

# Non-Regenerative Sampling Distributions of $\Xi$ -Coalescents: Möhle's Conjecture with a Kingman Component, for Finitely Many Atoms or Boundedly Many Dust-Free Boxes, and Off One Mutation Rate

Alper Ferudun\*

October 19, 2026

## Abstract

M. Möhle (Oberwolfach Report 40/2005) conjectured that the sampling distributions of the infinitely-many-alleles model along a coalescent with simultaneous multiple collisions (a  $\Xi$ -coalescent), with mutation rate  $\rho > 0$ , are never regenerative in the sense of Gnedin and Pitman, except for Kingman's coalescent and the star-shaped coalescent. He had proved this for coalescents with multiple but not simultaneous collisions ( $\Lambda$ -coalescents; *Combin. Probab. Comput.* 16 (2007)). We prove the conjecture for three classes of  $\Xi$  and for all but at most one mutation rate in general; the conjecture is not settled. Write  $\Xi = a\delta_0 + \Xi_0$ . The main tool is an identity: the regeneration equation at the partition  $(2, 2, 1^{n-4})$  holds if and only if  $\Phi(n-1)X(n) = \text{Bin}(n)W(n-1)$ , where  $X(n)$ ,  $\text{Bin}(n)$  and  $W(n)$  are the total rates of the mergers of two pairs, of one pair, and of all other kinds, among  $n$  lineages, and  $\Phi(n)$  is  $n\rho$  plus the total merger rate. Consequences: (A) if  $\Xi$  is not a  $\Lambda$ -coalescent, the partition structure is non-regenerative for every  $\rho > 0$  except possibly one explicit value  $\rho^*(\Xi)$ , and Möhle's theorem is recovered at sample size four; (B) every  $\Xi$ -coalescent with a Kingman component ( $a > 0$ ), other than Kingman's coalescent, is non-regenerative for every  $\rho > 0$ ; (C, D) the same holds for  $a = 0$ , the star-shaped coalescent excepted, when  $\Xi_0$  is carried by paintboxes without dust with boundedly many boxes, or is a finite sum of point masses at paintboxes with finitely many boxes. The measure  $\Xi = \frac{17}{1024}\delta_0 + \delta_{(1/4, 1/8, 0, \dots)}$  with  $\rho = \frac{13833}{5120}$  is regenerative at all sample sizes  $n \leq 5$  but not at  $n = 6$ , so finitely many sample sizes do not decide the question. Open remains the case  $a = 0$ ,  $\Xi_0$  charging  $\{x_2 > 0\}$  and outside the classes of (C) and (D), at the single value  $\rho = \rho^*(\Xi)$ . The conjecture and the  $\Lambda$  case are Möhle's; his recursion for the sampling probabilities is re-derived. No printed answer to the question was found in the literature searched, which is not a proof of novelty. This is an unrefereed note.

## 1 Introduction

### 1.1 The question

Let  $\nabla := \{x = (x_1, x_2, \dots) : x_1 \geq x_2 \geq \dots \geq 0, \sum_i x_i \leq 1\}$  be the infinite simplex, with  $p_2(x) := \sum_i x_i^2$ , and let  $\Xi$  be a finite measure on  $\nabla$ , written  $\Xi = a\delta_0 + \Xi_0$  with  $a \geq 0$  and  $\Xi_0(\{0\}) = 0$ . The  $\Xi$ -coalescent of Schweinsberg [Sch00] and Möhle–Sagitov [MS01] is the exchangeable coalescent in which, when there are  $b$  lineages, every specific collision in which  $r \geq 1$  groups of  $k_1, \dots, k_r \geq 2$

---

\*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: <https://github.com/AlperTheKing>.  
Unrefereed preprint. CC BY 4.0.

lineages merge simultaneously, while  $s = b - \sum_i k_i$  lineages are not involved, happens at rate

$$\lambda_{b;k_1,\dots,k_r;s} = \int_{\nabla} \sum_{l=0}^s \binom{s}{l} \sum_{i_1,\dots,i_{r+l} \text{ distinct}} x_{i_1}^{k_1} \cdots x_{i_r}^{k_r} x_{i_{r+1}} \cdots x_{i_{r+l}} \left(1 - \sum_j x_j\right)^{s-l} \frac{\Xi_0(dx)}{p_2(x)} + a \mathbf{1}\{r=1, k_1=2\} \quad (1)$$

(Schweinsberg's formula, in the notation of [Don07, (8)]). If  $\Xi_0$  is carried by  $\{x : x_2 = 0\}$ , only one group merges at a time and (1) becomes  $\lambda_{b;k;b-k} = \int_{[0,1]} y^{k-2} (1-y)^{b-k} \Lambda(dy)$  with  $\Lambda := a\delta_0 + \Xi_0 \circ (x \mapsto x_1)^{-1}$ : this is the  $\Lambda$ -coalescent of Pitman [Pit99] and Sagitov [Sag99]. Kingman's coalescent [Kin82] is  $\Xi = c\delta_0$ , and the *star-shaped* coalescent, in which all lineages merge at once, is  $\Xi = c\delta_{(1,0,0,\dots)}$ , i.e.  $\Lambda = c\delta_1$  ( $c > 0$ ).

Along the genealogy of a sample of  $n$  individuals, mutations occur on every lineage at the points of a Poisson process of rate  $\rho > 0$ , each creating a new type (infinitely-many-alleles model). The *allelic partition*  $\Pi_n$  of the sample groups individuals of the same type. The laws of  $\Pi_n$ ,  $n \geq 1$ , are exchangeable and consistent [Don07, p. 4], so they form a partition structure; the *sampling distribution* of [Mö06a] is the law of the ranked block sizes of  $\Pi_n$ . Following Gnedin and Pitman [GP05b, Definition 1], the partition structure is *regenerative* if for every  $n$  a part  $X_n$  of the ranked partition can be selected at random so that, for every  $1 \leq x < n$ , given  $X_n = x$  the remaining partition of  $n - x$  has the law of the ranked  $\Pi_{n-x}$ . These are exactly the partition structures obtained by forgetting the order of the parts of a regenerative composition structure [GP05b, Corollary 6], [GP05a]; this is why [Mö06a] speaks of composition structures. For  $\Lambda$ -coalescents Möhle proved that the structure is regenerative if and only if  $\Lambda$  is concentrated in 0 or in 1 [Mö07], [Mö06a, p. 2281], and he conjectured [Mö06a, §3, p. 2281] the same for the wider class of  $\Xi$ -coalescents, parametrised in [Mö06a] by Möhle and Sagitov's sequence of measures  $\Lambda_j$  on the simplices  $\Delta_j$ ;  $\Lambda_j = 0$  for all  $j \geq 2$  is the  $\Lambda$ -coalescent case. In Schweinsberg's parametrisation, used here, the question reads as follows;  $\Xi = 0$  is excluded, since without coalescence  $\Pi_n$  consists of singletons and is trivially regenerative.

**Question 1.1** (Möhle [Mö06a]). *Let  $\Xi \neq 0$  and  $\rho > 0$ . Is the allelic partition structure of the  $\Xi$ -coalescent with mutation rate  $\rho$  non-regenerative whenever  $\Xi$  is neither of the form  $c\delta_0$  nor of the form  $c\delta_{(1,0,\dots)}$ ?*

This is record OWR-1061-006 of the UnsolvedMath dataset [ula26]. The scale is immaterial:  $(c\Xi, c\rho)$  gives the same partition structure as  $(\Xi, \rho)$ . The two exceptional families are regenerative: Kingman's coalescent gives the Ewens sampling formula [Ewe72], [Mö06a, (5)], regenerative under size-biased deletion [GP05b, (10)–(12)], and the star-shaped coalescent gives a hook partition structure [GP05b, §6], [Mö06a, p. 2281]. We do not settle Question 1.1; the results below prove the conjectured answer in the cases stated and isolate what is left.

## 1.2 Results

For  $b \geq 2$  lineages let  $T(b)$  be the total collision rate and  $\Phi(b) := b\rho + T(b)$  (so  $\Phi(1) = \rho$ ), let  $t_b := \lambda_{b;2;b-2}$ ,  $\tau_b := \lambda_{b;3;b-3}$ ,  $j_b := \lambda_{b;2,2;b-4}$ , and let

$$\text{Bin}(b) := \binom{b}{2} t_b, \quad X(b) := \frac{1}{2} \binom{b}{2} \binom{b-2}{2} j_b, \quad W(b) := T(b) - \text{Bin}(b)$$

be the total rates of the collisions of  $b$  lineages in which one pair merges and nothing else, in which two pairs merge and nothing else, and of all other collisions. The key observation is Lemma 3.1: if the partition structure is regenerative, then for every  $n \geq 4$

$$\Phi(n-1)X(n) = \text{Bin}(n)W(n-1); \quad (2)$$

more precisely, the defect of the regeneration equation at the partition  $(2, 2, 1^{n-4})$  is a positive multiple of  $\Phi(n-1)X(n) - \text{Bin}(n)W(n-1)$ .

