Representation of multivariate Bernoulli distributions with a given set of specified moments

Roberto Fontana, Patrizia Semeraro

PII: S0047-259X(18)30024-1
DOI: https://doi.org/10.1016/j.jmva.2018.08.003
Reference: YJMVA 4398

To appear in: *Journal of Multivariate Analysis*

Received date : 12 January 2018

Representation of multivariate Bernoulli distributions with a given set of specified moments

Roberto Fontana[a,*], Patrizia Semeraro[a]

[a] *DISMA Dipartimento di Scienze Matematiche "Giuseppe Luigi Lagrange", Dipartimento di eccellenza 2018-2022, Politecnico di Torino, Torino, Italy*

Abstract

We propose a simple but new method of characterizing multivariate Bernoulli variables belonging to a given class, i.e., with some specified moments. Within a given class, this characterization allows us to generate easily a sample of mass functions. It also provides the bounds that all the moments must satisfy to be compatible and the possibility to choose the best distribution according to a certain criterion. For the special case of the Fréchet class of the multivariate Bernoulli distributions with given margins, we find a polynomial characterization of the class. Our characterization allows us to have bounds for the higher order moments. An algorithm is presented and illustrated.

Keywords: Algebraic statistics, Correlation, Fréchet class, Multivariate binary distribution, Simulation

1. Introduction

Dependent binary variables play a key role in many important scientific fields such as clinical trials and health studies. The problem of simulating correlated binary data is extensively addressed in the statistical literature; see, e.g., [3, 7, 9, 20]. Simulation studies are useful for analyzing extensions or alternatives to current estimating methods, such as generalized linear mixed models, or for the evaluation of statistical procedures for marginal regression models [18]. The simulation problem consists of constructing multivariate distributions for given Bernoulli marginal distributions and a given correlation matrix ρ. Frequently, assumptions are made about the correlation structure. Probably the most common is equicorrelation; see, e.g., [3]. A popular approach also uses working correlation matrices [11, 12], such as first-order moving average correlations or first-order autoregressive correlations; see [16] and references therein. An important issue for these simulation procedures is the compatibility of marginal binary variables and their correlations, since problems may arise when the margins and the correlation matrix are not compatible [2–4]. The range of admissible correlation matrices for binary variables is well-known in the bivariate case. For multivariate binary distributions with more than three variables, this problem has been widely studied in the literature; see, e.g., [10] and references therein. However, its solution from a practical point of view is still an open issue.

We propose a simple but new method to represent multivariate Bernoulli variables belonging to a given class, i.e., with some specified moments. This method represents the mass functions of the given class of multivariate Bernoulli distributions as points of the convex hull whose generators are mass functions which belong to the same class. Our main contribution is to provide a method and develop an algorithm to find the extreme rays of this convex hull. For the special case of the Fréchet class of the multivariate Bernoulli distributions with given one-dimensional margins, a polynomial-based representation of the mass functions is built. This representation is fully characterized, since our approach allows us to find necessary and sufficient conditions on the polynomial parameters to have a mass function in the Fréchet class.

Our new approach allows us to generate easily a sample of mass functions which belong to a given class and to find bounds for the non-specified moments of the distribution. It is worth noting that this method puts no restriction either on the number of variables or on the specified moments. The range of applications is limited only by the

*Corresponding author

amount of computational effort required, because the number of extremal rays increases very quickly as the dimension of the multivariate Bernoulli variables increases. This method provides a new computational procedure to simulate multivariate distributions of binary variables with some given moments. Theoretical results on the number of extremal rays can be found, e.g., in [14]. As we will see in Section 5.6 with respect to the problem of generating mass functions with given margins and correlations, our method performs better than other commonly used in real applications. In Section 4 we also show how this method offers the opportunity to choose the best distribution according to a certain criterion. For example, as the moments of multivariate Bernoulli are always positive, it could be of interest to find one of the distributions with the smallest sum of all the moments with order greater than 2. This problem can be efficiently solved using linear programming techniques [1].

The paper is organized as follows. After some preliminary notations in Section 2, Section 3 expresses any given Fréchet class of distributions with Bernoulli margins as points of the convex hull of the ray mass functions. The bivariate case is fully characterized in Section 3.2. The generalization to the class defined by a set of moments of any order is in Section 4. Some real examples and the algorithm are discussed in Section 5. Section 6 concludes.

2. Preliminaries

Let $\mathbb{F}_d$ be the set of d-dimensional distributions which have Bernoulli univariate marginal distributions. Let us consider the Fréchet class $\mathcal{F}(p_1, \ldots, p_d) \subseteq \mathbb{F}_d$ of distribution functions in $\mathbb{F}_d$ with Bernoulli marginal distributions $\mathcal{B}(p_1), \ldots, \mathcal{B}(p_d)$, where $p_1, \ldots, p_d \in (0,1)$. If $\boldsymbol{X} = (X_1, \ldots, X_d)$ is a random vector with joint distribution in $\mathcal{F}(p_1, \ldots, p_d)$, we denote its cumulative distribution function by $F_{\boldsymbol{p}}$ and its mass function by $f_{\boldsymbol{p}}$, where $\boldsymbol{p} = (p_1, \ldots, p_d)$. The column vectors which contain the values of $F_{\boldsymbol{p}}$ and $f_{\boldsymbol{p}}$ over $\mathcal{S}_d = \{0,1\}^d$ are denoted $\boldsymbol{F_p} = (F_{\boldsymbol{p}}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_d)$ and $\boldsymbol{f_p} = (f_{\boldsymbol{p}}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_d)$, respectively.

We make the non-restrictive hypothesis that the set $\mathcal{S}_d$ of 2^d binary vectors is ordered according to the reverse-lexicographical criterion. For example $\mathcal{S}_2 = \{00, 10, 01, 11\}$ and $\mathcal{S}_3 = \{000, 100, 010, 110, 001, 101, 011, 111\}$. For each $i \in \{1, \ldots, d\}$, the marginal cumulative distribution function and the marginal mass function of X_i are denoted $F_{\boldsymbol{p},i}$ and $f_{\boldsymbol{p},i}$, respectively; the values $f_{\boldsymbol{p},i}(0) \equiv F_{\boldsymbol{p},i}(0)$ and $f_{\boldsymbol{p},i}(1)$ are denoted q_i and p_i respectively. We observe that for each $i \in \{1, \ldots, d\}$, $q_i = 1 - p_i$ and that the expected value of X_i is p_i, viz. $\mathrm{E}(X_i) = p_i$.

Given two matrices $A \in \mathcal{M}(n \times m)$ and $B \in \mathcal{M}(d \times \ell)$, the matrix $A \otimes B = ((a_{ij}B)_{1 \le i \le n, 1 \le j \le m}) \in \mathcal{M}(nd \times m\ell)$ indicates their Kronecker product and

$$A^{\otimes n} = \underbrace{A \otimes \cdots \otimes A}_{n \text{ times}}.$$

If $\mathcal{B}(\tau)$ is a Bernoulli variable with $\tau \in (0,1)$, and F_τ and f_τ are its cumulative and mass function, respectively, then

$$\begin{pmatrix} f_\tau(0) \\ f_\tau(1) \end{pmatrix} = D \times \begin{pmatrix} F_\tau(0) \\ F_\tau(1) \end{pmatrix}, \quad \text{with } D = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix},$$

where D is the difference matrix. It follows that given $F_{\boldsymbol{p}}$ and $f_{\boldsymbol{p}}$ in $\mathcal{F}(p_1, \ldots, p_d)$, we have

$$\boldsymbol{f_p} = D^{\otimes d} \boldsymbol{F_p}. \tag{1}$$

We assume that vectors are column vectors. We denote the ith element of a vector a by $(a)_i \equiv a_i$ and its transpose by $a^\top$. We can thus write $\boldsymbol{f_p} \in \mathcal{F}(p_1, \ldots, p_d)$, $\boldsymbol{F_p} \in \mathcal{F}(p_1, \ldots, p_d)$ and $\boldsymbol{X} \in \mathcal{F}(p_1, \ldots, p_d)$.

2.1. Moments of multivariate Bernoulli variables

We observe that, given the Bernoulli variable $X \sim \mathcal{B}(\tau)$ with $\tau \in (0,1)$ and mass function f_τ, we can compute the vector of moments $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} = \begin{pmatrix} \mathrm{E}(1) \\ \mathrm{E}(X) \end{pmatrix} = M \begin{pmatrix} f_\tau(0) \\ f_\tau(1) \end{pmatrix} \quad \text{with } M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

It follows that given $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathcal{F}(p_1, \ldots, p_d)$ with multivariate joint mass $f_{\boldsymbol{p}}$, we can compute the vector of its moments $\boldsymbol{\mu} = (\mathrm{E}(X^{\boldsymbol{\alpha}}), \boldsymbol{\alpha} \in \mathcal{S}_d)$, where $\mathrm{E}(X^{\boldsymbol{\alpha}}) = \mathrm{E}(X_1^{\alpha_1} \cdots X_d^{\alpha_d})$, as $\boldsymbol{\mu} = M^{\otimes d} \boldsymbol{f_p}$. We use $\mu_{i_1, \ldots, i_k}$ to denote the moment $E(X_1^{\alpha_1}\cdots X_d^{\alpha_d})$ where $\{i_1,\ldots,i_k\} = \{1 \le i \le d : \alpha_i = 1\}$. We also observe that the correlation ρ_{ij} between two Bernoulli variables $X_i \sim \mathcal{B}(p_i)$ and $X_j \sim \mathcal{B}(p_j)$ is related to the second-order moment $\mu_{ij} \equiv E(X_iX_j)$ as follows:

$$E(X_iX_j) = \rho_{ij}\sqrt{p_iq_ip_jq_j} + p_ip_j. \tag{2}$$

The use of the Krockener product to represent a vectorized version of multivariate Bernoulli distributions can also be found in [22].

3. Representation of multivariate Bernoulli distributions with given margins

This section represents the Fréchet class of multivariate d-dimensional Bernoulli distributions with given margins, $d \ge 2$, as the points of a convex hull. As a first step, in Proposition 1 we give a polynomial-based representation of the distributions of a given Fréchet class, which allow us to model all the possible dependence structures.

