Doctoral Dissertation
Doctoral Program in Pure and Applied Mathematics (38th cycle)

Negative Dependence in Bernoulli Distributions

Algebraic and geometric representations with applications to copula modeling

Alessandro Mutti

* * * * *

Supervisors
Prof. Patrizia Semeraro, Supervisor
Prof. Giuseppe D'Onofrio, Co-Supervisor

Doctoral examination committee
Prof. Carole Bernard
Prof. Fabrizio Durante

Politecnico di Torino
July 21, 2026

I hereby declare that the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .

Alessandro Mutti

Turin, July 21, 2026

Summary

This thesis investigates multidimensional Bernoulli distributions through algebraic and geometric representations, which prove to be effective tools to address open issues in the theory of negative dependence. Bernoulli distributions play a central role in many applied contexts, such as statistical physics, finance, and insurance, yet characterizing their joint behavior under negative dependence remains a significant challenge. By focusing on Fréchet classes, i.e., sets of joint distributions with fixed univariate marginals, we treat these structures as convex polytopes. This setting allows for a study of dependence properties independently of marginal effects. In this analysis, we employ two main algebraic frameworks: the classical probability generating functions and a recent representation for identically distributed marginals based on ideals of polynomials.

A primary focus of this work is the characterization of extremal negative dependence. Unlike positive dependence, which is bounded by the comonotonic distribution, negative dependence does not have a unique, well-defined lower bound in dimensions $d > 2$. We concentrate on Σ-countermonotonicity, a property of extremal negative dependence that generalizes the concept of countermonotonicity to high dimensions. Our first major theoretical result proves that, in the Bernoulli setting, Σ-countermonotonicity coincides with the minimality of aggregate risk in the convex order. This simplifies the task of finding minimal risk bounds, which is essential for credit risk and insurance portfolios.

Furthermore, we integrate the strongly Rayleigh property into the framework of extremal dependence. As a form of stochastic repulsion based on the geometry of zeros of probability generating functions, the strongly Rayleigh property is stronger than negative association. Our second main result demonstrates that every Bernoulli Fréchet class contains at least one distribution satisfying both Σ-countermonotonicity and the SR property. We identify this distribution through entropy maximization within the class of Σ-countermonotonic Bernoulli distributions.

These theoretical findings are applied to specific subclasses, namely symmetric Bernoulli and Conway–Maxwell-binomial distributions, where we explicitly characterize extremal negative dependence distributions and parameter ranges that guarantee strongly Rayleigh, respectively. Finally, we extend our results to copula theory. Using stochastic representations, we construct extremal mixture and Farlie–Gumbel–Morgenstern copulas, and we also consider a generalization of the latter. This extension allows us to move beyond Bernoulli marginals and provide analytical tools for determining lower bounds in high-dimensional and continuous settings for risk measures, such as Value-at-Risk, expected shortfall, and the entropic risk measure.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Patrizia Semeraro, for her invaluable guidance, constant support, and the immense time she dedicated to my academic training. I am truly grateful for her meticulous care and for the insightful advice that consistently challenged me to improve. Her mentorship has been fundamental not only to the completion of this thesis but also to my growth as a researcher and as a person. I would also like to thank my co-supervisor, Professor Giuseppe D'Onofrio, for his support during this journey. I deeply appreciate his availability and the fresh perspective he brought to our discussions, which have been a significant source of encouragement. I am also profoundly grateful to both of them for their guidance regarding my future professional path.

A special and sincere thank you goes to Professor Hélène Cossette and Professor Etienne Marceau for hosting me during my research period at Université Laval. I am deeply grateful for their warm hospitality, which made me feel at home even when far from it, and for the intellectual generosity they showed in collaborating on this research. Much of the material presented in these pages has benefited immensely from the stimulating discussions and the work we did together.

Contents

Introduction

The main subject of this thesis is the study of multidimensional Bernoulli distributions through algebraic and geometrical representations, which have proven to be effective tools for solving relevant open issues in the theory of negative dependence. Bernoulli distributions and negative dependence naturally play a central role in many contexts, including probability (e.g., [17, 58]), statistics (e.g., [19, 53]), combinatorics, and statistical physics (e.g., percolation, Ising models, or interacting systems; see [37, 57, 72, 82]). To study dependence properties, it is important to consider Fréchet classes, defined as the set of all joint distributions with fixed univariate marginals, because they allow us not to consider the effect of marginals when comparing dependence structures (see [63]). From a geometrical perspective, Fréchet classes of multidimensional Bernoulli distributions are convex polytopes ([46]). Beyond this geometrical representation, we also employ algebraic characterizations. The first algebraic representation we consider is via probability generating functions; this approach is relevant for defining a strong property of negative dependence, the strongly Rayleigh property ([15]). Moreover, we also adopt the representation introduced in [45] for Fréchet classes with identically distributed marginals. This latter representation allows for an easy identification of a set of generators for these Fréchet classes in any dimension. These algebraic tools are useful for the characterization of negative dependence throughout this thesis.

The theory of negative dependence is much more challenging than its positive counterpart. The fundamental ingredients of the dependence theory are dependence properties and dependence orders (see [69, 93]). Dependence orders are partial orders defined on Fréchet classes and allow for the comparison of the strength of dependence between different distributions. In particular, negative dependence tries to capture the idea of random variables that repel each other; that is, high values of one random variable tend to inhibit high values in others. One of the strongest and well-known properties in this sense is negative association; see [62]. Most negative dependence properties introduced in the literature are defined via comparison with the element of their Fréchet class with independent marginals, according to a specific dependence order; see [69] for the positive dependence case. From this comparison, the concepts of extremal negative dependence arise, identifying dependence structures that are minimal with respect to a given order or that minimize functionals of the type $\mathbb{E}[\phi(\boldsymbol{X})]$ for specific classes of functions ϕ. There is a substantial difference between extremal positive and extremal negative dependence. Within any Fréchet class, dependence orders are bounded by the upper and the lower Fréchet bounds. However, while the upper bound always belongs to the class, the lower Fréchet bound does not in general. The upper Fréchet bound is the joint distribution of a vector with components that are increasing transformations of a common random variable. This structure is usually named comonotonicity ([68]), and it corresponds to the extremal positive dependence. In contrast, extremal negative dependence is an intuitive concept only in the bivariate case. A two-dimensional random vector is countermonotonic if its components are oppositely ordered, and in this case, its distribution corresponds with the lower Fréchet bound. In dimensions $d > 2$, instead, the lower Fréchet bound often fails to be a distribution function; see [32]. Therefore, generalizing the concept of countermonotonicity to higher dimensions remains an open challenge with significant implications for optimization problems in different fields, such as applied mathematics [71], mathematical finance [6, 43], and actuarial science [25, 70]. In [87], the authors review the most popular extremal dependence properties, and they introduce a new one: Σ-countermonotonicity, the notion of extremal dependence we focus on. Its definition is not directly linked to a dependence order: a random vector is Σ-countermonotonic if the sum of any subset of its components is countermonotonic with the sum of the other components. However, this notion is strictly linked to the convex order of the sum of all its components. This order is relevant, e.g., in insurance and finance, where it indicates which of two aggregate risks has the lower variability. Our first main theoretical result is indeed the characterization of Σ-countermonotonicity: we prove that, in the Bernoulli setting, Σ-countermonotonicity coincides with minimality of aggregate risk in the convex order. This also implies that the class of Σ-countermonotonic Bernoulli distributions has the geometrical structure of a convex polytope.

A further objective of this work is to include the strongly Rayleigh property within the framework of extremal negative dependence. Introduced in [15], the strongly Rayleigh property implies negative association. However, it is defined using a completely different approach, based on the geometry of zeros of probability generating functions, and on a property called stability. A crucial distinction is that the strongly Rayleigh property is not defined through a comparison with independence through a standard dependence order. To the best of our knowledge, there does not exist any dependence order that can compare the strength of the strongly Rayleigh property in two distributions within the same Fréchet class. Consequently, a notion of extremal negative dependence cannot be directly derived from this property. We therefore study the strongly Rayleigh property within the class of Σ-countermonotonic Bernoulli distributions. Another fundamental result of this thesis demonstrates that in every Bernoulli Fréchet class, there exists a distribution that satisfies both Σ-countermonotonicity and the strongly Rayleigh property. We identify this distribution by maximizing the entropy within the convex polytope of Σ-countermonotonic random variables. We also show that usually there is more than one distribution with both these properties in some Fréchet classes.

The theoretical framework developed in this thesis is applied to two specific subclasses of multivariate Bernoulli distributions. First, we focus on the Fréchet class of symmetric Bernoulli distributions, where all marginal means are equal to $1/2$. To analyze this class, we adopt the algebraic representation introduced by [45]. This representation allows us to map the probability mass functions into an ideal of polynomials, providing a robust algebraic framework for analyzing dependence structures. By utilizing this map, we can identify the generators of the class more efficiently, without losing essential statistical properties. This algebraic approach proves effective in characterizing extremal negative dependence within the symmetric Fréchet class, simplifying a problem that would otherwise be computationally complex. Second, we investigate the Conway–Maxwell-binomial family and its underlying exchangeable Bernoulli distributions. Within this framework, we study the strongly Rayleigh property, identifying values of the parameters that ensure this strong form of dependence. We show that it is possible to parametrize the family of exchangeable Bernoulli distributions such that this specific class is contained in a given Fréchet class. Moreover, we also prove that this family spans the full spectrum of dependence, from positive to negative supermodular dependence.

The findings of this thesis are further extended to the continuous framework of copula theory through the use of specific stochastic representations. Through these representations, we can construct several families of copulas starting from the Bernoulli random vectors studied in the previous chapters. This step allows us to analyze dependence structures and aggregate risk of portfolios with any marginal distribution, not just Bernoulli. Since copulas are, by definition, joint cumulative distribution functions belonging to the Fréchet class of distributions with uniform marginals, we can study extremal negative dependence also in this class. We construct and analyze three main families of copulas derived from Bernoulli vectors: extremal mixture copulas, Farlie–Gumbel–Morgenstern (FGM) copulas, and generalized FGM (GFGM) copulas. From symmetric Bernoulli random vectors, for which we have already established a full characterization of extremal dependence, we study the associated extremal mixture and FGM copulas. This allows us to identify copulas that satisfy properties of extremal negative dependence. We focus on their bivariate dependence structure using classical measures such as Pearson's correlation, Spearman's rho, and Kendall's tau, showing how the extremal properties of the underlying Bernoulli distributions are inherited by the copulas.

Furthermore, we study a generalization of FGM copulas. This generalization allows us to move beyond symmetric Bernoulli distributions and consider Bernoulli random vectors of any Fréchet class. Indeed, a family of generalized FGM (GFGM) copulas is characterized by d parameters $p_1, \dots, p_d$, which correspond to the marginal means of the Bernoulli random vector used to construct these copulas. We investigate the general geometric structure of the GFGM families, showing that they inherit the convex polytope structure of the corresponding Fréchet classes. Finally, we provide specific results for the case of identically distributed risks and GFGM copulas with a single parameter p. Under these assumptions, a key part of our analysis is the study of the convex order of the sum of identically distributed random variables $Y_1, \dots, Y_d$, whose dependence structure is described by a copula of this class. By establishing conditions for the convex ordering of these sums, we provide analytical tools to find lower bounds for risk measures such as Value-at-Risk, expected shortfall, and entropic risk measure, effectively demonstrating how the geometric properties of discrete Bernoulli vectors can be used in risk management problems in a continuous setting.

The thesis is organized as follows. We present the general structure of Bernoulli distributions in Chapter 1. We examine the classical parameterizations and their representation as an exponential family. Moreover, we describe the geometrical structure of this class and an algebraic approach to its characterization by defining the probability generating function. We also consider Fréchet classes, and we study their geometrical structure and their algebraic representation in the case of identically distributed marginals. A conclusive section investigates the connection between Bernoulli distributions and univariate distributions with finite support. This analysis is particularly significant for the study of aggregate risk, represented by the sum of the risks in a portfolio. In this context, we introduce the exchangeable Bernoulli random vectors, and we examine, from a geometrical point of view, the class of Bernoulli random vectors whose sums follow a specific univariate distribution. Different forms of negative dependence are presented and studied in Chapter 2. We introduce negative dependence notions, extremal negative dependence properties, and the strongly Rayleigh property, and we study their link. In this chapter, we also present our main theoretical results regarding the characterization of extremal negative dependence and the strongly Rayleigh property in this context. In Chapter 3, we focus on symmetric Bernoulli distributions and on the Conway–Maxwell-binomial distribution (and the underlying exchangeable Bernoulli). We study extremal negative dependence properties in the first class, using the representation introduced by [45], and the strongly property in the latter family. In Chapter 4, we present the main application of the results developed in the thesis, with the construction of copulas and the study of their negative dependence properties. This allows us to extend the study of aggregate risk beyond the Bernoulli case. Furthermore, in Appendix A, we recall some known results on the exponential families of distributions, which we use to describe certain classes of Bernoulli distributions and in some proofs throughout the thesis. Finally, in Appendix B, we consider the three-dimensional case, and we provide a simplified criterion to easily verify the strongly Rayleigh property.

Chapter 1 and Chapter 2 (up to Section 2.4) mainly review established results from the literature, providing the theoretical framework for the thesis. From Section 2.5 onwards, the dissertation presents original research written in collaboration with other researchers. These original contributions are based on the following research papers:

1. Cossette H., Marceau E., Mutti A., and Semeraro P., Generalized FGM dependence: geometrical representation and convex bounds on sums, Statistical Papers 66(7) (2025) 149.

2. Mutti A., and Semeraro P., Symmetric Bernoulli distributions and minimal dependence copulas, Journal of Multivariate Analysis 212 (2026) 105545.

3. Cossette H., Marceau E., Mutti A., and Semeraro P., Extremal negative dependence and the strongly Rayleigh property, *forthcoming in* Bernoulli, 2025, arXiv preprint arXiv:2504.17679.

4. Cossette H., Marceau E., Mutti A., and Semeraro P., Conway–Maxwell multivariate Bernoulli distribution, 2026, arXiv preprint arXiv:2604.23367.

Chapter 1

Multidimensional Bernoulli distributions

In this chapter, we provide the necessary theoretical background for multidimensional Bernoulli distributions, focusing on their statistical parameterizations, geometrical structures, and algebraic representations. These tools form the essential foundation for the analysis conducted in the subsequent chapters.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a Bernoulli random vector and $f : \mathcal{X}_d \to [0,1]$ its probability mass function (pmf), where $\mathcal{X}_d \coloneqq \{0,1\}^d$. We denote by $\mathcal{B}_d$ the class of all d-dimensional Bernoulli pmfs, and by $\mathcal{B}_d(\boldsymbol{p}) = \mathcal{B}_d(p_1, \ldots, p_d)$ the Fréchet class of pmfs with fixed marginal means $\boldsymbol{p} = (p_1, \ldots, p_d)$. We assume that the set $\mathcal{X}_d$ of 2^d binary vectors is ordered according to the reverse lexicographical criterion ($\prec_{\text{r-lex}}$). Specifically, for $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^d$, we have $\boldsymbol{x} \prec_{\text{r-lex}} \boldsymbol{y}$ if $x_j < y_j$ at the largest index $j \in \{1, \ldots, d\}$ where they differ. For example,

$$\begin{aligned}
\mathcal{X}_2 &= \{(0,0), (1,0), (0,1), (1,1)\}\\
\mathcal{X}_3 &= \{(0,0,0), (1,0,0), (0,1,0), (1,1,0), (0,0,1), (1,0,1), (0,1,1), (1,1,1)\}.
\end{aligned}$$

This order allows for a natural recursive partition of $\mathcal{X}_d$ into two disjoint subsets based on the last component. For each $\boldsymbol{i} \in \mathcal{X}_{d-1}$, we define the vector $\boldsymbol{s_i} \coloneqq (\boldsymbol{i},0) \in \mathcal{X}_d$, and we have

$$\mathcal{X}_d = \{\boldsymbol{s_i} \colon \boldsymbol{i} \in \mathcal{X}_{d-1}\} \cup \{\mathbf{1}_d - \boldsymbol{s_i} \colon \boldsymbol{i} \in \mathcal{X}_{d-1}\}, \tag{1.1}$$

where $\mathbf{1}_d$ is the d-dimensional vector of ones.

We assume that vectors $\boldsymbol{x} = (x_1, \ldots, x_d)$ are column vectors. Let $\mathbb{R}^{n\times m}$ be the space of real-valued matrices with n rows and m columns. The transpose of a matrix $A \in \mathbb{R}^{n\times m}$ is denoted by $A^\top \in \mathbb{R}^{m\times n}$. Given two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{d\times \ell}$, we denote their Kronecker product $A \otimes B = ((a_{ij}B) \colon i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}) \in \mathbb{R}^{nd\times m\ell}$, and

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

Moreover, if $n = d$, $A||B \in \mathbb{R}^{n\times(m+\ell)}$ denotes the row concatenation of A and B, while, if $m = \ell$, $A//B \in \mathbb{R}^{(n+d)\times m}$ denotes their column concatenation.

A fundamental aspect of this framework is the one-to-one correspondence between the set of binary vectors $\mathcal{X}_d$ and the power set of $\{1, \dots, d\}$. Specifically, for any vector $\boldsymbol{x} \in \mathcal{X}_d$, we define the set of indices corresponding to 1's as

$$J_{\boldsymbol{x}} \coloneqq \{j \in \{1, \dots, d\} : x_j = 1\}.$$

For any vector $\boldsymbol{x} \in \mathcal{X}_d$, we denote by $|\boldsymbol{x}|$ the sum of its components, namely $|\boldsymbol{x}| = x_1 + \cdots + x_d$. Note that $|\boldsymbol{x}|$ coincides exactly with the cardinality of the set $J_{\boldsymbol{x}}$, i.e., $|\boldsymbol{x}| = |J_{\boldsymbol{x}}|$, justifying the use of the same notation. By analogy, if we have a vector of marginal means $\boldsymbol{p} \in (0,1)^d$, we denote by $|\boldsymbol{p}| = p_1 + \cdots + p_d$ the sum of the means. Finally, we use the multi-index notation $\boldsymbol{z}^{\boldsymbol{x}} = z_1^{x_1} \cdots z_d^{x_d} = z_{j_1} \cdots z_{j_k}$, for $\boldsymbol{x} \in \mathcal{X}_d$, where $k = |\boldsymbol{x}|$ and $(j_1, \dots, j_k) = J_{\boldsymbol{x}}$.

The chapter is organized as follows. In Section 1.1, we introduce the fundamental representations of Bernoulli distributions, focusing on their geometric and algebraic properties via the probability generating function. Section 1.2 extends this analysis to Fréchet classes, characterizing them as convex polytopes and developing an algebraic framework for the case of common marginal means. Finally, Section 1.3 explores the relationship between a Bernoulli vector and the distribution of its sum, with a specific focus on the exchangeable case, which provides a one-to-one correspondence between multivariate and univariate discrete distributions.

1.1 Representations of Bernoulli distributions

In this section, we introduce several representations of the class $\mathcal{B}_d$ of all d-variate Bernoulli distributions. We begin by characterizing the geometrical structure of this class. Every pmf $f \in \mathcal{B}_d$ can be represented as a point in $\mathbb{R}^{2^d}$ by the vector

$$\boldsymbol{f} = (f_{\boldsymbol{x}} \colon \boldsymbol{x} \in \mathcal{X}_d) \in \mathbb{R}^{2^d},$$

where $f_{\boldsymbol{x}} \coloneqq f(\boldsymbol{x})$. This representation is unique since we assume that $\mathcal{X}_d = \{0,1\}^d$ is ordered according to the reverse-lexicographical order; therefore, we use the term pmfs to denote also the vectors $\boldsymbol{f}$. This implies that the set of all d-variate Bernoulli distributions $\mathcal{B}_d$ is in a one-to-one correspondence with the standard simplex Δ_{2^d-1} in $\mathbb{R}^{2^d}$, i.e.,

$$\mathcal{B}_d \leftrightarrow \Delta_{2^d-1} = \{\boldsymbol{f} \in \mathbb{R}^{2^d}_+ : \mathbf{1}^\top \boldsymbol{f} = 1\}$$

The notations $\boldsymbol{X} \in \mathcal{B}_d$ or $\boldsymbol{f} \in \mathcal{B}_d$ mean that the random vector $\boldsymbol{X}$ has pmf $f \in \mathcal{B}_d$. Building on this, we introduce different statistical parameterizations of $\mathcal{B}_d$ in Section 1.1.1, including the representation as an exponential family in Section 1.1.2. Finally, Section 1.1.3 provide an algebraic representation via the probability generating function.

1.1.1 Linear and log-linear parameterizations

We introduce two parameterizations of $\mathcal{B}_d$ that will be used throughout this work: the log-linear and the linear parameterization. These parameterizations are indexed by vectors $\boldsymbol{i} \in \mathcal{X}_d$, and each uniquely determines the pmf $f \in \mathcal{B}_d$. A third parameterization, named multivariate-logistic, was introduced by [56]; we refer to that work for details. Recall the design matrices

$$T_d = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{\otimes d}, \qquad T_d^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}^{\otimes d},$$

which are $2^d \times 2^d$ lower-triangular matrices, and the $2^d \times 2^d$ matrices

$$H_d = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes d}, \qquad H_d^{-1} = 2^{-d}\, H_d.$$

The *log-linear parameters* $\boldsymbol{\theta} = (\theta_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d)$ are defined by

$$\boldsymbol{\theta} = T_d^{-1} \log \boldsymbol{f}, \tag{1.2}$$

where $\log \boldsymbol{f} = (\log f_{\boldsymbol{x}} \colon \boldsymbol{x} \in \mathcal{X}_d)$. Componentwise, for every $\boldsymbol{i} \in \mathcal{X}_d$, we have

$$\theta_{\boldsymbol{i}} = \sum_{\boldsymbol{x} \in \mathcal{X}_d \colon \boldsymbol{x} \leq \boldsymbol{i}} (-1)^{|\boldsymbol{i}-\boldsymbol{x}|} \log f_{\boldsymbol{x}}, \tag{1.3}$$

where $\boldsymbol{x} \leq \boldsymbol{i}$ means $x_j \leq i_j$ for all $j \in \{1, \ldots, d\}$ and $|\boldsymbol{i} - \boldsymbol{x}| = \sum_{j=1}^{d} (i_j - x_j)$. The inverse map is $\log \boldsymbol{f} = T_d\, \boldsymbol{\theta}$, i.e., $\log f_{\boldsymbol{x}} = \sum_{\boldsymbol{i} \in \mathcal{X}_d \colon \boldsymbol{i} \leq \boldsymbol{x}} \theta_{\boldsymbol{i}}$ for every $\boldsymbol{x} \in \mathcal{X}_d$. The parameter $\theta_{\mathbf{0}_d}$ is determined by the normalization constraint and the free parameters are $(\theta_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d^-)$, ranging freely over $\mathbb{R}^{2^d - 1}$, where $\mathcal{X}_d^- \coloneqq \mathcal{X}_d \setminus \{\mathbf{0}_d\}$.

The *linear* (or *moment*) *parameters* $\boldsymbol{\mu} = (\mu_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d)$ are defined by

$$\boldsymbol{\mu} = M_d \boldsymbol{f}, \tag{1.4}$$

where M_d is the transpose of the matrix T_d,

$$M_d = T_d^{\top} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{\otimes d},$$

and, componentwise,

$$\mu_{\boldsymbol{i}} = \sum_{\boldsymbol{x} \in \mathcal{X}_d \colon \boldsymbol{i} \leq \boldsymbol{x}} f_{\boldsymbol{x}} = \mathbb{E}[\boldsymbol{X}^{\boldsymbol{i}}] = \mathbb{E}\Big[\prod_{j=1}^{d} X_j^{i_j}\Big].$$

In particular, $\mu_{\mathbf{0}_d} = 1$, $\mu_{\boldsymbol{e}_j} = \mathbb{P}(X_j = 1)$ for $j \in \{1, \ldots, d\}$, and $\mu_{\boldsymbol{e}_{j_1} + \boldsymbol{e}_{j_2}} = \mathbb{P}(X_{j_1} = 1, X_{j_2} = 1)$ for $1 \leq j_1 < j_2 \leq d$. The inverse map is given by $\boldsymbol{f} = M_d^{-1} \boldsymbol{\mu}$, i.e.,

$$f_{\boldsymbol{x}} = \sum_{\boldsymbol{i} \in \mathcal{X}_d \colon \boldsymbol{i} \geq \boldsymbol{x}} (-1)^{|\boldsymbol{i}-\boldsymbol{x}|}\, \mu_{\boldsymbol{i}}, \qquad \boldsymbol{x} \in \mathcal{X}_d.$$

Again, the free parameters are $(\mu_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d^-)$. Notice that not every vector $\boldsymbol{\mu} \in \mathbb{R}^{2^d}$ with $\mu_{\mathbf{0}_d} = 1$ corresponds to a valid pmf; the admissible region is a convex polytope determined by the constraints that $f_{\boldsymbol{x}} \geq 0$ for all $\boldsymbol{x} \in \mathcal{X}_d$.

The parameterizations introduced above use the indicator coding, but it is possible to define the same parameterizations using the effect coding; see [105]. To switch to the effect coding, we replace $x_j \in \{0,1\}$ with $x'_j \in \{-1,1\}$ via the linear transformation $x'_j = 2x_j - 1$.

By so doing, the two categories are symmetric around zero. Under effect coding, the design matrices T_d and T_d^{-1} are replaced by the matrix H_d and its inverse $H_d^{-1} = 2^{-d} H_d$. In particular, the corresponding log-linear and linear parameters under the effect coding are

$$\boldsymbol{\theta}' = 2^{-d} H_d \log \boldsymbol{f} = 2^{-d} H_d T_d \boldsymbol{\theta}, \qquad \boldsymbol{\mu}' = H_d \boldsymbol{f} = H_d M_d^{-1} \boldsymbol{\mu}.$$

Componentwise, we have, for $\boldsymbol{i} \in \mathcal{X}_d$,

$$\theta'_{\boldsymbol{i}} = 2^{-d} \sum_{\boldsymbol{x} \in \mathcal{X}_d} (-1)^{\boldsymbol{x} \cdot \boldsymbol{i}} \log f_{\boldsymbol{x}},$$

and

$$\mu'_{\boldsymbol{i}} = \sum_{\boldsymbol{x} \in \mathcal{X}_d} (-1)^{\boldsymbol{x} \cdot \boldsymbol{i}} f_{\boldsymbol{x}},$$

where $\boldsymbol{x} \cdot \boldsymbol{i} = x_1 i_1 + \cdots + x_d i_d$ denotes the scalar product between $\boldsymbol{x}$ and $\boldsymbol{i}$. Since H_d is not triangular, both the log-linear parameters and the linear ones under effect coding depend on the entire probability vector $\boldsymbol{f}$. We refer to [105] for a comparison of the two codings. The parametrization through effect coding is relevant for the characterization of the class of palindromic Bernoulli distributions, which will be relevant in the results of Section 3.1. A Bernoulli distribution with pmf f is palindromic if $f(\boldsymbol{x}) = f(\mathbf{1}_d - \boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}_d$.

Proposition 1.1.1 ([74, Proposition 2.2]). *A Bernoulli distribution is palindromic if and only if all odd-order linear or log-linear parameters under the effect coding vanish, i.e., if and only if $\mu'_{\boldsymbol{i}} = 0$ for all $\boldsymbol{i} \in \mathcal{X}_d$ with $|\boldsymbol{i}|$ odd, or $\theta'_{\boldsymbol{i}} = 0$ for all $\boldsymbol{i} \in \mathcal{X}_d$ with $|\boldsymbol{i}|$ odd.*

1.1.2 Bernoulli distributions as an exponential family

Following [31], the log-linear parameters in (1.2) coincide with the natural parameters of the exponential family representation of $\mathcal{B}_d$. The main results we need on exponential families of distributions are recalled in Appendix A. In the context of multidimensional Bernoulli distributions, taking as sufficient statistics

$$t_{\boldsymbol{i}}(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{i}} = \prod_{j=1}^{d} x_j^{i_j}, \qquad \boldsymbol{i} \in \mathcal{X}_d^-,$$

and the counting measure m on $\mathcal{X}_d$ as the carrying measure, any pmf $f \in \mathcal{B}_d$ can be written as

$$f(\boldsymbol{x}) = \exp\Big(\sum_{\boldsymbol{i} \in \mathcal{X}_d^-} \theta_{\boldsymbol{i}}\, \boldsymbol{x}^{\boldsymbol{i}} - \psi(\boldsymbol{\theta}) \Big), \tag{1.5}$$

where, here, $\boldsymbol{\theta}$ is the $2^d - 1$-vector of free log-linear parameters in (1.3), $\boldsymbol{\theta} = (\theta_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d^-)$, and the log-partition function is

$$\psi(\boldsymbol{\theta}) = -\log f_{\mathbf{0}_d} = \log \Big[1 + \sum_{\boldsymbol{x} \in \mathcal{X}_d^-} e^{\eta_{\boldsymbol{x}}} \Big], \qquad \eta_{\boldsymbol{x}} = \sum_{\boldsymbol{i} \in \mathcal{X}_d^- \colon \boldsymbol{i} \leq \boldsymbol{x}} \theta_{\boldsymbol{i}}.$$

Note that $e^{\eta_{\boldsymbol{x}}} = f_{\boldsymbol{x}} / f_{\mathbf{0}_d}$ for every $\boldsymbol{x} \in \mathcal{X}_d^-$.

Lemma 1.1.2. *The class $\mathcal{B}_d$ of d-dimensional Bernoulli distributions is a regular exponential family and the representation in (1.5) is minimal.*

Proof. The natural parameter space is

$$\Theta = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{2^d-1} \colon \int_{\mathcal{X}_d} \exp\Bigg(\sum_{\boldsymbol{i} \in \mathcal{X}_d^-} \theta_{\boldsymbol{i}} \boldsymbol{x}^{\boldsymbol{i}} \Bigg) m(\mathrm{d}\boldsymbol{x}) < +\infty \right\}.$$

Since m is the counting measure on the finite set $\mathcal{X}_d$, the integral reduces to a finite sum and is finite for every $\boldsymbol{\theta} \in \mathbb{R}^{2^d-1}$. Hence, $\Theta = \mathbb{R}^{2^d-1}$, which is open.

For minimality, it is sufficient to show that the sufficient statistics $\{t_{\boldsymbol{i}}(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d^-\}$ are affinely independent functions on $\mathcal{X}_d$. We observe that, given some constants $(a_{\boldsymbol{i}}, \boldsymbol{i} \in \mathcal{X}_d) \subseteq \mathbb{R}$, the condition

$$a_{\mathbf{0}_d} + \sum_{\boldsymbol{i} \in \mathcal{X}_d^-} a_{\boldsymbol{i}} \boldsymbol{x}^{\boldsymbol{i}} = 0, \quad \text{for every } \boldsymbol{x} \in \mathcal{X}_d,$$

implies $a_{\boldsymbol{i}} = 0$, for every $\boldsymbol{i} \in \mathcal{X}_d$. Therefore, the representation in (1.5) is minimal. □

The moment parameters are denoted by $\boldsymbol{\mu} = (\mu_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d^-)$, where

$$\mu_{\boldsymbol{i}} = \mathbb{E}[t_{\boldsymbol{i}}(\boldsymbol{X})] = \mathbb{E}[\boldsymbol{X}^{\boldsymbol{i}}] = \mathbb{E}\Bigg[\prod_{j=1}^{d} X_j^{i_j} \Bigg],$$

and can be derived from $\psi(\boldsymbol{\theta})$, by

$$\mu_{\boldsymbol{i}} = \frac{\partial \psi(\boldsymbol{\theta})}{\partial \theta_{\boldsymbol{i}}},$$

for every $\boldsymbol{i} \in \mathcal{X}_d^-$. Moreover, Lemma 1.1.2 and Proposition A.0.1 in Appendix A imply that the map between the log-linear parameters $\boldsymbol{\theta}$ and the linear parameter $\boldsymbol{\mu}$ is one-to-one and differentiable.

In conclusion, this representation proves effective in characterizing independence among components of multidimensional Bernoulli random vectors. The following theorem collects the results of Theorems 3.1 and 3.2 in [31].

Theorem 1.1.3. *Let $\boldsymbol{X}$ be a Bernoulli random vector with pmf $f \in \mathcal{B}_d$. Then,*

a) The random vector $\boldsymbol{X}$ is a vector of independent Bernoulli random variables if and only if, for any $\boldsymbol{i} \in \mathcal{X}_d$ such that $|\boldsymbol{i}| \geq 2$,

$$\theta_{\boldsymbol{i}} = 0,$$

or, equivalently, for any $\boldsymbol{x} \in \mathcal{X}_d$,

$$\eta_{\boldsymbol{x}} = \sum_{j=1}^{d} x_j \theta_{\boldsymbol{e}_j},$$

where $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_d)$ is the canonical basis of $\mathbb{R}^d$.

b) Consider two disjoint subsets $J_1, J_2 \subseteq \{1, \dots, d\}$, and let $\boldsymbol{X}' = (X_j \colon j \in J_1)$ and $\boldsymbol{X}'' = (X_j \colon j \in J_2)$. Then, $\boldsymbol{X}'$ and $\boldsymbol{X}''$ are independent if and only if $\theta_{\boldsymbol{i}} = 0$, for every $\boldsymbol{i} \in \mathcal{X}_d$ such that there exist $j_1 \in J_1$ and $j_2 \in J_2$ satisfying $i_{j_1} = i_{j_2} = 1$.

Theorem 1.1.3 can be formulated with analogous statements in terms of the log-linear parameters $\boldsymbol{\theta}'$ under effect coding. Indeed, since $\boldsymbol{\theta}' = 2^{-d} H_d T_d \boldsymbol{\theta}$ and the structure of $H_d T_d$ is such that each $\theta'_{\boldsymbol{i}}$ depends only on $(\theta_{\boldsymbol{k}} \colon \boldsymbol{k} \geq \boldsymbol{i})$, the independence conditions are preserved under this change of parameters.

1.1.3 Probability generating functions

We introduce an algebraic representation of the class $\mathcal{B}_d$ via the probability generating function (pgf). This representation will be extensively exploited to investigate a notion of negative dependence that is based on the study of the zeros of this function; see Section 2.4. The pgf of a Bernoulli random vector $\boldsymbol{X}$ with pmf $f \in \mathcal{B}_d$ is the polynomial $\mathcal{P} \in \mathbb{R}[z_1, \dots, z_d]$ defined by

$$\mathcal{P}(\boldsymbol{z}) = \mathbb{E}[\boldsymbol{z}^{\boldsymbol{X}}] = \mathbb{E}\left[\prod_{j=1}^{d} z_j^{X_k}\right] = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f_{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}},$$

where $\boldsymbol{z} = (z_1, \dots, z_d)$ is the vector of indeterminates. The pgf of a Bernoulli random vector is a multiaffine polynomial with positive coefficients, i.e., each indeterminate appears with degree at most one. This reflects the fact that each random variable can take only the values 0 or 1. In this context, the degree of a polynomial, defined as the maximum sum of the exponents in any of its terms, corresponds to the maximum possible number of successes within the random vector. Moreover, we say that a polynomial is *homogeneous* of degree n if all its monomials have the same degree n.

For $j \in \{1, \dots, d\}$, we denote by ∂_j the partial derivative taken with respect to z_j,

$$\partial_j \mathcal{P}(\boldsymbol{z}) = \frac{\partial \mathcal{P}(\boldsymbol{z})}{\partial z_j}.$$

All linear parameters $\mu_{\boldsymbol{i}}$, for $\boldsymbol{i} \in \mathcal{X}_d^-$, can be derived from the pgf $\mathcal{P}$ of $\boldsymbol{X}$ by

$$\mu_{\boldsymbol{i}} = \mathbb{E}[\boldsymbol{X}^{\boldsymbol{i}}] = \partial_{\boldsymbol{i}} \mathcal{P}(\boldsymbol{z})\big|_{\boldsymbol{z}=\boldsymbol{1}_d}, \tag{1.6}$$

where $\partial_{\boldsymbol{i}} = \partial_{j_1} \cdots \partial_{j_k}$, and $(j_1, \dots, j_k) = J_{\boldsymbol{i}} \subseteq \{1, \dots, d\}$ are the positions of 1's in $\boldsymbol{i}$. For any $j \in \{1, \dots, d\}$, it is possible to decompose a multiaffine polynomial as the sum

$$\mathcal{P}(\boldsymbol{z}) = \mathcal{P}(\boldsymbol{z})\Big|_{z_j=0} + z_j\, \partial_j \mathcal{P}(\boldsymbol{z}). \tag{1.7}$$

This decomposition has a probabilistic meaning; indeed, if for any $j \in \{1, \dots, d\}$ we denote $p_j = 1 - q_j = \mathbb{E}[X_j]$, the polynomial

$$\frac{1}{q_j} \mathcal{P}(\boldsymbol{z})\big|_{z_j=0} = \frac{1}{1 - \partial_j \mathcal{P}(\boldsymbol{z})\big|_{\boldsymbol{z}=\boldsymbol{1}_d}} \mathcal{P}(\boldsymbol{z})\Big|_{z_j=0} \tag{1.8}$$

is the pgf of the random vector $(\boldsymbol{X}_{-j}|X_j = 0)$, while the polynomial

$$\frac{1}{p_j}\partial_j\mathcal{P}(\boldsymbol{z}) = \frac{1}{\partial_j\mathcal{P}(\boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{1}_d}}\partial_j\mathcal{P}(\boldsymbol{z})\Big|_{z_j=0} \tag{1.9}$$

is the pgf of $(\boldsymbol{X}_{-j}|X_j = 1)$, where $\boldsymbol{X}_{-j} = (X_1,\dots,X_{j-1},X_{j+1},\dots,X_d)$. Therefore, for $j \in \{1,\dots,d\}$, the pgf of the marginal random vector $\boldsymbol{X}_{-j}$ is given by

$$\begin{aligned}
\mathbb{E}\Big[\prod_{h\neq j} z_h^{X_h}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big(\prod_{h\neq j} z_h^{X_h}\Big|X_j\Big)\Big] \\
&= q_j\mathbb{E}\Big(\prod_{h\neq j} z_h^{X_h}\Big|X_j = 0\Big) + p_j\mathbb{E}\Big(\prod_{h\neq j} z_h^{X_h}\Big|X_j = 1\Big) \\
&= q_j\frac{1}{q_j}\mathcal{P}(\boldsymbol{z})\big|_{z_j=0} + p_j\frac{1}{p_j}\partial_j\mathcal{P}(\boldsymbol{z}) \\
&= \mathcal{P}(\boldsymbol{z})\big|_{z_j=0} + \partial_j\mathcal{P}(\boldsymbol{z}) = \mathcal{P}(\boldsymbol{z})\big|_{z_j=1}.
\end{aligned} \tag{1.10}$$

In particular, the pgf of the univariate Bernoulli random variable X_j, for $j \in \{1,\dots,d\}$, is $\mathcal{P}(\boldsymbol{1}_{j-1}, z, \boldsymbol{1}_{d-j})$. Finally, the pgf $\mathcal{P}_k$ of the conditional Bernoulli $(\boldsymbol{X}|S_{\boldsymbol{X}} = k)$, where $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ and $k \in \{0,\dots,d\}$ is given by

$$\mathcal{P}_k(\boldsymbol{z}) = \frac{1}{\sum_{\boldsymbol{x}\in\mathcal{X}_d^k} f_{\boldsymbol{x}}}\sum_{\boldsymbol{x}\in\mathcal{X}_d^k} f_{\boldsymbol{x}}\boldsymbol{z}^{\boldsymbol{x}}, \tag{1.11}$$

where $\mathcal{X}_d^k := \{\boldsymbol{x} \in \mathcal{X}_d \colon x_1 + \cdots + x_d = k\}$

1.2 Fréchet classes

Given a vector $\boldsymbol{p} = (p_1,\dots,p_d) \in (0,1)^d$, the Fréchet class $\mathcal{B}_d(\boldsymbol{p}) \subseteq \mathcal{B}_d$ is the set of all d-dimensional pmfs with univariate marginal means $p_1,\dots,p_d$. Thus, for any Bernoulli random vector $\boldsymbol{X} = (X_1,\dots,X_d)$ with pmf $f \in \mathcal{B}_d(\boldsymbol{p})$, we have $\mathbb{E}[X_j] = p_j$, for all $j \in \{1,\dots,d\}$. The study of pmfs within a given Fréchet class allows us to investigate the underlying dependence structures, since the marginal distributions are the same for all elements of the class. This framework is crucial for the study of dependence orders and dependence properties, which are naturally defined within Fréchet classes, as we recall in Section 2.2.

1.2.1 Parameterizations of Fréchet classes

By Lemma 1.1.2, the class $\mathcal{B}_d$ is a regular exponential family, and a natural question is whether the Fréchet class inherits this structure. Consider the partition of the sufficient statistics $\boldsymbol{t}(\boldsymbol{x}) = (\boldsymbol{x}^{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d^-)$ into $\boldsymbol{t}^{(1)}(\boldsymbol{x}) = (x_j \colon j \in \{1,\dots,d\})$ and $\boldsymbol{t}^{(2)}(\boldsymbol{x}) = (\boldsymbol{x}^{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d,\ |\boldsymbol{i}| \geq 2)$, and the corresponding partition $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$ of the natural parameters. Define the mixed parameter vector

$$\zeta(\boldsymbol{\theta}) = (\boldsymbol{\mu}^{(1)}(\boldsymbol{\theta}),\ \boldsymbol{\theta}^{(2)}) = (\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{t}^{(1)}(\boldsymbol{X})],\ \boldsymbol{\theta}^{(2)}). \tag{1.12}$$

By Proposition A.0.2, the map $\boldsymbol{\theta} \mapsto \zeta(\boldsymbol{\theta})$ is a homeomorphism between $\Theta = \mathbb{R}^{2^d-1}$ onto $(0,1)^d \times \mathbb{R}^{2^d-d-1}$. Under this mixed representation, we find the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ by fixing the linear parameters $\boldsymbol{\mu}^{(1)} = \boldsymbol{p}$. We can write

$$\mathcal{B}_d(\boldsymbol{p}) = \left\{ f(\boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\mu}^{(1)}, \boldsymbol{\theta}^{(2)}) \in \mathcal{B}_d(\boldsymbol{p}) \colon \boldsymbol{\mu}^{(1)} = \boldsymbol{p}, \boldsymbol{\theta}^{(2)} \in \Theta^{(2)} \subseteq \mathbb{R}^{2^d-d-1} \right\}$$

However, fixing $\boldsymbol{\mu}^{(1)} = \boldsymbol{p}$ imposes d constraints on $\boldsymbol{\theta}$ that are nonlinear in the natural parameters. Specifically, the first-order natural parameters $\boldsymbol{\theta}^{(1)}$ are determined by $\boldsymbol{\theta}^{(2)}$ and $\boldsymbol{p}$ through the moment equations

$$\nabla_{\boldsymbol{\theta}^{(1)}} \psi(\boldsymbol{\theta}) = \boldsymbol{p},$$

which are nonlinear in $\boldsymbol{\theta}^{(1)}$. In particular, the natural parameters of any pmf $f \in \mathcal{B}_d(\boldsymbol{p})$ are given by $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}(\boldsymbol{p}, \boldsymbol{\theta}^{(2)}), \boldsymbol{\theta}^{(2)})$, for $\boldsymbol{\theta}^{(2)} \in \Theta^{(2)} \subseteq \mathbb{R}^{2^d-d-1}$. We have

$$f(\boldsymbol{x}) = \exp\Big(\boldsymbol{\theta}^{(1)}(\boldsymbol{p}, \boldsymbol{\theta}^{(2)}) \cdot \boldsymbol{x}^{(1)} + \boldsymbol{\theta}^{(2)} \cdot \boldsymbol{x}^{(2)} - \psi(\boldsymbol{\theta}^{(1)}(\boldsymbol{p}, \boldsymbol{\theta}^{(2)}), \boldsymbol{\theta}^{(2)}) \Big).$$

We notice that the representation via natural parameters is not the most convenient choice for representing Fréchet classes. Therefore, we introduce a parameterization of the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, as proposed by [46]. Let us denote by $\tilde{\mathcal{X}}_d$ the set $\{0,1\}^d$ ordered according to the lexicographical criterion ($\prec_{\text{lex}}$), where, for $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^d$, we have $\boldsymbol{x} \prec_{\text{lex}} \boldsymbol{y}$ if $x_j < y_j$ at the smallest index $j \in \{1, \dots, d\}$ where they differ. Let us define the following $2^d \times 2^d$ matrices, for $\boldsymbol{p} \in (0,1)^d$,

$$U_{\boldsymbol{p}} = U_{p_1} \otimes \cdots \otimes U_{p_d} = \begin{pmatrix} 1 & p_1 \\ 1 & 0 \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 1 & p_d \\ 1 & 0 \end{pmatrix},$$

and $\Lambda_{\boldsymbol{p}} = \operatorname{diag}(\boldsymbol{q}^{\mathbf{1}_d - \boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_d)$, where $\boldsymbol{q} = \mathbf{1}_d - \boldsymbol{p}$.

Proposition 1.2.1 ([46, Proposition 1])**.** *Any pmf $f \in \mathcal{B}_d(\boldsymbol{p})$, for $\boldsymbol{p} \in (0,1)^d$, admits the representation*

$$\boldsymbol{f} = T_d^{-1} \Lambda_{\boldsymbol{p}} U_{\boldsymbol{p}} \boldsymbol{\xi},$$

where $\boldsymbol{\xi} = (\xi_{\boldsymbol{i}} \colon \boldsymbol{i} \in \tilde{\mathcal{X}}_d)$ is the vector of parameters such that $\xi_{\mathbf{0}_d} = 1$ and $\xi_{\boldsymbol{e}_j} = 0$, for $j \in \{1, \dots, d\}$. In particular, for every $\boldsymbol{i} \in \mathcal{X}_d$, the parameter $\xi_{\boldsymbol{i}} = (-1)^{|\boldsymbol{i}|} \mathbb{E}[\boldsymbol{Y}^{\boldsymbol{i}}]$, where $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ with, for $j \in \{1, \dots, d\}$,

$$Y_j = \frac{X_j - p_j}{p_j(1 - p_j)}.$$

The proof of Proposition 1.2.1 is provided in [46], and it is based on a polynomial representation. In particular, any cumulative distribution function corresponding to a pmf in $\mathcal{B}_d(\boldsymbol{p})$ is represented as the polynomial

$$g(\boldsymbol{u}) = \left(\prod_{j=1}^{d} u_j \right) \left[1 + \sum_{\boldsymbol{i} \in \mathcal{X}_d \colon |\boldsymbol{i}| \geq 2} \xi_{\boldsymbol{i}} (\mathbf{1}_d - \boldsymbol{u})^{\boldsymbol{i}} \right],$$

for $\boldsymbol{u} \in \{1-p_1,1\} \times \cdots \times \{1-p_d,1\}$. The polynomial g is reminiscent of the expression of the Farlie–Gumbel–Morgenstern (FGM) copula. The link between Bernoulli distributions and FGM copulas has been studied in [11], and we will recall this link in Chapter 4. This representation is a useful tool for describing the geometrical structure of a Fréchet class, as shown in Theorem 1 in [46]. In the next section, we recall its statement and discuss some of its main consequences.

1.2.2 Geometrical structure

For every $j \in \{1,\ldots,d\}$, let $\boldsymbol{x}_j$ be the 2^d-vector that contains the jth components of the elements of $\mathcal{X}_d$. For example, for $d=3$, we have $\boldsymbol{x}_1 = (0,1,0,1,0,1,0,1)$, $\boldsymbol{x}_2 = (0,0,1,1,0,0,1,1)$, and $\boldsymbol{x}_3 = (0,0,0,0,1,1,1,1)$. Taking a Bernoulli random vector $\boldsymbol{X} = (X_1,\ldots,X_d)$ with pmf in the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, for every $j \in \{1,\ldots,d\}$ we can write

$$\mathbb{E}[X_j] = \boldsymbol{x}_j^\top \boldsymbol{f} = p_j, \quad \text{and} \quad \mathbb{E}[1-X_j] = (\mathbf{1}_{2^d} - \boldsymbol{x}_j)^\top \boldsymbol{f} = q_j,$$

where $q_j = 1 - p_j$. If we consider the odds of the event $X_j = 0$, that is, $c_j = q_j/p_j$, the d conditions on the marginal means can be summarized by the following homogeneous linear system,

$$H_{\boldsymbol{p}} \boldsymbol{f} = \mathbf{0}_d,$$

where $H_{\boldsymbol{p}} \in \mathbb{R}^{d \times 2^d}$ is the matrix with rows are given by $[(\mathbf{1}_{2^d} - \boldsymbol{x}_j)^\top - c_j \boldsymbol{x}_j^\top]$, for $j \in \{1,\ldots,d\}$. By Theorem 1 in [46], the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ is in a one-to-one correspondence with a convex polytope $B_d(\boldsymbol{p}) \subseteq \Delta_{2^d-1}$ of dimension $2^d - 1 - d$,

$$\mathcal{B}_d(\boldsymbol{p}) \leftrightarrow B_d(\boldsymbol{p}) = \{\boldsymbol{f} \in \mathbb{R}_+^{2^d} \colon H_{\boldsymbol{p}} \boldsymbol{f} = \mathbf{0}_d,\ \mathbf{1}^\top \boldsymbol{f} = 1\}. \tag{1.13}$$

Every point in the convex polytope $B_d(\boldsymbol{p})$ can be expressed as a convex combination of a finite number $n_d^{\boldsymbol{p}}$ of extremal points $\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{n_d^{\boldsymbol{p}}} \in B_d(\boldsymbol{p})$, i.e., $\boldsymbol{f} \in B_d(\boldsymbol{p})$ if and only if there exists $\lambda_1,\ldots,\lambda_{n_d^{\boldsymbol{p}}} \geq 0$ summing up to 1 such that

$$\boldsymbol{f} = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k \boldsymbol{r}_k.$$

From the one-to-one correspondence in (1.13), we conclude that the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ is a convex polytope in $\mathcal{B}_d$ with extremal pmfs $r_1,\ldots,r_{n_d^{\boldsymbol{p}}} \in \mathcal{B}_d(\boldsymbol{p})$; therefore, $f \in \mathcal{B}_d(\boldsymbol{p})$ if and only if there exists $\lambda_1,\ldots,\lambda_{n_d^{\boldsymbol{p}}} \geq 0$ summing up to 1 such that

$$f(\boldsymbol{x}) = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k r_k(\boldsymbol{x}), \quad \text{for every } \boldsymbol{x} \in \mathcal{X}_d. \tag{1.14}$$

When $p_1,\ldots,p_d$ are rational numbers, the extremal points $\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{n_d^{\boldsymbol{p}}} \in B_d(\boldsymbol{p})$, and thus the extremal pmfs $r_1,\ldots,r_{n_d^{\boldsymbol{p}}} \in \mathcal{B}_d(\boldsymbol{p})$, can be found using, e.g., the software `4ti2` ([98]).

Example 1.2.1. *Consider $d=3$ and $\boldsymbol{p} = (1/2,\ 1/3,\ 2/3)$. We have*

$$H_{\boldsymbol{p}} = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -2 & -2 & 1 & 1 & -2 & -2 \\ 1 & 1 & 1 & 1 & -1/2 & -1/2 & -1/2 & -1/2 \end{pmatrix}$$

$\boldsymbol{x}$	$\boldsymbol{r}_1$	$\boldsymbol{r}_2$	$\boldsymbol{r}_3$	$\boldsymbol{r}_4$	$\boldsymbol{r}_5$	$\boldsymbol{r}_6$	$\boldsymbol{r}_7$	$\boldsymbol{r}_8$	$\boldsymbol{r}_9$	$\boldsymbol{r}_{10}$	$\boldsymbol{r}_{11}$	$\boldsymbol{r}_{12}$
(0,0,0)	0	0	0	0	0	0	1/6	1/6	1/6	1/3	1/3	1/4
(1,0,0)	0	0	1/6	1/3	1/3	1/4	0	1/6	1/6	0	0	0
(0,1,0)	0	1/3	0	0	0	1/12	1/6	0	0	0	0	0
(1,1,0)	1/3	0	1/6	0	0	0	0	0	0	0	0	1/12
(0,0,1)	1/2	1/6	1/2	1/6	1/3	5/12	0	0	1/3	0	1/6	0
(1,0,1)	1/6	1/2	0	1/6	0	0	1/2	1/3	0	1/3	1/6	5/12
(0,1,1)	0	0	0	1/3	1/6	0	1/6	1/3	0	1/6	0	1/4
(1,1,1)	0	0	1/6	0	1/6	1/4	0	0	1/3	1/6	1/3	0

Table 1.1: Extremal points of $B_3(1/2,\, 1/3,\, 2/3)$.

The extremal points of $B_d(\boldsymbol{p})$ can be found using, for example, the software **4ti2**; *see [98]. These extremal points are reported in Table 1.1.*

Finding all the extremal points of a Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ becomes computationally unfeasible when d grows. For example, the class $\mathcal{B}_6(p_1, \ldots, p_d)$, with $p_1 = \cdots = p_d = 1/2$, has $n_6^{\boldsymbol{p}} = 707{,}264$ extremal pmfs. A necessary condition for a pmf $r \in \mathcal{B}_d(\boldsymbol{p})$ to be an extremal pmf is that the support of r, $\mathrm{supp}(r) = \{\boldsymbol{x} \in \mathcal{X}_d \colon r(\boldsymbol{i}) > 0\}$, has cardinality at most $d+1$, or, equivalently, the corresponding vector $\boldsymbol{r}$ has at most $d+1$ non-zero elements; see Proposition 2.1 in [45]. To overcome this issue in high dimensions, there are different possible approaches. One can, for example, assume exchangeability ([43]) or, when $p_1 = \cdots = p_d = p$, one can use a specific algebraic representation, which we recall in the next section.

1.2.3 Algebraic representation

This section presents an algebraic representation of the Bernoulli Fréchet class $\mathcal{B}_d(p) = \mathcal{B}_d(p, \ldots, p)$, with $p \in (0{,}1)$, as originally introduced in [45]. For the purpose of this representation, we assume $p \in \mathbb{Q}$; since $\mathbb{Q}$ is dense in $\mathbb{R}$, this assumption is not a limitation in applications. Furthermore, we may restrict our analysis to the case $p \leq 1/2$ without loss of generality. Indeed, there exists a one-to-one correspondence between $\mathcal{B}_d(p)$ and $\mathcal{B}_d(1-p)$, induced by the linear mapping $\boldsymbol{X} \in \mathcal{B}_d(p) \mapsto \mathbf{1}_d - \boldsymbol{X} \in \mathcal{B}_d(1-p)$. As shown in Section 1.2.2, the class $\mathcal{B}_d(p)$ is a convex polytope with a finite number n_d^p of extremal points,

$$\mathcal{B}_d(p) \leftrightarrow B_d(p) = \{\boldsymbol{f} \in \mathbb{Q}_+^{2^d} \colon H_p \boldsymbol{f} = \mathbf{0}_d, \mathbf{1}^\top \boldsymbol{f} = 1\}, \tag{1.15}$$

where H_p is the $d \times 2^d$ matrix whose rows are given by $[(\mathbf{1}_{2^d} - \boldsymbol{x}_j)^\top - c\, \boldsymbol{x}_j^\top]$, for $j \in \{1, \ldots, d\}$, with $c = q/p$ and $q = 1-p$. Let $\boldsymbol{z} = (z_1, \ldots, z_{d-1})^\top$ be the row vectors of indeterminates. We define the row vectors $\boldsymbol{m}^+(\boldsymbol{z}) = (m_1^+(\boldsymbol{z}), \ldots, m_{2^{d-1}}^+(\boldsymbol{z}))^\top = (1, z_1)^\top \otimes \cdots \otimes (1, z_{d-1})^\top$ and $\boldsymbol{m}_p^-(\boldsymbol{z}) = (-m_{2^{d-1}}^+(\boldsymbol{z}) + \gamma_p, \ldots, -m_1^+(\boldsymbol{z}) + \gamma_p)^\top$, with $\gamma_p = 2 - 1/p$. We define the map $\mathcal{H}_p$ that associates with each point in $B_d(p)$ a polynomial with real coefficients, by

$$\begin{aligned} \mathcal{H}_p \colon B_d(p) &\to \mathbb{Q}[z_1, \ldots, z_{d-1}], \\ \boldsymbol{f} &\mapsto \mathcal{H}_p(\boldsymbol{f}) = P_{\boldsymbol{f}}(\boldsymbol{z}) = \boldsymbol{m}_p(\boldsymbol{z}) \boldsymbol{f}, \end{aligned} \tag{1.16}$$

where $\mathbb{Q}[z_1, \ldots, z_{d-1}]$ denotes the polynomial ring with rational coefficients, and $\boldsymbol{m}_p(\boldsymbol{z}) = (\boldsymbol{m}^+(\boldsymbol{z})||\boldsymbol{m}_p^-(\boldsymbol{z}))$. The map is linear; indeed $\mathcal{H}_p(\alpha \boldsymbol{f}_1 + \beta \boldsymbol{f}_2) = \alpha\mathcal{H}_p(\boldsymbol{f}_1) + \beta\mathcal{H}_p(\boldsymbol{f}_2)$ holds for every $\boldsymbol{f}_1, \boldsymbol{f}_2 \in B_d(p)$, $\alpha \in \mathbb{Q}$, and $\beta = 1-\alpha$. By rearranging the coefficients, we get

$$P_{\boldsymbol{f}}(\boldsymbol{z}) = \boldsymbol{m}_p(\boldsymbol{z})\boldsymbol{f} = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}},$$

where $a_{\boldsymbol{i}} \in \mathbb{Q}$, for $\boldsymbol{i} \in \mathcal{X}_{d-1}$, are linear combinations of the elements of $\boldsymbol{f}$. In particular, we consider $\boldsymbol{a} = (a_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathcal{X}_{d-1})$ and we get

$$\boldsymbol{a} = Q_p \boldsymbol{f}, \tag{1.17}$$

with Q_p is the matrix of order $2^{d-1} \times 2^d$ given by

$$Q_p = (I_{2^{d-1}}||\tilde{I}_{2^{d-1}} + \tilde{A}_p), \tag{1.18}$$

where $I_{2^{d-1}}$ is the identity matrix of order 2^{d-1}, $\tilde{I}_{2^{d-1}}$ is the square matrix of order 2^{d-1} with -1 on the anti-diagonal and 0 elsewhere, and $\tilde{A}_p$ is the square matrix of order 2^{d-1} with the components of the first row equal to $\gamma_p = 2 - 1/p$ and 0 elsewhere. By the correspondence $\mathcal{B}_d(p) \leftrightarrow B_d(p)$, we extend $\mathcal{H}_p$ to a map on $\mathcal{B}_d(p)$, by setting $\mathcal{H}_p(f) \coloneqq \mathcal{H}_p(\boldsymbol{f})$, for every $f \in \mathcal{B}_d(p)$.

Example 1.2.2. *We consider the case $d = 3$ and $p = 2/5$. Then, $\gamma_{2/5} = -1/2$ and the matrix associated with the convex polytope $P_3(2/5)$ is given by*

$$H_{2/5} = \begin{pmatrix} 1 & -3/2 & 1 & -3/2 & 1 & -3/2 & 1 & -3/2 \\ 1 & 1 & -3/2 & -3/2 & 1 & 1 & -3/2 & -3/2 \\ 1 & 1 & 1 & 1 & -3/2 & -3/2 & -3/2 & -3/2 \end{pmatrix}.$$

Moreover, the coefficients of the polynomial $\mathcal{H}_{2/5}(f)$ are given by $\boldsymbol{a} = Q_{2/5}\boldsymbol{f}$, i.e.,

$$\begin{pmatrix} a_{00} \\ a_{10} \\ a_{01} \\ a_{11} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{3}{2} \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f(0,0,0) \\ f(1,0,0) \\ f(0,1,0) \\ f(1,1,0) \\ f(0,0,1) \\ f(1,0,1) \\ f(0,1,1) \\ f(1,1,1) \end{pmatrix}.$$

Table 1.2 shows that if we evaluate the vector $\mathbf{m}(z_1, z_2)$ in the points $\mathbf{c}_1 = (-3/2,1)$, $\mathbf{c}_2 = (1, -3/2)$, and $\mathbf{1}_2 = (1,1)$, we obtain the first, the second, and the third row of H_p, respectively. This intuition is formalized in Theorem 1.2.2.

For the purposes of this algebraic representation, it is convenient to introduce the following notion of equivalence for polynomials with real coefficients. The motivation for this definition will be clarified by Proposition 1.2.4.

Definition 1.2.1. *Two polynomials $P_1(\boldsymbol{z})$ and $P_2(\boldsymbol{z})$ belonging to $\mathbb{Q}[z_1, \ldots, z_{d-1}]$ are equivalent (denoted by $P_1(\boldsymbol{z}) \simeq P_2(\boldsymbol{z})$) if there exists a constant $\alpha \in (0, +\infty)$ such that $P_1(\boldsymbol{z}) = \alpha P_2(\boldsymbol{z})$.*

Table 1.2: *Each row of $H_{2/5}$ is equal to the vector $\boldsymbol{m}(z_1, z_2)$ evaluated at* $(-3/2,1)$, $(1, -3/2)$, *and* $(1,1)$.

$\boldsymbol{m}(z_1,z_2)$	1	z_1	z_2	z_1z_2	$-z_1z_2-\frac{1}{2}$	$-z_2-\frac{1}{2}$	$-z_1-\frac{1}{2}$	$-\frac{3}{2}$
$\boldsymbol{m}(-3/2,1)$	1	-3/2	1	-3/2	1	-3/2	1	-3/2
$\boldsymbol{m}(1,-3/2)$	1	1	-3/2	-3/2	1	1	-3/2	-3/2
$\boldsymbol{m}(1,1)$	1	1	1	1	-3/2	-3/2	-3/2	-3/2

We denote by $[P_1(\boldsymbol{z})]$ the equivalent class of all polynomials equivalent to $P_1(\boldsymbol{z})$, namely,

$$[P_1(\boldsymbol{z})] = \{P_2(\boldsymbol{z}) \in \mathbb{Q}[\boldsymbol{z}] \colon P_1(\boldsymbol{z}) \simeq P_2(\boldsymbol{z})\}.$$

Furthermore, we define $\mathcal{C}_{\mathcal{H}_p}$ as the closure of the image $\mathcal{H}_p(\mathcal{B}_d(p))$ under the equivalence relation $\simeq$, viz.

$$\mathcal{C}_{\mathcal{H}_p} = \bigcup_{f \in \mathcal{B}_d(p)} [\mathcal{H}_p(f)].$$

We report here Theorem 3.1 in [45] that proves that if $P(\boldsymbol{z}) \in \mathcal{C}_{\mathcal{H}_p}$ then $P(\mathbf{1}_{d-1}) = P(\boldsymbol{c}_j) = 0$, where $\boldsymbol{c}_j$ is the $(d-1)$-dimensional vector whose jth element is equal to $-c = -q/p$ and 1 elsewhere, for every $j \in \{1, \dots, d-1\}$.

