Title
Some Problems on the Convex Geometry of Probability Measures

Author
Zhu, Theodore

Publication Date
2022

Peer reviewed|Thesis/dissertation

Some Problems on the Convex Geometry of Probability Measures

by

Theodore Zhu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Mathematics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jim Pitman, Chair
Professor Fraydoun Rezakhanlou
Professor Nike Sun

Summer 2022

Some Problems on the Convex Geometry of Probability Measures

Abstract

Some Problems on the Convex Geometry of Probability Measures

by

Theodore Zhu

Doctor of Philosophy in Mathematics

University of California, Berkeley

Professor Jim Pitman, Chair

This thesis consists of three main topics in which we explore the geometry and other features of certain convex sets arising in probabilistic contexts.

We first consider the set of laws of K_n, the number of distinct values among the first n terms of a sequence, for infinite exchangeable sequences of random variables $(X_1, X_2, \ldots)$. We prove for $n = 3$ that the extreme points of the convex set of all possible laws of K_3 are those derived from i.i.d. sampling from discrete uniform distributions and the limit case with $\mathbb{P}(K_3 = 3) = 1$. We also consider the problem in higher dimensions and variants of the problem for finite exchangeable sequences and exchangeable random partitions.

Second, we introduce the notion of a coherent pair of random variables, or two conditional probabilities of the same event, and study the convex set of laws on $[0,1]^2$ arising in this manner. We classify all extreme laws with a certain restriction on the support. We also discover a large class of extreme laws with finite and countably infinite support.

Third, we study the convex set of polynomial probability densities on $[0,1]$ of degree at most n. We review some known results, including characterization of the extreme points, a representation theorem, properties of the Bernstein polynomial basis, and the Lorentz degree which measures in some sense the representability of positive polynomials in the Bernstein basis. We map out the geometry for $n = 2$, consider the uniform random sampling model, and compute the upper envelope for this set of polynomials.

Contents

Acknowledgments

First and foremost, I am extremely fortunate to have been advised by Jim Pitman. He guided me with incredible enthusiasm, generosity, and patience since the very beginning of my journey, when I was just an undergraduate student. His mentorship was invaluable to my growth as a mathematician and as a person. I will remember with great fondness the countless hours we spent discussing mathematics on the whiteboard, hiking along the scenic trails of Berkeley, and competing on the tennis courts.

I am thankful to Chris Burdzy and Yuri Yakubovich for collaborations which were fruitful in the development of this thesis.

I am thankful to Fraydoun Rezakhanlou and Nike Sun, who kindly accepted their roles on my dissertation committee.

I am thankful to the many amazing professors at Berkeley whom I had the privilege of learning from. Some of my favorite courses were delivered by Michael Christ, Mark Haiman, Olga Holtz, and Ken Ribet. I also benefited from interactions and advice from Ani Adhikari, Steve Evans, Noah Giansiracusa, Alex Hening, and Allan Sly.

I am thankful to the math department staff, especially Vicky Lee, who always went above and beyond her duties as a graduate advisor to assist me and my peers.

I am thankful to my fellow graduate students in and around the math department, especially Johnny, Patrick, Hong, Wenpin, and Zhiyi, for many helpful conversations.

I am thankful to the many other wonderful people who have supported me through this endeavor, especially Min, Alex, Chris, Mike, Steven, Bon-Soon, Alvin, Kiho, Fu, Jennifer, and of course my family.

Chapter 1

Introduction

The theory of convex geometry has applications to nearly every area of mathematics as well as many other disciplines, with no shortage of contributions in probability theory. One popular theme is the characterization of extreme points of various convex sets of probability measures, such as moment sets, invariant measures, and many more. Extreme points are of intrinsic interest due to the elementary notion that any point in a convex body should be representable as a weighted average of extreme points. This is formally a theorem due to Carathéodory [10] for compact convex sets in Euclidean space, and was generalized to infinite-dimensional spaces via Choquet theory, which studies representations of points in compact convex sets as a generalized weighted average of extreme points in the form of an integral. This perspective has proven to be of great value to probabilists; for example, using Choquet theory, Hewitt and Savage [38] proved a more general version of de Finetti's theorem [30] for exchangeable sequences, and Johansen [41] gave a new proof of the Lévy-Khintchine formula [48] for infinitely divisible distributions. Thus, two of the most celebrated theorems in modern probability theory are instances of extreme point problems for convex sets.

