Exchangeable Bernoulli distributions: High dimensional simulation, estimation, and testing

(Article begins on next page)

Exchangeable Bernoulli distributions: high dimensional simulation, estimation, and testing

R.FONTANA[1]
Department of Mathematical Sciences G. Lagrange,
Politecnico di Torino.
P. SEMERARO
Department of Mathematical Sciences G. Lagrange,
Politecnico di Torino.

October 25, 2022

[1]Corresponding author: Roberto Fontana Department of Mathematical Sciences G. Lagrange, Politecnico di Torino. Email: roberto.fontana@polito.it

Abstract

High dimensional simulation of exchangeable multivariate Bernoulli distributions is a challenging and important issue in applications, for example in credit risk models. The main contributions of this paper are, even for high dimensions, algorithms to sample from exchangeable multivariate Bernoulli distributions and to determine the distributions and the bounds of a wide class of indices and measures of probability mass functions. Unlike the algorithms present in the literature the proposed method gives the possibility to simulate also from negatively correlated distributions. Such a method is based on the geometrical structure of the class of exchangeable Bernoulli probability mass functions, which are points in a convex polytope whose extremal points are analytically known. Estimation and testing are also addressed.

Keywords: Exchangeable Bernoulli distribution, convex polytope, extremal points, uniform sampling, simulation.

1 Introduction

Multivariate Bernoulli distributions are important in many fields such as human sciences, finance, biology and medicine. Binary data come out from social survey responses with "yes/no" questions, from responses to treatments in clinical trial, from measurements of genetic or epigenetic variations among individuals and from credit risk models, see [1]. In many applications correlations among variables cannot be ignored. As a consequence simulation of dependent multivariate Bernoulli distributions is an important issue in applications. In this paper we address this issue under the condition of exchangeability and also for high dimension. Exchangeable multivariate Bernoulli variables are very important for example in finance, where they represent the indicators of default of obligors in a credit risk portfolio. Usually credit risk portfolios include a large number of obligors, thus the necessity to work in high dimension (see, e.g. [2]). All credit risk models can be represented as Bernoulli mixture models and the simplest Bernoulli mixture model implies infinite exchangeability, see [3] as a standard reference. De Finetti's representation theorem asserts that if we have an infinite sequence of exchangeable Bernoulli variables, they can be seen as a mixture of independent and identically distributed Bernoulli variables. Using this representation it is possible to easily estimate, test and simulate exchangeable Bernoulli variables also in high dimension. The only drawback of De Finetti's representation is that it requires an infinite sequence. A finite form of De Finetti's theorem has been given in [4], based on the geometrical structure of the class $\mathcal{E}_d$ of d-dimensional exchangeable Bernoulli variables. In fact, $\mathcal{E}_d$ is proven to be a d-dimensional simplex and therefore each probability mass function in $\mathcal{E}_d$ has a unique representation as a mixture of the $d+1$ extremal points. This result has been extended in [5] where it is proved that the class of multivariate Bernoulli probability mass functions with some given moments is a convex polytope, i.e. a convex hull of extremal points. Furthermore, [6] provides an analytical expression of the extremal probability mass functions under exchangeability. The representation of exchangeable Bernoulli probability mass functions as points in a convex polytope and the ability of explicitly finding the extremal points are key steps in simulating, estimating and testing, also in high dimension.

The class $\mathcal{E}_d(p)$ of exchangeable multivariate Bernoulli probability mass functions with given mean p is extremely relevant for the applications. For example in homogeneous credit risk portfolios the obligors belong to the same class of rating, and therefore have the same given marginal default probability. In this paper we study the statistical properties of $\mathcal{E}_d$ and $\mathcal{E}_d(p)$, building on their geometrical structure. We show that we can find the values of a wide class of measures as convex combinations of their values on the extremal points. As a consequence we are not only able to find their extremal values on the class, but we can also numerically find their distribution across the class. For low dimensions the distribution is exact while for high dimension it is based on sampling.

Two important issues in the statistical literature are the simulation of high dimensional binary data with given correlation and the simulation of negatively correlated binary data. The geometrical structure of $\mathcal{E}_d(p)$ allows us to easily construct parametrical families of probability mass functions able to cover the whole correlation range. We can therefore select a multivariate Bernoulli probability mass function with any given correlation in the whole range of admissible correlations. This overcomes the limit of the models used in the literature to simulate exchangeable binary data, that only cover positive correlations. This is not a big issue in high dimensions, since at the limit negative correlation is not possible, but it comes out to be a strong limitation for low dimensions: as an example the three dimensional exchangeable Bernoulli random variables with mean $p = \frac{1}{3}$ admit negative correlations up to $\rho = -\frac{1}{2}$. For example, negatively correlated Bernoulli variables are of interest in insurance where they model safe dependent structures among risks, [7].

The geometrical structure also turns out to be crucial to perform uniform sampling from a convex polytope. Even for high dimensionality, uniform sampling can be used to find the approximate distribution across the class of general statistical indices, that cannot be expressed as combinations of their extremal values.

The explicit form of the extremal points is important in estimation and testing, which are also addressed. We find the maximum likelihood estimator for a probability mass function in the classes $\mathcal{E}_d$ and $\mathcal{E}_d(p)$ and provide a generalized likelihood test for the null hypothesis of exchangeability or exchangeability with a given mean.

The results can be extended to the more general framework of partially exchangeable multivariate Bernoulli distributions. To give an overall idea, we show that partially exchangeable Bernoulli distributions can also be seen as points in a convex polytope, but we leave their investigation to future research.

The paper is organized as follows. Section 2 introduces the polytope of exchangeable Bernoulli distributions and studies the distribution of statistical indices and measures across the class. Section 3 addresses high dimensional simulations and presents an application to credit risk. Section 4 finds the maximum likelihood estimator of exchangeable distributions using the representation of a probability mass function as linear combinations of the extremal points and provides a generalized likelihood ratio test for exchangeability. Section 5 opens the way to the generalization of this work to partially exchangeable Bernoulli distributions. Section 6 concludes the paper. A collection of SAS/IML modules and R scripts have been developed and it is available in the supplementary material. It can be used for working with different dimensions d and probabilities p.

2 Exchangeable Bernoulli distributions

Let $\mathcal{B}_d$ and $\mathcal{E}_d \subset \mathcal{B}_d$ be the classes of d-dimensional Bernoulli distributions and d-dimensional exchangeable Bernoulli distributions, respectively. Let $\mathcal{E}_d(p) \subset \mathcal{E}_d$ be the class of d-dimensional exchangeable Bernoulli distributions with the same Bernoulli marginal distributions $B(p)$. If $\boldsymbol{X} = (X_1, \dots, X_d)$ is a random vector with joint distribution in $\mathcal{E}_d$, we denote

- its cumulative distribution function (cdf) by F and its probability mass function (pmf) by f;
- the column vector which contains the values of f over $\mathcal{X}_d := \{0,1\}^d$, by $\boldsymbol{f}^{\mathcal{X}} = (f^{\mathcal{X}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}_d)$; we make the non-restrictive hypothesis that the set $\mathcal{X}_d$ of 2^d binary vectors is ordered according to the reverse-lexicographical criterion. For example $\mathcal{X}_2 = \{00, 10, 01, 11\}$ and $\mathcal{X}_3 = \{000, 100, 010, 110, 001, 101, 011, 111\}$. We assume that vectors are column vectors;
- the expected value of X_i as p, $\mathrm{E}[X_i] = p$, $i = 1, \dots, d$ and $q = 1 - p$;

Let us consider a pmf $f \in \mathcal{E}_d$. Since, by exchangeability, $f(\boldsymbol{x}) = f(\sigma(\boldsymbol{x}))$ for any permutation σ on $\{1, \dots, d\}$, any mass function f in $\mathcal{E}_d$ defines $f_i := f(\boldsymbol{x})$ if $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathcal{X}_d$ and $\#\{x_j : x_j = 1\} = i$, $i = 0, 1, \dots, d$. Therefore we identify a mass function f in $\mathcal{E}_d$ with the corresponding vector $\boldsymbol{f} := (f_0, \dots, f_d)$.

Let $\mathcal{S}_d$ be the class of distributions p_Y of the sum $Y = \sum_{i=1}^d X_i$ with $\boldsymbol{X} \in \mathcal{E}_d$. The pmf p_Y is a discrete distribution on $\{0, \dots, d\}$. Let $P(Y = j) = p_Y(j) = p_j$ and $\boldsymbol{p}_Y = (p_0, \dots, p_d)$. The map:

$$\begin{aligned} H : \mathcal{E}_d &\to \mathcal{S}_d \\ f_j &\to p_j = \binom{d}{j} f_j. \end{aligned} \tag{2.1}$$

is a one-to-one correspondence between $\mathcal{E}_d$ and $\mathcal{S}_d$. It is also a one-to-one correspondence between $\mathcal{E}_d(p)$ and $\mathcal{S}_d(p)$. Let $\mathcal{D}_d$ and $\mathcal{D}_d(dp)$ be the classes of discrete distributions and of discrete distributions with mean dp, respectively. The paper [6] proves that the three classes $\mathcal{E}_d$, $\mathcal{S}_d$ and $\mathcal{D}_d$ ($\mathcal{E}_d(p)$, $\mathcal{S}_d(p)$ and $\mathcal{D}_d(dp)$) are essentially the same class, i.e.

$$\mathcal{E}_d \leftrightarrow \mathcal{S}_d \equiv \mathcal{D}_d. \tag{2.2}$$

and

$$\mathcal{E}_d(p) \leftrightarrow \mathcal{S}_d(p) \equiv \mathcal{D}_d(dp). \tag{2.3}$$

For the sake of simplicity we write $\boldsymbol{X} \in \mathcal{E}_d$ or $\boldsymbol{X} \in \mathcal{E}_d(p)$ meaning that the distribution of $\boldsymbol{X}$ belongs to $\mathcal{E}_d$ or $\mathcal{E}_d(p)$, respectively. Analogously for $Y \in \mathcal{S}_d$ or $Y \in \mathcal{S}_d(p)$.

2.1 Polytope of Exchangeable Bernoulli distributions

This section summarizes the geometrical structure of the family of exchangeable Bernoulli distributions. We consider here the class $\mathcal{E}_d(p)$ of exchangeable Bernoulli distributions with given mean p. In [6] further details can be found including the proof of Proposition 2.1 (which is also reported in Appendix A) and the analysis of exchangeable Bernoulli distributions with given mean and correlation.

We recall that a polytope (or more specifically a d-polytope) is the convex hull of a finite set of points in $\mathbb{R}^d$ called the extremal points of the polytope. We say that a set of k points is affinely independent if no one point can be expressed as a linear convex combination of the others. For example, three points are affinely independent if they are not on the same line, four points are affinely independent if they are not on the same plane, and so on. The convex hull of $k+1$ affinely independent points is called a simplex or k-simplex. For example, the line segment joining two points is a 1-simplex, the triangle defined by three points is a 2-simplex, and the tetrahedron defined by four points is a 3-simplex. A complete reference on computational geometry is [8].

