Low-Dimensional Faces of Random 0/1-Polytopes*

Volker Kaibel

DFG Research Center *Mathematics for key technologies*
TU Berlin MA 6–2
Straße des 17. Juni 136
10623 Berlin
Germany
kaibel@math.tu-berlin.de

Abstract. Let P be a random 0/1-polytope in $\mathbb{R}^d$ with $n(d)$ vertices, and denote by $\varphi_k(P)$ the *k-face density* of P, i.e., the quotient of the number of k-dimensional faces of P and $\binom{n(d)}{k+1}$. For each $k \geq 2$, we establish the existence of a sharp threshold for the k-face density and determine the values of the threshold numbers τ_k such that, for all $\varepsilon > 0$,

$$\mathbb{E}\left[\varphi_k(P)\right] \;=\; \begin{cases} 1 - \mathrm{o}(1) & \text{if } n(d) \leq 2^{(\tau_k - \varepsilon)d} \text{ for all } d \\ \mathrm{o}(1) & \text{if } n(d) \geq 2^{(\tau_k + \varepsilon)d} \text{ for all } d \end{cases}$$

holds for the expected value of $\varphi_k(P)$. The threshold for $k = 1$ has recently been determined in [1].
In particular, these results indicate that the high face densities often encountered in polyhedral combinatorics (e.g., for the cut-polytopes of complete graphs) should be considered more as a phenomenon of the general geometry of 0/1-polytopes than as a feature of the special combinatorics of the underlying problems.

1 Introduction and Results

Over the last decades, investigations of various special classes of 0/1-polytopes (convex hulls of sets of 0/1-points) have not only lead to beautiful structural results on combinatorial optimization problems, but also to powerful algorithms. Consequently, there has been some effort to learn more about the general class of 0/1-polytopes (see [2]).

In the 1980's, e.g., several results on the graphs of 0/1-polytopes have been obtained, most notably Naddef's proof [3] showing that they satisfy the Hirsch-conjecture. A quite spectacular achievement in 2000 was Bárány and Pór's theorem [4] stating that random 0/1-polytopes (within a certain range of vertex numbers) have super-exponentially (in the dimension) many facets. Their proof

* This work was done while the author was a member of the *Mathematical Sciences Research Institute* at Berkeley, CA, during Oct/Nov 2003.

is based on the methods developed in the early 1990's by Dyer, Füredi, and McDiarmid [5], in order to show that the expected volume of a random d-dimensional 0/1-polytope with n vertices drops from (almost) zero to (almost) one very quickly with n passing the threshold $2^{(1-(\log e)/2)d}$.

While Bárány and Pór's result sheds some light on the highest-dimensional faces of random 0/1-polytopes, we investigate their lower dimensional faces in this paper. For a polytope P with n vertices and some $k \in [\dim P]$ (with $[a] := \{1, 2, \ldots, \lfloor a \rfloor\}$), we call

$$\varphi_k(P) \ := \ \frac{f_k(P)}{\binom{n}{k+1}}$$

the *k-face density* of P, where $f_k(P)$ is the number of k-dimensional faces of P. Clearly, we have $0 < \varphi_k(P) \leq 1$, and $\varphi_k(P) = 1$ holds if and only if P is $(k+1)$-neighbourly in the usual polytope theoretical sense (see, e.g., [6]).

The 1-face density $\varphi_1(P)$ is the density of the graph of P. In this case, a threshold result for random 0/1-polytopes has recently been obtained in [1]. However, for specific classes of 0/1-polytopes, high k-face densities have been observed also for larger values of k. For example, the cut-polytopes of complete graphs have 2-face density equal to one (and thus, also 1-face density equal to one), i.e., every triple of vertices makes a triangle-face (see [7, 8]). Note that the cut-polytopes of complete graphs have $2^{\Theta(\sqrt{d})}$ vertices.

Here, we obtain that there is a sharp threshold for the k-face density of random 0/1-polytopes for all (fixed) k. The threshold values nicely extend the results for $k = 1$, while the proof becomes more involved and needs a heavier machinery (the one developed in the above mentioned paper by Dyer, Füredi, and McDiarmid). As a pay-back, the proof, however, reveals several interesting insights into the geometry of (random) 0/1-polytopes.

1.1 Results

Let us fix some $k \in \{1, 2, \ldots\}$, set $r := k + 1$, and let $n : \mathbb{N} \to \mathbb{N}$ be a function (with $n(d) \in [2^d]$).

Define

$$\mathrm{V}_d \ := \ \{0,1\}^d \qquad \text{and} \qquad \mathrm{Q}_d \ := \ [0,1]^d \ = \ \operatorname{conv} \mathrm{V}_d \ ,$$

and consider the following two models of random 0/1-polytopes.

For the first one, choose W uniformly at random from the $n(d)$-element subsets of V_d, and define $P_1 := \operatorname{conv} W$. This is the model referred to in the abstract.

For the second one, choose $S_1, \ldots, S_r, X_1, \ldots, X_{n(d)-r} \in \mathrm{V}_d$ independently uniformly at random, and define

$$S := \{S_1, \ldots, S_r\} \ , \quad X := \{X_1, \ldots, X_{n(d)-r}\} \ , \quad P_2 := \operatorname{conv}(X \cup S) \ .$$

The main part of the paper will be concerned with the proof of a threshold result (Theorem 1) within the second model. If, for some $\varepsilon > 0$, $n(d) \leq 2^{(\frac{1}{2}-\varepsilon)d}$

holds for all d, then $S_1, \dots, S_r, X_1, \dots, X_{n(d)-r}$ are pairwise different with high probability:

$$\mathbb{P}\left[|S \cup X| = n(d)\right] \;=\; 1 - \mathrm{o}(1) \tag{1}$$

This will allow us to deduce from Theorem 1 the threshold result within the first model promised in the abstract.

