

## Metrics on sets of interval partitions with diversity\*

Noah Forman<sup>†</sup>   Soumik Pal<sup>‡</sup>   Douglas Rizzolo<sup>§</sup>   Matthias Winkel<sup>¶</sup>

### Abstract

We first consider interval partitions whose complements are Lebesgue-null and introduce a complete metric that induces the same topology as the Hausdorff distance (between complements). This is done using correspondences between intervals. Further restricting to interval partitions with  $\alpha$ -diversity, we then adjust the metric to incorporate diversities. We show that this second metric space is Lusin. An important feature of this topology is that path-continuity in this topology implies the continuous evolution of diversities. This is important in related work on tree-valued stochastic processes where diversities are branch lengths.

**Keywords:** interval partition; Poisson–Dirichlet distribution;  $\alpha$ -diversity.

**AMS MSC 2010:** Primary 60J25; 60J60; 60J80, Secondary 60G18; 60G52; 60G55.

Submitted to ECP on July 5, 2019, final version accepted on May 10, 2020.

Supersedes arXiv:1907.02132.

## 1 Introduction

We define interval partitions following Aldous [1, Section 17] and Pitman [26, Chapter 4].

**Definition 1.1.** *An interval partition is a set  $\beta$  of disjoint, open subintervals of some interval  $[0, L]$ , that cover  $[0, L]$  up to a Lebesgue-null set. We write  $\|\beta\|$  to denote  $L$ . We refer to the elements of  $\beta$  as its blocks. The Lebesgue measure of a block is called its mass.*

Interval partitions of  $[0, 1]$  appear naturally as representations of discrete distributions. Indeed, we can order the atoms of a discrete distribution and consider intervals whose lengths are the masses of atoms. This is useful e.g. to simulate from discrete distributions. More generally, an interval partition represents a totally ordered and summable collection of real numbers, for example, the interval partition generated naturally by the range of a subordinator (see Pitman and Yor [28]), or the partition of  $[0, 1]$  given by the

---

\*This research is partially supported by NSF grants DMS-1204840, DMS-1308340, DMS-1612483, DMS-1855568, UW-RRF grant A112251, and EPSRC grant EP/K029797/1

<sup>†</sup>Department of Mathematics, McMaster University, Hamilton, ON L8S 4K1, Canada.

E-mail: noahforman@gmail.com

<sup>‡</sup>Department of Mathematics, University of Washington, Seattle, WA 98195, USA.

E-mail: soumikpal@gmail.com

<sup>§</sup>Department of Mathematics, University of Delaware, Newark, DE 19716, USA.

E-mail: drizzolo@udel.edu

<sup>¶</sup>Department of Statistics, University of Oxford, 24–29 St Giles', Oxford OX1 3LB, UK.

E-mail: winkel@stats.ox.ac.uk

complement of the zero-set of a Brownian bridge (Gnedin and Pitman [18, Example 3]). They also arise from the so-called stick-breaking schemes; see [18, Example 2]. Furthermore, interval partitions occur as limits of *compositions* of natural numbers  $n$ , i.e. sequences of positive integers with sum  $n$ . Interval partitions serve as extremal points in paintbox representations of composition structures on  $\mathbb{N}$ ; see Gnedin [16].

The set of all interval partitions is denoted by  $\mathcal{I}_H$ . The subscript  $H$  indicates that this set is typically endowed with a metric  $d_H$  under which the distance between  $\beta$  and  $\gamma$  is the Hausdorff distance between their complements [16, 3]. Then  $(\mathcal{I}_H, d_H)$  is not complete: some Cauchy sequences such as  $\{((i-1)/2^n, i/2^n), 1 \leq i \leq 2^n\} \cup \{(1, 2)\}$ ,  $n \geq 0$ , do not converge in  $(\mathcal{I}_H, d_H)$ , since the complement of the “limiting interval partition”  $\{(1, 2)\}$  is not Lebesgue-null. This causes technical challenges for working with interval-partition-valued stochastic processes, where it is often convenient to work with processes taking values in complete spaces.

Our first aim is to define a complete metric  $d'_H$  on  $\mathcal{I}_H$  that induces the same topology as  $d_H$ , see Section 2. Moreover, we want  $d'_H$  to be well suited to the study of interval-partition-valued processes where the sizes of the subintervals evolve continuously over time and allowing for subintervals to vanish or grow from zero length from points in the complement of the interval partition. Such processes naturally arise when studying population models where the subintervals represent subpopulations. In [13, 14, 15] we apply this paper’s results to study particular examples of such processes that build towards resolving two long-standing conjectures.

Of particular interest in the study of interval partitions are random interval partitions formed by arranging the coordinates of an  $(\alpha, \theta)$ -Poisson–Dirichlet distributed random variable in a regenerative random order. These partitions arise both in the study of random trees and in genetics [21, 25, 27]. One of the important statistics of these partitions is the continuum analogue of the number of parts of an integer composition, called the diversity (see [17, 19]):

**Definition 1.2.** *If  $0 < \alpha < 1$ , we say that an interval partition  $\beta \in \mathcal{I}_H$  of an interval  $[0, L]$  has the  $(\alpha)$ -diversity property, or that  $\beta$  is an interval partition with  $(\alpha)$ -diversity, if the following limit exists for every  $t \in [0, L]$ :*

$$\mathcal{D}_\beta^\alpha(t) := \Gamma(1 - \alpha) \lim_{h \downarrow 0} h^\alpha \#\{(a, b) \in \beta : |b - a| > h, b \leq t\}. \tag{1.1}$$

We denote by  $\mathcal{I}_\alpha \subset \mathcal{I}_H$  the set of interval partitions  $\beta$  that possess the  $\alpha$ -diversity property.  $\alpha$ -diversities are also limits of suitably scaled ranked block masses, see e.g. [26, Lemma 3.11].

In the context of spinal decompositions of random trees the total diversity of an interval partition corresponds to the length of the spine and  $\mathcal{D}_\beta^\alpha(t)$  for  $t \in U \in \beta$  corresponds to the height at which a tree of mass  $\text{Leb}(U)$  branches off from the spine. In the context of genetic models, each block in the interval partition represents the number of individuals in a population with the same genetic type and the total diversity represents the genetic diversity.

In this paper we will fix  $0 < \alpha < 1$  and suppress it from the notation when doing so will not cause confusion. In particular, we will use  $\mathcal{D}_\beta(t)$  in place of  $\mathcal{D}_\beta^\alpha(t)$ . We call  $\mathcal{D}_\beta(t)$  the *diversity* of the interval partition up to  $t \in [0, L]$ . For  $U \in \beta$ ,  $t \in U$ , we write  $\mathcal{D}_\beta(U) = \mathcal{D}_\beta(t)$ , and we write  $\mathcal{D}_\beta(\infty) := \mathcal{D}_\beta(L)$  to denote the *total  $(\alpha)$ -diversity* of  $\beta$ .

When studying evolving population models [10, 31] or evolving random trees [22, 12] with connections to Poisson–Dirichlet distributions, it is natural to ask whether or not the total diversity evolves continuously. This provides challenges because  $\beta \mapsto \mathcal{D}_\beta(\infty)$  is not continuous on  $\mathcal{I}_\alpha$  with respect to the topology induced by  $d_H$ .

The second aim of this paper is to introduce a metric  $d_\alpha$  on  $\mathcal{I}_\alpha$  which generates the same Borel  $\sigma$ -algebra as  $d_H$  and with respect to which the diversity function is continuous.

In fact, we would like  $d_\alpha$  to be such that bead-splitting constructions of random trees as in [27] can be used to map continuously evolving interval partitions to continuously evolving trees. Specifically, we let  $\mathbb{M}$  be the set of (measure-preserving isometry classes of) compact metric measure spaces with the Gromov–Hausdorff–Prokhorov topology. We would like the map  $T: \mathcal{I}_\alpha \rightarrow \mathbb{M}$  defined by  $T(\beta) = ([0, \mathcal{D}_\beta(\infty)], d(\cdot, \cdot), M(\beta))$  to be Lipschitz continuous, where  $d(\cdot, \cdot)$  is the standard metric on  $\mathbb{R}$  and  $M(\beta) = \sum_{U \in \beta} \text{Leb}(U) \delta_{\mathcal{D}_\beta(U)}$  is the Stieltjes measure associated with the right inverse of  $\mathcal{D}_\beta$ . We ask for Lipschitz continuity as this will allow us to transfer Hölder continuity results from interval-partition-valued processes to  $\mathbb{M}$ -valued processes.

The structure of this paper is as follows. We define the metrics on  $\mathcal{I}_H$  and  $\mathcal{I}_\alpha$ , state main results and discuss applications in Section 2. We provide proofs in Section 3.

## 2 Definition of metrics and statement of main results

Fix  $0 < \alpha < 1$ . Our definitions of  $d'_H$  and  $d_\alpha$  are based on the following notion of correspondences between interval partitions, which is motivated by the correspondences that can be used to define the Gromov–Hausdorff metric [9]. We adopt the standard discrete mathematics notation  $[n] := \{1, 2, \dots, n\}$ .

For  $\beta, \gamma \in \mathcal{I}_H$ , a *correspondence* between  $\beta$  and  $\gamma$  is a finite sequence of ordered pairs of intervals  $(U_1, V_1), \dots, (U_n, V_n) \in \beta \times \gamma$ ,  $n \geq 0$ , where the sequences  $(U_j)_{j \in [n]}$  and  $(V_j)_{j \in [n]}$  are each strictly increasing in the left-to-right ordering of the interval partitions.

As in the case of the Gromov–Hausdorff metric, we need the notion of the distortion of a correspondence. Specifically the  $\alpha$ -*distortion* of a correspondence  $(U_j, V_j)_{j \in [n]}$  between  $\beta, \gamma \in \mathcal{I}_\alpha$ , denoted by  $\text{dis}_\alpha(\beta, \gamma, (U_j, V_j)_{j \in [n]})$ , is defined to be the maximum of the following four quantities:

- (i)  $\sum_{j \in [n]} |\text{Leb}(U_j) - \text{Leb}(V_j)| + \|\beta\| - \sum_{j \in [n]} \text{Leb}(U_j)$ ,
- (ii)  $\sum_{j \in [n]} |\text{Leb}(U_j) - \text{Leb}(V_j)| + \|\gamma\| - \sum_{j \in [n]} \text{Leb}(V_j)$ ,
- (iii)  $\sup_{j \in [n]} |\mathcal{D}_\beta(U_j) - \mathcal{D}_\gamma(V_j)|$ ,
- (iv)  $|\mathcal{D}_\beta(\infty) - \mathcal{D}_\gamma(\infty)|$ .

