



RESEARCH ARTICLE

Forbidden sparse intersections

Miltiadis Karamanlis¹ and Pandelis Dodos²

¹Department of Mathematics, National and Kapodistrian University of Athens, Panepistimiopolis, Athens, 157 84, Greece;
E-mail: kararemilt@gmail.com.

²Department of Mathematics, National and Kapodistrian University of Athens, Panepistimiopolis, Athens, 157 84, Greece;
E-mail: pdodos@math.uoa.gr (corresponding author).

Received: 22 June 2023; Revised: 19 March 2025; Accepted: 8 May 2025

2020 Mathematics Subject Classification: Primary – 05D05; Secondary – 05D40

Abstract

Let n be a positive integer, let $0 < p \leq p' \leq \frac{1}{2}$, and let $\ell \leq pn$ be a nonnegative integer. We prove that if $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ are two families whose cross intersections forbid ℓ —that is, they satisfy $|A \cap B| \neq \ell$ for every $A \in \mathcal{F}$ and every $B \in \mathcal{G}$ —then, setting $t := \min\{\ell, pn - \ell\}$, we have the subgaussian bound

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq 2 \exp\left(-\frac{t^2}{58^2 pn}\right),$$

where μ_p and $\mu_{p'}$ denote the p -biased and p' -biased measures on $\{0, 1\}^n$, respectively.

Contents

1	Introduction	1
2	Background material	3
3	Forbidding initial or final intervals	5
4	The widening lemma	7
5	The algorithm	11
6	Proof of Theorem 1.1	13
7	Extensions of the main estimate	19
8	Optimality	21
9	Supersaturation	23
	References	28

1. Introduction

Extremal set theory can be traced back to the seminal work¹ of Erdős, Ko and Rado [EKR61] who obtained sharp estimates of the cardinality of a family $\mathcal{A} \subseteq \binom{[n]}{k}$ that is *intersecting*; that is, it satisfies $A \cap B \neq \emptyset$ for every $A, B \in \mathcal{A}$. (Here and in the rest of this paper, $\binom{[n]}{k}$ denotes the set of all k -element subsets of the discrete interval $[n] := \{1, \dots, n\}$.) Since then, it is an active subfield of combinatorics; we refer the reader to [EI22, FT18] for recent expositions of this theory and its applications.

¹Although published in 1961, the Erdős–Ko–Rado theorem was actually discovered much earlier, in 1938.

1.1. The Erdős–Sós problem and related results

A more challenging problem was posed in 1971 by Erdős and Sós (see [Erd75]); it asks to determine, for a given triple $\ell \leq k \leq n$ of positive integers, the cardinality of the largest family $\mathcal{A} \subseteq \binom{[n]}{k}$ whose intersections forbid ℓ ; that is, it satisfies $|A \cap B| \neq \ell$ for every $A, B \in \mathcal{A}$. Early significant progress on the Erdős–Sós problem was made by Frankl–Wilson [FW81] and Frankl–Füredi [FF85]. Somewhat later, in 1987, a breakthrough was achieved by Frankl and Rödl, who obtained exponential estimates in the regime where ℓ is proportional to n . More precisely, it is shown in [FR87, Theorem 1.5] that for every $0 < \varepsilon < p \leq \frac{1}{2}$, there exists a constant $\gamma(\varepsilon, p) > 0$ such that for every pair of positive integers ℓ, n with $\varepsilon n \leq \ell \leq pn - \varepsilon n$, and every pair of families² $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ whose cross intersections forbid ℓ – that is, they satisfy $|A \cap B| \neq \ell$ for every $A \in \mathcal{F}$ and every $B \in \mathcal{G}$ – we have

$$\mu_p(\mathcal{F}) \mu_p(\mathcal{G}) \leq (1 - \gamma(\varepsilon, p))^n, \tag{1.1}$$

where μ_p denotes the p -biased measure on $\{0, 1\}^n$ – namely, the probability measure on $\{0, 1\}^n$ defined by setting for every $A \subseteq [n]$,

$$\mu_p(\{A\}) := p^{|A|} (1 - p)^{n - |A|}. \tag{1.2}$$

(For the case $\mathcal{F} = \mathcal{G}$, an alternative proof was given by Keevash and Long [KLo16]; see also [KSZ22].) The work of Frankl and Rödl has proven to be very influential, and it has found applications in a number of different areas such as discrete geometry [FR90], communication complexity [S99] and quantum computing [BCW99].

Further progress on the Erdős–Sós problem was made by several authors, including the very recent works of Ellis–Keller–Lifshitz [EKL24], Keller–Lifshitz [KLi21] and Kupavskii–Zaharov [KZ24] (see, also, [KLLM23, KLo20] for closely related developments). Collectively, the papers [EKL24, KLi21] obtain the sharp estimate $|\mathcal{A}| \leq \binom{n - (\ell + 1)}{k - (\ell + 1)}$ for every family $\mathcal{A} \subseteq \binom{[n]}{k}$ whose intersections forbid ℓ in the regime $2\ell < k \leq (\frac{1}{2} - \varepsilon)n$ with $n \geq n_0(\ell, \varepsilon)$ for some (unspecified) threshold function $n_0(\ell, \varepsilon)$. The more recent paper [KZ24] extends this sharp estimate in the regime $\ell = \lceil k^\beta \rceil$ and $n = \lceil k^\alpha \rceil$, where $\alpha, \beta > 0$ are positive reals with $\beta < \frac{1}{2}$ and $\alpha > 1 + 2\beta$ and k is sufficiently large in terms of α, β .

1.2. The main estimate

Our main result provides new estimates of the product of the biased measures of a pair of families with forbidden cross intersections, and in particular, it bridges the gap between the aforementioned results.

Theorem 1.1. *Let n be a positive integer, let $0 < p \leq p' \leq \frac{1}{2}$, and let $\ell \leq pn$ be a nonnegative integer. Also let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ be two families whose cross intersections forbid ℓ ; that is, $|A \cap B| \neq \ell$ for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$. Then, setting $t := \min\{\ell, pn - \ell\}$, we have the subgaussian bound*

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq 2 \exp\left(-\frac{t^2}{58^2 pn}\right). \tag{1.3}$$

Note that Theorem 1.1 extends the Frankl–Rödl theorem [FR87, Theorem 1.5]; indeed, the bound (1.3) is nontrivial if $80\sqrt{pn} \leq \ell \leq pn - 80\sqrt{pn}$ uniformly for $p \geq \frac{160^2}{n}$.

Remark 1.2. Theorem 1.1 can be extended to cover the case of all parameters p, p' in the regime $0 < p \leq p' \leq 1 - p$, and it also has a version for pairs of families \mathcal{F}, \mathcal{G} that are contained in two, possibly different, layers of the cube. We present these (standard) extensions in Section 7.

Remark 1.3. The subgaussian bound (1.3) is actually optimal, modulo universal constants, for various choices of p, p' and ℓ . We discuss these issues in Section 8.

²We identify every $A \subseteq [n]$ with its indicator function $\mathbf{1}_A \in \{0, 1\}^n$.

Remark 1.4. Theorem 1.1 also has a supersaturation version, which is in the spirit of [FR87, Theorem 1.14]; see Section 9 for details.

1.3. Outline of the proof of Theorem 1.1

The proof of Theorem 1.1 follows the general strategy invented by Frankl and Rödl [FR87]. The idea is to gradually ‘deform’ the given families and eventually arrive at a pair of families whose cross intersections forbid an initial or a final interval; the measures of these final families can then be estimated by standard probabilistic tools. This ‘deformation’ is entirely algorithmic, and it is the heart of the proof.

The algorithm in [FR87] takes as an input two families $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ whose cross intersections forbid an interval of $[n]$, and it starts by using a density increment argument in order to show that the two sections

$$\mathcal{F}_0 := \{A : n \notin A \in \mathcal{F}\} \quad \text{and} \quad \mathcal{F}_1 := \{A \setminus \{n\} : n \in A \in \mathcal{F}\}$$

of \mathcal{F} have roughly the same measure. Once this is done, the algorithm proceeds by comparing the measures of the union $\mathcal{G}_0 \cup \mathcal{G}_1$ and the intersection $\mathcal{G}_0 \cap \mathcal{G}_1$ of the sections of \mathcal{G} . Again, a density increment argument is used to ensure that the measures of $\mathcal{G}_0 \cup \mathcal{G}_1$ and $\mathcal{G}_0 \cap \mathcal{G}_1$ are roughly equal, which in turn implies that the two sections, \mathcal{G}_0 and \mathcal{G}_1 , are almost equal. One can then use this structural information to produce a pair of families whose cross intersections forbid a larger interval, while at the same time, one keeps control of the product of their biased measures.

While the algorithm of Frankl and Rödl is elegant and efficient, unfortunately it leads to suboptimal results as p gets smaller, and it hits a barrier³ at $p = o(1)$. The reason is rather simple: if p is small, then having (or not having) density increment for the section \mathcal{A}_1 of a family $\mathcal{A} \subseteq \{0, 1\}^n$ has negligible effect on the measure of the other section \mathcal{A}_0 .

We resolve this issue by introducing a new algorithm that also takes as an input two families $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ whose cross intersections forbid an interval of $[n]$, and it starts by seeking for a density increment for one of the pairs $(\mathcal{F}_1, \mathcal{G}_1)$, $(\mathcal{F}_0, \mathcal{G}_0 \cup \mathcal{G}_1)$ and $(\mathcal{F}_0 \cup \mathcal{F}_1, \mathcal{G}_0)$. However, the density increment the algorithm is searching for is not *uniform* and depends on the specific pair it is looking at (as well as the parameter p). The particular choice of the density increments is justified analytically: if the algorithm does not succeed in this search, then this yields a strong *lower* bound for the product of the measures of \mathcal{F}_1 and $\mathcal{G}_0 \cap \mathcal{G}_1$, or the product of the measures of $\mathcal{F}_0 \cap \mathcal{F}_1$ and \mathcal{G}_1 . This is the content of Lemma 4.1 (the ‘widening lemma’) in Section 4. With this information at hand, we may proceed as in the algorithm of Frankl and Rödl. The main novelty (and technical difficulty) of the proof of Theorem 1.1 is thus to show that this rough outline is actually feasible by appropriately selecting the various parameters.

2. Background material

2.1. General notation

For every pair a, b of integers with $0 \leq a \leq b$ by $[a, b]$, we denote the discrete interval $\{k \in \mathbb{Z} : a \leq k \leq b\}$. Also recall that for every positive integer n and every nonnegative integer $k \leq n$, we set $[n] := \{1, \dots, n\}$ and $\binom{[n]}{k} := \{A \subseteq [n] : |A| = k\}$; moreover, for every $t \geq 0$, we set $[n]^{\leq t} := \{A \subseteq [n] : |A| \leq t\}$, $[n]^{< t} := \{A \subseteq [n] : |A| < t\}$, $[n]^{\geq t} := \{A \subseteq [n] : |A| \geq t\}$ and $[n]^{> t} := \{A \subseteq [n] : |A| > t\}$.

2.2. Families of sets

Let n be a positive integer, and let $\mathcal{A} \subseteq \{0, 1\}^n$. We say that \mathcal{A} is *downwards closed* if for every $A \in \mathcal{A}$ and every $B \subseteq A$, we have that $B \in \mathcal{A}$; respectively, we say that \mathcal{A} is *upwards closed* if for every $A \in \mathcal{A}$

³It is also not clear if the algorithm works if $\min\{\ell, pn - \ell\} = o(pn)$ and $p = \Theta(1)$, but the obstacles in this regime seem somewhat less serious.

and every $B \supseteq A$, we have that $B \in \mathcal{A}$. If, in addition, $n \geq 2$, then we set

$$\mathcal{A}_0 := \{A : n \notin A \in \mathcal{A}\} \quad \text{and} \quad \mathcal{A}_1 := \{A \setminus \{n\} : n \in A \in \mathcal{A}\}, \tag{2.1}$$

and we view both \mathcal{A}_0 and \mathcal{A}_1 as subfamilies of $\{0, 1\}^{n-1}$.

It is also convenient to introduce the following definition.

Definition 2.1 (Forbidden intersections). Let n be a positive integer, let $L \subseteq [n]$, and let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$. We write $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, L)$ to denote the fact that the cross intersections of \mathcal{F} and \mathcal{G} forbid L ; that is, $|A \cap B| \notin L$ for every $A \in \mathcal{F}$ and $B \in \mathcal{G}$.

2.3. *p*-biased measures

Recall that for every positive integer n and every $0 < p < 1$ by μ_p , we denote the p -biased probability measure on $\{0, 1\}^n$ defined in (1.2). We record, for future use, the following elementary property of these measures.

Fact 2.2. Let n be a positive integer, let $0 < p \leq p' < 1$, and let $\mathcal{A} \subseteq \{0, 1\}^n$. If \mathcal{A} is upwards closed, then $\mu_p(\mathcal{A}) \leq \mu_{p'}(\mathcal{A})$; respectively, if \mathcal{A} is downwards closed, then $\mu_p(\mathcal{A}) \geq \mu_{p'}(\mathcal{A})$.

2.4. Chernoff bounds

We will need the following standard estimates of the biased measure of the tails of the binomial distribution (see, for example, [AS16, Appendix A]).

Lemma 2.3. Let n be a positive integer, let t be a nonnegative real, and let $0 < p < 1$.

(i) If $p \leq \frac{1}{2}$ and $pn \leq t \leq 2pn$, then

$$\mu_p([n]^{\geq t}) \leq \exp\left(-\frac{(t - pn)^2}{6p(1 - p)n}\right). \tag{2.2}$$

(ii) If $p \geq \frac{1}{2}$ and $t \geq pn$, then

$$\mu_p([n]^{\geq t}) \leq \exp\left(-\frac{(t - pn)^2}{2p(1 - p)n}\right). \tag{2.3}$$

2.5. Estimates of binomial coefficients

We will also need the following basic estimates of binomial coefficients that follow from a non-asymptotic version of Stirling’s approximation—see, for example, [Ro55]—and elementary computations.