Throughout the paper, $\log(\cdot)$ and $\ln(\cdot)$ will denote the binary and the natural logarithm, respectively. For $0 < \xi < 1$, define

$$h(\xi) \;:=\; \xi \log \frac{1}{\xi} + (1-\xi) \log \frac{1}{1-\xi}$$

(i.e., $h(\cdot)$ is the binary entropy function). Let us define

$$H_r \;:=\; \frac{1}{2^r - 2} \sum_{i \in [r-1]} \binom{r}{i} h\big(\frac{i}{r}\big)$$

and

$$\tilde{\tau}_r \;=\; 1 - (1 - 2^{1-r}) H_r \;.$$

Note that we have $H_2 = 1$ and $0 < H_r < 1$ for $r \geq 3$.

Theorem 1. *Let $r \in \{3, 4, \dots\}$ and $\varepsilon > 0$.*

1. *If $n(d) \leq 2^{(\tilde{\tau}_r - \varepsilon)d}$ holds for all d, then we have*

$$\mathbb{P}\left[\text{conv } S \text{ is a face of } P_2\right] \;=\; 1 - \mathrm{o}(1) \;.$$

2. *If $n(d) \geq 2^{(\tilde{\tau}_r + \varepsilon)d}$ holds for all d, then we have*

$$\mathbb{P}\left[P_2 \cap \text{aff } S \text{ is a face of } P_2\right] \;=\; \mathrm{o}(1) \;.$$

From the evolution result on the density of the graphs of random 0/1-polytopes obtained in [1] one readily derives that the statement of Theorem 1 is also true for $r = 2$ (note $\tilde{\tau}_2 = \frac{1}{2}$).

Using Theorem 1 (for $r \in \{2, 3, \dots\}$), we can now prove the main result of the paper, where for $k \in \{1, 2, \dots\}$ we denote

$$\tau_k \;:=\; \tilde{\tau}_{k+1} \;=\; 1 - (1 - 2^{-k}) H_{k+1} \;.$$

Theorem 2. *Let $k \in \{1, 2, \dots\}$, $\varepsilon > 0$, and $n : \mathbb{N} \to \mathbb{N}$ be any function. For each $d \in \mathbb{N}$, choose an $n(d)$-element subset W of $\{0,1\}^d$ uniformly at random, and set $P := \text{conv } W$. Then*

$$\mathbb{E}\left[\varphi_k(P)\right] \;=\; \begin{cases} 1 - \mathrm{o}(1) & \text{if } n(d) \leq 2^{(\tau_k - \varepsilon)d} \text{ for all } d \\ \mathrm{o}(1) & \text{if } n(d) \geq 2^{(\tau_k + \varepsilon)d} \text{ for all } d \end{cases}$$

holds for the expected k-face density of P.

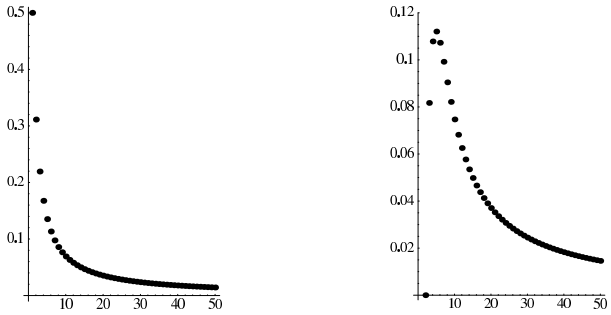

Fig. 1. The values τ_k for $k \geq 1$ and $1 - H_r$ (see Proposition 1) for $r \geq 2$.

Proof. Let us first consider the case $n(d) \leq 2^{(\tau_k - \varepsilon)d}$. We adopt the notation introduced in order to describe the first random model; in particular, $P_1 = P = \operatorname{conv} W$. Since $r = k + 1$ is constant, S (from the second random model) will consist of $k+1$ affinely independent points with (very) high probability for large d (see [9]). Thus, the first part of Theorem 1 here implies

$$\mathbb{P}\left[\operatorname{conv} S \text{ is a } k\text{-dimensional face of } P_2\right] \;=\; 1 - \mathrm{o}(1)\ .$$

Due to (1) (note $\tau_k \leq \frac{1}{2}$), this yields

$$\mathbb{P}\left[\operatorname{conv} T \text{ is a } k\text{-dimensional face of } P_1\right] \;=\; 1 - \mathrm{o}(1)$$

for T chosen uniformly at random from the $(k+1)$-subsets of (the random $n(d)$-set) W. But this probability obviously is a lower bound for $\mathbb{E}\left[\varphi_k(P_1)\right]$, which proves the first part of the theorem.

Now, we consider the case $n(d) \geq 2^{(\tau_k + \varepsilon)d}$. Similarly to the first case, the second part of Theorem 1 here implies

$$\mathbb{P}\left[P_2 \cap \operatorname{aff} S \text{ is a face of } P_2 \,\middle|\, |S| = k+1\right] \;=\; \mathrm{o}(1)\ . \tag{2}$$

Furthermore, it is easy to see that

$$\begin{aligned}\mathbb{P}\left[P_2 \cap \operatorname{aff} S \text{ is a face of } P_2 \,\middle|\, |S \cup X| = n(d)\right] \\ \leq\; \mathbb{P}\left[P_2 \cap \operatorname{aff} S \text{ is a face of } P_2 \,\middle|\, |S| = k+1\right]\end{aligned} \tag{3}$$

holds. From (2) and (3) one readily deduces

$$\mathbb{P}\left[P_1 \cap \operatorname{aff} T \text{ is a face of } P_1\right] \;=\; \mathrm{o}(1)$$

for T again chosen uniformly at random from the $(k+1)$-subsets of W. Since the number of k-faces of a polytope is at most the number of $(k+1)$-subsets of its vertex set for which the intersections of their affine hulls and the polytope are faces of the polytope, the latter probability is an upper bound for $\mathbb{E}\left[\varphi_k(P_1)\right]$. This proves the second part of the theorem.

