J. Comb. Algebra 6 (2022), 223–263
DOI 10.4171/JCA/63

Density of random subsets and applications to group theory

Tsung-Hsuan Tsai

Abstract. Developing an idea of M. Gromov (1993), we study the intersection formula for random subsets with density. The *density* of a subset A in a finite set E is defined by $\operatorname{dens} A := \log_{|E|}(|A|)$. The aim of this article is to give a precise meaning of Gromov's *intersection formula*: "Random subsets" A and B of a finite set E satisfy $\operatorname{dens}(A \cap B) = \operatorname{dens} A + \operatorname{dens} B - 1$.

As an application, we exhibit a phase transition phenomenon for random presentations of groups at density $\lambda/2$ for any $0 < \lambda < 1$, characterizing the $C'(\lambda)$-small cancellation condition. We also improve an important result of random groups by G. Arzhantseva and A. Ol'shanskii (1996) from density 0 to density $0 \leq d < 1/(120m^2 \ln(2m))$.

Introduction

Density of subsets. Let A be a subset of a finite set E. Denote $|E|$, $|A|$ their cardinalities. In [9, Section 9.A], M. Gromov defined the *density* of A in E as

$$\operatorname{dens}_E(A) := \log_{|E|}\big(|A|\big).$$

Namely, $\operatorname{dens}_E(A)$ is the number $d \in \{-\infty\} \cup [0, 1]$ such that $|A| = |E|^d$. Note that $d = -\infty$ if and only if $A = \emptyset$. If the set E is fixed, we omit the subscript and simply denote the density by $\operatorname{dens} A$.

In [9, p. 270], the *intersection formula* is stated as follows: *Random subsets A and B of a finite set E satisfy*

$$\operatorname{dens}(A \cap B) = \operatorname{dens} A + \operatorname{dens} B - 1$$

with the convention

$$\operatorname{dens} A < 0 \iff A = \emptyset.$$

If E is a finite-dimensional vector space over a finite field, every affine subspace A satisfies $\operatorname{dens} A = \dim A / \dim E$. The intersection formula is then a "random subset

---

2020 Mathematics Subject Classification. Primary 60C05; Secondary 20F05, 20F06.
Keywords. Random group, intersection formula, cancellation theory.

version" of the well-known result for affine subspaces: *Transversal affine subspaces A and B of a vector space E satisfy*

$$\dim(A \cap B) = \dim A + \dim B - \dim E$$

with the convention

$$\dim A < 0 \iff A = \emptyset.$$

Purpose of the paper. In [9, p. 270], Gromov's "explanation" did not give a precise definition of a random subset with density, neither a proof of the intersection formula. In [9, p. 272], he proposed that one can consider the class of random subsets defined by measures *invariant under the permutations of E*.

In this article, we discuss two basic models of random subsets that are contained in the permutation invariant model: The *uniform density model* and the *Bernoulli density model*. The first one is defined by the uniform distribution on all subsets of E with cardinality $\lfloor |E|^d \rfloor$. This model is used by Y. Ollivier in [14, 15, 17] to study the density model of random groups, and by A. Żuk in [19] to construct random triangular groups. For the Bernoulli density model, every element in E is taken independently with the same probability $|E|^{d-1}$. This model is considered by Antoniuk–Łuczak–Świątkowski in [1] to study random triangular groups.

The aim of this article is to establish a general framework for the study of random subsets with densities, and to prove the intersection formula for the class of random subsets that are *densable* and *permutation invariant*.

Random subsets and the intersection formula. In the first section, we introduce the notion of *densable sequences of random subsets*. Let E be a finite set. A *random subset* of E is a $\mathcal{P}(E)$-valued random variable, where $\mathcal{P}(E)$ is the set of subsets of E. Note that $|A|$ is a usual real-valued random variable. The *density* of A in E, defined by $\operatorname{dens}_E A := \log_{|E|}(|A|)$, is hence a random variable with values in $\{-\infty\} \cup [0, 1]$.

As our approach is asymptotic when $|E| \to \infty$, we consider a sequence of finite sets $\boldsymbol{E} = (E_n)_{n\in\mathbb{N}}$ where $|E_n| \to \infty$. A *sequence of random subsets* of $\boldsymbol{E}$ is a sequence $\boldsymbol{A} = (A_n)$ where A_n is a random subset of E_n for all $n \in \mathbb{N}$. Such a sequence is *densable* with density $d \in \{-\infty\} \cup [0, 1]$ if the sequence of random variables $\operatorname{dens}_{E_n}(A_n)$ converges weakly (i.e. converges in distribution) to the constant d (cf. [9, p. 272]). For a sequence of properties $\boldsymbol{Q} = (Q_n)$, we say that Q_n is *asymptotically almost surely* (a.a.s.) satisfied if the probability that Q_n is satisfied goes to 1 when $n \to \infty$. For example, for a sequence of random subsets $\boldsymbol{A} = (A_n)$, $\operatorname{dens} \boldsymbol{A} = -\infty$ if and only if a.a.s. $A_n = \emptyset$.

In Section 2, we work on the permutation invariant model in [9, p. 272]. Let E be a finite set. A random subset A of E is *permutation invariant* if its law is invariant under the permutations of E. Namely, for any subset $a \in \mathcal{P}(E)$ and any permutation $\sigma \in \mathcal{S}(E)$, we have $\mathbf{Pr}(A = a) = \mathbf{Pr}(A = \sigma(a))$.

Consider a sequence of finite sets $\boldsymbol{E} = (E_n)$ with $|E_n| \to \infty$. Denote by $\mathcal{D}(\boldsymbol{E})$ the class of *densable sequences of permutation invariant random subsets* of $\boldsymbol{E}$. We prove the intersection formula stated as follows:

Theorem 1 (The intersection formula, Theorem 2.9). *Let $\boldsymbol{A} = (A_n)$, $\boldsymbol{B} = (B_n)$ be independent sequences of random subsets in $\mathcal{D}(\boldsymbol{E})$ with densities α, β. If $\alpha + \beta \neq 1$, then the sequence of random subsets $\boldsymbol{A} \cap \boldsymbol{B}$ is also in $\mathcal{D}(\boldsymbol{E})$. In addition:*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1, \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

The density $-\infty$ means that a.a.s. the random subset is empty.

In Section 3, we study the intersection between a random subset and a *fixed* subset. We develop a generalized form: the *multi-dimensional intersection formula*. Let $\boldsymbol{E} = (E_n)$ be a sequence of finite sets with $|E_n| \to \infty$. Denote $E_n^{(k)}$ the set of pairwise distinct k-tuples of the set E_n. Let $\boldsymbol{A}$ be a sequence of random subsets in $\mathcal{D}(\boldsymbol{E})$ (densable and permutation invariant). We are interested in the intersection between $\boldsymbol{A}^{(k)}$ and a densable sequence of subsets $\boldsymbol{X}$ of $\boldsymbol{E}^{(k)}$.

For $k \geq 2$, the intersection formula is in general not correct (see Example 3.3). We show that by an additional *self-intersection condition* on $\boldsymbol{X}$, we can achieve an intersection formula.

Theorem 2 (The multi-dimensional intersection formula, Theorem 3.7). *Let $\boldsymbol{A} = (A_n)$ be a sequence of random subsets in $\mathcal{D}(\boldsymbol{E})$ with density $0 < d < 1$. Let $\boldsymbol{X} = (X_n)$ be a densable sequence of fixed subsets of $\boldsymbol{E}^{(k)}$ with density α.*

(i) *If $d + \alpha < 1$, then a.a.s.*

$$A_n^{(k)} \cap X_n = \emptyset.$$

(ii) *If $d + \alpha > 1$ and $\boldsymbol{X}$ satisfies the d-small self intersection condition (Definition 3.6), then the sequence of random subsets $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and*

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

The intersection formula in $\boldsymbol{E}$ between a random subset and a fixed subset is a special case of this theorem by taking $k = 1$.

Applications to group theory: Random groups. The last section is dedicated to applications to group theory, more precisely to small cancellation theory.

The first mention of generic property for finitely presented groups appears in the late '80s, in the works of V. S. Guba [10, Remark 2] and M. Gromov [8, Section 0.2]. In [10], the author showed that for "almost every" group presented by $m \geq 4$ generators and one "long" relator, any 2-generated subgroup is free. In [8, Section 0.2], Gromov defined two models of random group presentations with fixed number of generators and relators.

In 1993, Gromov introduced the *density model* of random groups in [9, Section 9.B]. The number of generators is still fixed, but the number of relators *grows exponentially* with the length of the relators, determined by a density parameter d. A *phase transition phenomenon* is then stated as follows: if $d < 1/2$, then a.a.s. the random group is infinite hyperbolic; whereas if $d > 1/2$, then a.a.s. the random group is trivial.

In a 1996 paper [5], G. Arzhantseva and A. Ol'shanskii generalized Guba's result. They proved that for "almost every" group presented by $m \geq 2$ generators and $k \geq 1$ long relators , any $(m-1)$-generated subgroup is free. In their model, the number of generators k is fixed, as in Gromov's 1987 model [8, Section 0.2]. This model is called the Arzhantseva–Ol'shanskii model, or the *few relator model* of random groups.

For more detailed surveys on random groups, see (in chronological order) [7] by E. Ghys, [16] by Y. Ollivier, [12] by I. Kapovich and P. Schupp and [6] by F. Bassino, C. Nicaud and P. Weil.

Fix a set of alphabets $X = \{x_1, \ldots, x_m\}$ as generators of groups. Denote by B_ℓ the set of cyclically reduced words of $X^{\pm}$ of lengths at most ℓ. If S_ℓ is the set of cyclically reduced words of length exactly ℓ, it is clear that

$$2m(2m-1)^{\ell-2}(2m-2) \leq |S_\ell| \leq 2m(2m-1)^{\ell-1}.$$

So,

$$\frac{2m}{2m-1}\left[(2m-1)^{\ell}-1\right] \leq |B_\ell| \leq \frac{2m}{2m-2}\left[(2m-1)^{\ell}-1\right].$$

As we are interested in asymptotic behaviors when $\ell \to \infty$, we can write $|B_\ell| = (2m-1)^{\ell+O(1)}$. Consider $\boldsymbol{B} = (B_\ell)_{\ell\geq 1}$ as our ambient sequence of sets. Let $d \in \{-\infty\} \cup [0,1]$. A *sequence of random groups* with density d, denoted by $\boldsymbol{G}(m,d) = (G_\ell(m,d))$, is defined by random presentations $G_\ell(m,d) = \langle X | R_\ell\rangle$ where $\boldsymbol{R} = (R_\ell)$ is a densable sequence of permutation invariant random subsets of $\boldsymbol{B}$ with density d.

The first mention of the $\lambda/2$ phase transition for the $C'(\lambda)$-small cancellation condition is by Gromov in [9, p. 273], showing that if $2d < \lambda$ then a random group at density d satisfies $C'(\lambda)$. He remarked also that, in particular, if $d < 1/12$ then the group is hyperbolic; and if $d > 1/12$ then the group is not $C'(1/6)$. Ollivier-Wise gave a detailed proof of $d < \lambda/2$ implying $C'(\lambda)$ in [18, Proposition 1.8]. In [16, p. 31], Ollivier stated the phase transition : if $d > \lambda/2$ then $C'(\lambda)$ does not hold. However, his "dimension reasoning" is the 2-dimensional intersection formula between a random subset (pairs of distinct relators in a random group) and a *fixed* subset (pairs of distinct relators denying $C'(\lambda)$), which does not hold in general (as Example 3.3 shows).

Bassino–Nicaud–Weil gave a proof of $d > \lambda/2$ implying non-$C'(\lambda)$ in [6, Theorem 2.1]. Their argument showed that the pairs of distinct relators *in a random group* denying $C'(\lambda)$ is not empty, but did not give its density.

The d-small self-intersection condition (Definition 3.6) for a fixed subset is introduced to rule out this difficulty. By the multi-dimensional intersection formula (Theorem 2), we show that if $d > \lambda/2$, then the pairs of distinct relators in a random group denying $C'(\lambda)$ is with density $d - \lambda/2$ and hence not empty.

Theorem 3 (Phase transition at density $\lambda/2$, Theorem 4.3). *Let* $\boldsymbol{G}(m,d) = (G_\ell(m,d))$ *be a random group with* m *generators and with density* d*. Let* $\lambda \in \,]0,1[$.

(1) *If* $d < \lambda/2$*, then a.a.s.* $G_\ell(m,d)$ *satisfies* $C'(\lambda)$.

(2) *If* $d > \lambda/2$*, then a.a.s.* $G_\ell(m,d)$ *does not satisfy* $C'(\lambda)$.

It was given as an "interesting problem" in [16, Section I.3.c] that some algebraic properties of random groups at density 0 (see [5] by Arzhantseva–Ol'shanskii, [2–4] by Arzhantseva, and [11] by Kapovich–Schupp) may extend to some positive density d. In [12, Theorem 7.5], Kapovich and Schupp extends Arzhantzeva's "all L-generated subgroups of infinite index are free" result [2] (for a fixed $L > 0$) to some density $d > 0$ independent of m. A property is called "low-density random" by Kapovich–Schupp in [12, p. 3] if the corresponding density $d(m)$ is positive but converges to 0 when m goes to infinity. They claimed that Arzhantseva–Ol'shanskii's "all $(m-1)$-generated subgroups are free" result [5] is a low-density random property ([12, Theorems 1.1 (2) and 5.4 (2)]), but the density $d(m)$ is not given.

In our study, the number of generators m is fixed, and we look for a density $d(m)$ such that the "all $(m-1)$-generated subgroups are free" property holds for a random group with m generators of density $d < d(m)$. Using Theorems 2 and 3, we give an explicit bound $d(m) = 1/(120m^2\ln(2m))$ that extends Arzhantseva–Ol'shanskii's result in [5] from density 0 to density $0 \le d < d(m)$.

Theorem 4 (Every $(m-1)$-generated subgroup is free, Theorem 4.5). *Let* $(G_\ell(m,d))$ *be a sequence of random groups with* m *generators and with density*

$$0 \le d < \frac{1}{120m^2\ln(2m)}.$$

Then a.a.s. every $(m-1)$*-generated subgroup of* $G_\ell(m,d)$ *is free.*

Ollivier remarked in [16, p. 71] that at density $d > 1 - \log_{2m-1}(2m-3)$, the rank of a random group with m generators with density d is at most $m-1$, so the "all $(m-1)$-generated subgroups are free" property fails. There is still a large gap between $\log_{2m-1}(2m-3) \sim 1/(m\ln(2m))$ and $1/(120m^2\ln(2m))$.

1. Definitions and basic models

1.1. Densable sequences of random subsets

Let E be a finite set, denote $|E|$ its cardinality. The following definition is due to M. Gromov in [9, p. 269].

Definition 1.1. Let E be a finite non-empty set and $A \subset E$. The density of A in E is defined by

$$\operatorname{dens}_E A := \log_{|E|} |A| = \frac{\log |A|}{\log |E|}.$$

So that $d \in [0, 1] \cup \{-\infty\}$ is a real number such that $|E|^d = |A|$.

We will omit the subscript E if the set is fixed and simply denote the density by $\operatorname{dens} A$. Note that $\operatorname{dens} A = -\infty$ if and only if $A = \emptyset$.

Definition 1.2. Let E be a finite set. Denote $\mathcal{P}(E)$ the set of subsets of E. A random subset A of E is a $\mathcal{P}(E)$-valued random variable.

In this article, we use upper-case letters $A, B, C, \ldots$ to denote random subsets and lower-case letters $a, b, c, \ldots$ to denote fixed subsets. The law of a random subset A is determined by instances $\mathbf{Pr}(A = a)$ through all subsets $a \in \mathcal{P}(E)$ (or $a \subset E$). Its cardinality $|A|$ is a usual real-valued random variable.