Fact 2.4. Let $n \geq 2$ be an integer, and let $k \in [n - 1]$. Then we have

$$\frac{24}{25\sqrt{2\pi}} \cdot \sqrt{\frac{n}{k(n - k)}} \cdot \frac{n^n}{k^k(n - k)^{n - k}} < \binom{n}{k} < \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{n}{k(n - k)}} \cdot \frac{n^n}{k^k(n - k)^{n - k}}; \tag{2.4}$$

in particular, for every $\mathcal{F} \subseteq \binom{[n]}{k}$ we have

$$\sqrt{2\pi} \cdot \sqrt{\frac{k(n - k)}{n}} \cdot \mu_{\frac{k}{n}}(\mathcal{F}) < \frac{|\mathcal{F}|}{\binom{n}{k}} < \frac{25\sqrt{2\pi}}{24} \cdot \sqrt{\frac{k(n - k)}{n}} \cdot \mu_{\frac{k}{n}}(\mathcal{F}). \tag{2.5}$$

Moreover, if $H: [0, 1] \rightarrow \mathbb{R}$ denotes the binary entropy function⁴, then

$$\frac{24}{25\sqrt{2\pi}} \cdot \sqrt{\frac{n}{k(n-k)}} \cdot 2^{nH(\frac{k}{n})} \leq \binom{n}{k} \leq \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{n}{k(n-k)}} \cdot 2^{nH(\frac{k}{n})}. \tag{2.6}$$

3. Forbidding initial or final intervals

Our goal in this section is to obtain estimates for the product of the biased measures of a pair of families $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ whose cross intersections forbid an initial or a final subinterval of $[n]$. This information is needed for the proof of Theorem 1.1.

We note that closely related problems have been studied extensively in extremal combinatorics; see [EI22, FT18] and the references therein. We shall obtain the desired estimates, however, from the following well-known concentration inequality for the biased measures. As usual, for a family $\mathcal{A} \subseteq \{0, 1\}^n$ and a nonnegative real $t \leq n$, we set $\mathcal{A}_t := \{H \subseteq [n] : \exists A \in \mathcal{A} \text{ such that } |H \triangle A| \leq t\}$.

Proposition 3.1. *Let n be a positive integer, let $0 < p < 1$, let $t \leq pn$ be a nonnegative real, and let $\mathcal{A} \subseteq \{0, 1\}^n$ such that $\mu_p(\mathcal{A}) \geq \frac{1}{2}$.*

(i) *If $0 < p \leq \frac{1}{2}$, then we have*

$$\mu_p(\mathcal{A}_t) \geq 1 - \exp\left(-\frac{t^2}{6p(1-p)n}\right). \tag{3.1}$$

(ii) *If $\frac{1}{2} < p < 1$, then we have*

$$\mu_p(\mathcal{A}_t) \geq 1 - \exp\left(-\frac{t^2}{2p(1-p)n}\right). \tag{3.2}$$

Proposition 3.1 follows from the proof of [BHT06, Proposition 2.4] in the work of Bobkov, Choudré and Tetali, which in turn is based on results of Bollobás–Leader [BL91], Jogdeo–Samuels [JS68] and Talagrand [Ta89]. Since Proposition 3.1 is not explicitly isolated in [BHT06], for the convenience of the reader, we briefly recall the argument.

Proof of Proposition 3.1. As it is mentioned in [BHT06], it is enough to prove the result under the additional assumption that \mathcal{A} is downwards closed. Indeed, Step 1 through Step 4 in the proof of [Ta89, Theorem 7] carry out this reduction. So, suppose that \mathcal{A} is downwards closed with $\mu_p(\mathcal{A}) \geq \frac{1}{2}$. By [JS68, Theorem 3.2 and Corollary 3.1], we have

$$\mu_p(\mathcal{A}) \geq \mu_p([n]^{\leq \lfloor pn \rfloor}); \tag{3.3}$$

that is, the median of the binomial distribution $\text{Bin}(n, p)$ is greater than or equal to $\lfloor pn \rfloor$. Moreover, since \mathcal{A} is downwards closed, by [BL91, Corollary 5], we have

$$\mu_p(\mathcal{A}_t) \geq \mu_p([n]^{\leq \lfloor pn \rfloor + t}), \tag{3.4}$$

which in turn implies that

$$1 - \mu_p(\mathcal{A}_t) \leq \mu_p([n]^{\gt \lfloor pn \rfloor + t}) \leq \mu_p([n]^{\geq pn + t}). \tag{3.5}$$

Therefore, if $0 < p \leq \frac{1}{2}$, then (3.1) follows from (3.5) and (2.2), while if $\frac{1}{2} < p < 1$, then (3.2) follows from (3.5) and (2.3). \square

⁴Recall that $H(0) = H(1) = 0$, and $H(x) = -x \log_2(x) - (1-x) \log_2(1-x)$ if $0 < x < 1$.

Proposition 3.1 will be used in the following form (the proof is straightforward, and it is left to the reader).

Corollary 3.2. *Let n be a positive integer, let $0 < p < 1$, let $t \leq pn$ be a nonnegative real, and let $\mathcal{A} \subseteq \{0, 1\}^n$.*

- (i) *If $0 < p \leq \frac{1}{2}$ and $\mu_p(\mathcal{A}) > \exp\left(-\frac{t^2}{6p(1-p)n}\right)$, then $\mu_p(\mathcal{A}_{2t}) > 1 - \exp\left(-\frac{t^2}{6p(1-p)n}\right)$.*
- (ii) *If $\frac{1}{2} < p < 1$ and $\mu_p(\mathcal{A}) > \exp\left(-\frac{t^2}{2p(1-p)n}\right)$, then $\mu_p(\mathcal{A}_{2t}) > 1 - \exp\left(-\frac{t^2}{2p(1-p)n}\right)$.*

We are now ready to state the main result in this section.

Lemma 3.3. *Let n be a positive integer, let $0 < p \leq p' \leq 1 - p$ with $p \leq \frac{1}{2}$, and let $\alpha \leq pn$ be a positive integer. Also let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$.*

- (i) *If $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, [0, \alpha])$, then*

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right). \tag{3.6}$$

- (ii) *If $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, [pn - \alpha, n])$, then*

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq 2 \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right). \tag{3.7}$$

Proof. We start with the proof of part (i). Clearly, we may assume that the pair $(\mathcal{F}, \mathcal{G})$ is optimal, in the sense that it maximizes the quantity in the left-hand side of (3.6); consequently, we may assume that \mathcal{F} and \mathcal{G} are both upwards closed. Next observe that if $\mu_p(\mathcal{F}) \leq \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$, then (3.6) is straightforward. Therefore, we may also assume that $\mu_p(\mathcal{F}) > \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$. By Corollary 3.2 applied for ‘ $t = \frac{\alpha}{2}$ ’, we obtain that $\mu_p(\mathcal{F}_\alpha) > 1 - \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$, where $\mathcal{F}_\alpha := \{H \subseteq [n] : \exists A \in \mathcal{F} \text{ with } |H \Delta A| \leq \alpha\}$. Set $\bar{\mathcal{G}} := \{[n] \setminus G : G \in \mathcal{G}\}$, and note that for every $F \in \mathcal{F}$ and every $G \in \mathcal{G}$, we have that $|F \Delta ([n] \setminus G)| > \alpha$. This yields that $\bar{\mathcal{G}} \cap \mathcal{F}_\alpha = \emptyset$, which in turn implies that $\mu_p(\bar{\mathcal{G}}) < \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$. Finally, since \mathcal{G} is upwards closed and $p' \leq 1 - p$, by Fact 2.2, we conclude that

$$\exp\left(-\frac{\alpha^2}{24p(1-p)n}\right) > \mu_p(\bar{\mathcal{G}}) = \mu_{1-p}(\mathcal{G}) \geq \mu_{p'}(\mathcal{G}) \geq \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}). \tag{3.8}$$

We proceed to the proof of part (ii). As before, we may assume that the pair $(\mathcal{F}, \mathcal{G})$ is optimal and, hence, that \mathcal{F} and \mathcal{G} are both downwards closed. Consequently, by Fact 2.2, it is enough to show that

$$\mu_p(\mathcal{F}) \mu_p(\mathcal{G}) \leq 2 \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right). \tag{3.9}$$

Setting

- $\mathcal{F}^{\leq pn - \frac{\alpha}{2}} := \{A \in \mathcal{F} : |A| \leq pn - \frac{\alpha}{2}\}$, $\mathcal{F}^{> pn - \frac{\alpha}{2}} := \{A \in \mathcal{F} : |A| > pn - \frac{\alpha}{2}\}$, and
- $\mathcal{G}^{\leq pn - \frac{\alpha}{2}} := \{B \in \mathcal{G} : |B| \leq pn - \frac{\alpha}{2}\}$, $\mathcal{G}^{> pn - \frac{\alpha}{2}} := \{B \in \mathcal{G} : |B| > pn - \frac{\alpha}{2}\}$,

by part (ii) of Lemma 2.3 applied for μ_{1-p} , we see that

$$\max \left\{ \mu_p(\mathcal{F}^{\leq pn - \frac{\alpha}{2}}), \mu_p(\mathcal{G}^{\leq pn - \frac{\alpha}{2}}) \right\} \leq \exp\left(-\frac{\alpha^2}{8p(1-p)n}\right); \tag{3.10}$$

thus, if $\mu_p(\mathcal{F}^{>p n - \frac{\alpha}{2}}) < \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$, then the result follows from (3.10). So, suppose that $\mu_p(\mathcal{F}^{>p n - \frac{\alpha}{2}}) \geq \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$. By Corollary 3.2 again applied for ‘ $t = \frac{\alpha}{2}$ ’, we obtain that

$$\mu_p(\mathcal{F}_\alpha^{>p n - \frac{\alpha}{2}}) \geq 1 - \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right), \tag{3.11}$$

where $\mathcal{F}_\alpha^{>p n - \frac{\alpha}{2}} := \{H \subseteq [n] : \exists A \in \mathcal{F}^{>p n - \frac{\alpha}{2}} \text{ with } |H \Delta A| \leq \alpha\}$. Now note that, since $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, [pn - \alpha, n])$, for every $F \in \mathcal{F}^{>p n - \frac{\alpha}{2}}$ and every $G \in \mathcal{G}^{>p n - \frac{\alpha}{2}}$, we have

$$|F \Delta G| = |F| + |G| - 2|F \cap G| > \alpha. \tag{3.12}$$

This observation yields that $\mathcal{G}^{>p n - \frac{\alpha}{2}} \cap \mathcal{F}_\alpha^{>p n - \frac{\alpha}{2}} = \emptyset$, and therefore, by (3.11), we obtain that $\mu_p(\mathcal{G}^{>p n - \frac{\alpha}{2}}) < \exp\left(-\frac{\alpha^2}{24p(1-p)n}\right)$. Inequality (3.7) follows from this estimate and (3.10). The proof of Lemma 3.3 is thus completed. \square

4. The widening lemma

The section is devoted to the proof of the following lemma. (Recall that for every family $\mathcal{A} \subseteq \{0, 1\}^n$ ($n \geq 2$) by \mathcal{A}_0 and \mathcal{A}_1 , we denote the sections of \mathcal{A} defined in (2.1); we also recall that we view \mathcal{A}_0 and \mathcal{A}_1 as families in $\{0, 1\}^{n-1}$.)

Lemma 4.1 (Widening lemma). *Let $n \geq 2$ be an integer, let $0 < p \leq p' \leq \frac{1}{2}$, and let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ be nonempty. Also let $0 < \delta < \frac{1}{10}$, and assume that*

$$\mu_p(\mathcal{F}_1) \mu_{p'}(\mathcal{G}_1) \leq (1 + \delta) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \tag{4.1}$$

and

$$\max\{\mu_p(\mathcal{F}_0) \mu_{p'}(\mathcal{G}_0 \cup \mathcal{G}_1), \mu_p(\mathcal{F}_0 \cup \mathcal{F}_1) \mu_{p'}(\mathcal{G}_0)\} \leq \left(1 + \frac{p}{1-p} \delta\right) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}). \tag{4.2}$$

Then we have

$$\max\{\mu_p(\mathcal{F}_1) \mu_{p'}(\mathcal{G}_0 \cap \mathcal{G}_1), \mu_p(\mathcal{F}_0 \cap \mathcal{F}_1) \mu_{p'}(\mathcal{G}_1)\} > \left(1 - \delta - 2\frac{p}{1-p} \delta^2\right) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}). \tag{4.3}$$

As we have noted in Subsection 1.3, Lemma 4.1 is a crucial ingredient of the proof of Theorem 1.1. That said, we advise the reader to skip its proof at first reading and return to this section once the basic steps of the proof of Theorem 1.1 have been properly understood.

We shall deduce Lemma 4.1 from the following, purely analytical, result.

Sublemma 4.2. *Let $x, x', y, y', z, z', w, w', p, p', \delta \in [-1, 1]$ be real numbers with $y, w \geq 0, y', w' \leq 0, \max\{x, x'\} \leq w, \max\{z, z'\} \leq y, 0 < p \leq p' \leq \frac{1}{2}$ and $0 < \delta < \frac{1}{10}$. Assume that the following identities*

$$x' = -\frac{p}{1-p}x, \quad w + w' = x + x' = \frac{1-2p}{1-p}x, \tag{4.4}$$

$$z' = -\frac{p'}{1-p'}z, \quad y + y' = z + z' = \frac{1-2p'}{1-p'}z, \tag{4.5}$$

as well as the following inequalities

$$(1 + x)(1 + z) \leq 1 + \delta, \tag{4.6}$$

$$(1 + x')(1 + y) \leq 1 + \frac{p}{1 - p} \delta, \tag{4.7}$$

$$(1 + z')(1 + w) \leq 1 + \frac{p}{1 - p} \delta, \tag{4.8}$$

are satisfied. Then at least one of the following inequalities

$$(1 + x)(1 + y') > 1 - \delta - 2 \frac{p}{1 - p} \delta^2, \tag{4.9}$$

$$(1 + z)(1 + w') > 1 - \delta - 2 \frac{p}{1 - p} \delta^2, \tag{4.10}$$

must also be satisfied.

We postpone the proof of Sublemma 4.2 to the end of this section. At this point, let us give the proof of Lemma 4.1.

