

The Bernoulli structure of discrete distributions

Original

The Bernoulli structure of discrete distributions / Fontana, R.; Semeraro, P.. - In: ELECTRONIC COMMUNICATIONS IN PROBABILITY. - ISSN 1083-589X. - 30:(2025), pp. 1-13. [10.1214/25-ECP741]

Availability:

This version is available at: 11583/3006646 since: 2026-01-16T11:21:42Z

Publisher:

Institute of Mathematical Statistics - IMS

Published

DOI:10.1214/25-ECP741

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

The Bernoulli structure of discrete distributions

Roberto Fontana* Patrizia Semeraro†

Abstract

Any discrete distribution with support on $\{0, \dots, d\}$ can be constructed as the distribution of sums of Bernoulli variables. We prove that the class of d -dimensional Bernoulli variables $\mathbf{X} = (X_1, \dots, X_d)$ whose sums $\sum_{i=1}^d X_i$ have the same distribution p is a convex polytope $\mathcal{P}(p)$ and we analytically find its extremal points. Our main result is to prove that the Hausdorff measure of the polytopes $\mathcal{P}(p), p \in \mathcal{D}_d$, is a continuous function $l(p)$ over \mathcal{D}_d and it is the density of a finite measure μ_s on \mathcal{D}_d that is Hausdorff absolutely continuous. We also prove that the measure μ_s normalized over the simplex \mathcal{D}_d belongs to the class of Dirichlet distributions. We observe that the symmetric binomial distribution is the mean of the Dirichlet distribution on \mathcal{D}_d and that when d increases it converges to the mode.

Keywords: multidimensional Bernoulli distribution; Dirichlet distribution; binomial distribution; extremal points; polytope.

MSC2020 subject classifications: 60E05; 62R01; 60A10.

Submitted to ECP on March 26, 2025, final version accepted on November 14, 2025.

1 Introduction

Sums of Bernoulli random variables model the number of occurrences of some events within d repeated trials. The case of d independent and identically distributed Bernoulli variables, the sum of which follows the binomial distribution, is often used in modeling across different areas, such as reliability (e.g. [14]) and finance (e.g. [6]). However, the binomial distribution also arises from sums of dependent Bernoulli variables in many ways ([17]), making it a possible model even when independence cannot be assumed, [16]. Actually, any discrete distribution with support on $\{0, \dots, d\}$ can be constructed as the distribution of sums of Bernoulli variables in many ways ([6]). Formally, let $\mathcal{D}_d \subset \mathbb{R}^{d+1}$ be the d -simplex of discrete probability mass functions on $\{0, \dots, d\}$ and $\mathcal{F}_d \subset \mathbb{R}^{2^d}$ be the $2^d - 1$ -simplex of d -dimensional Bernoulli probability mass functions. For any $p \in \mathcal{D}_d$, we define the class $\mathcal{P}(p)$ of probability mass functions $\mathbf{f} \in \mathcal{F}_d$ such that if $\mathbf{X} = (X_1, \dots, X_d)$ has probability mass function \mathbf{f} then $\sum_{i=1}^d X_i$ has probability mass function p .

In [2], the author proves that, as the dimension d increases, the normalized Hausdorff measure of Bernoulli sums with distribution close to the symmetric binomial distribution $Bin(1/2, d)$ converges to one. This means that the Bernoulli sums far from the binomial distributions are rare. It can be asked whether if this is related to the Hausdorff measure of $\mathcal{P}(b(1/2))$, where $b(1/2)$ is the probability mass function of $Bin(1/2, d)$ compared to

*Politecnico di Torino, Italy. E-mail: roberto.fontana@polito.it

†Politecnico di Torino, Italy. E-mail: patrizia.semeraro@polito.it

the Hausdorff measure of $\mathcal{P}(p)$ for any other $p \in \mathcal{D}_d$. Inspired by this question we characterize the class $\mathcal{P}(p)$ for any $p \in \mathcal{D}_d$. We prove that for any $p \in \mathcal{D}_d$, $\mathcal{P}(p) \subset \mathbb{R}^{2^d-d-1}$ is a convex polytope and we analytically find its extremal points. The geometrical structure of $\mathcal{P}(p)$ allows us to analytically study many statistical properties of dependent Bernoulli trials with a given distribution of their sum.

Our main result goes a step further. We prove that the Hausdorff measures of the polytopes $\mathcal{P}(p), p \in \mathcal{D}_d$ define a continuous function $l(p)$ over \mathcal{D}_d which is the density of an Hausdorff-absolutely continuous, positive, and finite measure μ_s on \mathcal{D}_d . We also prove that the normalized measure μ_s belongs to the class of Dirichlet distributions. The Dirichlet parameters are linked to the dimension of the polytopes $\mathcal{P}(p)$. This means that we have a geometrical interpretation of the Dirichlet distribution for a specific choice of its parameters.

We observe that the symmetric binomial distribution is the mean of the Dirichlet distribution on \mathcal{D}_d and that when d increases it converges to the mode. This answers our question: we have the density $l(p)$ for any p and find that the Binomial probability mass function $b(1/2)$ is close to its maximum value. For any dimension d , given $l(p)$ we can also find the Hausdorff measure \mathcal{H}^d of a neighborhood of $\mathcal{P}(b)$ in \mathcal{D}_d , and therefore we can measure the size of probability mass functions of Bernoulli variables with symmetric binomial sums even for low dimensions when the asymptotic result in [2] does not applies.

We point out that we use the Hausdorff measure that it is the proper analytical tool to measure m -dimensional objects (polytopes and simplices) embedded in \mathbb{R}^{2^d-1} , with $m < 2^d - 1$. We remark that computations remain exactly the same if, for each m -dimensional object, we use the Lebesgue measure on \mathbb{R}^m since the Hausdorff measure and the Lebesgue measure of any m -dimensional subset of \mathbb{R}^m coincide.

This paper is organized as follows. Section 2 introduces the class $\mathcal{P}(p)$, proves that it is a polytope, find its extremal points and studies its properties. Our main result is in Section 3, where we find the Hausdorff measure of $\mathcal{P}(p)$, $p \in \mathcal{D}_d$ and we prove that it defines a density on \mathcal{D}_d . Finally, Section 4 focuses on the Binomial distribution and answers our original question.

2 The convex polytope $\mathcal{P}(p)$

Let $\mathcal{X} = \{0, 1\}^d$, we make the non-restrictive hypothesis that the set \mathcal{X} of 2^d binary vectors is ordered according to the reverse-lexicographical criterion. For example for $d = 3$, $\mathcal{X} = \{000, 100, 010, 110, 001, 101, 011, 111\}$. Let $\mathcal{X}_k = \{\mathbf{x} \in \mathcal{X} : \sum_{i=1}^d x_i = k\}$ be the subset of \mathcal{X} that contains all the $\binom{d}{k}$ binary vectors with k ones and $d - k$ zeros, $k = 0, 1, \dots, d$. We observe that \mathcal{X}_k inherits the order of \mathcal{X} . Let \mathbf{x}_k^j be the j -th element of \mathcal{X}_k , $j = 1, \dots, \binom{d}{k}$. The first element is $\mathbf{x}_k := \mathbf{x}_k^1 = (\underbrace{1, \dots, 1}_k, 0, \dots, 0)$.

