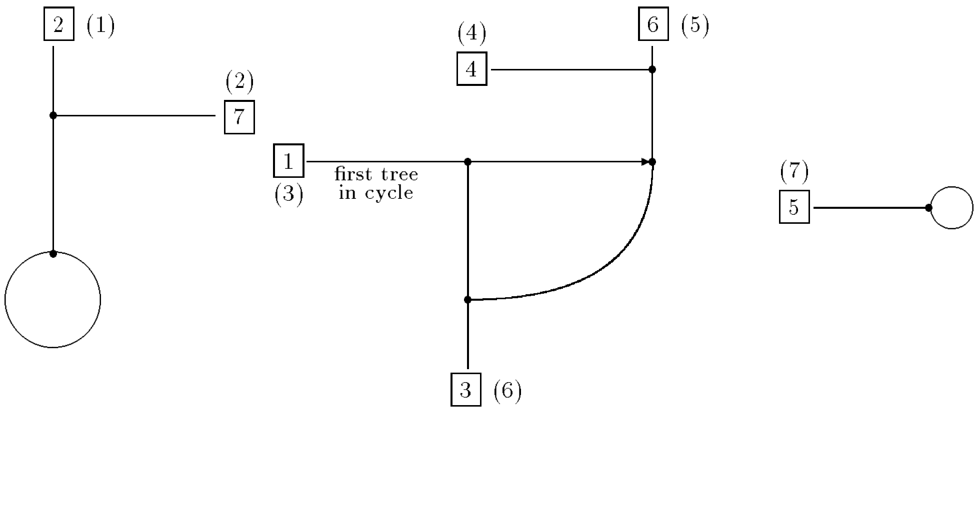


FIG. 5. *The rescaled walk $\bar{w}$ coding a certain mapping m on* [125]*, with equal weights, and the rescaled marks $\bar{d}(r)$. The points u_i identify seven vertices which can be labeled as $\{\kappa_1, \ldots, \kappa_7\}$. The set* $\mathrm{ORBITS}(\kappa_1, \ldots, \kappa_7; m)$ *is the digraph shown at the top of Figure 3. The walk steps corresponding to vertices in this sub-digraph are shown in thick lines. The top graph is the rescaled process $\bar{l}$ which counts cyclic vertices. Shown below is the graph-with-edge-lengths* $\mathbf{g}(u_1, \ldots, u_7; \bar{w}, \bar{l}, \bar{d})$*, with leaf labels* $\boxed{i}$ *and the corresponding leaf-ordered graph* $\mathbf{g}^{\mathrm{ord}}(u_1, \ldots, u_7; \bar{w}, \bar{l}, \bar{d})$*, with leaf labels* (i)*.*



(i) For each excursion of f from 0 which contains at least one of the u_i, construct the associated tree-with-edge-lengths $\mathbf{t}$ as defined above Lemma 8 (with the excursion interval I in place of $[0,1]$ and $\{i : u_i \in I\}$ in place of $[k]$ as the set of leaves). Write $l(\mathbf{t})$ for $l(\cdot)$ evaluated at the starting point $\mathrm{left}(\mathbf{t})$ of the excursion coding $\mathbf{t}$.
(ii) For trees $\mathbf{t}^j, \ldots, \mathbf{t}^J$ such that $\mathrm{left}(\mathbf{t}^j) < \ldots < \mathrm{left}(\mathbf{t}^J)$ and such that these are all the trees $\mathbf{t}$ with $\mathrm{left}(\mathbf{t}) \in [d(r-1), d(r))$ for some r, create a cyclic path between their roots, where the cyclic edge from $\mathrm{root}(\mathbf{t}^i)$ to $\mathrm{root}(\mathbf{t}^{i+1})$ has length $l(\mathbf{t}^{i+1}) - l(\mathbf{t}^i)$ for $j \le i < J$, and the cyclic edge from $\mathrm{root}(\mathbf{t}^J)$ to $\mathrm{root}(\mathbf{t}^j)$ has length $l(d(r)) - l(\mathbf{t}^J) + l(\mathbf{t}^j) - l(d(r-1))$.

Observe that external randomization appears twice in the statement of Theorem 15: in the hypothesis of sampling invariance, and in the conclusion where the mapping-walk is defined using q-biased random order of components. For the proof (but not the statement) it will be important to take these two randomization operations to be independent. So we now introduce i.i.d. uniform $(0,1)$ random variables (U_i), independent of the mappings and mapping-walks in section 4.1, and independent of $B^{|\mathrm{br}|}$ and $(\tilde{U}_i, D_i)$ in section 4.3.

Lemma 16 $\mathbf{g}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D) \stackrel{d}{=} G[k]$, *where $L(t)$ is local time for $B^{|\mathrm{br}|}$ at 0 and $D = (D_0, D_1, \ldots)$ are the marks defined at (22).*

Historically, Lemma 8 was first proved [4] as a consequence of the functional invariance principle for uniform random trees, and later reproved [26, 43] directly via excursion theory for Brownian excursion. Doubtless Lemma 16 could also be proved via excursion theory for Brownian bridge, but we shall deduce it during the course of the next proof from Theorem 14 for uniform random mappings. See [43] for further results in the spirit of Lemmas 8 and 16 for Brownian bridge and meander.

Lemma 17 *Under the hypotheses of Theorem 15,*

$$c\left(\bar{w}(V_1), \inf_{V_1 \le t \le V_2} \bar{w}(t), \bar{w}(V_2), \inf_{V_2 \le t \le V_3} \bar{w}(t), \ldots, \bar{w}(V_k)\right)$$

$$\stackrel{d}{\to} \left(2B^{|\mathrm{br}|}(V_1), \inf_{V_1 \le t \le V_2} 2B^{|\mathrm{br}|}(t), 2B^{|\mathrm{br}|}(V_2), \inf_{V_2 \le t \le V_3} 2B^{|\mathrm{br}|}(t), \ldots, 2B^{|\mathrm{br}|}(V_k)\right)$$

where $(V_1, \ldots, V_k)$ are the order statistics of $(U_1, \ldots, U_k)$.