The class of discrete distributions $\boldsymbol{p} = (p_0, \dots, p_d)$ on $\{0, \dots, d\}$ is the d-simplex $\Delta_d = \{\boldsymbol{p} : p_i \geq 0, \sum_{i=0}^{d} p_i = 1\}$, with extremal points $\boldsymbol{g}_j = (0, \dots, 1, \dots, 0)$, $j = 0, \dots, d$. By means of the equivalence $\mathcal{S}_d \equiv \mathcal{D}_d$ and the map H we have that the class $\mathcal{E}_d$ is a d-simplex. In 1977, [4] proved that $\mathcal{E}_d$ has $d+1$ extremal points $\boldsymbol{g}'_0, \dots, \boldsymbol{g}'_d$, where $\boldsymbol{g}'_j = (g'_j(\boldsymbol{x}); \boldsymbol{x} \in \chi_d)$ is the measure

$$g'_j(\boldsymbol{x}) = \begin{cases} \frac{1}{\binom{d}{j}} & \text{if } \#\{x_h : x_h = 1\} = j \\ 0 & \text{otherwise} \end{cases}.$$

The same result can be obtained by inverting the map H.

The class $\mathcal{S}_d(p)$ is a d-polytope, i.e. for any $Y \in \mathcal{S}_d(p)$ there exist $\lambda_1, \dots, \lambda_{n_p} \geq 0$ summing up to 1 and $\boldsymbol{r}_j \in \mathcal{S}_d(p)$ such that

$$\boldsymbol{p}_Y = \sum_{j=1}^{n_p} \lambda_j \boldsymbol{r}_j. \tag{2.4}$$

We call $\boldsymbol{r}_j = (r_j(0), \dots, r_j(d))$, $j = 1, \dots, n_p$ the extremal points or the extremal densities of $\mathcal{S}_d(p)$.

The extremal points of $\mathcal{S}_d(p)$ have been analytically found in [6] for any dimension. Proposition 2.1 provides them and their number explicitely. Because of the importance of this result to our aims, its proof, given in [6], is reported in Appendix A.

Proposition 2.1. *The extremal points $\boldsymbol{r}_j$ in (2.4) 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*

$$r_j(y) = r_{(j_1,j_2)}(y) = \begin{cases} \frac{j_2-pd}{j_2-j_1} & y = j_1 \\ \frac{pd-j_1}{j_2-j_1} & y = j_2 \\ 0 & otherwise \end{cases}. \tag{2.5}$$

If pd is integer the extremal densities contain also

$$r_{pd}(y) = \begin{cases} 1 & y = pd \\ 0 & otherwise \end{cases}. \tag{2.6}$$

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 $n_p = d^2p(1-p) + 1$ extremal points.

Thanks to the one-to-one correspondence H defined in Eq. (2.1) also $\mathcal{E}_d(p)$ is a d-polytope, i.e. for any $\boldsymbol{X} \in \mathcal{E}_d(p)$ there exist $\lambda_1, \ldots, \lambda_{n_p} \geq 0$ summing up to 1 and $\boldsymbol{e}_i \in \mathcal{E}_d(p)$ such that

$$\boldsymbol{f} = \sum_{j=1}^{n_p} \lambda_j \boldsymbol{e}_j. \tag{2.7}$$

We call $\boldsymbol{e}_j$ the extremal points of $\mathcal{E}_d(p)$. The map H allows us to explicitly find the extremal points $\boldsymbol{e}_j$. They are:

$$e_j(\boldsymbol{x}) = \begin{cases} \frac{r_j(k)}{\binom{d}{k}} & \text{if } \#\{x_h : x_h = 1\} = k \\ 0 & \text{otherwise} \end{cases}. \tag{2.8}$$

We denote by R_j and R_{pd} the random variables whose pmfs are r_j and r_{pd} respectively and by $\boldsymbol{E}_j$ and $\boldsymbol{E}_{pd}$ the random variables whose pmfs are e_j and e_{pd} respectively. We will refer to r_j, r_{pd}, e_j and e_{pd} as extremal densities.

Example 1. *As an illustrative example we consider the class of exchangeable distributions of dimension $d = 3$ and mean $p = 0.4$, i.e. $\mathcal{E}_3(0.4)$. In dimension $d = 3$, the polytope $\mathcal{S}_3(0.4)$ is 2-dimensional. The extremal densities of $\mathcal{S}_3(0.4)$ are the columns in Table 1. The extremal densities of $\mathcal{E}_3(0.4)$ can be derived by inverting H as defined in (2.1). The extremal densities are points in $\mathbb{R}^4$ which lie in a subspace of dimension $d - 1 = 2$. Using standard Principal Component Analysis we can project these 4 points to $\mathbb{R}^2$. The points inside the polygon in Figure 1 represent all the densities which belong to $\mathcal{S}_3(0.4) \leftrightarrow \mathcal{E}_3(0.4)$.*

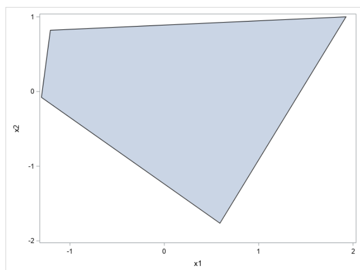
2.2 Expectation measures

Functionals Φ defined on a class $\mathcal{F}_0$ of d-dimensional distributions, $\Phi : \mathcal{F}_0 \to \mathbb{R}$, $\Phi(f) = E[\phi_d(\boldsymbol{X}_f)]$, where $\boldsymbol{X}_f \in \mathcal{F}_0$ has pmf f and ϕ_d is a real valued function of the random variable $\boldsymbol{X}_f$, are commonly used in applications to define for example measures of risk, see e.g. [9]. Examples of such functionals in our framework are:

Table 1: Extremal densities of $\mathcal{S}_3(0.4)$

y	$r_1(y)$	$r_2(y)$	$r_3(y)$	$r_4(y)$
0	0.4	0.6	0	0
1	0	0	0.8	0.9
2	0.6	0	0.2	0
3	0	0.4	0	0.1

Figure 1: 2-dimensional polytope $\mathcal{E}_3(0.4)$: x_1 and x_2 are the principal component coordinates of the densities.



1. Moments and cross moments of distributions in $\mathcal{E}_d$ or $\mathcal{E}_d(p)$ and moments of discrete distributions in $\mathcal{S}_d$ or $\mathcal{S}_d(p)$.

2. The entropic risk measure on $\mathcal{S}_d$ or $\mathcal{S}_d(p)$ for $\gamma \in (0, \infty)$:
$$\Gamma(f) = \frac{1}{\gamma} \log(E[e^{-\gamma Y}]), \tag{2.9}$$
where Y has pmf f.

3. Excess loss function on $\mathcal{S}_d$ or $\mathcal{S}_d(p)$, defined by
$$\Phi(Y) = E[(Y - k)^+], \;\; k \in \mathbb{R}, \tag{2.10}$$
where $(x - k)^+ = \max\{x - k, 0\}$.

4. Von Neumann-Morgestern expected utilities on $\mathcal{S}_d$ or $\mathcal{S}_d(p)$. These are expectation measures, where the function ϕ_d is some increasing utility function.

Proposition 2.2 allows us to have an analytical expression for a wide class of statistical indices defined as functionals on $\mathcal{E}_d$ ($\mathcal{E}_d(p)$) and $\mathcal{S}_d$ ($\mathcal{S}_d(p)$), such as all the moments of the Bernoulli exchangeable distributions.

Proposition 2.2. *1. Let $\boldsymbol{X} \in \mathcal{E}_d$ [$\boldsymbol{X} \in \mathcal{E}_d(p)$] and let be $\phi_d : \mathbb{R}^d \to \mathbb{R}$ a measurable function. Then*
$$E[\phi_d(\boldsymbol{X})] = \sum_{i=1}^{m} \lambda_i E[\phi_d(\boldsymbol{E}_j)], \tag{2.11}$$

where $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_m$ are the extremal points of $\mathcal{E}_d$ [$\mathcal{E}_d(p)$].

2. *Let $Y \in \mathcal{S}_d$ [$Y \in \mathcal{S}_d(p)$] and let be $\phi : \mathbb{R} \to \mathbb{R}$ a measurable function. Then*

$$E[\phi(Y)] = \sum_{i=1}^{m} \lambda_i E[\phi(R_j)], \tag{2.12}$$

where $R_1, \ldots, R_m$ are the extremal points of $\mathcal{S}_d$ [$\mathcal{S}_d(p)$].

Proof. We only prove part 1, because part 2 is analogous.

It holds

$$E[\phi_d(\boldsymbol{X})] = \sum_{\boldsymbol{x} \in \mathcal{X}_d} \phi_d(\boldsymbol{x}) f(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d} \phi_d(\boldsymbol{x}) \sum_{j=1}^{m} \lambda_j r_j(\boldsymbol{x}) = \sum_{j=1}^{m} \lambda_j E[\phi_d(\boldsymbol{E}_j)]. \tag{2.13}$$

□

We call expectation measures the measures defined by expectations and by their one-to-one transformations. Proposition 2.2 states that measures defined by expectation of the mass functions in a given class are themselves a convex polytope whose extremal points are the measures evaluated on the extremal points of the class. Therefore, they are bounded by their evaluations on the extremal points. Notice that also the entropic risk measure has its bounds on the extremal points, since the logarithm is a monotone transformation. Another important measure that reaches its bound on the extremal pmfs is the α-quantile of the sum Y as proved in [6]. Now we focus on the cross moments. By exchangeability the cross moments of $\boldsymbol{X} = (X_1, \ldots, X_d)$ depend only on their order. We therefore use μ_α to denote a moment of order $\alpha = \text{ord}(\boldsymbol{\alpha}) = \sum_{i=1}^{d} \alpha_i$, where $\boldsymbol{\alpha} \in \mathcal{X}_d$. We have

$$\mu_\alpha = E[X_1 \cdots X_\alpha] = \sum_{k=\alpha}^{d} \frac{\binom{d-\alpha}{k-\alpha}}{\binom{d}{k}} p_k,$$

where $p_k = P(\sum_{i=1}^{d} X_i = k)$. In particular the second order moment of the class $\mathcal{S}_d(p)$ are given by

$$E[Y^2] = E[(X_1 + \ldots + X_d)^2] = pd + d(d-1)\mu_2. \tag{2.14}$$

We want to determine the distribution $E[\phi_d(\boldsymbol{X})]$ where $\boldsymbol{X}$ is a random variable which has been chosen uniformly at random from $\mathcal{E}_d(p)$ or $\mathcal{E}_d$. Other distributions over $\mathcal{E}_d(p)$, different from the uniform, can be easily considered if necessary. Since classes of multivariate Bernoulli distributions with pre-specified moments, as e.g. $\mathcal{E}_d(p)$, are polytopes, the representation of each mass function as a convex linear combination of the extremal points is not unique. Therefore we need a partition of the convex polytope