Example 1.3. We give three examples of random subsets.

(i) (Dirac model) A fixed subset $c \subset E$ can be regarded as a constant random subset. Its law is

$$\mathbf{Pr}(A = a) = \begin{cases} 1 & \text{if } a = c, \\ 0 & \text{if } a \neq c. \end{cases}$$

(ii) (Uniform random subset) Fix an integer $k \leq |E|$. Let A be the uniform distribution on all subsets of E of cardinality k. Its law is

$$\mathbf{Pr}(A = a) = \begin{cases} \binom{|E|}{k}^{-1} & \text{if } |a| = k, \\ 0 & \text{if } |a| \neq k. \end{cases}$$

(iii) (Bernoulli random subset) Let A be the Bernoulli sampling of parameter $p \in [0, 1]$ on the set E: The events $\{x \in A\}$ through all $x \in E$ are independent of the same probability p. The law of A is

$$\mathbf{Pr}(A = a) = p^{|a|}(1 - p)^{|E|-|a|}.$$

In this case $|A|$ follows the binomial law $\mathcal{B}(|E|, p)$.

As usual random variables, a random subset can be constructed by other random subsets.

Example 1.4 (Set theoretic operations). The intersection of two random subsets A, B of a finite set E is another random subset. The law of $A \cap B$ is

$$\mathbf{Pr}(A \cap B = c) = \sum_{a,b \in \mathcal{P}(E);\, a \cap b = c} \mathbf{Pr}(A = a, B = b).$$

In particular, if A, B are *independent* random subsets, then

$$\mathbf{Pr}(A \cap B = c) = \sum_{a,b \in \mathcal{P}(E);\, a \cap b = c} \mathbf{Pr}(A = a)\mathbf{Pr}(B = b).$$

The union of two subsets and the complement of a subset are similarly defined.

We are interested in the asymptotic behavior of random subsets when $|E| \to \infty$. Consider a sequence of finite sets $\boldsymbol{E} = (E_n)_{n \in \mathbb{N}}$ with $|E_n| \xrightarrow[n\to\infty]{} \infty$. Recall that the density of a subset $a \subset E$ is defined by $\mathrm{dens}_E(a) := \log_{|E|} |a|$.

Definition 1.5. We define densable sequences of random subsets.

(i) A sequence of (fixed) subsets of $\boldsymbol{E} = (E_n)$ is a sequence $\boldsymbol{a} = (a_n)$ such that $a_n \subset E_n$ for all n.

A sequence of subsets $\boldsymbol{a}$ is *densable* with density $d \in [0, 1] \cup \{-\infty\}$ if

$$\mathrm{dens}_{E_n}(a_n) = \log_{|E_n|} |a_n| \xrightarrow[n\to\infty]{} d.$$

(ii) Similarly, a sequence of random subsets of $\boldsymbol{E}$ is a sequence $\boldsymbol{A} = (A_n)$ such that A_n is a random subset of E_n for all n.

A sequence of random subsets $\boldsymbol{A}$ is *densable* with density $d \in [0, 1] \cup \{-\infty\}$ if the sequence of real-valued random valuables $\mathrm{dens}_{E_n}(A_n) = \log_{|E_n|} |A_n|$ converges in distribution to the constant d.

(iii) Two sequences of random subsets $\boldsymbol{A} = (A_n)$, $\boldsymbol{B} = (B_n)$ of $\boldsymbol{E}$ are independent if A_n, B_n are independent random subsets of E_n for all n.

Example 1.6. Here are four examples of densable sequences of random subsets.

(i) For a fixed sequence of subsets $\boldsymbol{a} = (a_n)$, $\mathrm{dens}(\boldsymbol{a}) = -\infty$ if and only if $a_n = \emptyset$ for large enough n.

(ii) A densable sequence of subsets $\boldsymbol{a} = (a_n)$ can be regarded as a densable sequence of random subsets (Dirac model on each term). If we take $|a_n| = \lfloor |E_n|^d \rfloor$ with some $0 \le d \le 1$, then $\boldsymbol{a}$ is densable with density d.

(iii) (Uniform density model) Let $\boldsymbol{A} = (A_n)$ be a sequence of random subsets of $\boldsymbol{E}$. We call $\boldsymbol{A}$ a sequence of uniform random subsets with density d if A_n is the uniform distribution on all subsets of E_n of cardinality $\lfloor |E_n|^d \rfloor$. Its law is

$$\mathbf{Pr}(A_n = a) = \begin{cases} \binom{|E_n|}{\lfloor |E_n|^d \rfloor}^{-1} & \text{if } |a| = \lfloor |E_n|^d \rfloor, \\ 0 & \text{if } |a| \neq \lfloor |E_n|^d \rfloor. \end{cases}$$

(iv) (Bernoulli density model) Let $d > 0$. If A_n is a Bernoulli sampling of E_n with parameter $|E_n|^{d-1}$, then $\boldsymbol{A} = (A_n)$ is a sequence of densable random subsets of $\boldsymbol{E}$. It is rather not obvious that such sequences are densable (see Proposition 1.12).

Definition 1.7. Let $\boldsymbol{Q} = (Q_n)$ be a sequence of events. The event Q_n is *asymptotically almost surely* true if $\mathbf{Pr}(Q_n) \xrightarrow[n\to\infty]{} 1$.

Equivalently, for any $p < 1$ arbitrary close to 1 we have $\mathbf{Pr}(Q_n) > p$ for n large enough. We denote briefly a.a.s. Q_n.

For example, if $\boldsymbol{A}$ is a sequence of random subsets with dens $\boldsymbol{A} = -\infty$, then $\mathbf{Pr}(|A_n| = 0) \xrightarrow[n\to\infty]{} 1$, which is equivalent to a.a.s. $|A_n| = 0$, or a.a.s. $A_n = \emptyset$.

Proposition 1.8 (Characterization of densability). *Let $\boldsymbol{A}$ be a sequence of random subsets of $\boldsymbol{E}$. Let $d \geq 0$. $\boldsymbol{A}$ is densable with density d if and only if*

$$\forall \varepsilon > 0 \text{ a.a.s. } |E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}.$$

Proof. The convergence in distribution to a constant is equivalent to the convergence in probability. So $\log_{|E_n|} |A_n|$ converges in distribution to d if and only if

$$\forall \varepsilon > 0 \quad \mathbf{Pr}\big(|\log_{|E_n|} |A_n| - d| \leq \varepsilon\big) \xrightarrow[n\to\infty]{} 1,$$

which gives the estimation

$$\forall \varepsilon > 0 \text{ a.a.s. } |E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}. \qquad \blacksquare$$

In general, the intersection of two densable sequences is not necessarily densable. The intersection formula is then *not* satisfied by the class of densable sequences of random subsets. Here is a simple example.

Example 1.9. Let $\boldsymbol{E} = (E_n)$ be a sequence of sets with $|E_n| = 2n$. Let $\boldsymbol{a} = (a_n)$, $\boldsymbol{b} = (b_n)$ be sequences of subsets of $\boldsymbol{E}$ such that $b_n = E_n \backslash a_n$ and $|a_n| = |b_n| = n$. They are both densable subsets with density 1 because $\log(n)/\log(2n) \to 1$. Whereas $\text{dens}(\boldsymbol{a} \cap \boldsymbol{b}) = -\infty$. They do not verify the intersection formula.

Define another sequence of subset $\boldsymbol{c} = (c_n)$ by $c_n := a_n$ if n is odd and $c_n := b_n$ if n is even. By its definition, $\boldsymbol{c}$ is densable with density 1. But the intersection $\boldsymbol{b} \cap \boldsymbol{c}$ is empty when n is odd and non-empty when n is even, so $\boldsymbol{b} \cap \boldsymbol{c}$ is not densable.

1.2. The Bernoulli density model

Let $\boldsymbol{E} = (E_n)$ with $|E_n| \to \infty$ be the ambient sequence of sets.

Definition 1.10 (Bernoulli density model). Let $d \leq 1$. Let $\boldsymbol{A} = (A_n)$ be a sequence of random subsets of $\boldsymbol{E}$. It is a sequence of Bernoulli random subsets with density d if A_n is a Bernoulli sampling of E_n with parameter $|E_n|^{d-1}$.

This model is particularly easy to manipulate. We will see that it is densable, closed under intersection and verifies the intersection formula.

Recall that the real-valued random variable $|A_n|$ follows the binomial law

$$\mathcal{B}\big(|E_n|, |E_n|^{d-1}\big).$$

Thus, $\mathbb{E}(|A_n|) = |E_n|^d$.

Lemma 1.11 (Concentration lemma). *Let $\boldsymbol{A}$ be a sequence of Bernoulli random subsets with density $d > 0$. Then a.a.s.*

$$\big||A_n| - |E_n|^d\big| \leq \frac{1}{2}|E_n|^d.$$

Proof. By Chebyshev's inequality,

$$\mathbf{Pr}\Big(\big||A_n| - |E_n|^d\big| > \frac{1}{2}|E_n|^d\Big) \leq \frac{\operatorname{Var}(|A_n|)}{\frac{1}{4}|E_n|^{2d}} \leq \frac{4|E_n|^d(1 - |E_n|^{d-1})}{|E_n|^{2d}} \xrightarrow[n\to\infty]{} 0. \quad \blacksquare$$

Proposition 1.12. *Let $\boldsymbol{A}$ be a sequence of Bernoulli random subsets with density d. If $d \neq 0$, then $\boldsymbol{A}$ is densable and*

$$\operatorname{dens} \boldsymbol{A} = \begin{cases} d & \text{if } 0 < d \leq 1, \\ -\infty & \text{if } d < 0. \end{cases}$$

Proof. We separate the two cases $d < 0$ and $0 < d \leq 1$.

(i) If $d < 0$, by Markov's inequality

$$\mathbf{Pr}\big(|A_n| \geq 1\big) \leq |E_n|^d \to 0,$$

so $\mathbf{Pr}(A_n = \emptyset) \to 1$ and $\mathbf{Pr}(\log_{|E_n|}|A_n| = -\infty) \to 1$.

(ii) If $0 < d \leq 1$, by Lemma 1.11, a.a.s.

$$\frac{1}{2}|E_n|^d \leq |A_n| \leq \frac{3}{2}|E_n|^d.$$

For every $\varepsilon > 0$, the inequality $|E_n|^{d-\varepsilon} < \frac{1}{2}|E_n|^d < \frac{3}{2}|E_n|^d < |E_n|^{d+\varepsilon}$ holds for n large enough. Thus, a.a.s.

$$|E_n|^{d-\varepsilon} \leq |A_n| \leq |E_n|^{d+\varepsilon}.$$

Hence, $\boldsymbol{A} = (A_n)$ is densable with density d by Proposition 1.8. $\blacksquare$

Theorem 1.13 (The intersection formula for Bernoulli density model). *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of Bernoulli random subsets of $\boldsymbol{E} = (E_n)$ with densities α, β. Then $\boldsymbol{A} \cap \boldsymbol{B}$ is a sequence of Bernoulli random subsets of $\boldsymbol{E}$ with density $\alpha + \beta - 1$, and*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1, \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Proof. For every element $x \in E_n$, we have

$$\mathbf{Pr}(x \in A_n \cap B_n) = \mathbf{Pr}(x \in A_n)\mathbf{Pr}(x \in B_n) = |E_n|^{(\alpha+\beta-1)-1}.$$

In addition, for every pair of distinct elements x, y in E_n, we have

$$\begin{aligned} \mathbf{Pr}(x, y \in A_n \cap B_n) &= \mathbf{Pr}(x, y \in A_n)\mathbf{Pr}(x, y \in B_n) \\ &= \mathbf{Pr}(x \in A_n)\mathbf{Pr}(y \in A_n)\mathbf{Pr}(x \in B_n)\mathbf{Pr}(y \in B_n) \\ &= \mathbf{Pr}(x \in A_n \cap B_n)\mathbf{Pr}(y \in A_n \cap B_n). \end{aligned}$$

So $\boldsymbol{A} \cap \boldsymbol{B}$ is a sequence of Bernoulli random subsets with density $\alpha + \beta - 1$. Proposition 1.12 gives its density. ∎

As the theorem shows, the class of Bernoulli random subsets is *closed under intersections*. Thereby the intersection formula works for multiple independent sequences of random subsets. The formula is more concise in terms of *codensities*.

Definition 1.14 (cf. [9, p. 269]). Let $\boldsymbol{A}$ be a densable sequence of random subsets such that $\operatorname{dens} \boldsymbol{A} \in [0, 1]$. Then the codensity of $\boldsymbol{A}$ is defined by

$$\operatorname{codens} \boldsymbol{A} = 1 - \operatorname{dens} \boldsymbol{A}.$$

Theorem 1.13 can be rephrase as (compare [9, p. 270]):

Theorem 1.13′ (The intersection formula by codensities). *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of Bernoulli random subsets of $\boldsymbol{E}$ with positive densities. If* $\operatorname{codens} \boldsymbol{A} + \operatorname{codens} \boldsymbol{B} < 1$, *then*

$$\operatorname{codens}(\boldsymbol{A} \cap \boldsymbol{B}) = \operatorname{codens} \boldsymbol{A} + \operatorname{codens} \boldsymbol{B}.$$

If $\operatorname{codens} \boldsymbol{A} + \operatorname{codens} \boldsymbol{B} > 1$, *then* $\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = -\infty$.

Corollary 1.15 (Generalized intersection formula by codensities). *Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k$ be independent sequences of Bernoulli random subsets with positive densities. If* $\sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i < 1$, *then*

$$\operatorname{codens}\left(\bigcap_{i=1}^{k} \boldsymbol{A}_i\right) = \sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i.$$

If $\sum_{i=1}^{k} \operatorname{codens} \boldsymbol{A}_i > 1$, *then*

$$\operatorname{dens}\Big(\bigcap_{i=1}^{k} \boldsymbol{A}_i\Big) = -\infty.$$

As an exception, a Bernoulli sequence of random subsets with density $d = 0$ is *not densable*.

Proposition 1.16. *Let* $\boldsymbol{A}$ *be a Bernoulli sequence with density* $d = 0$*. Then* $\boldsymbol{A}$ *is* not *densable. In fact,*

$$\mathbf{Pr}(\operatorname{dens} A_n = -\infty) \xrightarrow[n\to\infty]{} 1/e.$$

Proof. $\mathbf{Pr}(|A_n| = 0) = (1 - |E_n|^{-1})^{|E_n|} \xrightarrow[n\to\infty]{} 1/e$, which gives

$$\mathbf{Pr}(\operatorname{dens} A_n = -\infty) \xrightarrow[n\to\infty]{} 1/e.$$

This justifies that the sequence of random variables $(\operatorname{dens}_{E_n} A_n)$ does not converge to any constant distribution. ■

1.3. The uniform density model

The uniform density model is the first example of densable sequences of random subsets. It was introduced by M.Gromov [9, p. 270] to construct random groups with fixed generators, and later developed by Y. Ollivier [15, p. 1]. It was also used by A. Żuk [19] to study random triangular groups.

Let $\boldsymbol{E} = (E_n)$ be a sequence of sets. To simplify, we assume that $|E_n| = n$ in this subsection. For an arbitrary sequence $\boldsymbol{E}$ with $|E_n| \to \infty$ we can proceed similar proofs by replacing n by $|E_n|$. Note that $|E_n|^d = n^d \sim \lfloor n^d \rfloor$, while $n \to \infty$ for $d \in [0, 1]$.

Recall that a sequence of uniform random subsets (Example 1.6 (iv)) of (E_n) with density d is a sequence of random subsets (A_n) with the following laws:

$$\mathbf{Pr}(A_n = a) = \begin{cases} \binom{n}{\lfloor n^d \rfloor}^{-1} & \text{if } |a| = \lfloor n^d \rfloor, \\ 0 & \text{if } |a| \neq \lfloor n^d \rfloor. \end{cases}$$

We give here a concentration lemma for uniform density model, similar to Lemma 1.11. For the proof, we will need Lemmas 2.2 and 2.3 given in the next section.