Similarly, the *Hausdorff distortion* of a correspondence  $(U_j, V_j)_{j \in [n]}$  between  $\beta, \gamma \in \mathcal{I}_H$ , denoted by  $\text{dis}_H(\beta, \gamma, (U_j, V_j)_{j \in [n]})$ , is defined to be the maximum of (i)–(ii).

We are now prepared to define  $d'_H$  and  $d_\alpha$ .

**Definition 2.1.** For  $\beta, \gamma \in \mathcal{I}_H$  we define

$$d'_H(\beta, \gamma) := \inf_{n \geq 0, (U_j, V_j)_{j \in [n]}} \text{dis}_H(\beta, \gamma, (U_j, V_j)_{j \in [n]}), \tag{2.1}$$

where the infimum is over all correspondences from  $\beta$  to  $\gamma$ .

For  $\beta, \gamma \in \mathcal{I}_\alpha$  we similarly define

$$d_\alpha(\beta, \gamma) := \inf_{n \geq 0, (U_j, V_j)_{j \in [n]}} \text{dis}_\alpha(\beta, \gamma, (U_j, V_j)_{j \in [n]}). \tag{2.2}$$

This correspondence-distortion approach to defining metrics is ideally suited to studying interval-partition-valued processes where blocks have distinct identities that persist over time.

**Example 2.2.** Consider the following construction. Let  $\beta$  be an interval partition and let  $((\varphi_U(t))_{t \geq 0}, U \in \beta)$  be a family of continuous, non-negative functions such that  $\varphi_U(0) = \text{Leb}(U)$  and  $\sum_{U \in \beta} \max_{t \geq 0} \varphi_U(t) < \infty$ . We can then define the process  $(\beta_t)_{t \geq 0}$  by

$$\beta_t = \{P_U(t) : U \in \beta\} \quad \text{where} \quad P_U(t) = \left( \sum_{V \in \beta: V < U} \varphi_V(t), \sum_{V \in \beta: V \leq U} \varphi_V(t) \right).$$

Heuristically, this is a process in which the size of each block  $U \in \beta$  evolves according to  $\varphi_U$ . This process, for particular distributions on  $\varphi_U$ , occurs in [13, 14, 15] where, in the population interpretation of those papers, it represents the evolution of the types present in the initial population. The models considered in [13, 14, 15] also allow for the creation of new types.

Showing that  $(\beta_t)_{t \geq 0}$  is continuous in  $d'_H$  is very natural; we will show that it is continuous at  $t = 0$ , continuity at other times being similar. Fix  $\epsilon > 0$ . Since  $\sum_{U \in \beta} \max_{t \geq 0} \varphi_U(t) < \infty$ , if  $\eta \subset \beta$  then  $\sum_{U \in \eta} \varphi_U$  is continuous. Since  $\|\beta\| < \infty$ , we can find a finite set  $\beta^\epsilon \subset \beta$  such that  $\sum_{U \in \beta \setminus \beta^\epsilon} \text{Leb}(U) < \epsilon/2$ . By the continuity of  $\sum_{U \in \beta \setminus \beta^\epsilon} \varphi_U$  and  $\varphi_U$  for  $U \in \beta^\epsilon$ , we can find  $\delta > 0$  such that  $t < \delta$  implies that

$$\sum_{U \in \beta^\epsilon} |\text{Leb}(U) - \varphi_U(t)| + \left| \sum_{U \in \beta \setminus \beta^\epsilon} \text{Leb}(U) - \sum_{U \in \beta \setminus \beta^\epsilon} \varphi_U(t) \right| < \epsilon/2. \tag{2.3}$$

Define a correspondence between  $\beta$  and  $\beta_t$  by letting  $U$  correspond to  $P_U(t)$  if  $U \in \beta^\epsilon$  and  $\varphi_U(t) > 0$  and leaving other blocks in  $\beta$  and  $\beta_t$  unmatched. The fact that the quantity in (i) is less than  $\epsilon$  follows directly from Equation (2.3) and that  $\sum_{U \in \beta \setminus \beta^\epsilon} \text{Leb}(U) < \epsilon/2$ . For the quantity in (ii), note that

$$\begin{aligned} & \sum_{U \in \beta^\epsilon: \varphi_U(t) > 0} |\text{Leb}(U) - \varphi_U(t)| + \|\beta_t\| - \sum_{U \in \beta^\epsilon: \varphi_U(t) > 0} \varphi_U(t) \\ & \leq \sum_{U \in \beta^\epsilon} |\text{Leb}(U) - \varphi_U(t)| + \left| \sum_{U \in \beta \setminus \beta^\epsilon} \text{Leb}(U) - \sum_{U \in \beta \setminus \beta^\epsilon} \varphi_U(t) \right| + \sum_{U \in \beta \setminus \beta^\epsilon} \text{Leb}(U) < \epsilon, \end{aligned}$$

which completes the proof. This method of constructing correspondences plays a central role in the continuity arguments of [13, 14].

It is more difficult to give a simple, natural example of a process continuous in  $d_\alpha$ . In [13], we develop a family of examples, similar to the example above, where the creation and disappearance of types are determined by the jumps of a Lévy process. The idea of the  $d_\alpha$ -continuity is that, using the results of [11], the diversity of these processes is given by the local time of the Lévy process and  $d_\alpha$ -continuity follows from the joint continuity of the local time.

We will relate  $d'_H$  to the Hausdorff metric on compact subsets of  $[0, \infty)$ . Specifically, when applied to the complements  $C_\beta := [0, \|\beta\|] \setminus \bigcup_{U \in \beta} U$ , the Hausdorff metric gives rise to a metric

$$d_H(\beta, \gamma) = \inf \{ \epsilon > 0: C_\beta \subseteq C_\gamma^\epsilon \text{ and } C_\gamma \subseteq C_\beta^\epsilon \},$$

on  $\mathcal{I}_H$ , where  $C^\epsilon = \{s \in [0, \infty): \inf_{t \in C} |t - s| \leq \epsilon\}$  is the  $\epsilon$ -thickening of  $C$ .

Our main results are as follows.

**Theorem 2.3.** (a)  $d'_H: \mathcal{I}_H^2 \rightarrow [0, \infty)$  is a metric on  $\mathcal{I}_H$ .

(b)  $d_H$  and  $d'_H$  generate the same separable topology.

(c)  $(\mathcal{I}_H, d'_H)$  is a complete metric space, while  $(\mathcal{I}_H, d_H)$  is not complete.

(d)  $\mathcal{I}_\alpha$  is a Borel subset of  $\mathcal{I}_H$  that is dense in  $\mathcal{I}_H$ .

**Theorem 2.4.** (a)  $d_\alpha: \mathcal{I}_\alpha^2 \rightarrow [0, \infty)$  is a metric on  $\mathcal{I}_\alpha$ .

(b) The topology on  $\mathcal{I}_\alpha$  generated by  $d_\alpha$  is strictly stronger than the subset topology generated by  $d_H$  or  $d'_H$ .

(c) The Borel  $\sigma$ -algebra generated by  $d_\alpha$  equals the one generated by  $d_H$  or  $d'_H$ .

(d)  $(\mathcal{I}_\alpha, d_\alpha)$  is Lusin, i.e. homeomorphic to a Borel subset of a compact metric space.

We prove these results in Section 3. Before we do so, let us note some of the consequences, which motivated us to introduce these metrics, and which also demonstrate some further connections to other metrics on interval partitions and related notions. Denote by  $\mathcal{M}$  the set of compactly supported finite Borel measures on  $[0, \infty)$ , equipped with the Prokhorov metric

$$d_P(\mu, \nu) = \inf\{\varepsilon > 0: \mu(C) \leq \nu(C^\varepsilon) + \varepsilon \text{ for all compact } C \subset [0, \infty)\},$$

and by  $S^\downarrow = \{(x_k)_{k \geq 1}: x_1 \geq x_2 \geq \dots \geq 0 \text{ and } \sum_{k \geq 1} x_k < \infty\}$  the space of summable decreasing sequences equipped with the  $\ell_1$  metric.

**Theorem 2.5.** (a) *The map  $M: \mathcal{I}_\alpha \rightarrow \mathcal{M}$ ,  $M(\beta) = \sum_{U \in \beta} \text{Leb}(U) \delta_{\mathcal{D}_\beta(U)}$  is  $d_\alpha$ -Lipschitz continuous.*

(b) *The diversity map  $\beta \mapsto \mathcal{D}_\beta(\infty)$  is  $d_\alpha$ -Lipschitz continuous on  $\mathcal{I}_\alpha$  but not  $d_H$ -continuous on  $\mathcal{I}_\alpha$ .*

(c) *The map  $\text{RANKED}: \mathcal{I}_\alpha \rightarrow S^\downarrow$ , that associates with  $\beta \in \mathcal{I}_\alpha$  the sequence of decreasing order statistics of  $(\text{Leb}(U), U \in \beta)$ , is continuous.*

The proof of (a) follows easily by comparing the  $d_\alpha$ -metric with the Prokhorov metric. Indeed, if  $d_\alpha(\beta, \gamma) < \varepsilon$ , then there is a correspondence of distortion at most  $\varepsilon$ . By (i) and (ii), this correspondence matches, up to  $\varepsilon$ , all mass of blocks of  $\beta$  and  $\gamma$ , which  $M(\beta)$  and  $M(\gamma)$  place onto  $[0, \infty)$  at locations that, by (iii) are at most  $\varepsilon$  apart. Taking into account (iv), this also entails the  $d_\alpha$ -Lipschitz continuity claimed in (b). The continuity claimed in (c) is elementary.

To see that  $d_H$ -continuity fails in (b), consider any  $\beta \in \mathcal{I}_\alpha$  with continuous  $\mathcal{D}_\beta$  and  $\mathcal{D}_\beta(\infty) > 0$ . Let  $\beta_n$  be the interval partition obtained from  $\beta$  by deleting all but the  $n$  longest intervals. Then  $d_H(\beta_n, \beta) \rightarrow 0$ , but  $\mathcal{D}_{\beta_n}(\infty) = 0$  does not converge to  $\mathcal{D}_\beta(\infty) > 0$ .