We consider the set $\tilde{\mathcal{S}}_d = \{0,1\}^d$ of binary vectors ordered according to the lexicographical criterion, e.g., $\tilde{\mathcal{S}}_3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$. Let $\boldsymbol{\theta} = \{\theta_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \tilde{\mathcal{S}}_d\}$ be a 2^d-vector parameter. We denote $\theta_{\boldsymbol{\alpha}}$ by $\theta_{i_1,\ldots,i_k}$, where $\{i_1,\ldots,i_k\} = \{i \in \{1,\ldots,d\} : \alpha_i = 1\}$. Thus we write $\boldsymbol{\theta} = (\theta_0, \theta_d, \theta_{d-1}, \theta_{d,d-1}, \ldots, \theta_{12\ldots d})$, e.g., in the bivariate case $\boldsymbol{\theta} = (\theta_0, \theta_2, \theta_1, \theta_{12})$. We also define $\mathcal{Q}_d = \{q_1, 1\} \times \cdots \times \{q_d, 1\}$, where $q_i = F_i(0)$ for all $i \in \{1,\ldots,d\}$. We make the non-restrictive hypothesis that $\{q_1, 1\} \times \cdots \times \{q_d, 1\}$ is ordered according to the reverse-lexicographical criterion.

In what follows, $U_{\boldsymbol{p}}$ is a $2^d \times 2^d$ matrix given by $U_{\boldsymbol{p}} = U_{p_1} \otimes \cdots \otimes U_{p_d}$, where for each $i \in \{1,\ldots,d\}$,

$$U_{p_i} = \begin{pmatrix} 1 & p_i \\ 1 & 0 \end{pmatrix}.$$

Proposition 1. *Any distribution $F_{\boldsymbol{p}} \in \mathcal{F}(p_1,\ldots,p_d)$ admits the representation $\boldsymbol{F_p} = \Lambda_{\boldsymbol{p}} U_{\boldsymbol{p}} \boldsymbol{\theta}$ over $\mathcal{Q}_d$ in terms of a $2^d \times 2^d$ diagonal matrix $\Lambda_{\boldsymbol{p}} = \mathrm{diag}(q_1^{(1-\alpha_1)}\cdots q_d^{(1-\alpha_d)}, (\alpha_1,\ldots,\alpha_d) \in \mathcal{S}_d)$, and $\boldsymbol{\theta} = (\theta_0, \theta_d, \theta_{d-1}, \theta_{d,d-1}, \ldots, \theta_{12\ldots d})^\top$. Necessary conditions for $F_{\boldsymbol{p}}$ being a distribution are $\theta_0 = 1$ and $\theta_i = 0$ for all $i \in \{1,\ldots,d\}$.*

Proof. Given $\boldsymbol{u} = (u_1,\ldots,u_d) \in \mathcal{Q}_d$, let us define the following polynomial:

$$\begin{aligned} g(\boldsymbol{u}) &= \left(\prod_{i=1}^d u_i\right)\left[1 + \sum_{k=2}^d \left\{\sum_{1\le i_1<i_2<\cdots<i_k\le d} \theta_{i_1i_2\ldots i_k}(1-u_{i_1})\cdots(1-u_{i_k})\right\}\right] \\ &= \left(\prod_{i=1}^d u_i\right)\left\{1 + \sum_{1\le i_1<i_2\le d} \theta_{i_1i_2}(1-u_j)(1-u_k) + \cdots + \theta_{12\ldots d}\prod_{i=1}^d (1-u_i)\right\}. \end{aligned}$$

For each $i \in \{1,\ldots,d\}$, define also the row vector $a_i = (1,\ 1-u_i)$. We can write $g(\boldsymbol{u}) \in \mathbb{R}$ as

$$g(u) = \left(\prod_{i=1}^d u_i\right)(a_1 \otimes \cdots \otimes a_d)\begin{pmatrix} \theta_0 \\ \theta_d \\ \theta_{d-1} \\ \vdots \\ \theta_{12\ldots d} \end{pmatrix}.$$

Considering all the $\boldsymbol{u} \in \mathcal{Q}_d$,we obtain the 2^d-vector $(g(\boldsymbol{u}), \boldsymbol{u} \in \mathcal{Q}_d) = \Lambda_{\boldsymbol{p}} U_{\boldsymbol{p}} \boldsymbol{\theta}$.

We observe that the determinant of U_{p_i} is $\det(U_{p_i}) = -p_i \neq 0$. It follows that the determinant of $U_{\boldsymbol{p}}$, which is $(p_1\cdots p_d)^2$, is also different from zero. Because the determinant of $\Lambda_{\boldsymbol{p}} \neq 0$, the determinant of $\Lambda_{\boldsymbol{p}} U_{\boldsymbol{p}}$ is different from zero. It follows that the rank of $\Lambda_{\boldsymbol{p}} U_{\boldsymbol{p}}$ is 2^d and then any vector $\boldsymbol{y} \in \mathbb{R}^{2^d}$, and in particular any distribution $F_{\boldsymbol{p}}$, can be written as $\boldsymbol{F_p} = \Lambda_{\boldsymbol{p}} U_{\boldsymbol{p}} \boldsymbol{\theta}$.

If $F_{\boldsymbol{p}}$ is a distribution in $\mathcal{F}(p_1,\ldots,p_d)$, it must be such that $F_{\boldsymbol{p}}(1,\ldots,1) = 1$ and $F_{\boldsymbol{p}}(1,\ldots 1, 0, 1, \ldots, 1) = q_i$ for all $i \in \{1,\ldots,d\}$. It follows that the vector parameter $\boldsymbol{\theta}$ must satisfy the following necessary conditions:

(i) $\theta_0 = 1$: The condition $F_{\boldsymbol{p}}(1,\ldots,1) = 1$ implies $\theta_0 = 1$, since $F_{\boldsymbol{p}}(1,\ldots,1) = \theta_0$.

(ii) $\theta_1 = \cdots = \theta_d = 0$: The condition $F_{\boldsymbol{p}}(1,\ldots 1, 0, 1, \ldots, 1) = q_i$ implies $\theta_1 = \cdots = \theta_d = 0$ because $F_{\boldsymbol{p}}(1,\ldots 1, 0, 1, \ldots, 1) = q_i\{1 + \theta_i(1 - q_i)\}$. □

As an example, in the bivariate case

$$\Lambda_{\boldsymbol{p}} = \begin{pmatrix} q_1q_2 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad U_{\boldsymbol{p}} = \begin{pmatrix} 1 & p_2 & p_1 & p_1p_2 \\ 1 & 0 & p_1 & 0 \\ 1 & p_2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

We observe that the polynomial $g(\boldsymbol{u})$ in the proof of Proposition 1 is restricted to the finite set $\mathcal{Q}_d$. Note also that its expression is reminiscent of the Farlie–Gumbel–Morgenstern [15] copula defined, for all $\boldsymbol{u} \in [0, 1]^d$, by

$$C(\boldsymbol{u}) = \left(\prod_{i=1}^{d} u_i\right)\left\{1 + \sum_{1\leq i_1<i_2\leq d} \theta_{i_1i_2}(1 - u_j)(1 - u_k) + \cdots + \theta_{12\ldots d}\prod_{i=1}^{d}(1 - u_i)\right\}.$$

This approach leads to a polynomial representation whose expression is similar to the Bahadur representation; see [13]. The main difference between the two approaches is given by the domain of the polynomial, which in our case is $\mathcal{Q}_d$. By contrast, the Bahadur representation depends on $\mathcal{Q}_d$ and on the support of the marginal distributions.

As a consequence of Proposition 1 and Eq. (1), any mass function $f_{\boldsymbol{p}} \in \mathcal{F}(p_1, \ldots, p_d)$ admits the following representation over $\mathcal{S}_d$:

$$\boldsymbol{f}_{\boldsymbol{p}} = Y_{\boldsymbol{p}}\boldsymbol{\theta}, \tag{3}$$

where $Y_{\boldsymbol{p}} = D^{\otimes d}\Lambda_{\boldsymbol{p}}U_{\boldsymbol{p}} \in \mathcal{M}(2^d \times 2^d)$ and D is the 2×2 difference matrix.

Remark 1. As in [19], we can interpret $\boldsymbol{\theta}$. Let us consider $d = 1$, $\boldsymbol{\theta} = (\theta_0, \theta_1)$. Simple computations give

$$\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1/(1-p) & -1/p \end{pmatrix}\begin{pmatrix} \boldsymbol{f}(0) \\ \boldsymbol{f}(1) \end{pmatrix}.$$

Thus $\theta_0 = \mathrm{E}(1)$ and $\theta_1 = -\mathrm{E}[(X - p)/\{p(1 - p)\}]$. If we consider $d \geq 2$ we obtain $\theta_{i_1,\ldots,i_k} = (-1)^k\,\mathrm{E}(Y_{i_1}\cdots Y_{i_k})$, where $Y_{i_j} = (X_{i_j} - p_{i_j})/\{p_{i_j}(1 - p_{i_j})\}$ for all $j \in \{1, \ldots, k\}$. We observe that Y_i is a zero-mean random variable which is obtained by the linear transformation of the Bernoulli $\mathcal{B}(p_i)$ that moves 0 to $-1/q_i$ and 1 to $1/p_i$ for each $i \in \{1, \ldots, d\}$ and that $\theta_{i_1,\ldots,i_k}$ are its moments, apart from the plus or minus sign.

We observe that given $f_{\boldsymbol{p}} \in \mathcal{F}(p_1, \ldots, p_d)$ we can write it as in Eq. (3). Reciprocally, Proposition 1 does not provide any condition on $\theta_{i_1,\ldots,i_k}$ for $k \geq 2$ such that $D^{\otimes d}\Lambda_{\boldsymbol{p}}U_{\boldsymbol{p}}\boldsymbol{\theta}$ represents a mass function $f_{\boldsymbol{p}}$ over $\mathcal{S}_d$. In the remaining part of this section we will provide a representation of all the mass functions $f_{\boldsymbol{p}} \in \mathcal{F}(p_1, \ldots, p_d)$. We denote by $H_{1\boldsymbol{p}}$ the $d \times 2^d$ sub-matrix of $Y_{\boldsymbol{p}}^{-1}$ obtained by selecting the rows corresponding to $\theta_1, \ldots, \theta_d$ and we recall that $\mathcal{S}_d = \{0, 1\}^d$ is ordered according to the reverse-lexicographical criterion.