Lemma 1.17 (Concentration lemma for uniform density model). *Let* $\boldsymbol{A}, \boldsymbol{B}$ *be independent sequences of uniform random subsets of* $\boldsymbol{E}$ *with densities* $\alpha, \beta \in [0, 1]$*. Then:*

(i) $n^{\alpha+\beta-1} - 2 \leq \mathbb{E}\big(|A_n \cap B_n|\big) \leq n^{\alpha+\beta-1}$.

(ii) *If* $\alpha < 1$ *and* $\beta < 1$, *then*

$$\mathrm{Var}\big(|A_n \cap B_n|\big) \sim n^{\alpha+\beta-1}.$$

Moreover, if $n \geq 3$, *then*

$$\mathrm{Var}\big(|A_n \cap B_n|\big) \leq 3n^{\alpha+\beta-1}.$$

(iii) *Let* $0 < c < 1$. *If* $\alpha + \beta - 1 > 0$ *and* $n \geq (\frac{4}{c})^{1/(\alpha+\beta-1)}$, *then*

$$\mathbf{Pr}\Big(\big||A_n \cap B_n| - n^{\alpha+\beta-1}\big| > cn^{\alpha+\beta-1}\Big) \leq \frac{12}{c^2 n^{\alpha+\beta-1}} \xrightarrow[n\to\infty]{} 0.$$

In particular, a.a.s.

$$\big||A_n \cap B_n| - n^{\alpha+\beta-1}\big| \leq cn^{\alpha+\beta-1}.$$

Proof. We consider each item.

(i) By Lemma 2.2, $A_n \cap B_n$ is a permutation invariant random set of E_n. Apply Lemma 2.3:

$$\begin{aligned}\mathbb{E}\big(|A_n \cap B_n|\big) &= n\mathbf{Pr}(x \in A_n \cap B_n) = n\mathbf{Pr}(x \in A_n)\mathbf{Pr}(x \in B_n)\\ &= n\frac{\mathbb{E}(|A_n|)}{n}\frac{\mathbb{E}(|B_n|)}{n} = \lfloor n^\alpha\rfloor\lfloor n^\beta\rfloor n^{-1} \sim n^{\alpha+\beta-1}.\end{aligned}$$

For the inequality, as $\alpha, \beta \leq 1$:

$$n^{\alpha+\beta-1} - 2 \leq n^{\alpha+\beta-1} - n^{\alpha-1} - n^{\beta-1} + n^{-1} \leq \lfloor n^\alpha\rfloor\lfloor n^\beta\rfloor n^{-1} \leq n^{\alpha+\beta-1}.$$

(ii) Let x, y be distinct elements in E. The number of subsets of E containing x, y of cardinality $\lfloor n^\alpha\rfloor$ is $\binom{n-2}{\lfloor n^\alpha\rfloor-2}$, so

$$\mathbf{Pr}(x, y \in A_n) = \frac{\binom{n-2}{\lfloor n^\alpha\rfloor-2}}{\binom{n}{\lfloor n^\alpha\rfloor}} = \frac{\lfloor n^\alpha\rfloor(\lfloor n^\alpha\rfloor - 1)}{n(n-1)}.$$

Similarly,

$$\mathbf{Pr}(x, y \in B_n) = \frac{\lfloor n^\beta\rfloor(\lfloor n^\beta\rfloor - 1)}{n(n-1)}.$$

Denote $k = \lfloor n^\alpha\rfloor$ and $l = \lfloor n^\beta\rfloor$ to simplify the notation. Note that $k = o(n)$ and $l = o(n)$ as $\alpha < 1$ and $\beta < 1$. Recall that $\mathbb{E}(|A_n \cap B_n|) = kln^{-1}$. Apply Lemma 2.3, the variance of $|A_n \cap B_n|$ is

$$\begin{aligned}\mathrm{Var}\big(|A_n \cap B_n|\big) &= kln^{-1} + n(n-1)\mathbf{Pr}(x, y \in A_n)\mathbf{Pr}(x, y \in B_n) - (kln^{-1})^2\\ &= kln^{-1} + \frac{k(k-1)l(l-1)}{n(n-1)} - (kln^{-1})^2\end{aligned}$$

$$= \frac{kl}{n^2(n-1)}(n^2 - n + nkl - nl - nk + n - nkl + kl)$$
$$\sim \frac{kl}{n^2(n-1)} \cdot n^2 \sim n^{\alpha+\beta-1}.$$

Moreover, if $n \geq 3$, then

$$\mathrm{Var}\big(|A_n \cap B_n|\big) = \frac{kl}{n^2(n-1)}(n^2 - nl - nk + kl) \leq \frac{2kl}{n-1} \leq \frac{2n^{\alpha+\beta}}{n-1} \leq 3n^{\alpha+\beta-1}.$$

(iii) By (i) if $n \geq (\frac{4}{c})^{1/(\alpha+\beta-1)} \geq 4$, then

$$\big|\mathbb{E}\big(|A_n \cap B_n|\big) - n^{\alpha+\beta-1}\big| \leq \frac{c}{2} n^{\alpha+\beta-1}.$$

If $\alpha = 1$ or $\beta = 1$ then the result is true as the $A_n = E_n$ or $B_n = E_n$. Otherwise by (ii) and Chebyshev's inequality, if $n \geq (\frac{4}{c})^{1/(\alpha+\beta-1)}$ then

$$\begin{aligned}
&\mathbf{Pr}\big(\big||A_n \cap B_n| - n^{\alpha+\beta-1}\big| > cn^{\alpha+\beta-1}\big) \\
&\quad \leq \mathbf{Pr}\Big(\Big||A_n \cap B_n| - \mathbb{E}\big(|A_n \cap B_n|\big)\Big| > \frac{c}{2} n^{\alpha+\beta-1}\Big) \\
&\quad \leq \frac{4\,\mathrm{Var}(|A_n \cap B_n|)}{c^2 n^{2\alpha+2\beta-2}} \leq \frac{12}{c^2 n^{\alpha+\beta-1}}.
\end{aligned}$$

■

Proposition 1.18 (The intersection formula for uniform density model). *Let* $\boldsymbol{A}, \boldsymbol{B}$ *be independent sequences of uniform random subsets of* $\boldsymbol{E}$ *with densities* α, β. *If* $\alpha + \beta \neq 1$, *then the sequence* $\boldsymbol{A} \cap \boldsymbol{B}$ *is densable and*

$$\mathrm{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1, \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Proof. We separate the two cases $\alpha + \beta < 1$ and $\alpha + \beta > 1$.

(i) If $\alpha + \beta < 1$, then by Markov's inequality and Lemma 1.17 (i):

$$\mathbf{Pr}\big(|A_n \cap B_n| \geq 1\big) \leq \mathbb{E}\big(|A_n \cap B_n|\big) \xrightarrow[n\to\infty]{} 0,$$

which implies a.a.s. $A_n \cap B_n = \emptyset$ and $\mathrm{dens}(A \cap B) = -\infty$.

(ii) If $\alpha + \beta > 1$, by Lemma 1.17 (iii) (with $c = 1/2$) a.a.s.

$$\big||A_n \cap B_n| - n^{\alpha+\beta-1}\big| \leq \frac{1}{2} n^{\alpha+\beta-1},$$

so for all $\varepsilon > 0$ a.a.s.

$$n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}.$$

Hence, by Proposition 1.8, $\boldsymbol{A} \cap \boldsymbol{B}$ is densable with density $\alpha + \beta - 1$. ■

The cardinality of $A_n \cap B_n$ is close to $n^{\alpha+\beta-1}$ with high probability, but not always. If $\alpha \neq 1$ and $\beta \neq 1$, then for n large enough $\lfloor n^\alpha \rfloor + \lfloor n^\beta \rfloor < n$, so

$$\mathbf{Pr}(A_n \cap B_n = \emptyset) \neq 0.$$

Which means that $\boldsymbol{A} \cap \boldsymbol{B}$ is not a sequence of uniform random subsets, so the class of sequences of uniform random subsets is *not* closed under intersection.

2. The general model: Densable and permutation invariant

2.1. Densable sequences of permutation invariant random subsets

Let E be a finite set with cardinality $|E| = n$. Denote $\mathcal{S}(E)$ as the group of permutations of E. The action of $\mathcal{S}(E)$ on E can be extended on $\mathcal{P}(E)$, defined by $\sigma(\{x_1, \ldots, x_k\}) := \{\sigma(x_1), \ldots, \sigma(x_k)\}$.

Note that this action has $(n+1)$ orbits of the form $\{a \in \mathcal{S}(E) \mid |a| = k\}$ for $k \in \{0, \ldots, n\}$. Moreover, the action commutes with set theoretic operations: $\sigma(E \backslash a) = E \backslash \sigma(a)$, $\sigma(a \cap b) = \sigma(a) \cap \sigma(b)$ and $\sigma(a \cup b) = \sigma(a) \cup \sigma(b)$.

Definition 2.1 (Permutation invariant random subsets). Let A be a random subset of E. It is permutation invariant if its law is invariant by the permutations of E. i.e.

$$\forall a \in \mathcal{P}(E)\ \forall \sigma \in \mathcal{S}(E) \quad \mathbf{Pr}(A = a) = \mathbf{Pr}(A = \sigma(a)).$$

Equivalently, subsets of E of the same cardinality are equiprobable. There exists real numbers $p_0, \ldots, p_n \in [0, 1]$ satisfying

$$\sum_{k=0}^{n} \binom{n}{k} p_k = 1$$

such that

$$\forall a \in \mathcal{P}(E) \quad |a| = k \Rightarrow \mathbf{Pr}(A = a) = p_k.$$

By definition, uniform random subsets and Bernoulli random subsets are permutation invariant. The advantage of such class of random subsets is that it is closed under set theoretic operations, especially under intersections.

Lemma 2.2 (Closed under set operations). *Let E be a finite set. The class of permutation invariant random subsets of E is closed under set theoretic operations (union, complement and intersection).*

Proof. We treat the three set theoretic operations (complement, intersection and union) separately.

(i) (Complement) Let A be a permutation invariant random subset. Let $a \in \mathcal{P}(E)$ and $\sigma \in \mathcal{S}(E)$. Then

$$\begin{aligned}\mathbf{Pr}(E\backslash A = a) &= \mathbf{Pr}(A = E\backslash a) = \mathbf{Pr}(A = \sigma(E\backslash a)) \\ &= \mathbf{Pr}(A = E\backslash\sigma(a)) = \mathbf{Pr}(E\backslash A = \sigma(a)).\end{aligned}$$

(ii) (Intersection) Let A, B be independent permutation invariant random subsets. Then for $\sigma \in \mathcal{S}(E)$,

$$\begin{aligned}\mathbf{Pr}(A \cap B = c) &= \sum_{a,b\in\mathcal{P}(E);\, a\cap b=c} \mathbf{Pr}(A = a)\mathbf{Pr}(B = b) \\ &= \sum_{a,b\in\mathcal{P}(E);\, \sigma(a)\cap\sigma(b)=\sigma(c)} \mathbf{Pr}(A = \sigma(a))\mathbf{Pr}(B = \sigma(b)) \\ &= \sum_{a',b'\in\mathcal{P}(E);\, a'\cap b'=\sigma(c)} \mathbf{Pr}(A = a')\mathbf{Pr}(B = b') \quad \text{(by substitution)} \\ &= \mathbf{Pr}(A \cap B = \sigma(c)).\end{aligned}$$

(iii) (Union) Let A, B be independent permutation invariant random subsets. Then $A \cup B = E\backslash((E\backslash A) \cap (E\backslash B))$. So $A \cup B$ is permutation invariant. ∎

We shall express the expectation and the variance of the random variable $|A|$ in terms of $\mathbf{Pr}(x \in A)$ and $\mathbf{Pr}(x \in A, y \in A)$, where x, y are distinct elements in E.

Lemma 2.3. *Let A be a permutation invariant random subset of E. Let x, y be distinct elements in E. Then:*

(i) $\mathbb{E}(|A|) = n\mathbf{Pr}(x \in A)$,

(ii) $\mathrm{Var}(|A|) = \mathbb{E}(|A|) + n(n-1)\mathbf{Pr}(x \in A, y \in A) - \mathbb{E}(|A|)^2$.

Proof. We consider each item.

(i) By definition the probability $\mathbf{Pr}(z \in A)$ does not depend on the choice of element $z \in E$. So,

$$\mathbb{E}(|A|) = \mathbb{E}\Big(\sum_{z\in E} \mathbb{1}_{z\in A}\Big) = \sum_{z\in E} \mathbf{Pr}(z \in A) = n\mathbf{Pr}(x \in A).$$

(ii) By the same argument, the probability $\mathbf{Pr}(z \in A, w \in A)$ does not depend on the choice of pair of distinct elements (z, w) in E. So,

$$\mathbb{E}(|A|^2) = \mathbb{E}\Big[\Big(\sum_{z\in E} \mathbb{1}_{z\in A}\Big)^2\Big]$$

$$= \sum_{z \in E} \mathbf{Pr}(z \in A) + \sum_{(z,w) \in E^2; z \neq w} \mathbf{Pr}(z \in A, w \in A)$$
$$= \mathbb{E}\big(|A|\big) + n(n-1)\mathbf{Pr}(x \in A, y \in A). \qquad \blacksquare$$

A permutation invariant random subset can be decomposed into uniform random subsets.

Proposition 2.4 (Decomposition into uniform random subsets). *Let A be a permutation invariant random subset of E.*

(i) *If $\mathbf{Pr}(|A| = k) \neq 0$, then the random subset A under the condition $\{|A| = k\}$ is a uniform random subset on all subsets of E of cardinality k.*

(ii) *Let Q be an event described by A (for example, $Q = \{x \in A\}$). Denote $\mathbb{N}_A = \{k \in \mathbb{N} \mid \mathbf{Pr}(|A| = k) \neq 0\}$, then*

$$\mathbf{Pr}(Q) = \sum_{k \in \mathbb{N}_A} \mathbf{Pr}\big(Q \mid |A| = k\big)\mathbf{Pr}\big(|A| = k\big).$$

Proof. Suppose that $\mathbf{Pr}(|A| = k) \neq 0$. Let $a \subset E$ of cardinal k. As A is permutation invariant,

$$\mathbf{Pr}\big(|A| = k\big) = \binom{n}{k}\mathbf{Pr}(A = a).$$

Hence,

$$\mathbf{Pr}\big(A = a \mid |A| = k\big) = \frac{\mathbf{Pr}(A = a)}{\mathbf{Pr}(|A| = k)} = \binom{n}{k}^{-1}.$$

If $|a| \neq k$ then $\mathbf{Pr}(A = a \mid |A| = k) = 0$.

The second assertion is the formula of total probability. $\blacksquare$

Definition 2.5. Let $\boldsymbol{A} = (A_n)$ be a sequence of random subsets of $\boldsymbol{E} = (E_n)$. It is a sequence of permutation invariant random subset if A_n is a permutation invariant random subset of E_n for all n.

Notation. Let $\boldsymbol{E} = (E_n)$ be a sequence of finite sets. Denote $\mathcal{D}(\boldsymbol{E})$ the class of densable sequences of permutation invariant random subsets of $\boldsymbol{E}$.

Example 2.6. We give three examples of sequences of random subsets that are densable and permutation invariant.

1. Sequences of Bernoulli random subsets of $\boldsymbol{E}$ with density $d \neq 0$ are in the class $\mathcal{D}(\boldsymbol{E})$.