Proof of Lemma 4.1. Notice, first, that

$$\mu_p(\mathcal{F}) = p\mu_p(\mathcal{F}_1) + (1 - p)\mu_p(\mathcal{F}_0) \quad \text{and} \quad \mu_{p'}(\mathcal{G}) = p'\mu_{p'}(\mathcal{G}_1) + (1 - p')\mu_{p'}(\mathcal{G}_0). \tag{4.11}$$

Next, define the real numbers $x, x', y, y', z, z', w, w'$ by setting

$$\begin{aligned} \frac{\mu_p(\mathcal{F}_1)}{\mu_p(\mathcal{F})} &= 1 + x, & \frac{\mu_{p'}(\mathcal{G}_1)}{\mu_{p'}(\mathcal{G})} &= 1 + z, \\ \frac{\mu_p(\mathcal{F}_0)}{\mu_p(\mathcal{F})} &= 1 + x', & \frac{\mu_{p'}(\mathcal{G}_0)}{\mu_{p'}(\mathcal{G})} &= 1 + z', \\ \frac{\mu_p(\mathcal{F}_0 \cup \mathcal{F}_1)}{\mu_p(\mathcal{F})} &= 1 + w, & \frac{\mu_{p'}(\mathcal{G}_0 \cup \mathcal{G}_1)}{\mu_{p'}(\mathcal{G})} &= 1 + y, \\ \frac{\mu_p(\mathcal{F}_0 \cap \mathcal{F}_1)}{\mu_p(\mathcal{F})} &= 1 + w', & \frac{\mu_{p'}(\mathcal{G}_0 \cap \mathcal{G}_1)}{\mu_{p'}(\mathcal{G})} &= 1 + y'. \end{aligned}$$

With these choices, the result follows from Sublemma 4.2 after taking into account the identities in (4.11). □

4.1. Proof of Sublemma 4.2

First observe that, by (4.4) and (4.5), we have

$$xy' + x'y = \frac{1 - 2p'}{1 - p'}xz - \frac{xy}{1 - p} \quad \text{and} \quad zw' + z'w = \frac{1 - 2p}{1 - p}xz - \frac{zw}{1 - p'} \tag{4.12}$$

that yields that

$$(1 + x)(1 + y') + (1 + x')(1 + y) = 2 + \frac{1 - 2p}{1 - p}x + \frac{1 - 2p'}{1 - p'}z(1 + x) - \frac{xy}{1 - p}, \tag{4.13}$$

$$(1 + z)(1 + w') + (1 + z')(1 + w) = 2 + \frac{1 - 2p}{1 - p}x(1 + z) + \frac{1 - 2p'}{1 - p'}z - \frac{zw}{1 - p'}. \tag{4.14}$$

Combining these equalities with (4.7) and (4.8), we obtain that

$$(1+x)(1+y') \geq 1 - \frac{p}{1-p}\delta + \frac{1-2p}{1-p}x + \frac{1-2p'}{1-p'}z(1+x) - \frac{xy}{1-p} \tag{4.15}$$

and

$$(1+z)(1+w') \geq 1 - \frac{p}{1-p}\delta + \frac{1-2p}{1-p}x(1+z) + \frac{1-2p'}{1-p'}z - \frac{zw}{1-p'}. \tag{4.16}$$

Next, observe that since y and w are nonnegative, by (4.7) and (4.8), we have

$$x' \leq \frac{p}{1-p}\delta \quad \text{and} \quad z' \leq \frac{p}{1-p}\delta \tag{4.17}$$

that yields that

$$x \geq -\delta \quad \text{and} \quad z \geq -\delta. \tag{4.18}$$

By (4.6), we have that either

- (A1) $x < \frac{\delta}{2}$, or
- (A2) $z < \frac{\delta}{2}$.

Claim 4.3. *If $x < \frac{\delta}{2}$, then*

$$y \leq 2\frac{p}{1-p}\delta. \tag{4.19}$$

However, if $z < \frac{\delta}{2}$, then

$$w \leq \frac{\frac{p}{1-p} + \frac{p'}{2(1-p')}}{1 - \frac{p'}{2(1-p')}}\delta. \tag{4.20}$$

Proof of Claim 4.3. First assume that $x < \frac{\delta}{2}$. Then, by (4.7), we have

$$1+y \leq \frac{1 + \frac{p}{1-p}\delta}{1+x'} = 1 + \frac{\frac{p}{1-p}\delta - x'}{1+x'} = 1 + \frac{\frac{p}{1-p}\delta + \frac{p}{1-p}x}{1 - \frac{p}{1-p}x}. \tag{4.21}$$

Since $-\frac{1}{10} < -\delta \leq x < \frac{\delta}{2} < \frac{1}{20}$ and the function $(-1, 1) \ni x \mapsto \frac{\frac{p}{1-p}\delta + \frac{p}{1-p}x}{1 - \frac{p}{1-p}x}$ is increasing, by (4.21), we see that (4.19) is satisfied.

Next assume that $z < \frac{\delta}{2}$. By (4.8), we have

$$1+w \leq \frac{1 + \frac{p}{1-p}\delta}{1+z'} = 1 + \frac{\frac{p}{1-p}\delta - z'}{1+z'} = 1 + \frac{\frac{p}{1-p}\delta + \frac{p'}{1-p'}z}{1 - \frac{p'}{1-p'}z}. \tag{4.22}$$

Thus, (4.20) follows from (4.22) using the fact that $-\frac{1}{10} < -\delta \leq z < \frac{\delta}{2} < \frac{1}{20}$ and the fact that the function $(-1, 1) \ni z \mapsto \frac{\frac{p}{1-p}\delta + \frac{p'}{1-p'}z}{1 - \frac{p'}{1-p'}z}$ is increasing. □

We proceed by considering the following cases.

Case 1: $x \geq 0$ and $z \geq 0$.

First assume that $x < \frac{\delta}{2}$. Then, since $p \leq \frac{1}{2}$, by (4.19), we see that $\frac{xy}{1-p} < 2\frac{p}{1-p}\delta^2$. Moreover, since $x, z \geq 0$, we have $\frac{1-2p}{1-p}x + \frac{1-2p'}{1-p'}z(1+x) \geq 0$ and so, by (4.15), we obtain that

$$(1+x)(1+y') > 1 - \frac{p}{1-p}\delta - 2\frac{p}{1-p}\delta^2 \geq 1 - \delta - 2\frac{p}{1-p}\delta^2; \tag{4.23}$$

that is, (4.9) is satisfied.

However, if $z < \frac{\delta}{2}$, then we consider the subcases ‘ $p \leq \frac{3}{7}$ ’ and ‘ $\frac{3}{7} < p$ ’. First observe that, by (4.20) and the fact that $\delta < \frac{1}{10}$, we have that $w \leq 2\delta$. If $0 < p \leq \frac{3}{7}$ then, since $w \geq 0$, this yields that $\frac{zw}{1-p'} < 2\delta^2 < \frac{1-2p}{1-p}\delta$. Using again our starting assumption that $x, z \geq 0$, we see that $\frac{1-2p}{1-p}x(1+z) + \frac{1-2p'}{1-p'}z \geq 0$. By (4.16) and the previous observations, we obtain that

$$(1+z)(1+w') \geq 1 - \frac{p}{1-p}\delta - \frac{1-2p}{1-p}\delta = 1 - \delta > 1 - \delta - 2\frac{p}{1-p}\delta^2; \tag{4.24}$$

in other words, in this subcase, (4.10) is satisfied. Finally, assume that $\frac{3}{7} < p \leq \frac{1}{2}$. Then observe that $\frac{p'}{2(1-p')} < \frac{2}{3}\frac{p}{1-p}$, which in turn implies, by (4.20) and the fact that $\delta < \frac{1}{10}$, that $w < 2\frac{p}{1-p}\delta$. Therefore, $zw < \frac{p}{1-p}\delta^2 < 2\frac{p}{1-p}\delta^2$. By the previous discussion, (4.16) and using once again the estimate $\frac{1-2p}{1-p}x(1+z) + \frac{1-2p'}{1-p'}z \geq 0$, we conclude that

$$(1+z)(1+w') > 1 - \delta - 2\frac{p}{1-p}\delta^2, \tag{4.25}$$

and so, in this subcase, (4.10) is satisfied.

Case 2: $x < 0$ and $z \geq 0$.

Then we have $-\frac{xy}{1-p} \geq 0$ and $\frac{1-2p}{1-p}x + \frac{1-2p'}{1-p'}z(1+x) \geq \frac{1-2p}{1-p}x$, and so, by (4.15) and (4.18), we obtain that

$$(1+x)(1+y') \geq 1 - \frac{p}{1-p}\delta - \frac{1-2p}{1-p}\delta = 1 - \delta > 1 - \delta - 2\frac{p}{1-p}\delta^2. \tag{4.26}$$

Thus, in this case, (4.9) is satisfied.

Case 3: $x \geq 0$ and $z < 0$.

It is similar to Case 2. Indeed, observe that $-\frac{zw}{1-p} \geq 0$ and $\frac{1-2p}{1-p}x(1+z) + \frac{1-2p'}{1-p'}z \geq \frac{1-2p'}{1-p'}z$. Hence, by (4.16), (4.18) and the fact that $\frac{1-2p}{1-p} \geq \frac{1-2p'}{1-p'}$, we obtain that

$$(1+z)(1+w') \geq 1 - \frac{p}{1-p}\delta - \frac{1-2p'}{1-p'}\delta > 1 - \delta - 2\frac{p}{1-p}\delta^2; \tag{4.27}$$

thus, in this case, (4.10) is satisfied.

Case 4: $x < 0$ and $z < 0$.

First observe that $-\frac{xy}{1-p} > 0$. Moreover, since $z' \leq y$, by (4.7),

$$1+x'+z'+x'z' = (1+x')(1+z') \leq (1+x')(1+y) \leq 1 + \frac{p}{1-p}\delta \tag{4.28}$$

that implies that $x' + z' + x'z' \leq \frac{p}{1-p}\delta$. Noticing that $\frac{1-2p'}{p'} \leq \frac{1-2p}{p}$, by (4.4) and (4.5),

$$\begin{aligned} \frac{1-2p}{1-p}x + \frac{1-2p'}{1-p'}z(1+x) &= -\frac{1-2p}{p}x' - \frac{1-2p'}{p'}z' + \frac{1-2p'}{p'}\frac{1-p}{p}x'z' \\ &\geq -\frac{1-2p}{p}x' - \frac{1-2p}{p}z' - \frac{1-2p}{p}x'z' \\ &\geq -\frac{1-2p}{p}(x' + z' + x'z') \geq -\frac{1-2p}{1-p}\delta. \end{aligned} \tag{4.29}$$

Hence, by (4.15) and (4.29), we conclude that

$$(1+x)(1+y') \geq 1 - \frac{p}{1-p}\delta - \frac{1-2p}{1-p}\delta = 1 - \delta > 1 - \delta - 2\frac{p}{1-p}\delta^2; \tag{4.30}$$

thus, in this case, (4.9) is satisfied.

The above cases are exhaustive, and so the proof of Sublemma 4.2 is completed.

5. The algorithm

In this section, we present the formal description and the basic properties of the algorithm that is used in the proof of Theorem 1.1; we shall also discuss in more loose terms its main features. (Again, we recall that for every family $\mathcal{A} \subseteq \{0, 1\}^n$ ($n \geq 2$) by $\mathcal{A}_0, \mathcal{A}_1 \subseteq \{0, 1\}^{n-1}$, we denote the sections of \mathcal{A} defined in (2.1).)

For the analysis of the algorithm, we will need the following elementary, but crucial, fact that originates in the work of Frankl and Rödl [FR87].

Fact 5.1. Let $n \geq 2$ be an integer, let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ and let $a, b \in [n]$ with $a \leq b$. Assume that $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, [a, b])$. Then we have

$$(\mathcal{F}_1, \mathcal{G}_1) \in \text{Forbid}(n-1, [a-1, b-1]), \tag{5.1}$$

$$(\mathcal{F}_0, \mathcal{G}_0 \cup \mathcal{G}_1) \in \text{Forbid}(n-1, [a, b]), \quad (\mathcal{F}_0 \cup \mathcal{F}_1, \mathcal{G}_0) \in \text{Forbid}(n-1, [a, b]), \tag{5.2}$$

$$(\mathcal{F}_1, \mathcal{G}_0 \cap \mathcal{G}_1) \in \text{Forbid}(n-1, [a-1, b]), \quad (\mathcal{F}_0 \cap \mathcal{F}_1, \mathcal{G}_1) \in \text{Forbid}(n-1, [a-1, b]). \tag{5.3}$$

The algorithm takes as an input

- (I1) an integer $n \geq 2$, two reals $0 < p \leq p' \leq \frac{1}{2}$ and a positive integer $\ell < pn$,
- (I2) a real $0 < \delta < \frac{1}{10}$, and
- (I3) two nonempty families $\mathcal{F}_{\text{init}}, \mathcal{G}_{\text{init}} \subseteq \{0, 1\}^n$ whose cross intersections forbid ℓ ,

and outputs

- (O1) three nonnegative integers $a^* \leq b^* \leq m^*$ with $m^* \geq 1$, and
- (O2) two families $\mathcal{F}^*, \mathcal{G}^* \subseteq \{0, 1\}^{m^*}$ such that

$$(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [0, b^*]) \quad \text{or} \quad (\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [a^*, m^*]). \tag{5.4}$$

It also uses six counters $S_{d_1}, S_{d_2}, S_w, a, b, m$ that serve different purposes. The first three counters, S_{d_1}, S_{d_2} and S_w , give us the total number of iterations, and they are used for bookkeeping the operations performed by the algorithm (we shall comment on these operations in due course). The counters a, b encode the interval that is forbidden for the families \mathcal{F} and \mathcal{G} . Finally, the counter m keeps track of the dimension of \mathcal{F} and \mathcal{G} ; in particular, m starts from n and drops by one at each iteration. Thus, we have

$$S_{d_1} + S_{d_2} + S_w = n - m^*. \tag{5.5}$$

Algorithm.

Input: $n, p, p', \ell, 0 < \delta < \frac{1}{10}, (\mathcal{F}_{\text{init}}, \mathcal{G}_{\text{init}}) \in \text{Forbid}(n, \{\ell\})$ with $\mathcal{F}_{\text{init}}, \mathcal{G}_{\text{init}} \neq \emptyset$

Output: $a^*, b^*, m^*, \mathcal{F}^*, \mathcal{G}^*$

Initialize: $S_{d_1} \leftarrow 0, S_{d_2} \leftarrow 0, S_w \leftarrow 0, a \leftarrow \ell, b \leftarrow \ell,$
 $m \leftarrow n, \mathcal{F} \leftarrow \mathcal{F}_{\text{init}}, \mathcal{G} \leftarrow \mathcal{G}_{\text{init}}.$

Step 1: If $a = 0$, then $a^* \leftarrow a, b^* \leftarrow b, m^* \leftarrow m, \mathcal{F}^* \leftarrow \mathcal{F}, \mathcal{G}^* \leftarrow \mathcal{G}$ and terminate;
 else go to **Step 2**.

