High dimensional Bernoulli distributions: algebraic representation and applications

Bernoulli **30**(1), 2024, 825–850
https://doi.org/10.3150/23-BEJ1618

High dimensional Bernoulli distributions: Algebraic representation and applications

ROBERTO FONTANA[a] and PATRIZIA SEMERARO[b]

Department of Mathematical Sciences, Politecnico di Torino, corso Duca Degli Abruzzi 24, 10129 Torino, Italy,
[a]*roberto.fontana@polito.it,* [b]*patrizia.semeraro@polito.it*

The main contribution of this paper is to find a representation of the class $\mathcal{F}_d(p)$ of multivariate Bernoulli distributions with the same mean p that allows us to find its generators analytically in any dimension. We map $\mathcal{F}_d(p)$ to an ideal of points and we prove that the class $\mathcal{F}_d(p)$ can be generated from a finite set of simple polynomials. We present two applications. Firstly, we show that polynomial generators help to find extremal points of the convex polytope $\mathcal{F}_d(p)$ in high dimensions. Secondly, we solve the problem of determining the lower bounds in the convex order for sums of multivariate Bernoulli distributions with given margins, but with an unspecified dependence structure.

Keywords: Convex order; extremal points; ideal of points; multidimensional Bernoulli distribution

1. Introduction

This paper proposes an algebraic representation of the class $\mathcal{F}_d(p)$ of d-dimensional Bernoulli distributions with Bernoulli univariate marginal distributions with parameter p - the class of joint distributions of $(X_1, \dots, X_d)$, where $X_i \sim \text{Bernoulli}(p)$, $i = 1, \dots, d$. This representation turns out to be fundamental for solving some open issues regarding this class. Our main theoretical contribution is to find a set of generators of the class $\mathcal{F}_d(p)$ analytically in any dimension. We build on the geometrical structure of $\mathcal{F}_d(p)$. The class $\mathcal{F}_d(p)$ is a convex polytope (see [12]) that we map into an ideal of points. Using the Gröebner basis of the ideal we find an analytical set of polynomials that generates the class $\mathcal{F}_d(p)$. These polynomial generators are also very simple in high dimensions. Through two applications we show that this novel representation allows us to address and solve some open problems in applied probability and statistics.

One open problem in applied probability is to find the lower bounds in the convex order for sums $S = X_1 + \ldots + X_d$ of the components of random vectors in a given Fréchet class, i.e. with given margins and an unspecified dependence structure (see, e.g. [26]). The upper bound is known to be the upper Fréchet bound, while the problem to find the lower bound has been solved only under specific and restrictive conditions. Sums of Bernoulli random variables have received much attention among researchers in computer science, optimization, engineering and finance because of its wide applicability, see e.g. [25] and references therein. The minimum convex order corresponds to the minimum risk distribution and it is relevant for example in credit risk, where univariate marginal distributions represent the probability of default of obligors in a credit portfolio (a classical reference for credit risk modelling is [24]). If obligors belong to the same class of rating, they have the same marginal default probability. Therefore it is important to assess the risk associated to possible dependence among obligors, especially nowadays since the global economy and risks are highly interconnected. For sums of Bernoulli variables this problem is solved in the class of exchangeable Bernoulli variables or only if $pd < 1$, which is a restrictive hypothesis for credit portfolios because banks usually handle hundreds of obligors.

1350-7265

The main application of our theoretical results is to solve the problem of finding lower bounds for sums of the components of random vectors with probability mass function (pmf) in $\mathcal{F}_d(p)$ for any p and d. We proceed in two steps. First we find the minimal distribution of sums in the convex order by building on the geometrical structure of discrete distributions. In this first step we immediately obtain one corresponding pmf in the class of exchangeable Bernoulli distributions (see [11]). To have a minimal pmf that is not-exchangeable is not trivial, and this is our second step. We find one pmf in $\mathcal{F}_d(p)$ that is minimal and not-exchangeable by combining what we will define as fundamental polynomials. The proof of this result shows how we can work with polynomials to obtain pmfs that satisfy some conditions. We are confident that this approach can be used to address many other issues. We also prove that the minimum convex order corresponds to the minimal mean correlation. It is well known that the minimal convex order is associated to negative dependence. In fact for $pd < 1$ the minimum convex order of sums is the lower Fréchet bound of the class, that is the distribution of the mutually exclusive random variables. Mutual exclusivity was first studied by [8] in the context of insurance and finance. It was called the safest dependence structure and was characterized as the strongest negative dependence structure by [4]. We also propose a generalization of mutual exclusivity for $pd > 1$ and we study its relationship with the minimum convex order, to attempt to find a general notion of the safest dependence structure.

Statistical research extensively investigates classes of multivariate Bernoulli distributions and its properties (see e.g. [6,23] and [3]) because of the importance of binary data in applications. Issues such as high dimensional simulation - see e.g. [17,20] and [30] - and estimate - [21] - are very important for many applications and our novel representation is particularly convenient for working in high dimensions. High dimensional simulation and testing is possible for some classes and under some conditions, for example see [10,20,27,30]. High dimensional simulation for exchangeable Bernoulli pmfs is addressed in [14]. Exchangeable Bernoulli pmfs are points in a convex polytope whose extremal points are analytically known ([11,13]) and high dimensional simulations is possible because we know how to sample from a polytope [14]. To use this approach extremal points are necessary. In [12] the authors provide a method to explicitly find the extremal points of $\mathcal{F}_d(p)$, but computational complexity increases very quickly and they can not be found in high dimensions. We prove that fundamental polynomials are associated to extremal points in the polytope, thus this representation also helps in finding analytically a huge number of extremal points in any dimension. We provide an algorithm. In this application it is evident that working with polynomials is simpler than with pmfs.

The paper is organized as follows. Section 2 introduces the setting and formalizes the problem. The main theoretical results are given in Section 3, where we provide an algebraic representation of $\mathcal{F}_d(p)$. Section 4 uses the algebraic representation to address the issue to find the extremal points and provides an algorithm to find them in any dimension. Section 5 applies the theoretical results to solve the problem of finding bounds in the convex order for sums of variables with pmf in $\mathcal{F}_d(p)$ if $pd > 1$. We conclude this section proving that for $pd > 1$ the minimal convex order corresponds to the minimal mean correlation. We also extend the notion of mutually exclusive random vectors to the case $pd > 1$ and we study its connection with minimality in the convex order in Section 6. Section 7 concludes.

2. Preliminaries and motivation

We build on the geometrical representation of the class $\mathcal{F}_d(p)$ as a convex polytope to map it into a ring of polynomials. This approach requires the introduction of some notation from computational geometry, algebra and algebraic geometry. A standard reference for these topics is [5].

2.1. Notation

Let $\mathcal{F}_d$ be the set of d-dimensional probability mass functions (pmfs) which have Bernoulli univariate marginal distributions. Let us consider the Fréchet class $\mathcal{F}_d(p) \subseteq \mathcal{F}_d$ of pmfs in $\mathcal{F}_d$ which have the same Bernoulli marginal distributions of parameter p, $B(p)$. We assume throughout the paper that p is rational, i.e. $p \in \mathbb{Q}$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, this is not a limitation in applications.

If $\boldsymbol{X} = (X_1, \dots, X_d)$ is a random vector with pmf in $\mathcal{F}_d$, we denote

- its cumulative distribution function by F and its mass function by f;
- the column vector which contains the values of F and f over $\mathcal{X} = \{0,1\}^d$, by $\boldsymbol{F} = (F_1, \dots, F_D) = (F_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}) := (F(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X})$, and $\boldsymbol{f} = (f_1, \dots, f_D) = (f_{\boldsymbol{x}} : \boldsymbol{x} \in \mathcal{X}) := (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X})$, $D = 2^d$, respectively; 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 we have $\mathcal{S}_3 = \{000, 100, 010, 110, 001, 101, 011, 111\}$.

The notation $\boldsymbol{X} \in \mathcal{F}_d(p)$ or $\boldsymbol{F} \in \mathcal{F}_d(p)$ indicates that $\boldsymbol{X}$ has pmf $f \in \mathcal{F}_d(p)$. Given a vector $\boldsymbol{x}$ we denote by $|\boldsymbol{x}|$ the number of elements of $\boldsymbol{x}$ which are different from 0. If $\boldsymbol{x}$ is binary, $|\boldsymbol{x}| = \sum_{i=1}^d x_i$. We assume that vectors are column vectors and we denote by $A^\top$ the transpose of a matrix A. Given two matrices $A \in \mathcal{M}(n \times m)$ and $B \in \mathcal{M}(d \times l)$:

- the matrix $A \otimes B = ((a_{ij}B)_{1 \le i \le n, 1 \le j \le m}) \in \mathcal{M}(nd \times ml)$ indicates their Kronecker product and $A^{\otimes n}$ is $\underbrace{A \otimes \cdots \otimes A}_{n \text{ times}}$;
- if $n = d$, $A||B$ denotes the row concatenation of A and B;
- if $m = l$, $A//B$ denotes the column concatenation of A and B.

Finally, we denote by $P(\boldsymbol{x}) = \sum_{\alpha \in \mathcal{A}} a_\alpha \boldsymbol{x}^\alpha$ a polynomial in $\mathbb{Q}[x_1, \dots, x_d]$, where $\boldsymbol{x}^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathcal{A}$ and $\mathcal{A}$ is a proper set of multi-indexes. We often use $\mathcal{A} = \{0,1\}^d$. In some cases we write $a_{\text{supp}(\alpha)}$, where $\text{supp}(\alpha) = \{i \in \{1, \dots, d\} : \alpha_i \neq 0\}$ to simplify the notation, i.e $a_{\{1,4\}} \equiv a_{1,4}$ instead of $a_{(1,0,0,1)}$ and $a_\varnothing$ instead of $a_{(0,\dots,0)}$. Finally, if no confusion arises $a_j, j = 1, \dots D$ stands for a_α where $\alpha \in \{0,1\}^d$ is the j-th term in reverse lexicographic order, i.e. a_{10} instead of $a_{1,4} = a_{(1,0,0,1)}$.

2.2. The convex polytope $\mathcal{F}_d(p)$

Let $\boldsymbol{X}$ be a multivariate Bernoulli random vector whose distribution belongs to the Fréchet class $\mathcal{F}_d(p) \subseteq \mathcal{F}_d$, $\boldsymbol{X} \in \mathcal{F}_d(p)$. As described in [12], we can write

$$\mathrm{E}(X_i) = \boldsymbol{x}_i^\top \boldsymbol{f},$$

where $\boldsymbol{x}_i$ is the vector which contains only the i-th element of $\boldsymbol{x} \in \mathcal{X}$, $i \in \{1, \dots, d\}$, e.g for the bivariate case $\boldsymbol{x}_1^T = (0,1,0,1)$ and $\boldsymbol{x}_2^T = (0,0,1,1)$. For $\boldsymbol{f} \in \mathcal{F}_d(p)$ we have

$$\begin{cases} \boldsymbol{x}_i^\top \boldsymbol{f} = p \\ (\boldsymbol{1} - \boldsymbol{x}_i)^\top \boldsymbol{f} = q, \end{cases}$$

where $\boldsymbol{1}$ is the vector with all the elements equal to 1 and $q = 1 - p$. If we consider the odds of the event $X_i = 0$, $c = \frac{1}{\gamma} = q/p$, we have

$$((\boldsymbol{1} - \boldsymbol{x}_i)^\top - c\boldsymbol{x}_i^\top)\boldsymbol{f} = 0.$$

Let H be the $d \times D$ matrix whose rows, up to a non-influential multiplicative constant, are $((\mathbf{1} - \boldsymbol{x}_i)^\top - c\boldsymbol{x}_i^\top)$, $i \in \{1, \dots, d\}$. The probability mass functions $\boldsymbol{f}$ in $\mathcal{F}_d(p)$ are the positive normalized solutions of the linear system, i.e. $f_i \geq 0$, $\sum_{i=1}^D f_i = 1$

$$H\boldsymbol{f} = \mathbf{0}. \tag{1}$$

From the standard theory of linear equations we know that all the positive, normalized solutions of (1) are the elements of a convex polytope, thus

$$\mathcal{F}_d(p) = \{\boldsymbol{f} \in \mathbb{R}^{2^d} : H\boldsymbol{f} = 0, f_i \geq 0, \sum_{i=1}^{D} f_i = 1\}. \tag{2}$$

The solutions of the linear system in Equation (1) are convex combinations of a set of generators which are referred to as extremal points of the convex polytope. Formally, for any $\boldsymbol{f} \in \mathcal{F}_d(p)$ there exist $\lambda_1 \geq 0, \dots, \lambda_n \geq 0$ summing up to one and extremal points $\boldsymbol{r} \in \mathcal{F}_d(p)$ such that

$$\boldsymbol{f} = \sum_{i=1}^{n} \lambda_i \boldsymbol{r}_i.$$

We call $\boldsymbol{r}_i$ extremal points or extremal pmfs. We denote with $\boldsymbol{R}_i$ a random vector with pmf $\boldsymbol{r}_i$. The proof of the Proposition 2.1 follows from Carathéodory's Theorem, see [28] for a general reference on convex analysis and Lemma 2.3 in [31].

Proposition 2.1. *Let us consider the linear system*

$$Az = 0,\ z \in \mathbb{R}^m,$$

where A is a $n \times m$ matrix, $n \leq m$ and $\operatorname{rank}(A) = n$. *The extremal points of the polytope*

$$\mathcal{P}_A := \{z \in \mathbb{R}^m : Az = 0, z_i \geq 0, \sum_{i=1}^{m} z_i = 1\} \tag{3}$$

have at most $n + 1$ non-zero components.