2. Sequences of uniform random subsets of $\boldsymbol{E}$ are in the class $\mathcal{D}(\boldsymbol{E})$.

3. Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences of uniform random subsets. By Lemma 2.2, the sequence $\boldsymbol{A} \cap \boldsymbol{B}$ is permutation invariant. By Proposition 1.18, if $\operatorname{dens} \boldsymbol{A} + \operatorname{dens} \boldsymbol{B} \neq 1$, then $\boldsymbol{A} \cap \boldsymbol{B}$ is densable. In this case the sequence $\boldsymbol{A} \cap \boldsymbol{B}$ is in the class $\mathcal{D}(\boldsymbol{E})$.

Except for some special cases, the class $\mathcal{D}(\boldsymbol{E})$ is closed under set theoretic operations:

Proposition 2.7. *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{D}(\boldsymbol{E})$ with densities α, β. Then the union $\boldsymbol{A} \cup \boldsymbol{B}$ is in $\mathcal{D}(\boldsymbol{E})$ and* $\operatorname{dens}(\boldsymbol{A} \cup \boldsymbol{B}) = \max(\alpha, \beta)$.

Proof. By Lemma 2.2, the sequence of random subset $\boldsymbol{A} \cup \boldsymbol{B}$ is permutation invariant. The cases $\alpha = -\infty$ or $\beta = -\infty$ can be easily shown. Without loss of generality, assume that $\alpha \geq \beta \geq 0$.

Let $\varepsilon > 0$. By densabilities of $\boldsymbol{A}$ and $\boldsymbol{B}$, a.a.s.

$$n^{\alpha-\varepsilon/2} \leq |A_n| \leq n^{\alpha+\varepsilon/2},$$
$$n^{\beta-\varepsilon/2} \leq |B_n| \leq n^{\beta+\varepsilon/2}.$$

Thus, a.a.s.

$$n^{\alpha-\varepsilon} \leq |A_n| \leq |A_n \cup B_n| \leq n^{\alpha+\varepsilon/2} + n^{\beta+\varepsilon/2} \leq 2n^{\alpha+\varepsilon/2} \leq n^{\alpha+\varepsilon}. \qquad \blacksquare$$

Proposition 2.8. *Let $\boldsymbol{A} \in \mathcal{D}(\boldsymbol{E})$ with density $\alpha < 1$. Then the complement $\boldsymbol{E} \backslash \boldsymbol{A}$ is in $\mathcal{D}(\boldsymbol{E})$ and* $\operatorname{dens}(\boldsymbol{E} \backslash \boldsymbol{A}) = 1$.

Proof. Again by Lemma 2.2 the sequence of random subset $\boldsymbol{E} \backslash \boldsymbol{A}$ is permutation invariant.

Let $0 < \varepsilon < (1-\alpha)/2$. By densablility of $\boldsymbol{A}$, a.a.s.

$$|A_n| \leq n^{\alpha+\varepsilon}.$$

As $n^{\alpha+\varepsilon} + n^{1-\varepsilon} \leq n$ for n large enough, a.a.s.

$$|E_n \backslash A_n| \geq n - n^{\alpha+\varepsilon} \geq n^{1-\varepsilon}. \qquad \blacksquare$$

2.2. The intersection formula

In this subsection we shall prove the intersection formula for the class of densable sequences of permutation invariant random subsets.

Theorem 2.9 (The intersection formula). *Let $\boldsymbol{A}, \boldsymbol{B}$ be independent sequences in $\mathcal{D}(\boldsymbol{E})$ with densities α, β. If $\alpha + \beta \neq 1$, then the sequence $\boldsymbol{A} \cap \boldsymbol{B}$ is in $\mathcal{D}(\boldsymbol{E})$ and*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{B}) = \begin{cases} \alpha + \beta - 1 & \text{if } \alpha + \beta > 1, \\ -\infty & \text{if } \alpha + \beta < 1. \end{cases}$$

Lemma 2.10. *Let $\alpha, \beta \in [0,1]$ such that $\alpha + \beta > 1$. Let $0 < \varepsilon < \alpha + \beta - 1$. Let $\boldsymbol{A}$, $\boldsymbol{B}$ independent sequences of uniform random subsets of $\boldsymbol{E}$ with densities α', β' with $\alpha' \in [\alpha - \varepsilon/3, \alpha + \varepsilon/3]$ and $\beta' \in [\beta - \varepsilon/3, \beta + \varepsilon/3]$. If $n \geq \max\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\}$, then*

$$\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\big) \geq 1 - \frac{48}{n^{\alpha+\beta-1-\varepsilon}} \xrightarrow[n\to\infty]{} 1.$$

Proof. By hypothesis $\alpha' + \beta' - 1 \geq \alpha + \beta - 2\varepsilon/3 - 1 > 0$. Apply Lemma 1.17 (iii) with $c = \frac{1}{2}$, for $n \geq 8^{1/(\alpha+\beta-1-\varepsilon)} \geq 8^{1/(\alpha'+\beta'-1)}$, we have

$$\mathbf{Pr}\Big(\big||A_n \cap B_n| - n^{\alpha'+\beta'-1}\big| \geq \frac{1}{2}n^{\alpha'+\beta'-1}\Big) \leq \frac{48}{n^{\alpha'+\beta'-1}}.$$

Considering the complement event, this inequality is equivalent to

$$\mathbf{Pr}\Big(\frac{1}{2}n^{\alpha'+\beta'-1} < |A_n \cap B_n| < \frac{3}{2}n^{\alpha'+\beta'-1}\Big) > 1 - \frac{48}{n^{\alpha'+\beta'-1}}.$$

Again by hypothesis,

$$\alpha + \beta - 1 - 2\varepsilon/3 \leq \alpha' + \beta' - 1 \leq \alpha + \beta - 1 + 2\varepsilon/3.$$

If $n \geq 2^{3/\varepsilon}$, then

$$n^{\alpha+\beta-1-\varepsilon} \leq \frac{1}{2}n^{\alpha+\beta-1-2\varepsilon/3} \leq \frac{3}{2}n^{\alpha+\beta-1+2\varepsilon/3} \leq n^{\alpha+\beta-1+\varepsilon},$$

so,

$$\begin{aligned}
&\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\big)\\
&\quad \geq \mathbf{Pr}\Big(\frac{1}{2}n^{\alpha+\beta-1-2\varepsilon/3} \leq |A_n \cap B_n| \leq \frac{3}{2}n^{\alpha+\beta-1+2\varepsilon/3}\Big)\\
&\quad \geq \mathbf{Pr}\Big(\frac{1}{2}n^{\alpha'+\beta'-1} < |A_n \cap B_n| < \frac{3}{2}n^{\alpha'+\beta'-1}\Big).
\end{aligned}$$

Combine two estimations on n. If $n \geq \max\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\}$, then:

$$\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\big) \geq 1 - \frac{48}{n^{\alpha'+\beta'-1}} \geq 1 - \frac{48}{n^{\alpha+\beta-1-\varepsilon}}.$$

As $\alpha + \beta - 1 - \varepsilon > 0$, when n goes to infinity

$$\frac{48}{n^{\alpha+\beta-1-\varepsilon}} \xrightarrow[n\to\infty]{} 0.$$

∎

Proof of Theorem 2.9. By Lemma 2.2, the intersection $\boldsymbol{A} \cap \boldsymbol{B}$ is a sequence of permutation invariant random subsets. In either case, denote (Q_n) the sequence of events defined by

$$Q_n = \big\{n^{\alpha-\varepsilon/3} \leq |A_n| \leq n^{\alpha+\varepsilon/3} \text{ and } n^{\beta-\varepsilon/3} \leq |B_n| \leq n^{\beta+\varepsilon/3}\big\}$$

for some small $\varepsilon > 0$. By the densabilities of $\boldsymbol{A}$ and $\boldsymbol{B}$, a.a.s. Q_n is true. Note that Q_n is a union of events of type $\{|A_n| = k, |B_n| = l\}$. Denote

$$\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon} := \Big\{(k,l) \in \mathbb{N}^2 \mid n^{\alpha-\varepsilon/3} \le k \le n^{\alpha+\varepsilon/3},\ n^{\beta-\varepsilon/3} \le l \le n^{\beta+\varepsilon/3} \\ \text{and } \mathbf{Pr}(|A_n| = k, |B_n| = l) \neq 0\Big\}.$$

For $(k,l) \in \mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}$, we may do a change of variables $k = n^{\alpha'}$, $l = n^{\beta'}$ so that

$$\alpha - \varepsilon/3 \le \alpha' \le \alpha + \varepsilon/3 \quad \text{and} \quad \beta - \varepsilon/3 \le \beta' \le \beta + \varepsilon/3.$$

(i) Suppose that $\alpha + \beta < 1$. Let $0 < \varepsilon < 1 - \alpha - \beta$. We shall prove that a.a.s. $A_n \cap B_n = \emptyset$. By the formula of total probability and Markov's inequality,

$$\begin{aligned}\mathbf{Pr}\big(A_n \cap B_n \neq \emptyset\big) &\le \mathbf{Pr}\big(|A_n \cap B_n| \ge 1 \mid Q_n\big)\mathbf{Pr}(Q_n) + \mathbf{Pr}(\overline{Q_n}) \\ &\le \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big[\mathbf{Pr}\big(|A_n \cap B_n| \ge 1 \mid |A_n| = k, |B_n| = l\big) \\ &\qquad \mathbf{Pr}\big(|A_n| = k, |B_n| = l|\big)\Big] + \mathbf{Pr}(\overline{Q_n}) \\ &\le \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big[\mathbb{E}\big(|A_n \cap B_n| \mid |A_n| = k, |B_n| = l\big) \\ &\qquad \mathbf{Pr}\big(|A_n| = k, |B_n| = l|\big)\Big] + \mathbf{Pr}(\overline{Q_n}).\end{aligned}$$

For any $(k,l) \in \mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}$, by Lemma 1.17 (i), we have

$$\begin{aligned}\mathbb{E}\big(|A_n \cap B_n| \mid |A_n| = k, |B_n| = l\big) &= \mathbb{E}\big(|A_n \cap B_n| \mid |A_n| = n^{\alpha'}, |B_n| = n^{\beta'}\big) \\ &\le n^{\alpha'+\beta'-1} \le n^{\alpha+\beta+2/3\varepsilon-1} \le n^{-1/3\varepsilon}.\end{aligned}$$

Hence,

$$\mathbf{Pr}(A_n \cap B_n \neq \emptyset) \le n^{-1/3\varepsilon}\mathbf{Pr}(Q_n) + \mathbf{Pr}(\overline{Q_n}) \xrightarrow[n\to\infty]{} 0.$$

(ii) Suppose that $\alpha + \beta > 1$. Let $0 < \varepsilon < \alpha + \beta - 1$. We shall prove that a.a.s.

$$n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon}.$$

By the formula of total probability,

$$\begin{aligned}&\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon}\big) \\ &\ge \mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon} \mid Q_n\big)\mathbf{Pr}(Q_n) \\ &= \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big[\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \le |A_n \cap B_n| \le n^{\alpha+\beta-1+\varepsilon} \mid |A_n| = k, |B_n| = l\big) \\ &\qquad \mathbf{Pr}\big(|A_n| = k, |B_n| = l|\big)\Big].\end{aligned}$$

By Lemma 2.10 and Proposition 2.4. If $n \geq \max\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\}$, then for any $(k,l) \in \mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}$, we have

$$\begin{aligned}
&\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon} \mid |A_n| = k, |B_n| = l\big) \\
&\quad = \mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon} \mid |A_n| = n^{\alpha'}, |B_n| = n^{\beta'}\big) \\
&\quad \geq 1 - \frac{48}{n^{\alpha+\beta-1+\varepsilon}} \xrightarrow[n\to\infty]{} 1.
\end{aligned}$$

Hence, for $n \geq \max\{2^{3/\varepsilon}, 8^{1/(\alpha+\beta-1-\varepsilon)}\}$, we obtain

$$\begin{aligned}
&\mathbf{Pr}\big(n^{\alpha+\beta-1-\varepsilon} \leq |A_n \cap B_n| \leq n^{\alpha+\beta-1+\varepsilon}\big) \\
&\quad \geq \sum_{(k,l)\in\mathbb{N}^2_{\boldsymbol{A},\boldsymbol{B},n,\varepsilon}} \Big(1 - \frac{48}{n^{\alpha+\beta-1+\varepsilon}}\Big)\mathbf{Pr}\big(|A_n| = k, |B_n| = l\big) \\
&\quad \geq \Big(1 - \frac{48}{n^{\alpha+\beta-1+\varepsilon}}\Big)\mathbf{Pr}(Q_n) \xrightarrow[n\to\infty]{} 1.
\end{aligned}$$

■

Remark that when $\alpha + \beta = 1$, the density is not determined, as Proposition 1.16 showed for Bernoulli random subsets. As the class is closed under intersection, we can conclude on multiple intersections.

Corollary 2.11. *Let $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k$ be independent sequences in $\mathcal{D}(\boldsymbol{E})$ of positive densities. If $\sum_{i=1}^k \operatorname{codens} \boldsymbol{A}_i < 1$, then*

$$\operatorname{codens}\Big(\bigcap_{i=1}^k \boldsymbol{A}_i\Big) = \sum_{i=1}^k \operatorname{codens} \boldsymbol{A}_i.$$

If $\sum_{i=1}^k \operatorname{codens} \boldsymbol{A}_i > 1$, then

$$\operatorname{dens}\Big(\bigcap_{i=1}^k \boldsymbol{A}_i\Big) = -\infty.$$

2.3. Another model: Random functions

We give here another natural model of random subsets: image of a random function, which can be found in [9, p. 270] “Self-intersection formula” by Gromov. This is also a variance of random groups considered by Ollivier in [16, Lemma 59]. In this subsection, we prove that such a model is densable and permutation invariant.

Definition 2.12. Let E, F be finite subsets of cardinalities n, m. Denote E^F the set of functions from F to E. A random function Φ from F to E is a E^F-valued random variable.

Let Φ be a random function from F to E. Its law is determined by

$$\mathbf{Pr}(\Phi = \varphi)$$

through all $\varphi \in E^F$.

The random function Φ can be regarded as a vector of E-valued random variables (or random elements of E) $(\Phi(y))_{y\in F}$ indexed by F. Note that these random elements are not necessarily independent. The image $\mathrm{Im}(\Phi) = \Phi(F) := \{\Phi(y) | y \in F\}$ is then a random subset of E.

Example 2.13 (Uniform random function). Let Φ be the uniform distribution on all functions from F to E. Its law is

$$\mathbf{Pr}(\Phi = \varphi) = \frac{1}{|E^F|} = \frac{1}{n^m}$$

through all $\varphi \in E^F$.

Proposition 2.14. *Let Φ be a uniform random function from F to E. Then the random elements $(\Phi(y))_{y\in F}$ are independent (identical) uniform distributions on E.*

Proof. Let $x \in E$, $y \in F$. The number of functions φ from F to E such that $\varphi(y) = x$ is n^{m-1}. So the law of $\Phi(y)$ is

$$\mathbf{Pr}(\Phi(y) = x) = \frac{n^{m-1}}{n^m} = \frac{1}{n}.$$

Which is a uniform distribution on E.