**Theorem A** (sample size four). *Let  $\Xi \neq 0$  and  $\rho > 0$ . The partition  $\Pi_4$  is regenerative if and only if  $j_4 \Phi(3) = 2t_4\tau_3$ . If  $\Xi$  is not a  $\Lambda$ -coalescent, then  $j_4 > 0$ , and the partition structure is non-regenerative for every  $\rho > 0$  except possibly  $\rho = \rho^*(\Xi) := \frac{1}{3}(2t_4\tau_3/j_4 - 3t_3 - \tau_3)$  (and for every  $\rho > 0$  if  $\rho^*(\Xi) \leq 0$ ).*

For  $\Lambda$ -coalescents ( $j_4 = 0$ ) Theorem A gives Möhle's theorem (Corollary 3.2).

**Theorem B** (Kingman component). *If  $a = \Xi(\{0\}) > 0$  and  $\Xi_0 \neq 0$ , the partition structure is non-regenerative for every  $\rho > 0$ .*

**Theorem C** (no dust, boundedly many boxes). *Let  $a = 0$ , let  $\Xi_0$  be carried by the points  $x$  with  $\sum_i x_i = 1$  and at most  $K_{\max} < \infty$  positive coordinates, and let  $\Xi_0$  not be a multiple of  $\delta_{(1,0,\dots)}$ . Then the partition structure is non-regenerative for every  $\rho > 0$ .*

**Theorem D** (finitely many atoms). *Let  $a = 0$  and  $\Xi_0 = \sum_{l=1}^L c_l \delta_{x^l}$  with  $c_l > 0$ , where each  $x^l$  has finitely many positive coordinates, and let  $\Xi_0$  not be a multiple of  $\delta_{(1,0,\dots)}$ . Then the partition structure is non-regenerative for every  $\rho > 0$ .*

Together with Corollary 3.2, the conjectured answer is proved for every  $\Lambda$ -coalescent [Mö07], every  $\Xi$  with  $\Xi(\{0\}) > 0$ , the classes of Theorems C and D, and, for every other  $\Xi$ , every  $\rho > 0$  except  $\rho^*(\Xi)$ . Open is the case  $a = 0$ ,  $\Xi_0$  charging  $\{x_2 > 0\}$ ,  $\Xi_0$  outside the classes of Theorems C and D (for instance with infinitely many atoms, or with a diffuse part outside the class of Theorem C), at the single value  $\rho = \rho^*(\Xi)$  when it is positive. For it, Corollary E in Section 6 gives a further necessary condition when  $M := \int \Xi_0(dx)/p_2(x) < \infty$ . Proposition 7.1 shows that no finite set of sample sizes can decide the question in general: an explicit  $\Xi$  with  $a > 0$ , covered by Theorem B, is regenerative at all levels  $n \leq 5$ .

The proof of Theorem B compares two consequences of (2) as  $n \rightarrow \infty$ . With a Kingman component, (2) forces  $X(n)/W(n-1) \rightarrow 1$ ; for every finite  $\Xi_0 \neq 0$ , a multiscale argument on the paintbox representation shows that  $X(m) \leq \gamma W(m)$  with a constant  $\gamma < 1$  for infinitely many  $m$ . Without a Kingman component both sides of (2) may tend to zero, and the method gives only the classes of Theorems C and D.

### 1.3 Related work

Möhle derived the recursion for the sampling probabilities of  $\Xi$ -coalescents [Mö06b, Theorem 5.1] and proved the  $\Lambda$  case of the conjecture [Mö07]. Dong, Gnedin and Pitman [DGP07] developed the theory of these partitions for  $\Lambda$ -coalescents by analogy with regenerative compositions, and Dong [Don07] did so for  $\Xi$ -coalescents; neither paper decides regenerativity in the  $\Xi$  case. Lagerås [Lag07, Theorem 1] showed that, for every  $\Lambda$ , the composition into litters ordered by age is regenerative; it is a different composition, and he restates Möhle's theorem for the partition into types [Lag07, Proposition 4]. Regenerative compositions and  $\Lambda$ -coalescents also meet in [GIM08, GIM11]; see [Gne10] for a survey. None of these treats the  $\Xi$  allelic partition. The question and the  $\Lambda$  case are Möhle's, the recursion we use is his [Mö06b] in Dong's form [Don07], and the notion of regenerativity is Gnedin and Pitman's. No printed answer to Question 1.1 was found in the sources searched (listed in the paragraph "Scope and priority" in Section 8); a search that finds nothing is not a proof of novelty.

## 2 Rates, the recursion and the one-level test

### 2.1 Paintboxes and rates

For  $x \in \nabla$  and  $b \geq 1$  let  $b$  draws be made independently from the *paintbox*  $x$ : a draw falls into box  $k$  with probability  $x_k$  and into the dust with probability  $1 - \sum_k x_k$ . Let  $\Pi_b^x$  be the partition of

$[b] = \{1, \dots, b\}$  in which two draws lie in the same block if and only if they fall into the same box (draws in the dust are singletons), and write  $\mathbb{P}_x$  for the probabilities. For a collision  $C$  of  $b$  lineages, i.e. a set partition of  $[b]$  with a block of size at least 2, of type  $(b; k_1, \dots, k_r; s)$ , we have  $\Pi_b^x = C$  if and only if the draws of each non-singleton block of  $C$  fall into one box, different blocks use different boxes, and every draw forming a singleton block falls into the dust or into a box that no other draw uses. Summing over the set of the  $l$  singletons that fall into boxes, one sees that the integrand in (1) is  $\mathbb{P}_x(\Pi_b^x = C)/p_2(x)$ . With  $\mu(dx) := \Xi_0(dx)/p_2(x)$  on  $\nabla \setminus \{0\}$ ,

$$\lambda_C = a \mathbf{1}\{C \text{ is a single pair}\} + \int \mathbb{P}_x(\Pi_b^x = C) \mu(dx). \quad (3)$$

Let  $Z_b$  be the number of pairs among the first  $b$  draws that fall into the same box. Then  $\Pi_b^x$  is a collision if and only if  $Z_b \geq 1$ ; it consists of one pair and singletons if and only if  $Z_b = 1$ , and of two pairs and singletons if and only if  $Z_b = 2$  (three draws in one box give three pairs). Summing (3) over the corresponding collisions,

$$\begin{aligned} T(b) &= a \binom{b}{2} + R(b), \quad \text{Bin}(b) = a \binom{b}{2} + Y(b), \quad X(b) = \int \pi_2(x; b) \mu(dx), \quad W(b) = \int F(x; b) \mu(dx), \\ R(b) &:= \int \mathbb{P}_x(Z_b \geq 1) \mu(dx), \quad Y(b) := \int \pi_1(x; b) \mu(dx), \end{aligned} \quad (4)$$

with  $\pi_i(x; b) := \mathbb{P}_x(Z_b = i)$  and  $F(x; b) := \mathbb{P}_x(Z_b \geq 2)$ . We also put  $G(x; b) := \mathbb{P}_x(Z_b \geq 3) = F(x; b) - \pi_2(x; b)$ . Since  $\mathbb{P}_x(Z_b \geq 1) \leq \mathbb{E}_x Z_b = \binom{b}{2} p_2(x)$ , all these rates are at most  $\binom{b}{2} (a + \Xi_0(\nabla)) < \infty$ .

Coupling the first  $b$  of  $b+1$  draws gives two identities. First,  $\mathbb{P}_x(Z_{b+1} \geq 1) - \mathbb{P}_x(Z_b \geq 1) = \mathbb{P}_x(Z_b = 0, Z_{b+1} \geq 1)$ , and on this event draw  $b+1$  shares its box with exactly one earlier draw  $i$ , so that  $\Pi_{b+1}^x$  is the single pair  $\{i, b+1\}$ ; with the Kingman part  $a \binom{b+1}{2} - a \binom{b}{2} = ab$  this gives

$$T(b+1) - T(b) = b t_{b+1} \quad (b \geq 1), \quad \text{hence} \quad T(b) = \sum_{m=2}^b (m-1) t_m. \quad (5)$$

Second, if  $\Pi_{b-1}^x$  is the single pair  $\{1, 2\}$ , then draw  $b$  joins the box of draws 1, 2, or the box of one of the draws 3,  $\dots$ ,  $b-1$ , or neither, so that

$$t_{b-1} = t_b + \tau_b + (b-3) j_b \quad (b \geq 3), \quad (6)$$

where  $j_3 := 0$ . Both identities are instances of the consistency relation of  $\Xi$ -coalescents [Don07, (9)]. By (6),  $t_b$  is nonincreasing in  $b$ . By (3),

$$j_4 = \int \sum_{i \neq k} x_i^2 x_k^2 \mu(dx), \quad \tau_3 = W(3) = \int \sum_i x_i^3 \mu(dx), \quad (7)$$

so  $j_4 = 0$  if and only if  $\Xi_0$  is carried by  $\{x_2 = 0\}$ , i.e.  $\Xi$  is a  $\Lambda$ -coalescent, and  $\tau_3 = 0$  if and only if  $\Xi_0 = 0$ .