Let $\mathbf{X} = (X_1, \dots, X_d)$ be a d -dimensional Bernoulli random variable with probability mass function (pmf) f . We identify f with the column vector which contains the values of f over $\mathcal{X} = \{0, 1\}^d$, by $\mathbf{f} = (f_1, \dots, f_{2^d}) = (f_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}) := (f(\mathbf{x}) : \mathbf{x} \in \mathcal{X})$. Let \mathcal{F}_d the $2^d - 1$ -simplex of d -dimensional pmfs \mathbf{f} of Bernoulli vectors. In this paper we identify random variables with their distributions, therefore the notation $\mathbf{X} \in \mathcal{F}_d$ means that \mathbf{X} has pmf $\mathbf{f} \in \mathcal{F}_d$. Let \mathcal{D}_d be the d -simplex of discrete pmfs on $\{0, \dots, d\}$. The notation $D \in \mathcal{D}_d$ means that D has pmf $p = (p_0, \dots, p_d) \in \mathcal{D}_d$.

Any pmf $p \in \mathcal{D}_d$ is the distribution of the sum of the components of at least one d -dimensional Bernoulli random vector $\mathbf{X} \in \mathcal{F}_d$ (see e.g. [6]). Actually, in general, behind any discrete pmf there are infinite Bernoulli vectors $\mathbf{X} \in \mathcal{F}_d$. Formally, we define the

following map between \mathcal{F}_d and \mathcal{D}_d .

$$\begin{aligned} s : \mathcal{F}_d &\rightarrow \mathcal{D}_d \\ \mathbf{f} &\rightarrow p_{\mathbf{f}}, \end{aligned} \tag{2.1}$$

where $p_{\mathbf{f}} := s(\mathbf{f})$ is the distribution of the sum $S := \sum_{i=1}^d X_i$ and $\mathbf{X} \sim \mathbf{f}$, i.e. \mathbf{X} has pmf $\mathbf{f} \in \mathcal{F}_d$. For any $p \in \mathcal{D}_d$, we define

$$\mathcal{P}(p) = \{\mathbf{f} \in \mathcal{F}_d : p_{\mathbf{f}} = p\}. \tag{2.2}$$

The next Theorem 2.1 proves that for any choice of p , the class $\mathcal{P}(p)$ is a convex polytope and provides an analytical expression for its extremal pmfs.

Given p , we define the n -dimensional simplex

$$\{\mathbf{x} \in \mathbb{R}^{n+1} : x_j \geq 0, j = 0, \dots, n, \sum_{h=0}^n x_h = p\}, \tag{2.3}$$

whose vertices are $e_0 = (p, 0, \dots, 0), e_1 = (0, p, 0, \dots, 0), \dots, e_n = (0, \dots, 0, p)$. The length $s_{i,j}$ of the edge between e_i and e_j , with $i, j = 0, \dots, n, i \neq j$ of the simplex is defined as $s_{i,j} = \sqrt{\sum_{k=1}^n (e_{ik} - e_{jk})^2}$. Since all lengths are equal, $s_{i,j} = \sqrt{2}p$, their common value $s = \sqrt{2}p$ is called the side of the simplex. We denote the n -dimensional simplex with side $\sqrt{2}p$ by $\Delta_{n, \sqrt{2}p}$.

Theorem 2.1. For any $p \in \mathcal{D}_d$ the class $\mathcal{P}(p) = \{\mathbf{f} \in \mathcal{F}_d : p_{\mathbf{f}} = p\}$ is the convex polytope $\mathcal{P}(p) = \prod_{k=0}^d \Delta_{n_k, \sqrt{2}p_k}$, where $n_k = \binom{d}{k} - 1$ and its extremal points are

$$f^{\sigma}(\mathbf{x}) = \begin{cases} p_k & \text{if } \mathbf{x} = \mathbf{x}_k^{\sigma_k} \\ 0 & \text{otherwise,} \end{cases} \tag{2.4}$$

where $\sigma = (\sigma_0, \dots, \sigma_k, \dots, \sigma_d), \sigma_k = 1, \dots, \binom{d}{k}, k = 0, \dots, d$.

Proof. Let $p \in \mathcal{D}_d$ and let $\mathbf{X} \sim \mathbf{f} \in \mathcal{F}_d$. We have $p_{\mathbf{f}} = p$ if and only if \mathbf{f} is a positive solution of the linear system:

$$\sum_{\mathbf{x} \in \mathcal{X}_k} f(\mathbf{x}) = p_k, \quad k = 0, \dots, d. \tag{2.5}$$

Each equation of the system (2.5) defines a $\binom{d}{k} - 1$ -simplex with side $\sqrt{2}p_k$. It is well known - see [10] - that the $\binom{d}{k}$ extremal points of the simplex are $(p_k, 0, \dots, 0), (0, p_k, \dots, 0), \dots, (0, \dots, 0, p_k)$. Since $\mathcal{X}_k \cap \mathcal{X}_j = \emptyset$, for any $k \neq j$, the extremal solutions of the system are the ones in (2.4). \square

Corollary 2.2. The number n_p of extremal points of $\mathcal{P}(p)$ is $n_p = \prod_{k \in \text{Supp}(p)} \binom{d}{k}$, where $\text{Supp}(p) \subseteq \{0, \dots, d\}$ is the support of p .

Proof. The proof follows from Theorem 2.1 since $\#\mathcal{X}_k = \binom{d}{k}$. \square

From Theorem 2.1 and Corollary 2.2 it follows that for any $\mathbf{f} \in \mathcal{P}(p)$ there exist $\lambda_i \geq 0$ summing up to one such that

$$\mathbf{f} = \sum_{i=1}^{n_p} \lambda_i \mathbf{r}_i,$$

where $\mathbf{r}_i \in \mathcal{P}(p), i = 1, \dots, n_p$ are the extremal points in Equation (2.4). We call \mathbf{r}_i extremal points or extremal pmfs of $\mathcal{P}(p)$. We denote with \mathbf{R}_i a d -dimensional random variable with distribution \mathbf{r}_i .

Notice that n_p depends only on the support and not on the values of p . If $\{1, \dots, d-1\} \subseteq \text{Supp}(p)$, since $\binom{d}{0} = \binom{d}{d} = 1$ we have $n_p = \prod_{k=0}^d \binom{d}{k}$.

Example 2.3. As an illustrative example we consider the polytope $\mathcal{P}(p)$ in dimension $d = 3$ for a given $p = (p_0, p_1, p_2, p_3) \in \mathcal{D}_3$ with full support. The extremal points of $\mathcal{P}(p)$ are $n_p = \binom{3}{1} \binom{3}{2} = 9$ and they are reported in Table 1.