Denote $F = \{y_1, \dots, y_m\}$. Let $(x_1, \dots, x_m)$ be a vector of m elements in E. Let $\varphi \in E^F$ such that $\varphi(y_i) = x_i$ for all $1 \le i \le m$. Then

$$\mathbf{Pr}\Big(\bigwedge_{i=1}^{m} \Phi(y_i) = x_i\Big) = \mathbf{Pr}(\Phi = \varphi) = \frac{1}{n^m} = \prod_{i=1}^{m} \mathbf{Pr}(\Phi(y_i) = x_i). \qquad \blacksquare$$

Proposition 2.15. *The image of a uniform random function is a permutation invariant random subset.*

Proof. Let Φ be a uniform random function from F to E. Let $\sigma \in \mathcal{S}(E)$, then for all $\varphi \in E^F$, we have

$$\mathbf{Pr}(\Phi = \varphi) = \mathbf{Pr}(\Phi = \sigma \circ \varphi) = \mathbf{Pr}(\sigma^{-1} \circ \Phi = \varphi).$$

The random function $\sigma^{-1} \circ \Phi$ has the same law of Φ. Now let $a \subset E$, then

$$\mathbf{Pr}(\mathrm{Im}(\Phi) = a) = \mathbf{Pr}(\mathrm{Im}(\sigma^{-1} \circ \Phi) = a) = \mathbf{Pr}(\mathrm{Im}(\Phi) = \sigma(a)). \qquad \blacksquare$$

3. The multi-dimensional intersection formula

Let $\boldsymbol{E} = (E_n)$ be a sequence of finite sets with $|E_n| = n$ and $k \geq 2$ be an integer. The set of pairwise different k-tuples of E_n is

$$E_n^{(k)} := \{(x_1, \dots, x_k) \in E_n^k \mid x_i \neq x_j \ \forall i \neq j\}.$$

Denote $\boldsymbol{E}^{(k)} = (E_n^{(k)})_{n \in \mathbb{N}}$.

Similarly, for a sequence of random subsets $\boldsymbol{A} = (A_n)$ of $\boldsymbol{E}$, we can define

$$A_n^{(k)} := \{(x_1, \dots, x_k) \in A_n^k \mid x_i \neq x_j \ \forall i \neq j\},$$

which is a random subset of $E_n^{(k)}$. Denote also $\boldsymbol{A}^{(k)} = (A_n^{(k)})$. We will establish an intersection formula between a sequence of random subsets of type $\boldsymbol{A}^{(k)}$ and a sequence of fixed subsets $\boldsymbol{X} = (X_n)$ of $\boldsymbol{E}^{(k)}$.

Proposition 3.1. *Let $\boldsymbol{A}$ be a densable sequence of random subsets of $\boldsymbol{E}$ with density $d > 0$. Then $\boldsymbol{A}^{(k)}$ is a densable sequence of random subsets of $\boldsymbol{E}^{(k)}$ with density d. Namely,*

$$\operatorname{dens}_{\boldsymbol{E}^{(k)}}(\boldsymbol{A}^{(k)}) = \operatorname{dens}_{\boldsymbol{E}}(\boldsymbol{A}).$$

Proof. Note that $n^k - k^2(n-1)^k \leq |E_n^{(k)}| \leq n^k$, so $|E_n^{(k)}| = n^{k+o(1)}$.

Let $\varepsilon > 0$. By densability a.a.s. $n^{d-\varepsilon/2} \leq |A_n| \leq n^{d+\varepsilon/2}$. By the same argument above, a.a.s. $|A_n^{(k)}| = |A_n|^{k+o(1)}$ as random variables. Hence, a.a.s.

$$(n^k)^{d-\varepsilon/2+o(1)} \leq |A_n^{(k)}| \leq (n^k)^{d+\varepsilon/2+o(1)},$$

so a.a.s.

$$|E_n^{(k)}|^{d-\varepsilon} \leq |A_n^{(k)}| \leq |E_n^{(k)}|^{d+\varepsilon}.$$ ∎

Although the densability is preserved, it is not the case for being permutation invariant. Given a permutation invariant random subset A_n of E_n, the random subset $A_n^{(k)}$ is *not* permutation invariant in $E_n^{(k)}$ for $k \geq 2$. See the following example.

Example 3.2. Let (A_n) be a sequence of Bernoulli random subsets of (E_n) with density $0 < d < 1$. Recall that subsets of the same cardinality have the same probability to be included in a permutation invariant random subset. Let $x_1, \dots, x_4$ be distinct elements in E_n.

$$\mathbf{Pr}\big(\{(x_1, x_2), (x_3, x_4)\} \subset A_n^{(2)}\big) = \mathbf{Pr}\big(\{x_1, x_2, x_3, x_4\} \subset A_n\big) = n^{4(d-1)},$$

while

$$\mathbf{Pr}\big(\{(x_1, x_2), (x_2, x_3)\} \subset A_n^{(2)}\big) = \mathbf{Pr}\big(\{x_1, x_2, x_3\} \subset A_n\big) = n^{3(d-1)}.$$

As a result the classical intersection formula (Theorem 2.9) can not be applied in this context. Actually, for $k \geq 2$ the intersection formula *does not* work for some choices of $\boldsymbol{X}$. We give here a counter example.

Example 3.3. Let $\boldsymbol{A}$ be a sequence of random subsets in $\mathcal{D}(\boldsymbol{E})$ with density $3/4$. Let $\boldsymbol{X} = (X_n)$ be a sequence of subsets defined by

$$X_n = \{x_n\} \times \big(E_n \backslash \{x_n\}\big) \subset E_n^{(2)}$$

with some $x_n \in E_n$. By its construction $\mathrm{dens}_{\boldsymbol{E}^{(2)}}(\boldsymbol{X}) = 1/2$, so we expected that $\mathrm{dens}(\boldsymbol{A}^{(2)} \cap \boldsymbol{X}) = 3/4 + 1/2 - 1 = 1/4$. However, we have

$$\mathrm{dens}(\boldsymbol{A}^{(2)} \cap \boldsymbol{X}) = 0$$

because a.a.s. $A_n \cap \{x_n\} = \emptyset$.

For the intersection formula between $\boldsymbol{A}^{(k)}$ and $\boldsymbol{X}$, we need an additional condition on $\boldsymbol{X}$. More precisely, $\boldsymbol{X}$ can not have too much "self-intersection". We will discuss this condition in Section 3.1.

Following the path for proving the intersection formula (Theorem 2.9), we shall study the case that $\boldsymbol{A}$ is a sequence of Bernoulli random subsets with density d (Section 3.2). We then adapt the proof for the uniform density model by estimating the probabilities $\mathbf{Pr}(\{x_1, \ldots, x_r\} \subset A_n)$ (Section 3.3).

For the general case (Section 3.4), according to Proposition 2.4, we can decompose a permutation invariant random subset into uniform random subsets. We then need to bound $|A_n^{(k)} \cap X_n|$ for sequences of uniform random subsets, uniformly in a small neighborhood of densities $d' \in [d - \varepsilon, d + \varepsilon]$.

3.1. Statement of the theorem

Definition 3.4 (Self-intersection partition). Let $\boldsymbol{X} = (X_n)$ be a sequence of fixed subsets of $\boldsymbol{E}^{(k)}$ with density α. For $0 \leq i \leq k$, the ith self-intersection of X_n is

$$Y_{i,n} := \big\{(x, y) \in X_n^2 \mid |x \cap y| = i\big\},$$

where $|x \cap y|$ is the number of common elements in $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$.

In particular, $Y_{0,n}$ is the set of pairs (x, y) in X_n^2 having no intersection; $Y_{k,n}$ is the set of pairs (x, x) in X_n^2. Note that $(Y_{i,n})_{0 \leq i \leq k}$ is a partition of X_n^2, called the *self-intersection partition* of X_n. Namely,

$$X_n^2 = \bigsqcup_{i=0}^{k} Y_{i,n}.$$

The sequence $\boldsymbol{Y}_i = (Y_{i,n})_{n\in\mathbb{N}}$ is called the ith self intersection of $\boldsymbol{X}$. The family of sequences $(\boldsymbol{Y}_i)_{0\le i\le k}$ is called the *self-intersection partition* of $\boldsymbol{X}$. Namely,

$$\boldsymbol{X}^2 = \bigsqcup_{i=0}^{k} \boldsymbol{Y}_i.$$

Remark that the sequences $\boldsymbol{X}^2$ and $\boldsymbol{Y}_i$ are sequences of fixed subsets of $(\boldsymbol{E}^{(k)})^2 = ((E_n^{(k)})^2)_{n\in\mathbb{N}}$. Note that $\mathrm{dens}_{(\boldsymbol{E}^{(k)})^2}(\boldsymbol{X}^2) = \mathrm{dens}_{\boldsymbol{E}^{(k)}}(\boldsymbol{X}) = \alpha$. To give a condition on $\boldsymbol{Y}_i$, we need the notion of *upper density*, defined by an upper limit:

Definition 3.5. Let $\boldsymbol{Y} = (Y_n)$ be a sequence of subsets of $\boldsymbol{E} = (E_n)$. The upper density of $\boldsymbol{Y}$ in $\boldsymbol{E}$ is

$$\overline{\mathrm{dens}}_{\boldsymbol{E}} \boldsymbol{Y} := \overline{\lim_{n\to\infty}} \log_{|E_n|}\big(|Y_n|\big).$$

We introduce here, for a sequence of densable fixed subsets $\boldsymbol{X}$ of $\boldsymbol{E}^{(k)}$ with density α, the small self-intersection condition.

Definition 3.6. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α and let $(\boldsymbol{Y}_i)_{0\le i\le k}$ be its self-intersection partition. Let $d > 1-\alpha$. We say that $\boldsymbol{X}$ satisfies the d-small self-intersection condition if, for every $1 \le i \le k-1$,

$$\overline{\mathrm{dens}}_{(\boldsymbol{E}^{(k)})^2}(\boldsymbol{Y}_i) < \alpha - (1-d) \times \frac{i}{2k}.$$

Remark that the right-hand side of the inequality is between 0 and α because $\alpha > 1-d > 0$. Note that $|Y_{k,n}| = |\{(x,y) \in X_n^2 \mid x = y\}| = |X_n|$, so

$$\mathrm{dens}_{(\boldsymbol{E}^{(k)})^2} \boldsymbol{Y}_k = \frac{\alpha}{2} < \alpha - (1-d)\frac{k}{2k},$$

which satisfies automatically the inequality. On the other hand, as the upper densities of $\boldsymbol{Y}_i$ for $1 \le i \le k$ are all smaller than α and $|Y_{0,n}| = |X_n^2| - \sum_{i=1}^{k} |Y_{i,n}|$, by Proposition 2.8, the sequence $\boldsymbol{Y}_0$ is densable and

$$\mathrm{dens}\, \boldsymbol{Y}_0 = \mathrm{dens}\, \boldsymbol{X}^2 = \alpha.$$

The purpose of this section is to demonstrate the following theorem.

Theorem 3.7 (Multi-dimensional intersection formula). *Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{E}$ with density $0 < d < 1$. Let $\boldsymbol{X} = (X_n)$ be a sequence of (fixed) subsets of $\boldsymbol{E}^{(k)}$ with density α.*

(i) *If $d + \alpha < 1$, then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and*

$$\mathrm{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = -\infty.$$

(ii) *If $d + \alpha > 1$ and $\boldsymbol{X}$ satisfies the d-small self intersection condition (Definition 3.6), then $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ is densable and*

$$\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = \alpha + d - 1.$$

Note that when $k = 1$, we have the intersection formula between a random subset and a fixed subset. In this case, the self intersection partition of $\boldsymbol{X}$ contains only $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_1$, and we do not need to check the self intersection condition.

Corollary 3.8 (Random-fixed intersection formula). *Let $\boldsymbol{A}$ be a densable sequence of permutation invariant random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X}$ be a sequence of (fixed) subsets of $\boldsymbol{E}$ with density α. If $d + \alpha \neq 1$, then the sequence of random subsets $\boldsymbol{A} \cap \boldsymbol{X}$ is densable and*

$$\operatorname{dens}(\boldsymbol{A} \cap \boldsymbol{X}) = \begin{cases} d + \alpha - 1 & \text{if } d + \alpha > 1, \\ -\infty & \text{if } d + \alpha < 1. \end{cases}$$

We shall first represent the expected value and the variance of the random variable $|A_n^{(k)} \cap X_n|$ by probabilities of the type $\mathbf{Pr}(\{x_1, \ldots, x_r\} \subset A_n)$. The following result generalizes Lemma 2.3.

Lemma 3.9. *Let $\boldsymbol{E}$, $\boldsymbol{A}$ and $\boldsymbol{X}$ be given as in Theorem 3.7 and let $(\boldsymbol{Y}_i)_{0 \le i \le k}$ be the self-intersection partition of $\boldsymbol{X}$. Let $x_1, \ldots, x_{2k}$ be distinct $2k$ elements of E_n.*

(i) $\mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) = |X_n| \mathbf{Pr}\big(\{x_1, \ldots, x_k\} \subset A_n\big)$.

(ii) $\operatorname{Var}\big(|A_n^{(k)} \cap X_n|\big)$

$$= |X_n|^2 \Big(\mathbf{Pr}\big(\{x_1, \ldots, x_{2k}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \ldots, x_k\} \subset A_n\big)^2\Big) + \sum_{i=1}^{k} |Y_{i,n}| \Big(\mathbf{Pr}\big(\{x_1, \ldots, x_{2k-i}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \ldots, x_{2k}\} \subset A_n\big)\Big).$$

Proof. We consider each item.

(i) As A_n is permutation invariant, the probability $\mathbf{Pr}(\{x_1, \ldots, x_k\} \subset A_n)$ does not depend on the choice of $\{x_1, \ldots, x_k\}$. So,

$$\mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) = \mathbb{E}\Bigg(\sum_{x \in X_n} \mathbb{1}_{x \in A_n^{(k)}}\Bigg) = \sum_{x \in X_n} \mathbf{Pr}\big(x \in A_n^{(k)}\big) = |X_n| \mathbf{Pr}\big(\{x_1, \ldots, x_k\} \subset A_n\big).$$

(ii) By the same reason $\mathbf{Pr}(\{x_1, \ldots, x_r\} \subset A_n)$ does not depend on the choice of $\{x_1, \ldots, x_r\}$ for all $r \in \mathbb{N}$. Note that

$$\operatorname{Var}\big(|A_n^{(k)} \cap X_n|\big) = \mathbb{E}\big(|A_n^{(k)} \cap X_n|^2\big) - \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big)^2.$$

If $(x, y) \in Y_{i,n}$, then there are $2k - i$ different elements of E_n in x and y, so

$$\mathbf{Pr}\big(x, y \in A_n^{(k)}\big) = \mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big).$$

Hence,

$$\begin{aligned}\mathbb{E}\big(|A_n^{(k)} \cap X_n|^2\big) &= \mathbb{E}\Bigg(\Big(\sum_{x \in X_n} \mathbb{1}_{x \in A_n^{(k)}}\Big)^2\Bigg) = \sum_{x,y \in X_n} \mathbf{Pr}\big(x, y \in A_n^{(k)}\big) \\ &= \sum_{i=0}^{k} \sum_{(x,y)\in Y_{i,n}} \mathbf{Pr}\big(x, y \in A_n^{(k)}\big) \\ &= \sum_{i=0}^{k} |Y_{i,n}| \mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big).\end{aligned}$$

Recall that $|Y_{0,n}| = |X_n^2| - \sum_{i=1}^{k} |Y_{i,n}|$. The above can be rewritten as

$$\begin{aligned}\mathbb{E}\big(|A_n^{(k)} \cap X_n|^2\big) &= \Big(|X_n^2| - \sum_{i=1}^{k} |Y_{i,n}|\Big)\mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) \\ &\quad + \sum_{i=1}^{k} |Y_{i,n}| \mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big) \\ &= |X_n^2| \mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) \\ &\quad + \sum_{i=1}^{k}\Big(\mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big)\Big).\end{aligned}$$

Combined with $\mathbb{E}(|A_n^{(k)} \cap X_n|)^2 = |X_n|^2 \mathbf{Pr}(\{x_1, \dots, x_k\} \subset A_n)^2$, we have

$$\begin{aligned}\mathrm{Var}\big(|A_n^{(k)} \cap X_n|\big) &= |X_n|^2\Big(\mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \dots, x_k\} \subset A_n\big)^2\Big) \\ &\quad + \sum_{i=1}^{k} |Y_{i,n}|\Big(\mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big)\Big).\end{aligned}$$ ∎

Remark that Lemma 2.3 is a special case of Lemma 3.9, by taking $k = 1$ and $X_n = E_n$. Note that if $k = 1$, then $\boldsymbol{X}^2 = \boldsymbol{Y}_0 \sqcup \boldsymbol{Y}_1$ and there is no need to introduce condition (3.6).