## 2.2 The recursion

Let  $p$  be the exchangeable partition probability function of the allelic partition:  $\mathbb{P}(\Pi_n = \pi) = p(n_1, \dots, n_\ell)$  for every set partition  $\pi$  of  $[n]$  with blocks of sizes  $n_1, \dots, n_\ell$ . A collision  $C$  of  $n$  lineages is *compatible* with  $\pi$  if every block of  $C$  lies in a block of  $\pi$ ; then  $\pi/C$  denotes the partition of the lineages after the collision induced by  $\pi$  (a block of  $\pi$  of size  $n_i$  becomes a block whose size is the number of blocks of  $C$  inside it).

**Lemma 2.1** (Möhle's recursion).  $p(1) = 1$ , and for  $n \geq 2$  and every set partition  $\pi$  of  $[n]$ ,

$$\Phi(n)p(\pi) = \rho \sum_{\{i\} \in \pi} p(\pi \setminus \{i\}) + \sum_{C \text{ compatible with } \pi} \lambda_C p(\pi/C),$$

where the first sum runs over the singleton blocks of  $\pi$ . In particular  $p(1^n) = \prod_{m=2}^n m\rho/\Phi(m) > 0$ .

*Proof.* Trace the  $n$  lineages back to the first event. By the strong Markov property it is a mutation on a given lineage with probability  $\rho/\Phi(n)$  and a given collision  $C$  with probability  $\lambda_C/\Phi(n)$ . After a mutation on lineage  $i$ , the individual  $i$  carries a type that no other individual of the sample carries, so  $\{i\}$  is a block of  $\Pi_n$ , and the other  $n-1$  lineages, with their mutations, evolve as the genealogy of a sample of size  $n-1$  (consistency of the  $\Xi$ -coalescent); by exchangeability their partition has the law of  $\Pi_{n-1}$ . After a collision  $C$ , the merged lineages have a common ancestral line and hence the same type, so  $\Pi_n$  is compatible with  $C$ , and the ancestral lineages evolve as a sample of size  $n - \sum_i (k_i - 1)$ . For the partition into singletons no collision is compatible. This is [Mö06b, Theorem 5.1] in the form of [Don07, (12) and Theorem 3].  $\square$

### 2.3 The one-level test

For a partition  $\lambda = (n_1, \dots, n_\ell)$  of  $n$  (ranked, or a list of block sizes) and a vector  $q(n; \cdot)$  on  $\{1, \dots, n\}$  consider

$$p(n_1, \dots, n_\ell) = \sum_{j=1}^{\ell} \frac{q(n; n_j)}{\binom{n}{n_j}} p(n_1, \dots, \widehat{n_j}, \dots, n_\ell), \quad p(\emptyset) := 1. \quad (R_n)$$

A partition of  $n$  is a *hook* if it has the form  $(m, 1^{n-m})$ .

**Lemma 2.2.** Fix  $n \geq 2$ , put  $q(n; 1) := p(1^n)/p(1^{n-1}) = n\rho/\Phi(n)$  and, for  $2 \leq m \leq n$ ,

$$q(n; m) := \binom{n}{m} \frac{p(m, 1^{n-m}) - (n-m) \frac{\rho}{\Phi(n)} p(m, 1^{n-m-1})}{p(1^{n-m})}. \quad (8)$$

Then  $q(n; m) \geq 0$  for all  $m$ ,  $(R_n)$  holds at every hook, and  $\Pi_n$  is regenerative in the sense of [GP05b, Definition 1] if and only if  $(R_n)$  holds at every partition of  $n$  that is not a hook. In that case  $\sum_m q(n; m) = 1$  and  $q(n; m)$  is the probability that the selected part has size  $m$ . In particular  $\Pi_2$  and  $\Pi_3$  are always regenerative.

*Proof.* Write  $P(\lambda) := \mathbb{P}(\Pi_n \text{ has ranked block sizes } \lambda) = N(\lambda)p(\lambda)$ , where  $N(\lambda) = n!/\prod_i n_i! \prod_k a_k!$  is the number of set partitions of shape  $\lambda$  and  $a_k$  is the number of parts of size  $k$ . By [GP05b, (4)–(6)],  $\Pi_n$  is regenerative if and only if there are a deletion kernel  $d(\lambda, x) \geq 0$  ( $x$  a part of  $\lambda$ ,  $\sum_x d(\lambda, x) = 1$ ) and numbers  $q(n; x)$  with  $P(\lambda)d(\lambda, x) = q(n; x)P(\lambda - \{x\})$  for all  $\lambda$  and all parts  $x$  of  $\lambda$ , where  $\lambda - \{x\}$  is  $\lambda$  with one part  $x$  deleted and  $P(\emptyset) = 1$ ; then  $q(n; x)$  is the probability that the selected part has size  $x$ . Summing over the distinct parts  $x$  of  $\lambda$  gives  $P(\lambda) = \sum_x q(n; x)P(\lambda - \{x\})$  with  $q \geq 0$ . Conversely, if  $q \geq 0$  satisfies this equation for all  $\lambda$ , then  $d(\lambda, x) := q(n; x)P(\lambda - \{x\})/P(\lambda)$  (any deletion kernel if  $P(\lambda) = 0$ ) has the required properties, and summing the equation over all  $\lambda$  gives  $\sum_x q(n; x) = 1$ , since  $\lambda \mapsto \lambda - \{x\}$  is a bijection from the partitions of  $n$  with a part  $x$  onto the partitions of  $n-x$ . Since  $N(\lambda - \{x\})/N(\lambda) = a_x/\binom{n}{x}$ , the equation in terms of  $p$  is  $(R_n)$ . At  $(1^n)$  and  $(m, 1^{n-m})$ ,  $(R_n)$  reads  $p(1^n) = q(n; 1)p(1^{n-1})$  and  $p(m, 1^{n-m}) = \frac{q(n; m)}{\binom{n}{m}} p(1^{n-m}) + (n-m) \frac{q(n; 1)}{n} p(m, 1^{n-m-1})$ ; as  $p(1^{n-m}) > 0$ , these hold exactly for the values defined above, so any vector satisfying  $(R_n)$  at all partitions is this one. Finally, Lemma 2.1 at the set partition  $\{1, \dots, m\}, \{m+1\}, \dots, \{n\}$  shows that the numerator in (8) equals  $\Phi(n)^{-1} \sum_C \lambda_C p(\pi/C) \geq 0$ , the sum over the collisions inside  $\{1, \dots, m\}$ ; so  $q(n; m) \geq 0$ . For  $n \leq 3$  every partition is a hook.  $\square$

By [GP05b, Theorem 3 and Corollary 4(i)], if  $\Pi_n$  is regenerative, then so is  $\Pi_m$  for every  $m \leq n$ . The partition structure is regenerative if and only if  $\Pi_n$  is regenerative for every  $n$ , so a single  $n$  at which  $(R_n)$  fails proves non-regenerativity.