Step 2: If $b = m$, then $a^* \leftarrow a, b^* \leftarrow b, m^* \leftarrow m, \mathcal{F}^* \leftarrow \mathcal{F}, \mathcal{G}^* \leftarrow \mathcal{G}$ and terminate;
 else go to **Step 3**.

Step 3: If $\mu_p(\mathcal{F}_1) \mu_{p'}(\mathcal{G}_1) > (1 + \delta) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G})$, then $S_{d_1} \leftarrow S_{d_1} + 1,$
 $a \leftarrow a - 1, b \leftarrow b - 1, m \leftarrow m - 1, \mathcal{F} \leftarrow \mathcal{F}_1, \mathcal{G} \leftarrow \mathcal{G}_1$ and go to **Step 1**;
 else go to **Step 4**.

Step 4: If $\mu_p(\mathcal{F}_0) \mu_{p'}(\mathcal{G}_0 \cup \mathcal{G}_1) > (1 + \frac{p}{1-p} \delta) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G})$, then $S_{d_2} \leftarrow S_{d_2} + 1,$
 $m \leftarrow m - 1, \mathcal{F} \leftarrow \mathcal{F}_0, \mathcal{G} \leftarrow \mathcal{G}_0 \cup \mathcal{G}_1$ and go to **Step 1**;
 else go to **Step 5**.

Step 5: If $\mu_p(\mathcal{F}_0 \cup \mathcal{F}_1) \mu_{p'}(\mathcal{G}_0) > (1 + \frac{p}{1-p} \delta) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G})$, then $S_{d_2} \leftarrow S_{d_2} + 1,$
 $m \leftarrow m - 1, \mathcal{F} \leftarrow \mathcal{F}_0 \cup \mathcal{F}_1, \mathcal{G} \leftarrow \mathcal{G}_0$ and go to **Step 1**;
 else go to **Step 6**.

Step 6: If $\mu_p(\mathcal{F}_1) \mu_{p'}(\mathcal{G}_0 \cap \mathcal{G}_1) > (1 - \delta - 2\frac{p}{1-p} \delta^2) \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G})$,
 then $S_w \leftarrow S_w + 1, a \leftarrow a - 1, m \leftarrow m - 1, \mathcal{F} \leftarrow \mathcal{F}_1, \mathcal{G} \leftarrow \mathcal{G}_0 \cap \mathcal{G}_1$
 and go to **Step 1**;
 else go to **Step 7**.

Step 7: $S_w \leftarrow S_w + 1, a \leftarrow a - 1, m \leftarrow m - 1, \mathcal{F} \leftarrow \mathcal{F}_0 \cap \mathcal{F}_1, \mathcal{G} \leftarrow \mathcal{G}_1$
 and go to **Step 1**.

Moreover, by Fact 5.1, at each iteration of ‘type’ S_{d_1} or S_w (that is, at each iteration where one of the counters S_{d_1} or S_w is increased by one), the lower bound a of the forbidden interval $[a, b]$ is reduced by one; since the algorithm starts with $a = \ell$, we obtain that

$$S_{d_1} + S_w \leq \ell. \tag{5.6}$$

Next, observe that S_{d_1} and S_{d_2} count the number of iterations where we have ‘density increment’. Note, however, that this increment is not uniform: at each iteration of ‘type’ S_{d_1} , the product of the measures is increased by a factor $(1 + \delta)$, while at each iteration of ‘type’ S_{d_2} , the product of the measures is increased by a factor $(1 + \frac{p}{1-p} \delta)$. However, if at a certain iteration the algorithm reaches **Step 6** and then moves to **Step 1**, then the product of the measures of the new families is comparable to the product of the measures of the previous families by a factor $(1 - \delta - 2\frac{p}{1-p} \delta^2)$; in other words, the product of the measures may possibly drop, but not significantly. Finally, if at a certain iteration the algorithm reaches **Step 7**, then the widening lemma (Lemma 4.1) ensures that the product of the measures of the new families is also at least $(1 - \delta - 2\frac{p}{1-p} \delta^2)$ times the product of the measures of the previous families.

Summing up the previous observations, we arrive at the following basic estimate

$$\mu_p(\mathcal{F}^*) \mu_{p'}(\mathcal{G}^*) > (1 + \delta)^{S_{d_1}} \left(1 + \frac{p}{1-p} \delta\right)^{S_{d_2}} \left(1 - \delta - 2\frac{p}{1-p} \delta^2\right)^{S_w} \mu_p(\mathcal{F}_{\text{init}}) \mu_{p'}(\mathcal{G}_{\text{init}}) \tag{5.7}$$

that will be used in the analysis of the algorithm in the next section.

6. Proof of Theorem 1.1

Let $n, p, p', \ell, \mathcal{F}, \mathcal{G}$ be as in the statement of the theorem. Clearly, we may assume that \mathcal{F} and \mathcal{G} are nonempty. Notice that (1.3) is straightforward if $\ell = 0$ or $\ell = pn$. Thus, we may also assume that ℓ is a positive integer with $\ell < pn$ and, consequently, $n \geq 2$; moreover, setting

$$\delta := \min \left\{ \frac{\ell}{58pn}, \frac{pn - \ell}{51pn} \right\}, \tag{6.1}$$

we have that $0 < \delta < \frac{1}{10}$. We will actually show the slightly stronger estimate

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq 2 \exp(-pn\delta^2). \tag{6.2}$$

Assume, towards a contradiction, that this is not the case; that is,

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) > 2 \exp(-pn\delta^2). \tag{6.3}$$

We run the algorithm described in Section 5 for n, p, p', ℓ, δ and the families \mathcal{F}, \mathcal{G} . Let $a^*, b^*, m^*, \mathcal{F}^*, \mathcal{G}^*$ denote the output of the algorithm. By (5.4), we see that either

- (A1) $(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [0, b^*])$, or
- (A2) $(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [a^*, m^*])$.

The contradiction will be derived by showing that none of these cases can occur.

To this end, we first observe that, by (5.7) and (6.3), we have

$$1 > (1 + \delta)^{S_{d_1}} \left(1 + \frac{p}{1-p} \delta \right)^{S_{d_2}} \left(1 - \delta - 2 \frac{p}{1-p} \delta^2 \right)^{S_w} \exp(-pn\delta^2), \tag{6.4}$$

where S_{d_1}, S_{d_2}, S_w are the counters used in the algorithm. We will need the following estimates for S_{d_1}, S_{d_2} and S_w .

Lemma 6.1. *We have*

$$S_{d_1} - S_w < 5pn\delta, \tag{6.5}$$

$$S_{d_2} - \frac{1-p}{p} S_w < 3n\delta. \tag{6.6}$$

In the proof of Lemma 6.1, as well as in the rest of this section, we will repeatedly use the following elementary observation, which we isolate for the convenience of the reader.

Observation 6.2. *The following hold.*

- (i) *We have $\frac{1}{1-x} = 1 + \frac{x}{1-x}$ for every $x \neq 1$.*
- (ii) *We have $x - \frac{x^2}{2} \leq \ln(1+x) \leq x$ for every $x \geq 0$.*

In particular, for every $0 \leq x < 1$, we have

$$\frac{x}{1-x} - \frac{x^2}{2(1-x)^2} \leq \ln \left(\frac{1}{1-x} \right) \leq \frac{x}{1-x}. \tag{6.7}$$

We are ready to proceed to the proof of Lemma 6.1.

Proof of Lemma 6.1. We start with the proof of (6.5). Notice first that, by (6.4),

$$1 > (1 + \delta)^{S_{d_1}} \left(1 - \delta - 2 \frac{p}{1-p} \delta^2 \right)^{S_w} \exp(-pn\delta^2) \tag{6.8}$$

or, equivalently,

$$1 > (1 + \delta)^{S_{d_1} - S_w} \left((1 + \delta) \left(1 - \delta - 2 \frac{p}{1-p} \delta^2 \right) \right)^{S_w} \exp(-pn\delta^2). \tag{6.9}$$

Since $(1 + \delta) \left(1 - \delta - 2 \frac{p}{1-p} \delta^2 \right) = 1 - \frac{1+p}{1-p} \delta^2 - 2 \frac{p}{1-p} \delta^3$, after taking logarithms and rearranging we find that

$$S_{d_1} - S_w < \frac{1}{\ln(1 + \delta)} \left(S_w \ln \left(\frac{1}{1 - \frac{1+p}{1-p} \delta^2 - 2 \frac{p}{1-p} \delta^3} \right) + pn\delta^2 \right) \tag{6.10}$$

that implies, by Observation 6.2, that

$$S_{d_1} - S_w < \frac{1}{\delta(1 - \frac{\delta}{2})} \left(S_w \frac{\frac{1+p}{1-p} \delta^2 + 2 \frac{p}{1-p} \delta^3}{1 - \frac{1+p}{1-p} \delta^2 - 2 \frac{p}{1-p} \delta^3} + pn\delta^2 \right). \tag{6.11}$$

By (5.6), the fact that $\ell < pn$ and (6.11), we conclude that

$$S_{d_1} - S_w < pn\delta \left(\frac{\frac{1+p}{1-p} + 2 \frac{p}{1-p} \delta}{(1 - \frac{\delta}{2}) \left(1 - \frac{1+p}{1-p} \delta^2 - 2 \frac{p}{1-p} \delta^3 \right)} + \frac{1}{(1 - \frac{\delta}{2})} \right). \tag{6.12}$$

The desired estimate (6.5) follows from (6.12) and the fact that $0 < \delta < \frac{1}{10}$ and $p \leq \frac{1}{2}$.

We proceed to show that inequality (6.6) is also satisfied. As before, we first observe that (6.4) yields that

$$1 > \left(1 + \frac{p}{1-p} \delta \right)^{S_{d_2} - S_w} \left(\left(1 + \frac{p}{1-p} \delta \right) \left(1 - \delta - 2 \frac{p}{1-p} \delta^2 \right) \right)^{S_w} \exp(-pn\delta^2). \tag{6.13}$$

However, since $0 < \delta < \frac{1}{10}$, we have

$$\left(1 + \frac{p}{1-p} \delta \right) \left(1 - \delta - 2 \frac{p}{1-p} \delta^2 \right) \geq 1 - \frac{1-2p}{1-p} \delta - \frac{16}{5} \frac{p}{1-p} \delta^2 \tag{6.14}$$

that combined with (6.13) yields that

$$1 > \left(1 + \frac{p}{1-p} \delta \right)^{S_{d_2} - S_w} \left(1 - \frac{1-2p}{1-p} \delta - \frac{16}{5} \frac{p}{1-p} \delta^2 \right)^{S_w} \exp(-pn\delta^2). \tag{6.15}$$

Now after taking logarithms and rearranging, we have

$$S_{d_2} - S_w < \frac{1}{\ln \left(1 + \frac{p}{1-p} \delta \right)} \left(S_w \ln \left(\frac{1}{1 - \frac{1-2p}{1-p} \delta - \frac{16}{5} \frac{p}{1-p} \delta^2} \right) + pn\delta^2 \right); \tag{6.16}$$

by Observation 6.2, this yields that

$$S_{d_2} - S_w < \frac{1-p}{p\delta \left(1 - \frac{p}{2(1-p)} \delta \right)} \left(S_w \frac{\frac{1-2p}{1-p} \delta + \frac{16}{5} \frac{p}{1-p} \delta^2}{1 - \frac{1-2p}{1-p} \delta - \frac{16}{5} \frac{p}{1-p} \delta^2} + pn\delta^2 \right) \tag{6.17}$$

that can be further simplified to

$$S_{d_2} - S_w < S_w \frac{\frac{1-2p}{p} + \frac{16}{5}\delta}{\left(1 - \frac{p}{2(1-p)}\delta\right)\left(1 - \frac{1-2p}{1-p}\delta - \frac{16}{5}\frac{p}{1-p}\delta^2\right)} + \frac{(1-p)n\delta}{1 - \frac{p}{2(1-p)}\delta}. \tag{6.18}$$

However, since $p \leq \frac{1}{2}$, for every $0 < \delta < \frac{1}{10}$ we have

$$\left(1 - \frac{p}{2(1-p)}\delta\right)\left(1 - \frac{1-2p}{1-p}\delta - \frac{16}{5}\frac{p}{1-p}\delta^2\right) \geq 1 - \delta; \tag{6.19}$$

indeed, after noticing that

$$1 - \frac{1-2p}{1-p}\delta - \frac{16}{5}\frac{p}{1-p}\delta^2 = 1 - \delta + \frac{p}{1-p}\delta - \frac{16}{5}\frac{p}{1-p}\delta^2, \tag{6.20}$$

the desired estimate (6.19) follows from the elementary inequality

$$\frac{p}{1-p}\delta + \frac{16}{5}\frac{p^2}{2(1-p)^2}\delta^3 - \frac{p}{2(1-p)}\delta(1-\delta) - \frac{p^2}{2(1-p)^2}\delta^2 - \frac{16}{5}\frac{p}{1-p}\delta^2 \geq 0. \tag{6.21}$$

By (6.18) and (6.19), we obtain that

$$S_{d_2} - S_w < S_w \left(\frac{1-2p}{p} + \frac{16}{5}\delta\right)\left(1 + \frac{10}{9}\delta\right) + (1-p)n\delta(1+\delta). \tag{6.22}$$

We then expand (6.22) to

$$S_{d_2} - S_w < S_w \left(\frac{1-2p}{p} + \frac{16}{5}\delta + \frac{1-2p}{p}\frac{10}{9}\delta + \frac{10}{9}\frac{16}{5}\delta^2\right) + (1-p)n\delta(1+\delta). \tag{6.23}$$

By (5.6), we see that $S_w \leq \ell < pn$, and so (6.23) yields that

$$S_{d_2} - \frac{1-p}{p}S_w < n\delta\left(p\frac{16}{5} + (1-2p)\frac{10}{9} + p\frac{10}{9}\frac{16}{5}\delta + (1-p)(1+\delta)\right). \tag{6.24}$$

Inequality (6.6) follows from (6.24) and the fact that $0 < \delta < \frac{1}{10}$ and $0 < p \leq \frac{1}{2}$. □

After these preliminary steps, we are ready to consider cases.