This thesis covers three main topics in this general framework of convex geometry in probabilistic settings. The remainder of this introductory chapter contains a background section covering some standard concepts in convex geometry and probability, and an outline of the main chapters.

1.1 Preliminaries

This section provides a minimal treatment of standard definitions and theorems in convex geometry and probability which are fundamental to the content of this thesis.

Convex geometry

Let V be a real vector space. Recall that a set $C \subseteq V$ is called *convex* if it contains the line segment connecting any pair of points in C, or formally

$$\alpha x + (1-\alpha)y \in C \text{ for every } x, y \in C \text{ and every } \alpha \in [0,1]. \tag{1.1}$$

It easily follows from the definition that if C is convex then any *convex combination*

$$\alpha_1 x_1 + \ldots + \alpha_n x_n, \qquad \alpha_1, \ldots, \alpha_n \geq 0,\ \alpha_1 + \ldots + \alpha_n = 1 \tag{1.2}$$

of elements $x_1, \ldots, x_n \in C$ also belongs to C.

A point p in a convex set C is called an *extreme point* of C if it does not admit the representation (1.1) for any pair of points x and y not equal to p. We denote the set of extreme points of C by $\text{ext}(C)$.

The *convex hull* of a set S, denoted $\text{conv}(S)$, is the set of all convex combinations of elements in S. It is the smallest convex set containing S, or formally the intersection of all convex sets containing S. The *closed convex hull* of S is the (topological) closure of the convex hull.

The *dimension* of a convex set C, denoted $\dim C$, is the smallest dimension of an affine subspace containing C.

The convex hull of a finite set of points S is called a *convex polytope*, whose set of extreme points is a subset of S. Equivalently, a convex polytope is a closed and bounded convex set with a finite number of extreme points, which are called *vertices*. A k-dimensional convex polytope with $k+1$ vertices is called a *simplex*.

A *convex cone* is a convex set which is closed under multiplication by positive scalars, or equivalently, a set which is closed under linear combinations with positive coefficients. A closed convex cone has the origin as its only extreme point.

If V is finite-dimensional, i.e. V is isomorphic to $\mathbb{R}^n$, then a *hyperplane* is an affine subspace of dimension $n-1$. A *closed half-space* is the set of points which lie on and to one side of a hyperplane.

A *supporting hyperplane* of a set S is a hyperplane H such that

- S is contained in a closed half-space bounded by the hyperplane, and
- $H \cap \partial S \neq \emptyset$, where ∂S denotes the boundary of S.

Theorem 1.1.1 (Supporting hyperplane theorem)**.** *Let C be a convex set in $\mathbb{R}^n$. If $x \in \partial C$, then there exists a supporting hyperplane of C containing x.*

A consequence of the supporting hyperplane theorem is that closed convex sets can be characterized as intersections of closed half-spaces. For example, convex polytopes can also be characterized as bounded convex sets which are finite intersections of closed half-spaces. If H is a supporting hyperplane of a convex polytope C, then $C \cap H$ is called a *facet* of C if $\dim C \cap H = \dim C - 1$.

Theorem 1.1.2 (Minkowski-Carathéodory theorem)**.** *Let C be a k-dimensional compact convex subset of $\mathbb{R}^n$. Then every point in C is a convex combination of at most $k+1$ elements of its extreme points. In particular, C is the convex hull of its extreme points.*

The Minkowski-Carathéodory theorem generalizes to infinite-dimensional topological vector spaces V which are *Hausdorff*, meaning that the topology separates points, and *locally convex*, meaning that it a neighborhood base consisting of convex open sets. For example, any normed vector space satisfies these criteria.

Theorem 1.1.3 (Choquet's theorem)**.** *Let C be a compact convex subset of a Hausdorff locally convex vector space V. Then for each $x \in C$, there exists a probability measure μ on C with $\mu(C \setminus \operatorname{ext}(C)) = 0$ such that*

$$x = \int_{\operatorname{ext}(C)} y \; d\mu(y) \tag{1.3}$$

in the weak sense, i.e. for every continuous linear functional f on C, we have

$$f(x) = \int_{\operatorname{ext}(C)} f(y) \; d\mu(y). \tag{1.4}$$

Theorem 1.1.4 (Krein-Milman theorem)**.** *Let C be a compact convex subset of a Hausdorff locally convex vector space V. Then C is the closed convex hull of its extreme points.*