Proof of Lemmas 16 and 17. First fix n. For $\kappa(u)$ defined at (9), $\kappa(U_1), \ldots, \kappa(U_k)$ are independent random vertices chosen from distribution

q, as required for the sampling invariance principle. Moreover, from their definitions ORBITS$(\kappa(U_1),\ldots,\kappa(U_k);M)$ is the same graph-with-edge-lengths (except for heights of branchpoints differing by 1) as $\mathbf{g}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d})$, where $\bar{w}$ is the rescaled mapping-walk defined in section 4.2, and $\bar{l}$ and $\bar{d}$ are also as defined there. Letting $n\to\infty$, the sampling invariance principle implies

$$c\otimes\mathbf{g}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d}) \stackrel{d}{\to} G[k] \text{ as } n\to\infty. \tag{29}$$

Now in the special case of uniform random mappings, it is not hard to see that the argument in [7] to prove Theorem 14 implies

$$c\otimes\mathbf{g}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d}) \stackrel{d}{\to} \mathbf{g}(U_1,\ldots,U_k;2B^{|\mathrm{br}|},L,D). \tag{30}$$

Thus the right sides of (29) and (30) must be equal in distribution, establishing Lemma 16. Returning to the general case of mappings satisfying the sampling invariance principle, we have now established (30). Now the order statistics $(V_1,\ldots,V_k)$ are related to $(U_1,\ldots,U_k)$ via some permutation $V_i=U_{\pi(i)}$. Construct $\mathbf{g}^{\mathrm{ord}}(U_1,\ldots,U_k;2B^{|\mathrm{br}|},L,D)$ and $\mathbf{g}^{\mathrm{ord}}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d})$ from $\mathbf{g}(U_1,\ldots,U_k;2B^{|\mathrm{br}|},L,D)$ and $\mathbf{g}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d})$ by relabeling leaf $\pi(i)$ as leaf i, to get a "leaf-ordered" graph (illustrated in figure 5 with leaf i labeled (i)). Clearly $\mathbf{g}^{\mathrm{ord}}(U_1,\ldots,U_k;2B^{|\mathrm{br}|},L,D)$ determines the $(2k-1)$-vector appearing on the right side in Lemma 17, and $\mathbf{g}^{\mathrm{ord}}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d})$ determines the $(2k-1)$-vector appearing on the left side. So to prove Lemma 17 it is enough to prove the "ordered" version of (30):

Lemma 18 *As $n\to\infty$*

$$c\otimes\mathbf{g}^{\mathrm{ord}}(U_1,\ldots,U_k;\bar{w},\bar{l},\bar{d}) \stackrel{d}{\to} \mathbf{g}^{\mathrm{ord}}(U_1,\ldots,U_k;2B^{|\mathrm{br}|},L,D).$$

Proof. First fix n. The external randomization used to order components of a mapping M as part of the walk coding can be implemented as follows. Take i.i.d. $(\tilde{\kappa}_i)$ with distribution q. Take C_1 to be the component containing $\tilde{\kappa}_1$ and then take C_r to be the component containing $\tilde{\kappa}_{I_r}$ where

$$I_r=\min\{i:\tilde{\kappa}_i\not\in\cup_{1\le j<r}C_j\}.$$

Fixing K and k, and letting $n\to\infty$, the sampling invariance principle with $K+k$ sampled vertices implies

$$c\otimes\mathrm{ORBITS}(\tilde{\kappa}_1,\ldots,\tilde{\kappa}_K,\kappa(U_1),\ldots,\kappa(U_k);M) \stackrel{d}{\to} G[K+k].$$

But from a realization of ORBITS$(\tilde{\kappa}_1,\ldots,\tilde{\kappa}_K,\kappa(U_1),\ldots,\kappa(U_k);M)$ we can not only derive (exactly) the realization of ORBITS$(\kappa(U_1),\ldots,\kappa(U_k);M)$, but also we get information about the permutation π taking U_i to $V_i = U_{\pi(i)}$. For instance, if a component contains $\tilde{\kappa}_1$ then it is the first component. In fact, the order of tree-components is determined exactly *unless*

$$\exists 1 \leq j \leq k: \ \kappa(U_j) \text{ in different component from each } \tilde{\kappa}_i, \ 1 \leq i \leq K.$$

Moreover the distribution of within-tree leaves is determined by the random depth-first ordering imposed by that aspect of external randomization. So to prove that the limit distribution in Lemma 18 does not depend on the model of random mappings, it is enough to prove

$$\lim_K \lim_n P\left(\exists 1 \leq j \leq k: \ \kappa_j \text{ in different component from each } \tilde{\kappa}_i, \ 1 \leq i \leq K\right) = 0. \tag{31}$$

The probability in question is bounded by

$$kP(\tilde{\kappa}_1 \text{ in different component from each } \tilde{\kappa}_i, \ 2 \leq i \leq K+1). \tag{32}$$

By the sampling invariance principle, the $n \to \infty$ limit of the probability in (32) does not depend on the model of random mapping. Using Theorem 14, the limiting distribution of component weights is $(D_r - D_{r-1}, r \geq 1)$. So the quantity in (31) is bounded by

$$\lim_K \ kE(1 - D_1)^K$$

which clearly equals 0. So we have shown that the limit distribution in Lemma 18 does not depend on the model of random mappings. Use Theorem 14 again to identify the limit distribution as $\mathbf{g}^{\mathrm{ord}}(U_1, \ldots, U_k; 2B^{|\mathrm{br}|}, L, D)$.

Proof of Theorem 15. Having established Lemma 17, the same argument as used in the proof of Proposition 7 shows $c^{(n)}\bar{W}_n \ \stackrel{d}{\to} \ 2B^{|\mathrm{br}|}$. Regarding (27), we will prove

$$(q(C_1), c \cdot \mathrm{cycle}(C_1)) \ \stackrel{d}{\to} \ (D_1, L(D_1)) \tag{33}$$

and the full version is similar. Recalling $(V_i, 1 \leq i \leq k)$ are the order statistics of $(U_i, 1 \leq i \leq k)$, write

$$I_k := \max\{i : \kappa(V_i) \text{ in same component as } \kappa(V_1)\}.$$

Then $V_{I_k} \leq q(C_1) \leq V_{I_k+1}$. Now (30) implies that as $n \to \infty$ for fixed k,

$$(V_{I_k}, V_{I_k+1}) \ \stackrel{d}{\to} \ (V_{J_k}, V_{J_k+1}) \tag{34}$$

where $J_k := \max\{i : V_i \leq D_1\}$. But as $k \to \infty$ we clearly have $(V_{J_k}, V_{J_k+1}) \stackrel{d}{\to} (D_1, D_1)$ and this establishes

$$q(C_1) \stackrel{d}{\to} D_1. \tag{35}$$

Now the sampling invariance principle (for the orbit of a single vertex), and our "q-biased component order" convention, immediately imply that $c{\cdot}\mathrm{cycle}(C_1)$ converges in distribution to a limit which does not depend on the model; so by Theorem 14 the limit distribution is that of $L(D_1)$. Repeating the argument above, convergence (34,35) holds jointly with convergence of $c \cdot \mathrm{cycle}(C_1)$ and the joint limit does not depend on the model; so again the joint limit distribution is that arising in the uniform model, $(D_1, L(D_1))$, and (33) is established.

