Polynomial Distribution Functions on Bounded Closed Intervals

by

Andrey Chirikhin

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Statistics

University of Warwick, Department of Statistics

December 2007

Contents

Acknowledgements

First and foremost, I would like to thank my wife Elena, for her love and inspiration.

This work would not be possible if my parents, Alexander Chirikhin, D.Sc, and Vera Chirikhina had not brought me up as a scholar.

I would also like to express my deepest gratitude to my PhD advisors Professor Wilfrid Kendall and Professor Saul Jacka, as well as to my colleagues Dr. Hans-Juergen Brasch and Dr. Mikhail Soloveitchik for numerous discussions, suggestions, and motivation.

1 Summary

The thesis explores several topics, related to polynomial distribution functions and their densities on $[0, 1]^M$, including polynomial copula functions and their densities. The contribution of this work can be subdivided into two areas.

- Studying the characterization of the extreme sets of polynomial densities and copulas, which is possible due to the Choquet theorem.
- Development of statistical methods that utilize the fact that the density is polynomial (which may or may not be an extreme density).

With regard to the characterization of the extreme sets, we first establish that in all dimensions the density of an extreme distribution function is an extreme density. As a consequence, characterizing extreme distribution functions is equivalent to characterizing extreme densities, which is easier analytically. We provide the full constructive characterization of the Choquet-extreme polynomial densities in the univariate case, prove several necessary and sufficient conditions for the extremality of densities in arbitrary dimension, provide necessary conditions for extreme polynomial copulas, and prove characterizing duality relationships for polynomial copulas. We also introduce a special case of *reflexive* polynomial copulas.

Most of the statistical methods we consider are restricted to the univariate case. We explore ways to construct univariate densities by mixing the extreme ones, propose non-parametric and ML estimators of polynomial densities. We introduce a new procedure to calibrate the mixing distribution and propose an extension of the standard method of moments to *pinned density moment matching.* As an application of the multivariate polynomial copulas, we introduce *polynomial coupling* and explore its application to convolution of coupled random variables.

The introduction is followed by a summary of the contributions of this thesis and the sections, dedicated first to the univariate case, then to the general multivariate case, and then to polynomial copula densities. Each section first presents the main results, followed by the literature review.

2 Introduction

Polynomials (and power series) are an indispensable tool in both pure and applied probability and in statistics. However, except in the case of Fourier-based kernel methods, polynomials have not been considered good candidates to model distribution functions or densities themselves. This is understandable: the most traditionally attractive feature of the set of polynomials is its structure as a vector space, from which the approximation methods ultimately derive. And it is this very structure that makes (general) polynomials awkward candidates for densities (or distribution functions): probability densities must not be negative.

Our main study subjects are polynomial densities on the M dimensional "unit cube" $[0,1]^M = I^M$ and polynomial copula densities of bounded degree, a special case of the former. This problem leads to the consideration of issues in both real algebraic geometry and convex analysis, since boundedness of the degree of the polynomials and the requirement that the polynomial integrates to 1 over I^M makes the set of such polynomials stable under formation of convex combinations. Our main goals will be a characterization of the extreme subsets of the set of polynomial and polynomial copula densities of a given bounded degree, and developing applied numerical methods involving polynomial models.

This introductory section contains the definitions and available results that will be used in the sequel. We start by reviewing the basic definitions, and then summarizing the relevant results from convex analysis, the theory of copulas, real algebraic geometry, and optimal transport. Most of these concepts will be used throughout the thesis; hence we collect them all together into this section.

2.1 Multivariate Distribution Functions and Copulas

Definition 2.1 . *Given a point* $\mathbf{u} = (u_1, \dots, u_M) \in \mathbb{R}^M$, *consider a sequence of open rectangles, with fixed vertex* $\mathbf{u}$ *and contracting from above*

$$R_n = \left\{ \begin{array}{c} (x_1, \dots, x_M) : x_i \in (u_i, r_{i,n}), \\ \text{such that } u_i < r_{i,n} < r_{i,n-1}, \forall n, i = 1 \dots M \end{array} \right\}. \quad (2.1)$$

Consider a function $f(x) : \mathbb{D} \to \mathbb{R}$, *where* $\mathbb{D} \subseteq \mathbb{R}^M$. *We call* A *the* ***right limit*** *of the function* $f(x)$:

$$A = \lim_{x \to \mathbf{u}+} f(x),$$

if given any sequence $R_n \subset \mathbb{D}$, *satisfying (2.1),* $\forall \varepsilon > 0$, $\exists \delta > 0$ *and* $N_\delta \in \mathbb{N}$ *such that* $\forall n > N_\delta$, $\max(r_{i,n} - u_i) < \delta \Rightarrow |f(r_{1,n}, \dots, r_{M,n}) - A| < \varepsilon$. *We say that* $f(x)$ *is* ***right-continuous*** *in* $\mathbf{u}$ *if* $\lim_{x \to \mathbf{u}+} f(x) = f(\mathbf{u})$.

Definition 2.2 *An* M*-dimensional (multivariate)* ***distribution function*** $f(\mathbf{x})$ *is a right-continuous function* $f(\mathbf{x}) : R^M \to [0, 1]$, *such that*

f is grounded: $\lim_{x_i \to -\infty} f(x_1, \dots, x_i, \dots, x_M) = 0,$ *for each index i individually* (2.2)

f is normalized: $\lim_{\|x\| \to +\infty} f(x_1, \dots, x_i, \dots, x_M) = 1,$ *for some norm* $\|x\|$ (2.3)

f is n-increasing: for any n-dimensional box with corners $(u_1^1, \dots, u_n^1)$ *and* $(u_1^2, \dots, u_n^2)$, *such that* $u_i^1 \le u_i^2, i = 1 \dots n$:

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1 + \dots i_n} f(u_1^{i_1}, \dots, u_n^{i_n}) \ge 0. \quad (2.4)$$

A distribution function $f(\mathbf{x})$ can be interpreted probabilistically as a cumulative distribution function of some M-dimensional random variable $\mathbf{X}$, i.e.$f(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x})$, where $\mathbf{X} \leq \mathbf{x}$ is interpreted coordinate-wise. In particular, condition (2.2) means that zero probability of one coordinate of the random vector implies the zero probability of the whole vector, and (2.4) implies that f induces a non-negative measure on the Borel $\sigma-$algebra of R^M.

If $M = 1$, then a distribution function is merely a non-decreasing right-continuous $f : R \to [0, 1]$.

Definition 2.3 *Denoting by $x^{(i)} \to +\infty$ the action of letting all but i-th component of the M-dimensional vector tend to $+\infty$, we define the $i-th$ marginal distribution function as*

$$m_i(x) = \lim_{x^{(i)} \to +\infty} f(x_1, \dots x_{i-1}, x, x_{i+1}, \dots, x_M). \tag{2.5}$$

Clearly, $m_i : R \to [0, 1]$. We denote $Ran(m_i) \subseteq [0, 1]$ the image of $m_i(x)$.

The setup where we study the properties of a family of multivariate distribution functions with fixed margins is quite common (e.g., optimal transport problems, or coupling). The fundamental result on the interplay between the multivariate distribution functions and their margins is Sklar's theorem, representing any multivariate distribution function as a superposition of the margins and a *copula*. Before we state that theorem, we need to make few more definitions.

Definition 2.4 *The **uniform distribution function** on $[0, 1]$ is the identity mapping $[0, 1] \to [0, 1]$. In other words, $f(x) = x$, for $x \in [0, 1]$.*

Definition 2.5 *An M-dimensional **copula function** is an M-dimensional distribution function with uniform margins.*

That is, a copula c is a mapping $I^M \to [0,1]$, which in addition to the conditions (2.2) and (2.4) satisfies:

$$c(1,1,\ldots,x_i,\ldots,1,1) = x_i. \tag{2.6}$$

The definition of a copula only has content for $M > 1$.

Theorem 2.6 (Sklar) *Given a joint distribution function $f : R^N \to [0,1]$ and M margins $m_i : R \to [0,1]$ there exists a copula $C : [0,1]^M \to [0,1]$ such that*

$$f(x_1,\ldots,x_i,\ldots x_M) = C(m_1(x_1),\ldots,m_i(x_i),\ldots,m_M(x_M)) \tag{2.7}$$

If all m_i are continuous, then C is unique; otherwise, C is uniquely determined on $Ran(m_1)\times Ran(m_2)\times\ldots\times Ran(m_M)$, the direct product of the ranges of the margins. Conversely, if C is a copula and $m_i(x_i)$ are univariate distribution functions then

$$C(m_1(x_1),\ldots,m_i(x_i),\ldots,m_M(x_M))$$

is a multivariate distribution function with margins $\{m_i(x_i), i = 1\ldots M\}$.

For further discussion see [Nel06].

In the theory of coupling methods, the construction involved in the Sklar theorem is known as *quantile* coupling (see, e.g., [Tho00], Section 3). That is, given the family of independent random variables $\{x_i\}$, we define a family of uniform $[0,1]$ random variables $u_i = m_i(x_i)$, and then define the joint distribution function of the u_i.

Definition 2.7 *The copula density is the M-fold mixed partial derivative of*

C with respect to all coordinates, if such a derivative exists:

$$\partial C(u_1, \ldots, u_M) = \frac{\partial^M C}{\partial u_1 \ldots \partial u_M}. \tag{2.8}$$

An immediate (and most practically useful) consequence of the Sklar theorem is that it allows the factorization of the density of f into the product of the densities of the margins and the copula density (if all such densities exist):

$$\partial f(x_1, \ldots, x_M) = \partial C|_{m_1(x_1), \ldots, m_M(x_M)} \prod_i m_i'(x_i) \tag{2.9}$$

For example, this allows us to implement the procedure of parametric estimation of f via likelihood maximization as the two-step procedure: first estimation of the margins (independently of the copula) and only then fitting the copula.

To complete the section, we note that condition (2.6) for copulas implies that

$$\int_{I^{M-1}} \partial C(u_1, , , u_M) du_1 \ldots du_{i-1} du_{i+1} \ldots du_M = 1, \tag{2.10}$$

which is just another way to say that the copula's margin is uniform.

Example 2.8 *Below are the archetypal examples of copulas.*

1. *Independence copula:*

$$C_{ind}(u_1, \ldots, u_M) = \prod_{i=1}^{M} u_i.$$

As the name implies, if a joint distribution of a family of random variables has the independence copula then the random variables are independent.

2. *Maximum copula (also known as the "upper Frechet-Hoeffting bound"):*

$$C_{\max}(u_1, \dots, u_M) = \min(u_1, \dots, u_M).$$

This copula characterizes the "highest" degree of dependency between the margins, technically "full co-monotonicity".

3. *Lower Frechet-Hoeffting bound:*

$$C_{\min}(u_1, \dots, u_M) = \max(1 - M + \sum_{i=1}^{M} u_i).$$

This is a copula only when $M = 2$*.*

Frechet-Hoeffting bounds are called so because for any copula C*,* $\forall (u_1, \dots, u_M)$: $C_{\min}(u_1, \dots, u_M) \leq C(u_1, \dots, u_M) \leq C_{\max}(u_1, \dots, u_M)$.

An important class of copulas is *factor copulas* (see, e.g., [LG05]), which are particular handy when our goal is to convolve the coupled random variables (in other words, to construct the distribution function of their sum). In general, a factor copula can be viewed as a pair $\{G(\xi), C_\xi\}$, where ξ is a random vector with the probability distribution function $G(\mathbf{x}) = \mathbb{P}(\xi \leq \mathbf{x})$, and C_ξ is a family of copulas that, conditionally on ξ, have simple structure, e.g. become conditionally independent copulas.

The motivation for such a construction is that for each C_ξ the probabilistic problem at hand simplifies, and hence the original problem can be solved by applying the Bayes formula using $G(\xi)$ as the conditioning measure. For example, if the goal is to convolve a family of dependent random variables and if their joint distribution function has conditionally independent copula, then conditioned on the realization of the conditioning variable one will have to convole conditionally independent variables.

Example 2.9 (One-factor Gaussian copula) *Given a correlation matrix that has a structure* $\rho_{i,j} = \beta_i \beta_j (1 - \delta_{i,j}) + \delta_{i,j}$ *and taking* ξ *as a standard Gaussian random variable,* $G(\xi) = \Phi(\xi)$*, the standard Gaussian distribution function, we can define a one-factor Gaussian copula as*

$$C_\xi(u_1, \dots u_M) = \prod_{i=1}^{M} \Phi\left(\frac{\Phi^{-1}(u_i) - \beta_i \xi}{\sqrt{1 - \beta_i^2}}\right).$$

The following example illustrates an equivalent way of representing the conditionally independent factor copulas: we consider M transformations $F_i(., \xi) : [0,1] \to [0,1]$, $i = 1 \dots M$, of the marginal distribution transformations, depending on ξ, such that the joint distribution function can be represented as

$$f(x_1, \dots, x_M) = \int \prod_{i=1}^{M} F_i(m_i(x_i), \xi) dG(\xi),$$

and such that the boundary condition is fulfilled:

$$\int F_k(m_k(x_k), \xi) dG(\xi) = m_k(x_k).$$

The main sources on copulas are [Nel06] and [Joe97].

2.2 Convexity and Choquet's Representation Theorem

Definition 2.10 *A subset* S *of a linear space* X *over field* R *is* ***convex*** *if* $c_1, c_2 \in S, \lambda \in [0,1]$ *implies*

$$\lambda c_1 + (1 - \lambda) c_2 \in S \tag{2.11}$$

Given a sequence of vectors x_i of a vector space, their ***convex hull*** *is defined as*

$$\sum_i \lambda_i x_i, \text{ such that } \lambda_i \geq 0, \quad \sum_i \lambda_i = 1. \tag{2.12}$$

If S is convex, a point $x \in S$ is extreme if

$$x = \lambda y + (1 - \lambda) z, \text{ with } y, z \in S \Rightarrow y = z = x. \tag{2.13}$$

The union of all extreme points of X, denoted as $\mathcal{E}(X)$, is called the ***extreme set of*** *X.*

We will use the fundamental result from convex analysis known as Choquet's theorem [Phe01].

Theorem 2.11 *If C is a closed convex subset of a locally convex compact metrizable topological space, then each $c \in C$ is the barycentre of a probability measure supported by $\mathcal{E}(C)$.*

Given the set C, fulfilling the conditions of Choquet's theorem, we will refer to its extreme points as *Choquet-extreme.*

Proposition 2.12 *The set of copulas is convex.*

Proof. Conditions (2.2), (2.4) and (2.6) can be directly verified to be stable under the action of forming the convex combination. ■

In what follows, we will also find the following simple observation useful.

Proposition 2.13 *If the density c can be represented as a finite sum of non-negative functions g_i, then it can be represented as a convex combination of densities.*

Proof. Since $c = \sum_{i=1}^n g_i$, and $\int c\mathrm{d}x = 1$, denoting $\int g_i \mathrm{d}x = \alpha_i > 0$, we observe that $\sum_{i=1}^n \alpha_i = 1$, hence

$$c = \sum_{i=1}^n g_i = \sum_{i=1}^n g_i \frac{\alpha_i}{\alpha_i} = \alpha_i \sum_{i=1}^n \frac{g_i}{\alpha_i} = \alpha_i \sum_{i=1}^n \widetilde{g}_i,$$

where each $\widetilde{g}_i$ is non-negative and integrates to 1 and hence is a density. ■

2.3 The Monge-Kantorovich Problem of Optimal Transport

Let (X, μ) and (Y, ν) be measure spaces and c be a non-negative measurable function on $X \times Y$. Consider the linear functional

$$\pi(c) = \int_{X\times Y} c(x, y)\mathrm{d}\pi(x, y) \tag{2.14}$$

defined on the non-empty convex set $\Pi(\mu, \nu)$ that consists of all probability measures on $X \times Y$ with margins μ and ν (i.e. $\pi[A \times Y] = \mu[A]$ and $\pi[X \times B] = \nu[B]$ for all measurable $A \in X$ and $B \in Y$).

The Monge-Kantorovich Problem of Optimal Transport (we abbreviate to "MK problem" in the following) amounts to finding the minimum of $\pi(c)$ over $\Pi(\mu, \nu)$, for details see [Vil03]. Clearly, if we assume that the margins are uniform then $\Pi(\mu, \nu)$ will be the set of bivariate copulas.

We consider the MK problem, because its functional $\pi(c)$ provides a set of supporting hyperplanes for the relevant set of copulas, the set of hyperplanes parametrized by c. In Section 6 we use this observation to formulate the duality relation for the extreme polynomial copulas.

2.4 Polynomial Densities and Copulas

Our main subject of study are bounded degree polynomial densities on I^M and bounded degree polynomial copula densities.

In what follows, when speaking of a polynomial measure or density we will always assume that they are restricted to I^M. We will also assume that the degree of the measure, density or copula is given and fixed and we are always considering the densities of this fixed bounded degree.

We will also find the following notation useful.

Definition 2.14 *For a multi-index $\alpha = \{\alpha_1, \dots, \alpha_M\} \in \mathbb{Z}_+^M$ define*

$$\|\alpha\| = \sum_{i=1}^{M} \alpha_i, \qquad \alpha^{(i)} = \{0, \dots, 0, \alpha_i, 0, \dots, 0\}, \tag{2.15}$$

$$|\alpha| = \text{"number of non-zero elements in } \alpha\text{"}$$

Remark 2.15 *Before continuing, we need to clarify an ambiguity in our usage of the term **polynomial**. It will be used in two contexts: when speaking about the elements of the ring of polynomials (over $\mathbb{R}$ or $\mathbb{C}$) and when speaking of polynomial functions, i.e. functions $f : \mathbb{C}^N \to \mathbb{C}$, $f(x) = \sum_i \theta_i x^i$, where i is a multi-index. In the latter case we will still refer to polynomial functions as just polynomials. Also when discussing the geometric properties of the set of bounded degree polynomial functions everywhere (except Section 6), we will identify this set with the set of coefficients of polynomial functions, i.e. the subset of $\mathbb{R}^N$ or $\mathbb{C}^N$, not as a subset of a functional space. We do this primarily for brevity; otherwise we can always re-formulate the geometric considerations in terms of bijections between the set of polynomial functions and the set of their coefficients.*

For a fixed positive integer M consider a tuple of M nonnegative integer numbers $\{p_1, \dots p_M\}$, such that $\sum_{i=1}^{M} p_i = N$. The object of our study will be polynomials in M real variables x_i, non-negative on I^M, such that the maximum power of the $i-$th coordinate is p_i, and such that the polynomials can be interpreted as densities of probability distributions. In other words, we consider polynomials of the

$$Q(\mathbf{x}) = Q(x_1, \dots, x_M) = \sum_{i_1=0}^{p_1} \dots \sum_{i_M=0}^{p_M} \theta_{i_1 \dots i_M} \prod_{j=1}^{M} x_j^{i_j}, \tag{2.16}$$

$$\text{such that} \int_{I^M} Q(\mathbf{x}) \mathrm{d}\mathbf{x} = 1, \tag{2.17}$$

$$\text{and } Q(\mathbf{x}) \geq 0. \tag{2.18}$$

Definition 2.16 *We will refer to the tuple $\{p_1, \dots p_M\}$ in (2.16) as the **type** of the polynomial.*

Thus, having fixed the type, we will be dealing with polynomial densities of that type, i.e. polynomials of that type, which are non-negative on I^M and integrate to 1 over I^M.