3.2. The Bernoulli density model

Let $\boldsymbol{X}$ be a fixed sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α. In this subsection, we study the intersection $\boldsymbol{A}^{(k)} \cap \boldsymbol{X}$ in the case that $\boldsymbol{A}$ is a sequence of Bernoulli random subsets of $\boldsymbol{E}$ with density $0 < d < 1$. Note that for any integer $r \in \mathbb{N}$ and any distinct elements $x_1, \ldots, x_r$ in E_n, we have

$$\begin{aligned}\mathbf{Pr}\big(\{x_1, \ldots, x_r\} \subset A_n\big) &= \mathbf{Pr}\big(\{x_1 \in A_n\}, \ldots, \{x_r \in A_n\}\big)\\ &= \prod_{i=1}^{r} \mathbf{Pr}(x_i \in A_n) = n^{r(d-1)}\end{aligned}$$

by independence of the events $\mathbf{Pr}(x_i \in A_n)$. Because of this equality, the proof of Theorem 3.7 for the Bernoulli density model is much simpler.

Proof of Theorem 3.7 for Bernoulli density model. We consider each item.

(i) Suppose that $\alpha + d < 1$. To prove that $\operatorname{dens}(\boldsymbol{A}^{(k)} \cap \boldsymbol{X}) = -\infty$, it is enough to prove that

$$\mathbf{Pr}\big(A_n^{(k)} \cap X_n \neq \emptyset\big) \xrightarrow[n\to\infty]{} 0.$$

By Markov's inequality and Lemma 3.9,

$$\begin{aligned}\mathbf{Pr}\big(A_n^{(k)} \cap X_n \neq \emptyset\big) &= \mathbf{Pr}\big(|A_n^{(k)} \cap X_n| \geq 1\big)\\ &\leq \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) = |X_n| \mathbf{Pr}\big(\{x_1, \ldots, x_k\} \subset A_n\big)\\ &\leq n^{k\alpha + o(1)} n^{k(d-1)}\\ &\leq n^{k(\alpha+d-1)+o(1)} \xrightarrow[n\to\infty]{} 0\end{aligned}$$

as $\alpha + d - 1 < 0$.

(ii) Suppose that $\alpha + d > 1$. To simplify the notation, denote $B_n = A_n^{(k)} \cap X_n$ and $\boldsymbol{B} = \boldsymbol{X} \cap \boldsymbol{A}^{(k)}$.

We shall prove that $\operatorname{dens} \boldsymbol{B} = \alpha + d - 1$. Let $\varepsilon > 0$ be an arbitrary small real number. We need prove that a.a.s.

$$n^{k(\alpha+d-1-\varepsilon)} \leq |B_n| \leq n^{k(\alpha+d-1+\varepsilon)}.$$

By Lemma 3.9,

$$\begin{aligned}\mathbb{E}\big(|B_n|\big) &= |X_n| \mathbf{Pr}\big(\{x_1, \ldots, x_k\} \subset A_n\big) = |X_n| n^{k(d-1)}\\ &= n^{k(\alpha+d-1)+o(1)}.\end{aligned}$$

For n large enough,

$$n^{k(\alpha+d-1-\varepsilon)} < \frac{1}{2} n^{k(\alpha+d-1)+o(1)} < \frac{3}{2} n^{k(\alpha+d-1)+o(1)} < n^{k(\alpha+d-1+\varepsilon)}.$$

So it is enough to prove that a.a.s.

$$\frac{1}{2}\mathbb{E}\big(|B_n|\big) < |B_n| < \frac{3}{2}\mathbb{E}\big(|B_n|\big),$$

which means that a.a.s.

$$\big||B_n| - \mathbb{E}\big(|B_n|\big)\big| < \frac{1}{2}\mathbb{E}\big(|B_n|\big).$$

By Chebyshev's inequality,

$$\mathbf{Pr}\Big(\big||B_n| - \mathbb{E}\big(|B_n|\big)\big| \geq \frac{1}{2}\mathbb{E}\big(|B_n|\big)\Big) \leq \frac{4\,\mathrm{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2}.$$

We shall prove that this quantity goes to zero when n goes to infinity. By Lemma 3.9,

$$\begin{aligned}\mathrm{Var}\big(|B_n|\big) &= |X_n|^2\big(\mathbf{Pr}\big(\{x_1,\dots,x_{2k}\}\subset A_n\big) - \mathbf{Pr}\big(\{x_1,\dots,x_k\}\subset A_n\big)^2\big) \\ &\quad + \sum_{i=1}^{k}|Y_{i,n}|\big(\mathbf{Pr}\big(\{x_1,\dots,x_{2k-i}\}\subset A_n\big) - \mathbf{Pr}\big(\{x_1,\dots,x_{2k}\}\subset A_n\big)\big) \\ &= \sum_{i=1}^{k}|Y_{i,n}|\big(n^{(2k-i)(d-1)} - n^{2k(d-1)}\big) \\ &\leq \sum_{i=1}^{k}|Y_{i,n}|n^{(2k-i)(d-1)}.\end{aligned}$$

Note that $n^{(2k-i)(d-1)} > n^{2k(d-1)}$ because $d < 1$. By the d-small self-intersection condition (3.6), there exists $\varepsilon > 0$ such that for all $1 \leq i \leq k$,

$$|Y_{i,n}| \leq n^{2k(\alpha+(d-1)i/2k)-\varepsilon}$$

for n large enough. Hence, for n large enough

$$\mathrm{Var}\big(|B_n|\big) \leq k n^{2k(\alpha+d-1)-\varepsilon}.$$

Recall that $\mathbb{E}(|B_n|)^2 = n^{2k(\alpha+d-1)+o(1)}$, so

$$\frac{4\,\mathrm{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2} \xrightarrow[n\to\infty]{} 0. \qquad \blacksquare$$

3.3. The uniform density model

Note that when $\boldsymbol{A}$ is a sequence of Bernoulli random subsets with density d, we have

$$\mathbf{Pr}\big(\{x_1,\dots,x_r\}\subset A_n\big) = n^{r(d-1)},$$

and consequently,

$$\mathbf{Pr}\big(\{x_1,\dots,x_k\}\subset A_n\big)^2 - \mathbf{Pr}\big(\{x_1,\dots,x_{2k}\}\subset A_n\big) = 0.$$

In order to proceed the same proof, we shall estimate these two quantities for the uniform density model.

Lemma 3.10. *Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density d. Let $0 < \varepsilon < d$ be a small real number and let $k \geq 1$ be an integer. If $n \geq (1 + 2k)^{1/\varepsilon}$, then*

(i) *for all integers $1 \leq r \leq 2k$,*

$$n^{r(d-1-\varepsilon)} \leq \mathbf{Pr}\big(\{x_1, \dots, x_r\} \subset A_n\big) \leq n^{r(d-1+\varepsilon)};$$

(ii) $0 \leq \mathbf{Pr}\big(\{x_1, \dots, x_k\} \subset A_n\big)^2 - \mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) \leq n^{2k(d-1+\varepsilon)-d}.$

Proof. Recall that $|E_n| = n$ and that A_n is uniform on all subsets of E_n of cardinality $\lfloor n^d \rfloor$.

(i) Note that $\lfloor n^d \rfloor \geq n^\varepsilon - 1 \geq 2k \geq r$. Among all subsets of E_n of cardinality $\lfloor n^d \rfloor$, there are $\binom{n-r}{\lfloor n^d \rfloor - r}$ subsets that include $\{x_1, \dots, x_r\}$. So,

$$\mathbf{Pr}\big(\{x_1, \dots, x_r\} \subset A_n\big) = \frac{\binom{n-r}{\lfloor n^d \rfloor - r}}{\binom{n}{\lfloor n^d \rfloor}} = \frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - r + 1)}{n \dots (n - r - 1)}.$$

We estimate that

$$\left(\frac{n^d - r}{n}\right)^r \leq \frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - r + 1)}{n \dots (n - r - 1)} \leq \left(\frac{n^d}{n - r}\right)^r.$$

The condition $n \geq (1 + 2k)^{1/\varepsilon} \geq (1 + r)^{1/\varepsilon}$ implies

$$\begin{cases} n \geq n^{1-\varepsilon}(1 + r), \\ n^d \geq n^{d-\varepsilon}(1 + r), \end{cases}$$

so

$$\begin{cases} n^{1-\varepsilon} \leq n - r, \\ n^{d-\varepsilon} \leq n^d - r. \end{cases}$$

Hence,

$$\big(n^{d-1-\varepsilon}\big)^r \leq \frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - r + 1)}{n \dots (n - r - 1)} \leq \big(n^{d-1+\varepsilon}\big)^r.$$

(ii) By the same argument,

$$\begin{aligned}
&\mathbf{Pr}\big(\{x_1, \dots, x_k\} \subset A_n\big)^2 - \mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) \\
&\quad = \left(\frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - k + 1)}{n \dots (n - k - 1)}\right)^2 - \frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - 2k + 1)}{n \dots (n - 2k - 1)} \\
&\quad = \left(\frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - k + 1)}{n \dots (n - k - 1)}\right) \\
&\qquad \cdot \left(\frac{\lfloor n^d \rfloor \dots (\lfloor n^d \rfloor - k + 1)}{n \dots (n - k - 1)} - \frac{(\lfloor n^d \rfloor - k) \dots (\lfloor n^d \rfloor - 2k + 1)}{(n - k) \dots (n - 2k - 1)}\right).
\end{aligned}$$

This quantity is positive because $\frac{\lfloor n^d \rfloor - i}{n-i} \geq \frac{\lfloor n^d \rfloor - i - k}{n-i-k}$ for every $0 \leq i \leq k-1$.

Now we estimate that

$$\begin{aligned}
&\mathbf{Pr}\big(\{x_1,\ldots,x_k\} \subset A_n\big)^2 - \mathbf{Pr}\big(\{x_1,\ldots,x_{2k}\} \subset A_n\big) \\
&\quad \leq \left(\frac{n^d}{n-k}\right)^k \left(\frac{n^{dk}}{(n-k)^k} - \frac{(n^d-2k)^k}{(n-k)^k}\right) \\
&\quad \leq \frac{n^{dk}}{(n-k)^{2k}} \left(n^{dk} - \sum_{i=0}^{k} \binom{k}{i} n^{d(k-i)} (-2k)^i\right) \\
&\quad \leq \frac{n^{dk}}{(n-k)^{2k}} (1+2k)^k n^{d(k-1)} = \left(\frac{n^d \sqrt{1+2k}}{n-k}\right)^{2k} n^{-d}.
\end{aligned}$$

As $n^\varepsilon \geq 1+2k$, we have

$$\begin{aligned}
n-k &\geq n^{1-\varepsilon}(1+2k) - k \\
&\geq n^{1-\varepsilon}(1+k) \\
&\geq n^{1-\varepsilon}\sqrt{1+2k}.
\end{aligned}$$

Hence,

$$\mathbf{Pr}\big(\{x_1,\ldots,x_k\} \subset A_n\big)^2 - \mathbf{Pr}\big(\{x_1,\ldots,x_{2k}\} \subset A_n\big) \leq n^{2k(d-1+\varepsilon)-d}. \qquad \blacksquare$$

Notation. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α and let $(\boldsymbol{Y}_i)_{0\leq i\leq k}$ be its self-intersection partition. Denote the density difference

$$\varepsilon_0(d) = \min_{1\leq i\leq k} \left\{\alpha + (d-1)\frac{i}{2k} - \overline{\text{dens}}\boldsymbol{Y}_i\right\}.$$

Remark that $\boldsymbol{X}$ has d-small self-intersection if and only if $\varepsilon_0(d) > 0$. In addition, for every small real number $0 < \varepsilon < \varepsilon_0(d)/10$ there exists $n_\varepsilon \in \mathbb{N}$ such that for all $n \geq n_\varepsilon$ we have, simultaneously for all $1 \leq i \leq k$,

$$|Y_{n,i}| \leq n^{2k(\alpha+(d-1)i/2k-10\varepsilon)} = n^{2k\alpha+(d-1)i-2k\times 10\varepsilon}.$$

By densability of $\boldsymbol{X}$, we can choose n_ε such that at the same time

$$n^{k(\alpha-\varepsilon)} \leq |X_n| \leq n^{k(\alpha+\varepsilon)}.$$

Combined with Lemma 3.10, we can now estimate the expected value and the variance of $|A_n^{(k)} \cap X_n|$ for the uniform density model.

Lemma 3.11. *Let $\boldsymbol{A}$ be a sequence of uniform random subsets of $\boldsymbol{E}$ with density d. Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ with density α. Let $0 < \varepsilon < \min\{\varepsilon_0(d)/10, d\}$ be a small real number. If $n \geq \max\{n_\varepsilon, (1+2k)^{1/\varepsilon}\}$, then*

(i) $n^{k(\alpha+d-1-2\varepsilon)} \leq \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) \leq n^{k(\alpha+d-1+2\varepsilon)}$.

(ii) *If in addition* $\alpha + d - 1 > 2\varepsilon > 0$ *and* $\boldsymbol{X}$ *has* d*-small self-intersection, then*

$$\mathrm{Var}\big(|A_n^{(k)} \cap X_n|\big) \leq k n^{2k(\alpha+d-1-9\varepsilon)}.$$

Proof. We consider each item.

(i) By Lemma 3.9,

$$\mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) = |X_n| \mathbf{Pr}\big(\{x_1, \dots, x_k\} \subset A_n\big).$$

So by Lemma 3.10 and $n^{k(\alpha-\varepsilon)} \leq |X_n| \leq n^{k(\alpha+\varepsilon)}$, we have

$$n^{k(\alpha-\varepsilon)} n^{k(d-1-\varepsilon)} \leq \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) \leq n^{k(\alpha+\varepsilon)} n^{k(d-1+\varepsilon)}.$$

(ii) By Lemma 3.10 (ii)

$$\mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \dots, x_k\} \subset A_n\big)^2 \leq 0.$$

Apply Lemma 3.9, eliminate negative parts:

$$\begin{aligned}
\mathrm{Var}\big(|A_n^{(k)} \cap X_n|\big) &= |X_n|^2 \Big(\mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \dots, x_k\} \subset A_n\big)^2\Big) \\
&\quad + \sum_{i=1}^{k} |Y_{i,n}| \Big(\mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big) - \mathbf{Pr}\big(\{x_1, \dots, x_{2k}\} \subset A_n\big)\Big) \\
&\leq \sum_{i=1}^{k} |Y_{i,n}| \mathbf{Pr}\big(\{x_1, \dots, x_{2k-i}\} \subset A_n\big).
\end{aligned}$$

By Lemma 3.10 (i) and $|Y_{i,n}| \leq n^{2k\alpha+i(d-1)+2k\times 10\varepsilon}$

$$\begin{aligned}
\mathrm{Var}\big(|A_n^{(k)} \cap X_n|\big) &\leq \sum_{i=1}^{k} n^{2k\alpha+i(d-1)-2k\times 10\varepsilon} n^{(2k-i)(d-1+\varepsilon)} \\
&\leq k n^{2k(\alpha+d-1-9\varepsilon)}.
\end{aligned}$$

∎

Proof of Theorem 3.7 for uniform density model. We consider each item.