Theorem 1. *Let f be a multivariate d-dimensional Bernoulli distribution, $f \in \mathbb{F}_d$. Then f is a mass with margins p, i.e., $f \in \mathcal{F}(p_1, \ldots, p_d)$ if and only if there exist $\lambda_1, \ldots, \lambda_{n_{C_{1p}}} \geq 0$ summing up to 1 such that*

$$\boldsymbol{f} = \sum_{i=1}^{n_{C_{1p}}} \lambda_i \boldsymbol{R}_{\boldsymbol{p}}^{(i)}, \tag{4}$$

where $\boldsymbol{R}_{\boldsymbol{p}}^{(i)} = (R_{\boldsymbol{p}}^{(i)}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_d) \in \mathcal{F}(p_1, \ldots, p_d)$ are the normalized extremal rays of the cone $C_{1\boldsymbol{p}} = \{\boldsymbol{z} : H_{1\boldsymbol{p}}\boldsymbol{z} = \boldsymbol{0}, \boldsymbol{z} \in \mathbb{R}_+^d\}$ and $n_{C_{1p}}$ is the number of extremal rays in $C_{1\boldsymbol{p}}$.

Proof. Let us prove that any $f_{\boldsymbol{p}} \in \mathcal{F}(p_1, \ldots, p_d)$ satisfies Eq. (4). From Proposition 1 and Eq. (3), we have $\boldsymbol{f}_{\boldsymbol{p}} = Y_{\boldsymbol{p}}\boldsymbol{\theta}$ with $\theta_0 = 1$ and $\theta_1 = \cdots = \theta_d = 0$. The conditions $\theta_1 = \cdots = \theta_d = 0$ can be written as

$$H_{1\boldsymbol{p}}\boldsymbol{f}_{\boldsymbol{p}} = 0, \tag{5}$$

where $H_{1\boldsymbol{p}}$ is the $d \times 2^d$ sub-matrix of $Y_{\boldsymbol{p}}^{-1}$ obtained by selecting the rows corresponding to $\theta_1, \ldots, \theta_d$. It follows that $\boldsymbol{f}_{\boldsymbol{p}}$ is a point of the cone $C_{1\boldsymbol{p}} = \{\boldsymbol{z} : H_{1\boldsymbol{p}}\boldsymbol{z} = \boldsymbol{0}, \boldsymbol{z} \in \mathbb{R}_+^d\}$ and can be written as

$$\boldsymbol{f}_{\boldsymbol{p}} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \tilde{\lambda}_i \tilde{\boldsymbol{R}}_{\boldsymbol{p}}^{(i)}$$

for some choices of non-negative $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_{n_{C_{1\boldsymbol{p}}}}$, where for each $i \in \{1, \ldots, n_{C_{1\boldsymbol{p}}}\}$, $\tilde{\boldsymbol{R}}_{\boldsymbol{p}}^{(i)} = (\tilde{R}_{\boldsymbol{p},k}^{(i)}, k \in \{1, \ldots, 2^d\}) \in \mathbb{R}_+^{2^d}$ is an extremal ray of the cone $C_{1\boldsymbol{p}}$ [8, 21].

If we define the normalized extremal rays by dividing each $\tilde{\boldsymbol{R}}_{\boldsymbol{p}}^{(i)}$ by the sum of its elements, viz.

$$\tilde{R}_{\boldsymbol{p},+}^{(i)} = \sum_{k=1}^{2^d} \tilde{R}_{\boldsymbol{p},k}^{(i)}$$

we can write $\boldsymbol{f}_{\boldsymbol{p}}$ in the form Eq. (4), where $\lambda_i = \tilde{\lambda}_i \tilde{R}_{\boldsymbol{p},+}^{(i)}$ and $\boldsymbol{R}_{\boldsymbol{p}}^{(i)} = \tilde{\boldsymbol{R}}_{\boldsymbol{p}}^{(i)} / \tilde{R}_{\boldsymbol{p},+}^{(i)}$ for all $i \in \{1, \ldots, n_{C_{1\boldsymbol{p}}}\}$. We observe that $\lambda_1, \ldots, \lambda_d$ are non-negative and that

$$\sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \tilde{\lambda}_i \tilde{R}_{\boldsymbol{p},+}^{(i)} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \tilde{\lambda}_i \sum_{k=1}^{2^d} \tilde{R}_{\boldsymbol{p},k}^{(i)} = \sum_{k=1}^{2^d} \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \tilde{\lambda}_i \tilde{R}_{\boldsymbol{p},k}^{(i)} = \sum_{k=1}^{2^d} (\boldsymbol{f}_{\boldsymbol{p}})_k = 1$$

Let us now prove the reciprocal, i.e., that if there exist non-negative $\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}}$ summing up to 1 such that $f \in \mathbb{F}$ is such that Eq. (4) holds, then $f \in \mathcal{F}(p_1, \ldots, p_d)$.

In Eq. (4), one can take $\boldsymbol{R}_{\boldsymbol{p}}^{(i)} = (R_{\boldsymbol{p}}^{(i)}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_d) \in \mathcal{F}(p_1, \ldots, p_d)$ to be the normalized extremal rays of the cone $C_{1\boldsymbol{p}} = \{z : H_{1\boldsymbol{p}}z = 0, z \in \mathbb{R}_+^d\}$, where $\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}}$ are non-negative and sum up to 1. Given that f is a convex linear combination of mass functions, it is therefore a mass function; we have to prove that its corresponding margins are $p_1, \ldots, p_d$. We have

$$\mathrm{E}(X_j) = \sum_{(x_1,\ldots,x_d)\in\mathcal{S}_d} x_j f_{\boldsymbol{p}}(x_1, \ldots, x_d) = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \sum_{(x_1,\ldots,x_d)\in\mathcal{S}_d} x_j \boldsymbol{R}_{\boldsymbol{p}}^{(i)}.$$

Now we observe that each $\boldsymbol{R}_{\boldsymbol{p}}^{(i)}$ is a point of the cone $C_{1\boldsymbol{p}}$ i.e., $H_{1\boldsymbol{p}}\boldsymbol{R}_{\boldsymbol{p}}^{(i)} = 0$. It follows that

$$\sum_{(x_1,\ldots,x_d)\in\mathcal{S}_d} x_j \boldsymbol{R}_{\boldsymbol{p}}^{(i)} = p_j.$$

Then we have

$$\sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \sum_{(x_1,\ldots,x_d)\in\mathcal{S}_d} x_j \boldsymbol{R}_{\boldsymbol{p}}^{(i)} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i p_j = p_j \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i = p_j.$$

The assertion is thus proved. □

We refer to the normalized extremal rays of the cone $C_{1\boldsymbol{p}}$ as the ray mass functions of $\mathcal{F}(p_1, \ldots, p_d)$.

Remark 2. The bivariate case is fully described from an analytical point of view; see Section 3.2. In general the number of extremal rays of the convex cone $C_{1\boldsymbol{p}}$ depends on $\boldsymbol{p}$ and d. For example for trivariate Bernoulli distributions, in Example 5.3.1 we consider the Fréchet class $\mathcal{F}(1/2, 1/2, 1/2)$ and we explicitly find six ray mass functions and the corresponding $\boldsymbol{\theta}$ vectors. In Example 5.3.2 we consider the Fréchet class $\mathcal{F}(1/4, 1/7, 1/3)$ and we find 11 ray mass functions. We also observe that the number of rays can become very large as d increases; see Section 5.5. The problem of determining bounds for the number of extremal rays of a convex cone without explicitly computing them is interesting and is studied in algebra and geometry; see, e.g., [14].

Theorem 1 allows us to give a characterization of the space of the parameters θ. From Eq. (3) we know that any $f_{\boldsymbol{p}} \in \mathcal{F}(p_1, \ldots, p_d)$ can be written as $Y_{\boldsymbol{p}}\boldsymbol{\theta}$.

Proposition 2. *Let $f \in \mathbb{F}_d$. Then f is a mass function with margins $\boldsymbol{p}$, i.e., $\boldsymbol{f} = Y_{\boldsymbol{p}}\boldsymbol{\theta} \in \mathcal{F}(p_1, \ldots, p_d)$ if and only if there exist non-negative numbers $\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}}$ summing up to 1 such that $\boldsymbol{f}_{\boldsymbol{p}} = Y_{\boldsymbol{p}}\boldsymbol{\theta}$, with*

$$\boldsymbol{\theta} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \boldsymbol{\theta}_{\boldsymbol{p}}^{(i)}, \tag{6}$$

where $\boldsymbol{\theta}_{\boldsymbol{p}}^{(i)} = Y_{\boldsymbol{p}}^{-1}\boldsymbol{R}_{\boldsymbol{p}}^{(i)}$ are the parameters of the ray mass functions of the cone $C_{1\boldsymbol{p}}$ and $n_{C_{1\boldsymbol{p}}}$ is the number of the extremal rays of $C_{1\boldsymbol{p}}$.

Proof. $\Rightarrow$) From Theorem 1, using $\boldsymbol{f}_{\boldsymbol{p}} = Y_{\boldsymbol{p}}\boldsymbol{\theta}$ we get

$$\boldsymbol{\theta} = Y_{\boldsymbol{p}}^{-1}\boldsymbol{f}_{\boldsymbol{p}} = Y_{\boldsymbol{p}}^{-1}\left(\sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \boldsymbol{R}_{\boldsymbol{p}}^{(i)}\right) = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i Y_{\boldsymbol{p}}^{-1}\boldsymbol{R}_{\boldsymbol{p}}^{(i)} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \boldsymbol{\theta}_{\boldsymbol{p}}^{(i)},$$

where $\boldsymbol{\theta}_{p}^{(i)}$ are the parameters of the ray mass functions of the cone $C_{1\boldsymbol{p}}$ and $\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}}$ are non-negative numbers summing up to 1.