5.1 Distributional limits implied by the sampling invariance principle

One can immediately use Theorem 15 to see that the sampling invariance principle implies convergence of rescaled cumulative height profiles, as for trees in Corollary 10.

Corollary 19 *Consider a model of random weighted mappings (M, q) on $[n]$ which satisfies the sampling invariance principle with scaling constants c. Define the* rescaled cumulative height profile *by*

$$H_n(u) := \sum_{i=1}^{n} q_i 1_{(\mathrm{height}(i,M) \leq u/c)}.$$

Then

$$H_n \stackrel{d}{\to} H$$

as random elements of $\mathcal{D}$, where

$$H(u) := \int_0^1 1_{(2B_t^{|\mathrm{br}|} \leq u)}\, dt; \quad 0 \leq u < \infty.$$

See [28, 44] for discussion of the explicit distribution of H.

It is true that the sampling invariance principle implies convergence of rescaled tree-sizes, but this cannot be deduced from Theorem 15. Instead we can give a direct proof.

Proposition 20 *Consider a model of random weighted mappings (M, q) on $[n]$ which satisfies the sampling invariance principle with scaling constants c. Let*

$$\Delta_1^n \geq \Delta_2^n \geq \ldots; \quad \sum_i \Delta_i^n = n$$

be the q-measures of the tree-components of the digraph of M^n. Then as $n \to \infty$

$$(\Delta_i^n, i \geq 1) \ \stackrel{d}{\to} \ (\Delta_i^\infty, i \geq 1) \tag{36}$$

where the right side denotes the lengths of excursions from 0 for $B^{|\mathrm{br}|}$, in decreasing order.

See [46] for discussion of the limit distribution.
Proof. Let $(U_j^n, j \geq 1)$ be independent vertices of M^n with distribution q. For each n and k there is a random partition $\mathcal{P}_k^n$ of $[k]$ defined by

j_1 and j_2 are in the some component of the partition if $U_{j_1}^n$ and $U_{j_2}^n$ are in the same tree-component of the digraph of M^n.
By results on exchangeable random partitions going back to Kingman [33], convergence (36) is equivalent to the assertion that, for each fixed k,

$$\mathcal{P}_k^n \ \stackrel{d}{\to} \ \mathcal{P}_k^\infty \text{ as } n \to \infty \tag{37}$$

where $\mathcal{P}_k^\infty$ is the partition defined analogously in terms of i.i.d. uniform $(0, 1)$ random variables and the excursions of $B^{|\mathrm{br}|}$; and where "convergence in distribution" has its elementary interpretation because the number of possible partitions is finite. Now implicit in the arguments of [7] is that (36) and hence (37) holds for the uniform model of random mappings with uniform weights. It is therefore sufficient to show that the $n \to \infty$ limit of $\mathcal{P}_k^n$ is the same in all models satisfying the sampling invariance principle. But this is clear from the definition: $c \otimes \mathrm{ORBITS}(\kappa_1, \ldots, \kappa_k; M) \ \stackrel{d}{\to} \ G[k]$ implies the shapes converge, and the shape determines the partition.

5.2 The functional invariance principle for random mappings

As we have already mentioned, it is not easy to decide exactly how to define the functional invariance principle for random mappings. Here is our tentative definition.

Definition 21 *A model of random weighted mappings (M, q) on $[n]$ satisfies the functional invariance principle with scaling constants $c = c^{(n)}$ if, as $n \to \infty$, the assertions of Theorem 14 hold (with $n^{-1/2}$ replaced by $c^{(n)}$) and the assertion of Proposition 20 holds.*

With this definition it is true – though we omit details – that the functional invariance principle implies the sampling invariance principle.

As with trees in Corollary 11, we can deduce convergence of maximal heights from the functional, but not sampling, invariance principle.

Corollary 22 *Consider a model of random weighted mappings (M, q) on $[n]$ which satisfies the functional invariance principle with scaling constants c. Then*

$$c \cdot \max_i \operatorname{height}(i, M) \ \xrightarrow{d} \ \sup_{0 \le t \le 1} 2B_t^{|\mathrm{br}|}.$$

The limit distribution of $\sup_t B_t^{|\mathrm{br}|}$ here is of course the classical Kolmogorov-Smirnov statistic ([23] equation 7.8.11).

To visualize the distinction between the sampling and functional invariance principles, consider mappings $M^{(n)}$ satisfying the functional invariance principle with respect to uniform weights. Take $m = o(n)$ and make a mapping $\hat{M}^{(m+n)}$ by including an arbitrary mapping $\tilde{M}^{(m)}$ on $\{n+1, \ldots, n+m\}$. Then $\hat{M}^{(m+n)}$ will still satisfy the *sampling* invariance principle, but in general not the functional invariance principle. One can easily make examples of $\tilde{M}^{(m)}$ so that the convergence of maximum heights assertion in Corollary 22 fails, or the convergence of total cycle length (26) fails. Indeed this highlights a subtlety of the joint convergence of component weights and cycle lengths in (27); saying that the cycle lengths of the largest components converge is not saying that the longest cycle lengths converge.

Example 26 later provides a more concrete example.

6 Models for non-uniform random mappings

As already mentioned, the definition of the functional invariance principle for random mappings was motivated by Theorem 14, the case of uniform random mappings. There is no obvious mappings analog of CBP(n) trees, but p-trees and P-trees do have natural mapping analogs. In the next Sections we show that direct analysis of these models can lead to proofs of the sampling invariance principle.