Theorem 1.2.2 ([45, Theorem 3.1])**.** *We have $\mathcal{C}_{\mathcal{H}_p} \subseteq \mathcal{I}_p$, where $\mathcal{I}_p$ is the ideal of polynomials that vanish at points $\{\mathbf{1}_{d-1}, \boldsymbol{c}_1, \dots, \boldsymbol{c}_{d-1}\}$.*

The authors of [45] notice that the map $\mathcal{H}_p$ is not injective, i.e., there exist different pmfs in the same Fréchet class $\mathcal{B}_d(p)$ whose images through $\mathcal{H}_p$ coincide. Since the mapping $\mathcal{H}_p$ is linear, it is relevant to characterize its kernel. Thus, we first consider the set of pmfs $\boldsymbol{f} \in B_d(p)$ such that $\mathcal{H}_p(\boldsymbol{f})$ is the null polynomial. Specifically, we define the kernel $\ker(\mathcal{H}_p)$ as

$$\ker(\mathcal{H}_p) = \{\boldsymbol{f} \in B_d(p) \colon \mathcal{H}_p(\boldsymbol{f}) \equiv 0\}.$$

Proposition 1.2.3. *The kernel $\ker(\mathcal{H}_p)$ of the map $\mathcal{H}_p$ is a convex polytope whose extremal points are the elements of the set*

$$\begin{aligned}\mathcal{K}_p = \{&(1-p,0,\dots,0,p), (1-2p,p,0,\dots,0,p,0),\\ &(1-2p,0,p,0,\dots,0,p,0,0), \dots (1-2p,0,\dots,0,p,p,0,\dots,0)\}\\ = \{&\boldsymbol{k}_h \coloneqq (1-2p)\boldsymbol{e}_1 + p\boldsymbol{e}_h + p\boldsymbol{e}_{2^d+1-h},\ h \in \{1,\dots,2^{d-1}\}\}\end{aligned} \tag{1.19}$$

where $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{2^d}\}$ is the canonical basis of $\mathbb{R}^{2^d}$. For $h \in \{1,\dots,2^{d-1}\}$, the vector $\boldsymbol{k}_h$ is an extremal points of $B_d(p)$.

Proof. We have

$$\ker(\mathcal{H}_p) = \{\boldsymbol{f} \in \mathbb{Q}_+^{2^d} \colon \tilde{H}_p\boldsymbol{f} = \mathbf{0}_{d+2^{d-1}}\},$$

where $\tilde{H}_p = (H_p//Q_p) \in \mathbb{Q}^{(d+2^{d-1})\times 2^d}$ is the column concatenation of the matrices H_p in (1.15) and Q_p in (1.18). We now show that, for any $h \in \{1,\dots,2^{d-1}\}$, the vector $\boldsymbol{k}_h$ is an extremal point of $B_d(p)$. Let us consider $h \in \{2,\dots,2^{d-1}\}$ and $\boldsymbol{k}_h = (k_{h,1}, \dots, k_{h,2^d})$.

Algorithm 1 Algorithm to find the type-0 pmf

Require: A polynomial $P(\boldsymbol{z}) \in \mathcal{I}_p$ of the form $P(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} \not\equiv 0$
Ensure: The type-0 pmf $f^P \in \mathcal{B}_d(p)$ corresponding to $P(\boldsymbol{z})$

```
for i ∈ X⁻_{d−1} do
    if a_i ≥ 0 then
        f(s_i) = a_i and f(1_d − s_i) = 0
    else
        f(s_i) = 0 and f(1_d − s_i) = −a_i
    end if
end for
Set c_0 := a_{0_{d−1}} + Σ_{i∈X⁻_{d−1} : a_i<0} γ_p a_i
if c_0 ≥ 0 then
    f(0_{d−1}) = c_0 and f(1_d) = 0
else
    f(0_{d−1}) = 0 and f(1_d) = −(c_0·p)/(1−p)
end if
for x ∈ X_d do                                   ▷ Normalization
    f^P(x) = f(x)/(Σ_{y∈X_d} f(y))
end for
```

By construction, $k_{h,j} > 0$ only if $j \in \{1, h, 2^d - h\}$. Consider the matrix $(H_p // I_h^*)$, where $I_h^* \in \mathbb{Q}^{(2^d-3)\times 2^d}$ is obtained from the identity matrix I_{2^d} by removing the rows in positions $\{1, h, 2^d - h\}$. It can be shown that $\operatorname{rank}(H_p // I_h^*) = 2^d - 1$; therefore, by Lemma 2.3 in [99], $\boldsymbol{k}_h$ is an extremal point of $B_d(p)$. An analogous argument holds for $\boldsymbol{k}_1$. Finally, by Proposition 3.1 in [45], the set $\mathcal{K}_p$ is a basis of $\ker(\mathcal{H}_p)$. Since $\ker(\mathcal{H}_p) \subseteq B_d(p)$, this implies that the vectors $\boldsymbol{k}_h$ are the extremal points of $\ker(\mathcal{H}_p)$. □

Having characterized the kernel, we now turn to the general case and focus on the set of pmfs in $\mathcal{B}_d(p)$ whose images under $\mathcal{H}_p$ coincide with a given non-zero polynomial $P \in \mathcal{I}_p$. Define the set $\mathcal{H}_p^{-1}[P(\boldsymbol{z})]$ of all pmfs $\boldsymbol{f} \in B_d(p)$ such that $\mathcal{H}_p(\boldsymbol{f}) \in [P(\boldsymbol{z})]$. This set is non-empty if we consider a polynomial $P(\boldsymbol{z}) \in \mathcal{I}_p$ of the form

$$P(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}, \tag{1.20}$$

with $a_{\boldsymbol{i}} \in \mathbb{R}$ for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$. Indeed, for such polynomials, Algorithm 1 guarantees the existence of a specific pmf $f^P \in \mathcal{B}_d(p)$, with $\mathcal{H}_p(f^P) = P(\boldsymbol{z})$. Following the terminology in [45], where this algorithm was first proposed, the resulting pmf is referred to as the *type-0* pmf associated with $P(\boldsymbol{z})$. Recall that $\boldsymbol{s}_{\boldsymbol{i}} = (\boldsymbol{i}//0) \in \mathcal{X}_d$, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$. The following proposition justifies the notion of equivalence for polynomials introduced in Definition 1.2.1.

Proposition 1.2.4 (Proposition 3 in [78])**.** *Two equivalent polynomials generate the same type-0 pmf.*

Proof. Let $P_1(\boldsymbol{z})$ and $P_2(\boldsymbol{z})$ be two equivalent polynomials and let f^{P_1} and f^{P_2} be the two corresponding type-0 pmfs. By Definition 1.2.1, there exists a constant $\alpha \in (0, +\infty)$ such that $P_1(\boldsymbol{z}) = \alpha P_2(\boldsymbol{z})$. We apply Algorithm 1 to both $P_1(\boldsymbol{z})$ and $P_2(\boldsymbol{z})$. The coefficients of the two polynomials have the same sign; therefore, after the first for-loop of the algorithm, we have $f^{P_1}(\boldsymbol{x}) = \alpha f^{P_2}(\boldsymbol{x})$, for every $\boldsymbol{x} \in \mathcal{X}_d \setminus \{\boldsymbol{0}_d, \boldsymbol{1}_d\}$. Moreover, also $c_0^{P_1} = \alpha c_0^{P_2}$, thus $f^{P_1}(\boldsymbol{x}) = \alpha f^{P_2}(\boldsymbol{x})$, for $\boldsymbol{x} \in \{\boldsymbol{0}_d, \boldsymbol{1}_d\}$. After normalization, the two type-0 pmfs are identical. □

Algorithm 1 also ensures that all polynomials $P(\boldsymbol{z}) \in \mathcal{I}_p$ with the structure in (1.20) belong to $\mathcal{C}_{\mathcal{H}_p}$. In particular, for such polynomials, the set $\mathcal{H}_p^{-1}[P(\boldsymbol{z})]$ is a convex polytope and Proposition 3.1 in [45] shows that

$$\mathcal{H}^{-1}[P(\boldsymbol{z})] = \{\boldsymbol{f} \in B_d(p) \colon \exists \boldsymbol{f}^K \in \ker(Q_p), \lambda \in (0,1] | \boldsymbol{f} = \lambda \boldsymbol{f}^P + (1-\lambda)\boldsymbol{f}^K\}, \tag{1.21}$$

where $\boldsymbol{f}^P$ is the type-0 pmf corresponding to $P(\boldsymbol{z})$, and $\ker(Q_p)$ is the kernel corresponding to the matrix Q_p, that is,

$$\ker(Q_p) = \{\boldsymbol{f} \in \mathbb{Q}^{2^d} \colon Q_p \boldsymbol{f}^K = \boldsymbol{0}_{2^{d-1}}\}. \tag{1.22}$$

Notice that $\ker(Q_p) \supseteq \ker(\mathcal{H}_p)$ and that the set $\mathcal{K}_p$ in (1.19) is a basis of $\ker(Q_p)$. The vector $\boldsymbol{f}^K \in \ker(Q_p)$ is a linear, not necessarily convex, combination of the elements of $\mathcal{K}_p$, and can have negative entries. Therefore, the extremal points of $\mathcal{H}_p^{-1}$ are not always $\boldsymbol{f}^P$ and the elements of $\mathcal{K}_p$. This is true for the case $p = 1/2$, as we discuss in detail in Section 3.1.

Example 1.2.3. *Consider the class $\mathcal{B}_3(2/5)$ of 3-dimensional Bernoulli distributions with common marginal mean $p = 2/5$. From Example 1 in [45], the point*

$$\boldsymbol{r} = (0, 1/5, 1/5, 1/5, 2/5, 0, 0, 0)$$

is an extremal point of $B_3(2/5)$ and it is not the type-0 pmf associated to its polynomial. Indeed, we have $\mathcal{H}(\boldsymbol{r}) = Q(\boldsymbol{z}) = -1/5 + z_1/5 + \ z_2/5 - z_1z_2/5$ and, by Algorithm 1, the type-0 pmf associated to $Q(\boldsymbol{z})$ is $\boldsymbol{f}^Q = (0, 3/10, 3/10, 0, 3/10, 0, 0, 1/10)$. Thus, $\boldsymbol{r}$ is an extremal point, but it is not a type-0 pmf. It is straightforward to see that we can write

$$\boldsymbol{r} = \frac{2}{3}\boldsymbol{f}^Q + \frac{1}{3}\boldsymbol{f}^K,$$

where $2\boldsymbol{f}^Q/3$ is the non-normalized type-0 pmf associated with $Q(\boldsymbol{z})$ and $\boldsymbol{f}^K/3$ is an element of $\ker(Q_p)$ in (1.22). However, it can be verified that $\boldsymbol{f}^K$ has negative components; therefore, $\boldsymbol{f}^K \notin B_3(2/5)$ and it does not identify a pmf in $\mathcal{B}_3(2/5)$.

Building on algebraic tools such as Gröbner bases, the authors of [45] provide a precise characterization of the set $\mathcal{C}_{\mathcal{H}_p}$, for every $p \in (0,1)$. First, let us introduce the fundamental polynomials $F_{\boldsymbol{i}}(\boldsymbol{z})$, for $\boldsymbol{i} \in \mathcal{X}_{d-1}^-$, defined by

$$\begin{aligned} F_{\boldsymbol{i}}(\boldsymbol{z}) &= \boldsymbol{z}^{\boldsymbol{i}} - \boldsymbol{z} \cdot \boldsymbol{i} + (|\boldsymbol{i}| - 1) \\ &= \prod_{j \in J_{\boldsymbol{i}}} z_j - \sum_{j \in J_{\boldsymbol{i}}} z_j + (|\boldsymbol{i}| - 1) \\ &= z_{j_1} \cdots z_{j_k} - (z_{j_1} + \cdots + z_{j_k}) + (k-1), \end{aligned} \tag{1.23}$$

where $(j_1, \ldots, j_k) = J_{\boldsymbol{i}} \subseteq \{1, \ldots, d-1\}$ are the positions of ones in $\boldsymbol{i}$, and $k = |\boldsymbol{i}|$. Note that $F_{\boldsymbol{i}}(\boldsymbol{z}) \equiv 0$, for all $\boldsymbol{i} \in \mathcal{X}_d$ such that $|\boldsymbol{i}| = 1$. To give an example, when $d = 5$, the fundamental polynomial corresponding to $\boldsymbol{i} = (1,0,1,0) \in \mathcal{X}_{d-1}$ is $F_{\boldsymbol{i}}(z_1, \ldots, z_4) = z_1 z_3 - z_1 - z_3 + 1$. The set $\mathcal{C}_{\mathcal{H}_p}$ is generated by the fundamental polynomials.

Theorem 1.2.5 ([45, Corollary 3.1])**.** *For every* $p \in (0,1)$*,*

$$\mathcal{C}_{\mathcal{H}_p} = \left\{ P(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}^-} \gamma_{\boldsymbol{i}} F_{\boldsymbol{i}}(\boldsymbol{z}) \colon \gamma_{\boldsymbol{i}} \in \mathbb{Q}, \text{ for every } \boldsymbol{i} \in \mathcal{X}_d^- \right\}.$$

An important consequence of Theorem 1.2.5 is that the images of all Fréchet classes $\mathcal{B}_d(p)$, for $p \in (0,1)$, coincide. Therefore, for every linear combination of the fundamental polynomials, it is possible to find a type-0 pmf in each class $\mathcal{B}_d(p)$ using Algorithm 1. Moreover, by considering the kernel $\ker(Q_p)$, we can, at least in principle, find all pmfs in $\mathcal{B}_d(p)$ whose image through $\mathcal{H}_p$ is that linear combination.

Finally, fundamental polynomials are also relevant because the type-0 pmfs corresponding to them are extremal pmfs.

Proposition 1.2.6. *Let* $F_{\boldsymbol{i}}(\boldsymbol{z})$*, for* $\boldsymbol{i} \in \mathcal{X}_{d-1}$ *with* $|\boldsymbol{i}| \geq 2$*. For every* $p \in (0,1)$*, the type-0 pmf* $f^{F_{\boldsymbol{i}}} \in \mathcal{B}_d(p)$ *corresponding to* $F_{\boldsymbol{i}}(\boldsymbol{z})$*, and the type-0 pmf* $f^{-F_{\boldsymbol{i}}} \in \mathcal{B}_d(p)$ *corresponding to* $-F_{\boldsymbol{i}}(\boldsymbol{z})$ *are extremal pmfs of* $\mathcal{B}_d(p)$*.*

1.3 Bernoulli and discrete distributions

The connection between exchangeable Bernoulli distributions and the distribution of the sum of Bernoulli random variables is particularly deep. In this section, we investigate the consequences of this linkage on the geometrical structures of these classes, while its implication for negative dependence will be further explored in Chapter 2. We now introduce the classes of distributions that play a central role in this section. First, let $\mathcal{D}_d$ be the set of univariate discrete pmfs with support $\{0, \ldots, d\}$, and $\mathcal{D}_d(\mu)$ the subset of pmfs in $\mathcal{D}_d$ with mean $\mu \in (0, d)$. We observe that the distribution of the sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ belongs to $\mathcal{D}_d(\mu)$ whenever $\boldsymbol{X} = (X_1, \ldots, X_d)$ is a d-dimensional Bernoulli random vector with marginal means $\boldsymbol{p} = (p_1, \ldots, p_d)$ with $|\boldsymbol{p}| = p_1 + \cdots + p_d = \mu$. Second, we recall the concept of exchangeability. We say that a Bernoulli random vector $\boldsymbol{X}$ and its pmf $f \in \mathcal{B}_d$ are exchangeable if f is invariant under any permutation of its components, i.e., $f(x_1, \ldots, x_d) = f(x_{\sigma(1)}, \ldots, x_{\sigma(d)})$ for all $(x_1, \ldots, x_d) \in \mathcal{X}_d$ and any permutation σ of $\{1, \ldots, d\}$. We denote by $\mathcal{E}_d \subseteq \mathcal{B}_d$ the class of all exchangeable Bernoulli pmfs, and by $\mathcal{E}_d(p) \subseteq \mathcal{B}_d(p)$ the family of all exchangeable Bernoulli pmfs with univariate marginals equal to p.

The connections between the class of d-dimensional Bernoulli distributions and the family $\mathcal{D}_d(\mu)$ can be summarized as follows:

$$\mathcal{E}_d(p) \leftrightarrow \mathcal{S}_d(\mu) \equiv \mathcal{D}_d(\mu), \quad \text{with } \mu = pd, \tag{1.24}$$

where $\mathcal{S}_d(\mu)$ denotes the class of univariate pmfs f^S such that f^S is the distribution of $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ for some $\boldsymbol{X} \in \mathcal{B}_d$, and $\mathbb{E}[S_{\boldsymbol{X}}] = \mu$. Obviously, if $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$, the condition $\mathbb{E}[S_{\boldsymbol{X}}] = \mu$ implies $|\boldsymbol{p}| = \mu$. The relations in (1.24) highlight two fundamental properties of these families. First, the distribution of the sum of any Bernoulli random vector $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$ coincides with the distribution of an exchangeable Bernoulli random vector $\boldsymbol{X}^e \in \mathcal{E}_d(|\boldsymbol{p}|/d)$. Second, the equivalence $\mathcal{S}_d(\mu) \equiv \mathcal{D}_d(\mu)$ confirms that any discrete distribution with support $\{0, \ldots, d\}$ and mean μ can be realized as the sum of at least one Bernoulli random vector. This vector can always be found in the exchangeable class $\mathcal{E}_d(\mu/d) \subseteq \mathcal{B}_d(\mu/d)$, but it is generally not unique. Moreover, existence is not guaranteed within a general Fréchet class: for an arbitrary vector of marginal means $\boldsymbol{p}$, there may not exist a pmf in $\mathcal{B}_d(\boldsymbol{p})$ whose sum coincides with $f^D \in \mathcal{D}_d(\mu)$, even if $|\boldsymbol{p}| = \mu$. Further details can be found in [43]. It should be noted that while their definition of $\mathcal{S}_d(\cdot)$ is restricted to sums arising from exchangeable Bernoulli distributions, the overall results remain consistent with this broader definition.

This section is structured as follows. We begin by characterizing the geometry of the class $\mathcal{D}_d(\mu)$ of univariate distributions with mean μ. We show that this class is a convex polytope with analytical extremal points. Building on the one-to-one correspondence between discrete distributions and exchangeable Bernoulli distributions, we then characterize the geometrical structure of $\mathcal{E}_d(p)$, for $p \in (0,1)$. Finally, we investigate the class of all the pmfs of Bernoulli distributed random vectors, whose sums of the components coincide with a given discrete distribution on $\mathcal{D}_d$.

1.3.1 Discrete distribution

In parallel with the Bernoulli case, any pmf $f^D \in \mathcal{D}_d$ can be represented as a point in $\mathbb{R}^{d+1}$ via the vector

$$\boldsymbol{f}^D = (f_y^D \colon y \in \{0, \ldots, d\}),$$

where $f_y^D := f^D(y)$. The class $\mathcal{D}_d$ is an exponential family with sufficient statistics given by the indicator functions $\mathbb{1}_{\{1\}}(y), \ldots, \mathbb{1}_{\{d\}}(y)$, where, for every $k \in \{1, \ldots, d\}$, $\mathbb{1}_{\{k\}}(y) = 1$ if $y \in \{k\}$, and $\mathbb{1}_{\{k\}}(y) = 0$ otherwise. Indeed, any pmf $f^D \in \mathcal{D}_d$ with $f_0^D > 0$, can be expressed as

$$f^D(y) = \exp\left\{ \sum_{k=1}^{d} \left[\ln\left(\frac{f_k^D}{f_0^D}\right) \mathbb{1}_{\{k\}}(y) \right] + \ln(f_0^D) \right\}.$$

Consider a random variable D with pmf $f^D \in \mathcal{D}_d(\mu) \subseteq \mathcal{D}_d$, with $\mu \in (0, d)$. The condition $\mathbb{E}[D] = \mu$ is equivalent to the linear condition

$$\sum_{y=0}^{d} (y - \mu) f^D(y) = \sum_{y=0}^{d} (y - \mu) f_y^D = 0.$$

This implies that the family $\mathcal{D}_d(\mu)$ is in a one-to-one correspondence with a convex polytope in $\mathbb{R}^{d+1}$, namely,

$$\mathcal{D}_d(\mu) \leftrightarrow \left\{ \boldsymbol{f}^D \in \mathbb{R}_+^{d+1} \colon \mathbf{1}_{d+1}^\top \boldsymbol{f}^D = 1, \sum_{y=0}^{d} (y - \mu) f_y^D = 0 \right\}. \tag{1.25}$$

The extremal pmfs of $\mathcal{D}_d(\mu)$ are analytical, as stated by the following proposition.

Proposition 1.3.1 ([43, Proposition 4.5]). *The class $\mathcal{D}_d(\mu)$ in (1.25) is a convex polytope with extremal pmfs $r^D_{j_1,j_2}$ given by*

$$r^D_{j_1,j_2}(y) = \begin{cases} \frac{j_2-\mu}{j_2-j_1}, & \text{for } y = j_1, \\ \frac{\mu-j_1}{j_2-j_1}, & \text{for } y = j_2, \\ 0, & \text{elsewhere}, \end{cases} \tag{1.26}$$

for every $j_1 \in \{0, \dots, \lfloor\mu\rfloor\}$ and $j_2 \in \{\lfloor\mu\rfloor + 1, \dots, d\}$, where $\lfloor\mu\rfloor \in \{0, \dots, d\}$ is the largest integer lower than or equal to μ.

It follows that the number of different extremal pmfs of $\mathcal{D}_d(\mu)$ is given by $(\lfloor\mu\rfloor + 1)(d - \lfloor\mu\rfloor)$ if μ is not integer, and by $\mu(d-\mu)+1$ if μ is integer.

We conclude by defining the pgf of a discrete random variable D with pmf $f_D \in \mathcal{D}_d$ as

$$\mathcal{P}_{f_D}(z) = \mathbb{E}[z^D] = \sum_{k=0}^{d} f_D(k) z^k.$$

Given a Bernoulli random vector $\boldsymbol{X}$ with pgf $\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z})$, the pgf of the sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ can be derived by evaluating the multivariate pgf along the diagonal, namely,

$$\mathcal{P}_{S_{\boldsymbol{X}}}(z) = \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z})\Big|_{z_1=\cdots=z_d=z}$$

The algebraic properties of $\mathcal{P}_{\boldsymbol{X}}$ translate directly into distributional characteristics of $S_{\boldsymbol{X}}$. Specifically, if $\mathcal{P}_{\boldsymbol{X}}$ is a polynomial of degree $n \le d$, then $\mathcal{P}_{S_{\boldsymbol{X}}}$ is a polynomial in z of degree n, confirming that the maximum number of 1's that $\boldsymbol{X}$ can take is n. Furthermore, if $\mathcal{P}_{\boldsymbol{X}}$ is homogeneous of degree n, then $S_{\boldsymbol{X}}$ takes value n almost surely.

1.3.2 Exchangeable Bernoulli distributions

The authors of [43, Section 3.2] show that the class $\mathcal{E}_d(p)$ of exchangeable Bernoulli distributions with univariate means p is in a one-to-one correspondence with the class $\mathcal{D}_d(\mu)$, for $\mu = pd$. The geometrical structure of $\mathcal{D}_d(\mu)$ has been described in Section 1.3.1; thus, we can easily characterize the geometrical structure of $\mathcal{E}_d(p)$.

Proposition 1.3.2. *The class $\mathcal{E}_d(p) \subseteq \mathcal{B}_d(p)$ in (1.25) is a convex polytope with extremal pmfs $r^e_{j_1,j_2}$ given by*

$$r^e_{j_1,j_2}(\boldsymbol{i}) = \begin{cases} \binom{d}{j_1}^{-1} \frac{j_2-pd}{j_2-j_1}, & \text{if } |\boldsymbol{i}| = j_1, \\ \binom{d}{j_2}^{-1} \frac{pd-j_1}{j_2-j_1}, & \text{if } |\boldsymbol{i}| = j_2, \\ 0, & \text{otherwise}, \end{cases} \tag{1.27}$$

for every $j_1 \in \{0, \dots, \lfloor pd\rfloor\}$ and $j_2 \in \{\lfloor pd\rfloor + 1, \dots, d\}$.

Every exchangeable Bernoulli pmf $f^e \in \mathcal{E}_d$ can be written in the form of an exponential family, namely,

$$f^e(\boldsymbol{x}) = \exp\left\{ \sum_{k=1}^{d} \left[\ln\left(\frac{f^D_k}{f^D_0}\right) \mathbb{1}_{\{k\}}\left(\sum_{j=1}^{d} x_j\right)\right] + \ln(f^D_0) \right\},$$

where $f^D \in \mathcal{D}_d$ is the discrete pmf corresponding to f^e via the bijection $\mathcal{E}_d \leftrightarrow \mathcal{D}_d$. The assumption of exchangeability reduces the complexity of the model, decreasing the number of free parameters from 2^d-1 in the general class $\mathcal{B}_d$ to only d. In this setting, the sufficient statistics $t_k(\boldsymbol{x}) = \mathbb{1}_{\{k\}}(x_1 + \cdots + x_d)$ depends exclusively on the sum of the components. This is consistent with the fact that, for exchangeable random vectors, the sum contains all the information regarding the joint distribution.

We finally consider the pgf of an exchangeable Bernoulli random vector. The condition of exchangeability translates directly into a specific algebraic structure of the pgf. A Bernoulli random vector $\boldsymbol{X}^e$ with pmf f^e is exchangeable if and only if its pgf $\mathcal{P}^e(\boldsymbol{z})$ is a symmetric polynomial in the variables $z_1, \dots, z_d$, that is, if

$$\mathcal{P}^e(z_1, \dots, z_d) = \mathcal{P}^e(z_{\sigma(1)}, \dots, z_{\sigma(d)}),$$

for every permutation σ of the set $\{1, \dots, d\}$. The fundamental theorem of symmetric polynomials states that every symmetric polynomial has a unique representation in terms of the elementary symmetric polynomials $\mathcal{E}_{d,k}(\boldsymbol{z}) \in \mathbb{R}[z_1, \dots, z_d]$, which are given by

$$\mathcal{E}_{d,k}(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathcal{X}_d^k} \boldsymbol{z}^{\boldsymbol{i}}, \tag{1.28}$$

for every $k \in \{0, \dots, d\}$. To visualize the elementary symmetric polynomials, for $d = 3$ we have

$$\begin{aligned}
\mathcal{E}_{3,1}(z_1, z_2, z_3) &= z_1 + z_2 + z_3, \\
\mathcal{E}_{3,2}(z_1, z_2, z_3) &= z_1 z_2 + z_1 z_3 + z_2 z_3, \\
\mathcal{E}_{3,3}(z_1, z_2, z_3) &= z_1 z_2 z_3.
\end{aligned}$$

Every symmetric polynomial $\mathcal{P}^e(\boldsymbol{z})$ can be written as

$$\mathcal{P}^e(\boldsymbol{z}) = \mathcal{Q}(\mathcal{E}_{d,0}(\boldsymbol{z}), \mathcal{E}_{d,1}(\boldsymbol{z}), \dots, \mathcal{E}_{d,d}(\boldsymbol{z})),$$

for some $\mathcal{Q}(z_1, \dots, z_d) \in \mathbb{R}[z_1, \dots, z_d]$. Notice that, when $\mathcal{P}^e$ is symmetric and multiaffine, then it is a linear combination of the elementary symmetric polynomials. Thus, when $\mathcal{P}^e(\boldsymbol{z})$ is a symmetric pgf, we have

$$\mathcal{P}^e(\boldsymbol{z}) = \sum_{k=0}^{k} \alpha_k \mathcal{E}_{d,k}(\boldsymbol{z}),$$

where, for $k \in \{0, \dots, d\}$, $\alpha_k = f^e(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathcal{X}_d^k$. This algebraic representation is consistent with the reduction of complexity previously discussed. Just as the exchangeability assumption reduces the number of free parameters of the pmf to d, it similarly constrains the pgf to be a linear combination of the d elementary symmetric polynomials (the constant term $\alpha_0 \mathcal{E}_{d,0}(\boldsymbol{z}) = \alpha_0 = f^e(\boldsymbol{0}_d)$ is fixed by the normalization $\mathcal{P}^e(\boldsymbol{1}_d) = 1$).

1.3.3 Polytope of Bernoulli behind a discrete distribution

In this section we characterize the family of all d-dimensional Bernoulli distributions whose sum of the components follows a given univariate distribution $g \in \mathcal{D}_d$. In principle, one can consider any surjective map $\phi\colon \mathcal{X}_d \to \{0,\dots,d\}$ and characterize the class of pmfs of Bernoulli random vectors $\boldsymbol{X}$ such that $\phi(\boldsymbol{X})$ has pmf g. However, we focus exclusively on the mapping $\phi(\boldsymbol{x}) = s(\boldsymbol{x}) = x_1 + \dots + x_d$ due to its relevance in applications, such as aggregate risk modeling. If a Bernoulli random vector $\boldsymbol{X}$ has pmf $f \in \mathcal{B}_d$, we denote by $s(f)$ the pmf of $s(\boldsymbol{X}) = S_{\boldsymbol{X}} = X_1 + \dots + X_d$; clearly, we have $s(f) \in \mathcal{D}_d$. We first give the definition of what we mean by Bernoulli distributions behind a given discrete distribution $g \in \mathcal{D}_d$.

Definition 1.3.1. *We say that a Bernoulli pmf $f \in \mathcal{B}_d$ is* behind *a discrete pmf $g \in \mathcal{D}_d$ if $s(f) = g$.*

We can then define the set of all Bernoulli pmfs behind a given distribution g on $\{0,\dots,d\}$ by

$$s^{-1}(g) \coloneqq \{f \in \mathcal{B}_d\colon s(f) = g\}.$$

The geometrical structure the family $s^{-1}(g)$ has been explicitly studied in [47]. For $k \in \{0,\dots,d\}$, we note that the set of binary vectors $\mathcal{X}_d^k = \{\boldsymbol{x} \in \mathcal{X}_d\colon |\boldsymbol{x}| = k\}$ inherits the reverse-lexicographical order of $\mathcal{X}_d$. Let $\boldsymbol{x}_{j,d}^k$ be the jth element of $\mathcal{X}_d^k$, for $j \in \{1,\dots,\binom{d}{k}\}$.

Proposition 1.3.3. *For every $g \in \mathcal{D}_d$, the class $s^{-1}(g)$ of Bernoulli pmfs behind g is a convex polytope whose extremal points are $f^{\boldsymbol{\sigma}}$, for $\boldsymbol{\sigma} = (\sigma_1,\dots,\sigma_d) \in \prod_{j=1}^d \{1,\dots,\binom{d}{j}\}$, which are given by*

$$f^{\boldsymbol{\sigma}}(\boldsymbol{x}) = \begin{cases} g(j), & \text{if } \boldsymbol{x} = \boldsymbol{x}_{\sigma_j,d}^j, \\ 0, & \text{otherwise.} \end{cases}$$

For any pmf $g \in \mathcal{D}_d$, $s^{-1}(g)$ is not empty. Indeed, as we discussed in Section 1.3.2, for every $g \in \mathcal{D}_d$, there exists a unique exchangeable Bernoulli pmf $f^e \in \mathcal{E}_d(\mu/d)$, where μ is the mean of any discrete random variable Y with pmf g. The exchangeable pmf occupies a central position within the polytope $s^{-1}(g)$, as it can be expressed as a convex combination of the extremal pmfs $f^{\boldsymbol{\sigma}}$ with equal weights.

Chapter 2

Negative Dependence in Bernoulli Distributions

In actuarial science and mathematical finance, a central object of interest is the aggregate risk associated with a portfolio of individual positions. This is naturally modeled as the sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$, where the X_j are Bernoulli random variables representing, for example, default events. A key question is to quantify how far the aggregate risk $S_{\boldsymbol{X}}$ may deviate from its expected value $\mathbb{E}S_{\boldsymbol{X}}$, since such deviations correspond to extreme but practically relevant risk scenarios. The objective is to identify dependence structures that control these deviations, ensuring that the aggregate risk remains concentrated around its mean. This motivates the study of negative dependence, which describes the idea of diversification, a key concept of mathematical finance and actuarial science. Negative dependence represents the idea that large values of some components tend to inhibit large values of others. If, for example, the variables are pairwise negatively correlated, i.e., $\mathrm{Cov}(X_{j_1}, X_{j_2}) \leq 0$ for all $1 \leq j_1 < j_2 \leq d$, we obtain

$$\mathbb{P}(|S - \mathbb{E}S| \geq a) \leq \frac{d}{4a^2},$$

which follows from Chebyshev's inequality and the reduction of variance induced by negative covariances. Stronger notions of negative dependence lead to even more powerful results. For example, under negative association (see Definition 2.2.1 in Section 2.2), the sum $S_{\boldsymbol{X}}$ satisfies the Chernoff–Hoeffding bounds. In particular, for $\delta > 0$ we have

$$\mathbb{P}(S_{\boldsymbol{X}} \geq (1+\delta)\mathbb{E}S_{\boldsymbol{X}}) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mathbb{E}S_{\boldsymbol{X}}},$$
$$\mathbb{P}(S_{\boldsymbol{X}} \leq (1-\delta)\mathbb{E}S_{\boldsymbol{X}}) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mathbb{E}S_{\boldsymbol{X}}}.$$

For other forms of negative dependences, see, e.g., [38, 62, 80, 84].

The study of negative dependence is intrinsically linked to the search for distributions that minimize aggregate risk. In this chapter, we describe how negative dependence properties, as a form of repulsion among random variables or diversification, are linked to the concept of minimization of risk measures (e.g., expected shortfall, entropic risk measure, and value-at-risk), or of functionals $\mathbb{E}[\phi(\boldsymbol{x})]$, for some classes of functions ϕ.

This chapter is organized as follows. Section 2.1 demonstrates how the convex polytope structure, a recurring characteristic of classes of distributions considered in this work, facilitates the identification of the maximum and minimum values of the aforementioned risk functionals. Section 2.2 reviews the existing literature on negative dependence, providing an introduction to negative dependence properties, also through some dependence orders. In Section 2.3, we focus on extremal negative dependence, investigating the dependence structures that minimize specific dependence orders and functionals of the form $\mathbb{E}[\phi(\boldsymbol{X})]$ for determined classes of functions ϕ. Section 2.4 introduces the Strongly Rayleigh property, a strong notion of negative dependence characterized by the stability of the pgf. Finally, Section 2.5 presents original results that characterize extremal dependence within Bernoulli distributions, and studies the Strongly Rayleigh property among these extremal configurations.

2.1 Optimization over convex polytopes

Many of the classes of distributions encountered throughout this thesis exhibit the geometrical structure of a convex polytope. An important consequence of this geometrical structure is that for any linear functional of the form $\mathbb{E}[\phi(\boldsymbol{X})]$, the extremal values are necessarily attained at the extremal pmf of the class.

Proposition 2.1.1. *Let $\boldsymbol{X}$ be a d-dimensional random vector with pmf $f \in \mathcal{F}$, where $\mathcal{F}$ is a class of pmfs with the structure of a convex polytope. Moreover, let $r_1, \dots, r_n \in \mathcal{F}$ be the extremal pmfs. Then, for any measurable function $\phi\colon \mathbb{R}^d \to \mathbb{R}$, there exist $\lambda_1, \dots, \lambda_n \geq 0$, with $\lambda_1 + \dots + \lambda_n = 1$, such that*

$$\mathbb{E}[\phi(\boldsymbol{X})] = \sum_{k=1}^{n} \lambda_k \mathbb{E}[\phi(\boldsymbol{R}_k)], \tag{2.1}$$

where, for every $k \in \{1, \dots, n\}$, $\boldsymbol{R}_k$ is a random vector with pmf r_k. Moreover,

$$\min_{k \in \{1,\dots,n\}} \mathbb{E}[\phi(\boldsymbol{R}_k)] \leq \mathbb{E}[\phi(\boldsymbol{X})] \leq \max_{k \in \{1,\dots,n\}} \mathbb{E}[\phi(\boldsymbol{R}_k)]. \tag{2.2}$$

Proof. The structure of $\mathcal{F}$ as a convex polytope implies that, for every $f \in \mathcal{F}$, there exist $\lambda_1, \dots, \lambda_n \geq 0$, with $\lambda_1 + \dots + \lambda_n = 1$, such that $f(\boldsymbol{x}) = \lambda_1 r_1(\boldsymbol{x}) + \dots + \lambda_n r_n(\boldsymbol{x})$, for every $\boldsymbol{x} \in \mathcal{X}$, where $\mathcal{X}$ is the support of $\mathcal{F}$. Therefore,

$$\mathbb{E}[\phi(\boldsymbol{X})] = \sum_{\boldsymbol{x} \in \mathcal{X}_d} \phi(\boldsymbol{x}) f(\boldsymbol{x}) = \sum_{k=1}^{n} \sum_{\boldsymbol{x} \in \mathcal{X}_d} \phi(\boldsymbol{x}) \lambda_k r_k(\boldsymbol{x}) = \sum_{k=1}^{n} \lambda_k \mathbb{E}[\phi(\boldsymbol{R}_k)]$$

holds, and (2.2) immediately follows. □

This property plays a central role in the study of dependence structures, as it reduces optimization problems over a high-dimensional class of distributions to a finite, albeit possibly large, set of the class, the extremal points. Many risk measures widely employed in financial mathematics and actuarial science can be expressed as linear functionals of the form $\mathbb{E}[\phi(\boldsymbol{X})]$. This implies that to find the most or least risky distribution within a polytope, we only need to evaluate the risk at its extremal configurations. A particularly important case arises when the function ϕ depends on $\boldsymbol{X}$ only through the sum of its components, that is, $\phi(\boldsymbol{x}) = \varphi(x_1 + \cdots + x_d)$. In this setting, maximizing or minimizing the expectation corresponds to finding the dependence structures that lead to the maximum or minimum aggregate risk ([5]).

In this thesis, we focus on three specific risk measures: the Value-at-Risk, the expected shortfall, and the entropic risk measure.

Definition 2.1.1. *Let Y be a random variable representing a loss with finite mean. Then, the value-at-risk (VaR) at level $\alpha \in (0,1)$ (VaR_α) is given by quantile of level α of the cumulative distribution function of Y, i.e.,*

$$VaR_\alpha(Y) = \inf\{y \in \mathbb{R} : \mathbb{P}(Y \leq y) \geq \alpha\}.$$

Definition 2.1.2. *Let Y be a random variable representing a loss with finite mean. The expected shortfall (ES) at level $\alpha \in (0,1)$ is defined as*

$$ES_\alpha(Y) = \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(Y)\mathrm{d}u. \tag{2.3}$$

The expected shortfall defined in (2.3) also admits the following representation

$$\mathrm{ES}_\alpha(Y) = \mathrm{VaR}_\alpha(Y) + \frac{1}{1-\alpha} E[\max(Y - \mathrm{VaR}_\alpha(Y), 0)], \; for \;\; \alpha \in (0,1).$$

Definition 2.1.3. *The entropic risk measure (ERM) is defined by*

$$\Psi_\gamma(Y) = \frac{1}{\gamma}\log(E[e^{\gamma Y}]),$$

assuming that there exists a real number $\gamma_0 > 0$ such that $E[e^{\gamma Y}]$ is finite for $\gamma \in (0, \gamma_0)$.

It is important to distinguish that both the ES and the ERM belong to the class of convex risk measures. Specifically, they can be represented as an expectation $\mathbb{E}[\phi(\boldsymbol{X})] = \mathbb{E}[\varphi(S_{\boldsymbol{X}})]$ where φ is a convex function. Therefore, we can apply Proposition 2.1.1, and their extremal values are attained at the extremal points of the polytope. Furthermore, the convexity of φ implies that the minimum risk is attained by distributions that are minimal with respect to the convex order, a connection that we will explore in detail in Section 2.3.

In contrast, the value-at-risk does not admit a representation as a linear functional $\mathbb{E}[\phi(\boldsymbol{X})]$ and, in general, is not convex. Despite this, the geometrical structure as a convex polytope still helps in finding bounds. As stated in the following proposition (proved in [43]), the maximum and minimum VaR are also attained at the extremal points of the polytope.

Proposition 2.1.2. *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector with pmf $f \in \mathcal{F}$, where $\mathcal{F}$ is a class of pmfs with the structure of a convex polytope, and let $r_1, \ldots, r_n \in \mathcal{F}$ be the extremal pmfs. Then,*

$$\min_{k\in\{1,\ldots,n\}} VaR\Big(\sum_{j=1}^d R_{k,j}\Big) \leq VaR\Big(\sum_{j=1}^d X_j\Big) \leq \max_{k\in\{1,\ldots,n\}} VaR\Big(\sum_{j=1}^d R_{k,j}\Big),$$

where, for every $k \in \{1, \ldots, n\}$, $\boldsymbol{R}_k = (R_{k,1}, \ldots, R_{k,d})$ is a random vector with pmf r_k.

In literature, numerous results provide bounds for the Value-at-Risk with fixed marginal distributions; see, e.g., [6, 41, 34, 89]. Moreover, [86] proposes an algorithm, the rearrangement algorithm, to approximate VaR bounds in a given Fréchet class. However, when the class of distributions admits an analytical characterization of its extremal points, these sharp bounds can be explicitly derived. For example, the extremal points of $\mathcal{D}_d(\mu)$ are given in (1.26). Therefore, we can determine the maximum and minimum aggregate VaR for Bernoulli random vectors with pmfs in the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ with $|\boldsymbol{p}| = \mu$. By Proposition 2.1.2 and the fact that the extremal points of $\mathcal{D}_d(\mu)$ are analytical, we can find bounds for the Value-at-Risk for distributions in $\mathcal{D}_d(\mu)$.

Proposition 2.1.3 ([43, Proposition 5.4])**.** *Let Y be a random variable with pmf in $\mathcal{D}_d(\mu)$, for $p \in (0,1)$. Then, for every $\alpha \in (0,1)$,*

$$\max\left\{0; \left\lceil \frac{\mu - d(1-\alpha)]}{\alpha} \right\rceil\right\} \leq VaR_\alpha(Y) \leq \min\left\{d; \left\lceil \frac{\mu}{1-\alpha} \right\rceil - 1\right\},$$

where $\lceil y \rceil$ is the smallest integer bigger than or equal to $y \in \mathbb{R}$.

2.2 Dependence orders and negative dependence

As observed by Pemantle in his seminal paper [83], the mathematical formalization of negatively dependent random variables is much harder than defining positive dependence. While positive dependence can be described by a few consistent properties, negative dependence is more fragmented and involves many different and non-equivalent definitions. A first intuitive idea of this asymmetry is clear in the trivariate case. Consider three random variables (X_1, X_2, X_3). If X_1 and X_2 are positively related (in the sense that a high value of one random variable implies a high value of the other, and vice versa) and X_2 and X_3 are positively related as well, there is no structural obstacle to X_1 and X_3 being positively related. In contrast, negative dependence is subject to frustration. Consider the case of a 3-dimensional Bernoulli random vector (X_1, X_2, X_3) being Markov, i.e., conditionally on X_2, X_1 and X_3 are independent. In this case, negative correlations between X_1 and X_2 and between X_2 and X_3 imply that X_1 and X_3 are positively correlated.

To rigorously show this asymmetry within the class $\mathcal{B}_d$, we first recall the standard definitions for positive dependence. A Bernoulli random vector $\boldsymbol{X}$ and its pmf f are said to be *positively associated* if

$$\mathrm{Cov}\big(h_1(\boldsymbol{X}), h_2(\boldsymbol{X})\big) \geq 0, \tag{2.4}$$

for all non-decreasing functions h_1 and h_2 on $\mathcal{X}_d$, or, equivalently, if

$$\sum_{\boldsymbol{x}\in\mathcal{X}_d} h_1(\boldsymbol{x})h_2(\boldsymbol{x})f(\boldsymbol{x}) \geq \Bigg(\sum_{\boldsymbol{x}\in\mathcal{X}_d} h_1(\boldsymbol{x})f(\boldsymbol{x})\Bigg)\Bigg(\sum_{\boldsymbol{x}\in\mathcal{X}_d} h_2(\boldsymbol{x})f(\boldsymbol{x})\Bigg).$$

This inequality is a version of the FKG inequalities for binary random variables that is useful, e.g., for proving distributional limit theorems; see [81]. It captures the idea that high values of some components tend to occur together with high values of others. As proved in [48], positive association is implied by a stronger pointwise property, named *positive lattice condition* (PLC) or *multivariate totally positivity of order 2* (MTP_2). A pmf $f \in \mathcal{B}_d$ satisfies the PLC if

$$f(\boldsymbol{x})f(\boldsymbol{y}) \leq f(\boldsymbol{x} \wedge \boldsymbol{y})f(\boldsymbol{x} \vee \boldsymbol{y}), \tag{2.5}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_d$, where $\wedge$ and $\vee$ denote the componentwise minimum and maximum operators, respectively. An important property of PLC is that it is closed under marginalization; see [63, Theorem 2.5].

Moving to negative dependence, it is not possible to define negative association by simply reversing the inequality in (2.4) for all pairs of non-decreasing function h_1 and h_2. Indeed, such a definition would imply $\mathrm{Var}(X_j) \leq 0$ for every $j \in \{1, \dots, d\}$. To avoid contradictions, we restrict to functions depending on disjoint subsets of variables. This definition was introduced in Section 4 of [67]; for its properties and applications, see [62].

Definition 2.2.1. *A Bernoulli random vector $\boldsymbol{X}$ and its pmf $f \in \mathcal{B}_d$ are said to be* negatively associated *(NA) if for every pair of disjoint index sets $\Lambda_1, \Lambda_2 \subseteq \{1, \dots, d\}$ and for all non-decreasing functions $h_1 : \mathcal{X}_{|\Lambda_1|} \to \mathbb{R}$ and $h_2 : \mathcal{X}_{|\Lambda_2|} \to \mathbb{R}$, the inequality*

$$\mathbb{E}[h_1(X_j : j \in \Lambda_1)h_2(X_j : j \in \Lambda_2)] \leq \mathbb{E}[h_1(X_j : j \in \Lambda_1)]\mathbb{E}[h_2(X_j : j \in \Lambda_2)] \tag{2.6}$$

holds, provided that the expectations are finite.

As a counterpart to the PLC, we can also define the negative lattice condition.

Definition 2.2.2. *A random vector $\boldsymbol{X}$ and its pmf $f \in \mathcal{B}_d$ satisfy the* negative lattice condition *(NLC), or* multivariate reverse rule of order 2 *(MRR_2), if*

$$f(\boldsymbol{x})f(\boldsymbol{y}) \geq f(\boldsymbol{x} \wedge \boldsymbol{y})f(\boldsymbol{x} \vee \boldsymbol{y}), \tag{2.7}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_d$.

It is well-known in the theory of negative dependence that, unlike the PLC, which is closed under marginalization and implies positive association, the NLC does not guarantee negative association, and it is not preserved when taking sub-vectors. See Section 9.2.1 in [63] and Example 2.5.2 in Section 2.5.1 for examples of MRR_2 distribution with positively associated marginals.

To overcome these structural inconsistencies, we shift from an absolute definition to a comparative framework. By operating within a Fréchet class $\mathcal{B}_d(\mathbf{p})$, we isolate the interaction among variables from their marginal behaviors; following the spirit of the axiomatic approach proposed for positive dependence in [69] and the formalizations of Joe [63], negative dependence properties are then rigorously defined by comparing a Bernoulli pmf to its independent counterpart through specific dependence orders. We denote by $f_{\boldsymbol{p}}^{\perp}$ the pmf in the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ of a Bernoulli random vector $\boldsymbol{X}^{\perp}$ with independent components. As we shall see in Section 2.3, dependence orders also provide an important mathematical framework to define concepts of extremal negative dependence by the identification of the distributions in a Fréchet class that are minimal with respect to these orders. We introduce in this section dependence orders using the standard notation for random vectors, i.e., $\boldsymbol{X} \preceq_* \boldsymbol{X}'$. Since these orders depend on the underlying pmfs, with a slight abuse of notation, we write $f \preceq_* f'$ to indicate that any pair of Bernoulli random vectors $\boldsymbol{X}$ and

$\boldsymbol{X}'$, with pmfs f and f', satisfies the specified dependence order. The NA property can be defined through the weak-association order, which was introduced in [93, pag. 419–420], and we recall in the following definition.

Definition 2.2.3. *Let $\boldsymbol{X}$ and $\boldsymbol{X}'$ be Bernoulli random vectors with pmfs $f, f' \in \mathcal{B}_d(\boldsymbol{p})$, respectively. We say that $\boldsymbol{X}$ is smaller than $\boldsymbol{X}'$ in the weak association order, denoted $\boldsymbol{X} \preceq_{wa} \boldsymbol{X}'$, if*

$$Cov(h_1(X_j : j \in \Lambda_1), h_2(X_j : j \in \Lambda_2)) \leq Cov(h_1(Y_j : j \in \Lambda_1), h_2(Y_j : j \in \Lambda_2)),$$

for all disjoint subsets $\Lambda_1, \Lambda_2 \subseteq \{1, \ldots, d\}$ and for all monotone increasing functions h_1 and h_2, for which the above covariances are defined.

In particular, we can define NA as a subset of $\mathcal{B}_d(\boldsymbol{p})$ as follows

$$\mathrm{NA} = \{f \in \mathcal{B}_d(\boldsymbol{p}) : f \preceq_{\mathrm{wa}} f_{\boldsymbol{p}}^{\perp}\}.$$

A weaker negative dependence property than NA is negative supermodular dependence, which is defined through the supermodular order; see [3]. A supermodular function is a function $\phi\colon \{0,1\}^d \to \mathbb{R}$ such that $\phi(\boldsymbol{i}) + \phi(\boldsymbol{k}) \leq \phi(\boldsymbol{i} \wedge \boldsymbol{k}) + \phi(\boldsymbol{i} \vee \boldsymbol{k})$, where $\wedge$ and $\vee$ respectively denote the componentwise minimum and maximum operators.

Definition 2.2.4. *Given two Bernoulli random vectors $\boldsymbol{X}$ and $\boldsymbol{X}'$ with pmfs $f, f' \in \mathcal{B}_d(\boldsymbol{p})$, we say that $\boldsymbol{X}$ is smaller than $\boldsymbol{X}'$ in the supermodular order, denoted by $\boldsymbol{X} \preceq_{sm} \boldsymbol{X}'$, if $\mathbb{E}[\phi(\boldsymbol{X})] \leq \mathbb{E}[\phi(\boldsymbol{X}')]$, for all supermodular functions $\phi\colon \{0,1\}^d \to \mathbb{R}$ such that the expectations are finite.*

Theorem 9.E.8 of [93] proves that $\boldsymbol{X} \preceq_{\mathrm{wa}} \boldsymbol{X}'$ implies $\boldsymbol{X} \preceq_{\mathrm{sm}} \boldsymbol{X}'$. The property of negative supermodular dependence is defined by comparing $\boldsymbol{X}$ in supermodular order with $\boldsymbol{X}^{\perp}$, see [59].

Definition 2.2.5. *A Bernoulli random vector $\boldsymbol{X}$ and its pmf $f \in \mathcal{B}_d(\boldsymbol{p})$ satisfy the negative supermodular dependence property (NSD) if $f \preceq_{sm} f_{\boldsymbol{p}}^{\perp}$, i.e.,*

$$NSD = \{f \in \mathcal{B}_d(\boldsymbol{p}) : f \preceq_{sm} f_{\boldsymbol{p}}^{\perp}\}.$$

In [27], the authors prove that NA $\subseteq$ NSD, that is also a consequence of the fact that the weak-association order implies the supermodular one.

Finally, the weakest form of negative dependence is the pairwise negative correlation (p-NC), which simply requires $\mathrm{Cov}(X_{j_1}, X_{j_2}) \leq 0$ for all $1 \leq j_1 < j_2 \leq d$, capturing only the bivariate interactions. In the class of Bernoulli random vectors, p-NC is equivalent to requiring that all pairs are negatively associated, corresponding to p-NA. Therefore, this property is the minimal requirement for negative dependence.

2.3 Extremal negative dependence

We define extremal negative dependence as the property of those pmfs that are minimal within their respective Fréchet class, according to a specified dependence order. In dimension two, this framework admits a unique solution: the Fréchet lower bound, which corresponds to the distribution of a countermonotonic bivariate random vector.

Definition 2.3.1. *The bivariate random vector* $\boldsymbol{X} = (X_1, X_2)$ *defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is countermonotonic if* $(X_1(\omega) - X_1(\omega'))(X_2(\omega) - X_2(\omega')) \leq 0$*, for* $(\mathbb{P} \times \mathbb{P})$*-almost every* $(\omega, \omega') \in \Omega^2$*.*

See, for example, [35] and [87] for details about the concept of countermonotonicity. Theorem 3.1 in [87] shows that countermonotonic random vectors are minimal in their Fréchet class under the supermodular order. Furthermore, since for bivariate random vectors the supermodular and the weak-association orders coincide (see, e.g., [77, Theorem 3.8.2]), it follows that countermonotonicity also identifies the minimal element under the weak-association order in the bivariate case.

However, as discussed previously for negative dependence properties, the generalization to $d \geq 3$ is non-trivial. Consider the notion of pairwise countermonotonicity, an intuitive extension of countermonotonicity that requires each pair of random variables within a Bernoulli random vector to be countermonotonic. Pairwise countermonotonicity was introduced in [33] and recently studied in [70]; it is also known as mutual exclusivity in the actuarial literature ([26]). As also shown in [87], the distribution of a pairwise countermonotonic random vector is the lower Fréchet bound; moreover, such a distribution is minimal within its Fréchet class with respect to the supermodular order. Nevertheless, a Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ admits a pairwise countermonotonic random vector if and only if $|\boldsymbol{p}| \leq 1$ or $|\boldsymbol{p}| \geq d - 1$, as proved in Lemma 3.5 in [26]. We do not examine this property in detail because we do not want to have constraints on the marginal means. We consider three other notions of extremal negative dependence introduced in the literature to generalize countermonotonicity in dimensions higher than two: minimality in convex sums, joint mixability, and Σ-countermonotonicity; see [87] and references therein for a discussion on these notions in a general framework.

As highlighted at the end of Section 1.2.2, evaluating a functional of the form $\mathbb{E}[\phi(\boldsymbol{X})]$ over a Fréchet class of Bernoulli distributions implies that the minimum is attained at an extremal point of the polytope. However, an extremal negative dependence concept requires a single distribution to minimize $\mathbb{E}[\phi(\boldsymbol{X})]$ for an entire class of functions (e.g., all supermodular functions). In a multidimensional setting, the specific extremal point that minimizes the expectation typically changes depending on the chosen function ϕ, meaning that a universal minimum for the entire class of functions often does not exist. Therefore, we focus on a specific subclass of supermodular functions of the form $\phi(\boldsymbol{x}) = \varphi(x_1 + \cdots + x_d)$, where $\varphi \colon \mathbb{R} \to \mathbb{R}$ is a convex function. This motivates the characterization of Bernoulli random vectors $\boldsymbol{X} = (X_1, \ldots, X_d)$ whose sums $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ are minimal in the convex order, which we recall in the following definition.

Definition 2.3.2. *Given two random variables* Y *and* Y'*, we say that* Y *is smaller than* Y' *under the convex order, denoted by* $Y \preceq_{cx} Y'$*, if* $\mathbb{E}[\phi(Y)] \leq \mathbb{E}[\phi(Y')]$ *for every convex function* $\phi : \mathbb{R} \to \mathbb{R}$ *for which the expectations are finite.*

The first concept of minimality in convex sums is formally presented as follows.

Definition 2.3.3. *A Bernoulli random vector* $\boldsymbol{X}$ *and its pmf* $f \in \mathcal{B}_d(\boldsymbol{p})$ *are said to be* Σ_{cx}*-smallest elements of* $\mathcal{B}_d(\boldsymbol{p})$ *if*

$$\sum_{j=1}^{d} X_j \preceq_{cx} \sum_{j=1}^{d} X'_j,$$

for all $\boldsymbol{X}' = (X'_1, \dots, X'_d) \in \mathcal{B}_d(\boldsymbol{p})$.

We denote by $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p}) \subset \mathcal{B}_d(\boldsymbol{p})$ the subclass of Bernoulli pmfs with marginal means $\boldsymbol{p}$ that are Σ_{cx}-smallest elements in $\mathcal{B}_d(\boldsymbol{p})$. It is worth noting that the supermodular order between random vectors implies the convex order of their sums, since any function $\phi(\boldsymbol{x}) = \varphi(x_1 + \cdots + x_d)$ is supermodular, whenever φ is convex. Therefore, any distribution that is minimal in the supermodular order is necessarily a Σ_{cx}-smallest element as well. Specifically, if a Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ admits a minimal element under the supermodular order, as is the case when $|\boldsymbol{p}| \leq 1$ or $|\boldsymbol{p}| \geq d-1$, where a pairwise countermonotonic vector exists, then $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$ contains only such a minimal distribution.

In general, not all Fréchet classes admit a Σ_{cx}-smallest element; however, under the Bernoulli framework, for all $\boldsymbol{p} \in (0,1)^d$, the set $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$ is non empty. Indeed, Lemma 3.1 in [6] provides a constructive method to find a Bernoulli random vector with pmf in $\mathcal{B}_d(\boldsymbol{p})$, for every $\boldsymbol{p} \in (0,1)^d$.

Lemma 2.3.1. *For every* $\boldsymbol{p} \in (0,1)^d$*, the Fréchet class* $\mathcal{B}_d(\boldsymbol{p})$ *admits a* Σ_{cx}*-smallest element.*

Moreover, for Fréchet classes with a common marginal mean p, the existence of an exchangeable Σ_{cx}-smallest element is guaranteed by the one-to-one correspondence between $\mathcal{E}_d(p)$ and the univariate distributions of their sums. An algorithm to construct a non-exchangeable Σ_{cx}-smallest pmf in $\mathcal{B}_d(p)$ is provided in Theorem 3.1 in [45].

Since any discrete distribution with support on $\{0, \dots, d\}$ can be constructed as the distribution of a sum of Bernoulli random variables in many ways, usually several Σ_{cx}-smallest elements exist. Let $\boldsymbol{X} \in \mathcal{B}_d$, then the sum $S_{\boldsymbol{X}}$ is a discrete random variable with support on $\{0, \dots, d\}$. The Bernoulli random vector $\boldsymbol{X}$ is a Σ_{cx}-smallest element in $\mathcal{B}_d(\boldsymbol{p})$ if $S_{\boldsymbol{X}}$ is minimal in convex order in $\mathcal{D}_d(|\boldsymbol{p}|)$. For any $|\boldsymbol{p}| \in (0, d)$, the minimal pmf in convex order $s_{|\boldsymbol{p}|}$ in $\mathcal{D}_d(|\boldsymbol{p}|)$ is known (see, e.g., [6, 45]), and it is the extremal point in (1.26) given by

$$r^D_{|\boldsymbol{p}|}(y) \coloneqq r^D_{m,m+1} = \begin{cases} m + 1 - |\boldsymbol{p}|, & y = m, \\ |\boldsymbol{p}| - m, & y = m+1, \\ 0, & \text{otherwise}, \end{cases} \tag{2.8}$$

where m is the largest integer smaller than or equal to $|\boldsymbol{p}|$. Then, the class $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$ is the family of Bernoulli random vectors $\boldsymbol{X}$ such that the pmf of $S_{\boldsymbol{X}}$ is equal to $r^D_{|\boldsymbol{p}|}(y) \in \mathcal{D}_d(|\boldsymbol{p}|)$.

Proposition 2.3.2. *The class of* $\mathcal{B}_d^{cx}(\boldsymbol{p})$ *is a convex polytope. Its extremal pmfs are the extremal pmfs of* $\mathcal{B}_d(\boldsymbol{p})$ *that are* Σ_{cx}*-smallest elements in* $\mathcal{B}_d(\boldsymbol{p})$.

Proof. The class $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$ is a convex polytope, as it arises from the intersection of two convex polytopes. Specifically, it can be characterized as $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p}) = \mathcal{B}_d(\boldsymbol{p}) \cap s^{-1}(r^D_{|\boldsymbol{p}|})$, where $\mathcal{B}_d(\boldsymbol{p})$ is the Fréchet class with fixed marginal means $\boldsymbol{p}$ and $s^{-1}(r^D_{|\boldsymbol{p}|})$ is the set of Bernoulli pmfs behind the minimal pmf in convex order in $\mathcal{D}_d(|\boldsymbol{p}|)$.

Consider the extremal pmfs of $\mathcal{B}_d(\boldsymbol{p})$, $r_1, \dots, r_{n_d^{\boldsymbol{p}}}$, and suppose that the first $\tilde{n}$ extremal pmfs are Σ_{cx}-smallest elements, with $\tilde{n} \in \{1, \dots, n_d^{\boldsymbol{p}}\}$, namely, $r_1, \dots, r_{\tilde{n}} \in \mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$. This means that, for every $k \in \{1, \dots, \tilde{n}\}$, the sum of the components of any random vector $\boldsymbol{R}_k$ with pmf r_k has pmf equal to $r^D_{|\boldsymbol{p}|}$. If $f \in \mathcal{B}_d(\boldsymbol{p})$ is a convex combination of the extremal points $r_1, \dots, r_{\tilde{n}}$, then it easily follows that $f \in \mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$. Conversely, suppose that $f \in \mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p}) \subseteq \mathcal{B}_d(\boldsymbol{p})$. Then, there exist $\alpha_1, \dots, \alpha_{n_d^{\boldsymbol{p}}} \geq 0$ summing up to one such that, for every $\boldsymbol{x} \in \mathcal{X}_d$,

$$f(\boldsymbol{x}) = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \alpha_k r_k(\boldsymbol{x}),$$

where r_k, for $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$, are the extremal points of $\mathcal{B}_d(\boldsymbol{p})$, and $n_d^{\boldsymbol{p}}$ is their number. Consider a random vector $\boldsymbol{X}$ with pmf f and, for each $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$, the random vector $\boldsymbol{R}_k$ with pmf r_k. Since f is a Σ_{cx}-smallest element in $\mathcal{B}_d(\boldsymbol{p})$, the sum $S_{\boldsymbol{X}}$ has pmf $r_{|\boldsymbol{p}|}^D$. Then, for every $y \in \{0, \dots, d\}$ we have

$$\begin{aligned} r_{|\boldsymbol{p}|}^D(y) = \mathbb{P}(S_{\boldsymbol{X}} = y) &= \sum_{\boldsymbol{x} \in \mathcal{X}_d^k} f(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d^k} \sum_{k=1}^{n_d^{\boldsymbol{p}}} \alpha_k r_k(\boldsymbol{x}) \\ &= \sum_{k=1}^{n_d^{\boldsymbol{p}}} \alpha_k \sum_{\boldsymbol{x} \in \mathcal{X}_d^k} r_k(\boldsymbol{x}) = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \alpha_k \mathbb{P}(S_{\boldsymbol{R}_k} = y), \end{aligned}$$

where, for every $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$, the random variable $S_{\boldsymbol{R}_k}$ is the sum of the components of $\boldsymbol{R}_k$ and has distribution in $\mathcal{D}_d(|\boldsymbol{p}|)$. However, since $r_{|\boldsymbol{p}|}^D$ is an extremal pmf of $\mathcal{D}_d(|\boldsymbol{p}|)$, it follows that $\alpha_k = 0$ if the pmf of $S_{\boldsymbol{R}_k}$ is not equal to $r_{|\boldsymbol{p}|}^D$. We can conclude that $f \in \mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p})$ if and only if , for every $\boldsymbol{x} \in \mathcal{X}_d$,

$$f(\boldsymbol{x}) = \sum_{k=1}^{\tilde{n}} \alpha_k r_k(\boldsymbol{x}).$$

□

From (2.8), $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$ is a Σ_{cx}-smallest element if and only if it has support on $\mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}$, when $|\boldsymbol{p}| \in (m, m+1)$. If $|\boldsymbol{p}|$ is an integer, any minimal Σ_{cx}-smallest Bernoulli random vector has support on $\mathcal{X}_d^{|\boldsymbol{p}|}$ and it is a joint mix, according to the following definition.

Definition 2.3.4. *A Bernoulli random vector $\boldsymbol{X}$ and its pmf $f \in \mathcal{B}_d(\boldsymbol{p})$ are joint mixes if there exists $m \in \{1, \dots, d-1\}$ such that $\mathbb{P}(S_{\boldsymbol{X}} = m) = 1$, or, equivalently, if $f(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \notin \mathcal{X}_d^m$.*

Notice that if $\boldsymbol{X}$ is a joint mix, then its pgf $\mathcal{P}_{\boldsymbol{X}}$ is a homogeneous polynomial of degree m. It immediately follows that if $|\boldsymbol{p}|$ is not integer the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ does not admit any joint mix element. Introduced in [87], the last notion of extremal negative dependence is a generalization of countermonotonicity weaker than minimality in convex sums, but which exists in every Fréchet class, not only for Bernoulli random variables. In the following definition, we use the convention $\sum_{j \in \varnothing} X_j = 0$.