**Lemma 2.3.** For  $m \geq 2$ ,  $H_m := p(2, 1^{m-2})/p(1^m) = T(m)/(\rho m(m-1))$ .

*Proof.* Lemma 2.1 at  $\{1, 2\}, \{3\}, \dots, \{m\}$  and at the singletons gives  $\Phi(m)p(2, 1^{m-2}) = (m-2)\rho p(2, 1^{m-3}) + t_m p(1^{m-1})$  and  $\Phi(m)p(1^m) = m\rho p(1^{m-1})$ . Dividing,  $H_m = \frac{m-2}{m}H_{m-1} + \frac{t_m}{m\rho}$  with  $H_1 := 0$ , so  $m(m-1)\rho H_m = \sum_{k=2}^m (k-1)t_k = T(m)$  by (5).  $\square$

### 3 The identity at the partition $(2, 2, 1^{n-4})$ and level four

**Lemma 3.1.** For  $n \geq 4$  let  $q(n; \cdot)$  be the vector of Lemma 2.2 and let

$$\varepsilon_n := p(2, 2, 1^{n-4}) - (n-4)\frac{q(n; 1)}{n}p(2, 2, 1^{n-5}) - 2\frac{q(n; 2)}{\binom{n}{2}}p(2, 1^{n-4})$$

be the defect of  $(R_n)$  at the partition  $(2, 2, 1^{n-4})$ . Then

$$\varepsilon_n = \frac{p(1^n)\Delta_n}{\rho^2 \binom{n}{2}^2 \binom{n-2}{2}}, \quad \Delta_n := \Phi(n-1)X(n) - \text{Bin}(n)W(n-1).$$

Consequently, if the partition structure is regenerative, then (2) holds for every  $n \geq 4$ .

*Proof.* Write  $P_k := p(1^k)$ ,  $\kappa_k := P_k/P_{k-1} = k\rho/\Phi(k) = q(k; 1)$ ,  $E_k := p(2, 2, 1^{k-4})/P_k$ , and let  $H_k$  be as in Lemma 2.3.

*Step 1.* By Lemma 2.2,  $(R_n)$  holds at the hook  $(2, 1^{n-2})$ :  $p(2, 1^{n-2}) = (n-2)\frac{\kappa_n}{n}p(2, 1^{n-3}) + \frac{q(n; 2)}{\binom{n}{2}}P_{n-2}$ . Dividing by  $P_n = \kappa_n\kappa_{n-1}P_{n-2}$  gives  $H_n = \frac{n-2}{n}H_{n-1} + \frac{q(n; 2)}{\binom{n}{2}\kappa_n\kappa_{n-1}}$ , and the recursion in the proof of Lemma 2.3 yields  $q(n; 2)/\binom{n}{2} = \kappa_n\kappa_{n-1}t_n/(n\rho)$ .

*Step 2.* Dividing the definition of  $\varepsilon_n$  by  $P_n$ , with  $p(2, 2, 1^{n-5}) = E_{n-1}P_{n-1}$  and  $p(2, 1^{n-4}) = H_{n-2}P_{n-2}$ ,

$$\varepsilon_n/P_n = E_n - \frac{n-4}{n}E_{n-1} - \frac{2t_n}{n\rho}H_{n-2}.$$

*Step 3.* At the set partition  $\{1, 2\}, \{3, 4\}, \{5\}, \dots, \{n\}$  the compatible collisions are the merger of  $\{1, 2\}$  alone and of  $\{3, 4\}$  alone (rate  $t_n$  each, leading to the shape  $(2, 1^{n-3})$ ) and the simultaneous merger of both pairs (rate  $j_n$ , leading to  $1^{n-2}$ ). By Lemma 2.1,  $\Phi(n)p(2, 2, 1^{n-4}) = (n-4)\rho p(2, 2, 1^{n-5}) + 2t_n p(2, 1^{n-3}) + j_n P_{n-2}$ . Dividing by  $P_{n-1}$  and using  $\Phi(n)\kappa_n = n\rho$  and  $P_{n-2}/P_{n-1} = \Phi(n-1)/((n-1)\rho)$ ,

$$n\rho E_n = (n-4)\rho E_{n-1} + 2t_n H_{n-1} + \frac{j_n \Phi(n-1)}{(n-1)\rho}.$$

*Step 4.* Inserting Step 3 into Step 2,  $n\rho \varepsilon_n/P_n = j_n \Phi(n-1)/((n-1)\rho) - 2t_n(H_{n-2} - H_{n-1})$ . By Lemma 2.3 and (5),  $T(n-1) = T(n-2) + (n-2)t_{n-1}$ , and since  $\binom{n-1}{2} - (n-2) = \binom{n-2}{2}$ ,

$$\begin{aligned} (n-1)(n-2)(n-3)\rho(H_{n-2} - H_{n-1}) &= (n-1)T(n-2) - (n-3)T(n-1) \\ &= 2\left[T(n-2) - \binom{n-2}{2}t_{n-1}\right] = 2W(n-1). \end{aligned} \tag{9}$$

Hence

$$\frac{\varepsilon_n}{P_n} = \frac{\binom{n-2}{2}j_n \Phi(n-1) - 2t_n W(n-1)}{n(n-1)\rho^2 \binom{n-2}{2}},$$

and  $\binom{n-2}{2}j_n = 2X(n)/\binom{n}{2}$ ,  $2t_n = 2\text{Bin}(n)/\binom{n}{2}$  and  $n(n-1) = 2\binom{n}{2}$  give the formula. If the structure is regenerative, then  $\varepsilon_n = 0$  by Lemma 2.2, and  $P_n > 0$  gives  $\Delta_n = 0$ .  $\square$

The formula for  $\varepsilon_n$  was also checked in exact rational arithmetic for  $n \leq 8$ , on random consistent rate arrays and on rates of atomic  $\Xi$  (Section 8). For a  $\Lambda$ -coalescent  $X \equiv 0$ , and (2) says  $\text{Bin}(n)W(n-1) = 0$  for all  $n \geq 4$ , i.e. star-shaped ( $\text{Bin}(n) = 0$  for  $n \geq 3$ ) or Kingman ( $W \equiv 0$ ); both satisfy (2) for every  $n$ .

*Proof of Theorem A.* By Lemma 2.2,  $\Pi_2$  and  $\Pi_3$  are regenerative, and  $\Pi_4$  is regenerative if and only if  $(R_n)$  holds at  $(2, 2)$ , the only partition of 4 that is not a hook, i.e. if and only if  $\Delta_4 = 0$  (Lemma 3.1). Here  $X(4) = 3j_4$ ,  $\text{Bin}(4) = 6t_4$  and  $W(3) = \tau_3$  (the triple merger is the only collision of three lineages that is not a single pair), so  $\Delta_4 = 3(j_4\Phi(3) - 2t_4\tau_3)$ . As  $\Phi(3) = 3\rho + 3t_3 + \tau_3$ , the condition  $j_4\Phi(3) = 2t_4\tau_3$  is affine in  $\rho$  with slope  $3j_4$ . If  $\Xi$  is not a  $\Lambda$ -coalescent, then  $j_4 > 0$  by (7), and the condition holds for at most one  $\rho$ , namely  $\rho^*(\Xi)$ .  $\square$

**Corollary 3.2** (Möhle [Mö07]). *Let  $\Xi \neq 0$  be a  $\Lambda$ -coalescent and  $\rho > 0$ . The partition structure is regenerative if and only if  $\Lambda = c\delta_0$  or  $\Lambda = c\delta_1$ ; otherwise  $\Pi_4$  is not regenerative.*

*Proof.* Here  $j_4 = 0$ , so by Theorem A  $\Pi_4$  is regenerative if and only if  $t_4\tau_3 = 0$ . By the  $\Lambda$ -coalescent rates of Section 1,  $\tau_3 = \int y \Lambda(dy)$  and  $t_4 = \int (1-y)^2 \Lambda(dy)$ , which vanish if and only if  $\Lambda = c\delta_0$ , resp.  $\Lambda = c\delta_1$ . These two families are regenerative (Section 1).  $\square$

## 4 Three estimates for paintboxes

Let  $\lambda_m(x) := \mathbb{E}_x Z_m = \binom{m}{2} p_2(x)$  and  $M := \mu(\nabla \setminus \{0\}) \in (0, \infty]$  for  $\Xi_0 \neq 0$ . We use three elementary facts.

- (Q1)  $Z_m$  is nondecreasing in  $m$  (add draws), so  $F(x; m)$ ,  $G(x; m)$  and  $W(m)$  are nondecreasing in  $m$ , and  $W(m) \geq W(3) = \tau_3 > 0$  for  $m \geq 3$  if  $\Xi_0 \neq 0$ .
- (Q2)  $R(m)/\binom{m}{2} \leq \int \min\{1, 1/\lambda_m(x)\} \Xi_0(dx) \rightarrow 0$  as  $m \rightarrow \infty$ , by dominated convergence ( $\Xi_0$  is finite and  $p_2 > 0$  on  $\nabla \setminus \{0\}$ ); moreover  $Y(m), W(m) \leq R(m)$ .
- (Q3) For  $x \neq 0$ , box 1 receives at least three of  $m$  draws with probability tending to 1; hence  $\pi_1(x; m), \pi_2(x; m) \rightarrow 0$  and  $F(x; m), G(x; m) \rightarrow 1$  as  $m \rightarrow \infty$ .