1.2 Overview of the proof of Theorem 1

The structure of the proof is as follows: First, we will (in Section 2) reduce the proof of Theorem 1 to a statement (Proposition 1) about the event that S is not contained in a proper face of the cube, i.e., S is *spanning*. (A *proper* face of a polytope is any face that is not the entire polytope, which is considered a face of itself here.) This statement finally is proved in Section 5. There we need the results of Section 3 (for treating the cases *behind* the threshold) and Section 4 (for the cases *below* the threshold).

We will use only basic facts from polytope theory (such as in the proof of Theorem 2). Consult [6] in case of doubts – or for background information.

Throughout the paper, $r \in \{3, 4, \dots\}$ will be a constant.

Acknowledgments

I am grateful to the Mathematical Sciences Research Institute at Berkeley for the generous support and the excellent conditions I enjoyed during my visit in October/November 2003, when this work was done. I thank Günter M. Ziegler for comments on an earlier version of the paper.

2 Reduction to the spanning case

From now on, we stick to the second model of randomness. Thus, for some function $n : \mathbb{N} \to \mathbb{N}$, we choose the points $S_1, \dots, S_r, X_1, \dots, X_{n(d)-r} \in \mathrm{V}_d$ independently uniformly at random, and let $S := \{S_1, \dots, S_r\}$, $X := \{X_1, \dots, X_{n(d)-r}\}$, and $P := \mathrm{conv}\,(X \cup S)$. Denote by $F(S)$ the smallest face of the cube Q_d that contains S. Clearly, $P \cap F(S)$ is a face of P. Let $d(S)$ be the dimension of $F(S)$ (i.e., $d(S)$ is the number of coordinates where not all elements of S agree). If $F(S) = \mathrm{Q}_d$ (i.e., $d(S) = d$), then we call S *spanning*.

In Section 5, we will prove the following result (where ∂ denotes the boundary operator).

Proposition 1. *Let $r \in \{3, 4, \dots\}$ and $\varepsilon > 0$.*

1. *If $n(d) \le 2^{(1-H_r-\varepsilon)d}$ holds for all d, then we have*

$$\mathbb{P}\left[\mathrm{conv}\ S \text{ is a face of } P \,|\, S \text{ is spanning}\right] \;=\; 1 - \mathrm{o}(1)\ .$$

2. *If $n(d) \ge 2^{(1-H_r+\varepsilon)d}$ holds for all d, then we have*

$$\mathbb{P}\left[\mathrm{conv}\ S \subseteq \partial\, P \,|\, S \text{ is spanning}\right] \;=\; \mathrm{o}(1)\ .$$

Figure 1 illustrates the threshold values $1-H_r$. The aim of the current section is to show that Proposition 1 implies Theorem 1.

2.1 Preliminaries

Let A be the $r \times d$ matrix whose rows are $S_1, \ldots, S_r$. Clearly, $d(S)$ equals the number of columns of A which are neither $\mathbb{0}$ (the all-zero vector) nor $\mathbb{1}$ (the all-one vector).

The random matrix A is distributed in the same way as an $r \times d$ matrix is distributed whose columns are chosen independently uniformly at random from $\{0,1\}^r$. For $t \in \{0,1\}^r$ chosen uniformly at random, we have $\mathbb{P}[t \notin \{\mathbb{0}, \mathbb{1}\}] = 1 - 2^{1-r}$.

The de Moivre-Laplace Theorem (see, e.g., [10, Chap. 7]) yields that, for every $\delta > 0$, there is a $B_\delta > 0$ such that

$$\mathbb{P}\left[|d(S) - (1-2^{1-r})d| \le B_\delta \sqrt{d}\right] \;\ge\; 1 - \delta \tag{4}$$

holds for all large enough d.

For each $\delta > 0$, define

$$J_\delta(d) \;:=\; \left\{j \in [d] \,:\, |j - (1-2^{1-r})d| \le B_\delta \sqrt{d}\right\} .$$

Thus, by (4) we have

$$\mathbb{P}[d(S) \in J_\delta] \;\ge\; 1 - \delta \tag{5}$$

for all large enough d.

Let us denote

$$n(S) \;:=\; \left|\{i \in [n] \,:\, X_i \in F(S)\}\right| .$$

2.2 The case $\mathbf{n(d) \le 2^{(\tilde{\tau}_r - \varepsilon)d}}$

From elementary polytope theory one derives

$$\text{conv } S \text{ is a face of } P \;\Leftrightarrow\; \text{conv } S \text{ is a face of } P \cap F(S) . \tag{6}$$

Let $\delta > 0$ be fixed and let $j_{\min} \in J_\delta$ such that

$$\begin{aligned}&\mathbb{P}[\text{conv } S \text{ is a face of } P \,|\, d(S) = j_{\min}]\\ &\qquad = \;\min\left\{\mathbb{P}[\text{conv } S \text{ is a face of } P \,|\, d(S) = j] \,:\, j \in J_\delta(d)\right\} .\end{aligned}$$

Then we have

$$\left|d - \frac{j_{\min}}{1-2^{1-r}}\right| \;=\; \mathrm{o}(j_{\min}) .$$

We therefore obtain

$$\begin{aligned}\mathbb{E}\left[n(S) \,\middle|\, d(S) = j_{\min}\right] &= \quad 2^{j_{\min}-d}(n(d) - r)\\ &\le \quad 2^{j_{\min}-d+(\tilde{\tau}_r-\varepsilon)d+\mathrm{o}(j_{\min})}\\ &\le 2^{\frac{1-2^{1-r}+\tilde{\tau}_r-1-\varepsilon}{1-2^{1-r}} j_{\min}+\mathrm{o}(j_{\min})} .\end{aligned} \tag{7}$$