Definition 2.3.5. *A Bernoulli random vector $\boldsymbol{X}$ and its pmf $f \in \mathcal{B}_d(\boldsymbol{p})$ are said to be Σ-countermonotonic if, for every subset $\Lambda \subseteq \{1, \dots, d\}$, the random vector*

$$\left(\sum_{j \in \Lambda} X_j, \sum_{j \notin \Lambda} X_j \right)$$

is countermonotonic.

Let us denote by $\mathcal{B}_d^{\Sigma} \subseteq \mathcal{B}_d$ the class of Σ-countermonotonic d-dimensional Bernoulli pmfs, and let $\mathcal{B}_d^{\Sigma}(\boldsymbol{p}) \coloneqq \mathcal{B}_d(\boldsymbol{p}) \cap \mathcal{B}_d^{\Sigma}$ be the set of Σ-countermonotonic Bernoulli pmf in the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$. The authors of [87] proved in Theorem 3.8 that Σ_{cx}-smallest Bernoulli random vectors are Σ-countermonotonic,i.e., $\mathcal{B}_d^{\mathrm{cx}}(\boldsymbol{p}) \subseteq \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. In Section 2.5, we will show that these two classes coincide.

We conclude this section by considering the class of exchangeable Bernoulli distributions $\mathcal{E}_d(p) \subseteq \mathcal{B}_d(p)$. In this case, the following result by [49] shows that the supermodular order coincides with the convex order of sums.

Proposition 2.3.3 ([49])**.** *Let $\boldsymbol{X}$ and $\boldsymbol{X}'$ be two exchangeable Bernoulli random vectors with pmfs $f, f' \in \mathcal{E}_d(p)$, respectively. Then,*

$$\boldsymbol{X} \preceq_{sm} \boldsymbol{X}' \iff \sum_{j=1}^{d} X_j \preceq_{cx} \sum_{j=1}^{d} X'_j.$$

This equivalence has an immediate consequence for the extremal properties we have identified. Specifically, if we consider the extremal pmf $r^e_{m,m+1}$ defined in (1.27), where $m = \lfloor pd \rfloor$, it follows that the associated random vector $\boldsymbol{X}$ is not only a Σ_{cx}-smallest element of $\mathcal{E}_d(p)$, but it also represents the minimum within the exchangeable class with respect to the supermodular order.

2.4 Strongly Rayleigh Bernoulli distributions

In [15], the authors introduce a strong property of negative dependence for Bernoulli random vectors through a property of their probability generating function (pgf), called stability. We recall the definition and the main properties of stable polynomials in Section 2.4.1 and we formally define the strongly Rayleigh property in Section 2.4.2.

2.4.1 Stable polynomials

We denote by $\mathbb{R}[\boldsymbol{z}] = \mathbb{R}[z_1, \dots, z_d]$ and $\mathbb{C}[\boldsymbol{z}] = \mathbb{C}[z_1, \dots, z_d]$ the rings of polynomials in d indeterminates $\boldsymbol{z} = (z_1, \dots, z_d)$ with real and complex coefficients, respectively.

Definition 2.4.1. *A polynomial $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ is stable if $\mathcal{P}(\boldsymbol{z}) \equiv 0$ or if*

$$\mathcal{P}(z_1, \dots, z_d) \neq 0$$

whenever $\boldsymbol{z} = (z_1, \dots, z_d) \in \mathcal{H}^d$, where $\mathcal{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$, and $\Im(z)$ denotes the imaginary part of a complex number z. If the coefficients of $\mathcal{P}$ are real, then $\mathcal{P}$ is real stable.

Since a univariate polynomial with real coefficients is real stable if and only if it is real-rooted, stability can be interpreted as an extension of this property to the multivariate case. Some examples of real stable polynomials are the monomials $\prod_{j=1}^{n} z_j$, for $n \in \{1,2,\dots\}$, and linear polynomials $\sum_{i=1}^{d} a_i z_i$ with coefficients $a_i \geq 0$. The polynomial $1 - z_1 z_2$ is stable, in fact we observe that if $(z_1, z_2) \in \mathcal{H}^2$, then $z_1 z_2 < 0$ or $z_1 z_2 \in \mathbb{C}$. For the same reason, the polynomial $1 + z_1 z_2$ is not stable. For every $d \in \{1,2,\dots\}$ and

$m \in \{0, \dots, d\}$, the elementary symmetric polynomial $\mathcal{E}_{d,m}$, defined in (1.28), is real stable. Finally, another example of real stable polynomials are the so-called determinantal polynomials. If $A_1, \dots, A_d \in \mathbb{C}^{d\times d}$ are positive semidefinite and $B \in \mathbb{C}^{d\times d}$ is Hermitian, then the polynomial $\det(A_1 z_1 + \cdots + A_d z_d + B)$ is real stable; see Proposition 2.1 in [104].

Given the central role of this property, there exist various approaches in the literature to determine whether a specific polynomial satisfies the stability condition. First, since $\mathcal{H}^d = \{\boldsymbol{a} + \boldsymbol{b}z \colon \boldsymbol{a} \in \mathbb{R}^d, \boldsymbol{b} \in \mathbb{R}^d_+, z \in \mathcal{H}\}$, it directly follows from the definition that a polynomial $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ is stable if and only if for every $\boldsymbol{a} \in \mathbb{R}^d$ and $\boldsymbol{b} \in \mathbb{R}^d_+$, the univariate polynomial $\tilde{\mathcal{P}}(z) = \mathcal{P}(\boldsymbol{a} + \boldsymbol{b}z)$ is real-rooted.

As previously noted, the pgf of a Bernoulli distribution is a multiaffine polynomial, meaning that the maximum degree of each indeterminate is at most one. The following proposition is particularly important in the context of Bernoulli distributions, as it applies to multiaffine polynomials.

Theorem 2.4.1 ([18, Theorem 5.6])**.** *A multiaffine polynomial $\mathcal{P}(\boldsymbol{z})$ with real coefficients is real stable if and only if, for every $j, k \in \{1, \dots, d\}$,*

$$(\partial_j \mathcal{P}(\boldsymbol{y}))(\partial_k \mathcal{P}(\boldsymbol{y})) \geq (\partial_j \partial_k \mathcal{P}(\boldsymbol{y}))\mathcal{P}(\boldsymbol{y}), \quad \textit{for all } \boldsymbol{y} \in \mathbb{R}^d. \tag{2.9}$$

Given a multiaffine polynomial $\mathcal{P} \in \mathbb{C}[\boldsymbol{z}]$, using the shorthand notations $\mathcal{P}^j = \mathcal{P}|_{z_j=0}$ and $\mathcal{P}_j = \partial_j \mathcal{P}$, the decomposition in (1.7) reads $\mathcal{P} = \mathcal{P}^j + z_j \mathcal{P}_j$, for all $j \in \{1, \dots, d\}$. Since $\mathcal{P}^j$ and $\mathcal{P}_j$ are still multiaffine polynomials, we can write

$$\begin{aligned} \mathcal{P} = \mathcal{P}^j + z_j \mathcal{P}_j &= \mathcal{P}^{jk} + z_k \mathcal{P}^j_k + z_j(\mathcal{P}^k_j + z_k \mathcal{P}_{jk}) \\ &= \mathcal{P}^{jk} + z_j \mathcal{P}^k_j + z_k \mathcal{P}^j_k + z_j z_k \mathcal{P}_{jk}, \end{aligned}$$

for all $j, k \in \{1, \dots, d\}$. When $d > 2$, define the polynomial $\Delta_{jk}\mathcal{P} \coloneqq (\partial_j \mathcal{P})(\partial_k \mathcal{P}) - (\partial_j \partial_k \mathcal{P})\mathcal{P}$ for all $j, k \in \{1, \dots, d\}$, with $j \neq k$. Notice that the condition in (2.9) becomes $\Delta_{jk}\mathcal{P}(\boldsymbol{y}) \geq 0$ for all $j, k \in \{1, \dots, d\}$ and $\boldsymbol{y} \in \mathbb{R}^d$. Moreover,

$$\begin{aligned} \Delta_{jk}\mathcal{P} &= \mathcal{P}_j \mathcal{P}_k - \mathcal{P}_{jk}\mathcal{P} \\ &= (\mathcal{P}^k_j + z_k \mathcal{P}_{jk})(\mathcal{P}^j_k + z_j \mathcal{P}_{jk}) - \mathcal{P}_{jk}(\mathcal{P}^{jk} + z_j \mathcal{P}^k_j + z_k \mathcal{P}^j_k + z_j z_k \mathcal{P}_{jk}) \\ &= \mathcal{P}^k_j \mathcal{P}^j_k - \mathcal{P}^{jk}\mathcal{P}_{jk}. \end{aligned} \tag{2.10}$$

In Appendix B, we examine the 3-dimensional case, showing how (2.10) provides a straightforward criterion for verifying the Strongly Rayleigh property. We can simplify the conditions in (2.9) and state the following theorem.

Theorem 2.4.2 ([104, Theorem 3.1])**.** *Let $\mathcal{P}(\boldsymbol{z}) \in \mathbb{R}[z_1, \dots, z_d]$ be a multiaffine polynomial. The following are equivalent:*

(a) The polynomial $\mathcal{P}$ is real stable.

(b) For all $j, k \in \{1, \dots, d\}$ and for all $\boldsymbol{y} \in \mathbb{R}^d$,

$$(\partial_j \mathcal{P}|_{z_k=0}(\boldsymbol{y}))(\partial_k \mathcal{P}|_{z_j=0}(\boldsymbol{y})) - (\partial_j \partial_k \mathcal{P}(\boldsymbol{y}))(\mathcal{P}|_{z_j=z_k=0}(\boldsymbol{y})) \geq 0$$

(c) Either $d = 1$, or there exist $j, k \in \{1, \dots, d\}$ such that $\mathcal{P}_j$, $\mathcal{P}^j$, $\mathcal{P}_k$, and $\mathcal{P}^k$ are real stable, and $\Delta_{jk}\mathcal{P}(\boldsymbol{y}) \geq 0$ for all $\boldsymbol{y} \in \mathbb{R}^d$.

In the remainder of this section, we consider two fundamental aspects. First, we introduce the concept of *proper position* between two polynomials, a property that describes the geometry of zeros of two polynomials. Second, we characterize stability preservers, which are operators defined on polynomials that maintain the stability property. In both cases, we first consider the case of univariate polynomials, then we move to the generalizations in the multivariate case.

Interlacing polynomials

As mentioned above, univariate polynomials with real coefficients are stable if and only if are real-rooted. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two such polynomials and let $\alpha_1 \leq \cdots \leq \alpha_{n_1}$ and $\beta_1 \leq \cdots \leq \beta_{n_2}$ be their zeros, respectively. These zeros are *interlacing* if they can be ordered such that either $\alpha_1 \leq \beta_1 \leq \alpha_2 \leq \beta_2 \leq \cdots$ or $\beta_1 \leq \alpha_1 \leq \beta_2 \leq \alpha_2 \leq \cdots$. If the zeros of $\mathcal{P}_1$ and $\mathcal{P}_2$ interlace the Wronskian $W[\mathcal{P}_1, \mathcal{P}_2] = \mathcal{P}_1\mathcal{P}_2' - \mathcal{P}_1'\mathcal{P}_2$ is either non-negative or non-negative on the real line $\mathbb{R}$; see, e.g., [104]. We say that $\mathcal{P}_1$ and $\mathcal{P}_2$ are in a *proper position*, denoted $\mathcal{P}_1 \ll \mathcal{P}_2$, if their zeros interlace and $W[\mathcal{P}_1, \mathcal{P}_2] \leq 0$. For convenience, we also say that $\mathcal{P} \ll 0$ and $0 \ll \mathcal{P}$, for any real-rooted polynomial $\mathcal{P}$. Two classical results on interlacing polynomials state that $\mathcal{P}_2 \ll \mathcal{P}_1$ if and only if $\mathcal{G} = \mathcal{P}_1 + i\mathcal{P}_2 \in \mathbb{C}[z]$ is a stable polynomial and that $\mathcal{P}_1 \ll \mathcal{P}_2$ or $\mathcal{P}_2 \ll \mathcal{P}_1$ if and only if $\alpha\mathcal{P}_1 + \beta\mathcal{P}_2$ is real stable for all $\alpha, \beta \in \mathbb{R}$. The following definition generalizes this notion to the multivariate case; see [13].

Definition 2.4.2. *Let $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{R}[z_1, \ldots, z_d]$. We say that $\mathcal{P}_1$ and $\mathcal{P}_2$ are in a* proper position*, denoted $\mathcal{P}_1 \ll \mathcal{P}_2$, if for all $\boldsymbol{a} \in \mathbb{R}^d$ and $\boldsymbol{b} \in \mathbb{R}^d_+$, the univariate polynomials $\mathcal{P}_1(\boldsymbol{a} + \boldsymbol{b}z)$ and $\mathcal{P}_2(\boldsymbol{a} + \boldsymbol{b}z)$ are in a proper position, i.e., $\mathcal{P}_1(\boldsymbol{a} + \boldsymbol{b}z) \ll \mathcal{P}_2(\boldsymbol{a} + \boldsymbol{b}z)$.*

The importance of this property lies in its connection with the stability of linear combinations of two polynomials. Let $W_j[\mathcal{P}_1, \mathcal{P}_2] = \partial_j\mathcal{P}_1\mathcal{P}_2 - \mathcal{P}_1\partial_j\mathcal{P}_2$ be the jth Wronskian of the pair $(\mathcal{P}_1, \mathcal{P}_2)$.

Theorem 2.4.3 ([14, Theorem 2.9])**.** *Let $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{R}[z_1, \ldots, z_d]$. The following are equivalent.*

- *(a) The polynomial $\mathcal{P}_2 + i\mathcal{P}_1 \in \mathbb{C}[z_1, \ldots, z_d]$ is stable.*
- *(b) The polynomial $\mathcal{P}_2 + z_{d+1}\mathcal{P}_1 \in \mathbb{R}[z_1, \ldots, z_d, z_{d+1}]$ is real stable, where z_{d+1} is a new indeterminate.*
- *(c) $\mathcal{P}_1 \ll \mathcal{P}_2$.*
- *(d) $\alpha\mathcal{P}_1 + \beta\mathcal{P}_2$ is real stable for all $\alpha, \beta \in \mathbb{R}$, and $W_j[\mathcal{P}_1, \mathcal{P}_2](\boldsymbol{y}) \leq 0$ for all $j \in \{1, \ldots, d\}$ and $\boldsymbol{y} \in \mathbb{R}^d$.*

We can check that $\mathcal{E}_{d,m+1}(\boldsymbol{z}) + z_{d+1}\mathcal{E}_{d,m}(\boldsymbol{z}) = \mathcal{E}_{d+1,m+1}(\boldsymbol{z}, z_{d+1})$. Therefore, all the statements of Theorem 2.4.3 are verified by two consecutive elementary symmetric polynomials.

Stability preservers

In the univariate setting, stability preservers coincide with those operators that preserve the real-rootedness of polynomials. Consider an infinite sequence of real numbers $\lambda\colon \mathbb{N} \to \mathbb{R}$ and the associated diagonal linear operator $T_\lambda\colon \mathbb{C}[z] \to \mathbb{C}[z]$ defined by $T_\lambda(z^n) = \lambda(n)z^n$, for all $n \in \mathbb{N}$. The sequence λ is a *multiplier sequence* if T_λ preserves real-rootedness, i.e., if $T_\lambda(\mathcal{P})$ is real stable, whenever $\mathcal{P} \in \mathbb{R}[z]$ is real stable. The Pólya–Schur theorem characterizes multiplier sequences.

Theorem 2.4.4 (Pólya–Schur theorem)**.** *Let $\lambda\colon \mathbb{N} \to \mathbb{R}$ be an infinite sequence and T_λ be the associated diagonal linear operator. Then, λ is a multiplier sequence if and only if, for all $n \in \mathbb{N}$, the polynomial $T_\lambda((1+z)^n)$ is real stable with all roots of the same sign.*

Moving to the multivariate case, the following lemma summarizes several operators that preserve stability.

Lemma 2.4.5 ([104, Lemma 2.3])**.** *Let $\mathcal{P}(\boldsymbol{z}) \in \mathbb{C}[z_1, \dots, z_d]$ be a stable polynomial.*

*a) **Permutation.** The polynomial $\mathcal{P}(z_{\sigma(1)}, \dots, z_{\sigma(d)})$ is stable, for any permutation σ of $\{1, \dots, d\}$.*

*b) **Scaling.** For $c \in \mathbb{C}$ and $\boldsymbol{a} \in \mathbb{R}^d$, the polynomial $c\mathcal{P}(a_1 z_1, \dots, a_d z_d)$ is stable.*

*c) **Diagonalization.** The polynomial $\mathcal{P}(\boldsymbol{z})|_{z_1=z_2}$ is stable.*

*d) **Specialization.** For $a \in \mathbb{C}$ with $\Im(a) \geq 0$, the polynomial $\mathcal{P}(a, z_2, \dots, z_d)$ is stable.*

*e) **Inversion.** The polynomial $z_1^{-m}\mathcal{P}(-z_1^{-1}, z_2, \dots, z_d)$ is stable, where m is the maximum degree of z_1 in the polynomial.*

*f) **Differentiation.** The polynomial $\partial_1\mathcal{P}(\boldsymbol{z})$ is stable.*

Obviously, by permutation, items c), d), e), and f) hold for all indexes in $\{1, \dots, d\}$. The operators listed above represent only specific examples of stability preservers. A characterization of all finite order linear differential operators that preserve stability is provided in Theorems 1.2 and 1.3 of [13].

Beyond these operators, there are other two fundamental geometric constructions that preserve stability: polarization and homogenization. These techniques are particularly powerful and have a probabilistic meaning, as we shall see in Proposition 2.4.12 in Section 2.4.2. First, let $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ be a polynomial of degree n_j in z_j, for $j \in \{1, \dots, d\}$. The polarization $\pi(\mathcal{P})$ of $\mathcal{P}$ is the unique polynomial in $n_1 + \dots + n_d$ variables z_{jk}, for $j \in \{1, \dots, d\}$ and $j \in \{1, \dots, n_d\}$, such that:

(a) $\pi(\mathcal{P})$ is multiaffine;

(b) $\pi(\mathcal{P})$ is symmetric in the indeterminates $z_{j1}, \dots, z_{jn_j}$, for every $j \in \{1, \dots, d\}$;

(c) We recover $\mathcal{P}$ by letting $z_{jk} = z_j$ in $\pi(\mathcal{P})$, for all j, k.

We report next the famous Grace–Walsh–Szegö coincidence theorem. Recall that a circular region in $\mathbb{C}$ is an open or closed interior or exterior of a circle, or an open or closed affine half-plane.

Theorem 2.4.6 (Grace–Walsh–Szegö, Theorem 4.6 in [15])**.** *Let $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ be symmetric and multiaffine, and let $C \subseteq \mathbb{C}$ be a circular region containing the points $\zeta_1, \dots, \zeta_d$. Suppose that either $\mathcal{P}$ has total degree d or C is convex (or both). Then there exists at least one point $\zeta \in C$ such that $\mathcal{P}(\zeta_1, \dots, \zeta_d) = \mathcal{P}(\zeta, \dots, \zeta)$.*

This result has an important consequence for the preservation of stability under polarization. Since $\mathcal{H} = \{z \in \mathbb{C} \colon \Im(z) > 0\}$ is a circular region, we have the following proposition.

Proposition 2.4.7 ([15, Corollary 4.7])**.** *Let $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$. Then, $\mathcal{P}$ is stable if and only if $\pi(\mathcal{P})$ is stable.*

It is also possible to prove a result similar to Proposition 2.4.7.

Proposition 2.4.8. *Let $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ be a multiaffine polynomial. Then, for every $n \in \mathbb{N}$, $\mathcal{P}$ is stable if and only if $\pi_n(\mathcal{P})$ is stable, where $\pi_n(\mathcal{P}) \in \mathbb{C}[z_{11}, \dots, z_{1n}, z_2, \dots, z_d]$ is defined as*

$$\pi_n(\mathcal{P})(z_{11}, \dots, z_{1n}, z_2, \dots, z_d) = \mathcal{P}(\mathcal{E}_{n,1}(z_{11}, \dots, z_{1n}), z_2, \dots, z_d),$$

where $\mathcal{E}_{n,1}(z_{11}, \dots, z_{1n}) = (z_{11} + \cdots + z_{1n})/n$.

Proof. Since $\mathcal{P}(z_1, \dots, z_d) = \pi_n(\mathcal{P})(z_1, \dots, z_1, z_2, \dots, z_d)$, by the diagonalization property in Lemma 2.4.5, if $\pi_n(\mathcal{P})$ is stable, then so is $\mathcal{P}$. Suppose that $\pi_n(\mathcal{P})$ is not stable. This means that there exists $\boldsymbol{\zeta} = (\zeta_{11}, \dots, \zeta_{1n}, \zeta_2, \dots, \zeta_d) \in \mathcal{H}^{d+n-1}$ such that $\pi_n(\mathcal{P})(\boldsymbol{\zeta}) = 0$. By applying Theorem 2.4.6 to the multiaffine and symmetric polynomial $\pi_n(\mathcal{P})(z_{11}, \dots, z_{1n}, \zeta_2, \dots, \zeta_d) \in \mathbb{C}[z_{11}, \dots, z_{1n}]$, we find that there exists $\zeta_1 \in \mathcal{H}$ such that

$$\mathcal{P}(\zeta_1, \zeta_2, \dots, \zeta_d) = \pi_n(\mathcal{P})(\zeta_1, \dots, \zeta_1, \zeta_2, \dots, \zeta_d) = 0,$$

implying that $\mathcal{P}$ is not stable. □

A polynomial $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ is said to be *homogeneous* of degree n if all its terms have the same total degree n. This implies the scaling property $\mathcal{P}(a\boldsymbol{z}) = a^n \mathcal{P}(\boldsymbol{z})$, for any $a \in \mathbb{C}$. Any non-homogeneous polynomial $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$ can be transformed in a homogeneous polynomial. Let n be the degree of the polynomial $\mathcal{P}$. The homogenization of $\mathcal{P}$ is the polynomial of degree n given by

$$H(\mathcal{P})(z_1, \dots, z_d, z_{d+1}) = z_{d+1}^n \mathcal{P}\left(\frac{z_1}{z_{d+1}}, \dots, \frac{z_d}{z_{d+1}}\right).$$

Proposition 2.4.9 ([15, Theorem 4.5])**.** *Let $\mathcal{P} \in \mathbb{R}[z_1, \dots, z_d]$ be a polynomial with non-negative coefficients. Then $\mathcal{P}$ is real stable if and only if its homogenization $H(\mathcal{P})$ is real stable.*

In conclusion, a particular combination of homogenization and polarization leads to the so-called *symmetric homogenization* ([15, Section 2.3]), that is a construction that preserves stability. In particular, given $\mathcal{P} \in \mathbb{C}[z_1, \dots, z_d]$, its symmetric homogenization is the polynomial $\mathcal{P}_{\text{sh}} \in \mathbb{C}[z_1, \dots, z_{2d}]$ defined as

$$\mathcal{P}_{\text{sh}}(z_1, \dots, z_{2d}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f_{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}} \mathcal{E}_{d, d-|\boldsymbol{x}|}(z_{d+1}, \dots, z_{2d}),$$

where $\boldsymbol{z} = (z_1, \dots, z_d)$. Standard computations show that $\mathcal{P}_{\text{sh}} = \pi(z_{d+1}^{d-n} H(\mathcal{P}))$, therefore Propositions 2.4.7 and 2.4.9 imply the following result.

Proposition 2.4.10 ([15, Theorem 4.2])**.** *Let us consider a multiaffine polynomial $\mathcal{P} \in \mathbb{R}[z_1, \dots, z_d]$. Then, $\mathcal{P}$ is real stable if and only if its symmetric homogenization $\mathcal{P}_{sh} = \pi(z_{d+1}^{d-n} H(\mathcal{P})) \in \mathbb{R}[z_1, \dots, z_{2d}]$ is real stable.*

2.4.2 The strongly Rayleigh property

We are now ready to formally define the strongly Rayleigh property.

Definition 2.4.3. *A random vector $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ taking values on $\mathbb{N}^d$ and its pmf $f^{\boldsymbol{Y}}$ are strongly Rayleigh if the pgf $\mathcal{P}_{\boldsymbol{Y}}$ of the random vector $\boldsymbol{Y}$ is real stable.*

This definition naturally applies to the case of Bernoulli random vectors $\boldsymbol{X}$ with pmf $f \in \mathcal{B}_d$, as well as to discrete random variables D with pmf $f^D \in \mathcal{D}_d$. In the latter case, we recall that it is sufficient to verify whether the univariate pgf $\mathcal{P}_Y(z)$ of a discrete distribution is real-rooted.

Since the pgf of a Bernoulli distribution is a multiaffine polynomial with real coefficients, we can apply Theorem 2.4.1, and a Bernoulli random vector is strongly Rayleigh if and only if, for every $j, k \in \{1, \dots, d\}$,

$$(\partial_j \mathcal{P}(\boldsymbol{y}))(\partial_k \mathcal{P}(\boldsymbol{y})) \geq (\partial_j \partial_k \mathcal{P}(\boldsymbol{y}))\mathcal{P}(\boldsymbol{y}), \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^d. \tag{2.11}$$

By (1.6), it is easy to check that p-NC is equivalent to requiring that (2.11) holds for $\boldsymbol{y} = \mathbf{1}_d$; this immediately clarifies that the strongly Rayleigh property implies p-NC. In Section 4.2 of [15], it is proven that the strongly Rayleigh property implies NA and that NA does not imply it. Therefore, it is the strongest negative dependence property among those analyzed in this work. In the same paper, it is shown that the strongly Rayleigh property is stronger than the NLC.

Our main interest is the study of Bernoulli distributions within a Fréchet class with fixed marginals. As discussed in Section 2.2, some negative dependence properties can be formally expressed by comparing a given distribution to the state of independence through a suitable dependence ordering. This comparative framework further allows us to define concepts of extremal negative dependence as the minimal elements within a Fréchet class relative to these orders (see Section 2.3). The strongly Rayleigh property is a negative dependence property that is not defined through a dependence order, as the NSD property in Definition 2.2.5. A natural question now arises: can this strong form of negative dependence be characterized in a similar way? We aim to investigate whether there exists a dependence order that allows us to define the class of strongly Rayleigh Bernoulli distributions in a given Fréchet class via comparison with the independent case. In Section 4.3.1 of [15], the authors introduce a partial order $\trianglelefteq$ on the set of strongly Rayleigh pmfs, based on the property of proper position between polynomials (Definition 2.4.2).

Definition 2.4.4. *[15, Definition 4.1] Let $\mathcal{SR}_d \subset \mathcal{B}_d$ be the class of strongly Rayleigh Bernoulli pmfs. First, we define a partial order $\trianglelefteq'$ on $\mathcal{SR}_d$ by setting $f_1 \trianglelefteq' f_2$ if there exists a sequence of strongly Rayleigh pmfs $f_1 = g_0, g_1, \dots, g_n = f_2$ such that $\mathcal{P}_{g_0} \ll \mathcal{P}_{g_1} \ll \dots \ll \mathcal{P}_{g_n}$. Then, we define the partial order $\trianglelefteq$ on $\mathcal{SR}_d$ as the closure of $\trianglelefteq'$. In particular, we set $f_1 \trianglelefteq f_2$ if there exist two sequences, $\{f_{1,n}\}_{n=0}^{\infty}, \{f_{2,n}\}_{n=0}^{\infty} \subset \mathcal{SR}_d$, such that*

$$\lim_{n\to\infty} f_{1,n} = f_1, \quad \lim_{n\to\infty} f_{1,n} = f_1, \quad \textit{and } f_{1,n} \trianglelefteq' f_{2,n} \textit{ for all } n \in \mathbb{N}.$$

The following proposition shows that the partial order $\trianglelefteq$ implies the usual stochastic order $\preceq_{\mathrm{st}}$. Given two Bernoulli random vectors $\boldsymbol{X}_1, \boldsymbol{X}_2$ with pmfs f_1, f_2, we say that $\boldsymbol{X}_1$ are f_1 are smaller than $\boldsymbol{X}_2$ and f_2, respectively, in the usual stochastic order, denoted $\boldsymbol{X}_1 \preceq_{\mathrm{st}} \boldsymbol{X}_2$ or $f_1 \preceq_{\mathrm{st}} f_2$, if $\mathbb{P}(\boldsymbol{X}_1 \geq \boldsymbol{x}) \leq \mathbb{P}(\boldsymbol{X}_2 \geq \boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$. See [93] for the definition of the usual stochastic order and as a standard reference on stochastic orders.

Proposition 2.4.11. *Let $f_1, f_2 \in \mathcal{B}_d$ be two strongly Rayleigh pmfs. Then $f_1 \trianglelefteq f_2$ implies $f_1 \preceq_{st} f_2$.*

It is well known that any two Bernoulli random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same joint distribution, if their pmfs are in the same Fréchet class and are such that $f_1 \preceq_{\mathrm{st}} f_2$; see Theorem 6.B.19 of [93]. This means that two Bernoulli random vectors $\boldsymbol{X}$ and $\boldsymbol{X}'$ in the same Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ have the same distribution or are not comparable in $\trianglelefteq$. This implies that $\trianglelefteq$ is not a dependence order as defined in [63]. Consequently, within a Fréchet class, it is impossible to compare the strength of the strongly Rayleigh property, at least with $\trianglelefteq$. This means that $\trianglelefteq$ is not the right partial order to define a class of extremal negative dependent Bernoulli random vectors, i.e., minimal with respect to the corresponding partial order. The best we can do is to define a class of extremal negative dependence distributions with respect to the traditional dependence (partial) orders, such as $\preceq_{\mathrm{wa}}$ or $\preceq_{\mathrm{sm}}$, and look for the distribution within the extremal negative dependence class that satisfies the strongly Rayleigh property.

As previously argued, if we consider two Bernoulli random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ whose pmfs f_1, f_2 belong to the same Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, their pgfs $\mathcal{P}_1$ and $\mathcal{P}_2$ can never be in a proper position. This observation has a significant impact on the study of convex combinations of strongly Rayleigh distributions. Specifically, we are unable to invoke Theorem 2.4.3 to guarantee that $\lambda\mathcal{P}_1 + (1-\lambda)\mathcal{P}_2$ is a stable polynomial for every $\lambda \in (0,1)$. From a probabilistic perspective, this implies that we cannot use this approach to ensure that the mixture $f_\lambda = \lambda f_1 + (1-\lambda) f_2 \in \mathcal{B}_d(\boldsymbol{p})$ remains strongly Rayleigh for all $\lambda \in (0,1)$.

The following proposition collects the main closure properties of the strongly Rayleigh property. These results follow directly from the polynomial transformations that preserve stability, such as the operators discussed in Lemma 2.4.5, as well as the operations of polarization and homogenization.

Proposition 2.4.12. *Let $\boldsymbol{X}$ be a d-dimensional Bernoulli random vector with pmf $f \in \mathcal{B}_d(\boldsymbol{p})$ and pgf $\mathcal{P}_{\boldsymbol{X}}$. If $\boldsymbol{X}$ satisfies the strongly Rayleigh property, then:*

(a) The Bernoulli random vector $(X_{\sigma(1)}, \ldots, X_{\sigma(d)})$ satisfies the strongly Rayleigh property, for every permutation σ of $\{1, \ldots, d\}$.

(b) The sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ satisfies the strongly Rayleigh property.

(c) The strongly Rayleigh property is closed under marginalization, i.e., for any $k \in \{1, \ldots, d-1\}$ and for every $1 \leq j_1 < \cdots < j_k \leq d$, the Bernoulli random vector $(X_{j_1}, \ldots, X_{j_k}) \in \mathcal{B}_k(p_{j_1}, \ldots, p_{j_k})$ satisfies the strongly Rayleigh property.

(d) The random vector $(X_1, \ldots, X_d, Y)$, where $Y = n - S_{\boldsymbol{X}}$, is strongly Rayleigh.

(e) The Bernoulli random vector $(X_1, \dots, X_d, Y_1, \dots, Y_d)$ is strongly Rayleigh, where the random vector $(Y_1, \dots, Y_d)$ is an exchangeable Bernoulli random vector such that

$$\sum_{j=1}^{d} Y_j = d - \sum_{j=1}^{d} X_j.$$

(f) For any $n \in \mathbb{N}$, the Bernoulli random vector $(X_1K_1, \dots, X_1K_n, X_2, \dots, X_d)$ satisfies the strongly Rayleigh property, where $(K_1, \dots, K_n)$ is an exchangeable Bernoulli random vector independent of $\boldsymbol{X}$ such that $\mathbb{P}(K_1 + \cdots + K_n = 1) = 1$.

Proof. (a) It follows directly from the permutation property of Lemma 2.4.5.

(b) Since the pgf $\mathcal{P}_S$ of $S_{\boldsymbol{X}}$ is given by

$$\mathcal{P}_S(z) = \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z})\big|_{z_1=\cdots=z_d=z},$$

this statement follows by Item c) of Lemma 2.4.5.

(c) As recalled in Section 1.1.3, for every $k \in \{1, \dots, d-1\}$, the pgf of the marginal random vector $(X_1, \dots, X_k)$ is given by

$$\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z})\big|_{z_{k+1}=\cdots=z_d=1}.$$

Thus, by Items a) and d) of Lemma 2.4.5, the strongly Rayleigh property is closed under marginalization.

(d) The pgf of the random vector $(X_1, \dots, X_d, Y)$ with $Y = n - \sum_{j=1}^{d} X_j$ is the homogenization of $\mathcal{P}_{\boldsymbol{X}}$. Therefore, this statement follows by Proposition 2.4.9.

(e) The pgf of the random vector $(X_1, \dots, X_d, Y_1, \dots, Y_d)$ is the symmetric homogenization of $\mathcal{P}_{\boldsymbol{X}}$. Thus, this statement follows from Proposition 2.4.10.

(f) For every $h \in \{1, \dots, n\}$, let $X_{1h} = X_1K_h$. Since the Bernoulli random vector $\boldsymbol{K} = (K_1, \dots, K_n)$ is exchangeable and its components sum to 1 almost surely, its distribution is uniquely determined. Specifically, the distribution of $\boldsymbol{K}$ corresponds to the lower Fréchet bound within the class $\mathcal{B}_n(1/n)$, and its associated pgf is given by $\mathcal{P}_{\boldsymbol{K}} = \mathcal{E}_{n,1}$. Let $p_1 = \mathbb{E}[X_1]$ and $q_1 = 1 - p_1$. To compute the pgf $\tilde{\mathcal{P}}$ of $\tilde{\boldsymbol{X}}$, consider $\boldsymbol{z} = (z_{11}, \dots, z_{1n}, z_2, \dots, z_d)$. Then, we have

$$\begin{aligned}
\tilde{\mathcal{P}}(\boldsymbol{z}) &= \mathbb{E}[z_{11}^{X_1K_1} \cdots z_{1n}^{X_1K_n} z_2^{X_2} \cdots z_d^{X_d}] \\
&= \mathbb{E}[\mathbb{E}(z_{11}^{X_1K_1} \cdots z_{1n}^{X_1K_n} z_2^{X_2} \cdots z_d^{X_d} | X_1)] \\
&= q_1\mathbb{E}(z_2^{X_2} \cdots z_d^{X_d} | X_1 = 0) + p_1\mathbb{E}(z_{11}^{K_1} \cdots z_{1n}^{K_n} z_2^{X_2} \cdots z_d^{X_d} | X_1 = 1) \\
&= q_1\mathbb{E}(z_2^{X_2} \cdots z_d^{X_d} | X_1 = 0) + p_1\mathbb{E}(z_{11}^{K_1} \cdots z_{1n}^{K_n})\mathbb{E}(z_2^{X_2} \cdots z_d^{X_d} | X_1 = 1),
\end{aligned}$$

where the last equality follows from the independence of $\boldsymbol{K}$ and $\boldsymbol{X}$. By (1.8) and (1.9), we have

$$\tilde{\mathcal{P}}(\boldsymbol{z}) = \mathcal{P}_{\boldsymbol{X}}(z_1, \dots, z_d)\big|_{z_1=0} + \mathcal{E}_{n,1}(z_{11}, \dots, z_{1n})\, \partial_1\mathcal{P}_{\boldsymbol{X}}(z_1, \dots, z_d).$$

From (1.7), it follows that

$$\tilde{\mathcal{P}}(z_{11}, \dots, z_{1n}, z_2, \dots, z_d) = \mathcal{P}_{\boldsymbol{X}}(\mathcal{E}_{n,1}(z_{11}, \dots, z_{1n}), z_2, \dots, z_d),$$

and, by Proposition 2.4.8, the random vector $\tilde{\boldsymbol{X}}$ is strongly Rayleigh.

$\square$

Obviously, from the permutation property in Item (a) of Lemma 2.4.5, Item (f) of the above proposition holds for any index j, and not only for $j = 1$.

A fundamental question in this framework is whether the Strongly Rayleigh property can be simplified under the assumption of exchangeability. The strong connection between the class of exchangeable Bernoulli distributions and the class of univariate discrete distributions discussed in Section 1.3 is also underlined by the following theorem. It establishes that an exchangeable Bernoulli distribution is strongly Rayleigh if and only if its sum is strongly Rayleigh.

Theorem 2.4.13 ([15, Theorem 3.8])**.** *Let $\boldsymbol{X}^e$ be an exchangeable random vector with pmf $f^e \in \mathcal{E}_d$ and let $S^e = X_1^e + \dots + X_d^e$. Then, the pgf of $\boldsymbol{X}^e$ is real stable if and only if the pgf of S^e is real-rooted, i.e., $\boldsymbol{X}^e$ is strongly Rayleigh if and only if S^e is strongly Rayleigh.*

2.5 Main results

This section characterizes the class of Σ-countermonotonic Bernoulli random vectors and their relationship with the strongly Rayleigh property.

2.5.1 Characterization of Σ-countermonotonicity

In this section, Theorem 2.5.1 proves that Σ-countermonotonic Bernoulli random vectors are Σ_{cx}-smallest. This result is not obvious, even if Lemma 2.3.1 proves that every Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ admits a Σ_{cx}-smallest element. In fact, the following example exhibits a Fréchet class with a Σ_{cx}-smallest element and a Σ-countermonotonic Bernoulli random vector that is not Σ_{cx}-smallest, proving that in general the two notions do not coincide. See also Example 4.1.2 in Section 4.1.

Example 2.5.1. *Let us consider the Fréchet class $\mathcal{F}$ of 3-dimensional distributions with standard uniform continuous univariate marginals. The random vector $(U, U, 1-U)$, with $U \sim Unif(0,1)$, has distribution that belongs to $\mathcal{F}$, and it is easy to see that it enjoys the Σ-countermonotonicity property. However, it is not a Σ_{cx}-smallest element in $\mathcal{F}$, even if a Σ_{cx}-smallest element exists in the Fréchet class. A joint mix that belongs to $\mathcal{F}$ is constructed, e.g., in Example 3 in [50].*

This result is significant for two reasons. First, the two extensions of the lower Fréchet bound in higher dimensions lead to the same class of multidimensional Bernoulli distributions; second, this class is a convex polytope. Then, we prove that Σ-countermonotonic Bernoulli pmfs are not always strongly Rayleigh.

Theorem 2.5.1. *Let $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$. Then, the following statements are equivalent:*

1. *$\boldsymbol{X} \in \mathcal{B}_d^{cx}(\boldsymbol{p})$, i.e., $\boldsymbol{X}$ is Σ_{cx}-smallest in $\mathcal{B}_d(\boldsymbol{p})$.*

2. *$\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, i.e., $\boldsymbol{X}$ is Σ-countermonotonic.*

3. *X_h and $\sum_{j \neq h} X_j$ are countermonotonic, for every $h \in \{1, \ldots, d\}$.*

Proof. Theorem 3.8 of [87] states that $1 \Longrightarrow 2$, and it is clear that $2 \Longrightarrow 3$. We will prove that if $\boldsymbol{X} = (X_1, \ldots, X_d)$ is a Bernoulli random vector, then $3 \Longrightarrow 1$. Let $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$ be a Bernoulli random vector such that X_h and $\sum_{j \neq h} X_j$ are countermonotonic, for every $h \in \{1, \ldots, d\}$, and let $f_{\boldsymbol{X}}$ be its pmf. Let $p_j = \mathbb{E}[X_j]$, for $j \in \{1, \ldots, d\}$, and $|\boldsymbol{p}| = \sum_{j=1}^d p_j$. Suppose that the pmf of $\boldsymbol{X}$ is not a Σ_{cx}-smallest element, or, in other terms, that the pmf $s \in \mathcal{D}(|\boldsymbol{p}|)$ of the sum $S_{\boldsymbol{X}} = \sum_{j=1}^d X_j$ is different from $s_{|\boldsymbol{p}|}$ given in (2.8). Therefore, there exist $m_1, m_2 \in \{0,1, \ldots, d\}$, with $m_2 - m_1 \geq 2$, such that $s(m_1) > 0$ and $s(m_2) > 0$. This means that there exist $\boldsymbol{x}_1 \in \mathcal{X}_d^{m_1}$ and $\boldsymbol{x}_2 \in \mathcal{X}_d^{m_2}$ such that $f_{\boldsymbol{X}}(\boldsymbol{x}_1) > 0$ and $f_{\boldsymbol{X}}(\boldsymbol{x}_2) > 0$. Since $m_2 \geq m_1 + 2$, there exist $h^* \in \{1, \ldots, d\}$ such that $x_{1,h^*} = 0$ and $x_{2,h^*} = 1$. In the context of Definition 2.3.1, since X_{h^*} and $\sum_{j \neq h^*} X_j$ are countermonotonic, we have

$$\left(\sum_{j \neq h^*} X_j(\omega) - \sum_{j \neq h^*} X_j(\omega') \right) \Big(X_{h^*}(\omega) - X_{h^*}(\omega') \Big) \leq 0,$$

for $(\mathbb{P} \times \mathbb{P})$-almost every $(\omega, \omega') \in \Omega^2$. By considering $(\omega, \omega') \in \Omega^2$ such that $\boldsymbol{X}(\omega) = \boldsymbol{x}_1$ and $\boldsymbol{X}(\omega') = \boldsymbol{x}_2$, we have $(m_1 - (m_2 - 1))(0 - 1) \leq 0$, that implies $m_2 - m_1 \leq 1$. However, the latter contradicts the initial assumption that $m_2 - m_1 \geq 2$. This is absurd and therefore $\boldsymbol{X}$ is Σ_{cx}-smallest. $\square$

Remark 2.5.1. *Theorem 2.5.1 has a direct interest for applications. In fact, the third statement of Theorem 2.5.1 is the stopping condition of the rearrangement algorithm introduced in [86]. Since the three statements are equivalent in the Bernoulli framework, the rearrangement algorithm always returns a Σ_{cx}-smallest and Σ-countermonotonic pmf. Moreover, we can analytically find, without the algorithm, a Σ-countermonotonic Bernoulli random vector in any Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ by finding a Σ_{cx}-smallest element, and this can be done as in [6] and in [47]. We therefore have a perfect correspondence between extremal negative dependence and minimal aggregate risk. This has consequences, for example, in credit risk applications; indeed, we can find lower bounds for aggregate risk of credit portfolios (see [43]) and the corresponding best-case scenarios analytically.*

If $\boldsymbol{X}$ has pmf $f \in \mathcal{B}_d^{\Sigma}$, then there exists $m \in \{0,1, \ldots, d-1\}$ such that the corresponding pgf is of the form

$$\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d^m} f_{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}} + \sum_{\boldsymbol{x} \in \mathcal{X}_d^{m+1}} f_{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}}. \tag{2.12}$$

If $\boldsymbol{X}$ is a joint mix, then (2.12) simplifies to

$$\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d^m} f_{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}}, \tag{2.13}$$

for $m \in \{1, \ldots, d-1\}$.

We now characterize the geometrical structure of $\mathcal{B}_d^{\Sigma}$ and its intersection with the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, with $\boldsymbol{p} \in (0,1)^d$. Proposition 2.5.2 proves that the class $\mathcal{B}_d^{\Sigma}$ of pmfs of Σ-countermonotonic Bernoulli random vectors is the union of standard simplexes.

$\boldsymbol{i}$	(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)
$f_{\boldsymbol{I}}(\boldsymbol{i})$	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	0	0	0

Table 2.1: *Values of a pmf* $f_{\boldsymbol{I}} \in \mathcal{B}_3^{\Sigma}(2/5,2/5,2/5)$.

Proposition 2.5.2. *There is a one-to-one map between the class* $\mathcal{B}_d^{\Sigma}$ *and the union of the standard simplexes, i.e.,* $\mathcal{B}_d^{\Sigma} \longleftrightarrow \cup_{m=0}^{d} \Delta_{n_m,1}$*, where* $\Delta_{n_m,1}$ *is the standard* n_m*-simplex and* $n_m = \binom{d+1}{m+1}$*, for* $m \in \{0,\dots,d\}$*.*

Proof. From Theorem 2.5.1, if $\boldsymbol{X}$ is a Bernoulli random vector with pmf $f_{\boldsymbol{X}} \in \mathcal{B}_d^{\Sigma}$, then the pmf s of the sum of its components $S_{\boldsymbol{X}}$ has support on two consecutive points. Therefore, there exists $m \in \{0,\dots,d\}$ such that $f_{\boldsymbol{X}}$ has support on $\mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}$. The space of pmfs on $\mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}$ is in one-to-one correspondence with the standard simplex in dimension $\binom{d}{m} + \binom{d}{m+1} = \binom{d+1}{m+1}$. $\square$

The strength of dependence of two Bernoulli random vectors can be compared if their joint pmfs belong to the same Fréchet class $\mathcal{B}_d(\boldsymbol{p})$. From Theorem 2.5.1 and Proposition 2.3.2, we can prove that the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ is a convex polytope.

Corollary 2.5.3. *The class* $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ *is a convex polytope. Its extremal points are the extremal points of* $\mathcal{B}_d(\boldsymbol{p})$ *that are* Σ*-countermonotonic.*

Remark 2.5.2. *An important consequence of Corollary 2.5.3 for applications is that it enables the simulation of scenarios with extremal negative dependence using the technique introduced in [44]. In low dimensions, the extremal points of* $\mathcal{B}_d(\boldsymbol{p})$ *can be determined analytically for any given* $\boldsymbol{p} \in (0,1)^d \cap \mathbb{Q}^d$ *(see [46]). This provides an explicit characterization of the generators of* $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$*, and hence of all* Σ*-countermonotonic Bernoulli random vectors. In higher dimensions, however, identifying all extremal points becomes computationally challenging and remains only partially resolved; see, e.g., [44, 47]. A characterization of this class for the case of symmetric Bernoulli distributions will be provided in Section 3.1.*

We now study the link between $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ and the strongly Rayleigh property. The strongly Rayleigh property does not imply the Σ-countermonotonicity property; in fact, any vector of independent Bernoulli random variables satisfies the strongly Rayleigh property. The following counterexample shows that the reverse implication is not true either. In particular, Σ-countermonotonicity does not imply p-NC.

Example 2.5.2. *Consider a Bernoulli random vector* $\boldsymbol{X} = (X_1, X_2, X_3)$ *with* $f \in \mathcal{B}_3(\boldsymbol{p})$*, with* $\boldsymbol{p} = (2/5,\, 2/5,\, 2/5)$*, whose values are provided in Table 2.1. We observe that* $\boldsymbol{X}$ *is* Σ*-countermonotonic and satisfies the NLC. However,* $\boldsymbol{X}$ *does not satisfy the p-NC property, indeed* $Cov(X_1, X_2) = 1/25 > 0$*. It follows that* Σ*-countermonotonicity does not imply all the negative dependence properties stronger than p-NC, including NA and the strongly Rayleigh property.*

Example 2.5.2 shows that Σ-countermonotonic Bernoulli random vectors can have positively correlated pairs of random variables. This can be explained by observing that if

$\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$, the sum of all the cross moments of order two depends only on the distribution of the sum $S_{\boldsymbol{X}}$, i.e.,

$$\sum_{1 \leq j_1 < j_2 \leq d} \mathbb{E}[X_{j_1} X_{j_2}] = \sum_{y=2}^{d} \binom{y}{2} \mathbb{P}(S_{\boldsymbol{X}} = y) = \mathbb{E}\left[\frac{S_{\boldsymbol{X}}(S_{\boldsymbol{X}} - 1)}{2}\right]. \tag{2.14}$$

This expression takes the same value for every Bernoulli random vector $\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, that is,

$$\sum_{1 \leq j_1 < j_2 \leq d} \mathbb{E}[X_{j_1} X_{j_2}] = \sum_{y=2}^{d} \frac{y(y-1)}{2} s_{|\boldsymbol{p}|}(y) = \frac{1}{2} m(2|\boldsymbol{p}| - m - 1), \tag{2.15}$$

where m is the largest integer smaller than or equal to $|\boldsymbol{p}|$. Since the binomial coefficient in (2.14) is a convex function in y, the expression in (2.15) is the minimum value in the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$. Equations (2.14) and (2.15) also imply that the mean of the bivariate correlations is constant for every Bernoulli random vector $\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. For this reason, if $\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ has some countermonotonic pairs, it will have other pairs with higher (possibly positive) correlation, in order to keep the mean correlation equal to the constant value of the class. This discussion highlights the importance of studying negative dependence within the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. We conclude this section with the following result, which shows that the NLC property is implied both from the Σ-countermonotonicity and the strongly Rayleigh properties.

Proposition 2.5.4. *If $\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}$, then $\boldsymbol{X}$ has the NLC property.*

Proof. If $\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}$, then there exists $m \in \{0, \ldots, d-1\}$ such that $f_{\boldsymbol{X}}(\boldsymbol{x}) = 0$ for every $\boldsymbol{x} \notin \mathcal{X}_d^* = \mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}$. Let $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^d$ and let $\boldsymbol{w} = \boldsymbol{x} \wedge \boldsymbol{y}$ and $\boldsymbol{z} = \boldsymbol{x} \vee \boldsymbol{y}$. We prove that the condition in (2.7) is verified for every $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^d$. Without any restriction, we can assume $|\boldsymbol{x}| \leq |\boldsymbol{y}|$. We have $|\boldsymbol{w}| \leq |\boldsymbol{x}| \leq |\boldsymbol{y}| \leq |\boldsymbol{z}|$, and $\boldsymbol{w} \leq \boldsymbol{z}$, where $\leq$ denotes the usual partial order in $\mathbb{R}^d$. If $\boldsymbol{w} \notin \mathcal{X}_d^*$ or $\boldsymbol{z} \notin \mathcal{X}_d^*$, then $f(\boldsymbol{w}) f(\boldsymbol{z}) = 0$ and the condition in (2.7) is verified. Consider now $\boldsymbol{w}, \boldsymbol{z} \in \mathcal{X}_d^*$. If $|\boldsymbol{w}| = |\boldsymbol{z}|$, then $\boldsymbol{w} \leq \boldsymbol{z}$ implies $\boldsymbol{w} = \boldsymbol{z}$, and $\boldsymbol{w} = \boldsymbol{i} = \boldsymbol{k} = \boldsymbol{z}$; thus, the condition in (2.7) is verified. The condition is also verified if $|\boldsymbol{w}| = m$ and $|\boldsymbol{z}| = m+1$, since $\boldsymbol{w} \leq \boldsymbol{z}$ implies $\boldsymbol{w} = \boldsymbol{x}$ and $\boldsymbol{z} = \boldsymbol{y}$. □

We have verified that the Σ-countermonotonicity property and the strongly Rayleigh property are neither equivalent nor hierarchically related with proper counterexamples. In Figure 2.1, we summarize the many links between dependence properties, dependence orders, and the strongly Rayleigh property. In the next section, we study the strongly Rayleigh property within the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$.

2.5.2 Strongly Rayleigh and extremal negative dependence

This section discusses the strongly Rayleigh property within the convex polytope $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, with $\boldsymbol{p} \in (0,1)^d$, of Σ-countermonotonic Bernoulli random vectors. We have established that these two properties are not hierarchically ordered; thus, it becomes natural to explore their intersection. Since the elements of $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ are characterized by having a specific distribution for their sum $S_{\boldsymbol{X}}$, namely $r_{|\boldsymbol{p}|}^{D}$ in (2.8), and since the Strongly Rayleigh property of a random vector implies the same property for its sum (as noted in Proposition 2.4.12),

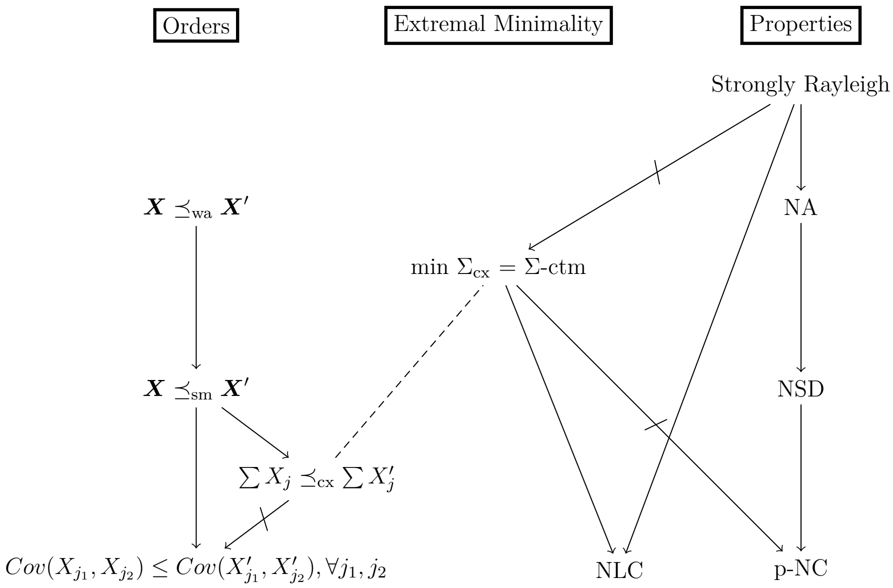

Figure 2.1: *Links among negative dependence notions and extremal negative dependence. The arrow $\rightarrow$ means strictly implies, and $\not\rightarrow$ means does not imply. The dashed line means that there is a link but not an implication.*

it is necessary to verify whether this univariate distribution is itself Strongly Rayleigh. The pgf corresponding to $r^D_{|\boldsymbol{p}|}$ is given by

$$\mathcal{P}^D_{|\boldsymbol{p}|}(z) = r^D_{|\boldsymbol{p}|}(m)\, z^m + r^D_{|\boldsymbol{p}|}(m+1)\, z^{m+1}, \tag{2.16}$$

or by

$$\mathcal{P}^D_{|\boldsymbol{p}|}(z) = r^D_{|\boldsymbol{p}|}(m)\, z^m. \tag{2.17}$$

In both cases, it has only real zeros, and thus $\mathcal{P}^D_{|\boldsymbol{p}|}$ is real stable. These considerations also show that the class $\mathcal{B}^{\Sigma}_d$ contains pmfs that satisfy the strongly Rayleigh property. The following proposition is a consequence of the previous discussion and Theorem 3.1.10.

Proposition 2.5.5. *If $\boldsymbol{X}^e$ is an exchangeable Bernoulli random vector with pmf $f^e \in \mathcal{B}^{\Sigma}_d$, then it satisfies the strongly Rayleigh property.*

Proposition 2.5.5 also implies that, for any $p \in (0,1)$, there is a pmf in $\mathcal{B}^{\Sigma}_d(p) := \mathcal{B}^{\Sigma}_d \cap \mathcal{B}_d(p)$ that satisfies the strongly Rayleigh property. Consider now $\mathcal{B}_d(\boldsymbol{p})$, and suppose that $|\boldsymbol{p}| \leq 1$. This implies $\mathcal{B}^{\Sigma}_d(\boldsymbol{p}) = \{f^C\}$, where $f^C \in \mathcal{B}_d(\boldsymbol{p})$ denotes the pmf of a pairwise countermonotonic Bernoulli random vector $\boldsymbol{X}^C$. The sum $S^C = X^C_1 + \cdots + X^C_d$ has distribution with support on $\{0,1\}$. Therefore, the pgf of $\boldsymbol{X}^C$ is a linear polynomial $\mathcal{P}^C(\boldsymbol{z}) = a_0 + \sum_{j=1}^d a_j z_j$, with $a_0 = f^C(\boldsymbol{0}_d)$ and $a_j = f^C(\boldsymbol{e}_j)$, for $j \in \{1, \dots, d\}$, where $\boldsymbol{0}_d$ is a vector of length d with all zeros and $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_d)$ is the canonical basis of $\mathbb{R}^d$. As mentioned in Section 2.4.1, linear polynomials are stable. Since the case $|\boldsymbol{p}| \geq d-1$ is similar, we conclude that any pairwise countermonotonic Bernoulli random vector $\boldsymbol{X}^C$ is strongly Rayleigh.

Example 2.5.3. *We present a simple example in the special case $|\boldsymbol{p}| \leq 1$, where the Σ-countermonotonic Bernoulli random vector is the pairwise countermonotonic Bernoulli random vector $\boldsymbol{X}^C$, with the lower Fréchet bound as distribution. Consider $\mathcal{B}_5(\boldsymbol{p})$, with $\boldsymbol{p} = (3/20,\ 1/20,\ 1/4,\ 3/10,\ 1/10)$; the pgf $\mathcal{P}^C$ of $\boldsymbol{X}^C$ is given by $\mathcal{P}^C(z_1, \dots, z_5) = (3 + 3z_1 + z_2 + 5z_3 + 6z_4 + 2z_5)/20$, and it is stable. Then, $\boldsymbol{X}^C$ is pairwise countermonotonic and satisfies the strongly Rayleigh property.*

The general case is more challenging as $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ is a convex polytope, meaning that there are several Σ-countermonotonic random Bernoulli pmfs in this class. We prove below that if $f \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ maximizes the entropy in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, then it satisfies the strongly Rayleigh property. We recall that the entropy $H(f)$ of a pmf $f \in \mathcal{B}_d$ is defined as

$$H(f) = -\sum_{\boldsymbol{x} \in \mathcal{X}_d} f(\boldsymbol{x}) \ln(f(\boldsymbol{x})). \tag{2.18}$$

If $\boldsymbol{X}$ has pmf f, we say that $\boldsymbol{X}$ has entropy $H(f)$. We prove that the distribution of $\boldsymbol{X}$ with maximal entropy pmf in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ belongs to the family of conditional Bernoulli distributions defined as follows. Consider a vector $\boldsymbol{K} = (K_1, \dots, K_d)$ of independent Bernoulli random variables with marginal means $\boldsymbol{\pi} = (\pi_1, \dots, \pi_d) \in (0,1)^d$.

1. For $m \in \{0, \dots, d\}$, let $\boldsymbol{X}^{(m)} = (X_1^{(m)}, \dots, X_d^{(m)})$ be a Bernoulli random vector with pmf $f_{\boldsymbol{\pi}}^{(m)}$ given by

$$f_{\boldsymbol{\pi}}^{(m)}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{K} = \boldsymbol{x} | K_1 + \dots K_d = m), \quad \boldsymbol{x} \in \mathcal{X}_d. \tag{2.19}$$

2. For $m \in \{0, \dots, d-1\}$, let $\boldsymbol{X}^{(m+)} = (X_1^{(m+)}, \dots, X_d^{(m+)})$ be a Bernoulli random vector with pmf $f_{\boldsymbol{\pi}}^{(m+)}$ given by

$$f_{\boldsymbol{\pi}}^{(m+)}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{K} = \boldsymbol{x} | m \leq K_1 + \dots K_d \leq m+1), \quad \boldsymbol{x} \in \mathcal{X}_d. \tag{2.20}$$

As in [22] (and references therein) or Example 5.1 (Conditioned Bernoulli sampling) of [54], we provide the following definition.

Definition 2.5.1. *A Bernoulli pmf $f \in \mathcal{B}_d$ that can be written in the form of (2.19) or (2.20) is called a* conditional Bernoulli pmf. *Any Bernoulli random vector $\boldsymbol{X}$ associated with a conditional Bernoulli pmf is referred to as a* conditional Bernoulli random vector.

Conditional Bernoulli distributions have been extensively studied in the framework of sampling with unequal probabilities without replacement. In the literature of this research field, they are also called conditional Poisson sampling; see Chapter 5 of [101]. By construction, it is clear that the Bernoulli random vector $\boldsymbol{X}^{(m)}$ is a joint mix and the Bernoulli random vector $\boldsymbol{X}^{(m+)}$ is Σ-countermonotonic.

Theorem 2.5.6. *For any $\boldsymbol{p} \in (0,1)^d$, there exists a unique pmf $f \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ that maximizes the entropy in the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ and it is the pmf of the unique conditional Bernoulli distribution in the class. Furthermore, this pmf satisfies the strongly Rayleigh property.*

Proof. We first prove that any conditional Bernoulli random vector satisfies the strongly Rayleigh property. Since $\boldsymbol{K} = (K_1, \dots, K_d)$ is a vector of independent Bernoulli random variables, it satisfies the strongly Rayleigh property. Therefore, from Corollary 4.18 of [15], the conditional Bernoulli random variables $\boldsymbol{X}^{(m)}$ and $\boldsymbol{X}^{(m+)}$ built from $\boldsymbol{K}$ satisfy the strongly Rayleigh property.

It is easily proved that a pmf $f_{\boldsymbol{\pi}}^{(m)}$ is the unique pmf that maximizes the entropy in the class $\mathcal{B}_d^{\Sigma}(q_1, \dots, q_d)$, where $q_j = \mathbb{E}[X_j^{(m)}]$, for $j \in \{1, \dots, d\}$ (see Section 1 of [23]). A straightforward generalization of this result shows that this also holds for a pmf $f_{\boldsymbol{\pi}}^{(m+)} \in \mathcal{B}_d^{\Sigma}(q_1', \dots, q_d')$, where $q_j' = \mathbb{E}[X_j^{(m+)}]$, for $j \in \{1, \dots, d\}$. It only remains to be proved that for every $\boldsymbol{p} \in (0,1)^d$, there is a unique conditional Bernoulli pmf in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ of the form of (2.19) or (2.20). Fix $\boldsymbol{p} \in (0,1)^d$. If $|\boldsymbol{p}| = m \in \{1, \dots, d-1\}$, Theorem 1 of [23] ensures that there exists $\boldsymbol{\pi} = (\pi_1, \dots, \pi_d) \in (0,1)^d$, unique up to rescaling the ratios $\frac{\pi_j}{1-\pi_j}$, and a vector $\boldsymbol{K} = (K_1, \dots, K_d)$ of independent Bernoulli random variables with marginal means $(\pi_1, \dots, \pi_d)$ such that $f_{\boldsymbol{\pi}}^{(m)}$ given by (2.19) is unique and $f_{\boldsymbol{\pi}}^{(m)} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. We prove that the same holds for $f_{\boldsymbol{\pi}}^{(m+)}$. A conditional Bernoulli pmf $f_{\boldsymbol{\pi}}^{(m+)}$ of the form of (2.20) can be written as

$$f_{\boldsymbol{\pi}}^{(m+)}(\boldsymbol{x}) = f_{\boldsymbol{\theta}}^{(m+)}(\boldsymbol{x}) = \exp\left\{\sum_{j=1}^{d} \theta_j x_j - \psi(\boldsymbol{\theta})\right\},$$

for $\boldsymbol{x} \in \mathcal{X}_d^* = \mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, with $\theta_j = \log \frac{\pi_j}{1-\pi_j}$, for $j \in \{1, \dots, d\}$, and $\psi(\boldsymbol{\theta}) = \log\left(\sum_{\boldsymbol{i} \in \mathcal{A}^*} \exp\{\sum_{j=1}^{d} \theta_j i_j\}\right)$. This shows that the class of conditional Bernoulli distributions belongs to the family of exponential distributions. The sufficient statistics $t_j(\boldsymbol{x}) = x_j$ are affinely independent, and since for every $\boldsymbol{\theta} \in \mathbb{R}^d$ we have

$$\sum_{\boldsymbol{x} \in \mathcal{X}_d} \exp\left\{\sum_{j=1}^{d} \theta_j x_j\right\} < +\infty,$$

it follows that the natural parameter space Θ coincides with $\mathbb{R}^d$. Thus, this class is a minimally represented and regular exponential family. Therefore, by Proposition A.0.1 in Appendix A, the map between the natural parameters $\theta_1, \dots, \theta_d$ and the marginal means $p_1, \dots, p_d$ is one-to-one. This implies that there exists a unique $\boldsymbol{\pi}^* = (\pi_1^*, \dots, \pi_d^*) \in (0,1)^d$ such that $f_{\boldsymbol{\pi}^*}^{(m+)} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. $\square$

Theorem 2.5.6 highlights that there is a Σ-countermonotonic Bernoulli random vector in $\mathcal{B}_d(\boldsymbol{p})$ that satisfies the strongly Rayleigh property. We already know that independence in $\mathcal{B}_d(\boldsymbol{p})$ satisfies the strongly Rayleigh property. Therefore, there are different levels of negative dependence in the same Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ that are strongly Rayleigh. Since two pmfs in $\mathcal{B}_d(\boldsymbol{p})$ are not comparable in $\trianglelefteq$, we do not have a way of comparing the strength of the strongly Rayleigh property within $\mathcal{B}_d(\boldsymbol{p})$, i.e., for given margins. The definition of a partial order allowing to compare the strength of the strongly Rayleigh property is an interesting open issue beyond the scope of this thesis.