**Lemma 4.1.**  $\pi_2(x; m) \leq \lambda_m(x)^2/2$ .

*Proof.* On  $\{Z_m = 2\}$  the two coincident pairs are disjoint, since three draws in one box give three pairs. Coincidences of disjoint pairs of draws are independent events of probability  $p_2(x)$  each, and there are  $\binom{m}{2} \binom{m-2}{2}/2 \leq \binom{m}{2}^2/2$  unordered pairs of disjoint pairs of draws. The expected number of unordered pairs of disjoint coincident pairs is therefore at most  $\lambda_m(x)^2/2$ , and it bounds  $\pi_2(x; m)$ .  $\square$

**Lemma 4.2.** For  $m \geq 5$ ,  $\pi_2(x; m) \leq \frac{m}{m-4} \pi_2(x; m-1)$ .

*Proof.* On  $\{Z_m = 2\}$  exactly four draws lie in coincident pairs. The draws are exchangeable, so the probability of  $\{Z_m = 2\}$  and that draw  $i$  lies in no coincident pair is the same for all  $i$ ; summing over  $i$ , it equals  $\frac{m-4}{m} \pi_2(x; m)$ . For  $i = m$ , the first  $m-1$  draws then have  $Z_{m-1} = 2$ .  $\square$

**Lemma 4.3.** For every  $\eta \in (0, 1]$  there is  $c(\eta) > 0$  such that  $G(x; m) \geq c(\eta)$  for all  $x \in \nabla \setminus \{0\}$  and  $m \geq 3$  with  $\lambda_m(x) \geq \eta$ .

*Proof.* Let  $y := x_1 = \max_i x_i$ ; then  $p_2(x) \leq y \sum_i x_i \leq y$ . Let  $m_0 := \min\{m' \geq 3 : \lambda_{m'}(x) \geq \eta\} \leq m$ ; by (Q1),  $G(x; m) \geq G(x; m_0)$ , so we may assume  $m = m_0$ .

Case  $m = 3$ . Then  $y \geq p_2 \geq \eta/3$  and  $G(x; 3) = \sum_i x_i^3 \geq y^3 \geq (\eta/3)^3$ .

Case  $m \geq 4$ . Then  $\lambda_{m-1} < \eta$  and  $\lambda := \lambda_m = \frac{m}{m-2}\lambda_{m-1} \in [\eta, 2\eta)$ . Fix  $\theta \in (0, \frac{1}{2}]$ , to be chosen depending on  $\eta$  only.

Subcase  $m\eta \geq \theta$ . Box 1 receives a binomial( $m, y$ ) number of draws, which is stochastically larger than a binomial( $m, \theta/m$ ) variable  $\beta$ , so  $G(x; m) \geq \mathbb{P}(\beta \geq 3) \geq \binom{m}{3}(\theta/m)^3(1-\theta/m)^m \geq \frac{\theta^3}{16}(1-\theta) \geq \frac{\theta^3}{32}$ , since  $\binom{m}{3}/m^3 = \frac{(m-1)(m-2)}{6m^2} \geq \frac{1}{16}$  for  $m \geq 4$ .

Subcase  $m\eta < \theta$ . Then  $\eta \leq \lambda \leq \binom{m}{2}y < (m-1)\theta/2$ , so  $m > 2\eta/\theta$ . Let  $N$  be a Poisson( $m/2$ ) variable independent of an i.i.d. sequence of draws, and consider the first  $N$  draws. The numbers of these draws in the boxes are independent Poisson( $\nu_k$ ) variables,  $\nu_k = mx_k/2 < \theta/2 \leq \frac{1}{4}$ . Let  $B$  be the number of boxes receiving at least two of them; then  $Z_N \geq B$ , and  $B$  is a sum of independent Bernoulli( $\pi_k$ ) variables with  $e^{-1/4}\nu_k^2/2 \leq \pi_k \leq \nu_k^2/2 < \theta^2/8$ . Hence  $s := \sum_k \pi_k$  satisfies  $e^{-1/4}m^2p_2/8 \leq s \leq m^2p_2/8$ , and  $m^2p_2/8 = \lambda m/(4(m-1)) \in [\lambda/4, \lambda/3]$ ; so  $s_0 := e^{-1/4}\eta/4 \leq s < 2\eta/3$ . Le Cam's inequality [LC60] (applied to finite subsums, then letting their number tend to infinity) bounds the total variation distance between  $B$  and a Poisson( $s$ ) variable by  $\sum_k \pi_k^2 \leq \frac{\theta^2}{8}s \leq \frac{\theta^2\eta}{12}$ . With  $p_0 := \mathbb{P}(\text{Poisson}(s_0) \geq 3) > 0$ ,  $\mathbb{P}(B \geq 3) \geq p_0 - \theta^2\eta/12$ . On  $\{N \leq m\}$  we have  $Z_m \geq Z_N$ , and  $\mathbb{P}(N \geq m) \leq (e/4)^{m/2} < (e/4)^{\eta/\theta}$  by the Chernoff bound  $\mathbb{P}(\text{Poisson}(\nu) \geq 2\nu) \leq (e/4)^\nu$ . Thus  $G(x; m) \geq p_0 - \theta^2\eta/12 - (e/4)^{\eta/\theta}$ . Choose  $\theta \in (0, \frac{1}{2}]$  with  $\theta^2\eta/12 \leq p_0/2$  and  $(e/4)^{\eta/\theta} \leq p_0/4$ ; then  $G(x; m) \geq p_0/4$ .

So  $c(\eta) := \min\{(\eta/3)^3, \theta^3/32, p_0/4\}$  works.  $\square$

## 5 A Kingman component: proof of Theorem B

The hypothesis of Theorem B does not depend on the parametrisation:  $a = \lim_n t_n$ , since  $t_n - a = \int \mathbb{P}_x(\Pi_n^x = \{1, 2\} \text{ and singletons}) \mu(dx)$  has an integrand at most  $\pi_1(x; n)$  and at most  $p_2(x)$ , so it tends to 0 by (Q3) and dominated convergence.

**Proposition 5.1.** *For every finite measure  $\Xi_0 \neq 0$  on  $\nabla \setminus \{0\}$  there are  $\gamma < 1$  and infinitely many  $m \geq 4$  with  $X(m) \leq \gamma W(m)$ . Consequently  $\liminf_{n \rightarrow \infty} X(n)/W(n-1) \leq \gamma$ .*

*Proof.* Let  $c := c(1/4)$  be the constant of Lemma 4.3 and  $\gamma := 3/(3+2c)$ . For  $j \geq 0$  let  $B_j := \{x : 2^{-j-1} \leq p_2(x) < 2^{-j}\}$ , with the points  $p_2(x) = 1$  added to  $B_0$ , and  $h(j) := \Xi_0(B_j)$ ; then  $\sum_j h(j) = \Xi_0(\nabla) < \infty$ . Let  $u_m := 1/\binom{m}{2}$ , so that  $\lambda_m(x) = p_2(x)/u_m$ .

(a) *Good scales.* Call  $J$  good if  $\sum_{i \geq 1} 2^{-i}h(J+i) \leq h(J)$ . For every  $J_0$  there is a good  $J \geq J_0$ . Otherwise  $\sum_{i \geq 1} 2^{-i}h(J+i) > h(J)$  for all  $J \geq J_0$ ; summing these inequalities between convergent series gives  $\sum_{J \geq J_0} \sum_{i \geq 1} 2^{-i}h(J+i) > \sum_{J \geq J_0} h(J)$ , whereas the left side equals  $\sum_{j > J_0} h(j) \sum_{i=1}^{j-J_0} 2^{-i} \leq \sum_{j > J_0} h(j)$ , a contradiction.

(b) *The matching  $m$ .* For  $J \geq 3$  let  $m = m(J)$  be the largest  $m \geq 4$  with  $u_m \geq 2^{-J}$  (it exists, as  $u_4 = \frac{1}{6}$ ). Since  $u_{m+1}/u_m = \frac{m-1}{m+1} \geq \frac{1}{2}$ , we have  $u_m \in [2^{-J}, 2^{-J+1})$ , and  $m(J) \rightarrow \infty$  as  $J \rightarrow \infty$ .