Although  $\beta \mapsto \mathcal{D}_\beta(\infty)$  is  $d_\alpha$ -continuous, we remark that the map from an interval partition to its right-continuous diversity process  $\beta \mapsto \mathcal{D}_\beta(\cdot)$  is *not* continuous from  $(\mathcal{I}_\alpha, d_\alpha)$  to Skorokhod space  $(\mathcal{D}, d_{\mathcal{D}})$ . We state and prove a stronger claim in Proposition 3.9.

As noted in the introduction, we can combine (a) and (b) by representing  $\beta$  as a (single-branch) tree  $([0, \mathcal{D}_\beta(\infty)], d, M(\beta))$  in the space  $\mathbb{T} \subseteq \mathbb{M}$  of isometry classes of compact, weighted, R-trees equipped with the Gromov–Hausdorff–Prokhorov metric. Here,  $d$  is the Euclidean metric of  $\mathbb{R}$  restricted to  $[0, \mathcal{D}_\beta(\infty)]$ . With reference to the correspondence definition of this metric in [23, Proposition 6], this can be expressed as follows.

**Corollary 2.6.** *The map  $T: \mathcal{I}_\alpha \rightarrow \mathbb{T}$  defined by  $T(\beta) = ([0, \mathcal{D}_\beta(\infty)], d, M(\beta))$  is  $d_\alpha$ -Lipschitz continuous.*

This entails, in particular, that for  $(\beta_t)_{t \geq 0}$  evolving  $d_\alpha$ -continuously in  $\mathcal{I}_\alpha$ , the associated evolution  $(T(\beta_t))_{t \geq 0}$  in  $\mathbb{T}$  is Gromov–Hausdorff–Prokhorov-continuous. This result when suitably iterated by replacing atoms by further branches (cf. the bead-splitting constructions of [27]) is a key step in our construction of the Aldous diffusion [12] as a  $\mathbb{T}$ -valued diffusion that has Aldous’s Brownian Continuum Random Tree [2] as its stationary distribution.

Further key steps towards this goal are certain  $\mathcal{I}_\alpha$ -valued diffusions [13, 14, 15], which are of independent interest and are related to Petrov’s [24] diffusions on spaces of decreasing sequences by a projection via RANKED onto the ranked sequence of block masses. In connection with Theorem 2.5(b) this entails continuously evolving diversity processes for Petrov’s diffusions, which does not appear to follow from previous constructions [10, 24, 30, 8, 5]. Indeed, other processes have been constructed by directly modelling a continuously evolving diversity process [31].

### 3 Proofs of Theorems 2.3 and 2.4

For the ease of the reader, we will restate all parts of the theorems as propositions/corollaries.

**Proposition 3.1.** *The maps  $d'_H: \mathcal{I}_H^2 \rightarrow [0, \infty)$  and  $d_\alpha: \mathcal{I}_\alpha^2 \rightarrow [0, \infty)$  are metrics.*

*Proof.* Symmetry is built into the definition, and we leave positive-definiteness as an exercise for the reader. We will prove that  $d_\alpha$  satisfies the triangle inequality. The reader will then easily simplify this proof to obtain the triangular inequality for  $d'_H$ .

Suppose that  $d_\alpha(\eta, \beta) = a$  and  $d_\alpha(\beta, \gamma) = b$ . Then

$$|\mathcal{D}_\eta(\infty) - \mathcal{D}_\gamma(\infty)| \leq |\mathcal{D}_\eta(\infty) - \mathcal{D}_\beta(\infty)| + |\mathcal{D}_\beta(\infty) - \mathcal{D}_\gamma(\infty)| \leq a + b. \quad (3.1)$$

Now take  $\epsilon > 0$ . It suffices to show that  $d_\alpha(\eta, \gamma) \leq a + b + 2\epsilon$ .

There exist correspondences  $(U_j, V_j)_{j \in [m]}$  and  $(W_j, X_j)_{j \in [n]}$ , from  $\eta$  to  $\beta$  and from  $\beta$  to  $\gamma$  respectively, with distortions less than  $a + \epsilon$  and  $b + \epsilon$  respectively. We will split these two sequences into two parts each. Let  $(\hat{V}_j)_{j \in [k]} = (\hat{W}_j)_{j \in [k]}$  denote the subsequence of intervals that appear in both  $(V_j)_{j \in [m]}$  and  $(W_j)_{j \in [n]}$ ; note that  $k$  may equal zero, i.e. the overlap may be empty. For each  $j \in [k]$ , let  $\hat{U}_j$  and  $\hat{X}_j$  denote the intervals in  $\eta$  and  $\gamma$  respectively that are paired with  $\hat{V}_j = \hat{W}_j$  in the two correspondences. Then, let  $(\hat{U}_j, \hat{V}_j)_{j \in [m] \setminus [k]}$  denote the remaining terms in the first correspondence not accounted for in the intersection, and let  $(\hat{W}_j, \hat{X}_j)_{j \in [n] \setminus [k]}$  denote the remaining terms in the second correspondence. So overall, the sequences  $(\hat{U}_j, \hat{V}_j)_{j \in [m]}$  and  $(\hat{V}_j, \hat{W}_j)_{j \in [n]}$  are reorderings of the two correspondences.

We will show that the correspondence  $(\hat{U}_j, \hat{X}_j)_{j \in [k]}$  has distortion less than  $a + b + 2\epsilon$ . There are four quantities, listed in Definition 2.1, that we must bound. Quantity (iv) has already been bounded in (3.1). To bound (iii), observe that

$$\sup_{j \in [k]} |(\mathcal{D}_\eta(\hat{U}_j) - \mathcal{D}_\gamma(\hat{X}_j))| \leq \sup_{j \in [k]} \left( |\mathcal{D}_\eta(\hat{U}_j) - \mathcal{D}_\beta(\hat{V}_j)| + |\mathcal{D}_\beta(\hat{W}_j) - \mathcal{D}_\gamma(\hat{X}_j)| \right) < a + b + 2\epsilon.$$

We now go about bounding (i), which is more involved. By the triangle inequality,

$$\begin{aligned} & \sum_{j \in [k]} |\text{Leb}(\hat{U}_j) - \text{Leb}(\hat{X}_j)| + \|\eta\| - \sum_{j \in [k]} \text{Leb}(\hat{U}_j) \\ & \leq \sum_{j \in [m]} |\text{Leb}(U_j) - \text{Leb}(V_j)| + \sum_{j \in [n]} |\text{Leb}(W_j) - \text{Leb}(X_j)| \\ & \quad - \sum_{j \in [m] \setminus [k]} |\text{Leb}(\hat{U}_j) - \text{Leb}(\hat{V}_j)| + \left( \|\eta\| - \sum_{j \in [m]} \text{Leb}(\hat{U}_j) \right) + \sum_{j \in [m] \setminus [k]} \text{Leb}(\hat{U}_j). \end{aligned}$$

Since the  $(\hat{V}_j)_{j \in [m] \setminus [k]}$  are members of  $\beta$  not listed in  $(W_j)_{j \in [n]}$ ,

$$\sum_{j=k+1}^m \text{Leb}(\hat{U}_j) - \sum_{j=k+1}^m |\text{Leb}(\hat{U}_j) - \text{Leb}(\hat{V}_j)| \leq \sum_{j=k+1}^m \text{Leb}(\hat{V}_j) \leq \|\beta\| - \sum_{j=1}^n \text{Leb}(W_j),$$

again by the triangle inequality. Thus,

$$\begin{aligned} & \sum_{j \in [k]} |\text{Leb}(\hat{U}_j) - \text{Leb}(\hat{X}_j)| + \|\eta\| - \sum_{j \in [k]} \text{Leb}(\hat{U}_j) \\ & \leq \sum_{j \in [m]} |\text{Leb}(U_j) - \text{Leb}(V_j)| + \|\eta\| - \sum_{j \in [m]} \text{Leb}(U_j) + \sum_{j \in [n]} |\text{Leb}(W_j) - \text{Leb}(X_j)| + \|\beta\| - \sum_{j \in [n]} \text{Leb}(W_j) \\ & \leq a + \epsilon + b + \epsilon. \end{aligned}$$

This is the desired bound on quantity (i) in Definition 2.1. The same argument bounds (ii):

$$\sum_{j \in [k]} |\text{Leb}(\hat{U}_j) - \text{Leb}(\hat{X}_j)| + \|\gamma\| - \sum_{j \in [k]} \text{Leb}(\hat{X}_j) \leq a + b + 2\epsilon.$$

Therefore  $d_\alpha(\eta, \gamma) \leq a + b + 2\epsilon$ , as desired. □

There are some natural operations for interval partitions:

**Definition 3.2.** We define a scaling map  $\odot: (0, \infty) \times \mathcal{I}_H \rightarrow \mathcal{I}_H$  by setting, for  $c > 0$  and  $\beta \in \mathcal{I}_H$

$$c \odot \beta = \{(ca, cb) : (a, b) \in \beta\}.$$

Let  $(\beta_a)_{a \in \mathcal{A}}$  denote a family of interval partitions indexed by a totally ordered set  $(\mathcal{A}, \preceq)$ . For the purpose of this definition, let  $S_\beta(a-) := \sum_{b \prec a} \|\beta_b\|$  for  $a \in \mathcal{A}$ . If  $S_\beta(a-) < \infty$  for every  $a \in \mathcal{A}$ , then we define the concatenation

$$\star_{a \in \mathcal{A}} \beta_a := \{(x + S_\beta(a-), y + S_\beta(a-)) : a \in \mathcal{A}, (x, y) \in \beta_a\}. \tag{3.2}$$

When  $\mathcal{A} = \{a_1, a_2\}$ , we denote this by  $\beta_{a_1} \star \beta_{a_2}$ . We call  $(\beta_a)_{a \in \mathcal{A}}$  summable if  $\sum_{a \in \mathcal{A}} \|\beta_a\| < \infty$ . It is then strongly summable if the concatenated partition satisfies the diversity property (1.1).