Let us denote by f^H the pmf of the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ with the maximum entropy in Theorem 2.5.6 and by $\boldsymbol{X}^H = (X_1^H, \dots, X_d^H)$ the corresponding Bernoulli random vector. Sampling $\boldsymbol{X}^H$ fits within the framework of sampling with unequal probabilities in a maximum entropy design. The authors of [23] have suggested two draw-by-draw (or sequential) without replacement sampling procedures in the case where $|\boldsymbol{p}| = m \in \{1,\dots,d-1\}$. See [24] and Chapter 5 of [101] for additional details. The procedures are implemented in the `R` package called `sampling` [102]. However, in the case where $|\boldsymbol{p}| \in (m, m+1)$, with $m \in \{0,\dots,d-1\}$, one cannot directly apply the two procedures of [23]. To define a new sampling procedure that applies to the case $|\boldsymbol{p}|$ not integer, we first state the following theorem.

Theorem 2.5.7. *Let $\boldsymbol{X}$ be a Bernoulli random vector with pmf $f \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, where $\boldsymbol{p} \in (0,1)^d$ is such that $|\boldsymbol{p}| \in (m, m+1)$ with $m \in \{0,\dots,d-1\}$. Let $\boldsymbol{X}' = (X_1,\dots,X_d,X'_{d+1})$ where $X'_{d+1} = (m+1) - \sum_{j=1}^{d} X_j$. Then, $\boldsymbol{X}'$ is a joint mix with pmf $f' \in \mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$, where $\boldsymbol{p}' = (\boldsymbol{p}, p'_{d+1})$ and $p'_{d+1} = (m+1) - |\boldsymbol{p}|$. Moreover, $\boldsymbol{X}$ is the conditional Bernoulli random vector in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ with the maximal entropy distribution in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ if and only if $\boldsymbol{X}'$ is the conditional Bernoulli random vector in $\mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$ with the maximal entropy distribution in $\mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$.*

Proof. By construction, it easily follows that $\boldsymbol{X}'$ is a joint mix with pmf in the class $\mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$. Moreover, we have that $f(\boldsymbol{x}) = f'(\boldsymbol{x}//0)$ if $\boldsymbol{x} \in \mathcal{X}_d^{m+1}$, and $f(\boldsymbol{x}) = f'(\boldsymbol{x}//1)$ if $\boldsymbol{x} \in \mathcal{X}_d^{m}$. Since we can write

$$\mathcal{X}_{d+1}^{m+1} = \{(\boldsymbol{x},1) \colon \boldsymbol{x} \in \mathcal{X}_d^{m}\} \cup \{(\boldsymbol{x},0) \colon \boldsymbol{x} \in \mathcal{X}_d^{m+1}\},$$

we have $H(f) = H(f')$. Assume that f is the maximum entropy pmf in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Suppose also that there exists a Bernoulli random vector $\boldsymbol{Y}'$ with pmf $f_{\boldsymbol{Y}'} \in \mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$ such that $H(f_{\boldsymbol{Y}'}) > H(f')$, and let $\boldsymbol{Y} = (Y'_1,\dots,Y'_d)$. Then $\boldsymbol{Y}$ has pmf $f_{\boldsymbol{Y}} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ and $H(f_{\boldsymbol{Y}}) = H(f_{\boldsymbol{Y}'})$. This implies that $H(f_{\boldsymbol{Y}}) > H(f)$, which contradicts the assumption that f is the maximum entropy pmf in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Therefore, f' is the maximum entropy pmf in $\mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$. A similar argument shows that if f' is the maximum entropy pmf in $\mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}')$, then f is the maximum entropy pmf in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. □

Note that the construction of $\boldsymbol{X}'$ in Theorem 2.5.7 from $\boldsymbol{X}$ follows the procedure called homogenization in Section 2.4.1; this construction preserves the strongly Rayleigh property as shown in Proposition 2.4.12(d). Building on this result, when $|\boldsymbol{p}|$ is not integer, we can simulate samples of $\boldsymbol{X} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ in two steps. First, using Procedure 2 of [23], we simulate the jth sample $(X_1'^{(j)},\dots,X_d'^{(j)},X_{d+1}'^{(j)})$ of the Bernoulli random vector $\boldsymbol{X}'$ with pmf in $\mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}, p'_{d+1})$ with maximum entropy, defined as in Theorem 2.5.7. Then, we simulate the jth sample $(X_1^{(j)},\dots,X_d^{(j)})$ of $\boldsymbol{X}$ by letting $X_1^{(j)} = X_1'^{(j)},\dots,X_d^{(j)} = X_d'^{(j)}$.

Since every Fréchet class supports a Σ-countermonotonic Bernoulli random vector, Theorem 2.5.6 implies that every Fréchet class admits a Bernoulli random vector that satisfies both the strongly Rayleigh and the Σ-countermonotonicity property. We investigate below the construction of f^H as the pmf of a conditional Bernoulli random vector. We explicitly derive the expression of a conditional Bernoulli pmf in the case specified by (2.19). Let $\boldsymbol{K} = (K_1,\dots,K_d)$ be a vector of d independent Bernoulli random variables with $K_j \sim \text{Bern}(\pi_j)$, with $\pi_j \in (0,1)$, for $j \in \{1,\dots,d\}$. The pmf g of the Bernoulli random vector $\boldsymbol{K}$ is given by

$$g(\boldsymbol{x};\boldsymbol{\pi}) = \prod_{j=1}^{d} \pi_j^{x_j}(1-\pi_j)^{(1-x_j)},$$

for $\boldsymbol{x} \in \mathcal{X}_d$. Moreover, we denote by γ the univariate pmf of $S_{\boldsymbol{K}} = K_1 + \cdots + K_d$, that is, for $m \in \{0, \ldots, d\}$,

$$\gamma(m; \boldsymbol{\pi}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d^m} g(\boldsymbol{x}; \boldsymbol{\pi}).$$

Fix $m \in \{0, \ldots, d\}$. From (2.19), the values of the conditional Bernoulli pmf $f_{\boldsymbol{\pi}}^{(m)}$ of a conditional Bernoulli random vector $\boldsymbol{X}^{(m)}$ with the joint mixability property is given by

$$f_{\boldsymbol{\pi}}^{(m)}(\boldsymbol{x}) = \begin{cases} \frac{1}{\gamma(m;\boldsymbol{\pi})} g(\boldsymbol{x}; \boldsymbol{\pi}), & \boldsymbol{x} \in \mathcal{X}_d^m, \\ 0, & \boldsymbol{x} \notin \mathcal{X}_d^m. \end{cases} \tag{2.21}$$

The conditional Bernoulli pmf $f_{\boldsymbol{\pi}}^{(m)}$ that we obtain in (2.21) belongs to the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, for some $\boldsymbol{p} = (p_1, \ldots, p_d)$ such that $|\boldsymbol{p}| = m$. By Theorem 1 in [23], given an integer m, the map between the marginal means $\boldsymbol{\pi}$ of the vector $\boldsymbol{K}$ of independent Bernoulli random variables and the marginal means $\boldsymbol{p}$ of the conditional Bernoulli distribution is one-to-one up to rescaling. Given $\boldsymbol{\pi} \in (0,1)^d$, the values of $\boldsymbol{p}$ can be computed by evaluating the marginal means of the conditional Bernoulli pmfs in (2.21). However, for a fixed class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, finding the values of the marginal means $\boldsymbol{\pi}^*$ that generate the conditional Bernoulli that belongs to this class is not straightforward. Given any $\boldsymbol{p} \in (0,1)^d$, such that $|\boldsymbol{p}| = m$, we define the following maximization problem,

$$\boldsymbol{\pi}^* = \arg \max_{\boldsymbol{\pi} \in (0,1)^d} H(f_{\boldsymbol{\pi}}^{(m)}), \tag{2.22}$$

subject to the constraints

$$\sum_{\boldsymbol{x} \in \mathcal{X}_d} x_j f_{\boldsymbol{\pi}}^{(m)}(\boldsymbol{x}) = p_j, \quad j \in \{1, \ldots, d\}. \tag{2.23}$$

Notice that for $f_{\boldsymbol{\pi}}^{(m)} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, the expression of the entropy in (2.18) becomes

$$H(f_{\boldsymbol{\pi}}^{(m)}) = - \sum_{\boldsymbol{x} \in \mathcal{X}_d^m} f_{\boldsymbol{\pi}}^{(m)}(\boldsymbol{x}) \ln(f_{\boldsymbol{\pi}}^{(m)}(\boldsymbol{x})). \tag{2.24}$$

Consider now $\boldsymbol{p}$ such that $|\boldsymbol{p}| \in (m, m+1)$, for $m \in \{0, \ldots, d-1\}$. Using the methodology described above, we find $f_{\boldsymbol{\pi}^*}^{(m+1)} \in \mathcal{B}_{d+1}^{\Sigma}(\boldsymbol{p}, p_{d+1})$ with $p_{d+1} = m+1-|\boldsymbol{p}|$. By Theorem 2.5.7, we know that the conditional Bernoulli pmf in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ with maximal entropy in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ is the marginalization of $f_{\boldsymbol{\pi}^*}^{(m+1)}$.

Finally, since Corollary 2.5.3 proves that $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ is a convex polytope with $n_{\boldsymbol{p}}^{\Sigma}$ extremal points, we have also the following representation:

$$f^H(\boldsymbol{x}) = \sum_{k=1}^{n_{\boldsymbol{p}}^{\Sigma}} \alpha_k r_k(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{X}_d, \tag{2.25}$$

where r_k, for $k \in \{1, \ldots, n_{\boldsymbol{p}}^{\Sigma}\}$, are the extremal points of $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, and $\sum_{k=1}^{n_{\boldsymbol{p}}^{\Sigma}} \alpha_k = 1$, with $\alpha_k \geq 0$ for every $k \in \{1, \ldots, n_{\boldsymbol{p}}^{\Sigma}\}$. We illustrate this methodology in the following example.

$\boldsymbol{x}$	(1,1,0,0)	(1,0,1,0)	(0,1,1,0)	(1,0,0,1)	(0,1,0,1)	(0,0,1,1)	H
$r_1(\boldsymbol{x})$	0	0	0.3	0.35	0.15	0.2	1.3351
$r_2(\boldsymbol{x})$	0	0.3	0	0.05	0.45	0.2	1.1922
$r_3(\boldsymbol{x})$	0.3	0	0	0.05	0.15	0.5	1.1421
$f^H(\boldsymbol{x})$	0.0803	0.0927	0.1270	0.1770	0.2427	0.2803	1.6917

Table 2.2: *Values of the 3 extremal pmfs of $\mathcal{B}_4^{\Sigma}(\boldsymbol{p})$, where $\boldsymbol{p} = (\frac{7}{20}, \frac{9}{20}, \frac{1}{2}, \frac{7}{10})$, and values of $f^H \in \mathcal{B}_4^{\Sigma}(\boldsymbol{p})$.*

Example 2.5.4. *Consider the family $\mathcal{B}_4^{\Sigma}(\boldsymbol{p})$ with $\boldsymbol{p} = (7/20, 9/20, 1/2, 7/10)$. Given that $|\boldsymbol{p}| = 2$, the class $\mathcal{B}_4^{\Sigma}(\boldsymbol{p})$ is the set of joint pmfs of the vectors of four Bernoulli random variables, with marginal means $\boldsymbol{p}$, satisfying the joint mixability property, i.e., whose sum is equal to 2 with probability one. This convex polytope has three extremal points, denoted by r_1, r_2, and r_3, that are provided in Table 2.2. The pmf f^H of the Bernoulli random vector $\boldsymbol{X}^H$ satisfying both the joint mixability property and the strongly Rayleigh property is obtained with (2.21); it maximizes (2.24) under the constraints in (2.23). The values of f^H are provided in Table 2.2. Any pmf $f \in \mathcal{B}_4^{\Sigma}(\boldsymbol{p})$ admits the following representation:*

$$f(\boldsymbol{x}) = \alpha_1 r_1(\boldsymbol{x}) + \alpha_2 r_2(\boldsymbol{x}) + \alpha_3 r_3(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{A}_{d,m}, \tag{2.26}$$

where $\alpha_1, \alpha_2, \alpha_3 \geq 0$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1$. For f^H, the values of the coefficients in (2.26) are $(\alpha_1, \alpha_2, \alpha_3) = (0.4235, 0.3090, 0.2675)$.

The following proposition considers the case with $p_1 = \cdots = p_d = p$ leading to an explicit form of the pmf maximizing the entropy in (2.18), that is, the pmf of an exchangeable Bernoulli random vector in the class $\mathcal{B}_d^{\Sigma}(p)$, already proven to satisfy the strongly Rayleigh property in Proposition 2.5.5.

Proposition 2.5.8. *For every $p \in (0,1)$, the maximal entropy pmf $f^H \in \mathcal{B}_d^{\Sigma}(p)$ is the exchangeable pmf in $\mathcal{B}_d^{\Sigma}(p)$. In particular, if $pd = m \in \{1, \ldots, d-1\}$, for any $\boldsymbol{\pi} = (\pi, \ldots, \pi)$ with $\pi \in (0,1)$, we have $f^H = f_{\boldsymbol{\pi}}^{(m)}$, where*

$$f_{\boldsymbol{\pi}}^{(m)}(\boldsymbol{x}) = \begin{cases} \binom{d}{m}^{-1}, & \boldsymbol{x} \in \mathcal{X}_d^m, \\ 0, & \boldsymbol{x} \notin \mathcal{X}_d^m, \end{cases} \tag{2.27}$$

and, if $pd \in (m, m+1)$ for some $m \in \{0, \ldots, d-1\}$, there exists a unique $\boldsymbol{\pi} = (\pi, \ldots, \pi)$ with $\pi \in (0,1)$ such that $f^H = f_{\boldsymbol{\pi}}^{(m+)}$, where

$$f_{\boldsymbol{\pi}}^{(m+)}(\boldsymbol{x}) = \begin{cases} \frac{1-\pi}{\eta}, & \boldsymbol{x} \in \mathcal{X}_d^m, \\ \frac{\pi}{\eta}, & \boldsymbol{x} \in \mathcal{X}_d^{m+1}, \\ 0, & \boldsymbol{x} \notin (\mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}), \end{cases} \tag{2.28}$$

where $\eta = \binom{d}{m}(1-\pi) + \binom{d}{m+1}\pi$.

Proof. For every $\boldsymbol{\pi} = (\pi, \ldots, \pi)$ with $\pi \in (0,1)$, if $K_j \sim \text{Bern}(\pi)$, the expression for the pmfs $f_{\boldsymbol{\pi}}^{(m)}$ and $f_{\boldsymbol{\pi}}^{(m+)}$ in (2.27) and (2.28) follows from standard computations, implying that $f_{\boldsymbol{\pi}}^{(m)}$ and $f_{\boldsymbol{\pi}}^{(m+)}$ are exchangeable. Suppose that $pd = m \in \{1, \dots, d-1\}$. By Theorem 2.5.6, the maximal entropy pmf in the class of joint mixes $\mathcal{B}_d^{\Sigma}(p)$ is the pmf of a conditional Bernoulli distribution, i.e., $f^H = f_{\boldsymbol{\pi}}^{(m)}$, for some $\boldsymbol{\pi} \in (0,1)^d$, and, by Lemma 1 in [23], we have that $\boldsymbol{\pi} = (\pi, \dots, \pi)$ for some $\pi \in (0,1)$. Moreover, from (2.27), we have that $f_{\boldsymbol{\pi}}^{(m)}$, with $\boldsymbol{\pi} = (\pi, \dots, \pi)$ does not depend on $\pi \in (0,1)$. Therefore, $f^H = f_{\boldsymbol{\pi}}^{(m)}$, for every $\boldsymbol{\pi} = (\pi, \dots, \pi)$ with $\pi \in (0,1)$ (see also Theorem 1 in [23]). We now consider the case $pd \in (m, m+1)$, for some $m \in \{0, \dots, d-1\}$. As shown in the proof of Theorem 2.5.6, the map between $\boldsymbol{\pi} = (\pi_1, \dots, \pi_d) \in (0,1)^d$ and $\boldsymbol{p} = (p_1, \dots, p_d)$ is one-to-one. We have seen from (2.28) that if $\boldsymbol{\pi} = (\pi, \dots, \pi)$, the conditional Bernoulli pmf $f_{\boldsymbol{\pi}}^{(m+)}$ is exchangeable and has all marginal means equal. Therefore, the marginal means of $f_{\boldsymbol{\pi}}^{(m+)}$ are all equal if and only if $\pi_1 = \dots = \pi_d = \pi$. We can conclude that the maximum entropy distribution in $\mathcal{B}_d^{\Sigma}(p)$ is $f^H = f_{\boldsymbol{\pi}}^{(m+)}$, for a unique $\boldsymbol{\pi} = (\pi, \dots, \pi) \in (0,1)^d$. □

The following corollary characterizes the pgf associated to the maximal entropy pmf f^H in terms of its pgf in (2.12) and (2.13).

Corollary 2.5.9. *Let us consider the maximal entropy pmf f^H in $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Then, there exists $\boldsymbol{a} = (a_1, \dots, a_d)$, with $a_j > 0$, such that the pgf is given by*

$$\mathcal{P}^H(\boldsymbol{z}) \propto \sum_{\boldsymbol{i} \in \mathcal{X}_d^*} \boldsymbol{a}^{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}, \tag{2.29}$$

where $\mathcal{X}_d^* = \mathcal{X}_d^{|\boldsymbol{p}|}$ if $|\boldsymbol{p}|$ is integer, or $\mathcal{X}_d^* = \mathcal{X}_d^{m} \cup \mathcal{X}_d^{m+1}$ if $|\boldsymbol{p}| \in (m, m+1)$, for $m \in \{0, \dots, d-1\}$.

Proof. From Theorem 2.5.6, there exists a vector $\boldsymbol{K} = (K_1, \dots, K_d)$ of d independent Bernoulli random variables with marginal means $\boldsymbol{\pi} = (\pi_1, \dots, \pi_d)$ such that f^H is given by (2.19) or by (2.20). The pgf of $\boldsymbol{K}$ is given by

$$\mathcal{P}^{\perp}(\boldsymbol{z}) = \prod_{j=1}^{d} (\pi_j z_j + 1 - \pi_j) = \prod_{j=1}^{d} (1 - \pi_j) \sum_{\boldsymbol{x} \in \mathcal{X}_d} \boldsymbol{a}^{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}},$$

where $\boldsymbol{a} = (a_1, \dots, a_d)$, with $a_j = \frac{\pi_j}{1-\pi_j}$ for $j \in \{1, \dots, d\}$. By (1.11), the pgf $\mathcal{P}^H$ of f^H is proportional to the terms of $\mathcal{P}^{\perp}$ of order m if $|\boldsymbol{p}| = m$, or of orders m and $m+1$ if $|\boldsymbol{p}| \in (m, m+1)$. Moreover, $\mathcal{P}^H$ is normalized such that the sum of the coefficients is one. Therefore, we have (2.29), with $a_j = \frac{\pi_j}{1-\pi_j}$, for $j \in \{1, \dots, d\}$. □

Example 2.5.5. *Let $a_1, a_2, a_3 > 0$. The polynomial $\mathcal{P}^{\perp}$, that is such that*

$$\begin{aligned}
\mathcal{P}^{\perp}(z_1, z_2, z_3) &\propto (a_1 z_1 + 1)(a_2 z_2 + 1)(a_3 z_3 + 1) \\
&= 1 + a_1 z_1 + a_2 z_2 + a_1 a_2 z_1 z_2 + a_3 z_3 + a_1 a_3 z_1 z_3 + a_2 a_3 z_2 z_3 \\
&\quad + a_1 a_2 a_3 z_1 z_2 z_3 \\
&= \sum_{\boldsymbol{x} \in \mathcal{X}_3} \boldsymbol{a}^{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}},
\end{aligned}$$

is the pgf of a vector of independent Bernoulli random variables. Consider the Fréchet class $\mathcal{B}_3(p_1, p_2, p_3)$ with $p_1 + p_2 + p_3 \in (1,2)$. By Theorem 2.5.6, the maximum entropy *pmf* $f^H \in \mathcal{B}_3^\Sigma(p_1, p_2, p_3)$ *is the conditional Bernoulli pmf* $f^{(1+)}_{\boldsymbol{\pi}^*}$, *with* $\boldsymbol{\pi}^*$ *the solution to the optimization problem in (2.22) subject to the three constraints specified by (2.23). Let* $a_j^* = \frac{\pi_j^*}{1-\pi_j^*}$, *for* $j \in \{1, 2, 3\}$. *Then, the pgf associated to* f^H *is given by*

$$\mathcal{P}^H(\boldsymbol{z}) \propto a_1^* z_1 + a_2^* z_2 + a_1^* a_2^* z_1 z_2 + a_3^* z_3 + a_1^* a_3^* z_1 z_3 + a_2^* a_3^* z_2 z_3 = \sum_{\boldsymbol{x} \in \mathcal{X}_d^*} (\boldsymbol{a}^*)^{\boldsymbol{x}} \boldsymbol{z}^{\boldsymbol{x}},$$

where $\mathcal{X}_d^* = \mathcal{X}_3^1 \cup \mathcal{X}_3^2$.

Remark 2.5.3. *The pgf* $\mathcal{P}^H$ *associated to the maximum entropy pmf* f^H *is such that* $\mathcal{P}^H(\frac{z_1}{a_1}, \ldots, \frac{z_d}{a_d}) \propto \sum_{\boldsymbol{x} \in \mathcal{X}_d^*} \boldsymbol{z}^{\boldsymbol{x}} = \mathcal{E}_{d,|\boldsymbol{p}|}$, *where* $\mathcal{E}_{d,|\boldsymbol{p}|} = \mathcal{E}_{d,|\boldsymbol{p}|}$ *if* $|\boldsymbol{p}|$ *is integer, otherwise* $\mathcal{E}_{d,|\boldsymbol{p}|} = \mathcal{E}_{d,m} + \mathcal{E}_{d,m+1}$ *if* $|\boldsymbol{p}| \in (m, m+1)$, *and where* $\mathcal{E}_{d,m}$ *is the elementary symmetric polynomial of order* m *given in (1.28). This is another way to show that* f^H *satisfies the strongly Rayleigh property. As recalled in Section 2.4.1,* $\mathcal{E}_{d,m}$ *is real stable for every* $m \in \{0, \ldots, d\}$. *Moreover, by Theorem 2.4.3,* $\mathcal{E}_{d,m} + \mathcal{E}_{d,m+1}$ *is real stable because* $z_{d+1}\mathcal{E}_{d,m} + \mathcal{E}_{d,m+1} = \mathcal{E}_{d+1,m+1}$ *is real stable. Therefore,* $\mathcal{E}_{|\boldsymbol{p}|}$ *is a real stable polynomial for every* $|\boldsymbol{p}| \in (0, d)$, *and from Item b) of Lemma 2.4.5, the polynomial* $\mathcal{P}^H$ *is real stable.*

In general, the maximal entropy pmf in $\mathcal{B}_d^\Sigma(\boldsymbol{p})$ is not the unique pmf in $\mathcal{B}_d^\Sigma(\boldsymbol{p})$ that satisfies the strongly Rayleigh property. Beyond Theorem 2.5.6, there are other approaches to find Bernoulli random vectors that enjoy both the strongly Rayleigh and the Σ-countermonotonicity properties. The following example shows that it is possible to find strongly Rayleigh pmfs in $\mathcal{B}_d^\Sigma(\boldsymbol{p})$ by slightly modifying f^H, using its representation as convex combination of extremal points in (2.25).

Example 2.5.6. *Consider the class* $\mathcal{B}_3^\Sigma(\frac{2}{5})$ *for which the extremal points* r_k, *with* $k \in \{1, 2, 3\}$ *are reported in Table 2.3. The maximum entropy pmf* $f^H \in \mathcal{B}_3^\Sigma(\frac{2}{5})$ *is the exchangeable pmf in Table 2.3, and it is such that*

$$f^H(\boldsymbol{x}) = \frac{1}{3} r_1(\boldsymbol{x}) + \frac{1}{3} r_2(\boldsymbol{x}) + \frac{1}{3} r_3(\boldsymbol{x}), \tag{2.30}$$

for all $\boldsymbol{x} \in \mathcal{X}_3$. *To find another pmf in the class* $\mathcal{B}_3^\Sigma(\frac{2}{5})$ *that satisfies the strongly Rayleigh property, we slightly perturb the coefficient of the convex linear combination in (2.30). Consider the pmf,* $\tilde{f} \in \mathcal{B}_3^\Sigma(\frac{2}{5})$ *with*

$$\tilde{f}(\boldsymbol{x}) = \sum_{k=1}^{3} \alpha_k r_k(\boldsymbol{x}),$$

for $\boldsymbol{x} \in \mathcal{X}_3$, *and* $\alpha_1 = 0.333$, $\alpha_2 = 0.335$, $\alpha_3 = 0.332$ *given in Table 2.3. The pgf associated to* $\tilde{f}$ *is*

$$\tilde{\mathcal{P}}(\boldsymbol{z}) = 0.2664 z_1 + 0.267 z_2 + 0.2666 z_3 + 0.0666 z_1 z_2 + 0.067 z_1 z_3 + 0.0664 z_2 z_3,$$

and it satisfies the conditions for the strongly Rayleigh property in (2.9) for all $j_1, j_2 \in \{1, 2, 3\}$ *and* $\boldsymbol{x} \in \mathbb{R}^3$. *Therefore,* $\tilde{f}$ *is strongly Rayleigh and* Σ*-countermonotonic, but it does not have the maximum entropy in the class* $\mathcal{B}_3^\Sigma(\frac{2}{5})$.

$\boldsymbol{x}$	$r_1(\boldsymbol{x})$	$r_2(\boldsymbol{x})$	$r_3(\boldsymbol{x})$	$f^H(\boldsymbol{x})$	$\tilde{f}(\boldsymbol{x})$
(0,0,0)	0	0	0	0	0
(1,0,0)	0.2	0.2	0.4	$\frac{4}{15}$	0.2664
(0,1,0)	0.2	0.4	0.2	$\frac{4}{15}$	0.267
(1,1,0)	0.2	0	0	$\frac{1}{15}$	0.0666
(0,0,1)	0.4	0.2	0.2	$\frac{4}{15}$	0.2666
(1,0,1)	0	0.2	0	$\frac{1}{15}$	0.067
(0,1,1)	0	0	0.2	$\frac{1}{15}$	0.0664
(1,1,1)	0	0	0	0	0
Entropy	1.33217904	1.33217904	1.33217904	1.599014712	1.599012963

Table 2.3: *r_1, r_2, and r_3 are the extremal points of $\mathcal{B}_3^{\Sigma}(\frac{2}{5})$, f^H is the maximum entropy pmf in $\mathcal{B}_3^{\Sigma}(\frac{2}{5})$, and $\tilde{f} \in \mathcal{B}_3^{\Sigma}(\frac{2}{5})$ is strongly Rayleigh and Σ-countermonotonic.*

A second approach is based on the polarization of a stable polynomial, but it requires some conditions on the Fréchet class. In particular, its proof relies on Proposition 2.4.8.

Proposition 2.5.10. *Consider $\boldsymbol{p} \in (0,1)^d$ and suppose that there exists a partition of $\{1, \dots, d\}$, denoted $(\Lambda_1, \dots, \Lambda_D)$, with cardinality $(\lambda_1, \dots, \lambda_D)$ such that we have $p_j = \tilde{p}_h \leq \frac{1}{\lambda_h}$ for every $j \in \Lambda_h$, for all $h \in \{1, \dots, D\}$. For every $h \in \{1, \dots, D\}$, let $\boldsymbol{K}_h \in \mathcal{B}_{\lambda_h}(\frac{1}{\lambda_h})$ be a Bernoulli random vector with joint cdf being equal to the lower Fréchet bound of $\mathcal{B}_{\lambda_h}(\frac{1}{\lambda_h})$. Let $\boldsymbol{X}^H$ be the maximum entropy D-dimensional Bernoulli random vector of $\mathcal{B}_D^{\Sigma}(\lambda_1\tilde{p}_1, \dots, \lambda_D\tilde{p}_D)$ and assume that $\boldsymbol{K}_1, \dots, \boldsymbol{K}_D$, and $\boldsymbol{X}^H$ are independent. Then, the Bernoulli random vector $\boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_D)$, where $\boldsymbol{Y}_h = (Y_{h,1}, \dots, Y_{h,\lambda_h})$ and $Y_{h,j} = X_h^H K_{h,j}$, for every $h \in \{1, \dots, D\}$ and $j \in \{1, \dots, \lambda_h\}$, has distribution in the class $\mathcal{B}_d(\boldsymbol{p})$ and satisfies the strongly Rayleigh and the Σ-countermonotonicity properties.*

Proof. The Fréchet class $\mathcal{B}_{\lambda_h}(\frac{1}{\lambda_h})$ always admits the lower Fréchet bound as a distribution because the sum of the marginal means is equal to 1. Since $\mathbb{P}(\sum_{l=1}^{\lambda_h} K_{h,l} = 1) = 1$, we have $\mathbb{P}(\sum_{l=1}^{\lambda_h} Y_{h,l} = X_h^H) = 1$ for every $h \in \{1, \dots, D\}$, and $\mathbb{P}(\sum_{j=1}^{d} Y_j = \sum_{h=1}^{D} X_h^H) = 1$. Therefore, since $p_1 + \dots + p_d = \lambda_1\tilde{p}_1 + \dots + \lambda_D\tilde{p}_D$, the Σ-countermonotonicity property of $\boldsymbol{X}^H$ implies that also $\boldsymbol{Y}$ is Σ-countermonotonic. By iteratively applying Proposition 2.4.8, it follows that $\boldsymbol{Y}$ satisfies the strongly Rayleigh property. □

We notice that for every Fréchet class $\mathcal{B}_d(p)$ with $p \leq \frac{1}{2}$ and d even, the construction in Proposition 2.5.10 is always feasible. We present the details of this construction in Example 2.5.7.

Example 2.5.7. *Let $\boldsymbol{X}^H = (X_1^H, \dots, X_D^H)$ be a Σ-countermonotonic Bernoulli random vector with the maximum entropy pmf $f^H \in \mathcal{B}_D^{\Sigma}(p')$, with $p' \in (0,1)$. By Theorem 2.5.6, it also satisfies the strongly Rayleigh property. Let $\boldsymbol{K}_h = (K_{h,1}, K_{h,2})$ be a countermonotonic Bernoulli random vector with marginal means equal to $1/2$, for every $h \in \{1, \dots, D\}$. Assume that $\boldsymbol{K}_1, \dots, \boldsymbol{K}_D$, and $\boldsymbol{X}^H$ are independent. Define $\boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_D)$, where $\boldsymbol{Y}_h = (Y_{h,1}, Y_{h,2})$ and $Y_{h,j} = X_h^H K_{h,j}$, for $h \in \{1, \dots, D\}$ and $j \in \{1, 2\}$. We have*

$\boldsymbol{Y} \in \mathcal{B}_d^{\Sigma}(p)$, where $d = 2D$ and $p = p'/2$, and the pgf of $\boldsymbol{Y}$ is given by

$$\mathcal{P}_{\boldsymbol{Y}}(z_1, \dots, z_d) = \mathbb{E}\Big[\prod_{j=1}^{d} z_j^{Y_j}\Big] = \mathbb{E}\Big[\prod_{h=1}^{D} z_{2h-1}^{X_h^H K_{h,1}} z_{2h}^{X_h^H K_{h,2}}\Big].$$

Conditioning on $\boldsymbol{X}^H$, and given the assumption of independence, we obtain

$$\begin{aligned}\mathcal{P}_{\boldsymbol{Y}}(z_1, \dots, z_d) &= \mathbb{E}\Big[\mathbb{E}\Big[\prod_{h=1}^{D} z_{2h-1}^{X_h^H K_{h,1}} z_{2h}^{X_h^H K_{h,2}}\Big|(X_1^H, \dots, X_D^H)\Big]\Big] \\ &= \mathbb{E}\Big[\prod_{h=1}^{D} \mathbb{E}\Big[z_{2h-1}^{K_{h,1}} z_{2h}^{K_{h,2}}\Big]^{X_h^H}\Big] \\ &= \mathcal{P}^H(\mathcal{P}_{\boldsymbol{K}_1}(z_1, z_2), \dots, \mathcal{P}_{\boldsymbol{K}_D}(z_{d-1}, z_d)).\end{aligned}$$

From Proposition 2.5.8, we can derive the expression for the pgf of $\boldsymbol{I}^H$ in the following two cases. If $p'D$ is integer, we have

$$\mathcal{P}^H(z_1, \dots, z_D) = \frac{1}{\binom{D}{p'D}} \mathcal{E}_{d,p'D}(z_1, \dots, z_D).$$

If $p'D \in (m, m+1)$, for some $m \in \{0, \dots, D-1\}$, the expression of $\mathcal{P}^H$ is

$$\mathcal{P}^H(z_1, \dots, z_D) = \frac{(1-\pi)\mathcal{E}_{D,m}(z_1, \dots, z_D) + \pi \mathcal{E}_{D,m+1}(z_1, \dots, z_D)}{\binom{D}{m}(1-\pi) + \binom{D}{m+1}} \pi$$

where $\pi \in (0,1)$ is the value of the marginal means of the Bernoulli random vector $\boldsymbol{K}$ with independent marginals such that f^H is given by (2.20). Therefore, since $p'D = pd$, we find

$$\mathcal{P}_{\boldsymbol{Y}}(z_1, \dots, z_d) = \frac{1}{\binom{d/2}{pd}} \mathcal{E}_{d/2,pd}\left(\frac{1}{2}(z_1 + z_2), \dots, \frac{1}{2}(z_{d-1} + z_d)\right),$$

if pd is integer, and

$$\begin{aligned}\mathcal{P}_{\boldsymbol{Y}}(z_1, \dots, z_d) =& \frac{1-\pi}{\binom{d/2}{m}(1-\pi) + \binom{d/2}{m+1}} \pi \mathcal{E}_{d/2,m}\left(\frac{1}{2}(z_1 + z_2), \dots, \frac{1}{2}(z_{d-1} + z_d)\right) \\ &+ \frac{\pi}{\binom{d/2}{m}(1-\pi) + \binom{d/2}{m+1}\pi} \mathcal{E}_{d/2,m+1}\left(\frac{1}{2}(z_1 + z_2), \dots, \frac{1}{2}(z_{d-1} + z_d)\right),\end{aligned}$$

if $pd \in (m, m+1)$. In both cases, by Proposition 2.5.10, $\mathcal{P}_{\boldsymbol{Y}}$ is a real stable polynomial. $\mathcal{P}_{\boldsymbol{Y}}$ is not symmetric and $\boldsymbol{Y}$ is not exchangeable, thus, by Proposition 2.5.8, its pmf $f_{\boldsymbol{Y}}$ does not have the maximal entropy in $\mathcal{B}_d^{\Sigma}(p)$.

2.5.3 Dependence ordering

Since two Bernoulli random vectors $\boldsymbol{X}$ and $\boldsymbol{X}'$ in the same Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ are not comparable under $\trianglelefteq$, we study the link between the strongly Rayleigh property and the dependence orders $\preceq_{\text{wa}}$ and $\preceq_{\text{sm}}$. Proposition 2.5.11 and Corollary 2.5.12 consider the simpler case of equal means, i.e., the class $\mathcal{B}_d^{\Sigma}(p)$.

Proposition 2.5.11. *The maximum entropy pmf $f^H \in \mathcal{B}_d^\Sigma(p)$ that satisfies the strongly Rayleigh property is minimal under the supermodular order in the class of exchangeable Bernoulli pmfs with mean p.*

Proof. The maximum entropy pmf $f^H \in \mathcal{B}_d^\Sigma(p)$ is exchangeable. By Proposition 5 in [49] it is minimal under the supermodular order in the class of exchangeable pmfs in $\mathcal{B}_d(p)$. $\square$

The strongly Rayleigh property is closed under marginalization (see Item (c) of Proposition 2.4.12), but the Σ-countermonotonicity property is not in general. We provide an example of a Bernoulli random vector that satisfies the strongly Rayleigh property, that is not Σ-countermonotonic, and that is different from independence. The following corollary states that by marginalization of the d-dimensional maximum entropy pmf, $f_d^H \in \mathcal{B}_d^\Sigma(p)$, we obtain a k-dimensional pmf $f_{d|k}^H \in \mathcal{B}_k(p)$ that satisfies the strongly Rayleigh property but that is not Σ-countermonotonic and greater under the supermodular order than the k-dimensional maximum entropy pmf $f_k^H \in \mathcal{B}_k^\Sigma(p)$.

Corollary 2.5.12. *Let f_d^H be the maximum entropy pmf in the d-dimensional class $\mathcal{B}_d^\Sigma(p)$. Then, for $k \leq d$, the k-dimensional marginal pmf $f_{d|k}^H \in \mathcal{B}_k(p)$ of f_d^H satisfies the strongly Rayleigh property and $f_k^H \preceq_{sm} f_{d|k}^H$, where $f_k^H \in \mathcal{B}_k^\Sigma(p)$ is the maximum entropy pmf in $\mathcal{B}_k^\Sigma(p)$.*

Proof. If $pd \leq 1$ or $pd \geq d-1$, the result is trivial since the lower Fréchet bound is closed under marginalization. Otherwise, Item (c) of Proposition 2.4.12 implies that $f_{d|k}^H \in \mathcal{B}_k(p)$ satisfies the strongly Rayleigh property. By Proposition 2.5.8, the pmf f_d^H is exchangeable, thus $f_{d|k}^H \in \mathcal{B}_k(p)$ is also exchangeable. By Proposition 2.5.11, we have $f_k^H \preceq_{\mathrm{sm}} f_{d|k}^H$. $\square$

Corollary 2.5.12 proves that for any dimension d we can find a pmf in $\mathcal{B}_d(p) \setminus \mathcal{B}_d^\Sigma(p)$ that satisfies the strongly Rayleigh property and it is not the independence pmf. We therefore have at least three different levels of strongly Rayleigh dependence: independence, extremal negative dependence and at least one pmf in between. This result supports the idea that it should be possible to find a partial order, stronger than the weak association order, to compare the strongly Rayleigh property within a Fréchet class $\mathcal{B}_d(p)$.

The following theorem proves that the class $\mathcal{B}_d^\Sigma(\boldsymbol{p})$ verifies a minimality property with respect to the supermodular order and, therefore, also with respect to the weak association order. We recall that an antichain in a partially ordered set $(\mathcal{X}, \preceq_*)$ is a subset $A \subseteq \mathcal{X}$ such that any two distinct elements in A are not comparable according to $\preceq_*$. A chain is a subset $C \subseteq \mathcal{X}$ such that if $x, y \in C$ then $x \preceq_* y$ or $y \preceq_* x$ (see Chapter 12 in [106]).

Theorem 2.5.13. *The following holds true:*

1. *The class $\mathcal{B}_d^\Sigma(\boldsymbol{p})$ is an antichain in the partially ordered set $(\mathcal{B}_d(\boldsymbol{p}), \preceq_{sm})$.*

2. *For any Bernoulli pmf $f \in \mathcal{B}_d(\boldsymbol{p}) \setminus \mathcal{B}_d^\Sigma(\boldsymbol{p})$, either there exists $f^\Sigma \in \mathcal{B}_d^\Sigma(\boldsymbol{p})$ such that $f^\Sigma \preceq_{sm} f$ or f is not comparable with any f^Σ in the supermodular order.*

Proof. To prove Item 1, consider two Bernoulli pmfs $f, f' \in \mathcal{B}_d^\Sigma(\boldsymbol{p})$, and suppose $f \neq f'$. By Theorem 2.5.4 in [76], $f \preceq_{\mathrm{sm}} f'$ if and only if f' can be obtained from f by a finite number of supermodular transfers. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}_d$, with $|\boldsymbol{x}| \leq |\boldsymbol{y}|$, and define $\boldsymbol{w} = \boldsymbol{x} \wedge \boldsymbol{y}$ and $\boldsymbol{z} = \boldsymbol{x} \vee \boldsymbol{y}$. A simple supermodular transfer is a transfer given by $\eta(\frac{1}{2}\delta_{\boldsymbol{x}} + \frac{1}{2}\delta_{\boldsymbol{y}}) \rightarrow \eta(\frac{1}{2}\delta_{\boldsymbol{w}} + \frac{1}{2}\delta_{\boldsymbol{z}})$, where

$\delta_{\boldsymbol{x}}$ denotes the point mass in $\boldsymbol{x} \in \mathcal{X}_d$. If $f, f' \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$, they have support contained on $\mathcal{X}^* = \mathcal{X}_d^m \cup \mathcal{X}_d^{m+1}$. Thus, since we must have $\boldsymbol{w}, \boldsymbol{z} \in \mathcal{X}^*$, and $\boldsymbol{w} \leq \boldsymbol{z}$, it follows that $\boldsymbol{w} = \boldsymbol{x}$ and $\boldsymbol{z} = \boldsymbol{y}$. Therefore, the only admissible supermodular transfer is the identity. Item 2 is a direct consequence of Theorem 2.5.1. In fact, the class of Σ-countermonotonic pmfs is the class of Σ_{cx}-smallest pmfs defined by minimizing a class of supermodular functions. $\square$

Remark 2.5.4. *Notice that Theorem 2.5.13 cannot be improved. In fact, there exists $f \in \mathcal{B}_d(\boldsymbol{p}) \setminus \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ such that f is not comparable with any $f^{\Sigma} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Indeed, the author of [49] proposes an algorithm to find sequences of Bernoulli random vectors decreasing in supermodular order, and in Example 3 of [49], the algorithm stops at a Bernoulli pmf that is not Σ_{cx}-smallest. The proposed algorithm considers all supermodular transfers of Theorem 2.5.4 of [76]. Thus, the algorithm stops at a pmf that is not comparable with any $f^{\Sigma} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$.*

From Theorem 2.5.13 it follows that the Bernoulli random vector $\boldsymbol{Y} \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$ in Proposition 2.5.10, that satisfies the strongly Rayleigh property, is not comparable in supermodular order with the maximum entropy Bernoulli random vector $\boldsymbol{X}^H \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Using f^H and $f_{\boldsymbol{p}}^{\perp}$ within a given Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, we can construct a parametric family of Bernoulli random vectors ordered in supermodular order, from a Σ-countermonotonic pmf to independence. As we have seen in Example 2.5.2, since the Σ-countermonotonicity property does not imply negative association, the inequalities $f \preceq_{\mathrm{wa}} f_{\boldsymbol{p}}^{\perp}$ and $f \preceq_{\mathrm{sm}} f_{\boldsymbol{p}}^{\perp}$ do not hold for all $f \in \mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Obviously, both $f^H \preceq_{\mathrm{wa}} f_{\boldsymbol{p}}^{\perp}$ and $f^H \preceq_{\mathrm{sm}} f_{\boldsymbol{p}}^{\perp}$ hold since the strongly Rayleigh property implies negative association.

Proposition 2.5.14. *Consider a Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ and, for $\alpha \in [0,1]$, let $f^{(\alpha)} \in \mathcal{B}_d(\boldsymbol{p})$ be a Bernoulli pmf defined by*

$$f^{(\alpha)}(\boldsymbol{x}) = (1-\alpha) f^H(\boldsymbol{x}) + \alpha f_{\boldsymbol{p}}^{\perp}(\boldsymbol{x}), \quad \text{for } \boldsymbol{i} \in \{0,1\}^d, \tag{2.31}$$

where $f_{\boldsymbol{p}}^{\perp} \in \mathcal{B}_d(\boldsymbol{p})$ is the pmf of the Bernoulli random vector with independent components and f^H is the maximum entropy pmf of the class $\mathcal{B}_d^{\Sigma}(\boldsymbol{p})$. Then, the class $\{f^{(\alpha)} \colon \alpha \in [0,1]\}$ is a chain in the partially ordered set $(\mathcal{B}_d(\boldsymbol{p}), \preceq_{sm})$. In particular, $f^{(\alpha)}$ satisfies the NSD property for every $\alpha \in [0,1]$, and, for every $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, we have

$$f^H \preceq_{sm} f^{(\alpha_1)} \preceq_{sm} f^{(\alpha_2)} \preceq_{sm} f_{\boldsymbol{p}}^{\perp}. \tag{2.32}$$

Proof. Consider three Bernoulli random vectors $\boldsymbol{X}^{(\alpha)}$, $\boldsymbol{X}^H$, and $\boldsymbol{X}^{\perp}$ with pmfs $f^{(\alpha)}$, f^H, and $f_{\boldsymbol{p}}^{\perp}$, respectively. From the definition of $f^{(\alpha)}$ in (2.31), it follows that $\mathbb{E}[\phi(\boldsymbol{X}^{(\alpha)})] = (1-\alpha)\mathbb{E}[\phi(\boldsymbol{X}^H)] + \alpha\mathbb{E}[\phi(\boldsymbol{X}^{\perp})]$, for every real function $\phi \colon \{0,1\}^d \to \mathbb{R}$, such that the expectations are finite. Given that $\boldsymbol{X}^H$ satisfies the strongly Rayleigh property, and also enjoys the NSD property, if ϕ is also supermodular, then $\mathbb{E}[\phi(\boldsymbol{X}^H)] \leq \mathbb{E}[\phi(\boldsymbol{X}^{\perp})]$. This implies that $\mathbb{E}[\phi(\boldsymbol{X}^{(\alpha)})]$ increases as α increases. Therefore, we have (2.32) and $f^{(\alpha)} \in \mathrm{NSD}$ for every $\alpha \in [0,1]$. $\square$

The Bernoulli random vector $\boldsymbol{X}^{(\alpha)}$ with pmf $f^{(\alpha)}$ is not Σ-countermonotonic for $\alpha \in (0,1]$. The following counterexample shows that $\boldsymbol{X}^{(\alpha)}$ is not necessarily negatively associated; it does not, hence, always satisfy the strongly Rayleigh property.

Example 2.5.8. *Let us consider the Fréchet class* $\mathcal{B}_5(\frac{3}{10})$. *The pmf* $f^{\perp} \in \mathcal{B}_5(\frac{3}{10})$ *of the vector* $\boldsymbol{X}^{\perp}$ *of independent random variables is*

$$f^{\perp}(\boldsymbol{x}) = \left(\tfrac{3}{10}\right)^{x_1+\cdots+x_5}\left(\tfrac{7}{10}\right)^{5-(x_1+\cdots+x_5)},$$

for $\boldsymbol{x} \in \mathcal{X}_5$, *while the maximum entropy pmf* f^H *in the class* $\mathcal{B}_5^{\Sigma}(\frac{3}{10})$ *is the pmf of the Bernoulli exchangeable random vector* $\boldsymbol{X}^H$ *of the class* $\mathcal{B}_5^{\Sigma}(\frac{3}{10})$. *Since* $|\boldsymbol{p}| = \frac{3}{2} \in (1{,}2)$, *the values of* f^H *are given by*

$$f^H(\boldsymbol{x}) = \begin{cases} \frac{1}{10}, & \text{for } \boldsymbol{x} \in \mathcal{X}_5^1, \\ \frac{1}{20}, & \text{for } \boldsymbol{x} \in \mathcal{X}_5^2, \\ 0, & \text{elsewhere.} \end{cases}$$

Let us consider the Bernoulli random vector $\boldsymbol{X}^{(0.9)}$, *whose pmf* $f^{(0.9)}$ *is defined by* (2.31) *with* $\alpha = 0.9$, *and consider the increasing functions* $h_1(x_1, x_2) = \exp\{x_1 + 2x_2\}$ *and* $h_2(x_3, x_4, x_5) = \exp\{3x_3 + 4x_4 + 6x_5\}$. *Note that* $f^{(0.9)}$ *is neither NA nor strongly Rayleigh, since*

$$Cov(h_1(X_1^{(0.9)}, X_2^{(0.9)}), h_2(X_3^{(0.9)}, X_4^{(0.9)}, X_5^{(0.9)})) = 12.6715 > 0.$$

Chapter 3

Families of Bernoulli distributions

While the previous chapters established the general theoretical framework for negative dependence and its geometric implications, this chapter focuses on the detailed study of two subclasses of multivariate Bernoulli distributions. Our goal is to show how the algebraic representation and the strongly Rayleigh property can be used to identify distributions with negative dependence structures. The chapter is organized into two main parts. Section 3.1 investigates the class of symmetric Bernoulli distributions $\mathcal{SB}_d \coloneqq \mathcal{B}_d(1/2)$, characterized by marginal means $p_j = 1/2$ for all $j \in \{1, \dots, d\}$. By leveraging the algebraic representation introduced in Section 1.2.3, we provide a complete characterization of the distributions that satisfy the extremal negative dependence properties defined in Section 2.3. Section 3.2 introduces and analyzes the Conway–Maxwell multivariate Bernoulli distributions. This family, derived from the Conway–Maxwell-binomial distribution, is particularly flexible: we demonstrate that it is possible to parametrize this family such that the marginal distributions remain fixed, allowing for a study of its properties within a Fréchet class. We show that this family spans the full spectrum of dependence, from positive to negative supermodular dependence. Moreover, we focus on the Strongly Rayleigh property, identifying some parameter ranges that ensure this strong form of negative dependence.

3.1 Symmetric Bernoulli distributions

In this section, we consider the class of symmetric Bernoulli distributions, namely the Fréchet class $\mathcal{SB}_d$. We employ the algebraic representation of Fréchet classes introduced in Section 1.2.3, which proves to be particularly effective for the symmetric case. In particular, using this representation, we provide a complete characterization of the Σ_{cx}-smallest elements within $\mathcal{SB}_d$. As established in Section 2.5.1, these elements coincide with the set of Σ-countermonotonic Bernoulli distributions.

As recalled in Section 1.2.2, the class $\mathcal{SB}_d$ is a convex polytope, viz.

$$\mathcal{SB}_d \leftrightarrow B_d(1/2) = \left\{ \boldsymbol{f} \in \mathbb{R}_+^{2^d} : H_d \boldsymbol{f} = \boldsymbol{0}, \boldsymbol{1}_{2^d}^\top \boldsymbol{f} = 1 \right\},$$

where H_d is a $d \times 2^d$ matrix whose rows are $(\mathbf{1}_{2^d} - 2\boldsymbol{x}_h)^\top$, for $h \in \{1,\dots,d\}$, where $\mathbf{1}_{2^d}$ is the 2^d-vector with all elements equal to 1, and $\boldsymbol{x}_h$ is the 2^d-vector that contains the hth components of all the d-vectors $\boldsymbol{x} \in \mathcal{X}_d$. Therefore, $\mathcal{SB}_d$ is the convex hull of a finite set of pmfs $r_k \in \mathcal{SB}_d$, for $k \in \{1,\dots,n_d\}$, called extremal extremal pmfs. In other terms, for any $f \in \mathcal{SB}_d$, there exist n_d positive weights $\lambda_1,\dots,\lambda_{n_d}$ summing up to one such that

$$f(\boldsymbol{x}) = \sum_{k=1}^{n_d} \lambda_k r_k(\boldsymbol{x}), \quad \text{for every } \boldsymbol{x} \in \mathcal{X}_d.$$

When the dimension d is sufficiently small, the extremal points of the convex polytope can be found using, for example, `4ti2`; see [46]. However, this representation has computational limitations. When the dimension d increases, due to the growth of the number n_d, finding all the extremal pmfs becomes computationally infeasible. To overcome this limitation, we shall consider the algebraic representation introduced by [45] and recalled in Section 1.2.3.

In this section, we consider the map $\mathcal{H}_{1/2}$ defined in (1.16) and extended to $\mathcal{SB}_d$, hereafter denoted simply by $\mathcal{H}$. Since $\gamma_{1/2} = 0$, the image of $f \in \mathcal{SB}_d$ through $\mathcal{H}$ reads

$$\mathcal{H}(f) = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} \{f(\boldsymbol{s}_{\boldsymbol{i}}) - f(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}})\}\boldsymbol{z}^{\boldsymbol{i}}, \tag{3.1}$$

where $\boldsymbol{s}_{\boldsymbol{i}} = (\boldsymbol{i}//0)$.

We recall the definition of equivalent polynomial given in Definition 1.2.1, and we define

$$\mathcal{C}_{\mathcal{H}} = \bigcup_{f \in \mathcal{SB}_d} [\mathcal{H}(f)].$$

The set $\mathcal{C}_{\mathcal{H}}$ is included in the ideal $\mathcal{I}$ of polynomials that vanish at the points $\mathbf{1}_{d-1}$ and $\mathbf{1}_{d-1}^{-j}$, for $j \in \{1,\dots,d-1\}$, where $\mathbf{1}_{d-1}^{-j}$ is a vector with -1 in position j and 1 elsewhere. By Theorem 1.2.5, the set $\mathcal{C}_{\mathcal{H}}$ is generated by the fundamental polynomials $F_{\boldsymbol{i}}(\boldsymbol{z})$, for $\boldsymbol{i} \in \mathcal{X}_d$ with $|\boldsymbol{i}| \geq 2$, defined in (1.23). The following example provides an application for $d = 3$.

Example 3.1.1. *We consider the case $d = 3$, $\mathcal{SB}_3$. From (1.17) with $p = 1/2$, we have $\boldsymbol{a} = Q\boldsymbol{f}$ with*

$$\boldsymbol{a} = \begin{pmatrix} a_{00} \\ a_{10} \\ a_{01} \\ a_{11} \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

Therefore, for $d = 3$, the polynomials in $\mathcal{C}_{\mathcal{H}}$ are of the form $P(\boldsymbol{z}) = a_{00} + a_{10}z_1 + a_{01}z_2 + a_{11}z_1z_2$, where

$$a_{00} = f_{000} - f_{111}, \quad a_{10} = f_{100} - f_{011}, \quad a_{01} = f_{010} - f_{101}, \quad a_{11} = f_{110} - f_{001}.$$

Finally, let us consider the following pmfs and their corresponding polynomials:

$$\begin{aligned}
&\boldsymbol{f}_1 = (0.3,0.1,0.1,0,0.1,0,0,0.4), && P_1(\boldsymbol{z}) = -0.1 + 0.1z_1 + 0.1z_2 - 0.1z_1z_2;\\
&\boldsymbol{f}_2 = (0.1,0.1,0.1,0.2,0.3,0,0,0.2), && P_2(\boldsymbol{z}) = -0.1 + 0.1z_1 + 0.1z_2 - 0.1z_1z_2;\\
&\boldsymbol{f}_3 = (0,0.25,0.25,0,0.25,0,0,0.25), && P_3(\boldsymbol{z}) = -0.25 + 0.25z_1 + 0.25z_2 - 0.25z_1z_2;\\
&\boldsymbol{f}_4 = (0.25,0,0,0.25,0,0.25,0.25,0), && P_4(\boldsymbol{z}) = 0.25 - 0.25z_1 - 0.25z_2 + 0.25z_1z_2.
\end{aligned}$$

Algorithm 2 *Algorithm to find the type-0 pmf in the class* $\mathcal{SB}_d$

Require: A polynomial $P(\boldsymbol{z}) \in \mathcal{I}$ of the form $P(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}} a_{\boldsymbol{i}}\boldsymbol{z}^{\boldsymbol{i}} \not\equiv 0$
Ensure: The type-0 pmf $f^P \in \mathcal{SB}_d$ corresponding to $P(\boldsymbol{z})$

```
for i ∈ X_{d−1} do
    if a_i ≥ 0 then
        f(s_i) = a_i and f(1_d − s_i) = 0
    else
        f(s_i) = 0 and f(1_d − s_i) = −a_i
    end if
end for
for x ∈ X_d do                                   ▷ Normalization
    f^P(x) = f(x)/(Σ_{y∈X_d} f(y))
end for
```

The polynomials $P_1(\boldsymbol{z})$ *and* $P_2(\boldsymbol{z})$ *are identical, while* $P_3(\boldsymbol{z}) = 5P_1(\boldsymbol{z})/2 \in [P_1(\boldsymbol{z})]$, *and* $P_4(\boldsymbol{z}) = -P_3(\boldsymbol{z}) \notin [P_1(\boldsymbol{z})]$.

As Example 3.1.1 shows, the map $\mathcal{H}$ is again not injective. The basis of the kernel $\ker(\mathcal{H}) = \{f \in \mathcal{SB}_d \colon \mathcal{H}(f) = 0\}$ provided in (1.19), in the case $p = 1/2$, reads

$$\mathcal{K} = \{f \in \mathcal{SB}_d \colon \exists \boldsymbol{x} \in \mathcal{X}_d \text{ such that } f(\boldsymbol{x}) = f(\boldsymbol{1}_d - \boldsymbol{x}) = \tfrac{1}{2}\}, \tag{3.2}$$

which can be described as a set of points in $B_d(1/2)$ as

$$\{(\tfrac{1}{2},0,0,\ldots,0,0,\tfrac{1}{2}), (0,\tfrac{1}{2},0,\ldots,0,\tfrac{1}{2},0), (0,0,\tfrac{1}{2},\ldots,\tfrac{1}{2},0,0), \ldots\}.$$

We denote by $\mathcal{PB}_d$ the class of d-dimensional palindromic Bernoulli pmfs, i.e., the pmfs f of Bernoulli random vectors such that $f(\boldsymbol{x}) = f(\boldsymbol{1}_d - \boldsymbol{x})$, for every $\boldsymbol{x} \in \mathcal{X}_d$; see [74] as a reference for palindromic distributions. The proofs of the following propositions are straightforward, yet the results are important for our purposes, because palindromic Bernoulli distributions generate the class of extremal mixture copulas [75], which we will introduce in Chapter 4. The proof of Proposition 3.1.1 is trivial and omitted, while Proposition 3.1.2 follows from Proposition 1.2.3.

Proposition 3.1.1. *The kernel of* $\mathcal{H}$ *coincides with the set of palindromic Bernoulli distributions,* $\ker(\mathcal{H}) \equiv \mathcal{PB}_d$.

Proposition 3.1.2. *The pmfs of the basis* $\mathcal{K}$ *in* (3.2) *are extremal points of the polytope* $\mathcal{SB}_d$.

The basis $\mathcal{K}$ has 2^{d-1} pmfs; therefore, there are 2^{d-1} extremal points of $\mathcal{SB}_d$ that have null polynomial. The kernel $\ker(\mathcal{H})$ is now fully characterized. For characterizing the counter-image of a non-null polynomial, we report the version of Algorithm 1 adapted to the class $\mathcal{SB}_d$.

Propositions 1.2.4 and 3.1.3 are the main ingredients to prove Theorem 3.1.10, one of our main results on negative dependence in the class $\mathcal{SB}_d$. Proposition 3.1.3 improve the result in (1.21), by characterizing the set

$$\mathcal{H}^{-1}[P(\boldsymbol{z})] := \{f \in \mathcal{SB}_d \colon \mathcal{H}(\boldsymbol{f}) \in [P(\boldsymbol{z})]\},$$

i.e., the collection of pmfs that $\mathcal{H}$ maps into a polynomial equivalent to $P(\boldsymbol{z}) \in \mathcal{C}_{\mathcal{H}}$. This proposition is crucial for identifying all extremal negative dependent Bernoulli random vectors, which will be characterized through their polynomials.

Proposition 3.1.3. *Consider a polynomial $P(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} \in \mathcal{I}$, such that $P(\boldsymbol{z}) \not\equiv 0$. Then,*

$$\mathcal{H}^{-1}[P(\boldsymbol{z})] = \{f \in \mathcal{SB}_d : f = \lambda f^P + (1-\lambda) f^K, \text{ for } f^K \in \mathcal{PB}_d, \lambda \in (0,1]\},$$

where f^P is the type-0 pmf of $P(\boldsymbol{z})$.

Proof. Let $A := \{f \in \mathcal{SB}_d : f = \lambda f^P + (1-\lambda) f^K, \text{ for } f^K \in \mathcal{PB}_d, \lambda \in (0,1]\}$, where f^P is the type-0 pmf of $P(\boldsymbol{z})$ and, without restriction, suppose that $P(\boldsymbol{z})$ is such that $\mathcal{H}(f^P) = P(\boldsymbol{z})$. The map $\mathcal{H}$ is linear, therefore $\mathcal{H}(\lambda f^P + (1-\lambda) f^K) = \lambda \mathcal{H}(f^P) + (1-\lambda)\mathcal{H}(f^K) = \lambda P(\boldsymbol{z})$, which implies $A \subseteq \mathcal{H}^{-1}[P(\boldsymbol{z})]$. Let $f \in \mathcal{H}^{-1}[P(\boldsymbol{z})]$ and let $\mathcal{H}(f) = Q(\boldsymbol{z})$. By definition of $\mathcal{H}^{-1}$, there exists $\mu \in (0, +\infty) \cap \mathbb{Q}$ such that $Q(\boldsymbol{z}) = \mu P(\boldsymbol{z})$ and, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$, we have

$$a_{\boldsymbol{i}}^Q = \mu a_{\boldsymbol{i}}^P, \tag{3.3}$$

where $a_{\boldsymbol{i}}^Q$ and $a_{\boldsymbol{i}}^P$ are the coefficients of the polynomials $Q(\boldsymbol{z})$ and $P(\boldsymbol{z})$, respectively. Our first goal is to show that $\mu \leq 1$. By (1.1), for every $\boldsymbol{x} \in \mathcal{X}_d$, there exists a unique $\boldsymbol{i} \in \mathcal{X}_{d-1}$ such that $\boldsymbol{x} = \boldsymbol{s}_{\boldsymbol{i}}$ or $\boldsymbol{x} = \mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}$. By Algorithm 2, for each $\boldsymbol{i} \in \mathcal{X}_{d-1}$, we have either $|a_{\boldsymbol{i}}^P| = f^P(\boldsymbol{s}_{\boldsymbol{i}})$ or $|a_{\boldsymbol{i}}^P| = f^P(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}})$, so that

$$\sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} |a_{\boldsymbol{i}}^P| = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f^P(\boldsymbol{x}) = 1.$$

Moreover, (3.1) implies

$$\sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} |a_{\boldsymbol{i}}^Q| \leq \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} |f(\boldsymbol{s}_{\boldsymbol{i}})| + |f(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}})| = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f(\boldsymbol{x}) = 1 = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} |a_{\boldsymbol{i}}^P|. \tag{3.4}$$

By (3.3) and (3.4), we conclude that $\mu \leq 1$. We now aim to prove that $f \in A$, i.e., that there exist $\lambda \in (0,1]$ and a pmf $f^K \in \mathcal{PB}_d$ such that $f = \lambda f^P + (1-\lambda) f^K$. If we show that

$$f^K = \frac{1}{1-\mu} f - \frac{\mu}{1-\mu} f^P$$

is a pmf with null polynomial, the claim follows by taking $\lambda = \mu$. First, we observe that

$$\mathcal{H}(f^K) = \mathcal{H}\left(\frac{1}{1-\mu} f - \frac{\mu}{1-\mu} f^P\right) = \frac{1}{1-\mu}\mathcal{H}(f) - \frac{\mu}{1-\mu}\mathcal{H}(f^P) = 0.$$

Also, since $1/(1-\mu) - \mu/(1-\mu) = 1$, we have $\sum_{\boldsymbol{x} \in \mathcal{X}_d} f^K(x) = 1$. It remains to show that the components of $f^K(\boldsymbol{x}) \geq 0$ for every $\boldsymbol{x} \in \mathcal{X}_d$. From (3.1), $f(\boldsymbol{s}_{\boldsymbol{i}}) - f(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = \mu f^P(\boldsymbol{s}_{\boldsymbol{i}}) - \mu f^P(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}})$. Thus, we have

$$f(\boldsymbol{s}_{\boldsymbol{i}}) - \mu f^P(\boldsymbol{s}_{\boldsymbol{i}}) = f(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) - \mu f^P(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}). \tag{3.5}$$

By construction, the type-0 pmf satisfies either $f^P(\boldsymbol{s}_{\boldsymbol{i}}) = 0$ or $f^P(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = 0$. In the first case, (3.5) becomes $f(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) - \mu f^P(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = f(\boldsymbol{s}_{\boldsymbol{i}}) \geq 0$, while in the other case, (3.5) becomes $f(\boldsymbol{s}_{\boldsymbol{i}}) - \mu f^P(\boldsymbol{s}_{\boldsymbol{i}}) = f(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) \geq 0$. Thus, for every $\boldsymbol{x} \in \mathcal{X}_d$, we have $f(\boldsymbol{x}) - \mu f^P(\boldsymbol{x}) \geq 0$, and $f^K(\boldsymbol{x}) = (f(\boldsymbol{x}) - \mu f^P(\boldsymbol{x}))/(1-\mu) \geq 0$. Hence, f^K is a pmf of the kernel of $\mathcal{H}$ and we have $\mathcal{H}^{-1}[P(\boldsymbol{z})] \subseteq A$. It follows that $\mathcal{H}^{-1}[P(\boldsymbol{z})] \equiv A$. $\square$

Therefore, the set $\mathcal{H}^{-1}[P(\boldsymbol{z})]$ is a convex polytope, whose extremal points are given by the pmfs corresponding to the elements of $\mathcal{K}$ and by the type-0 pmf of $P(\boldsymbol{z})$, f^P. We conclude this section with the following proposition that highlights the importance of the type-0 pmfs and their link with the generators of $\mathcal{SB}_d$ as a convex polytope.