Example 1. As an illustrative example we consider the class $\mathcal{F}_3(2/5)$, i.e. $d = 3$ and $p = \frac{2}{5}$. We have $c = \frac{3}{2}$ and

$$H = \begin{pmatrix} 1 & -\frac{3}{2} & 1 & -\frac{3}{2} & 1 & -\frac{3}{2} & 1 & -\frac{3}{2} \\ 1 & 1 & -\frac{3}{2} & -\frac{3}{2} & 1 & 1 & -\frac{3}{2} & -\frac{3}{2} \\ 1 & 1 & 1 & 1 & -\frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} \end{pmatrix}. \tag{4}$$

In this case the extremal points can be found using 4ti2 (see [1]). Giving H as input, we generate the matrix R whose columns are the extremal pmfs $\boldsymbol{r}_i$, $i = 1, \dots, 9$, which are reported in Table 1.

The dimension of the system (1) increases very fast because the number of unknowns is $D = 2^d$, and finding all the extremal generators of the system becomes computationally infeasible. For example for the moderate-size case $d = 6$ and $p = \frac{2}{5}$ there are 1,251,069 extremal pmfs. Proposition 2.1 states that the support of an extremal pmf has at most $d + 1$ points. One possible approach for finding some extremal pmfs could be based on the selection of $d + 1$ components of $\boldsymbol{f}$ and set the others equal to

x_1	x_2	x_3	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
0	0	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	0	0	0	0
1	0	0	0	0	$\frac{2}{5}$	0	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{10}$
0	1	0	0	$\frac{2}{5}$	0	0	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{3}{10}$
1	1	0	$\frac{2}{5}$	0	0	$\frac{1}{5}$	0	$\frac{1}{5}$	0	0	0
0	0	1	$\frac{2}{5}$	0	0	0	0	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{3}{10}$
1	0	1	0	$\frac{2}{5}$	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$	0	0
0	1	1	0	0	$\frac{2}{5}$	$\frac{1}{5}$	0	0	0	$\frac{1}{5}$	0
1	1	1	0	0	0	0	$\frac{2}{5}$	0	0	0	$\frac{1}{10}$

Table 1. Extremal pmfs of $\mathcal{F}_d(p)$ where $d = 3$ and $p = \frac{2}{5}$. There are nine extremal pmfs.

zero. This is equivalent to consider a submatrix H_1 of H made by $d+1$ columns of H. The extremal points of H_1 must be computed. The problem is that we are interested in the *positive* solutions of the system $H_1 z$, $z \in \mathbb{R}_+^{d+1}$. As we show in Example 2 this approach could be extremely inefficient because many choices of H_1 do not have positive solutions.

Example 2. We consider the class $\mathcal{F}_5(2/5)$, i.e. the case $d = 5$ and $p = \frac{2}{5}$. In this case the matrix H has $d = 5$ rows and $D = 32$ columns. We can compute all the extremal pmfs of H. There are 5,162 extremal pmfs. For 1,000 times we repeat the random selection of the submatrix H_1 of H (H_1 is made by $d+1 = 6$ columns of H) and the computation of the extremal pmfs of H_1. We find a non-empty set of extremal pmfs only in 162 cases, providing an estimate of the success of this simple approach equal to 16.2%.

Given the computational limitation of the above representation, we propose a different approach. The next section describes an algebraic representation of the class $\mathcal{F}_d(p)$ that reduces the complexity of the problem and provides a new analytical class of generators. The new generators are extremal points. They are very simple and allow us to easily find pmfs in the class.

3. Main results

We make the non-restrictive hypothesis that $p \leq 1/2$. Since $p \in \mathbb{Q}$, let $p = s/t$ with $s \leq t$. We get $c = q/p = (t-s)/s$. Let $\boldsymbol{x} = (x_1, \ldots, x_{d-1})^T$ and $\boldsymbol{x}_i = (1, x_i)^T$. Let us consider the row vectors $\boldsymbol{m}_+(\boldsymbol{x}) = (m_1(\boldsymbol{x}), \ldots, m_{D/2}(\boldsymbol{x}))^T := \boldsymbol{x}_{d-1} \otimes \ldots \otimes \boldsymbol{x}_1$ and $\boldsymbol{m}_-(\boldsymbol{x}) = (-m_{D/2}(\boldsymbol{x}) + \frac{2s-t}{s}, \ldots, -m_1(\boldsymbol{x}) + \frac{2s-t}{s})^T$ and finally let $\boldsymbol{m}(\boldsymbol{x}) = (\boldsymbol{m}_+(\boldsymbol{x}) || \boldsymbol{m}_-(\boldsymbol{x}))$, the row vector obtained concatenating $\boldsymbol{m}_+(\boldsymbol{x})$ and $\boldsymbol{m}_-(\boldsymbol{x})$.

We define the map $\mathcal{H}$ from $\mathcal{F}_d(p)$ to the polynomial ring with rational coefficients $\mathbb{Q}[x_1, \ldots, x_d]$ as:

$$\begin{aligned} &\mathcal{H} : \mathcal{F}_d(p) \to \mathbb{Q}[x_1, \ldots, x_{d-1}] \\ &\mathcal{H} : \boldsymbol{f} \to P_f(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x})\boldsymbol{f}. \end{aligned} \tag{5}$$

We call $C_{\mathcal{H}}$ the image of $\mathcal{F}_d(p)$ through H. By construction, a rearrangement of the coefficient of $P_f(\boldsymbol{x})$ gives

$$P_f(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x})\boldsymbol{f} = \sum_{\alpha \in \{0,1\}^d} a_\alpha x^\alpha, \tag{6}$$

Table 2. The $\boldsymbol{m}_+(\boldsymbol{x})$ matrix for the case $d = 3$ and $p = 2/5$.

$\boldsymbol{m}_+(\boldsymbol{x}) =$	1	x_1	x_2	x_1x_2	
$\boldsymbol{m}_+(-3/2,1) =$	1	$-\frac{3}{2}$	1	$-\frac{3}{2}$	$= H[1,\{1,2,3,4\}]$
$\boldsymbol{m}_+(1,-3/2) =$	1	1	$-\frac{3}{2}$	$-\frac{3}{2}$	$= H[2,\{1,2,3,4\}]$
$\boldsymbol{m}_+(1,1) =$	1	1	1	1	$= H[3,\{1,2,3,4\}]$

where $a_\alpha \in \mathbb{Q}$ are linear combinations of the elements of $\boldsymbol{f}$. Specifically, it holds

$$\boldsymbol{a} = Q\boldsymbol{f},$$

where $\boldsymbol{a} = (a_\alpha : \alpha \in \{0,1\}^{d-1})$,

$$Q = (I(2^{d-1})||\tilde{I}(2^{d-1}) + \tilde{A}), \tag{7}$$

$I(2^{d-1})$ is the 2^{d-1} dimensional identity matrix, $\tilde{I}(2^{d-1})$ is the 2^{d-1} dimensional matrix with -1 on the secondary diagonal and all other entries equal to zero and $\tilde{A}$ is a 2^{d-1} square matrix whose entries are $\tilde{a}_{1j} = (2s-t)/s$, $j = 1,\dots 2^{d-1}$ and $\tilde{a}_{ij} = 0$, $j = 2,\dots,2^{d-1}$.

Example 3. As in Example 1 we consider $\mathcal{F}_3(2/5)$, thus $d = 3$, $s = 2$, $t = 5$, $p = \frac{2}{5}$, $c = \frac{3}{2}$ and $\frac{2s-t}{s} = -\frac{1}{2}$. Given $\boldsymbol{f} \in \mathcal{F}_3(2/5)$, we have $\boldsymbol{a} = Q\boldsymbol{f}$, where

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -1-\frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

For example the image of $\boldsymbol{r}_9$ (one of the extremal pmfs listed in Table 1) is $P_{\boldsymbol{r}_9}(\boldsymbol{x}) = -0.3(1 - x_1 - x_2 + x_1x_2)$. In the view of the proof of Theorem 3.1, we observe that the elements of $\boldsymbol{m}_+(\boldsymbol{x}) = \boldsymbol{m}_+(x_1,x_2) = (1,x_1) \otimes (1,x_2) = (1\ \ x_1\ \ x_2\ \ x_1x_2)$ computed in $\boldsymbol{x} = (-3/2,1)$, $\boldsymbol{x} = (1,-3/2)$ and $\boldsymbol{x} = (1,1)$ can be put in a one-to-one correspondence with the first $D/2 = 4$ columns of the matrix H of Equation (4) as represented in Table 2, where $H[i,\{1,2,3,4\}]$ is the i-th row of the matrix containing the columns $\{1,2,3,4\}$ of H, $i = 1,2,3$. Similarly we observe that the elements of $\boldsymbol{m}_-(\boldsymbol{x}) = \boldsymbol{m}_-(x_1,x_2) = (-x_1x_1 - 1/2\ \ -x_2 - 1/2\ \ -x_1 - 1/2\ \ -1-1/2)$ computed in $\boldsymbol{x} = (-3/2,1)$, $\boldsymbol{x} = (1,-3/2)$ and $\boldsymbol{x} = (1,1)$ can be put in a one-to-one correspondence with the last $D/2 = 4$ columns of the matrix H of Equation (4) as reported in Table 3, where $H[i,\{5,6,7,8\}]$ is the i-th row of the matrix containing the columns $\{5,6,7,8\}$ of H, $i = 1,2,3$.

Table 3. The $\boldsymbol{m}_-(\boldsymbol{x})$ matrix for the case $d = 3$ and $p = 2/5$.

$\boldsymbol{m}_-(\boldsymbol{x}) =$	$-x_1x_2-\frac{1}{2}$	$-x_2-\frac{1}{2}$	$-x_1-\frac{1}{2}$	$-\frac{3}{2}$	
$\boldsymbol{m}_-(-3/2,1) =$	1	$-\frac{3}{2}$	1	$-\frac{3}{2}$	$= H[1,\{5,6,7,8\}]$
$\boldsymbol{m}_-(1,-3/2) =$	1	1	$-\frac{3}{2}$	$-\frac{3}{2}$	$= H[2,\{5,6,7,8\}]$
$\boldsymbol{m}_-(1,1) =$	$-\frac{3}{2}$	$-\frac{3}{2}$	$-\frac{3}{2}$	$-\frac{3}{2}$	$= H[3,\{5,6,7,8\}]$

Theorem 3.1. *The map $\mathcal{H}$ maps $\mathcal{F}_d(p)$ into polynomials $\sum_\alpha a_\alpha \boldsymbol{x}^\alpha \in \mathbb{Q}[x_1,...,x_{d-1}]$, $\alpha \in \{0,1\}^{d-1}$ of the ideal $\mathcal{I}_\mathcal{P}$ of polynomials which vanish at points $\mathcal{P} = \{\mathbf{1}_{d-1}, \boldsymbol{c}_j, j = 1,\ldots,d-1\}$, where $\mathbf{1}_{d-1} = (1,\ldots,1)$ and $\boldsymbol{c}_j = (1,\ldots,1,\underbrace{-c}_{j\text{-th element}},1,\ldots 1)$.*

Proof. Let $H(d)$, $\boldsymbol{m}_+^d(\boldsymbol{x})$ and $\boldsymbol{m}_+^d(\boldsymbol{x})$ be the matrices H associated to $\mathcal{F}_d(p)$, $\boldsymbol{m}_+(\boldsymbol{x})$ and $\boldsymbol{m}_-(\boldsymbol{x})$ in dimension d, respectively. Let $\boldsymbol{f} \in \mathcal{F}_d(p)$ and $H_{j.}(d)$ be the j-th row of $H(d)$. We show that $H_{j.}(d)\boldsymbol{f}$ is the polynomial $P_f(\boldsymbol{x})$ evaluated at $\boldsymbol{x} = \boldsymbol{c}_j$, that is $H_{j.}(d)\boldsymbol{f} = P_f(\boldsymbol{c}_j)$, $j = 1,\ldots,d-1$, and $H_{d.}(d)\boldsymbol{f} = P_f(\mathbf{1}_{d-1})$. Let $H^+(d)$ be the submatrix of the first 2^{d-1} columns of $H(d)$ and $H^-(d)$ the submatrix of the last 2^{d-1} columns of $H(d)$, i.e. $H(d) = (H^+(d)||H^-(d))$.

We consider the following two steps. Step one proves by induction that $H_{j.}^+(d) = \boldsymbol{m}_+^d(\boldsymbol{c}_j)$, for $j = 1,\ldots,d-1$ and $H_{d.}^+(d) = \boldsymbol{m}_+^d(\mathbf{1}_{d-1})$. We consider $d = 2$. It holds $\boldsymbol{c}_1 = -c$, $\mathbf{1}_{d-1} = 1$ and $\boldsymbol{m}_+^2(\boldsymbol{x}) = (1,x)^T$. We have $\boldsymbol{m}_+^2(\boldsymbol{c}_1) = (1,-c)^T$ and $\boldsymbol{m}_+^2(\mathbf{1}_1) = (1,1)^T$. Since $H_{1.}^+ = (1,-c)^T$ and $H_{2.}^+ = (1,1)^T$, the case $d = 2$ is proved. Let us assume that the assert is true for d. We prove it for $d+1$. We consider separately three cases: $j = 1,\ldots d-1$, $j = d$ and $j = d+1$. For $j = 1,\ldots d-1$, we have $x_j = -c$ and then it must be $x_d = 1$ by construction. Therefore,

$$H_{j.}^+(d+1) = (1,1)^T \otimes H_{j.}^+(d) = (1,1) \otimes \boldsymbol{m}_+^d(\boldsymbol{c}_j) = (\boldsymbol{m}_+^d(\boldsymbol{c}_j)||1) = \boldsymbol{m}_+^{d+1}(\boldsymbol{c}_j).$$

For $j = d$ we have $x_d = -c$ and

$$H_{d.}^+(d+1) = (1,-c)^T \otimes H_d^+(d) = (1,-c) \otimes \boldsymbol{m}_+^d(\mathbf{1}_d) = \boldsymbol{m}_+^{d+1}(\boldsymbol{c}_d).$$