It will be useful to separate the diversity of a partition from most of its mass in the following sense. For  $\eta \in \mathcal{I}_\alpha$  and  $\epsilon > 0$ , let

$$\delta(\eta, \epsilon) := \sup \left\{ m > 0 : \sum_{U \in \eta} \mathbf{1}\{\text{Leb}(U) < m\} \text{Leb}(U) < \epsilon \right\} \tag{3.3}$$

For the purpose of the following, let  $A := \{U \in \eta : \text{Leb}(U) \geq \delta(\eta, \epsilon)\}$ ,

$$S_A(x) := \sum_{(a,b) \in A} (b-a) \mathbf{1}\{b \leq x\}, \quad \text{and} \quad S_{\eta \setminus A}(x) := \sum_{(a,b) \in \eta \setminus A} (b-a) \mathbf{1}\{b \leq x\} \quad \text{for } x \geq 0.$$

We define

$$\begin{aligned} \eta_\epsilon^\mathcal{D} &:= \{(a - S_A(a), b - S_A(a)) : (a, b) \in \eta \setminus A\} \\ \text{and } \eta_\epsilon^L &:= \{(a - S_{\eta \setminus A}(a), b - S_{\eta \setminus A}(a)) : (a, b) \in A\}. \end{aligned} \tag{3.4}$$

Effectively, we form  $\eta_\epsilon^L$  by taking the large blocks of  $\eta$  and sliding them down to sit next to each other, and correspondingly for  $\eta_\epsilon^\mathcal{D}$  with the small blocks. These partitions have the properties

$$\mathcal{D}_{\eta_\epsilon^\mathcal{D}}(\infty) = \mathcal{D}_\eta(\infty), \quad \|\eta_\epsilon^\mathcal{D}\| \leq \epsilon, \quad \mathcal{D}_{\eta_\epsilon^L}(\infty) = 0, \quad \|\eta_\epsilon^L\| \geq \|\eta\| - \epsilon.$$

We note the following easy lemma.

**Lemma 3.3.** For  $c > 0$ , the scaling functions  $\beta \mapsto c \odot \beta$  are bijections on  $\mathcal{I}_\alpha$  and on  $\mathcal{I}_H$ ; specifically, partitions in the image of  $\mathcal{I}_\alpha$  possess the diversity property with

$$\mathcal{D}_{c \odot \beta}(ct) = c^\alpha \mathcal{D}_\beta(t) \quad \text{for } \beta \in \mathcal{I}_\alpha, t > 0, c > 0. \tag{3.5}$$

Moreover, for  $\beta, \gamma \in \mathcal{I}_\alpha$ ,

$$d'_H(\beta, c \odot \beta) = |c - 1| \|\beta\|, \quad d'_H(c \odot \beta, c \odot \gamma) = cd'_H(\beta, \gamma), \tag{3.6}$$

$$d_\alpha(\beta, c \odot \beta) \leq \max\{|c^\alpha - 1| \mathcal{D}_\beta(\infty), |c - 1| \|\beta\|\}, \tag{3.7}$$

$$\text{and } \min\{c, c^\alpha\} d_\alpha(\beta, \gamma) \leq d_\alpha(c \odot \beta, c \odot \gamma) \leq \max\{c, c^\alpha\} d_\alpha(\beta, \gamma). \tag{3.8}$$

**Proposition 3.4** ( $d_H$  is equivalent to  $d'_H$  and weaker than  $d_\alpha$ ). (i) For every  $\epsilon > 0$ , there exist some  $\beta, \gamma \in \mathcal{I}_\alpha$  for which  $d_H(\beta, \gamma) < \epsilon$  and  $d_\alpha(\beta, \gamma) > 1/\epsilon$ .

(ii) For  $\beta, \gamma \in \mathcal{I}_\alpha$ , we have  $d'_H(\beta, \gamma) \leq d_\alpha(\beta, \gamma)$ .

(iii) The metrics  $d_H$  and  $d'_H$  generate the same topology on  $\mathcal{I}_H$ .

The related claim that each of  $d_\alpha$ ,  $d_H$ , and  $d'_H$  generates the same Borel  $\sigma$ -algebra on  $\mathcal{I}_\alpha$  will be proved at the end of this paper, in Proposition 3.16.

*Proof.* (i) Fix  $\epsilon > 0$ . Consider an arbitrary  $\eta \in \mathcal{I}_\alpha$  with  $\mathcal{D}_\eta(\infty) > 1/\epsilon$ . The pair  $(\eta, \eta_{\epsilon/2}^L)$  defined in (3.4) has the desired property.

(ii) This is immediate from Definition 2.1 of  $d'_H$ .

(iii) First, we show  $d_H(\beta, \gamma) \leq 3d'_H(\beta, \gamma)$  for every  $\beta, \gamma \in \mathcal{I}_H$ . Suppose  $d'_H(\beta, \gamma) < x$  for some  $x > 0$ . Then there is some correspondence  $(U_i, V_i)_{i \in [n]}$  from  $\beta$  to  $\gamma$  with Hausdorff distortion less than  $x$ . Recall from before Definition 2.1 that, in a correspondence, the  $(U_i)$  and  $(V_i)$  are each listed in left-to-right order. Let

$$\beta' := \star_{i \in [n]} \{(0, \text{Leb}(U_i))\}, \quad \gamma' := \star_{i \in [n]} \{(0, \text{Leb}(V_i))\}.$$

By definition of Hausdorff distortion before Definition 2.1,  $\|\beta\| - \|\beta'\| < x$ , and likewise for  $\gamma$  and  $\gamma'$ . Thus, for each  $j \in [n - 1]$ , the right endpoint of  $U_j$  and the left endpoint of  $U_{j+1}$  are within distance  $x$  of the corresponding point in  $\beta'$ , and similarly for the left endpoint of  $U_1$  and the right endpoint of  $U_n$ . Thus,  $d_H(\beta, \beta') < x$  and correspondingly for  $\gamma$ . By definition of distortion, we also find  $d_H(\beta', \gamma') < x$ . By the triangle inequality,  $d_H(\beta, \gamma) < 3x$ , as desired.

Now let  $\beta \in \mathcal{I}_H$  and  $\epsilon > 0$ . Take  $\delta_0 > 0$  small enough that  $\sum_{U \in \beta: \text{Leb}(U) \leq 2\delta_0} \text{Leb}(U) < \epsilon/3$ . Let  $K$  denote the number of blocks in  $\beta$  with mass at least  $2\delta_0$ . Take  $\delta := \min\{\delta_0, \epsilon/(6K+3)\}$ . It suffices to show that for  $\gamma \in \mathcal{I}_H$ , if  $d_H(\beta, \gamma) < \delta$  then  $d'_H(\beta, \gamma) < \epsilon$ .

Suppose  $d_H(\beta, \gamma) < \delta$  for some  $\gamma \in \mathcal{I}_H$ . Then for each  $U \in \beta$  with  $\text{Leb}(U) > 2\delta_0 \geq 2\delta$ , the midpoint of  $U$  must lie within some block  $V$  of  $\gamma$ . Consider the correspondence from  $\beta$  to  $\gamma$  that matches each such  $(U, V)$ . Then, by the bound on  $d_H(\beta, \gamma)$ , for each such pair,  $|\text{Leb}(U) - \text{Leb}(V)| < 2\delta \leq \epsilon/3K$ . Moreover, by our choice of  $\delta_0$ , the total mass in  $\beta$  excluded from the blocks in the correspondence is at most  $\epsilon/3$ . Similarly, the reader may confirm that the mass in  $\gamma$  excluded from the correspondence is at most  $(\epsilon/3) + 2K\delta + \delta \leq 2\epsilon/3$ . Thus, by Definition 2.1 of  $d'_H$ , we have  $d'_H(\beta, \gamma) < \epsilon$ , as desired.  $\square$

**Lemma 3.5.**  $(\mathcal{I}_\alpha, d_\alpha)$  is path-connected and separable.

*Proof.* For path-connectedness, just note that  $c \mapsto c \odot \eta$ ,  $c \in [0, 1]$ , is a path from  $\emptyset \in \mathcal{I}_\alpha$  to  $\eta \in \mathcal{I}_\alpha$ . Specifically, continuity holds since Lemma 3.3 yields for  $0 \leq a < b \leq 1$

$$d_\alpha(a \odot \eta, b \odot \eta) = d_\alpha\left(\frac{a}{b} \odot b \odot \eta, b \odot \eta\right) \leq \max\{|b^\alpha - a^\alpha| \mathcal{D}_\eta(\infty), |b - a| \|\eta\|\}.$$

For separability, we fix a partition  $\eta \in \mathcal{I}_\alpha$  with  $\mathcal{D}_\eta(\infty) > 0$  and such that  $t \mapsto \mathcal{D}_\eta(t)$  is continuous on  $[0, \|\eta\|]$ . For the purpose of this proof we abbreviate our scaling notation from  $c \odot \eta$  to  $c\eta$ . We will construct a countable  $S \subset \mathcal{I}_\alpha$  in which each element is formed by taking  $(c\eta)_\epsilon^\mathcal{D}$ , as in (3.4), for some  $c \geq 0$  and  $\epsilon > 0$ , and inserting finitely many large blocks into the middle, via the following operation. For  $s \in [0, \mathcal{D}_\eta(\infty)]$  and  $m > 0$ , we define

$$\eta \oplus_s m := \left(\{U \in \eta: \mathcal{D}_\eta(U) \leq s\} \star \{(0, m)\}\right) \cup \{(a + m, b + m): (a, b) \in \eta, \mathcal{D}_\eta(a) > s\}.$$



This operation inserts a new interval  $V$  of length  $m$  into the middle of  $\eta$  in such a way that  $\mathcal{D}_{\eta \oplus_s m}(V) = s$ . Let

$$S := \left\{ (c\eta)_{\epsilon}^{\mathcal{D}} \oplus_{s_1} m_1 \cdots \oplus_{s_r} m_r \mid \begin{array}{l} r \in \mathbb{N}, s_1, \dots, s_r \in [0, \mathcal{D}_{c\eta}(\infty)] \cap \mathbb{Q}, \\ c, \epsilon, m_1, \dots, m_r \in (0, \infty) \cap \mathbb{Q} \end{array} \right\}.$$

By Lemma 3.3,  $\mathcal{D}_{c\eta}(\infty) = c^\alpha \mathcal{D}_\eta(\infty)$  for  $c \geq 0$ . Thus, any  $\beta \in \mathcal{I}_\alpha$  can be approximated in  $S$  by the partitions constructed from the following rational sequences. First, take rational

$$c_n \rightarrow \left( \frac{\mathcal{D}_\beta(\infty)}{\mathcal{D}_\eta(\infty)} \right)^{1/\alpha}, \quad \epsilon_n = \frac{1}{n} \downarrow 0, \quad \text{and} \quad r_n = \#\beta_{\epsilon_n}^L.$$