Proposition 3.1.4. *Every extremal point of $\mathcal{SB}_d$ is either a type-0 pmf or an element of $\mathcal{K}$.*

Proof. Let f be an extremal pmf of $\mathcal{SB}_d$ and let $\mathcal{H}(f) = P(\boldsymbol{z})$. If $P(\boldsymbol{z}) \equiv 0$, then $f \in \mathcal{K}$, which is the set of extremal points that generate $\ker(\mathcal{H})$. Suppose $P(\boldsymbol{z}) \not\equiv 0$. By Proposition 3.1.3, there exists an element of the kernel of $\mathcal{H}$, i.e., a palindromic Bernoulli distribution $\boldsymbol{f}^K \in \mathcal{PB}_d$, such that

$$f = \lambda f^P + (1-\lambda) f^K,$$

for some $\lambda \in (0,1]$, where f^P is the type-0 pmf of $P(\boldsymbol{z})$. Since f is an extremal point, we have $f = f^P$. □

This structure implies that, if f is an extremal pmf of $\mathcal{SB}_d$, then it is the unique extremal pmf associated with this polynomial. In other words, there do not exist two distinct extremal points of $\mathcal{SB}_d$ associated with the same non-zero polynomial. Example 1.2.3 shows that these results hold only in the case $p = 1/2$. It is important to note, however, that while f^P is by construction an extremal pmf of $\mathcal{H}^{-1}[P(\boldsymbol{z})]$, it is not necessarily an extremal point of the entire Fréchet class $\mathcal{SB}_d$.

3.1.1 Characterization of extremal negative dependence

Our main result is to completely characterize the class of Σ_{cx}-smallest elements in $\mathcal{SB}_d$. By Theorem 2.5.1 in Section 2.5.1, this also yields a complete characterization of Σ-countermonotonic random vectors in $\mathcal{SB}_d$. Indeed, Theorem 2.5.1 in Section 2.5.1 states that a Bernoulli random vector is Σ-countermonotonic if and only if it is a Σ_{cx}-smallest element in its Fréchet class.

As anticipated in Section 2.3, the problem of finding Σ_{cx}-smallest elements is trivial if we restrict the analysis to exchangeable Bernoulli random vectors with marginal mean p, for any $p \in (0,1)$. In this case, there is only one Σ_{cx}-smallest element in the class, and, by Proposition 2.3.3, it is also minimal in the supermodular order. In [45], Theorem 5.2 provides an algorithm to construct a non-exchangeable Σ_{cx}-smallest element in the class, but the general problem, even with a common marginal mean p, is still open. If $p = 1/2$, we establish a stronger result by providing an explicit characterization of all such elements.

We define the sets

$$\mathcal{X}_d^* = \begin{cases} \mathcal{X}_d^{d/2}, & \text{if } d \text{ is even,} \\ \mathcal{X}_d^{(d-1)/2} \cup \mathcal{X}_d^{(d+1)/2}, & \text{if } d \text{ is odd,} \end{cases}$$

and

$$\mathscr{I}_{d-1}^* = \begin{cases} \mathcal{X}_{d-1}^{d/2}, & \text{if } d \text{ is even,} \\ \mathcal{X}_{d-1}^{(d-1)/2} \cup \mathcal{X}_{d-1}^{(d+1)/2}, & \text{if } d \text{ is odd.} \end{cases}$$

Moreover, we denote by M_d the largest integer smaller than or equal to $d/2$, i.e.,

$$M_d = \left\lfloor \frac{d}{2} \right\rfloor = \begin{cases} \frac{d}{2}, & \text{if } d \text{ is even,} \\ \frac{d-1}{2}, & \text{if } d \text{ is odd.} \end{cases}$$

We restate the characterization of the class of Σ_{cx}-smallest elements discussed in Section 2.3 in the case $p = 1/2$. The following proposition identifies the distribution of the sum of a symmetric Bernoulli random vector that is minimal in the convex order.

Proposition 3.1.5. *A Bernoulli random vector $\boldsymbol{X} \in \mathcal{SB}_d$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$ if and only if the sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ has distribution given by*

$$r^D_{d/2} = \begin{cases} M_d + 1 - \frac{d}{2}, & y = M_d, \\ \frac{d}{2} - M_d, & y = M_d + 1, \\ 0, & \textit{otherwise.} \end{cases}$$

Proposition 3.1.6 characterizes the support of the Σ_{cx}-smallest pmfs in $\mathcal{SB}_d$.

Proposition 3.1.6. *A pmf $f \in \mathcal{SB}_d$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$ if and only if the support of f is contained in $\mathcal{X}_d^*$.*

Proof: Consider a Bernoulli random vector $\boldsymbol{X}$ with pmf $f \in \mathcal{SB}_d$. If f is a Σ_{cx}-smallest element in $\mathcal{SB}_d$, the distribution of the sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ is specified by Proposition 3.1.5, and we can easily prove that $f(\boldsymbol{x}) = 0$, for all $\boldsymbol{x} \notin \mathcal{X}_d^*$. Conversely, if the support of f is contained in $\mathcal{X}_d^*$, $S_{\boldsymbol{X}}$ is equal to $(d-1)/2$ or $(d+1)/2$ with probability one, when d is odd, and $S_{\boldsymbol{X}}$ is equal to $d/2$ with probability one, when d is even. Moreover, $\boldsymbol{X} \in \mathcal{SB}_d$ implies that $\mathbb{E}(S_{\boldsymbol{X}}) = d/2$. Therefore, the distribution of $S_{\boldsymbol{X}}$ is the distribution specified in Proposition 3.1.5, and we conclude that f is a Σ_{cx}-smallest element in $\mathcal{SB}_d$. □

Remark 3.1.1. *When d is odd and $\mathcal{X}_d^* = \mathcal{X}_d^{(d-1)/2} \cup \mathcal{X}_d^{(d+1)/2}$, if the support of $f \in \mathcal{SB}_d$ is contained in $\mathcal{X}_d^*$, Proposition 3.1.5 implies that there exist $\boldsymbol{x}_1 \in \mathcal{X}_d^{(d-1)/2}$ and $\boldsymbol{x}_2 \in \mathcal{X}_d^{(d+1)/2}$, such that $f(\boldsymbol{x}_1) > 0$ and $f(\boldsymbol{x}_2) > 0$.*

It is worth noting that there does not exist any joint mix in $\mathcal{SB}_d$ when d is odd. When d is even, instead, a Bernoulli random vector is a Σ_{cx}-smallest element in $\mathcal{SB}_d$ if and only if it is a joint mix; in this case, the definitions of Σ-countermonotonic random vector, Σ_{cx}-smallest element, and joint mix coincide in $\mathcal{SB}_d$. Therefore, building on Proposition 3.1.6, we can characterize extremal negative dependence by considering the class of Σ_{cx}-smallest elements. We first identify the Σ_{cx}-smallest palindromic pmfs.

Proposition 3.1.7. *A pmf $f^{K*} \in \mathcal{PB}_d$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$ if and only if it is a convex linear combination of the Σ_{cx}-smallest pmfs corresponding to the elements of $\mathcal{K}$ in (3.2).*

Proof. Any pmf $f^{K*} \in \mathcal{PB}_d \equiv \ker(\mathcal{H})$ can be expressed as a convex linear combination of the pmfs corresponding to $\mathcal{K}$,

$$f^{K*} = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} \lambda_{\boldsymbol{i}} g_{\boldsymbol{i}},$$

where, for $\boldsymbol{i} \in \mathcal{X}_{d-1}$, the extremal pmf $g_{\boldsymbol{i}}$ satisfies $g_{\boldsymbol{i}}(\boldsymbol{s}_{\boldsymbol{i}}) = g_{\boldsymbol{i}}(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = 1/2$. Let f^{K*} be a Σ_{cx}-smallest element in $\mathcal{SB}_d$. By Proposition 3.1.6, f^{K*} has support on $\mathcal{X}_d^*$, and it is easy to show that $\lambda_{\boldsymbol{i}} = 0$ for every $g_{\boldsymbol{i}}$ that is not a Σ_{cx}-smallest element in $\mathcal{SB}_d$. The converse implication holds since any convex linear combination of Σ_{cx}-smallest elements is itself a Σ_{cx}-smallest element. □

We denote by $\mathcal{PB}_d^* \subseteq \mathcal{PB}_d$ the set of Σ_{cx}-smallest palindromic Bernoulli distributions. Proposition 3.1.7 states that $\mathcal{PB}_d^*$ is a convex polytope whose extremal points are the Σ_{cx}-smallest pmfs in $\mathcal{K}$, which are easy to identify. Indeed, they have support on exactly two points, $\boldsymbol{x}$ and $\mathbf{1}_d - \boldsymbol{x}$, where $\boldsymbol{x}$ is such that $x_1 + \cdots + x_d = M_d$. We now consider the entire class $\mathcal{SB}_d$. The following theorem characterizes the coefficients of the polynomials corresponding to the Σ_{cx}-smallest pmfs of the class $\mathcal{SB}_d$.

Theorem 3.1.8. *Let $f^* \in \mathcal{SB}_d$ be a Σ_{cx}-smallest element in $\mathcal{SB}_d$. Then, the coefficients of the polynomial $\mathcal{H}(f^*) = P^*(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} \in \mathcal{I}$ are such that:*

1. *$a_{\boldsymbol{i}} = 0$, for every $\boldsymbol{i} \notin \mathscr{I}_{d-1}^*$,*
2. *the sum of the coefficients of the monomials of the same order is equal to 0, i.e.,*
$$\sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^k} a_{\boldsymbol{i}} = 0, \quad \text{for every } k \in \{0, \dots, d-1\},$$
3. *for each $j \in \{1, \dots, d-1\}$, the sum of the coefficients $a_{\boldsymbol{i}}$ involving z_j is zero, i.e.,*
$$\sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}: i_j=1} a_{\boldsymbol{i}} = 0, \quad \text{for each } j \in \{1, \dots, d-1\}.$$

Proof. By construction, $\boldsymbol{x} \in \mathcal{X}_d^*$ if and only if $\mathbf{1}_d - \boldsymbol{x} \in \mathcal{X}_d^*$, and, by Proposition 3.1.6, if f^* is a Σ_{cx}-smallest pmf in $\mathcal{SB}_d$, then $f^*(\boldsymbol{x}) = 0$ for every $\boldsymbol{x} \notin \mathcal{X}_d^*$. Moreover, for $\boldsymbol{i} = (i_1, \dots, i_{d-1}) \notin \mathscr{I}_{d-1}^*$, we have $\boldsymbol{s}_{\boldsymbol{i}} = (i_1, \dots, i_{d-1}, 0) \notin \mathcal{X}_d^*$ and $\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}} = (1 - i_1, \dots, 1 - i_{d-1}, 1) \notin \mathcal{X}_d^*$. It follows that, for every $\boldsymbol{i} \notin \mathscr{I}_{d-1}^*$, $f^*(\boldsymbol{s}_{\boldsymbol{i}}) = f^*(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = 0$, and $a_{\boldsymbol{i}} = f^*(\boldsymbol{s}_{\boldsymbol{i}}) - f^*(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = 0$. This proves item 1. We now prove the other two statements of the theorem with the assumption that d is odd; the case where d is even follows analogously. By item 1, we can write

$$P^*(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathscr{I}_{d-1}^*} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^{M_d}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} + \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^{M_d+1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}. \tag{3.6}$$

Theorem 1.2.5 states that the image of any pmf through the map $\mathcal{H}$ is a linear combination of the fundamental polynomials, which are given by, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}^-$,

$$F_{\boldsymbol{i}}(\boldsymbol{z}) = F_{j_1,\dots,j_{n_{\boldsymbol{i}}}}(\boldsymbol{z}) = \prod_{h=1}^{n_{\boldsymbol{i}}} z_{j_h} - \sum_{h=1}^{n_{\boldsymbol{i}}} z_{j_h} + (n_{\boldsymbol{i}} - 1) = \boldsymbol{z}^{\boldsymbol{i}} - \sum_{h=1}^{n_{\boldsymbol{i}}} z_{j_h} + (n_{\boldsymbol{i}} - 1),$$

where $(j_1, \dots, j_{n_{\boldsymbol{i}}})$ are the position of the ones in $\boldsymbol{i}$, and $n_{\boldsymbol{i}} := \sum_{j=1}^{d-1} i_j$ is their number. Note that, differing from (1.23), we have also defined $F_{\boldsymbol{i}}(\boldsymbol{z})$ for $\boldsymbol{i} \in \mathcal{X}_{d-1}^1$, as it coincides with the null polynomial. We can write $P^*(\boldsymbol{z}) = \mathcal{H}(f^*)$ as a linear combination of the fundamental polynomials

$$P^*(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}\setminus\{\boldsymbol{0}\}} \gamma_{\boldsymbol{i}} F_{\boldsymbol{i}}(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}\setminus\{\boldsymbol{0}\}} \gamma_{\boldsymbol{i}} \left\{ \boldsymbol{z}^{\boldsymbol{i}} - \sum_{h=1}^{n_{\boldsymbol{i}}} z_{j_h} + (n_{\boldsymbol{i}} - 1) \right\}, \tag{3.7}$$

with $\gamma_{\boldsymbol{i}} \in \mathbb{Q}$, for every $\boldsymbol{i} \in \mathcal{X}_{d-1} \setminus \{\boldsymbol{0}\}$. Equating the two representation of $P^*(\boldsymbol{z})$ in (3.6) and (3.7), for every $\boldsymbol{i} \in \cup_{k=2}^{d-1} \mathcal{X}_{d-1}^k$, we have $\gamma_{\boldsymbol{i}} = a_{\boldsymbol{i}}$ if $\boldsymbol{i} \in \mathscr{I}_{d-1}^*$, and $\gamma_{\boldsymbol{i}} = 0$ otherwise. Moreover, since $M_d \geq 1$, the expression in (3.6) has no constant term. Therefore, the constant term in (3.7) equals zero, i.e.,

$$\sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}\setminus\{\boldsymbol{0}\}} a_{\boldsymbol{i}}(n_{\boldsymbol{i}} - 1) = (M_d - 1) \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^{M_d}} a_{\boldsymbol{i}} + M_d \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^{M_d+1}} a_{\boldsymbol{i}} = 0. \tag{3.8}$$

Recall that the polynomial $P^*(\boldsymbol{z}) \in \mathcal{I}$ vanishes at points $\mathbf{1}_{d-1}$ and $\mathbf{1}_{d-1}^{-j}$, for $j \in \{1, \ldots, d-1\}$; in particular,

$$P^*(\mathbf{1}_{d-1}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^{M_d}} a_{\boldsymbol{i}} + \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}^{M_d+1}} a_{\boldsymbol{i}} = 0. \tag{3.9}$$

Item 2 follows by plugging (3.9) in (3.8). In addition, for every $j \in \{1, \ldots, d-1\}$, we can rewrite $P^*(\mathbf{1}_{d-1})$ as

$$P^*(\mathbf{1}_{d-1}) = \sum_{\boldsymbol{i}\in\mathscr{I}_{d-1}^* :\, i_j=0} a_{\boldsymbol{i}} + \sum_{\boldsymbol{i}\in\mathscr{I}_{d-1}^* :\, i_j=1} a_{\boldsymbol{i}} = 0, \tag{3.10}$$

and, for every $j \in \{1, \ldots, d-1\}$, we also have

$$P^*(\mathbf{1}_{d-1}^{-j}) = \sum_{\boldsymbol{i}\in\mathscr{I}_{d-1}^* :\, i_j=0} a_{\boldsymbol{i}} - \sum_{\boldsymbol{i}\in\mathscr{I}_{d-1}^* :\, i_j=1} a_{\boldsymbol{i}} = 0. \tag{3.11}$$

Equations (3.10) and (3.11) imply item 3. □

From Theorem 3.1.8, all the polynomials $P^*(\boldsymbol{z})$ with a Σ_{cx}-smallest pmf in their counter-image $\mathcal{H}^{-1}[P^*(\boldsymbol{z})]$ are of the form

$$P^*(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathscr{I}_{d-1}^*} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}, \tag{3.12}$$

where the coefficients $a_{\boldsymbol{i}}$, for $\boldsymbol{i} \in \mathscr{I}_{d-1}^*$, verify the conditions 2 and 3 of the above theorem. The next corollary to Theorem 3.1.8 states that the coefficients of the polynomials of the Σ_{cx}-smallest pmfs in the class $\mathcal{SB}_d$ are the solutions of a homogeneous linear system. The number n_d^* vectors in $\mathscr{I}_{d-1}^*$ is given by

$$n_d^* = \begin{cases} \binom{d-1}{M_d} + \binom{d-1}{M_d+1}, & \text{if } d \text{ is odd,} \\ \binom{d-1}{M_d}, & \text{if } d \text{ is even.} \end{cases}$$

Corollary 3.1.9. *Let $f \in \mathcal{SB}_d$ be a Σ_{cx}-smallest element in $\mathcal{SB}_d$. Then, the coefficients of the polynomial $\mathcal{H}(f) = P(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}} a_{\boldsymbol{i}}\boldsymbol{z}^{\boldsymbol{i}} \in \mathcal{I}$ are the solutions of*

$$A_d\boldsymbol{a} = \boldsymbol{0}, \tag{3.13}$$

where $\boldsymbol{a} = (a_{\boldsymbol{i}} \colon \boldsymbol{i} \in \mathscr{I}^*_{d-1})$ and A_d is obtained from the matrix $A_{\mathscr{I}^*_{d-1}} = (\boldsymbol{i} \colon \boldsymbol{i} \in \mathscr{I}^*_{d-1}) \in \mathbb{R}^{(d-1)\times n^*_d}$, whose columns are the vectors $\boldsymbol{i} \in \mathscr{I}^*_{d-1}$. In particular,

- *if d is even, $A_d = (\boldsymbol{1}^\top_{n^*_d}//A_{\mathscr{I}^*_{d-1}}) \in \mathbb{R}^{d\times n^*_d}$;*
- *if d is odd, $A_d = (R_1//R_2//A_{\mathscr{I}^*_{d-1}}) \in \mathbb{R}^{(d+1)\times n^*_d}$, where $R_1 \in \mathbb{R}^{1\times n^*_d}$ is a row vector with 1s in correspondence of the indexes $\boldsymbol{i}$ with sum M_d and zeros elsewhere, and $R_2 \in \mathbb{R}^{1\times n^*_d}$ is a row vector with 1s in correspondence of the indexes $\boldsymbol{i}$ with sum $M_d + 1$ and zeros elsewhere.*

Proof. By item 1, the condition in item 2 of Theorem 3.1.8 simplifies to

$$\begin{cases} \sum_{\boldsymbol{i}\in\mathcal{X}^{M_d}_{d-1}} a_{\boldsymbol{i}} = \sum_{\boldsymbol{i}\in\mathcal{X}^{M_d+1}_{d-1}} a_{\boldsymbol{i}} = 0, & \text{if } d \text{ is odd,} \\ \sum_{\boldsymbol{i}\in\mathcal{X}^{M_d}_{d-1}} a_{\boldsymbol{i}} = 0, & \text{if } d \text{ is even.} \end{cases} \tag{3.14}$$

Moreover, since

$$\sum_{\boldsymbol{i}\in\mathscr{I}^*_{d-1}: i_j=1} a_{\boldsymbol{i}} = \sum_{\boldsymbol{i}\in\mathscr{I}^*_{d-1}} i_j a_{\boldsymbol{i}},$$

item 3 of Theorem 3.1.8 reads

$$\sum_{\boldsymbol{i}\in\mathscr{I}^*_{d-1}} i_j a_{\boldsymbol{i}} = 0, \quad j \in \{1,\dots,d-1\}. \tag{3.15}$$

We consider two cases. If d is odd, (3.14) and (3.15) lead to the linear system

$$\begin{cases} \sum_{\boldsymbol{i}\in\mathcal{X}^{M_d}_{d-1}} a_{\boldsymbol{i}} = 0, \\ \sum_{\boldsymbol{i}\in\mathcal{X}^{M_d+1}_{d-1}} a_{\boldsymbol{i}} = 0, \\ \sum_{\boldsymbol{i}\in\mathscr{I}^*_{d-1}} i_j a_{\boldsymbol{i}} = 0, \quad j \in \{1,\dots,d-1\}, \end{cases}$$

whose coefficient matrix is $A_d = (R_1//R_2//A_{\mathscr{I}^*_{d-1}})$. If instead d is even, the linear system becomes

$$\begin{cases} \sum_{\boldsymbol{i}\in\mathcal{X}^{M_d}_{d-1}} a_{\boldsymbol{i}} = 0, \\ \sum_{\boldsymbol{i}\in\mathscr{I}^*_{d-1}} i_j a_{\boldsymbol{i}} = 0, \quad j \in \{1,\dots,d-1\}, \end{cases}$$

whose coefficient matrix is $A_d = (\boldsymbol{1}^\top_{n^*_d}//A_{\mathscr{I}^*_{d-1}})$. Hence, the assertion follows. □

The following two examples characterize the polynomials of Σ_{cx}-smallest pmfs in dimensions $d = 3$ and $d = 4$, respectively.

Example 3.1.2. *We consider $d = 3$. Since d is odd, we have $M_d = (d-1)/2 = 1$ and $\mathscr{I}_2^* = \{(1,0),(0,1),(1,1)\}$. For $(i_1,i_2) \in \mathscr{I}_2^*$, the first row of A_3 is equal to 1 if $i_1+i_2 = M_d = 1$ and 0 otherwise, and the second row is the opposite, viz.*

$$A_3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The matrix $A_3 \in \mathbb{R}^{4\times 3}$ has rank 3. Consequently, the linear system in (3.13) admits only the trivial solution $a_{10} = a_{01} = a_{11} = 0$, and all Σ_{cx}-smallest pmfs in $\mathcal{SB}_3$ have null polynomials.

Example 3.1.3. *We consider $d = 4$. Since d is even, we have $M_d = d/2 = 2$ and $\mathscr{I}_3^* = \{(1,1,0),(1,0,1),(0,1,1)\}$. The first row of A_4 is a vector of all 1s, viz.*

$$A_4 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The matrix $A_4 \in \mathbb{R}^{4\times 3}$ has rank 3. Consequently, the linear system in (3.13) admits only the trivial solution $a_{110} = a_{101} = a_{011} = 0$, and all Σ_{cx}-smallest pmfs in $\mathcal{SB}_4$ have null polynomials.

Remark 3.1.2. *As shown in Examples 3.1.2 and 3.1.3, in the cases $d = 3$ and $d = 4$, all Σ_{cx}-smallest pmfs have null polynomial, i.e., $\mathcal{H}(f) = 0$, if f is Σ_{cx}-smallest. Therefore, the set of Σ_{cx}-smallest pmfs is included in $\ker(\mathcal{H}) \equiv \mathcal{PB}_d$. Thus, for $d \leq 4$, both the Σ_{cx}-smallest pmfs and the "Σ_{cx}-maximal" pmf (the upper Fréchet bound) are palindromic Bernoulli distributions.*

Corollary 3.1.9 states that the coefficients of the polynomials of all Σ_{cx}-smallest pmfs of $\mathcal{SB}_d$ are solutions of the homogeneous linear system in (3.13). However, there exist pmfs in $\mathcal{SB}_d$ that are not Σ_{cx}-smallest elements yet generate a polynomial of the form in (3.12). This is a consequence of the fact that the map $\mathcal{H}$ is not injective. For example, let $f^* \in \mathcal{SB}_d$ be a Σ_{cx}-smallest Bernoulli pmf, and denote by $P^*(\boldsymbol{z})$ its corresponding polynomial. We know that $P^*(\boldsymbol{z})$ verifies the three properties of Theorem 3.1.8. Consider also $g = (f^* + \bar{f})/2$, where $\bar{f}(0,\ldots,0) = \bar{f}(1,\ldots,1) = 1/2$. Since $\mathcal{H}(\bar{f}) \equiv 0$, it follows by linearity of the map $\mathcal{H}$ that $\mathcal{H}(g) = P^*(\boldsymbol{z})$, although g is not a Σ_{cx}-smallest element in $\mathcal{SB}_d$. Theorem 3.1.10 states the key result of this section, because it completely characterizes the class of Σ_{cx}-smallest pmfs.

Theorem 3.1.10. *Let $P^*(\boldsymbol{z}) = \sum_{\boldsymbol{i}\in\mathcal{X}_{d-1}} a_{\boldsymbol{i}}\boldsymbol{z}^{\boldsymbol{i}} \in \mathbb{Q}[z_1,\ldots,z_{d-1}]$ be a non-null polynomial that verifies the three properties of Theorem 3.1.8. Then, $P^*(\boldsymbol{z}) \in \mathcal{I}$ and the type-0 pmf f^* corresponding to $P^*(\boldsymbol{z})$ is a Σ_{cx}-smallest pmf of $\mathcal{SB}_d$. Moreover, the family*

$$\{f = \lambda f^* + (1-\lambda)f^{K*} \colon f^{K*} \in \mathcal{PB}_d^*,\ \lambda \in (0,1]\},$$

constitutes the set of all Σ_{cx}-smallest pmfs corresponding to polynomials equivalent to $P^*(\boldsymbol{z})$.

Proof. We first consider the case d odd. Our first goal is to show that $P^*(\boldsymbol{z}) \in \mathcal{I}$, i.e., $P^*(\mathbf{1}_{d-1}) = 0$ and $P^*(\mathbf{1}_{d-1}^{-j}) = 0$, for every $j \in \{1, \ldots, d\}$. From item 1 of Theorem 3.1.8, we can write

$$P^*(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} = \sum_{\boldsymbol{i} \in \mathcal{X}^{M_d}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}} + \sum_{\boldsymbol{i} \in \mathcal{X}^{M_d+1}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}.$$

Therefore,

$$P^*(\mathbf{1}_{d-1}) = \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1}} a_{\boldsymbol{i}} = \sum_{\boldsymbol{i} \in \mathcal{X}^{M_d}_{d-1}} a_{\boldsymbol{i}} + \sum_{\boldsymbol{i} \in \mathcal{X}^{M_d+1}_{d-1}} a_{\boldsymbol{i}} = 0,$$

where the last equality follows from item 2 of Theorem 3.1.8. Moreover, for each $j \in \{1, \ldots, d-1\}$, it holds

$$0 = \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1}} a_{\boldsymbol{i}} = \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1} :\, i_j = 0} a_{\boldsymbol{i}} + \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1} :\, i_j = 1} a_{\boldsymbol{i}}. \tag{3.16}$$

Hence, by item 3 of Theorem 3.1.8 and by (3.16), it follows that, for each $j \in \{1, \ldots, d-1\}$,

$$P^*(\mathbf{1}_{d-1}^{-j}) = \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1} :\, i_j = 0} a_{\boldsymbol{i}} - \sum_{\boldsymbol{i} \in \mathscr{I}^*_{d-1} :\, i_j = 1} a_{\boldsymbol{i}} = 0.$$

Then, $P^*(\boldsymbol{z}) \in \mathcal{I}$, and we can apply Algorithm 2 to find the corresponding type-0 pmf f^*. For every $\boldsymbol{i} \in \mathcal{X}_{d-1}$ such that $a_{\boldsymbol{i}} = 0$, the type-0 pmf has $f^*(\boldsymbol{s}_{\boldsymbol{i}}) = f^*(\mathbf{1}_d - \boldsymbol{s}_{\boldsymbol{i}}) = 0$. This implies that, given $k \in \{0, \ldots, d-1\}$, if $a_{\boldsymbol{i}} = 0$ for all $\boldsymbol{i} \in \mathcal{X}^k_{d-1}$, the sum of the components of a Bernoulli random vector with the type-0 as pmf does not have support on k or $d-k$. Therefore, the condition $a_{\boldsymbol{i}} = 0$ for every $\boldsymbol{i} \notin \mathscr{I}^*_{d-1}$ implies that the type-0 pmf f^* has support only on points with sum equal to M_d or $M_d + 1$. Thus, by Proposition 3.1.6, f^* is a Σ_{cx}-smallest element in $\mathcal{SB}_d$. It remains to prove the last part of the theorem. First, we note that if $f = \lambda f^* + (1-\lambda) f^{K*}$, for $\lambda \in (0,1]$ and $f^{K*} \in \mathcal{PB}^*_d$, then f is a Σ_{cx}-smallest pmf in $\mathcal{SB}_d$. Conversely, suppose that f is a Σ_{cx}-smallest pmf such that $\mathcal{H}(f) = Q(\boldsymbol{z}) \in [P^*(\boldsymbol{z})]$. From Proposition 3.1.3, we have that $f = \lambda f^* + (1-\lambda) f^K$, for some $\lambda \in (0,1]$ and $f^K \in \mathcal{PB}_d$. Since $f(\boldsymbol{x}) = f^*(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \notin \mathcal{X}^*_d$, we have that f^K is a Σ_{cx}-smallest pmf, i.e., $f^K \in \mathcal{PB}^*_d$. The arguments for the case of even d are analogous. □

The algebraic representation and the proofs of the main results are technical; however, their strength lies in their simplicity of use. In what follows, we illustrate how to apply these results to find a Σ_{cx}-smallest element in $\mathcal{SB}_d$ and, at least in principle, how they allow us to determine all Σ_{cx}-smallest elements. To find a Σ_{cx}-smallest element in the class $\mathcal{SB}_d$, we proceed as follows:

1. Choose a polynomial $P(\boldsymbol{z}) = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} a_{\boldsymbol{i}} \boldsymbol{z}^{\boldsymbol{i}}$, with coefficients that satisfies the conditions in Theorem 3.1.8.

2. Apply Algorithm 2 to find the type-0 pmf f^*.

The type-0 pmf f^* is a Σ_{cx}-smallest element in $\mathcal{SB}_d$. To find all Σ_{cx}-smallest elements, we rely on Corollary 3.1.9, which provides a method to determine the coefficients of all polynomials satisfying the conditions of Theorem 3.1.8. Then, for each polynomial, we construct the type-0 pmf f^*. By Theorem 3.1.10, all Σ_{cx}-smallest pmfs can be written as

$$f = \lambda f^* + (1-\lambda) f^{K*}, \quad \lambda \in (0,1],$$

where f^{K*} is a convex combination of pmfs in $\mathcal{K}$ with support in $\mathcal{X}_d^*$. We recall that this is equivalent to finding all Σ-countermonotonic elements in $\mathcal{SB}_d$, which also satisfy the joint mixability property, when d is even.

We conclude this section with Examples 3.1.4 and 3.1.5, which characterize the Σ_{cx}-smallest elements of the classes $\mathcal{SB}_5$ and $\mathcal{SB}_6$, respectively.

Example 3.1.4. *In this example, we show how to find all Σ_{cx}-smallest elements in the class $\mathcal{SB}_5$. By Corollary 3.1.9, we build the matrix*

$$A_5 = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

that is given by $A_5 = (R_1//R_2//A_{I_4^*})$. Since rank$(A_5) = 5$, the solution space of the system in (3.13) has dimension $n_5^* - \text{rank}(A_5) = 5$. A basis of the space of the solutions of $A_5\boldsymbol{a} = \mathbf{0}$ is

$$\begin{aligned} \mathcal{A} = \{&\boldsymbol{a}^{(1)} = (0,1,-1,0,-1,1,0,0,0,0), \boldsymbol{a}^{(2)} = (0,1,0,-1,-1,0,1,0,0,0), \\ &\boldsymbol{a}^{(3)} = (1,0,-1,0,-1,0,0,1,0,0), \boldsymbol{a}^{(4)} = (1,0,0,-1,-1,0,0,0,1,0), \\ &\boldsymbol{a}^{(5)} = (1,1,-1,-1,-1,0,0,0,0,1)\}. \end{aligned}$$

Thus, every polynomial whose coefficients

$$\boldsymbol{a} = (a_{1100}, a_{1010}, a_{0110}, a_{1110}, a_{1001}, a_{0101}, a_{1101}, a_{0011}, a_{1011}, a_{0111})$$

are a linear combination of the vectors of the basis $\mathcal{A}$ verify the three assumptions of Theorem 3.1.8 and its type-0 pmf is a Σ_{cx}-smallest element in $\mathcal{SB}_5$. For example, the polynomial corresponding to the vector $\boldsymbol{a}^{(1)}$ is $P_1(\boldsymbol{z}) = z_1z_3 - z_2z_3 - z_1z_4 + z_2z_4$ and the corresponding type-0 pmf $f^{(1)}$ is such that $f^{(1)}((1,0,1,0,0)) = f^{(1)}((1,0,0,1,1)) = f^{(1)}((0,1,1,0,1)) = f^{(1)}((0,1,0,1,0)) = 1/4$ and it is zero elsewhere. Following Lemma 2.3 in [99], which gives the conditions for a pmf to be an extremal point, it can be proved that $f^{(1)}$ is an extremal pmf of the polytope $\mathcal{SB}_5$. A general Σ_{cx}-smallest pmf in $\mathcal{H}^{-1}[P_1(\boldsymbol{z})]$ can be found as $f = \lambda f^{(1)} + (1-\lambda)f^{K*}$, where $f^{K*} \in \mathcal{PB}_6^*$.

Example 3.1.5. *In this example, we show how to find all Σ_{cx}-smallest elements of the class $\mathcal{SB}_6$. By Corollary 3.1.9, we build the matrix $A_6 = (\mathbf{1}_{10}^\top//A_{I_5^*})$. We find that the basis $\mathcal{A}$ of the solution space in Example 3.1.4 is also a basis of the space of solutions of the system $A_6\boldsymbol{a} = \mathbf{0}$. Thus, every polynomial that have a Σ_{cx}-smallest pmf in its counter-image has coefficients*

$$\boldsymbol{a} = (a_{11100}, a_{11010}, a_{10110}, a_{01110}, a_{11001}, a_{10101}, a_{01101}, a_{10011}, a_{01011}, a_{00111})$$

that are a linear combination of $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \boldsymbol{a}^{(3)}, \boldsymbol{a}^{(4)}$, and $\boldsymbol{a}^{(5)}$. For example, the polynomial with coefficients $\boldsymbol{a}^{(1)}$ is $P_1(\boldsymbol{z}) = z_1z_2z_4 - z_1z_3z_4 - z_1z_2z_5 + z_1z_3z_5$ and the corresponding type-0 pmf $f^{(1)}$ is such that $f^{(1)}((1,1,0,1,0,0)) = f^{(1)}((0,1,0,0,1,1)) = f^{(1)}((0,0,1,1,0,1)) =$

$f^{(1)}((1,0,1,0,1,0)) = 1/4$ and it is zero elsewhere. As in Example 3.1.4, it can be proved that $f^{(1)}$ is an extremal pmf of the polytope $\mathcal{SB}_6$. Finally, we consider the linear combination $\tilde{\boldsymbol{a}} = \boldsymbol{a}^{(1)} - \boldsymbol{a}^{(2)} - \boldsymbol{a}^{(3)} + \boldsymbol{a}^{(4)}$. The resulting polynomial is $\tilde{P}(\boldsymbol{z}) = P_1(\boldsymbol{z}) - P_2(\boldsymbol{z}) - P_3(\boldsymbol{z}) + P_4(\boldsymbol{z}) = z_1z_3z_5 - z_2z_3z_5 - z_1z_4z_5 + z_2z_4z_5$ and its type-0 pmf $\tilde{f}$ is such that $\tilde{f}((1,0,1,0,1,0)) = \tilde{f}((1,0,0,1,0,1)) = \tilde{f}((0,1,1,0,0,1)) = \tilde{f}((0,1,0,1,1,0)) = 1/4$ and zero elsewhere. It can be proved that also $\tilde{f}$ is an extremal point of $\mathcal{SB}_6$.

Proposition 3.1.11 states an interesting property of the linear system in (3.13).

Proposition 3.1.11. *If d is odd, we have rank(A_d) = rank$(A_{d+1}) = d$.*

Proof. Let d be odd. Since $n_d^* = n_{d+1}^*$, it is clear that $A_d = (R_1//R_2//A_{\mathscr{I}_{d-1}^*})$ and $A_{d+1} = (\mathbf{1}_{n_{d+1}^*}^\top//A_{\mathscr{I}_d^*}) = (\mathbf{1}_{n_d^*}^\top//A_{\mathscr{I}_d^*})$ are matrices of order $(d+1)\times n_d^*$. We now prove that $A_{\mathscr{I}_d^*} = (A_{\mathscr{I}_{d-1}^*}//R_1)$. We have $\boldsymbol{i}_{d-1} \in \mathscr{I}_{d-1}^*$ if and only if $\sum_{k=1}^{d-1} i_k = M_d$ or $\sum_{j=1}^{d-1} i_j = M_d + 1$ and $\boldsymbol{i}_d \in \mathscr{I}_d^*$ if and only if $\sum_{j=1}^{d} j_k = \frac{d+1}{2} = M_d+1$. Let $\boldsymbol{i}_{d-1} \in \mathscr{I}_{d-1}^*$. If $\sum_{j=1}^{d-1} i_{d-1,j} = M_d$, then $(1//\boldsymbol{i}_{d-1}) \in \mathscr{I}_d^*$, while if $\sum_{j=1}^{d-1} i_{d-1,j} = M_d + 1$, then $(0//\boldsymbol{i}_{d-1}) \in \mathscr{I}_d^*$. Therefore, $A_{\mathscr{I}_d^*} = (R_1//A_{\mathscr{I}_{d-1}^*})$, where $R_1 \in \mathbb{R}^{1\times n_d^*}$ with ones in correspondence of the indexes $\boldsymbol{i}$ with sum M_d and zeros elsewhere. Thus, $A_d = (R_1//R_2//A_{\mathscr{I}_{d-1}^*})$ and $A_{d+1} = (\mathbf{1}_{n_d^*}//R_1//A_{\mathscr{I}_{d-1}^*})$. Since $\mathbf{1}_{n_d^*} = R_1 + R_2$ by construction, then

$$\text{rank}(A_d) = \text{rank}(A_{d+1}). \tag{3.17}$$

Let $r_d := \text{rank}(A_d) = \text{rank}(A_{d+1})$. We first observe that $r_d \leq d$, indeed the sum of the elements of the columns of $A_{\mathscr{I}_d^*}$ is $\frac{d+1}{2}$, thus $\mathbf{1}_{n_d^*}$ is a linear combination of the other rows. We prove that $r_d = d$ by induction. From Example 3.1.2 and Example 3.1.3 we have that $r_3 = 3$. Let assume that $r_d = d$. We know that $r_{d+2} \leq d+2$ and we want to prove that the equality holds. From (3.17) we have rank$(A_{d+1}) = d$. The matrix $A_{d+1} = (\mathbf{1}_{n_d^*}//A_{\mathscr{I}_d^*})$ and the columns of $A_{\mathscr{I}_d^*}$ have exactly $M_d + 1 = (d+1)/2$ ones. Let B_{d+1} be a submatrix extracted from A_{d+1} by choosing d linearly independent columns of A_{d+1}. We have $B_{d+1} = (\mathbf{1}_d//\tilde{A}_{d+1})$, where $\tilde{A}_{d+1}$ is a submatrix of $A_{\mathscr{I}_d^*}$. Define $B_{d+2} = (\mathbf{1}_d//\mathbf{0}_d//\mathbf{0}_d//\tilde{A}_{d+1})$, where $\mathbf{0}_d$ is a row vector of all zeros. Let $\tilde{A}_{d+2} = (\mathbf{0}_d//\tilde{A}_{d+1})$. This is a submatrix of $A_{\mathscr{I}_{d+2}^*}$. Since $d+2$ is odd, the sum of the elements of the columns of $\tilde{A}_{d+2}$ is $M_d+1 = M_{d+2}$. We have that B_{d+2} is a submatrix of A_{d+2} and $r_{d+2} = \text{rank}(A_{d+2}) \geq d$. For $d \geq 3$, we can always find other two columns in A_{d+2} as follows: $(1,0,1,\boldsymbol{i}_1)$, where $\boldsymbol{i}_1 \in \mathcal{X}_d$ with $\sum_{k=1}^{d} i_{1,k} = M_{d+2} - 1$ and $(0,1,0,\boldsymbol{i}_2)$ where $\boldsymbol{i}_2 \in \mathcal{X}_d$ with $\sum_{k=1}^{d} i_{2,k} = m_{d+2}$. These two columns are independent from B_{d+2} thus $r_{d+2} = \text{rank}(A_{d+2}) \geq d+2$. Thus, $r_{d+2} = d+2$ and this proves the proposition. □

3.2 The Conway–Maxwell multivariate Bernoulli distributions

In this section, we introduce the Conway–Maxwell-binomial (CMB) distributions, a two-parameter family of univariate discrete distributions with finite support. We then focus on the Conway–Maxwell multivariate Bernoulli (CMMB) distributions, defined as the exchangeable Bernoulli distributions whose sums belong to the CMB family. We refer to these two classes as the CM families. In our analysis, we investigate the strongly Rayleigh property and dependence order properties within these families.

The pmf of a discrete random variable W with support on $\{0,\dots,d\}$ having the CMB distribution, denoted $W \sim \mathrm{CMB}_d(r,\nu)$, for $r \in (0,1)$ and $\nu \in \mathbb{R}$, is given by

$$f_W(k) = \mathbb{P}(W = k) = \frac{\binom{d}{k}^{\nu} r^k (1-r)^{d-k}}{S_d(r,\nu)}, \quad k \in \{0,1,\dots,d\}, \tag{3.18}$$

where

$$S_d(r,\nu) = \sum_{k=0}^{d} \binom{d}{k}^{\nu} r^k (1-r)^{d-k}.$$

When $\nu = 1$, the binomial distribution results. As $\nu \to +\infty$, W piles up at $d/2$ if d is even, or at $(d-1)/2$ and $(d+1)/2$ if d is odd, while, as $\nu \to -\infty$, the mass is at $W = 0$ and $W = d$. The expected value of W is given by

$$\mathbb{E}[W] = \sum_{k=0}^{d} k\mathbb{P}(W = k) = \sum_{k=0}^{d} \frac{\binom{d}{k}^{\nu} k r^k (1-r)^{d-k}}{S_d(r,\nu)}.$$

Finally, let us introduce the univariate polynomial

$$\mathcal{G}_{d,\nu}(z) = \sum_{k=0}^{d} \binom{d}{k}^{\nu} z^k, \quad z \in \mathbb{C}. \tag{3.19}$$

The pgf of a discrete random variable $W \sim \mathrm{CMB}_d(r,\nu)$ has been derived in Section 5 of [65], and can be expressed in terms of the polynomial $\mathcal{G}_{d,\nu}$ in (3.19) as follows,

$$\mathcal{P}_W(z) = \frac{(1-r)^d}{S_d(r,\nu)} \sum_{k=0}^{d} \binom{d}{k}^{\nu} \left(\frac{r}{1-r}\right)^k z^k = \frac{(1-r)^d}{S_d(r,\nu)} \mathcal{G}_{d,\nu}\left(\frac{r}{1-r} z\right), \quad z \in \mathbb{C}.$$

Let $\boldsymbol{X}_W^e \in \mathcal{E}_d$ be the exchangeable Bernoulli vector behind $W \sim \mathrm{CMB}_d(r,\nu)$ (see Definition 1.3.1). We name its distribution the Conway–Maxwell multivariate Bernoulli (CMMB) distribution with parameters $r \in (0,1)$ and $\nu \in \mathbb{R}$, denoted by $\mathrm{CMMB}_d(r,\nu)$. For the exchangeable Bernoulli random vectors, the two limit distributions for $\nu \to +\infty$ and $\nu \to -\infty$ respectively represent extreme negative and positive dependence, as we discussed in Section 2.3. For a fixed d and given such a parametrization of r in terms of ν and d, ν is a true dependence parameter. Obviously, if $\mathbb{E}[W] = \mu$, then $\boldsymbol{X}_W^e \in \mathcal{E}_d(\mu/d)$. As mentioned in [65], when we modify either r or ν, the marginals are not preserved. However, for a fixed p, we show that it is possible to find $r = \varphi(p,\nu,d)$ such that $\mathrm{CMMB}_d(r,\nu)$ is a subclass of the Fréchet class $\mathcal{B}_d(p)$, making it more attractive for modelling purposes.

Considering the case $r = 1/2$, we obtain

$$\mathbb{E}[W] = \sum_{k=0}^{d} k\mathbb{P}(W = k) = \sum_{k=0}^{d} \frac{\binom{d}{k}^{\nu} k (1/2)^d}{S_d(1/2,\nu)} = \frac{\sum_{k=0}^{d} \binom{d}{k}^{\nu} k (1/2)^d}{\sum_{k=0}^{d} \binom{d}{k}^{\nu} (1/2)^d} = \frac{\sum_{k=0}^{d} \binom{d}{k}^{\nu} k}{\sum_{k=0}^{d} \binom{d}{k}^{\nu}},$$

and, using the symmetry of the binomial coefficient, we find

$$2\mathbb{E}[W] = \frac{\sum_{k=0}^{d} \binom{d}{k}^{\nu} k}{\sum_{k=0}^{d} \binom{d}{k}^{\nu}} + \frac{\sum_{k=0}^{d} \binom{d}{k}^{\nu} (d-k)}{\sum_{k=0}^{d} \binom{d}{k}^{\nu}} = \frac{\sum_{k=0}^{d} \binom{d}{k}^{\nu} d}{\sum_{k=0}^{d} \binom{d}{k}^{\nu}} = d. \tag{3.20}$$

Therefore, $\mathbb{E}[W] = d/2$ for every $W \sim \mathrm{CMB}_d(1/2,\nu)$, for any $\nu \in \mathbb{R}$. The exchangeable Bernoulli vector $\boldsymbol{X}_W^e$ behind $W \sim \mathrm{CMB}_d(1/2,\nu)$ has Bernoulli marginals with mean 1/2, that is $\boldsymbol{X}_W^e \in \mathcal{E}_d(1/2)$; thus, the $\mathrm{CMMB}_d(1/2,\nu)$ distributions belong to the same class $\mathcal{E}_d(1/2)$. This implies that we can compare the strength of association between two $\mathrm{CMMB}_d(1/2,\nu)$ distributions using a dependence order, as these are partial orders defined on Fréchet classes of distributions with the same one-dimensional margins; see Section 2.2.

3.2.1 The CM families and the strongly Rayleigh property

The authors of [66] state that random vectors following a $\mathrm{CMMB}_d(r,\nu)$ distribution exhibit positive pairwise correlation if $\nu < 1$ and negative pairwise correlation if $\nu > 1$, with the special case $\nu = 1$ recovering independence. As already observed, the parameter ν drives the dependence that is, unusually, decreasing in ν. We want to study whether the CMMB distribution satisfies the strongly Rayleigh property for $\nu \geq 1$. The first step is to study whether the CMB distribution does. We first notice that if $\nu = 1$, then W satisfies the strongly Rayleigh property as $\mathcal{P}_W(z)$ has only real zeros. The next example shows that the CMB distribution is not always strongly Rayleigh for $\nu \geq 1$.

Example 3.2.1. *We consider the case $d = 4$, $r = 1/2$, and $\nu \geq 1$. The pgf of the random variable $W \sim CMB_4(1/2,\nu)$ is given by the univariate polynomial*

$$\mathcal{P}_W(z) = \frac{1}{2^4 S_4(1/2,\nu)}(1 + 4^\nu z + 6^\nu z^2 + 4^\nu z^3 + z^4).$$

As $\mathcal{P}_W(z)$ has strictly positive coefficients, it follows that its value is strictly positive for every non-negative real argument, i.e., no non-negative real number can be a zero of the polynomial. Considering $z < 0$, we can write

$$\frac{\mathcal{P}_W(z)}{z^2} = \frac{1}{2^4 S_4(1/2,\nu)}\left[\left(z^2 + \frac{1}{z^2}\right) + 4^\nu\left(z + \frac{1}{z}\right) + 6^\nu\right].$$

Using the change of variable $y = z + 1/z$, we define the function $Q(y,\nu) = y^2 + 4^\nu y + (6^\nu - 2)$, and we notice that if $y_1 \in \mathbb{R}$ is such that $Q(y_1,\nu) = 0$, then we find two zeros of $\mathcal{P}_W(z)$ by

$$z_{1,j} = \frac{y_1 + (-1)^j\sqrt{y_1^2 - 4}}{2},$$

for $j \in \{1,2\}$. If $|y_1| < 2$, it follows that $z_{1,1}$ and $z_{1,2}$ are complex roots. Moreover, by Bolzano's theorem, if $Q(2,\nu)Q(-2,\nu) < 0$, then there exists $y_1 \in (-2,2)$ such that $Q(y_1,\nu) = 0$ and, therefore, $\mathcal{P}_W(\boldsymbol{z})$ admits complex roots. Since $Q(2,\nu) = 2 + 2\cdot 4^\nu + 6^\nu > 0$ for every $\nu \in \mathbb{R}$, we need to study the sign of the continuous function $H(\nu) := Q(-2,\nu) = 2 - 2\cdot 4^\nu + 6^\nu$, as ν varies in $\mathbb{R}$. Since $H(1) = 0$, $H(2) > 0$, and $H'(1) = \log(6^6/2^{16}) < 0$, where $H'(\nu) = \mathrm{d}H(\nu)/\mathrm{d}\nu$, it follows that there exists $\bar{\nu} \in (1,2)$ such that $H(\bar{\nu}) = 0$ and $H(\nu) < 0$ for every $\nu \in (1,\bar{\nu})$. Therefore, for every $\nu \in (1,\bar{\nu})$, $\mathcal{P}_W(z)$ admits complex roots and $W \sim CMB_4(1/2,\nu)$ does not satisfy the strongly Rayleigh property. Numerically, one finds $\bar{\nu} \cong 1.14772$.

To prove our main result on the strongly Rayleigh property, that is to give the conditions on the parameters r and ν for the distribution $\mathrm{CMB}_d(r,\nu)$ to be strongly Rayleigh, we need some preliminary propositions.

Proposition 3.2.1. *If $W \sim CMB_d(r,\nu)$ satisfies the strongly Rayleigh property, then $W' \sim CMB_d(r',\nu)$ satisfies the strongly Rayleigh, for every choice of $r' \in (0,1)$.*

Proof. The pgfs of W and W' are respectively given by

$$\mathcal{P}_W(z) = \frac{(1-r)^d}{S_d(r,\nu)}\mathcal{G}_{d,\nu}\left(\frac{r}{1-r}z\right)$$

and

$$\mathcal{P}_{W'}(z) = \frac{(1-r')^d}{S_d(r',\nu)}\mathcal{G}_{d,\nu}\left(\frac{r'}{1-r'}z\right) = \left(\frac{1-r'}{1-r}\right)^d \frac{S_d(r,\nu)}{S_d(r',\nu)}\mathcal{P}_W\left(\frac{1-r}{r}\frac{r'}{1-r'}z\right).$$

Since W satisfies the strongly Rayleigh property, its pgf $\mathcal{P}_W(z)$ is real-rooted, it is easy to show that $\mathcal{P}_{W'}(z)$ is real-rooted and thus that W' satisfies the strongly Rayleigh property; see the stability preservers in Lemma 2.4 in [104]. □

A consequence of Proposition 3.2.1 is that the strongly Rayleigh property is independent of the parameter r. Hence, the conclusions in Example 3.2.1 can be extended to any $r \in (0,1)$. Moreover, we can prove our results on the strongly Rayleigh property for $r = 1/2$ without loss of generality. This choice simplifies the proofs, thanks to the following corollary.

Corollary 3.2.2. *The random variable $W \sim CMB_d(r,\nu)$ with $r \in (0,1)$ and $\nu \in \mathbb{R}$ satisfies the strongly Rayleigh property if and only if the polynomial $\mathcal{G}_{d,\nu}(z)$ in (3.19) is real-rooted.*

Proof. By Proposition 3.2.1 the strongly Rayleigh property is independent of r, therefore the random variable W satisfies the strongly Rayleigh property if and only if $W' \sim \mathrm{CMB}_d(1/2,\nu)$ satisfies the strongly Rayleigh property. The pgf of W' is given by

$$\mathcal{P}_{W'}(z) = \frac{(1/2)^d}{S_d(1/2,\nu)}\mathcal{G}_{d,\nu}(z),$$

that is real-rooted if and only if the polynomial $\mathcal{G}_{d,\nu}(z)$ in (3.19) is real-rooted. □

The proof of our main result is based on the following lemma.

Lemma 3.2.3. *If $W \sim CMB_d(r,\nu)$ with $r \in (0,1)$ and $\nu \in \mathbb{R}$ satisfies the strongly Rayleigh property, then $W' \sim CMB_d(r,\nu+1)$ satisfies the strongly Rayleigh property.*

Proof. Using Corollary 3.2.2, the statement reduces to show that if $\mathcal{G}_{d,\nu}(z)$ is real-rooted, then $\mathcal{G}_{d,\nu+1}(z)$ is real-rooted. Observe that $\mathcal{G}_{d,\nu+1}(z) = T[\mathcal{G}_{d,\nu}(z)]$, where T is the linear operator on $\mathbb{C}[z]$ defined by $T[z^k] = \lambda(k)z^k$, with $\lambda\colon \mathbb{N} \to \mathbb{R}$ given by

$$\lambda(k) = \begin{cases} \binom{d}{k}, & \text{if } 0 \le k \le d, \\ 0, & \text{if } k > d. \end{cases}$$

It is therefore sufficient to prove that the operator T preserves real-rootedness. In [91] (see also Theorem 1.7 in [13]), Pólya and Schur show that T is a real-rootedness preserver if and only if $T[(1+z)^n]$ is real-rooted with all zeros of the same sign, for all nonnegative integers n. We have

$$T[(1+z)^n] = T\left[\sum_{k=0}^{n}\binom{n}{k}z^k\right] = \sum_{k=0}^{n}\binom{n}{k}T[z^k] = \sum_{k=0}^{\min\{n,d\}}\binom{n}{k}\binom{d}{k}z^k.$$

Consider $n \in \{0,\dots,d-1\}$. Theorem 2(v) in [36] states that the hypergeometric polynomial, defined by

$$F(-n,b;c;z) = \sum_{k=0}^{\infty}\frac{(-n)_k(b)_k}{(c)_k}\frac{x^k}{k!},$$

with $n \in \mathbb{N}$, $b,c \in \mathbb{R}$, $c > 0$, and $b < -n$, is real-rooted with all negative zeros; we write $(a)_k$ to denote the Pochhammer's symbol, i.e., $(a)_k = a(a+1)\cdots(a+k-1)$ for $k \in \mathbb{N}$, and $(a)_0 = 1$, for every $a \in \mathbb{C}$. By setting $b = -d$ and $c = 1$, we have

$$\begin{aligned}
F(-n,-d;1;z) &= \sum_{k=0}^{\infty}\frac{(-n)_k(-d)_k}{(1)_k}\frac{x^k}{k!}\\
&= \sum_{k=0}^{n}\frac{(-n)(-n+1)\cdots(-n+k-1)(-d)(-d+1)\cdots(-d+k-1)}{k!}\frac{x^k}{k!}\\
&= \sum_{k=0}^{n}(-1)^{2k}\frac{n!}{k!(n-k)!}\frac{d!}{k!(d-k)!}x^k = \sum_{k=0}^{n}\binom{n}{k}\binom{d}{k}x^k = T[(1+z)^n].
\end{aligned}$$

Therefore, for $n \in \{0,\dots,d-1\}$, the polynomial $T[(1+z)^n]$ is real-rooted with all negative zeros. When $n \in \{d+1,d+2,\dots\}$, by applying again Theorem 2(v) of [36], we have that $T[(1+z)^n] = F(-d,-n;1;z)$ is real-rooted with all negative zeros. It remains to show that $T[(1+n)^d]$ is real-rooted with all negative zeros. For every $z \neq 1$, we can write

$$T[(1+z)^d] = \sum_{k=0}^{d}\binom{d}{k}^2 z^k = (1-z)^d\mathcal{L}_d\left(\frac{1+z}{1-z}\right), \tag{3.21}$$

where $\mathcal{L}_d(x)$ is the Legendre polynomial of order d. Indeed, from [88, p.162], the Legendre polynomial $\mathcal{L}_d(x)$ can be written as

$$\mathcal{L}_d(x) = \sum_{k=0}^{d}\binom{d}{k}^2\left(\frac{x-1}{2}\right)^{d-k}\left(\frac{x-1}{2}\right)^{k},$$

and, after standard computations, (3.21) follows. It is well-known that $\mathcal{L}_d(x)$ has d real and distinct zeros, all of which lie in the interval $(-1,1)$; see, e.g., [103]. We denote the zeros of $\mathcal{L}_d(x)$ by $x_1,\dots,x_d$. Therefore, the zeros of $T[(1+z)^d]$ are given by

$$z_j = \frac{x_j-1}{x_j+1},$$

for $j \in \{1,\dots,d\}$, and we can conclude that $z_1,\dots,z_d$ are real, distinct, and lies in $(-\infty,0)$. Thus, $T[(1+z)^d]$ is real-rooted with all negative zeros, and this completes the proof. □

We can now state our main result about the strongly Rayleigh property.

Theorem 3.2.4. *The random variable $W \sim CMB_d(r,\nu)$ with $r \in (0,1)$ and $\nu \in \mathbb{N}_+$ satisfies the strongly Rayleigh property. The random vector $\boldsymbol{X}^e_W \sim CMMB_d(r,\nu)$ behind satisfies the strongly Rayleigh property.*

Proof. Using Lemma 3.2.3, the claim follows by induction since the random variable $W' \sim \mathrm{CMB}_d(r,1)$, with $r \in (0,1)$, satisfies the strongly Rayleigh property. The exchangeable vector behind any strongly Rayleigh random variable satisfies the strongly Rayleigh property by Theorem 3.8 in [15]. □

In the following proposition, we provide conditions to construct a chain with respect to the convex order in $\mathcal{D}(\mu)$, that is the first step to build a chain with respect to the supermodular order in $\mathcal{B}_d(\mu/d)$.

Proposition 3.2.5. *Let $W_1 \sim CMB_d(r_1,\nu_1)$ and $W_2 \sim CMB_d(r_2,\nu_2)$ such that $\mathbb{E}[W_1] = \mathbb{E}[W_2] = \mu$ and $\nu_1 \leq \nu_2$. Then, $W_2 \preceq_{cx} W_1$ holds.*

Proof. We denote by f_1 and f_2 the pmfs and by F_1 and F_2 the cdfs of W_1 and W_2, respectively. Let also denote $g := f_1 - f_2$ and $G := F_1 - F_2$. Since $\mathbb{E}[W_1] = \mathbb{E}[W_2]$, the cdfs F_1 and F_2 cross at least once (see the proof of Theorem 3.A.5 in [93]), and G changes sign at least once. Since $G(n) = \sum_{k=0}^{n} g(k)$ and $G(d) = 0$, if G changes sign at least once, then g changes sign at least twice. If we prove that g changes sign exactly twice, with sign sequence $+,-,+$, by Theorem 3.A.44 in [93], we have $W_2 \preceq_{cx} W_1$. The sign changes of g coincide with those of $\ell := \ln(f_1/f_2)$. Indeed,

$$\ell(k) > 0 \iff \ln\frac{f_1(k)}{f_2(k)} > 0 \iff \frac{f_1(k)}{f_2(k)} > 1 \iff f_1(k) > f_2(k) \iff g(k) > 0.$$

Moreover, $\ell(k)$ is a discrete convex function, i.e., $\ell(k+1) + \ell(k-1) - 2\ell(k) \geq 0$. Indeed, from (3.18), for any $k \in \{0,1,\dots,d\}$, we have

$$\frac{f_1(k)}{f_2(k)} = \left(\frac{1-r_1}{1-r_2}\right)^d \frac{S_d(r_2,\nu_2)}{S_d(r_1,\nu_1)} \binom{d}{k}^{\nu_1-\nu_2} \left(\frac{r_1(1-r_2)}{r_2(1-r_1)}\right)^k,$$

thus

$$\ell(k) = d\ln\left(\frac{1-r_1}{1-r_2}\right) + \ln\left(\frac{S_d(r_2,\nu_2)}{S_d(r_1,\nu_1)}\right) + (\nu_1-\nu_2)\ln\binom{d}{k} + k\ln\left(\frac{r_1(1-r_2)}{r_2(1-r_1)}\right),$$

and

$$\begin{aligned}\ell(k+1) + \ell(k-1) - 2\ell(k) &= (\nu_1-\nu_2)\ln\left(\binom{d}{k+1}\binom{d}{k-1}\binom{d}{k}^{-2}\right)\\ &= (\nu_1-\nu_2)\ln\left(\frac{k}{k+1}\frac{d-k}{d-k+1}\right) \geq 0\end{aligned}$$

for any k. Therefore, since ℓ is a convex function, it changes sign at most twice with sign sequence $+,-,+$, and so does g. We can conclude that g changes sign exactly twice, with sign sequence $+,-,+$, and $W_2 \preceq_{cx} W_1$. □

Corollary 3.2.6. *Let $\boldsymbol{X}_1^e$ and $\boldsymbol{X}_2^e$ be two exchangeable Bernoulli random vectors such that*

$$\sum_{j=1}^{d} X_{1,j}^e \overset{\mathcal{L}}{=} W_1 \quad \text{and} \quad \sum_{j=1}^{d} X_{2,j}^e \overset{\mathcal{L}}{=} W_2,$$

where $W_1 \sim CMB_d(r_1, \nu_1)$ and $W_2 \sim CMB_d(r_2, \nu_2)$, with $\mathbb{E}[W_1] = \mathbb{E}[W_2] = \mu$ and $\nu_1 \leq \nu_2$, and $\overset{\mathcal{L}}{=}$ means equality in distribution. Then, $\boldsymbol{X}_1^e, \boldsymbol{X}_2^e \in \mathcal{B}_d(p)$, with $p = \mu/d$, and $\boldsymbol{X}_2^e \preceq_{sm} \boldsymbol{X}_1^e$.