Beyond intrinsic appeal from a geometric perspective, convex sets and their extreme points are of fundamental importance in convex optimization. If C is a convex set, a function $f : C \to \mathbb{R}$ is called *convex* if for all $x, y \in C$ and $0 \leq \alpha \leq 1$,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y). \tag{1.5}$$

Theorem 1.1.5. *Let $C \subseteq V$ be a compact convex set and let $f : C \to \mathbb{R}$ be a convex function. Then*

$$\max_{x \in C} f(x) = \max_{x \in \operatorname{ext}(C)} f(x). \tag{1.6}$$

In words, the problem of maximizing a convex function over a compact convex set can be reduced to maximizing the function over just the set of extreme points. In particular, if f is linear, then the maximization or minimization problem over a compact convex set can be simplified in this manner.

For a more thorough treatment of the material in this section, see e.g. [70].

Probability

We assume that the reader is familiar with basic terminology in measure theory, topology, and probability theory. For background, refer to e.g. [31] and [22].

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

The *law* or *distribution* of a random variable $X : \Omega \to \mathbb{R}$ is the probability measure μ on $\mathbb{R}$ specified by

$$\mu(B) = \mathbb{P}(X \in B) \tag{1.7}$$

for every (Borel) measurable subset $B \subseteq \mathbb{R}$.

The *support* of a probability measure μ on $\mathbb{R}$, denoted $\text{supp}(\mu)$, is the smallest closed set $S \subseteq \mathbb{R}$ such that $\mu(S) = 1$. We say μ is *supported on* a set T if $\text{supp}(\mu) \subseteq T$.

A sequence of probability measures (μ_n) on $\mathbb{R}$ is said to *converge weakly* to μ if for every bounded, continuous function $f : \mathbb{R} \to \mathbb{R}$ we have

$$\int f(x) d\mu_n(x) \to \int f(x) d\mu(x). \tag{1.8}$$

The topology of weak convergence on the set of probability measures on $\mathbb{R}$ is called the *weak topology*. It is a well-known result in functional analysis that if K is a compact subset of $\mathbb{R}$, then the set of probability measures supported on K is *weakly compact*. (See [31], specifically the Riesz representation theorem and Alaoglu's theorem.)

If $\mathcal{C}$ is a collection of probability measures on $\mathbb{R}$, let $\lambda(\mathcal{C})$ denote the σ-field on $\mathcal{C}$ generated by sets of form

$$\{\nu \in \mathcal{C} : \nu(B) \in A\} \tag{1.9}$$

for measurable $A, B \subseteq \mathbb{R}$. This is the smallest σ-field which guarantees measurability of functions on $\mathcal{C}$ of the form $\nu \mapsto \nu(B)$ for measurable $B \subseteq \mathbb{R}$. A probability measure μ on $\mathbb{R}$ is called a *mixture* over $\mathcal{C}$ if there exists a probability measure ρ on $(\mathcal{C}, \lambda(\mathcal{C}))$ such that

$$\mu(B) = \int_{\mathcal{C}} \nu(B) d\rho(\nu) \tag{1.10}$$

for every measurable subset $B \subseteq \mathbb{R}$. In particular, if $\mathcal{C}$ is a finite set, then the integral in (1.10) reduces to a convex combination of the probability measures in $\mathcal{C}$.

Note that the preceding definitions can be extended from real-valued random variables to vector-valued random variables, as well as infinite sequences of random variables by way of Kolmogorov's extension theorem; see [22].

A finite sequence $(X_1, \ldots, X_n)$ of random variables is called *exchangeable* if its law is invariant under permutations of the indices, i.e. for every permutation σ of $\{1, \ldots, n\}$, we have the equality in distribution

$$(X_{\sigma(1)}, \ldots, X_{\sigma(n)}) \stackrel{d}{=} (X_1, \ldots, X_n). \tag{1.11}$$

An infinite sequence $(X_1, X_2, \ldots)$ of random variables is called exchangeable if $(X_1, \ldots, X_n)$ is exchangeable for every $n \geq 1$.