Case 1: $(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [0, b^*])$

Note that, in this case, the cardinality of the final forbidden interval $[0, b^*]$ is $b^* + 1$; however, the forbidden interval for the initial families \mathcal{F}, \mathcal{G} was a singleton. By Fact 5.1, the cardinality of the forbidden interval increases by 1 if and only if the algorithm executes an iteration of ‘type’ S_w . Thus,

$$b^* = S_w. \tag{6.25}$$

Next observe that the initial value of the lower bound a of the forbidden interval is equal to ℓ , and it is equal to 0 when the algorithm terminates. Using Fact 5.1 again, we see that the counter a decreases by 1 if and only if an iteration of ‘type’ S_{d_1} or an iteration of ‘type’ S_w is executed. Therefore, we also have that

$$S_{d_1} + S_w = \ell. \tag{6.26}$$

By (6.5) and (6.26), we obtain that

$$S_w \geq \frac{\ell}{2} - \frac{5}{2}pn\delta. \tag{6.27}$$

However, since $(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [0, b^*])$ and $b^* = S_w$, by part (i) of Lemma 3.3, we have

$$\mu_p(\mathcal{F}^*) \mu_{p'}(\mathcal{G}^*) \leq \exp\left(-\frac{S_w^2}{24pn}\right). \tag{6.28}$$

Combining (5.7), (6.3), (6.26) and (6.28), we obtain in particular that

$$\exp\left(-\frac{S_w^2}{24pn}\right) > (1 + \delta)^{\ell - S_w} \left(1 - \delta - 2\frac{p}{1-p}\delta^2\right)^{S_w} \exp(-pn\delta^2) \tag{6.29}$$

that implies, after taking logarithms and using Observation 6.2, that

$$(S_w - \ell) \left(\delta - \frac{\delta^2}{2}\right) + S_w \left(\frac{\delta + 2\frac{p}{1-p}\delta^2}{1 - \delta - 2\frac{p}{1-p}\delta^2}\right) + pn\delta^2 > \frac{S_w^2}{24pn}. \tag{6.30}$$

Moreover, since $0 < \delta < \frac{1}{10}$ and $0 < p \leq \frac{1}{2}$, we have

$$\delta + \frac{11}{3}\delta^2 \geq \frac{\delta + 2\frac{p}{1-p}\delta^2}{1 - \delta - 2\frac{p}{1-p}\delta^2}, \tag{6.31}$$

which, combined with (6.30), implies that

$$24pn(2S_w - \ell)\delta - 12pn(2S_w - \ell)\delta^2 + 100pnS_w\delta^2 + 24p^2n^2\delta^2 > S_w^2. \tag{6.32}$$

We now consider the following subcases.

Subcase 1.1: $S_w \leq \frac{2}{3}\ell$

In this subcase, by (6.27) and the choice of δ in (6.1), we have

$$-\frac{pn}{2} \leq 2S_w - \ell \leq \frac{S_w}{2} \quad \text{and} \quad \delta \leq \frac{2}{53} \frac{S_w}{pn}; \tag{6.33}$$

indeed, by (6.27) and (6.1), $S_w \geq \frac{\ell}{2} - \frac{5}{2}pn\delta \geq \ell\left(\frac{1}{2} - \frac{5}{116}\right) \geq 58pn\delta\left(\frac{1}{2} - \frac{5}{116}\right) \geq \frac{53}{2}pn\delta$. Hence, by (6.32), (6.33) and the fact that $\delta < \frac{1}{10}$, we obtain that

$$S_w^2 < 12pnS_w\delta + 30p^2n^2\delta^2 + 100pnS_w\delta^2 \leq \left(\frac{24}{53} + \frac{30 \cdot 4}{53^2} + \frac{200}{53} \cdot \frac{1}{10}\right)S_w^2 < S_w^2, \tag{6.34}$$

which is clearly a contradiction.

Subcase 1.2: $S_w > \frac{2}{3}\ell$

By (6.26) and (6.1), we have

$$0 \leq \frac{S_w}{2} \leq 2S_w - \ell \leq S_w \quad \text{and} \quad \delta \leq \frac{3}{116} \frac{S_w}{pn}. \tag{6.35}$$

Therefore, by (6.32) and (6.35), we get that

$$S_w^2 < 24pnS_w\delta + 100pnS_w\delta^2 + 24p^2n^2\delta^2 < \left(\frac{24 \cdot 3}{116} + \frac{100 \cdot 3}{116} \cdot \frac{1}{10} + \frac{24 \cdot 9}{116^2}\right)S_w^2 < S_w^2, \tag{6.36}$$

which leads, again, to a contradiction.

Case 2: $(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [a^*, m^*])$

The proof in this case is slightly more involved. We start by observing that the initial value of the upper bound b of the forbidden interval is equal to ℓ , and it is equal to m^* when the algorithm terminates. Moreover, by Fact 5.1, the counter b decreases by 1 if and only if an iteration of ‘type’ S_{d_1} is executed. Thus,

$$m^* = \ell - S_{d_1} \leq \ell. \tag{6.37}$$

Combining (5.5) and (6.37), we obtain that

$$S_{d_2} + S_w = n - \ell. \tag{6.38}$$

This identity together with (6.6) yields that $S_w \geq pn - 3pn\delta - p\ell$ that we rewrite as

$$S_w \geq (1 - p)\ell + ((pn - \ell) - 3pn\delta). \tag{6.39}$$

Moreover, as we have already noted in the previous case, the cardinality of the forbidden interval increases by 1 if and only if the algorithm executes an iteration of ‘type’ S_w . Therefore, we also have that

$$m^* - a^* = S_w. \tag{6.40}$$

Next, we introduce the quantity

$$\alpha := S_w - (1 - p)\ell, \tag{6.41}$$

and we observe that, by (6.39), we have the lower bound

$$\alpha \geq (pn - \ell) - 3pn\delta; \tag{6.42}$$

notice that $\alpha > 0$ by the choice of δ in (6.1). Also notice that, by (6.40) and (6.41),

$$a^* + (\ell - m^*) = p\ell - \alpha. \tag{6.43}$$

However, setting

$$\widehat{\mathcal{F}} := \{A \subseteq [\ell] : A \cap [m^*] \in \mathcal{F}^*\} \subseteq \{0, 1\}^\ell, \tag{6.44}$$

$$\widehat{\mathcal{G}} := \{B \subseteq [\ell] : B \cap [m^*] \in \mathcal{G}^*\} \subseteq \{0, 1\}^\ell, \tag{6.45}$$

by (6.43) and the fact that $(\mathcal{F}^*, \mathcal{G}^*) \in \text{Forbid}(m^*, [a^*, m^*])$, we see that

- (i) $\mu_p(\mathcal{F}^*) = \mu_p(\widehat{\mathcal{F}})$ and $\mu_{p'}(\mathcal{G}^*) = \mu_{p'}(\widehat{\mathcal{G}})$, and
- (ii) $(\widehat{\mathcal{F}}, \widehat{\mathcal{G}}) \in \text{Forbid}(\ell, [p\ell - \alpha, \ell])$.

Hence, by part (ii) of Lemma 3.3, we obtain that

$$\mu_p(\mathcal{F}^*) \mu_{p'}(\mathcal{G}^*) \leq 2 \exp\left(-\frac{\alpha^2}{24p\ell}\right). \tag{6.46}$$

In particular, by (5.7), (6.3) and (6.46), we have that

$$\exp\left(-\frac{\alpha^2}{24p\ell}\right) > \left(1 + \frac{p}{1-p}\delta\right)^{S_{d_2}} \left(1 - \delta - 2\frac{p}{1-p}\delta^2\right)^{S_w} \exp(-pn\delta^2). \tag{6.47}$$

As in the previous case, we will show that (6.47) leads to a contradiction.

To this end, it is enough to show that, by the choice of δ in (6.1), we have

$$\exp\left(-\frac{\alpha^2}{24p\ell}\right) \leq \left(1 + \frac{p}{1-p}\delta\right)^{S_{d_2}} \left(1 - \delta - 2\frac{p}{1-p}\delta^2\right)^{S_w} \exp(-pn\delta^2). \tag{6.48}$$

After taking logarithms and using Observation 6.2, it is enough to show that

$$pn\delta^2 + S_w \frac{\delta + 2\frac{p}{1-p}\delta^2}{1 - \delta - 2\frac{p}{1-p}\delta^2} \leq S_{d_2} \left(\frac{p}{1-p}\delta - \frac{p^2}{2(1-p)^2}\delta^2\right) + \frac{\alpha^2}{24p\ell}. \tag{6.49}$$

Since $0 < \delta < \frac{1}{10}$ and $p \leq \frac{1}{2}$, we have

$$\frac{1}{1 - \delta - 2\frac{p}{1-p}\delta^2} < \left(1 + \frac{3}{2}\delta\right), \tag{6.50}$$

$$\left(1 + 2\frac{p}{1-p}\delta\right) \left(1 + \frac{3}{2}\delta\right) \leq (1 + 4\delta). \tag{6.51}$$

Thus, by (6.49)–(6.51), it is enough to show that

$$pn\delta^2 + S_w \delta(1 + 4\delta) \leq S_{d_2} \left(\frac{p}{1-p}\delta - \frac{p^2}{2(1-p)^2}\delta^2\right) + \frac{\alpha^2}{24p\ell}, \tag{6.52}$$

which is equivalent to saying, after rearranging, that

$$\left(pn + 4S_w + \frac{p^2}{2(1-p)^2} S_{d_2}\right)\delta^2 + \left(S_w - \frac{p}{1-p} S_{d_2}\right)\delta \leq \frac{\alpha^2}{24p\ell}. \tag{6.53}$$

By (5.6), we have $S_w \leq \ell < pn$ and, clearly, $S_{d_2} \leq n$. Hence, by (6.53) and the fact that $0 < p \leq \frac{1}{2}$, it is enough to show that

$$6pn\delta^2 + \left(S_w - \frac{p}{1-p} S_{d_2}\right)\delta \leq \frac{\alpha^2}{24p\ell}. \tag{6.54}$$

Observe that

$$\begin{aligned} S_w - \frac{p}{1-p} S_{d_2} &\stackrel{(6.38)}{=} S_w - \frac{p}{1-p} (n - \ell - S_w) = \frac{1}{1-p} (S_w - pn + p\ell) \\ &\stackrel{(6.41)}{=} \frac{1}{1-p} (\alpha + (1-p)\ell - pn + p\ell) = \frac{1}{1-p} (\alpha - (pn - \ell)). \end{aligned} \tag{6.55}$$

In order to verify (6.54), we consider the following subcases.

Subcase 2.1: $\alpha \leq pn - \ell$

By (6.55), we have $S_w - \frac{p}{1-p} S_{d_2} \leq 0$, and so it is enough to show that

$$144p^2n^2\delta^2 \leq \alpha^2. \tag{6.56}$$

Since $0 < \delta < \frac{pn-\ell}{51pn}$, by the choice of δ in (6.1), the estimate (6.56) – and, consequently, (6.54) – follows from (6.42).

Subcase 2.2: $\alpha > pn - \ell$

In this subcase, by (6.55), we have $S_w - \frac{p}{1-p}S_{d_2} > 0$. Hence, using again the fact that $0 < \delta < \frac{pn-\ell}{51pn}$, it is enough to show that

$$\frac{144}{51^2}(pn - \ell)^2 + \frac{24}{51}\left(S_w - \frac{p}{1-p}S_{d_2}\right)(pn - \ell) \leq \alpha^2, \tag{6.57}$$

which is equivalent to saying, by (6.55), that

$$\left(\frac{144}{51^2} - \frac{24}{51(1-p)}\right)(pn - \ell)^2 + \frac{24}{51(1-p)}\alpha(pn - \ell) \leq \alpha^2. \tag{6.58}$$

Since $0 < p \leq \frac{1}{2}$, it is enough to show that

$$\frac{48}{51}(pn - \ell) \leq \alpha \tag{6.59}$$

that follows from our starting assumption that $\alpha > pn - \ell$.

Summing up, we conclude that (6.54) is satisfied, and as we have already indicated, this contradicts (6.47). This completes the proof that Case 2 cannot occur, and so the entire proof of Theorem 1.1 is completed.

7. Extensions of the main estimate

We start with the following proposition, which is the analogue of Theorem 1.1 for families of sets contained in layers of the cube.

Proposition 7.1. *Let $\ell \leq k \leq m \leq n$ be positive integers, and let $\mathcal{F} \subseteq \binom{[n]}{k}$ and $\mathcal{G} \subseteq \binom{[n]}{m}$ with $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, \{\ell\})$.*

(i) *If $k \leq \frac{n}{2}$ and $m \leq n - k$, then, setting $t := \min\{\ell, k - \ell\}$, we have*

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{m}} \leq 50 \sqrt{\frac{k(n-k)m(n-m)}{n^2}} \exp\left(-\frac{t^2}{58^2 k}\right). \tag{7.1}$$

(ii) *If $k \leq \frac{n}{2} \leq n - k < m \leq n - k + \ell$, then, setting $\bar{t} := \min\{k - \ell, n - m - (k - \ell)\}$,*

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{m}} \leq 50 \sqrt{\frac{k(n-k)m(n-m)}{n^2}} \exp\left(-\frac{\bar{t}^2}{58^2(n-m)}\right). \tag{7.2}$$

Proof. We start with the proof of part (i). Assume, first, that $m \leq \frac{n}{2}$. Then,

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{m}} \stackrel{(2.5)}{\leq} 25 \sqrt{\frac{k(n-k)m(n-m)}{n^2}} \mu_{\frac{k}{n}}(\mathcal{F}) \mu_{\frac{m}{n}}(\mathcal{G}). \tag{7.3}$$

Thus, in this case, (7.1) follows from (7.3) and (1.3) applied for ‘ $p = \frac{k}{n}$ ’ and ‘ $p' = \frac{m}{n}$ ’. Next, assume that $\frac{n}{2} \leq m \leq n - k$ and set $\bar{\mathcal{G}} := \{[n] \setminus G : G \in \mathcal{G}\} \subseteq \binom{[n]}{n-m}$. Notice that $(\mathcal{F}, \bar{\mathcal{G}}) \in \text{Forbid}(n, \{k - \ell\})$ and, moreover, $k - \ell \leq k \leq n - m \leq \frac{n}{2}$. Therefore, applying the estimate obtained in the first part of the proof to the pair $(\mathcal{F}, \bar{\mathcal{G}})$ and invoking the choice of t , we obtain that

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{m}} = \frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\overline{\mathcal{G}}|}{\binom{n}{n-m}} \leq 50 \sqrt{\frac{k(n-k)m(n-m)}{n^2}} \exp\left(-\frac{t^2}{58^2 k}\right). \tag{7.4}$$

We proceed to the proof of part (ii). As before, we set $\overline{\mathcal{G}} := \{[n] \setminus G : G \in \mathcal{G}\} \subseteq \binom{[n]}{n-m}$, and we observe that $(\mathcal{F}, \overline{\mathcal{G}}) \in \text{Forbid}(n, \{k - \ell\})$ and $k - \ell \leq n - m < k \leq \frac{n}{2}$. Thus, applying part (i) to the pair $(\mathcal{F}, \overline{\mathcal{G}})$ and using the fact that $\bar{t} = \min\{k - \ell, n - m - (k - \ell)\}$, we conclude that

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{m}} = \frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\overline{\mathcal{G}}|}{\binom{n}{n-m}} \leq 50 \sqrt{\frac{k(n-k)m(n-m)}{n^2}} \exp\left(-\frac{\bar{t}^2}{58^2 (n-m)}\right). \tag{7.5}$$

□

The next result supplements Theorem 1.1 and extends the subgaussian bound (1.3) to a wider range of parameters p, p' .