6.1 The sampling invariance principle for p-mappings

Our proof will use Poisson approximation to evaluate asymptotics in the exact distribution of ORBITS($[k]; M$), where $[k] = \{1, 2, \ldots, k\}$. We start by isolating the Poisson approximation result we need.

For a probability distribution p on $[n]$ write

$$c_p := \sqrt{\sum_i p_i^2}.$$

It is well known (see e.g. [18] and papers cited there) that the elementary analysis of the uniform "birthday problem" extends to the non-uniform case under condition (7), as follows.

Lemma 23 *Let $(p^{(n)})$ satisfy (7). Suppose $m = m^{(n)}$ satisfies $mc_p \to s$ for some $0 < s < \infty$. For each n let $(\xi_i, 1 \le i \le m)$ be i.i.d.(p). Then*

$$P(\xi_i, 1 \le i \le m \textit{ are all distinct }) \to \exp(-s^2/2).$$

The precise result we need is the following variant.

Corollary 24 *In the setting of Lemma 23, fix $k \ge 1$. For each n let $i_1, j_1, i_2, j_2, \ldots, i_k, j_k$ be distinct elements of $[m]$. Then*

$$P(\xi_i \notin [k], i \in [m];\ \xi_i \textit{ distinct for all } i \in [m]\backslash\{j_1, \ldots, j_k\};\ \xi_{j_u} = \xi_{i_u}, 1 \le u \le k)$$
$$\sim c_p^{2k} \exp(-s^2/2).$$

Proof. First observe

$$P(\xi_i \notin [k], i \in [m]) \to 1 \tag{38}$$

because the complementary probability is at most $km \max_i p_i = o(mc_p) = o(1)$. Conditioning on the event in (38) is equivalent to replacing $(p^{(n)})$ by conditioned probabilities which satisfy the same hypotheses, so we can reduce the problem to proving

$$P(\xi_i \text{ distinct for all } i \in [m]\backslash\{j_1, \ldots, j_k\};\ \xi_{j_u} = \xi_{i_u}, 1 \le u \le k) \sim c_p^{2k} \exp(-s^2/2). \tag{39}$$

By a similar conditioning argument and Lemma 23,

$$P(\xi_i, i \in [m] \setminus \{j_1, \ldots, j_k\} \text{ are distinct } | \xi_{i_1}, \ldots, \xi_{i_k}) \sim \exp(-s^2/2)$$

uniformly on $\{\xi_{i_1}, \ldots, \xi_{i_u}$ distinct $\}$. From the definition of c_p and (7),

$$P(\xi_{j_u} = \xi_{i_u}, 1 \le u \le k;\ \xi_{i_u} \text{ distinct}, 1 \le u \le k) \sim c_p^{2k}.$$

Combine these two relations to obtain (39).

Theorem 25 (Weak invariance principle for p-mappings) *For a sequence of p-mappings M, if the sequence $(p^{(n)})$ satisfies (7), then as $n \to \infty$*

$$c_p \otimes \mathrm{ORBITS}([k]; M) \stackrel{d}{\to} G[k], \quad k \geq 1.$$

In particular, the sampling invariance principle holds for any weight functions q, with scaling constants c_p.

The proof of the first assertion will be given soon. The second assertion follows from the first because, by relabeling vertices, the first holds for any deterministic choice of k vertices, and hence for the q-random choices in the sampling invariance principle.

Note in particular that Theorem 25 implies that, under (7), the rescaled cumulative height profile converges (Corollary 19). O'Cinneide and Pokrovskii [42] Theorem 2.1 used loosely similar methods to prove convergence (under the same hypotheses) of the rescaled cumulative height process to an unspecified limit. The next example shows that (7) is not enough to imply the functional invariance principle, though we speculate that a very slight strengthening would be enough.

Example 26 *Condition (7) is not sufficient to imply the functional invariance principle for p-mappings.*

Take

$$p_i^{(n,m)} = \tfrac{1}{2n}, 1 \leq i \leq n; \quad p_i^{(n,m)} = \tfrac{1}{2m}, n+1 \leq i \leq n+m.$$

Write $M = M_{n,m}$ for the associated p-mapping on $[n+m]$. Take $m = m(n)$ such that $m/n \to \infty$, and then (7) holds with $c(p^{(n,m)}) \sim \frac{1}{2}n^{-1/2}$. Now fix n. We shall show that, as $m \to \infty$,

$$\max_i \mathrm{height}(i, M) \to \infty \text{ in probability.} \tag{40}$$

By taking $m = m(n) \to \infty$ sufficiently fast, (40) implies that Corollary 22 fails for $M_{n,m}$, and hence the functional invariance principle fails.

Fix n. Choose $L = L(m) \to \infty$ so that

$$\tfrac{m}{2L} 4^{-L} \to \infty.$$

Take $i_1 = m+1$ and define $A_1 = \{i_1, M(i_1), M^{(2)}(i_1), \ldots, M^{(L-1)}(i_1)\}$. Inductively, for $2 \leq r \leq m/(2L)$ let i_r be the minimum element of $\{n+1, \ldots, n+m\} \setminus \cup_{1 \leq s < r} A_s$ and let $A_r = \{i_r, M(i_r), M^{(2)}(i_r), \ldots, M^{(L-1)}(i_r)\}$. Then the conditional probability of the event

A_r consists of L elements, distinct from each other and from $[n] \cup \cup_{1 \leq s < r} A_s$

is at least $(1/4)^L$. So the number $N(m)$ of such events (for some $r \leq m/(2L)$) tends in probability to infinity as $m \to \infty$, by specification of $L(m)$. Now (40) follows easily.

Proof of Theorem 25. Fix $k \geq 1$. Fix $\mathbf{g} \in \mathbf{G}[k]$ with total edge-length $\sum \ell_i = s$. We shall prove Theorem 25 by proving the corresponding, formally stronger, *local* limit theorem

$$\text{if } c_p \otimes \mathbf{g}^{(n)} \to \mathbf{g} \text{ then } P(\text{ORBITS}([k]; M) = \mathbf{g}^{(n)}) \sim c_p^{2k} \exp(-s^2/2). \quad (41)$$

Here $\mathbf{g}^{(n)}$ denotes a possible value of ORBITS$([k]; M)$ which is in $\mathbf{G}[k]$. Note that, although for finite n the graph ORBITS$([k]; M)$ may not be in $\mathbf{G}[k]$ (for example, because it may have degree-4 vertices or non-leaf labeled vertices), combining (41) with the fact that $G[k]$ is a *probability* distribution will imply $P(\text{ORBITS}([k]; M) \in \mathbf{G}[k]) \to 1$ and hence we need only consider the case $\mathbf{g}^{(n)} \in \mathbf{G}[k]$.