$\mathcal{E}_d(p)$ into simplices, i.e. a triangulation of $\mathcal{E}_d(p)$. See [8] as a general reference in computational geometry. To reduce dimensionality we work on $\mathcal{D}_d(dp)$ instead of $\mathcal{E}_d(p)$. This can be done using the map H defined in (2.1) under the exchangeability condition $f(\boldsymbol{x}) = f(\sigma(\boldsymbol{x}))$ for $f \in \mathcal{E}_d(p)$. As a consequence the distribution of $E[\phi_d(\boldsymbol{X})]$ can be studied in $\mathcal{D}_d(dp)$ if $\phi_d(\boldsymbol{x}) = \phi_d(\sigma(\boldsymbol{x}))$; we make this assumption in this section. We perform a triangulation of $\mathcal{D}_d(dp)$ that is equivalent to perform a triangulation of the polytope $\mathcal{C} = \{\boldsymbol{p} : p_i \geq 0, \sum_{i=0}^d p_i = 1, \sum_{i=0}^d ip_i = dp\}$. The dimension of $\mathcal{C}$ is $d-1$ because $\mathcal{C}$ is defined by two constraints. We can partition $\mathcal{C}$ into simplices $\mathcal{T}_i$, $i \in \mathcal{I}$ (e.g. using a Delaunay triangulation)

$$\mathcal{C} = \bigcup_{i \in \mathcal{I}} \mathcal{T}_i, \tag{2.15}$$

where $\mathcal{I}$ is a proper set of indices, and $\mathcal{T}_i \cap \mathcal{T}_j = \emptyset$ for $i \neq j$. We observe that from a geometric point of view the intersection between two simplices $\mathcal{T}_i$ and $\mathcal{T}_j$ is not empty, being the common part of their borders. But this common part has zero probability of being selected and so we can neglect it assuming that each $\mathcal{T}_i,\ i \in \mathcal{I}$ coincides with its interior part.

Let $\phi(y) = \phi_d(\boldsymbol{x})$, where $y = \sum_{i=1}^d x_i,\ y = \{0, \ldots, d\}$. Let us denote by F_ϕ the distribution of $E[\phi(Y)]$, where Y is a random variable with pmf p_Y from $\mathcal{S}_d(p)$ or $\mathcal{S}_d$. We get

$$F_\phi(t) = P(E[\phi(Y_p)] \leq t) = \sum_{i \in \mathcal{I}} P(\mathcal{T}_i) P(E[\phi(Y_p)] \leq t | \mathcal{T}_i). \tag{2.16}$$

If we assign a uniform measure on the space $\mathcal{E}_d(p)$ the probability $P(\mathcal{T}_i)$ of sampling a probability mass function in the simplex $\mathcal{T}_i$ is simply the ratio between the volume of $\mathcal{T}_i$ and the total volume of $\mathcal{C}$, i.e.

$$P(\mathcal{T}_i) = \frac{\mathrm{vol}(\mathcal{T}_i)}{\mathrm{vol}(\mathcal{C})}. \tag{2.17}$$

The volume of each $\mathcal{T}_i$ and consequently that of $\mathcal{C}$ can be easily computed because the volume of an n-simplex in n-dimensional space with vertices $(v_0, \ldots, v_n)$ is

$$\left| \frac{1}{n!} \det \begin{pmatrix} v_1 - v_0, & v_2 - v_0, & \ldots, & v_n - v_0 \end{pmatrix} \right|$$

where each column of the $n \times n$ determinant is the difference between the vectors representing two vertices [10].

The probability $P(E[\phi(Y)] \leq t | \mathcal{T}_i)$ is the ratio between the volume of the region

$$\mathcal{R}_{i,t} = \{p_Y \in \mathcal{T}_i : E[\phi(Y)] \leq t\} \tag{2.18}$$

and the volume of $\mathcal{T}_i$, i.e.

$$P(E[\phi(Y)] \leq t | \mathcal{T}_i) = \frac{\mathrm{vol}(\mathcal{R}_{i,t})}{\mathrm{vol}(\mathcal{T}_i)}. \tag{2.19}$$

The computation of the volume of $\mathcal{R}_{i,t}$ will depend on the definition of ϕ in the expectation measure $E[\phi(Y)]$.

We now consider the k-order moments $\mu_k^{(Y)}$ of the random variable $Y = X_1 + \ldots + X_d$ whose pmf is denoted by p_Y

$$\mu_k^{(Y)} = E[Y^k] = \sum_{i=0}^{d} i^k p_Y(i) = \sum_{i=0}^{d} i^k p_i$$

From Eq. (2.1) we have $p_Y(i) = \binom{d}{i} f_i$ and $f_i := f_p(\boldsymbol{x})$ for $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathcal{X}_d$ and $\#\{x_k : x_k = 1\} = i$.

The pmf p_Y will lie in exactly one of the simplices $\mathcal{T}_i, i \in \mathcal{I}$. Let's denote this simplex by $\mathcal{T}_{i^\star}$. We can write $p_Y = \sum_{j \in \mathcal{J}^*} \lambda_j r_j$, where $\mathcal{J}^*$ is the set of indexes that defines the subset of points which generate the simplex $\mathcal{T}_{i^\star}$, i.e. $\mathcal{T}_{i^\star} = \text{simplex}(\{r_j : j \in \mathcal{J}^*\})$. As a corollary of Proposition 2.2 we can write

$$\mu_k^{(Y)} = \sum_{j \in \mathcal{J}^*} \lambda_j \mu_k^{(j)}$$

where $\mu_k^{(j)}$ are the k-moments of the extremal random variables R_j, $\mu_k^{(j)} = E[R_j^k]$. For k-order moments the region $\mathcal{R}_{i,t} = \{p_Y \in \mathcal{T}_i : \mu_k(Y) \leq t\}$ is the subset of the standard simplex defined as $\{(\lambda_j; j \in \mathcal{J}^*) : \lambda_j \geq 0, \sum_{j \in \mathcal{J}^*} \lambda_j = 1, \sum_{j \in \mathcal{J}^*} \lambda_j \mu_k^{(j)} \leq t\}$. For k-order moments the ratio of the volumes in Eq. (2.19) can be computed using an exact and iterative formula, see [11] and [12]. It follows that the distribution of $E[Y^k]$ can be computed exactly.

The same methodology can be applied also to other measures defined as function of expected values like the entropic risk measures. In the next section, using Proposition 2.2 and convexity, we numerically find the distribution of the above measures across the class of pmf where they are defined.

3 High dimensional simulation

The representation of Bernoulli pmfs as convex combinations of extremal densities allows us to build high dimensional exchangeable Bernoulli distributions in $\mathcal{E}_d(p)$. The simplest way to do that is to generate directly from an extremal density. Extremal densities have support on two points. We can also use combinations of extremal points to choose a distribution in the interior of the polytope. As an example we could choose $\lambda_i = \frac{1}{n_p}$ in (2.7). Such a choice identifies a pmf inside the polytope.

3.1 Simulation of the whole correlation range

The extremal densities allow us to simulate from a family of distributions that cover the whole range spanned by a measure of dependence, meaning that we can construct a parametrical family of mass functions such that for any admissible value of the measure there is a mass function in the class for which the measure takes that value. Let M and m the maximum and minimum values of an expectation measure. If r_M and r_m are two corresponding extremal densities, the parametrical family $f = \lambda r_m + (1-\lambda) r_M$ span the whole range $[m, M]$. We can therefore use this family to simulate a sample of binary data from a pmf with a given value of the measure or we can simulate binary variables with different values of the measure, by moving λ.

An important example is correlation. The problem to simulate from multivariate Bernoulli distributions with given correlations and in particular negative correlations is of interest in many applications and is addressed in the statistical literature, see [13]. The geometrical structure of $\mathcal{E}_d(p)$ allows us to solve this problem for exchangeable random vectors. Let us suppose that pd is not an integer. We can choose a pmf with the required correlation simply by using $r_{\rho_m} = r_{(j_1^M, j_2^m)}$ and $r_{\rho_M} = r_{(0,d)}$, the two extremal densities in $\mathcal{E}_d(p)$ with the minimum and maximum of allowed correlations, respectively. The pmfs in the family $r_\lambda = \lambda r_{\rho_m} + (1-\lambda) r_{\rho_M}$ span the whole correlation range. If the desired value of the correlation is ρ_0, $\rho_m \leq \rho_0 \leq \rho_M$ it is enough to choose $\lambda = \frac{\rho_M - \rho_0}{\rho_M - \rho_m}$. If pd is integer we must use $r_{\rho_m} = r_{pd}$. We observe that $\rho_M = 1$.

This case is interesting because the families of multivariate Bernoulli variables commonly used for simulation of exchangeable binary variables incorporate only positive correlations. We consider here two families of exchangeable Bernoulli models used in the literature to simulate binary data. The first family is proposed in [1] and we term it family of one-factor models, taking the name from the one-factor models used in credit risk, that have a similar dependence structure. The second family is the mixture model based on De Finetti's representation theorem. The construction proposed in [1] provides an algorithm to generate binary data with given marginal Bernoulli distributions with means $(p_1, \ldots, p_d)$ and exchangeable dependence structure, meaning that they are equicorrelated. However, by assuming that the marginal parameters are equal to a common parameter p, their construction gives a vector $\boldsymbol{X} \in \mathcal{E}_d(p)$. This is the case considered here. We therefore define the multivariate Bernoulli variable in this framework. Let

$$X_i = (1 - U_i) Y_i + U_i Z, i = 1, \ldots, d, \tag{3.1}$$

where $U_i \sim B(\sqrt{\rho})$, $Y_i \sim B(p)$, $i = 1, \ldots, d$ and $Z \sim B(p)$ and they are independent. We say that $\boldsymbol{X} = (X_1, \ldots, X_d)$ and its pmf $f \in \mathcal{E}_d(p)$ have a one-factor structure. Clearly, $\boldsymbol{X}$ is exchangeable, has distribution $B(p)$ and correlation ρ. By construction we have $\rho \geq 0$ and the case $\rho = 0$ implies that U_i, $i = 1, \ldots, n$ put all the mass on 0.

According to De Finetti's Theorem if $f \in \mathcal{E}_d(p)$ is the pmf of a random vector

$(X_1, \dots, X_d)$ extracted from an exchangeable infinite sequence then $f = (f_0, \dots, f_d)$ has the representation

$$f_j = \int_0^1 p^j (1-p)^{d-j} d\Psi(p),\ j = 0, \dots, d,$$

where $\Psi(p)$ is a pdf on $[0,1]$. Clearly, these vectors can only have positive correlations. One of the most used mixed Bernoulli model is the β-mixing models, where $\Psi \sim \beta(a,b)$ is a β-distributed mixing variable. In this case we have

$$\begin{aligned} p &= E[\Psi] \\ \mu_2 &= E[\Psi^2]. \end{aligned}$$

Therefore we determine the β parameters a and b by solving the equations

$$\begin{aligned} p &= \frac{a}{a+b} \\ \mu_2 &= \frac{a(a+1)}{(a+b)(a+b+1)}. \end{aligned}$$

For this model $\rho = 0$ is not admissible, therefore the model cannot include independence. Notice that the β model has two parameters and, chosen a β-mixture model in $\mathcal{E}_d(p)$ it can be parameterized by ρ.