Proof. By Proposition 3.2.5, we have $W_2 \preceq_{cx} W_1$, which implies $\boldsymbol{X}_2^e \preceq_{sm} \boldsymbol{X}_1^e$; see Section 3 of [49]. □

The above corollary proves that, for any given $\mu \in (0, d)$, there is a chain with respect to $\preceq_{sm}$, included in the Fréchet class $\mathcal{B}_d(p)$, with $p = \mu/d$, from the minimum under the supermodular order to the upper Fréchet bound.

Example 3.2.2. *For a fixed triple (d, ν, p), the value r is obtained such that $\mathbb{E}[W] = dp$ by optimization. For example, assume that $p = 1/3$ and $d = 9$. For $\nu \in \{2,3,4,5\}$, the corresponding values of r and $Var(W)$ are reported in Table 3.1. The values of $Var(W)$*

ν	r	Var(W)
2	0.21367747	1.0748125
3	0.12810820	0.7319828
4	0.07352747	0.5547126
5	0.04105203	0.4452527

Table 3.1: Values of r and Var(W) for different ν.

are decreasing in ν as a consequence of Corollary 3.2.6. When $\nu = 1$ (independence), $r = 1/3$ and $Var(W) = 2$. Figure 3.1 shows the (approximated) values of the pmfs, which, as expected, are more concentrated around the mean as ν increases.

Since we do not have an analytical expression for r as a function of p, r, and ν, we cannot explicitly find the chain of pmfs unless we choose $r = 1/2$. For this reason, we now consider the CM families with $r = 1/2$. Indeed, as discussed at the end of Section 3.2, if $r = 1/2$ the exchangeable Bernoulli vector $\boldsymbol{X}_W^e$ behind $W \sim \text{CMB}_d(1/2, \nu)$ belongs to $\mathcal{E}_d(1/2)$. The above corollary proves that CMMB$(1/2, \nu)$ is a chain with respect to the supermodular order in $\mathcal{B}_d(1/2)$. As $\nu \to +\infty$, the CMB distribution converges to a distribution with support on $\{(d-1)/2, (d+1)/2\}$ if d is odd, and on $\{d/2\}$ if d is even. Conversely, as $\nu \to -\infty$, it converges to a distribution with support on $\{0, d\}$. Consequently, the CMMB$(1/2, \nu)$ converges to the minimum under the supermodular order as $\nu \to +\infty$ and to the upper Fréchet bound as $\nu \to -\infty$. Therefore, CMMB$_d(1/2, \nu)$ is a well-ordered chain with respect to the supermodular order in $\mathcal{B}_d(1/2)$. This means that the family is able to model each level of dependence with only one parameter. Furthermore, the subfamily CMMB$_d(1/2, n)$, for $n \in \mathbb{N}_+$, is a chain under the supermodular order of pmfs that satisfy the strongly Rayleigh property.

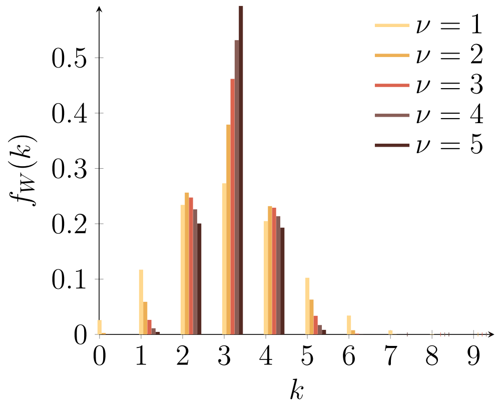

Figure 3.1: Histogram of the values of pmfs from Example 3.2.2.

3.2.2 Non-exchangeable extensions of the CMMB family

This section presents two ways to move from the $\mathrm{CMMB}_d(1/2, \nu)$ distribution to obtain a d-dimensional strongly Rayleigh and non-exchangeable Bernoulli vector. The first approach relies on a geometrical representation of the Bernoulli pmfs and allows us to include in the family the limit cases $\nu \to \pm\infty$. The second approach is based on a stochastic representation and allows us to choose the new vector belonging to a given Fréchet class, i.e., with given marginal means.

Geometrical approach

Consider a random variable $W \sim \mathrm{CMB}_d(1/2, \nu)$, with $\nu \in \mathbb{R}$. Since its expected value is $d/2$, we have $W \in \mathcal{D}(d/2)$. As described in Section 1.3.1, the class $\mathcal{D}(d/2)$ is a convex polytope with extremal points

$$r^D_{j_1,j_2}(k) = \begin{cases} \frac{j_2-\mu}{j_2-j_1}, & \text{for } k = j_1, \\ \frac{\mu-j_1}{j_2-j_1}, & \text{for } k = j_2, \\ 0, & \text{elsewhere}, \end{cases}$$

for any $j_1 \in \{0, \dots, M_d\}$ and $j_2 \in \{M_d+1, \dots, d\}$, where $M_d = \lfloor d/2 \rfloor$ is the highest integer lower than or equal to $d/2$. Therefore, we have

$$f_W = \sum_{j_1=0}^{M_d} \sum_{j_2=M_d+1}^{d} \lambda_{j_1,j_2} r^D_{j_1,j_2},$$

where λ_{j_1,j_2}, for $j_1 \in \{0, \dots, M_d\}$ and $j_2 \in \{M_d + 1, \dots, d\}$, are non-negative real numbers summing to 1. Moreover, we also know that f_W is symmetric, i.e., $f_W(k) = f_W(d-k)$, for every $k \in \{0, \dots, d\}$. This implies that we can write f_W as a convex combination involving only the symmetric extremal points,

$$f_W(k) = \sum_{j=0}^{M_d} \lambda_j r^D_j(k), \tag{3.22}$$

where $\lambda_j \geq 0$, $\lambda_0 + \cdots + \lambda_{M_d} = 1$, and, for $j \in \{0, \ldots, M_d\}$,

$$r_j^D(k) := r_{j,d-j}^D(k) = \begin{cases} \frac{1}{2}, & \text{for } k \in \{j, d-j\}, \\ 0, & \text{otherwise.} \end{cases}$$

By equating (3.22) and (3.18) with $r = 1/2$, we get, for $j \in \{0, \ldots, M_d\}$,

$$\lambda_j = \frac{2\binom{d}{j}^\nu}{\sum_{y=0}^d \binom{d}{y}^\nu}. \tag{3.23}$$

Since there exists a one-to-one correspondence between exchangeable Bernoulli distributions and discrete distributions, the exchangeable Bernoulli distribution f^e of a random vector $\boldsymbol{X}^e$ such that $X_1^e + \cdots + X_d^e \overset{\mathcal{L}}{=} W$ can be written as

$$f^e = \sum_{j=0}^{M_d} \lambda_j r_j^e, \tag{3.24}$$

where, for every $\boldsymbol{x} \in \mathcal{X}_d$,

$$r_j^e(\boldsymbol{x}) = \begin{cases} \frac{1}{2}\binom{d}{j}^{-1}, & \text{for } \boldsymbol{x} \in \mathcal{X}_d^j, \\ \frac{1}{2}\binom{d}{j}^{-1}, & \text{for } \boldsymbol{x} \in \mathcal{X}_d^{d-j}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.25}$$

The above representation includes the exchangeable pmf that is minimal under the supermodular order $r_{M_d}^e$ when $\lambda_{M_d} = 1$, and the upper Fréchet bound r_0^e when $\lambda_0 = 1$. The extremal pmf $r_{M_d}^e$ lies in a polytope contained in the Fréchet class $\mathcal{SB}_d$: the convex polytope of all Σ_{cx}-smallest elements in $\mathcal{SB}_d$ that we have characterized in Section 3.1. As we illustrate at the end of the following example, one can select pmfs in this sub-polytope distinct from $r_{M_d}^e$ to construct strongly Rayleigh distributions that are no longer exchangeable. This approach follows a procedure similar to the one described in Example 2.5.6.

Example 3.2.3. Three-dimensional case. *For $d = 3$, there are only two symmetric extremal points of $\mathcal{D}_3(3/2)$, which are given by*

$$r_0^D(k) = \begin{cases} \frac{1}{2}, & \text{if } k \in \{0,3\}, \\ 0, & \text{if } k \in \{1,2\}, \end{cases} \quad \text{and} \quad r_1^D(k) = \begin{cases} 0, & \text{if } k \in \{0,3\}, \\ \frac{1}{2}, & \text{if } k \in \{1,2\}, \end{cases}$$

and for $W \sim CMB_3(1/2, \nu)$, from (3.22) we have

$$f_{\lambda,W}(k) := f_W(k) = \lambda r_0^D(k) + (1-\lambda) r_1^D(k), \quad \lambda \in [0,1], \tag{3.26}$$

where $\lambda = 1/(1+3^\nu)$ derives from (3.23). The extremal points corresponding to $\lambda = 0$ and $\lambda = 1$ are the limit cases for the CMMB distribution $f_{\lambda,w}^e$ underlying W, that can be included in the family using this representation. With $\lambda = 0$, we have the minimal pmf under the supermodular order, if $\lambda = 1$ we have the upper Fréchet bound, and if $\lambda = 0.25$ we have independence. The extremal point r_1^D is the minimal convex pmf in $\mathcal{D}_3(3/2)$. Therefore, $f_{\lambda,W}^e$ is a chain with respect to the supermodular order.

In what follows, we prove that the strongly Rayleigh property holds for any $\nu \geq 1$, when $d = 3$. The pgf of W is given by

$$\mathcal{P}_W(z) = \frac{1}{2(1+3^\nu)}(1+z^3) + \frac{3^\nu}{2(1+3^\nu)}(z+z^2).$$

This polynomial is real-rooted if and only if the joint pgf $\mathcal{P}^e$ of the corresponding exchangeable Bernoulli distribution is stable, where

$$\mathcal{P}^e(z) = \frac{1}{2(1+3^\nu)}(1+z_1z_2z_3) + \frac{3^\nu}{6(1+3^\nu)}(z_1+z_2+z_3+z_1z_2+z_1z_3+z_2z_3).$$

Since this polynomial is multi-affine and exchangeable, it is stable if and only if condition (2.9) holds for $j_1 = 1$ and $j_2 = 2$. This condition simplifies to

$$(3^{\nu-1}+3^{\nu-1}x_3)^2 - (3^{\nu-1}+x_3)(1+3^{\nu-1}x_3) \geq 0,$$

for every $x_3 \in \mathbb{R}$. After standard computations, the condition is verified if and only if $\nu \geq 1$. By Proposition 3.2.1, any CMB distribution with $d = 3$, $r \in (0,1)$, and $\nu \geq 1$ satisfies the strongly Rayleigh property.

We conclude this example showing how the geometrical representation can be used to construct non-exchangeable strongly Rayleigh pmfs. From (3.24) we have that a CMMB pmf in $\mathcal{B}_3(1/3)$ has the representation

$$f^e_{\lambda,W}(\boldsymbol{x}) := \lambda r^e_0(\boldsymbol{x}) + (1-\lambda)r^e_1(\boldsymbol{x}), \quad \lambda \in [0,1],$$

where r^e_j are given in (3.25). Moreover, r^e_1 corresponds to the minimal convex extremal point r^D_1, and, as already highlighted in Remark 3.1.2, it belongs to $\mathcal{PB}_d$. Therefore, it belongs to $\mathcal{PB}^*_d$, that is, the set of Σ_{cx}-smallest palindromic Bernoulli pmfs. This class is a convex polytope whose extremal points are r^{cx}_1, r^{cx}_2, and r^{cx}_3, given by $r^{cx}_k(\boldsymbol{e}_k) = r^{cx}_k(\mathbf{1}_3 - \boldsymbol{e}_k) = 1/2$, for $k \in \{1, 2, 3\}$, where $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ is the canonical basis of $\mathbb{R}^3$. In particular, we have

$$f^e_{\lambda,W}(\boldsymbol{x}) = \lambda r^e_0(\boldsymbol{x}) + (1-\lambda)(r^{cx}_1(\boldsymbol{x})/3 + r^{cx}_2(\boldsymbol{x})/3 + r^{cx}_3(\boldsymbol{x})/3), \quad \lambda \in [0,1].$$

Starting from $f^e_{\lambda,W}$, we can find a new pmf f^* in a neighborhood of $f^e_{\lambda,W} \in \mathcal{B}_3(1/2)$ that is no more exchangeable, but still SR. For example,

$$f^*(k) = 0.1r^e_0(k) + 0.9((1/3-0.05)r^{cx}_1(k) + 1/3r^{cx}_2(k) + (1/3+0.05)r^{cx}_3(k))$$

belongs to $\mathcal{B}_3(1/2)$ and satisfies the NLC. Moreover, by Proposition B.0.1 in Appendix B, one can easily show that f^* satisfies the strongly Rayleigh property.

Stochastic approach

Consider the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, i.e., the class of d-dimensional Bernoulli pmfs with marginal means $\boldsymbol{p} = (p_1, \dots, p_d)$. This stochastic approach aims to find a non-exchangeable Bernoulli random vector that belongs to a given Fréchet class $\mathcal{B}_d(\boldsymbol{p})$, with $0 \leq p_1 \leq \cdots \leq p_d \leq 1/2$, and that inherits some dependence property of an exchangeable Bernoulli random vector $\boldsymbol{Y}$ with $\text{CMMB}_d(1/2, \nu)$ distribution. Consider a vector $\boldsymbol{K} = (K_1, \dots, K_d)$

of d independent Bernoulli random variables, with $\boldsymbol{K} \in \mathcal{B}_d(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1 \dots, \theta_d)$ with $\theta_j = 2p_j$, for every $j \in \{1, \dots, d\}$. Let us define a Bernoulli random vector $\boldsymbol{X}$ such that $X_j = K_j Y_j$, for every $j \in \{1, \dots, d\}$. By construction, it follows that $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$. To study negative dependence properties of the random vector $\boldsymbol{X}$, we state the following proposition.

Proposition 3.2.7. *Let $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ be a multivariate Bernoulli random vector that satisfies the strongly Rayleigh property, and let K be a Bernoulli random variable independent of $\boldsymbol{Y}$ with mean $p_K \in (0,1)$. The random vector $\boldsymbol{X}_1 = (Y_1 \cdot K, Y_2, \dots, Y_d)$ satisfies the strongly Rayleigh property.*

Proof. The pgf of $\boldsymbol{Y}$ can be written as

$$\mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z}) = \mathbb{E}[\boldsymbol{z}^{\boldsymbol{Y}}] = q_1 \mathbb{E}(\boldsymbol{z}^{\boldsymbol{Y}} | Y_1 = 0) + p_1 \mathbb{E}(\boldsymbol{z}^{\boldsymbol{Y}} | Y_1 = 1) = \mathcal{P}_{\boldsymbol{J}}(\boldsymbol{z})\big|_{z_1=0} + z_1 \frac{\partial}{\partial z_1} \mathcal{P}_{\boldsymbol{J}}(\boldsymbol{z}),$$

where $p_1 = 1 - q_1 = \Pr(J_1 = 1)$. Therefore, the pgf of $\boldsymbol{X}_1$ is given by

$$\begin{aligned}
\mathcal{P}_{\boldsymbol{X}_1}(\boldsymbol{z}) &= \mathbb{E}[\boldsymbol{z}^{\boldsymbol{X}_1}] = \mathbb{E}[\mathbb{E}(\boldsymbol{z}^{\boldsymbol{X}_1} | K)] = p_K \mathbb{E}[\boldsymbol{z}^{\boldsymbol{X}_1} | K = 1] + q_K \mathbb{E}[\boldsymbol{z}^{\boldsymbol{X}_1} | K = 0] \\
&= p_K \mathbb{E}[\boldsymbol{z}^{\boldsymbol{Y}}] + q_K \mathbb{E}[z_2^{Y_2} \cdots z_d^{Y_d}] = p_K \mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z}) + q_K \mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z})\big|_{z_1=1} \\
&= p_K \left(\mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z})\big|_{z_1=0} + z_1 \frac{\partial}{\partial z_1} \mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z}) \right) + q_K \left(\mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z})\big|_{z_1=0} + 1 \cdot \frac{\partial}{\partial z_1} \mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z}) \right) \\
&= \mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z})\big|_{z_1=0} + (q_K + p_K z_1) \frac{\partial}{\partial z_1} \mathcal{P}_{\boldsymbol{Y}}(\boldsymbol{z}) = \mathcal{P}_{\boldsymbol{Y}}(q_K + p_K \cdot z_1, z_2, \dots, z_d),
\end{aligned}$$

where $q_K = 1 - p_K$. If $(z_1, \dots, z_d) \in \mathcal{H}^d$, then $(q_K + p_K \cdot z_1, z_2, \dots, z_d) \in \mathcal{H}^d$, and

$$\mathcal{P}_{\boldsymbol{X}_1}(z_1, z_2, \dots, z_d) = \mathcal{P}_{\boldsymbol{Y}}(q_K + p_K \cdot z_1, z_2, \dots, z_d) \neq 0,$$

and $\boldsymbol{X}_1$ is SR. □

By iteratively applying Proposition 3.2.7 to each component of the random vector $\boldsymbol{Y}$, we deduce that $\boldsymbol{X}$ satisfies the strongly Rayleigh property whenever $\boldsymbol{J}$ does. Moreover, the pgf of $\boldsymbol{X}$ reads

$$\mathcal{P}_{\boldsymbol{X}}(\boldsymbol{z}) = \mathcal{P}_{\boldsymbol{Y}}(\mathcal{P}_{K_1}(z_1), \dots, \mathcal{P}_{K_d}(z_d)) = \mathcal{P}_{\boldsymbol{Y}}((1 - \theta_1 + \theta_1 z_1), \dots, (1 - \theta_d + \theta_d z_d)).$$

Chapter 4

Bernoulli and copulas

This chapter explores the connection between multivariate Bernoulli distributions and the construction of copulas. Through specific stochastic representations (introduced in [11, 7, 75]), we show how the dependence properties investigated in the previous chapters can be transferred to the continuous setting, providing new insights into the structure of aggregate risk for general random variables.

We first recall the definition and the fundamental properties of copulas. For a detailed discussion see the comprehensive works [39, 79]. A d-dimensional copula $C\colon [0,1]^d \to [0,1]$ is a multivariate cumulative distribution function (cdf) whose one-dimensional margins are uniform on the unit interval [0,1]. The cornerstone of this theory is Sklar's Theorem [95], which states that any d-dimensional cdf F with univariate marginals $F_1, \dots, F_d$ can be written as

$$F(y_1, \dots, y_d) = C(F_1(y_1), \dots, F_d(y_d)), \quad \text{for all } \boldsymbol{y} \in \mathbb{R}^d,$$

for some copula C. A crucial aspect is the uniqueness of the copula C. While the copula is uniquely determined on $[0,1]^d$ if the margins are strictly continuous, for discrete distributions, the copula is uniquely defined only on its support. However, our focus is not on studying the copulas of Bernoulli random vectors. We use Bernoulli distributions and their underlying dependence structure to construct copulas through stochastic representations. Furthermore, copulas are naturally bounded. Since copulas are d-dimensional cdfs belonging to a Fréchet class, specifically, the class with fixed uniform marginals on [0,1], the definitions of extremal negative dependence introduced in Section 2.3 apply perfectly to this continuous setting. In this chapter, we characterize copulas that are Σ_{cx}-smallest within their Fréchet classes, as well as Σ-countermonotonic copulas. Furthermore, we will provide an illustrative example demonstrating that, while the equivalence between being Σ_{cx}-smallest and Σ-countermonotonicity holds for multivariate Bernoulli vectors, it fails to hold in the general framework of copulas.

We also study how the choice of the copula directly determines the behavior of the aggregate risk $S_{\boldsymbol{Y}} = Y_1 + \dots + Y_d$, when the distributions of Y_j are fixed. In particular, we discuss in detail the case of identically distributed risks with dependence described by a specific class of copulas. Indeed, establishing dependence orders between copulas (such as the concordance order) allows for the comparison of aggregate risks via the convex order, and for the study of bounds of linear functionals $\mathbb{E}[\phi(\boldsymbol{Y})]$ and convex risk measures like the expected shortfall and the entropic risk measure.

The chapter is structured as follows: Section 4.1 focuses on the construction of copulas with symmetric Bernoulli distributions, which we have previously studied in Section 3.1. We investigate extremal copulas, extremal mixture copulas, and the Farlie–Gumbel–Morgenstern (FGM) family. By applying our previous results on Σ_{cx}-smallest symmetric Bernoulli random vectors, we identify classes of extremal negative dependent copulas. Furthermore, we provide a detailed analysis of their pairwise dependence structures. Section 4.3 considers a generalization of FGM copulas that can be constructed from any multivariate Bernoulli vector via a stochastic representation. This framework allows us to extend the study of aggregate risk beyond the Bernoulli case: we investigate the convex order of the sums of random variables $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ whose dependence is described by these generalized copulas.

4.1 Extremal negative dependent copulas

Two classes of copulas can be constructed from symmetric Bernoulli distributions: extremal mixture copulas and FGM copulas. Both of these classes inherit some dependence properties from $\mathcal{SB}_d$. In particular, the results on extremal negative dependence within the Bernoulli class allow us to find a sub-family of the class of extremal copulas that are Σ-countermonotonic, i.e., they represent extremal negative dependence in the entire class of copulas, and a class of copulas with the joint mixability property.

Extremal Mixture Copulas

In this section, we study the class of extremal mixture copulas. These copulas are in a one-to-one correspondence with the palindromic Bernoulli distributions (see [75]) that coincide with the kernel of the map $\mathcal{H}$ (Proposition 3.1.1).

Definition 4.1.1. *Given a standard uniform random variable U, an extremal copula with index set $J \subseteq \{1, \dots, d\}$ is the distribution function of the d-dimensional random vector $\boldsymbol{V} = (V_1, \dots, V_d)$ where $V_j \overset{d}{=} U$ if $j \in J$, and $V_j \overset{d}{=} 1 - U$ if $j \notin J$, for every $j \in \{1, \dots, d\}$.*

There exist 2^{d-1} different d-dimensional extremal copulas. It is possible to infer the explicit form of the copulas, indeed, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$, one has

$$C_{\boldsymbol{i}}(\boldsymbol{u}) = (\min_{j \in J_{\boldsymbol{i}}} u_j + \min_{j \in \bar{J}_{\boldsymbol{i}}} u_j - 1)^+, \quad \boldsymbol{u} \in [0,1]^d,$$

where we recall that, for $\boldsymbol{i} \in \mathcal{X}_{d-1}$, $J_{\boldsymbol{i}} = \{j \in \{1, \dots, d-1\} : i_j = 1\}$ is the set of indexes corresponding to 1s in $\boldsymbol{i}$, and where $\bar{J}_{\boldsymbol{i}} = \{1, \dots, d\} \setminus J_{\boldsymbol{i}}$ and $y^+ = \max(0, y)$; we use the convention $\min_{j \in \emptyset} u_j = 1$.

One may consider a wider class of copulas by taking mixtures of extremal copulas; see [100].

Definition 4.1.2. *An extremal mixture copula C is a copula of the form*

$$C = \sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} w_{\boldsymbol{i}} C_{\boldsymbol{i}},$$

where, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$, $C_{\boldsymbol{i}}$ is the extremal copula with index set $J_{\boldsymbol{i}}$ and the weights $w_{\boldsymbol{i}}$ are such that $w_{\boldsymbol{i}} \geq 0$, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$, and $\sum_{\boldsymbol{i} \in \mathcal{X}_{d-1}} w_{\boldsymbol{i}} = 1$.

We denote by $\mathcal{C}_d^{\mathrm{EM}}$ the class of extremal mixture copulas. The following proposition has been proved in [75] and states that there exists a non-injective map between the class of multivariate Bernoulli distributions and the class of extremal mixture copulas.

Proposition 4.1.1. *Let U be a standard uniform random variable and $\boldsymbol{X}$ a d-dimensional multivariate Bernoulli random vector with pmf f. Let $\boldsymbol{X}$ and U be independent. Then the cdf of the uniform random vector*

$$\boldsymbol{V} = U\boldsymbol{X} + (1-U)(\mathbf{1}_d - \boldsymbol{X}) \tag{4.1}$$

is an extremal mixture copula with weights given, for each $\boldsymbol{i} \in \mathcal{X}_{d-1}$, by

$$w_{\boldsymbol{i}} = f(\boldsymbol{s_i}) + f(\mathbf{1}_d - \boldsymbol{s_i}), \tag{4.2}$$

where $\boldsymbol{s_i} = (\boldsymbol{i}//0)$.

Given an extremal mixture copula with weights $w_{\boldsymbol{i}}$, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$, there exist infinitely many Bernoulli distributions satisfying (4.2). However, it is possible to identify a unique Bernoulli distribution by considering the class of palindromic Bernoulli distributions $\mathcal{PB}_d$, characterized by the constraint $f(\boldsymbol{s_i}) = f(\mathbf{1}_d - \boldsymbol{s_i})$, for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$. Therefore, the class $\mathcal{PB}_d$ is in a one-to-one correspondence with the family of extremal mixture copulas $\mathcal{C}_d^{\mathrm{EM}}$, see [75]:

$$\mathcal{PB}_d \longleftrightarrow \mathcal{C}_d^{\mathrm{EM}}. \tag{4.3}$$

In particular, the extremal copulas correspond to the pmfs of the basis $\mathcal{K}$ in (3.2).

The results of Section 3.1.1 are useful to explore the concept of negative dependence in the class of extremal mixture copulas. We conclude this section with three results within the class $\mathcal{C}_d^{\mathrm{EM}}$.

Proposition 4.1.2. *Let $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{SB}_d$ and let $\boldsymbol{V}$ and $\boldsymbol{V}'$ be the uniform random vectors with extremal mixture copulas constructed from $\boldsymbol{X}$ and $\boldsymbol{X}'$, respectively, via the stochastic representation in (4.1). Then,*

$$\sum_{j=1}^d X_j \leq_{cx} \sum_{j=1}^d X_j' \iff \sum_{j=1}^d V_j \leq_{cx} \sum_{j=1}^d V_j'.$$

Proof: Given a Bernoulli random vector $\boldsymbol{X} \in \mathcal{SB}_d$, from the stochastic representation in (4.1), we have

$$\sum_{j=1}^d V_j|(U=u) = \sum_{j=1}^d uX_j + (1-u)(1-X_j) = d(1-u) + (2u-1)\sum_{j=1}^d X_j.$$

Since $f(ax+b)$, with $a, b \in \mathbb{R}$, is convex if f is a convex function, it follows that

$$\sum_{j=1}^d X_j \leq_{\mathrm{cx}} \sum_{j=1}^d X_j' \implies \sum_{j=1}^d V_j|(U=u) \leq_{\mathrm{cx}} \sum_{j=1}^d V_j'|(U=u) \implies \sum_{j=1}^d V_j \leq_{\mathrm{cx}} \sum_{j=1}^d V_j',$$

where the last implication follows from Theorem 3.A.12(b) in [93]. The converse follows from (4.1) by observing that

$$\sum_{j=1}^d X_j|(U=u) = \frac{d(1-u) + \sum_{j=1}^d V_j}{(2u-1)},$$

and applying the same arguments. $\square$

Proposition 4.1.2 implies that if $\boldsymbol{X} \in \mathcal{SB}_d$ is a Σ_{cx}-smallest element, then $\boldsymbol{V}$ in (4.1) is a Σ_{cx}-smallest element in $\mathcal{C}_d^{\text{EM}}$. An important consequence of the one-to-one map in (4.3) is that we can construct all the extremal mixture copulas from $\mathcal{PB}_d$. Therefore, the Σ_{cx}-smallest elements in $\mathcal{C}_d^{\text{EM}}$ can be constructed from Σ_{cx}-smallest palindromic Bernoulli random vectors. We recall that Proposition 3.1.7 identifies all Σ_{cx}-smallest pmfs in $\mathcal{PB}_d$.

The last two results of this section characterize the extremal negative dependence in the entire class of copulas, not only in $\mathcal{C}_d^{\text{EM}}$. Proposition 4.1.3 states that the extremal copulas built from Σ_{cx}-smallest Bernoulli random vectors in $\mathcal{K}$ are Σ-countermonotonic, but, in general, are not Σ_{cx}-smallest in the entire Fréchet class of copulas. Proposition 4.1.4, instead, implies that, when d is even, the extremal mixture copulas constructed from Σ_{cx}-smallest pmfs of $\mathcal{PB}_d$ have the joint mixability property, hence they are Σ_{cx}-smallest elements in the entire class of copulas.

Proposition 4.1.3. *Let $\boldsymbol{X}$ be a Bernoulli random vector with pmf $f \in \mathcal{K}$, where $\mathcal{K}$ is the basis of* $\ker(\mathcal{H}) \equiv \mathcal{PB}_d$ *and is given in (3.2). If $\boldsymbol{X}$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$, then the uniform random vector $\boldsymbol{V}$ from (4.1) is a Σ_{cx}-smallest in $\mathcal{C}_d^{EM}$ and Σ-countermonotonic.*

Proof: We assume d odd, as the case d even is analogous. Since $\boldsymbol{X}$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$ (and also in $\mathcal{PB}_d$), it follows from Proposition 4.1.2 that the random vector $\boldsymbol{V}$ is a Σ_{cx}-smallest element in $\mathcal{C}_d^{\text{EM}}$. We now prove that $\boldsymbol{V}$ is Σ-countermonotonic. Let $J \subset \{1, \dots, d\}$ be a set of indexes, $J \neq \emptyset$. Since $f \in \mathcal{K}$, there exists $\boldsymbol{x} \in \mathcal{X}_d$ such that the random vector $\boldsymbol{X}$ can only take the values $\boldsymbol{x}$ or $\mathbf{1}_d - \boldsymbol{x}$. Moreover, since $\boldsymbol{X}$ is a Σ_{cx}-smallest element, Proposition 3.1.6 implies that $\boldsymbol{x} \in \mathcal{X}_d^*$. Recall that the sum of the components of $\boldsymbol{x} \in \mathcal{X}_d^*$ is equal to $M_d = (d-1)/2$ or $M_d + 1$. Therefore, we have two options: either $\sum_{j=1}^d x_j = M_d$ and $\sum_{j=1}^d (1 - x_j) = M_d + 1$, or $\sum_{j=1}^d x_j = M_d + 1$ and $\sum_{j=1}^d (1 - x_j) = M_d$. We consider the case $\sum_{j=1}^d x_j = M_d$; the other case follows by taking $\boldsymbol{y} = \mathbf{1}_d - \boldsymbol{x}$. Let $k := \sum_{j \in J} x_j$. We have

$$\sum_{j \in \bar{J}} x_j = M_d - k, \quad \sum_{j \in J} (1 - x_j) = |J| - k, \quad \sum_{j \in \bar{J}} (1 - x_j) = M_d + 1 - (|J| - k),$$

where $\bar{J}$ denotes the complement of J and $|J|$ its cardinality. Let us define two random variables

$$\begin{aligned} A &= \sum_{j \in J} V_j = U \sum_{j \in J} X_j + (1 - U) \sum_{j \in J} (1 - X_j), \\ B &= \sum_{j \in \bar{J}} V_j = U \sum_{j \in \bar{J}} X_j + (1 - U) \sum_{j \in \bar{J}} (1 - X_j). \end{aligned}$$

By conditioning on the two possible outcomes of the random variable $\boldsymbol{X}$, we have

$$\begin{aligned} A|(\boldsymbol{X} = \boldsymbol{x}) &= |J| - k + (2k - |J|)U, \\ B|(\boldsymbol{X} = \boldsymbol{x}) &= M_d + 1 - (|J| - k) - (1 + 2k - |J|)U, \\ A|(\boldsymbol{X} = \mathbf{1}_d - \boldsymbol{x}) &= k - (2k - |J|)U, \\ B|(\boldsymbol{X} = \mathbf{1}_d - \boldsymbol{x}) &= M_d - k + (1 + 2k - |J|)U. \end{aligned}$$

Let (A_1, B_1) and (A_2, B_2) be two independent copies of (A, B). We have, for $h \in \{1,2\}$,

$$A_h = U_h \sum_{j \in J} X_j^{(h)} + (1 - U_h) \sum_{j \in J} (1 - X_j^{(h)}), \quad B_h = U_h \sum_{j \in \bar{J}} X_j^{(h)} + (1 - U_h) \sum_{j \in \bar{J}} (1 - X_j^{(h)}),$$

where U_1 and U_2 are independent standard uniform random variables, and $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are iid copies of $\boldsymbol{X}$, independent of U_1 and U_2. We want to prove that (A, B) is countermonotonic, viz.

$$\begin{aligned}&\mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0\} \\ &= \sum_{\boldsymbol{x}_1}\sum_{\boldsymbol{x}_2} \mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0 | \boldsymbol{X}^{(1)} = \boldsymbol{x}_1, \boldsymbol{X}^{(2)} = \boldsymbol{x}_2\} f(\boldsymbol{x}_1) f(\boldsymbol{x}_2).\end{aligned} \tag{4.4}$$

The pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ can take one of four possible values: $(\boldsymbol{x}, \boldsymbol{x}), (\boldsymbol{x}, \mathbf{1}_d - \boldsymbol{x}), (\mathbf{1}_d - \boldsymbol{x}, \boldsymbol{x})$, or $(\mathbf{1}_d - \boldsymbol{x}, \mathbf{1}_d - \boldsymbol{x})$.

Case 1. Let $(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}, \boldsymbol{x})$. We have

$$\begin{aligned}&\mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0 | \boldsymbol{X}^{(1)} = \boldsymbol{x}, \boldsymbol{X}^{(2)} = \boldsymbol{x}\} \\ &= \mathbb{P}\{(2k - |J|)(U_1 - U_2) \cdot (-1)(1 + 2k - |J|)(U_1 - U_2) \le 0\} \\ &= \mathbb{P}\{-(2k - |J|)(1 + 2k - |J|)(U_1 - U_2)^2 \le 0\} = 1,\end{aligned}$$

where the last equality follows because $2k - |J| \in \mathbb{Z}$ and, if $2k - |J| \ge 0$ then $1 + 2k - |J| > 0$, while if $2k - |J| < 0$ then $1 + 2k - |J| \le 0$.

Case 2. Let $(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}, \mathbf{1}_d - \boldsymbol{x})$. We have

$$\begin{aligned}&\mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0 | \boldsymbol{X}^{(1)} = \boldsymbol{x}, \boldsymbol{X}^{(2)} = \mathbf{1}_d - \boldsymbol{x}\} \\ &= \mathbb{P}\{-(2k - |J|)(1 + 2k - |J|)(U_1 + U_2 - 1)^2 \le 0\} = 1,\end{aligned}$$

for the same argument in Case 1.

Cases 3 and 4 are analogous. From (4.4), we have

$$\mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0\} = \sum_{\boldsymbol{x}_1}\sum_{\boldsymbol{x}_2} 1 \cdot f(\boldsymbol{x}_1) f(\boldsymbol{x}_2) = 1.$$

Therefore, $(\sum_{j \in J} V_j, \sum_{j \in \bar{J}} V_j)$ is countermonotonic, and $\boldsymbol{V}$ is Σ-countermonotonic. $\square$

The following example shows that Proposition 4.1.3 holds only for extremal copulas and not for extremal mixture copulas.

Example 4.1.1. *Let $d = 5$. Consider $\boldsymbol{X}$ with pmf $f \in \mathcal{PB}_5$ such that $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2) = f(\mathbf{1}_5 - \boldsymbol{x}_1) = f(\mathbf{1}_5 - \boldsymbol{x}_2) = 1/4$, with $\boldsymbol{x}_1 = (1,0,0,0,1)$ and $\boldsymbol{x}_2 = (0,0,0,1,1)$. Since f has support in $\mathcal{X}_5^*$, by Proposition 3.1.6, $\boldsymbol{X}$ is a Σ_{cx}-smallest element in $\mathcal{SB}_5$. Let $\boldsymbol{V}$ be the uniform random vector with extremal mixture copula obtained from (4.1), and define $A = \sum_{j \in J} V_j$ and $B = \sum_{j \in \bar{J}} V_j$, with $J = \{1,5\}$. Let (A_1, B_1) and (A_2, B_2) be two independent copies of (A, B). We have, for $h \in \{1,2\}$,*

$$\begin{aligned}A_h &= U_h \left(X_1^{(h)} + X_5^{(h)}\right) + (1 - U_h)\left(2 - X_1^{(h)} - X_5^{(h)}\right), \\ B_h &= U_h \left(X_2^{(h)} + X_3^{(h)} + X_4^{(h)}\right) + (1 - U_h)\left(3 - X_2^{(h)} - X_3^{(h)} - X_4^{(h)}\right),\end{aligned}$$

where U_1 and U_2 are independent standard uniform random variables, and $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are iid copies of $\boldsymbol{X}$, independent of U_1 and U_2. It holds that,

$$\begin{aligned}&\mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0 | \boldsymbol{X}^{(1)} = \boldsymbol{x}_1, \boldsymbol{X}^{(2)} = \boldsymbol{x}_2\} \\ &= \mathbb{P}\{(2U_1 - 1)(1 - 3U_1 + U_2) \le 0\} < 1.\end{aligned}$$

Therefore, $\mathbb{P}\{(A_1 - A_2)(B_1 - B_2) \le 0\} < 1$ and $\boldsymbol{V}$ is not Σ-countermonotonic.

Proposition 4.1.4. *Let d be even. If $\boldsymbol{X}^* \in \mathcal{PB}_d$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$, then $\boldsymbol{V}^* = U\boldsymbol{X}^* + (1-U)(1-\boldsymbol{X}^*)$ is a joint mix.*

Proof: By Proposition 3.1.5, the sum $X_1^* + \cdots + X_d^*$ takes value $d/2$ with probability 1. We have

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^d V_j^* = \frac{d}{2}\right) &= \mathbb{P}\left\{d(1-U) + (2U-1)\sum_{j=1}^d X_j^* = \frac{d}{2}\right\} \\
&= \int_0^1 \mathbb{P}\left\{\sum_{j=1}^d X_j^* = \frac{d/2 - d(1-u)}{2u-1}\right\} \mathrm{d}u \\
&= \int_0^1 \mathbb{P}\left(\sum_{j=1}^d X_j^* = \frac{d}{2}\right) \mathrm{d}u \\
&= 1,
\end{aligned}$$

thus, $\boldsymbol{V}^*$ is a joint mix. □

We conclude this section with an example of a Σ_{cx}-smallest (thus Σ-countermonotonic) copula and an example of a Σ-countermonotonic copula that is not Σ_{cx}-smallest.

Example 4.1.2. *Let $d = 100$. Let $\boldsymbol{V}^{(1)}$ be a random vector with uniform marginals corresponding to the copula*

$$C_1(u_1, \ldots, u_{100}) = (\min_{j \le 50} u_j + \min_{j \ge 51} u_j - 1)^+. \tag{4.5}$$

The copula C_1 in (4.5) is build from the symmetric Σ_{cx}-smallest Bernoulli distribution with support on $\boldsymbol{x}^{(1)} = (\mathbf{1}_{50}//\mathbf{0}_{50})$, where $\mathbf{0}_{50}$ is a vector of length 50 with all zeros, and $\mathbf{1}_{100} - \boldsymbol{x}^{(1)}$. By Proposition 4.1.4, $\boldsymbol{V}^{(1)}$ is a joint mix; therefore, it is Σ-countermonotonic and a Σ_{cx}-smallest element in its Fréchet class.

Let $d = 103$. Let $\boldsymbol{V}^{(2)}$ be a random vector with uniform marginals corresponding to the copula

$$C_2(u_1, \ldots, u_{103}) = (\min_{j \le 51} u_j + \min_{j \ge 52} u_j - 1)^+. \tag{4.6}$$

The copula C_2 in (4.6) corresponds to the symmetric Σ_{cx}-smallest Bernoulli distribution with support on $\boldsymbol{x}^{(2)} = (\mathbf{1}_{51}//\mathbf{0}_{52})$ and $\mathbf{1}_{103} - \boldsymbol{x}^{(2)}$. According to Proposition 4.1.3, the random vector $\boldsymbol{V}^{(2)}$ is Σ-countermonotonic and a Σ_{cx}-smallest element in $\mathcal{C}_{103}^{EM}$. However, $\boldsymbol{V}^{(2)}$ is not a Σ_{cx}-smallest element in its Fréchet class. In fact, we can see that $\boldsymbol{V}^{(2)}$ is not a joint mix, as the sum of its components varies in the interval (51,52); yet, its Fréchet class admits a joint mix. Consider, for example, $\boldsymbol{V}^{(JM)} := (\boldsymbol{V}^{(1)}//\boldsymbol{V}^{(3)})$, where $\boldsymbol{V}^{(3)}$ is a 3-dimensional joint mix with uniform marginals, independent of $\boldsymbol{V}^{(1)}$, and with the dependence structure specified in Example 3 in [50]. $\boldsymbol{V}^{(1)}$ and $\boldsymbol{V}^{(3)}$ are joint mixes, thus $\boldsymbol{V}^{(JM)}$ is a joint mix and

$$\sum_{j=1}^d V_j^{(JM)} \le_{cx} \sum_{j=1}^d V_j^{(2)}.$$

FGM copulas

Another class of copulas that can be built from Bernoulli random vectors via a stochastic representation is the class $\mathcal{C}_d^{\mathrm{FGM}}$ of multivariate FGM copulas. In this section, we recall their definition and the stochastic representation introduced in [12], which provides a one-to-one correspondence with the class $\mathcal{SB}_d$. In the class of FGM copulas, the elements corresponding to Σ_{cx}-smallest Bernoulli random vectors are Σ_{cx}-smallest in the class $\mathcal{C}_d^{\mathrm{FGM}}$, but not in the whole class of copulas. It is therefore interesting to compare the minimal negative dependence in the class of FGM copulas, with the minimal dependence in the whole class of copulas. This is the focus of Section 4.2. Here, we present some known, but necessary, results on FGM copulas.

Definition 4.1.3. *A multivariate copula C belongs to the class of FGM copulas if it has the following expression:*

$$C(\boldsymbol{u}) = u_1 \cdots u_d \left(1 + \sum_{k=2}^{d} \sum_{1 \leq j_1 < \ldots < j_k \leq d} \theta_{j_1 \ldots j_k} \bar{u}_{j_1} \cdots \bar{u}_{j_k}\right), \quad \boldsymbol{u} \in [0,1]^d,$$

where, for each $j \in \{1, \ldots, d\}$, $\bar{u}_j = 1 - u_j$.

There exist 2^d constraints on the parameters for the existence of a FGM copula, which are

$$1 + \sum_{k=2}^{d} \sum_{1 \leq j_1 < \ldots < j_k \leq d} \theta_{j_1 \ldots j_k} \epsilon_{j_1} \epsilon_{j_2} \cdots \epsilon_{j_k} \geq 0,$$

for every $(\epsilon_1, \ldots, \epsilon_d) \in \{-1,1\}^d$; see [21]. When $d = 2$, the admissible set for the unique parameter is the interval $[-1,1]$. However, as the dimension increases, the shape of the set of admissible parameters becomes more and more complex. A useful tool to address this problem is the stochastic representation provided in [12]. Let $\boldsymbol{Z}_0 = (Z_{1,0}, \ldots, Z_{d,0})$ be a vector of independent exponential random variables with mean $1/2$ and let $\boldsymbol{Z}_1 = (Z_{1,1}, \ldots, Z_{d,1})$ be a vector of independent exponential random variables with mean 1. Let $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$ be independent. The following theorem, proved in [12], shows the existence of a one-to-one correspondence between the class $\mathcal{SB}_d$ and the class $\mathcal{C}_d^{\mathrm{FGM}}$.

Theorem 4.1.5. *Let $\boldsymbol{X} \in \mathcal{SB}_d$ be a d-dimensional symmetric Bernoulli random vector. Let $\boldsymbol{U} = (U_1, \ldots, U_d)$ be a random vector such that, for each $j \in \{1, \ldots, d\}$,*

$$U_j = 1 - \exp\{-(Z_{j,0} + X_j Z_{j,1})\}. \tag{4.7}$$

Then, $\boldsymbol{U}$ has a d-variate distribution with standard uniform marginals, and its cdf is a FGM copula given, for each $\mathbf{u} \in [0,1]^d$, by

$$C(\boldsymbol{u}) = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f_{\boldsymbol{X}}(\boldsymbol{x}) \prod_{h=1}^{d} u_h \Big(1 + (-1)^{x_h}(1 - u_h)\Big).$$

In [12], the authors derive the parameters of the FGM copula in terms of the centered moments of its corresponding symmetric Bernoulli distribution, viz.

$$\theta_{j_1 \ldots j_k} = (-2)^k \, \mathbb{E}_{\boldsymbol{X}} \left\{ \prod_{\ell=1}^{k} \left(X_{j_\ell} - \frac{1}{2} \right) \right\}, \tag{4.8}$$

for $k \in \{2, \dots, d\}$ and $1 \le j_1 < \dots < j_k \le d$.

Regarding extremal negative dependence in $\mathcal{C}_d^{\mathrm{FGM}}$, in [8], the authors explicitly find the FGM copula that corresponds to the Σ_{cx}-smallest exchangeable Bernoulli distribution. In Section 4.3.2, we will discuss a generalization of FGM copulas that can be built from a generic Bernoulli random vector $\boldsymbol{X}$ with pmf in $\mathcal{B}_d$. Moreover, Theorem 4.3.8 show that the one-to-one map between the classes $\mathcal{SB}_d$ and $\mathcal{C}_d^{\mathrm{FGM}}$ preserves the convex order of the sums of the components, viz.,

$$\sum_{j=1}^{d} X_j \le_{cx} \sum_{j=1}^{d} X'_j \implies \sum_{j=1}^{d} U_j \le_{cx} \sum_{j=1}^{d} U'_j.$$

This implies that FGM copulas constructed from the Σ_{cx}-smallest element of $\mathcal{SB}_d$ are Σ_{cx}-smallest in the class of FGM copulas. Using the characterization of all Σ_{cx}-smallest elements of $\mathcal{SB}_d$ in Section 3.1.1, we can investigate some properties of a Σ_{cx}-smallest FGM copula.

Remark 4.1.1. *We have already seen in Examples 3.1.2 and 3.1.3 that the Σ_{cx}-smallest pmfs are palindromic in dimension $d \le 4$. By Proposition 3.4 in [12], we know that the FGM copulas constructed from the class $\mathcal{PB}_d$ are radially symmetric and, in particular, have $\theta_{j_1\dots j_k} = 0$ for $1 \le j_1 < \dots < j_k \le d$, for every odd value of k.*

4.2 Minimal pairwise dependence measures

The most common dependence measures are Pearson's correlation ρ_P, Spearman's rho ρ_S, and Kendall's tau τ_K, defined by

$$\rho_P(X,Y) = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}, \quad \rho_S(X,Y) = \rho_P\{F_X(X), F_Y(Y)\},$$
$$\tau_K(X,Y) = \mathbb{P}\{(X - X')(Y - Y') \ge 0\} - \mathbb{P}\{(X - X')(Y - Y') \le 0\}.$$

When the marginals are continuous, Spearman's rho and Kendall's tau are measures that depend only on the copula of the two random variables and not on their marginals; see [79]. This is not true when the marginals are discrete; see Section 4.2 in [52]. It is evident that $\rho_P(U,V) = \rho_S(U,V)$ for uniform random variables U and V. It is also easy to verify that Spearman's rho and Pearson's correlation are equal for Bernoulli random variables; therefore, we consider only Pearson's correlation ρ_P.

In this section, we compare these measures for the Σ-countermonotonic random vectors in the class $\mathcal{SB}_d$, the Σ-countermonotonic copulas in the family of extremal mixture copulas, and the Σ_{cx}-smallest elements in the class of FGM copulas. The following proposition follows from direct computations.

Proposition 4.2.1. *Let X_{j_1} and X_{j_2} be two Bernoulli random variables of means p_{j_1} and p_{j_2}, respectively. Then, we have*

$$\rho_P(X_{j_1}, X_{j_2}) = \frac{\tau_K(X_{j_1}, X_{j_2})}{2\sqrt{p_{j_1}(1-p_{j_1})p_{j_2}(1-p_{j_2})}}.$$

In [46], the authors prove that the upper and lower bounds for correlations in a Fréchet class of multivariate Bernoulli pmfs are attained at the extremal points of the Fréchet class, making it possible to determine the range of admissible correlations. From Proposition 4.2.1, it follows that Kendall's tau also reaches its bounds on the extremal points.

For dimensions higher than 2, we adopt a simple approach to generalize these dependence measures, namely by averaging over all pairwise measures; see Section 3.1 in [55]. In particular, we denote the mean of Pearson's correlation and the mean of Kendall's tau of a d-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ as follows

$$\bar{\rho}_P(\boldsymbol{Y}) = \frac{2}{d(d-1)} \sum_{1 \le j_1 < j_2 \le d} \rho_P(Y_{j_1}, Y_{j_2}), \quad \bar{\tau}_K(\boldsymbol{Y}) = \frac{2}{d(d-1)} \sum_{1 \le j_1 < j_2 \le d} \tau_K(Y_{j_1}, Y_{j_2}).$$

Proposition 4.2.2. *Let $\boldsymbol{X} \in \mathcal{SB}_d$ and let $\boldsymbol{V} \in \mathcal{C}_d^{EM}$ and $\boldsymbol{U} \in \mathcal{C}_d^{FGM}$ be the uniform random vectors built from $\boldsymbol{X}$ with extremal mixture copula and FGM copula, respectively. We have*

$$\begin{aligned} &\rho_P(X_{j_1}, X_{j_2}) = \rho_P(V_{j_1}, V_{j_2}) = 3\rho_P(U_{j_1}, U_{j_2}) \\ &= 2\tau_K(X_{j_1}, X_{j_2}) = 2\tau_K(V_{j_1}, V_{j_2}) = \frac{9}{2}\tau_K(U_{j_1}, U_{j_2}), \end{aligned}$$

for every $j_1, j_2 \in \{1, \ldots, d\}$, $j_1 \neq j_2$, and

$$\bar{\rho}_P(\boldsymbol{X}) = \bar{\rho}_P(\boldsymbol{V}) = 3\bar{\rho}_P(\boldsymbol{U}) = 2\bar{\tau}_K(\boldsymbol{X}) = 2\bar{\tau}_K(\boldsymbol{V}) = \frac{9}{2}\bar{\tau}_K(\boldsymbol{U}).$$

Proof: Let $j_1, j_2 \in \{1, \ldots, d\}$, $j_1 \neq j_2$. We have $\rho_P(X_{j_1}, X_{j_2}) = 2\tau_K(X_{j_1}, X_{j_2})$ f rom Proposition 4.2.1. Moreover, from the stochastic representation in (4.1), standard computations give $\rho_P(V_{j_1}, V_{j_2}) = \rho_P(X_{j_1}, X_{j_2})$ and $\tau_K(V_{j_1}, V_{j_2}) = \tau_K(X_{j_1}, X_{j_2})$. Regarding the FGM copula, it is possible to show that $\rho_P(X_{j_1}, X_{j_2}) = 3\rho_P(U_{j_1}, U_{j_2})$. Since $\boldsymbol{U}$ has distribution in $\mathcal{C}_d^{\mathrm{FGM}}$, the cdf of the pair (U_{j_1}, U_{j_2}) is a bivariate FGM copula with parameter $\theta_{j_1 j_2}$, given by (4.8). It is well known that $\rho_P(U_{j_1}, U_{j_2}) = \theta_{j_1 j_2}/3$ and $\tau_K(U_{j_1}, U_{j_2}) = 2\theta_{j_1 j_2}/9$; see [90]. Thus, $\rho_P(U_{j_1}, U_{j_2}) = 3\tau_K(U_{j_1}, U_{j_2})/2$ and the conclusion follows since, from (4.8), $\theta_{j_1 j_2} = \rho_P(X_{j_1}, X_{j_2})$. □

As a consequence of Proposition 4.2.2, the analysis of the pairwise dependence measures of extremal mixture copulas and FGM copulas is fully described by the pairwise dependence measures of the corresponding symmetric Bernoulli distributions. Moreover, Corollary 4.2.3 shows that it is sufficient to consider palindromic Bernoulli distributions, and the corresponding copulas, to describe all the possible structures of pairwise dependence measures. Indeed, the authors of [61] proved that for every $\boldsymbol{X} \in \mathcal{SB}_d$ there exists $\boldsymbol{X}' \in \mathcal{PB}_d$ with the same bivariate Pearson's correlation structure.

Corollary 4.2.3. *Let $\boldsymbol{X} \in \mathcal{SB}_d$ and let $\boldsymbol{U}$ be a uniform random vector with the FGM copula constructed from $\boldsymbol{X}$. Then there exists $\boldsymbol{X}' \in \mathcal{PB}_d$ such that $\rho_P(U_{j_1}, U_{j_2}) = \rho_P(U'_{j_1}, U'_{j_2})$ and $\tau_K(U_{j_1}, U_{j_2}) = \tau_K(U'_{j_1}, U'_{j_2})$, for every $j_1, j_2 \in \{1, \ldots, d\}$, with $j_1 \neq j_2$, where $\boldsymbol{U}'$ is a uniform random vector with FGM copula constructed from $\boldsymbol{X}'$. Moreover, the extremal mixture copulas constructed from $\boldsymbol{X}$ and $\boldsymbol{X}'$ via the stochastic representation in (4.1) coincide.*

Proof: Given $\boldsymbol{X} \in \mathcal{SB}_d$, from Theorem 1 in [61], there exists $\boldsymbol{X}' \in \mathcal{PB}_d$ such that $\rho_P(X_{j_1}, X_{j_2}) = \rho_P(X'_{j_1}, X'_{j_2})$, for every $j_1, j_2 \in \{1, \dots, d\}$, with $j_1 \neq j_2$. Moreover, by Proposition 4.2.2, one has $\rho_P(U_{j_1}, U_{j_2}) = \rho_P(U'_{j_1}, U'_{j_2})$ and $\tau_K(U_{j_1}, U_{j_2}) = \tau_K(U'_{j_1}, U'_{j_2})$, for every $j_1, j_2 \in \{1, \dots, d\}$, with $j_1 \neq j_2$. From the construction of $\boldsymbol{X}' \in \mathcal{PB}_d$ in the proof of Theorem 1 in [61], $\boldsymbol{X}$ and $\boldsymbol{X}'$ are such that

$$f(\boldsymbol{s_i}) + f(\mathbf{1}_d - \boldsymbol{s_i}) = f'(\boldsymbol{s_i}) + f'(\mathbf{1}_d - \boldsymbol{s_i}),$$

for every $\boldsymbol{i} \in \mathcal{X}_{d-1}$. Therefore, the weights $w_{\boldsymbol{i}}$, given in (4.2), of the corresponding extremal mixture copulas are equal and these copulas coincide. $\square$

The following corollary to Proposition 4.2.2 states that extremal mixture copulas and FGM copulas built from Σ_{cx}-smallest Bernoulli random vectors have minimal mean correlation and Kendall's tau. Its proof relies on the following known fact: the mean Pearson correlation of a Bernoulli random vector $\boldsymbol{X} \in \mathcal{SB}_d$ can be expressed as the expectation of a convex function of the sum $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$, i.e.,

$$\bar{\rho}_P(\boldsymbol{X}) = \mathbb{E}\{\phi(S_{\boldsymbol{X}})\}, \tag{4.9}$$

where

$$\phi(y) = \begin{cases} \frac{8}{d(d-1)}\binom{y}{2} - 1, & \text{if } y \geq 2, \\ -1, & \text{otherwise;} \end{cases} \tag{4.10}$$

see [16]. Since ϕ is a convex function, if $\boldsymbol{X}$ is a Σ_{cx}-smallest element in $\mathcal{SB}_d$, then $\bar{\rho}_P(\boldsymbol{X})$ is minimal and, by Proposition 3.1.5, we have

$$\bar{\rho}_P(\boldsymbol{X}) = \begin{cases} -\frac{1}{d-1}, & \text{if } d \text{ is even,} \\ -\frac{1}{d}, & \text{if } d \text{ is odd.} \end{cases} \tag{4.11}$$

We observe that if d is odd, the minimal mean Pearson correlation in the class $\mathcal{SB}_d$ is equal to the minimal mean Pearson correlation in the class $\mathcal{SB}_{d+1}$.

Corollary 4.2.4. *Let $\boldsymbol{X} \in \mathcal{SB}_d$ be a Σ_{cx}-smallest element in $\mathcal{SB}_d$ and let $\boldsymbol{V} \in \mathcal{C}_d^{EM}$ and $\boldsymbol{U} \in \mathcal{C}_d^{FGM}$ be the uniform random vectors built from $\boldsymbol{X}$ with extremal mixture copula and FGM copula, respectively. We have*

$$\bar{\rho}_P(\boldsymbol{V}) \leq \bar{\rho}_P(\boldsymbol{V}'), \quad \bar{\tau}_K(\boldsymbol{V}) \leq \bar{\tau}_K(\boldsymbol{V}'),$$

for any $\boldsymbol{V}' \in \mathcal{C}_d^{EM}$, and

$$\bar{\rho}_P(\boldsymbol{U}) \leq \bar{\rho}_P(\boldsymbol{U}'), \quad \bar{\tau}_K(\boldsymbol{U}) \leq \bar{\tau}_K(\boldsymbol{U}'),$$

for any $\boldsymbol{U}' \in \mathcal{C}_d^{FGM}$. Moreover,

$$\bar{\rho}_P(\boldsymbol{V}) = \begin{cases} -\frac{1}{d-1}, & \text{if } d \text{ is even,} \\ -\frac{1}{d}, & \text{if } d \text{ is odd,} \end{cases} \quad \bar{\tau}_K(\boldsymbol{V}) = \begin{cases} -\frac{1}{2(d-1)}, & \text{if } d \text{ is even,} \\ -\frac{1}{2d}, & \text{if } d \text{ is odd,} \end{cases}$$

and

$$\bar{\rho}_P(\boldsymbol{U}) = \begin{cases} -\frac{1}{3(d-1)}, & \text{if } d \text{ is even,} \\ -\frac{1}{3d}, & \text{if } d \text{ is odd,} \end{cases} \quad \bar{\tau}_K(\boldsymbol{U}) = \begin{cases} -\frac{2}{9(d-1)}, & \text{if } d \text{ is even,} \\ -\frac{2}{9d}, & \text{if } d \text{ is odd.} \end{cases}$$

Proof: Let $\boldsymbol{X} \in \mathcal{SB}_d$ be a Σ_{cx}-smallest element in $\mathcal{SB}_d$. From (4.9) and (4.10), since ϕ in (4.10) is a convex function, $\bar{\rho}_P(\boldsymbol{X}) \leq \bar{\rho}_P(\boldsymbol{X}')$, for any $\boldsymbol{X}' \in \mathcal{SB}_d$. Since every extremal mixture copula and every FGM copula can be built from a Bernoulli random vector $\boldsymbol{X}' \in \mathcal{SB}_d$, the thesis follows from Proposition 4.2.2 and from (4.11). □

The mean Pearson correlation of a Σ_{cx}-smallest Bernoulli random vector, or of a copula built from it, depends only on the dimension of the class. However, Σ_{cx}-smallest Bernoulli random vectors in the same class $\mathcal{SB}_d$ have different dependence structures. We now study pairwise dependence measures of different Σ_{cx}-smallest Bernoulli random vectors in the class $\mathcal{SB}_d$. We first consider a Σ_{cx}-smallest Bernoulli random vector $\boldsymbol{X}^K$ with pmf $f^K \in \mathcal{K}$. We know that there exists $\boldsymbol{x} \in \mathcal{X}_d^*$ such that $f^K(\boldsymbol{x}) = f^K(\mathbf{1}_d - \boldsymbol{x}) = 1/2$. It follows that there are n_d^+ comonotonic pairs (a pair is comonotonic if one random variable is a deterministic non-decreasing transformation of the other; see [87]), where

$$n_d^+ = \begin{cases} \frac{d(d-2)}{4}, & \text{if } d \text{ is even,} \\ \frac{(d-1)^2}{4}, & \text{if } d \text{ is odd,} \end{cases}$$

and n_d^- countermonotonic pairs, where

$$n_d^- = \binom{d}{2} - n_d^+ = \begin{cases} \frac{d^2}{4}, & \text{if } d \text{ is even,} \\ \frac{(d-1)(d+1)}{4}, & \text{if } d \text{ is odd.} \end{cases}$$

Obviously, countermonotonic pairs have correlation $\rho_P = -1$ and comonotonic pairs have correlation $\rho_P = 1$, and the mean Pearson correlation of a random vector $\boldsymbol{X}^K \in \mathcal{K}$ is given by $2(n_d^+ - n_d^-)/\{d(d-1)\}$, which is equal to (4.11). Since the extremal copulas are in a one-to-one relationship with the elements in $\mathcal{K}$, and $\rho_P(X_{j_1}, X_{j_2}) = \rho_P(V_{j_1}, V_{j_2})$, we can conclude that extremal copulas that are built from Σ_{cx}-smallest Bernoulli random vectors have n_d^+ comonotonic pairs and n_d^- countermonotonic pairs, and they are Σ-countermonotonic by Proposition 4.1.3.

We can then consider the unique exchangeable Σ_{cx}-smallest Bernoulli random vector $\boldsymbol{X}^e \in \mathcal{SB}_d$. In this case, for every $j_1, j_2 \in \{1, \ldots, d\}$, with $j_1 \neq j_2$, we have $\rho_P(X^e_{j_1}, X^e_{j_2}) = \bar{\rho}_P(\boldsymbol{X}^e)$, where $\bar{\rho}_P(\boldsymbol{X}^e)$ is given by (4.11). The Bernoulli random vector $\boldsymbol{X}^e$ and the uniform random vectors $\boldsymbol{V}^e$ and $\boldsymbol{U}^e$, associated with the extremal mixture copula and FGM copula defined by $\boldsymbol{X}^e$, are pairwise negatively correlated (all pairwise Pearson's correlations are non-positive). Notice that the distribution of $\boldsymbol{X}^e$ is the maximum entropy pmf discussed in Section 2.5.2.

Corollary 4.2.3 implies that palindromic Bernoulli distributions are sufficient to construct FGM copulas with any admissible bivariate dependence measures ρ_P and τ_K. Let $\boldsymbol{X}$ be a Bernoulli random vector with pmf $f \in \mathcal{SB}_d$, but $f \notin \mathcal{PB}_d$, and let $\boldsymbol{V}$ and $\boldsymbol{U}$ be the uniform random vectors built from $\boldsymbol{X}$ with extremal mixture copula and FGM copula, respectively. By Corollary 4.2.3, there exists $\boldsymbol{X}'$ with pmf $f' \in \mathcal{PB}_d$ such that $\boldsymbol{V}$ has the same distribution of $\boldsymbol{V}'$ and $\rho_P(U_{j_1}, U_{j_2}) = \rho_P(U'_{j_1}, U'_{j_2})$, for every $j_1, j_2 \in \{1, \ldots, d\}$, with $j_1 \neq j_2$, where $\boldsymbol{V}'$ and $\boldsymbol{U}'$ are the uniform random vectors with the extremal mixture copula and FGM copula constructed from $\boldsymbol{X}'$, respectively. To compare the dependence structures of $\boldsymbol{X}$, $\boldsymbol{X}'$, and their corresponding FGM copulas, we need to consider higher order moments. We define the centered cross moments of order three of a d-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ as

$$\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{Y}) = \mathbb{E}\left[\prod_{h=1}^{3}\left\{\frac{Y_{j_h} - \mathbb{E}(Y_{j_h})}{\sqrt{\mathrm{Var}(Y_{j_h})}}\right\}\right],$$

for $j_1, j_2, j_3 \in \{1, \ldots, d\}$, with j_1, j_2, j_3 all distinct.