Table 1: Extremal pmfs of $\mathcal{P}(p)$, case $d = 3$

\mathbf{x}_1	\mathbf{x}_2	\mathbf{x}_3	\mathbf{r}_1	\mathbf{r}_2	\mathbf{r}_3	\mathbf{r}_4	\mathbf{r}_5	\mathbf{r}_6	\mathbf{r}_7	\mathbf{r}_8	\mathbf{r}_9
0	0	0	p_0	p_0	p_0	p_0	p_0	p_0	p_0	p_0	p_0
1	0	0	p_1	p_1	p_1	0	0	0	0	0	0
0	1	0	0	0	0	p_1	p_1	p_1	0	0	0
1	1	0	p_2	0	0	p_2	0	0	0	0	p_2
0	0	1	0	0	0	0	0	0	p_1	p_1	p_1
1	0	1	0	p_2	0	0	p_2	0	p_2	0	0
0	1	1	0	0	p_2	0	0	p_2	0	p_2	0
1	1	1	p_3	p_3	p_3	p_3	p_3	p_3	p_3	p_3	p_3

Remark 2.4. Theorem 2.1 can be easily generalized to any surjective map $h : \mathcal{X} \rightarrow \{0, \dots, d\}$. Let $p_f^h \in \mathcal{D}_d$ the pmf associated to $h(\mathbf{X})$ with $\mathbf{X} \sim \mathbf{f} \in \mathcal{F}_d$. For any $p \in \mathcal{D}_d$ the class $\mathcal{P}^h(p) = \{\mathbf{f} \in \mathcal{F}_d : p_f^h = p\}$ is a convex polytope and its extremal points are

$$f^\sigma(\mathbf{x}) = \begin{cases} p_y & \text{if } \mathbf{x} = \mathbf{x}_y^{\sigma_y}, \\ 0 & \text{otherwise,} \end{cases}$$

where for any $y \in \{0, \dots, d\}$, $\mathbf{x}_y^{\sigma_y}$ is the σ_y -th element of $h^{-1}(y)$. If $h(\mathbf{x}) = \sum_{i=1}^d x_i$ we have Theorem 2.1.

2.1 Moments and entropy

In [7], the authors prove that the bounds of the moments of a pmf in a convex polytope are sharp and reached on the extremal points, that in this case are explicitly known.

Proposition 2.5. Let \mathbf{X} with pmf $\mathbf{f} \in \mathcal{P}(p)$, then for any $\{j_1, \dots, j_k\} \subseteq \{1, \dots, d\}$,

$$p_d \leq E[X_{j_1} \cdots X_{j_k}] \leq \sum_{h=k}^d p_h,$$

and the bounds are sharp.

Proof. Since $X_{j_1} \cdots X_{j_k} = 1$ is contained in the event $\sum_{i=1}^d X_i \geq k$. It follows

$$E[X_{j_1} \cdots X_{j_k}] = \sum_{\{x \in \mathcal{X} : x_{j_1} \cdots x_{j_k} = 1\}} \mathbf{f}_x \leq \sum_{\{x \in \mathcal{X} : \sum_{i=1}^d x_i \geq k\}} \mathbf{f}_x = \sum_{h=k}^d p_h$$

Similarly, the event $X_1 \cdots X_d = 1$ is contained in the event $X_{j_1} \cdots X_{j_k} = 1$. It follows

$$p_d = f_{\mathbf{x}_d} \leq \sum_{\{x \in \mathcal{X} : x_{j_1} \cdots x_{j_k} = 1\}} \mathbf{f}_x = E[X_{j_1} \cdots X_{j_k}]$$

The bounds are sharp because we can consider the extremal pmf $\tilde{r}(\mathbf{x})$ defined as

$$\tilde{r}(\mathbf{x}) = \begin{cases} p_k & \text{if } \mathbf{x} = \mathbf{x}_k, \\ 0 & \text{otherwise,} \end{cases}$$

and the associated random variable \tilde{R} with pmf \tilde{r} . We have $E[\tilde{R}_1 \cdots \tilde{R}_k] = \tilde{r}(\mathbf{x}_k) + r(\mathbf{x}_{k+1}) + \dots + \tilde{r}(\mathbf{x}_{d-1}) + \tilde{r}(\mathbf{x}_d) = \sum_{h=k}^d p_h$ and $E[\tilde{R}_{d-k+1} \cdots \tilde{R}_d] = \tilde{r}(\mathbf{x}_d) = p_d$. \square

We now consider the definition of the Shannon entropy for a discrete random variable W pmf $p_w := P[W = w]$ as $H(W) = -\sum_{w \in \mathcal{W}} p_w \log p_w$ with the convention $0 \log(0) = 0$. It is easy to verify that, given a random variable $W \sim p \in \mathcal{D}_d$, the extremal random variables $\mathbf{R}_1, \dots, \mathbf{R}_{n_p}$ of the polytope $\mathcal{P}(p)$ have all the same entropy, which is equal to the entropy of W , $H(\mathbf{R}_i) = H(W), i = 1, \dots, n_p$. Moreover, we can identify the multivariate Bernoulli variables whose distributions lie within the polytope $\mathcal{P}(p)$, and which have the maximum and minimum entropy.

Proposition 2.6. *Given pmf $p \in \mathcal{D}_d$, let \mathbf{X}_M be a multivariate Bernoulli random variable with the exchangeable pmf $\mathbf{f}_M \in \mathcal{P}(p)$, whose expression is:*

$$\mathbf{f}_M(\mathbf{x}) = \begin{cases} \frac{p_k}{\binom{d}{k}} & \text{if } \mathbf{x} \in \mathcal{X}_k, k = 0, \dots, d \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

Then the following hold:

1. $\mathbf{X}_M = \operatorname{argmax}_{\mathbf{X} \in \mathcal{P}(p)} H(\mathbf{X})$
2. $\mathbf{R}_i = \operatorname{argmin}_{\mathbf{X} \in \mathcal{P}(p)} H(\mathbf{X}), i = 1, \dots, n_p,$

where $\mathbf{R}_i \sim r_i$ and r_i are the extremal pmfs of $\mathcal{P}(p)$, $i = 1, \dots, n_p$.

Proof. Both statements can be proved by noting that for any $\mathbf{X} \in \mathcal{F}_d$ we can express its entropy as: $H(\mathbf{X}) = -\sum_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x}) \log(\mathbf{f}(\mathbf{x})) = -\sum_{k=0}^d \sum_{\mathbf{x} \in \mathcal{X}_k} \mathbf{f}(\mathbf{x}) \log(\mathbf{f}(\mathbf{x}))$.