In [6] the authors analytically found the correlation bounds for exchangeable pmf for each dimension d and the minimum attainable correlation. The minimal correlation ρ_m goes to zero if the dimension increases, according to De Finetti's representation theorem. Therefore, the capability to generate binary data with negative correlations is more important in low dimensions. Nevertheless, simulation of a pmf in the family $r_\lambda = \lambda r_{\rho_m} + (1-\lambda) r_{\rho_M}$ is easy and can be performed also in high dimension. In Section 3.4.1 we discuss an example where the three methods are compared.

3.2 A simple algorithm for high dimensionality

Now we describe a general and simple algorithm to simulate also high dimensional exchangeable Bernoulli distribution from $\mathcal{E}_d(p)$, given the vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{n_p})$ in (2.4) or equivalently in (2.7). The vector $\boldsymbol{\lambda}$ can be chosen as we discussed above ($\lambda_i = 1/n_p,\ \ i = 1, \dots n_p$) or it can be chosen by giving a distribution P_Λ on the simplex $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^{n_p} : \sum_{i=1}^{n_p} \lambda_i = 1, \lambda_i > 0, i = 1, \dots, n_p\}$ in (2.4) or as an extremal density. Once $\boldsymbol{\lambda}$ is selected, it represents a probability distribution on the set of extremal densities, i.e. λ_j is the probability to extract the extremal density r_j. Given the pmf $\boldsymbol{\lambda}$ on the set of extremal points, Algorithm 1 allows us to simulate also in high dimension. We observe that Algorithm 1 does not require to store any big structure of data and then it can be easily used for large d, e.g. $d = 10^5$. The Algorithm 1 together with the algorithms "one-factor model" (Algorithm 2) and "β-mixture Bernoulli model" (Algorithm

Algorithm 1
Input: the expected value p, the dimension d, the vector $\lambda \in \Lambda$.
1) Select an extremal density r_j with probability λ_j; r_j has support on j_1, j_2, see (2.5).
2) Select $j_* \in \{j_1, j_2\}$ with probability $r_j(j_1)$ and $r_j(j_2)$ respectively, see (2.5) .
3) Select a binary vector with j_* ones among the combinations $\binom{d}{j_*}$.
Output: One realization of a d dimensional binary variable with pmf as in (2.7).

Table 2: Computational times [seconds. hundredths of a second]. Algorithm 1 has been implemented without specifiying correlation (Algorithm 1 "norho") and with specifying correlation (Algorithm 1 "rho").

Algorithm	$d = 100$ $n = 100$	$d = 100$ $n = 10000$	$d = 1000$ $n = 100$	$d = 1000$ $n = 10000$
Algorithm 1 "norho"	0.12	0.31	0.18	7.17
Algorithm 1 "rho"	0.9	0.15	0.8	0.50
Algorithm 2	0.12	0.14	0.39	31.25
Algorithm 3	0.8	0.13	0.7	0.47

2) have been implemented in SAS/IML. It is worth noting that in the case $d = 100$, $p = 0.4$ and $\rho = 0.1$ the computational time for getting a sample of size $n = 100$ is very small for all the algorithms (some hundredths of a second on a standard laptop). We also tested the cases $d = 100$ and $n = 10000$, $d = 1000$ and $n = 100; 10000$. We always obtain the output in a few seconds. The computational times are reported in Table 2.

The advantage of Algorithm 1 is that while keeping low the computational time allows the users to explore all the extremal densities of $\mathcal{E}_d(p)$ including those with negative correlation. For example, we observe that in the case $d = 100$ and $p = 0.4$ there are 85 extremal densities with negative correlations and these densities cannot be simulated using Algorithms 2 and 3. We also observe that Algorithm 1 with given correlation performs similarly to Algorithm 2 and better than Algorithm 3.

3.3 Uniform simulation

There are famous measures that are not expectation measures and for which Proposition 2.2 does not apply, for example the α-quantile or value at risk of a balanced portfolio of exchangeable Bernoulli variables, widely used in applications as a measure of risk. In particular, the distribution of the value at risk of the loss of a large portfolio is important to evaluate the risk associated to the choice of a specific model. Another famous measure of risk is the entropy. For a discrete pmf $p_Y \in \mathcal{S}_d$ it is given by:

$$E(p_Y) = -\sum_{i=0}^{d} p_i \log p_i.$$

The entropy does not satisfy Proposition 2.2 and does not reach its bound on the extremal densities. We therefore are not able to use the geometry of a pmf in $\mathcal{S}_d$ ($\mathcal{S}_d(p)$) to numerically find its exact distribution across the $\mathcal{S}_d$ ($\mathcal{S}_d(p)$). However, we can address this goal using simulations. For other measures for which the exact value of the ratio in (2.19) cannot be computed, an estimate of it can be simply obtained by sampling uniformly at random over $\mathcal{T}_i$ and determining the relative frequency of the points that fall in the region $\mathcal{R}_{i,t}$, as defined in (2.18)

$$\widehat{\left(\frac{\mathrm{vol}(\mathcal{R}_{i,t})}{\mathrm{vol}(\mathcal{T}_i)}\right)} = \frac{\#\{p_k \in \mathcal{R}_{i,t},\ k = 1, \ldots, N\}}{N},$$

where N is the size of the sample. In these cases an estimate $\hat{F}_\phi$ of the distribution F_ϕ will be obtained. Being able to perform uniform sampling allow us to sample also for any given distribution on $\mathcal{E}_d$ ($\mathcal{E}_d(p)$).

Let's start considering uniform sampling from $\mathcal{E}_d$. From Eq.(5.1) we know $\mathcal{E}_d \leftrightarrow \mathcal{S}_d \equiv \mathcal{D}_d$. It follows that sampling uniformly at random from $\mathcal{E}_d$ is equivalent to sampling uniformly at random from $\mathcal{D}_d$ and then it is equivalent to sampling uniformly at random from the d-simplex $\Delta_d = \{\boldsymbol{p} : p_i \geq 0, \sum_{i=0}^{d} p_i = 1\}$. Sampling from a simplex is a standard topic in the statistical literature.

Let's now consider uniform sampling from $\mathcal{E}_d(p)$. From Eq.(2.3) we know $\mathcal{E}_d(p) \leftrightarrow \mathcal{S}_d(p) \equiv \mathcal{D}_d(dp)$. It follows that sampling uniformly at random from $\mathcal{E}_d(p)$ is equivalent to sampling uniformly at random from $\mathcal{D}_d(dp)$ and then is equivalent to sampling uniformly at random from the polytope $\mathcal{C} = \{\boldsymbol{p} : p_i \geq 0, \sum_{i=0}^{d} p_i = 1, \sum_{i=0}^{d} ip_i = dp\}$. We therefore have to consider one possible triangulation of $\mathcal{C}$ as defined in (2.15). We can consider sampling from $\mathcal{E}_d$ as a special case of sampling from $\mathcal{E}_d(p)$. For this purpose when sampling from $\mathcal{E}_d$ we denote the d-simplex by $\mathcal{T}_1$, $\mathcal{T}_1 \equiv \Delta_d$. We have $\mathcal{I} = \{1\}$.

Remark 1. *In case we want to sample from a distribution on the politope $\mathcal{C}$ with pdf $\tilde{f}$ different from the uniform we can adapt the algorithm. We move from a point $\boldsymbol{p}_0 \in \mathcal{C}$ to another point $\boldsymbol{p}_1 \in \mathcal{C}$ with probability $\min\{1, \frac{\tilde{f}\boldsymbol{p}_1}{\tilde{f}(\boldsymbol{p}_0)}\}$. Further details can be found in the classical reference [14].*

3.4 Application

We discuss a possible application in credit risk modelling. As an example we consider an homogeneous portfolio of credit card holders extracted from the Kaggle database on credit card defaulters[1]. This dataset is a collection of data from 30000 clients of a bank issuing credit cards in Taiwan, from April to September 2005. In this dataset, the proportion of default customer is almost 40% for the ones that have already registered one or more defaults in their history, while it is less than 20% for those who have never had a default. Consider a d dimensional portfolio of credit card holders belonging to the former homogenous group. A random vector in $\boldsymbol{X} \in \mathcal{E}_d(0.4)$ represents the vector of default indicators of each card holder. With our choice of $p = 0.4$ we are considering a high marginal default probability, therefore we imagine to handle a high risk portfolio. Nevertheless, the following analysis can be performed for any $p \in (0, 1)$, since the proposed method does not put any restriction on p. To evaluate the risk of a credit portfolio P of d card holders we consider the number of defaults $Y \in \mathcal{S}_d(0.4)$

$$Y = \sum_{i=1}^{d} X_i.$$

The number of defaults Y represents the loss of a balanced portfolio. Banks are interested to measure the risk associated to a given portfolio. One of the most used measures of risk is the so called value at risk (VaR), that is the α- quantile of the distribution of the loss for a given level $\alpha \in (0, 1)$. The VaR of Y for exchangeable portfolios has been analysed in [6] where analytical bounds have been found, since they are reached on the extremal points of the class. Nevertheless, the main factors influencing the VaR of Y are the default probability and the correlation among defaults [3]. For this reason, we consider the second order moment and the correlation distribution across the class $\mathcal{E}_d(0.4)$. As a measure of risk and randomness of the portfolio, we consider the entropy because it does not reach its bounds on the extremal points and therefore it is necessary to simulate its distribution across the class. We present two different scenarios. The first scenario consider a low dimension portfolio, $d = 3$, with the purpose to explicitly see the triangulation and the exact numerical distribution of the moments. The second example considers a portfolio of 100 card holders, therefore it is performed in $\mathcal{E}_{100}(0.4)$.

3.4.1 Scenario 1

In this first application we use both an ad-hoc module written in SAS/IML [15] and the R package uniformly [16] which uses a triangulation based method to provide uniform samples from a given convex polytope. We consider the class of all the possible default scenarios in $\mathcal{E}_3(0.4)$, the class of exchangeable distributions of dimension $d = 3$ and mean $p = 0.4$. We get:

1. the exact distribution of the second-order moment;

[1]https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset, accessed on 07 July 2022

2. a family of pmfs that spans the whole range of correlation, showing that we can simulate a portfolio with any admissible correlation and not only positively correlated portfolios;

3. the sampling distribution of the entropy.

We also analyse the joint distribution of the first-order moment and correlation in $\mathcal{E}_3$, the class of exchangeable distributions of dimension $d = 3$.