Finally, for $j = d+1$ we have

$$H_{d+1.}^+(d+1) = (1,1)^T \otimes H_{d.}^+(d) = (1,1) \otimes \boldsymbol{m}_+^d(\mathbf{1}_d) = \boldsymbol{m}_+^{d+1}(\mathbf{1}_d)$$

and the step one is proved. Step two proves that $H_{j.}^-(d) = \boldsymbol{m}_-^d(\boldsymbol{c}_j)$, for $j = 1,\ldots,d-1$ and $H_{d.}^-(d) = \boldsymbol{m}_-^d(\mathbf{1}_{d-1})$. It is sufficient to observe that $\boldsymbol{m}_-^d(\boldsymbol{x}) = (-m_{D/2}(\boldsymbol{x}) + \frac{2s-t}{s},\ldots - m_1(\boldsymbol{x}) + \frac{2s-t}{s})^T$ and $H^- = (-H_{1.}^+ + \frac{2s-t}{s},\ldots,-H_{d.}^+ + \frac{2s-t}{s})$ and the assert of step two easily follows. As a consequence $H_j(d) = \boldsymbol{m}^d(\boldsymbol{c}_j)$, $j = 1,\ldots,d-1$ and $H_d(d) = \boldsymbol{m}^d(\mathbf{1}_{d-1})$. Since $\boldsymbol{f} \in \mathcal{F}_d(p)$ iff $H(d)\boldsymbol{f} = 0$ we have $\boldsymbol{m}^d(\boldsymbol{c}_j)\boldsymbol{f} = 0$ and $\boldsymbol{m}^d(\mathbf{1}_{d-1})\boldsymbol{f} = 0$. Therefore $P_f(\boldsymbol{x}) = \boldsymbol{m}^d(\boldsymbol{x})\boldsymbol{f} \in \mathcal{I}_\mathcal{P}$ and by construction (see (6)) $\boldsymbol{m}^d(\boldsymbol{x})\boldsymbol{f} = \sum_\alpha a_\alpha \boldsymbol{x}^\alpha$, $\alpha \in \{0,1\}^{d-1}$, with $\boldsymbol{a} = Q\boldsymbol{f}$. □

Remark 1. The map $\mathcal{H}$ is not injective. In Example 1, we have $\mathcal{H}(\boldsymbol{r}_1) = 1/5 + 2/5xy - 2/5(xy + 1/2) = 0$, $\mathcal{H}(\boldsymbol{r}_2) = 1/5 + 2/5y - 2/5(y + 1/2) = 0$ and similarly $\mathcal{H}(\boldsymbol{r}_3) = \mathcal{H}(\boldsymbol{r}_4) = \mathcal{H}(\boldsymbol{r}_5) = 0$, thus $\{\boldsymbol{r}_1,\boldsymbol{r}_2,\boldsymbol{r}_3,\boldsymbol{r}_4,\boldsymbol{r}_5\} \subseteq Ker(\mathcal{H})$.

Since $\mathcal{H}$ is not injective, as observed in Remark 1, given a polynomial $P(\boldsymbol{x}) = \sum_\alpha a_\alpha \boldsymbol{x}^\alpha \in \mathcal{I}_\mathcal{P}$, $\alpha \in \{0,1\}^{d-1}$ there are many pmfs in $\mathcal{H}^{-1}(P(\boldsymbol{x}))$ as we study in the remaining part of this section. We provide a three steps algorithm to find one pmf $\boldsymbol{f}_P = (f_1,\ldots,f_D) \in \mathcal{F}_d(p)$ associated to a given $P(\boldsymbol{x}) = \sum_{\alpha\in\{0,1\}^{d-1}} a_\alpha \boldsymbol{x}^\alpha \in \mathcal{I}_\mathcal{P}$. It is easy to verify that $\boldsymbol{f}_p \in \mathcal{F}_d(p)$ and that $\mathcal{H}(\boldsymbol{f}_p) = P(\boldsymbol{x})$. We call $\boldsymbol{f}_p$ the type-0 pmf associated to the polynomial P. We describe the use of this algorithm in Example 4.

Example 4. We consider two polynomials: the first one has a positive constant term, the second one a negative constant term. Consider the case of Example 1 ($d = 3, p = 2/5$) and let $P(\boldsymbol{x})$ be the polynomial $P(\boldsymbol{x}) = x_1x_2 - x_1 - x_2 + 1 \in C_\mathcal{H}$. We have $a_1 = 1$, $a_2 = a_3 = -1$, $a_4 = 1$ and then:

Algorithm 1

Input: A polynomial $P(\boldsymbol{x}) = \sum_{\alpha \in \{0,1\}^{d-1}} a_\alpha \boldsymbol{x}^\alpha \in \mathcal{I}_\mathcal{P}$.
1)For each $j \in \{2, \dots, D/2\}$: if $a_j \geq 0$, then $f_j = a_j$ and $f_{D+1-j} = 0$; if $a_j < 0$ then $f_j = 0$ and $f_{D+1-j} = -a_j$. 2) Let $c_0 = \sum_{j=1, a_j<0}^{d/2} a_j \frac{2s-t}{s} + a_1$. If $c_0 \geq 0$ then $f_1 = c_0$ and $f_D = 0$; if $c_0 < 0$ then $f_1 = 0$ and $f_D = -\frac{c_0}{c}$, where $c = \frac{q}{p}$; 3) normalize $\boldsymbol{f}_p$ getting, with a small abuse of notation $\boldsymbol{f}_p := \boldsymbol{f}_p / (\sum_{j=1}^D f_j)$.
Output: a pmf $\boldsymbol{f}_P = (f_1, \dots, f_D) \in \mathcal{F}_d(p)$.

1. step 1 immediately gives $\boldsymbol{f}_P = (f_1, 0, 0, 1, 0, 1, 1, f_8)$;
2. from step 2 we get $c_0 = \frac{1}{2} + \frac{1}{2} + 1 = 2$, thus $f_1 = c_0 = 2$, $f_8 = 0$ and $\boldsymbol{f}_P = (2, 0, 0, 1, 0, 1, 1, 0)$;
3. the normalization step gives $\boldsymbol{f}_P = (\frac{2}{5}, 0, 0, \frac{1}{5}, 0, \frac{1}{5}, \frac{1}{5}, 0)$. We observe that $\boldsymbol{f}_P = \boldsymbol{r}_4$, see Table 1.

Consider now the polynomial $P(\boldsymbol{x}) = -x_1 x_2 + x_1 + x_2 - 1 \in \mathcal{C}_\mathcal{H}$. We have $a_1 = -1$, $a_2 = a_3 = 1$, $a_4 = -1$ and then:

1. step 1 immediately gives $\boldsymbol{f}_P = (f_1, 1, 1, 0, 1, 0, 0, f_8)$;
2. from step 2 we get $c_0 = -1/2$, thus $f_1 = 0$, $f_D = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$ and $\boldsymbol{f}_P = (0, 1, 1, 0, 1, 0, 0, \frac{1}{3})$;
3. the normalization step gives $\boldsymbol{f}_P = (0, \frac{3}{10}, \frac{3}{10}, 0, \frac{3}{10}, 0, 0, \frac{1}{10})$. We observe that $\boldsymbol{f}_P = \boldsymbol{r}_9$, see Table 1.

In Proposition 3.1 given a polynomial $P(\boldsymbol{x}) \in \mathcal{I}_\mathcal{P}$ we determine all the pmfs such that $\mathcal{H}(f) = P(\boldsymbol{x})$.

Proposition 3.1. *Given a polynomial* $P(\boldsymbol{x}) = \sum_\alpha a_\alpha \boldsymbol{x}^\alpha \in \mathcal{I}_\mathcal{P}$, $\alpha \in \{0,1\}^{d-1}$,

$$\mathcal{H}^{-1}(P(\boldsymbol{x})) = \{\boldsymbol{f}_p + \boldsymbol{e}_k, \ \ \boldsymbol{e}_k \in Ker(\mathcal{H})\},$$

where $\boldsymbol{f}_p$ *is the type-0 pmf associated to* $P(\boldsymbol{x})$*. A basis of* $Ker(\mathcal{H})$ *is*

$$\mathcal{B}_K = \{(q, 0, \dots, 0, p); (1-2p, p, 0, \dots, 0, p, 0); (1-2p, 0, p, 0, \dots 0,, p, 0, 0);$$
$$\dots (1-2p, 0, \dots, 0, p, p, 0 \dots, 0)\}.$$

Proof. Notice that $Ker(\mathcal{H})$ in (5) coincides with $Ker(Q)$, where Q is the matrix given in (7) which is the coefficient matrix of the linear application $\mathcal{H}$ between $\mathbb{R}^D$ and $\mathbb{R}^{D/2}$. Since rank$(Q) = D/2$ (it is enough to observe that the first $D/2$ columns of Q are the identity matrix), we have rank$(Ker(Q)) = D/2$. Now it is sufficient to observe that $\mathcal{B}_K$ is a set of $D/2$ linearly independent vectors. □

We observe that type-0 pmfs are characterized by the condition that only one of the components f_i or f_{D-i+1} of $\boldsymbol{f}_p$ can be different from zero. We can now classify the pmfs in $\mathcal{F}_d(p)$ as follows.

Definition 3.2. We say that $\boldsymbol{f} = (f_1, \dots, f_D)$ is of type-0 if it is the particular solution $\boldsymbol{f}_P$ corresponding to a polynomial $P(\boldsymbol{x}) \in \mathcal{I}_\mathcal{P}$, it is of type-1K if $\boldsymbol{f} \in Ker(\mathcal{H})$ and it is of type-1 otherwise.

Example 5. In the Example 1 the extremal point $\boldsymbol{r}_9 = (0, \frac{3}{10}, \frac{3}{10}, 0, \frac{3}{10}, 0, 0, \frac{1}{10})$ is of type-0 and $\boldsymbol{r}_5 = (\frac{3}{5}, 0, 0, 0, 0, 0, 0, \frac{2}{5})$ is of type-1K, while $\boldsymbol{r}_6 = (0, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5}, 0, 0, 0)$ is of type-1, since $f_4 = f_{8-3} = f_5$. Notice that $\mathcal{H}(\boldsymbol{r}_6) = 1/5(1 - x_1 + x_2 - x_1x_2)$ and $\mathcal{H}(\boldsymbol{r}_5) = 0$.

Proposition 3.2. *The Gröbner basis of $\mathcal{I}_{\mathcal{P}}$ with respect to the lexicographic order is $GB = \{G_i(\boldsymbol{x}) = (x_i - 1)(x_i + c), G_{ik}(\boldsymbol{x}) = 1 - x_i - x_x + x_i x_k, i, k =, \dots, 2^{d-1}, i < k\}$. A basis of the quotient space $\mathbb{Q}[x_1, \dots, x_{d-1}]/\mathcal{I}_{\mathcal{P}}$ is $\{1, x_1, \dots, x_{d-1}\}$.*

Proof. Let $\mathcal{I}_{\mathcal{P}} = \langle G_i(\boldsymbol{x}) \rangle$. We prove that the variety $V(\mathcal{I}_{\mathcal{P}})$ of $\mathcal{I}_{\mathcal{P}}$ is $\mathcal{P}$. $\mathcal{P} \subset V(\mathcal{I}_{\mathcal{P}})$ is easy to verify, so we prove the other inclusion. If $\boldsymbol{x} \in V(\mathcal{I}_{\mathcal{P}})$, then $x_j = 1$ or $x_j = -c$. In fact if there is $k : x_k \neq 1, -c$ then $G_k(\boldsymbol{x}) = (x_k - 1)(x_k + c) \neq 0$. If $x_k = -c$ then $x_j = 1, \forall j \neq k, j = 1, \dots, d-1$. In fact if there is $i : x_i = -c$, then $G_{ik}(\boldsymbol{x}) = x_i x_k - x_1 - x_k + 1 \neq 0$ and $\boldsymbol{x} \notin V(\mathcal{I}_{\mathcal{P}})$. Thus $\boldsymbol{x} = \boldsymbol{1}_{d-1}$ or $\boldsymbol{x} = \boldsymbol{c}_j, j = 1, \dots d-1$ that is $\boldsymbol{x} \in \mathcal{P}$.

GB is a Gröebner basis for $\mathcal{I}_{\mathcal{P}}$. Let $A(\boldsymbol{x}) \in \mathcal{I}_{\mathcal{P}}$ and let $\tilde{a}x^{\alpha}$ its leading term. If $x_i^2 \nmid \tilde{a}x^{\alpha}$ and $x_i x_j \nmid \tilde{a}x^{\alpha}$, that is $x_i x_j$ does not divide $\tilde{a}x^{\alpha}$, for any i, j then $\tilde{a}x^{\alpha} = \tilde{a}x_k$ for a $k = 1, \dots d-1$. Thus $A(\boldsymbol{x})$ has degree one and cannot be zero both in $\boldsymbol{c}_k$ and $\boldsymbol{1}_{d-1}$. Thus, since $A(\boldsymbol{x}) \in \mathcal{I}_{\mathcal{P}}$, its leading term $LT(A(\boldsymbol{x}))$ must be divisible by one of the $LT(G(\boldsymbol{x})), G(\boldsymbol{x}) \in GB$ and GB is a Gröebner basis. □

Proposition 3.3. *For each $n \in \mathbb{N}$, $2 \le n \le d-1$, the monomials $\pi_{j_1,\dots,j_n}(\boldsymbol{x}) = \prod_{i=1}^n x_{j_i}$ have remainder $R_{j_1,\dots,j_n}(\boldsymbol{x}) = \sum_{i=1}^n x_{j_i} - (n-1)$.*