Consider the chance that, for a random p-mapping, ORBITS$([7]; M)$ is exactly the graph $\mathbf{g}$ at the top of figure 3. One can construct ORBITS$([7]; M)$ from an i.i.d.(p) sequence $(\xi_j, 1 \leq j \leq n)$ as follows. Declare the iterates of 1, that is $M(1), M^2(1), M^3(1), \ldots$, to be the values $\xi_1, \xi_2, \ldots, \xi_{r_1}$ for $r_1 := \min\{j : \xi_j \in \{1, \xi_1, \ldots, \xi_{j-1}\}\}$. Then, if $2 \notin \{1, \xi_1, \ldots, \xi_{r_1 - 1}\}$, declare the iterates of 2 to be the subsequent ξ-values $\xi_{r_1+1}, \xi_{r_1+2}, \ldots, \xi_{r_2}$ until $r_2 := \min\{j > r_1 : \xi_j \in \{1, 2, \xi_1, \ldots, \xi_{j-1}\}\}$; and so on. So the probability under consideration is exactly the chance that the i.i.d.(p) sequence (ξ_i) starts with a "pattern" of the form

$$\overset{[1]}{\cdot}\ \cdot\cdot a\cdot\cdot\cdot b\cdot\cdot\cdot\cdot\cdot c\cdot\cdot a\ \overset{[2]}{\cdot}\ d\cdot\cdot e\cdot\cdot\cdot\cdot e\ \overset{[3]}{\cdot}\ c\ \overset{[4]}{\cdot}\ \cdot\cdot f\cdot b\ \overset{[5]}{\cdot}\ \cdot g\cdot\cdot g\ \overset{[6]}{f}\ \overset{[7]}{\cdot}\ \cdot\cdot d$$

whose meaning we now explain. The orbit of vertex 1 in figure 3 consists of a path of length 4 attached to a cycle of length 13. To create such an orbit we need $\xi_4 = \xi_{17}$ and we need the other values $\xi_1, \ldots, \xi_3, \xi_5, \ldots, \xi_{16}$ to be distinct and different from the former common value. This makes $r_1 = 17$. Similarly, the orbit of vertex 2 consists of a path of length 5 attached to a cycle of length 5; To create such an orbit we need $\xi_{27} = \xi_{22}$ and we need the other values $\xi_{18}, \ldots, \xi_{21}, \xi_{23}, \ldots, \xi_{26}$ to be distinct from each other and from previous ξ-values. This makes $r_2 = 27$. Vertex 3 is attached by a path of length 2 to a particular point of the cycle from vertex 1; this requires $\xi_{29} = \xi_k$ for a particular k (which turns out to be $k = 14$). And so on. The upshot is that the graph in figure 3 corresponds exactly to the case where

$\xi_1, \ldots, \xi_{46}$ fit the "pattern" shown:
(i) the successive ξ-valuse denoted as $\cdots\cdot$ are distinct except where indicated otherwise; and they are distinct from [7];
(ii) $\xi_{i_u} = \xi_{j_u}$ for each of the 7 pairs (i_u, j_u) whose positions are denoted by symbols $abcdefg$: for instance, symbol a at positions $(4, 17)$ indicates that $\xi_{17} = \xi_4$, and symbol e at positions $(22, 27)$ indicates that $\xi_{22} = \xi_{27}$. In the pattern, superscript $[i]$ denotes position r_{i+1}, that is (because $M(i) = \xi_{r_{i+1}}$) the position of the ξ-value giving $M(i)$.

Applying this analysis to a general possible $\mathbf{g}^{(n)} \in \mathbf{G}[k]$, we see that $P(\mathrm{ORBITS}([k]; M) = \mathbf{g}^{(n)})$ is precisely equal to a probability of the form appearing in Corollary 24, and so that Corollary implies (41).

6.2 Random walk P-mappings

Recall that a P-mapping is a random mapping M on $[n]$ such that $M(1), \ldots, M(n)$ are independent with $P(M(i) = j) = p_{ij}$ for a Markov transition matrix P. One can imagine qualitatively different hypotheses on P which would lead to the sampling invariance principle – note that in this context the stationary distribution of P would be a natural choice of weight function.

Here we shall indicate one possible type of hypothesis. Suppose (for some subsequence of $n \to \infty$) we have size-n groups G_n with group operation denoted by $*$. Suppose each $P = P^{(n)}$ is of the form $p_{ij} = \mu(i^{-1} * j)$ for some probability distribution $\mu = \mu^{(n)}$ on G_n. That is, P is the transition matrix of a random walk $X_m = \xi_1 * \xi_2 * \ldots * \xi_m$ whose steps ξ have distribution μ. Suppose there exist constants $\rho = \rho^{(n)}, t = t^{(n)}$ and γ not depending on n such that the following hold as $n \to \infty$.

(i) $t \to \infty,\ t/\rho \to 0,\ n^{-1/2}\rho \to 0$.

(ii) $n^{1/2} \sum_{i=1}^{\rho} P(X_i = \mathrm{identity}) \to 0$.

(iii) $E(N(\rho) | N(\rho) \geq 1) \to \gamma$, where $N(\rho) := \sum_{i=0}^{\rho} \sum_{j=0}^{\rho} 1_{(X_i = Y_j)}$ for independent random walks (X_i) and (Y_i) with Y_0 independent uniform on G_n.

(iv) $\max_{g \in G_n} n^2 |P(X_t = g) - n^{-1}| \to 0$.