(i) Suppose that $\alpha + d < 1$. We shall prove that

$$\mathbf{Pr}\big(A_n^{(k)} \cap X_n \neq \varnothing\big) \xrightarrow[n\to\infty]{} 0.$$

Let $\varepsilon > 0$ such that

$$\varepsilon < \min\Big\{\frac{1-d-\alpha}{2}, \frac{\varepsilon_0(d)}{10}, d\Big\}.$$

By Markov's inequality and Lemma 3.11: If $n \geq \max\{n_\varepsilon, (1+2k)^{1/\varepsilon}\}$, then

$$\begin{aligned}\mathbf{Pr}\big(A_n^{(k)} \cap X_n \neq \emptyset\big) &= \mathbf{Pr}\big(|A_n^{(k)} \cap X_n| \geq 1\big) \\ &\leq \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) \\ &\leq n^{k(\alpha+d-1+2\varepsilon)} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

(ii) Suppose that $\alpha + d > 1$. Denote $B_n = A_n^{(k)} \cap X_n$. Let $\varepsilon > 0$ be an arbitrary small number, with

$$\varepsilon < \min\Big\{\frac{\alpha+d-1}{3}, \frac{\varepsilon_0(d)}{10}, d\Big\}.$$

We shall prove that a.a.s.

$$n^{k(\alpha+d-1-3\varepsilon)} \leq |B_n| \leq n^{k(\alpha+d-1+3\varepsilon)}.$$

By Lemma 3.11, if $n \geq \max\{n_\varepsilon, (1+2k)^{1/\varepsilon}\}$, then

$$n^{k(\alpha+d-1-2\varepsilon)} \leq \mathbb{E}\big(|B_n|\big) \leq n^{k(\alpha+d-1+2\varepsilon)}.$$

In addition, if $n \geq 2^{1/k\varepsilon}$, we have

$$n^{k(\alpha+d-1-3\varepsilon)} \leq \frac{1}{2}n^{k(\alpha+d-1-2\varepsilon)} \leq \frac{1}{2}\mathbb{E}\big(|B_n|\big)$$

and

$$\frac{3}{2}\mathbb{E}\big(|B_n|\big) \leq \frac{3}{2}n^{k(\alpha+d-1+2\varepsilon)} \leq n^{k(\alpha+d-1+3\varepsilon)}.$$

So, it is enough to prove that a.a.s.

$$\big||B_n| - \mathbb{E}\big(|B_n|\big)\big| \leq \frac{1}{2}\mathbb{E}\big(|B_n|\big).$$

By Chebyshev's inequality,

$$\mathbf{Pr}\Big(\big||B_n| - \mathbb{E}\big(|B_n|\big)\big| > \frac{1}{2}\mathbb{E}\big(|B_n|\big)\Big) \leq \frac{4\,\mathrm{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2}.$$

Combined with Lemma 3.11, if $n \geq \max\{n_\varepsilon, (1+2k)^{1/\varepsilon}, 2^{1/k\varepsilon}\}$, then

$$\frac{4\,\mathrm{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2} \leq \frac{4k\,n^{2k(\alpha+d-1-9\varepsilon)}}{n^{2k(\alpha+d-1-2\varepsilon)}} \leq \frac{4k}{n^{14k\varepsilon}} \xrightarrow[n\to\infty]{} 0. \qquad \blacksquare$$

3.4. The general model (densable and permutation invariant)

Let $\boldsymbol{X}$ be a sequence of subsets of $\boldsymbol{E}^{(k)}$ having the d-small intersection condition. Recall that

$$\varepsilon_0(d) = \min_{1\le i\le k}\left\{\alpha + (d-1)\frac{i}{2k} - \overline{\operatorname{dens}}\boldsymbol{Y}_i\right\} > 0.$$

Note that if $d' < d$ then $\varepsilon(d') < \varepsilon(d)$.

In order to apply Lemma 3.11 in a small interval $[d-\varepsilon, d+\varepsilon]$, we choose

$$0 < \varepsilon < \min\left\{\frac{\varepsilon_0(d)}{20}, \frac{d}{2}\right\},$$

so that

$$\varepsilon < \min\left\{\frac{\varepsilon_0(d-\varepsilon)}{10}, d-\varepsilon\right\} \le \min\left\{\frac{\varepsilon_0(d')}{10}, d'\right\}$$

for every $d' \in [d-\varepsilon, d+\varepsilon]$.

By the definition of ε_0 and the densability of $\boldsymbol{X}$, we choose again $n_\varepsilon \in \mathbb{N}$ such that for all $n \ge n_\varepsilon$,

$$|Y_{n,i}| \le n^{2k\alpha+(d-1)i-2k\times 20\varepsilon} \le n^{2k\alpha+(d'-1)i-2k\times 10\varepsilon} \quad \forall 1 \le i \le k$$

and

$$n^{k(\alpha-\varepsilon)} \le |X_n| \le n^{k(\alpha+\varepsilon)}.$$

Lemma 3.12. *Let* $0 < \varepsilon < \min\{\varepsilon_0(d)/20, d/2\}$ *be a small real number. Let* $\boldsymbol{A}$ *be a sequence of uniform random subsets of* $\boldsymbol{E}$ *with density* $d' \in [d-\varepsilon, d+\varepsilon]$. *Let* $\boldsymbol{X}$ *be a sequence of subsets of* $\boldsymbol{E}^{(k)}$ *with density* α. *If* $n \ge \max\{n_\varepsilon, (1+2k)^{1/\varepsilon}\}$, *then*

(i) $n^{k(\alpha+d-1-3\varepsilon)} \le \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) \le n^{k(\alpha+d-1+3\varepsilon)}$.

(ii) *If in addition* $\alpha + d - 1 > 3\varepsilon > 0$ *and* $\boldsymbol{X}$ *has* d*-small self-intersection, then*

$$\operatorname{Var}\big(|A_n^{(k)} \cap X_n|\big) \le kn^{2k(\alpha+d-1-8\varepsilon)}.$$

Proof. We consider each item.

(i) Recall from the above discussion that $\varepsilon < \min\{\varepsilon_0(d')/10, d'\}$. By Lemma 3.11,

$$n^{k(\alpha+d'-1-2\varepsilon)} \le \mathbb{E}\big(|A_n^{(k)} \cap X_n|\big) \le n^{k(\alpha+d'-1+2\varepsilon)}.$$

We then have the inequality by $d-\varepsilon \le d' \le d+\varepsilon$.

(ii) Because $\varepsilon_0(d') > 0$, $\boldsymbol{X}$ has d'-small self-intersection. By Lemma 3.11 and the fact that $d' \le d+\varepsilon$,

$$\operatorname{Var}\big(|A_n^{(k)} \cap X_n|\big) \le kn^{2k(\alpha+d'-1-9\varepsilon)} \le kn^{2k(\alpha+d-1-8\varepsilon)}. \qquad \blacksquare$$

Lemma 3.13 (Concentration lemma). *Let $\varepsilon > 0$ be an arbitrary small real number. Let $\boldsymbol{A}$ and $\boldsymbol{X}$ given as the previous lemma, with $\alpha + d - 1 > 4\varepsilon > 0$ and $\boldsymbol{X}$ having d-small self-intersection. If $\varepsilon < \min\{\varepsilon_0/10, d/2\}$ and $n \geq \max\{n_\varepsilon, (1+2k)^{1/\varepsilon}\}$, then*

$$\mathbf{Pr}\big(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\big) > 1 - kn^{-10k\varepsilon}.$$

Proof. Denote $B_n = A_n^{(k)} \cap X_n$. By Lemma 3.12 (i) and $n^{k\varepsilon} \geq 2$, we have

$$n^{k(\alpha+d-1-4\varepsilon)} \leq \frac{1}{2} n^{k(\alpha+d-1-3\varepsilon)} \leq \frac{1}{2}\mathbb{E}\big(|B_n|\big)$$

and

$$\frac{3}{2}\mathbb{E}\big(|B_n|\big) \leq \frac{3}{2} n^{k(\alpha+d-1+3\varepsilon)} \leq n^{k(\alpha+d-1+4\varepsilon)}.$$

By Chebyshev's inequality,

$$\begin{aligned}\mathbf{Pr}\big(n^{k(\alpha+d-1-4\varepsilon)} \leq |B_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\big) &\geq \mathbf{Pr}\Big(\big||B_n| - \mathbb{E}\big(|B_n|\big)\big| \leq \frac{1}{2}\mathbb{E}\big(|B_n|\big)\Big) \\ &\geq 1 - \frac{4\,\mathrm{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2}.\end{aligned}$$

Again by Lemma 3.12,

$$\frac{4\,\mathrm{Var}(|B_n|)}{\mathbb{E}(|B_n|)^2} \leq \frac{kn^{2k(\alpha+d-1-8\varepsilon)}}{n^{2k(\alpha+d-1-3\varepsilon)}} \leq kn^{-10k\varepsilon}.$$ ∎

Proof of the Theorem 3.7. Let $\varepsilon > 0$ be an arbitrary small number with $\varepsilon < \min\{d/2, \varepsilon_0(d)/20\}$ as given in Lemma 3.12. Denote $Q_n = \{n^{d-\varepsilon} \leq |A_n| \leq n^{d+\varepsilon}\}$ and

$$\mathbb{N}_{\boldsymbol{A},\varepsilon,n} := \big\{\ell \in \mathbb{N} \mid n^{d-\varepsilon} \leq \ell \leq n^{d+\varepsilon} \text{ and } \mathbf{Pr}\big(|A_n| = \ell\big) > 0\big\}.$$

By densability of $\boldsymbol{A}$, we have $\mathbf{Pr}(Q_n) \xrightarrow[n\to\infty]{} 1$. Denote by $\mathbf{Pr}_{Q_n} := \mathbf{Pr}(\cdot \mid Q_n)$ the probability measure under the condition Q_n. Define similarly $\mathbb{E}_{Q_n}$ and Var_{Q_n}.

In order to prove that some sequence of properties (R_n) is a.a.s. true, by the inequality

$$\mathbf{Pr}(\overline{R_n}) \leq \mathbf{Pr}(Q_n)\mathbf{Pr}_{Q_n}(\overline{R_n}) + \mathbf{Pr}(\overline{Q_n}),$$

it is enough to prove that $\mathbf{Pr}_{Q_n}(\overline{R_n}) \xrightarrow[n\to\infty]{} 0$.

(i) Suppose that $\alpha + d < 1$. Assume in addition that $\varepsilon < (1-d-\alpha)/3$. We shall prove that

$$\mathbf{Pr}_{Q_n}\big(A_n^{(k)} \cap X_n \neq \emptyset\big) = \mathbf{Pr}_{Q_n}\big(|A_n^{(k)} \cap X_n| \geq 1\big) \xrightarrow[n\to\infty]{} 0.$$

By the formula of total probability and Markov's inequality,

$$\begin{aligned}\mathbf{Pr}_{Q_n}\big(|A_n^{(k)} \cap X_n| \geq 1\big) &\leq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l)\mathbf{Pr}\big(|A_n^{(k)} \cap X_n| \geq 1 \mid |A_n| = l\big) \\ &\leq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l)\mathbb{E}\big(|A_n^{(k)} \cap X_n| \mid |A_n| = l\big).\end{aligned}$$

By a change of variable $l = n^{d'}$ with $d - \varepsilon \leq d' \leq d + \varepsilon$, apply Lemma 3.12:

$$\begin{aligned}\mathbf{Pr}_{Q_n}\big(|A_n^{(k)} \cap X_n| \geq 1\big) &\leq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l = n^{d'})n^{\alpha+d-1+3\varepsilon} \\ &\leq n^{\alpha+d-1+3\varepsilon} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

(ii) Suppose that $\alpha + d > 1$. Assume in addition that $\varepsilon < (\alpha + d - 1)/4$, so that we can apply Lemma 3.13. We shall prove that

$$\mathbf{Pr}_{Q_n}\big(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\big) \xrightarrow[n\to\infty]{} 1.$$

By the formula of total probability, Lemma 3.13 and a change of variables $l = n^{d'}$:

$$\begin{aligned}&\mathbf{Pr}_{Q_n}\big(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)}\big) \\ &\quad = \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l) \\ &\qquad\quad \cdot \mathbf{Pr}\big(n^{k(\alpha+d-1-4\varepsilon)} \leq |A_n^{(k)} \cap X_n| \leq n^{k(\alpha+d-1+4\varepsilon)} \mid |A_n| = l\big) \\ &\quad \geq \sum_{l \in \mathbb{N}_{A,\varepsilon,n}} \mathbf{Pr}_{Q_n}(A_n = l = n^{d'})\big(1 - kn^{-10k\varepsilon}\big) \\ &\quad \geq 1 - kn^{-10k\varepsilon} \xrightarrow[n\to\infty]{} 1.\end{aligned}$$

■

4. Applications to group theory

Fix an alphabet $X = \{x_1, \ldots, x_m\}$ as generators of groups. Let B_ℓ be the set of cyclically reduced words of length at most ℓ on $X^{\pm}$. Recall that $|B_\ell| = (2m-1)^{\ell+O(1)}$.

We are interested in asymptotic behaviors, when ℓ goes to infinity, of group presentations $\langle X | R_\ell \rangle$ where R_ℓ is a random subset of B_ℓ.

Definition 4.1 (Random groups with density). Let $d \in]0, 1]$. Let $\boldsymbol{R} = (R_\ell)$ be a densable sequence of permutation invariant random subsets with density d of the sequence $\boldsymbol{B} = (B_\ell)$.

Denote $G_\ell = G_\ell(m, d)$ the random presentation defined by $\langle X | R_\ell \rangle$. The sequence $\boldsymbol{G} = \boldsymbol{G}(m, d) = (G_\ell(m, d))_{\ell \in \mathbb{N}}$ is called a sequence of random groups with density d.

For example, if $d = 1$, then $G_\ell(m, 1)$ is isomorphic to the trivial group.

A sequence of events $\boldsymbol{Q} = (Q_\ell)$ described by $\boldsymbol{G}(m, d)$ is *asymptotically almost surely* satisfied if $\mathbf{Pr}(Q_\ell) \xrightarrow[m\to\infty]{} 1$. We denote briefly a.a.s. Q_ℓ.

4.1. Phase transition at density 1/2

Theorem 4.2 (Gromov, phase transition at density 1/2). *Let $(G_\ell(m,d))$ be a sequence of random groups with density d.*

(i) *If $d > 1/2$, then a.a.s $G_\ell(m, d)$ is isomorphic to the trivial group.*

(ii) *If $d < 1/2$, then a.a.s $G_\ell(m, d)$ is a hyperbolic group.*

In [15, Section 2.1] (or [16, Section I.2.b]), Ollivier proved the first assertion by probabilistic pigeon-hole principle. We give a proof here by the intersection formula (Theorem 2.9 and Corollary 3.8).

Proof of Theorem 4.2 (i). Let $x \in X$. Let A_ℓ be the set of cyclically reduced words that does not start or end by x, of lengths at most $\ell - 1$ (so that $xA_\ell \subset B_\ell$). It is easy to check that the sequences (A_ℓ) and (xA_ℓ) are sequences of fixed subsets of $\boldsymbol{B} = (B_\ell)$ of density 1. By the random-fixed intersection formula (Corollary 3.8), the sequences $(x(R_\ell \cap A_\ell))$ and $(R_\ell \cap xA_\ell)$ are sequences of permutation invariant random subsets of (xA_ℓ) of density d.

By the intersection formula (Theorem 2.9), their intersection $(xR_\ell \cap R_\ell \cap xA_\ell)$ is a sequence of permutation invariant random subsets of (xA_ℓ) of density $(2d - 1) > 0$, which is a.a.s. not empty. Thus, a.a.s. there exists a word $w \in A_\ell$ such that $w \in R_\ell$ and $xw \in R_\ell$, so a.a.s. $x = 1$ in G_ℓ by canceling w.