(c) *Splitting.* Let  $L := \{x : \lambda_m(x) \geq \frac{1}{4}\} = \{p_2 \geq u_m/4\}$  and  $S := (\nabla \setminus \{0\}) \setminus L$ . As  $u_m/4 \in [2^{-J-2}, 2^{-J-1})$ ,  $L \supseteq B_J$  and  $S \subseteq \{p_2 < 2^{-J-1}\} = \bigcup_{i \geq 1} B_{J+i}$ . On  $L$ , Lemma 4.3 gives  $\pi_2 \leq 1 \leq G/c$ , so  $\int_L \pi_2 d\mu \leq c^{-1} \int_L G d\mu$  ( $\mu(L) \leq 4\Xi_0(\nabla)/u_m$  is finite), and

$$\int_L G d\mu \geq c \mu(B_J) \geq c 2^J h(J) \geq c h(J)/u_m.$$

On  $S$ , Lemma 4.1 gives

$$\int_S \pi_2 d\mu \leq \frac{1}{2u_m^2} \int_S p_2 d\Xi_0 \leq \frac{1}{2u_m^2} \sum_{i \geq 1} 2^{-J-i} h(J+i) \leq \frac{1}{2u_m} \sum_{i \geq 1} 2^{-i} h(J+i),$$

using  $2^{-J} \leq u_m$ .

(d) *Conclusion.* If  $J$  is good, then  $\int_S \pi_2 d\mu \leq h(J)/(2u_m) \leq (2c)^{-1} \int_L G d\mu$ . (If  $h(J) = 0$ , goodness forces  $h(J+i) = 0$  for all  $i$ , so  $\Xi_0(S) = 0$  and the bound holds trivially.) Hence  $X(m) = \int \pi_2 d\mu \leq \frac{3}{2c} \int G d\mu = \frac{3}{2c}(W(m) - X(m))$  by (4), i.e.  $X(m) \leq \gamma W(m)$ . By (a) and (b) this holds for infinitely many  $m$ . For each of them, Lemma 4.2 gives  $X(m+1) \leq \frac{m+1}{m-3} X(m) \leq \frac{m+1}{m-3} \gamma W(m)$ , so  $\liminf_n X(n)/W(n-1) \leq \gamma$ .  $\square$

*Proof of Theorem B.* Suppose the structure is regenerative. For  $n \geq 4$ ,  $W(n-1) \geq \tau_3 > 0$  by (Q1), and (2) gives

$$\frac{X(n)}{W(n-1)} = \frac{\text{Bin}(n)}{\Phi(n-1)} = \frac{1 + Y(n)/(a\binom{n}{2})}{\frac{n-2}{n} + \frac{2\rho}{an} + R(n-1)/(a\binom{n}{2})} \rightarrow 1$$

by (Q2), since  $\text{Bin}(n) = a\binom{n}{2} + Y(n)$  and  $\Phi(n-1) = (n-1)\rho + a\binom{n-1}{2} + R(n-1)$  by (4). This contradicts Proposition 5.1.  $\square$

If  $M < \infty$ , Proposition 5.1 is immediate:  $X(n) \rightarrow 0$  by (Q3) and dominated convergence, while  $W(n-1) \rightarrow M$ . In general only the  $\liminf$  can be bounded (Section 8).

## 6 No Kingman component: proofs of Theorems C and D

For  $a = 0$ , (2) reads  $[(n-1)\rho + R(n-1)]X(n) = Y(n)W(n-1)$ , and both sides may tend to 0; the normalisation of the proof of Theorem B is not available. A *box* of  $x$  is a positive coordinate.

*Proof of Theorem C.* Since  $\Xi_0$  is not a multiple of  $\delta_{(1,0,\dots)}$ , the largest number  $K$  of boxes carried by a set of positive  $\Xi_0$ -measure satisfies  $2 \leq K \leq K_{\max}$ , and  $\Xi_0(\{x \text{ has exactly } K \text{ boxes}\}) > 0$ . For a point  $x$  without dust and with at most  $K$  boxes,  $Z_{K+2} = 1$  would require one box with exactly two of the  $K+2$  draws and all other draws in different boxes, i.e. at least  $K+1$  boxes; so  $\text{Bin}(K+2) = Y(K+2) = 0$ . For  $x$  with exactly  $K$  boxes,  $\pi_2(x; K+2) \geq x_1^2 x_2^2 x_3 \cdots x_K > 0$  (draws 1,2 in box 1, draws 3,4 in box 2, draw  $4+i$  in box  $2+i$ ). Hence  $X(K+2) > 0$  and  $\Delta_{K+2} = \Phi(K+1)X(K+2) > 0$ , so (2) fails at  $n = K+2$ .  $\square$

For example, for  $\Xi = \delta_{(1/3, 1/3, 1/3, 0, \dots)}$  one has  $K = 3$ ,  $\text{Bin}(5) = 0$  and  $X(5) = \frac{10}{9}$ . For  $K = 2$  one has  $t_4 = 0$  (four draws into two boxes give at least two pairs), so  $\Pi_4$  is already not regenerative by Theorem A.

For Theorem D we need the rates of one paintbox  $x$  with  $K < \infty$  boxes and dust  $d = 1 - \sum_i x_i$ . For a set  $\mathcal{B}$  of boxes let  $S_r(\mathcal{B})$  be the probability that each of  $r$  draws falls into the dust or into a box of  $\mathcal{B}$ , no two into the same box:  $S_r(\mathcal{B}) = \sum_{k=0}^{\min(r, |\mathcal{B}|)} \binom{r}{k} k! e_k(\mathcal{B}) d^{r-k}$ , with  $e_k$  the elementary symmetric polynomial of the box sizes. Then

$$\pi_1(x; n) = \binom{n}{2} \sum_i x_i^2 S_{n-2}(\mathcal{B}_i), \quad \pi_2(x; n) = \frac{1}{2} \binom{n}{2} \binom{n-2}{2} \sum_{i \neq k} x_i^2 x_k^2 S_{n-4}(\mathcal{B}_{ik}), \quad (10)$$

where  $\mathcal{B}_i, \mathcal{B}_{ik}$  are the sets of boxes other than  $i$ , resp. other than  $i, k$ .

**Lemma 6.1.** (i) If  $d = 0$ , then  $\pi_1(x; n) = 0$  for  $n \geq K+2$  and  $\pi_2(x; n) = 0$  for  $n \geq K+3$ . (ii) If  $d > 0$ , then for  $n \geq K+2$  one has  $\pi_1(x; n) = d^n P_x(n)$  and  $\pi_2(x; n) = d^n Q_x(n)$  with polynomials  $P_x$  of degree  $K+1$  and, if  $K \geq 2$ ,  $Q_x$  of degree  $K+2$ , both with positive leading coefficients;  $Q_x = 0$  if  $K = 1$ .

*Proof.* If  $d = 0$ , then  $S_r(\mathcal{B}) = 0$  for  $r > |\mathcal{B}|$ . If  $d > 0$  and  $r \geq |\mathcal{B}|$ , then  $d^{-r} S_r(\mathcal{B}) = \sum_{k=0}^{|\mathcal{B}|} r(r-1) \cdots (r-k+1) e_k(\mathcal{B}) d^{-k}$  is a polynomial in  $r$  of degree  $|\mathcal{B}|$  with leading coefficient  $e_{|\mathcal{B}|}(\mathcal{B}) d^{-|\mathcal{B}|} > 0$ . Insert this into (10), with  $|\mathcal{B}_i| = K-1$  and  $|\mathcal{B}_{ik}| = K-2$ .  $\square$

*Proof of Theorem D.* Here  $M = \sum_l c_l/p_2(x^l) < \infty$ , so  $R(n), W(n) \leq M$ , and  $W(n) \geq \tau_3 > 0$  for  $n \geq 3$ . Let  $K_l$  and  $d_l$  be the number of boxes and the dust of  $x^l$ . If all  $d_l = 0$ , Theorem C applies (some  $x^l$  has at least two boxes, as  $\Xi_0$  is not a multiple of  $\delta_{(1,0,\dots)}$ ). Otherwise let  $d^* := \max_l d_l > 0$  and  $K^* := \max\{K_l : d_l = d^*\}$ . All terms below are nonnegative, atoms with  $d_l < d^*$  contribute  $O(n^A d_l^n)$  for some  $A$ , and atoms with  $d_l = 0$  contribute nothing for large  $n$  (Lemma 6.1).

*Case  $K^* \geq 2$ .* By Lemma 6.1,  $Y(n) \leq C_1 n^{K^*+1} d^{*n}$  and  $X(n) \geq c_1 n^{K^*+2} d^{*n}$  for large  $n$ , with constants  $C_1, c_1 > 0$ . Then  $\Phi(n-1)X(n) \geq (n-1)\rho c_1 n^{K^*+2} d^{*n}$  exceeds  $MC_1 n^{K^*+1} d^{*n} \geq \text{Bin}(n)W(n-1)$  for large  $n$ .