To maximize (minimize) $H(\mathbf{X})$, it is sufficient to maximize (minimize) each term $-\sum_{\mathbf{x} \in \mathcal{X}_k} \mathbf{f}(\mathbf{x}) \log(\mathbf{f}(\mathbf{x}))$, which represents the entropy restricted to $\mathcal{X}_k, k = 0, \dots, d$. It is well known that entropy is maximized by choosing a uniform distribution see, e.g., [11], and then we get \mathbf{f}_M in (2.6). The entropy is minimized by choosing a Dirac delta distribution centered at any point in $\mathcal{X}_k, k = 0, \dots, d$. \square

3 The induced measure on \mathcal{D}_d

The next Theorem 3.2 proves our main result that Bernoulli sums induce a Dirichlet distribution on the simplex \mathcal{D}_d . We need some preliminaries. Since $\mathcal{F}_d \subset \mathbb{R}^{2^d}$ is the standard 2^{d-1} -simplex, $\mathcal{D}_d \subset \mathbb{R}^d$ is the standard d -simplex and $\mathcal{P}(p) \subset \mathbb{R}^{2^d}$ is a $(2^d - d - 1)$ -convex polytope we consider the Hausdorff measure \mathcal{H}^n for any $n \in \{0, \dots, 2^{d-1}\}$. We recall that $\mathcal{H}^0(x) = 1$, for any $x \in \mathbb{R}^m, m \in \mathbb{N}$ (a standard reference for Hausdorff measures is [4]). The Corollary 3.1 finds the Hausdorff measure of $\mathcal{P}(p)$ in $\mathbb{R}^{2^d - d - 1}$ for any pmf $p \in \mathcal{D}_d$. It is well known (see, e.g., [10]) that the Hausdorff measure of the n -simplex with side $\sqrt{2}p, \Delta_{n, \sqrt{2}p} \subset \mathbb{R}^{n+1}$ is

$$\mathcal{H}^n(\Delta_{n, \sqrt{2}p}) = \frac{(\sqrt{2}p)^n \sqrt{n+1}}{n! \sqrt{2^n}} = \frac{p^n \sqrt{n+1}}{n!}. \tag{3.1}$$

We have the following corollary of Theorem 2.1.

Corollary 3.1. *For any $p = (p_0, \dots, p_d) \in \mathcal{D}_d$, it holds*

$$\mathcal{H}^{2^d - d - 1}(\mathcal{P}(p)) = \prod_{k=0}^d \mathcal{H}^{n_k}(\Delta_{n_k, \sqrt{2}p_k}),$$

where $n_k = \binom{d}{k} - 1$.

We can now prove our main result.

Theorem 3.2. Let μ_s the measure on $(\mathcal{D}_d, \mathcal{B}(\mathcal{D}_d))$, where $\mathcal{B}(\mathcal{D}_d)$ is the Borel σ -algebra on \mathcal{D}_d , induced from the function s in Equation (2.1). It holds

$$\mu_s(A) = \mathcal{H}^{2^d-1}(s^{-1}(A)) = \int_A \prod_{k=0}^d \frac{p_k^{n_k}}{n_k!} d\mathcal{H}^d(p), \quad A \in \mathcal{B}(\mathcal{D}_d), \quad (3.2)$$

where $n_k = \binom{d}{k} - 1$. Then μ_s is a positive finite measure on \mathcal{D}_d such that $\mu_s(\mathcal{D}_d) = \mathcal{H}^{2^d-1}(\mathcal{F}_d)$. The measure μ_s is absolutely continuous with respect the Hausdorff measure on \mathcal{D}_d .

Proof. We start proving the following

$$\mathcal{H}^{2^d-1}(\mathcal{F}_d) = \int_{\mathcal{D}_d} \prod_{k=0}^d \frac{1}{\sqrt{n_k+1}} \mathcal{H}^{n_k}(\Delta_{n_k, \sqrt{2}p_k}) d\mathcal{H}^d(p), \quad (3.3)$$

We explicitly build an isometry α^{ort} on \mathcal{F}_d by orthonormalizing the following transformation α

$$\begin{aligned} \alpha : \mathcal{F}_d &\rightarrow \mathcal{F}_d \\ \mathbf{f} &\rightarrow (p_k^j), \end{aligned} \quad (3.4)$$

where $k = 0, \dots, d, j = 1, \dots, \binom{d}{k}$ and

$$p_k^j = \begin{cases} \sum_{\mathbf{x} \in \mathcal{X}_k} f(\mathbf{x}), & j = 1, \\ f(\mathbf{x}_k^j), & j \neq 1. \end{cases}$$

Since the resulting isometry does not modify p_k^1 , we write $p_k = p_k^1, k = 0, \dots, d$.

Let $I_{2^d} = (i_{\mathbf{x},j})_{\mathbf{x} \in \mathcal{X}, j \in \{1, \dots, 2^d\}}$ be the 2^d - identity matrix and let $\mathbf{i}_{\mathbf{x}}$ be the row vector $\mathbf{i}_{\mathbf{x}} = (i_{\mathbf{x},j}, j = 1, \dots, 2^d)$. Let \mathbf{x}_k be the first element in the reverse lexicographic order of \mathcal{X}_k , and $\mathbf{a}_{\mathbf{x}_k} := \frac{1}{\sqrt{n_k+1}}(\mathbf{1}_{\mathcal{X}_k}(\mathbf{x}), \mathbf{x} \in \mathcal{X}), k = 1, \dots, d$, where $\mathbf{1}_B()$ is the indicator function of B .

Let $A := I_{2^d}(\mathbf{i}_{\mathbf{x}_k} \rightarrow \mathbf{a}_{\mathbf{x}_k})$ be the matrix obtained from I_{2^d} by replacing the row $\mathbf{i}_{\mathbf{x}_k}$ with the row $\mathbf{a}_{\mathbf{x}_k}, k = 1, \dots, d$. Let A^{ort} be the matrix obtained from A by the Gram-Schmidt orthonormalization process applied to the rows of A , considered from the first to the last. It holds $\mathbf{a}_{\mathbf{x}_k}^{ort} = \mathbf{a}_{\mathbf{x}_k}$. In fact, since \mathbf{x}_k is the first element in the reverse lexicographic order of \mathcal{X}_k , all the preceding rows $\mathbf{a}_{\mathbf{x}_h}$ of $A, \mathbf{x}_h < \mathbf{x}_k$, refer to \mathcal{X}_h with $\mathcal{X}_h \cap \mathcal{X}_k = \emptyset$. It follows that $\langle \mathbf{a}_{\mathbf{x}_k}, \mathbf{a}_{\mathbf{x}_h} \rangle = 0$, for any row $\mathbf{a}_{\mathbf{x}_h}$ with $\mathbf{x}_h < \mathbf{x}_k$. We also observe that if \mathbf{x}_k^2 is the second element of \mathcal{X}_k we still have $\langle \mathbf{a}_{\mathbf{x}_k^2}, \mathbf{a}_{\mathbf{x}_h} \rangle = 0$, for any row $\mathbf{a}_{\mathbf{x}_h}$ with $\mathbf{x}_h < \mathbf{x}_k$ and $\mathbf{x}_h \notin \mathcal{X}_k$. On the other hand, the product $\langle \mathbf{a}_{\mathbf{x}_k^2}, \mathbf{a}_{\mathbf{x}_k} \rangle$ is different from zero. For our purposes, it is not necessary to make explicit the result of the orthonormalization, $\mathbf{a}_{\mathbf{x}_k}^{ort}$ but it is enough to notice that the orthonormalization process of $\mathbf{a}_{\mathbf{x}_k^2}$ will produce a vector $\mathbf{a}_{\mathbf{x}_k^2}^{ort}$ with zeros in all the positions $\mathbf{x}_j \notin \mathcal{X}_k$. A similar argument holds for the subsequent rows $\mathbf{a}_{\mathbf{x}_k^j}$ of $A, j = 3, \dots, n_k + 1$. We show the matrices \mathcal{X}, A , and A^{ort} for $d = 2$ and $d = 3$. in Remark 3.3. Since A^{ort} is an orthonormal matrix the application