Definition 2.17 *Given a family of polynomials $h_i(x_1, \dots, x_n) \in \mathbb{R}[x_1, \dots, x_n], i \in \mathbb{N}$, a subset $W(x_1, \dots, x_n) = W \in \mathbb{R}^n$ is called **semialgebraic** if it is defined by the family of polynomial inequalities:*

$$W(x_1 \dots, x_n) = \{(x_1, \dots, x_n) : h_i(x_1, \dots, x_n) \geq 0\}. \tag{2.19}$$

Apparently, I^M is semialgebraic, defined by a family of polynomial inequalities

$$\{h_{i.1}(x_i) = x_i \geq 0; h_{i.2}(x_i) = 1 - x_i \geq 0; i = 1 \dots M\}. \tag{2.20}$$

The positivity condition and the normalization condition merely impose restrictions on the coefficients of $Q(\mathbf{x})$. Restriction (2.17) is linear in the coefficients, hence it defines a hyperplane. Condition (2.18) is more complex, as it is parametrized by the points of I^M; it therefore can be viewed as a whole continuum of conditions on the coefficients of $Q(\mathbf{x})$. However, as we review in the following subsection, there exits a characterization of the form of polynomials that are non-negative on I^M, which refers to I^M only through the polynomials $h_{i.1}(x_i), h_{i.2}(x_i)$.

In addition to being grounded and non-negative, a polynomial copula density has to reproduce margins, i.e. to fulfil (2.10). Note that in its general form (for all copulas) this restriction is also of a continuum nature, as it is a restriction on the whole continuous function.

The situation is simpler for the polynomial copulas, as in this case (2.10) reduces to the system of linear equality constraints for the coefficients of $Q(\mathbf{x})$. Indeed, if we integrate out all but one coordinate (e.g., all except x_1) in (2.16), then we obtain

$$\sum_{i_1=0}^{p_1} \cdots \sum_{i_M=0}^{p_M} \theta_{i_1 \ldots i_M} x_1^{i_1} \prod_{j=2}^{M} \frac{1}{i_j + 1} = 1,$$

for $x_1 \in [0, 1]$. This means that the coefficients of all the powers of x_1 should be zero, and that the constant term equals 1, i.e.

$$\sum_{i_2=0}^{p_2} \sum_{i_M=0}^{p_M} \theta_{i_1 \ldots i_M} \prod_{j=2}^{M} \frac{1}{i_j + 1} = 0, \forall i_1 > 0; \quad \sum_{i_2=0}^{p_2} \sum_{i_M=0}^{p_M} \theta_{0, i_2 \ldots i_M} \prod_{j=2}^{M} \frac{1}{i_j + 1} = 1$$

These formulae provide a finite set of *linear* conditions on the coefficients $\theta_{i_1 \ldots i_M}$ of a polynomial. This is a major simplification from the practical standpoint, as it means that numerical algorithms (e.g. model parameter estimation) formulated for polynomial densities need only be augmented by this set of linear constraints on the coefficients to become applicable to polynomial copula densities.

The set of polynomial distribution functions is not empty, as we can consider an arbitrary sum of squares and normalize it over I^M. The set of polynomial copulas is not empty either.

Example 2.18 *Farlie-Gumbel-Morgenstern (FGM) family.*

$$C(\mathbf{u}) = \prod_{i=1}^{M} u_i \left(1 + \sum_{k=2}^{M} \sum_{1 \leq j_1 < \ldots j_k \leq n} \theta_{j_1 j_2 \ldots j_k} (1 - u_{j_1})(1 - u_{j_2}) \ldots (1 - u_{j_k})\right)$$

There are 2^M *constraints on the parameters such that the above expression defines a copula*

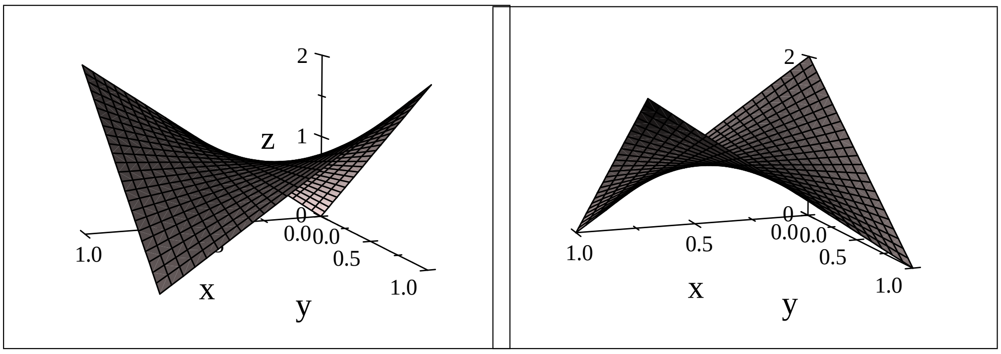
$$\sum_{k=2}^{M} \sum_{1 \leq j_1 < \ldots j_k \leq n} \theta_{j_1 j_2 \ldots j_k} \varepsilon_{j_1} \varepsilon_{j_2} \ldots \varepsilon_{j_k} \geq -1, \text{ for each } \varepsilon_{j_1}, \varepsilon_{j_2}, \ldots, \varepsilon_{j_k} \in \{-1, 1\}.$$

This conditions results from the observation that the density is linear in each u_1*, therefore it must reach its minima only in the vertexes of* I^M*. The above condition merely states that the value of the density on the vertexes should be non-negative.*

In the two-dimensional case we have for the FGM copula and its density:

$$C_\theta(x, y) = xy(1 + \theta(1 - x)(1 - y)); \qquad \partial C_\theta(x, y) = 1 + \theta(1 - 2x)(1 - 2y).$$

It can be shown that this is the only two-dimensional quadratic family of copulas, and that the extreme copulas correspond to the cases $\theta = \pm 1$*. The extreme densities are plotted below.*



In the sections below we will provide the form the polynomial copula (and general polynomial distribution function, correspondingly) must have. These forms are the consequences of groundedness, normalization and, in the case of copulas, of reproduction of the margins. Together with the semialgebraic results referred to in the previous section, this fully characterizes the set of both polynomial distribution functions and copulas. It is possible however to further narrow down the characterization.

Proposition 2.19 *For all N, the set of polynomial densities/copulas of a degree bounded by N is compact and metrizable.*

Proof. This follows from the fact that the set of polynomial copulas of bounded degree is a finite-dimensional linear span within the set of bounded-degree polynomial distribution functions. ■

The last two propositions have established the facts that the sets of polynomial densities and polynomial copula densities of bounded degree are convex, compact and metrizable subsets of $\mathbb{R}^n$. This allows us to use the Choquet theorem to reduce the problem of characterizing those sets to characterizing their extreme sets. One of the main subjects of this thesis is to generate results and conjectures concerning this problem.

2.5 Polynomials that are Non-Negative on I^M

Characterization of the extreme polynomial densities of a bounded degree on I^M amounts to characterizing a subset of polynomials that are non-negative on I^M and integrate to 1 over this set. As we have already noted, the set I^M is semialgebraic. Thus the characterization problem we consider is one of characterizing a sub-set of polynomials of a bounded degree, non-negative on a semialgebraic set. Except for the boundedness of the degree, this is a standard problem of *real algebraic geometry*. In this section we therefore introduce some basic constructs of real algebraic geometry and briefly analyze their applicability to our problem given that the boundedness of the degree of polynomials is essential for us.

Definition 2.20 *Consider a commutative ring K with operations denoted as $+$ and $\cdot$ (we will write for brevity $a \cdot b = ab$). A subset T of a commutative ring K is called a* ***quadratic module*** *if*

$$T + T \subset T,\ K^2 T \subset T,\ 1 \in T, -1 \notin T. \tag{2.21}$$

A quadratic module is a generalization of the concept of "sum of squares". An example of a quadratic module is $K[x_1, \ldots x_N]^2$, the set of sums of squares of multivariate polynomials over K.

Definition 2.21 *A quadratic module T of a commutative ring K is called a* ***preorder*** *or* ***pre-positive cone*** *if it is multiplicatively closed.*

This amounts to replacing the condition $K^2 \cdot T \subset T$ in (2.21) for $T \cdot T = T^2 \subset T$ and $K^2 \subset T$.

An example of a preorder would be the set of positive continuous functions: $f \in C(R)$, such that $f(x) \geq 0$.

Definition 2.22 *If the preorder T is such that $T \cup -T = K$ then T is called a **positive cone**.*

Definition 2.23 *If $\leq$ linearly orders the underlying set of K (that is the ordering $\leq$ is defined for all pairs of elements of K), then it is called a **field ordering** if*

$$a \leq b \Rightarrow a + c \leq b + c,$$

$$0 \leq a, 0 \leq b \Rightarrow 0 \leq ab.$$

It can be proven that P being a positive cone of K is equivalent to $\leq_P$, defined by $a \leq_P b \Leftrightarrow b - a \subset P$ being an ordering of K.

Definition 2.24 *If $-1 \notin \sum K^2$, which denotes sums of squares of the elements from K, then K is called **real**. A field is called real closed if it has no proper real algebraic extension.*

A characterization of a real field is that it has an ordering; furthermore such an ordering is unique if and only if $\sum K^2$ (sums of squares of elements of K) is a positive cone of K.

Definition 2.25 *Given a ring R, write $R[x_1, \dots x_n]$ for the ring of $n-$variate polynomials over R, and $R(x_1, \dots x_n)$ for the ring of polynomial fractions over R. Furthermore, $R[x_1, \dots x_n]^2$ and $R(x_1, \dots x_n)^2$ will denote the sums of squares of elements of $R[x_1, \dots x_n]$ and $R(x_1, \dots x_n)$ correspondingly.*

Real algebraic geometry evolved from Artin's solution to Hilbert's 17^{th} problem (see, e.g., [PD01]), which was to prove that each non-negative multivariate polynomial over $\mathbb{R}$ can be represented as sum of squares of rational polynomial functions. In the one-dimensional case the result is stronger: The Fundamental Theorem of Algebra implies that a polynomial that is non-negative on $\mathbb{R}$ is a sum of squares of polynomials. From the general real-algebraic point of view this is because for $\mathbb{R}[x]$ the natural preorder is the set of polynomials that can be represented as *sums of squares* (SOS): $T = \mathbb{R}[x]^2$. This property does not hold for higher dimension: the preorder in $\mathbb{R}[x_1 \ldots, x_n]$ fails to be $\mathbb{R}[x_1 \ldots, x_n]^2$. The Hilbert conjecture was that a multivariate polynomial non-negative on R^M could be represented as a sum of squares of rational functions, i.e. fractions of polynomials, and this conjecture was proven to hold.

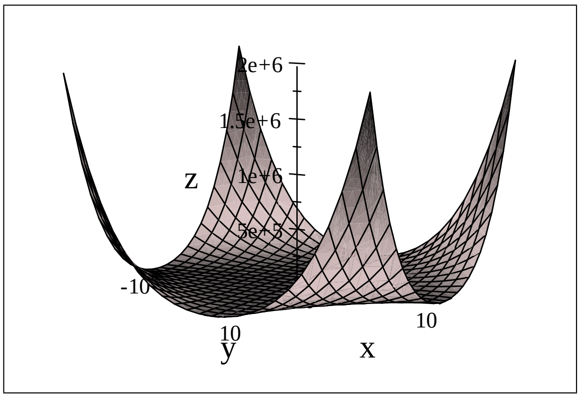
Example 2.26 *A polynomial $p(x_1 \ldots, x_n) \geq 0$, such that p is not SOS can be easily constructed using the arithmetic-algebraic inequality*

$$\frac{1}{k}\sum_{i=1}^{k} a_i \geq \left(\prod_{i=1}^{k} a_i\right)^{1/k},$$

except in the cases ruled out by Hilbert's results. For example, let $n = 2$, and $k = 3$ and consider polynomials $\{x^4y^2, x^2y^4, 1\}$. Then the arithmetic-algebraic inequality implies that

$$x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \geq 0.$$

The graph of this polynomial (known as the Motzkin polynomial) is shown on the chart below for $(x, y) \in [-10, 10]^2$:



Motzkin Polynomial

The fact that it is not SOS follows from coordinate-wise examination.

Hilbert's 17^{th} problem dealt with polynomials that are non-negative on the whole of $\mathbb{R}^n$. In the past few decades, the results related to that problem were generalized for the setup requiring the polynomial to be non-negative only on a semialgebraic set.

Definition 2.27 *Given a family of polynomial inequalities* $\{h_i \geq 0, i = 1 \ldots M\}$, *define the sets of polynomials* $T_{h_1,\ldots,h_M}[x_1,\ldots,x_n]$ *and* $T_{h_1,\ldots,h_M}(x_1,\ldots,x_n)$ *as*

$$T_{h_1,\ldots,h_M}[x_1,\ldots,x_n] = \left\{ \begin{array}{c} \sum_{\sigma_j=\{0,1\}} \prod_{j=1}^{M} h_j^{\sigma_j} m^{+}_{\sigma_1\ldots\sigma_M}, \\ \textit{such that } m^{+}_{\sigma_1\ldots\sigma_M} \in \mathbb{R}[x_1\ldots,x_n]^2 \end{array} \right\}, \tag{2.22}$$

and

$$T_{h_1,\ldots,h_M}(x_1,\ldots,x_n) = \left\{ \begin{array}{c} \sum_{\sigma_j=\{0,1\}} \prod_{j=1}^{M} h_j^{\sigma_j} m_{\sigma_1\ldots\sigma_M}, \\ \textit{such that } m_{\sigma_1\ldots\sigma_M} \in \mathbb{R}(x_1\ldots,x_n)^2 \end{array} \right\}. \tag{2.23}$$

These are called the ***preorders*** *generated by the polynomials* h_j*, defining the reference semialgebraic set.*

These constructions are quite similar to the ideals in $\mathbb{R}[x_1, \ldots, x_n]$ that would be generated by h_j. The only difference is the nature of "multiples", which would be

- arbitrary polynomials in the case of ideals,
- polynomials representable as squares in the case of $T_{h_1,\ldots,h_M}[x_1, \ldots, x_n]$,
- polynomials representable as sums of squares of rational functions in the case of $T_{h_1,\ldots,h_M}(x_1, \ldots, x_n)$.

The main relevant results are the following two representation theorems.

Theorem 2.28 ("Schmuedgen's Theorem", [Sch91]) *If* $p(x_1, \ldots, x_n) > 0$*, for* $(x_1, \ldots, x_n) \in W(x_1, \ldots, x_n)$*, then* $p(x_1, \ldots, x_n) \in T_{h_1,\ldots,h_M}[x_1, \ldots, x_n]$.

Theorem 2.29 ("PD Theorem", [PD01], Theorem 3.5.8) *If* $p(x_1, \ldots, x_n) \geq 0$*, for* $(x_1, \ldots, x_n) \in W(x_1, \ldots, x_n)$*, then* $p(x_1 \ldots, x_n) \in T_{h_1,\ldots,h_M}(x_1, \ldots, x_n)$.

The theorems generalize the classical results, preserving the main observation: definite polynomials can be represented as a sum of squares of polynomials, while semidefinite polynomials may only be represented as sums of squares of rational functions.

These results, in their most general form, are obtained using the theory of quadratic forms over real fields (see the first 3 Sections of [PD01]). Let K denote a field of characteristic not 2 (i.e., $1 + 1 \neq 0$), and consider the set of quadratic forms

$$f(x_1, \ldots x_M) = \sum_{i=1}^{M} \sum_{j=1}^{M} b_{ij} x_i x_j.$$

First, the set of quadratic forms is split into the equivalence classes that are represented by different diagonal quadratic forms. Then the set of such equivalence classes is endowed with the structure of a commutative ring (making it the so-called "Witt ring of K", $W(K)$). Recall that a ring is real closed if and only if it has an unique order. So if K is real then it is possible to define positive definiteness of a quadratic form from $W(K)$. The key result is Pfister's Local-Global Principle. This can be loosely stated as follows: if K is real, then if the signature of a quadratic form f with respect to the unique ordering of the real closure of K is zero, then f is a torsion element of $W(K)$, i.e. there exists $m \in \mathbb{N}$, such that $mf = 0$ (which means that if we add f to itself m times using the addition operation in $W(K)$, then we obtain the zero element of $W(K)$). Note that none of those addition operations have anything to do with polynomial addition, it is the addition in $W(K)$. This result is then applied to the case when K is a field of polynomial rational functions (field of fractions of $K[x_1, \dots x_M]$). This implies that should a polynomial p be non-negative (or "*positive semi-definite*" in the terminology of the theory of quadratic forms), then for some $m \in \mathbb{N}$ the quadratic form $(s_1 - f \cdot s_2)$ should be a torsion element of $W(K(x_1, \dots x_M))$. Here $s_1, s_2 \in K(x_1, \dots x_M)^2$, the order of $K(x_1, \dots x_M)$. This implies that there should exist s_1, s_2 such that $s_1 - f \cdot s_2 = 0$, implying Artin's solution. The same logic is applied to prove the semialgebraic case, only the preorder in the field of the rational polynomial functions is defined by (2.23).

The relevance of these characterizations to our problem is apparent. There are two complications however. Firstly, we consider polynomials of bounded degree. This means that we cannot directly use the ring structure of the polynomial ring; only the group structure is available. Therefore the representation results above are useful, but only as a starting point of a more precise characterization. Furthermore, as we will prove below, extreme densities will always have to have zeros on I^M, which implies that we are in the situation described by Theorem 2.29, hence the characterizing class of functions is too rich for our purposes.

One will therefore need more restrictive characterizations. As we show below in Section 5, we essentially need a characterization of polynomials, non-negative on I^M, that cannot be represented as sums of other polynomials, non-negative on I^M. This problem is not quite solved yet. Partial results are now being obtained with regards to characterizing what have come to be known as *distinguished representations* of non-negative polynomials. Given $h_i(x_1, \ldots, x_n) \in \mathbb{R}[x_1, \ldots, x_n], i \in \mathbb{N}$ as before, a distinguished representation of a non-negative polynomial is that of the form:

$$f = s_0 + \sum_{i=1}^{n} s_i h_i, \tag{2.24}$$

where $s_i, i = 0 \ldots n$ are sums of squares. This is a stronger characterization than the one given by Theorems 2.28 and 2.29. However to prove the distinguished representation theorems one needs to consider quadratic modules, generated by the h_i's (which is given by the expression above), as opposed to preorderings generated by h_i. This makes a difference for the general underlying field K, when a quadratic module is not necessarily a preordering.

Distinguished representations are not entirely suitable for our case, because they do not consider summands of the form $s_{i,j} h_i h_j$. Therefore in this thesis we sometimes are only able to formulate conjectures of the multidimensional characterizations of extreme polynomial densities.

3 Problem Set-up and Contribution

Although studying polynomial distribution functions with applications to applied probability is an obvious idea, this subject has never really been researched. Therefore many relatively simple results contained in this work are original.