Proposition 7.2. *Let n be a positive integer, let $0 < p < \frac{1}{2} < p' \leq 1 - p$, and let $\ell \leq pn$ be a nonnegative integer. Also let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ be two families whose cross intersections forbid ℓ . Set $t := \min\{\ell, pn - \ell\}$ and assume that $t \geq 3$. Then we have*

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq t \cdot \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right). \tag{7.6}$$

Combining Theorem 1.1 and Proposition 7.2, we obtain the following corollary.

Corollary 7.3. *Let n be a positive integer, let $\frac{6}{n} \leq p \leq p' \leq 1 - p$, and let $\ell \leq pn$ be a nonnegative integer. Also let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ be two families whose cross intersections forbid ℓ . Set $t := \min\{\ell, pn - \ell\}$, and assume that $t \geq 210\sqrt{pn \ln(pn)}$. Then we have*

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq \exp\left(-\frac{t^2}{90^2 pn}\right). \tag{7.7}$$

We proceed to the proof of Proposition 7.2.

Proof of Proposition 7.2. Set $m := \frac{t}{30}$, and observe that, by Lemma 2.3,

$$\max\{\mu_p([n]^{<pn-m}), \mu_p([n]^{>pn+m})\} \leq \exp\left(-\frac{m^2}{6pn}\right) \leq \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right). \tag{7.8}$$

By (7.8) and the choice of m , there is a nonnegative integer i_0 with $pn - m \leq i_0 \leq pn + m$ such that, setting $\mathcal{F}_{i_0} := \mathcal{F} \cap \binom{[n]}{i_0}$, we have

$$\mu_p(\mathcal{F}) \leq 2 \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right) + \frac{t}{15} \mu_p(\mathcal{F}_{i_0}). \tag{7.9}$$

Set $\overline{\mathcal{G}} := \{[n] \setminus G : G \in \mathcal{G}\}$, and notice that

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) = \mu_p(\mathcal{F}) \mu_{1-p'}(\overline{\mathcal{G}}) \stackrel{(7.9)}{\leq} \frac{t}{15} \mu_p(\mathcal{F}_{i_0}) \mu_{1-p'}(\overline{\mathcal{G}}) + 2 \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right). \tag{7.10}$$

Next, observe that $i_0 \geq \ell$ and $(\mathcal{F}_{i_0}, \overline{\mathcal{G}}) \in \text{Forbid}(n, \{i_0 - \ell\})$. Since $0 < p \leq 1 - p' \leq \frac{1}{2}$ and $i_0 - \ell \leq pn$, by Theorem 1.1 and (7.10), we obtain that

$$\mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq 2 \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right) + \frac{2t}{15} \exp\left(-\frac{\bar{t}^2}{58^2 pn}\right), \tag{7.11}$$

where $\bar{t} := \min\{i_0 - \ell, pn - i_0 + \ell\}$.

Claim 7.4. We have that $|\bar{t} - t| \leq m = \frac{t}{30}$.

Proof of Claim 7.4. Suppose, towards a contradiction, that $|\bar{t} - t| > m$; that is, either $t + m < \bar{t}$ or $\bar{t} < t - m$. We recall that $pn - m \leq i_0 \leq pn + m$.

Assume, first, that $t + m < \bar{t}$. By the definition of t and \bar{t} , we see that

- (i) $\min\{\ell, pn - \ell\} + m < \bar{t} \leq i_0 - \ell$ and
- (ii) $\min\{\ell, pn - \ell\} + m < \bar{t} \leq pn - i_0 + \ell$.

If $\ell \leq pn - \ell$, then, by (ii), we obtain that $\ell + m < pn - i_0 + \ell$, which is a contradiction; however, if $pn - \ell \leq \ell$, then, by (i), we have $pn - \ell + m < i_0 - \ell$ which leads, again, to a contradiction.

Next assume that $\bar{t} < t - m$. Then,

- (iii) $\min\{i_0 - \ell, pn - i_0 + \ell\} + m < t \leq \ell$ and
- (iv) $\min\{i_0 - \ell, pn - i_0 + \ell\} + m < t \leq pn - \ell$.

Consequently, if $i_0 - \ell \leq pn - i_0 + \ell$, then, by (iv), we have $i_0 - \ell + m < pn - \ell$, which is a contradiction; finally, if $pn - i_0 + \ell \leq i_0 - \ell$, then, by (iii), we have $pn - i_0 + \ell + m < \ell$, which is also a contradiction. \square

By (7.11) and Claim 7.4, we conclude that

$$\begin{aligned} \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) &\leq 2 \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right) + \frac{2t}{15} \exp\left(-\frac{\bar{t}^2}{4 \cdot 30^2 pn}\right) \\ &\leq t \cdot \exp\left(-\frac{t^2}{6 \cdot 30^2 pn}\right). \end{aligned} \quad \square$$

8. Optimality

We proceed to discuss the optimality of the bounds obtained by Theorem 1.1 and its extension, Corollary 7.3. Specifically, fix a positive integer n , $0 < p \leq p' \leq 1 - p$ and a nonnegative integer $\ell \leq pn$, set

$$\varepsilon_n(p, p', \ell) := \max\{\varepsilon > 0 : \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) \leq e^{-\varepsilon} \text{ for every pair of nonempty families } (\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, \{\ell\})\}, \quad (8.1)$$

and observe that our goal reduces to that of obtaining appropriate upper bounds for $\varepsilon_n(p, p', \ell)$. To this end, we shall additionally assume that⁵

$$\frac{16}{n} \leq p \quad \text{and} \quad 2\sqrt{pn \ln(pn)} \leq \ell \leq pn - 2\sqrt{pn \ln(pn)}; \quad (8.2)$$

we will also use the following standard lower bounds of the biased measures of the tails of the binomial distribution (see, for example, [Ash65, p. 115]).

Lemma 8.1. Let k, n be positive integers, and let $0 < p \leq \frac{1}{2}$. If $k \leq pn$, then

$$\mu_p\left([n]^{\leq k}\right) \geq \frac{1}{\sqrt{8n\left(1 - \frac{k}{n}\right)\frac{k}{n}}} \exp\left(-\frac{(k - pn)^2}{p(1 - p)n}\right), \quad (8.3)$$

⁵Notice (8.2) slightly narrows down the regime where the bound (1.3) is nontrivial.

while if $pn \leq k < 2pn$, then

$$\mu_p([n]^{\geq k}) \geq \frac{1}{\sqrt{8n(1 - \frac{k}{n})\frac{k}{n}}} \exp\left(\frac{(k - pn)^2}{p(1 - p)n}\right). \tag{8.4}$$

8.1. The high-intersection case: $\ell \geq cpn$ for some constant $c > 0$

Set $\mathcal{F} := [n]^{<\ell}$ and $\mathcal{G} := \{0, 1\}^n$. Then $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, \{\ell\})$ and, moreover,

$$\begin{aligned} \mu_p(\mathcal{F}) \mu_{p'}(\mathcal{G}) = \mu_p([n]^{<\ell}) &\stackrel{(8.3)}{\geq} \frac{1}{\sqrt{8(\ell - 1)(1 - \frac{\ell - 1}{n})}} \exp\left(-\frac{(pn - \ell + 1)^2}{p(1 - p)n}\right) \\ &\geq \frac{1}{\sqrt{8pn}} \exp\left(-\frac{(pn - \ell)^2 + 2(pn - \ell) + 1}{p(1 - p)n}\right). \end{aligned} \tag{8.5}$$

Next, set $C := \max\{2, \frac{pn}{\ell}\} \leq \max\{2, \frac{1}{c}\}$ and $t := \min\{\ell, pn - \ell\}$; notice that if $\ell \geq \frac{pn}{2}$, then $C = 2$ and $t = pn - \ell$, while if $\ell < \frac{pn}{2}$, then $C \leq \frac{1}{c}$ and $t = \ell$. By (8.2), (8.5) and taking into account the previous observations, it is easy to see that

$$\varepsilon_n(p, p', \ell) \leq 4(C - 1)^2 \frac{t^2}{pn}. \tag{8.6}$$

In particular, under (8.2), if $\ell \geq \frac{pn}{2}$, then $\varepsilon_n(p, p', \ell) \leq 4 \frac{t^2}{pn}$.

Remark 8.2. Note that if $\mathcal{F} := [n]^{<\ell}$, then $|A \cap B| \neq \ell$ for every $A, B \in \mathcal{F}$ and, by (8.2) and (8.5), we have $\mu_p(\mathcal{F}) \geq \exp(-4(C - 1)^2 \frac{t^2}{pn})$, where $C = \max\{2, \frac{pn}{\ell}\}$. Thus, Theorem 1.1 and Corollary 7.3 are optimal in the regime $\ell \geq cpn$ also in the non-crossing case.

8.2. The symmetric case: $p' = 1 - p$

We will show that, under (8.2), we have

$$\varepsilon_n(p, 1 - p, \ell) \leq 4 \frac{t^2}{pn}, \tag{8.7}$$

where, as usual, $t := \min\{\ell, pn - \ell\}$; note that (8.7) includes the important special case $p = p' = \frac{1}{2}$ that corresponds to the uniform probability measure on $\{0, 1\}^n$.

The subcase ' $\ell \geq \frac{pn}{2}$ ' follows of course from (8.6), and so we may assume that $\ell \leq \frac{pn}{2}$. Set $\mathcal{F} := [n]^{>pn+\frac{\ell}{2}}$ and $\mathcal{G} := [n]^{\geq(1-p)n+\frac{\ell}{2}}$. Notice that $(\mathcal{F}, \mathcal{G}) \in \text{Forbid}(n, \{\ell\})$ and

$$\mu_p(\mathcal{F}) \geq \mu_p([n]^{\geq pn+\frac{\ell}{2}+1}) \stackrel{(8.4)}{\geq} \frac{1}{\sqrt{8pn}} \exp\left(-\frac{\ell^2}{4p(1 - p)n} - \frac{\ell + 1}{p(1 - p)n}\right), \tag{8.8}$$

$$\mu_{1-p}(\mathcal{G}) = \mu_p([n]^{\leq pn-\frac{\ell}{2}}) \stackrel{(8.3)}{\geq} \frac{1}{\sqrt{8pn}} \exp\left(-\frac{\ell^2}{4p(1 - p)n}\right). \tag{8.9}$$

Therefore, using (8.2) and observing that in this case we have $t = \ell$, we obtain that

$$\mu_p(\mathcal{F}) \mu_{1-p}(\mathcal{G}) \geq \exp\left(-2 \frac{t^2}{p(1 - p)n}\right), \tag{8.10}$$

which clearly yields (8.7).

Remark 8.3. It is unclear whether the subgaussian bound (1.3) is optimal (modulo universal constants) in the low-intersection and asymmetric case – namely, when $\ell = o(pn)$ and $p' < 1 - p$. The optimality of Theorem 1.1 in this regime is closely related⁶ to the problem of obtaining sharp estimates of the product $\mu_p(\mathcal{F})\mu_p(\mathcal{G})$ of the biased measures of a pair of families $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$ that are *cross- ℓ -intersecting*; that is, they satisfy $|A \cap B| \geq \ell$ for every $A \in \mathcal{F}$ and every $B \in \mathcal{G}$. The non-crossing case, $\mathcal{F} = \mathcal{G}$, is completely understood thanks to the seminal work of Ahlswede–Khachatrian [AK97, AK99] and the more recent work of Filmus [Fi17]; see, also, [Fi13, Theorem 3.28] for some progress for general cross-intersecting families.

9. Supersaturation

The main result in this section is a supersaturation version of Proposition 7.1, which is the analogue of [FR87, Theorem 1.14]. To state it we need, first, to introduce some pieces of notation. Let n be a positive integer, let $\mathcal{F}, \mathcal{G} \subseteq \{0, 1\}^n$, and let $S \subseteq [n]$. Given a nonnegative integer $\ell \leq n$, we set

$$I_\ell(\mathcal{F}, \mathcal{G}) := \{(F, G) \in \mathcal{F} \times \mathcal{G} : |F \cap G| = \ell\} \quad \text{and} \quad i_\ell(\mathcal{F}, \mathcal{G}) := |I_\ell(\mathcal{F}, \mathcal{G})| \tag{9.1}$$

and, respectively,

$$I_\ell(S, \mathcal{G}) := \{G \in \mathcal{G} : |S \cap G| = \ell\} \quad \text{and} \quad i_\ell(S, \mathcal{G}) := |I_\ell(S, \mathcal{G})|. \tag{9.2}$$

We have the following theorem.

Theorem 9.1. *Let ℓ, k, n be positive integers with $\ell < k \leq \frac{n}{2}$, and set $T := \max\{\ell, k - \ell\}$. Also let $\delta > 0$ and assume that $10^5 \sqrt{k}(\ln n)^{3/2} \leq \delta \leq \min\{\ell, k - \ell\}$. Finally, set*

$$\varepsilon(\delta) := \frac{\delta^4}{CT^2 \ell \left(\ln\left(\frac{n}{\delta}\right)\right)^4}, \tag{9.3}$$

where $C := 2^8 58^6 60^4$. If $\mathcal{F} \subseteq \binom{[n]}{k}$ and $\mathcal{G} \subseteq \binom{[n]}{n-k}$ satisfy

$$\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{n-k}} > \exp(-\varepsilon(\delta)), \tag{9.4}$$

then we have

$$\frac{i_\ell(\mathcal{F}, \mathcal{G})}{i_\ell\left(\binom{[n]}{k}, \binom{[n]}{n-k}\right)} > \exp(-\delta). \tag{9.5}$$

By Theorem 9.1, we obtain the following corollary.