$$\begin{aligned} \alpha^{ort} : \mathcal{F}_d &\rightarrow \mathcal{F}_d \\ \mathbf{f} &\rightarrow \mathbf{f}^{ort} = A^{ort} \mathbf{f}, \end{aligned}$$

is an isometry. Then it holds

$$\begin{aligned} \mathcal{H}^{2^d-1}(\mathcal{F}_d) &= \int_{\mathcal{F}_d} d\mathcal{H}^{2^d-1}(\mathbf{f}) = \int_{\Delta_{2^d, \sqrt{2}}} d\mathcal{H}^{2^d-1}(\mathbf{f}) = \\ &= \int_{\Delta_{d, \sqrt{2}}} \prod_{k=0}^d \frac{1}{\sqrt{n_k+1}} \int_{\Delta_{n_k, \sqrt{2}p_k}} d\mathcal{H}^{n_k}(\mathbf{f}_{\mathcal{X}_k}^{ort}) d\mathcal{H}^d(p) \\ &= \int_{\mathcal{D}_d} \prod_{k=0}^d \frac{1}{\sqrt{n_k+1}} \mathcal{H}^{n_k}(\Delta_{n_k, \sqrt{2}p_k}) d\mathcal{H}^d(p), \end{aligned} \tag{3.5}$$

where $\mathbf{f}_{\mathcal{X}_k}^{ort} = (A^{ort} \mathbf{f}, \mathbf{f} \in \mathcal{X}_k)$, that is (3.3). Plugging Equation (3.1) in (3.5) we have

$$\mathcal{H}^{2^d-1}(\mathcal{F}_d) = \int_{\mathcal{D}_d} \prod_{k=0}^d \frac{p_k^{n_k}}{n_k!} d\mathcal{H}^d(p).$$

Thus, for any $A \in \mathcal{B}(\mathcal{D}_d)$,

$$\mu_s(A) = \mathcal{H}^{2^d-1}(s^{-1}(A)) = \int_{s^{-1}(A)} d\mathcal{H}^{2^d-1}(\mathbf{f}) = \int_A \prod_{k=0}^d \frac{p_k^{n_k}}{n_k!} d\mathcal{H}^d(p),$$

Let $l : \mathcal{D}_d \rightarrow \mathbb{R}^+$ defined by $l(p) = \prod_{k=0}^d \frac{p_k^{n_k}}{n_k!}$. The function is almost surely continuous on \mathcal{D}_d . Therefore, the measure μ_s on \mathcal{D}_d defined by Equation (2.1) is a positive, finite and Hausdorff absolutely continuous measure on \mathcal{D}_d . By construction $\mu_s(\mathcal{D}_d) = \mathcal{H}^{2^d-1}(\mathcal{F}_d)$. \square

Remark 3.3. For illustration purposes, we show the matrices \mathcal{X} , A , and A^{ort} introduced in the proof of Theorem 3.2 for $d = 2$ and $d = 3$. For $d = 2$, we have

$$\mathcal{X} = \begin{pmatrix} 00 \\ 10 \\ 01 \\ 11 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad A^{ort} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

and for $d = 3$, we have

$$\begin{aligned} \mathcal{X} &= \begin{pmatrix} 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{pmatrix} \quad A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\ A^{ort} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{6}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\sqrt{6}} & 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Remark 3.4. We observe that Theorem 3.2 can be generalized to any surjective map h as discussed in Remark 2.4.

The function $l : \mathcal{D}_d \rightarrow \mathbb{R}^+$ defined by

$$l(p) = \prod_{k=0}^d \frac{p_k^{n_k}}{n_k!}$$

is the density of μ_s with respect the Hausdorff measure \mathcal{H}^d on \mathcal{D}_d . The following Corollary 3.5 provides a useful formula for practical computations.

Corollary 3.5. *It holds*

$$\mathcal{H}^{2^d-1}(\mathcal{F}_d) = \sqrt{2^d} \int_{\Sigma_d} \prod_{k=0}^{d-1} \frac{p_k^{n_k}}{n_k!} dp_0 \dots dp_{d-1}, \tag{3.6}$$

where $\Sigma_d = \{\mathbf{x} \in \mathbb{R}^d : x_j \geq 0, j = 0, \dots, d-1, \sum_{k=0}^{d-1} x_k \leq 1\}$, and $n_k = \binom{d}{k} - 1$.

Remark 3.6. Theorem 3.2 provides a link between the Hausdorff measure of the polytopes $\mathcal{P}(p)$ and the Hausdorff measure of \mathcal{F}_d . It holds $\mathcal{F}_d = \cup_{p \in \mathcal{D}_d} \mathcal{P}(p)$, and

$$\mathcal{H}^{2^d-1}(\mathcal{F}_d) = N \int_{\mathcal{D}_d} \mathcal{H}^{2^d-d-1}(\mathcal{P}(p)) d\mathcal{H}^d(p),$$

where $N = \prod_{k=0}^d \frac{1}{\sqrt{n_k+1}}$.

The following proposition states that the Hausdorff measure of $\mathcal{P}(p)$, $p \in \mathcal{D}_d$, induces the Dirichlet distribution on the simplex of discrete distributions \mathcal{D}_d (see [13] for an overview on the Dirichlet distribution). Let $Dirichlet(\alpha_0, \dots, \alpha_d)$ be the Dirichlet distributions with parameters $\alpha_0, \dots, \alpha_d$.

Proposition 3.7. *The density $l(p)$ normalized over the simplex \mathcal{D}_d is the Dirichlet density with parameters $\alpha_k = \binom{d}{k}$, $k = 0, \dots, d$, on the d -simplex \mathcal{D}_d .*

Proof. It is sufficient to observe $l(p) \sim \prod_{k=0}^d p_k^{n_k}$ and $n_k = \binom{d}{k} - 1$ with $n_0 = n_d = 0$. \square

Proposition 3.7 gives a geometrical interpretation for the parameters of this Dirichlet distribution as the Hausdorff dimensions of the simplexes corresponding to each p_j , $j = 0, \dots, d$.