The extremal densities of $\mathcal{E}_3(0.4)$ are the columns in Table 1. The polygon in Figure 1 represent all the densities which belong to $\mathcal{S}_3(0.4) \leftrightarrow \mathcal{E}_3(0.4)$.

Figure 2: Triangulation of $\mathcal{S}_3(0.4)$.

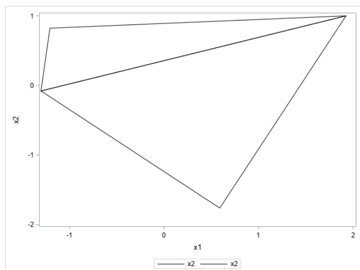


Figure 3 shows one possible triangulation. We have two triangles: $\mathcal{T}_1$, the largest one with area 3.74 and $\mathcal{T}_2$, the smallest one with area 1.40. Then, the sampling probabilities are $P(\mathcal{T}_1) = 0.73$ and $P(\mathcal{T}_2) = 0.27$.

For 2-order moments the region $\mathcal{R}_{i,t} = \{p_Y \in \mathcal{T}_i : \mu_2(Y) \leq t\}$ is the subset of the standard simplex defined as $\{(\lambda_j; j \in \mathcal{J}^*) : \lambda_j \geq 0, \sum_{j \in \mathcal{J}^*} \lambda_j = 1, \sum_{j \in \mathcal{J}^*} \lambda_j \mu_2^{(j)} \leq t\}$. We recall that in this case the ratio of the volumes in Eq. (2.19) can be computed using an exact and iterative formula, see [11] and [12].

Figure 3 (left side) exhibits the exact numerical cumulative distribution function (cdf) $P(\mu_2(Y) \leq t | \mathcal{T}_1)$ of $\mu_2(Y)$ across $\mathcal{T}_1$. The distribution of $\mu_2(Y)$ across $\mathcal{T}_2$ is similar. The cdf F_{μ_2} of $\mu_2(Y)$ across the whole polytope is obtained by mixing the conditional cdfs as in (2.16). Figure 3 - right side - shows the probability density function (pdf) of the mixture obtained from the cdf of μ_2 by $f_{\mu_2}(t) = \frac{F_{\mu_2}(t+\Delta) - F_{\mu_2}(t)}{\Delta}$, where Δ has been chosen equal to $(\max(\mu_2) - \min(\mu_2))/10000$.

The pmfs in the family $r_\lambda = \lambda r_{\rho_m} + (1 - \lambda) r_{\rho_M}$ span the whole correlation range and can be simulated according to the proposed Algorithm. Figure 4 shows the family in the polytope $\mathcal{E}_3(0.4)$ together with the families of β-mixture models and one-factor models in the same class. It is evident that our approach allows us to consider also pmfs with negative correlations (green straight line in the figure), while the other two approaches provide only positive correlations (blue and red lines in the figure, that in this case are almost overlapped). In dimension three the range of negative correlation $[-0.39, 0]$ is wide, as shown in the figure.

Figure 3: Distribution of the 2-order moment across $\mathcal{E}_3(0.4)$.

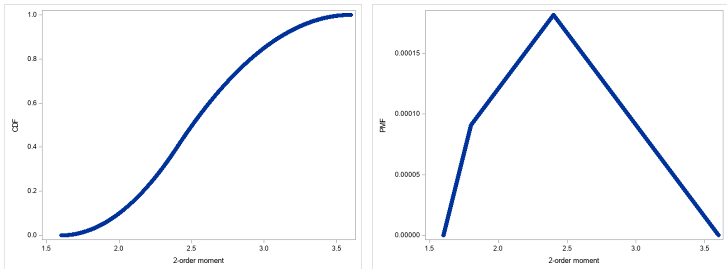
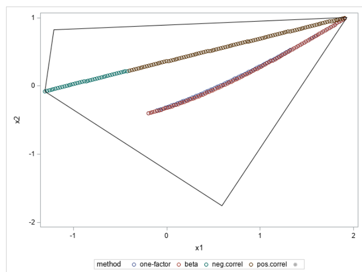
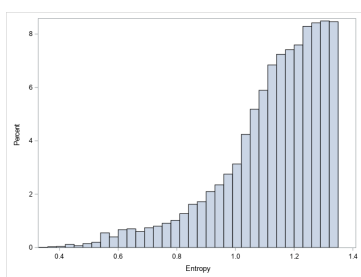


Figure 4: Families of distributions: r_λ (straight line: yellow line are positive correlations, green line are negative correlations), β-mixture (red line) and one-factor (blue line) across $\mathcal{E}_3(0.4)$. The blue and red line are almost overlapped.



The simulated pdf of the entropy in the class $\mathcal{E}_3(0.4)$, where the entropy of $\boldsymbol{X} \in \mathcal{E}_3(0.4)$ is defined to be the entropy of the number of defaults $Y = \sum_{i=1}^{d} X_i$, is found by sampling using the methodology in Section 3.3. It is shown in Figure 5.

Figure 5: Empirical distribution of the entropy across $\mathcal{E}_3(0.4)$.



The simulated pdf can also be found for the second order moment $\mu_2(Y)$. It is shown in Figure 6 for completeness. The simulated pdf is obviously in agrement with the exact numerical one (right side of Figure 3).

Since mean and correlation are the relevant factors influencing the tail of the number of defaults, the pairs (p, ρ) correspond to different risk profiles. Figure 7 shows the simulated bivariate distribution of the mean p and the correlation ρ across $\mathcal{E}_3$. The joint behavior of p and ρ is in accordance with the theoretical bounds found in [6]. In this case, where $d = 3$, the minimal correlation is -0.5 and it is attained for $p = \frac{1}{3}$ and $p = \frac{2}{3}$.

Figure 6: Empirical distribution of the second order moment across $\mathcal{E}_3(0.4)$.

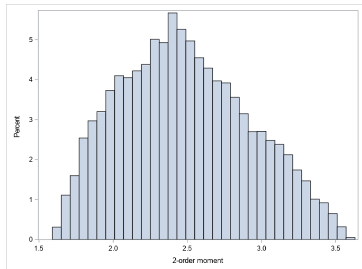


Figure 7: Bivariate distribution of the first order moment and correlation across $\mathcal{E}_3$.

3.4.2 Scenario 2 - high dimensional portfolio

In real credit risk applications the dimensionality is high and the triangulation based method becomes unfeasible from a computational point of view. In this second scenario we use both an ad-hoc module written in SAS/IML [15] and the R package volesti [17] which does not use a triangulation but a random-walk-based method to provide uniform samples from a given convex polytope.

We generate a sample of size 2000 portfolios from $\mathcal{E}_{100}(0.4)$ and we study the empirical distribution of the correlation, the VaR and the entropy. In this case we have 2401 extremal points. The simulated pdf of the VaR in the class $\mathcal{E}_{100}(0.4)$ is shown in Figure 8. As Figure 8 shows we are dealing with a high risk portfolio (the VaR is at least 83), and this is a consequence of having a high marginal default probability. The analytical bounds for the VaR can be computed using Proposition 5.4 in [6]. Since we consider $\alpha = 0.95$ we find min(VaR)=40 and max(VaR)=100. The empirical distribution shows that the observed VaR are all significantly higher than the minimum possible value, that is reached on an extreml point of the class. This fact shows that the empirical distribution can be useful to properly evaluate the risk of a portfolio. The distribution of the correlation across $\mathcal{E}_{100}(0.4)$ is in Figure 8.

Figure 8 also shows the simulated pdf of the entropy in the class $\mathcal{E}_{100}(0.4)$; the entropy does not reach its bounds on the extremal points.

Figure 8: From left to right: empirical distribution of VaR, corrlation and entropy across $\mathcal{E}_{100}(0.4)$.

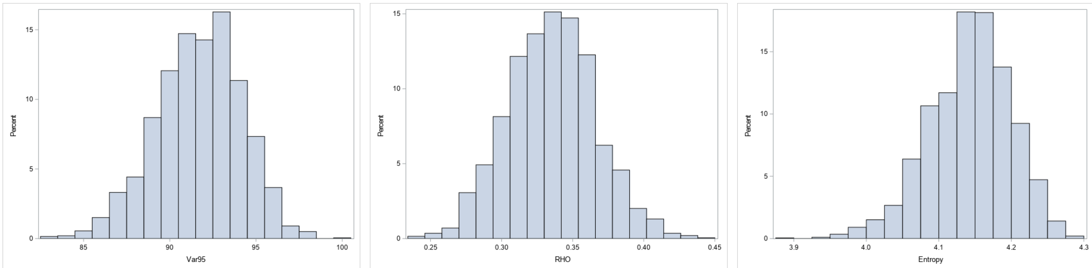
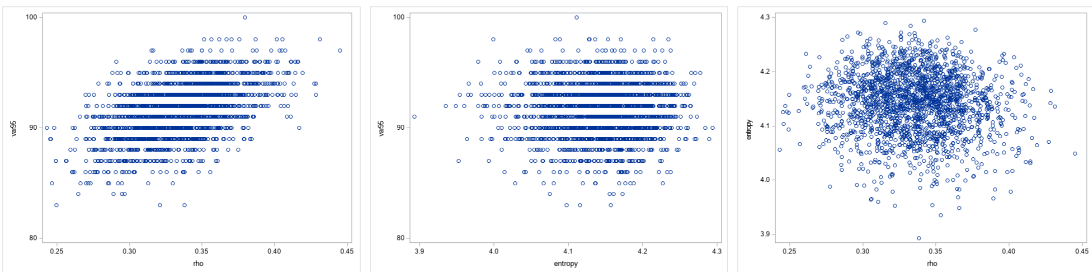


Figure 9: From left to right: joint empirical distribution of VaR vs correlation, VaR vs entropy and correlation vs entropy across $\mathcal{E}_{100}(0.4)$.



We conclude this section with an empirical analysis of the joint behavior of correlation, VaR and entropy, see Figure 9. Figure 9 - left side- shows that the range spanned by correlation is different for different values of VaR. Specifically correlations are concentrated on higher values in correspondence of higher values of the portfolio VaR. This fact empirically exhibits that correlations among obligors is an important issue to take into account in the portfolio management. On the contrary, the entropy seems to be not linked to correlation - Figure 9 right side- and it also seems to be a measure of risk independent of VaR - Figure 9, middle side.

4 Estimate and testing

4.1 Maximum likelihood estimation

In many applications it is necessary to fit exchangeable Bernoulli models to historical data. For example in credit risk this is done on default data for a single homogeneous group of obligors with some common credit rating, see [18]. We focus on the maximum likelihood estimation in the classes of exchangeable distributions and exchangeable distributions with given margins, $\mathcal{E}_d$ and $\mathcal{E}_d(p)$ respectively.

Proposition 4.1. *Let $\boldsymbol{f} \in \mathcal{F}_d$. A maximum likelihood estimator of $\hat{\boldsymbol{f}}$ which belongs to*

$\mathcal{E}_d$ ($\mathcal{E}_d(p)$) *always exists.*

Proof. $\mathcal{E}_d$ ($\mathcal{E}_d(p)$) is a closed convex sets in $\mathbb{R}^d$, hence it is compact and the likelihood functions for the models in (4.6) are continuous. $\square$

The maximum likelihood estimator (MLE) in the class $\mathcal{E}_d$ can be found analytically using the map in (2.1).