Corollary 9.2. *Let ℓ, k, n, T, δ be as in Theorem 9.1 and set $\varepsilon'(\delta) := \frac{\delta^4}{CT^2 (k-\ell) \left(\ln\left(\frac{n}{\delta}\right)\right)^4}$, where $C := 2^8 58^6 60^4$. If $\mathcal{F}, \mathcal{G} \subseteq \binom{[n]}{k}$ satisfy $\frac{|\mathcal{F}|}{\binom{n}{k}} \cdot \frac{|\mathcal{G}|}{\binom{n}{k}} > \exp(-\varepsilon'(\delta))$, then*

$$\frac{i_\ell(\mathcal{F}, \mathcal{G})}{i_\ell\left(\binom{[n]}{k}, \binom{[n]}{k}\right)} > \exp(-\delta). \tag{9.6}$$

Proof. Set $\overline{\mathcal{G}} := \{[n] \setminus G : G \in \mathcal{G}\} \subseteq \binom{[n]}{n-k}$, and notice that $i_\ell(\mathcal{F}, \mathcal{G}) = i_{k-\ell}(\mathcal{F}, \overline{\mathcal{G}})$ and $i_\ell\left(\binom{[n]}{k}, \binom{[n]}{k}\right) = i_{k-\ell}\left(\binom{[n]}{k}, \binom{[n]}{n-k}\right)$. The result follows from these observations and Theorem 9.1 applied to \mathcal{F} and $\overline{\mathcal{G}}$. \square

We proceed to the proof of Theorem 9.1.

⁶In fact, the techniques developed in this paper show that these two problems are essentially equivalent.

Proof of Theorem 9.1. We argue as in the proof of [FR87, Theorem 1.14] with the main new ingredients being Theorem 1.1 and Proposition 7.1. For the reader’s convenience, we will first give a high level overview of the proof.

Our analysis is focused on the way the elements of \mathcal{F} and \mathcal{G} are correlated with arbitrary sets of size 2ℓ . More precisely, for every $A \in \binom{[n]}{2\ell}$, we define \mathcal{F}_A and \mathcal{G}_A to be the sets $F \in \mathcal{F}$ and $G \in \mathcal{G}$, respectively, whose intersection with A is roughly equal to ℓ . We shall informally refer to these sets as ‘good’.

The proof is then divided into three parts. In the first part, we show that there are many $A \in \binom{[n]}{2\ell}$ for which both \mathcal{F}_A and \mathcal{G}_A are large – this is the content of Claim 9.3. In the second part, we work towards a contradiction, and we show that if $I_\ell(\mathcal{F}, \mathcal{G})$ is small, then there exists $A_0 \in \binom{[n]}{2\ell}$ for which both \mathcal{F}_{A_0} and \mathcal{G}_{A_0} are large, and at the same time, there are few number of pairs $(F, G) \in \mathcal{F}_{A_0} \times \mathcal{G}_{A_0}$ whose intersection is of size ℓ and the size of its trace on A_0 is a specific proportion of ℓ – this is achieved in Claim 9.4. Finally, in the third step of the proof, we arrive to a contradiction. Specifically, the properties of A_0 imply that most of the ‘good’ $F \in \mathcal{F}_{A_0}$ and $G \in \mathcal{G}_{A_0}$ are also bad in the sense that they do not form a pair whose intersection is of size ℓ and its trace on A_0 has size specified by the previous step; this tension is enough to derive the contradiction.

We proceed to the details. We fix \mathcal{F}, \mathcal{G} that satisfy (9.4), and we assume, towards a contradiction, that (9.5) does not hold true. We set $d(\mathcal{F}) := \frac{|\mathcal{F}|}{\binom{n}{k}}$, $d(\mathcal{G}) := \frac{|\mathcal{G}|}{\binom{n}{n-k}}$ and

$$\alpha := 116 \sqrt{\ell \varepsilon(\delta)} \stackrel{(9.3)}{=} \frac{\delta^2}{(2^3 58^2 60^2) T(\ln(\frac{n}{\delta}))^2}. \tag{9.7}$$

We also notice that for every $S \in \binom{[n]}{2\ell-\alpha}$, we have

- $i_\ell(\binom{[n]}{k}, \binom{[n]}{n-k}) = \binom{n}{k} \binom{k}{\ell} \binom{n-k}{n-k-\ell}$,
- $i_\ell(S, \binom{[n]}{k}) = \binom{2\ell-\alpha}{\ell} \binom{n-2\ell+\alpha}{k-\ell}$,
- $i_\ell(S, \binom{[n]}{n-k}) = \binom{2\ell-\alpha}{\ell} \binom{n-2\ell+\alpha}{n-k-\ell}$.

Next, we set

- $\mathcal{S}_{\mathcal{F}} := \{S \in \binom{[n]}{2\ell-\alpha} : i_\ell(S, \mathcal{F}) \geq K_{\mathcal{F}}\}$ and $\mathcal{S}_{\mathcal{G}} := \{S \in \binom{[n]}{2\ell-\alpha} : i_\ell(S, \mathcal{G}) \geq K_{\mathcal{G}}\}$,

where $K_{\mathcal{F}} := \frac{d(\mathcal{F})}{2} i_\ell(S, \binom{[n]}{k})$ and $K_{\mathcal{G}} := \frac{d(\mathcal{G})}{2} i_\ell(S, \binom{[n]}{n-k})$. By [FR87, Lemma 4.1], we have

$$|\mathcal{S}_{\mathcal{F}}| \geq \frac{d(\mathcal{F})}{2} \binom{n}{2\ell-\alpha} \quad \text{and} \quad |\mathcal{S}_{\mathcal{G}}| \geq \frac{d(\mathcal{G})}{2} \binom{n}{2\ell-\alpha}. \tag{9.8}$$

Moreover, for every $A \in \binom{[n]}{2\ell}$, set

- $\mathcal{F}_A := \{F \in \mathcal{F} : \ell \leq |F \cap A| \leq \ell + \alpha\}$ and $\mathcal{G}_A := \{G \in \mathcal{G} : \ell \leq |G \cap A| \leq \ell + \alpha\}$,

and define the family

$$\mathcal{A} := \left\{ A \in \binom{[n]}{2\ell} : |\mathcal{F}_A| \geq K_{\mathcal{F}} \text{ and } |\mathcal{G}_A| \geq K_{\mathcal{G}} \right\}. \tag{9.9}$$

Claim 9.3. *We have*

$$|\mathcal{A}| \geq \max \left\{ \frac{|\mathcal{S}_{\mathcal{F}}|}{2 \binom{2\ell}{\alpha}}, \frac{|\mathcal{S}_{\mathcal{G}}|}{2 \binom{2\ell}{\alpha}} \right\} \geq \frac{1}{4} \frac{\binom{n}{2\ell-\alpha}}{\binom{2\ell}{\alpha}} \exp \left(- \frac{\varepsilon(\delta)}{2} \right). \tag{9.10}$$

Proof of Claim 9.3. We set $\mathcal{S}_{\mathcal{F}}^* := \{S \in \mathcal{S}_{\mathcal{F}} : \exists S' \in \mathcal{S}_{\mathcal{G}} \text{ with } |S' \setminus S| \leq \alpha\}$, and we observe that $(\mathcal{S}_{\mathcal{F}} \setminus \mathcal{S}_{\mathcal{F}}^*, \mathcal{S}_{\mathcal{G}}) \in \text{Forbid}(n, \{2\ell - 2\alpha\})$. Also note that $\varepsilon(\delta) \leq \frac{\ell}{2^8 58^6 60^4}$, and so, by (9.7), $\alpha = \min\{2\ell - 2\alpha, \alpha\}$.

(In fact, α is significantly smaller than ℓ .) By part (i) of Proposition 7.1, we obtain that⁷

$$\frac{|\mathcal{S}_{\mathcal{F}} \setminus \mathcal{S}_{\mathcal{F}}^*|}{\binom{n}{2\ell-\alpha}} \cdot \frac{|\mathcal{S}_{\mathcal{G}}|}{\binom{n}{2\ell-\alpha}} \leq 50 \frac{(2\ell - \alpha)(n - 2\ell + \alpha)}{n} \exp\left(-\frac{\alpha^2}{58^2(2\ell - \alpha)}\right). \tag{9.11}$$

Plugging (9.4) and (9.8) into (9.11), we see that

$$\frac{|\mathcal{S}_{\mathcal{F}} \setminus \mathcal{S}_{\mathcal{F}}^*|}{|\mathcal{S}_{\mathcal{F}}|} \leq 200 \frac{(2\ell - \alpha)(n - 2\ell + \alpha)}{n} \exp\left(\varepsilon(\delta) - \frac{\alpha^2}{58^2(2\ell - \alpha)}\right); \tag{9.12}$$

since $\varepsilon(\delta) - \frac{\alpha^2}{58^2(2\ell-\alpha)} < -\varepsilon(\delta)$, this in turn implies that $|\mathcal{S}_{\mathcal{F}}^*| \geq \frac{|\mathcal{S}_{\mathcal{F}}|}{2}$. Next, for every $S \in \mathcal{S}_{\mathcal{F}}^*$, we select $A \in \binom{[n]}{2\ell}$ and $S' \in \mathcal{S}_{\mathcal{G}}$ such that $S, S' \subseteq A$, and we observe that for each such A , there exist at most $\binom{2\ell}{\alpha}$ such S 's. Therefore,

$$|\mathcal{A}| \geq \frac{|\mathcal{S}_{\mathcal{F}}|}{2\binom{2\ell}{\alpha}}. \tag{9.13}$$

With identical arguments, we verify that $|\mathcal{A}| \geq \frac{|\mathcal{S}_{\mathcal{G}}|}{2\binom{2\ell}{\alpha}}$. The last inequality in (9.10) follows from the previous estimates, (9.4) and (9.8). The proof of Claim 9.3 is completed. \square

Next, set

$$\beta := \sqrt{2^3 58^2 T \alpha} \stackrel{(9.7)}{=} \sqrt{2^4 58^3 T \sqrt{\ell} \varepsilon(\delta)} \stackrel{(9.3)}{=} \frac{\delta}{60 \ln\left(\frac{n}{\delta}\right)}, \tag{9.14}$$

and note that $116\alpha \leq \beta \leq \frac{\ell}{30}$ and $10\beta \ln\left(\frac{n}{\beta}\right) \leq \delta$. (Indeed, $T = \max\{\ell, k - \ell\} \geq \ell \geq \alpha$, and so, by (9.14), we see that $\beta \geq 116\alpha$; however, since $\delta \leq \min\{\ell, k - \ell\} \leq \ell$, by (9.14) again, we obtain that $\beta \leq \frac{\ell}{30}$.) Moreover, for every $A \in \binom{[n]}{2\ell}$, set

$$y_A := |\{(F, G) \in I_{\ell}(\mathcal{F}, \mathcal{G}) : (F, G) \in \mathcal{F}_A \times \mathcal{G}_A \text{ and } |F \cap G \cap A| = \ell - \beta\}|. \tag{9.15}$$

For every $(F, G) \in I_{\ell}(\mathcal{F}, \mathcal{G})$, we can bound the number of $A \in \mathcal{A}$ for which (F, G) contributes to y_A by counting the ways we can first select $F \cap G \cap A$, then $F \cap ([n] \setminus G) \cap A$, then $([n] \setminus F) \cap G \cap A$, and finally $([n] \setminus F) \cap ([n] \setminus G) \cap A$. In particular, we have

$$\sum_{A \in \mathcal{A}} y_A \leq i_{\ell}(\mathcal{F}, \mathcal{G}) \binom{\ell}{\ell - \beta} \sum_{0 \leq i, j \leq \alpha} \binom{k - \ell}{i + \beta} \binom{n - k - \ell}{j + \beta} \binom{\ell}{i + j + \beta}. \tag{9.16}$$

Claim 9.4. Given our starting assumption that (9.5) does not hold true, we may select $A_0 \in \mathcal{A}$ such that

$$y_{A_0} \leq \binom{2\ell}{\ell} \binom{n - 2\ell}{k - \ell} \exp\left(-\frac{\delta}{3}\right). \tag{9.17}$$

Proof of Claim 9.4. First observe that

$$\circ i_{\ell}\left(\binom{[n]}{k}, \binom{[n]}{n-k}\right) = \binom{n}{k} \binom{k}{\ell} \binom{n-k}{n-k-\ell} = \binom{2\ell}{\ell} \binom{n-2\ell}{k-\ell} \binom{n}{2\ell}.$$

Next, using (i) our starting assumption that (9.5) does not hold true, (ii) the fact that $\alpha + \beta \leq 2\alpha + \beta \leq \frac{1}{2} \min\{\ell, k - \ell, n - k - \ell\}$, and (iii) the fact that the function $x \mapsto \binom{y}{x}$ is increasing for $x \leq \frac{y}{2}$, we obtain that

⁷If $2\ell - \alpha > \frac{n}{2}$, then (9.11) follows by applying Proposition 7.1 to the complementary families.

$$\sum_{A \in \mathcal{A}} y_A \leq \alpha^2 \binom{\ell}{\beta} \binom{k-\ell}{\alpha+\beta} \binom{n-k-\ell}{\alpha+\beta} \binom{\ell}{2\alpha+\beta} \binom{2\ell}{\ell} \binom{n-2\ell}{k-\ell} \binom{n}{2\ell} \exp(-\delta). \tag{9.18}$$

Therefore, by (9.10), there exists $A_0 \in \mathcal{A}$ such that

$$y_{A_0} \leq \binom{2\ell}{\ell} \binom{n-2\ell}{k-\ell} 4\alpha^2 \frac{\binom{2\ell}{\alpha} \binom{n}{2\ell}}{\binom{n}{2\ell-\alpha}} \binom{k-\ell}{\alpha+\beta} \times \binom{n-k-\ell}{\alpha+\beta} \binom{\ell}{2\alpha+\beta} \binom{\ell}{\beta} \exp\left(\frac{\varepsilon(\delta)}{2} - \delta\right). \tag{9.19}$$

Now we claim that

$$4\alpha^2 \frac{\binom{2\ell}{\alpha} \binom{n}{2\ell}}{\binom{n}{2\ell-\alpha}} \binom{k-\ell}{\alpha+\beta} \binom{n-k-\ell}{\alpha+\beta} \binom{\ell}{2\alpha+\beta} \binom{\ell}{\beta} \leq \binom{n}{\beta}^5. \tag{9.20}$$

To this end, notice that

$$\frac{\binom{2\ell}{\alpha} \binom{n}{2\ell}}{\binom{n}{2\ell-\alpha}} = \binom{n-2\ell+\alpha}{\alpha}, \tag{9.21}$$

and consequently,

$$\frac{\binom{2\ell}{\alpha} \binom{n}{2\ell}}{\binom{n}{2\ell-\alpha}} \binom{k-\ell}{\alpha+\beta} \binom{n-k-\ell}{\alpha+\beta} \binom{\ell}{2\alpha+\beta} \binom{\ell}{\beta} \leq \binom{n}{\alpha} \binom{n}{\alpha+\beta}^2 \binom{n}{2\alpha+\beta} \binom{n}{\beta}. \tag{9.22}$$

Moreover, by (2.6) and the fact that $2\alpha + \beta \leq \frac{\ell}{10} \leq \frac{n}{4}$,

$$\frac{\binom{n}{\alpha+\beta}^2 \binom{n}{2\alpha+\beta}}{\binom{n}{\beta}^3} \leq \left(\frac{50}{24}\right)^3 2^{6\alpha \log_2(n)}, \tag{9.23}$$

and similarly, since $\alpha \leq \frac{\beta}{116}$,

$$\frac{\binom{n}{\alpha}}{\binom{n}{\beta}} \leq \frac{25}{24} \sqrt{\frac{\beta}{\alpha}} 2^{-\frac{\beta}{2} \log_2(n)}. \tag{9.24}$$

After observing that

$$\frac{\beta}{2} \log_2(n) > 2 \log_2(2\alpha) + 6\alpha \log_2(n) + \frac{1}{2} \log_2\left(\frac{\beta}{\alpha}\right) + \log_2\left(\frac{25}{24}\right) + 3 \log_2\left(\frac{50}{24}\right), \tag{9.25}$$

we conclude that (9.20) is satisfied.