Remark 3.8. Notice that $\mathcal{H}^{2^d-1}(\mathcal{F}_d) = \frac{\sqrt{2^d}}{(2^d-1)!}$ is the normalizing constant for $l(p)$ to be the Dirichlet density.

The size of the class of multivariate Bernoulli distributions the sums of which have pmf close to a given $p \in \mathcal{D}_d$ depends on the behavior of $l(p)$ in a neighborhood of p . The next Corollary 3.9 explicitly provides the pmf $p^M \in \mathcal{D}_d$ that maximizes the Hausdorff measure $\mathcal{H}^{2^d-d-1}(\mathcal{P}(p))$.

Corollary 3.9. *Let $p^M = (p_0^M, \dots, p_d^M) \in \mathcal{D}$ be such that*

$$p_k^M = \frac{\binom{d}{k} - 1}{2^d - d - 1}, \quad k = 0, \dots, d$$

then $p^M = \operatorname{argmax}_{p \in \mathcal{D}_d} \mathcal{H}^{2^d-d-1}(\mathcal{P}(p))$.

Proof. The proof follows directly from Proposition 3.7 observing that p^M is the mode of the Dirichlet distribution. \square

We name p^M the maximal pmf in \mathcal{D}_d .

3.1 Measuring a neighborhood of $\mathcal{P}(p)$

This section finds the Hausdorff measure of the Bernoulli sums whose pmf is close to a given $p \in \mathcal{D}_d$ according to a given metrics d . In practice, using the measure μ_s we measure a neighborhood of a pmf $p \in \mathcal{D}_d$. This means finding the Hausdorff measure in \mathcal{F}_d of the set of multivariate Bernoulli distributions f such that p_f is close to p .

Formally, let d be a distance on \mathcal{D}_d and define a neighborhood of p in \mathcal{D}_d by

$$I_d(p, \epsilon) = \{\tilde{p} \in \mathcal{D}_d : d(\tilde{p}, p) \leq \epsilon\}, \quad \epsilon > 0,$$

and a corresponding neighborhood of $\mathcal{P}(p)$ in \mathcal{F}_d as its counter-image through the map s is (2.1)

$$I_d^{\mathcal{F}}(p, \epsilon) = s^{-1}(I_d(p, \epsilon)) = \{\tilde{f} \in \mathcal{F} : d(\tilde{p}, p) \leq \epsilon\}, \quad \epsilon > 0,$$

where $\tilde{p} = p_{\tilde{f}}$. Using Equation 3.2, the Hausdorff measure in \mathcal{F}_d of $I_d^{\mathcal{F}}(p, \epsilon)$ is the measure μ_s of $I_d(p, \epsilon)$, and this can be found by integration of $l(p)$ over $I_d(p, \epsilon)$. Following [2] we consider the maximum distance on \mathcal{D}_d and show as to estimate $I_S(p, \epsilon)$. Given two probability measures P and Q the maximum distance d_S defined by

$$d_S(\tilde{p}, p) := \max_{0 \leq k \leq d} |\tilde{p}_k - p_k|.$$

By definition of $I_S(p, \epsilon)$ it holds

$$I_S(p, \epsilon) = \{\mathbf{x} \in \mathbb{R}^{d+1} : \sum_{i=0}^d x_i = 1, \max\{p_j - \epsilon, 0\} \leq x_j \leq \min\{p_j + \epsilon, 1\}, j = 0, \dots, d\},$$

therefore $I_S(p, \epsilon) \subseteq \mathcal{D}_d$ is a convex polytope. From Corollary 3.5 it follows that

$$\mathcal{H}^{2^d-1}(I_S^{\mathcal{F}}(p, \epsilon)) = \mu_s(I_S(p, \epsilon)) = \sqrt{2^d} \int_{\Sigma_S(p, \epsilon)} \prod_{j=0}^d \frac{p_j^{n_j}}{n_j!} dp_0 \dots dp_{d-1}, \quad (3.7)$$

where

$$\Sigma_S(p, \epsilon) = \{\mathbf{x} \in \mathbb{R}^d : \sum_{i=0}^{d-1} x_i \leq 1, \max\{p_j - \epsilon, 0\} \leq x_j \leq \min\{p_j + \epsilon, 1\}, j = 0, \dots, d-1\}.$$

To compute $\mu_s(I_S(p, \epsilon))$ we can find an estimate $\hat{\mu}_s(I_S(p, \epsilon))$ of $\mu_s(I_S(p, \epsilon))$ by using (3.7)

$$\hat{\mu}_s^d(I_S(p, \epsilon)) = \hat{\mathcal{H}}^{2^d-1}(I_S^{\mathcal{F}}(p, \epsilon)),$$

where $\hat{\mathcal{H}}^{2^d-1}(I_S^{\mathcal{F}}(p, \epsilon))$ is an estimate of $\mathcal{H}^{2^d-1}(I_S^{\mathcal{F}}(p, \epsilon))$ computed as

$$\hat{\mathcal{H}}^{2^d-1}(I_S^{\mathcal{F}}(p, \epsilon)) = \hat{E}_U[\prod_{j=0}^d \mathcal{H}^{n_j}(\Delta_{p_j})] \mathcal{H}^d(I_S(p, \epsilon)) = \frac{\sum_{j=1}^N \prod_{j=0}^d \mathcal{H}^{n_j}(\Delta_{\hat{p}_j})}{N} \mathcal{H}^d(I_S(p, \epsilon)),$$

where $\hat{p}_j, j = 1, \dots, N$ are uniformly extracted from $I_S(p, \epsilon)$, the expectation E_U is relative to a uniform distribution on the simplex, and $\hat{\mathcal{H}}(I_S(p, \epsilon))$ is computed using package `volesti` [5] which uses a random-walk-based method to provide uniform samples from a given convex polytope.

4 The binomial distribution

This section focuses on the Bernoulli structure behind the discrete distribution corresponding to the most important independence model: the binomial distribution.

Let $b(\theta) \in \mathcal{D}_d$ be the pmf of the binomial distribution with parameters θ and d ($B(\theta, d)$) and let $\mathcal{P}(b(\theta)) = \{\mathbf{f} \in \mathcal{F} : p_{\mathbf{f}} = b(\theta)\}$. From Theorem 2.1 its extremal points are

$$f_B^\sigma(\mathbf{x}) = \begin{cases} \binom{d}{k} \theta^k (1-\theta)^{d-k} & \text{if } \mathbf{x} = \mathbf{x}_k^{\sigma_k}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\sigma = (\sigma_0, \dots, \sigma_k, \dots, \sigma_d)$, $\sigma_k = 1, \dots, \binom{d}{k}$, $k = 0, \dots, d$. Since the binomial distributions have full support on $\{0, \dots, d\}$ from Corollary 2.2 the number of extremal points is $n_b := n_p = \prod_{k=0}^d \binom{d}{k}$.