$\Leftarrow$) Let us consider $\boldsymbol{\theta}$ given by Eq. (6), where for each $i \in \{1, \ldots, n_{C_{1\boldsymbol{p}}}\}$, $\boldsymbol{\theta}_{\boldsymbol{p}}^{(i)}$ is the parameter of the ray mass function $\boldsymbol{R}_{\boldsymbol{p}}^{(i)}$, and $\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}}$ are non-negative numbers summing up to 1. We get

$$Y_{\boldsymbol{p}}\boldsymbol{\theta} = Y_{\boldsymbol{p}} \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \boldsymbol{\theta}_{\boldsymbol{p}}^{(i)} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i Y_{\boldsymbol{p}}\boldsymbol{\theta}_{p}^{(i)} = \sum_{i=1}^{n_{C_{1\boldsymbol{p}}}} \lambda_i \boldsymbol{R}_{\boldsymbol{p}}^{(i)}.$$

From Theorem 1 we find $\boldsymbol{f}_{\boldsymbol{p}} = Y_{\boldsymbol{p}}\boldsymbol{\theta} \in \mathcal{F}(p_1, \ldots, p_d)$. □

Note that the above proposition provides the set of admissible parameters for the polynomial $g(\boldsymbol{u})$, which is the region of the convex hull of the parameters $\boldsymbol{\theta}_{p}^{(i)}$ of the polynomials of the ray mass functions obtained using non-negative scalars $\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}}$ summing up to 1. We also point out that the parameters of the ray mass functions depend on the marginal parameters $\boldsymbol{p}$, which implies that the admissible range for the parameters $\boldsymbol{\theta}$ depends on the marginal parameters. This fact is made explicit in Section 3.2 and in the first example of Section 5. Further note that Theorem 1 makes it extremely easy to generate any mass function $f_{\boldsymbol{p}}$ of the Fréchet class $\mathcal{F}(p_1, \ldots, p_d)$. It is enough to take a positive vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}})$ whose components add up to 1 and to build $\boldsymbol{f}_{\boldsymbol{p}}$ using Eq. (4).

Using Theorem 1 we represent each Fréchet class $\mathcal{F}(p_1, \ldots, p_d)$ as the intersection of the convex hull of the ray mass functions with the probability simplex. We observe that the ray mass functions depend only on the marginal distributions $F_1, \ldots, F_d$, i.e., by $p_1, \ldots, p_d$. Building the $(2^d \times n_{C_{1\boldsymbol{p}}})$-ray matrix

$$R_{\boldsymbol{p}} = \begin{pmatrix} R_{p,1}^{(1)} & \ldots & R_{p,1}^{(n_{C_{1\boldsymbol{p}}})} \\ & \ldots & \\ R_{p,2^d}^{(1)} & \ldots & R_{p,2^m}^{(n_{C_{1\boldsymbol{p}}})} \end{pmatrix}$$

whose columns are the ray mass functions $\boldsymbol{R}_{p}^{(1)}, \ldots, \boldsymbol{R}_{p}^{n_{C_{1\boldsymbol{p}}}}$ we write Eq. (4) simply as $\boldsymbol{f}_{\boldsymbol{p}} = R_{\boldsymbol{p}}\boldsymbol{\lambda}$ with the vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{n_{C_{1\boldsymbol{p}}}})$ whose components are non-negative and add up to 1.

In practice the extremal rays $\tilde{\boldsymbol{R}}_{p}^{(i)}$ of the cone $C_{1\boldsymbol{p}}$ and therefore the ray mass functions $\boldsymbol{R}_{p}^{(i)}$ can be found using the software `4ti2` [21]. In Section 5 and in the Appendix we use `SAS` and `4ti2` to show some numerical examples.

3.1. Second-order moments of multivariate Bernoulli variables with given margins

This section focuses on the problem of studying second-order moments of multivariate Bernoulli variables with given first-order moments. Given $\boldsymbol{f}_{\boldsymbol{p}} \in \mathcal{F}(p_1, \ldots, p_d)$, we observe from Theorem 1 that each moment $\mathrm{E}(X^{\alpha})$ with $\boldsymbol{\alpha} \in \mathcal{S}_d$ can be computed as

$$\boldsymbol{\mu} = M^{\otimes d}\boldsymbol{f}_{\boldsymbol{p}} = M^{\otimes d}R_{\boldsymbol{p}}\lambda.$$

Let $A_{k\boldsymbol{p}} = (M^{\otimes d})_k R_{\boldsymbol{p}}$ where $(M^{\otimes d})_k$ is the sub-matrix of $M^{\otimes d}$ obtained by selecting the rows corresponding to the kth order moments and $R_{\boldsymbol{p}}$ is the ray matrix. We observe that the columns of the matrix $A_{k\boldsymbol{p}}$ contain the moments of the ray mass functions. We denote by $A_{\boldsymbol{p}}$ the matrix whose columns contain all the moments of the ray mass functions, viz. $A_{\boldsymbol{p}} = M^{\otimes d} R_{\boldsymbol{p}}$.

In particular for the second-order moments $\boldsymbol{\mu}_2 = \mathrm{E}(X^{\alpha} : \boldsymbol{\alpha} \in \mathcal{S}_d, ||\boldsymbol{\alpha}||_0 = 2)$, where $||\boldsymbol{\alpha}||_0 = \alpha_1 + \cdots + \alpha_d$, we get the following result, which is crucial for the solution of the problem of simulating multivariate binary distributions with a given correlation matrix.

Proposition 3. *One has* $\boldsymbol{\mu}_2 = A_{2p}\boldsymbol{\lambda}$*, where* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{n_{C_{1p}}})$ *is a vector whose non-negative components sum to* 1.

It follows that the target second-order moments are compatible with the means if they belong to the part of the convex hull generated by the points which are the columns of the $A_{2\boldsymbol{p}} = (M^{\otimes d})_2 R_{\boldsymbol{p}}$ matrix restricted to $\lambda_1, \ldots, \lambda_d \geq 0$ and $\lambda_1 + \cdots + \lambda_{n_{C_{1p}}} = 1$. As a direct consequence of Proposition 3 we also get the univariate bounds for the second-order moments and the correlations.

Proposition 4. *For each* $\boldsymbol{\alpha} \in \mathcal{S}_d$*,* $||\boldsymbol{\alpha}||_0 = 2$*, the second-order moment* $\boldsymbol{\mu}_2^{(\alpha)}$ *must satisfy the following bounds*

$$\min A_{2\boldsymbol{p}}^{(\alpha)} \leq \boldsymbol{\mu}_2^{(\alpha)} \leq \max A_{2\boldsymbol{p}}^{(\alpha)}$$

and the correlations ρ_{ij} *must satisfy the following bounds*

$$\frac{\min A_{2\boldsymbol{p}}^{(\alpha)} - p_i p_j}{\sqrt{p_i q_i p_j q_j}} \leq \rho_{ij} \leq \frac{\max A_{2\boldsymbol{p}}^{(\alpha)} - p_i p_j}{\sqrt{p_i q_i p_j q_j}},$$

where $A_{2\boldsymbol{p}}^{(\alpha)}$ *is the row of the matrix* $A_{2\boldsymbol{p}}$ *such that* $\boldsymbol{\mu}_2^{(\alpha)} = A_{2\boldsymbol{p}}^{(\alpha)}\boldsymbol{\lambda}$ *and* $\{i, j\} = \{k : \alpha_k = 1\}$.

Proof. From Proposition 3 using the the proper row of $A_{2\boldsymbol{p}}$ we get

$$\boldsymbol{\mu}_2^{(\alpha)} = A_{2\boldsymbol{p}}^{(\alpha)}\boldsymbol{\lambda}.$$

To prove (4) it is enough to observe that

a) because the λ_i's are non-negative and add up to 1, it follows that the minimum (maximum) value of $\boldsymbol{\mu}_2^{(\alpha)}$ is obtained by choosing $\boldsymbol{\lambda}$ equal to one of the $\boldsymbol{e}_i$ vectors, where for each $i \in \{1, \ldots, n_{C_{1p}}\}$, $\boldsymbol{e}_i \in \{0, 1\}^{n_{C_{1p}}}$ is the binary vector with all the elements equal to 0 apart from the ith which is equal to 1;

b) the product $A_{2\boldsymbol{p}}^{(\alpha)}\boldsymbol{e}_i$ gives the ith element of $A_{2\boldsymbol{p}}^{(\alpha)}$.

To prove (4) we simply observe that using Eq. (2) the bounds in (4) can be transformed to those suitable for correlations. □

Generalization to kth order moments ($k > 2$) is straightforward.

Proposition 5. *For each* $\boldsymbol{\alpha} \in \mathcal{S}_d$*,* $||\boldsymbol{\alpha}||_0 = k$*, the kth order moment* $\boldsymbol{\mu}_k^{(\alpha)}$ *must satisfy the following bounds*

$$\min A_{k\boldsymbol{p}}^{(\alpha)} \leq \boldsymbol{\mu}_k^{(\alpha)} \leq \max A_{k\boldsymbol{p}}^{(\alpha)},$$

where $A_{k\boldsymbol{p}}^{(\alpha)}$ *is the row of the matrix* $A_{k\boldsymbol{p}}$ *such that* $\boldsymbol{\mu}_k^{(\alpha)} = A_{k\boldsymbol{p}}^{(\alpha)}\boldsymbol{\lambda}$*.*

In Section 3.2 we recover these bounds for the bivariate case. In the examples we exhibit these bounds for other specified Fréchet classes.

3.2. Bivariate Bernoulli distributions with given margins

In this section we consider bivariate distributions, i.e., the class $\mathcal{F}(p_1, p_2)$ of 2-dimensional random variables (X_1, X_2) which have Bernoulli marginal distributions $F_1 \sim \mathcal{B}(p_1)$, $F_2 \sim \mathcal{B}(p_2)$. This class has been extensively studied and we can provide the analytical expression of the ray matrix R_p.

In the bivariate case, two key distributions are F_L and F_U, the lower and upper Fréchet–Hoeffding bound of $\mathcal{F}(p_1, p_2)$, respectively defined, for all $\boldsymbol{x} = (x_1, x_2) \in \{0, 1\}^2$, by

$$F_L(\boldsymbol{x}) = \max\{F_1(x_1) + F_2(x_2) - 1, 0\}, \quad F_U(\boldsymbol{x}) = \min\{F_1(x_1), F_2(x_2)\}. \tag{7}$$

For any $F_{\boldsymbol{p}} \in \mathcal{F}(p_1, p_2)$, one has

$$F_L(\boldsymbol{x}) \leq F_{\boldsymbol{p}}(\boldsymbol{x}) \leq F_U(\boldsymbol{x}) \tag{8}$$

for all $\boldsymbol{x} \in \{0, 1\}^2$. For an overview of Fréchet classes and their bounds, see [5].