The fraction in the exponent equals $1 - H_r - \varepsilon'$ where $\varepsilon' := \frac{\varepsilon}{1-2^{1-r}} > 0$. By Markov's inequality, we obtain

$$\mathbb{P}\left[n(S) \leq 2^{(1-H_r-\varepsilon'/2)j_{\min}} \,|\, d(S) = j_{\min}\right] \;=\; 1 - \mathrm{o}(1)\;. \tag{8}$$

Proposition 1 implies

$$\mathbb{P}\left[\text{conv } S \text{ is a face of } P \cap F(S) \,|\, d(S) = j_{\min}, n(S) \leq 2^{(1-H_r-\varepsilon'/2)j_{\min}}\right] \;=\; 1 - \mathrm{o}(1)\;.$$

Together with (8), the definition of $j_{\min}$, and (5), this implies

$$\mathbb{P}\left[\text{conv } S \text{ is a face of } P \cap F(S)\right] \;=\; 1 - \mathrm{o}(1)\;,$$

which, by (6), proves the first part of Theorem 1.

2.3 The case $\mathbf{n(d) \geq 2^{(\tilde{\tau}_r+\varepsilon)d}}$

Again, elementary polytope theory tells us

$$P \cap \text{aff } S \text{ is a face of } P \;\Rightarrow\; P \cap \text{aff } S = P \cap F(S) \text{ or } \text{conv } S \subseteq \partial(P \cap F(S))\;. \tag{9}$$

We omit the calculations that are necessary to prove the following lemma.

Lemma 1. *Let $\alpha, \beta, \gamma > 0$ with $\alpha + \beta > 1 + \beta\gamma$, $\tilde{n}(d) := \lfloor 2^{\alpha d} \rfloor$, $j(d) = \beta d + \mathrm{o}(d)$, and let F be any $j(d)$-dimensional face of Q_d. If $X_1, \ldots, X_{\tilde{n}(d)}$ are chosen independently uniformly at random from V_d, then we have*

$$\mathbb{P}\left[\left|\{i \in [\tilde{n}(d)] \,:\, X_i \in F\}\right| \geq 2^{\gamma j(d)}\right] \;=\; 1 - \mathrm{o}(1)\;.$$

Now we can prove the second part of Theorem 1 (using Proposition 1). Let $\delta > 0$ be fixed and let $j_{\max} \in J_\delta$ such that

$$\mathbb{P}\left[\text{conv } S \subseteq \partial(P \cap F(S)) \,|\, d(S) = j_{\max}\right] \;=\; \max\left\{\mathbb{P}\left[\text{conv } S \subseteq \partial(P \cap F(S)) \,|\, d(S) = j\right] \,:\, j \in J_\delta(d)\right\}\;.$$

With $\alpha := \tilde{\tau}_r + \varepsilon$, $\beta := 1 - 2^{1-r}$, and $\gamma := 1 - H_r + \varepsilon$, one easily verifies $\alpha + \beta > 1 + \beta\gamma$. Since $j_{\max} = (1 - 2^{1-r})d + \mathrm{o}(d)$ we thus obtain from Lemma 1

$$\mathbb{P}\left[n(S) \geq 2^{(1-H_r+\varepsilon)j_{\max}} \,|\, d(S) = j_{\max}\right] \;=\; 1 - \mathrm{o}(1)\;. \tag{10}$$

The second part of Proposition 1 implies

$$\mathbb{P}\left[\text{conv } S \subseteq \partial(P \cap F(S)) \,|\, d(S) = j_{\max}, n(S) \geq 2^{(1-H_r+\varepsilon)j_{\max}}\right] \;=\; \mathrm{o}(1)\;.$$

Furthermore, since $\dim(\text{aff } S)$ is constant, we obviously have

$$\mathbb{P}\left[P \cap \text{aff } S = P \cap F(S) \,|\, d(S) = j_{\max}, n(S) \geq 2^{(1-H_r+\varepsilon)j_{\max}}\right] \;=\; \mathrm{o}(1)\;.$$

Together with (10), the definition of $j_{\max}$, and (5), the latter two equations even hold for the corresponding unconditioned probabilities. Thus, we have

$$\mathbb{P}\left[\text{conv } S \subseteq \partial(P \cap F(S)) \text{ or } P \cap \text{aff } S = P \cap F(S)\right] \;=\; \mathrm{o}(1)\;,$$

which, due to (9), proves the second part of Theorem 1.

3 Membership probabilities

Here, we derive (from Dyer, Füredi, and McDiarmid's paper [5]) suitable lower bounds on $n(d)$ that, for specified points of Q_d, guarantee their membership in our random 0/1-polytopes with high probability.

For any $z \in \mathrm{Q}_d$, let us define

$$p(z) \;:=\; \frac{1}{2^d} \min \left\{ |U \cap \mathrm{V}_d \,| \,:\, U \subset \mathbb{R}^d \text{ (closed affine) halfspace}, z \in U \right\} .$$

For each $\alpha > 0$, denote

$$\mathrm{Q}_d^\alpha \;:=\; \{z \in \mathrm{Q}_d \,:\, p(z) \geq 2^{-\alpha d}\} .$$

For $z = (\zeta_1, \ldots, \zeta_d) \in \operatorname{int} \mathrm{Q}_d$ (the interior of Q_d), define

$$H(z) \;:=\; \frac{1}{d} \sum_{j \in [d]} h(\zeta_j) .$$

From Lemmas 2.1 and 4.1 of [5] one can deduce the following fact. Let us mention that in particular the proof of Lemma 4.1 (needed for part (2) of Lemma 2) is quite hard. It is the core of Dyer, Füredi, and McDiarmid's beautiful paper.