Theorem 1.1.6 (de Finetti's theorem [30],[38])**.** *Let $(X_1, X_2, \ldots)$ be an infinite exchangeable sequence of random variables and let $\mathcal{C}$ denote the set of (Borel) probability measures on $\mathbb{R}$. Then there exists a unique probability measure ρ on $(\mathcal{C}, \lambda(\mathcal{C}))$ such that for every $n \geq 1$ and measurable sets $B_1, \ldots, B_n$,*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \int_{\mathcal{C}} \mu(B_1) \cdots \mu(B_n) d\rho(\mu). \tag{1.12}$$

In words, de Finetti's theorem asserts that the law of every exchangeable sequence of random variables is a mixture of laws of i.i.d. sequences of random variables. Geometrically, the assertion looks the part of a Choquet extreme point representation theorem; indeed, Hewitt and Savage [38] proved the general version of de Finetti's theorem taking this approach, showing that the extreme points of the set of exchangeable laws are precisely those derived from i.i.d. sequences.

1.2 Organization

The rest of this thesis is organized as follows:

- In Chapter 2, we study the set of laws of K_n for an exchangeable sequence of random variables, where K_n is the number of distinct values among the first n terms. We prove for $n = 3$ that the extreme points of the convex set of all possible laws of K_3 are those derived from i.i.d. sampling from discrete uniform distributions and the limit case with $\mathbb{P}(K_3 = 3) = 1$. We also consider variants of the problem for finite exchangeable sequences and exchangeable random partitions. In particular, we discover a remarkable symmetry for laws of K_3 corresponding to the Ewens-Pitman [64] two-parameter partition model. We also discuss recent contributions by Yakubovich [77], in response to our main result for $n = 3$, toward former conjectures for the higher dimensional problem. The material in this chapter was published in [78].

- In Chapter 3, we introduce the notion of a coherent pair of random variables (X, Y), or two conditional probabilities of the same event, and study the convex set of laws on $[0, 1]^2$ arising in this manner. We respond to a number of open problems posed in [20] and [9]; specifically, we classify all extreme points of this set which are supported on the corners of a rectangle, which leads to an alternate proof of a maximization problem recently solved in [9]. We also discover a large class of extreme points with finite and countably infinite support. The general problem of describing all extreme points remains open.

- In Chapter 4, we study the convex set $\mathcal{D}_n$, the set of polynomial probability densities on $[0, 1]$ of degree at most n, drawing connections to known results in algebra, geometry, analysis, and probability. We introduce the Lorentz degree of a polynomial, which measures the representability of positive polynomials as a positive linear combination of Bernstein basis polynomials, and map out the geometry of $\mathcal{D}_2$ which uncovers a collection of ellipses related to this notion. We also consider a novel model for random polynomial densities and compute the upper envelope of $\mathcal{D}_n$ using orthogonal polynomials.

Chapter 2

Clustering in exchangeable processes

For an infinite sequence of real-valued random variables $(X_1, X_2, \ldots)$, let

$$K_n = K_n(X_1, \ldots, X_n) := \#\{X_i : 1 \leq i \leq n\}, \tag{2.1}$$

the number of distinct values appearing in the first n terms. We focus our discussion on the case in which the sequence $(X_1, X_2, \ldots)$ is *exchangeable*, meaning that its distribution is invariant under finite permutations of the indices. It is a well-known and celebrated result of de Finetti that every infinite exchangeable sequence is a mixture of i.i.d. sequences. We explore ideas related to the following central question:

> Given a probability distribution $(a_1, \ldots, a_n)$ on $[n] := \{1, \ldots, n\}$, is there an infinite exchangeable sequence of random variables $(X_1, X_2, \ldots)$ such that $\mathbb{P}(K_n = k) = a_k$ for $1 \leq k \leq n$?

The functional K_n has been studied extensively in the context of the *occupancy problem* as well as other closely related formulations including the birthday problem, the coupon collector's problem, and random partition structures [29, 45, 63]. Much of the literature pertains to the asymptotic behavior of K_n in the classical version in which the X_i are i.i.d. discrete uniform random variables, as well as the general i.i.d. case. See [33] for a recent survey with many references. Asymptotics of K_n have also been studied for a random walk $(X_1, X_2, \ldots)$ with stationary increments [71],[22, Section 7.3].

Let us first consider the problem for small values of n. For $n = 1$, the random variable K_1 is just the constant 1. Next, it is easy to see that every probability distribution on $\{1, 2\}$ can be achieved as the law of K_2 for some exchangeable sequence; indeed, for $a \in [0, 1]$, i.i.d. sampling from a distribution with a single atom having weight $\sqrt{a}$ yields $\mathbb{P}(K_2 = 1) = a$. However, the problem is not trivial for $n = 3$, as evident by the following bound due to Jim Pitman (personal communication.) The proof is presented in Section 2.2.