The number of rays is independent of the Fréchet class $\mathcal{F}(p_1, p_2)$. We have two ray mass functions. We make the non-restrictive hypothesis $p_1 \geq p_2$ which implies $q_1 \leq q_2$ and we consider separately the cases $q_1 + q_2 \leq 1$ and $q_1 + q_2 > 1$.

3.2.1. Case 1: $q_1 + q_2 \leq 1$

From (8) we get

$$\boldsymbol{R_p} = \begin{pmatrix} 0 & q_1 \\ q_2 & q_2 - q_1 \\ q_1 & 0 \\ 1 - q_1 - q_2 & p_2 \end{pmatrix}.$$

If we denote by $\boldsymbol{\theta}^{(i)}$ the vector of parameters corresponding to $\boldsymbol{R}_{\boldsymbol{p}}^{(i)}$, $i \in \{1, 2\}$ we obtain $\boldsymbol{\theta}^{(1)} = (1, 0, 0, -1/(p_1 p_2))$ and $\boldsymbol{\theta}^{(2)} = (1, 0, 0, 1/(p_1 q_2))$, which yields $-1/(p_1 p_2) \leq \theta_{12} \leq 1/(p_1 q_2)$. Finally we get the following well-known bounds for the correlation:

$$-\sqrt{\frac{q_1 q_2}{p_1 p_2}} \leq \rho_{12} \leq \sqrt{\frac{p_2 q_1}{p_1 q_2}}.$$

3.2.2. Case 2: $q_1 + q_2 > 1$

From (8) we get

$$\boldsymbol{R_p} = \begin{pmatrix} q_1 + q_2 - 1 & q_1 \\ p_1 & q_2 - q_1 \\ p_2 & 0 \\ 0 & p_2 \end{pmatrix}.$$

If we denote by $\boldsymbol{\theta}^{(i)}$ the vector of parameters corresponding to $\boldsymbol{R}_{\boldsymbol{p}}^{(i)}$, $i \in \{1, 2\}$ we get

$$\boldsymbol{\theta}^{(1)} = (1, 0, 0, -(q_1 + q_2 - 1 - q_1 q_2)/(q_1 q_2 p_1 p_2)) \quad \text{and} \quad \boldsymbol{\theta}^{(2)} = (1, 0, 0, 1/(p_1 q_2)),$$

which yields

$$-\frac{q_1 + q_2 - 1 - q_1 q_2}{q_1 q_2 p_1 p_2} \leq \theta_{12} \leq \frac{1}{p_1 q_2}.$$

Finally we get the following well-known bounds for the correlation:

$$-\sqrt{\frac{p_1 p_2}{q_1 q_2}} \leq \rho_{12} \leq \sqrt{\frac{p_2 q_1}{p_1 q_2}}.$$

It is simple to verify that, in Cases 1 and 2, the ray mass functions are the lower and upper Fréchet–Hoeffding bounds of the Fréchet class $\mathcal{F}(p_1, p_2)$.

4. Multivariate Bernoulli variables with some given moments

We now extend our method to characterize the class of multivariate Bernoulli variables with a given set of moments of any order. The constraints $\mathrm{E}(X_1) = p_1, \dots, \mathrm{E}(X_d) = p_d$ allow us to obtain an interesting interpretation of the matrix $H_{1\boldsymbol{p}}$ of (5). We have

$$\mathrm{E}(X_i) = \sum_{(x_1,\dots,x_d)\in\mathcal{S}_d} x_i f_{\boldsymbol{p}}(x_1, \dots, x_d).$$

For each $\boldsymbol{f}_{\boldsymbol{p}} \in \mathcal{F}(p_1, \dots, p_d)$ and all $i \in \{1, \dots, d\}$, we have $\boldsymbol{x}_i^\top \boldsymbol{f}_{\boldsymbol{p}} = p_i$ and $(\boldsymbol{1} - \boldsymbol{x}_i)^\top \boldsymbol{f}_{\boldsymbol{p}} = q_i$, where $\boldsymbol{1}$ is the vector with all the elements equal to 1 and $\boldsymbol{x}_i$ is the vector which contains only the ith element of $\boldsymbol{x} \in \mathcal{S}_d$, e.g., for the bivariate case $\boldsymbol{x}_1 = (0, 1, 0, 1)$ and $\boldsymbol{x}_2 = (0, 0, 1, 1)$. If we consider the odds of the event $X_i = 1$, $\gamma_i = p_i/q_i$ we have $\gamma_i q_i - p_i = 0$. We can write

$$\gamma_i = p_i/q_i, \quad \{\gamma_i(\boldsymbol{1} - \boldsymbol{x}_i)^\top - \boldsymbol{x}_i^\top\}\boldsymbol{f}_{\boldsymbol{p}} = 0.$$

Then we observe that $H_{1\boldsymbol{p}}$ is simply the $d \times 2^d$ matrix whose rows, up to a non-influential multiplicative constant, are $\gamma_i(\boldsymbol{1} - \boldsymbol{x}_i)^\top - \boldsymbol{x}_i^\top$ for $i \in \{1, \dots, d\}$. This approach can be easily generalized to solve the problem of studying the class of Bernoulli variables with given kth order moments, $k \in \{1, \dots, d\}$. Let us consider the class $\mathcal{F}_k(\boldsymbol{\mu}_k)$ of multivariate Bernoulli distributions with given kth order moments $\boldsymbol{\mu}_k = (\mu_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{S}_d, \|\boldsymbol{\alpha}\|_0 = k)$; with this notation $\mathcal{F}_1(\boldsymbol{\mu}_1) \equiv \mathcal{F}(p_1, \dots, p_d)$. Let us denote by f_k a mass function of $\mathcal{F}_k(\boldsymbol{\mu}_k)$.

We observe that

$$\mathrm{E}(X^{\boldsymbol{\alpha}}) = \sum_{(x_1,\dots,x_d)\in\mathcal{S}_d} x_1^{\alpha_1} \cdots x_d^{\alpha_d} f_k(x_1, \dots, x_d),$$

for all $\boldsymbol{\alpha} \in \mathcal{S}_d$ with $\|\boldsymbol{\alpha}\|_0 = k$. It follows that

$$\boldsymbol{x}_{\boldsymbol{\alpha}}^\top \boldsymbol{f}_k = \mu_{\boldsymbol{\alpha}}, \quad (\boldsymbol{1} - \boldsymbol{x}_{\boldsymbol{\alpha}})^\top \boldsymbol{f}_k = 1 - \mu_{\boldsymbol{\alpha}},$$

where $\boldsymbol{x}_{\boldsymbol{\alpha}}$ is the vector which contains the product $x^{\boldsymbol{\alpha}} \equiv x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, $\boldsymbol{x} \in \mathcal{S}_d$, e.g., for the bivariate case $\boldsymbol{x}_{(1,1)} = (0, 0, 0, 1)$. If we consider the odds of the event $\boldsymbol{X}^{\boldsymbol{\alpha}} = 1$, $\gamma_{\boldsymbol{\alpha}} = \mu_{\boldsymbol{\alpha}}/(1 - \mu_{\boldsymbol{\alpha}})$, we have $\gamma_{\boldsymbol{\alpha}}(1 - \mu_{\boldsymbol{\alpha}}) - \mu_{\boldsymbol{\alpha}} = 0$, i.e.,

$$\gamma_{\boldsymbol{\alpha}} = \mu_{\boldsymbol{\alpha}}/(1 - \mu_{\boldsymbol{\alpha}}), \quad \{\gamma_{\boldsymbol{\alpha}}(\boldsymbol{1} - \boldsymbol{x}_{\boldsymbol{\alpha}})^\top - \boldsymbol{x}_{\boldsymbol{\alpha}}^\top\}\boldsymbol{f}_k = 0.$$

Let us denote by $H_{k\boldsymbol{\mu}_k}$ the $\binom{d}{k} \times 2^d$ matrix whose rows are $\gamma_{\boldsymbol{\alpha}}(\boldsymbol{1} - \boldsymbol{x}_{\boldsymbol{\alpha}})^\top - \boldsymbol{x}_{\boldsymbol{\alpha}}^\top$, $\boldsymbol{\alpha} \in \mathcal{S}_d$ for all $\|\boldsymbol{\alpha}\|_0 = k$.

Theorem 2. *Let $f \in \mathbb{F}_d$ be a multivariate d-dimensional Bernoulli distribution. Then f is a mass function with kth order moments $\boldsymbol{\mu}_k$, i.e., $f \in \mathcal{F}_k(\boldsymbol{\mu}_k)$ if and only if there exist non-negative $\lambda_1, \dots, \lambda_{n_{C_{k\boldsymbol{\mu}_k}}}$ adding up to 1 such that*

$$\boldsymbol{f} = \sum_{i=1}^{n_{C_{k\boldsymbol{\mu}_k}}} \lambda_i \boldsymbol{R}_{\boldsymbol{\mu}_k}^{(i)},$$

where $\boldsymbol{R}_{\boldsymbol{\mu}_k}^{(i)} = (R_{\boldsymbol{\mu}_k}^{(i)}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_d) \in \mathcal{F}_k(\boldsymbol{\mu}_k)$ are the normalized extremal rays of the cone $C_{k\boldsymbol{\mu}_k} = \{\boldsymbol{z} : H_{k\boldsymbol{\mu}_k}\boldsymbol{z} = 0, \boldsymbol{z} \in \mathbb{R}_+^d\}$ and $\lambda_1, \dots, \lambda_{n_{C_{1\boldsymbol{\mu}_k}}}$ are non-negative constants adding up to 1.

Proof. The proof is analogous to that of Theorem 1, where $H_{1\boldsymbol{p}}$ is replaced by $H_{k\boldsymbol{\mu}_k}$. □

The generalization of this approach to any given set of moments is straightforward. We describe this generalization in a specific case which is extremely relevant from the point of view of applications. We suppose that first- and second-order moments are specified, viz. $\boldsymbol{p} = (p_1, \dots, p_d)$ and $\boldsymbol{\mu}_2 = (\mu_{12}, \dots, \mu_{d-1,d})$. If the correlations $\boldsymbol{\rho} = (\rho_{12}, \dots, \rho_{d-1,d})$ are specified instead of the second-order moments, the desired correlations ρ_{ij} are transformed into the corresponding desired second-order moments $\mathrm{E}(X_i X_j)$ for all $i, j \in \{1, \dots, d\}$ with $i < j$ using Eq. (2).