Proof. Without loss of generality we consider the monomials $\pi_n(\boldsymbol{x}) := \pi_{1,\dots,n}(\boldsymbol{x}) = x_1 \cdots x_n$ and $R_n(\boldsymbol{x}) := R_{1,\dots,n}(\boldsymbol{x})$. We prove this proposition by induction. If $n = 2$ $G_{ij}(\boldsymbol{x}) = x_i x_j - x_i - x_j + 1 \in \mathcal{I}_{\mathcal{P}}$ and therefore $x_i x_j$ has the remainder $x_i + x_j - 1$. We now prove that $n \Rightarrow n+1$. By inductive hypothesis

$$\prod_{i=1}^{n+1} x_j = x_{n+1} \prod_{j=1}^{n} x_j = x_{n+1}(I(x) + \sum_{j=1}^{n} x_j - (n-1)),$$

where $I(\boldsymbol{x}) = x_1 \cdot \ldots \cdot x_n - \sum_{i=1}^n x_i + n - 1 \in \mathcal{I}_{\mathcal{P}}$. Thus

$$\begin{aligned}\prod_{i=1}^{n+1} x_j &= x_{n+1}(I(\boldsymbol{x}) + \sum_{j=1}^{n} x_j - (n-1)) \\ &= x_{n+1} I(x) + \sum_{j=1}^{n} x_{n+1} x_j - x_{n+1}(n-1) \\ &= \tilde{I}(x) + \sum_{j=1}^{n} (G_{n+1,j}(\boldsymbol{x}) + x_{n+1} + x_j - 1) - x_{n+1}(n-1),\end{aligned}$$

where $\tilde{I}(\boldsymbol{x}) \in I$. It follows that

$$\begin{aligned}\prod_{i=1}^{n+1} x_j &= \tilde{I}(x) + \sum_{j=1}^{n} G_{n+1,j}(\boldsymbol{x}) + n x_{n+1} + \sum_{j=1}^{n} x_j - n - n x_{n+1} + x_{n+1} \\ &= \tilde{I}^*(x) + \sum_{j=1}^{n} x_j - n + x_{n+1} = \tilde{I}^*(x) + \sum_{j=1}^{n+1} x_j - n,\end{aligned}$$

where $\tilde{I}^*(\boldsymbol{x}) = \tilde{I}(x) + \sum_{j=1}^{n} G_{n+1,j}(\boldsymbol{x}) \in I$, and the assert is proved. □

We call the polynomials

$$F_{j_1,\dots,j_n}(\boldsymbol{x}) = \pi_{j_1,\dots,j_n}(\boldsymbol{x}) - R_{j_1,\dots,j_n}(\boldsymbol{x}) = x_{j_1}\cdots x_{j_n} - \sum_{i=1}^{n} x_{j_i} + n - 1$$

fundamental polynomials of the ideal $I_{\mathcal{P}}$. In particular we denote by $F_n(\boldsymbol{x})$ the fundamental polynomial $F_n(\boldsymbol{x}) := F_{1,\dots,n}(\boldsymbol{x})$, $n = 2, \dots, d-1$. The proof of Corollary 3.1 is in the supplemetary material [15].

Corollary 3.1. *The polynomials $P_f(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x})\boldsymbol{f} \in C_{\mathcal{H}}$ are linear combinations of the fundamental polynomials. In particular they have the form $P_f(\boldsymbol{x}) = Q(\boldsymbol{x}) - R(\boldsymbol{x})$, where*

$$Q(\boldsymbol{x}) = \sum_{k=2,\dots,d-1, j_1<j_2\dots<j_k} a_{j_1\dots j_k} x_{j_1}\dots x_{j_k}$$

and $R(\boldsymbol{x}) = b_0 + \sum_{j=1}^{d-1} b_j x_j$ is the remainder.

Notice that the polynomials in $C_{\mathcal{H}}$ are the same for each marginal probability p. The probability p defines the points $\mathcal{P}$ of the variety $V(I_{\mathcal{P}})$. We have proved that the set of fundamental polynomials is a set of generators of $C_{\mathcal{H}}$ and therefore we can use them to generate $\mathcal{F}_d(p)$. All pmfs can be generated as linear combinations of the fundamental polynomials, computing the corresponding type-0 pmf and eventually adding an element of $Ker(\mathcal{H})$. We now use fundamental polynomials to address two open issues. The first is the search of the extremal generators of the convex polytope $\mathcal{F}_d(p)$ and the second is the study of bounds for the distributions in $\mathcal{F}_d(p)$. We provide an algorithm to find extremal points, all in principle and many in practice. We find an analytical solution of the second problem under the convex order of sums.

4. An algorithm to find extremal pmfs

The first step to construct an algorithm for generating extremal pmfs is to prove that type-0 pmfs associated to fundamental polynomials are extremal themselves.

Proposition 4.1. *Let $\boldsymbol{f} \in \mathcal{F}_d(p)$ be the type-0 pmf associated to a fundamental polynomial. The pmf $\boldsymbol{f}$ is an extremal probability mass function.*

Proof. Let $\boldsymbol{f}$ such that $\mathcal{H}(\boldsymbol{f}) = F_n(\boldsymbol{x}), 2 \le n \le d-1$ is a fundamental polynomial, $F_n(\boldsymbol{x}) = \prod_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i + (n-1)$. Then $\boldsymbol{f}$ has support on $n+2$ points $\text{supp}(\boldsymbol{f}) = \{i_1, \dots, i_{n+2}\} \subset \{1, \dots, D\}$, corresponding to the monomials $-x_j - \frac{2s-t}{s}, j = 1, \dots, n$, $\prod_{i=1}^{n} x_i$ and the constant term. Let $I^* \in \mathcal{M}((2^d - (n+2)) \times 2^d)$ be the submatrix of I_D so that $I^*\boldsymbol{f} = 0$ (it contains the rows $I_D[i,]$ of I_D such that $f_i = 0$). Let us define the matrix

$$H/I := H//I^* = \begin{bmatrix} H \\ I^* \end{bmatrix}.$$

Let us consider the columns of H/I. The pmf $\boldsymbol{f}$ has mass on $n+2$ points and the corresponding $n+2$ columns $H/I[,j]$ of H/I, that we call $H/I_j, j = i_1, \ldots, i_{n+2}$ have all zeros below the rows of H,

$$H/I_j = \begin{bmatrix} H_j \\ \mathbf{0} \end{bmatrix},$$

where $\mathbf{0}$ is the $2^d - (n+2)$-vector with all zeros. Among them, we consider the columns corresponding to $-x_j - \frac{2s-t}{s}, j = 1, \ldots, n$ and constant term. These are $n+1$ independent columns of H/I because it can be proved that $\sum_{j=1}^{n} \gamma_j(-x_j - \frac{2s-t}{s}) + \gamma_{n+1} = 0$ for $\boldsymbol{x} \in \mathcal{P}$ iff $\gamma_1 = \ldots = \gamma_{n+1} = 0$. The column corresponding to $\prod_{i=1}^{n} x_i$ is linearly dependent by these $n+1$ independent columns of H/I because $F_n(\boldsymbol{x})$ belongs to the ideal $\mathcal{I}_{\mathcal{P}}$, that is $F_n(\boldsymbol{x}) = 0,\ \boldsymbol{x} \in \mathcal{P}$. It follows that $\prod_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i - n + 1$ and then $\prod_{i=1}^{n} x_i = -\sum_{i=1}^{n}(-x_i - \frac{2s-t}{s}) - n\frac{2s-t}{s} - n + 1$. The $2^d - (n+2)$ remaining columns $H/I_j, j \in \{1, \ldots 2^d\}, j \neq i_1, \ldots, i_{n+2}$ of H/I are independent because $I^*[, j \in \{1, \ldots 2^d\}, j \neq i_1, \ldots, i_{n+2}]$ is an identity matrix. Therefore we have $2^d - (n+2) + n + 1 = 2^d - 1$ independent columns. It follows that

$$\operatorname{rank}(H/I) = 2^d - 1.$$

From Lemma 2.3 in [31], $\boldsymbol{f}$ is an extremal point. □

Example 6. Let us consider $d = 3$, $p = 2/5$ and $F_2(x_1, x_2) = x_1x_2 - x_1 - x_2 + 1$. As shown in Example 4, the corresponding pmf is $\boldsymbol{f} = (\frac{2}{5}, 0, 0, \frac{1}{5}, 0, \frac{1}{5}, \frac{1}{5}, 0)$ and it is an extremal pmf. The H/I matrix is

$$H/I = \begin{pmatrix} 1 & -\frac{3}{2} & 1 & -\frac{3}{2} & 1 & -\frac{3}{2} & 1 & -\frac{3}{2} \\ 1 & 1 & -\frac{3}{2} & -\frac{3}{2} & 1 & 1 & -\frac{3}{2} & -\frac{3}{2} \\ 1 & 1 & 1 & 1 & -\frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} & -\frac{3}{2} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and $\operatorname{rank}(H/I) = 2^3 - 1 = 7$.

Analogously, we can be prove the following.

Corollary 4.1. *If $\mathcal{H}(\boldsymbol{f}) = -F(\boldsymbol{x})$, where $F(\boldsymbol{x})$ is a fundamental polynomial, then $\boldsymbol{f}$ is an extremal mass function.*

Remark 2. The extremal points associated to $F(\boldsymbol{x})$ and $-F(\boldsymbol{x})$ are different and they are not symmetric. We can see from Example 1 that $\mathcal{H}(\boldsymbol{r}_4) = x_1x_2 - x_1 - x_2 + 1$ and $\mathcal{H}(\boldsymbol{r}_9) = -x_1x_2 + x_1 + x_2 - 1$. This is a consequence of the role played by the constant terms arising from the monomials with negative coefficients, i.e. the monomials in $\boldsymbol{m}_-(\boldsymbol{x})$.

We proved that to find a point in $\mathcal{F}_d(p)$ it is sufficient to pick a polynomial in $C_{\mathcal{H}}$. From Corollary 3.1 we have that any $P_f(\boldsymbol{x}) \in C_{\mathcal{H}}$ can be obtained as a linear combination of fundamental polynomials. Proposition 2.1 gives a necessary condition for a point to be an extremal point, it must have support on at most $d+1$ points. Therefore to find an extremal point not associated to a fundamental polynomial we have to choose a linear combination of at most $d+1$ fundamental polynomials, find a corresponding pmf in $\mathcal{F}_d(p)$ and finally check if it is an extremal point by means of Lemma 2.3 in [31].

Proposition 3.1 provides a way to easily build extremal points of type-1 as $\boldsymbol{f}+\boldsymbol{e}$ where $\boldsymbol{f}$ is a type-0 pmf and $\boldsymbol{e}\in\ker(\mathcal{H})$. It is enough to choose $\boldsymbol{e}$ in a way that all the components of $\boldsymbol{f}+\boldsymbol{e}$ are non-negative and then the corresponding type-1 pmf are obtained normalizing $\boldsymbol{f}+\boldsymbol{e}$. To find type-0 extremal points not associated to the fundamental polynomials is more challenging. We show an algorithm to find type-0 extremal points which requires the construction of a matrix whose columns are the coefficients of the remainders of the monomials of degree greater than one, $x_{j_1}\dots x_{j_k}$, with $k\geq 2$.

Let $\boldsymbol{\pi}=(x_1x_2,\dots,x_1\dots x_{d-1})^T$ be the row vector of the $2^{d-1}-d$ monomials of degree greater than one ordered according to the reverse-lexicographical criterion. Let B be the matrix whose elements of the j-th column are the coefficients of the remainders of $\boldsymbol{\pi}_j$ corresponding to the basis $\{1,x_1,\dots,x_{d-1}\}$ of the quotient space, as illustrated in Table 4. We look for a (column) vector $\boldsymbol{a}=(a_{12},\dots,a_{12,\dots d-1})$, so that

$$P(\boldsymbol{x})=\sum_{k=2}^{d-1}\sum_{j_1<\ldots<j_k}a_{j_1\ldots j_k}F_{j_1\ldots j_k}(\boldsymbol{x})=b_\varnothing+\sum_{k=1}^{d-1}b_kx_k+\sum_{k=2}^{d-1}\sum_{j_1<\ldots<j_k}a_{j_1\ldots j_k}x_{j_1}\dots x_{j_k}$$

is associated to an extremal point. The term $-(b_\varnothing+\sum_{k=1}^{d-1}b_kx_k)$ is the remainder of

$$\sum_{k=2}^{d-1}\sum_{j_1<\ldots<j_k}a_{j_1\ldots j_k}x_{j_1}\dots x_{j_k}.$$

By construction, the product $B\boldsymbol{a}$ is the column vector of the coefficients of the remainder of $\sum_{k=2}^{d-1}\sum_{j_1<\ldots<j_k}a_{j_1\ldots j_k}x_{j_1}\dots x_{j_k}$, that is $B\boldsymbol{a}=-(b_\varnothing,b_1,\dots,b_{d-1})$. If the k-th row of $B\boldsymbol{a}$ is zero, the k-th term of the remainder of $\sum_{k=2}^{d-1}\sum_{j_1<\ldots<j_k}a_{j_1\ldots j_k}x_{j_1}\dots x_{j_k}$ is zero. The k-th row of $(B\boldsymbol{a})_{k\cdot}$ is $B_{k\cdot}\boldsymbol{a}$, where $B_{k\cdot}$ is the k-th row of B, thus $B_{k\cdot}\boldsymbol{a}$ is $-b_{k-1}$, and $B_{1\cdot}\boldsymbol{a}$ is $-b_\varnothing$.