As noted in the Summary, the contribution of the work comprises dealing with the two sets of problems: proving or conjecturing characterizations of the extreme sets of polynomial densities and copulas, and developing applied statistical methods that utilize the fact that the density is polynomial (whether or not it is an extreme).

Section 4 is devoted to the univariate case. We first fully and constructively characterize the extreme polynomial densities on $[0, 1]$ in terms of their factorizations. Since polynomials in one variable are characterized (up to a multiplicative constant) by their roots, Choquet's theorem implies that each polynomial density can be associated with a copula of the marginal measures of the roots of the extreme densities of same degree. We then identify the special role of the case of the independent copula of the roots, in which case the coefficients in the polynomial factorization can be simply interpreted as the first and second moments of the measures of the roots. This yields a fast numerical algorithm for calculation of the normalization constant, which in general would be a bottleneck in any numerical method involving construction of a mixture. We build on the case of independent roots by naturally generalizing the independence copula to a factor copula of the roots, which, numerically, reduces the problem of constructing the mixture to a finite number of conditionally independent problems. We observe that in this case a factor copula can be formulated in an entirely non-parametric way, by specifying only the first few moments of the conditional measures of the roots. Finally, we formulate a Pinned Density Moment Matching parameter estimation procedure in the case of a factor copula of the roots. The key point here is that working with polynomials necessitates solving systems of non-linear equations for the first two moments of the conditional measures of the roots. Therefore we can augment the set of equations/constraints by another linear constraint (or set of them) that would prescribe the value of the density in a given point. Since the constraint is linear it does not substantially change the original estimation setup.

Section 5 is devoted to the case of polynomial densities on I^M. We first generalize the univariate observation of existence of an affine bijection between distribution functions and their densities. The general form of a polynomial distribution function is derived, which by construction satisfies all conditions except non-negativity. Several results from the univariate case are extended, up to formulation partial necessary and sufficient conditions for the extreme multivariate polynomial, generalizing the univariate factorization approach by using standard results of real algebraic geometry. Our characterization results are only practical, however, as they only reach the point where one needs to distinguish between the $\mathbb{C}$-irreducible real semi-definite polynomials that can be represented as sums of other real semi-definite polynomials and those that cannot. As we have already mentioned, existing results from semialgebraic geometry are not strong enough to do such differentiation in the necessary generality and we were not able to obtain them ourselves yet. The section therefore ends with a partial conjecture on the characterization of the two-dimensional extreme polynomial densities.

Section 6 is devoted to polynomial copulas. As before, we make use of the affine bijection between polynomial copulas and their densities to restrict our study to the polynomial densities. The general form of a polynomial copula is derived, satisfying all restrictions, except non-negativity. We are still able to extend certain necessary conditions for the extreme copula from the more general case of multivariate densities. However, the necessity to reproduce the uniform margin renders using the factorization approach awkward. Therefore we follow a different path and use the Monge-Kantorovich (MK) optimal transport problem setup to identify a duality correspondence between the set of extreme polynomial copulas and positive measures having a set of finite moments. Using this correspondence we then propose an approximate but numerically tractable method to identify the extreme polynomial copulas as the limits of MK problems with properly selected weight functions. We finish the section with proposing a new *polynomial coupling* method, which may be used to convolve coupled random variables. We show that polynomial copulas can be viewed as linear combinations of conditionally independent copulas operating on the powers of the marginal distributions, which appears useful in certain applied problems.

4 Polynomial Densities on $[0, 1]$

In this section we study the univariate case. We find the univariate case crucial to build up intuition for the multivariate case and for copulas, as certain abstract constructions of the univariate case also apply in higher dimensions.

First we establish that identifying extreme polynomial distribution functions is equivalent to identifying extreme polynomial densities. Then we prove the main result: the constructive characterization of the Choquet-extreme densities. Then we study certain properties following from the form of the extreme densities and go on to propose several applied statistical methods for polynomial models of the bounded random variables.

4.1 General Form and Affine Bijection

Fix the order N and consider the sets $\mathbb{P}_N$ and $\mathbb{Q}_N$ of polynomial distribution functions and densities respectively.

Proposition 4.1 *A polynomial distribution function on* $[0, 1]$, *i.e. an element of* $\mathbb{P}_N$ *has the form*

$$P(x) = x\left(1 + \sum_{i=1}^{N-1} \theta_i (x^i - 1)\right). \tag{4.1}$$

Proof. Consider a general polynomial of bounded degree $P(x) = \sum_{i=0}^{N} \eta_i x^i$.

Groundedness (2.2) implies $P(0) = 0$, therefore $\eta_0 = 0$. Therefore we can write

$$P(x) = x \sum_{i=0}^{N-1} \theta_i x^i. \tag{4.2}$$

Furthermore, normalization implies $P(1) = 1$, therefore $\sum_{i=0}^{N-1} \theta_i = 1$, which implies $\theta_0 = 1 - \sum_{i=0}^{N-1} \theta_i$. Substituting this into (4.2) yields (4.1). ∎

This means that we do not "lose" the free term when differentiating a polynomial distribution function, hence there is a bijection between polynomial distribution functions and their densities. Since differentiation is a linear operation, the bijection is affine. Therefore, an extreme polynomial density will be the density of an extreme distribution function.

4.2 Characterization of Extreme Densities

In this section we will denote by $\mathcal{E}(\mathbb{Q}_N)$ the extreme set of $\mathbb{Q}_N$.

4.2.1 Main Theorem

In what follows we assume that $\deg(Q(x)) > 0$, since the case of $Q(x) = 1$ (the uniform density) is trivial.

Consider an arbitrary polynomial density $Q(x)$, i.e. just a polynomial, non-negative on $[0, 1]$ and integrating on this segment to 1. The Fundamental Theorem of Algebra implies such a polynomial can be factorized into:

$$Q(x) = Cx^p(1-x)^q \prod_i (x-a_i)^2 \prod_k (x^2 + 2b_k x + c_k), \tag{4.3}$$

where $a_i, b_k, c_, \in \mathbb{R}$ and $b_k^2 < c_k$.

In this section we obtain our main univariate constructive result, given by the following.

Theorem 4.2 *For the density $Q_N^{\mathcal{E}}$ to be extreme, i.e. $Q_N^{\mathcal{E}} \in \mathcal{E}(\mathbb{Q}_{\mathcal{N}})$, it is necessary and sufficient for it to have a factorization of the form*

$$Q_N^{\mathcal{E}}(x) = Cx^{p_0}(1-x)^{p_1} \prod_{j>1}(x-a_j)^{2p_j} \tag{4.4}$$

such that $a_j \in (0,1)$, *for* $j > 1$ *and* $N = p_0 + p_1 + 2\sum_{j\geq 0} p_j$ *where* $p_j \in Z_+$.

In other words, $Q_N^{\mathcal{E}}$ *must have only real roots and all roots are located on* $[0,1]$.

The most important observation is that the first degree polynomials, of which the polynomial in (4.4) is the product are polynomials with real coefficients that are irreducible over $\mathbb{C}$. This will serve as the source of generalization of this result to higher dimensions.

Representation (4.4) implies that extreme polynomial densities of a given degree naturally split into equivalence classes. Such classes are characterized by the density behavior at 0 and 1, which is specified by powers p_0 and p_1. The classes are stable under formation of convex combinations. In this section, when necessary, we will refer to the corresponding pair of (p_0, p_1) as *univariate type* of Q.

Before proving Theorem 4.2 we will prove several lemmas that we find useful in building up intuition about the properties of an extreme density. Some of these lemmas are important enough to be presented individually because they can be generalized for the multivariate case, keeping the proof essentially same.

Lemma 4.3 *$Q(x)$ is not Choquet-extreme if any of the following conditions hold (in terms of notation of (4.3)):*

1. *At least one of the factors in $Q(x)$ is strictly greater than 0 on I.*

2. $\deg Q(x) < N$.

Proof. 1. Assume $Q(x) = R(x)U(x)$, such that $U(x) > 0$ on I. Then selecting $\varepsilon \in (0, \quad \min_I U(x))$ we can write

$$Q(x) = \frac{1}{2}R(x)\left[(U(x) - \varepsilon) + (U(x) + \varepsilon)\right] = Q_-(x) + Q_+(x),$$

and applying Proposition 2.13 yields the representation of $Q(x)$ in terms of other densities; hence $Q(x) \notin \mathcal{E}(\mathbb{Q}_N)$.

2. Denote $\mathbb{E}_Q()$ to be the expectation with respect to density $Q()$ and consider $R(x) = C(x - \mathbb{E}_Q(x))$. Then $\int_I R(x)Q(x) = 0$. Select a constant C such that $|R(x)| < 1$. Then $Q(x)(1 \pm R(x))$ is a density, necessarily belonging to $\mathbb{Q}_N$, but also $Q(x) = \frac{Q(x)+R(x)}{2} + \frac{Q(x)-R(x)}{2}$, hence $Q(x) \notin \mathcal{E}(\mathbb{Q}_N)$. ■

Corollary 4.4 (1-d Necessary Condition) *$Q(x) \in \mathcal{E}(\mathbb{Q}_N)$ implies that $Q(x)$ has the form (4.4).*

Now we turn to the sufficient conditions and start with

Lemma 4.5 *If densities $P_N(x), Q_N(x), N \geq 2$ are such that $P_N(x) = (x - a)^p R_{N-p}(x)$ and $Q_N(x) = (x - a)^q L_{N-q}(x)$ for some p and q such that $p > q$, $R_{N-p}(a) > 0$ and $L_{N-q}(a) > 0$ then $P_N(x) < Q_N(x)$ in some $\overset{\circ}{\mathrm{U}}(a) \bigcap [0, 1]$, where $\overset{\circ}{\mathrm{U}}(a)=\mathrm{U}(a)\backslash\{a\}$ and $U(a)$, is an open set containing a.*

Proof. Consider

$$\frac{P_N(x)}{Q_N(x)} = (x - a)^{p-q}\frac{R_{N-p}(x)}{L_{N-q}(x)}$$

For some $\delta > 0$ consider $\overset{\circ}{\mathrm{U}}_\delta(a)$ which is small enough for $\frac{P_N(x)}{Q_N(x)}$ not to change the sign over it. Such selection is possible because both numerator and denominator are continuous at a. Then

$$\frac{P_N(x)}{Q_N(x)} = (x-a)^{p-q}\frac{R_{N-p}(x)}{L_{N-q}(x)}$$

does not change sign over $\overset{\circ}{\mathrm{U}}_\delta(a)\bigcap[0,1]$. This is because if a is an inner point of $[0,1]$, then both p and q are even, and otherwise either $x>a$ or $x<a$ on $\overset{\circ}{\mathrm{U}}_\delta(a)\bigcap[0,1]$ (for end roots). Now making $\delta\to 0$, select

$$\delta^* = \sup\left\{\delta : (x-a)^{p-q}\frac{R_{N-p}(x)}{L_{N-q}(x)}\right\} < 1,$$

which exists as the ratio stays positive and $(x-a)^{p-q}$ becomes sufficiently small. Any subset of $\overset{\circ}{\mathrm{U}}_{\delta^*}(a)$ fulfills the conditions of the Lemma. ■

Proof (of Theorem 4.2). Necessity follows from Corollary 4.4, so we prove sufficiency.

Suppose $Q_N(x) = \lambda R_N(x) + (1-\lambda)L_N(x), \lambda\in(0,1)$. The zero set of Q_N, should the intersection of zero sets of R_N and L_N, therefore it is sufficient to show that the multiplicities of the common roots of R_N, L_N and $Q_N(x)$ are same.

Suppose, on the contrary, that for some root a (for any $a\in[0,1]$)

$$\begin{aligned}Q_N(x) &= (x-a)^m Q_{N-m}(x) = \lambda(x-a)^p R_{N-p}(x) + (1-\lambda)(x-a)^q L_{N-q}(x)\\ &= (x-a)^q\left(\lambda(x-a)^{p-q}R_{N-p}(x) + (1-\lambda)L_{N-q}(x)\right) = (x-a)^q T_{N-q}(x),\end{aligned}$$

assuming $p\geq q$. Note that $T_{N-q}(a)\neq 0$ and $Q_{N-m}(a)\neq 0$. Now if $q\neq m$

then applying ***Lemma*** 4.5 we get that

$$Q_N(x) < (x-a)^q T_{N-q}(x) \text{ or}$$
$$Q_N(x) > (x-a)^q T_{N-q}(x)$$

in some $\overset{\circ}{\mathrm{U}}(a)$, if $m > q$ and $m < q$ correspondingly, yielding a contradiction, hence $m = q$. Continuing each root we obtain that the multiplicity of each root of $Q_N(x)$ equals to the lowest multiplicity of that root in R_N and L_N. As the order of the density is bounded this means that $p = q$ for all roots, which completes the proof. ■

4.2.2 Reconciliation with the Semialgebraic Approach

Theorem 4.2 provides a more precise representation of extreme polynomial densities in terms of 1-d natural preorder in the ring of polynomials in one variable, positive on $[0, 1]$ than the one suggested by the standard representation theorems like Theorem 2.29. The relevant preorder is generated by the following system of irreducible polynomial inequalities:

$$h_1(x) = x \geq 0,$$
$$h_2(x) = 1 - x \geq 0.$$

This allows rewriting the main result of Theorem 4.2 as

$$Q_n(x) \in \mathcal{E}(\mathbb{Q}_n) \Leftrightarrow$$
$$Q_n(x) = h_1^{\epsilon_1}(x) h_2^{\epsilon_2}(x) q(x), \text{ where } \epsilon_1, \epsilon_1 = \{0, 1\},$$

where $q(x)$ is an even degree polynomial having only real roots of even multiplicity only on $[0, 1]$. Therefore we can say that $Q_n(x)$ is extreme if it is a full degree polynomial belonging to $T_{h_1 h_2}[x]$ (as in (2.22)) and having only real roots only on $[0, 1]$. This characterization is indeed much stronger the one suggested by (2.22). We not only require that $m^+_{\sigma_1 \ldots \sigma_M}$ is a square, but that it has only real roots that must be concentrated on $[0, 1]$.

4.3 Independent Copula of the Roots and Normalization

In this section consider the class of extreme univariate densities of order N with fixed powers p and q of x and $(1 - x)$ (following the notation of (4.3)) and denote $\widetilde{N} = (N - p - q)/2$.

Since extreme densities of bounded degree are fully characterized by their roots, the parameter space of the densities of the given univariate type is $[0, 1]^{\widetilde{N}}$ (this will automatically include multiplicities). We therefore can say that the probability measure referred to in Choquet's theorem (herein referred to as the root "mixing measure") is in our case fully characterized by the marginal root distribution functions and their copula. The roots can therefore be viewed as random variables, taking values in the range $[0, 1]$.

Let $f_i(a_i), i = 1 \ldots \widetilde{N}$ be the family of the marginal root distribution functions and consider the case when multiplicities of all the inner roots are 2 and they are coupled with the independence copula, i.e.

$$M(\mathbf{a}) = \prod_{i}^{\widetilde{N}} f_i(a_i). \tag{4.5}$$

Since all f_i have bounded support, they will also have finite moments of all degrees, including the first two, which we denote $\mu_{i,1}$ and $\mu_{i,2}$. Now integration of (4.4) with respect to (4.5) yields:

$$Q(x) = Cx^p(1-x)^q \prod_{i=1}^{\widetilde{N}} (x^2 - 2\mu_{i,1}x + \mu_{i,2}). \tag{4.6}$$

Thus we obtain:

Proposition 4.6 *If the mixing measure of the roots of (4.4), where* $p_k = 2$, $k = 1 \ldots \widetilde{N}$, *is obtained by coupling the marginal distribution functions of the roots with the independence copula, then the density characterized by such a mixing measure is characterized by only the first two moments of the marginal measures.*

Proposition 4.7 *If the mixing measure of the roots of (4.4) is obtained by coupling the marginal distribution functions of the roots with the independence copula, then the density characterized by such a mixing measure is characterized by only the first* $2p_k$ *moments of the marginal measures (where the* p_k *are defined in (4.4)).*

If we construct a numerical method based on representations of the extreme densities, we need a method of fast computation of the normalization constant. This is because we cannot really make use of the form of the density obtained by differentiating (4.1). The representation of an extreme density can trivially be reconciled with (4.1), however this will yield a system of equations, containing squares of the roots, which will have to be solved numerically.

As an alternative, we note the fastest practical way to compute the normalization constant in the case when, given the marginal root distributions, the root mixing measure has independence copula. An important observation here is that the inverse of the constant is linear in the moments of the marginal measures. This allows a fast analytical way of arriving at the gradient of the constant with respect to the moments of the measures of the roots that may be useful in applied numerical methods.

To simplify the exposition we formulate the following straightforward observation as follows.

Proposition 4.8 *Write an extreme density as*

$$Q(x) = \sum_l p_l x^l = C\, x^n (1-x)^m \prod_{j>0} (x - a_j)^2 \tag{4.7}$$

$$= (-1)^m C \prod_{k=1}^{N} (x + b_k), \tag{4.8}$$

writing

$$\begin{aligned} b_k &= 0 & k &= 1 \dots n, \\ b_k &= -1 & k &= n+1 \dots n+m, \\ b_k &= -a_j & k &= n+m+2(j-1)+1 \dots n+m+2j. \end{aligned}$$

Then the normalization constant C can be computed as

$$C(\overline{a}) = \frac{(-1)^m}{\sum_{l=1}^{N} p_l / l}. \tag{4.9}$$

Proof. The normalization constant is given by the condition $\int Q(x)\mathrm{d}x = 1$, which implies that the normalization constant is a function of the vector of the roots of the extreme density, $\overline{a} = \{a_j\}$.

To compute the integral we therefore need to convolve (possibly multiple) terms of the form $x + b_j$ and then integrate the resulting polynomial term-wise. The coefficients $p_l, l = 0, \dots, N$ of the product polynomial can be obtained by one of the methods outlined in the Appendix, taking into consideration that here we are dealing with binomials $(x + b_k)$ and that we only really need to multiply the terms corresponding to $(x - a_j)^2$ and $(1 - x)$, as multiplication by x^m merely amounts to renumbering $p_l = p_{l-m}$.

Integration of the product polynomial term-wise and substituting into the normalization condition implies that the constant satisfies (4.9). ■

We here note that the same method of calculation of the normalization constant can be applied in a more general case of a density with the following decomposition:

$$Q(x) = Cx^p(1-x)^q \prod_{k=1}^{M}(x^2 + f_k x + g_k). \tag{4.10}$$

To conclude, we note that (4.10) implies that p_l are linear in f_k and g_k. This implies that one can efficiently compute the analytical gradient of C with regards to f_k and g_k, Indeed, denoting

$$C = \frac{1}{\int x^p(1-x)^q \prod_{k=1}^{M}(x^2 + f_k x + g_k)\mathrm{dx}} = \frac{1}{\int P(x)\mathrm{dx}},$$

and letting $u_k = \{f_k, g_k\}$ we obtain

$$\frac{\partial C}{\partial u_k} = -C^2 \int \frac{\partial P(x)}{\partial u_k}\mathrm{dx},$$

with $\partial P(x)/\partial u_k$ readily available using the approach described in the Appendix, specifically formulas (8.9)-(8.10).