Then let  $\{U \in \beta : \text{Leb}(U) > \delta(\beta, \epsilon_n)\} = \left\{ \left( a_j^{(n)}, a_j^{(n)} + k_j^{(n)} \right), j \in [r_n] \right\}$  with  $a_1^{(n)} \leq \dots \leq a_{r_n}^{(n)}$ , where  $\delta$  is as in (3.3). This is the sequence of blocks of  $\beta$  that comprise  $\beta_{\epsilon_n}^L$ . Finally, we take rational sequences  $\left( \left( s_j^{(n)}, m_j^{(n)} \right), j \in [r_n] \right)$  so that

$$\sup_{j \in [r_n]} \left| s_j^{(n)} - \mathcal{D}_\beta \left( a_j^{(n)} \right) \right| \leq \epsilon_n \quad \text{and} \quad \sum_{j \in [r_n]} \left| k_j^{(n)} - m_j^{(n)} \right| \leq \epsilon_n. \quad \square$$

**Corollary 3.6.** *There is a metric on  $\mathcal{I}_\alpha$  that generates the same topology as  $d_\alpha$ , for which  $\mathcal{I}_\alpha$  is isometric to a subset of a compact metric space.*

*Proof.* Since  $(\mathcal{I}_\alpha, d_\alpha)$  is a separable metric space, Dudley’s [7, Theorem 2.8.2] applies.  $\square$

Unfortunately, this argument is unsuitable to show that the subset can be chosen as a Borel subset. Indeed, the argument can be applied to non-Borel subsets of a compact metric space. To prove this, we introduce a larger metric space  $(\mathcal{J}, d_{\mathcal{J}})$ , on pairs  $(\eta, f)$ , where  $\eta \in \mathcal{I}_H$  is an interval partition and where  $f$  is a right-continuous increasing function that is not necessarily  $f = \mathcal{D}_\eta(\cdot+)$ , which may not even exist, but which shares the property of  $\mathcal{D}$  to be constant on intervals  $U \in \eta$ . Then  $(\beta, \mathcal{D}_\beta(\cdot+)) \in \mathcal{J}$  for all  $\beta \in \mathcal{I}_\alpha$ .

The reader may wonder why we take the process of right limits  $\mathcal{D}_\beta(\cdot+)$  associated with  $\mathcal{D}_\beta$ . First note that, in general,  $\mathcal{D}_\beta$  may be neither left- nor right-continuous. E.g., take any interval partition  $\beta$  with positive diversity  $D = \mathcal{D}_\beta(\infty)$  and reorder the blocks in ranked order of mass. Then the resulting interval partition has zero diversity function, jumping to  $D$  at  $\|\beta\|$ . If we instead arrange intervals of even rank from the left and of odd rank from the right, accumulating in the “middle”, at  $t$ , say, then the diversity function of the resulting interval partition is constant 0 on  $(0, t)$ , constant  $D$  on  $(t, \infty)$  and  $D/2$  at  $t$ .

We use right-continuous functions in  $\mathcal{J}$  to be definite. We actually only care about the values that  $f$  takes on the intervals of constancy, but we prefer to work with representatives in a familiar class of functions.

**Definition 3.7.** *Let  $\mathcal{J}$  be the set of pairs  $(\eta, f)$ , where  $\eta$  is an interval partition of  $[0, \|\eta\|]$  with  $\text{Leb}([0, \|\eta\|] \setminus \bigcup_{U \in \eta} U) = 0$ , and where  $f: [0, \infty) \rightarrow [0, \infty)$  is a right-continuous increasing function that is constant on every interval  $U \in \eta$  and on  $[\|\eta\|, \infty)$ . We replace  $\mathcal{D}_\eta$  and  $\mathcal{D}_\beta$  in Definition 2.1, the definition of  $d_\alpha(\eta, \beta)$ , by  $f$  and  $g$ , to define  $d_{\mathcal{J}}((\eta, f), (\beta, g))$ .*

Recall the Skorokhod metric of [4, equations (14.12), (14.13)]; we denote this by  $d_{\mathcal{D}}$ . For  $n \geq 1$ , let  $\mathcal{J}_n \subseteq \mathcal{J}$  denote the set of  $(\beta, f) \in \mathcal{J}$  for which  $\beta$  has exactly  $n$  blocks. For  $n \geq 1$  and  $\beta \in \mathcal{I}_H$ , let  $\beta_n$  denote the interval partition formed by deleting all but the  $n$  largest blocks from  $\beta$  (breaking ties via left-to-right order) and sliding these large blocks together, as in the construction of  $\eta_\epsilon^L$  in (3.4). For  $(\beta, f) \in \mathcal{J}$ , equip  $\beta_n$  with the function  $f_n$  that is constant on each block of  $\beta_n$  with the value that  $f$  takes on the corresponding block of  $\beta$ .

**Lemma 3.8.** (i) *The distance function  $d_{\mathcal{J}}$  is a metric on  $\mathcal{J}$ .*

(ii) *For  $n \geq 1$ , the metric  $d_{\mathcal{J}}$  on  $\mathcal{J}_n$  is topologically equivalent to the maximum of  $d'_H$  in the first coordinate and  $d_{\mathcal{D}}$  in the second.*

(iii) *The maps  $\beta \mapsto \beta_n$  and  $(\beta, f) \mapsto (\beta_n, f_n)$  are Borel under  $d'_H$  and  $d_{\mathcal{J}}$  respectively.*

(iv) *The map  $(\eta, f) \mapsto f$  is Borel from  $(\mathcal{J}, d_{\mathcal{J}})$  to  $(\mathcal{D}, d_{\mathcal{D}})$ .*

*Proof.* (i) Given the proof of Proposition 3.1, the only change needed for this part of the lemma is in proving positive-definiteness, since now  $f$  is not determined by  $\eta$ . However, this follows easily since we assume that  $f$  is right-continuous and constant on each  $U \in \eta$  and on  $[\|\eta\|, \infty]$ , and  $f$  is therefore determined by the values it takes on these sets.

(ii) Fix  $(\beta, f) \in \mathcal{J}_n$ . We denote the blocks of  $\beta$  by  $U_1, \dots, U_n$ , in left-to-right order. Take  $r \in (0, \min_{j \in [n]} \text{Leb}(U_j))$ . We will show that, for  $(\gamma, g) \in \mathcal{J}_n$ , we get  $d_{\mathcal{J}}((\beta, f), (\gamma, g)) < r$  if and only if both  $d'_H(\beta, \gamma) < r$  and  $d_{\mathcal{D}}(f, g) < r$ .

Consider  $(\gamma, g) \in \mathcal{J}_n$  with  $d'_H(\beta, \gamma) < r$  and  $d_{\mathcal{D}}(f, g) < r$ . Since we have required  $r$  to be smaller than all block masses in  $\beta$ , the only correspondence from  $\beta$  to  $\gamma$  that can have Hausdorff distortion less than  $r$  is  $(U_i, V_i)_{i \in [n]}$ , where  $V_1, \dots, V_n$  denote the blocks of  $\gamma$  in left-to-right order. In particular,  $\sum_{i \in [n]} |\text{Leb}(V_i) - \text{Leb}(U_i)| < r$ . Thus, in order for a continuous time-change  $\lambda: [0, \|\beta\|] \rightarrow [0, \|\gamma\|]$  to never deviate from the identity by  $r$ , it must map some time in each  $U_i$  to a time in the corresponding  $V_i$ . Therefore, by our bound on  $d_{\mathcal{D}}$ , we have  $\max_{i \in [n]} |g(V_i) - f(U_i)| < r$ . We conclude that  $d_{\mathcal{J}}((\beta, f), (\gamma, g)) < r$ .

Now, consider  $(\gamma, g) \in \mathcal{J}_n$  with  $d_{\mathcal{J}}((\beta, f), (\gamma, g)) < r$ . Following our earlier notation, the only correspondence that can give distortion less than  $r$  is  $(U_i, V_i)_{i \in [n]}$ . It follows immediately from Definitions 2.1 and 3.7 of  $d'_H$  and  $d_{\mathcal{J}}$  that  $d'_H(\beta, \gamma) \leq d_{\mathcal{J}}((\beta, f), (\gamma, g)) < r$ . We define  $\lambda: [0, \|\beta\|] \rightarrow [0, \|\gamma\|]$  by mapping the left and right endpoints of each  $U_j$  to the corresponding left and right endpoints of  $V_j$  and interpolating linearly. Since  $\sum_{i \in [n]} |\text{Leb}(V_i) - \text{Leb}(U_i)| < r$ , it follows that  $|\lambda(t) - t| < r$  for  $t \in [0, \|\beta\|]$  as well. By definition of  $d_{\mathcal{J}}$ , we have  $|g(V_i) - f(U_i)| < r$  for each  $i \in [n]$ . Thus,  $|g(\lambda(t)) - f(t)| < r$  for  $t \in [0, \|\beta\|]$ . This gives  $d_{\mathcal{D}}(f, g) < r$ .

(iii) The map RANKED that sends  $\beta \in \mathcal{I}_H$  to the vector of its order statistics is continuous under  $d'_H$ . The restriction map  $(\beta, t) \mapsto \beta|_{[0, t]} := \{U \cap (0, t) : U \in \beta, U \cap (0, t) \neq \emptyset\}$  is continuous from  $d'_H$  plus the Euclidean metric to  $d'_H$ . If  $\text{RANKED}(\beta) = (x_1, x_2, \dots)$ , then we determine whether the block of mass  $x_1$  is to the right of the block of mass  $x_2$  by finding the least  $t_1, t_2 \in x_2\mathbb{N}$  for which  $\beta|_{[0, t_1]}$  has  $x_1$  as its first order statistic and  $\beta|_{[0, t_2]}$  has  $(x_1, x_2)$  as its first two order statistics. If  $t_1 < t_2$  then  $\beta_2 = \{(0, x_1), (x_1, x_1 + x_2)\}$ ; otherwise,  $\beta_2 = \{(0, x_2), (x_2, x_2 + x_1)\}$ . This method extends to give the measurability of  $\beta \mapsto \beta_n$ .