The solutions of $B_{k\cdot}\boldsymbol{a}=0$ give the coefficients of all the polynomials $P(\boldsymbol{x})$ which are associated to pmfs in $\mathcal{F}_d(p)$ without the components corresponding to x_{k-1} and $-x_{k-1}+\frac{2s-t}{s}$. We observe that $b_\varnothing$ is not immediately related to f_1 and f_D. For example, in the case $d=4$, $p=2/5$, the polynomial $\frac{1}{5}x_1x_2+\frac{1}{5}x_1x_3+\frac{1}{5}x_2x_3-\frac{2}{5}x_1x_2x_3-\frac{1}{5}$ corresponds to the pmf $\boldsymbol{f}=(0,0,0,\frac{1}{5},0,\frac{1}{5},\frac{1}{5},0,\frac{2}{5},0,0,0,0,0,0,0)$. We have $b_\varnothing=-1/5\neq 0$ and $f_1=f_{16}=0$. Or the polynomial $x_1x_2-x_1x_3+x_2+x_3$ corresponds to the pmf $\boldsymbol{f}=(\frac{1}{5},0,0,\frac{1}{5},\frac{1}{5},0,0,0,0,0,\frac{1}{5},0,0,\frac{1}{5},0,0)$. We have $b_\varnothing=0$ and $f_1=1/5\neq 0$. For this reason we do not consider the equation $B_{1\cdot}\boldsymbol{a}=0$. Let us suppose that we are interested in polynomials $P(\boldsymbol{x})$ where only some $a_{j_1\ldots j_k}$ can be different from zero. We define J as the corresponding set of indexes, $j_1\dots j_k\in J\leftrightarrow a_{j_1\ldots j_k}\neq 0$. Let $B_{\cdot J}$ be the matrix whose columns are $B_{\cdot j}$, $j\in J$ and $\boldsymbol{a}_J$ the sub-vector of $\boldsymbol{a}$ whose elements are $a_j, j\in J$. The elements of the kernel of the linear application

$$(B_{\cdot J}\boldsymbol{a}_J)_{K\cdot},$$

where $(B_{\cdot J}\boldsymbol{a}_J)_{K\cdot}$ are the rows $(B_{\cdot J}\boldsymbol{a}_J)_{k\cdot}$ of $(B_{\cdot J}\boldsymbol{a}_J)$, $k\in K\subseteq\{2,\dots,d\}$, are the coefficients $\boldsymbol{a}$ such that the polynomial $P(\boldsymbol{x})$ does not have the terms x_{k-1}, $k\in K\subseteq\{2,\dots,d\}$.

The polynomials $P(\boldsymbol{x})$ associated to extremal points must have at most $d+2$ non-zero coefficient ($d+2$ because the constant term $b_\varnothing$ does not always correspond to have f_1 or f_D in the support of the corresponding pmf). Therefore, we choose $\#J$ monomials in $\boldsymbol{\pi}$, with $\#J\leq d+2$, that means $\#J$ columns of the matrix in Table 4. Then we look for the polynomials $P(\boldsymbol{x})$ so that the remainder has $1,\dots,d-1$ terms equal zero. Formally we have the following steps:

1. choose J and K with $\#J+d-\#K\leq d+2$;
2. find a set of kernel generators of

$$(B_{J\cdot}\boldsymbol{a}^T)_{K\cdot};$$

Table 4. Matrix representation of the remainders of the monomials $x_{j1} \dots x_{jk}$, general case.

$$\frac{\quad | \ \boldsymbol{\pi}(x)}{\boldsymbol{R}(x) \ | \quad \mathrm{B}} = \begin{array}{c|cccc} & x_1x_2 & x_1x_3 & \cdots & x_1 \cdots x_{d-1} \\ \hline 1 & -1 & -1 & \dots & -(d-1) \\ x_1 & 1 & 1 & \dots & 1 \\ x_2 & 1 & 0 & \dots & 1 \\ \dots & \dots & & & \\ x_{d-1} & 0 & 0 & 0 & 1 \end{array}$$

3. for each kernel generator $\boldsymbol{a}^{(i)}$ consider the corresponding polynomial $P^{(i)}(\boldsymbol{x})$;
4. construct the corresponding pmf in $\mathcal{F}_d(p)$;
5. check if it is an extremal point of the polytope.

The above steps find all the type-0 extremal points corresponding to each generator of the set of kernel generators found in point 3. If we use linear combinations of the kernel generators we can potentially find all the type-0 extremal points, but an efficient choice of a good linear combination is not considered in this algorithm and it will be part of our future research. It is worth noting that the set of generators does not depend on p (up to point 3 the algorithm is independent of p). Then the kernel generators and the corresponding polynomials can be computed once for all. Another positive aspect is that the available algorithms for finding set of generators of linear system kernels are extremely efficient. On the other hand if the dimensionality d increases the computational effort increases also for the high number of choices of J and K in step 1.

To the only purpose of illustrating the procedure, in Example 7 we look for the type-0 extremal points for $d = 4$, because the case $d = 3$ has only two fundamental polynomials (opposite signs) and it is trivial.

Example 7. Consider $\mathcal{F}_4(2/5)$, $d = 4$ and $p = 2/5$.

The extremal pmfs have support on at most $d + 1 = 5$ points. Since the remainder has 4 terms, if we decide to combine two fundamental polynomials we have to eliminate at least one monomial of the remainder. By so doing, the remainder has at most 3 terms and the polynomial has at most 5 coefficients. As an example we choose the first two columns, i.e. $J = \{1,2\}$. We have $\boldsymbol{a} = (a_{12}, a_{13})$ and

$$B_{\cdot J} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Table 5. Matrix representation of the remainders of the monomials x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$, case $d = 4$.

$$\frac{\quad | \ \boldsymbol{\pi}(x)}{\boldsymbol{R}(x) \ | \quad \mathrm{B}} = \begin{array}{c|cccc} & x_1x_2 & x_1x_3 & x_2x_3 & x_1x_2x_3 \\ \hline 1 & -1 & -1 & -1 & -2 \\ x_1 & 1 & 1 & 0 & 1 \\ x_2 & 1 & 0 & 1 & 1 \\ x_3 & 0 & 1 & 1 & 1 \end{array}$$

We have

$$B_{\cdot J}\boldsymbol{a} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{a} = \begin{pmatrix} -a_{12} - a_{23} \\ a_{12} + a_{23} \\ a_{12} \\ a_{23} \end{pmatrix}.$$

We look for the polynomial without the term corresponding to x_1, that is $K = 2$. These are the solutions of the homogeneous system:

$$(B_{\cdot J}A^T)_2 = a_{12} + a_{23} = 0.$$

We can choose $\boldsymbol{a} = (1,-1)$ and we obtain $P(x) = x_1x_2 - x_1x_3 + x_2 - x_3$. The associated pmf is $(\frac{1}{5},0,0,\frac{1}{5},\frac{1}{5},0,0,0,0,0,0,\frac{1}{5},0,0,\frac{1}{5},0,0)$ and we can verify that is an extremal point.

In the Section 5 we use fundamental polynomials to find a pmf in $\mathcal{F}_d(p)$ that satisfies a given condition. We show that it is convenient to set the condition in terms of coefficient of a polynomial in C_H and find one corresponding pmf.

5. Lower bounds for the convex order

Finding upper and lower bounds for sums $S = X_1 + \ldots + X_d$ of random variables X_i of which the marginal distributions are known but the joint distribution is unspecified is a problem extensively addressed in statistics and applied probability [19]. These bounds are linked to the highest and lowest dependence structure and they are actually bounds in the sense of the convex order. We recall the definition of the convex order.

Definition 5.1. Given two random variables X and Y with finite means, X is said to be smaller than Y in the convex order (denoted $X \leq_{cx} Y$) if

$$E[\phi(X)] \leq E[\phi(Y)]$$

for all real-valued convex functions ϕ for which the expectations exist.

The convex order is a variability order, in fact it is easy to verify that $X \leq_{cx} Y$ implies $E[X] = E[Y]$, and $V[X] \leq V[Y]$. It can also be proved, see e.g. [29], that

$$X \leq_{cx} Y \text{ iff } E[X] = E[Y] \text{ and } E[(X-l)^+] \leq E[(Y-l)^+] \text{ for all } l \in \mathbb{R}_+,$$

where $x^+ = max\{x,0\}$. The last inequality defines the so called stop-loss order, that is important in insurance. See [7] for a discussion on the relationship between convex and stop-loss orders. We look for the minimum convex order for sums of Bernoulli variables with mean p, when the joint distribution is unspecified. The problem to find the upper bound is solved and it is well known that the upper bound is the upper Fréchet bound of the class, that is the extremal point $\boldsymbol{r}^U$ with support on the two points $(0,\ldots,0)$ and $(1,\ldots,1)$. We look for $\boldsymbol{X} \in \mathcal{F}_d(p)$ such that $S = \sum_{i=1}^d X_i$ is minimal in the sense of the convex order. Following the notation in [26], we call the pmf $\boldsymbol{f}$ of a vector $\boldsymbol{X}$ in $\mathcal{F}_d(p)$ whose sums are minimal in the sense of the convex order a Σ_{cx}-smallest element of $\mathcal{F}_d(p)$. Since we consider $\boldsymbol{X} \in \mathcal{F}_d(p)$ the sums have all the same mean dp, thus our problem reduce to find $\boldsymbol{X}^*$ such that for any $\boldsymbol{Y} \in \mathcal{F}_d(p)$

$$E[(S^* - l)^+] \leq E[(S_{\boldsymbol{Y}} - l)^+], \ \ \forall l \in \mathbb{R}_+, \tag{8}$$

where $S^* = \sum_{i=1}^d X_i^*$ and $S_{\boldsymbol{Y}} = \sum_{i=1}^d Y_i$

Let $\mathcal{D}(dp)$ be the class of discrete distributions on $\{0, \dots, d\}$ with mean dp, clearly $S \in \mathcal{D}(dp)$. The paper [11] proves that the class of sums of exchangeable Bernoulli distributions with the same mean p coincides with the entire class of discrete distributions with mean dp, $\mathcal{D}_d(dp)$. Therefore the map

$$H : \mathcal{F}_d(p) \to \mathcal{D}(dp)$$
$$\boldsymbol{f} \to \boldsymbol{p},$$

where $\boldsymbol{p}$ is the pmf of $S = \sum_{i=1}^d X_i$, is onto on $\mathcal{D}(dp)$.

Thanks to the above result we can look for the convex order bounds in $\mathcal{D}_d(dp)$ to find firstly the bounds for the sums and then the corresponding multivariate Bernoulli distributions. Formally, we look for S^* such that for any $S \in \mathcal{D}(dp)$

$$E[(S^* - l)^+] \le E[(S - l)^+], \ \forall l \in \mathbb{R}_+. \tag{9}$$

Then, by means of the results in previous sections, we characterize the points $\boldsymbol{X} \in \mathcal{F}_d(p)$ so that $\sum_{i=1}^d X_i = S^*$. In [11] the authors prove that the class $\mathcal{D}_d(dp)$ is a convex polytope and they explicitly found its generators. This result is stated in Proposition 5.1.

Proposition 5.1. *The extremal pmfs of $\mathcal{D}(dp)$, s_{j_1,j_2} have support on two points (j_1, j_2) with $j_1 = 0, 1, \dots, j_1^M$, $j_2 = j_2^m, j_2^m + 1, \dots, d$, j_1^M is the largest integer less than pd and j_2^m is the smallest integer greater than pd. They are*

$$s_{j_1,j_2}(y) = \begin{cases} \frac{j_2 - pd}{j_2 - j_1} & y = j_1 \\ \frac{pd - j_1}{j_2 - j_1} & y = j_2 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

If pd is integer the extremal densities contain also

$$s_{pd}(y) = \begin{cases} 1 & y = pd \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

If pd is not integer there are $n_p = (j_1^M + 1)(d - j_1^M)$ extremal points. If pd is integer there are $n_p = d^2 p(1 - p) + 1$ extremal points.

For any ϕ the extremal values for $E[\phi(S)]$ are reached on the extremal points (see [14]). Thus, the bounds for the convex order are reached on the extremal points. In particular, the upper bound of convex order is on the extremal point $s^U = s_{0,d}$. This is a straightforward consequence of the well known fact that the upper bound for $S = \sum_{i=1}^d X_i$ is the upper Frechét bound $\boldsymbol{r}^U \in \mathcal{F}_d(p)$ and $S^U = \sum_{i=1}^d R_i^U$ with pmf $s_{0,d}$. On the contrary the lower bound in $\mathcal{F}_d(p)$ is still an open issue. [8] proved that if $pd < 1$ the lower Fréchet bound belongs to $\mathcal{F}_d(p)$ and corresponds to the lower bound for convex order. Here we generalize their result for each p and d. [16] found the lower bound in the subclass of exchangeable pmfs. We first find the solution of Equation (9) and then using the algebraic representation we find a corresponding pmf in $\mathcal{F}_d(p)$. The proofs of Proposition 5.2 and 5.3 are in the supplementary material.

Proposition 5.2. *The solution of equation (9), i.e.*

$$E[(S^* - l)^+] \le E[(S - l)^+], \ \forall l \in \mathbb{R}_+,$$

for any $S \in \mathcal{D}(dp)$, is $S^* = S_{j^M, j^m}$, where S_{j^M, j^m} is the random variable with pmf s_{j^M, j^m}.

Therefore the solutions of (8), that is

$$E[(S^* - l)^+] \le E[(S - l)^+], \ \forall l \in \mathbb{R}_+,$$

for any $S \in \mathcal{D}(dp)$, are the pmfs of $\boldsymbol{X}$ with $\boldsymbol{X}$ in $\chi^* = \{\boldsymbol{X} \in \mathcal{F}_d(p) : \sum_{i=1}^d X_i = S^* = S_{j^M, j^m}\}$. As usual with a small abuse of notation if $\boldsymbol{f}$ is the vector pmf of $\boldsymbol{X} \in \chi^*$ we also write $\boldsymbol{f} \in \chi^*$. A pmf $\boldsymbol{f} \in \chi^*$ is a Σ_{cx}-smallest element in $\mathcal{F}_d(p)$.

Proposition 5.3. *If $\boldsymbol{X} \in \mathcal{X}^*$, then its pmf $\boldsymbol{f}$ is such that*

$$\boldsymbol{f} = \sum_{i_1}^{n*} \lambda_i \boldsymbol{r}_i^*, \ \lambda_i \neq 0,$$

where $\boldsymbol{r}_i^*$ are the extremal points of $\mathcal{F}_d(p)$ in $\mathcal{X}^*$.

Remark 3. By Jensen's inequality it is known that for any $l \in \mathbb{R}^+$, $\min_{S \in \mathcal{D}(pd)} E[(S-l)^+] \ge (pd-l)^+$. The proof of Proposition 5.2 shows that this bound is not sharp for all the values of $l \in \mathbb{R}^+$, namely $l \in (j^M, j^m)$. Indeed, if $l \le j^M$ we have $\min_{S \in \mathcal{D}(dp)} E[(S-l)^+] = pd - l$, if $l \in (j^M, j^m)$ we have $\min_{S \in \mathcal{D}(dp)} E[(S-l)^+] = (pd - j^M)(j^m - l)$ and if $l \ge j^m$ we have $\min_{S \in \mathcal{D}}(dp) E[(S-l)^+] = 0$. In the supplementary material (S.1) we provide an example where we explicitly compare the sharp bound with Jensen's one.