The argument above works for any generator $x \in X$. By intersecting a finite number of a.a.s. satisfied events, a.a.s. all generators $x \in X$ are trivial in G_ℓ. Hence, a.a.s. G_ℓ is isomorphic to the trivial group. ■

The proof of Theorem 4.2 (ii) needs *van Kampen diagrams* and will not be treated here. See [9, Section 9.B] for the original idea by Gromov, and [15, Section 2.2] or [16, Section V] for a detailed proof by Ollivier.

4.2. Phase transition at density $\lambda/2$

Theorem 4.3. *Let $\boldsymbol{G}(m, d) = (G_\ell(m, d))$ be a sequence of random groups with density d. Let $\lambda \in]0, 1[$.*

(i) *If $d < \lambda/2$, then a.a.s. $G_\ell(m, d)$ satisfies $C'(\lambda)$.*

(ii) *If $d > \lambda/2$, then a.a.s. $G_\ell(m, d)$* does not *satisfy $C'(\lambda)$.*

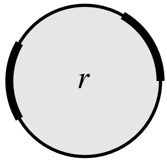

Figure 1. A cyclically reduced word having a piece that appears twice.

Proof. We consider each item.

(i) Recall that (Lyndon–Schupp [13, p. 240]) a *piece* with respect to a set of relators is a cyclic sub-word that appears at least twice. There are two cases to verify.

(a) Let A_ℓ be the set of cyclically reduced words of length at most ℓ having a piece appearing twice on itself (Figure 1) that is longer than λ times itself. We shall prove that a.a.s. the intersection $A_\ell \cap R_\ell$ is empty. We estimate first the number of relators of length $t \leq \ell$ with a piece of length $s \geq \lambda t$. There are $2t$ ways (including orientations) to choose the first position of the piece, and $2t - s$ ways the choose the second position (note that because r is reduced, it can not overlay the first one if they are with opposite orientations). For each way of positioning we can determine freely $t - s$ letters, each with $(2m - 1)$ choices, except for the first letter and the last letter having respectively $2m$ and $2m - 2$ or $2m - 1$ choices. So this number is $2t(2t - s)C(m)(2m - 1)^{t-s}$, where $C(m)$ is a real number that depends only on m. Hence,

$$|A_\ell| = \sum_{t=1}^{\ell} \sum_{s=\lfloor \lambda t \rfloor}^{t} 2t(2t - s)C(m)(2m - 1)^{t-s} = (2m - 1)^{(1-\lambda)\ell + o(\ell)},$$

which means that (A_ℓ) is a sequence of fixed subsets of (B_ℓ) with density $1 - \lambda$. By the intersection formula (Corollary 3.8), because $1 - \lambda + d < 1$, we have a.a.s.

$$A_\ell \cap R_\ell = \emptyset.$$

(b) Let X_ℓ be the set of distinct pairs of relators r_1, r_2 in B_ℓ having a piece (Figure 2) longer than $\lambda \min\{|r_1|, |r_2|\}$. It is a fixed subset of $B_\ell^{(2)}$. We shall prove that a.a.s the intersection $X_\ell \cap R_\ell^{(2)}$ is empty. There are $4\ell^2$ possible positions for pieces, $(2m - 1)^{\ell + o(\ell)}$ choices for r_1 and $(2m - 1)^{\ell - \lambda\ell + o(\ell)}$ choices for r_2. So,

$$|X_\ell| = (2m - 1)^{(2-\lambda)\ell + o(\ell)},$$

which means that (X_ℓ) is a sequence of fixed subsets of $(B_\ell^{(2)})$ with density $1 - \frac{\lambda}{2}$. By the multi-dimension intersection formula (Theorem 3.7 (i)), because $1 - \frac{\lambda}{2} + d < 1$, we have a.a.s.

$$X_\ell \cap R_\ell^{(2)} = \emptyset.$$

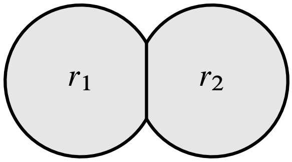

Figure 2. A pair of relators sharing a common subword (a piece).

(ii) Take the sequence of sets $\boldsymbol{X} = (X_\ell)$ constructed in part 1 (b). We shall prove that a.a.s. the intersection $X_\ell \cap R_\ell^{(2)}$ is *not* empty. We have already

$$\text{dens}\, \boldsymbol{X} + \text{dens}\, \boldsymbol{R}^{(2)} > 1.$$

To apply Theorem 3.7 (ii), we need to calculate the size of the self-intersection

$$Y_{1,\ell} = \big\{(x_1, x_2) \in X_\ell^2 \mid |x_1 \cap x_2| = 1\big\}.$$

Take $x_1 = (r_1, r_2)$ and $x_2 = (r_1, r_3)$, where r_1, r_2, r_3 are three different relators in B_ℓ. There are $(2m-1)^{\ell+o(\ell)}$ choices for r_1, $(2m-1)^{\ell-\lambda\ell+o(\ell)}$ choices for r_2 and $(2m-1)^{\ell-\lambda\ell+o(\ell)}$ choices for r_3. The other three cases ($x_2 = (r_2, r_3)$, (r_3, r_1), or (r_3, r_2)) are symmetric. Multiply these numbers, we have

$$|Y_{1,\ell}| = (2m-1)^{3\ell-2\lambda\ell+o(\ell)}.$$

The density of $\boldsymbol{Y}_1 = (Y_{1,\ell})$ is $(3-2\lambda)/4$ in $(B_\ell^{(2)})^2$. As $d > 0$, we have $(3-2\lambda)/4 < 1 - \frac{\lambda}{2} + \frac{1}{4}(d-1)$, which implies

$$\text{dens}\, \boldsymbol{Y}_1 < \text{dens}\, \boldsymbol{X} + (d-1)\frac{1}{2 \times 2}.$$

Thus, we have the d-small self intersection condition (Definition 3.6). By the multi-dimensional intersection formula, a.a.s.

$$X_\ell \cap R_\ell^{(2)} \neq \varnothing. \qquad \blacksquare$$

4.3. Every $(m-1)$-generated subgroup is free

Fix the set of m generators $X = \{x_1, \ldots, x_m\}$. Recall that B_ℓ is the set of $(2m-1)^{\ell+o(\ell)}$ cyclically reduced words on $X^\pm = \{x_1^\pm, \ldots, x_m^\pm\}$ of length at most ℓ. The few relator model of random groups is constructed as follows: fix a number $k \in \mathbb{N}$ and let

$$G_\ell = \langle x_1, \ldots, x_m \mid r_1, \ldots, r_k \rangle,$$

where $R_\ell = \{r_1, \dots, r_k\}$ is a random subset of B_ℓ given by the uniform probability on all subsets of B_ℓ with cardinality k.

The sequence $(G_\ell)_{\ell \in \mathbb{N}}$ is called a *sequence of random groups with k relators*. As k is independent of ℓ, the sequence (G_ℓ) is a sequence of random groups with density $d = 0$. By Proposition 4.3, a.a.s. G_ℓ satisfies $C'(\lambda)$ for arbitrary small $\lambda > 0$.

In [5], Arzhantseva and Ol'shanskii proved the following result:

Theorem 4.4 (Arzhantseva–Ol'shanskii [5, Theorem 1]). *Let (G_ℓ) be a sequence of random groups with k relators. Then a.a.s. every $(m-1)$-generated subgroup of G_ℓ is free.*

Combining the intersection formula and their arguments, we prove the following theorem.

Theorem 4.5. *Let $(G_\ell(m,d))$ be a sequence of random groups with density $0 \le d < 1/(120m^2 \ln(2m))$. Then a.a.s. every $(m-1)$-generated subgroup of $G_\ell(m,d)$ is free.*

Let us recall the definition of "μ-readable words" in [5, Section 2].

Definition 4.6 ([5, Section 2]). Let $0 < \mu \le 1$. A cyclically reduced word w of length ℓ on $X^\pm$ is μ-readable if there exists a graph Γ marked by $X^\pm$ with the following properties:

(a) the number of edges of Γ is less than $\mu\ell$;

(b) the rank of Γ is at most $m-1$;

(c) the word w can be read along some path of Γ.

Note that the condition (b) is essential, because every word on $X^\pm$ can be read along the wedge of m circles of length 1 marked by $x_1, \dots, x_m$, respectively.

Let M_ℓ^μ be the set of words $r \in B_\ell$ having a cyclic sub-word $w < r$ such that $|w| \ge \frac{1}{2}|r|$ and w is μ-readable. We admit the following two lemmas in [5].

Lemma 4.7 ([5, Lemma 4]). *If $\mu < \log_{2m}(1 + 1/(4m-4))$, then there exists a constant $C(\mu, m)$ such that*

$$|M_\ell^\mu| \le C(\mu, m)\ell^2\Big(2m - \frac{5}{4}\Big)^\ell.$$

Recall that $|B_\ell| = (2m-1)^{\ell + O(1)}$, so (M_ℓ^μ) is a densable sequence of subsets of (B_ℓ) with density $\log_{2m-1}(2m - 5/4)$.

Lemma 4.8 ([5, Section 4]). *Let $G = \langle X | R \rangle$ be a group presentation where $X = \{x_1, \dots, x_m\}$ and R is a subset of B_ℓ. Suppose that*

$$\mu < \log_{2m}\Big(1 + \frac{1}{4m-4}\Big) \quad \text{and} \quad \lambda \le \frac{\mu}{15m + 3\mu}.$$

If R does not intersect M_ℓ^μ, has no true powers, and satisfies $C'(\lambda)$, then every $(m-1)$-generated subgroup of G is free. ■

Proof of Theorem 4.5. We look for a density $d(m) \leq 1/2$ such that for any $d < d(m)$ a.a.s. the random group $G_\ell(m,d) = \langle X | R_\ell \rangle$ satisfies the conditions of Lemma 4.8 with

$$\mu = \log_{2m}\Big(1 + \frac{1}{4m-4}\Big) - \varepsilon \quad \text{and} \quad \lambda = \frac{\mu}{15m + 3\mu}$$

with an arbitrary small $\varepsilon > 0$.

The set of true powers in B_ℓ is with density $1/2$. By the intersection formula (Corollary 3.8), because $d < 1/2$, a.a.s. R_ℓ has no true powers. By Lemma 4.7 and the intersection formula, we need

$$d(m) < 1 - \operatorname{dens}(M_\ell^\mu) < 1 - \log_{2m-1}\Big(2m - \frac{5}{4}\Big),$$

so that a.a.s. R_ℓ does not intersect M_ℓ^μ by the intersection formula.

At the end we need a.a.s. R_ℓ satisfies $C'(\lambda)$ with

$$\lambda = \frac{\log_{2m}(1 + \frac{1}{4m-4}) - \varepsilon}{15m + 3\log_{2m}(1 + \frac{1}{4m-4}) - 3\varepsilon}.$$

By Theorem 4.3, we need $d(m) < \lambda/2$. Note that this inequality implies the previous one. For ε small enough, we have $\lambda > 1/(60m^2 \ln(2m))$. It is enough to take

$$d(m) = \frac{1}{120m^2 \ln(2m)}.$$ ■

Acknowledgments. I would like to thank my supervisor, Thomas Delzant, for his guidance and interesting discussions on the subject, especially for his patience with me while completing this article. I would also like to thank the referee for his/her thorough review of the manuscript and highly appreciate the comments and suggestions, which significantly contributed to improving the quality of this work.

References

[1] S. Antoniuk, T. Łuczak, and J. Świątkowski, Random triangular groups at density 1/3. *Compos. Math.* **151** (2015), no. 1, 167–178 Zbl 1334.20076 MR 3305311

[2] G. N. Arzhantseva, On groups in which subgroups with a fixed number of generators are free. *Fundam. Prikl. Mat.* **3** (1997), no. 3, 675–683 Zbl 0929.20025 MR 1794135

[3] G. N. Arzhantseva, Generic properties of finitely presented groups and Howson's theorem. *Comm. Algebra* **26** (1998), no. 11, 3783–3792 Zbl 0911.20027 MR 1647075

[4] G. N. Arzhantseva, A property of subgroups of infinite index in a free group. *Proc. Amer. Math. Soc.* **128** (2000), no. 11, 3205–3210 Zbl 0976.20014 MR 1694447

[5] G. N. Arzhantseva and A. Y. Ol'shanskiĭ, The class of groups all of whose subgroups with lesser number of generators are free is generic. *Mat. Zametki* **59** (1996), no. 4, 489–496, 638 Zbl 0877.20021 MR 1445193

[6] F. Bassino, C. Nicaud, and P. Weil, Random presentations and random subgroups: A survey. In *Complexity and Randomness in Group Theory*, pp. 45–76, GAGTA 1, de Gruyter, Berlin, 2020 Zbl 7205680

[7] É. Ghys, Groupes aléatoires (d'après Misha Gromov,. . .). *Astérisque* (2004), no. 294, 173–204 Zbl 1134.20306 MR 2111644

[8] M. Gromov, Hyperbolic groups. In *Essays in group theory*, pp. 75–263, Math. Sci. Res. Inst. Publ. 8, Springer, New York, 1987 Zbl 0634.20015 MR 919829

[9] M. Gromov, Asymptotic invariants of infinite groups. In *Geometric group theory, Vol. 2 (Sussex, 1991)*, pp. 269–282, London Math. Soc. Lecture Note Ser. 182, Cambridge Univ. Press, Cambridge, 1993 MR 1253544

[10] V. S. Guba, Conditions under which 2-generated subgroups in small cancellation groups are free. *Izv. Vyssh. Uchebn. Zaved. Mat.* **87** (1986), no. 7, 12–19 Zbl 0614.20022 MR 867603

[11] I. Kapovich and P. Schupp, Genericity, the Arzhantseva-Ol'shanskii method and the isomorphism problem for one-relator groups. *Math. Ann.* **331** (2005), no. 1, 1–19 Zbl 1080.20029 MR 2107437

[12] I. Kapovich and P. Schupp, On group-theoretic models of randomness and genericity. *Groups Geom. Dyn.* **2** (2008), no. 3, 383–404 Zbl 1239.20047 MR 2415305

[13] R. C. Lyndon and P. E. Schupp, *Combinatorial group theory*. Ergeb. Math. Grenzgeb. (3) 89, Springer, Berlin-New York, 1977 Zbl 0368.20023 MR 0577064

[14] Y. Ollivier, Critical densities for random quotients of hyperbolic groups. *C. R. Math. Acad. Sci. Paris* **336** (2003), no. 5, 391–394 Zbl 1050.20048 MR 1979351

[15] Y. Ollivier, Sharp phase transition theorems for hyperbolicity of random groups. *Geom. Funct. Anal.* **14** (2004), no. 3, 595–679 Zbl 1064.20045 MR 2100673

[16] Y. Ollivier, *A January 2005 invitation to random groups*. Ensaios Mat. 10, Soc. Brasil. Mat., Rio de Janeiro, 2005 Zbl 1163.20311 MR 2205306

[17] Y. Ollivier, Some small cancellation properties of random groups. *Internat. J. Algebra Comput.* **17** (2007), no. 1, 37–51 Zbl 1173.20025 MR 2300404

[18] Y. Ollivier and D. T. Wise, Cubulating random groups at density less than $1/6$. *Trans. Amer. Math. Soc.* **363** (2011), no. 9, 4701–4733 Zbl 1277.20048 MR 2806688

[19] A. Żuk, Property (T) and Kazhdan constants for discrete groups. *Geom. Funct. Anal.* **13** (2003), no. 3, 643–670 Zbl 1036.22004 MR 1995802

Received 22 April 2021; revised 13 December 2021.

Tsung-Hsuan Tsai
Institut de Recherche Mathématique Avancée, UFR de Mathématiques (Office 105),
7 rue René Descartes, 67000 Strasbourg, France; tsung-hsuan.tsai@math.unistra.fr