Proposition 27 ([2] Prop. 33) *Under the hypotheses above, as $n \to \infty$*

$$(\gamma n)^{-1/2} \otimes \mathrm{ORBITS}(\mathrm{identity}; M) \ \stackrel{d}{\to}\ G[1].$$

Though we shall not give details, the analysis in [2] can be extended to show

Proposition 28 *Under the hypotheses above,* $(M, \mathrm{uniform})$ *satisfies the sampling invariance principle with constants $(\gamma n)^{-1/2}$.*

As an illustrative example ([2] Example 34) take G_n to be the cyclic group on $\{0, 1, \ldots, n-1\}$ and take $\mu(i) = \frac{1}{2n}$ for $i \neq 1$ and $\mu(1) = \frac{1}{2} + \frac{1}{2n}$. Here one can show the hypotheses hold with $\gamma = 4/3$. Note $P(M(1) = 2) = 1/2 + 1/(2n)$; this "immediate dependence" implies we do not have

$$(\gamma n)^{-1/2} \otimes \mathrm{ORBITS}(1, 2; M) \stackrel{d}{\to} G[2]$$

in contrast to Theorem 25. This explains why we use randomly-chosen vertices κ_i in Definition 12.

7 Remarks

(a) The coding of trees as walks in this paper is via "depth first search" or the "exploration process", used also in e.g. [35]. See [38] for further references and the asymptotic equivalence of variant definitions. Note that a different family of *breadth-first* walks are used for other purposes, e.g. [5].

(b) The literature on asymptotics for p-trees [18, 11, 9, 8] develops a complete theory of all possible limits of p-trees without assumption (7), in which setting the limit tree-with-edge-lengths $T[k]$ will have some different distribution. This is the sense in which condition 7 is analogous to Lindeberg's condition in the central limit theorem. The method of deriving random p-mapping asymptotics from p-trees asymptotics via Joyal's bijection, mentioned in Section 1.1, should lead to a parallel complete description of all possible limits of p-mappings. We plan to investigate this elsewhere [6]. At a technical level, note that Theorem 25 holds for arbitrary weights $q^{(n)}$ whereas the method of [6] seems tied to the choice $q^{(n)} = p^{(n)}$.

(c) Conceptually, one can think of using $B^{|\mathrm{br}|}$ to construct a *continuum random mapping* analogously to the *continuum random trees* constructed from Brownian excursion. This idea also may be developed elsewhere.

(d) It is intriguing, and easy to check, that the distribution of $G[k]$ is (up to scaling constants) the *maximum entropy* distribution on its state space $\mathbf{G}[k]$ subject to the constraint $E(\text{ sum of edge-lengths})^2 = \text{constant}$. The corresponding assertion is not true for $T[k]$ because of the prefactor in (2).

Acknowledgement. We thank Grégory Miermont for a careful reading of a draft.

References

[1] D. J. Aldous and J. Pitman. Two recursive decompositions of Brownian bridge related to the asymptotics of random mappings. Technical Report 595, Dept. Statistics, U.C. Berkeley, 2002.

[2] D.J. Aldous. Self-intersections of random walks on discrete groups. *Math. Proc. Cambridge Philos. Soc.*, 98:155–177, 1985.

[3] D.J. Aldous. A random tree model associated with random graphs. *Random Structures Algorithms*, 1:383–402, 1990.

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

[5] D.J. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25:812–854, 1997.

[6] D.J. Aldous, G. Miermont, and J. Pitman. Limit distributions for non-uniform random mappings, involving Brownian bridge and inhomogeneous continuum random trees. In preparation, 2002.

[7] D.J. Aldous and J. Pitman. Brownian bridge asymptotics for random mappings. *Random Structures Algorithms*, 5:487–512, 1994.

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

[9] D.J. Aldous and J. Pitman. Inhomogeneous continuum random trees and the entrance boundary of the additive coalescent. *Probab. Th. Rel. Fields*, 118:455–482, 2000.

[10] G. Baron, M. Drmota, and L. Mutafchiev. Predecessors in random mappings. *Combin. Probab. Comput.*, 5(4):317–335, 1996.

[11] J. Bennies and J. Pitman. Asymptotics of the Hurwitz binomial distribution related to mixed Poisson Galton-Watson trees. *Combinatorics, Probability and Computing*, 10:203–211, 2001.

[12] S. Berg and L. Mutafchiev. Random mappings with an attracting center: Lagrangian distributions and a regression function. *J. Appl. Probab.*, 27:622 – 636, 1990.

[13] J. Bertoin and J. Pitman. Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math. (2)*, 118:147–166, 1994.

[14] P. Biane, J. Pitman, and M. Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bull. Amer. Math. Soc.*, 38:435–465, 2001.

[15] P. Biane and M. Yor. Valeurs principales associees aux temps locaux Browniens. *Bull. Sci. Math. (2)*, 111:23–101, 1987.

[16] P. Billingsley. *Convergence of Probability Measures.* Wiley, second edition, 1999.

[17] Y. D. Burtin. On a simple formula for random mappings and its applications. *J. Appl. Probab.*, 17:403 – 414, 1980.

[18] M. Camarri and J. Pitman. Limit distributions and random trees derived from the birthday problem with unequal probabilities. *Electron. J. Probab.*, 5:Paper 2, 1–18, 2000.

[19] H. Cremers and D. Kadelka. On weak convergence of integral functions of stochastic processes with applications to processes taking paths in $L(p, E)$. *Stochastic Process. Appl.*, 21:305–317, 1986.

[20] M. Drmota and B. Gittenberger. On the profile of random trees. *Random Structures and Algorithms*, 10:421–451, 1997.

[21] T. Duquesne. A limit theorem for the contour process of conditioned Galton-Watson trees. Unpublished, 2000.

[22] T. Duquesne and J.-F. Le Gall. Random trees, Lévy processes and spatial branching processes. Technical Report DMA-01-23, Ecole Normale Superieure, 2001.

[23] R. Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.

[24] S.N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley, New York, 1986.

[25] A.M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Appl. Math.*, 10:47–56, 1985.

[26] J.-F. Le Gall. The uniform random tree in a Brownian excursion. *Probab. Th. Rel. Fields*, 96:369–384, 1993.

[27] B. Gittenberger and G. Louchard. The Brownian excursion multidimensional local time density. *J. Appl. Probab.*, 36:350–373, 1999.

[28] B. Gittenberger and G. Louchard. On the local time density of the reflecting Brownian bridge. *J. Appl. Math. Stochastic Anal.*, 13:125–136, 2000.

[29] L.S. Grinblat. A limit theorem for measurable random processes and its applications. *Proc.Amer. Math. Soc.*, 61:371–376, 1976.