Proposition 2.0.1. *For K_3 the number of distinct values in the first 3 terms of an infinite exchangeable sequence of random variables $(X_1, X_2, \ldots)$,*

$$\mathbb{P}(K_3 = 2) \leq \frac{3}{4}. \tag{2.2}$$

Here we present the main result of this chapter. Let $\boldsymbol{v}_{n,m}$ denote the law of $K_{n,m} := K_n(X_{m,1}, \ldots, X_{m,n})$ where $X_{m,i}$ are i.i.d. with uniform distribution on m elements, i.e.

$$\boldsymbol{v}_{n,m} = \big(\mathbb{P}(K_{n,m} = k) : 1 \leq k \leq n\big) \tag{2.3}$$

and let $\boldsymbol{v}_{n,\infty} = (0, \ldots, 0, 1)$, corresponding to the limit case $m = \infty$ since

$$\mathbb{P}(K_{n,m} = n) = \frac{m(m-1)\cdots(m-n+1)}{m^n} \longrightarrow 1 \quad \text{as } m \to \infty. \tag{2.4}$$

Let

$$V_n := \{\boldsymbol{v}_{n,m} : m = 1, 2, \ldots, \infty\} \tag{2.5}$$

and let H_n denote the convex hull of V_n.

Theorem 2.0.2. *For $n = 3$,*

(i) The set of extreme points of H_n is V_n.

(ii) The set of possible laws of K_n for an infinite exchangeable sequence $(X_1, X_2, \ldots)$ is H_n.

It is natural to conjecture that the assertions in Theorem 2.0.2 hold true for larger values of n. Yuri Yakubovich [77] proved that (i) holds for all $n \geq 3$. However, Yakubovich exhibits a counterexample to (ii) for $n = 7$. The results in [77] are further discussed in Section 2.3. It remains a conjecture that (ii) holds for $n = 4, 5$ and fails for all $n \geq 6$, and more generally it remains an open problem to characterize the the set of possible laws of K_n for $n \geq 4$.

The rest of this chapter is organized as follows. Section 2.1 establishes notation and the fundamentals of our approach. Section 2.2 covers some properties of the law of K_3 leading to a proof of Theorem 2.0.2, and Section 2.3 extends some of these results to higher dimensions. Section 2.4 considers a variant of the main problem for finite exchangeable sequences by appealing to the framework of exchangeable random partitions, and Section 2.5 explores a remarkable symmetry for K_3 in the Ewens-Pitman two-parameter partition model.

2.1 Preliminaries

For an i.i.d. sequence $(X_1, X_2, \ldots)$, there is an associated *ranked discrete distribution* $(p_1, p_2, \ldots)$ with $p_1 \geq p_2 \geq \ldots \geq 0$ and $\sum_{i=1}^{\infty} p_i \leq 1$ where the p_i are the weights of the atoms for the law of X_i in decreasing order, and $1 - \sum_{i=1}^{\infty} p_i$ is the weight of the continuous component.

Consider the set

$$\nabla_\infty := \Big\{(p_1, p_2, \ldots) : p_1 \geq p_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} p_i \leq 1\Big\}, \tag{2.6}$$

sometimes referred to as the infinite dimensional *Kingman simplex* as in [62]. The uniform distribution on m elements corresponds to

$$\boldsymbol{u}_m := \Big(\underbrace{\frac{1}{m}, \ldots, \frac{1}{m}}_{m \text{ times}}, 0, 0, \ldots\Big) \in \nabla_\infty. \tag{2.7}$$

and every non-atomic law corresponds to $\boldsymbol{u}_\infty := (0, 0, \ldots) \in \nabla_\infty$. With Theorem 2.0.2 in mind, note that

$$\big\{\boldsymbol{u}_m : m = 1, 2, \ldots, \infty\big\} \tag{2.8}$$

is precisely the set of extreme points of ∇_∞ [16, Theorem 4.1]. Every $(p_1, p_2, \ldots) \in \nabla_\infty$ has a unique representation as a convex combination of $\boldsymbol{u}_m$, $m = 1, 2, \ldots, \infty$ given by

$$(p_1, p_2, \ldots) = p_* \boldsymbol{u}_\infty + \sum_{i=1}^{\infty} (p_i - p_{i+1}) \boldsymbol{u}_i, \qquad p_* = 1 - \sum_{i=1}^{\infty} p_i. \tag{2.9}$$

This is a discrete version of Khintchine's representation theorem for unimodal distributions [48].