Lemma 2. *Let $\alpha, \varepsilon > 0$.*

1. *If $\tilde{n}(d) \geq 2^{(\alpha+\varepsilon)d}$ holds for all d, and $X_1, \ldots, X_{\tilde{n}(d)} \in \mathrm{V}_d$ are chosen independently uniformly at random, then we have*

$$\mathbb{P}\left[\mathrm{Q}_d^\alpha \subseteq \operatorname{conv}\{X_1, \ldots, X_{\tilde{n}(d)}\}\right] \;=\; 1 - \mathrm{o}(1) .$$

2. *For large enough d,*

$$\{z \in \operatorname{int} \mathrm{Q}_d \,:\, H(z) \geq 1 - \alpha + \varepsilon\} \;\subseteq\; \mathrm{Q}_d^\alpha$$

holds.

The following straight consequence (choose $\alpha := 1 - \beta + \varepsilon/2$) of Lemma 2 is the key to the proof of the second part of Proposition 1.

Corollary 1. *If $\beta > 0$, $\tilde{n}(d) \geq 2^{(1-\beta+\varepsilon)d}$ for all d, and $X_1, \ldots, X_{\tilde{n}(d)} \in \mathrm{V}_d$ are chosen independently uniformly at random, then we have*

$$\mathbb{P}\left[\{z \in \operatorname{int} \mathrm{Q}_d \,:\, H(z) \geq \beta\} \subseteq \operatorname{conv}\{X_1, \ldots, X_{\tilde{n}(d)}\}\right] \;=\; 1 - \mathrm{o}(1) .$$

4 Shallow cuts of the cube

This section is the heart of the proof of (the first part of) Proposition 1.

For $m \in \{1, 2, \dots\}$, let $A(m)$ be an $r \times M$ matrix with $M := (2^r - 2)m$ that has as its columns m copies of each vector $v \in \{0,1\}^r \setminus \{\mathbb{0}, \mathbb{1}\}$. This choice is motivated by the following fact (which is, however, irrelevant in this section): If $S_1, \dots, S_r$ are chosen independently uniformly at random from V_M, then the multiplicity m of each vector $v \in \{0,1\}^r$ among the columns of $A(m)$ equals the expected number of appearances of v as a column of the matrix with rows $S_1, \dots, S_r$ — conditioned on the event that S is spanning.

Let $s_1, \dots, s_r \in \{0,1\}^M$ be the rows of $A(m)$, and let, for $1 \le i \le r-1$, $L(i)$ be the set of indices of columns that have precisely i ones. We have $|L(i)| = \binom{r}{i} m$. Denote by $\sigma(i)$ the number of ones that any of the rows has in columns indexed by $L(i)$ (these numbers are equal for all rows). Obviously, we have $\sigma(i) = \frac{i}{r}\binom{r}{i} m$.

Let $b := (\beta_1, \dots, \beta_M)$ be the barycenter of the rows $s_1, \dots, s_r$. For each $j \in [M]$ we thus have $\beta_j = \frac{i(j)}{r}$, if $j \in L(i(j))$. Consequently (with the definition of $H(\cdot)$ from Section 3),

$$H(b) \;=\; \frac{1}{M} \sum_{i \in [r-1]} \binom{r}{i} m\, h\big(\frac{i}{r}\big) \;=\; H_r \;. \tag{11}$$

From Section 3 (see Lemma 2) we know that no hyperplane in $\mathbb{R}^M$ that contains b can therefore cut off significantly *less* than $2^{H_r M}$ points from V_M, and that there are indeed hyperplanes containing b that do also not cut off significantly *more* than $2^{H_r M}$ cube vertices. However, for our purposes, it will be necessary to know that there is a hyperplane containing not only b, but even the entire set $\{s_1, \dots, s_r\}$, and nevertheless cutting off not significantly more than $2^{H_r M}$ cube vertices.

The next result guarantees the existence of such a hyperplane, i.e., a certain shallow cut of the cube. Its proof will also reveal the basic reason for the appearance of the entropy function $h(\cdot)$: It is due to the well-known fact that, for any constant $\alpha > 0$,

$$\sum_{p \in [\alpha q]} \binom{q}{p} \;=\; 2^{h(\alpha) q + \mathrm{o}(q)} \tag{12}$$

(see, e.g., [11, Chap. 9,Ex. 42]).

Proposition 2. *There are coefficients $\alpha_1, \dots, \alpha_{r-1} \in \mathbb{R}$, such that the inequality*

$$\sum_{i \in [r-1]} \sum_{j \in L(i)} \alpha_i \xi_j \;\le\; \sum_{i \in [r-1]} \alpha_i \sigma(i) \tag{13}$$

has at most $2^{H_r M + \mathrm{o}(M)}$ 0/1-solutions $(\xi_1, \dots, \xi_M) \in \{0,1\}^M$. (By construction, the 0/1-points $s_1, \dots, s_r$ satisfy (13) with equality.)

Proof. Throughout the proof, we denote the components of any vectors $a, l, z \in \mathbb{R}^{r-1}$ by α_i, λ_i, and ζ_i, respectively.