4.4 Representation of an Arbitrary Density

The Choquet theorem only asserts existence of a representation of a point of a convex compact set in terms of the set's extreme points, but does not provide a way to construct such a representation. Since we are dealing with polynomials, addition and multiplication of which can easily be handled analytically, and we have a constructive characterization of the extreme polynomial densities, a natural question is how a given non-extreme polynomial density can be represented in terms of the extreme ones. In this section we find an explicit representation of $Q(x)$ as the convex combination of the boundary terms (4.4).

By the Fundamental Theorem of Algebra any polynomial over $\mathbb{R}$ can be represented as in (4.3). Our approach therefore is to start with (4.3) and then show how each of the factors, if necessary, can be represented as sums of products of $x, (1-x), (x-a_i)^2$ in some powers (which would be the terms like in (4.4)). Once we obtain such a representation for each original factor in (4.3), we can open the brackets to obtain an expression involving only the sum of terms as in (4.4) and then apply Lemma 2.13 to turn each term in the sum into a density.

First we make (4.3) a bit more specific and write $Q(x)$ as:

$$Q(x) = Cx^p(1-x)^q \prod (x-a_i)^2 \prod (x^2+2b_kx+c_k^2)\times \\ \prod \operatorname{sgn}(d_j)(d_j-x) \prod \operatorname{sgn}(g_l)(x^2+2f_lx+g_l), \tag{4.11}$$

and we assume that

$a_i \in [0,1]$	this corresponds to the terms in the extreme density
$d_j \notin [0,1]$	i.e. real roots outside $[0,1]$
$0 < b_k, b_k^2 < c_k^2 \leq 1$	a term with no real roots, where b_k and c_k^2 can be interpreted as the first two moments of some marginal measure of the *k-th* root
$f_l^2 < g_l,$	all other terms, not having real roots.

Since the terms $Cx^p(1-x)^q\prod(x-a_i)^2$ are already non-negative, we have to deal only with the last three cases. First observe that since $d_j \notin [0,1]$, we have two cases to consider. Either $d_j < 0$, in which case $-d_j > 0$, and hence

$$R(x)\mathrm{sgn}(d_j)(d_j - x) = R(x)(x - d_j) = xR(x) + |d_j|R(x)$$

where both terms are non-negative. Otherwise, if $d_j > 1$ then

$$\begin{aligned} \mathrm{sgn}(d_j)(d_j - x) = d_j - x = 1 - 1 + d_j - x = (1-x) + (d_j - 1) \Rightarrow \\ R(x)\mathrm{sgn}(d_j)(d_j - x) = (1-x)R(x) + (d_j - 1)R(x), \end{aligned} \tag{4.12}$$

where both terms are non-negative again.

For the terms $x^2 + 2f_l x + g_l$ we extract the full square

$$x^2 + 2f_l x + g_l = x^2 + 2f_l x + f_l^2 + (g_l - f_l^2),$$

where all terms are positive, as $f_l^2 - g_l < 0$ is the discriminant of the initial quadratic equation.

Finally there is nothing to be done with the terms $(x^2+2b_k x+c_k^2)$, as, given the properties of b_k they are merely sums of terms like in (4.4), and hence

$$\begin{aligned} P(x) &= (x^2 + 2b_k x + c_k^2)\sum_s R_s(x) \\ &= \sum_s \left(x^2 R_s(x) + 2b_k x R_s(x) + c_k^2 R_s(x)\right), \end{aligned} \tag{4.13}$$

where all summands are of the form in (4.4).

Note that all steps but the last result in obtaining lower degree polynomials in the convex representation. The last step, given by (4.13), is not really needed at all, since b_k and c_k can be interpreted as the first and second moments of some marginal measure on the corresponding root(s).

Proposition 4.9 *In representation (4.6),* $0 \le \mu_{i,1}^2 \le \mu_{i,2} \le \mu_{i,1} \le 1, \forall i \Leftrightarrow$ *the root mixing measure yielding* $Q(x)$ *must have independence copula, i.e. must have the form (4.5).*

Proof. The necessity follows from the observation that should the marginal root measures $\eta_i(a_i)$ be coupled with the independence copula then:

$$\int \prod_i (x-a_i)^2 \prod_i \mathrm{d}\eta_i(a_i) = \int \left(\prod_i (x-a_i)^2 \mathrm{d}\eta_i(a_i) \right) = \tag{4.14}$$
$$\int \left(\prod_i (x^2 - 2xa_i + a_i^2)\mathrm{d}\eta_i(a_i) \right) = \prod_i (x^2 - 2x\mu_{i,1} + \mu_{i,2}),$$

and since the roots of the extreme densities lie on $[0,1]$ we have for the moments of the extreme densities $0 \le \mu_{i,1}^2 \le \mu_{i,2} \le \mu_{i,1} \le 1$, equalities between the moments corresponding to the measures being delta functions.

Conversely, as follows from the solution to the moment problem for $[0,1]$, there exists a probability measure η_i on $[0,1]$, such that two nonnegative numbers $\mu_{i,1}, \mu_{i,2}$ satisfying $0 \le \mu_{i,1}^2 \le \mu_{i,2} \le \mu_{i,1} \le 1$ are the measure's first and second moments. Coupling such measures η_i with the independence copula provides a Choquet probability measure that appears in the first integral of (4.14). ■

4.5 Applications

We envisage the main application of the polynomial densities on I to be in modelling of bounded random variables. The traditional ways of constructing such families with a large number of (relatively) free parameters are: (i) either performing mappings of unbounded variables onto $[0, 1]$, which rarely yields analytically tractable densities or distribution functions, or (ii) mixtures of already bounded models. In the case of a bounded model, its parameters will usually have to satisfy some non-linear restriction, ultimately due to the fact that the probability mass is concentrated on the bounded interval.

The situation is fundamentally easier with polynomials. Normalization of the density is only one linear condition on the coefficients. Throughout this section we assume that the relevant normalization constant of the polynomial density (or conditional polynomial density) can be computed according to Proposition 4.8 at the relevant stage of the numerical procedure. Non-negativity in general would be difficult. However knowing the explicit form of Choquet-extreme polynomial densities allows us to formulate several methods only in terms of the extreme densities, hence no additional conditions are needed to ensure non-negativity.

Firstly we note that if the mixing measure of the roots has a factor copula, then specification of such measure is possible in a simple and tractable way. We also observe that extreme polynomial densities are particularly attractive for the construction of non-parametric density estimators and to be used in the likelihood maximization procedure (ML) of parameter estimation (in this case, ML estimation of the set of the roots). Finally, calibration of the mixtures of the extreme densities is considered.

In this section we will also assume that we are given a sample $\mathbf{x} = \{x_j, j = 1 \dots N_S\}$ of numbers in $[0, 1]$, being the observation of i.i.d random variables, taking values on $[0, 1]$.

4.5.1 Factor Copula of the Root Measures

Section 4.3 established that if the mixing measure on the roots of the extreme densities has a factor copula, then to obtain the mixture polynomial density we need not specify the marginal root distribution functions explicitly. Instead of that, one only needs to know the first several moments of the marginal distribution functions of the roots, as implied by Corollary 4.7. Also, a simple and efficient way of computing the normalization constant will be available.

A natural generalization of the independence copula of the roots is a conditionally independence factor copula of the roots. Such copulas were defined in Section 2.1.

Proposition 4.10 *In the notation of (4.4), if all roots a_i have multiplicities 2 and mixed with with a factor copula, then the full joint measure of the roots is fully characterized by the first two moments of the conditional marginal distributions of the roots.*

Proof. Suppose the marginal root probability distributions are $\eta_i(a_i)$, and the factor copula of their mixing measure is given by

$$u(x_1, \ldots, x_n) = \int \prod_i F_i(x, \xi) \mathrm{d}G(\xi).$$

Then we have for the density of the joint distribution function

$$f(a_1, \ldots a_n) = \int \prod_i F_i(\eta_i(a_i), \xi) \mathrm{d}G(\xi),$$

and for the density

$$\partial f(a_1, \ldots a_n) = \int \prod_i \partial_\eta F_i(\eta_i(a_i), \xi) \eta_i'(a_i) \mathrm{d}G(\xi).$$

This implies that the polynomial resulting from mixing the extreme polynomials with such a measure will be

$$\begin{aligned} Q(x) &= Cx^m(1-x)^n \int \prod_i (x-a_i)^2 \partial f(a_1, \dots a_n) \\ &= Cx^m(1-x)^n \int \prod_i (x-a_i)^2 \int \prod_i \partial_\eta F_i(\eta_i(a_i), \xi)\eta_i'(a_i) \mathrm{d}G(\xi) \\ &= \int \mathrm{d}G(\xi) C(\xi) x^m(1-x)^n \int \prod_i (x-a_i)^2 \prod_i \mathrm{d}F_i(\eta_i(a_i), \xi) \\ &= \int \mathrm{d}G(\xi) C(\xi) x^m(1-x)^n \int \prod_i (x^2 - 2a_i x + a_i^2) \mathrm{d}F_i(\eta_i(a_i), \xi) \\ &= \int \mathrm{d}G(\xi) C(\xi) x^m(1-x)^n \prod_i \left(x^2 - 2\mu_{i,1}(\xi)x + a_i^2\mu_{i,2}(\xi)\right), \end{aligned}$$

where $\mu_{i,1}(\xi)$ and $\mu_{i,2}(\xi)$ are the first two conditional moments of the measure of the roots. ■

This simplifies the matter considerably: we do not need to know either the marginal root measures, or copula, let alone the multivariate distribution of the roots. All we need are the first two moments of the root distributions, conditional on factor ξ. In the above we denoted by $C(\xi)$ the normalization constant of the density

$$x^m(1-x)^n \prod_i \left(x^2 - 2\mu_{i,1}(\xi)x + a_i^2\mu_{i,2}(\xi)\right).$$

Such $C(\xi)$ can be interpreted as the normalization constant, conditional on ξ; hence our notation. Note that $C(\xi)$ will also be readily available using the procedure described in Proposition 4.8.

4.5.2 Non-parametric Estimators

Consider constructing the non-parametric estimators first. As a result of this subsection indicates, the main advantage of the non-parametric extreme polynomial density estimator is the simplicity of the numerical method to construct such estimator. Such estimator can be mixed later with another estimator using the method, described in Section 4.5.4 below.

Expression (4.4) implies that the minima of the probability density are at the roots that we are trying to estimate. Therefore, a most naive but straightforward estimator would be the one that has the roots somewhere between the sample points, e.g.

$$Q(x|\mathbf{x}) = C\prod_{j=2}^{N_S}\left(x - \frac{x_j + x_{j-1}}{2}\right)^2. \tag{4.15}$$

Despite the simplicity of the estimator, it has an apparent disadvantage: if several x_j in the sample happen to be located close to each other ("clustered" in other words), then the resulting polynomial will oscillate around them, having several minima, while we probably would want the polynomial to have one maximum around the "most probable" point.

Therefore, an alternative idea would be to put the maxima of the density in x_i or in the middle of the clusters of x_i if such can be identified. Consider the family

$$Q(x,\mathbf{u}) = C\prod_{j=1}^{N_u}(x-u_j)^2 = C(\mathbf{u})\prod_{j=1}^{N_u}(x-u_j)^2, \tag{4.16}$$

such that $N_u \leq N_s, .\mathbf{u} = \{u_1, \ldots, u_{N_u}\}$. Then

$$\partial_x Q(x, \mathbf{u}) = 2C(\mathbf{u}) \sum_{j=1}^{N_u} (u_j - x) \prod_{k \neq j} (x - u_k)^2 =$$
$$= 2 \sum_{j=1}^{N_u} \frac{Q(x, \mathbf{u})}{(u_j - x)} = 2Q(x, \mathbf{u}) \sum_{j=1}^{N_u} \frac{1}{(u_j - x)}. \tag{4.17}$$

This expression is intuitive: $Q(x, \mathbf{u})$ has minima at its roots, while locations of the maxima are given by the condition

$$\sum_{j=1}^{N_u} \frac{1}{(u_j - x)} = 0 \Leftrightarrow \sum_{j=1}^{N_u} \prod_{k \neq j} (u_j - x) = 0, \tag{4.18}$$

the equivalence being due to the fact that we cannot have a maximum in a root of $Q(x, \mathbf{u})$.

The values of u_j can now be obtained from the requirement that x_i are the roots of (4.18), in other words

$$\sum_{j=1}^{N_u} \prod_{k \neq j} (u_j - x) = \prod_{l=1}^{N_u} (x - \widetilde{x}_l), \tag{4.19}$$

where $\{\widetilde{x}_l\}$ are the points where we would like to place the maxima of the density. In general, this yields a system of non-linear equations for u_j. In case we solve for just three roots u_1, u_2, u_3 the solution will be available analytically, as (4.18) is quadratic form in $(u_j - x)$. We would like to stress that practically it would be preferable to first aggregate the sample points into clusters, and then solve (4.19) for example by setting $\widetilde{x}_l =$ "average in the *l-th*" cluster. The limiting case is obtained by setting $N_u = N_s$ and $\widetilde{x}_l = x_l$

Note that by construction of (4.16), it will always have a maximum inside $[0, 1]$ if $N_u > 1$ and the roots are different.

4.5.3 Max-likelihood Estimator for the Extreme Density Roots

Through this section we assume that the normalization constant can be computed according to Proposition 4.8 at the relevant stage of the numerical procedure.

When dealing with ML estimation of the density parameters, it is clear that the polynomial form of the density makes it more useful to work with the likelihood itself, not log-likelihood. The reason is that the log likelihood function of a polynomial density is not differentiable everywhere, which would unnecessarily complicate calculation.

Staying with parametrization (4.16) first, the likelihood of an extreme density given the sample x_i will be

$$L = C(\mathbf{u})^{N_s} \prod_{l=1}^{N_S} \prod_{j=1}^{Nu} (x_l - u_j)^2. \tag{4.20}$$

Given the analytical tractability of the likelihood, the straightforward way to locate the maxima is to compute the gradient of the above expression and solve the resulting system of equations for the critical points. The $i-th$ component of the gradient of L with respect to $\mathbf{u}$ is (denoting $\partial_i \triangleq \partial_{u_i}$)

$$\begin{aligned} \partial_i L(x) = N_s C(\mathbf{u})^{N_s - 1} \partial_i C(\mathbf{u}) \prod_{l=1}^{N_S} \prod_{j=1}^{Nu} (x_l - u_j)^2 - \\ 2C(\mathbf{u})^{N_s} \sum_{k=1}^{N_s} (x_k - u_i) \prod_{l=1}^{N_S} \prod_{j \neq i}^{Nu} (x_l - u_j)^2 = \\ = N_s \frac{\partial_i C(\mathbf{u})}{C(\mathbf{u})} L - 2 \sum_{k=1}^{N_s} \frac{L}{(x_k - u_i)} = L \left(\frac{\partial_i C(\mathbf{u})}{C(\mathbf{u})} - 2 \sum_{k=1}^{N_s} \frac{1}{(x_k - u_i)} \right). \end{aligned} \tag{4.21}$$

The last equation has an appealing interpretation. The leading L accounts for all local minima, while the value in the parenthesis can be viewed as a Lagrangian, $\partial_i C(\mathbf{u})/C(\mathbf{u})$ playing the role of the constraint on the set of feasible $\mathbf{u}$. It is this term that complicates the problem somewhat, which otherwise would be a simple algebraic equation for u_i. We also note that the second term in the parenthesis is exactly the same as (4.19), if we consider the x_k as the roots of (4.19). This makes perfect sense, since the difference between the non-parametric estimator proposed and MLE is, broadly speaking, that the former cares only about the location of the local maxima of the density, respectively of the values of those maxima, so the normalization constant is assumed constant. In MLE estimator, the values at the local maxima matter, so we are effectively solving the conditional extremum problem.

Write

$$\begin{aligned} P(x,\mathbf{u}) &= \prod_j (x-u_j)^2 = Q(x,\mathbf{u})/C, \\ P^{-k}(x,\mathbf{u}) &= \prod_{j\neq k} (x-u_j)^2 = P(x,\mathbf{u})/(x-u_k)^2, \\ \mu_n^{-k} &= \int_0^1 x^n \prod_{j\neq k}(x-u_j)^2 \mathrm{dx} = \int_0^1 x^n P^{-k}(x,\mathbf{u})\mathrm{dx}. \end{aligned}$$

Therefore if we have already computed the coefficients of $P(x,\mathbf{u})$, then we can easily obtain those for $P^{-k}(x,\mathbf{u})$ and compute all μ_n^{-k} by term-wise integration.

Then, since L cannot have maxima at the roots of the density, we can cancel L in (4.21) and rewrite it as

$$\begin{aligned} \partial_i L(x) &= \frac{\partial_i C(\mathbf{u})}{C(\mathbf{u})} - 2\sum_{k=1}^{N_s} \frac{1}{(x_k-u_i)} = 2\frac{C^2(\mathbf{u})}{C(\mathbf{u})}(\mu_1^{-i} - u_i\mu_0^{-i}) - 2\sum_{k=1}^{N_s}\frac{1}{(x_k-u_i)} = \\ &= 2\left(\frac{\mu_1^{-i} - u_i\mu_0^{-i}}{\mu_2^{-i} - 2u_i\mu_1^{-i} + u_i^2\mu_0^{-i}} - \sum_{k=1}^{N_s}\frac{1}{(x_k-u_i)}\right). \end{aligned} \tag{4.22}$$

Thus the critical points can be obtained, in principle, by solving the system of algebraic equations (4.22). Note that both analytical gradient and Hessian are available for this system, and computing them analytically is faster than doing numerical differentiation.

One can expect from the form of (4.22) that the system may have multiple solutions. Since L is obviously *Lipschitz*, if N_s is not too large (˜10), it may be more beneficial to use a deterministic global optimization method (e.g. a "Direct" algorithm for box-constrained problems; see [CHB$^+$01] for a review) to localize the global maxima of L first or to use "Interval Arithmetic" to analyze (4.22) directly using "divide and conquer"; and then to solve (4.22) locally.

Despite the apparent complexity of the optimization problem at hand, some qualitative analysis is possible due to the symmetry of (4.20). Indeed, writing

$$P(u) = \prod_{l=1}^{N_S} (x_l - u)^2, \tag{4.23}$$

we can rewrite (4.20) as

$$L = C(\mathbf{u})^{N_s} \prod_{j=1}^{Nu} P(u_j), \tag{4.24}$$

thus recognizing that (4.20) can also be viewed as the product of N_u univariate $2N_J-$order polynomials in each u_i, with roots x_l, and the constant $C(\mathbf{u})^{N_s}$, dependent on a $\mathbf{u}$. So for a given $\mathbf{u}$, L is proportional to exactly the same polynomial $P(u)$ evaluated in the components of $\mathbf{u}$. The components of the optimal $\mathbf{u}$ are not the maxima of $P(u)$, because of the presence of $C(\mathbf{u})$.