The class of binomial distributions describes a parametrical curve on the simplex \mathcal{D}_d , given by $b(\theta) = (b_0(\theta), \dots, b_d(\theta))$, $\theta \in [0, 1]$. The following Proposition proves that the density l restricted to the binomial class is a concave function in the parameter space $[0, 1]$ and maximal for $\theta = 1/2$, i.e. $l(\theta) = \mathcal{H}^{2^d-d-1}(\mathcal{P}(b(\theta)))$ is maximal for $\theta = 1/2$.

Proposition 4.1. *The map*

$$\begin{aligned} l : [0, 1] &\rightarrow \mathbb{R}_+ \\ \theta &\rightarrow \mathcal{H}^{2^d-d-1}(\mathcal{P}(b(\theta))), \end{aligned} \tag{4.1}$$

is a concave function in θ and

$$\operatorname{argmax}_{\theta \in [0,1]} \mathcal{H}^{2^d-d-1}(\mathcal{P}(b(\theta))) = \frac{1}{2}.$$

Proof. We have

$$\mathcal{H}^{2^d-d-1}(\mathcal{P}(b(\theta))) = \prod_{k=0}^d \frac{\binom{d}{k} \theta^k (1-\theta)^{d-k} n_k \sqrt{n_k + 1}}{n_k!},$$

thus

$$\begin{aligned} \log(\mathcal{H}^{2^d-d-1}(\mathcal{P}(b(\theta)))) &= \log \prod_{k=0}^d \frac{\binom{d}{k} \theta^k (1-\theta)^{d-k} n_k \sqrt{n_k + 1}}{n_k!} \\ &= \log \prod_{k=0}^d \frac{\binom{d}{k} n_k \sqrt{n_k + 1}}{n_k!} + \log \prod_{k=0}^d (\theta^k (1-\theta)^{d-k})^{n_k}. \end{aligned}$$

It is sufficient to find the maximum of $f(\theta) = \log \prod_{k=0}^d (\theta^k (1-\theta)^{d-k})^{n_k}$. Straightforward computations lead to

$$f'(\theta) = \sum_{k=0}^d \frac{n_k}{\theta(1-\theta)} (k - d\theta) = \sum_{k=0}^{d\theta^-} \frac{n_k}{\theta(1-\theta)} (k - d\theta) + \sum_{k=d\theta^+}^d \frac{n_k}{\theta(1-\theta)} (k - d\theta),$$

where $d\theta^-$ is the largest integer smaller than $d\theta$ and $d\theta^+$ is the smallest integer bigger than $d\theta$. $f'(\theta) = 0$ iff

$$\sum_{k=0}^{d\theta^-} \frac{n_k}{\theta(1-\theta)} (k - d\theta) = \sum_{k=d\theta^+}^d \frac{n_k}{\theta(1-\theta)} (k - d\theta)$$

and, since $n_k = n_{d-k}$ this is true iff $\theta = 1/2$. If $\theta > 1/2$ we have $f'(\theta) < 0$ and if $\theta < 1/2$ we have $f'(\theta) > 0$, and the maximum is reached on $\theta = 1/2$. \square

Remark 4.2. Notice that the symmetric binomial distribution is the mean of the Dirichlet distribution $Dirichlet(n_0 + 1, \dots, n_d + 1)$ on the simplex \mathcal{D}_d .

The following proposition proves that if the dimension d increases, the pmf $b(1/2)$ converges to the distribution p^M .

Proposition 4.3. *Let $b(1/2)$ be the pmf of the $B(1/2, d)$ and p^M is the maximal polytope pmf in \mathcal{F}_d . We have*

$$\lim_{d \rightarrow \infty} d_S(b(1/2), p^M) = 0, \quad \text{and} \quad \lim_{d \rightarrow \infty} \mathcal{H}^{2^d - d - 1} \mathcal{P}(b(1/2)) = l^M,$$

where $l^M = \mathcal{H}^{2^d - d - 1}(\mathcal{P}(p^M))$.

Proof. We have

$$|b(1/2)_k - p_k^M| = \frac{|2^d \binom{d}{k} - 1 - (2^d - d - 1) \binom{d}{k}|}{2^d(2^d - d - 1)} = \frac{|(d + 1) \binom{d}{k} - 2^d|}{2^d(2^d - d - 1)}.$$

Since $\binom{d}{k}$ is maximal for $k = \frac{d-1}{2}$ and $k = \frac{d+1}{2}$ if d is odd, we have

$$\max_k |b(1/2)_k - p_k^M| = \frac{|(d + 1) \frac{d!}{\frac{d-1}{2}! \frac{d+1}{2}!} - 2^d|}{2^d(2^d - d - 1)},$$

that converges to 0 as d goes to ∞ . Similarly $\max_k |b(1/2)_k - p_k^M|$ converges to 0 as d goes to ∞ if d is even, since $\binom{d}{k}$ is maximal for $k = \frac{d}{2}$. Since $\lim_{d \rightarrow \infty} d_S(b(1/2), p^M) = 0$ implies $\lim_{d \rightarrow \infty} |b(1/2)_k - p_k^M| = 0$ for any $k \in \{0, \dots, d\}$. We also have $\lim_{d \rightarrow \infty} d_E(b(1/2), p^M) = 0$, where E is the usual Euclidean norm and therefore $l(b(1/2)) = \mathcal{H}^{2^d - d - 1}(\mathcal{P}(b(1/2)))$ converges to the maximum $l^M = l(p^M)$ of the density $l(p)$. \square

Since the Bernoulli distribution is close to the maximal pmf p^M the Hausdorff measure of $\mathcal{P}(b(1/2))$ is close to the maximal one both in low and high dimension. As a consequence we expect that for a given ϵ , $\mathcal{H}^{2^d - 1}(I_S^{\mathcal{F}}(b(1/2), \epsilon))$ converges to the maximal one. The following Theorem proved in [2] provides asymptotic lower bounds for the size in \mathcal{F}_d of $I_S^{\mathcal{F}}(b(1/2), \epsilon)$ and shows that its normalized Hausdorff measure goes to one when d increases.

Theorem 4.4. [2] *There exists a constant A such that for all positive integers d and all positive numbers ϵ ,*

$$\mu(I_S^{\mathcal{F}}(b(1/2), \epsilon)) \geq 1 - \frac{A\sqrt{d}}{\epsilon^2 2^{d-1}}, \quad \text{and} \quad \mu(I_{TV}^{\mathcal{F}}(b(1/2), \epsilon)) \geq 1 - \frac{Ad^{5/2}}{\epsilon^2 2^{d-1}},$$

where μ is the normalized Hausdorff measure on the probability simplex \mathcal{F}_d .