Let us assume to observe a sample of size n drawn from a d-dimensional Bernoulli distribution $\boldsymbol{X}$ and let $Y = \sum_{i=1}^{d} X_i$ have pmf $\boldsymbol{p}_Y = (p_0, \ldots, p_d)$ that gives rise to counts $\boldsymbol{N}$. The count $\boldsymbol{N} = (N_0, \ldots, N_d)$ has a multinomial distribution with parameters $n, \boldsymbol{p}$, i.e. $\boldsymbol{N} \sim$ Multinomial$(n, \boldsymbol{p})$, where the parameter $\boldsymbol{p}$ belongs to the d-simplex Δ_d. The likelihood function is

$$L(\boldsymbol{n}; \boldsymbol{p}) = P(N_i = n_i, i \in \{0, \ldots, d\}) = \binom{n}{n_0 \cdots n_d} \prod_{j=0}^{d} (p_j)^{n_j}, \tag{4.1}$$

where we set $0^0 := 1$. The MLE is the solution of the constrained maximization problem

$$\begin{gathered} max_{\boldsymbol{p}} \log L(\boldsymbol{n}; \boldsymbol{p}), \\ \text{sub} \\ \sum_{i=0}^{d} p_j - 1 = 0 \end{gathered} \tag{4.2}$$

By using the Lagrange multipliers we find:

$$\hat{p_j} = \frac{N_j}{n}, \;\; j = 0, \ldots, d.$$

The MLE in the class $\mathcal{E}_d$ is

$$\hat{f}_j = \frac{\hat{p_j}}{\binom{d}{j}} = \frac{\frac{N_j}{n}}{\binom{d}{j}}, \;\; i = 0, \ldots, d$$

We now consider the class $\mathcal{E}_d(p)$. The MLE estimator in $\mathcal{D}_d(dp)$ can be numerically found by solving the constrained maximization problem:

$$\begin{gathered} max_{\boldsymbol{p}} \log L(\boldsymbol{n}; \boldsymbol{p}), \\ \text{sub} \\ \sum_{i=0}^{d} p_j - 1 = 0 \\ \sum_{i=0}^{d} j p_j - pd = 0, \end{gathered} \tag{4.3}$$

Then we use the map H again to have the MLE in $\mathcal{E}_d(p)$.

We can also use a direct approach and look for the MLE in $\mathcal{E}_d(p)$. Let us assume to observe a sample of size n drawn from a d-dimensional Bernoulli distribution $\boldsymbol{X}$ with pmf $(f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}_d)$ that gives rise to counts $\boldsymbol{N} = (N_1, \dots N_m)$, where $m = 2^d$. Then the count $\boldsymbol{N}$ has a multinomial distribution with parameters $n, \boldsymbol{f}^{\mathcal{X}}$, i.e. $\boldsymbol{N} \sim \text{Multinomial}(n, \boldsymbol{f}^{\mathcal{X}})$, where the parameter $\boldsymbol{f}^{\mathcal{X}} = (f_i^{\mathcal{X}} : i = 1, \dots, m) := (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}_d)$ belongs to the $m-1$-simplex Δ_{m-1}. The likelihood function is

$$L(\boldsymbol{n}; \boldsymbol{f}^{\mathcal{X}}) = P(N_i = n_i, i \in I) = \binom{n}{n_1 \cdots n_m} \prod_{j=1}^{m} (f_j^{\mathcal{X}})^{n_j}, \tag{4.4}$$

where we set $0^0 := 1$.

If we assume that $\boldsymbol{X} \in \mathcal{E}_d(p)$, then $\boldsymbol{f}^{\mathcal{X}} = \boldsymbol{f}^{\mathcal{X}}(\lambda)$ has the form:

$$\boldsymbol{f}^{\mathcal{X}} = \sum_{i=1}^{k} \lambda_i \boldsymbol{e}_i^{\mathcal{X}}, \tag{4.5}$$

where $\boldsymbol{e}_i^{\mathcal{X}} = (e(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}_d)$ are the extremal points of $\mathcal{E}_d(p)$ and n_p is their number. The likelihood function of the count $\boldsymbol{N}$ is

$$L(\boldsymbol{n}; \boldsymbol{f}^{\mathcal{X}}) = \binom{n}{n_1 \cdots n_m} \prod_{j=1}^{m} (\sum_{i=1}^{n_p} \lambda_i e_{ij}^{\mathcal{X}})^{n_j} \tag{4.6}$$

where $\lambda = (\lambda_1, \dots, \lambda_{n_p}) \in \Delta_{n_p}$. The MLE can be found by maximizing the log-likelihood function in the simplex.

Example 2. *Let $\boldsymbol{X} \in \mathcal{E}_2(1/2)$. We have two extremal densities: the upper and lower Fréchet bound ([5]) $\boldsymbol{e}_U$ and $\boldsymbol{e}_L$. The count $\boldsymbol{N}$ has support on four points and the likelihood becomes:*

$$L(\boldsymbol{n}; \boldsymbol{f}^{\mathcal{X}}) = \binom{n}{n_1 \cdots n_4} \prod_{j=1}^{4} (\lambda_1 e_{Uj} + \lambda_2 e_{Lj})^{n_j}, \lambda \in \Delta_2. \tag{4.7}$$

By standard computations we find

$$L(\boldsymbol{n}; \boldsymbol{f}^{\mathcal{X}}) = \binom{n}{n_1 \cdots n_4} (\frac{\lambda_1}{2})^{n_1+n_4} (\frac{\lambda_2}{2})^{n_2+n_3}, \lambda \in \Delta_2. \tag{4.8}$$

The MLE can be found using the log-likelihood and the Lagrange multipliers and it is:

$$\hat{\lambda}_1 = \frac{N_1 + N_4}{n}, \quad \hat{\lambda}_2 = \frac{N_2 + N_3}{n}.$$

The same result has been found in [19] for the palindromic Bernoulli distributions, that in the 2-dimensional case coincide with the whole Fréchet class $\mathcal{E}_d(1/2)$.

4.2 Testing

Let $\mathcal{E}_d^*$ be one of the classes $\mathcal{E}_d$ or $\mathcal{E}_d(p)$. This section provides a generalized likelihood ratio (GLR) test for

$$H_0 : \boldsymbol{f}^{\mathcal{X}} \in \mathcal{E}_d^*$$

versus

$$H_A : \boldsymbol{f}^{\mathcal{X}} \in \mathcal{B}_d \setminus \mathcal{E}_d^*,$$

where in this case the class $\mathcal{E}_d^*$ is a d-simplex or a d-polytope and $\mathcal{B}_d$ is a (2^d-1)-simplex.

Let $m = 2^d$ and $\hat{\boldsymbol{f}}^{\mathcal{X}} = (\frac{N_1}{n}, \dots, \frac{N_m}{n})$ be the MLE estimator of $\boldsymbol{N} \sim \text{Multinomial}(\boldsymbol{n}, \boldsymbol{f}^{\mathcal{X}})$, where the parameter $\boldsymbol{f}^{\mathcal{X}} = (f_i^{\mathcal{X}} : i = 1, \dots, m) := (f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}_d)$ belongs to the $(m-1)$-simplex Δ_{m-1} and let $\hat{\lambda}$ be the MLE estimator for $\boldsymbol{X} \in \mathcal{E}_d^*$ with pmf $\boldsymbol{f}^{\mathcal{X}} = \boldsymbol{f}^{\mathcal{X}}(\lambda)$ as determined in the previous section. The GLR statistics is

$$\Lambda(\boldsymbol{N}) = \frac{\prod_{i=1}^{m} (f^{\mathcal{X}})_j(\hat{\lambda})^{N_j}}{\prod_{i=1}^{m} (\frac{N_j}{n})^{N_j}}, \tag{4.9}$$

where $\boldsymbol{N}$ is the count arised from $\boldsymbol{X}$. The α-level critical region is defined by

$$\alpha = P_0(\Lambda(\boldsymbol{N}) < c), \tag{4.10}$$

where P_0 is the probability measure under H_0. The value c is obtained observing that $-2\log(\Lambda(\boldsymbol{N}))$ is approximatively a χ_k^2 distribution with $k = m - 1 - \dim(\mathcal{E}_d^*)$, where $\dim(\mathcal{E}_d^*) = d$ if $\mathcal{E}_d^* = \mathcal{E}_d$ and $\dim(\mathcal{E}^*) = d - 1$ if $\mathcal{E}_d^* = \mathcal{E}_d(p)$.

Example 3. *Consider the MLE of $\boldsymbol{X} \in \mathcal{E}_2(1/2)$ in Example 2. The test statistics for*

$$H_0 : \boldsymbol{f}^{\mathcal{X}} \in \mathcal{E}_2(1/2)$$

versus

$$H_A : \boldsymbol{f}^{\mathcal{X}} \in \mathcal{B}_2 \setminus \mathcal{E}_2(1/2),$$

is

$$\Lambda(\boldsymbol{N}) = \frac{\left(\frac{N_1+N_4}{2n}\right)^{N_1+N_4} \left(\frac{N_2+N_3}{2n}\right)^{Y_2+Y_3}}{\prod_{i=1}^{4} \frac{N_j}{n}^{N_j}}, \tag{4.11}$$

and $-2\log(\Lambda(\boldsymbol{N}))$ has approximatively a χ_2^2 distribution. If we consider $\alpha = 0.05$ and $c_1 = 5.991$ the upper 0.95 quantile of a χ_1^2 distribution the critical region is defined by $c = e^{-2c_1}$.

Table 3: Extremal densities of $\mathcal{S}_5(\frac{1}{2}) \equiv \mathcal{E}_5(\frac{1}{2})$ and estimated MLE pmf

y	$r_1(y)$	$r_2(y)$	$r_3(y)$	$r_4(y)$	$r_5(y)$	$r_6(y)$	$r_7(y)$	$r_8(y)$	$r_9(y)$	$p_{MLE}(y)$
0	0.167	0.375	0.5	0	0	0	0	0	0	0.031
1	0	0	0	0.25	0.5	0.625	0	0	0	0.153
2	0	0	0	0	0	0	0.5	0.75	0.833	0.318
3	0.833	0	0	0.75	0	0	0.5	0	0	0.311
4	0	0.625	0	0	0.5	0	0	0.25	0	0.156
5	0	0	0.5	0	0	0.375	0	0	0.167	0.031

4.3 Example

Over the Spring 2009 semester, two Berkeley undergraduates undertook 40,000 tosses of a coin. The dataset and the description of the protocol which followed are available at `https://www.stat.berkeley.edu/~aldous/Real-World/coin_tosses.html`. Here, we rearrange this dataset as if the tosses had been undertaken five at a time and we use this dataset to find the MLE in the class $\mathcal{E}_5(\frac{1}{2})$. After finding the MLE in $\mathcal{E}_5(\frac{1}{2})$, we perform the GLR test described in Section 4.2.