This representation hints that there should exist two extreme configurations of $\mathbf{u}$: when all components are same, and when they are all different. The first case where $u_1 = \ldots = u_{N_u} = u$ admits some further analysis. Write

$$L = C(u)^{N_s} \prod_{l=1}^{N_S} (x_l - u)^{2N_u}. \tag{4.25}$$

Proposition 4.11 *If, for a given sample, N_u is such that*

$$x_{\min} > \frac{1}{2N_u + 1}, \tag{4.26}$$

then (4.25) has a minimum on $[0, 1]$.

Proof. This can be proved by direct calculations: differentiating (4.25) and finding the conditions on the derivative changing signs at the ends of $[0, 1]$. It is important for the prove that $u, x_l \in [0, 1]$. ■

Corollary 4.12 *For a sufficiently large sample of continuous* $[0, 1]$ *random variables, an optimizer to (4.25) exists on* $[0, 1]$ *for fixed N_u with probability* 1.

The feasibility of non-numerical analysis of the general case requires further exploration.

4.5.4 Mixture Calibration

In the previous section we concentrated on the MLE approach to determine the parameters of the extreme density. The main motivation to use the extreme density, as opposed to a mixture directly, was the ability to keep the corresponding numerical procedure only box-constrained: an extreme density is automatically non-negative on $[0, 1]$, while if we allow terms with a linear term we will have to consider cases in a fashion similar to that of Section 4.4 and if needed we would have to incorporate inequality constraints to assure non-negativity.

We first observe that neither likelihood, nor log-likelihood functions appear useful in analyzing the extremal properties of a convex mixture.

Recall the following fact from the convex analysis.

Lemma 4.13 *If $f(x)$ is concave, i.e. $f''(x) \leq 0$, then $g(x) = \exp(-f(x))$ is convex and the maximum of $f(x)$ is the minimum of $g(x)$.*

Proof. Follows from $g' = -f' \exp(-f)$, $g'' = \exp(-f)\left(f'^2 - f''\right) \geq 0$. ∎

We therefore find the following function of likelihood useful in our situation:

$$\Lambda(\xi, \mathbf{x}) = \exp\left(-L(\xi, \mathbf{x})\right) = \exp\left(-\prod_{i=1}^{N_S} f(x_i, \xi)\right). \tag{4.27}$$

Obviously, the critical point of $L(\xi, \mathbf{x})$ is the critical point of $\Lambda(\xi, \mathbf{x})$ and in case $L(\xi, \mathbf{x})$ is concave then $\Lambda(\xi, \mathbf{x})$ is convex. Therefore the maximizer of $L(\xi, \mathbf{x})$ will be the minimizer of $\Lambda(\xi, \mathbf{x})$.

Now observe that if we have a family of parametric families $f_s(x, \varphi_s)$, where φ_s are the parameter vectors of the $s - th$ family, then for the Λ of the convex combination,

$$f(x, \mathbf{\Phi}) = \sum_s \lambda_s f_s(x, \varphi_s), suchthat\ \lambda_s \in [0, 1], \sum_s \lambda_s = 1, \mathbf{\Phi} = \bigcup_s \varphi_s$$

we have

$$\begin{aligned}\Lambda(\mathbf{\Phi}, \mathbf{x}) &= \exp\left(-\sum_s \lambda_s f_s(\mathbf{x}, \varphi_s)\right) = \prod_s \exp\left(-\lambda_s f_s(\mathbf{x}, \varphi_s)\right) = \\ &= \prod_s (\varphi_s, \mathbf{x})^{\lambda_s}\end{aligned} \tag{4.28}$$

Suppose we have obtained the MLE parameter estimates for all families $f_s(x, \varphi_s)$. This will imply that each $\Lambda_s(\varphi_s, \mathbf{x})$ will be minimized by its own MLE parameter estimate. Now, (4.28) implies that

$$\nabla_{\varphi_s} \Lambda_s = 0, \ \forall s \ \Rightarrow \nabla_{\mathbf{\Phi}} \Lambda(\mathbf{\Phi}, \mathbf{x}) = 0,$$

i.e. the critical points of $\Lambda(\mathbf{\Phi}, \mathbf{x})$ are Cartesian products of the critical points of Λ_s. The final necessary condition for the minimum of $\Lambda(\mathbf{\Phi}, \mathbf{x})$ is (denoting by β the Lagrange multiplier):

$$\partial_{\lambda_s} \Lambda(\mathbf{\Phi}, \mathbf{x}) - \beta = \mathbf{0}, \, \forall s \Leftrightarrow$$
$$\partial_{\lambda_1} \Lambda(\mathbf{\Phi}, \mathbf{x}) = \partial_{\lambda_\mathbf{2}} \mathbf{\Lambda}(\mathbf{\Phi}, \mathbf{x}) = \ldots = \partial_{\lambda s} \Lambda(\mathbf{\Phi}, \mathbf{x}).$$

4.5.5 Pinned Density Moment Matching (PDMM)

Moment matching methods are widely used to create semi-parametric estimators of the L^2 functions; in particular, orthogonal moment matching, i.e. when the measures are orthogonal polynomials, is well researched.

However, the standard setup for these methods usually does not require strict non-negativity of the approximating polynomial(s). This makes the standard setup not very useful when the approximating polynomial is intended to be a probability measure.

The constructive characterization of extreme points of univariate polynomial measures, given the density of polynomial measures in $\mathcal{L}^2$ measures, allows one to solve the moment matching problem, at least in the $\mathcal{L}^2$ approximation sense. Furthermore, in addition to fixing the moments, it is possible to fix the value of the density function in the set of points.

Fix the type (p,q) of the density (4.4),

$$Q(x) = Cx^p(1-x)^q \prod_i (x^2 + 2b_i x + c_i^2),$$

such that $b_i^2 \le c_i$, the equality case for all terms yielding an extreme density, as implied by (4.4).

Suppose we are matching first N moments, $\mu_p, p = 1 \ldots N$. The moment matching system of equations can be written as

$$\begin{aligned} &\int x^n C x^p (1-x)^q \prod_i (x-u_i)^2 \mathrm{d}x = \\ &C \int x^{n+p}(1-x)^q \prod_i (x-u_i)^2 \mathrm{d}x = \mu_n \Rightarrow \end{aligned} \tag{4.29}$$

$$\int x^{n+p}(1-x)^q \prod_i (x-u_i)^2 \mathrm{d}x - \mu_n \int x^p (1-x)^q \prod_i (x-u_i)^2 \mathrm{d}x = 0, \tag{4.30}$$

with the normalizing constant cancelling out from all equations but for $n = 0$. Now all the integrals can be computed by term-wise integration of the product polynomials, the coefficients of which can be obtained in the same way as that of Proposition 4.8. Note that should a numerical method with analytical gradient be selected, the gradient (and Hessian) will be available analytically, as described in the Appendix.

As we have already noted, the polynomial form of the density allows generalizing the above moment matching problem by imposing extra restrictions on the density or the cumulative distribution function directly. E.g. we can require that the density has some prescribed value at a given point $Q(\xi) = \eta$, which gives rise to the following constraint:

$$C\xi^n(1-\xi)^m \prod_i (\xi - u_i)^2 - \eta = 0.$$

Such a generalization does not introduce any considerable computational overhead to the initial setup. The normalization constant will be computed as part of the above system on each step anyway.

4.5.6 A Final Note: Allowing Roots Outside $[0, 1]$

In this section we have presented several applied statistical methods that were explicitly using the characteristic representation of an extreme density (4.4). However, we utilized the fact that the roots of the density belong to $[0, 1]$ only once, when considering an extreme case of equal ML estimator for all the roots. Except for that, we never really need to bound the roots on $[0, 1]$. All methods considered will still work, as the normalization constant will ensure that the resulting non-negative polynomial is a density.

Note that going beyond $[0, 1]$ is a simpler alternative to using the representation of a non-extreme density

$$Cx^p(1-x)^q \prod_i (x^2 + 2b_i x + c_i). \tag{4.31}$$

Indeed, in the above we need to fulfill inequality constraints on c_i and b_i, while with (4.4) with roots going outside $[0, 1]$ there are no constraints. We know that the density (4.4) with roots outside $[0, 1]$ is necessarily not extreme, therefore the parametrization (4.4) without restriction on the roots can be regarded simply as a more tractable alternative to (4.31).

With such formulation, (4.19) can be solved with and without constraints on the roots and the better density approximator selected based on the ratio of likelihoods or χ^2 test. The same procedure can also be used to obtain an ML estimator.

4.6 Literature Review

As was noted in Section 3, the problem of identifying the Choquet-extreme polynomial distributions functions on I is new and appears not to have been discussed before.

There is vast literature on semialgebraic methods of characterizing polynomials, which are positive or non-negative on both bounded and unbounded intervals, however they are mostly valuable in the multivariate case. In univariate setting we note the works of Powers an Reznick, [PR00], [PR05].

There is a strong interplay between (a) the characterization of polynomials, non-negative on semialgebraic sets, and (b) moment problems, culminating in the Schmuedgen's result [Sch91], which is, ultimately, a generalization of the classic result characterizing the moment sequences on $\mathbb{R}$ in terms of Hankel determinants (see, e.g., [Wal04]), going back to Stieltjes.

Factor copulas are very popular in mathematical finance, especially in basket credit modelling. One of the pioneering articles on the subject is [LG05]. A notable source on using polynomial multiplication methods to convolve Bernoulli variables, coupled with factor copulas, is [LAB03].

The PDMM method appears to be novel; we were not able to find any comparable framework in the literature.

A construction similar to our exponential non-likelihood is utilized in [CH74], only however in the case of a mixture of two distributions and only from the hypothesis testing point of view (whether one of the additives in the mixture should be preferred to another).

5 Polynomial Densities on $[0, 1]^M$

In this section we extend results obtained in the previous section for the univariate polynomial densities over $[0, 1]$ to the case of polynomial densities over $[0, 1]^M$. Some results in this Section hold in the general multivariate case, however certain conditions for the extreme density are proven only for $M = 2$. We expect the generalization to the higher dimensional case should be straightforward and will have essentially the same qualitative results; however, the proof will be more technical.

Consider a general multivariate polynomial of bounded degree (introducing the multi-index $\mathbf{i} = \{i_1 \dots i_M\}$):

$$P(x_1, \dots, x_M) = \sum_{i_1 \dots i_M} \theta_{i_1 \dots i_M} \prod_{i=1}^{M} x_i^{n_i} = \sum_{\mathbf{i}} \theta_{\mathbf{i}} \prod_{i=1}^{M} x_i^{n_i}. \tag{5.1}$$

If it fulfils (2.2) - (2.4) on I^M, then it can be interpreted as a polynomial distribution function on I^M, which we will refer to as a *polynomial distribution function*. The partial derivative $\partial_{x_1, \dots, x_M} P(x_1, \dots, x_M)$ will be referred to as the *polynomial density*. In what follows we will omit references to I^M and will always assume we are talking about polynomial distribution functions or polynomial densities on I^M.

In this section we will be mostly interested in characterizing the Choquet-extreme polynomial densities. In what follows, we first establish an affine bijection between polynomial distribution functions and densities, the bijection preserving convexity. Similar to the univariate case, densities are easier to deal with, because the M increasing condition (2.4) is replaced by the non-negativity of the density. Since we are working with polynomials, non-negativity of a polynomial density can be analyzed using the tools of real algebraic geometry.

5.1 General Form and Affine Bijection

Imposing normalization and groundness conditions on (5.1) yields the following.

Proposition 5.1 *Polynomial distribution functions on I^M have the form (using the notation of (2.15))*

$$P(x_1, \ldots, x_M) = \prod_{i=1}^{M} x_i \left[1 + \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} \left(\prod_{k=1}^{M} x_k^{i_k} - 1 \right) \right]. \tag{5.2}$$

Proof. For the general polynomial (5.1), groundedness (2.2) implies that for any single index l, we set $x_l = 0$ then $P(\ldots, x_{l-1}, 0, x_{l+1}, \ldots) = 0$. This implies:

$$P = \sum_{i_1 \ldots i_M} \theta_{i_1 \ldots i_M} \prod_{k \neq l} x_k^{i_k} = \sum_{\substack{i_1 \ldots i_M, \\ i_l = 0}} \theta_{i_1 \ldots i_M} \prod_{k=1}^{M} x_k^{i_k} \equiv 0 \;\Rightarrow$$

$$\theta_{i_1 \ldots i_M} = 0 \text{ if any } i_l = 0.$$

Repeating the same argument for all $l = 1 \ldots M$ we obtain that (5.1) does not contain terms that would miss at least one x_l; hence all terms in (5.1) have a factor of each x_l, which is another way of saying that a polynomial distribution function must take the following form (up to relabeling of coefficients):

$$P(x_1, \ldots, x_M) = \prod_{i=1}^{M} x_i \left(\sum_{i_1 \ldots i_M} \theta_{i_1 \ldots i_M} \prod_{i=1}^{M} x_i^{n_i} \right). \tag{5.3}$$

Now, normalization implies that

$$
\begin{aligned}
P(1,\dots,1) = 1 = \sum_{i_1\dots i_M} \theta_{i_1\dots i_M} \Rightarrow \\
\theta_{0\dots0} = 1 - \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}}.
\end{aligned} \tag{5.4}
$$

In other words, (5.4) imposes a restriction on the free term in the brackets of (5.3). Plugging (5.4) into (5.3) yields:

$$
\begin{aligned}
P(x_1,\dots,x_M) &= \prod_{i=1}^{M} x_i \left(\sum_{i_1\dots i_M} \theta_{i_1\dots i_M} \prod_{i=1}^{M} x_i^{n_i} \right) = \prod_{i=1}^{M} x_i \left(\theta_{0\dots0} + \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} \prod_{i=1}^{M} x_i^{n_i} \right) \\
&= \prod_{i=1}^{M} x_i \left(1 - \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} + \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} \prod_{i=1}^{M} x_i^{n_i} \right) \\
&= \prod_{i=1}^{M} x_i \left[1 + \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} \left(\prod_{k=1}^{M} x_k^{i_k} - 1 \right) \right],
\end{aligned}
$$

which is (5.2). ∎

Corollary 5.2 *Differentiation of (5.2) establishes an affine bijection, i.e. it transfers the convexity characteristics of the set of polynomial distribution functions to the set of their densities.*

Proof. Differentiation is a linear operation, therefore a convex combination of distribution functions is a convex combination of densities.

For the converse, observe that an integral of a convex combination of a family of polynomials $R_k(x_1,\dots,x_M)$ will be

$$
\int \sum_k \lambda_i R_k(x_1,\dots,x_M) \mathrm{d}\mathbf{x} = \sum_k \lambda_i P_k(x_1,\dots,x_M) + \sum_{i=1}^{M} C_i \prod_{j=1}^{i-1} x_j,
$$

and Proposition 5.1 will imply that all the C_i are necessarily zero. ∎

Corollary 5.3 *If* $\mathbf{p} = \{p_1, \ldots, p_M\}$ *is a type of a distribution function then* $p_i > 0$, $i = 1 \ldots M$.

Proof. Follows from the fact that $\prod_{i=1}^{M} x_i$ divides the expression in (5.2). ■

Differentiating (5.2) with respect to all x_i yields for the density

$$\begin{aligned} \mathrm{d}P(x_1, \ldots, x_M) &= \mathrm{d}\left(\prod_{i=1}^{M} x_i \left[1 + \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} \left(\prod_{k=1}^{M} x_k^{i_k} - 1\right)\right]\right) \\ &= 1 + \sum_{\|\mathbf{i}\|>0} \theta_{\mathbf{i}} \left(\prod_{k=1}^{M} (i_k + 1) x_k^{i_k} - 1\right). \end{aligned} \tag{5.5}$$

Therefore we can characterize a polynomial density as a polynomial of the form (5.5), which is non-negative on I^M.

5.2 On Characterization of Extreme Densities

The result of the last section implies that characterizing the extreme polynomial densities is equivalent to characterizing the extreme polynomial distribution functions. That is, one needs to characterize polynomials, non-negative on I^M, and integrating on I^M to one. Characteristic representation of the polynomials that are non-negative on I^M is given ultimately by the PD Theorem (Theorem 2.29); however this theorem gives only a representation in terms of the rational functions. Only fulfilling the normalization requirement is trivial and amounts to rescaling by a constant. Furthermore, we immediately observe (as we did in the univariate case) that an extreme density should have a zero on I^M, which implies that it may be represented as the sum of squares of rational functions only, which is useless in our case. Therefore direct utilization of the available real algebraic results is not possible and we must attempt to generalize to higher dimensions the univariate approach based on the analysis in term of factorization into the irreducibles.

In this section we will be mostly using two results from real algebraic geometry that hold in two dimensions, i.e. for polynomials from $R[x, y]$, which are non-negative on I^2. The first is a specialization of a general Positivstellensatz (see, e.g., [PD01]).

Definition 5.4 *Given a polynomial* $f \in \mathbb{R}[x, y]$ *we define the zero set of* f *as* $V_{\mathbb{R}}(f) = \{(x, y) : \ f(x, y) = 0\}$.

Lemma 5.5 ([dlP02], Real Study Lemma) *Given that* $f, h \in \mathbb{R}[x, y]$ *are of positive degree, such that* f *is irreducible in* $R[x, y]$ *and indefinite and* $V_{\mathbb{R}}(f) \subseteq V_{\mathbb{R}}(h)$ *then* $f|h$.

Another one is an example of a distinguished representation.

Lemma 5.6 ([Sch05], Corollary 3.8) *Let* $h_1, \dots h_n$ *be linear polynomials in* $\mathbb{R}[x, y]$ *such that the convex polygon* $K = \{h_1 \geq 0, \dots, h_n \geq 0\}$ *is compact with non-empty interior. Let* f *be a polynomial which is non-negative on* K. *If* f *does not vanish in any vertex of* K, *then* f *has a representation.*

$$f = s_0 + \sum_{i=1}^{n} s_i h_i, \tag{5.6}$$

where $s_i, i = 0...n$ *are sums of squares of polynomials.*

5.2.1 A Sufficient Condition

In this section we restrict our consideration to the two dimensional case, i.e. we consider a polynomial from $\mathbb{R}[x, y]$.

Lemma 5.7 *Suppose that* $P \in \mathbb{R}[x, y]$ *is such that*

- *P^{2n} is of a type* $\mathbf{p}$*, for some* $n > 0$,
- *P is indefinite and irreducible in* $\mathbb{R}[x, y]$,
- *the algebraic curve defined by P crosses the interior of I^2, i.e. $V(P) \cap I^2 \neq \varnothing$.*

Then the density $A \cdot P^{2n}$ is extreme, where A is a constant.

Proof. Since P is indefinite and $V(P) \cap I^2 \neq \varnothing$ then this intersection is infinite.

Suppose that $P^{2n} = \lambda q + (1 - \lambda)r$, $\lambda \in [0, 1]$.