Proposition 4.2.5. *Let $\boldsymbol{X}' \in \mathcal{PB}_d$ and let $\boldsymbol{V}'$ and $\boldsymbol{U}'$ be the uniform random vectors constructed from $\boldsymbol{X}' \in \mathcal{PB}_d$ with extremal mixture copula and FGM copula, respectively. Then, $\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{X}') = \tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{V}') = \tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{U}') = 0$, for every $j_1, j_2, j_3 \in \{1, \ldots, d\}$, with j_1, j_2, j_3 all distinct.*

Proof: It is easy to show that if a random vector $\boldsymbol{Y}$ has the same distribution as $\mathbf{1}_d - \boldsymbol{Y}$, then $\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{Y}) = 0$ for every $j_1, j_2, j_3 \in \{1, \ldots, d\}$, with j_1, j_2, j_3 all distinct. A palindromic Bernoulli random vector is such that $\boldsymbol{X}'$ has the same distribution as $\mathbf{1}_d - \boldsymbol{X}'$, and every extremal mixture copula is radially symmetric, i.e., $\boldsymbol{V}'$ has the same distribution as $\mathbf{1}_d - \boldsymbol{V}'$. Finally, by Proposition 3.4 in [12], if $\boldsymbol{X}' \in \mathcal{PB}_d$, the corresponding FGM copula $\boldsymbol{U}'$ is radially symmetric. □

The above proposition shows that the palindromic Bernoulli random vectors and the copulas built from them have centered cross moments of order three equal to zero. This is not true for every $f \in \mathcal{SB}_d$. Proposition 4.2.6 establishes the relation between centered cross moments of order three of a general Bernoulli random vector and of the copulas constructed from it. We conclude this section by providing an example of a symmetric Bernoulli random vector that is non-palindromic, and whose centered cross moments of order three are different from zero. The proof of the following proposition follows from standard computation and is therefore omitted.

Proposition 4.2.6. *Let $\boldsymbol{X}$ with pmf $f \in \mathcal{SB}_d$ and let $\boldsymbol{V}$ and $\boldsymbol{U}$ be the uniform random vectors constructed from $\boldsymbol{X}$ with extremal mixture copula and FGM copula, respectively. Then, $\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{X}) = \tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{V}) = -\sqrt{3}/9\,\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{U})$, for every $j_1, j_2, j_3 \in \{1, \ldots, d\}$, with j_1, j_2, j_3 all distinct.*

We conclude this section with an example in dimension $d = 6$.

Example 4.2.1. *Consider $f^{(1)} \in \mathcal{SB}_6$ from Example 3.1.5, which is clearly not palindromic. Let $\boldsymbol{X}$ be a Bernoulli random vector with pmf $f^{(1)}$, and let $\boldsymbol{V}$ and $\boldsymbol{U}$ be the uniform random vectors built from $\boldsymbol{X}$ with extremal mixture copula and FGM copula, respectively. From Corollary 4.2.3, there exists a Bernoulli random vector $\boldsymbol{X}'$, with pmf $f' \in \mathcal{PB}_6$, such that $\rho_P(U_{j_1}, U_{j_2}) = \rho_P(U'_{j_1}, U'_{j_2})$, for every $j_1, j_2 \in \{1, \ldots, d\}$, with $j_1 \neq j_2$, where $\boldsymbol{U}'$ is a uniform random vector with the FGM copula defined by $\boldsymbol{X}'$. Moreover, $\boldsymbol{V}$ has the same distribution as $\boldsymbol{V}'$, where $\boldsymbol{V}'$ is a uniform random vector with the extremal mixture copula constructed from $\boldsymbol{X}'$. However, it is not true that $\boldsymbol{U}$ has the same distribution as $\boldsymbol{U}'$. Indeed, by Proposition 4.2.5, we have $\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{X}') = \tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{U}') = 0$ for every $j_1, j_2, j_3 \in \{1, \ldots, 6\}$, with j_1, j_2, j_3 all distinct, while*

$$\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{X}) = \begin{cases} 1, & \text{if } (j_1, j_2, j_3) \in \{(1,2,4), (1,3,5), (2,5,6), (3,4,6)\}, \\ -1, & \text{if } (j_1, j_2, j_3) \in \{(3,5,6), (2,4,6), (1,3,4), (1,2,5)\}, \\ 0, & \text{otherwise}, \end{cases}$$

and, by Proposition 4.2.6, $\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{U}) = -3\sqrt{3}\,\tilde{\mu}_{j_1,j_2,j_3}(\boldsymbol{X})$*, for every* $j_1, j_2, j_3 \in \{1,\ldots,6\}$*, with* j_1, j_2, j_3 *all distinct.*

4.3 Generalized FGM copulas

The authors of [7] introduced a generalization of FGM copulas, which we call GFGM copulas, through a stochastic representation that we provide below. This stochastic representation builds on a multivariate Bernoulli vector $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$, establishing a link with the class $\mathcal{B}_d(\boldsymbol{p})$ on which we establish our results.

Let $\boldsymbol{X}$ be a d-dimensional Bernoulli random vector with pmf in the Fréchet class $\mathcal{B}_d(\boldsymbol{p})$ and let $\boldsymbol{U}_0 = (U_{0,1}, \ldots, U_{0,d})$ and $\boldsymbol{U}_1 = (U_{1,1}, \ldots, U_{1,d})$ be vectors of d independent uniform random variables. Assume the random vectors $\boldsymbol{X}$, $\boldsymbol{U}_0$, and $\boldsymbol{U}_1$ to be independent and define the random vector $\boldsymbol{U}$ with the following representation:

$$\boldsymbol{U} \overset{\mathcal{L}}{=} \boldsymbol{U}_0^{1-\boldsymbol{p}} \boldsymbol{U}_1^{\boldsymbol{X}} = (U_{0,1}^{1-p_1} U_{1,1}^{X_1}, \ldots, U_{0,d}^{1-p_d} U_{1,d}^{X_d}). \tag{4.12}$$

The joint cdf of the random vector $\boldsymbol{U}$ in (4.12) is the GFGM copula C with vector of parameters $\boldsymbol{p}$ as defined below.

Definition 4.3.1. *A d-variate GFGM copula C with parameter vector $\boldsymbol{p} = (p_1, \ldots, p_d) \in (0,1)^d$ has the following expression:*

$$C(\boldsymbol{u}) = \prod_{m=1}^{d} u_m \left(1 + \sum_{k=2}^{d} \sum_{1\leq j_1 < \cdots < j_k \leq d} \nu_{j_1\ldots j_k} \left(1 - u_{j_1}^{\frac{p_{j_1}}{1-p_{j_1}}}\right) \cdots \left(1 - u_{j_k}^{\frac{p_{j_k}}{1-p_{j_k}}}\right)\right), \tag{4.13}$$

for $\boldsymbol{u} \in [0,1]^d$*, where, for* $1 \leq j_1 < \cdots < j_k \leq d$*,* $k \in \{2, \ldots, d\}$*,*

$$\nu_{j_1\ldots j_k} = \mathbb{E}\left[\prod_{n=1}^{k} \frac{X_{j_n} - p_{j_n}}{p_{j_n}}\right],$$

where $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathcal{B}_d(\boldsymbol{p})$*.*

Let $b_j = \frac{p_j}{1-p_j}$, for $j \in \{1, \ldots, d\}$; the expression of the copula C in (4.13) becomes

$$C(\boldsymbol{u}) = \prod_{m=1}^{d} u_m \left(1 + \sum_{k=2}^{d} \sum_{1\leq j_1 < \cdots < j_k \leq d} \nu_{j_1\ldots j_k} \left(1 - u_{j_1}^{b_{j_1}}\right) \cdots \left(1 - u_{j_k}^{b_{j_k}}\right)\right), \tag{4.14}$$

for $\boldsymbol{u} \in [0,1]^d$. The shape parameters p_j (or b_j), $j \in \{1, \ldots, d\}$, govern the dependence relation between the components of $\boldsymbol{U}$ in regard to the symmetry or asymmetry. When $p_j = p$, implying that $b_j = b = \frac{p}{1-p}$, for every $j \in \{1, \ldots, d\}$, the copula C in (4.14) corresponds to the multivariate version of the symmetric bivariate Huang-Kotz FGM copula introduced and studied in Section 2 of [60]. In [1], the authors propose an asymmetric bivariate Huang-Kotz FGM copula, while [4] suggests a multivariate version of the symmetric Huang-Kotz FGM copula; without however providing lower and upper bounds on the set of dependence parameters. The stochastic representation in (4.12) allows the authors of [7] to provide a multivariate extension of the Huang-Kotz FGM family of copulas with constraints on the dependence parameters; it is a key result to investigate dependence properties within this copula family and it easily provides a sampling algorithm. In this section, we build on the stochastic representation of the (4.12) to provide results on bounds of dependent aggregated risks. For a review on various extensions of the family of FGM copulas, including the Huang-Kotz FGM copulas, see [90] and [7].

We denote by $\mathcal{C}_d^{\boldsymbol{p}}$ and $\mathcal{C}_d^{p}$ the classes of d-variate GFGM copulas with parameters $\boldsymbol{p} = (p_1, \dots, p_d)$ and with a common parameter p, respectively. The notation $\boldsymbol{U} \in \mathcal{C}_d^{\boldsymbol{p}}$ indicates that the joint cdf C of the random vector $\boldsymbol{U}$ with uniform margins is a copula $C \in \mathcal{C}_d^{\boldsymbol{p}}$. In the special case $p = 1/2$, the class $\mathcal{C}_d^{1/2}$ coincides with the class $\mathcal{C}_d^{\mathrm{FGM}}$ of FGM copulas introduced in Section 4.1.

The authors of [7] provide a preliminary analysis of the possible dependence structures in each class $\mathcal{C}_d^{\boldsymbol{p}}$ in the bivariate case. In Figures 3 and 4 of their paper, the authors show the density functions of the bivariate GFGM copulas corresponding to maximal and minimal dependence for different values of $p_1, p_2 \in (0{,}1)$. Their analysis concludes that for $p_1, p_2 \in (0.5{,}1)$ the maximal dependence copulas have more mass in the upper tail, while for $p_1, p_2 \in (0{,}0.5)$ in the lower tail. On the contrary, due to the moderate dependence of the FGM copula, the minimal dependence copulas have mass almost uniformly on the unit square, with less mass in the upper tail when $p_1, p_2 \in (0{,}0.5)$, and less mass in the lower tail when $p_1, p_2 \in (0.5{,}1)$. In Figure 1 of [7], instead, they analyze the case of a common p and compare the maximal and minimal values of Spearman's rho in classes $\mathcal{C}_d^{p}$ for different values of the common parameter p, as the dimension d varies. They show that the range of correlation spanned by the copula increases as p increases.

In [7], the authors mention in Remark 1 that the class $\mathcal{C}_d^{\boldsymbol{p}}$ shares the geometrical structure of $\mathcal{B}_d(\boldsymbol{p})$. Recall that $\mathcal{B}_d(\boldsymbol{p})$ is a convex polytope with extremal pmfs r_k, for $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$; see Section 1.2.2. Let $\boldsymbol{R}_k$ be a Bernoulli random vector with pmf r_k. Any cdf $F_{\boldsymbol{U}} \in \mathcal{C}_d^{\boldsymbol{p}}$ is a convex combination of $F_{\boldsymbol{U}^{(\boldsymbol{R}_k)}}$, where, for every $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$, the uniform random vector $\boldsymbol{U}^{(\boldsymbol{R}_k)}$ is built from $\boldsymbol{R}_k$ according to (4.12). Indeed, by using the stochastic representation in (4.12), for any $\boldsymbol{U} \in \mathcal{C}_d^{\boldsymbol{p}}$ there is a multivariate Bernoulli vector $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$ such that $\boldsymbol{U} = \boldsymbol{U}^{(\boldsymbol{X})} = \boldsymbol{U}_0^{1-\boldsymbol{p}} \boldsymbol{U}_1^{\boldsymbol{X}}$, and we have

$$
\begin{aligned}
F_{\boldsymbol{U}^{(\boldsymbol{x})}}(\boldsymbol{u}) &= \mathbb{P}\Big(U_1^{(\boldsymbol{X})} \le u_1, \dots, U_d^{(\boldsymbol{X})} \le u_d\Big) \\
&\overset{*}{=} \sum_{\boldsymbol{x} \in \mathcal{X}_d} \mathbb{P}\Big(U_{0,1}^{1-p_1} U_{1,1}^{x_1} \le u_1, \dots, U_{0,d}^{1-p_d} U_{1,d}^{x_d} \le u_d\Big) f_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k \sum_{\boldsymbol{x} \in \mathcal{X}_d} \mathbb{P}\Big(U_{0,1}^{1-p_1} U_{1,1}^{x_1} \le u_1, \dots, U_{0,d}^{1-p_d} U_{1,d}^{x_d} \le u_d\Big) r_k(\boldsymbol{x}) \\
&\overset{**}{=} \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k F_{\boldsymbol{U}^{(\boldsymbol{R}_k)}}(\boldsymbol{u}), \qquad \boldsymbol{u} \in [0{,}1]^d,
\end{aligned}
$$

where equalities $\overset{*}{=}$ and $\overset{**}{=}$ respectively follow from the independence of $\boldsymbol{U}_0$, $\boldsymbol{U}_1$ and $\boldsymbol{X}$ and of $\boldsymbol{U}_0$, $\boldsymbol{U}_1$ and $\boldsymbol{R}_k$, $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$. It follows that any GFGM copula C admits the representation

$$C(\boldsymbol{u}) = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k C_{\boldsymbol{R}_k}(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d, \tag{4.15}$$

where $C_{\boldsymbol{R}_k}$ is the copula associated to $\boldsymbol{U}^{(\boldsymbol{R}_k)}$, for every $k \in \{1, \ldots, n_d^{\boldsymbol{p}}\}$. The class $\mathcal{C}_d^{\boldsymbol{p}}$ of copulas is therefore a convex polytope.

We conclude this section with a simple but general result for a class of multivariate distributions with dependence structure built with a family of copulas being a convex polytope. Our aim is to later investigate the specific class $\mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$ of distributions with marginal distributions $F_1, \ldots, F_d$ and with a copula in the class $\mathcal{C}_d^{\boldsymbol{p}}$.

Proposition 4.3.1. *Let $\mathcal{C}$ be a class of copulas. Let $\mathcal{F}_d(F_1, \ldots, F_d)$ be a class of multivariate distributions with marginal distributions $F_1, \ldots, F_d$ and with a copula in the class $\mathcal{C}$. If $\mathcal{C}$ is a convex polytope with extremal points $\widetilde{C}_1, \ldots, \widetilde{C}_n$, then $\mathcal{F}_d(F_1, \ldots, F_d)$ is a convex polytope with extremal points $\widetilde{F}_1, ..., \widetilde{F}_n$, where*

$$\widetilde{F}_k(\boldsymbol{u}) = \widetilde{C}_k(F_1(u_1), ..., F_d(u_d)),$$

for $\boldsymbol{u} \in \mathbb{R}^d$ and $k \in \{1, \ldots, n\}$.

Proof. Consider $F \in \mathcal{F}_d(F_1, \ldots, F_d)$. Then, for $\boldsymbol{u} \in \mathbb{R}^d$, we have

$$F(\boldsymbol{u}) = C(F_1(u_1), ..., F_d(u_d)) = \sum_{k=1}^{n} \lambda_k \widetilde{C}_k(F_1(u_1), ..., F_d(u_d)),$$

for some $\lambda_1 \geq 0, \ldots, \lambda_n \geq 0$ such that $\sum_{k=1}^n \lambda_k = 1$. After defining

$$\widetilde{F}_k(\boldsymbol{x}) = \widetilde{C}_k(F_1(x_1), \ldots, F_d(x_d)),$$

for $k \in \{1, \ldots, n\}$, the desired result directly follows. □

Corollary 4.3.2. *The class $\mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$ of distributions is a convex polytope with extremal points the distributions associated to the extremal pmfs r_k of $\mathcal{B}_d(\boldsymbol{p})$, $k \in \{1, \ldots, n_d^{\boldsymbol{p}}\}$.*

4.3.1 A new representation theorem and consequences

We generalize the results of the previous section and introduce a stochastic representation for any random vector $\boldsymbol{Y}$ with distribution in $\mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$. We then use this representation in two particular cases, more precisely with exponential and discrete marginals, for which the expression of the distribution of the sum is analytical. The family of mixed Erlang distributions could also be considered. The notation $\boldsymbol{Y} \in \mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$ indicates that the cdf of $\boldsymbol{Y}$, $F_{\boldsymbol{Y}}$, belongs to the class $\mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$.

Theorem 4.3.3. *Let $V_{0,1}, \ldots, V_{0,d}, V_{1,1}, \ldots, V_{1,d}$ be independent random variables, where $V_{0,j} \sim Beta\left(\frac{1}{1-p_j}, 1\right)$, $p_j \in (0,1)$, and $V_{1,j} \sim U(0,1)$, for $j \in \{1, \ldots, d\}$. Fix some margins $F_1, \ldots, F_d$ and, for $h \in \{0,1\}$, let $\boldsymbol{Z}_h = (Z_{h,1}, \ldots, Z_{h,d})$ be vectors of independent random variables with $Z_{0,j} \stackrel{\mathcal{L}}{=} F_j^{-1}(V_{0,j})$ and $Z_{1,j} \stackrel{\mathcal{L}}{=} F_j^{-1}(V_{0,j}V_{1,j})$, for all $j \in \{1, \ldots, d\}$. Define the random vector*

$$\boldsymbol{Y} = (\boldsymbol{1} - \boldsymbol{X})\boldsymbol{Z}_0 + \boldsymbol{X}\boldsymbol{Z}_1, \tag{4.16}$$

where $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$, for $\boldsymbol{p} = (p_1, \ldots, p_d) \in (0,1)^d$. Then, the distribution of the random vector $\boldsymbol{Y}$ has margins $F_1, \ldots, F_d$ and copula $C \in \mathcal{C}_d^{\boldsymbol{p}}$, i.e., $F_{\boldsymbol{Y}} \in \mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$.

Proof. Let $\boldsymbol{U} \in \mathcal{C}_d^{\boldsymbol{p}}$ and hence $\boldsymbol{U} = \boldsymbol{U}_0^{1-\boldsymbol{p}}\boldsymbol{U}^{\boldsymbol{X}}$, for some $\boldsymbol{X} \in \mathcal{B}_d(\boldsymbol{p})$, with $\boldsymbol{p} = (p_1, \dots, p_d)$. In the proof of Theorem 2 of [7], we have

$$F_{U_j}(u) = \mathbb{P}(U_{0,j}^{1-p_j}U_{1,j}^{X_j} \le u) = \mathbb{P}(X_j = 0)F_{U_j|X_j=0}(u) + \mathbb{P}(X_j = 1)F_{U_j|X_j=1}(u),$$

where

$$F_{U_j|X_j=0}(u) = u^{\frac{1}{1-p_j}},$$

and

$$F_{U_j|X_j=1}(u) = \frac{u}{p_j} - \frac{1-p_j}{p_j}u^{\frac{1}{1-p_j}}, \quad j \in \{1, \dots, d\}.$$

Thus, $(U_j|X_j = 0) \stackrel{\mathcal{L}}{=} V_{0,j}$ and $(U_j|X_j = 1) \stackrel{\mathcal{L}}{=} V_{0,j}V_{1,j}$, $j \in \{1, \dots, d\}$, and the assert follows. □

Obviously, the special case of uniform margins leads to a stochastic representation of the GFGM family of copulas equivalent to (4.12). We notice that for $p = \frac{1}{2}$ we find the stochastic representation of FGM copulas in [8].

Theorem 4.3.3 provides a very useful representation of the random vector $\boldsymbol{Y}$. This helps us to derive plenty of results such as examining the distribution of any integrable function of $\boldsymbol{Y}$ and analyzing pairwise dependence properties of $\boldsymbol{Y}$. The following corollary considers the expectation of functionals of $\boldsymbol{Y}$ for which we derive bounds, in Section 4.3.3, building on the geometrical structure of Bernoulli vectors, the distribution of the sum $S_{\boldsymbol{Y}} = Y_1 + \dots + Y_d$ which may represent aggregate risks in a portfolio and some results on correlation between components of $\boldsymbol{Y}$ to analyze dependence corresponding to minimal aggregate risks.

Corollary 4.3.4. *Let $\boldsymbol{Y} \in \mathcal{G}_d^{\boldsymbol{p}}(F_1, \dots, F_d)$, and let r_k, for $k \in \{1, \dots, n_d^{\boldsymbol{p}}\}$, be the extremal pmfs of $\mathcal{B}_d(\boldsymbol{p})$. The following holds.*

1. *Let $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ be a real-valued function for which the following expectations exist; we have*

$$E[\varphi(Y_1, \dots, Y_d)] = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k \sum_{\boldsymbol{x} \in \mathcal{X}_d} r_k(\boldsymbol{x}) E[\varphi(Z_{x_1,1}, \dots, Z_{x_d,d})]. \tag{4.17}$$

2. *The cdf of the sum $S_{\boldsymbol{Y}} = \sum_j^d Y_j$ is given by*

$$F_{S_{\boldsymbol{Y}}}(y) = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k \sum_{\boldsymbol{x} \in \mathcal{X}_d} r_k(\boldsymbol{x}) F_{S_{\boldsymbol{Z}_{\boldsymbol{x}}}}(y), \quad y \in \mathbb{R}. \tag{4.18}$$

where, for every $\boldsymbol{x} \in \mathcal{X}_d$, $F_{S_{\boldsymbol{Z}_{\boldsymbol{x}}}}$ is the cdf of the random variable $S_{\boldsymbol{Z}_{\boldsymbol{x}}} = \sum_{j=1}^d Z_{x_j,j}$.

3. *The covariance between each pair of components (Y_{j_1}, Y_{j_2}) of $\boldsymbol{Y}$ is*

$$Cov(Y_{j_1}, Y_{j_2}) = Cov(Y_{j_1}, Y_{j_2})\gamma_{j_1,j_2} = \nu_{j_1,j_2} p_{j_1} p_{j_2} \gamma_{j_1,j_2}, \tag{4.19}$$

where, for every $1 \le j_1 < j_2 \le d$,

$$\gamma_{j_1,j_2} = (E[Z_{1,j_1}] - E[Z_{0,j_1}])(E[Z_{1,j_2}] - E[Z_{0,j_2}]).$$

The sharp bounds of the covariance are

$$(\max(p_{j_1} + p_{j_2} - 1{,}0) - p_{j_1}p_{j_2})\gamma_{j_1,j_2} \leq Cov(Y_{j_1}, Y_{j_2}) \leq p_{j_1}(1 - p_{j_2})\gamma_{j_1,j_2},$$

for every $1 \leq j_1 < j_2 \leq d$.

4. *If* $\boldsymbol{Y}$ *has continuous marginal distributions, Spearman's rho between each pair of components* (Y_{j_1}, Y_{j_2}) *of* $\boldsymbol{Y}$*, for* $1 \leq j_1 < j_2 \leq d$*, is*

$$\rho_S(Y_{j_1}, Y_{j_2}) = \frac{3\,Cov(X_{j_1}, X_{j_2})}{(2 - p_{j_1})(2 - p_{j_2})} = \frac{3\,\nu_{j_1,j_2}p_{j_1}p_{j_2}}{(2 - p_{j_1})(2 - p_{j_2})},$$

where $\boldsymbol{X} = (X_1, \dots, X_d)$ *is the Bernoulli random vector corresponding to* $\boldsymbol{Y}$ *of Theorem 4.3.3. The sharp bounds of the Spearman's rho are given by*

$$\frac{3(\max(p_{j_1} + p_{j_2} - 1{,}0) - p_{j_1}p_{j_2})}{(2 - p_{j_1})(2 - p_{j_2})} \leq \rho_S(Y_{j_1}, Y_{j_2}) \leq \frac{3p_{j_1}(1 - p_{j_2})}{(2 - p_{j_1})(2 - p_{j_2})}.$$

Proof. From the stochastic representation in (4.16) of Theorem 4.3.3, we have, for $j \in \{1, \dots, d\}$,

$$Y_j = \begin{cases} Z_{1,j}, & \text{if } X_j = 1, \\ Z_{0,j}, & \text{if } X_j = 0. \end{cases}$$

Thus, $Y_j = Z_{X_j,j}$, $j \in \{1, \dots, d\}$, and by conditioning on $\boldsymbol{X}$, the expectation of a function φ of $\boldsymbol{Y}$ is given by

$$E[\varphi(Y_1, \dots, Y_d)] = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f_{\boldsymbol{X}}(\boldsymbol{x}) E[\varphi(Z_{x_1,1}, \dots, Z_{x_d,d})], \tag{4.20}$$

assuming that the expectations exist.

1. Since $\mathcal{B}_d(\boldsymbol{p})$ is a convex polytope, by replacing (1.14) in (4.20), (4.17) follows.

2. By choosing $\varphi(\boldsymbol{x}) = \mathbb{1}\{x_1 + \cdots + x_d \leq y\}$, where $\mathbb{1}\{A\} = 1$, if A is true, and $\mathbb{1}\{A\} = 0$, otherwise, (4.18) follows.

3. For every $1 \leq j_1 < j_2 \leq d$, the covariance between Y_{j_1} and Y_{j_2} in (4.19) follows by considering $\varphi(\boldsymbol{x}) = x_{j_1}x_{j_2}$ in (4.20).

4. Spearman's rho ρ_S for any pair of continuous rvs (Y_{j_1}, Y_{j_2}), for $1 \leq j_1 < j_2 \leq d$, is given by

$$\rho_S(Y_{j_1}, Y_{j_2}) = \rho_P(U_{j_1}, U_{j_2}) = \frac{E[U_{j_1}U_{j_2}] - E[U_{j_1}]E[U_{j_2}]}{\sqrt{Var(U_{j_1})Var(U_{j_2})}}, \tag{4.21}$$

where ρ_P indicates the Pearson's correlation and $Var(U_{j_1}) = Var(U_{j_2}) = \frac{1}{12}$. Using the representation in (4.12), the expression for $E[U_{j_1}U_{j_2}]$, $1 \leq j_1 < j_2 \leq d$, is

$$E[U_{j_1}U_{j_2}] = E[U_{0,j_1}^{1-p_{j_1}} U_{0,j_2}^{1-p_{j_2}} U_{1,j_1}^{X_{j_1}} U_{1,j_2}^{X_{j_2}}] = E[U_{0,j_1}^{1-p_{j_1}}]E[U_{0,j_2}^{1-p_{j_2}}]E[U_{1,j_1}^{X_{j_1}} U_{1,j_2}^{X_{j_2}}],$$

where

$$E[U_{0,j_1}^{1-p_{j_1}}]E[U_{0,j_2}^{1-p_{j_2}}] = \frac{1}{2-p_{j_1}}\frac{1}{2-p_{j_2}},$$

and

$$E[U_{1,j_1}^{X_{j_1}}U_{1,j_2}^{X_{j_2}}] = \sum_{(i_{j_1},i_{j_2})\in\mathcal{X}_2} f_{\boldsymbol{X}}^{(j_1,j_2)}(i_{j_1},i_{j_2})E[U_{1,j_1}^{i_{j_1}}]E[U_{1,j_2}^{i_{j_2}}], \tag{4.22}$$

where $f_{\boldsymbol{X}}^{(j_1,j_2)}$ denotes the (j_1,j_2)-marginal pmf of $\boldsymbol{X}=(X_1,\dots,X_d)$. Finally, replacing (4.22) in (4.21), for every $1\le j_1<j_2\le d$, the expression of the Spearman's rho is given by

$$\begin{aligned}\rho_S(Y_{j_1},Y_{j_2}) &= \frac{12f_{\boldsymbol{X}}^{(j_1,j_2)}(0{,}0)+6f_{\boldsymbol{X}}^{(j_1,j_2)}(0{,}1)+6f_{\boldsymbol{X}}^{(j_1,j_2)}(1{,}0)+3f_{\boldsymbol{X}}^{(j_1,j_2)}(1{,}1)}{(2-p_{j_1})(2-p_{j_2})}-3\\ &= \frac{3(f_{\boldsymbol{X}}^{(j_1,j_2)}(1{,}1)-p_1p_2)}{(2-p_{j_1})(2-p_{j_2})} = \frac{3\,Cov(X_{j_1},X_{j_2})}{(2-p_{j_1})(2-p_{j_2})}.\end{aligned}$$

Assume $p_{j_1}\le p_{j_2}$. In Section 3.2 of [46], the authors found the bounds for the covariance between (X_{j_1},X_{j_2}) for any pair of Bernoulli variables. The maximum value is $Cov(X_{j_1},X_{j_2})=p_{j_1}(1-p_{j_2})$, while the minimum value is $Cov(X_{j_1},X_{j_2})=\max(p_{j_1}+p_{j_2}-1{,}0)-p_{j_1}p_{j_2}$. In the first case, $E[X_{j_1}X_{j_2}]=\mathbb{P}(X_{j_1}=1,X_{j_2}=1)=p_{j_1}$ and $Cov(X_{j_1},X_{j_2})=p_{j_1}(1-p_{j_2})$. In the second case, $E[X_{j_1}X_{j_2}]=\mathbb{P}(X_{j_1}=1,X_{j_2}=1)=\max(p_{j_1}+p_{j_2}-1{,}0)$ and $Cov(X_{j_1},X_{j_2})=\max(p_{j_1}+p_{j_2}-1{,}0)-p_{j_1}p_{j_2}$. Therefore, we have the following bounds for Spearman's rho of any pair (Y_{j_1},Y_{j_2}) with $1\le j_1<j_2\le d$,

$$\frac{3(\max(p_{j_1}+p_{j_2}-1{,}0)-p_{j_1}p_{j_2})}{(2-p_{j_1})(2-p_{j_2})}\le\rho_S(Y_{j_1},Y_{j_2})\le\frac{3p_{j_1}(1-p_{j_2})}{(2-p_{j_1})(2-p_{j_2})}.$$

□

The relation (4.17) in particular holds for $\varphi(\boldsymbol{Y})=\phi(S_{\boldsymbol{Y}})$ and implies that the bounds of $E[\phi(S_{\boldsymbol{Y}})]$, for any ϕ for which the expectation exists, can be found by enumerating their values on the sums $S_{\boldsymbol{R}_k}$, $k\in\{1,\dots,n_p^{\mathcal{B}}\}$, and this is computationally feasible in low dimension.

We end this section by considering two examples for $\mathcal{G}_d^{\boldsymbol{p}}(F_1,\dots,F_d)$, where the margins are discrete in the first one and the margins are exponential in the second one.

Example 4.3.1 (Discrete margins)**.** *Let $\boldsymbol{Y}=(Y_1,\dots,Y_d)$ be defined as a d-dimensional random vector, where, for every $j\in\{1,\dots,d\}$, Y_j is a discrete random variable taking values on the set $A_j=\{0{,}1,\dots,n_j\}$, $n_j\in\mathbb{N}$, and with cdf F_j. When the joint distribution of $\boldsymbol{Y}$ belongs to the class $\mathcal{G}_d^{\boldsymbol{p}}(F_1,\dots,F_d)$, $\boldsymbol{Y}$ admits the representation in (4.16). Moreover, for all $j\in\{1,\dots,d\}$, to derive the values of the pmf of the discrete rvs $Z_{i,j}$, $i=0{,}1$, we observe that, for each $k\in A_j$, we have*

$$f_{Z_{0,j}}(k)=\mathbb{P}(Z_{0,j}=k)=\mathbb{P}(F_j^{-1}(V_{0,j})=k)=F_{V_{0,j}}(F_j(k))-F_{V_{0,j}}(F_j(k-1)),$$

and

$$f_{Z_{1,j}}(k) = \mathbb{P}(Z_{1,j} = k) = \mathbb{P}(F_j^{-1}(V_{0,j}V_{1,j}) = k)$$
$$= F_{V_{0,j}V_{1,j}}(F_j(k)) - F_{V_{0,j}V_{1,j}}(F_j(k-1)),$$

where

$$F_{V_{0,j}}(u) = u^{\frac{1}{1-p_j}}, \quad u \in (0,1),$$

$$F_{V_{0,j}V_{1,j}}(u) = \frac{u}{p_j} - \frac{1-p_j}{p_j} u^{\frac{1}{1-p_j}}, \quad u \in (0,1),$$

and $F_j(-1) := 0$. *For all* $j \in \{1, \ldots, d\}$, *it follows that the expectations of the discrete rvs* $Z_{i,j}$, $i = 0,1$, *are*

$$E[Z_{0,j}] = n_j - \sum_{k=0}^{n_j-1} F_{V_{0,j}}(F_j(k)),$$

and

$$E[Z_{1,j}] = n_j - \sum_{k=0}^{n_j-1} F_{V_0 V_{1,j}}(F_j(k)).$$

Finally, we consider the sum $S_{\boldsymbol{Y}} = Y_1 + \cdots + Y_d$, *which can be rewritten using the representation in* (4.16), *as*

$$S_{\boldsymbol{Y}} = \sum_{j=1}^{d} (1 - X_j) Z_{0,j} + X_j Z_{1,j}. \tag{4.23}$$

From (4.23), *the expression of the probability generating function (pgf) of* $S_{\boldsymbol{Y}}$ *is given by*

$$\mathcal{P}_{S_{\boldsymbol{Y}}}(s) = \sum_{\boldsymbol{x} \in \mathcal{X}_d} f_{\boldsymbol{X}}(\boldsymbol{x}) \prod_{j=1}^{d} \mathcal{P}_{Z_{x_j,j}}(s), \quad s \in [-1,1], \tag{4.24}$$

where the pgf of $Z_{x_j,j}$ *is*

$$\mathcal{P}_{Z_{x_j,j}}(s) = E[s^{Z_{x_j,j}}] = \sum_{k=0}^{n_j} f_{Z_{x_j,j}}(k) s^k, \quad s \in [-1,1],$$

for $x_j \in \{0,1\}$ *and* $j \in \{1, \ldots, d\}$. *One uses the Fast Fourier Transform (**FFT**) algorithm of [28] to extract the values of the pmf of* $S_{\boldsymbol{Y}}$ *from its pgf in* (4.24). *Details about that efficient approach is explained in Chapter 30 of [29]. See also [40] for FFT applications in actuarial science and quantitative risk management. This procedure is illustrated in Example 4.3.6, within Section 4.3.5.* □

Example 4.3.2 (Exponential margins)**.** *Assume that* $\boldsymbol{Y} \in \mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$, *where* F_j *is the cdf of an exponential distribution with mean* $\frac{1}{\lambda_j}$, $j \in \{1, \ldots, d\}$. *In [7], the authors present an alternative stochastic representation of* $\boldsymbol{Y}$ *equivalent to* (4.16) *that allows finding an analytical expression for the distribution of the sum of the components of the random vector* $\boldsymbol{Y}$. *This other stochastic representation is*

$$\boldsymbol{Y} = \boldsymbol{W}_1 + \boldsymbol{X}\boldsymbol{W}_2, \tag{4.25}$$

where the components $W_{1,j}$ *within* $\boldsymbol{W}_1$ *and* $W_{2,j}$ *within* $\boldsymbol{W}_2$ *are independent, with* $W_{1,j}$ *following an exponential distribution with mean* $\frac{1-p_j}{\lambda_j}$ *and* $W_{2,j}$ *also following an exponential distribution but with mean* $\frac{1}{\lambda_j}$*, for* $j \in \{1, \ldots, d\}$*. The random vectors* $\boldsymbol{X}$*,* $\boldsymbol{W}_1$ *and* $\boldsymbol{W}_2$ *are independent.*

From (4.25)*, Pearson's coefficient becomes*

$$\rho_P(Y_{j_1}, Y_{j_2}) = \frac{Cov(X_{j_1}, X_{j_2})E[W_{2,j_1}]E[W_{2,j_2}]}{\sqrt{Var(Y_{j_1})Var(Y_{j_2})}} = Cov(X_{j_1}, X_{j_2}) \tag{4.26}$$

since $E[W_{2,j_1}] = \frac{1}{\lambda_{j_1}}$*,* $E[W_{2,j_2}] = \frac{1}{\lambda_{j_2}}$*,* $Var(Y_{j_1}) = \frac{1}{\lambda_{j_1}^2}$*, and* $Var(Y_{j_2}) = \frac{1}{\lambda_{j_2}^2}$*, for* $1 \leq j_1 < j_2 \leq d$*.*

Considering now the sum $S_{\boldsymbol{Y}}$ *of the components of* $\boldsymbol{Y}$*, the representation in* (4.25) *allows us to express it as follows:*

$$S_{\boldsymbol{Y}} \stackrel{\mathcal{L}}{=} W_{1,1} + X_1 W_{2,1} + \ldots + W_{1,d} + X_d W_{2,d} = \sum_{j=1}^{d} W_{1,j} + \sum_{j=1}^{d} X_j W_{2,j}. \tag{4.27}$$

To identify the distribution of $S_{\boldsymbol{Y}}$*, we find from* (4.27)*, the expression of the Laplace-Stieltjes transform (LST) of* $S_{\boldsymbol{Y}}$*, denoted by* $\mathcal{L}_{S_{\boldsymbol{Y}}}$ *and given by*

$$\mathcal{L}_{S_{\boldsymbol{Y}}}(t) = \prod_{j=1}^{d} \mathcal{L}_{W_{1,j}}(t) \left(\sum_{\boldsymbol{x} \in \mathcal{X}_d} f_{\boldsymbol{X}}(\boldsymbol{x}) \prod_{j=1}^{d} \left(\mathcal{L}_{W_{2,j}}(t)\right)^{x_j} \right), \quad t \geq 0. \tag{4.28}$$

Using techniques explained in [107] and [30], the LST of S *admits the representation given by*

$$\mathcal{L}_{S_{\boldsymbol{Y}}}(t) = \sum_{k=0}^{\infty} \eta_k \left(\frac{\beta}{\beta + t}\right)^{d+k}, \quad t \geq 0$$

where $\beta = \max\left(\lambda_1, \ldots, \lambda_d, \frac{\lambda_1}{1-p_1}, \ldots, \frac{\lambda_d}{1-p_d}\right)$*,* $\eta_k \geq 0$ *for* $k \geq 0$*, and* $\sum_{k=0}^{\infty} \eta_k = 1$*. From the expression of its LST in* (4.28)*, it follows that the rv* S *follows a mixed Erlang distribution. An application of* (4.28) *is provided in Example 4.3.5. Details about the computation of the sequence of probabilities* $\{\eta_k, k \in \mathbb{N}_0\}$ *are explained in [107] and [30].* □

In Example 4.3.2, we find the distribution of S, which comes with an analytical expression for F_S because we assume that the margins are exponential. In most cases, if the margins do not belong to a class of distributions closed under convolution, one must resort to numerical approximations. One of them is to use discretization techniques as explained in Section 4.3 of [10] jointly with the method explained in Example 4.3.1.

4.3.2 Minimal convex sums under GFGM dependence

Based on Sections 1.2.2 and 1.3.1, we provide in this section the geometrical structure embedded within the class of sums of components of random vectors whose joint distribution belongs to the class of d-variate GFGM copulas with a common parameter p, that is, the class $\mathcal{C}_d^p$. Then, we are able to present a new result about convex order, as stated in

Theorem 4.3.8. Let us begin with the following proposition, which is a direct consequence of the one-to-one relationship between exchangeable Bernoulli distributions and discrete distributions, discussed in Section 1.3, viz.,

$$\mathcal{E}_d(p) \leftrightarrow \mathcal{D}_d(dp) \tag{4.29}$$

Proposition 4.3.5. *Let $\boldsymbol{X} \in \mathcal{B}_d(p)$ and let $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$. Then, there exists an exchangeable Bernoulli random vector $\boldsymbol{X}^e \in \mathcal{E}_d(p)$ such that $S_{\boldsymbol{X}^e} = \sum_{j=1}^d X_j^e \in \mathcal{D}_d(dp)$ has the same distribution as $S_{\boldsymbol{X}}$.*

Proof. Given $\boldsymbol{X} \in \mathcal{B}_d(p)$, the sum of the components $S_{\boldsymbol{X}} = X_1 + \cdots + X_d$ is a random variable that takes values in the set $\{0, \dots, d\}$ and whose mean is $\mathbb{E}[S_{\boldsymbol{X}}] = dp$, that is $S_{\boldsymbol{X}} \in \mathcal{D}_d(dp)$. Therefore, from (4.29), it follows that there exists one exchangeable Bernoulli random vector $\boldsymbol{X}^e \in \mathcal{E}_d(p)$ such that $S_{\boldsymbol{X}^e} = \sum_{j=1}^d X_j^e \stackrel{\mathcal{L}}{=} S_{\boldsymbol{X}}$. □

Let $\mathcal{S}_d^p$ denote the class of sums of components of random vectors with multivariate distributions in $\mathcal{C}_d^p$. Let us indicate with $S_{(\boldsymbol{U},\boldsymbol{X})} = \sum_{j=1}^d U_j^{(\boldsymbol{X})}$, where $\boldsymbol{U}^{(\boldsymbol{X})}$ is the random vector whose joint cdf is the copula C associated to $\boldsymbol{X} \in \mathcal{B}_d(p)$, as defined in (4.12). In order to study the geometrical structure of $\mathcal{S}_d^p$, we need to investigate the correspondence between distributions belonging to this class and multivariate Bernoulli distributions of the class $\mathcal{B}_d(p)$.

Lemma 4.3.6. *Let $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{B}_d(p)$ and let $S_{\boldsymbol{X}} = \sum_{j=1}^d X_j$ and $S_{\boldsymbol{X}'} = X_1' + \cdots + X_d'$. Let C and C' be the GFGM(p) copulas corresponding to $\boldsymbol{X}$ and $\boldsymbol{X}'$, respectively. Finally, let $\boldsymbol{U}$ and $\boldsymbol{U}'$ be uniform random vectors with joint cdfs C and C', respectively. If $S_{\boldsymbol{X}} \stackrel{\mathcal{L}}{=} S_{\boldsymbol{X}'}$, then $\sum_{j=1}^d U_j \stackrel{\mathcal{L}}{=} \sum_{j=1}^d U_j'$.*

Proof. Let F and F' be the cdfs of $\sum_{j=1}^d U_j$ and $\sum_{j=1}^d U_j'$, respectively. From (4.12), we have the following stochastic representation:

$$U_j = U_{0,j}^{1-p} U_{1,j}^{X_j}, \quad j \in \{1, \dots, d\},$$

where $\boldsymbol{U}_0 = (U_{0,1}, \dots, U_{0,d})$ and $\boldsymbol{U}_1 = (U_{1,1}, \dots, U_{1,d})$ are vectors of independent standard uniform random variables and $\boldsymbol{U}_0$, $\boldsymbol{U}_1$, and $\boldsymbol{X} = (X_1, \dots, X_d)$ are independent. Then, we have

$$\begin{aligned}
F(u) &= \mathbb{P}\Big(\sum_{j=1}^d U_j \leq u\Big) = \mathbb{P}\Big(\sum_{j=1}^d U_{0,j}^{1-p} U_{1,j}^{X_j} \leq u\Big) \\
&= \sum_{\boldsymbol{x} \in \mathcal{X}_d} \mathbb{P}\Big(\sum_{j=1}^d U_{0,j}^{1-p} U_{1,j}^{x_j} \leq u\Big) f_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= \sum_{\boldsymbol{x} \in \mathcal{X}_d} \mathbb{P}\Big(\sum_{j:x_j=1} U_{0,j}^{1-p} U_{1,j} + \sum_{j:x_j=0} U_{0,j}^{1-p} \leq u\Big) f_{\boldsymbol{X}}(\boldsymbol{x}), \quad u \in [0, d].
\end{aligned} \tag{4.30}$$

Since $\boldsymbol{U}_0$ and $\boldsymbol{U}_1$ are independent vectors of iid random variables, the distribution of the sum $\sum_{j:x_j=1} U_{0,j}^{1-p} U_{1,j} + \sum_{j:x_j=0} U_{0,j}^{1-p}$ does not depend on the position of the 1's in $\boldsymbol{x}$, but only on the number of 1's, that is, on the sum of the components $|\boldsymbol{x}|$. Hence, the cdf in (4.30) becomes

$$F(u) = \sum_{k=0}^{d} \sum_{\boldsymbol{x}\in\mathcal{X}_d^k} \mathbb{P}\bigg(\sum_{j=1}^{k} U_{0,j}^{1-p}U_{1,j} + \sum_{j=k+1}^{d} U_{0,j}^{1-p} \leq u\bigg) f_{\boldsymbol{X}}(\boldsymbol{x}), \tag{4.31}$$

where $\mathcal{X}_d^k = \{\boldsymbol{x} \in \mathcal{X}_d : \sum_{j=1}^d x_j = k\}$ and we set $\sum_{j=1}^0 U_{0,j}^{1-p}U_{1,j} := 0$ and $\sum_{j=d+1}^d U_{0,j}^{1-p} := 0$. It follows that

$$F(u) = \sum_{k=0}^{d} \mathbb{P}\bigg(\sum_{j=1}^{k} U_{0,j}^{1-p}U_{1,j} + \sum_{j=k+1}^{d} U_{0,j}^{1-p} \leq u\bigg) f_{S_{\boldsymbol{X}}}(k), \quad u \in [0, d]. \tag{4.32}$$

Therefore, given that the first probability in the summation of (4.32) does not depend on $\boldsymbol{x}$ for $k \in \{0, \dots, d\}$, we conclude that $S_{\boldsymbol{X}} \overset{\mathcal{L}}{=} S_{\boldsymbol{X}'}$ implies $F(u) = F'(u)$, for every $u \in [0, d]$. □

We can now prove the following theorem that characterizes $\mathcal{S}_d^p$. The geometrical structure of $\mathcal{S}_d^p$ is based on the geometrical structure of $\mathcal{E}_d(p)$, described in Section 1.3.2. Recall that the extremal pmfs of $\mathcal{E}_d(p)$ are r_{j_1,j_2}^e, which are given in (1.27). To simplify the notation, in this section we enumerate the extremal pmfs $\{r_{j_1,j_2}^e \colon 1 \leq j_1 \leq \lfloor pd \rfloor < j_2 \leq d\}$ as $\{r_k^e \colon k \in \{1, \dots, m_d^p\}$, where $m_d^p \in \mathbb{N}$ is the total number of extremal pmfs of $\mathcal{E}_d(p)$ ([43, Corollary 4.6]).

Theorem 4.3.7. *The class $\mathcal{S}_d^p$ is a convex polytope, and its extremal cdfs are the cdfs $F_{S_{(\boldsymbol{U},\boldsymbol{E}_k)}}$ of the random variables $S_{(\boldsymbol{U},\boldsymbol{E}_k)} = \sum_{j=1}^d U_j^{(\boldsymbol{E}_k)}$, where $\boldsymbol{U}^{(\boldsymbol{E}_k)} \overset{\mathcal{L}}{=} \boldsymbol{U}_0^{1-p}\boldsymbol{U}_1^{\boldsymbol{E}_k}$, and $\boldsymbol{E}_k$ is a Bernoulli random vector with pmf the extremal pmf r_k^e of the class $\mathcal{E}_d(p)$, for every $k \in \{1, \dots, m_d^p\}$.*

Proof. Consider any $S \in \mathcal{S}_d^p$, then there exists $\boldsymbol{X} \in \mathcal{B}_d(p)$ such that $S = \sum_{i=1}^d U_i^{(\boldsymbol{X})}$. By Proposition 4.3.5, there exists a unique $\boldsymbol{X}^e \in \mathcal{E}(p)$ with $\sum_{j=1}^d X_j^e \overset{\mathcal{L}}{=} \sum_{j=1}^d X_j$. Thus by Lemma 4.3.6, $S \overset{\mathcal{L}}{=} \sum_{j=1}^d U_j^e$, where $\boldsymbol{U}^e \in \mathcal{C}_d^p$ is the uniform random vector with copula corresponding to $\boldsymbol{X}^e$. In other words, $\boldsymbol{U}^e$ admits the representation $\boldsymbol{U}^e \overset{\mathcal{L}}{=} \boldsymbol{U}_0^{1-p}\boldsymbol{U}_1^{\boldsymbol{X}^e}$, where $\boldsymbol{X}^e \in \mathcal{E}(p)$. Let F_S be the cdf of S. Then, we have

$$\begin{aligned}
F_S(u) &= \mathbb{P}\bigg(\sum_{j=1}^{d} U_j^e \leq u\bigg) = \mathbb{P}\bigg(\sum_{j=1}^{d} U_{0,j}^{1-p}U_{1,j}^{X_j^e} \leq u\bigg) \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}_d} \mathbb{P}\bigg(\sum_{j=1}^{d} U_{0,j}^{1-p}U_{1,j}^{x_j} \leq u\bigg) f_{\boldsymbol{X}^e}(\boldsymbol{x}) \\
&= \sum_{k=1}^{m_d^p} \lambda_k \sum_{\boldsymbol{x}\in\mathcal{X}_d} \mathbb{P}\bigg(\sum_{j=1}^{d} U_{0,j}^{1-p}U_{1,j}^{x_j} \leq u\bigg) f_{\boldsymbol{E}_k}(\boldsymbol{x}) = \sum_{k=1}^{m_d^p} \lambda_k \mathbb{P}\bigg(\sum_{j=1}^{d} U_j^{(\boldsymbol{E}_k)} \leq u\bigg) \\
&= \sum_{k=1}^{m_d^p} \lambda_k F_{S_{(\boldsymbol{U},\boldsymbol{E}_k)}}(u),
\end{aligned}$$

where $F_{S_{(\boldsymbol{U},\boldsymbol{E}_k)}}$ is the cdf of $S_{(\boldsymbol{U},\boldsymbol{E}_k)} = \sum_{j=1}^d U_j^{(\boldsymbol{E}_k)}$. □

From Theorem 4.3.7, we can complete the relationship between classes in (4.29) as follows:

$$\mathcal{S}_d^p \leftrightarrow \mathcal{E}_d(p) \leftrightarrow \mathcal{D}_d(dp).$$

We now study the convex order of sums of the components of random vectors with joint distribution described by a GFGM(p) copula.

Theorem 4.3.8. *Let $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{B}_d(p)$ be such that $\sum_{j=1}^d X_j \preceq_{cx} \sum_{j=1}^d X'_j$. Let C and C' be the GFGM copulas associated with $\boldsymbol{X}$ and $\boldsymbol{X}'$, and let $\boldsymbol{U}$ and $\boldsymbol{U}'$ be uniform random vectors with joint cdf C and C', respectively. Then,*

$$\sum_{j=1}^d U_j \preceq_{cx} \sum_{j=1}^d U'_j.$$

Proof. By Proposition 4.3.5, there exist two exchangeable Bernoulli random vectors $\boldsymbol{X}^e$ and $\boldsymbol{X}^{e'}$ of the class $\mathcal{E}_d(p)$ such that $\sum_{j=1}^d X_j \stackrel{\mathcal{L}}{=} \sum_{j=1}^d X_j^e$ and $\sum_{j=1}^d X'_j \stackrel{\mathcal{L}}{=} \sum_{j=1}^d X_j^{e'}$. Therefore, by Lemma 4.3.6, we have

$$\sum_{j=1}^d U_j \stackrel{\mathcal{L}}{=} \sum_{j=1}^d U_j^e \quad \text{and} \quad \sum_{j=1}^d U'_j \stackrel{\mathcal{L}}{=} \sum_{j=1}^d U_j^{e'}, \tag{4.33}$$

where $\boldsymbol{U}^e$ and $\boldsymbol{U}^{e'}$ are uniform random vectors with joint distributions given by the GFGM copulas associated to $\boldsymbol{X}^e$ and $\boldsymbol{X}^{e'}$, respectively. Moreover, since by hypothesis $\sum_{j=1}^d X_j \preceq_{\text{cx}} \sum_{j=1}^d X'_j$, then $\sum_{j=1}^d X_j^e \preceq_{\text{cx}} \sum_{j=1}^d X_j^{e'}$. Moreover, by Proposition 2.3.3 the double implication

$$\sum_{j=1}^d X_j^e \preceq_{\text{cx}} \sum_{j=1}^d X_j^{e'} \iff \boldsymbol{X}^e \preceq_{\text{sm}} \boldsymbol{X}^{e'}$$

holds, and, by Theorem 4.2 of [12], $\boldsymbol{X}^e \preceq_{\text{sm}} \boldsymbol{X}^{e'}$ implies $\boldsymbol{U}^e \preceq_{\text{sm}} \boldsymbol{U}^{e'}$. Since, given a convex function $\phi\colon \mathbb{R} \to \mathbb{R}$, the function $\psi(\boldsymbol{x}) = \phi(x_1 + \ldots + x_d)$ is supermodular, $\boldsymbol{X}^e \preceq_{\text{sm}} \boldsymbol{X}^{e'}$ implies, in particular, $\sum_{j=1}^d U_j^e \preceq_{\text{cx}} \sum_{j=1}^d U_j^{e'}$. Finally, by the equality in distribution in (4.33), we have $\sum_{j=1}^d U_j \preceq_{\text{cx}} \sum_{j=1}^d U'_j$. □

Theorem 4.3.8 obviously holds for FGM copulas by setting $p = \frac{1}{2}$.

Class $\mathcal{G}_d^p(F)$ of distributions

Let us introduce the class $\mathcal{G}_d^p(F)$ of joint cdfs with a copula in the class $\mathcal{C}_d^p$ and with the same marginal cdfs F. Consider $\boldsymbol{Y} \in \mathcal{G}_d^p(F)$. From Theorem 4.3.3, there exists $\boldsymbol{X} \in \mathcal{B}_d(p)$ such that $\boldsymbol{Y}$ is built from the Bernoulli random vector $\boldsymbol{X}$, according to the stochastic representation (4.16). Using this representation, we can generalize Lemma 4.3.6 as follows.

Lemma 4.3.9. *Let $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{B}_d(p)$ be such that $S_{\boldsymbol{X}} = \sum_{j=1}^d X_j$ and $S_{\boldsymbol{X}'} = \sum_{j=1}^d X'_j$ are equal in distribution. Let $\boldsymbol{Y}, \boldsymbol{Y}' \in \mathcal{G}_d^p(F)$ be the random vectors corresponding to $\boldsymbol{X}$ and $\boldsymbol{X}'$, respectively. Then, $\sum_{j=1}^d Y_j$ and $\sum_{j=1}^d Y'_j$ have the same distribution.*

Proof. Similarly to the proof of Lemma 4.3.6, Lemma 4.3.9 is proved observing that $\boldsymbol{Z}_0$ and $\boldsymbol{Z}_1$ have independent and identically distributed components. □

Again, the stochastic representation in (4.16) helps us to find the following generalization of Theorem 4.3.7.

Theorem 4.3.10. *The class $\mathcal{S}_d^p(F)$ of distributions of sums of components of vectors with distribution in $\mathcal{G}_d^p(F)$ is a convex polytope and its extremal points are the distributions of $S_{(\boldsymbol{Y},\boldsymbol{E}_k)} = \sum_{j=1}^d Y_j^{(\boldsymbol{E}_k)}$ where*

$$\boldsymbol{Y}^{(\boldsymbol{E}_k)} = (\mathbf{1} - \boldsymbol{E}_k)\boldsymbol{Z}_0 + \boldsymbol{E}_k\boldsymbol{Z}_1, \tag{4.34}$$

where $\boldsymbol{E}_k$ is an extremal point of $\mathcal{E}_d(p)$, for $k \in \{1, \dots, m_d^p\}$.

Proof. The proof is similar to the one of Theorem 4.3.7. □

Theorem 4.3.10 completes the relationship in (4.3.2) as follows:

$$\mathcal{S}_d^p(F) \leftrightarrow \mathcal{S}_d^p \leftrightarrow \mathcal{E}_d(p) \leftrightarrow \mathcal{D}_d(dp).$$

From this relationship, it follows that the number of extremal points in $\mathcal{S}_d^p(F)$ is $n_p^{\mathcal{D}}$, that is significantly lower than the number of extremal points in $\mathcal{G}_d^p(F)$ and they are analytical. Therefore we can also find them in high dimensions.

Finally, we have the following generalization of Theorem 4.3.8.

Theorem 4.3.11. *Let $\boldsymbol{X}, \boldsymbol{X}' \in \mathcal{B}_d(p)$ and let $\boldsymbol{Y} \in \mathcal{G}_d^p(F)$ and $\boldsymbol{Y}' \in \mathcal{G}_d^p(F)$, for any cdf F, be respectively built from $\boldsymbol{X}$ and $\boldsymbol{X}'$, as in (4.16). Then,*

$$\sum_{j=1}^d X_j \preceq_{cx} \sum_{j=1}^d X'_j \implies \sum_{j=1}^d Y_j \preceq_{cx} \sum_{j=1}^d Y'_j.$$

Proof. The proof of this theorem is along the same lines at the one of Theorem 4.3.8. □

We conclude this section by considering the two examples with exponential and discrete margins previously discussed in Example 4.3.1 and Example 4.3.2, but here in the special case of identically distributed risks.

Example 4.3.1 (Discrete margins, **continued**). *For the case of identically distributed discrete margins, we obtain the following expression for the pmf of $S_{\boldsymbol{Y}}$:*

$$f_{S_{\boldsymbol{Y}}}(k) = \sum_{j=0}^d f_{S_{\boldsymbol{X}}}(j) f_{Z_0}^{*(d-j)} * f_{Z_1}^{*j}(k), \quad k \in \mathbb{N}, \tag{4.35}$$

*where * denotes the convolution product and f^{*j} denotes the j-fold convolution product of the probability density function measure f with itself. It follows from (4.24) that the probability generating function (pgf) of S is given by*

$$\mathcal{P}_{S_{\boldsymbol{Y}}}(s) = \sum_{j=0}^d f_{S_{\boldsymbol{X}}}(j) \mathcal{P}_{Z_0}^{d-j}(s) \mathcal{P}_{Z_1}^j(s), \quad s \in [0,1], \tag{4.36}$$

*where $\mathcal{P}_{Z_0}$ and $\mathcal{P}_{Z_1}$ are the pgf of Z_0 and Z_1, respectively. As previously mentioned, it is more convenient to use the Fast Fourier Transform (**FFT**) algorithm to extract the values of the pmf of $S_{\boldsymbol{Y}}$ from the pgf in (4.36) rather than finding those values directly from (4.35). The procedure is illustrated in Example 4.3.6 provided in Section 4.3.5.* □

Example 4.3.2 (Exponential margins, **continued**). *First, let* $\boldsymbol{Y} \in \mathcal{G}_d^p(Exp(\lambda))$*, i.e.,* $Y_j \sim Exp(\lambda)$ *for every* $j \in \{1, \ldots, d\}$*, with GFGM(p) copula. The LST of* $S_{\boldsymbol{Y}}$ *in* (4.28) *becomes*

$$\mathcal{L}_{S_{\boldsymbol{Y}}}(t) = (\mathcal{L}_{W_1}(t))^d \left(\sum_{j=0}^{d} f_{S_{\boldsymbol{X}}}(j) \left(\mathcal{L}_{W_2}(t)\right)^j \right), \quad t \geq 0,$$

where $S_{\boldsymbol{X}} = \sum_{j=1}^{d} X_j$*. By identification of the LST transform, we conclude*

$$F_{S_{\boldsymbol{Y}}}(y) = f_{S_{\boldsymbol{X}}}(0) F_{W_1}^{*d}(y) + \sum_{j=1}^{d} f_{S_{\boldsymbol{X}}}(j) F_{W_1}^{*d} * F_{W_2}^{*j}(y), \quad y \geq 0,$$

where $F_{W_1}^{*d}$ *corresponds to the cdf of an Erlang*$(d, \frac{\lambda}{1-p})$ *distribution and* $F_{W_2}^{*j}$ *corresponds to the cdf of an Erlang*(j, λ) *distribution for* $j \in \{1, \ldots, d\}$*.* □

Minimal convex sums

The solution to the problem of finding the distribution of the vectors with minimal convex sums is a corollary of our main theorems, Theorems 4.3.8 and 4.3.11. Let $\boldsymbol{X} \in \mathcal{B}_d(p)$ and let $\boldsymbol{U}$ be the corresponding uniform random vector with GFGM(p) copula. The following corollary is a straightforward but important consequence of Theorem 4.3.8 and Theorem 4.3.11.

Corollary 4.3.12. *Let* $\boldsymbol{X} \in \mathcal{B}_d(p)$ *be a* Σ_{cx}*-smallest element.*

1. *Let* $\boldsymbol{U}$ *be a uniform random vector whose joint cdf is the GFGM(*p*) copula corresponding to* $\boldsymbol{X}$*. Then,* $\boldsymbol{U}$ *is a* Σ_{cx}*-smallest element in* $\mathcal{C}_d^p$*.*

2. *Let* $\boldsymbol{Y} \in \mathcal{G}_d^p(F)$ *with joint cdf defined with the GFGM(*p*) copula corresponding to* $\boldsymbol{X}$*. Then,* $\boldsymbol{Y}$ *is a* Σ_{cx}*-smallest element in* $\mathcal{G}_d^p(F)$*.*

Consequently, distributions, for which the lower bounds of a convex functional are reached, are built using a Σ_{cx}-smallest element of $\mathcal{B}_d(p)$. Obviously, using the upper Fréchet bound of $\mathcal{B}_d(p)$, we build the distributions of vectors with maximal convex sums.

4.3.3 Sharp bounds for risk measures

This section finds sharp bounds for functions of aggregated risks represented by random variables with cdf in $\mathcal{S}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$. As a consequence of Corollary 4.3.2, to derive sharp bounds for risk measures in the class $\mathcal{S}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$, we can proceed by enumerating their values on the extremal points. This is computationally expensive because the number of extremal points n_d^p explodes and becomes larger, as highlighted by the authors of [45]. In this section, we illustrate how this problem can be solved by finding sharp bounds for risk measures in the classes $\mathcal{C}_d^p$ and $\mathcal{G}_d^p(F_X)$, for any $p \in (0,1)$.