To obtain all the multivariate Bernoulli mass functions f which have the given $\boldsymbol{p}$ and $\boldsymbol{\mu}_2$, it is enough to build the matrix $H_{1\boldsymbol{p},2\boldsymbol{\mu}_2}$ as the row-juxtaposition of the matrices $H_{1\boldsymbol{p}}$ and $H_{2\boldsymbol{\mu}_2}$, viz.

$$H_{1\boldsymbol{p},2\boldsymbol{\mu}_2} = \begin{pmatrix} H_{1\boldsymbol{p}} \\ H_{2\boldsymbol{\mu}_2} \end{pmatrix}$$

and to compute the normalized extremal rays of the cone $\{z : H_{1\boldsymbol{p},2\boldsymbol{\mu}_2} z = 0, z \in \mathbb{R}_+^m\}$. If we denote the corresponding ray matrix by $R_{\boldsymbol{p}\boldsymbol{\mu}_2}$, we obtain $\boldsymbol{f} = R_{\boldsymbol{p}\boldsymbol{\mu}_2}\boldsymbol{\lambda}$, with $\boldsymbol{\lambda} = (\lambda_i)$ a vector whose non-negative components add up to 1.

We observe that the choice of a particular solution does not modify the distributions of the sample means and of the sample second-order moments, which depend only on $p_1, \ldots, p_d$ and $\boldsymbol{\mu}_2 = (\mu_{12}, \ldots, \mu_{d-1,d})$ respectively. To explain this point let us consider a random sample $\{(X_{k1}, \ldots, X_{kd}),\ k = 1, \ldots, N\}$ extracted from a randomly selected m-dimensional Bernoulli variable with given first-order moments $\boldsymbol{p} = (p_1, \ldots, p_d)$ and with given second-order moments $\boldsymbol{\mu}_2 = (\mu_{12}, \ldots, \mu_{d-1,d})$. For each $i \in \{1, \ldots, m\}$, the sample mean $\overline{X}_i$ is N^{-1} times a Binomial(N, p_i) and for each $i, j \in \{1, \ldots, n\}$ with $i < j$, the sample second-order moments $\overline{X_i X_j} = \sum_{k=1}^{N}(X_{ki}X_{kj})/N$ is N^{-1} times Binomial(N, μ_{ij}).

Remark 3. In general different distributions which have the same $\boldsymbol{p}$ and $\boldsymbol{\mu}_2$, will have different kth order moments, with $k \geq 3$. As a consequence of Proposition 5 we observe that the minimum (maximum) value of the kth order moment $\boldsymbol{\mu}_k^{\alpha}$ is obtained choosing $\boldsymbol{\lambda}$ equal to one of the vectors $\boldsymbol{e}_i$. It follows that the ray matrix $R_{\boldsymbol{p}\boldsymbol{\mu}_2}$ provides interesting mass functions because they are extremal from the point of view of kth order moments.

This method offers the opportunity to choose the best distribution according to a certain criterion. For example, as the moments of multivariate Bernoulli are always positive, it could be of interest to find one of the distributions with the smallest sum of all the moments with order greater than 2. This problem can be efficiently solved using linear programming techniques [1]. It can be simply stated as

$$\min_{f \in \mathbb{F}_d} \{\mathbf{1}^\top (M^{\otimes d})_{3\ldots d} \boldsymbol{f}\}$$

subject to $H_{1\boldsymbol{p}2\boldsymbol{\mu}_2}\boldsymbol{f} = 0$, where $(M^{\otimes d})_{3\ldots d}$ is the sub-matrix of $M^{\otimes d}$ obtained by selecting the rows corresponding to the kth moments, with $k \geq 3$.

If the margins $\boldsymbol{p}$ and the second order moments $\boldsymbol{\mu}_2$ (or equivalently the correlations $\boldsymbol{\rho}$) are not compatible, the system $H_{1\boldsymbol{p}2\boldsymbol{\mu}_2} f = 0$ does not have any solution. In this case it is possible to search for a feasible $\boldsymbol{\rho}^*$ which is the correlation closest to the desired $\boldsymbol{\rho}$, according to a chosen distance.

Finally it is worth noting that this method provides a geometrical characterization of multivariate Bernoulli variables with specified moments. Given a set of moments S (S contains the chosen moments $\boldsymbol{\alpha} \in \mathcal{S}_d$ and their corresponding values $\boldsymbol{\mu}_\alpha \in \mathbb{R}_+$) and built the corresponding matrix H_S, the mass functions are the subset of the cone

$$C_S = \{z : H_s z = \mathbf{0}, z \in \mathbb{R}_+^m\} = \left\{ \sum_{i=1}^{n_{C_S}} \lambda_i \boldsymbol{R}_i^{(S)}, \lambda_1 \geq 0, \ldots, \lambda_{n_{C_S}} \geq 0 \right\}$$

obtained imposing $\lambda_1 + \cdots + \lambda_{n_{C_S}} = 1$. We show the algorithm and some examples in Section 5.

5. The algorithm and examples

5.1. The algorithm

In this section we briefly describe the algorithm that we use for the examples. The inputs are d, the dimension of the multivariate Bernoulli distributions, and S, the chosen moments $\boldsymbol{\alpha} \in \mathcal{S}_d$ and their corresponding values $\boldsymbol{\mu}_\alpha \in \mathbb{R}_+$.

The main output is the ray matrix R_S. If the chosen moments are not compatible, R_S is empty and the algorithm stops. Otherwise, the algorithm also gives the moments of the ray mass functions A_S, and the bounds for the moments of the distribution which belong to the class of multivariate Bernoulli distributions specified by S.

The algorithm has the following main steps:

Step 1: Construction of the matrix H_S whose rows are given, for all $(\boldsymbol{\alpha}, \boldsymbol{\mu}_\alpha) \in S$, by

$$\gamma_\alpha (1 - x_\alpha)^\top - x_\alpha, \quad \gamma_\alpha = \frac{\mu_\alpha}{1 - \mu_\alpha}.$$

Step 2: Generation of the ray matrix R_S.

Step 3: Computation of the moments of the ray mass functions, $A_S = M^{\otimes d} R_S$.

Step 4: Computation of the bounds of each moment μ_α as the minimum and the maximum values of the corresponding row of A_S.

Steps 1, 3 and 4 are implemented in `SAS/IML`. The rays of the cone C_S are generated using `4ti2` [21]. The software code is available on request. We performed the analysis using a standard laptop (CPU Intel core I7-2620M CPU 2.70GHz 2.70GHz, RAM 8GB).

In this section we show some results corresponding to different multivariate Bernoulli distributions. In Example 5.3.1 we show how to implement the algorithm at each step.

5.2. *Bivariate Bernoulli distribution*

In the bivariate case we have an analytical expression for the ray mass functions, the bounds of θ and the bounds of the linear correlation, as discussed in Section 3.2. Here, we show a numerical example. Let us choose $q_1 = 1/4$ and $q_2 = 1/2$. We have $q_1 + q_2 \leq 1$. Therefore from (3.2.1) we obtain $R_{\boldsymbol{p}}$

$$R_{\boldsymbol{p}} = \begin{pmatrix} 0 & 1/4 \\ 1/2 & 1/4 \\ 1/4 & 0 \\ 1/4 & 1/2 \end{pmatrix}.$$

If we denote by $\boldsymbol{\theta}^{(i)}$ the vector of parameters corresponding to $\boldsymbol{R}_{\boldsymbol{p}}^{(i)}$, $i \in \{1, 2\}$ we obtain $\boldsymbol{\theta}^{(1)} = (1, 0, 0, -8/3)$ and $\boldsymbol{\theta}^{(2)} = (1, 0, 0, 8/3)$, which means $-8/3 \leq \theta_{12} \leq 8/3$. For $\theta_{12} = 8/3$ we find the upper Fréchet–Hoeffding bound. In fact $\boldsymbol{R}_p^{(2)} = (1/4, 1/4, 0, 1/2)$ implies $\boldsymbol{F}^{(2)} = (1/4, 1/2, 1/4, 1)$, which coincides with the upper Fréchet–Hoeffding bound in (7). Finally we get the following bounds for the correlation $-\sqrt{1/3} \leq \rho_{12} \leq \sqrt{1/3}$.

5.3. *Trivariate Bernoulli distributions*

5.3.1. *Case* 1*:* $\boldsymbol{p} = (1/2, 1/2, 1/2)$

Let us consider the case where $d = 3$ and the chosen first-order moments are $\boldsymbol{p} = (1/2, 1/2, 1/2)$. The algorithm has the following main steps.

Step 1: Construction of the matrix H_S, which in this case has three rows and eight columns. For example the row corresponding to $\boldsymbol{\alpha} = (1, 0, 0)$ is $\gamma_\alpha(1 - \boldsymbol{x}_\alpha)^\top - \boldsymbol{x}_\alpha = (1, -1, 1, -1, 1, -1, 1, -1)$, as $\gamma_\alpha = (1/2)/(1 - 1/2) = 1$ and $\boldsymbol{x}_\alpha \equiv \boldsymbol{x}_1 = (0, 1, 0, 1, 0, 1, 0, 1)$. We get

$$H_S = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix}.$$

Step 2: Using `4ti2`, giving H_S as input, we generate the ray matrix R_S where the columns are the ray mass functions $\boldsymbol{R}^{(i)}$, after normalization.

Step 3: We compute the moments of the ray mass functions $\boldsymbol{R}^{(i)}$ as $A_S = M^{\otimes 3} R_S$.

Step 4: We compute the bounds of each moment μ_α as the minimum and the maximum values of the corresponding row of A_S.

The output is

a) the ray matrix R_S, which has 6 ray mass functions $\boldsymbol{R}^{(1)}, \ldots, \boldsymbol{R}^{(6)}$, in Table 1;

b) the moments of the ray mass functions $\boldsymbol{R}^{(1)}, \ldots, \boldsymbol{R}^{(6)}$ in Table 2;

c) the bounds for the moments of the distribution which are $0 \leq \mu_{ij} \leq 0.5$ for all $i, j \in \{1, 2, 3\}$ with $i < j$ and $0 \leq \mu_{123} \leq 0.5$.