First observe that both $V_{\mathbb{R}}(P) \subseteq V_{\mathbb{R}}(q)$ and $V_{\mathbb{R}}(P) \subseteq V_{\mathbb{R}}(r)$; otherwise $\lambda q + (1 - \lambda)r$ would have only finite number of the intersection points by the Bezout theorem. But then, by the Real Study Lemma, $P|q$ and $P|r$. Therefore $P^{2n-1} = \lambda q/P + (1 - \lambda)r/P$. Repeating this reasoning $2n$ times we obtain that $P^{2n}|q$ and $P^{2n}|r$ The lemma now follows because P, q and r are densities, and hence should integrate to 1. ■

Corollary 5.8 *Suppose that a polynomial P of type* $\mathbf{p}$ *can be represented as*

$$P = \prod_i P_i^{2n_i}, \tag{5.7}$$

where each $P_i \in \mathbb{R}[x, y]$ is indefinite and irreducible in $\mathbb{R}[x, y]$ and the algebraic curve defined by P_i crosses the interior of I^2: $V(P) \cap I^2 \neq \varnothing$. Then the density $A \cdot P$, is extreme, where A is a constant.

5.2.2 Necessary Conditions

Recall that in the univariate case an extreme density was a product of terms x, $(1 - x)$, and $(x - a)$, which are $\mathbb{C}$-irreducible polynomials with real coefficients.

In this section we partly generalize this observation to the multivariate case.

Lemma 5.9 ([dlP02, Lemma 23]) *Suppose that $f \in \mathbb{R}[x_1, \ldots, x_n]$ is irreducible in $\mathbb{R}[x_1, \ldots, x_n]$. Then f is reducible in $\mathbb{C}[x_1, \ldots, x_n]$, if and only if either f or $-f$ is a sum of two squares in $\mathbb{R}[x_1, \ldots, x_n]$.*

This immediately implies the following.

Proposition 5.10 *If $P \in \mathbb{R}[x_1, \ldots, x_n]$ is $\mathbb{R}$-irreducible, but $\mathbb{C}-$reducible then $P(x) \notin \mathcal{E}(\mathbf{p})$.*

Thus the building blocks of the factorization of the multivariate extreme polynomial can only be $\mathbb{C}$-irreducible polynomials with real coefficients.

Another basic observation in the univariate case was that a polynomial of the degree lower than the one we were considering could not be extreme (Lemma 4.3). In the multivariate case it is clear than if we fix the degree N and dimension M, then the allowed powers $p_i, i = 1 \ldots M$ for the arguments of the density must fulfil:

$$\sum_{i=1}^{M} p_i = N, \text{ such that} p_i > 0. \tag{5.8}$$

The whole set of multivariate polynomials of degree N will split into the classes, closed under the formation of convex combinations. Each such class can be identified by the *type,* as defined in Section 2.4, fulfilling the above condition. By construction, a convex combination of elements of different types cannot be an extreme density of degree N, therefore in what follows we can assume that the type $\mathbf{p}$ is fixed. Given two types $\mathbf{p}_1$ and $\mathbf{p}_2$, we will write $\mathbf{p}_1 \leq \mathbf{p}_2$ if $p_{i,1} \leq p_{i,2}, \forall i = 1 \ldots M$ (the second index being the index of the whole tuple) and $\mathbf{p}_1 < \mathbf{p}_2$ if the strict inequality holds coordinate-wise. The extreme set of densities of type $\mathbf{p}$ will be denoted by $\mathcal{E}(\mathbf{p})$.

Now we can formulate generalizations of Lemma 4.3 to the multivariate case.

Proposition 5.11 *If the type* $\mathbf{p}$ *of a density is such that* $\sum_{i=1}^{M} p_i < N$ *then it is not extreme.*

This can be proved by considering exactly the same construction as in the one-dimensional case in any variable; see proof of Lemma 4.3. Such a construction will keep the degree of the "tilted" multivariate density less than or equal to N.

Proposition 5.12 *If* $P(\mathbf{x}) > 0$ *on* I^M *then* $P \notin \mathcal{E}(\mathbf{p})$.

Proof. The case when $N = 0$, for which $P(\mathbf{x}) \equiv C$, is trivial, as the whole set of polynomial densities only contains one polynomial. Otherwise, assuming that the highest power of x_1 is positive (which will be fulfilled, perhaps after renumbering the arguments), consider $Q_{\pm}(\mathbf{x}) = P(\mathbf{x}) \pm (Cx_1 + D)$. The constants C and D need to be selected such that $Q_{\pm}(\mathbf{x}) \geq 0$, which can be done by requiring that $\min Q_{\pm}(x) = 0$ and $\int (Cx_1 + D)\mathrm{d}x = 0$. This yields a system with two linear equations with two unknowns C and D that always has a solution . ■

The next conditions are essentially two-dimensional.

Proposition 5.13 *If* $P(x, y) \in \mathbb{R}[x, y]$ *has only a single isolated zero in* I^2 *then* $P(x, y) \notin \mathcal{E}(\mathbf{p})$.

Proof. Withoutessential loss of generality, suppose that the zero is located at $(0,0)$. Rewriting the polynomial P in terms of the homogenous polynomials

$$P(x,y) = \sum_i \sum_{j=1}^{i} \theta_{i,j} x^i y^{j-i},$$

consider the value of $P_n(x,y)$ on the line $y = \beta x$:

$$P(x,\beta x) = \sum_i x^i \sum_{j=1}^{i} \theta_{i,j} \beta^{i-j}.$$

Since the polynomial has an single isolated real zero at $(0,0)$, it does not have any other zeroes, for all β. Therefore, considered as a polynomial in β (for the fixed x), it can be decomposed into the product of positive definite quadratic factors:

$$P(\beta) = c\, x^2 \prod_k (\alpha_{0,k}(x) + \beta \alpha_{1,k}(x) + \beta^2 \alpha_{2,k}(x)),$$

where $\alpha(x)$ are some polynomial functions. Substituting $\beta = y/x$ we immediately obtain that the original polynomial can be factorized and some of the factors are positive definite, which implies that the original polynomial is not extreme, which follows from Proposition 5.12. ■

Proposition 5.14 *If $P(x,y) \in \mathbb{R}[x,y]$ is extreme and has only a finite number of isolated zeroes on I^2 excluding the corners, then $P(x,y)$ can have zeroes only on the border of I^2.*

Proof. By Lemma 5.6, we have

$$P(x,y) = s_0 + \sum_{i=1}^{4} s_i h_i, \tag{5.9}$$

where $h_i \in \{x, y, 1-x, 1-y\}$,

and $s_i, i = 0, \ldots, 4$ are sums of squares.

If this representation contained more than one term than $P(x,y)$ then the $P(x,y)$ would not be extreme. Therefore, to fulfil our assumption, the above representation can only have one summand.

The representation of the form $P(x,y) = s_i h_i$, would mean that $P(x,y)$ has an isolated zero in the corner of I^2, which is prohibited by the theorem assumption. Therefore we can only have $P(x,y) = s_0 = r^2$, a single square, having several isolated zeroes inside I^2 except the corners.

Now we have three possibilities (since the case of an isolated zero in the corner is excluded):

1. r changes sign inside I^2,
2. r changes sign on the border of I^2,
3. r does not change sign on I^2.

If r does change the sign then r^2 cannot have isolated zeroes inside I^2, so case 1 is not possible. Consider case 3, and assume that it is true. In other words, r is non-negative on I^2 and has a finite number of isolated zeroes on I^2 except the corners. Clearly, if P were extreme, so should be r; otherwise, if r is not extreme, then $P = r^2$ is surely not extreme.

Therefore r fulfils the conditions of the lemma therefore we can apply all logic starting with decomposition (5.9) to it. This construction can be repeated infinitely many times, due to our assumption that P was extreme, and no minimal polynomial q, such that $P = q^{2^n}$ can be established, because on each step we should be able to repeat our construction. Thus case 3 yields a contradiction. Therefore only case 2 is possible, which implies that $P(x, y)$ will have an isolated zero on the border of I^2, except the corners. ■

5.2.3 Conjecture on Characterization

Consider a factorization of a polynomial density $P(x, y)$ into the product of powers of coprime polynomials with real coefficients, irreducible over $\mathbb{C}$:

$$Q = C \prod_i R_i^{n_i}. \tag{5.10}$$

We have established that all R_i must have zeroes on I^2 and if R_i changes the sign on I^2 then n_i should be even.

Consider a convex combination of polynomials that have factorizations of the form (5.7):

$$\lambda \prod_i P_i^{2n_i} + (1 - \lambda) \prod_j Q_j^{2n_j}.$$

Since all P_i and Q_j are irreducible over $\mathbb{C}$, we have two options: either the algebraic curves $V_{\mathbb{C}}(P_i)$ and $V_{\mathbb{C}}(Q_j)$ have common components or, by the Bezout theorem, they intersect in a finite number of complex points. Therefore, they will either have common real components or will intersect in a finite number of real points. In the former case, the Real Study Lemma will imply that $P_i{=}Q_j$ and $n_i = m_j$.

This observation indicates two difficulties of characterizing the extreme polynomial densities in the multivariate case compared to the univariate one. Firstly, the fact that a factor R_i in (5.10) is non-negative and has only a finite number of zeros on I^2 does not necessarily imply that it is a sum of squares. It does in the case when it has only one isolated root. But, as we figured out above, it may be extreme and have several isolated roots on the boundaries except the corners, or it may theoretically have several roots in the corners.

Another difficulty is that R_i may be non-negative on I^2 and define a non-trivial real algebraic curve on I^2, but it need not be some of squares. It will only belong to the preordering $T_{\{x,1-x,y,1-y\}}(x,y)$, i.e. it may be represented as sums of $h_i s_1^2/s_2^2$ or just s_1^2/s_2^2.

Therefore only a partial characterization could be formulated as follows.

Conjecture 5.15 (Multivariate Characterization) *Consider a factorization of $Q(x,y)$, which has order N, into components that are polynomials with real coefficients, irreducible over $\mathbb{C}$:*

$$Q = C \prod_i P_i^{2p_i} \prod_j R_{j,}^{q_j} \tag{5.11}$$

where $P_i(x,y)$ changes sign inside I^2 and R_j either does change sign on the interior of I^2 or changes sign on its border. Then Q is extreme if and only if the following holds:

1. R_j must have zeroes in I^2, but they cannot have a single isolated zero on I^2 and if they have a finite number of isolated zeroes on I^2 except its corners, then these zeroes must be on the border.

2. R_j cannot be represented as a sum of other polynomials, non-negative on I^2.

5.3 Literature Review

The problem considered in this section appears to be essentially new. We could not find any publications on this matter.

The nearest research subject available in the literature is the characterization of extreme semi-definite forms, e.g. as in [CL77]. The setup there is however much less restrictive, as the authors do not bound the degree of the form. The approach and results seem to be quite specific and the work does not appear to have any major follow ups.

At the same time, there is a vast literature on the characterization of polynomials (of unbounded degree) that are either definite or semi-definite of either bounded or unbounded semialgebraic sets. The two main research subjects and results, somewhat relevant to our subject, can be categorized into two groups:

- General characterizations, given by Theorems like 2.28 or 2.29, sometimes generalized for a general rings and cones (see, e.g. section on "formal" Positivstellensatz in [JBR98]).

- Characterizations of the cases when polynomials that are definite or semi-definite on the semialgebraic sets can be represented in terms of (2.22), and not (2.23). See [PD01] for details.

Corollary 3.10 from [Sch05] was cited in Section 3 and is most relevant for our consideration.

6 Polynomial Copulas

This section is devoted to studying polynomial copulas and their densities. As discussed in the introduction, copulas provide a way to construct (quantile) couplings of random variables. However, due to the poor analytical tractability of multivariate copulas, only factor copulas attracted attention from the practical perspective. In this section we show that polynomial copulas can be viewed as a generalization of factor copulas the way we defined them in Section 3, which makes them attractive in all applications where factor copulas would be. In addition to that, their polynomial nature allows one to simplify the characterization of the set of polynomial copulas compared to that of factor copulas, in particular, in the sense of Choquet extremality.

In line with the emphases of the current work, we first consider the characterization of extreme polynomial copula densities of bounded degree and then analyze the application of polynomial copulas to convolution. We also introduce a new class of *reflexive* copulas.

Definition 6.1 *Polynomial copulas are polynomials that fulfil conditions (2.2)-(2.6).*

In this section we will continue to use the notation we used for the general polynomial distribution functions. Having fixed the polynomial's degree N and type $\mathbf{p}$, we write

$\mathfrak{C}^N_{poly}(\mathbf{p})$ for the set of polynomial copulas of degree N and type $\mathbf{p}$, and

$\partial\mathfrak{C}^N_{poly}(\mathbf{p})$ for the set of densities of the elements of $\mathfrak{C}^N_{poly}(\mathbf{p})$.

6.1 General Form and Affine Bijection

Firstly we extend Proposition 5.1. Being polynomial distribution functions, polynomial copulas satisfy the conditions of that proposition, however the fact that margins are uniform narrows down the representation.

Proposition 6.2 *Polynomial copulas have the form (using the notation for* $\|\mathbf{i}\|$ *and* $\mathbf{i}^{(j)}$*from (2.15)):*

$$C(x_1,\ldots,x_M)=\prod_{i=1}^{M}x_i\left[1+\sum_{\|\mathbf{i}\|>0,\ \ |\mathbf{i}|>1}\theta_{\mathbf{i}}\left(\prod_{k=1}^{M}x_k^{i_k}-\sum_{k=1}^{M}x_k^{i_k}+M-1\right)\right]. \tag{6.1}$$

Proof. Reproduction of the margins implies that if in (5.2) we set all x_i but one (say x_j) to 1 then we obtain x_j, i.e.

$$x_j=x_j\left[1+\sum_{\|\mathbf{i}\|>0}\theta_{\mathbf{i}}\left(x_j^{i_j}-1\right)\right] \tag{6.2}$$

$$=x_j\left[1+\sum_{i_j}\sum_{\|\mathbf{i}\|>0,\ \ \mathbf{i}^{(j)}=i_j}\theta_{\mathbf{i}}\left(x_j^{i_j}-1\right)\right]$$

$$=x_j\left[1+\sum_{i_j}\left(x_j^{i_j}-1\right)\sum_{\|\mathbf{i}\|>0,\ \ \mathbf{i}^{(j)}=i_j}\theta_{\mathbf{i}}\right]\Rightarrow\sum_{\|\mathbf{i}\|>0,\ \ \mathbf{i}^{(j)}=i_j}\theta_{\mathbf{i}}\equiv 0.$$

The latter equality means the following. Given the index j, on the left-hand side we first fix the value of the j-th coordinate of the multi-index at i_j, and then sum over all other realizations of the coordinates in the multi-index. This gives us the sum over all $\theta_{\mathbf{i}}$ with one value of the coordinate vector fixed. Thus we can solve this equation for one of such $\theta_{\mathbf{i}}$. For example, if we pull out $\theta_{\mathbf{i}^{(j)}}$, i.e. the coefficient where the $j-$th coordinate is i_j, and all others are zero, then we can write:

$$\theta_{\mathbf{i}^{(j)}}=-\sum_{|\mathbf{i}|>1,\ \ \mathbf{i}^{(j)}=i_j}\theta_{\mathbf{i}}.$$

In other words, on the right hand side we are summing over all indices, with fixed positive $j-$th coordinate and at least one other coordinate larger than zero.

The coefficient $\theta_{\mathbf{i}(j)}$ corresponds to the term $x_j^{\mathbf{i}(j)} - 1$ in the brackets of (6.2), so collecting the terms yields (6.1). ■

Note that the above proposition implies that Proposition 5.3 also holds for polynomial copulas.

Also, the proposition implies that polynomial densities have the form

$$\mathrm{d}C(x_1, \ldots, x_M) = 1 + Q(x_1, \ldots, x_M), \tag{6.3}$$
$$\text{where } Q(x_1, \ldots, x_M) \geq -1 \text{ on } I^M,$$
$$\text{and} \int_0^1 Q(x_1, \ldots, x_M)\mathrm{d}x_i = 0.$$

The set of polynomial copulas/densities is not empty.

Example 6.3 *FGM density:*

$$c(x, y) = 1 + \theta(1 - 2x)(1 - 2y).$$

Proposition 6.4 *Let $G(u, v)$ be a copula. Then*

$$C(u, v) = uv\,(1 + \theta(1 - u)(1 - v)G(u, v))$$

is also a copula if $\theta \in [0, 1/2]$.

Proof. It is clear that $C(u, v)$ is grounded and reproduces the margins; therefore it is sufficient to check that the second cross derivative is greater than or equal to zero. Direct calculations yield

$$C_{uv} = 1 + \theta\{\begin{Bmatrix} (1 - 2u)(1 - 2v)G(u, v) + u(1 - u)v(1 - v)G_{uv}(u, v) + \\ (1 - 2u)v(1 - v)G_v(u, v) + u(1 - u)(1 - 2v)G_u(u, v) \end{Bmatrix}.$$

Since $G(u, v)$ is a copula, $G(u, v), G_u(u, v), G_v(u, v) \in [0, 1], G_{uv} \geq 0$ (see [Nel06], Theorem 2.2.7). Hence

$$\begin{aligned} &(1-2u)(1-2v)G(u,v) \geq -1, \\ &u(1-u)v(1-v)G_{uv}(u,v) \geq 0, \\ &(1-2u)v(1-v)G_v(u,v) \geq -1/4, \\ &u(1-u)(1-2v)G_u(u,v) \geq -1/4, \end{aligned} \tag{6.4}$$

and $1 + \theta[\ldots] \geq 1 - 2\theta$, which is greater than 0 once $\theta \in [0, 1/2]$. ■

6.2 On Extreme Polynomial Copula Densities

In this section we will use the form (6.3) of a polynomial copula density.

6.2.1 A Necessary Condition

First we note an elementary necessary condition for a polynomial copula density to be extreme.

Proposition 6.5 *An extreme density has a zero in* I^M $(\rho(x) \in \mathcal{E}(\partial\mathfrak{C}^n_{poly}(\mathbf{p})) \Rightarrow \exists \eta \in I^M : \rho(\eta) = 0)$.

Proof. Suppose that $\rho(\mathbf{x}) > 0$ on I^M. Set $\mu = \inf(\rho(\mathbf{x}) : \mathbf{x} \in I^M)$ and consider

$$\widetilde{\rho}_{\pm}(\mathbf{x}) = \rho(\mathbf{x}) \pm \mu \prod_i (x_i - 1/2).$$

Since $\rho(\mathbf{x})$ is a copula density, so is $\widetilde{\rho}_{\pm}(\mathbf{x})$. This is because $\widetilde{\rho}_{\pm}(\mathbf{x}) \geq 0$ on I^M and

$$\widetilde{C}_{\pm}(\mathbf{x}) = \int_0^{\mathbf{x}} \widetilde{\rho}_{\pm}(\mathbf{u})\mathrm{d}\mathbf{u} = C(\mathbf{x}) \pm \frac{\mu}{2^n} \prod_i x_i(x_i - 1)$$

is grounded and reproduces the margins.