Now let  $y_1(\beta, f) := f(U_1)$ , where  $U_1 \in \beta$  is the longest interval (the left-most of these, if there are ties). Then  $\{(\gamma, g) \in \mathcal{J} : y_1(\gamma, g) > z\}$  is open in  $(\mathcal{J}, d_{\mathcal{J}})$ . This extends to show the measurability of the functions  $y_{n,k}: \mathcal{J} \rightarrow [0, \infty)$ ,  $1 \leq k \leq n$ , that assign to  $(\beta, f)$  the values  $y_{n,k}(\beta, f)$  of  $f$  on the  $n$  longest intervals of  $\beta$ , in left-to-right order. This allows to measurably construct  $f_n$  from  $(\beta, f)$ , which entails the measurability of  $(\beta, f) \mapsto (\beta_n, f_n)$ .

(iv) As right-continuity of  $f$  yields  $\lim_{n \uparrow \infty} f_n(t) = f(t)$ , the measurability of  $(\eta, f) \mapsto f(t)$  for each  $t \in [0, \infty)$  follows from (iii). By [4, Theorem 12.5], the Borel  $\sigma$ -algebra on  $(\mathcal{D}, d_{\mathcal{D}})$  is generated by the evaluation maps, so the claimed measurability of  $(\eta, f) \mapsto f$  follows.  $\square$

We now digress from proving Theorems 2.3 and 2.4 to pursue an idea inspired by Lemma 3.8(ii). Consider the metric  $d'_H(\beta, \gamma) + d_{\mathcal{D}}(\mathcal{D}_{\beta}(\cdot +), \mathcal{D}_{\gamma}(\cdot +))$  as an alternative to  $d_{\alpha}$ , where  $d_{\mathcal{D}}$  is the Skorokhod metric. The following proposition shows that this would not be metrically equivalent to  $d_{\alpha}$ , and our counterexample illustrates why we prefer Definition 2.1: we prefer a metric where small blocks only contribute to diversity between large blocks.

**Proposition 3.9.** *For every  $\epsilon > 0$  there exist interval partitions  $\beta, \gamma \in \mathcal{I}_\alpha$  with  $d_\alpha(\beta, \gamma) \leq \epsilon$  and  $d_{\mathcal{D}}(\mathcal{D}_\beta(\cdot+), \mathcal{D}_\gamma(\cdot+)) \geq 1/\epsilon$ .*

*Proof.* Fix  $\epsilon > 0$ . Consider  $\eta \in \mathcal{I}_\alpha$  with the property that  $\mathcal{D}_\eta(t)$  equals 0 if  $t < \|\eta\|$ , and equals  $1/\epsilon$  if  $t \geq \|\eta\|$ . As noted above Definition 3.7,  $\eta$  can be formed by taking any interval partition with total diversity  $1/\epsilon$  and rearranging its block masses so that they appear in non-increasing order from left to right. Recall (3.4), which constructs  $\eta_\epsilon^\mathcal{D}$  by removing the large blocks from  $\eta$  so that it retains all of its diversity but has total mass bounded by  $\epsilon$ . Let

$$\eta' := (2^{1/\alpha} \odot \eta)_\epsilon^\mathcal{D} \quad \text{and} \quad \eta'' := \eta_{\epsilon/2}^\mathcal{D}. \tag{3.9}$$

Then  $\eta'$  (respectively,  $\eta''$ ) has total diversity  $2/\epsilon$  (resp.  $1/\epsilon$ ) and mass less than  $\epsilon$  (resp.  $\epsilon/2$ ). Both have the property that their diversity processes  $\mathcal{D}_{\eta'}$  and  $\mathcal{D}_{\eta''}$  are constant 0 until they jump up at the right ends of the respective partitions. We define

$$\beta := \eta' \star (0, 1) \quad \text{and} \quad \gamma := \eta'' \star \eta'' \star (0, 1). \tag{3.10}$$

Then we see  $d_\alpha(\beta, \gamma) \leq \epsilon$  via the obvious correspondence of matching the two unit mass blocks and leaving all others unmatched. At the same time,  $\mathcal{D}_\beta$  is constant 0 until it jumps to  $2/\epsilon$ , while  $\mathcal{D}_\gamma$  first jumps to  $1/\epsilon$  before jumping to  $2/\epsilon$ . Thus,  $d_{\mathcal{D}}(\mathcal{D}_\beta, \mathcal{D}_\gamma) \geq 1/\epsilon$ .  $\square$

For  $t \geq 0$ , let

$$D_{\beta,n}(t) := \Gamma(1 - \alpha)x_n^\alpha \#\{(a, b) \in \beta_n : b \leq t\}, \quad \text{where } x_n = \min\{\text{Leb}(U) : U \in \beta_n\}. \tag{3.11}$$

If the following two limits are equal, then we adapt Definition 1.2 to additionally define

$$\begin{aligned} \mathcal{D}_\beta^+(t) &:= \lim_{u \downarrow t} \limsup_{h \downarrow 0} \Gamma(1 - \alpha)h^\alpha \#\{(a, b) \in \beta : (b - a) > h, b \leq u\} \\ &= \lim_{u \downarrow t} \liminf_{h \downarrow 0} \Gamma(1 - \alpha)h^\alpha \#\{(a, b) \in \beta : (b - a) > h, b \leq u\}. \end{aligned} \tag{3.12}$$

**Lemma 3.10.** (i) *The map  $\beta_n \mapsto D_{\beta,n}$  is Borel under  $d'_H$ .*

(ii) *The set  $\{(\beta, t) \in \mathcal{I}_H \times [0, \infty) : \mathcal{D}_\beta(t) \text{ exists}\}$  is Borel under  $d'_H$  in the first coordinate plus the Euclidean metric in the second. The map  $(\beta, t) \mapsto \mathcal{D}_\beta(t)$  is measurable on this set, under the same  $\sigma$ -algebra. The same assertions hold with  $\mathcal{D}_\beta(t)$  replaced by  $\mathcal{D}_\beta^+(t)$ .*

(iii) *For  $\beta \in \mathcal{I}_\alpha$ , the pairs  $(\beta_n, D_{\beta,n})$  converge to  $(\beta, \mathcal{D}_\beta(\cdot+))$  under  $d_{\mathcal{J}}$ .*

*Proof.* (i) The measurability of  $\beta_n \mapsto D_{\beta,n}$  follows as in the proof of Lemma 3.8 (iii) from the measurability of RANKED and restrictions.

(ii) The set of interval partitions with finitely many blocks is Borel in  $(\mathcal{I}_H, d'_H)$ , and diversity is constant 0 for such interval partitions. It remains to check the claim for interval partitions with infinitely many blocks. Consider  $\beta \in \mathcal{I}_H$  with infinitely many blocks. For  $n \geq 1$ , let  $U_1, \dots, U_n$  denote the  $n$  largest blocks of  $\beta$ , in left-to-right order. Let  $\theta_{\beta,n} : [0, \|\beta\|] \rightarrow [0, \|\beta_n\|]$  denote the continuous time-change starting from  $\theta_{\beta,n}(0) = 0$ , increasing with slope 1 on  $\bigcup_{i \in [n]} U_i$ , and having slope 0 on  $[0, \|\beta\|] \setminus \bigcup_{i \in [n]} \bar{U}_i$ , where  $\bar{U}$  denotes closure. Note that  $\{\theta_{\beta,n}(U_1), \dots, \theta_{\beta,n}(U_n)\} = \beta_n$ . It follows from similar arguments to those in the proof of Lemma 3.8 (iii) that  $\beta \mapsto \theta_{\beta,n}$  is measurable from  $(\mathcal{I}_H, d'_H)$  to  $\mathcal{C}([0, \infty), [0, \infty))$ . Also,  $f_n(\beta, t) := D_{\beta,n}(\theta_{\beta,n}(t))$  is Borel since pre-images of  $(-\infty, x)$  are open for all  $x \in \mathbb{R}$ , i.e.  $f_n$  is upper semi-continuous.

By comparing (3.11) to Definition 1.2 of  $\mathcal{D}_\beta$ , we see that  $\lim_{n \uparrow \infty} f_n(\beta, t) = \mathcal{D}_\beta(t)$  for every  $t \geq 0$ , with each limit existing if and only if the other does. As  $f_n$  is Borel, this

proves the two claims for  $\mathcal{D}_\beta(t)$ . By monotonicity of the limiting terms in (3.12),  $\mathcal{D}_\beta^+(t)$  exists if and only if

$$\lim_{m \uparrow \infty} \limsup_{n \uparrow \infty} D_{\beta,n}(\theta_{\beta,n}([2^m t + 1]2^{-m})) = \lim_{m \uparrow \infty} \liminf_{n \uparrow \infty} D_{\beta,n}(\theta_{\beta,n}([2^m t + 1]2^{-m})). \quad (3.13)$$

If these limits are equal, then they equal  $\mathcal{D}_\beta^+(t)$ . This proves the two claims for  $\mathcal{D}_\beta^+(t)$ .

(iii) This follows from the previous argument by taking the correspondences from  $\beta$  to  $\beta_n$  that pair  $U_i$  with  $\theta_{\beta,n}(U_i)$ , for each  $i \in [n]$ .  $\square$

**Lemma 3.11.** Consider the map  $\iota: \mathcal{I}_\alpha \rightarrow \mathcal{J}$  given by  $\iota(\eta) = (\eta, \mathcal{D}_\eta(\cdot +))$ .

- (i) Both  $\iota(\mathcal{I}_\alpha)$  and  $\mathcal{J} \setminus \iota(\mathcal{I}_\alpha)$  are dense in  $(\mathcal{J}, d_{\mathcal{J}})$ .
- (ii) Both  $\iota(\mathcal{I}_\alpha)$  and  $\mathcal{J} \setminus \iota(\mathcal{I}_\alpha)$  are Borel subsets of  $\mathcal{J}$ .
- (iii) The space  $(\mathcal{J}, d_{\mathcal{J}})$  is a completion of  $(\mathcal{I}_\alpha, d_\alpha)$ , with respect to the isometric embedding  $\iota$ .