*Case  $K^* = 1$ .* Every atom with dust  $d^*$  has one box and contributes nothing to  $X$ ; the atoms with at least two boxes have  $d_l < d^*$ . Hence  $X(n) = O(n^A d'^n)$  with  $d' < d^*$ , or  $X(n) = 0$  for large  $n$ , while  $Y(n) \geq c_2 n^2 d^{*n}$  with  $c_2 > 0$ , by Lemma 6.1. Then  $\Phi(n-1)X(n) \leq ((n-1)\rho + M)X(n) < c_2 \tau_3 n^2 d^{*n} \leq \text{Bin}(n)W(n-1)$  for large  $n$ .

In both cases  $\Delta_n \neq 0$  for all large  $n$ , and Lemma 3.1 applies.  $\square$

**Corollary E.** *Let  $a = 0$  and  $M < \infty$ . If the partition structure is regenerative, then either  $\text{Bin}(n) = X(n) = 0$  for all large  $n$ , or  $n\rho X(n)/(M\text{Bin}(n)) \rightarrow 1$ ; in the latter case  $\rho = M \lim_n \text{Bin}(n)/(nX(n))$  is determined by  $\Xi$ , in addition to the value  $\rho = \rho^*(\Xi)$  required by Theorem A when  $\Xi$  is not a  $\Lambda$ -coalescent.*

*Proof.* By (Q3) and dominated convergence ( $\mu$  is finite),  $R(n-1) \rightarrow M$  and  $W(n-1) \rightarrow M$ . If  $\text{Bin}(n_0) = 0$  for some  $n_0$ , then  $t_n = 0$  for  $n \geq n_0$  ( $t_n$  is nonincreasing), and (2) forces  $X(n) = 0$ . Otherwise (2) gives  $n\rho X(n)/(M\text{Bin}(n)) = \frac{n\rho}{(n-1)\rho + R(n-1)} \cdot \frac{W(n-1)}{M} \rightarrow 1$ .  $\square$

## 7 Finitely many levels do not decide

**Proposition 7.1.** *Let  $\Xi = \frac{17}{1024}\delta_0 + \delta_{(1/4,1/8,0,\dots)}$  and  $\rho = \frac{13833}{5120}$ . Then  $\rho = \rho^*(\Xi)$ , and  $\Pi_n$  is regenerative for every  $n \leq 5$ , with*

$$q(5; \cdot) = \left( \frac{23055}{33517}, \frac{5300}{33517}, \frac{9805}{100551}, \frac{16183550}{367245769}, \frac{1245097}{100157937} \right),$$

*while  $\Pi_6$  is not regenerative:  $(R_n)$  fails at  $n = 6$  exactly at the partitions  $(4, 2)$ ,  $(3, 3)$ ,  $(3, 2, 1)$ ,  $(2, 2, 2)$  and  $(2, 2, 1, 1)$ . Moreover  $\Delta_4 = \Delta_5 = 0$ ,  $\Delta_n > 0$  for  $6 \leq n \leq 15$  and  $\Delta_n < 0$  for  $16 \leq n \leq 60$ .*

*Proof.* A finite computation in rational arithmetic: the rates from (1), the probabilities  $p$  from Lemma 2.1,  $q(n; \cdot)$  from Lemma 2.2 and the residuals of  $(R_n)$  at all partitions of  $n \leq 7$ , and  $\Delta_n$  from (10). The program of this note, that of the first version and those of the three verification runs give the same values (Section 8).  $\square$

Since  $a > 0$ , Theorem B predicts failure at some level; indeed  $\text{Bin}(n)W(n-1)/\binom{n}{2} \rightarrow aM = \frac{17}{80}$  and  $\Phi(n-1)X(n)/\binom{n}{2} \rightarrow 0$ . The pair  $(a, \rho)$  was found by solving exactly the condition of Theorem A and one more equation at  $n = 5$ .

*Remark 7.2* (a heuristic count). At level  $n$  the unknowns are  $\rho$  and the rates of the  $p(n) - 1$  collision types of  $n$  lineages ( $p(n)$  is the number of partitions of  $n$ ; the lower levels follow by consistency), i.e.  $p(n) - 1$  parameters up to scale, while  $(R_n)$  adds  $p(n) - n$  equations, one for each partition that is not a hook. Ignoring that the rates must come from a measure  $\Xi$ , one therefore expects solutions at every single level. This count is heuristic only and is not used; it suggests that any proof has to use the behaviour as  $n \rightarrow \infty$ , as the proofs of Theorems B to D do.

*The open case.* Let  $a = 0$ , let  $\Xi_0$  charge  $\{x_2 > 0\}$  and lie outside the classes of Theorems C and D, and let  $\rho^*(\Xi) > 0$ . Regenerativity at  $\rho^*(\Xi)$  requires (2) for all  $n$  and, if  $M < \infty$ , Corollary E; (2) can be balanced asymptotically by mixing paintboxes with one box, which contribute to Bin but not to  $X$ , with paintboxes with several boxes. The other regeneration equations, for instance at  $(3, 2, 1^{n-5})$  or  $(2, 2, 2, 1^{n-6})$ , give further necessary conditions of the same kind; we did not find a combination that decides this case.

## 8 Computations and reproducibility

Only Proposition 7.1, a finite exact computation, depends on a computation. All programs use exact rational arithmetic unless stated otherwise.

*The exact program of this note*, `check_note.py` (standard library only; about 8 seconds; 55 checks, all passed), tests the rates (1) of three atomic measures against a brute-force enumeration of the draws, the consistency relation, (5), (6), (4) and (10); Lemma 2.1 against the Ewens formula, and  $(R_n)$  for Kingman and star-shaped coalescents ( $n \leq 7$ ); Lemmas 2.2 and 2.3, (9) and Lemma 3.1 for  $4 \leq n \leq 8$  on random consistent arrays and atomic measures; Theorem A and Corollary 3.2 on seven measures; Proposition 7.1 as printed (and failure at 8 partitions of 7); the example after the proof of Theorem C; the eventual signs of  $\Delta_n$  predicted by the proofs of Theorems B and D for nine measures ( $40 \leq n \leq 60$ ); Lemmas 4.1 and 4.2 and (Q1) on random paintboxes; and the elementary constants in the proofs of Lemma 4.3 and Proposition 5.1.

*The programs of the first version* (`original/`) check Lemma 3.1 for  $n \leq 7$  and the forms (4) for  $m \leq 8$  (`verify_Dn.py`), the certificate of Proposition 7.1 (`certify_levels.py`, with `sympy`) and Lemmas 4.1 and 4.2 on random paintboxes (`check_lemmas.py`). The programs of the two verification runs of the first version are in `verification_run_A/` and `verification_run_B/`, and those of the further run on the text of this note in `independent_run_2/`. All recorded outputs were reproduced from the package (command lines and running times in its `README.md`).

**Verification.** The results were first obtained with written proofs and the programs in `original/`. Two independent verification runs, both AI-assisted, examined this first written version; each read the source [Mö06a] in the publisher’s open PDF, re-derived all proofs line by line and wrote its own programs, and neither read the other’s report. Both confirmed Lemma 3.1, Theorems A to D, Corollary E and Proposition 7.1, found no mathematical error, and classified the result as a partial answer to Question 1.1.

The first run also read [GP05b, Don07, DGP07, Lag07] and compared the statements used here with them. Its programs recompute the certificate with rates obtained by enumerating all  $3^b$  assignments of  $b$  draws to the two boxes and the dust, and solve at each level the full linear system  $(R_n)$  in the unknowns  $q(n; \cdot)$ : unique solutions with  $q \geq 0$  for  $n \leq 5$ , equal to those of Proposition 7.1, and no solution for  $n = 6$ . They check Lemma 3.1 for  $n \leq 8$  on random consistent arrays and determine the sign of  $\Delta_n$  for  $n \leq 60$  from closed forms of their own.

The second run wrote its programs from the definitions before reading those of the first version: rates by brute force over all assignments of the draws, the probabilities  $p$  by first-step analysis over set partitions. Kingman’s coalescent gave the Ewens formula, and Kingman and star-shaped coalescents satisfied  $(R_n)$  for  $n \leq 8$ ; Theorem A held in 56 random trials, 24 of them at  $\rho = \rho^*$ ; Lemma 3.1 held exactly in 51 cases with the rates of atomic measures and in 51 with arbitrary consistent arrays ( $n \leq 8$ ); Proposition 7.1 was reproduced, with failure also at  $n = 7$ ; deliberately wrong inputs (an inconsistent array,  $W(n)$  in place of  $W(n-1)$ , a shifted  $\rho$ ) were detected; and the predicted eventual sign of  $\Delta_n$  held for  $n \leq 60$  in 10 examples of Theorem D and 8 of Theorem B. A numerical test of Proposition 5.1 for a measure with very sparse scales gave values of  $X(m+1)/W(m)$  from below

0.01 to 0.96 for  $10^3 \leq m \leq 3 \cdot 10^4$ , so only the  $\liminf$  can be bounded in general. The run re-ran the programs of the first version, reproduced their recorded outputs, and checked the bibliography against Crossref, OpenAlex and zbMATH.