To simplify the computations we look for the ML estimates in $\mathcal{S}_5(\frac{1}{2})$. We have nine extremal densities, provided in Table 3. The MLE estimate $\hat{\boldsymbol{\lambda}} = (\lambda_i,\ i = 1, \ldots, 9)$ which is also the ML estimate in $\mathcal{E}_5(\frac{1}{2})$ is

$$\hat{\boldsymbol{\lambda}} = (0.1, 0.019, 0.015, 0.188, 0.202, 0.008, 0.173, 0.174, 0.121). \tag{4.12}$$

For completeness we also exhibit the estimated ML pmf p_{MLE} in $\mathcal{S}_5(\frac{1}{2})$ in the last column of Table 3.

We now perform the GLR test for

$$H_0 : \boldsymbol{f}^{\chi} \in \mathcal{E}_5(\frac{1}{2})$$

versus

$$H_A : \boldsymbol{f}^{\chi} \in \mathcal{B}_d \setminus \mathcal{E}_5(\frac{1}{2}).$$

Let $\hat{\boldsymbol{f}}^{\chi} = (\frac{N_1}{40000}, \ldots, \frac{N_{32}}{40000})$ be the MLE for $\boldsymbol{N} \sim$ Multinomial$(40000, \boldsymbol{f}^{\chi})$ and let $\hat{\boldsymbol{\lambda}}$ be the MLE estimator for $\boldsymbol{X} \in \mathcal{E}_5(\frac{1}{2})$ provided in (4.12). The GLR statistics is $\Lambda(\boldsymbol{N})$ in (4.9). Since $-2\log(\Lambda(\boldsymbol{N}))$ is approximatively a χ^2_k distribution with $k = dim(\mathcal{B}_5) - dim(\mathcal{S}_5(1/2)) = 31 - 4 = 27$ degree of freedom, its observed value is 39.49 and $\chi^2_{0.95} = 40.113$ we do not reject the null hypothesis at level 0.05.

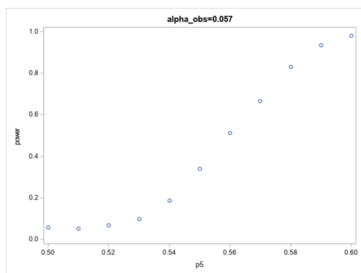
4.4 Power of the test

We test the alternative hypothesis that $\boldsymbol{f}$ has independent margins $g_1, \ldots, g_5$, where $g_i \sim B(1/2), i = 1, \ldots 4$ and $g_5 \in B(p)$, and we move p. We therefore consider the power of this test at level $\alpha = 0.05$ for the following alternative hypothesis:

$$H_A : \boldsymbol{f} = g_1 g_2 g_3 g_4 g_5; \quad g_i \sim B(1/2), i = 1, \dots 4,\ g_5 \in B(p),$$

where $p = 0.5 + 0.01n$, $n = 1, \dots 10$. Clearly if $p = 0.5$, then $\boldsymbol{f} \in \mathcal{E}_5(\frac{1}{2})$ otherwise $\boldsymbol{f} \notin \mathcal{E}_5(\frac{1}{2})$. This choice of H_A allows us to move from exchangeability by moving a single parameter. Thus, we study the power function $\gamma(p)$ as a function of p. The power fiunction $\gamma(p)$ has been estimated by simulation. For each value of p ($p = 0.5 + 0.01n$, $n = 1, \dots 10$), $N = 1000$ samples of size $n = 1000$ are taken from $\boldsymbol{f} = g_1 g_2 g_3 g_4 g_5$. The estimate $\hat{\gamma}(p)$ of $\gamma(p)$ is computed as $\frac{N_p}{N}$, where N_p is the number of times H_0 is rejected for the given p. The estimated power function $\hat{\gamma}(p)$ is shown in Figure 10. If $p = 0.5$, $\boldsymbol{f}$ is exchangeable and the observed α is 0.057, which is close to $\alpha = 0.05$. We can see that the power function is low if $p < 0.52$, than it increases and it becomes higher that 0.8 from $p = 0.58$. The power function is close to one when $p = 0.6$.

Figure 10: Power of the test $H_A : \boldsymbol{f} = g_1 g_2 g_3 g_4 g_5$, where $g_i \sim B(1/2), i = 1, \dots 4,\ g_5 \in B(p)$ and $p = 0.5 + 0.01n$, $n = 0, \dots 10$.



5 Further developments

This section shows that the geometrical structure of exchangeable Bernoulli pmfs holds in a more general framework, partial exchangeability. We also show that as well as exchangeable pmfs are in a one-to-one relationship with discrete distributions, partially exchangeable pmfs are in a one-to-one relationship with multivariate discrete distributions. The results we present here open the way to the study of this more general class of multivariate Bernoulli pmfs.

Definition 5.1. *Let $\mathcal{G}$ be a partition of $I = \{1, \dots, d\}$. A multivariate Bernoulli distribution $f(\boldsymbol{x})$ is partially exchangeable if $f(\sigma(\boldsymbol{x})) = f(\boldsymbol{x})$ for any $\sigma \in \mathcal{P}_d$ such that $\sigma(G) = G$ for any $G \in \mathcal{G}$. We say that σ and $f(\boldsymbol{x})$ are compatible with $\mathcal{G}$. We denote by $\mathcal{P}_d(\mathcal{G})$ the set of partitions compatible with $\mathcal{G}$ and $\mathcal{E}_d(\mathcal{G})$ the family of partially exchangeable distributions compatible with $\mathcal{G}$.*

Partial exchangeability is an extension of exchangeability, that is recovered by choosing the trivial partition $\mathcal{G} = \{I\}$.

Let $\mathcal{D}_{d_1,\ldots,d_n}$ be the class of multivariate discrete distributions with support on $J_1 \times \ldots \times J_n$ and $J_k = \{0,\ldots,d_k\}$ and $\mathcal{D}_{d_1,\ldots,d_n}(\boldsymbol{\mu}) = \mathcal{D}_{d_1,\ldots,d_n}(\mu_1,\ldots,\mu_n)$ the class of multivariate discrete distributions with support on $J_1 \times \ldots \times J_n$ and mean vector $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)$.

Theorem 5.1. *Let $\mathcal{G} = \{G_1,\ldots,G_n\}$, and $d_j = \#G_j$. There is a one to one map $F_{\mathcal{G}}$ between $\mathcal{E}_d(\mathcal{G})$ and $\mathcal{D}_{d_1,\ldots,d_n}$.*

Proof. Let $f \in \mathcal{E}_d(\mathcal{G})$. Since $f(\boldsymbol{x}) = f(\sigma(\boldsymbol{x}))$ for any $\sigma \in \mathcal{P}_d(\mathcal{G})$, any mass function $f(\boldsymbol{x})$ in $\mathcal{E}_d(\mathcal{G})$ uniquely defines a function $g : J_1 \times \ldots \times J_n \rightarrow$, where $J_k = \{0,\ldots,d_k\}$ and $d_k = \#G_k$ given by $g(j_1,\ldots,j_n) := f(\boldsymbol{x})$ if $\boldsymbol{x} = (x_1,\ldots,x_d) \in \mathcal{X}_d$ and $\#\{x_h \in G_i : x_h = 1\} = j_i,\ \ i = 1,\ldots n$. The map

$$
\begin{aligned}
F_{\mathcal{G}} : \mathcal{E}_d(\mathcal{G}) &\rightarrow \mathcal{D}_{d_1,\ldots,d_n} \\
f &\rightarrow p_D.
\end{aligned}
\tag{5.1}
$$

where $p_D(j_1,\ldots,j_n) = \binom{d_1}{j_1}\ldots\binom{d_n}{j_n} g(j_1,\ldots,j_n)$ is a bijection.

□

Notice that if $\boldsymbol{X}$ is partially exchangeable, each d_j-dimensional margin of the form $(X_i)_{i\in G_j}$ is a vector of exchangeable Bernoulli variables.

Remark 2. *Let $\mathcal{S}(\mathcal{G})$ be the class of random variables $\boldsymbol{Y} = (Y_1,\ldots,Y_n)$ defined by:*

$$
Y_j = \sum_{h\in G_j} X_h, \tag{5.2}
$$

then $p_{\boldsymbol{Y}}(j_1,\ldots,j_n) = p_D(j_1,\ldots,j_n)$. Thus, $\mathcal{S}(\mathcal{G}) = \mathcal{D}_{d_1,\ldots,d_n}$.

Corollary 5.1. *The class $\mathcal{E}_d(\mathcal{G})$ is a $d_{\mathcal{G}}$-simplex, where $d_{\mathcal{G}} = (d_1+1)\times\ldots\times(d_n+1)-1$. The class $\mathcal{E}_d(\mathcal{G})(\boldsymbol{\mu})$ of partially exchangeable distributions compatible with $\mathcal{G}$ and set of moments $\boldsymbol{\mu}$ is a $d_{\mathcal{G}}$-polytope, where $d_{\mathcal{G}} = (d_1+1)\times\ldots\times(d_n+1)-1$.*

This last result implies that all the analysis performed in the previous sections can be extended to partially exchangeable distributions.

Example 4. *Let $\boldsymbol{X} \in \mathcal{E}_4(\mathcal{G})$, where $\mathcal{G} = \{\{1,2\},\{3,4\}\}$. Let $\boldsymbol{Y} = (Y_1,Y_2)$ defined by*

$$
Y_1 = X_1 + X_2,\ \ Y_2 = X_3 + X_4. \tag{5.3}
$$

$\boldsymbol{Y} \in \mathcal{D}_{2,2}$ and $p_S(j_1,j_2) = \binom{2}{j_1}\binom{2}{j_2} f(j_1,j_2)$, $(j_1,j_2) \in J_1 \times J_2$. Therefore the vector $\boldsymbol{p}_Y = (p_Y(j_1,j_2))_{j_1,j_2\in J_1\times J_2}$ is a point in $\mathbb{R}^9$ and $\mathcal{E}_4(\mathcal{G})$ is a 8-simplex in $\mathbb{R}^9$.

5.1 Given means: the class $\mathcal{E}_d(\mathcal{G})(\boldsymbol{p})$

Let $\boldsymbol{X} \in \mathcal{E}_d(\mathcal{G})$, $\mathcal{G} = \{G_1, \ldots, G_n\}$ and assume that $E[X_i] = p_j$ if $i \in G_j$. Let $\boldsymbol{p} = (p_1, \ldots, p_n)$ the mean vector. We consider here the class $\mathcal{E}_d(\mathcal{G})(\boldsymbol{p})$ of Bernoulli pmf with mean vector $\boldsymbol{p}$. The map $F_{\mathcal{G}}$ induces a bijection between $\mathcal{E}_d(\mathcal{G})(\boldsymbol{p})$ and $\mathcal{D}_{d_1,\ldots,d_n}(\boldsymbol{dp}) = \mathcal{S}(\mathcal{G})(\boldsymbol{dp})$, where $\boldsymbol{dp} := (d_1p_1, \ldots, d_np_n)$.