Table 1: Ray mass functions $d = 3, \boldsymbol{p} = (1/21/2, 1/2)$.

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{R}^{(1)}$	$\boldsymbol{R}^{(2)}$	$\boldsymbol{R}^{(3)}$	$\boldsymbol{R}^{(4)}$	$\boldsymbol{R}^{(5)}$	$\boldsymbol{R}^{(6)}$
0	0	0	0	0	0	0	0.5	0.25
1	0	0	0	0	0.5	0.25	0	0
0	1	0	0	0.5	0	0.25	0	0
1	1	0	0.5	0	0	0	0	0.25
0	0	1	0.5	0	0	0.25	0	0
1	0	1	0	0.5	0	0	0	0.25
0	1	1	0	0	0.5	0	0	0.25
1	1	1	0	0	0	0.25	0.5	0

Table 2: Moments of the ray mass functions $d = 3, \boldsymbol{p} = (1/2, 1/2, 1/2)$.

α_1	α_2	α_3	Order	$\mu_{(1)}$	$\mu_{(2)}$	$\mu_{(3)}$	$\mu_{(4)}$	$\mu_{(5)}$	$\mu_{(6)}$
0	0	0	0	1	1	1	1	1	1
1	0	0	1	0.5	0.5	0.5	0.5	0.5	0.5
0	1	0	1	0.5	0.5	0.5	0.5	0.5	0.5
1	1	0	2	0.5	0	0	0.25	0.5	0.25
0	0	1	1	0.5	0.5	0.5	0.5	0.5	0.5
1	0	1	2	0	0.5	0	0.25	0.5	0.25
0	1	1	2	0	0	0.5	0.25	0.5	0.25
1	1	1	3	0	0	0	0.25	0.5	0

By construction all the mass functions have first-order moments equal to 0.5. The ray density $\boldsymbol{R}^{(5)}$ is the upper Fréchet–Hoeffding bound of the class and it has correlations $\rho_{12} = \rho_{13} = \rho_{23} = 1$. For dimension $d > 2$, the lower Fréchet–Hoeffding bound is not in general a distribution function. For $d > 2$, the lower and upper Fréchet–Hoeffding bounds F_L and F_U of $\mathcal{F}(p_1, \dots, p_d)$ are respectively defined by

$$F_L(\boldsymbol{x}) = \max\{F_1(x_1) + \cdots + F_d(x_d) - d + 1, 0\}, \quad F_U(\boldsymbol{x}) = \min\{F_1(x_1), \dots, F_d(x_d)\},$$

where $\boldsymbol{x} = (x_1, \dots, x_d) \in \{0,1\}^d$. For any $F_{\boldsymbol{p}} \in \mathcal{F}(p_1, p_2, \dots, p_d)$, one has $F_L(\boldsymbol{x}) \leq F_{\boldsymbol{p}}(\boldsymbol{x}) \leq F_U(\boldsymbol{x})$ for all $\boldsymbol{x} \in \{0,1\}^d$. Both the rays $\boldsymbol{R}^{(4)}$ and $\boldsymbol{R}^{(6)}$ are mass functions of not correlated random variables, but that have different third-order moments, 0.25 and 0 respectively. The independent distribution is inside the convex hull and it is the distribution $\boldsymbol{R}^{(4)}/2 + \boldsymbol{R}^{(6)}/2$. The rays $\boldsymbol{R}^{(1)}$, $\boldsymbol{R}^{(2)}$ and $\boldsymbol{R}^{(3)}$ exhibit both negative and positive correlations, for example for $\boldsymbol{R}^{(1)}$ we have $\rho_{12} = 1$ and $\rho_{13} = \rho_{23} = -1$.

Using Eq. (4) we get $-1 \leq \rho_{ij} \leq 1$ for all $i, j \in \{1,2,3\}$ with $i < j$. Finally by Proposition 2 we compute the $\boldsymbol{\theta}^{(i)}$ corresponding to the ray mass function $\boldsymbol{R}^{(1)}, \dots, \boldsymbol{R}^{(6)}$, which are reported in Table 3.

In the Appendix we give two examples where we also choose the second-order moments.

Table 3: $\boldsymbol{\theta}^{(i)}$ of the ray mass function $\boldsymbol{R}^{(i)}$, $i \in \{1, \dots, 6\}$ $d = 3, \boldsymbol{p} = (1/2, 1/2, 1/2)$.

α_1	α_2	α_3	$\theta_\alpha^{(1)}$	$\theta_\alpha^{(2)}$	$\theta_\alpha^{(3)}$	$\theta_\alpha^{(4)}$	$\theta_\alpha^{(5)}$	$\theta_\alpha^{(6)}$
0	0	0	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
1	1	0	4	–4	–4	0	4	0
0	0	1	0	0	0	0	0	0
1	0	1	–4	4	–4	0	4	0
0	1	1	–4	–4	4	0	4	0
1	1	1	0	0	0	–8	0	8

5.3.2. Case 2: $\boldsymbol{p} = (1/4, 1/7, 1/3)$

As the last example of trivariate Bernoulli distributions we consider $\boldsymbol{p} = (1/4, 1/7, 1/3)$. The ray matrix $R_{\boldsymbol{p}}$, rounded to the third decimal digit, has 11 ray mass functions, viz.

$$R_{\boldsymbol{p}} = \begin{pmatrix} 0.274 & 0.417 & 0.417 & 0.524 & 0.524 & 0.56 & 0.607 & 0.607 & 0.667 & 0.667 & 0.637 \\ 0.25 & 0.107 & 0.25 & 0 & 0 & 0.107 & 0 & 0.06 & 0 & 0 & 0 \\ 0.143 & 0 & 0 & 0 & 0.143 & 0 & 0.06 & 0 & 0 & 0 & 0 \\ 0 & 0.143 & 0 & 0.143 & 0 & 0 & 0 & 0 & 0 & 0 & 0.03 \\ 0.333 & 0.333 & 0.19 & 0.226 & 0.083 & 0.19 & 0 & 0 & 0 & 0.083 & 0 \\ 0 & 0 & 0 & 0.107 & 0.25 & 0 & 0.25 & 0.19 & 0.19 & 0.107 & 0.22 \\ 0 & 0 & 0.143 & 0 & 0 & 0 & 0.083 & 0.143 & 0.083 & 0 & 0.113 \\ 0 & 0 & 0 & 0 & 0 & 0.143 & 0 & 0 & 0.06 & 0.143 & 0 \end{pmatrix}.$$

Using Eq. (4) we get

$$-0.236 \leq \rho_{12} \leq 0.707, \quad -0.408 \leq \rho_{13} \leq 0.816, \quad -0.289 \leq \rho_{23} \leq 0.577.$$

We also report the correlations of random variables corresponding to each ray mass function in Table 4.

Table 4: Correlations ρ_{ij} of random variables whose mass functionis $\boldsymbol{R}^{(i)}$, $i \in \{1, \ldots, 11\}$ $d = 3$, $\boldsymbol{p} = (1/4, 1/7, 1/3)$.

i	j	$\boldsymbol{R}^{(1)}$	$\boldsymbol{R}^{(2)}$	$\boldsymbol{R}^{(3)}$	$\boldsymbol{R}^{(4)}$	$\boldsymbol{R}^{(5)}$	$\boldsymbol{R}^{(6)}$	$\boldsymbol{R}^{(7)}$	$\boldsymbol{R}^{(8)}$	$\boldsymbol{R}^{(9)}$	$\boldsymbol{R}^{(10)}$	$\boldsymbol{R}^{(11)}$
1	2	–0.236	0.707	–0.236	0.707	–0.236	0.707	–0.236	–0.236	0.157	0.707	–0.039
1	3	–0.408	–0.408	–0.408	0.117	0.816	0.292	0.816	0.525	0.816	0.816	0.671
2	3	–0.289	–0.289	0.577	–0.289	–0.289	0.577	0.217	0.577	0.577	0.577	0.397

The mass function $\boldsymbol{R}^{(10)}$ corresponds to the upper Fréchet–Hoeffding bound, while in this case the lower Fréchet–Hoeffding bound is also a distribution and it has mass function $\boldsymbol{R}^{(1)}$. Consequently the ray mass functions exhibit both the maximal and the minimal correlations.

5.4. Multivariate $d = 5$ Bernoulli distributions

Let us consider the case $p = (1/2, 1/2, 1/2, 1/2, 1/2)$. We obtain 2712 ray mass functions. If we choose the additional constraints $\rho_{12} = 0.3, \rho_{13} = 0.2, \rho_{14} = 0.2, \rho_{15} = 0.1, \rho_{23} = -0.2, \rho_{24} = 0.3, \rho_{25} = 0.2, \rho_{34} = 0.2, \rho_{35} = 0.1$ and $\rho_{45} = -0.2$, we obtain 25,100 ray mass functions.

5.5. Multivariate $d \geq 6$ Bernoulli distributions

For $d = 6$ and $\boldsymbol{p} = (1/21/2, 1/2, 1/2, 1/2, 1/2)$ we obtain 707,264 ray mass functions. In general we observe that if the number of rays is too large with respect to the available computer power and if the objective can be reduced to the problem of finding just one mass function $f \in \mathbb{F}_d$ with given margins p and second order moments $\boldsymbol{\mu}_2$, it is enough to find one solution of the system

$$(M^{\otimes d})_1 \boldsymbol{f} = \boldsymbol{p}, \quad (M^{\otimes d})_2 \boldsymbol{f} = \boldsymbol{\mu}_2,$$

using standard linear programming tools; see, e.g., [1].

In [3] trivariate Bernoulli distributions with respect to four scenarios relative to different choices of p_1, p_2, p_3, ρ_{12}, ρ_{23} are considered. In Table 5 as an example, we report their first scenario. The other scenarios give similar results. We observe that none of the methods considered can reach the theoretical bounds computed by [3]. On the contrary our method computes the ray mass functions which are extremal with respect to the moments. We get the four ray mass functions $\boldsymbol{R}^{(1)}, \ldots, \boldsymbol{R}^{(4)}$ that we report in Table 6.

5.6. Comparison with some methods for simulating correlated binary variables

We compare our method with those that have been considered on p. 201, in Section 4 of [3], namely (i) the discretized normal method of Emrich and Piedmonte [6]; (ii) the extension of the beta-binomial multivariate binary model [18]; and (iii) the method of Park [17].