As a motivation, we start with an example showing that the assumption of common margins in Lemma 4.3.9 and Theorem 4.3.11 is necessary. Note that there is no restriction on the chosen discrete distributions.

k	$F_1(k)$	$F_2(k)$	$F_3(k)$
0	0.1	0.1	0.8
1	0.2	0.4	1.0
2	0.3	0.7	1.0
3	1.0	1.0	1.0

Table 4.1: *Marginal cdfs.*

Example 4.3.3. *Consider the class $\mathcal{G}_3^{2/5}(F_1, F_2, F_3)$, where F_1, F_2, and F_3 are the discrete cdfs provided in Table 4.1. Let $\boldsymbol{Y}$, $\boldsymbol{Y}'$, and $\boldsymbol{Y}'' \in \mathcal{G}_3^{2/5}(F_1, F_2, F_3)$ with copulas respectively defined by the Bernoulli rvs $\boldsymbol{X}$, $\boldsymbol{X}'$, and $\boldsymbol{X}''$ respectively distributed according f, f', and $f'' \in \mathcal{B}_3(\frac{2}{5})$ given in Table 4.2. Table 4.3 exhibits the values of pmfs of the discrete random*

$\boldsymbol{x}$	(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)
$f(\boldsymbol{x})$	0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	0	0	0
$f'(\boldsymbol{x})$	$\frac{1}{5}$	0	$\frac{2}{5}$	0	0	$\frac{2}{5}$	0	0
$f''(\boldsymbol{x})$	0	$\frac{2}{5}$	$\frac{1}{5}$	0	$\frac{1}{5}$	0	$\frac{1}{5}$	0

Table 4.2: *Bernoulli pmfs.*

*variables $S_{\boldsymbol{Y}} = \sum_{j=1}^3 Y_j$, $S_{\boldsymbol{Y}'} = \sum_{j=1}^3 Y_j'$, and $S_{\boldsymbol{Y}''} = \sum_{j=1}^3 Y_j''$; those values are computed using the pgf and the **FFT** algorithm as explained in Example 4.3.1. As one can see by looking at the support of the Bernoulli distributions in Table 4.2, although $\boldsymbol{X}$ and $\boldsymbol{X}''$ have the same distribution of the sum, $S_{\boldsymbol{Y}}$ and $S_{\boldsymbol{Y}''}$ are not equal in distribution. Moreover, it is easy to show that $S_{\boldsymbol{X}} = \sum_{j=1}^3 X_j \preceq_{cx} S_{\boldsymbol{X}'} = \sum_{j=1}^3 X_j'$, but $S_{\boldsymbol{Y}} \npreceq_{cx} S_{\boldsymbol{Y}'}$, since $Var(S_{\boldsymbol{Y}}) = 2.0633 \geq 1.8865 = Var(S_{\boldsymbol{Y}'})$.* $\square$

k	0	1	2	3	4	5	6	7
$f_{S_{\boldsymbol{Y}}}(k)$	0.0080	0.0338	0.0640	0.1328	0.2467	0.2592	0.2312	0.0242
$f_{S_{\boldsymbol{Y}'}}(k)$	0.0032	0.0249	0.0602	0.1556	0.2636	0.2569	0.2004	0.0352
$f_{S_{\boldsymbol{Y}''}}(k)$	0.0029	0.0214	0.0549	0.1588	0.2798	0.2521	0.1976	0.0324

Table 4.3: *Pmfs of $S_{\boldsymbol{Y}}$ and $S_{\boldsymbol{Y}'}$.*

Using the results from Corollary 4.3.12, we can analytically find the lower bounds of the convex risk measures introduced in Section 2.1 for exponential margins and discrete margins, respectively. Although the VaR_α is not convex we can find its bounds in $\mathcal{S}_d^p(F)$. The authors of [43] prove that the bounds of the VaR_α in a class of univariate distributions that has a convex polytope structure are reached at the extremal points. We consider a random vector $\boldsymbol{Y}$ with distribution in the class $\mathcal{G}_d^p(F)$, whose sum $S_{\boldsymbol{Y}} = Y_1 + \cdots + Y_d$ have cdf in $\mathcal{S}_d^p(F)$. Despite the number of extremal points of $\mathcal{G}_d^p(F)$ being n_d^p, as a consequence of Theorem 4.3.10 we can restrict our attention to the m_d^p extremal points of $\mathcal{S}_d^p(F)$. By

y	$\boldsymbol{r}_1^D$	$\boldsymbol{r}_2^D$	$\boldsymbol{r}_3^D$	$\boldsymbol{r}_4^D$	$\boldsymbol{r}_5^D$	$\boldsymbol{r}_6^D$	$\boldsymbol{r}_7^D$	$\boldsymbol{r}_8^D$	$\boldsymbol{r}_9^D$
0	$\frac{1}{6}$	$\frac{3}{8}$	$\frac{1}{2}$	0	0	0	0	0	0
1	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{5}{8}$	0	0	0
2	0	0	0	0	0	0	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{5}{6}$
3	$\frac{5}{6}$	0	0	$\frac{3}{4}$	0	0	$\frac{1}{2}$	0	0
4	0	$\frac{5}{8}$	0	0	$\frac{1}{2}$	0	0	$\frac{1}{4}$	0
5	0	0	$\frac{1}{2}$	0	0	$\frac{3}{8}$	0	0	$\frac{1}{6}$

Table 4.4: *Extremal pmfs of the class $\mathcal{D}_5(\frac{5}{2})$.*

computing the VaR_α of the random variable $S_{(\boldsymbol{Y},\boldsymbol{E}_k)}$, for every $k \in \{1, \dots, m_d^p\}$, it is possible to find maximum and minimum values that $\mathrm{VaR}_\alpha(Y_1 + \cdots + Y_d)$ can reach.

Remark 4.3.1. *In some cases, the minimum VaR_α is reached on the minimal Σ_{cx}-element of the polytope. In fact, from the proof of Theorem 3.A.4 in [93], it follows that, if $S \preceq_{cx} S'$, then there exists $\tilde{\alpha} \in (0,1)$ such that, $VaR_\alpha(S) \leq VaR_\alpha(S')$, for every $\alpha \in (\tilde{\alpha},1)$. Therefore, if $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ is a Σ_{cx}-smallest element of its Fréchet class, then there exists $\tilde{\alpha} \in (0,1)$ such that, $VaR_\alpha(\sum_{j=1}^d Y_j) \leq VaR_\alpha(\sum_{j=1}^d Y_j')$, for every $\alpha \in (\tilde{\alpha},1)$, for every random vector $\boldsymbol{Y}' = (Y_1', \dots, Y_d')$ of the same Fréchet class of $\boldsymbol{Y}$.*

We present below two numerical illustrations of our results.

Example 4.3.4. *Consider the case $d = 5$, $p = \frac{1}{2}$. Then, $dp = \frac{5}{2}$ and $j_1^\vee = 2$, $j_2^\wedge = 3$. The class $\mathcal{D}_5(\frac{5}{2})$ has $m_f^{1/2} = 9$ extremal points provided in Table 4.4. Let us compute value-at-risk, expected shortfall, and entropic risk measures of the extremal pmfs for $\alpha = 0.8$ and $\gamma = 0.1$. Results are reported in Table 4.5. The choice of α has been made to exhibit the case where the minimum VaR_α is not the Σ_{cx}-smallest element of the class. Table 4.6 reports instead the same risk measures evaluated on the corresponding FGM copulas. Notice that the bounds for the sums are at the extremal copulas corresponding to the upper Fréchet bound and to the Σ_{cx}-smallest Bernoulli pmfs for the convex measures, as proved in Corollary 4.3.12. The minimum VaR_α in $\mathcal{C}_5^{1/2}$ is reached at C_{D_7}, where D_7 is the discrete random variable with pmf r_7^D; while in the Bernoulli case it is for $S_{\boldsymbol{X}_9}$ with pmf r_9^D. This proves that the minimum VaR_α of the sum of the components of $\boldsymbol{U}$ in $\mathcal{C}_5^{1/2}$ is not inherited from the underlying Bernoulli pmf.* □

	r_1^D	r_2^D	r_3^D	r_4^D	r_5^D	r_6^D	r_7^D	r_8^D	r_9^D
$\mathrm{VaR}_{0.8}(S_{\boldsymbol{X}})$	3	4	$\underline{5}$	3	4	$\underline{5}$	3	4	$\boxed{2}$
$\mathrm{ES}_{0.8}(S_{\boldsymbol{X}})$	$\boxed{3}$	4	$\underline{5}$	$\boxed{3}$	4	$\underline{5}$	$\boxed{3}$	4	4.5
$\Psi_{0.1}(S_{\boldsymbol{X}})$	2.558	2.680	$\underline{2.809}$	2.536	2.612	2.693	$\boxed{2.513}$	2.539	2.567

Table 4.5: *Values of risk measures of the sum of the components of Bernoulli random vectors with the extremal probability mass functions of the class $\mathcal{D}_5(\frac{5}{2})$. The minimum values are squared and the maximum ones are underlined.*

	$C_{\boldsymbol{r}_1^D}$	$C_{\boldsymbol{r}_2^D}$	$C_{\boldsymbol{r}_3^D}$	$C_{\boldsymbol{r}_4^D}$	$C_{\boldsymbol{r}_5^D}$	$C_{\boldsymbol{r}_6^D}$	$C_{\boldsymbol{r}_7^D}$	$C_{\boldsymbol{r}_8^D}$	$C_{\boldsymbol{r}_9^D}$
$\mathrm{VaR}_{0.8}(S_{\boldsymbol{U}})$	3.031	3.328	$\underline{3.493}$	3.016	3.171	3.264	$\boxed{2.973}$	3.018	3.048
$\mathrm{ES}_{0.8}(S_{\boldsymbol{U}})$	3.463	3.735	$\underline{3.840}$	3.364	3.516	3.585	$\boxed{3.275}$	3.323	3.348
$\Psi_{0.1}(S_{\boldsymbol{U}})$	2.521	2.535	$\underline{2.549}$	2.518	2.526	2.535	$\boxed{2.515}$	2.518	2.521

Table 4.6: *Values of risk measures of the sum of components of uniform vectors with joint cdf defined by FGM copulas corresponding to the extremal probability mass functions of the class $\mathcal{D}_5(\frac{5}{2})$. The minimum values are squared and the maximum ones are underlined.*

The following Example 4.3.5 is the main example of application of our results. We consider a high dimensional portfolio of risks for six scenarios: three different GFGM(p) dependence structures for two Fréchet classes, with exponential and discrete margins, discussed in a theoretical setting in Examples 4.3.1 and 4.3.2.

Example 4.3.5. *Consider the classes $\mathcal{G}^p_{100}\left(Exp\left(\frac{1}{10}\right)\right)$ and $\mathcal{G}^p_{100}(F)$, where F is the discrete cdf whose pmf is given by*

$$f(y) = \begin{cases} \frac{4}{5}, & y = 0,\\ \frac{1}{5}\left[\left(\frac{y}{100}\right)^3 - \left(\frac{y-1}{100}\right)^3\right], & y \in \{1, \dots, 100\}. \end{cases}$$

We consider three different cases of GFGM(p) dependencies for each class, that is, we consider $\mathcal{G}^p_{100}(Exp(1/10))$ and $\mathcal{G}^p_{100}(F)$, where each case is associated to a common $p \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}\}$.

The bounds for the convex measures are reached at the distributions of the two classes $\mathcal{G}^p_{100}\left(Exp\left(\frac{1}{10}\right)\right)$ and $\mathcal{G}^p_{100}(F)$ corresponding to the minimal and maximal convex sums in $\mathcal{D}_{100}(100p)$, for $p = \frac{1}{3}, p = \frac{1}{2}$ and $p = \frac{2}{3}$. The minimal convex sum is the pmf in $\mathcal{D}_{100}(100p)$ with support on the pair $(33, 34)$ when $p = \frac{1}{3}$, with support on the point 50 when $p = \frac{1}{2}$, and support on the pair $(66, 67)$ when $p = \frac{2}{3}$. Table 4.7 provides the sharp bounds for the convex risk measures with exponential and discrete margins, respectively.

	$\mathcal{G}^p_{100}\left(\mathrm{Exp}\left(\frac{1}{10}\right)\right)$			$\mathcal{G}^p_{100}(F)$		
	$p = \frac{1}{3}$	$p = \frac{1}{2}$	$p = \frac{2}{3}$	$p = \frac{1}{3}$	$p = \frac{1}{2}$	$p = \frac{2}{3}$
$\min \mathrm{ES}_{0.95}$	1191.274	1189.272	1192.332	2152.595	2122.718	2019.207
$\max \mathrm{ES}_{0.95}$	1858.185	1702.844	1540.619	2858.955	3448.241	4440.057
$\min \Psi_{0.001}$	1003.921	1003.822	1003.924	1555.710	1551.957	1546.627
$\max \Psi_{0.001}$	1124.634	1125.051	1101.526	1888.303	2216.540	2843.312

Table 4.7: *Bounds of convex risk measures in the classes $\mathcal{G}^p_{100}\left(Exp\left(\frac{1}{10}\right)\right)$ and $\mathcal{G}^p_{100}(F)$.*

The VaR_α is bounded by its evaluations on the extremal pmfs of the two classes of distributions $\mathcal{S}^p_{100}\left(Exp\left(\frac{1}{10}\right)\right)$ and $\mathcal{S}^p_{100}(F)$. When $d = 100$, the number of extremal points m^p_d is lower than or equal to 2501, see Corollary 4.6 of [43], and we find bounds by enumeration. Table 4.8 provides the bounds for the VaR_α in the above-mentioned classes and also the analytical bounds for the whole Fréchet classes given in Equation (4) of [6]. We *mention that the minimum $VaR_{0.95}$ in the class $\mathcal{G}^p_{100}(F)$ is not reached at the distribution corresponding to the minimal convex pmf in $\mathcal{D}_{100}\left(\frac{200}{3}\right)$, whose VaR_α is 1961.*

	$Exp\left(\frac{1}{10}\right)$			F		
	$p=\frac{1}{3}$	$p=\frac{1}{2}$	$p=\frac{2}{3}$	$p=\frac{1}{3}$	$p=\frac{1}{2}$	$p=\frac{2}{3}$
LB $\mathcal{F}_{100}(G)$	842.330	842.330	842.330	1045.963	1045.963	1045.963
$\min \mathcal{G}^p_{100}(G)$	1149.729	1147.012	1150.223	2016	1994	1960
$\max \mathcal{G}^p_{100}(G)$	1791.328	1645.054	1488.231	2688	3258	4225
UB $\mathcal{F}_{100}(G)$	3995.732	3995.732	3995.732	9606.61	9606.61	9606.61

Table 4.8: *$VaR_{0.95}$ Bounds: $\mathcal{F}_{100}(G)$ is the Fréchet class with 100 identically distributed risks, $Y_j \sim G$, for every $j \in \{1,\dots,100\}$, and $\mathcal{G}^p_{100}(G)$, where $G = Exp\left(\frac{1}{10}\right)$ or $G = F$.*

We conclude this example by considering Pearson's correlation of the exchangeable Σ_{cx}-smallest element $\boldsymbol{Y}^e_{cx}$ in $\mathcal{G}^{1/3}_{100}\left(Exp\left(\frac{1}{10}\right)\right)$ and Pearson's correlation matrix of a vector $\boldsymbol{Y}_{cx}$ corresponding to the Bernoulli Σ_{cx}-smallest element provided by the authors of [45] in Theorem 5.2. Using (4.26), Pearson's correlation $\rho_P(Y_{j_1}, Y_{j_2})$ (denoted by $\rho_{j_1 j_2}$) is equal to $Cov(X_{j_1}, X_{j_1})$, for $1 \le j_1 < j_2 \le d$.

We therefore have to find the covariance of the exchangeable Σ_{cx}-smallest element and the covariance matrix of the Σ_{cx}-smallest element f_{cx} in $\mathcal{B}_{100}(\frac{1}{3})$, obtained following Theorem 5.2 in [45], and given by

$$f_{cx}(\boldsymbol{x}) = \begin{cases} \frac{1}{3} & \boldsymbol{x} = (\underbrace{1,\dots,1}_{33},0,\dots,0,0,\dots,0), \\ \frac{1}{3} & \boldsymbol{x} = (0,\dots,0,\underbrace{1,\dots,1}_{34},0,\dots,0), \\ \frac{1}{3} & \boldsymbol{x} = (0,\dots,0,0,\dots,0,\underbrace{1,\dots,1}_{33}). \end{cases}$$

The equi-correlation of the exchangeable Σ_{cx}-smallest element in $\mathcal{G}^{1/3}_{100}\left(Exp\left(\frac{1}{10}\right)\right)$ is $\rho_e = -0.0022$, that is the minimal correlation in the subclass of exchangeable distributions in $\mathcal{G}^{1/3}_{100}\left(Exp\left(\frac{1}{10}\right)\right)$. Let $A = \{1,\dots,33\}^2 \cup \{34,\dots,67\}^2 \cup \{68,\dots,100\}^2$. The entries of Pearson's correlation matrix of $\boldsymbol{Y}_{cx}$ are given by

$$\rho_P(Y_{j_1}, Y_{j_2}) = \begin{cases} 1 & j_1 = j_2, \\ \frac{2}{9} & j_1 \ne j_2, (j_1, j_2) \in A, \\ -\frac{1}{9} & j_1 \ne j_2, (j_1, j_2) \in \{1,\dots,100\}^2 \setminus A. \end{cases}$$

The mean ρ_m of Pearson's correlations of the random vector $\boldsymbol{Y}_{cx}$ is given by

$$\rho_m = \frac{2}{99 \cdot 100} \sum_{j_1=1}^{99} \sum_{j_2=j_1+1}^{100} \rho_P(Y_{j_1}, Y_{j_2}) = -0.0022. \tag{4.37}$$

From (4.37) we notice that $\rho_m = \rho_e$, the equi-correlation of the exchangeable vector $\boldsymbol{Y}^e_{cx}$. This result is a consequence of Corollary 5.2 in [45] and of the fact that Pearson's correlations in the class with the exponential margins is equal to the covariance of the corresponding Bernoulli pair. □

4.3.4 Applications and estimation procedures

The FGM family of copulas is used to build dependent models in different fields of application. To mention a few, we find applications in reliability, e.g., [92], survival analysis, e.g., [73]), as well as insurance. For example, in healthcare insurance context, the authors of [94] study whether individuals' private information on health risks affects their medical care utilization. [85] also uses a copula-based approach to examine the severity of driver injury in motor vehicle crashes in a developing country. For an application with health data, see e.g. [96].

In the bidimensional case, the simple analytical formulation and the consequent statistical tractability of FGM copulas are some of the reasons for their success. In dimensions higher than two, although the analytical formula of the FGM copula remains simple, the parameter space becomes more complex. As discussed in Section 4.3, the geometrical representation allows us to prove that the parameter space is a convex polytope, whose generators are the parameters of the extremal copulas. Indeed, the geometrical representation in (4.15) is a parametric representation of FGM (or GFGM) copulas in terms of the extremal copulas $C_{\boldsymbol{R}_k}$, for $k \in \{1, \ldots, n_d^{\boldsymbol{p}}\}$. For any $C \in \mathcal{C}_d^{\boldsymbol{p}}$ and for $\boldsymbol{u} \in [0,1]^d$, we have

$$C(\boldsymbol{u}) = C(\boldsymbol{u}; \boldsymbol{\lambda}) = \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k C_{\boldsymbol{R}_k}(\boldsymbol{u}), \quad \text{for } \lambda_1, \ldots, \lambda_{n_d^{\boldsymbol{p}}} \geq 0, \text{ and } \sum_{k=1}^{n_d^{\boldsymbol{p}}} \lambda_k = 1. \tag{4.38}$$

From (4.38), it follows that, in any dimension, the space of the parameters $\boldsymbol{\lambda}$ is the standard simplex. For the bivariate FGM copula, [96] reviews several estimation approaches. Although the investigation of estimation procedures is outside the scope of this work, we mention some directions that are of interest for our future research development. Given a random sample of size n, and observations denoted by $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, we can implement, beyond the traditional maximum likelihood estimation, the inference functions for margins (IFM) method (see [64]) or the semi-parametric pseudo maximum likelihood (PML) technique, as proposed by [51]. The IFM and PML methods are both two-stage approaches in which the marginal distributions are estimated separately from the dependence structure. They differ in the estimation of the marginals, which is done parametrically in the IFM method and non-parametrically in the PML technique. Denote by $\hat{F}_j$ the estimated marginal cdfs obtained by one of the two methods. The second step of both methods consists of maximizing the pseudo-log-likelihood function

$$L(\boldsymbol{\lambda}) = \sum_{h=1}^{n} \log\Big(c(\hat{F}_1(x_{h,1}), \ldots, \hat{F}_d(x_{h,d}); \boldsymbol{\lambda})\Big), \tag{4.39}$$

under the constraints $\lambda_1 \geq 0, \ldots, \lambda_{n_d^{\boldsymbol{p}}} \geq 0$ and $\lambda_1 + \cdots + \lambda_{n_d^{\boldsymbol{p}}} = 1$, where $c(\boldsymbol{u}; \lambda)$ is the density of $C(\boldsymbol{u}; \lambda)$. As in [96], we can consider minimum-distance methods and perform goodness-of-fit tests. The geometrical representation of the FGM copulas facilitates the estimation and, consequently, their use in dimensions higher than two. However, some computational issues remain in high dimensions, where finding all the generators is still an open issue. In practice, using standard symbolic computation and polyhedral geometry software, exact enumeration becomes strictly infeasible as soon as the dimension reaches $d \geq 7$.

4.3.5 GFGM copulas with non-equal parameters

We conclude with one example in low dimension of the general class $\mathcal{G}_d(\boldsymbol{p})$, with $\boldsymbol{p} = (p_1, \ldots, p_d)$, and we leave its theoretical investigation to further research. We find the risk measures' sharp bounds for the sum $S_{\boldsymbol{Y}} = Y_1+Y_2+Y_3$, where Y_j have discrete distributions and $\boldsymbol{Y}$ has a GFGM copula with vector parameter $\boldsymbol{p}$. In fact, for $d = 3$ we are able to find the extremal points of $\mathcal{B}_d(\boldsymbol{p})$, to construct the corresponding copulas, and to find the generators of the convex polytope $\mathcal{G}_d^{\boldsymbol{p}}(F_1, \ldots, F_d)$. We evaluate the risk measures on the extremal point, and we find sharp bounds by enumeration. Furthermore, we find the expected allocation and the expected contribution of Y_j for each risk $\boldsymbol{Y}$ with extremal pmf, following [9].

Example 4.3.6. *We consider the class $\mathcal{B}_d(\boldsymbol{p})$ with $d = 3$ and $\boldsymbol{p} = (\frac{1}{2}, \frac{1}{3}, \frac{2}{3})$. In this case, there are $n_d^{\boldsymbol{p}} = 12$ extremal points that can be found by using 4ti2, [98]. The extremal points $\boldsymbol{r}_k$, $k \in \{1, \ldots, 12\}$, of the class $\mathcal{B}_3(\boldsymbol{p})$ are reported in Table 4.9. Note that there are three extremal pmfs whose sum is minimal under the convex order: r_1, r_2, and r_4. These three vectors have two couples of Bernoulli random variables with minimal covariance, and the remaining pair — (X_1, X_2) for r_1, (X_1, X_3) for r_2, and (X_2, X_3) for r_4 — has maximal covariance.*

$\boldsymbol{x}$	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}
(0,0,0)	0	0	0	0	0	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$
(1,0,0)	0	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$	0	$\frac{1}{6}$	$\frac{1}{6}$	0	0	0
(0,1,0)	0	$\frac{1}{3}$	0	0	0	$\frac{1}{12}$	$\frac{1}{6}$	0	0	0	0	0
(1,1,0)	$\frac{1}{3}$	0	$\frac{1}{6}$	0	0	0	0	0	0	0	0	$\frac{1}{12}$
(0,0,1)	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{5}{12}$	0	0	$\frac{1}{3}$	0	$\frac{1}{6}$	0
(1,0,1)	$\frac{1}{6}$	$\frac{1}{2}$	0	$\frac{1}{6}$	0	0	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{5}{12}$
(0,1,1)	0	0	0	$\frac{1}{3}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$	0	$\frac{1}{6}$	0	$\frac{1}{4}$
(1,1,1)	0	0	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{4}$	0	0	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	0

Table 4.9: *Values of the extremal pmfs r_k, $k \in \{1, \ldots, 12\}$, of the class $\mathcal{B}_3(\frac{1}{2}, \frac{1}{3}, \frac{2}{3})$ for every $\boldsymbol{x} \in \mathcal{X}_3$.*

We now consider the class $\mathcal{G}_3^{\boldsymbol{p}}(F_1, F_2, F_3)$, where F_j, $j \in \{1,2,3\}$, is the cdf whose pmf f_j is defined by

$$f_j(y) = \begin{cases} 1 - a_j & y = 0, \\ a_i[(\frac{y}{n})^{c_j} - (\frac{y-1}{n})^{c_j}] & y \in \{1, \ldots, n\}, \end{cases}$$

and we choose $n = 1000$, $a_1 = 0.2$, $a_2 = 0.1$, $a_3 = 0.3$ and $c_1 = 3$, $c_2 = 4$, $c_3 = 2$. Also, we have $E[Y_1] = 150.09995$, $E[Y_2] = 80.04997$, $E[Y_3] = 200.14995$ and $E[S] = 430.29987$.

The lower and upper bounds of convex risk measures are reached at the random vectors corresponding to the extremal pmfs r_1 and r_{11}, respectively. In Table 4.10, we provide Pearson's correlation coefficients $\rho(Y_1, Y_2)$, $\rho(Y_1, Y_3)$, and $\rho(Y_2, Y_3)$ for all of the twelve extremal dependence structures. Notice that the correlation matrix of $\boldsymbol{Y}_1$ also has negative entries, i.e., $(Y_{1,1}, Y_{1,3})$ and $(Y_{1,2}, Y_{1,3})$ are negatively correlated, while $(Y_{1,1}, Y_{1,2})$ has the same correlation as $(Y_{11,1}, Y_{11,2})$. This last equality follows observing that the correlation

$\rho_1(1,2)$ between $(r_{1,1}, r_{1,2})$ and the correlation $\rho_{11}(1,2)$ between $(r_{11,1}, r_{11,2})$ are equal, in fact

$$\rho_1(1,2) = \frac{r_1((1,1,0)) + r_1((1,1,1)) - \frac{1}{2}\frac{1}{3}}{\sqrt{\frac{1}{2}(1-\frac{1}{2})\frac{1}{3}(1-\frac{1}{3})}} = 0.0605$$

and

$$\rho_{11}(1,2) = \frac{r_{11}((1,1,0)) + r_{11}((1,1,1)) - \frac{1}{2}\frac{1}{3}}{\sqrt{\frac{1}{2}(1-\frac{1}{2})\frac{1}{3}(1-\frac{1}{3})}} = 0.0605$$

coincide since $r_1((1,1,0)) + r_1((1,1,1)) = r_{11}((1,1,0)) + r_{11}((1,1,1)) = \frac{1}{3}$.

	$\boldsymbol{r}_1$	$\boldsymbol{r}_2$	$\boldsymbol{r}_3$	$\boldsymbol{r}_4$	$\boldsymbol{r}_5$	$\boldsymbol{r}_6$
$\rho(Y_1, Y_2)$	0.0605	-0.0605	0.0605	-0.0605	0.0000	0.0302
$\rho(Y_1, Y_3)$	-0.1610	0.1610	-0.1610	-0.1610	-0.1610	-0.0805
$\rho(Y_2, Y_3)$	-0.1229	-0.1229	-0.0307	0.0615	0.0615	0.0154
	$\boldsymbol{r}_7$	$\boldsymbol{r}_8$	$\boldsymbol{r}_9$	$\boldsymbol{r}_{10}$	$\boldsymbol{r}_{11}$	$\boldsymbol{r}_{12}$
$\rho(Y_1, Y_2)$	-0.0605	-0.0605	0.0605	0.0000	0.0605	-0.0302
$\rho(Y_1, Y_3)$	0.1610	0.0000	0.0000	0.1610	0.1610	0.0805
$\rho(Y_2, Y_3)$	-0.0307	0.0615	0.0615	0.0615	0.0615	0.0154

Table 4.10: Pearson's coefficients of $\boldsymbol{X}$.

	$\boldsymbol{r}_1$	$\boldsymbol{r}_2$	$\boldsymbol{r}_3$	$\boldsymbol{r}_4$	$\boldsymbol{r}_5$	$\boldsymbol{r}_6$
$VaR_{0.95}(S)$	$\boxed{1219.00}$	1532.00	1360.00	1342.00	1403.00	1479.00
$ES_{0.95}(S)$	$\boxed{1590.08}$	1733.70	1665.46	1641.07	1683.14	1724.32
$\Psi_{0.001}(S)$	$\boxed{555.98}$	587.74	566.80	563.46	570.51	580.07
$Std(S)$	$\boxed{473.23}$	521.70	488.85	485.22	494.70	508.47
	$\boldsymbol{r}_7$	$\boldsymbol{r}_8$	$\boldsymbol{r}_9$	$\boldsymbol{r}_{10}$	$\boldsymbol{r}_{11}$	$\boldsymbol{r}_{12}$
$VaR_{0.95}(S)$	1561.00	1493.00	1567.00	1618.00	$\underline{1643.00}$	1535.00
$ES_{0.95}(S)$	1802.17	1771.05	1824.07	1888.55	$\underline{1906.84}$	1818.89
$\Psi_{0.001}(S)$	602.12	590.22	603.90	622.97	$\underline{629.61}$	601.55
$Std(S)$	535.91	518.49	536.10	558.13	$\underline{566.39}$	531.66

Table 4.11: *Values of risk measures of the sum of the components of vectors with joint cdf defined by the GFGM copulas corresponding to the extremal probability mass functions ($\boldsymbol{r}_j$, $j \in \{1, \ldots, 12\}$) of the class $\mathcal{B}_3(\frac{1}{2}, \frac{1}{3}, \frac{2}{3})$. The minimum values are squared and the maximum ones are underlined.*

We conclude this example by finding the contribution of risk Y_j, $j \in \{1,2,3\}$, to the standard deviation of the sum $S_{\boldsymbol{Y}} = Y_1 + Y_2 + Y_3$, to the VaR_α and to the ES_α, for all the extremal dependence structures. We recall the definitions of expected allocation and of *expected contribution of the risk Y_j to a total outcome $S = y$, for $y \in \{0,1,\dots,3000\}$. The expected allocation of each risk Y_j in Definition 1.1 of [9] is given by*

$$\mathbb{E}[Y_j \mathbf{1}\{S = y\}], \quad y \in \mathbb{N}_+,$$

where $\mathbf{1}$ is the indicator function, such that $\mathbf{1}\{A\} = 1$, if A is verified, and $\mathbf{1}\{A\} = 0$, otherwise. The expected contribution of each risk Y_j, $j \in \{1,2,3\}$, is provided in Equation (6) of [9], and is defined by

$$\mathbb{E}[Y_j | S = y] = \frac{\mathbb{E}[Y_j \mathbf{1}\{S = y\}]}{\mathbb{P}(S = y)}, \quad y \in \mathbb{N}_+,$$

assuming that $\mathbb{P}(S = y) > 0$. The expected contribution of Y_j to the VaR_α is given by $\mathbb{E}[Y_j | S = VaR_\alpha(S)]$. We now recall the expression for the contribution to the ES_α based on the Euler-based allocation rule provided in [97]:

$$CES_\alpha(Y_j, S) = \frac{\mathbb{E}[Y_j] - \mathbb{E}[Y_j \mathbf{1}\{S \leq VaR_\alpha(S)\}] + \beta_S \mathbb{E}[Y_j \mathbf{1}\{S = VaR_\alpha(S)\}]}{1 - \alpha},$$

where

$$\beta_S = \begin{cases} \frac{\mathbb{P}(S \leq VaR_\alpha(S)) - \alpha}{\mathbb{P}(S = VaR_\alpha(S))} & \text{if } \mathbb{P}(S = VaR_\alpha(S)) > 0, \\ 0 & \text{if } \mathbb{P}(S = VaR_\alpha(S)) = 0, \end{cases}$$

and

$$\mathbb{E}[Y_j \mathbf{1}\{S \leq k\}] = \sum_{y=1}^{k} \mathbb{E}[Y_j \mathbf{1}\{S = y\}], \quad k \in \mathbb{N}_+.$$

Finally, the contribution of Y_j to the standard deviation of $S_{\mathbf{Y}}$ based on Euler's rule is given by

$$CStd(Y_j, S_{\mathbf{Y}}) = \frac{Cov(Y_j, S_{\mathbf{Y}})}{\sqrt{Var(S_{\mathbf{Y}})}} = \frac{Var(Y_j) + \sum_{j' \neq j} Cov(Y_j, Y_{j'})}{\sqrt{Var(S_{\mathbf{Y}})}}, \quad j \in \{1, 2, 3\}. \tag{4.40}$$

Figure 4.1 reports the contributions to the $VaR_{0.95}$, the $ES_{0.95}$ and the standard deviation of S. □

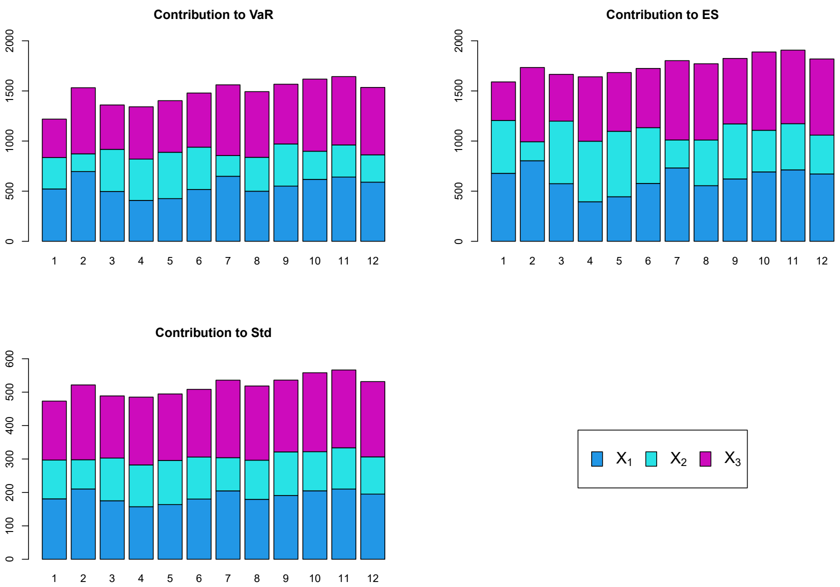

Figure 4.1: *Contributions of Y_j, $j \in \{1,2,3\}$ to the $VaR_{0.95}(S)$ (top-left), to the $ES_{0.95}(S)$ (top-right), and to the standard deviation of S (bottom-left), based on Euler's rule.*

Chapter 5

Conclusions

This thesis has investigated algebraic and geometrical representations of multidimensional Bernoulli distributions, establishing their role as effective frameworks for characterizing and solving open problems within the theory of negative dependence. By connecting algebraic geometry, commutative algebra, and applied probability theory, this work provides a unified framework to address open problems of negative dependence, specifically addressing the difficulties that arise when negative dependent structures are extended to high-dimensional cases.

Within this context, among the various generalizations of countermonotonicity to dimensions $d > 2$, we have focused on the notion of Σ-countermonotonicity. In the Bernoulli setting, we have characterized this property by establishing its equivalence with the minimality of aggregate risk under the convex order. This result provides a solid characterization of extremal negative dependence, proving that the class of Σ-countermonotonic Bernoulli distributions forms a convex polytope. Furthermore, in the specific case of symmetric marginals where all means are equal to $1/2$, a complete characterization of the corresponding class of distributions has been achieved. A substantial part of our analysis has also been devoted to the integration of the strongly Rayleigh property, defined via the geometry of zeros of probability generating functions, into the framework of extremal negative dependence. We proved that every Bernoulli Fréchet class contains at least one distribution that satisfies both the Σ-countermonotonicity and the strongly Rayleigh property, identifying this specific structure through entropy maximization within the Σ-countermonotonic polytope. We also investigated the Conway–Maxwell-binomial family, proving its capability to span the full spectrum of dependence, and we established parametric values that ensure the strongly Rayleigh property, consolidating its role as a flexible parametric family. Finally, these discrete properties have been transferred to the continuous framework of copula theory through specific stochastic representations. We constructed and analyzed negative properties of extremal mixture, Farlie–Gumbel–Morgenstern (FGM), and generalized FGM (GFGM) copulas. In particular, focusing on the family of GFGM copulas under the assumption of identically distributed risks, we studied the convex order of aggregate risk. This provides explicit tools to determine lower bounds for standard risk measures such as Value-at-Risk, Expected Shortfall, and entropic risk measures.

The findings of this thesis open several directions for future research, both from a theoretical and an applied perspective. A first open direction concerns the relationship between Σ-countermonotonicity and the strongly Rayleigh property. While we relied on entropy maximization to select a joint distribution that satisfies both conditions, a clear probabilistic interpretation of the strongly Rayleigh property is still missing. Furthermore, for those Fréchet classes where more than one such distribution exists, it would be valuable to analyze how to compare them under specific dependence orders.

Representing Fréchet classes through convex polytopes and polynomial ideals provides a key mechanism for isolating optimal negative dependence structures. However, these techniques become highly demanding in high-dimensional settings. Investigating their computational complexity in greater depth represents an important next step to understand how these tools scale as d grows. Specifically, it remains to be explored whether and how the algebraic approach can offer a computationally simpler or more efficient alternative compared to the purely geometric analysis.

Concerning the continuous framework and actuarial applications, the current analysis generally assumes the parameters of the proposed copula families (in particular of GFGM copulas) to be fixed and known. A natural next step is to incorporate parameter uncertainty directly into the evaluation of the risk bounds, allowing the copula parameters to vary within a specified uncertainty set. By extending the sharp bounds to account for this parameter uncertainty, these results can be integrated into the broader framework of model ambiguity (see, e.g., [42] and references therein). This would yield more conservative, uncertainty-aware risk management tools where the precise dependence structure is only partially known.

Finally, a natural next step is the practical implementation of these theoretical results to real-world credit or insurance portfolios. In a static setting, applying our negative dependence characterizations to empirical data would help optimize capital allocation and evaluate precise risk bounds for defaults or claims. Moreover, a natural extension would be to investigate how these negative dependence concepts can be adapted to dynamic frameworks to capture the temporal evolution of risk propagation.

Appendix A

Exponential families

An exponential family is a collection $\mathcal{G}$ of probability densities defined with respect to a σ-finite measure m on a measurable space $(\mathcal{X}, \mathcal{A})$ of the form

$$\mathcal{G} = \left\{ \frac{\mathrm{d}G_{\boldsymbol{\theta}}}{\mathrm{d}m}(\boldsymbol{x}) = e^{\boldsymbol{\theta}\cdot\boldsymbol{t}(\boldsymbol{x})-\psi(\boldsymbol{\theta})} h(\boldsymbol{x}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^n, \boldsymbol{x} \in \mathcal{X} \right\}, \tag{A.1}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, $\boldsymbol{t}(\boldsymbol{x}) = (t_1(\boldsymbol{x}), \ldots, t_n(\boldsymbol{x}))$ with $t_1, \ldots, t_n$ real-valued measurable functions, $h\colon \mathcal{X} \to \mathbb{R}$ is a non-negative measurable function, and ψ is a real-valued function. The parameters θ_j, for $j \in \{1, \ldots, n\}$, are called natural parameters, while $t_j(\boldsymbol{x})$, for $j \in \{1, \ldots, n\}$, are the sufficient statistics. Since $G_{\boldsymbol{\theta}}$ are probability measures, it follows that $\psi(\boldsymbol{\theta})$ assumes the role of normalizing function. The natural parameter space Θ is defined as the set of all $\boldsymbol{\theta} \in \mathbb{R}^n$ such that

$$e^{\psi(\boldsymbol{\theta})} = \int_{\mathcal{X}} e^{\boldsymbol{\theta}\cdot\boldsymbol{t}(\boldsymbol{x})} h(\boldsymbol{x}) m(\mathrm{d}\boldsymbol{x}) < +\infty,$$

and it can be shown that Θ is a convex subset of $\mathbb{R}^n$. An exponential family is said to be regular if Θ is open. If the functions $t_1, \ldots, t_n$ are affinely independent, i.e., if given some constants $c_0, c_1, \ldots, c_n$ such that $c_0 + c_1 t_1 + \cdots + c_n t_n = 0$ m-almost everywhere then $c_0 = c_1 = \cdots = c_n = 0$, we say that the representation of the exponential family in (A.1) is minimal. When $\mathcal{G}$ is a minimally represented and regular exponential family, the normalizing function $\psi(\boldsymbol{\theta})$ is infinitely differentiable on the interior of Θ, and its gradient and Hessian matrix coincide with the mean vector $\boldsymbol{\mu}(\boldsymbol{\theta}) \coloneqq \mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{t}(\boldsymbol{X})]$ and the covariance matrix $V(\boldsymbol{\theta}) \coloneqq \mathrm{Cov}_{\boldsymbol{\theta}}[\boldsymbol{t}(\boldsymbol{X})]$ of the sufficient statistics,

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \nabla\psi(\boldsymbol{\theta}) = \left(\frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_1}, \ldots, \frac{\partial\psi(\boldsymbol{\theta})}{\partial\theta_n} \right),$$

and

$$V(\boldsymbol{\theta}) = \nabla^2\psi(\boldsymbol{\theta}) = \begin{pmatrix} & \vdots & \\ \cdots & \frac{\partial^2\psi(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} & \cdots \\ & \vdots & \end{pmatrix}$$

The following proposition is a classical result; see, e.g., [20].

Proposition A.0.1. *If $\mathcal{G}$ is a minimally represented and regular exponential family, then the map $\boldsymbol{\theta} \to \boldsymbol{\mu}(\boldsymbol{\theta})$ is a homeomorphism (i.e., continuous, one-to-one, and onto) between Θ and the interior of the set of attainable means*

$$\mathcal{M} = \{\boldsymbol{\mu}(\boldsymbol{\theta})\colon \boldsymbol{\theta} \in \Theta\} \subseteq \mathbb{R}^n.$$

For a regular exponential family, an important result by [2] states that for any partition of $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$, the distribution $f \in \mathcal{G}$ is uniquely determined by the mixed vector $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\mu}^{(2)})$ or by $(\boldsymbol{\mu}^{(1)}, \boldsymbol{\theta}^{(2)})$, where $(\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)})$ is the partition of $\boldsymbol{\mu}$ that corresponds to that of $\boldsymbol{\theta}$.

Proposition A.0.2. *Let $\mathcal{G}$ be a regular exponential family with natural parameters $\boldsymbol{\theta} \in \Theta$ and mean-value parameterization $\boldsymbol{\mu} \in \mathcal{M}$. Consider a partition $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$ and the corresponding partition $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)})$. The map*

$$\boldsymbol{\theta} \mapsto (\boldsymbol{\mu}^{(1)}, \boldsymbol{\theta}^{(2)})$$

is a homeomorphism.

Thus, $(\boldsymbol{\mu}^{(1)}, \boldsymbol{\theta}^{(2)})$ defines a proper parameterization of a regular exponential family, called mixed parameterization.

Appendix B

The strongly Rayleigh property for $d = 3$

By (2.10), the condition to check if a 3-dimensional Bernoulli random vector $\boldsymbol{X} = (X_1, X_2, X_3)$ is strongly Rayleigh simplifies. Indeed, we have that $\boldsymbol{X}$ is strongly Rayleigh if and only if $\Delta_{jk}\mathcal{P}(\boldsymbol{y}) \geq 0$, for all $\boldsymbol{y} = (y_1, y_2, y_3) \in \mathbb{R}^3$ and for every $j, k \in \{1,2,3\}$, where

$$\Delta_{jk}\mathcal{P}(\boldsymbol{y}) = (\partial_j \mathcal{P}|_{z_k=0}(\boldsymbol{y}))(\partial_k \mathcal{P}|_{z_j=0}(\boldsymbol{y})) - (\partial_j\partial_k \mathcal{P}(\boldsymbol{y}))(\mathcal{P}|_{z_j=z_k=0}(\boldsymbol{y})) \geq 0. \tag{B.1}$$

If $j = k$, then $\Delta_{jj} = (\partial_j \mathcal{P}(\boldsymbol{y}))^2$ that is clearly non-negative for all $\boldsymbol{y} \in \mathbb{R}^3$. Consider $j = 1$ and $k = 2$. From (B.1), it follows that $\Delta_{12}\mathcal{P}(\boldsymbol{y})$ is a univariate polynomial in y_3 of degree 2. After standard computations, we find

$$\Delta_{12}\mathcal{P}(\boldsymbol{y}) = Ay_3^2 + By_3 + C,$$

where

$$\begin{aligned}
A &= f_{(1,0,1)}f_{(0,1,1)} - f_{(0,0,1)}f_{(1,1,1)},\\
B &= f_{(1,0,0)}f_{(0,1,1)} + f_{(0,1,0)}f_{(1,0,1)} - f_{(0,0,0)}f_{(1,1,1)} - f_{(1,1,0)}f_{(0,0,1)},\\
C &= f_{(1,0,0)}f_{(0,1,0)} - f_{(0,0,0)}f_{(1,1,0)}.
\end{aligned}$$

Therefore, $\Delta_{12}\mathcal{P}(\boldsymbol{y})$ is non-negative for every $\boldsymbol{y} \in \mathbb{R}^3$ if

$$A \geq 0, \tag{B.2}$$

and

$$D = B^2 - 4\,A\,C \leq 0. \tag{B.3}$$

If we assume that f satisfies the NLC, then (B.2) is verified. It remains to check if (B.3) holds. We can thus state the following proposition.

Proposition B.0.1. *Let $f \in \mathcal{B}_3$ be a Bernoulli pmf that satisfies the NLC. Let $\boldsymbol{f} = (f_1, \ldots, f_8) \in \mathbb{R}_+^{2^d}$ be the corresponding vector that contains the values $f(\boldsymbol{x})$ with $\mathcal{X}_3$. Define*

$$\begin{aligned}
D_1 &= (f_2f_7 + f_3f_6 - f_4f_5 - f_1f_8)^2 - 4(f_6f_7 - f_5f_8)(f_2f_3 - f_1f_4),\\
D_2 &= (f_2f_7 + f_4f_5 - f_3f_6 - f_1f_8)^2 - 4(f_4f_7 - f_3f_8)(f_2f_5 - f_1f_6),\\
D_3 &= (f_4f_5 + f_6f_3 - f_2f_7 - f_1f_8)^2 - 4(f_4f_6 - f_2f_8)(f_3f_5 - f_1f_7).
\end{aligned}$$

If $D_j \leq 0$ for every $j \in \{1, 2, 3\}$, the pmf f satisfies the strongly Rayleigh property.

Bibliography

[1] Ismihan Bairamov, Samuel Kotz, and Muhammet Bekci. "New generalized Farlie–Gumbel–Morgenstern distributions and concomitants of order statistics". In: *Journal of Applied Statistics* 28.5 (2001), pp. 521–536.

[2] Ole Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory.* John Wiley & Sons, 2014.

[3] Nicole Bäuerle and Alfred Müller. "Stochastic orders and risk measures: Consistency and bounds". In: *Insurance: Mathematics and Economics* 38.1 (2006), pp. 132–148.

[4] Hakim Bekrizadeh, Gholam Ali Parham, and Mohamad Reza Zadkarmi. "The new generalization of Farlie–Gumbel–Morgenstern copulas". In: *Applied Mathematical Sciences* 6.71 (2012), pp. 3527–3533.

[5] Carole Bernard, Xiao Jiang, and Ruodu Wang. "Risk aggregation with dependence uncertainty". In: *Insurance: Mathematics and Economics* 54 (2014), pp. 93–108.

[6] Carole Bernard, Ludger Rüschendorf, Steven Vanduffel, and Jing Yao. "How robust is the value-at-risk of credit risk portfolios?" In: *The European Journal of Finance* 23.6 (2017), pp. 507–534.

[7] Christopher Blier-Wong, Hélène Cossette, Sébastien Legros, and Etienne Marceau. "A new method to construct high-dimensional copulas with Bernoulli and Coxian-2 distributions". In: *Journal of Multivariate Analysis* 201 (2024), p. 105261.

[8] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. "Exchangeable FGM copulas". In: *Advances in Applied Probability* 56.1 (2024), pp. 205–234.

[9] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. "Generating function method for the efficient computation of expected allocations". In: *arXiv preprint arXiv:2207.02654* (2022).

[10] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. "Risk aggregation with FGM copulas". In: *Insurance: Mathematics and Economics* 111 (2023), pp. 102–120.

[11] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. "Stochastic Representation of FGM Copulas Using Multivariate Bernoulli Random Variables". In: *Computational Statistics & Data Analysis* 173 (2022).

[12] Christopher Blier-Wong, Hélène Cossette, and Etienne Marceau. "Stochastic representation of FGM copulas using multivariate Bernoulli random variables". In: *Computational Statistics & Data Analysis* 173 (2022), p. 107506.

[13] Julius Borcea and Petter Brändén. “Multivariate Pólya–Schur classification problems in the Weyl algebra”. In: *Proceedings of the London Mathematical Society* 101.1 (2010), pp. 73–104.

[14] Julius Borcea and Petter Brändén. “The Lee-Yang and Pólya-Schur programs. I. linear operators preserving stability”. In: *Inventiones mathematicae* 177.3 (2009), pp. 541–569.

[15] Julius Borcea, Petter Brändén, and Thomas Liggett. “Negative dependence and the geometry of polynomials”. In: *Journal of the American Mathematical Society* 22.2 (2009), pp. 521–567.

[16] Endre Boros and András Prékopa. “Closed form two-sided bounds for probabilities that at least r and exactly r out of n events occur”. In: *Mathematics of Operations Research* 14.2 (1989), pp. 317–342.

[17] Michael V. Boutsikas and Markos V. Koutras. “A bound for the distribution of the sum of discrete associated or negatively associated random variables”. In: *The Annals of Applied Probability* 10.4 (2000), pp. 1137–1150.

[18] Petter Brändén. “Polynomials with the half-plane property and matroid theory”. In: *Advances in Mathematics* 216.1 (2007), pp. 302–320.

[19] Petter Brändén and Johan Jonasson. “Negative dependence in sampling”. In: *Scandinavian Journal of Statistics* 39.4 (2012), pp. 830–838.

[20] Lawrence D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Institute of Mathematical Statistics, 1986.

[21] Stamatis Cambanis. “Some properties and generalizations of multivariate Eyraud-Gumbel-Morgenstern distributions”. In: *Journal of Multivariate Analysis* 7.4 (1977), pp. 551–559.

[22] Sean X Chen. “General properties and estimation of conditional Bernoulli models”. In: *Journal of Multivariate Analysis* 74.1 (2000), pp. 69–87.

[23] Xiang-Hui Chen, Arthur P Dempster, and Jun S Liu. “Weighted finite population sampling to maximize entropy”. In: *Biometrika* 81.3 (1994), pp. 457–469.

[24] Xianghui Chen. *Weighted Sampling, Statistical Applications and Generalizations*. Harvard University, 1995.

[25] Ka Chun Cheung and Ambrose Lo. “Characterizations of counter-monotonicity and upper comonotonicity by (tail) convex order”. In: *Insurance: Mathematics and Economics* 53.2 (2013), pp. 334–342.

[26] Ka Chun Cheung and Ambrose Lo. “Characterizing mutual exclusivity as the strongest negative multivariate dependence structure”. In: *Insurance: Mathematics and Economics* 55 (2014), pp. 180–190.

[27] Tasos C. Christofides and Eutichia Vaggelatou. “A connection between supermodular ordering and positive/negative association”. In: *Journal of Multivariate Analysis* 88.1 (2004), pp. 138–151.

[28] James W. Cooley and John W. Tukey. “An Algorithm for the Machine Calculation of Complex Fourier Series”. In: *Mathematics of Computation* 19.90 (1965), pp. 297–301.

[29] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, 2009.

[30] Hélène Cossette, Marie-Pier Côté, Etienne Marceau, and Khouzeima Moutanabbir. "Multivariate distribution defined with Farlie–Gumbel–Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation". In: *Insurance: Mathematics and Economics* 52.3 (2013), pp. 560–572.

[31] Bin Dai, Shilin Ding, and Grace Wahba. "Multivariate Bernoulli distribution". In: *Bernoulli* 19.4 (2013), pp. 1465–1483.

[32] G. Dall'Aglio, S. Kotz, and G. Salinetti. *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*. Vol. 67. Springer Science & Business Media, 2012.

[33] G Dall'Aglio. "Fréchet classes and compatibility of distribution functions". In: *Symposia Mathematica*. Vol. 9. 1972, pp. 131–150.

[34] Michel Denuit, Christian Genest, and Étienne Marceau. "Stochastic bounds on sums of dependent risks". In: *Insurance: Mathematics and Economics* 25.1 (1999), pp. 85–104.

[35] Jan Dhaene and Michel Denuit. "The safest dependence structure among risks". In: *Insurance: Mathematics and Economics* 25.1 (1999), pp. 11–21.

[36] Diego Dominici, Sarah Jane Johnston, and Kerstin Jordaan. "Real zeros of 2F1 hypergeometric polynomials". In: *Journal of Computational and Applied Mathematics* 247 (2013), pp. 152–161.

[37] Hugo Duminil-Copin, Dmitry Ioffe, and Yvan Velenik. "A quantitative Burton–Keane estimate under strong FKG condition". In: *The Annals of Probability* 44.5 (2016), pp. 3335–3356.

[38] Sharmila Duppala, George Z Li, Juan Luque, Aravind Srinivasan, and Renata Valieva. "Concentration of Submodular Functions and Read-k Families Under Negative Dependence: S. Duppala, GZ Li, J. Luque, A. Srinivasan, R. Valieva". In: *Algorithmica* 88.2 (2026), p. 21.

[39] Fabrizio Durante and Carlo Sempi. *Principles of Copula Theory*. CRC press, 2015.

[40] P. Embrechts, R. Grübel, and S. M. Pitts. "Some Applications of the Fast Fourier Transform Algorithm in Insurance Mathematics This Paper Is Dedicated to Professor W. S. Jewell on the Occasion of His 60th Birthday". In: *Statistica Neerlandica* 47.1 (1993), pp. 59–75.

[41] Paul Embrechts, Giovanni Puccetti, and Ludger Rüschendorf. "Model uncertainty and VaR aggregation". In: *Journal of Banking & Finance* 37.8 (2013), pp. 2750–2764.

[42] Tolulope Fadina, Yang Liu, and Ruodu Wang. "A framework for measures of risk under uncertainty". In: *Finance and Stochastics* 28.2 (2024), pp. 363–390.

[43] Roberto Fontana, Elisa Luciano, and Patrizia Semeraro. "Model risk in credit risk". In: *Mathematical Finance* 31.1 (2021), pp. 176–202.

[44] Roberto Fontana and Patrizia Semeraro. "Exchangeable Bernoulli distributions: High dimensional simulation, estimation, and testing". In: *Journal of Statistical Planning and Inference* 225 (2023), pp. 52–70.

[45] Roberto Fontana and Patrizia Semeraro. "High dimensional Bernoulli distributions: Algebraic representation and applications". In: *Bernoulli* 30.1 (2024), pp. 825–850.

[46] Roberto Fontana and Patrizia Semeraro. "Representation of multivariate Bernoulli distributions with a given set of specified moments". In: *Journal of Multivariate Analysis* 168 (2018), pp. 290–303.

[47] Roberto Fontana and Patrizia Semeraro. "The Bernoulli structure of discrete distributions". In: *Electronic Communications in Probability* 30 (2025), pp. 1–13.

[48] Cees M. Fortuin, Pieter W. Kasteleyn, and Jean Ginibre. "Correlation inequalities on some partially ordered sets". In: *Communications in Mathematical Physics* 22 (1971), pp. 89–103.

[49] Esther Frostig. "Comparison of portfolios which depend on multivariate Bernoulli random variables with fixed marginals". In: *Insurance: Mathematics and Economics* 29.3 (2001), pp. 319–331.

[50] Norbert Gaffke and Ludger Rüschendorf. "On a class of extremal problems in statistics". In: *Mathematische Operationsforschung und Statistik. Series Optimization* 12.1 (1981), pp. 123–135.

[51] Christian Genest, Kilani Ghoudi, and Louis-Paul Rivest. "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions". In: *Biometrika* 82.3 (1995), pp. 543–552.

[52] Christian Genest and Johanna Nešlehová. "A primer on copulas for count data". In: *ASTIN Bulletin: The Journal of the IAA* 37.2 (2007), pp. 475–515.

[53] Mathieu Gerber, Nicolas Chopin, and Nick Whiteley. "Negative association, ordering and convergence of resampling methods". In: *The Annals of Statistics* 47.4 (2019), pp. 2236–2260.

[54] Subhroshekhar Ghosh, Thomas M. Liggett, and Robin Pemantle. "Multivariate CLT follows from strong Rayleigh property". In: *2017 Proceedings of the Fourteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*. SIAM. 2017, pp. 139–147.

[55] Irène Gijbels, Vojtěch Kika, and Marek Omelka. "On the specification of multivariate association measures and their behaviour with increasing dimension". In: *Journal of Multivariate Analysis* 182 (2021), p. 104704.

[56] Garique F.V. Glonek and Peter McCullagh. "Multivariate logistic models". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.3 (1995), pp. 533–546.

[57] Geoffrey R Grimmett. *Percolation*. Vol. 321. Springer Science & Business Media, 2013.

[58] Jonathan Hermon and Justin Salez. "Modified log-Sobolev inequalities for strong-Rayleigh measures". In: *The Annals of Applied Probability* 33.2 (2023), pp. 1501–1514.

[59] Taizhong Hu. “Negatively superadditive dependence of random variables with applications”. In: *Chinese Journal of Applied Probability and Statistics* 16.2 (2000), pp. 133–144.

[60] Jian Shan Huang and Samuel Kotz. “Modifications of the Farlie–Gumbel–Morgenstern distributions. A tough hill to climb”. In: *Metrika* 49.2 (1999), pp. 135–145.

[61] Mark Huber and Nevena Marić. “Admissible Bernoulli correlations”. In: *Journal of Statistical Distributions and Applications* 6.1 (2019), p. 2.

[62] Kumar Joag-Dev and Frank Proschan. “Negative association of random variables with applications”. In: *The Annals of Statistics* (1983), pp. 286–295.

[63] Harry Joe. *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC, 1997.

[64] Harry Joe and James Jianmeng Xu. “The estimation method of inference functions for margins for multivariate models”. In: *Working Paper, Department of Statistics, University of British Columbia* (1996).

[65] Joseph B. Kadane. “Sums of possibly associated Bernoulli variables: The Conway–Maxwell-binomial distribution”. In: *Bayesian Analysis* 11.2 (2016), pp. 403–420.

[66] Joseph B. Kadane and Ferdinand L. Næshagen. “The number of killings in southern rural Norway, 1300–1569”. In: *The Annals of Applied Statistics* 7.2 (2013), pp. 846–859.

[67] Alam Khursheed and K.M. Lai Saxena. “Positive dependence in multivariate distributions”. In: *Communications in Statistics-Theory and Methods* 10.12 (1981), pp. 1183–1196.

[68] George Kimeldorf and Allan R Sampson. “Monotone dependence”. In: *The Annals of Statistics* (1978), pp. 895–903.

[69] George Kimeldorf and Allan R. Sampson. “A framework for positive dependence”. In: *Annals of the Institute of Statistical Mathematics* 41.1 (1989), pp. 31–45.

[70] Jean-Gabriel Lauzier, Liyuan Lin, and Ruodu Wang. “Pairwise counter-monotonicity”. In: *Insurance: Mathematics and Economics* 111 (2023), pp. 279–287.

[71] Woojoo Lee and Jae Youn Ahn. “On the multidimensional extension of countermonotonicity and its applications”. In: *Insurance: Mathematics and Economics* 56 (2014), pp. 68–79.

[72] Thomas M Liggett. “Stochastic models of interacting systems”. In: *The Annals of Probability* 25.1 (1997), pp. 1–29.

[73] Francisco Louzada, Adriano Kamimura Suzuki, and Vicente Garibay Cancho. “The FGM long-term bivariate survival copula model: modeling, Bayesian estimation, and case influence diagnostics”. In: *Communications in Statistics-Theory and Methods* 42.4 (2013), pp. 673–691.

[74] Giovanni M. Marchetti and Nanny Wermuth. “Palindromic Bernoulli distributions”. In: *Electronic Journal of Statistics* 10 (2016), pp. 2435–2460.

[75] Alexander J. McNeil, Johanna G. Nešlehová, and Andrew D. Smith. “On attainability of Kendall’s tau matrices and concordance signatures”. In: *Journal of Multivariate Analysis* 191 (2022), p. 105033.

[76] Alfred Müller. "Duality theory and transfers for stochastic order relations". In: *Stochastic Orders in Reliability and Risk: In Honor of Professor Moshe Shaked* (2013), pp. 41–57.

[77] Alfred Müller and Dietrich Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons, 2002.

[78] Alessandro Mutti and Patrizia Semeraro. "Symmetric Bernoulli distributions and minimal dependence copulas". In: *Journal of Multivariate Analysis* (2025), p. 105545.

[79] Roger B. Nelsen. *An introduction to copulas*. Springer, 2006.

[80] Charles M. Newman. "Asymptotic independence and limit theorems for positively and negatively dependent random variables". In: *Lecture Notes-Monograph Series* (1984), pp. 127–140.

[81] Charles M. Newman. "Normal fluctuations and the FKG inequalities". In: *Communications in Mathematical Physics* 74.2 (1980), pp. 119–128.

[82] Andrei Okounkov. "The Ising model in our dimension and our times". In: *International Congress of Mathematicians*. European Mathematical Society-EMS-Publishing House GmbH. 2023, pp. 376–413.

[83] Robin Pemantle. "Towards a theory of negative dependence". In: *Journal of Mathematical Physics* 41.3 (2000), pp. 1371–1390.

[84] Robin Pemantle and Yuval Peres. "Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures". In: *Combinatorics, Probability and Computing* 23.1 (2014), pp. 140–160.

[85] Shahrior Pervaz, Tanmoy Bhowmik, and Naveen Eluru. "Examining driver injury severity in motor vehicle crashes: A copula-based approach considering temporal heterogeneity in a developing country context". In: *Accident Analysis & Prevention* 206 (2024), p. 107721.

[86] Giovanni Puccetti and Ludger Rüschendorf. "Computation of sharp bounds on the distribution of a function of dependent risks". In: *Journal of Computational and Applied Mathematics* 236.7 (2012), pp. 1833–1840.

[87] Giovanni Puccetti and Ruodu Wang. "Extremal Dependence Concepts". In: *Statistical Science* 30.4 (2015), pp. 485–517.

[88] Earl D. Rainville. *Special Functions*. The Macmillan Company, New York, 1960.

[89] Ludger Rüschendorf. "Random variables with maximum sums". In: *Advances in Applied Probability* 14.3 (1982), pp. 623–632.

[90] Susanne Saminger-Platz, Anna Kolesárová, Adam Šeliga, Radko Mesiar, and Erich Peter Klement. "The impact on the properties of the EFGM copulas when extending this family". In: *Fuzzy Sets and Systems* 415 (2021), pp. 1–26.

[91] Issai Schur and George Pólya. "Über zwei Arten von Faktorenfolgen in der Theorie der algebraischen Gleichungen." In: *Journal für die reine und angewandte Mathematik* (1914).

[92] Naijun Sha. "A copula approach of reliability analysis for hybrid systems". In: *Reliability: Theory & Applications* 16.1 (61) (2021), pp. 231–242.

[93] Moshe Shaked and J. George Shanthikumar. *Stochastic Orders*. Springer, 2007.

[94] Peng Shi and Wei Zhang. "Private information in healthcare utilization: specification of a copula-based hurdle model". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 178.2 (2015), pp. 337–361.

[95] M. Sklar. "Fonctions de répartition à n dimensions et leurs marges". In: *Annales de l'ISUP*. Vol. 8. 3. 1959, pp. 229–231.

[96] Selim Orhun Susam. "A flexible parameter estimation method for the Farlie–Gumbel–Morgenstern copula: a simulation study". In: *Journal of Statistical Computation and Simulation* (2025), pp. 1–19.

[97] Dirk Tasche. "Risk contributions and performance measurement". In: *Report of the Lehrstuhl für mathematische Statistik, TU München* (1999).

[98] 4ti2 team. *4ti2—A software package for algebraic, geometric and combinatorial problems on linear spaces.* Available at https://4ti2.github.io. 2018.

[99] Marco Terzer. "Large scale methods to enumerate extreme rays and elementary modes". PhD thesis. ETH Zurich, 2009.

[100] Ene-Margit Tiit. "Mixtures of multivariate quasi-extremal distributions having given marginals". In: *Lecture Notes-Monograph Series* (1996), pp. 337–357.

[101] Yves Tillé. *Sampling Algorithms*. Springer, 2006.

[102] Yves Tillé and Alina Matei. *sampling: Survey Sampling*. R package version 2.11. 2025. URL: https://cran.r-project.org/web/packages/sampling/index.html.

[103] Paul Turán. "On the zeros of the polynomials of Legendre". In: *Časopis pro pěstování matematiky a fysiky* 75.3 (1950), pp. 113–122.

[104] David Wagner. "Multivariate stable polynomials: theory and applications". In: *Bulletin of the American Mathematical Society* 48.1 (2011), pp. 53–84.

[105] Nanny Wermuth and David R. Cox. "On the ralation between interactions obtained with alternative codings of discrete variables". In: *Methodika* VI (1992), pp. 76–85.

[106] Douglas B. West. *Combinatorial mathematics*. Cambridge University Press, 2021.

[107] Gordon E. Willmot and Jae-Kyung Woo. "On the class of Erlang mixtures with risk theoretic applications". In: *North American Actuarial Journal* 11.2 (2007), pp. 99–115.

This Ph.D. thesis has been typeset by means of the TeX-system facilities. The typesetting engine was pdfLaTeX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TeX-system installation.