In [8] (Remark 2, Section 5) it is shown that even for moderate d , the lower bound of $\mu(I_S^{\mathcal{F}}(b(1/2), \epsilon))$ is close to one, meaning that the distributions of sums of Bernoulli random variables that are not close to the binomial $b(1/2)$ pmf are rare. Using Equations 3.6 and 3.1 we can find the Hausdorff measure of $I_S^{\mathcal{F}}(b(1/2), \epsilon)$ even for small d , where the asymptotic result can not be applied. The binomial class of pmfs $b(\theta)$ with $\theta \neq 1/2$ is not close to $b(1/2)$ even in high dimension. Notice that from Proposition 4.1 it follows that the closer θ is to $1/2$ the higher the size of the corresponding polytope $\mathcal{P}(b(\theta))$ is. We mention another important discrete distribution, the Poisson-binomial distribution. It is the law of the sum of independent and not identically distributed Bernoulli variables (see e.g. [15] and [1] for an example of its use in applications). The Poisson-binomial distribution with parameter θ , $PB(\theta)$, with $\theta = (\theta_1, \dots, \theta_d)$, is usually far from $b(1/2)$, as for example if $\sum_{i=1}^d \frac{\theta_i}{d} \neq 1/2$, where θ_i are the means of the independent Bernoulli variables. Even if $\sum_{i=1}^d \frac{\theta_i}{d} = 1/2$, [3] proved that in general $b(\theta)$ is not close to $b(1/2)$, see Remark 3) in [8].

The Shepp-Olkin entropy monotonicity conjecture proved in [12] asserts that if \mathbf{X} has independent components X_i with means θ_i , $i = 1, \dots, d$, the entropy $H(\theta)$ of their sum $S = \sum_{j=1}^d X_j$, that is a function of the parameters $\theta = (\theta_1, \dots, \theta_d)$, is non-decreasing in θ if all $\theta_j \leq 1/2$. In [9] the author proves that the binomial distribution $b(\frac{\mu}{d}) \in \mathcal{D}_d$, $\mu = \sum_{i=1}^d \theta_i$ is the maximal entropy distribution in the class of Poisson-binomial distributions $PB(\theta)$. Our Proposition 4.1 proves that the case $\theta_i = 1/2$, i.e. the symmetric binomial distribution, corresponds to the Polytope $\mathcal{P}(b(\theta))$ with maximal Hausdorff measure in the class of binomial distributions. Here, we prove that the symmetrical binomial distribution is the distribution of the sum S of the d -dimensional Bernoulli variable $\mathbf{X} = (X_1, \dots, X_d) \in \mathcal{F}_d$ with maximal entropy.

Proposition 4.5. *The multivariate Bernoulli random variable $U = (U_1, \dots, U_d) \sim \mathbf{f}_U$, where*

$$\mathbf{f}_U(\mathbf{x}) = \begin{cases} \frac{1}{2^d} & \text{if } \mathbf{x} \in \mathcal{X}, \\ 0 & \text{otherwise,} \end{cases}$$

achieves the maximum entropy within \mathcal{F}_d . The sum of its components follows a symmetric binomial distribution.

Proof. It is well known that the uniform random variable over \mathcal{X} has the highest entropy among the class \mathcal{F}_d . □

To further investigate the parallelism with the Shepp-Olkin conjecture, studying the concavity of $l(\theta) = \mathcal{H}^{2^d-d-1}(\mathcal{P}(\theta))$ as a function of the parameters θ is on the agenda of our future research since it is an interesting and nontrivial issue. Indeed, even in dimension two, $l(\theta)$ is not concave for any $\theta \in [0, 1]^2$, e.g. for $\theta_1 = 0.1$ and $\theta_2 = 0.4$ the determinant of the Hessian matrix of $l(\theta)$ is -0.0270 , which is negative.

References

- [1] Philip J Boland and Frank Proschan, *The reliability of k out of n systems*, The Annals of Probability (1983), 760–764. MR0704562
- [2] N Chevallier, *Law of the sum of bernoulli random variables*, Theory of Probability & Its Applications **55** (2011), no. 1, 27–41. MR2768519
- [3] Werner Ehm, *Binomial approximation to the Poisson binomial distribution*, Statistics & Probability Letters **11** (1991), no. 1, 7–16. MR1093412
- [4] Lawrence Craig Evans, *Measure theory and fine properties of functions*, Routledge, 2018.
- [5] Vissarion Fisikopoulos, Apostolos Chalkis, and contributors in file inst/AUTHORS, *volesti: Volume approximation and sampling of convex polytopes*, 2020, R package version 1.1.2.
- [6] Roberto Fontana, Elisa Luciano, and Patrizia Semeraro, *Model risk in credit risk*, Mathematical Finance (2020), 1–27. MR4205881
- [7] Roberto Fontana and Patrizia Semeraro, *Representation of multivariate Bernoulli distributions with a given set of specified moments*, Journal of Multivariate Analysis **168** (2018), 290–303. MR3858366
- [8] Ricardo Fraiman, Leonardo Moreno, and Thomas Ransford, *A quantitative heppes theorem and multivariate bernoulli distributions*, Journal of the Royal Statistical Society Series B: Statistical Methodology **85** (2023), no. 2, 293–314. MR4726968
- [9] Peter Harremoës, *Binomial and Poisson distributions as maximum entropy distributions*, IEEE Transactions on Information Theory **47** (2001), no. 5, 2039–2041. MR1842536
- [10] Martin Henk, Jürgen Richter-Gebert, and Günter M Ziegler, *Basic properties of convex polytopes*, Handbook of discrete and computational geometry, Chapman and Hall/CRC, 2017, pp. 383–413.
- [11] David J Hessen, *A family of discrete maximum-entropy distributions*, Journal of Statistical Planning and Inference **236** (2025), 106243. MR4804778

- [12] Erwan Hillion and Oliver T Johnson, *A proof of the shepp-olkin entropy monotonicity conjecture*, *Electronic Journal of Probability* **24** (2019), 126. MR4029429
- [13] Man-Lai Tang Kai Wang Ng, Guo-Liang Tian, *Dirichlet and related distributions: Theory, methods and applications (wiley series in probability and statistics)*, 1 ed., Wiley, 2011. MR2830563
- [14] HF Martz, RA Wailer, and ET Fickas, *Bayesian reliability analysis of series systems of binomial subsystems and components*, *Technometrics* **30** (1988), no. 2, 143–154.
- [15] Wenpin Tang and Fengmin Tang, *The Poisson binomial distribution—old & new*, *Statistical Science* **38** (2023), no. 1, 108–119. MR4534645
- [16] PAG Van Der Geest, *The binomial distribution with dependent bernoulli trials*, *Journal of Statistical Computation and Simulation* **75** (2005), no. 2, 141–154. MR2117011
- [17] P Vellaisamy and Abraham P Punnen, *On the nature of the binomial distribution*, *Journal of applied probability* **38** (2001), no. 1, 36–44. MR1816111

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System: <https://vtex.lt/services/ejms-peer-review/>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <https://imstat.org/shop/donation/>