However $\rho(\mathbf{x}) = (\rho_+(\mathbf{x}) + \rho_-(\mathbf{x}))/2$, so $\rho(\mathbf{x})$ is not extreme. Hence the proposition's claim follows by contradiction. ■

6.2.2 Characterization and Duality

A characterization of the extreme polynomial copula densities can be obtained in terms of a duality relationship that we derive in this section. We construct the supporting hyperplane using the functional from the Monge-Kantorovich (MK) problem. Note that in this section we consider polynomial copulas as a subset of some functional Banach space, as opposed to merely being an element of a polynomial ring; see Remark 2.15.

First we consider a more trivial case, relaxing the requirement that $c(\mathbf{x}) \geq 0$. Consider a generalized Monge-Kantorovich (MK) problem, restricted to the polynomial densities on I^M:

$$\begin{gathered} \text{find } \rho(\mathbf{x}) \in \partial\mathfrak{C}^N_{poly}(\mathbf{p}), \\ \text{which minimizes } I(\rho) = \int_{I^M} c(\mathbf{x})\rho(\mathbf{x})\mathrm{d}\mathbf{x}, \\ \text{provided that } I(\rho) \geq 0. \end{gathered} \tag{6.5}$$

The generalization here is that instead of requiring $c(\mathbf{x}) \geq 0$, we require only the whole functional $I(\rho)$ to be positive. This allows us to consider (6.5) for the cost functions taking both positive and negative values, while keeping $I(\rho)$ a non-negative linear functional on $\mathcal{E}\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\right)$, which we will need to apply the separation theorems.

Our goal is to characterize the set $\mathcal{E}\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\right)$ in terms of the cost functions $c(\mathbf{x})$. The following two lemmas provide the characterization.

Lemma 6.6 (Direct MK) *If the set of cost functions $c(\mathbf{x})$ is such that $I(\rho) < \infty, \forall\rho(\mathbf{x}) \in \partial\mathfrak{C}^N_{poly}(\mathbf{p})$, then the problem (6.5) has a solution, $\rho^{\mathcal{E}} \in \mathcal{E}\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\right)$.*

Proof. Since $I(\rho)$ is linear, continuous and bounded on a convex closed set $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$, by the Weierstrass theorem it achieves its minimum and maximum at an extreme point of the set, while these extrema need not be unique. ■

Therefore, to formulate the characterization we need to solve the "inverse" MK problem: given an extreme copula we prove the existence of a cost function of an MK problem, for which the given extreme copula is an optimizer.

We start with a weaker characterization of the equivalence classes of the polynomial copulas in terms of $c(\mathbf{x}) \in \mathcal{L}^2\left(I^M\right)$ that are densities of signed measures on I^M.

Lemma 6.7 *For each $\rho^{\mathcal{E}}(x,y) \in \mathcal{E}\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\right)$ there exists a regular signed Borel measure, absolutely continuous with respect to Lebesgue measure with the density $c(\mathbf{x}) \in \mathcal{L}^2\left(I^M\right)$, such that $\rho^{\mathcal{E}}(\mathbf{x})$ solves the problem (6.5) for the cost function $c(\mathbf{x})$.*

Proof. By the Separating Hyperplane Theorem, there exists an affine hyperplane $H(\rho) = 0$, going through $\rho^{\mathcal{E}}$, such that $H\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\backslash\rho^{\mathcal{E}}\right) > 0$. Considering $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$ as a subspace of $\mathcal{L}^2\left(I^M\right)$, by the Riesz representation theorem there will exist an element $c(\mathbf{x}) \in \mathcal{L}^2\left(I^M\right)$ such that

$$H(\rho) = \int c(\mathbf{x})\rho(\mathbf{x})\mathrm{d}\mathbf{x}.$$

In this case $c(\mathbf{x})$ can be considered a cost function in the problem 6.5 and by construction $\rho^{\mathcal{E}}$ is a solution of this problem over $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$. ■

Now we return to the standard setting of the MK problem, requiring that the cost function is non-negative. That is, our goal will be to associate with each extreme polynomial copula a cost function $c(\mathbf{x})$, non-negative on I^M, such that the copula is the solution to the following problem:

$$\begin{aligned} &\text{find } \rho(\mathbf{x}) \in \partial\mathfrak{C}^N_{poly}(\mathbf{p}), \\ &\text{which minimizes } I(\rho) = \int_{I^M} c(\mathbf{x})\rho(\mathbf{x})\mathrm{d}\mathbf{x}, \\ &\text{provided that } c(\mathbf{x}) \geq 0. \end{aligned} \tag{6.6}$$

First note that both the formulation and proof of necessity (Lemma 6.6) still hold in this case.

We now proceed to sufficiency. In the following lemma we use the following convention given the type $\mathbf{p}: \ \mathbf{p} - 1 = \{p_1 - 1, \dots p_M - 1\}$.

Lemma 6.8 (Inverse MK, general) *For each $\rho^{\mathcal{E}} \in \mathcal{E}\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\right)$, there exists a cost function $c(\mathbf{x})$ on I^M such that $\rho^{\mathcal{E}}$ is a solution of the problem (6.6) over $\rho \in \partial\mathfrak{C}^N_{poly}(\mathbf{p})$.*
Such $c(\mathbf{x})$ is a non-negative density of a regular Borel measure with moments

$$m_{\alpha_1 \dots \alpha_M}(\mu) = \int_{I^M} \prod_i x_i^{\alpha_i} \mathrm{d}\mu < \infty,$$

for all $\alpha_i < p_i$.

Proof. The plan is to use the separation theorems to construct a linear functional over $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$ attaining the minimum at $\rho^{\mathcal{E}}$, then find the representation of this functional as an integral by some measure and then utilize the fact that

$\partial\mathfrak{C}^N_{poly}(\mathbf{p})$ is a finite dimensional space to select the measure to be positive and have density.

Consider a cone $Con_{poly}(\mathbf{p}-1)$ of non-negative polynomials with marginal degrees of up to $\mathbf{p}-1=(p_1-1,\dots,p_M-1)$, which is closed and convex (see Section 3); $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$ is a convex and compact subset of $Con_{poly}(\mathbf{p}-1)$ and $\rho^{\mathcal{E}}(\mathbf{x})$ is an extreme point.

Note that the trivial zero polynomial belongs to $Con_{poly}(\mathbf{p}-1)$ and is also an extreme point. Consider a line connecting $l(t,\rho^{\mathcal{E}})=t+(1-t)\rho^{\mathcal{E}}$, connecting the trivial zero polynomial with $\rho^{\mathcal{E}}(\mathbf{x})$. By the Hyperplane Separation Theorem, there exist an affine hyperplane $H(\rho)=0$, going through $\rho^{\mathcal{E}}$ such that $Con_{poly}(\mathbf{p}-1)$ lies entirely on one side of that hyperplane, i.e. $H\left(Con_{poly}(\mathbf{p}-1)\right)\geq 0$. The line $l(t,\rho^{\mathcal{E}})$ lies on the boundary of $Con_{poly}(\mathbf{p}-1)$, so it must belong to that hyperplane H, i.e. $H((l(t,\rho^{\mathcal{E}}))=0$, hence $H(0)=0$, and therefore the functional H is linear and positive on $Con_{poly}(\mathbf{p}{-}1)$. Thus H is a positive continuous functional on the space of all $(\mathbf{p}-1)$ polynomials.

Consider the space of all polynomials $C_{poly}\left(I^M\right)$ as a subset of functions $C\left(I^M\right)$, continuous on $[0,1]$, denoting $C^{\mathbf{p}}_{poly}\left(I^M\right)$ the subset of $C_{poly}\left(I^M\right)$ of $(\mathbf{p}-1)$ polynomials. Since for $\forall p\in C_{poly}\left(I^M\right)$, $\exists q\in C^{\mathbf{p}}_{poly}\left(I^M\right)$, such that $q\geq p$, and H is a positive linear functional on $C^{\mathbf{p}}_{poly}\left(I^M\right)$, the conditions of the Riesz extension theorem of a linear positive functional (see e.g. [Cas05, Theorem 1]) are fulfilled, hence H can be extended as a positive functional from $C^{\mathbf{p}}_{poly}\left(I^M\right)$ to $C_{poly}\left(I^M\right)$. Note that such an extension will provide a continuous functional, if continuity on $C\left(I^M\right)$ is with respect to the supremum norm.

Since $C_{poly}\left(I^M\right)$ is dense in the set $C\left(I^M\right)$ of functions continuous on I^M, H can be extended to a positive continuous functional on $C\left(I^M\right)$.

As H is now a positive continuous functional on $C\left(I^M\right)$, and I^M is compact, by the Riesz representation theorem for positive continuous functionals (see, e.g. [How95, Theorems 8.14 and 8.15]) for all $f \in C\left(I^M\right)$, we have

$$H^*(f) = \int_{I^M} f \mathrm{d}\mu, \tag{6.7}$$

where μ is a regular positive Borel measure on I^M. Since H^* was obtained through the sequence of extensions of a functional, such that $H(\rho) > 0$ for $\rho \in \partial\mathfrak{C}^N_{poly}(\mathbf{p})$ and $H(\rho^{\mathcal{E}}) > 0$, $\rho^{\mathcal{E}}$ can be viewed as a solution to the optimization problem $H^*(\rho) \to \min$, over $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$.

Note that by the nature of ρ, μ must have finite moments $m_{\alpha_1\ldots\alpha_M}$ at least for all $\alpha_i < p_i$. Moreover, the value of H^* on $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$ is a linear combination of these moments. Therefore we can select any other measure ν, such that for $\overline{\alpha} < \mathbf{p}$: $m_{\alpha_1\ldots\alpha_n}(\mu) = m_{\alpha_1\ldots\alpha_n}(\nu)$, $\rho^{\mathcal{E}}$ is also the solution of the problem

$$I(\rho) = \int_{I^M} \rho(\mathbf{x})\mathrm{d}\nu \to \min .$$

We can select ν such that it has a polynomial density, i.e.

$$\mathrm{d}\nu = \sum_{\overline{\beta}} \eta_{\overline{\beta}} \mathbf{x}^{\overline{\beta}} \mathrm{d}\mathbf{x} = q(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{6.8}$$

The moment matching condition $m_{\alpha_1\ldots\alpha_n}(\mu) = m_{\alpha_1\ldots\alpha_n}(\nu)$ translates into a system of linear equations for $\eta_{\overline{\beta}}$. The set of multi-indices (and powers) $\overline{\beta}$ of such a polynomial can always be selected such that the moment-matching system has a non-trivial solution.

If the resulting polynomial $q(\mathbf{x}) \geq 0$ on I^M, then set $c(\mathbf{x}) = q(\mathbf{x})$. Otherwise, observe that if $\rho^{\mathcal{E}}$ is the solution of

$$I(\rho) = \int_{I^M} \rho(\mathbf{x}) q(\mathbf{x}) \mathrm{d}\mathbf{x} \to \min,$$

it is also the solution of

$$I_C(\rho) = \int_{I^M} \rho(\mathbf{x}) \left(q(\mathbf{x}) + C\right) \mathrm{d}\mathbf{x} \to \min,$$

for arbitrary positive constant C. Indeed, (local) extremality of $\rho^{\mathcal{E}}$ requires for any admissible variation $\delta\rho$ that

$$\begin{aligned} I(\rho^{\mathcal{E}} + \delta\rho) &= \int_{I^M} \left(\rho^{\mathcal{E}}(\mathbf{x}) + \delta\rho(\mathbf{x})\right) q(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= I(\rho^{\mathcal{E}}) + \int_{I^M} \delta\rho(\mathbf{x}) q(\mathbf{x}) \mathrm{d}\mathbf{x} \geq I(\rho^{\mathcal{E}}), \end{aligned}$$

implying that

$$\int_{I^M} \delta\rho(\mathbf{x}) q(\mathbf{x}) \mathrm{d}\mathbf{x} \geq 0, \tag{6.9}$$

for the admissible variations. Therefore

$$\begin{aligned} I_C(\rho^{\mathcal{E}} + \delta\rho) &= \int_{I^M} \left(\rho^{\mathcal{E}}(\mathbf{x}) + \delta\rho(\mathbf{x})\right) \left(q(\mathbf{x}) + C\right) \mathrm{d}\mathbf{x} \\ &= \int_{I^M} \rho^{\mathcal{E}}(\mathbf{x}) \left(q(\mathbf{x}) + C\right) \mathrm{d}\mathbf{x} + \int_{I^M} \delta\rho(\mathbf{x}) \left(q(\mathbf{x}) + C\right) \mathrm{d}\mathbf{x} \\ &= I_C(\rho^{\mathcal{E}}) + \int_{I^M} \delta\rho(\mathbf{x}) q(\mathbf{x}) \mathrm{d}\mathbf{x} \geq I_C(\rho^{\mathcal{E}}), \end{aligned}$$

where we have used (6.9) and the boundary condition on the admissible variation

$$\delta\rho(\mathbf{x}) : \int_{I^M} \delta\rho(\mathbf{x}) \mathrm{d}\mathbf{x} = 0$$

(which is needed to keep $\rho^{\mathcal{E}} + \delta\rho$ a probability measure density). Now $c(\mathbf{x}) =$

$q(\mathbf{x}) - \min_{\mathbf{x} \in I^M}(q(\mathbf{x}))$, which by construction is the density of a positive Borel measure on I^M. This completes the proof. ■

The two lemmas above imply the following

Theorem 6.9 *For a polynomial density ρ to be extreme, i.e. $\rho \in \mathcal{E}\left(\partial\mathfrak{C}^N_{poly}(\mathbf{p})\right)$, it is necessary and sufficient that there exists a positive cost function $c(\mathbf{x})$ (in the MK sense), such that ρ is the solution for the MK problem over $\partial\mathfrak{C}^N_{poly}(\mathbf{p})$. Such $c(\mathbf{x})$ is a density of a regular Borel measure μ on I^M, and such μ has finite moments*

$$m_{\alpha_1 \ldots \alpha_n}(\mu) = \int_{I^M} \prod_i x_i^{\alpha_i} \mathrm{d}\mu < \infty,$$

for all $\alpha_i < p_i$.

By the Weierstrass Theorem, a positive cost function can be approximated arbitrarily closely by a polynomial that will necessarily be **strictly** positive on I^M. The representation of such polynomials (of arbitrary degree) is given by the Schmuedgen Theorem (Theorem 2.28), which implies the following:

Corollary 6.10 *There is a bijection between the set of the cost functions $c(\mathbf{x})$ considered in Theorem 6.9 and a subset of preorder $T_{h_{11}h_{21}\ldots h_{1M}h_{2M}}(\mathbf{x})$, where $h_{1i}(x_i) = x_i, h_{2i}(x_i) = 1 - x_i$.*

6.3 Polynomial Copulas and Convolutions

Recall that the *probability generating function* (**pgf**) of an integer-valued random variable x with probability distribution $p_n = \mathbb{P}(x = n)$ is defined by

$$f(z) = \mathbb{E}(z^x) = \sum_n p_n z^n.$$

In particular, if x is bounded than its **pgf** can be written as

$$f(z) = z^{-x_{\min}} Q(z),$$

where $Q(z)$ is a polynomial. This property is used, in particular, to reduce convolving independent bounded integer valued random variables to multiplication of polynomials.

Somewhat stretching the analogy, we envisage potential applications of the multivariate polynomial copulas in convolving random variables. This is because the very form of a polynomial copula

$$C(x_1, \ldots, x_M) = \sum_{i_1 \ldots i_M} \theta_{i_1 \ldots i_M} \prod_{i=1}^{M} x_i^{n_i} = \sum_{\mathbf{i}} \theta_{\mathbf{i}} \prod_{i=1}^{M} x_i^{n_i} \tag{6.10}$$

can be interpreted as a "probability generating function" for the $\mathbb{N}^M$-valued random vector v, assuming that $\theta_{i_1 \ldots i_M} = \mathbb{P}\left(v = (i_1 \ldots i_M)\right)$. The vector v is bounded because the order of the copula is bounded. Recall that normalization of copula implies

$$C(1, \ldots, 1) = 1 = \sum_{i_1 \ldots i_M} \theta_{i_1 \ldots i_M},$$

and hence the vector of $\theta_{i_1 \ldots i_M}$ can indeed be interpreted as a signed probability measure.

Substituting the margins into the copula arguments yields, for a joint distribution, which we will refer to as the *polynomial coupling* of the margins, the following expression:

$$G(x_1, \ldots, x_M) = C(F_1(x_1), \ldots, F_M(x_M)) = \sum_{k} \theta_k \prod_{i=1}^{M} F_i(x_i)^{m_{ik}}, \tag{6.11}$$

and its density is

$$g(x_1,\ldots,x_M)=\partial_{x_1,\ldots x_M}G(x_1,\ldots,x_M)=\sum_k\theta_k\prod_{i=1}^M m_{ik}F_i(x_i)^{m_{ik}-1}f_i(x_i), \tag{6.12}$$

where $f_i(x_i)=F_i'(x_i)$ is i-th marginal density. Note that we have abbreviated the notation of the index of the summands, writing $\theta_k=\theta_{i_1\ldots i_M}$.

Expression (6.11) can now be interpreted as a factor copula, the factor being θ_k. There is, however, a difference to the typical construction of a factor copula. Typically, conditioned on the factor we obtain an independence copula of the margins. In (6.11), conditioned on the factor, we obtain an independence copula of the integer powers of the margins.

As an illustration, consider how this interpretation of a polynomial copula as a factor copula can be used to convolve random variables, coupled with this copula. Consider $z=\sum_{i=1}^M x_i$. To construct the distribution function for z, $F(t)=\mathbb{P}(z\le t)$ we have to integrate 6.12 over the region in $\mathbb{R}^M$, determined by $\sum_{i=1}^M x_i<t$:

$$\begin{aligned}\mathbb{P}(z<t)&=\int_{\sum_{i=1}^n x_i<t}\sum_k\theta_k\prod_{i=1}^M m_{ik}F_i(x_i)^{m_{ik}-1}f_i(x_i)\mathrm{d}x_i \qquad (6.13)\\ &=\sum_k\theta_k\int_{\sum_{i=1}^n x_i<t}\prod_{i=1}^n\mathrm{d}F_i(x_i)^{m_{ik}}\\ &=\sum_k\theta_k\int_{\sum_{i=1}^n x_i<t}\prod_{i=1}^n\mathrm{d}F_{i,k}^*(x_i),\\ &\text{where } F_{i,k}^*(x_i)=F_i(x_i)^{m_{ik}}\end{aligned}$$

Now each of the summands can be evaluated separately, for which we need to convolve M conditionally-independent random variables with probability distributions $F_i^*(x_i)$.

6.4 Reflexive Polynomial Copulas

In this section we introduce a class of *reflexive* polynomial copulas which is closed under forming the convex combinations and allows simple characterization of its extreme points. As we show below, this class is attractive, because it is quite easy to construct such copulas, as opposed to constructing general polynomial copulas.

Definition 6.11 *We call a polynomial copula density* $1 + Q(\mathbf{x})$ ***reflexive*** *if* $|Q(\mathbf{x})| \leq 1$ *on* I^M.

The subset of the reflexive copulas of type $\mathbf{p}$ will be denoted $\mathfrak{R}^N_{poly}(\mathbf{p})$; note that $\mathfrak{R}^N_{poly}(\mathbf{p}) \subset \mathfrak{C}^N_{poly}(\mathbf{p})$ the subset of reflexive polynomial copulas of given order and type.