We therefore look for the multivariate Bernoulli variables $\boldsymbol{X} \in \mathcal{F}_d(p)$ such that $P(S_{\boldsymbol{X}} = k) = 0$ for $k \neq j^M, j^m$, where $S_{\boldsymbol{X}} = \sum_{i=1}^d X_i$. This is equivalent to look for the pmfs with support $\chi_M \cup \chi_m$, where $\chi_k = \{x \in \chi : \sum_{i=1}^d x_i = k\}$.

We have proved that minimal convex sums correspond to a family χ^* of multivariate Bernoulli pmf. In [11] the authors prove that there is a one-to-one map between $\mathcal{D}_d(dp)$ and the class of exchangeable Bernoulli distributions with mean p, $\mathcal{E}_d(p) \subseteq \mathcal{F}_d(p)$. Therefore, there is exactly one pmf in $\boldsymbol{f}^* \in \mathcal{E}_d(p)$ with minimal convex sums. It is the pmf of the unique exchangeable random vector $\boldsymbol{X}^*$ such that $\sum_{i=1}^d X^* = S^*$, then we have $\boldsymbol{X}^* \in \chi^*$. In [11] the authors also proved that the exchangeable pmf $\boldsymbol{f}^*$ associated to S^* is the pmf with minimal correlation. Therefore, under exchangeability, minimal convex sums corresponds to minimal correlation.

Proposition 5.4 proves that if $\boldsymbol{f} \in \chi^*$ then its mean correlation, i.e. the mean of the correlations $\rho(X_i, X_j)$ of each pair of variables X_i and X_j, $i, j = 1, \dots, d$, $i < j$, is constant and equal to the correlation of $\boldsymbol{f}^*$. Therefore the extremal points belonging to χ^* generate the pmf with the lower mean correlation. In particular, Proposition 5.4 states that the sum of the second-order crossed moments of $\boldsymbol{X}$ is equal to the 2nd binomial moment of S, i.e. $S_2 = E[\binom{S}{2}]$. Proposition 5.5 and Corollary 5.1 are known results, see e.g. [2]. Nevertheless we report a proof of Proposition 5.4 in the supplementary material.

Proposition 5.4. *Let $f \in \mathcal{F}_d$ be the pmf of a d-dimensional multivariate Bernoulli random variable $\boldsymbol{X} = (X_1, \dots, X_d)$ and p_S the pmf of the sum $S = \sum_{i=1}^d X_i$, with $p_S(k) = p_k = P(S = k), k = 0, \dots, d$. The sum of the second-order crossed moments of $\boldsymbol{X}$ can be written as a linear combination of the values $p_k, k = 2, \dots, d$ of the pmf p_S of S:*

$$\sum_{1 \le i < j \le d} E[X_i X_j] = \sum_{k=2}^d \binom{k}{2} p_k.$$

Proposition 5.4 can be generalized to τ-order crossed moments, for $\tau = 2,\dots,d$ as stated in Proposition 5.5.

Proposition 5.5. *Let $f \in \mathcal{F}_d$ be the pmf of a d-dimensional multivariate Bernoulli random variable $\boldsymbol{X} = (X_1,\dots,X_d)$ and p_S the pmf of the sum $S = \sum_{i=1}^d X_i$, with $p_S(k) = p_k = P(S = k), k = 0,\dots,d$. The sum of the τ-order crossed moments of $\boldsymbol{X}$, $\tau \geq 2$, can be written as a linear combination of the values $p_k, k = \tau,\dots,d$ of the pmf p_S of S:*

$$\sum_{1\leq i_1<\ldots<i_\tau\leq d} E[X_{i_1}\cdots X_{i_\tau}] = \sum_{k=\tau}^{d} \binom{k}{\tau} p_k. \tag{12}$$

Corollary 5.1 of Proposition 5.4 characterizes the second order cross moments for pmfs of the sums of $\boldsymbol{X}$ with support included in $\{0,1\}$.

Corollary 5.1. *Let $f \in \mathcal{F}_d$ be the pmf of a d-dimensional multivariate Bernoulli random variable $\boldsymbol{X} = (X_1,\dots,X_d)$ and p_S the pmf of the sum $S = \sum_{i=1}^d X_i$, with $p_S(k) = p_k = P(S = k), k = 0,\dots,d$. If $p_k = 0$ for $k \geq 2$ then $E[X_iX_j] = 0$ for $1 \leq i < j \leq d$.*

The next Corollary provides the average second-order cross moment $\bar{\mu}_2$ of the sums of d-dimensional multivariate Bernoulli random variable $\boldsymbol{X} = (X_1,\dots,X_d) \in \mathcal{A}^*$.

Corollary 5.2. *Given S^*, a discrete random variable defined over $\{0,1,\dots,d\}$ with pmf p_S, $p_S(k) = p_k = P(S = k)$, $k = 0,\dots,d$, let $f \in \mathcal{F}_d$ be the pmf of a d-dimensional multivariate Bernoulli random variable $\boldsymbol{X} = (X_1,\dots,X_d) \in \mathcal{A}^*$, where $\mathcal{A}^* = \{\boldsymbol{X} \in \mathcal{F}_d : \sum_{i=1}^d X_i = S^*\}$. The average second-order cross moment $\bar{\mu}_2 = \frac{\sum_{1\leq i<j\leq d} E[X_iX_j]}{\binom{d}{2}}$ can be computed as*

$$\bar{\mu}_2 = \frac{1}{d(d-1)} \sum_{k=2}^{d} k(k-1)p_k.$$

We observe that for an exchangeable multivariate Bernoulli random variable the average second-order cross moment $\bar{\mu}_2$ coincides with any second-order cross moment $E[X_iX_j], 1 \leq i < j \leq d$.

Corollary 5.3. *If pd is not integer, given $S^* = S_{J^M,j^m} \in \mathcal{D}_d(dp)$, let $f \in \mathcal{F}_d$ be the pmf of a d-dimensional multivariate Bernoulli random variable $\boldsymbol{X} = (X_1,\dots,X_d) \in \mathcal{A}^*$, where $\mathcal{A}^* = \{\boldsymbol{X} \in \mathcal{F}_d : \sum_{i=1}^d X_i = S^*\}$. The average second-order cross moment $\bar{\mu}_2 = \frac{\sum_{1\leq i<j\leq d} E[X_iX_j]}{\binom{d}{2}}$ can be computed as*

$$\bar{\mu}_2 = \frac{1}{d(d-1)}(J^M(2pd - J^M - 1)). \tag{13}$$

If pd is integer we have $\bar{\mu}_2 = \frac{1}{d(d-1)}(pd(pd-1)) = \frac{1}{(d-1)}(p(pd-1))$.

Example 8. Let $p = 11/20$ and $d = 5$. We have $pd = 11/4 = 2.75$ and $J^M = 2$. Using Equation (13) we get $\bar{\mu}_2 = 0.25$. We also have $S^* = S_{J^M,j^m} \equiv S_{(2,3)}$. The pmf $\boldsymbol{p}_S = (p_0,\dots,p_5)$ of $S_{2,3}$ is defined as

$p_2 = 0.25$, $p_3 = 0.75$ and $p_k = 0$, $k = 0, 1, 4, 5$. It follows that all the 5-dimensional multivariate Bernoulli random variables $\boldsymbol{X} = (X_1, \ldots, X_5)$ such that $\sum_{i=1}^{5} X_i = S_{2,3}$ have the average of its second-order cross moments equal to 0.25.

The exchangeable case is quite simple because the geometrical representation is simpler than that of the general one. The generators of the exchangeable polytope are known in closed form and in a one-to-one correspondence with the generators of $\mathcal{D}_d(dp)$. Using the simple algebraic approach proposed here, we can solve the more challenging problem to explicitly find a polynomial corresponding to a non-exchangeable Bernoulli pmf $\boldsymbol{f} \in \chi^*$. This means that if $\boldsymbol{X}$ has pmf $\boldsymbol{f}$ then its sum $S = \sum_{i=1}^{d}$ has support on j^M, j^m. The vector $\boldsymbol{X}$ correspond to the minimum convex order and to the minimal mean correlation. Notice that pmfs corresponding to fundamental polynomials $F_{j_1,\ldots,j_n}(\boldsymbol{x})$ are not in χ^*, for $n \neq d-1$, because the pmfs of their sums have support on the three points: $\{0, n, d-1\}$. The pmf of the sum corresponding to F_{d-1} has support on $\{0, d-1, d-1\} \equiv \{0, d-1\}$ and then it is not of interest apart from the simple case $d = 2$. Similarly, pmfs corresponding to fundamental polynomials $-F_{j_1,\ldots,j_n}(\boldsymbol{x})$ are not in χ^*, for $n \neq d$, because their sums have support on three points: $\{1, d-n, d\}$. The pmf of the sum corresponding to $-F_{d-1}$ has support on $\{1, d-(d-1), d\} \equiv \{1, d\}$ and then it is not of interest apart from the simple case $d = 2$.

We would like to build a non-exchangeable random variable $\boldsymbol{X} \in \mathcal{F}_d(p)$ whose pmf $\boldsymbol{f}_\star$ has support only on the points $x \in \{0,1\}^d$ with $|x| = j^M$ or $|x| = j^m$. Consequently, the corresponding $\sum_{i=1}^{d} X_i$ has support only on j^M and j^m. A possible way for building $\boldsymbol{f}_\star$ is based on a well-known tool in algebraic statistics, see [9]. We build the exchangeable $\boldsymbol{f}_e$ corresponding to the discrete random variable S_{j^M, j^m}. This can be easily done taking into consideration Equation (10) and Equation (11). Then $\boldsymbol{f}_\star$ can be built as $\boldsymbol{f}_e + \epsilon \boldsymbol{m}$, $\epsilon \in (-1, +1)$, where the move m must satisfy the constraints $H\boldsymbol{m} = 0$, $\boldsymbol{f}_e + \epsilon \boldsymbol{m} \geq 0$, and $\sum_{i=1}^{D} (\boldsymbol{f}_e + \epsilon \boldsymbol{m})_i = 1$. The move $\boldsymbol{m}$ can be built using the Markov Basis of the matrix H. This method works well for small dimensions but become computationally unfeasible for large dimensions d. We now prove that making use of the polynomial structure of the generators of $\mathcal{F}_d(p)$ we find a method that provides an analytical solution of the problem and therefore works also for large dimensions. We find a non-exchangeable pmf $\boldsymbol{f} \in \chi^*$ as the type-0 pmf associated to a specific linear combination of fundamental polynomials, as stated in Theorem 5.2.

Theorem 5.2. *Let $p = s/t \leq 1/2$, $a = \frac{2s-t}{s}$, $a_1 = |2s - t| = t - 2s$ and $a_2 = s$. If $p = 1/2$ we obtain $a = 0$, $a_1 = 0$, $a_2 = 1$.*

1. *We first consider the case pd not integer.*
   a) *If $pd + p < j^m$ there are $\alpha_i, \beta_i \in \{0,1\}^d$ with $|\alpha_i| = j^M$, and $|\beta_i| = j^m$ such that the polynomial*

$$P_{d-j^M}(\boldsymbol{x}) = -a_2 \prod_{i=1}^{d-j^M} x_i + \sum_{i=1, |\alpha_i| = j^M}^{h} \boldsymbol{x}^{\alpha_i} + \sum_{i=1, |\beta_i| = j^m}^{k} \boldsymbol{x}^{\beta_i} - a_1,$$

where

$$k = a_2 d - 2a_2 j^M - a_1 j^M$$
$$h = a_1 + a_2 - k$$

belongs to the ideal $I_{\mathcal{P}}$ and the corresponding $\boldsymbol{X} \in \boldsymbol{F}_d(p)$ has sum $S^* = \sum_{i=1}^{d} X_i$ with support on $\{j^M, j^m\}$.

b) *If $pd+p \geq j^m$, there are $\alpha_i, \beta_i \in \{0,1\}^d$ with $|\alpha_i| = j^M$, and $|\beta_i| = j^m$ such that the polynomial*

$$P_{d-j^m}(\boldsymbol{x}) = -a_2 \prod_{i=1}^{d-j^m} x_i + \sum_{i=1, |\alpha_i|=j^M}^{h} \boldsymbol{x}^{\alpha_i} + \sum_{i=1, |\beta_i|=j^m}^{k} \boldsymbol{x}^{\beta_i} - a_1,$$

where

$$k = a_2 d - 2a_2 j^M - a_1 j^M - a_2$$
$$h = a_1 + a_2 - k$$

belongs to the ideal $I_{\mathcal{P}}$ and the corresponding $\boldsymbol{X} \in \mathcal{F}_d(p)$ has sum $S^* = \sum_{i=1}^d X_i$ with support on $\{j^M, j^m\}$.

2. *If pd is integer the polynomial*

$$P_{pd}(\boldsymbol{x}) = -a_2 \prod_{i=1}^{d-pd} x_i + \sum_{i=1, |\alpha_i|=j^M}^{a_1+a_2} \boldsymbol{x}^{\alpha_i} - a_1,$$

belongs to the ideal I and the corresponding $\boldsymbol{X} \in \mathcal{F}_d(p)$ has sum with support on $\{pd\}$.

Proof. Assume pd not integer. To simplify the notation let $j^M = m$. We have $j^m = m+1$ and $pd \in (m, m+1)$. The basic idea is to build $P(\boldsymbol{x})$, a linear combination of fundamental polynomials (which belongs to the ideal $I_{\mathcal{P}}$ by construction) and that could be rewritten as a polynomial $P_{d-m}(\boldsymbol{x})$ (or $P_{d-(m+1)}(\boldsymbol{x})$) whose corresponding type-0 pmf has support only on points $\boldsymbol{x}$, with $|\boldsymbol{x}| = m$ or $|\boldsymbol{x}| = m+1$. We consider two different cases, $pd + p < m+1$ and $pd + p \geq m+1$.