[30] B. Harris. A survey of the early history of the theory of random mappings. In *Probabilistic methods in discrete mathematics (Petrozavodsk, 1992)*, pages 1–22. VSP, Utrecht, 1993.

[31] K.M. Jansons. The distribution of the time spent by a standard excursion above a given level, with applications to ring polymers near a discontinuity potential. *Elect. Comm. in Probab.*, 2:53–58, 1997.

[32] J.Jaworski. On a random mapping (T, P_j). *J. Appl. Probab.*, 21:186 – 191, 1984.

[33] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc.*, 18:374–380, 1978.

[34] V.F. Kolchin. *Random Mappings.* Optimization Software, New York, 1986. (Translation of Russian original).

[35] J. F. Le Gall and Y. Le Jan. Branching processes in Lévy processes: the exploration process. *Ann. Probab.*, 26:213–252, 1998.

[36] P. Lévy. Sur certains processus stochastiques homogènes. *Compositio Math.*, 7:283–339, 1939.

[37] R. Lyons and Y. Peres. *Probability on Trees and Networks.* Cambridge University Press, 2001. In preparation.

[38] J.-F. Marckert and A. Mokkadem. The depth first processes of Galton-Watson trees converge to the same Brownian excursion. Univ. de Versailles, 2001.

[39] L. R. Mutafchiev. Random mappings and trees: relations between asymptotics of moments of certain characteristics. In *Probabilistic methods in discrete mathematics (Petrozavodsk, 1992)*, pages 353–365. VSP, Utrecht, 1993.

[40] L.R. Mutafchiev. On random mappings with a single attracting centre. *J. Appl. Probab.*, 24:258 – 264, 1987.

[41] L. Mutafciev. Probability distributions and asymptotics for some characteristics of random mappings. In W. Grossman et al., editor, *Proc. 4th Pannonian Symp. on Math. Statist.*, pages 227–238, 1983.

[42] C. A. O'Cinneide and A. V. Pokrovskii. Nonuniform random transformations. *Ann. Appl. Probab.*, 10(4):1151–1181, 2000.

[43] J. Pitman. Brownian motion, bridge, excursion and meander characterized by sampling at independent uniform times. *Electron. J. Probab.*, 4:Paper 11, 1–33, 1999.

[44] J. Pitman. The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest. *Ann. Probab.*, 27:261–283, 1999.

[45] J. Pitman. Random mappings, forests and subsets associated with Abel-Cayley-Hurwitz multinomial expansions. *Séminaire Lotharingien de Combinatoire*, Issue 46:45 pp., 2001. www.mat.univie.ac.at/˜slc/.

[46] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25:855–900, 1997.

[47] S. M. Ross. A random graph. *J. Appl. Probab.*, 18:309–315, 1981.

[48] V.E. Stepanov. Random mappings with a single attracting center. *Theory Probab. Appl.*, 16:155–162, 1971.

A Appendix: Weak convergence on L_0

Recall L_0 is the space of measurable functions $f : [0,1] \to R$ with metric

$$||f_1 - f_2||_0 := \int_0^1 \min(1, |f_1(t) - f_2(t)|)\ dt.$$

Weak convergence of stochastic processes with respect to the L_0 metric has been discussed in several places [19, 29]; in particular it is known that convergence of finite dimensional distributions implies convergence in L_0. We need the following, apparently new, result which characterizes L_0 weak convergence as convergence of *random* finite dimensional distributions.

Proposition 29 *Let X and X_n, $n = 1, 2, \ldots$, be random elements of L_0. For each k let $0 < V_{k,1} < V_{k,2} < \ldots < V_{k,k} < 1$ be the order statistics of k*

independent uniform$(0,1)$ *random variables. Then* $X_n \xrightarrow{d} X$ *in* L_0 *if and only if for each fixed* k,

$$(X_n(V_{k,1}), X_n(V_{k,2}), \ldots, X_n(V_{k,k})) \xrightarrow{d} (X(V_{k,1}), X(V_{k,2}), \ldots, X(V_{k,k})) \text{ as } n \to \infty.$$

The "only if" part is immediate. The proof of the "if" part occupies the rest of this Appendix. Fix B and set

$$F_B := \{f \in L_0 : \ \mathrm{esssup}_t |f(t)| \leq B\}.$$

On F_B the L_0 metric is equivalent to the L_1 metric

$$||f_1 - f_2||_1 := \int_0^1 |f_1(t) - f_2(t)| \ dt.$$

For $m \geq 1$ define $A_m : F_B \to F_B$ by averaging over dyadic rational intervals:

$$(A_m f)(t) = 2^m \int_{i2^{-m}}^{(i+1)2^{-m}} f(u) \ du, \quad i2^{-m} \leq t < (i+1)2^{-m}.$$

The following compactness criterion is straightforward.

Lemma 30 *A sequence* $\{f_n, n \geq 1\} \subset F_B$ *is relatively compact if and only if*

$$\lim_m \limsup_n ||f_n - A_m f_n||_1 = 0.$$

For $f \in F_B$ and $k \geq 1$ define a random element $\tilde{f}^k$ of F_B by

$$\tilde{f}^k(t) = f(V_{k,i}), \quad \tfrac{i-1}{k} \leq t < \tfrac{i}{k}.$$

Because a measurable function can be approximated by functions constant on intervals, it is straightforward to show

Lemma 31 $\lim_k E||\tilde{f}^k - f||_1 = 0, \quad f \in F_B.$

The technical heart of the argument is the following lemma, whose proof we defer.

Lemma 32 *Fix* B *and* m. *Then*

$$||f - A_m f||_1 \leq E||\tilde{f}^k - A_m \tilde{f}^k||_1 + \delta(k, B, m) \quad \forall f \in F_B, \ \forall k \geq 1$$

where the constants $\delta(k, B, m)$ *satisfy* $\lim_k \delta(k, B, m) = 0$.