Clearly, the set of the reflexive polynomial copulas is stable with regards to forming convex combinations, therefore it is characterized by its own set of extreme points. Note that since $|Q(\mathbf{x})| \leq 1$, $1 - Q(\mathbf{x}) \in \partial\mathfrak{R}^N_{poly}$. Now we have the following straightforward result:

Proposition 6.12 *If* $1 + Q(\mathbf{x})$ *is reflexive and an extreme point f* $\partial\mathfrak{R}^N_{poly}(\mathbf{p})$, *so is* $1 - Q(\mathbf{x})$.

Corollary 6.13 *If* $1+Q(\mathbf{x})$ *is reflexive and extreme of* $\partial\mathfrak{R}^N_{poly}(\mathbf{p})$ *then* $|Q(\mathbf{x})| - 1$ *must have a zero on* I^M

Proof. This follows from the fact that the construction of the above Proposition creates a reflexive density out of a reflexive density. ∎

The natural tool to analyze reflexive copulas are multivariate Legendre polynomials, shifted to I^M, which we now review.

Definition 6.14 *A* $[0,1]$ *shifted Legendre polynomial of degree* n *is defined by*

$$\xi_n(x) = (-1)^n \sum_{i=0}^{n} \binom{n}{k} \binom{n+k}{k} (-x)^k.$$

The shifted Legendre polynomials are orthogonal on $[0,1]$ with respect to $\mathcal{L}^2$ inner product:

$$\int_0^1 \xi_n(x)\xi_m(x)\mathrm{d}x = \frac{1}{2n+1}\delta_{nm}. \tag{6.14}$$

Definition 6.15 *A multivariate Legendre polynomial shifted to* I^M *is defined by*

$$P_{n_1 \dots n_M}(x_1, \dots, x_M) = \prod_{i=1}^{M} \xi_{n_i}(x_i)$$

Clearly the $P_{n_1 \dots n_M}$ are orthogonal on I^M. For further details on applications of Legendre and other orthogonal polynomials see [Wal04].

The reason Legendre polynomials are especially useful in the analysis of copulas is because $Q(\mathbf{x})$ must change sign on I^M. In all other settings we have considered, the polynomials of interest were non-negative on I^M, and the non-negativity was essential. Now the change of the sign is essential, as otherwise the uniform margins of $1+Q(\mathbf{x})$ will not be reproduced and hence it will not be a copula. However, orthogonality (6.14) implies that $\int_0^1 \xi_n(x)\mathrm{d}x = 0$; therefore if we put

$$C(x_1, \dots, x_M) = 1 + Q(x_1, \dots, x_M),$$

where Q is a multivariate Legendre polynomial, such that $Q(x_1, \dots, x_M) \geq -1$ on I^M, then C will be a copula.

With regard to reflexive copulas, representation in terms of the Legendre polynomials plays a special role.

Theorem 6.16 *Polynomials of the form*

$$\mathrm{d}C(\mathbf{x}) = 1 + Q(\mathbf{x}) = 1 + \prod_i \xi_i(x_i) \tag{6.15}$$

are extreme in $\partial\mathfrak{R}^N_{poly}(\mathbf{p})$*.*

Proof. This follows from the fact that univariate Legendre polynomials form a complete orthogonal basis on $[0,1]$ and that the representation of any $Q(\mathbf{x})$, a polynomial of bounded degree, in terms of linear combination of the Legendre polynomials is unique. ∎

Note that Legendre polynomials fulfil the necessary condition of extremality: they are bounded between $[-1,1]$ and reach values of either -1 or 1 on I^M.

6.5 Literature Review

Polynomial copulas are essentially a new research subject, the problem of characterizing the Choquet-extreme copulas being totally new. We believe this is because the fields of multivariate applied probability (including coupling methods) and real algebraic geometry did not generally overlap, perhaps, for historical reasons. Indeed, real algebraic geometry is a much younger field of mathematics, therefore at the time many classical results in multivariate applied probability were being obtained, including the development of the nomenclature for parametric models, the relevant real algebraic tools were not yet available, at least for the case of bounded random variables. Also, the practical implementation of algebraic methods only became feasible with the substantial increase in CPU power in recent decades.

There are only a few publications dealing with polynomial copulas per se. Firstly we mention a few sections in [Nel06] exploring FGM and dealing with the cases when *sections* of copula (e.g. a distribution function of one of the variables, conditioned on another variable) are polynomials of prescribed degree. There are several publications that generalize the FGM family in various ways (e.g. [FK07] and the publications it references), with the resulting copulas not being polynomials. The nearest work to our setup would probably be [SS04], where the Bernstein approximation to copula functions is studied, although coefficients that would make the polynomials under consideration a copula are not characterized. Still this work mentions ideas that are explored here in the section on polynomial coupling.

On the other hand, the characterization of Choquet-extreme probability distribution functions with fixed margins is a relatively old field, with an abundance of literature. The mainstream research in this area deals with the optimal transport problem (MK problem in the current work), both in the classical two-dimensional, and the generic case. It is essential for the problem setup that the margins of the transport plan are fixed, therefore, given the margins, identifying and characterizing the optimal transport plans is essentially characterizing some subset of copulas. However, the solution of the optimal transport problem will depend on the margins, so having margins uniform is, perhaps, a simpler setup from the optimal transport problem point of view. For the references in the two-dimensional cases see [Vil03], or the major reference [RR98].

The classical result in characterization of optimal transport plan is the Birkhoff theorem [Bir46], stating that permutation matrices are extreme doubly stochastic matrices. An extension of this to the case of continuous bounded random variables stems from the works of Sudakov [Sud79] . A characterization of the optimal transport plan in the MK problem is given in [WT07]. None of these results is actually useful in our setup, as the authors do not restrict themselves to considering only sets of polynomials.

7 Conclusions and Further Research

This thesis explored both theoretical and applied aspects of polynomial and copula densities, and one main theme of the work was the characterization of the Choquet-extreme densities.

From the theoretical standpoint, it is most interesting to complete and extend the results of Section 5 to a characterization of the extreme polynomial norms (homogenous multivariate polynomials of a given degree), non-negative either on real algebraic sets or on full underlying fields. Clearly, in the most general setting one will have to drop the assumptions of the set of such norm being compact, thus rendering the Choquet theorem inapplicable. The problem will then be to characterize the non-negative norms that cannot be represented as convex combinations of other non-negative norms only in the algebraic sense. For example, given non-negative forms x, y, z over a real field $\mathbb{K}$ with a preorder P (writing $a <_P b$, $a, b \Leftrightarrow b - a \in P$) and $\lambda \in \mathbb{K}$, such that $0 <_P \lambda <_P 1$, then

$$x = \lambda y + (1 - \lambda)z, \text{ with } y, z \in S \Rightarrow y = z = x. \tag{7.1}$$

It can be required that the forms involved are non-negative on the whole $\mathbb{K}^N$ or only on its semialgebraic subset.

As we noted in the literature review for Section 5, the problem of characterizing extreme non-negative forms has been considered, but not systematically. Also, it was only considered in the more complex (in our opinion) case of the form being non-negative on $\mathbb{K}^N$. However, if one requires the form to be non-negative only on a (simple) semialgebraic set, and considers only forms of bounded degree and endow the set of forms with some norm, this will introduce an equivalence relationship on the set of such forms, the equivalence classes being stable with respect to the formation of convex combinations. Then one can use the decomposition approach we employed in Section 5 to characterize the extreme norms in a given equivalence class. The fact that given the set of non-negative polynomials of a fixed degree, an extreme polynomial must be of that degree hints that the whole set of forms non-negative on a semialgebraic set will not have extreme points at all.

From the practical standpoint, polynomial densities both in the univariate and multivariate cases are attractive in modelling bounded random variables. In the multivariate case however the computational burden of the numerical methods may be too high.

A two-dimensional copula can be viewed as a Markov chain generator. Therefore another problem interesting from both the practical and applied points of views would be studying Markov chains generated by two dimensional polynomial copulas.

8 Appendix. Polynomial Multiplication

On several occasions the numerical algorithms we considered could be reduced to the multiplication of polynomials of equal and low degrees, e.g. the multiplication of a sequence of quadratic polynomials:

$$Q(x) = \prod_{k=1}^{M} (\alpha_k x^2 + \beta_k x + \gamma_k).$$

Since this operation for large M can become a bottleneck of a numerical algorithm, we briefly review the efficient numerical methods of computing such products.

Recall that given polynomials $\sum_i a_i x^i$ and $\sum_i b_i x^i$ their product is the polynomial with coefficients

$$c_i = \sum_j a_j b_{i-j}. \tag{8.1}$$

This operation is known as "coordinate-wise" multiplication.

Coordinate-wise multiplication of two dense polynomials (i.e. those with most coefficients $\neq 0$) of order N has complexity $O(N^2)$. In particular, if the powers of the polynomials are the same and equal N, then the complexity is approximately $2N^2$.

There is a method of multiplying two dense polynomials of the same degree N with complexity $N \log N$, which for large N will be faster than coordinate-wise multiplication. This method is based on the Fourier Transform, specifically on the Fast Fourier Transform (FFT) algorithm. The main idea is to replace the coordinate-wise multiplication by interpolation. Indeed, a univariate polynomial of order N is fully specified by its N zeroes, or, equivalently, its values at N different points. Suppose we need to multiply two polynomials f_1 and f_2 of orders N_1 and N_2 respectively. This means that if we select N_1+N_2 different values of $x_i, i = 1 \ldots N_1 + N_2$ and write $g(x) = f_1(x)f_2(x)$, then

$$g(x_i) = f_1(x_i)f_2(x_i).$$

Once the values $g(x_i)$ are obtained, the coefficients of $g(x)$ can be recovered by *interpolation*, which in this case amounts to solving a system of linear $N_1 + N_2$ equations for the coefficients. The problem is that direct solution of such system has complexity $O\left((N_1 + N_2)^2\right)$, equalling $O\left((2N)^2\right)$ if $N_1 = N_2 = N$. This is not better than the complexity of the coordinate-wise multiplication. Note that we also need to evaluate $f_1(x_i)$ and $f_2(x_i)$ and their product, but this has complexity $O\left((N_1 + N_2)\right)$, which is dominated by the complexity of solving the linear system.

It is possible, however, to drastically speed up both the initial evaluation and the inversion, using the Fast Fourier Transform algorithm. The FFT-based multiplication of the polynomials of the same degree N has the complexity $O(2N(1+\log 2N))$, which for large N is much better than $O(N^2)$. For a detailed treatment of the method see [CCLR01].

However, our problem is special in that we have to multiply low (and equal) degree polynomials, but there may be a lot of them. FFT can be applied in two ways here. One way is first to evaluate all polynomials in the common set of points (i.e. their representation in Fourier space is established), then to multiply the values, and then to invert. This is bad if the degree of the polynomials is much smaller than their number. Indeed, in this case the complexity will be $O\left(MN(1 + \log NM)\right)$. A better alternative is a "divide and conquer" strategy: first multiply pairs, then group the results of pair multiplication and multiply pair-wise, and so forth.

Yet another alternative is to implement the coordinate-wise multiplication of the whole sequence of polynomials via a "recursion " (which would be the preferable method of convoluting about up to 100-150 low-integer weighted independent Bernoulli variables). To implement the recursion we order the multiples somehow (the order is irrelevant to commutativity). The *k-th* step of the recursion will be multiplying the *k-th* trinomial $\alpha_k x^2 + \beta_k x + \gamma_k$ with the product of the first k-1 polynomials, which is the polynomial of order $3(k-1)$. Since polynomial multiplication is commutative, the order in which the polynomials are multiplied is irrelevant.

If we denote the coefficients of product of the first $k-1$ polynomials as $q_i^{k-1}, i = 0 \dots 3(k-1)$, in other words, the outcome of the $(k-1)-th$ step of the recursion, then we get for the $k-th$ step:

$$q_0^k = q_0^{k-1}\gamma_k, \tag{8.2}$$

$$q_1^k = q_0^{k-1}\beta_k + q_1^{k-1}\gamma_k, \tag{8.3}$$

$$q_i^k = q_{i-2}^{k-1}\alpha_k + q_{i-1}^{k-1}\beta_k + q_i^{k-1}\gamma_k, \qquad i > 2 \dots 3k, \tag{8.4}$$

assuming that $q_i^k = 0$ for $k < 0$.

The recursion is initialized by setting $q_0^0 = 1, q_{>0}^0 = 0$, representing a unit monomial of power 0. The complexity of the algorithm is approximately $4\sum_{i=1}^{3M} i = 2(3M-1)(3M)$, assuming all operations have the complexity of (8.4), which is the majorant, including two multiplications and two additions, and hence 4 operations. For large M it makes sense to implement a divide-and-conquer scheme, where coordinate-wise multiplication, including the above recursion, can be used to multiply a few initial trinomials, split into groups, and the resulting larger degree polynomials to be multiplied with some version of FFT. The obvious options would be to switch to "power of 3 " FFT after multiplying groups of 3 polynomials, or switching to a more traditional "power of 2 " FFT, after multiplying the groups of 10 polynomials to minimize zero padding. Clearly the fact that we have to multiply the trinomials plays in favour of the recursion or the non-standard FFT.

The importance of (8.2)-(8.4) is that they allow trivial inversion, which can be used to implement the coordinate-wise division (independently of how the polynomials were multiplied)

$$\prod_{k=1}^{M}(\alpha_k x^2 + \beta_k x + \gamma_k)/(\alpha_i x^2 + \beta_i x + \gamma_i) \tag{8.5}$$

for some $i = 1, \ldots, M$, independently of how the product has been obtained. Indeed, assuming that the $i-th$ polynomial was the last to have been multiplied with, we can solve (8.2)-(8.4) for q_i^{k-1}, which are the coefficient of the following fraction:

$$q_0^{k-1} = q_0^k/\gamma_i, \tag{8.6}$$

$$q_1^{k-1} = (q_1^k - q_0^{k-1}\beta_i)/\gamma_i, \tag{8.7}$$

$$q_{i-2}^{k-1} = \left(q_i^k - q_{i-1}^{k-1}\beta_i - q_i^{k-1}\gamma_i\right)/\alpha_i, \qquad i > 2 \ldots 3k. \tag{8.8}$$

Such division is needed mainly if one is interested in the gradient of $Q(x)$ with regard to the coefficients of the $i-th$ polynomial. Indeed, we have

$$\partial_{\alpha_i} Q(x) = x^2 \frac{Q(x)}{\alpha_i x^2 + \beta_i x + \gamma_i}, \tag{8.9}$$

$$\partial_{\beta_i} Q(x) = x \frac{Q(x)}{\alpha_i x^2 + \beta_i x + \gamma_i}, \tag{8.10}$$

$$\partial_{\gamma_i} Q(x) = \frac{Q(x)}{\alpha_i x^2 + \beta_i x + \gamma_i}. \tag{8.11}$$

Thus, having performed the division (8.5) we get the coefficients of the gradient components up to renumbering (which corresponds to multiplication by x and x^2). Therefore the complexity of the operation to compute the gradient is $O(4 \cdot 3M)$, while the straightforward numerical differentiation would require rebuilding the whole product polynomial with the corresponding complexity.

Note that the same logic may be used to compute the analytical Hessian.

References

[Bir46] G. Birkhoff, *Tres observaciones sobre el algebra lineal*, Univ. Nac. Tucumán Rev, Ser. A (1946), no. 5, 147–151.

[Cas05] R. E. Castillo, *A note on Krein theorem*, Lecturas Metematicas **26** (2005), 5–9.

[CCLR01] T. H. Cormen, S. Clifford, C. E. Leiserson, and R. L. Rivest, *Introduction to algorithms*, 2nd revised ed., The MIT Press, 2001.

[CH74] D.R. Cox and D.V. Hinkley, *Theoretical statistics*, Chapman & Hall, 1974.

[CHB+01] S. E. Cox, R. T. Haftka, C. A. Baker, B. Grossman, W. H. Mason, and L. T. Watson, *A comparison of global optimization methods for the design of a high-speed civil transport*, Journal of Global Optimization **21** (2001), 415–432.

[CL77] M.-D. Choi and T.-Y. Lam, *Extremal positive semidefimite forms*, Math. Ann. (1977), no. 231, 1–18.

[dlP02] M. J. de la Puente, *Real plane algebraic curves*, Real Algebraic Geometry Preprint Server, 2002, `http://www.uni-regensburg.de/Fakultaeten/nat_Fak_I/RAAG/`.

[FK07] M. Fischer and I. Klein, *Constructing symmetric generalized FGM copulas by means of certain univariate distributions*, Metrika **65** (2007), no. 2, 243–260.

[How95] N. R. Howes, *Modern analysis and topology*, Springer, 1995.

[JBR98] M. Coste J. Bochnak and M.-F. Roy, *Real albegraic geometry*, Springer, 1998.

[Joe97] H. Joe, *Multivariate models and dependence concepts*, Chapman & Hall, 1997.

[LAB03] J. Sidenius L. Andersen and S. Basu, *All your hedges in one basket*, Risk (2003), 67–72.

[LG05] J.-P. Laurent and J. Gregory, *Basket default swaps, CDO's and factor copulas*, Journal of Risk **7** (2005), no. 4, 103–122.

[Nel06] R. B. Nelsen, *An introduction to copulas*, 2nd ed., Springer, 2006.

[PD01] A. Prestel and C. N. Delzell, *Positive polynomials*, Springer, 2001.

[Phe01] R. R. Phelps, *Lectures on Choquet's theorem*, 2nd revised ed., Springer, 2001.

[PR00] V. Powers and B. Reznick, *Polynomials that are positive on an interval*, Trans. Amer. Math. Soc. (2000), no. 352, 4677–4692.

[PR05] ______, *Polynomials that are positive on unbounded intervals*, Lecture Notes in Control and Inform. Sci., no. 312, pp. 151–163, Springer, 2005.

[RR98] S. Rachev and L. Ruschendorf, *Mass transportation problems*, Springer, 1998.

[Sch91] K. Schmudgen, *The k-moment problem for compact semi-algebraic sets*, Mathematische Annalen (1991), no. 289, 203–206.

[Sch05] C. Scheiderer, *Distinguished representation of non-negative polynomials*, J. Algebra (2005), no. 289, 558–573.

[SS04] A. Sancetta and S. Satchell, *Bernstein approximations to the copula function and portfolio optimization*, Econometric Theory (2004), no. 20, 535–562.

[Sud79] V. N. Sudakov, *Geometric problems in the theory of infinite dimensional probability distributions*, Proceedings of the Steklov Institute of Mathematics, no. 2, American Mathematical Society, 1979.

[Tho00] H. Thorisson, *Coupling, stationarity and regeneration*, Springer, 2000.

[Vil03] C. Villani, *Topics in optimal transportation*, American Mathematical Society, 2003.

[Wal04] G. Walter, *Orthogonal polynomials. Computation and approximation*, Oxford University Press, 2004.

[WT07] W.Schachermayer and J. Teichmann, *Characterization of optimal transport plans of the Monge-Kantorovich problem*, Proceedings of the American Mathematical Society (2007).