For every $a \in \mathbb{R}^{r-1}$ and $l \in \mathbb{N}^{r-1}$, denote by $\omega_a(l)$ the number of 0/1-solutions to (13) with precisely λ_i ones in components indexed by $L(i)$ and define

$$\omega(l) \;:=\; \prod_{i\in[r-1]} \binom{\binom{r}{i}m}{\lambda_i} \,.$$

With

$$L_a \;:=\; \big\{ l \in \mathbb{N}^{r-1} \,:\, \sum_{i\in[r-1]} \alpha_i\lambda_i \;\le\; \sum_{i\in[r-1]} \alpha_i\sigma(i) \big\}$$

we thus have

$$\omega_a(l) \;=\; \begin{cases} \omega(l) & \text{if } l \in L_a \\ 0 & \text{otherwise} \end{cases} \,.$$

Consequently, the number of 0/1-points satisfying (13) is precisely

$$\sum_{l\in L_a} \omega(l) \,. \tag{14}$$

If, for some i, we have $\lambda_i > \binom{r}{i}m$, then clearly $\omega(l) = 0$. Thus, the number of nonzero summands in (14) is $\mathrm{O}(m^r)$. Below, we will exhibit a vector $a \in \mathbb{R}^{r-1}$ of (constant) coefficients that satisfies, with $z^\star := (\sigma(1), \ldots, \sigma(r-1))$,

$$\omega(l) \;\le\; \omega(z^\star)\, 2^{\mathrm{o}(M)} \tag{15}$$

for all $l \in L_a$. This will eventually prove the proposition, since we have

$$\begin{aligned} \omega(z^\star) &= \prod_{i\in[r-1]} \tbinom{\binom{r}{i}m}{\sigma(i)} \\ &= \prod_{i\in[r-1]} \tbinom{\binom{r}{i}m}{(i/r)\binom{r}{i}m} \\ &= \prod_{i\in[r-1]} 2^{h(i/r)\binom{r}{i}m+\mathrm{o}(m)} \\ &= 2^{\sum_{i\in[r-1]} h(i/r)\binom{r}{i}m+\mathrm{o}(m)} \\ &= 2^{H_r M+\mathrm{o}(M)} \end{aligned}$$

(where the third equation is due to (12), and for the the last one, see (11)).

We now approximate the function $\omega(\cdot)$ by Sterling's formula (see, e.g., [11, Eq. (9.40)])

$$N! \;=\; \Theta\Big(\sqrt{N}\frac{N^N}{e^N}\Big) \,.$$

For simplicity, we define $M_i := \binom{r}{i}m$. Thus we obtain

$$\omega(l) \;\le\; \mathrm{O}(M^r) \prod_{i\in[r-1]} \frac{M_i^{M_i}}{\lambda_i^{\lambda_i}(M_i-\lambda_i)^{M_i-\lambda_i}}$$

(with $0^0 = 1$). Let us define the closed box

$$B \;:=\; [0, M_1] \times [0, M_2] \times \cdots \times [0, M_{r-1}] \;,$$

the map $\eta : B \to \mathbb{R}$ via

$$\eta(z) \;:=\; \prod_{i \in [r-1]} \frac{M_i^{M_i}}{\zeta_i^{\zeta_i} (M_i - \zeta_i)^{M_i - \zeta_i}} \;,$$

and the halfspace

$$U_a \;:=\; \left\{ z \in \mathbb{R}^{r-1} \,:\, \sum_{i \in [r-1]} \alpha_i \zeta_i \;\leq\; \sum_{i \in [r-1]} \alpha_i \sigma(i) \right\} .$$

We have

$$\{ l \in L_a \,:\, \omega(l) > 0 \} \;\subseteq\; B \cap U_a \;.$$

By the continuity of η on B it hence suffices to determine $a \in \mathbb{R}^{r-1}$ such that $\eta(z^\star) \geq \eta(z)$ holds for all $z \in U_a \cap \operatorname{int} B$. Note that $z^\star$ itself is contained in the interior $\operatorname{int} B$ of the box B, where η is a differentiable function.

In fact, since $\ln(\cdot)$ is monotonically increasing, we may equivalently investigate the function $\tilde{\eta} : \operatorname{int} B \to \mathbb{R}$ defined via

$$\tilde{\eta}(z) \;:=\; \ln \eta(z) = \sum_{i \in [r-1]} M_i \ln M_i - \sum_{i \in [r-1]} \big(\zeta_i \ln \zeta_i + (M_i - \zeta_i) \ln (M_i - \zeta_i) \big) \;,$$

and thus find a vector $a \in \mathbb{R}^{r-1}$ of coefficients with

$$\tilde{\eta}(z^\star) \;\geq\; \tilde{\eta}(z) \qquad \text{for all } z \in U_a \cap \operatorname{int} B \;. \tag{16}$$

Now we choose the vector $a \in \mathbb{R}^{r-1}$ to be the gradient of $\tilde{\eta}$ at $z^\star$. One easily calculates

$$\alpha_i = \ln \frac{M_i - \sigma(i)}{\sigma(i)} \;.$$

In order to prove that, with this choice, (16) holds, let $z \in U_a \cap \operatorname{int} B$ be arbitrary ($z \neq z^\star$). Define $v := z - z^\star$, and consider the function $\tilde{\eta}_{z^\star, z} : [0, 1] \to \mathbb{R}$ defined via $\tilde{\eta}_{z^\star, z}(t) := \tilde{\eta}(z^\star + tv)$. The derivative of this function on $(0, 1)$ is

$$\tilde{\eta}'_{z^\star, z}(t) \;=\; \sum_{i \in [r-1]} v_i \ln \frac{M_i - \sigma(i) - t v_i}{\sigma(i) + t v_i} \;. \tag{17}$$

Consider any $i \in [r-1]$, and define $\varrho(t) := \frac{M_i - \sigma(i) - t v_i}{\sigma(i) + t v_i}$. If $v_i \geq 0$, then $\varrho(t) \leq \varrho(0)$, therefore, $v_i \ln \varrho(t) \leq v_i \ln \varrho(0) = \alpha_i v_i$. If $v_i < 0$, then $\varrho(t) > \varrho(0)$, and thus, $v_i \ln \varrho(t) < v_i \ln \varrho(0) = \alpha_i v_i$. Hence, in any case the i-th summand in (17) is at most as large as $\alpha_i v_i$. Therefore, we obtain

$$\tilde{\eta}'_{z^\star, z}(t) \;\leq\; \sum_{i \in [r-1]} \alpha_i v_i \;.$$

Since $z \in U_a$, we have $\sum_{i \in [r-1]} \alpha_i v_i \leq 0$. Thus, $\tilde{\eta}'_{z^\star, z}(t) \leq 0$ for all $t \in (0, 1)$. Since $\tilde{\eta}_{z^\star, z}$ is continuous on $[0, 1]$, we hence conclude $\tilde{\eta}(z^\star) \geq \tilde{\eta}(z)$.