It is easy to verify that if we consider $\boldsymbol{X}_{(1)} \sim \boldsymbol{R}^{(1)}$ we obtain $\rho_{13} \approx -0.13$ and if we consider $\boldsymbol{X}_{(4)} \sim \boldsymbol{R}^{(4)}$ we obtain $\rho_{13} \approx 0.61$ where -0.13 and 0.61 are the theoretical lower and upper bounds for this case.

Table 5: Range of ρ_{13} given $p_1, p_2, p_3, \rho_{12}, \rho_{23}$; CJ theoretical bounds, EP method of Emrich and Piedmonte, Q method of Qaqish and PPS method of Park.

p_1	p_2	p_3	ρ_{12}	ρ_{23}	CJ	EP	Q	PPS
0.2	0.3	0.4	0.1	0.7	$(-0.13, 0.61)$	$(-0.12, 0.33)$	$(-0.09, 0.15)$	$(0, 0.17)$

Table 6: Ray mass functions $d = 3, \boldsymbol{p} = (0.2, 0.3, 0.4), \rho_{12} = 0.1, \rho_{23} = 0.7$.

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{R}^{(1)}$	$\boldsymbol{R}^{(2)}$	$\boldsymbol{R}^{(3)}$	$\boldsymbol{R}^{(4)}$
0	0	0	0.455	0.455	0.577	0.577
1	0	0	0.122	0.122	0	0
0	1	0	0	0.023	0	0.023
1	1	0	0.023	0	0.023	0
0	0	1	0.123	0.123	0.001	0.001
1	0	1	0	0	0.122	0.122
0	1	1	0.222	0.199	0.222	0.199
1	1	1	0.055	0.078	0.055	0.078

6. Conclusion

The proposed approach can be applied to any given set of moments, even of different orders. The method has the advantage of making it possible to generate all the mass functions which belong to a given class of multivariate Bernoulli variables, where the class is specified by a set of moments of any order. It is clear that this approach is different from obtaining one solution using linear programming techniques.

The availability of all the mass functions has an important impact on practical applications. Let us suppose that a researcher needs a sample of N observations from a multivariate Bernoulli distributions with given moments to analyze extensions or alternatives to current estimating methods. She will get more robust results if she considers $k \in \{2, \ldots, N\}$ different distributions and a sample of proper size from each of them, instead of taking a sample of size N from a single distribution, which can be obtained using a linear programming approach. In this way she will consider a variety of distributions that will be different from each other, relative to the unspecified moments, i.e., the moments that do not belong to S. Depending on how large N is, the selected distributions could include the rays and some other randomly generated distributions, e.g., randomly selecting some λ vectors.

Acknowledgments

Roberto Fontana wishes to thank Prof. Antonio Di Scala (Politecnico di Torino) and Prof. Giovanni Pistone (Collegio Carlo Alberto, Moncalieri) for helpful discussions. The authors wish to thank Prof. Ilya Molchanov (Universität Bern) for having provided some useful references. They also thank Prof. Enrico Carlini (Politecnico di Torino) and Levent Tuncel (University of Waterloo) for providing a reference on the bounds of the number of extremal rays.

Appendix A. Examples with given higher order moments

Case 1*:* $\boldsymbol{p} = (1/2, 1/2, 1/2)$ *and* $\rho_{12} = \rho_{13} = \rho_{23} = 0$

Let us consider the case in which X_1, X_2, X_3 must have means equal to 1/2 and must also be not correlated, which means that the second-order moments are all equal to 1/4. We obtain the two ray mass functions $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ that we report in Table A.1. We know that one solution to this problem is the uniform distribution whose mass function is $f(\boldsymbol{x}) = 1/8$, $\boldsymbol{x} \in \mathcal{S}_3$. We find it as $\lambda_1 \boldsymbol{R}^{(1)} + \lambda_2 \boldsymbol{R}^{(2)}$ with $\lambda_1 = \lambda_2 = 1/2$.

The third-order moments of $\boldsymbol{X}_{(1)} \sim \boldsymbol{R}^{(1)}$ and $\boldsymbol{X}_{(2)} \sim \boldsymbol{R}^{(2)}$ are reported in Table A.2. Therefore any mass function in this class has the third-order moment between 0 and 0.25.

Table A.1: Ray mass functions $d = 3$, $\boldsymbol{p}_i = 1/2$, $\rho_{ij} = 0$, $i, j \in \{1, 2, 3\}$, $i < j$.

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{R}^{(1)}$	$\boldsymbol{R}^{(2)}$
0	0	0	0	0.25
1	0	0	0.25	0
0	1	0	0.25	0
1	1	0	0	0.25
0	0	1	0.25	0
1	0	1	0	0.25
0	1	1	0	0.25
1	1	1	0.25	0

Table A.2: Third-order moments of the ray mass functions $d = 3$, $\boldsymbol{p} = (1/2, 1/2, 1/2)$, $\rho_{ij} = 0$, $i, j \in \{1, 2, 3\}$, $i < j$.

α_1	α_2	α_3	Order	$\mu_{(1)}$	$\mu_{(2)}$
1	1	1	3	0.25	0

Case 2: $\boldsymbol{p} = (1/2, 1/2, 1/2)$ *and* $\rho_{12} = 0.2$, $\rho_{13} = -0.3$, $\rho_{23} = 0.4$

Let us consider the case in which $X_{(1)}, X_{(2)}, X_{(3)}$ must have means equal to 1/2 and correlations $\rho_{12} = 0.2, \rho_{13} = -0.3$ and $\rho_{23} = 0.4$. We obtain the 2 ray mass functions $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ reported in Table A.3. For example if we choose $\lambda_1 = \lambda_2 = 1/2$ the corresponding mass function is

$$\boldsymbol{f}_{\boldsymbol{p}}^{\top} = (0.1625,\ 0.1875,\ 0.0125,\ 0.1375,\ 0.1375,\ 0.0125,\ 0.1875,\ 0.1625)\,.$$

Table A.3: Ray mass functions $d = 3$, $\boldsymbol{p} = (1/2, 1/2, 1/2)$, $\rho_{12} = 0.2$, $\rho_{13} = -0.3$, $\rho_{23} = 0.4$.

$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{R}^{(1)}$	$\boldsymbol{R}^{(2)}$
0	0	0	0.15	0.175
1	0	0	0.2	0.175
0	1	0	0.025	0
1	1	0	0.125	0.15
0	0	1	0.15	0.125
1	0	1	0	0.025
0	1	1	0.175	0.2
1	1	1	0.175	0.15

The third-order moments of $\boldsymbol{X}_{(1)} \sim \boldsymbol{R}^{(1)}$ and $\boldsymbol{X}_{(2)} \sim \boldsymbol{R}^{(2)}$ are reported in Table A.4. Therefore any mass function in this class has the third-order moment between 0.150 and 0.175.

References

[1] M. Berkelaar, K. Eikland, P. Notebaert, lpsolve: Open source (mixed-integer) linear programming system, Eindhoven U. of Technology (2004).

[2] N.R. Chaganty, H. Joe, Efficiency of generalized estimating equations for binary responses, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 66 (2004) 851–860.

[3] N.R. Chaganty, H. Joe, Range of correlation matrices for dependent Bernoulli random variables, Biometrika 93 (2006) 197–206.

Table A.4: Third-order moments of the ray mass functions $d = 3$, $p_i = 1/2$, $\rho_{12} = 0.2$, $\rho_{13} = -0.3$, $\rho_{23} = 0.4$.

α_1	α_2	α_3	Order	$\mu_{(1)}$	$\mu_{(2)}$
1	1	1	3	0.175	0.15

[4] M. Crowder, On the use of a working correlation matrix in using generalised linear models for repeated measures, Biometrika 82 (1995) 407–410.
[5] G. Dall'Aglio, S. Kotz, G. Salinetti, Advances in Probability Distributions with Given Marginals: Beyond the Copulas, Springer, Berlin, 2012.
[6] L.J. Emrich, M.R. Piedmonte, A method for generating high-dimensional multivariate binary variates, Amer. Statist. 45 (1991) 302–304.
[7] M.E. Haynes, R.T. Sabo, N.R. Chaganty, Simulating dependent binary variables through multinomial sampling, J. Stat. Comput. Simul. 86 (2016) 510–523.
[8] R. Hemmecke, On the computation of Hilbert bases of cones, Math. Softw. ICMS (2002) 307–317.
[9] S.H. Kang, S.H. Jung, Generating correlated binary variables with complete specification of the joint distribution, Biometrical J. 43 (2001) 263–269.
[10] R. Lachieze-Rey, I. Molchanov, Regularity conditions in the realisability problem with applications to point processes and random closed sets, Ann. Appl. Probab. 25 (2015) 116–149.
[11] K.Y. Liang, S.L. Zeger, Longitudinal data analysis for discrete and continuous outcomes, Biometrics (1986) 121–130.
[12] K.Y. Liang, S.L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[13] S.R. Lipsitz, G.M. Fitzmaurice, L. Sleeper, L.P. Zhao, Estimation methods for the joint distribution of repeated binary observations, Biometrics (1995) 562–570.
[14] P. McMullen, The maximum numbers of faces of a convex polytope, Mathematika 17 (1970) 179–184.
[15] R.B. Nelsen, An Introduction to Copulas, 2nd ed., Springer, Berlin, 2006.
[16] S.D. Oman, Easily simulated multivariate binary distributions with given positive and negative correlations, Comput. Stat. Data Anal. 53 (2009) 999–1005.
[17] G.C. Park, T. Park, D.W. Shin, A simple method for generating correlated binary variates, Amer. Statist. 50 (1996) 306–310.
[18] B.F. Qaqish, A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations, Biometrika 90 (2003) 455–463.
[19] S. Sharakhmetov, R. Ibragimov, A characterization of joint distribution of two-valued random variables and its applications, J. Multivariate Anal. 83 (2002) 389–408.
[20] J. Shults, Simulating longer vectors of correlated binary random variables via multinomial sampling (2016).
[21] 4ti2 Team, `4ti2`—A Software Package for Algebraic, Geometric and Combinatorial Problems on Linear Spaces, Available at `www.4ti2.de`, 2015.
[22] J.L. Teugels, Some representations of the multivariate Bernoulli and binomial distributions, J. Multivariate Anal. 32 (1990) 256–268.