Summing up, we see that there exists $A_0 \in \mathcal{A}$ such that

$$y_{A_0} \leq \binom{2\ell}{\ell} \binom{n-2\ell}{k-\ell} \binom{n}{\beta}^5 \exp\left(\frac{\varepsilon(\delta)}{2} - \delta\right). \tag{9.26}$$

The claim follows from this estimate together with (2.6), and invoking the choices of $\varepsilon(\delta)$ and β in (9.3) and (9.14), respectively. □

Let $A_0 \in \mathcal{A}$ be as in Claim 9.4. We will show that

$$\frac{|\mathcal{F}_{A_0}|}{y_{A_0}} > 2 \quad \text{and} \quad \frac{|\mathcal{G}_{A_0}|}{y_{A_0}} > 2. \tag{9.27}$$

Indeed, by (9.9) and the choices of $K_{\mathcal{F}}$ and $K_{\mathcal{G}}$, we have

$$\frac{|\mathcal{F}_{A_0}|}{y_{A_0}} \geq \frac{\binom{2\ell-\alpha}{\ell} \binom{n-2\ell+\alpha}{k-\ell}}{2 \binom{2\ell}{\ell} \binom{n-2\ell}{k-\ell}} \exp\left(\frac{\delta}{3} - \varepsilon(\delta)\right) \tag{9.28}$$

and

$$\frac{|\mathcal{G}_{A_0}|}{y_{A_0}} \geq \frac{\binom{2\ell-\alpha}{\ell} \binom{n-2\ell+\alpha}{n-k-\ell}}{2 \binom{2\ell}{\ell} \binom{n-2\ell}{n-k-\ell}} \exp\left(\frac{\delta}{3} - \varepsilon(\delta)\right). \tag{9.29}$$

Noticing that

$$\frac{\binom{2\ell-\alpha}{\ell}}{\binom{2\ell}{\ell}} \geq 2^{-\alpha} \tag{9.30}$$

and using the previous two estimates and the choice of α in (9.7), we see that (9.27) is satisfied.

We introduce the families

$$\begin{aligned} \mathcal{D}_{\mathcal{F}} &:= \{F \in \mathcal{F}_{A_0} : \forall G \in \mathcal{G}_{A_0} (|F \cap G| \neq \ell \text{ or } |F \cap G \cap A_0| \neq \ell - \beta)\}, \\ \mathcal{D}_{\mathcal{G}} &:= \{G \in \mathcal{G}_{A_0} : \forall F \in \mathcal{F}_{A_0} (|F \cap G| \neq \ell \text{ or } |F \cap G \cap A_0| \neq \ell - \beta)\}, \\ \mathcal{D}_{\mathcal{F}}^* &:= \left\{B \subseteq A_0 : \left| \{F \in \mathcal{D}_{\mathcal{F}} : F \cap A_0 = B\} \right| > \frac{K_{\mathcal{F}}}{2^{2\ell+2}} \right\}, \\ \mathcal{D}_{\mathcal{G}}^* &:= \left\{B \subseteq A_0 : \left| \{G \in \mathcal{D}_{\mathcal{G}} : G \cap A_0 = B\} \right| > \frac{K_{\mathcal{G}}}{2^{2\ell+2}} \right\}. \end{aligned}$$

Note that $|\mathcal{D}_{\mathcal{F}}| > \frac{K_{\mathcal{F}}}{2}$ and $|\mathcal{D}_{\mathcal{G}}| > \frac{K_{\mathcal{G}}}{2}$, and so, there are $|\mathcal{D}_{\mathcal{F}}| > \frac{K_{\mathcal{F}}}{2}$ pairs (F, B) such that $F \cap A_0 = B$. For every $F \in \mathcal{D}_{\mathcal{F}}$, we have that $\ell \leq |F \cap A_0| \leq \ell + \alpha$; therefore, there are at most $2^{2\ell-1}$ such choices for $B = F \cap A_0$. However, for every such B , there at most $\binom{n-2\ell}{k-\ell}$ choices for $F \in \mathcal{D}_{\mathcal{F}}$. Hence,

$$\frac{K_{\mathcal{F}}}{2} < \frac{K_{\mathcal{F}}}{2^3} + |\mathcal{D}_{\mathcal{F}}^*| \binom{n-2\ell}{k-\ell}, \tag{9.31}$$

and similarly for \mathcal{G} . Consequently, we have

$$|\mathcal{D}_{\mathcal{F}}^*| > \frac{1}{4} \frac{K_{\mathcal{F}}}{\binom{n-2\ell}{k-\ell}} \quad \text{and} \quad |\mathcal{D}_{\mathcal{G}}^*| > \frac{1}{4} \frac{K_{\mathcal{G}}}{\binom{n-2\ell}{n-k-\ell}} \tag{9.32}$$

that implies that

$$|\mathcal{D}_{\mathcal{F}}^*| \cdot |\mathcal{D}_{\mathcal{G}}^*| \geq 2^{4\ell+1} \exp(-2(\varepsilon(\delta) + 2\alpha)) > 2^{4\ell+1} \exp\left(-\frac{\beta^2}{58^2\ell}\right). \tag{9.33}$$

By Theorem 1.1 applied for ‘ $p = p' = \frac{1}{2}$ ’, there exist $B_1 \in \mathcal{D}_{\mathcal{F}}^*$ and $B_2 \in \mathcal{D}_{\mathcal{G}}^*$ such that $|B_1 \cap B_2| = \ell - \beta$. Next, set $x := |B_1| - \ell$, $\mathcal{X} := \{F \setminus A_0 : F \in \mathcal{D}_{\mathcal{F}} \text{ and } F \cap A_0 = B_1\}$ and $y := |B_2| - \ell$, $\mathcal{Y} := \{G \setminus A_0 : G \in \mathcal{D}_{\mathcal{G}} \text{ and } G \cap A_0 = B_2\}$. Observe that $0 \leq x, y \leq \alpha$ and

$$|\mathcal{X}| \geq \frac{K_{\mathcal{F}}}{2^{2\ell+2}} \quad \text{and} \quad |\mathcal{Y}| \geq \frac{K_{\mathcal{G}}}{2^{2\ell+2}}. \tag{9.34}$$

Therefore, by our assumption on δ and the choice of α and β ,

$$\begin{aligned} |\mathcal{X}| \cdot |\mathcal{Y}| &\geq \binom{n-2\ell}{k-\ell-x} \cdot \binom{n-2\ell}{n-k-\ell-y} \cdot \frac{1}{2^{10}} \cdot \frac{1}{4\ell} \cdot e^\alpha \cdot \exp(-2(\varepsilon(\delta) + 2\alpha)) \\ &> \binom{n-2\ell}{k-\ell-x} \cdot \binom{n-2\ell}{n-k-\ell-y} \cdot 50n \cdot \exp\left(-\frac{\beta^2}{58^2(k-\ell-x)}\right). \end{aligned} \quad (9.35)$$

By part (i) of Proposition 7.1, there exist $H_1 \in \mathcal{X}$, $H_2 \in \mathcal{Y}$ such that $|H_1 \cap H_2| = \beta$. It follows that $(B_1 \cup H_1, B_2 \cup H_2) \in (\mathcal{F}_{A_0} \times \mathcal{G}_{A_0}) \cap (\mathcal{D}_{\mathcal{F}} \times \mathcal{D}_{\mathcal{G}})$, which clearly leads to a contradiction. \square

Acknowledgements. The authors would like to thank the anonymous referee for numerous comments, remarks and suggestions that helped us improve the exposition.

Competing interest. The authors have no competing interests to declare.

Funding statement. The research was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the ‘2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers’ (Project Number: HFRI-FM20-02717).

References

- [AK97] R. Ahlswede and L. H. Khachatrian, ‘The complete intersection theorem for systems of finite sets’, *European J. Combin.* **18** (1997), 125–136.
- [AK99] R. Ahlswede and L. H. Khachatrian, ‘A pushing-pulling method: new proofs of intersection theorems’, *Combinatorica* **19** (1999), 1–15.
- [AS16] N. Alon and J. H. Spencer, *The Probabilistic Method* (Wiley-Interscience Series in Discrete Mathematics and Optimization) (John Wiley & Sons, 2016).
- [Ash65] R. B. Ash, *Information Theory* (John Wiley & Sons, 1965).
- [BHT06] S. G. Bobkov, C. Houdré and P. Tetali, ‘The subgaussian constant and concentration inequalities’, *Israel J. Math.* **156** (2006), 255–283.
- [BL91] B. Bollobás and I. Leader, ‘Isoperimetric inequalities and fractional set systems’, *J. Combin. Theory Ser. A* **56** (1991), 63–74.
- [BCW99] H. Buhrman, R. Cleve and A. Wigderson, ‘Quantum vs. classical communication and computation’, in *STOC ’98 (Dallas, TX)* (ACM, New York, 1999), 63–68.
- [El22] D. Ellis, ‘Intersection problems in extremal combinatorics: theorems, techniques and questions old and new’, in *Surveys in Combinatorics 2022* (London Math. Soc. Lecture Note Series) vol. 481 (2022), 115–173.
- [EKL24] D. Ellis, N. Keller and N. Lifshitz, ‘Stability for the complete intersection theorem, and the forbidden intersection problem of Erdős and Sós’, *J. Eur. Math. Soc.* **26** (2024), 1611–1654.
- [Erd75] P. Erdős, ‘Problems and results in graph theory and combinatorial analysis’, in *Proc. 5th British Combinatorial Conference* (1975), 169–192.
- [EKR61] P. Erdős, C. Ko and R. Rado, ‘Intersection theorems for systems of finite sets’, *Quart. J. Math.* **12** (1961), 313–320.
- [Fi13] Y. Filmus, *Spectral Methods in Extremal Combinatorics*, Ph.D. dissertation, University of Toronto, 2013.
- [Fi17] Y. Filmus, ‘The weighted complete intersection theorem’, *J. Combin. Theory Ser. A* **151** (2017), 84–101.
- [FF85] P. Frankl and Z. Füredi, ‘Forbidding just one intersection’, *J. Combin. Theory Ser. A* **39** (1985), 160–176.
- [FT18] P. Frankl and N. Tokushige, *Extremal Problems for Finite Sets* (Student Mathematical Library) vol. 86 (American Mathematical Society, Providence, RI, 2018).
- [FR87] P. Frankl and V. Rödl, ‘Forbidden intersections’, *Trans. Amer. Math. Soc.* **300** (1987), 259–286.
- [FR90] P. Frankl and V. Rödl, ‘A partition property of simplices in Euclidean space’, *J. Amer. Math. Soc.* **3** (1990), 1–7.
- [FW81] P. Frankl and R. M. Wilson, ‘Intersection theorems with geometric consequences’, *Combinatorica* **1** (1981), 357–368.
- [JS68] K. Jogdeo and S. M. Samuels, ‘Monotone convergence of binomial probabilities and a generalization of Ramanujan’s equation’, *Ann. Math. Statist.* **39** (1968), 1191–1195.
- [KLLM23] P. Keevash, N. Lifshitz, E. Long and D. Minzer, ‘Forbidden intersections for codes’, *J. Lond. Math. Soc.* **108** (2023), 2037–2083.
- [KL016] P. Keevash and E. Long, ‘Frankl–Rödl-type theorems for codes and permutations’, *Trans. Amer. Math. Soc.* **369** (2016), 1147–1162.
- [KL020] P. Keevash and E. Long, ‘Forbidden vector-valued intersections’, *Proc. Lond. Math. Soc.* **121** (2020), 702–742.
- [KLi21] N. Keller and N. Lifshitz, ‘The junta method for hypergraphs and the Erdős–Chvátal simplex conjecture’, *Adv. Math.* **392** (2021), Article ID 107991, 95 p.
- [KSZ22] A. Kupavskii, A. Sagdeev and D. Zakharov, ‘Cutting corners’, *Preprint*, 2022, <https://arxiv.org/abs/2211.17150>.

- [KZ24] A. Kupavskii and D. Zaharov, 'Spread approximations for forbidden intersections problems', *Adv. Math.* **445** (2024), Article ID 109653, 29 p.
- [Ro55] H. Robbins, 'A remark on Stirling's formula', *Amer. Math. Monthly* **62** (1955), 26–29.
- [S99] J. Sgall, 'Bounds on pairs of families with restricted intersections', *Combinatorica* **19** (1999), 555–566.
- [Ta89] M. Talagrand, 'Isoperimetry and integrability of the sum of independent Banach-space valued random variables', *Ann. Probab.* **17** (1989), 1546–1570.