Let $pd + p < m+1$ and let $P(\boldsymbol{x})$ be the following linear combination of fundamental polynomials

$$P(\boldsymbol{x}) = -a_2 F_{d-m}(\boldsymbol{x}) + \sum_{i=1, |\alpha_i|=m}^{h} F_{\alpha_i}(\boldsymbol{x}) + \sum_{i=1, |\alpha_i|=m+1}^{k} F_{\beta_i}(\boldsymbol{x}),$$

where $F_{\alpha_i}(\boldsymbol{x}) \equiv F_{i_1,\ldots,i_n}(\boldsymbol{x})$, $i_1, \ldots, i_n$ are the positions where α_i is one, and $F_{\beta_i}(\boldsymbol{x})$ is similarly defined. We recall that $F_{i_1,\ldots,i_n}(\boldsymbol{x}) = x_{i_1} \cdots x_{i_n} - \sum_{j=1}^n x_{i_j} + (n-1)$. Clearly, $P(\boldsymbol{x}) \in I$.

We define $P_{d-m}(\boldsymbol{x}) = -a_2 \prod_{i=1}^{d-m} x_i + \sum_{i=1,|\alpha_i|=m}^{h} \boldsymbol{x}^{\alpha_i} + \sum_{i=1,|\beta_i|=m+1}^{k} \boldsymbol{x}^{\beta_i} - a_1$ and we look for h, k, α_i, β_i such that $P(\boldsymbol{x}) = P_{d-m}(\boldsymbol{x})$. It follows that:

1. the constant term of $P_{d-m}(\boldsymbol{x})$, that is $-a_1$ must be equal to the constant term of $P(\boldsymbol{x})$, that is $-a_2(d-m-1) + h(m-1) + km$;
2. because $P_{d-m}(\boldsymbol{x})$ has not linear terms, the linear terms of $P(\boldsymbol{x})$ must be zero. It follows that the (positive) linear terms of $-a_2 F_{d-m}(\boldsymbol{x})$ have to be cancelled by the (negative) linear terms of $F_{\alpha_i}(\boldsymbol{x})$ and $F_{\beta_i}(\boldsymbol{x})$. It is worth noting that in this proof given the term $\gamma_i x_i$, γ_i integer, we say that there are γ_i linear terms, i.e. we consider $\gamma_i x_i = \underbrace{x_i + \ldots + x_i}_{\gamma_i \text{times}}$ if $\gamma_i > 0$ and $\gamma_i x_i = \underbrace{-x_i - \ldots - x_i}_{\gamma_i \text{times}}$ if $\gamma_i < 0$.

Condition 1 is satisfied for any choice of $\alpha_i, i = 1, \ldots, k, |\alpha_i| = m$, and $\beta_i, i = 1, \ldots, k, |\beta_i| = m+1$ if h, and k are positive solutions of:

$$h(m-1) + km = a_2(d-m-1) - a_1.$$

To satisfy Condition 2 we first look for h, k such that the total number of linear terms in $F_{\alpha_i}(\boldsymbol{x})$ and $F_{\beta_i}(\boldsymbol{x})$ is equal to the total number of linear terms in $-a_2 F_{d-m}(\boldsymbol{x})$. Then we show that by properly choosing the h polynomials F_{α_i} of degree m and the k polynomials F_{β_i} of degree $m+1$ we can simplify all the linear terms in $P(\boldsymbol{x})$. Since all $F_{\alpha_i}(\boldsymbol{x})$ have the same number m of linear terms for any i, all $F_{\beta_i}(\boldsymbol{x})$ have the same number $m+1$ of linear terms for any i, and the number of linear terms of $-a_2 F_{d-m}(\boldsymbol{x})$ is $a_2(d-m)$, h and k must be positive solutions of:

$$hm + k(m+1) = a_2(d-m).$$

From Conditions 1 and 2 we have to find the solutions of

$$\begin{cases} h(m-1) + km = a_2(d-m-1) - a_1 \\ hm + k(m+1) = a_2(d-m). \end{cases}$$

Standard computations give

$$k = a_2 d - 2a_2 m - a_1 m$$
$$h = a_1 + a_2 - k.$$

We must check that solutions are positive integers. It is possible to verify that the solutions are integer since $(m-1)(m+1) - m^2 = -1$ and that the solutions are positive iff $pd + p < m + 1 = j^m$.

A possible not unique choice for α_i and β_i can be obtained using the following steps.

1. The a_2 copies of the linear terms $x_1, \ldots, x_{d-m}$ must be ordered, repeating the sequence "$x_1, \ldots, x_{d-m}$" a_2 times

$$\underbrace{x_1, \ldots, x_{d-m}}_{\text{1st time}}, \underbrace{x_1, \ldots, x_{d-m}}_{\text{2nd time}}, \ldots, \underbrace{x_1, \ldots, x_{d-m}}_{a_2\text{-th time}}. \tag{14}$$

2. The $\alpha_i, i = 1, \ldots, h$, $|\alpha_i| = m$ are determined. This is equivalent to build h mononomials of degree m. The first monomial is built as the product of the first m linear terms (i.e. variables) in the list shown in Equation (14), i.e. $x_1 \cdots x_m$, the second monomial with the subsequent m variables, i.e. $x_{m+1} \cdots x_{2m(mod(d-m))+1}$, and so on for the first h mononomials.
3. Then, in an analogous way, starting from the position $hm + 1$ to the end of the list shown in Equation (14) we build k monomials of degree $m + 1$.

Let us now consider the case $pd + p \geq j^m$. The proof is similar to the case $pd + p > j^m$ and we look for the positive solutions of

$$\begin{cases} h(m-1) + km = a_2(d-m-2) - a_1 \\ hm + k(m+1) = a_2(d-m-1). \end{cases}$$

Easy computations give

$$k = a_2 d - 2a_2 m - a_1 m - a_2$$
$$h = a_1 + a_2 - k.$$

If $pd + p \geq m + 1 = j^m$ then $pd - m \geq 1 - p$ and since $p < 1/2$, $pd - m > p$. It is easy to verify that if $pd - m > p$, h and k are both integers and positives. Notice that $p = 1/2$ implies $pd + p \geq j^m$ for $d \geq 2$, thus it is included in this case.

The case pd integer can be proved similarly by proving that $P_{pd}(\boldsymbol{x})$ is the following linear combination of fundamental polynomials:

$$P_{pd}(\boldsymbol{x}) = -a_2 F_{d-pd}(\boldsymbol{x}) + \sum_{i=1, |\alpha_i|=pd}^{a_1+a_2} F_{\alpha_i}(\boldsymbol{x}).$$

This completes the proof. □

The proof of Proposition 5.2 provides a way to find α_i and β_i with a very simple algorithm. Therefore, we can easily find a pmf $\boldsymbol{f} \in \mathcal{F}_d(p)$ minimal with respect to the convex order in any dimension d and for any p. From Lemma 2.3 in [31], we can also check if the density found is an extremal point - we know that the minimum is reached on at least one extremal point.

We conclude this section with some examples. Example 9 of $P_{d-j^M}(\boldsymbol{x})$ in a case where $pd + p < j^m$. Examples 10 and 12 show the polynomials corresponding to $\boldsymbol{X} \in \chi^*$ in high dimensions and with different choices for p. Example 11 shows the case pd integer. In all cases, given the final polynomial, the corresponding type-0 pmf can be easily found using the steps in the algorithm described in Section 3.

Example 9. Let $d = 7$, $s = 2$, and $t = 5$. We have $p = s/t = 2/5$, $pd = 14/5 = 2.8$, $j^M \equiv m = 2$, $j^m \equiv m + 1 = 3$, $a = (2s - t)/s = -1/2$, $a_1 = 1$, and $a_2 = 2$.

Since $pd + p = \frac{16}{5} > 3$, we have to consider $P_{d-j^m}(\boldsymbol{x}) \equiv P_{7-3}(\boldsymbol{x}) = P_4(\boldsymbol{x})$. Thus we have to find h, k such that

$$\begin{cases} h + 2k = 5 \\ 2k + 3k = 8. \end{cases}$$

We find $h = 1, k = 2$.

Since the linear term of $-2F_{1,\dots,4}(\boldsymbol{x})$ is $2\sum_{i=1}^{4} x_i$ we have split the $2 \cdot 4 = 8$ variables listed in the first row of Table 7 using $h = 1$ group with $m = 2$ variables and $k = 2$ groups with $m + 1 = 3$ variables. The second row of Table 7 reports the corresponding monomials.

The resulting polynomial is: $P_4(\boldsymbol{x}) = -2x_1x_2x_3x_4 + x_1x_2 + x_1x_3x_4 + x_2x_3x_4 - 1$.

Example 10. Consider the following two cases:

1. Let $d = 9$, $s = 2$, and $t = 5$. We have $p = s/t = 2/5$, $pd = 18/5 = 3.6$, $j^M \equiv m = 3$, $j^m \equiv m + 1 = 4$, $a = (2s - t)/s = -1/2$, $a_1 = 1$, and $a_2 = 2$. Since $pd + p = \frac{20}{5} = 4$, we have to consider $P_{d-j^m}(\boldsymbol{x}) = P_5(\boldsymbol{x})$. We find $h = 2, k = 1$ and $P_5(\boldsymbol{x}) = -2x_1 \cdots x_5 + x_1x_2x_3 + x_1x_4x_5 + x_2x_3x_4x_5 - 1$.
2. Let $d = 9$, $s = 2$, and $t = 7$. We have $p = s/t = 2/7$, $pd = 18/7 \approx 2.57$, $j^M \equiv m = 2$, $j^m \equiv m + 1 = 3$, $a = (2s - t)/s = -3/2$, $a_1 = 3$, and $a_2 = 2$. Since $pd + p = \frac{20}{7} < 3$, we have to consider $P_{d-m}(\boldsymbol{x}) = P_7(\boldsymbol{x})$. We find $h = 1, k = 4$ and

$$P_7(\boldsymbol{x}) = -2x_1 \cdots x_7 + x_1x_2 + x_3x_4x_5 + x_1x_6x_7 + x_2x_3x_4 + x_5x_6x_7 - 3.$$

S	$\boldsymbol{x}_{\alpha_i}$	constant	$d_j = d - j$	$\boldsymbol{x}_{d_j}$	constant	a
$S = 2$	$x_{i_1}x_{i_2}$	$+1$	5	$-x_1 \cdots x_5$	-4	$-\frac{1}{2}$
$S = 3$	$x_{i_1}x_{i_2}x_{i_3}$	$+2$	4	$-x_1 \cdots x_4$	-3	$-\frac{1}{2}$

Table 6. Monomial terms to be balanced in the case $d = 7$ and $p = 2/5$.

Linear terms	x_1 x_2	x_3 x_4 x_1	x_2 x_3 x_4
Monomials	x_1x_2	$x_3x_4x_1$	$x_2x_3x_4$

Table 7. Linear terms and the corresponding monomials for the case $d = 7$ and $p = 2/5$.

Example 11. Let $d = 5$, $s = 2$, and $t = 5$. We have $p = s/t = 2/5$, $pd = 2$ $a = (2s - t)/s = -1/2$, $a_1 = 1$, and $a_2 = 2$. The value of pd is integer and then we have to consider $P_{d-pd}(\boldsymbol{x}) = P_3(\boldsymbol{x})$.

$$P_3(\boldsymbol{x}) = -2x_1x_2x_3 + x_1x_2 + x_1x_3 + x_2x_3 - 1.$$

Example 12. Let $d = 216$, $s = 2$, and $t = 5$. We have $p = s/t = 2/5$, $pd = 432/5 = 86.4$, $j^M \equiv m = 86$, $j^m \equiv m + 1 = 87$, $a = (2s - t)/s = -1/2$, $a_1 = 1$, and $a_2 = 2$. Since $pd + p = 86.8$ is less than $m + 1 = 87$, we have to consider $P_{d-m}(\boldsymbol{x}) = P_{130}(\boldsymbol{x})$. We find $h = 1, k = 2$ and

$$P_{130}(\boldsymbol{x}) = -2x_1 \cdots x_{130} + x_1 \cdots x_{86} + x_1 \cdots x_{43} \cdot x_{87} \cdots x_{130} + x_{44} \cdots x_{130} - 1.$$

We have proved that $\boldsymbol{f} \in \chi^*$, i.e. $\boldsymbol{f}$ is a Σ_{cx}-smallest element in $\mathcal{F}_d(p)$, implies that $\boldsymbol{f}$ has the lowest mean correlation. The relationship between minimal convex sum and minimal dependence, also called strongest negative dependence, has been addressed in the literature (see [26] for a complete discussion on this matter). The main notion of negative dependence is countermonotonicity, that is defined for random vectors $\boldsymbol{X}$ of dimension $d = 2$ and that means that the components of $\boldsymbol{X}$ are oppositely ordered (see [26] and [22] for a formal definition of countermonotonicity). Minimal convex sums imply a negative dependence that is an extension of countermonotonicity to any d. The notion of Σ-countermonotonicity has been introduced in [26] (see also [18]) and exists for any Fréchet class.

Definition 5.3. A random vector $\boldsymbol{X}$ is Σ-countermonotonic if for any subset $i \subset \{1, \ldots, d\}$, we have that the random variables $\sum_{j \in I} X_j$ and $\sum_{j \notin I} X_j$ are countermonotonic.

In [26] it is proved that if $\boldsymbol{X}$ is a Σ_{cx}-smallest element than it is Σ-countermonotonic. Therefore the non-exchangeable pmf $\boldsymbol{f}^* \in \chi^*$ that we explicitly find in Theorem 5.2 is Σ-countermonotonic. We have also seen that it has the lowest mean correlation in Proposition 5.4. Section 6 further investigates the relationship between minimal convex sums and negative dependence for the class $\mathcal{F}_d(p)$.