Proposition 5.1. *Let $\boldsymbol{Y} \in \mathcal{D}_{d_1,\ldots,d_n}$ and let $p_{\boldsymbol{Y}}$ be its pmf. Then*

$$\boldsymbol{Y} \in \mathcal{S}_d(\boldsymbol{dp}) \iff \sum_{j_1=0}^{d_l} \cdots \sum_{j_n=0}^{d_n} (j_k - p_k d_k) p_Y(j_1, \ldots, j_n) = 0, \quad k = 1, \ldots, n.$$

Proof. Let $\boldsymbol{Y} \in \mathcal{D}_{d_1,\ldots,d_n}$. By Theorem 5.1 $\boldsymbol{Y} \in \mathcal{S}_d(\mathcal{G})(\boldsymbol{dp})$ iff $E[Y_k] = p_k d_k$. It holds

$$E[Y_k] = p_k d_k \Longleftrightarrow E[Y_k - p_k d_k] = 0 \Longleftrightarrow \sum_{j_1=0}^{d_l} \cdots \sum_{j_n=0}^{d_n} (j_k - p_k d_k) p_Y(j_1, \ldots, j_n) = 0.$$

□

From this proposition it can be proved that the extremal points of the polytope $\mathcal{E}_d(\mathcal{G})(\boldsymbol{p})$ have support on at most $n + 1$ points.

Example 5. *Let $\boldsymbol{X} \in \mathcal{E}_4(\mathcal{G})$, where $\mathcal{G} = \{\{1, 2\}, \{3, 4\}\}$ as in Example 4 and let $\boldsymbol{p} = (p_1, p_2)$ the mean vector. The convex polytope $\mathcal{E}_d(\mathcal{G})(\boldsymbol{p})$ is the set of solutions of the linear system:*

$$\begin{cases} -2p_1(p_{00} + p_{01} + p_{02}) + (1 - 2p_1)(p_{10} + p_{11} + p_{12}) + (2 - 2p_1)(p_{20} + p_{21} + p_{22}) = 0 \\ -2p_2(p_{00} + p_{10} + p_{20}) + (1 - 2p_2)(p_{01} + p_{11} + p_{21}) + (2 - 2p_2)(p_{02} + p_{12} + p_{22}) = 0 \end{cases},$$

therefore the extremal points have support on at most three points. As an example Table 4 provides the extremal points for $\boldsymbol{p} = (\frac{1}{2}, \frac{1}{4})$.

6 Conclusions

Building on the geometrical structure of the class of exchangeable Bernoulli distributions with given means as studied in [6], in this paper we have addressed different statistical issues: high dimensional simulation, estimation and testing. We provide an algorithm to simulate from a multivariate Bernoulli distribution with given mean even in high dimension. We also provide an algorithm to simulate a random sample with given mean from a parametrical family of distributions able to span the whole correlation range. This can be done virtually in any dimension with a small computational effort. Furthermore, we also perform uniform simulation on the class of exchangeable Bernoulli distributions

Table 4: Extremal densities of $\mathcal{S}(\mathcal{G})(1, \frac{1}{2}) \equiv \mathcal{E}_d(\mathcal{G})(\frac{1}{2}, \frac{1}{4})$

y	$r_1(y)$	$r_2(y)$	$r_3(y)$	$r_4(y)$	$r_5(y)$	$r_6(y)$	$r_7(y)$	$r_8(y)$	$r_9(y)$	$r_{10}(y)$	$r_{11}(y)$	$r_{12}(y)$	$r_{13}(y)$	$r_{14}(y)$
00	0	0	0	0	0	0	0	0	0.5	0.25	0.25	0.25	0.5	0.38
10	0	0.5	0.5	0.5	0.75	0.67	0.67	0.75	0	0	0	0.5	0	0
20	0.5	0	0	0.25	0	0	0	0	0	0.25	0.5	0	0.25	0.38
01	0.5	0	0.25	0	0	0	0.17	0	0	0	0	0	0	0
11	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0
21	0	0	0.25	0	0	0.17	0	0	0.5	0	0	0	0	0
02	0	0	0	0.25	0	0.17	0	0.13	0	0	0.25	0	0	0
12	0	0	0	0	0.25	0	0	0	0	0	0	0	0	0.25
22	0	0	0	0	0	0	0.17	0.13	0	0	0	0.25	0.25	0

with given mean in high dimension. The importance of high dimensional simulation of exchangeable Bernoulli variables is particularly evident in finance. The loss associated to a credit risk portfolio is a convex linear combination of indicators of defaults. Indicator of defaults are often assumed to be exchangeable, for example in homogeneous portfolios, where the obligors belong to the same class of rating. In a first application to credit risk, we show that using uniform simulation, we can investigate the relevant factors and measures of risk and their behavior across the class of possible joint defaults in a high dimensional portfolio ($d = 100$ in our application).

The extension to partially exchangeable distributions, that concludes this paper, opens the way to the simulation of partially homogeneous portfolios.

Acknowledgements

The authors gratefully acknowledge financial support from the Italian Ministry of Education, University and Research, MIUR, "Dipartimenti di Eccellenza" grant 2018-2022. The authors also thank the anonymous referees for their valuable suggestions.

A Proof of Proposition 2.1

Let Y be a discrete random variable defined over $\{0, \ldots, d\}$ and let p_Y be its pmf. Then it is easy to show that $Y \in \mathcal{S}_d(p)$ iff $E[Y] = pd$. It holds

$$E[Y] = pd \Longleftrightarrow E[Y - pd] = 0 \Longleftrightarrow \sum_{j=0}^{d} (j - pd) p_Y(j) = 0.$$

Therefore,

$$Y \in \mathcal{S}_d(p) \iff \sum_{j=0}^{d} (j - pd) p_Y(j) = 0.$$

Hence, the extremal points of $\mathcal{S}_d(p)$ are the positive, normalized extremal solutions $\boldsymbol{p}_S = (p_0, \ldots, p_d)$ of

$$\sum_{j=0}^{d} (j - pd) p_j = 0. \tag{A.1}$$

with the conditions $p_j \geq 0$, $j = 0, \ldots, d$ and $\sum_{j=0}^{d} p_j = 1$. Let $a_j = j - pd$ and consider the case pd not integer. Equation (A.1) becomes

$$\sum_{j=0}^{d} a_j p_j = 0.$$

Using the standard theory of linear systems [6] proved that the extremal points of $\mathcal{S}_d(p)$ have at most two non zero components, say j_1, j_2. Therefore the extremal solutions $\boldsymbol{r} = (r_0, \ldots, r_d)$ of A.1 can be found considering the equations

$$a_{j_1} r_{j_1} + a_{j_2} r_{j_2} = 0,$$

where we make the non restrictive assumption $j_1 < j_2$. The equation (A.1) has positive solutions only if $a_{j_1} a_{j_2} < 0$. We observe that $a_{j_1} < 0$ for $0 \leq j_1 \leq j_1^M$ and $a_{j_2} > 0$ for $j_2^m \leq j_2 \leq d$. In this case we have $j_2^m = j_1^M + 1$. It follows that for $0 \leq j_1 \leq j_1^M$ and $j_2^m \leq j_2 \leq d$ we have $a_{j_1} a_{j_2} < 0$. A positive solution of Equation (A.1) is

$$\begin{cases} \tilde{r}_y(j_1) = j_2 - pd \\ \tilde{r}_y(j_2) = pd - j_1 \end{cases} .$$

We have $\tilde{r}_y(j_1) + \tilde{r}_y(j_2) = j_2 - pd + pd - j_1 = j_2 - j_1$ and then the normalized extremal rays corresponding to j_1 and j_2 are given by (2.5). If pd is integer we have $a_{pd} = 0$. It follows that (2.8) is also an extremal solution. The number of extremal solutions easily follows.

References

[1] W. Jiang, S. Song, L. Hou, and H. Zhao, "A set of efficient methods to generate high-dimensional binary data with specified correlation structures," *The American Statistician*, pp. 1–13, 2020.

[2] İ. Basoğlu, W. Hörmann, and H. Sak, "Efficient simulations for a Bernoulli mixture model of portfolio credit risk," *Annals of Operations Research*, vol. 260, no. 1, pp. 113–128, 2018.

[3] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative risk management*, vol. 3. Princeton University Press, 2005.

[4] P. Diaconis, "Finite forms of De Finetti's theorem on exchangeability," *Synthese*, vol. 36, no. 2, pp. 271–281, 1977.

[5] R. Fontana and P. Semeraro, "Representation of multivariate Bernoulli distributions with a given set of specified moments," *Journal of Multivariate Analysis*, vol. 168, pp. 290–303, 2018.

[6] R. Fontana, E. Luciano, and P. Semeraro, "Model risk in credit risk," *Mathematical Finance*, pp. 1–27, 2020.

[7] J. Dhaene and M. Denuit, "The safest dependence structure among risks," *Insurance: Mathematics and Economics*, vol. 25, no. 1, pp. 11–21, 1999.

[8] M. De Berg, M. Van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational geometry*. Springer, 1997.

[9] R. Wang and Y. Wei, "Risk functionals with convex level sets," *Mathematical Finance*, 2020.

[10] P. Stein, "A note on the volume of a simplex," *The American Mathematical Monthly*, vol. 73, no. 3, pp. 299–301, 1966.

[11] G. Varsi, "The multidimensional content of the frustum of the simplex," *Pacific Journal of Mathematics*, vol. 46, no. 1, pp. 303–314, 1973.

[12] L. Calès, A. Chalkis, I. Z. Emiris, and V. Fisikopoulos, "Practical volume computation of structured convex bodies, and an application to modeling portfolio dependencies and financial crises," in *34th International Symposium on Computational Geometry (SoCG 2018)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[13] S. D. Oman, "Easily simulated multivariate binary distributions with given positive and negative correlations," *Computational Statistics & Data Analysis*, vol. 53, no. 4, pp. 999–1005, 2009.

[14] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," 1970.

[15] *SAS-IML 14.1 User's Guide.* Cary, NC: SAS Institute Inc., 2015.

[16] S. Laurent, *uniformly: Uniform Sampling*, 2018. R package version 0.1.0.

[17] V. Fisikopoulos, A. Chalkis, and contributors in file inst/AUTHORS, *volesti: Volume Approximation and Sampling of Convex Polytopes*, 2020. R package version 1.1.2.

[18] R. Frey and A. J. McNeil, "Dependent defaults in models of portfolio credit risk," *Journal of Risk*, vol. 6, pp. 59–92, 2003.

[19] G. M. Marchetti, N. Wermuth, *et al.*, "Palindromic Bernoulli distributions," *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 2435–2460, 2016.