The runs required textual corrections, all of which were made: the signs of  $\Delta_n$  in Proposition 7.1 for every  $n \leq 60$  (the first version had tested five values  $n \geq 16$  and omitted  $13 \leq n \leq 15$ ); the nonnegativity statement of Lemma 2.2; powers of 2 for the scales and the case  $h(J) = 0$  in the proof of Proposition 5.1; the exclusion of  $\Xi = 0$ ; the check of Lemma 3.1 for  $n \leq 8$ ; the heuristic status of Remark 7.2; one year for [Mö06a]; and the sources not read. Apart from Proposition 7.1, no proof depends on a computation.

A further independent verification run, also AI-assisted, then examined the text of this note as written. It read [Mö06a, GP05b, Don07, DGP07] again, re-derived every proof (in particular Lemmas 2.2 and 3.1 for all  $n$ , Lemma 4.3 with its constants, Proposition 5.1 including the case  $h(J) = 0$ , and Theorems A to D), and wrote its own exact programs with methods different from the earlier ones: the rates of atomic  $\Xi$  by a recursion over the boxes, the probabilities  $p$  by a first-step recursion over integer partitions, and the full linear system  $(R_n)$  solved in  $q(n; \cdot)$ . All 57 of its checks passed, among them Lemma 3.1 in 36 cases with  $4 \leq n \leq 8$ , Theorem A in 27 trials, and Proposition 7.1 with the printed vector  $q(5; \cdot)$ , the five partitions of 6 and the signs of  $\Delta_n$  for  $n \leq 60$ . It re-ran every program of the package from the source archive and obtained the recorded outputs, checked all DOIs against Crossref, and repeated the literature search. It found no mathematical error; its textual corrections (this paragraph, the final form of the sentence on the sources not read, the record of the searches, the description of the open case in Section 1 and one bibliography note) were made.

**Scope and priority.** The note gives a partial answer to Question 1.1: the conjectured answer holds for every  $\Xi$  with a Kingman component (Theorem B), for the classes of Theorems C and D, and for every other  $\Xi$  that is not a  $\Lambda$ -coalescent at every  $\rho > 0$  except possibly  $\rho^*(\Xi)$  (Theorem A); the case  $a = 0$ ,  $\Xi_0$  charging  $\{x_2 > 0\}$  and outside the classes of Theorems C and D, at  $\rho = \rho^*(\Xi)$ , remains open. The conjecture and the  $\Lambda$  case are due to Möhle [Mö06a, Mö07]; Corollary 3.2 recovers his theorem, and its proof was not compared with his. The recursion of Lemma 2.1 is Möhle’s [Mö06b, Theorem 5.1], used in the form [Don07, (12)], see also [DGP07], and re-derived here; regenerative partition and composition structures, with [GP05b, (4)–(6)] behind Lemma 2.2, are due to Gneden and Pitman [GP05b, GP05a];  $\Xi$ -coalescents are due to Möhle and Sagitov [MS01] and Schweinsberg [Sch00]. Lemma 3.1, Theorems A to D (beyond the  $\Lambda$  case), Corollary E and Proposition 7.1 were not found in print.

What was read: [Mö06a] in the publisher’s open PDF (Möhle’s abstract, pp. 2279–2282); [GP05b] in the journal’s PDF; [Don07], [DGP07] and [Lag07] in their arXiv versions. The full texts of [Mö07] and [Mö06b] were not accessible to us and were not read: [Mö07] is used only through its abstract (on the publisher’s page, in Crossref and in zbMATH) and the statements of its theorem in [Mö06a, DGP07, Lag07], and [Mö06b] only through the restatement of its Theorem 5.1 in [Don07, Theorem 3]. Of [GIM08, GIM11] the abstracts were read, and [Gne10] was searched for the relevant terms. No printed answer to the question was found in the sources searched: the papers named in this paragraph; the works citing [Mö07] listed by Semantic Scholar, OpenAlex, Crossref and zbMATH, the works citing [Mö06b] listed by OpenAlex and those citing [GP05b, DGP07] listed by Semantic Scholar (titles, and abstracts where available); queries to the arXiv, zbMATH, Crossref and OpenAlex interfaces, and four web searches, made on 18 and 19 October 2026 (the last round of them by the run on the text of this note). A search that finds nothing is not a proof of novelty, and no priority is claimed.

AI-assisted tools were used to search the problem corpus and the literature, to find the arguments, to write and run the verification programs, and to help prepare this manuscript. The author has

reviewed the proofs, the computations and the sources and is responsible for all claims. The note has not been peer reviewed.

## References

- [DGP07] Rui Dong, Alexander Gnedin, and Jim Pitman. Exchangeable partitions derived from Markovian coalescents. *Annals of Applied Probability*, 17(4):1172–1201, 2007. doi:10.1214/105051607000000069.
- [Don07] Rui Dong. Exchangeable partitions derived from Markovian coalescents with simultaneous multiple collisions. Preprint, arXiv:0707.1606 [math.PR], 2007. URL: <https://arxiv.org/abs/0707.1606>.
- [Ewe72] W. J. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3(1):87–112, 1972. doi:10.1016/0040-5809(72)90035-4.
- [GIM08] Alexander Gnedin, Alexander Iksanov, and Martin Möhle. On asymptotics of exchangeable coalescents with multiple collisions. *Journal of Applied Probability*, 45(4):1186–1195, 2008. doi:10.1239/jap/1231340242.
- [GIM11] Alexander Gnedin, Alexander Iksanov, and Alexander Marynych. On  $\Lambda$ -coalescents with dust component. *Journal of Applied Probability*, 48(4):1133–1151, 2011. doi:10.1239/jap/1324046023.
- [Gne10] Alexander Gnedin. Regeneration in random combinatorial structures. *Probability Surveys*, 7:105–156, 2010. doi:10.1214/10-PS163.
- [GP05a] Alexander Gnedin and Jim Pitman. Regenerative composition structures. *Annals of Probability*, 33(2):445–479, 2005. doi:10.1214/009117904000000801.
- [GP05b] Alexander Gnedin and Jim Pitman. Regenerative partition structures. *Electronic Journal of Combinatorics*, 11(2):Research Paper R12, 2005. doi:10.37236/1869.
- [Kin82] J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248, 1982. doi:10.1016/0304-4149(82)90011-4.
- [Lag07] Andreas Nordvall Lagerås. A population model for  $\Lambda$ -coalescents with neutral mutations. *Electronic Communications in Probability*, 12:9–20, 2007. doi:10.1214/ECP.v12-1245.
- [LC60] Lucien Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10(4):1181–1197, 1960. doi:10.2140/pjm.1960.10.1181.
- [MS01] Martin Möhle and Serik Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Annals of Probability*, 29(4):1547–1562, 2001. doi:10.1214/aop/1015345761.
- [Mö06a] Martin Möhle. Coalescent theory – simultaneous multiple collisions and sampling distributions. In *Mathematical Population Genetics (Report No. 40/2005)*, organised by E. Baake, W. J. Ewens and A. Wakolbinger, volume 2 of *Oberwolfach Reports*, pages 2279–2282. European Mathematical Society, 2006. Issue 3 (the whole report: pp. 2241–2304). doi:10.4171/OWR/2005/40.

- [Mö06b] Martin Möhle. On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli*, 12(1):35–53, 2006. zbMATH 1099.92052.
- [Mö07] Martin Möhle. On a class of non-regenerative sampling distributions. *Combinatorics, Probability and Computing*, 16(3):435–444, 2007. doi:10.1017/S0963548306008212.
- [Pit99] Jim Pitman. Coalescents with multiple collisions. *Annals of Probability*, 27(4):1870–1902, 1999. doi:10.1214/aop/1022874819.
- [Sag99] Serik Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *Journal of Applied Probability*, 36(4):1116–1125, 1999. doi:10.1239/jap/1032374759.
- [Sch00] Jason Schweinsberg. Coalescents with simultaneous multiple collisions. *Electronic Journal of Probability*, 5:1–50, 2000. doi:10.1214/EJP.v5-68.
- [ula26] ulamai. UnsolvedMath. Hugging Face dataset, <https://huggingface.co/datasets/ulamai/UnsolvedMath>, 2026. Version 1.8.0; record OWR-1061-006.