6. The safest dependence structure

Since convex order is a variability order the minimal convex order correspond to a minimal risk random variable and the minimal convex order can be thought as the safest dependence structure. If $pd < 1$ the latter is linked to the strongest negative dependence, i.e. mutual exclusivity. This section generalizes the notion of mutual exclusivity to the case $pd > 1$ and discusses its link with the minimal convex order and with other notions of extreme negative dependence. In [4] the authors characterize mutual exclusivity as the strongest negative dependence structure. When mutual exclusivity is possible, it corresponds to the minimal convex order and in this light it can be considered the safest dependence structure. A vector $\boldsymbol{X} \in \mathcal{F}_d(p)$ is mutually exclusive if

$$P(X_{i_1} = 1, X_{i_2} = 1) = 0, \ \forall i_1, i_2 \in \{1, \ldots, d\}.$$

Remark 4. In dimension $d = 2$ a random vector $\boldsymbol{X}$ is mutually exclusive if and only if it is countermonotonic. If $d > 2$ there is not any concept of negative dependence that satisfies all the properties of countermonotonicity. The most intuitive extension of countermonotonicity is the notion of pairwise countermonotonicity ([26]), that is equivalent to the above notion of mutual exclusivity.

In [4] the authors also show that the distribution of a mutually exclusive random vector is the lower Fréchet bound of its Fréchet class, that in our framework is $F_L(\boldsymbol{x}) = \max(\sum_{i=1}^d F_i(x_i) - d + 1, 0)$. Nevertheless, if $pd > 1$ mutual exclusivity is not possible and the lower Fréchet bound is not a cdf. We now generalize the notion of mutual exclusivity to any p and d and discuss its link with negative correlation. The proof of Proposition 6.1 is in the supplementary material.

Definition 6.1. Let $\boldsymbol{X} \in \mathcal{F}_d(p)$, $\boldsymbol{X}$ is mutually exclusive of order m if

$$P(X_{i_1} = 1, \dots, X_{i_m} = 1) = 0, \ \forall i_1, \dots, i_m \in \{1, \dots, d\}.$$

If $m = 2$, $\boldsymbol{X}$ is mutually exclusive.

Proposition 6.1. *Let $\boldsymbol{X} \in \mathcal{F}_d(p)$ then $\boldsymbol{X}$ is mutually exclusive of order m iff $P(S \geq m) = 0$.*

The following result is a straightforward consequence of the fact that, to preserve the condition on the mean, if $dp \in (j^M, j^m]$, then $P(S \geq j^m) \neq 0$.

Remark 5. Another notion of extreme negative dependence is the joint mixability ([26] and [22]). A random vector $\boldsymbol{X}$ is said to be a joint mix if

$$P(S = k) = 1, \tag{15}$$

where $S = \sum_{i=1}^d X_i$, for some $k \in \mathbb{R}$, called the joint center of $\boldsymbol{X}$. In our case, $\boldsymbol{X} \in \mathcal{F}_d(p)$, thus $k \in \{0, \dots, d\}$ and $\boldsymbol{X}$ has the same one dimensional margins (under this assumption we say that $\boldsymbol{X}$ is completely mixable [22]). If $\boldsymbol{X} \in \mathcal{F}_d(p)$, we have $E[S] = dp$, therefore $\boldsymbol{X}$ can be completely mixable only with joint center pd and pd integer. If pd is integer the completely mixable random vector $\boldsymbol{X}$ has sum $S_{pd,pd}$ that is the Σ_{cx}-smallest element of $\mathcal{F}_d(p)$, according to Theorem 5.2. The relation between Σ-countermonotonicity, mixability and mutual countermonotonicity is discussed in [26].

Proposition 6.2. *Let $\boldsymbol{X} \in \mathcal{F}_d(p)$. If $dp \in (j^M, j^m]$ then $\boldsymbol{X}$ cannot be k-exclusive for all $k \leq j^m$.*

This is in line with the condition that the lower Fréchet bound is a distribution and belongs to the class $\mathcal{F}_d(p)$ iff $pd < 1$. In fact $\boldsymbol{X}$ can be 2-exclusive iff $j^m < 2$ and therefore $pd \leq j^m \leq 1$. Under this condition if $\boldsymbol{X}$ is 2-exclusive it is also pairwise countermonotonic and its cdf is the Fréchet lower bound. A consequence of the above results is that if $pd \in (j^M, j^m]$ the vector in $\mathcal{F}_d(p)$ cannot be mutually exclusive of order lower than j^m. Thus $j^m + 1$-mutually exclusive pmfs have sums with support on $\{0, 1, \dots, j^m\}$. Since the minimum of the convex order is reached on χ^*, the safest dependence structure is $j^m + 1$-mutually exclusive. Nevertheless, Proposition 5.2 implies that if $pd > 1$ not all the $j^m + 1$-mutually exclusive pmfs are minimal with respect to the convex order, but only the ones with support on $\{j^M, j^m\}$. Only the pmfs with support on $\{j^M, j^m\}$ are Σ_{cx}-smallest elements in $\mathcal{F}_d(p)$. Therefore $j^m + 1$-mutually exclusivity does not imply minimal convex sums, on the contrary Σ-countermonotonicity does. Since Σ-countermonotonicity implies minimal convex sums that implies $j^m + 1$-mutually exclusivity, Σ-countermonotonicity is a stronger notion than $j^m + 1$-mutually exclusivity. As already observed, fundamental polynomials of degree n and their opposite have sums with support on three points: $\{0, n, d-1\}$ or $\{1, d-n, d\}$, respectively. This means that they cannot be m-exclusive of order lower than $m = d - 1$ or $m = d$. The peculiarity of their support let us wonder whether they are associated to specific dependence structures. Nevertheless, since they are generators of the whole class they are able to generate all the admissible dependencies in the class itself.

7. Conclusion

We map the class $\mathcal{F}_d(p)$ into an ideal of points and we show that in any dimension the class $\mathcal{F}_d(p)$ is generated by a set of polynomials that we call fundamental polynomials. Each pmf in $\mathcal{F}_d(p)$ is associated to a linear combination of fundamental polynomials. This representation turns out to be very important to address open issues in the study of multivariate Bernoulli distributions. As a first application, we prove that a specific linear combination of fundamental polynomials solves the open problem to find a minimal distribution with respect to the convex order of sums and with respect to a measure of negative dependence in $\mathcal{F}_d(p)$. Lower bounds in the convex order identify the safest dependence structure, that is important in many fields, such as insurance or finance. In particular we are interested in applying our results to credit portfolio management, where multivariate Bernoulli distributions are used to model indicators of default. In this framework sums represent aggregate defaults and their bounds in the convex order are helpful to identify bounds for the risk associated to credit portfolios. Nowadays, the real economy is highly interconnected and banks and financial intermediaries are exposed to losses arising from defaults of obligors that are not independent. Since there are usually hundreds of obligors, banks need to handle high dimensional portfolios, hence the importance of analytical results.

Our theoretical future research will focus on fundamental polynomials, as generators of the $\mathcal{F}_d(p)$, and in particular on two open issues. First, fundamental polynomials of degree n, $n < d-1$ and their opposites have sums with support on three points: $0, n, d-1$ or $1, d-n, d$, respectively. This means that they cannot be m-exclusive of order lower than $m = d-1$ or $m = d$. The peculiarity of their support and its connection with a dependence notion puts in question if pmfs associated to fundamental polynomials can be characterized in terms of their dependence structure and if we can then identify a class of dependencies that are able to generate all the admissible dependencies in the class. Second, fundamental polynomials correspond to extremal points of the polytope $\mathcal{F}_d(p)$ and we have used them in this work to provide an algorithm to find extremal points in high dimensions. The connection between algebraic and geometrical generators and also their connections with bounds for the class will be further investigated.

Acknowledgements

The authors thank the anonymous referees for their valuable suggestions.

Funding

The authors gratefully acknowledge financial support from the Italian Ministry of Education, University and Research, MIUR, "Dipartimenti di Eccellenza" grant 2018-2022.

Supplementary Material

Supplement to "High dimensional Bernoulli distributions: Algebraic representation and applications" (DOI: 10.3150/23-BEJ1618SUPP; .pdf). The supplement contains the proofs of some propositions.

References

[1] 4ti2team (2018). 4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de.

[2] Boros, E. and Prékopa, A. (1989). Closed form two-sided bounds for probabilities that at least r and exactly r out of n events occur. *Math. Oper. Res.* **14** 317–342. MR0997039 https://doi.org/10.1287/moor.14.2.317

[3] Chaganty, N.R. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* **93** 197–206. MR2277750 https://doi.org/10.1093/biomet/93.1.197

[4] Cheung, K.C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance Math. Econom.* **55** 180–190. MR3179810 https://doi.org/10.1016/j.insmatheco.2014.01.001

[5] Cox, D.A., Little, J. and O'Shea, D. (2015). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 4th ed. *Undergraduate Texts in Mathematics*. Cham: Springer. MR3330490 https://doi.org/10.1007/978-3-319-16721-3

[6] Dai, B., Ding, S. and Wahba, G. (2013). Multivariate Bernoulli distribution. *Bernoulli* **19** 1465–1483. MR3102559 https://doi.org/10.3150/12-BEJSP10

[7] Denuit, M. and Vermandele, C. (1998). Optimal reinsurance and stop-loss order. *Insurance Math. Econom.* **22** 229–233. MR1650288 https://doi.org/10.1016/S0167-6687(97)00039-5

[8] Dhaene, J. and Denuit, M. (1999). The safest dependence structure among risks. *Insurance Math. Econom.* **25** 11–21. MR1718527 https://doi.org/10.1016/S0167-6687(99)00009-8

[9] Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26** 363–397. MR1608156 https://doi.org/10.1214/aos/1030563990

[10] Emrich, L.J. and Piedmonte, M.R. (1991). A method for generating high-dimensional multivariate binary variates. *Amer. Statist.* **45** 302–304.

[11] Fontana, R., Luciano, E. and Semeraro, P. (2021). Model risk in credit risk. *Math. Finance* **31** 176–202. MR4205881 https://doi.org/10.1111/mafi.12285

[12] Fontana, R. and Semeraro, P. (2018). Representation of multivariate Bernoulli distributions with a given set of specified moments. *J. Multivariate Anal.* **168** 290–303. MR3858366 https://doi.org/10.1016/j.jmva.2018.08.003

[13] Fontana, R. and Semeraro, P. (2022). Computational and analytical bounds for multivariate Bernoulli distributions. *J. Stat. Theory Pract.* **16** Paper No. 6, 12. MR4366952 https://doi.org/10.1007/s42519-021-00231-x

[14] Fontana, R. and Semeraro, P. (2023). Exchangeable Bernoulli distributions: High dimensional simulation, estimation, and testing. *J. Statist. Plann. Inference* **225** 52–70. MR4520494 https://doi.org/10.1016/j.jspi.2022.11.005

[15] Fontana, R. and Semeraro, P. (2024). Supplement to "High dimensional Bernoulli distributions: algebraic representation and applications." https://doi.org/10.3150/23-BEJ1618SUPP

[16] Frostig, E. (2001). Comparison of portfolios which depend on multivariate Bernoulli random variables with fixed marginals. *Insurance Math. Econom.* **29** 319–331. MR1874627 https://doi.org/10.1016/S0167-6687(01)00092-0

[17] Haynes, M.E., Sabo, R.T. and Chaganty, N.R. (2016). Simulating dependent binary variables through multinomial sampling. *J. Stat. Comput. Simul.* **86** 510–523. MR3421742 https://doi.org/10.1080/00949655.2015.1020313

[18] Hofert, M., Memartoluie, A., Saunders, D. and Wirjanto, T. (2017). Improved algorithms for computing worst Value-at-Risk. *Stat. Risk Model.* **34** 13–31. MR3652295 https://doi.org/10.1515/strm-2015-0028

[19] Kaas, R., Dhaene, J. and Goovaerts, M.J. (2000). Upper and lower bounds for sums of random variables. *Insurance Math. Econom.* **27** 151–168. MR1801600 https://doi.org/10.1016/S0167-6687(00)00060-3

[20] Kang, S.-H. and Jung, S.-H. (2001). Generating correlated binary variables with complete specification of the joint distribution. *Biom. J.* **43** 263–269. MR1855743 https://doi.org/10.1002/1521-4036(200106)43:3<263::AID-BIMJ263>3.0.CO;2-5

[21] Kvam, P.H. (1996). Maximum likelihood estimation and the multivariate Bernoulli distribution: An application to reliability. In *Lifetime Data: Models in Reliability and Survival Analysis* 187–194. Springer.

[22] Lee, W., Cheung, K.C. and Ahn, J.Y. (2017). Multivariate countermonotonicity and the minimal copulas. *J. Comput. Appl. Math.* **317** 589–602. MR3606098 https://doi.org/10.1016/j.cam.2016.12.032

[23] Marchetti, G.M. and Wermuth, N. (2016). Palindromic Bernoulli distributions. *Electron. J. Stat.* **10** 2435–2460. MR3545465 https://doi.org/10.1214/16-EJS1175

[24] McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised ed. *Princeton Series in Finance*. Princeton, NJ: Princeton Univ. Press. MR3445371

[25] Padmanabhan, D. and Natarajan, K. (2021). Tree bounds for sums of Bernoulli random variables: A linear optimization approach. *INFORMS J. Optim.* **3** 23–45. MR4237329 https://doi.org/10.1287/ijoo.2019.0038

[26] Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Statist. Sci.* **30** 485–517. MR3432838 https://doi.org/10.1214/15-STS525

[27] Qaqish, B.F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* **90** 455–463. MR1986660 https://doi.org/10.1093/biomet/90.2.455

[28] Rockafellar, R.T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton, NJ: Princeton Univ. Press. MR0274683

[29] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. New York: Springer. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

[30] Shults, J. (2017). Simulating longer vectors of correlated binary random variables via multinomial sampling. *Comput. Statist. Data Anal.* **114** 1–11. MR3660834 https://doi.org/10.1016/j.csda.2017.04.002

[31] Terzer, M. (2009). Large scale methods to enumerate extreme rays and elementary modes. Ph.D. thesis, ETH Zurich.

Received September 2022 and revised April 2023