5 The spanning case

Using the material collected in Sections 3 and 4, we will now prove Proposition 1 (and thus, as shown in Section 2) Theorem 1.

Towards this end, let $S_1, \ldots, S_r, X_1, \ldots, X_{n(d)-r} \in \mathrm{V}_d$ be chosen according to the probability distribution induced by our usual distribution (choosing all points independently uniformly at random) on the event that $S := \{S_1, \ldots, S_r\}$ is spanning. As before, define $S := \{S_1, \ldots, S_r\}$, $X := \{X_1, \ldots, X_{n(d)-r}\}$, and $P := \mathrm{conv}\,(S \cup X)$.

Let A be the $r \times d$ matrix with rows $S_1, \ldots, S_r$. Then A is a random matrix that has the same distribution as the $r \times d$ random matrix A' which arises from choosing each column independently uniformly at random from $\{0,1\}^r \setminus \{\mathbb{0}, \mathbb{1}\}$. Therefore, if we denote the columns of A by $t_1, \ldots, t_d \in \{0,1\}^r$, then the t_j are (independently) distributed according to the distribution

$$\mathbb{P}\,[t_j = t] \;=\; \frac{1}{2^r - 2} \;=:\; \pi$$

for each $t \in \{0,1\}^r \setminus \{\mathbb{0}, \mathbb{1}\}$.

Define

$$T_r \;:=\; \{0,1\}^d \setminus \{\mathbb{0}, \mathbb{1}\}\ ,$$

and denote, for every $t \in T_r$,

$$J(t) \;:=\; \{j \in [d] \,:\, t_j = t\}\ .$$

Let $m \in \mathbb{N}$ be the largest number such that $m \leq |J(t)|$ holds for all $t \in T_r$. For each t, choose an arbitrary subset $\tilde{J}(t) \subseteq J(t)$ with $|\tilde{J}(t)| = m$.

Denote by

$$\Delta_{\max} \;:=\; \max\left\{\left||J(t)| - \pi d\right| \,:\, t \in T_r\right\}$$

the maximal deviation of any $|J(t)|$ from its expected value πd.

From the de Moivre-Laplace Theorem (see, e.g., [10, Chap. 7]) one deduces the following for each $t \in T_r$: For every $\gamma' > 0$ there is a $C'_{\gamma'} > 0$ such that

$$\mathbb{P}\,\left[\left||J(t)| - \pi d\right| \leq C'_{\gamma'}\sqrt{d}\right] \;\geq\; 1 - \gamma'$$

holds for all large enough d. Since $|T_r|$ is a constant, one can even derive the following stronger result from this: For every $\gamma > 0$ there is a constant $C_\gamma > 0$ such that

$$\mathbb{P}\,\left[\Delta_{\max} \leq C_\gamma\sqrt{d}\right] \;\geq\; 1 - \gamma \tag{18}$$

holds for all large enough d.

Let us define

$$\tilde{D} \;:=\; \bigcup_{t \in T_r} \tilde{J}(t)$$

and $\tilde{d} := |\tilde{D}| = m(2^r - 2)$. In case of $\Delta_{\max} \leq C_\gamma\sqrt{d}$, we can deduce

$$\tilde{d} \;\geq\; d - \mathrm{o}(d)\ . \tag{19}$$

5.1 The case $n(d) \leq 2^{(1-H_r-\varepsilon)d}$

Let $\tilde{S}_1, \ldots, \tilde{S}_r$ be the canonical projections of $S_1, \ldots, S_r$, respectively, to the coordinates in $\tilde{D}$. Then $\tilde{S}_1, \ldots, \tilde{S}_r$ form a matrix $A(m)$ as defined in Section 4. Denote, for each $i \in [r-1]$,

$$\tilde{L}(i) \;:=\; \bigcup_{t \in T_r \,:\, \mathbb{1}^T t = i} \tilde{J}(t) \;.$$

Due to Proposition 2, there are coefficients $\tilde{a}_1, \ldots, \tilde{a}_{r-1} \in \mathbb{R}$ such that the inequality

$$\sum_{i \in [r-1]} \tilde{a}_i \sum_{j \in \tilde{L}(i)} \xi_j \;\leq\; \sum_{i \in [r-1]} \tilde{a}_i \frac{i}{r} \binom{r}{i} m \;=:\; a_0 \tag{20}$$

has at most $2^{H_r \tilde{d} + \mathrm{o}(\tilde{d})}$ many 0/1-solutions (and $\tilde{S}_1, \ldots, \tilde{S}_r$ satisfy the inequality with equality).

For each $j \in [d]$ let

$$a_j \;:=\; \begin{cases} \tilde{a}_i & \text{if } j \in \tilde{L}(i) \\ 0 & \text{if } j \in [d] \setminus \tilde{D} \end{cases} \;,$$

i.e., $a_1, \ldots, a_d$ are the coefficients of (20) considered as an inequality for $\mathbb{R}^d$.

The inequality

$$\sum_{j \in [d]} a_j \xi_j \;\leq\; a_0 \tag{21}$$

is satisfied with equality by $S_1, \ldots, S_r$.