*Proof.* (i) By the definitions of  $d_\alpha$  and  $d_{\mathcal{J}}$ , the map  $\iota$  is an isometry. Take  $(\beta, g) \in \mathcal{J} \setminus \iota(\mathcal{I}_\alpha)$  and  $\eta \in \mathcal{I}_\alpha$  with  $\mathcal{D}_\eta(\infty) = g(\infty) \geq 0$  and such that  $t \mapsto \mathcal{D}_\eta(t)$  is continuous on  $[0, \|\eta\|]$ . Using the notation of the proof of Lemma 3.5, we consider

$$\beta^{(n)} := \eta_{1/n} \oplus_{g(a_1^{(n)})} k_1^{(n)} \cdots \oplus_{g(a_{r_n}^{(n)})} k_{r_n}^{(n)}.$$

Then  $d_{\mathcal{J}}((\beta^{(n)}, \mathcal{D}_{\beta^{(n)}}), (\beta, g)) \rightarrow 0$ , i.e.  $\beta$  is in the closure of  $\iota(\mathcal{I}_\alpha)$ . The same argument, with roles of  $(\eta, \mathcal{D}_\eta(\cdot +))$  and  $(\beta, g)$  swapped (now  $\eta \in \mathcal{I}_\alpha$  general and  $(\beta, g) \in \mathcal{J} \setminus \iota(\mathcal{I}_\alpha)$  and such that  $g$  is continuous), shows that  $(\eta, \mathcal{D}_\eta)$  is in the closure of  $\mathcal{J} \setminus \iota(\mathcal{I}_\alpha)$ .

(ii) Recall that for  $\eta \in \mathcal{I}_\alpha$  we have  $\mathcal{D}_\eta(\cdot +) = \mathcal{D}_\eta^+$  identically. Thus,

$$\iota(\mathcal{I}_\alpha) = \{(\eta, f) \in \mathcal{J} : \text{for all } t \in [0, \|\eta\|], \mathcal{D}_\eta(t) \text{ exists and } \mathcal{D}_\eta^+(t) = f(t)\}.$$

By Lemmas 3.8 (iv) and 3.10 (ii), the following set is Borel under  $d_{\mathcal{J}}$ :

$$A := \{(\eta, f) \in \mathcal{J} : \text{for all } t \in [0, \|\eta\|] \cap \mathbb{Q}, \mathcal{D}_\eta^+(t) \text{ exists and equals } f(t)\}.$$

For  $(\eta, f) \in A$ , writing  $\mathcal{D}_\eta^+(t)$  as in (3.13), we find that  $\mathcal{D}_\eta^+(t)$  exists for all  $t \in [0, \infty)$  and by the right-continuity and monotonicity of  $f$  and  $\mathcal{D}_\eta^+$  we have  $f = \mathcal{D}_\eta^+$  identically. By comparing Definition 1.2 of  $\mathcal{D}_\eta$  with (3.12), we see that if  $\mathcal{D}_\eta^+$  is continuous at some  $t \in [0, \|\eta\|]$  then  $\mathcal{D}_\eta(t)$  exists and equals  $\mathcal{D}_\eta^+(t)$ , by a sandwiching argument. Thus,  $\iota(\mathcal{I}_\alpha)$  is the set of  $(\eta, f) \in A$  for which  $\mathcal{D}_\eta(t)$  exists at each time  $t$  at which  $f$  jumps.

By Lemma 3.8 (iv),  $(\eta, f) \mapsto f$  is measurable from  $(A, d_{\mathcal{J}})$  into Skorokhod space. By [20, Proposition II.(1.16)], the map from  $f$  to the point process of its jumps is measurable; and by [6, Proposition 9.1.XII], we can measurably map the latter to a sequence  $(t_1, \Delta_1), (t_2, \Delta_2), \dots$  listing times and sizes of all jumps of  $f$ , though these may not be listed in chronological order. We write  $\tau_i(\eta, f) := t_i$ , or  $\tau_i(\eta, f) := -1$  if  $f$  has less than  $i$  jumps. Then

$$\iota(\mathcal{I}_\alpha) = \{(\eta, f) \in A : \text{for all } i \in \mathbb{N}, \tau_i(\eta, f) = -1 \text{ or } \mathcal{D}_\eta(\tau_i(\eta, f)) \text{ exists}\}.$$

By Lemma 3.10 (ii), this set is measurable.

(iii) It is clear from the definition of  $d_{\mathcal{J}}$ , based on that of  $d_\alpha$ , that  $\iota$  is an isometry. Now consider any Cauchy sequence  $((\eta_n, f_n), n \geq 1)$  in  $(\mathcal{J}, d_{\mathcal{J}})$ . Then  $(f_n(\infty), n \geq 1)$  is a Cauchy sequence in  $[0, \infty)$ ; let us denote the limit by  $f(\infty)$ . Consider  $(\beta^{(0)}, f^{(0)}) = (\emptyset, f(\infty)) \in \mathcal{J}$ , i.e. the empty partition with the increasing function that is constant

$f(\infty)$ . Recall the space  $\mathcal{S}^\downarrow$  introduced above Theorem 2.5. Let  $\mathbf{s}_n = (s_n^{(i)}, i \geq 1) = (\text{Leb}(U), U \in \eta_n)^\downarrow \in \mathcal{S}^\downarrow$  be the decreasing rearrangement of interval sizes. Then for all correspondences  $(U_j, V_j)_{j \in [k]}$ ,

$$\ell^1(\mathbf{s}_n, \mathbf{s}_m) = \sum_{i \geq 1} \left| s_n^{(i)} - s_m^{(i)} \right| \leq \sum_{j \in [k]} |\text{Leb}(U_j) - \text{Leb}(V_j)| + \|\eta_n\| + \|\eta_m\| - \sum_{j \in [k]} \text{Leb}(U_j) + \text{Leb}(V_j),$$

Let  $\epsilon > 0$ . By the Cauchy property of  $((\eta_n, f_n), n \geq 1)$ , there is some  $N_1 \geq 1$  so that  $d_{\mathcal{J}}((\eta_n, f_n), (\eta_m, f_m)) < \epsilon/2$  for all  $m, n \geq N_1$ . Taking the infimum over all correspondences on the RHS of the display, this yields  $\ell^1(\mathbf{s}_n, \mathbf{s}_m) < \epsilon$  for all  $m, n \geq N_1$ . By completeness of  $(\mathcal{S}^\downarrow, \ell^1)$ , we have convergence  $\mathbf{s}_n \rightarrow \mathbf{s} = (s^{(i)}, i \geq 1) \in \mathcal{S}^\downarrow$ .

Now consider any  $r \geq 1$  with  $s^{(r)} > s^{(r+1)}$  and  $\epsilon > 0$  with  $3\epsilon < s^{(r)} - s^{(r+1)}$ . Then there is  $N_2 \geq 1$  such that for all  $n \geq N_2$ , there are precisely  $r$  intervals  $(a_1^{(n)}, a_1^{(n)} + k_1^{(n)})$ ,  $\dots$ ,  $(a_r^{(n)}, a_r^{(n)} + k_r^{(n)}) \in \eta_n$  of length greater than  $s^{(r)} - \epsilon$ . We define  $\beta_n^{(r)} := \star_{j \in [r]} \left\{ (0, k_j^{(n)}) \right\}$  and associate to these intervals the  $f_n$ -values of the corresponding intervals in  $\eta_n$ :

$$f_n^{(r)} \left( k_1^{(n)} + \dots + k_j^{(n)} + x \right) = \begin{cases} f_n \left( a_j^{(n)} \right) & \text{for } x \in [0, k_j^{(n)}), j \in [0, r - 1], \\ f_n(\infty) & \text{for } x \geq k_1^{(n)} + \dots + k_r^{(n)}. \end{cases}$$

Then  $d_{\mathcal{J}}((\beta_n^{(r)}, f_n^{(r)}), (\beta_m^{(r)}, f_m^{(r)})) \leq d_{\mathcal{J}}((\eta_n, f_n), (\eta_m, f_m))$ , so  $(\beta_n^{(r)}, f_n^{(r)}), n \geq 1$ , is a Cauchy sequence in  $(\mathcal{J}, d_{\mathcal{J}})$ ; and since for  $n \geq N_2$

$$d_{\mathcal{J}}(\beta_n^{(r)}, \beta_m^{(r)}) = \max \left\{ \sup_{j \in [r]} \left| f_n(a_j^{(n)}) - f_m(a_j^{(m)}) \right|, \sum_{j \in [r]} \left| k_j^{(n)} - k_j^{(m)} \right| \right\},$$

the vector  $\left( \left( f_n \left( a_j^{(n)} \right), k_j^{(n)} \right), 1 \leq j \leq r \right)$  is a Cauchy sequence in the metric space  $(\mathbb{R}^{2r}, \|\cdot\|_\infty)$ . By completeness of  $(\mathbb{R}^{2r}, \|\cdot\|_\infty)$ , we have convergence to a limit  $((f_j, k_j), 1 \leq j \leq r)$ , which gives rise to a  $d_{\mathcal{J}}$ -limit  $(\beta^{(r)}, f^{(r)}) \in \mathcal{J}$  of  $((\beta_n^{(r)}, f_n^{(r)}), n \geq 1)$ . By construction,  $(\beta^{(r)}, f^{(r)})$  is consistent as  $r$  varies, in the sense that they are related by insertions of intervals of sizes from  $\mathbf{s}$ , and natural correspondences demonstrate that convergence  $(\beta^{(r)}, f^{(r)}) \rightarrow (\beta, f)$  holds in  $\mathcal{J}$  for a limiting  $(\beta, f) \in \mathcal{J}$  that incorporates intervals of all sizes  $s^{(i)}, i \geq 1$ .

Finally, let  $\epsilon > 0$  and  $N_1 \geq 1$  be as above. Then there is  $r$  large enough so that, following the notation of (3.4),

$$\|\beta_{\mathbf{s}^{(r)}}^\emptyset\| = \sum_{j \geq r+1} s_j < \epsilon/4.$$

Since  $\mathbf{s}_n \rightarrow \mathbf{s}$ , there is  $N_3 \geq N_1$  such that for all  $n \geq N_3$ , we have  $\ell^1(\mathbf{s}_n, \mathbf{s}) < \epsilon/4$ . Finally, there is  $N_4 \geq N_3$  so that for all  $n \geq N_4$  we have  $d_{\mathcal{J}} \left( \left( \beta_n^{(r)}, f_n^{(r)} \right), \left( \beta^{(r)}, f^{(r)} \right) \right) < \epsilon/4$ . Then for all  $n \geq N_4$ , we have

$$\begin{aligned} d_{\mathcal{J}}((\eta_n, f_n), (\beta, f)) &\leq d_{\mathcal{J}}((\eta_n, f_n), (\beta_n^{(r)}, f_n^{(r)})) + d_{\mathcal{J}}((\beta_n^{(r)}, f_n^{(r)}), (\beta^{(r)}, f^{(r)})) + d_{\mathcal{J}}((\beta^{(r)}, f^{(r)}), (\beta, f)) \\ &< \sum_{j=r+1}^\infty s_j + \ell^1(\mathbf{s}_n, \mathbf{s}) + \frac{\epsilon}{4} + \frac{\epsilon}{4} < \epsilon. \end{aligned}$$

Hence,  $((\eta_n, f_n), n \geq 1)$  converges to  $(\beta, f)$  in  $(\mathcal{J}, d_{\mathcal{J}})$ . Therefore,  $(\mathcal{J}, d_{\mathcal{J}})$  is complete.  $\square$

**Corollary 3.12.**  $(\mathcal{J}, d_{\mathcal{J}})$  and  $(\mathcal{I}_H, d'_H)$  are complete and separable metric spaces.