To prove the Proposition, first truncate at $\pm B$ to reduce to the case where X_n and X take values in F_B; the general case then follows by letting $B \uparrow \infty$. By hypothesis,

$$\widetilde{X}_n^k \stackrel{d}{\to} \widetilde{X}^k \text{ as } n \to \infty, \ k \text{ fixed}.$$

Since A_m is a continuous function, we deduce

$$\lim_n E||\widetilde{X}_n^k - A_m \widetilde{X}_n^k||_1 = E||\widetilde{X}^k - A_m \widetilde{X}^k||_1. \tag{42}$$

Now

$$\begin{aligned}
\limsup_n E||X_n - A_m X_n||_1 &\leq E||\widetilde{X}^k - A_m \widetilde{X}^k||_1 + \delta(k,B,m) \\
&\quad\ \text{by Lemma 32 and (42)} \\
&\leq E||X - A_m X||_1 \\
&\quad\ \text{letting } k \to \infty \text{, using Lemma 31.}
\end{aligned}$$

So

$$\lim_m \limsup_n E||X_n - A_m X_n||_1 \leq \lim_m E||X - A_m X||_1.$$

But the right side equals zero by applying Lemma 30 to X alone; and so Lemma 30 implies that the sequence $\{X_n, n \geq 1\}$ is tight in L_1. To prove convergence it therefore suffices to verify the *identifiability* result

$$\text{if } (X(V_{k,1}), X(V_{k,2}), \ldots, X(V_{k,k})) \stackrel{d}{=} (Y(V_{k,1}), Y(V_{k,2}), \ldots, Y(V_{k,k})), \quad k \geq 1$$

$$\text{then } X \stackrel{d}{=} Y$$

and this is straightforward.

Proof of Lemma 32. The desired inequality can be split into two parts:

$$||f - A_m f||_1 \leq E||\tilde{f}^k - A_m f||_1 + \delta_1(k,B,m) \tag{43}$$

$$E||A_m f - A_m \tilde{f}^k||_1 \leq \delta_2(k,B,m) \tag{44}$$

where we want $\lim_k \delta(k,B,m) = 0$ in each case. We will prove these for $m = 1$, the general case being similar. Note $A_1 f$ is constant on the intervals $[0,1/2)$ and $[1/2,1]$. To study (43), the contribution to $E||\tilde{f}^k - A_1 f||_1$ from the interval $[0,1/2)$ equals

$$E\int_0^{1/2} |\tilde{f}^k(t) - A_1 f(1/4)| \, dt.$$

By a Fubini argument, this equals

$$\int_0^1 |f(s) - A_1 f(1/4)|\ \rho_k(s) ds, \ \text{for } \rho_k(s) ds = k^{-1} \sum_{i=1}^{k/2} P(V_{k,i} \in [s, s+ds]). \tag{45}$$

And we can write the contribution to $||f - A_1 f||_1$ from the interval $[0, 1/2)$ as

$$\int_0^1 |f(s) - A_1 f(1/4)| 1_{(s<1/2)} ds. \tag{46}$$

Using the definition of the $V_{k,i}$ as uniform order statistics one can show $\rho_k(s) = P(V_{k-1,k/2} > s)$ and we lose nothing in assuming k is even. Now observe

$$\text{quantity (46)} - \text{quantity (45)} \ \leq \int_0^{1/2} |f(s) - A_1 f(1/4)|\ P(V_{k-1,k/2} < s) ds.$$

The integrand is bounded by $2B$, and combining with the symmetric contribution from $[1/2, 1]$ we see

$$||f - A_1 f||_1 - E||\tilde{f}^k - A_1 f||_1 \leq 2B\ E|V_{k-1,k/2} - 1/2|.$$

Since clearly $V_{k-1,k/2} \ \stackrel{d}{\to}\ 1/2$ as $k \to \infty$, we have a bound of the required form (43).

To argue (44), the value of $A_1 \tilde{f}^k$ on $[1/2, 1]$ equals

$$\tfrac{2}{k} \sum_{i=1}^{k/2} f(V_{k,i}) := \alpha_k, \ \text{say}.$$

So the contribution to (44) from $[1/2, 1]$ can be bounded by

$$\begin{aligned} & \left| \tfrac{1}{2} E\alpha_k - \int_0^{1/2} f(t) dt \right| + \tfrac{1}{2} E|\alpha_k - E\alpha_k| \\ \leq & \left| \tfrac{1}{2} E\alpha_k - \int_0^{1/2} f(t) dt \right| + \tfrac{1}{2} \sqrt{\text{var } \alpha_k}. \end{aligned} \tag{47}$$

But

$$E\alpha_k = 2 \int_0^1 f(s) \rho_k(s)\ ds \tag{48}$$

for $\rho_k(\cdot)$ as before. So the first term in (47) equals

$$\left| \int_0^1 f(s) (\rho_k(s) - 1_{(s \leq 1/2)})\ ds \right| \leq B \int_0^1 |\rho_k(s) - 1_{(s \leq 1/2)}|\ ds = BE|V_{k-1,k/2} - \tfrac{1}{2}|$$

and this bound is of the required form. For the second term of (47), a brief calculation (whose details we omit) gives

$$E\alpha_k^2 = \tfrac{2}{k}E\alpha_k + \tfrac{4(k-1)}{k}\int_0^1\int_0^1 f(s_1)f(s_2)P(V_{k-2,k/2-1} > \max(s_1,s_2))\ ds_1 ds_2.$$

So using (48),

$$\begin{aligned}
\mathrm{var}\ \alpha_k &= E\alpha_k^2 - (E\alpha_k)^2 \\
&= \tfrac{2}{k}E\alpha_k - \tfrac{4}{k}\int_0^1\int_0^1 f(s_1)f(s_2)P(V_{k-2,k/2-1} > \max(s_1,s_2))\ ds_1 ds_2 \\
+4 &\int_0^1\int_0^1 f(s_1)f(s_2)\Big(P(V_{k-2,k/2-1} > \max(s_1,s_2)) - \rho_k(s_1)\rho_k(s_2)\Big)\ ds_1 ds_2.
\end{aligned}$$

Because $|f| \le B$ we then see

$$\mathrm{var}\ \alpha_k \le \tfrac{2B}{k} + \tfrac{4B^2}{k} + 4B^2\int_0^1\int_0^1 \Big|P(V_{k-2,k/2-1} > \max(s_1,s_2)) - \rho_k(s_1)\rho_k(s_2)\Big|\ ds_1 ds_2.$$

Since $V_{k-2,k/2-1} \stackrel{d}{\to} 1/2$ as $k\to\infty$, this bound $\to 0$ as $k\to\infty$, as required.