Let us, for the moment, restrict our attention to the event $\Delta_{\max} \leq C_\gamma \sqrt{d}$. Then (21) has at most

$$2^{H_r \tilde{d} + \mathrm{o}(\tilde{d})} 2^{d - \tilde{d}} \;=\; 2^{H_r d + \mathrm{o}(d)}$$

solutions (due to (19).

Define the halfspace

$$U \;:=\; \big\{ (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d \,:\, \sum_{j \in [d]} a_j \xi_j \leq a_0 \big\} \;,$$

and let ∂U be its bounding hyperplane. Thus, we have

$$S_1, \ldots, S_r \;\in\; \partial U \qquad \text{and} \qquad \big| U \cap \mathrm{V}_d \big| \;\leq\; 2^{H_r d + \mathrm{o}(d)} \;. \tag{22}$$

Since $n(d) \leq 2^{(1-H_r-\varepsilon)d}$, the expected number of points from X lying in U is

$$\frac{2^{H_r d + \mathrm{o}(d)}}{2^d} (n(d) - r) \;\leq\; 2^{-\varepsilon d + \mathrm{o}(d)} \;.$$

Therefore, by Markov's inequality,

$$\mathbb{P}\left[X \cap U = \varnothing \,|\, \Delta_{\max} \leq C_\gamma \sqrt{d} \right] \;= \mathrm{o}(1) \tag{23}$$

From (23) and (18) we derive

$$\mathbb{P}\left[\partial U \cap P = \text{conv } S, X \cap U = \varnothing\right] \;=\; 1 - \text{o}(1) \;,$$

which proves the first part of Proposition 1.

5.2 The case $\mathbf{n(d) \geq 2^{(1-H_r+\varepsilon)d}}$

From the remarks in the introduction, we know

$$\mathbb{P}\left[|S| = r\right] \;=\; 1 - \text{o}(1) \;. \tag{24}$$

Let $\gamma > 0$ be fixed, and assume $|S| = r$, i.e., the points $S_1, \ldots, S_r$ are pairwise disjoint. Denote by $b(S) = (\beta_1, \ldots, \beta_d)$ the barycenter of S. For each $t \in T_r$ and $j \in \tilde{J}(t)$, we have

$$\beta_j \;=\; \frac{\mathbb{1}^T t}{r} \;.$$

If $\Delta_{\max} \leq C_\gamma \sqrt{d}$ holds, we thus have (where the last equation is due to (19))

$$\begin{aligned} H(b(S)) = \;& \tfrac{1}{d}\big(\textstyle\sum_{t \in T_r} mh\big(\frac{\mathbb{1}^T t}{r}\big) + \text{o}(d)\big) \\ =\; & \tfrac{1}{d}\big(\textstyle\sum_{i \in [r-1]} m\binom{r}{i} h(i/r) + \text{o}(d)\big) \\ =\; & \tfrac{m(2^r-2)}{d} H_r + \text{o}(1) \\ =\; & (1 - \text{o}(1)) H_r + \text{o}(1) \;. \end{aligned}$$

Hence, in this case

$$H(b(S)) \;\geq\; H_r - \tfrac{\varepsilon}{2}$$

holds for large enough d. Since H is continuous, there is a neighborhood N of $b(S)$ such that $H(x) \geq H_r - \varepsilon$ holds for all $x \in N$. Due to $n(d) \geq 2^{(1-H_r+\varepsilon)d}$, Corollary 1 implies

$$\mathbb{P}\;\left[N \subseteq \text{conv } X \;\middle|\; |S| = r, \Delta_{\max} \leq C_\gamma \sqrt{d}\right] \;\geq\; 1 - \text{o}(1) \;.$$

Together with (24) and (18), this shows

$$\mathbb{P}\left[b(S) \in \text{int } P\right] \;=\; 1 - \text{o}(1) \;,$$

which proves the second part of Proposition 1.

References

1. Kaibel, V., Remshagen, A.: On the graph-density of random 0/1-polytopes. In Arora, S., Jansen, K., Rolim, J., Sahai, A., eds.: Approximation, Randomization, and Combinatorial Optimization (Proc. RANDOM03). Volume 2764 of Lecture Notes in Computer Science., Springer (2003) 318–328

2. Ziegler, G.M.: Lectures on 0/1-polytopes. In: Polytopes—Combinatorics and Computation (Oberwolfach, 1997). Volume 29 of DMV Sem. Birkhäuser, Basel (2000) 1–41
3. Naddef, D.: The Hirsch conjecture is true for $(0, 1)$-polytopes. Math. Programming **45** (1989) 109–110
4. Bárány, I., Pór, A.: On 0-1 polytopes with many facets. Adv. Math. **161** (2001) 209–228
5. Dyer, M.E., Füredi, Z., McDiarmid, C.: Volumes spanned by random points in the hypercube. Random Structures Algorithms **3** (1992) 91–106
6. Ziegler, G.M.: Lectures on Polytopes. Volume 152 of Graduate Texts in Mathematics. Springer-Verlag, New York (1995) 2nd edition: 1998.
7. Barahona, F., Mahjoub, A.R.: On the cut polytope. Math. Programming **36** (1986) 157–173
8. Déza, M.M., Laurent, M.: Geometry of Cuts and Metrics. Volume 15 of Algorithms and Combinatorics. Springer-Verlag, Berlin (1997)
9. Kahn, J., Komlós, J., Szemerédi, E.: On the probability that a random ± 1-matrix is singular. J. Amer. Math. Soc. **8** (1995) 223–240
10. Feller, W.: An introduction to probability theory and its applications. Vol. I. Third edition. John Wiley & Sons Inc., New York (1968)
11. Graham, R.L., Knuth, D.E., Patashnik, O.: Concrete mathematics. Second edn. Addison-Wesley Publishing Company, Reading, MA (1994)