*Proof.* We have shown in the lemmas that  $(\mathcal{J}, d_{\mathcal{J}})$  is a complete metric space, and since the completion of a separable metric space is also separable,  $(\mathcal{J}, d_{\mathcal{J}})$  is also separable. As  $(\mathcal{I}_H, d'_H)$  has a natural isometrical embedding  $\mathcal{I}_H \times \{0\} \subset \mathcal{J}$  and  $\mathcal{I}_H \times \{0\}$  is closed in  $(\mathcal{J}, d_{\mathcal{J}})$ , completeness and separability of  $(\mathcal{I}_H, d'_H)$  follow.  $\square$

**Corollary 3.13.** *The sets  $\mathcal{I}_\alpha$  and  $\mathcal{I}_H \setminus \mathcal{I}_\alpha$  are dense Borel subsets of  $(\mathcal{I}_H, d'_H)$ .*

**Lemma 3.14.** *The space  $(\mathcal{J}, d_{\mathcal{J}})$  is not locally compact.*

*Proof.* Consider the interval partitions  $\eta_n = \{((k-1)2^{-n}, k2^{-n}), 1 \leq k \leq 2^n\}$  and  $f_n \equiv 0$ . For  $m < n$ , any correspondence for  $\eta_m$  and  $\eta_n$  that matches up any intervals of  $\eta_m$  and  $\eta_n$  attracts a term  $2^{-m} - 2^{-n} \geq 2^{-n}$ , so it is best to use the trivial correspondence which gives  $d_{\mathcal{J}}((\eta_m, f_m), (\eta_n, f_n)) = 1$ . Now assume that  $(\emptyset, 0) \in \mathcal{J}$  has a compact neighbourhood  $K$ . Then  $K$  contains an open ball of some radius  $2\epsilon > 0$ , which contains  $(\epsilon \odot \eta_n, 0)$  for all  $n \geq 1$ . Covering  $K$  with open balls of radius  $\epsilon/2$ , the open balls around  $(\epsilon \odot \eta_n, 0)$  are disjoint, so there cannot be a finite subcover. This contradicts the compactness of  $K$ . Hence  $(\emptyset, 0)$  does not have a compact neighbourhood, and  $(\mathcal{J}, d_{\mathcal{J}})$  is not locally compact.  $\square$

Even though  $(\mathcal{J}, d_{\mathcal{J}})$  is not locally compact, we can now deduce that  $(\mathcal{I}_\alpha, d_\alpha)$  is Lusin:

**Proposition 3.15.** *The metric space  $(\mathcal{I}_\alpha, d_\alpha)$  is isometric to a path-connected Borel subset of a complete separable metric space  $(\mathcal{J}, d_{\mathcal{J}})$ . Furthermore,  $(\mathcal{I}_\alpha, d_\alpha)$  is Lusin.*

*Proof.* Lemma 3.11 and Corollary 3.12 yield that  $(\mathcal{I}_\alpha, d_\alpha)$  is isometric to a Borel subset of the Polish space  $(\mathcal{J}, d_{\mathcal{J}})$ . By [29, Theorem II.82.5], this implies  $(\mathcal{I}_\alpha, d_\alpha)$  is Lusin.  $\square$

**Proposition 3.16.** *Each of  $d_\alpha$ ,  $d_H$ , and  $d'_H$  generate the same Borel  $\sigma$ -algebra on  $\mathcal{I}_\alpha$ .*

*Proof.* In light of Proposition 3.4 (ii)–(iii), we need only check that all  $d_\alpha$ -balls are Borel sets with respect to  $d'_H$ . Recall the notation  $\beta_n$  and  $D_{\beta,n}$  of introduced before Lemmas 3.8 and 3.10. By Lemma 3.10 (iii), the  $d_\alpha$ -ball of radius  $r > 0$  about  $\beta$  equals

$$\bigcup_{m \geq 1} \bigcup_{N \geq 1} \bigcap_{n > N} \{ \gamma \in \mathcal{I}_\alpha : d_{\mathcal{J}}((\beta_n, D_{\beta,n}), (\gamma_n, D_{\gamma,n})) < r - m^{-1} \}.$$

The claimed measurability now follows by Lemmas 3.10 (i) and 3.8 (ii).  $\square$

## References

- [1] David J. Aldous, *Exchangeability and related topics*, École d'été de probabilités de Saint-Flour, XIII—1983, Lecture Notes in Math., vol. 1117, Springer, Berlin, 1985, pp. 1–198. MR-0883646
- [2] David Aldous, *The continuum random tree. I*, Ann. Probab. **19** (1991), no. 1, 1–28. MR-1085326
- [3] Jean Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006. MR-2253162
- [4] Patrick Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999, A Wiley-Interscience Publication. MR-1700749
- [5] Cristina Costantini, Pierpaolo De Blasi, Stewart N. Ethier, Matteo Ruggiero, and Dario Spanó, *Wright–Fisher construction of the two-parameter Poisson–Dirichlet diffusion*, Ann. Appl. Probab. **27** (2017), no. 3, 1923–1950. MR-3678488
- [6] Daryl J. Daley and David Vere-Jones, *An introduction to the theory of point processes. Vol. II*, second ed., Probability and its Applications (New York), Springer, New York, 2008, General theory and structure. MR-2371524
- [7] Richard M. Dudley, *Real analysis and probability*, Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, 2002, Revised reprint of the 1989 original. MR-1932358
- [8] Stewart Ethier, *A property of Petrov’s diffusion*, Electron. Commun. Probab. **19** (2014), no. 65, 1–4. MR-3262071
- [9] Steven N. Evans, Jim Pitman, and Anita Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Related Fields **134** (2006), no. 1, 81–126. MR-2221786

- [10] Shui Feng and Feng-Yu Wang, *A class of infinite-dimensional diffusion processes with connection to population genetics*, J. Appl. Probab. **44** (2007), no. 4, 938–949. MR-2382936
- [11] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Uniform control of local times of spectrally positive stable processes*, Ann. Appl. Probab. **28** (2018), no. 4, 2592–2634. MR-3843837
- [12] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Aldous diffusion I: a projective system of continuum  $k$ -tree evolutions*, arXiv:1809.07756 [math.PR], 2018.
- [13] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Diffusions on a space of interval partitions: construction from marked Lévy processes*, arXiv:1909.02584 [math.PR], 2019.
- [14] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Diffusions on a space of interval partitions: Poisson–Dirichlet stationary distributions*, arXiv:1910.07626 [math.PR], 2019.
- [15] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Interval partition diffusions: connection with Petrov’s Poisson–Dirichlet diffusions*, Work in progress, revising parts of arXiv:1609.06706 [math.PR], 2020. MR-2693404
- [16] Alexander V. Gnedin, *The representation of composition structures*, Ann. Probab. **25** (1997), no. 3, 1437–1450. MR-1457625
- [17] Alexander V. Gnedin, *Regeneration in random combinatorial structures*, Probability Surveys **7** (2010), 105–156. MR-2684164
- [18] Alexander Gnedin and Jim Pitman, *Regenerative composition structures*, Ann. Probab. **33** (2005), no. 2, 445–479. MR-2122798
- [19] Alexander Gnedin, Jim Pitman, and Marc Yor, *Asymptotic laws for compositions derived from transformed subordinators*, Ann. Probab. **34** (2006), no. 2, 468–492. MR-2223948
- [20] Jean Jacod and Albert N. Shiryaev, *Limit theorems for stochastic processes*, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 288, Springer-Verlag, Berlin, 2003. MR-1943877
- [21] John Frank Charles Kingman, *Random partitions in population genetics*, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **361** (1978), no. 1704, 1–20. MR-0526801
- [22] Wolfgang Löhr, Leonid Mytnik, and Anita Winter, *The Aldous chain on cladograms in the diffusion limit*, arXiv:1805.12057 [math.PR], 2018.
- [23] Grégory Miermont, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 5, 725–781. MR-2571957
- [24] Leonid A. Petrov, *A two-parameter family of infinite-dimensional diffusions on the Kingman simplex*, Funktsional. Anal. i Prilozhen. **43** (2009), no. 4, 45–66. MR-2596654
- [25] Jim Pitman, *Exchangeable and partially exchangeable random partitions*, Probab. Theory Related Fields **102** (1995), no. 2, 145–158. MR-1337249
- [26] Jim Pitman, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. MR-2245368
- [27] Jim Pitman and Matthias Winkel, *Regenerative tree growth: binary self-similar continuum random trees and Poisson–Dirichlet compositions*, Ann. Probab. **37** (2009), no. 5, 1999–2041. MR-2561439
- [28] Jim Pitman and Marc Yor, *Arcsine laws and interval partitions derived from a stable subordinator*, Proc. London Math. Soc. (3) **65** (1992), no. 2, 326–356. MR-1168191
- [29] L. C. G. Rogers and David Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, second ed., Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester, 1994, Foundations. MR-1331599
- [30] Matteo Ruggiero and Stephen G. Walker, *Countable representation for infinite dimensional diffusions derived from the two-parameter Poisson–Dirichlet process*, Electron. Commun. Probab. **14** (2009), 501–517. MR-2564485

- [31] Matteo Ruggiero, Stephen G. Walker, and Stefano Favaro, *Alpha-diversity processes and normalized inverse-Gaussian diffusions*, Ann. Appl. Probab. **23** (2013), no. 1, 386–425. MR-3059239

**Acknowledgments.** We are grateful to the referee and to the Associate Editor for their suggestions improving the presentation of this paper.