It is easy to see that the law of K_n for an i.i.d sequence depends only on the ranked frequencies of the atoms. Let

$$q_{n,i}(p_1, p_2, \ldots) := \mathbb{P}\big(K_n = i\big) \tag{2.10}$$

where $K_n = K_n(X_1, \ldots, X_n)$ for i.i.d. X_i with ranked frequencies $(p_1, p_2, \ldots)$. Then for $n = 3$, it is easy to see that

$$q_{3,1}(p_1, p_2, \ldots) = \sum_{i=1}^{\infty} p_i^3 \tag{2.11}$$

$$q_{3,2}(p_1, p_2, \ldots) = \sum_{i=1}^{\infty} 3p_i^2(1 - p_i) \tag{2.12}$$

$$q_{3,3}(p_1, p_2, \ldots) = 1 - \sum_{i=1}^{\infty} \big[3p_i^2 - 2p_i^3\big]. \tag{2.13}$$

For the general exchangeable case, de Finetti's theorem guarantees that the law of K_n for an exchangeable sequence of random variables $(X_1, X_2, \ldots)$ is a mixture of laws of K_n for i.i.d. sequences. In other words, the set of laws of K_n derived from exchangeable sequences is the convex hull of those derived from i.i.d. sequences. This property allows us to focus on the i.i.d. case and simplify our treatment to ranked discrete distributions.

Note that there is an equivalent reformulation of the problem in the setting of exchangeable random partitions; see e.g. [63] for relevant background on the subject. For an exchangeable random partition $\Pi = (\Pi_n)$ of $\mathbb{N}$, let K_n denote the number of *clusters* in the restriction Π_n of Π to $[n]$. Through *Kingman's representation theorem* [50] for exchangeable random partitions of $\mathbb{N}$ in terms of random ranked discrete distributions, the possible laws of K_n in this setting are identical to the possible laws of K_n as defined earlier as the number of distinct values in the first n terms of an exchangeable sequence $(X_1, X_2, \ldots)$. In Sections 5 – 7, we explore some related problems in the framework of exchangeable random partitions.

Notations and conventions. If a ranked discrete distribution $(p_1, p_2, \ldots)$ has finitely many atoms, i.e. there exists m such that $p_i = 0$ for all $i > m$, we call it a *finite* distribution and abbreviate it as $(p_1, \ldots, p_m)$ when convenient. Since all of the functionals that we work with on ∇_∞ are symmetric functions of the arguments, we understand an equivalence between an unordered discrete distribution $(p_1, p_2, \ldots)$ and its ranked version. Unless otherwise stated, it is implicit in the appearance of $(p_1, p_2, \ldots)$ or $(p_1, \ldots, p_m)$ that the conditions $p_i \geq 0$ and $\sum p_i \leq 1$ hold.

2.2 Laws of K_3

To simplify notation in this section, let

$$q_i := q_{3,i} = \mathbb{P}(K_3 = i) \tag{2.14}$$

where q_i may be treated as a functional on ∇_∞.

Lemma 2.2.1. *For $(p_1, \ldots, p_m)$ with $m \geq 3$ and $p_1 \leq \ldots \leq p_m$,*

$$q_2(p_1 + p_2, p_3 \ldots, p_m) \geq q_2(p_1, p_2, p_3, \ldots, p_m). \tag{2.15}$$

Proof. Let $a = p_1$ and $b = p_2$. We have

$$q_2(a, b, p_3, \ldots, p_m) = 3a^2(1-a) + 3b^2(1-b) + \sum_{i=3}^{m} 3p_i^2(1-p_i) \tag{2.16}$$

and

$$q_2(a+b, p_3, \ldots, p_m) = 3(a+b)^2(1-a-b) + \sum_{i=3}^{m} 3p_i^2(1-p_i). \tag{2.17}$$

Then

$$\begin{aligned} q_2(a+b, p_3, \ldots, p_m) - q_2(a, b, p_3, \ldots, p_m) &= 3(a+b)^2(1-a-b) - 3a^2(1-a) - 3b^2(1-b) && (2.18) \\ &= 6ab(1-a-b) - 3a^2b - 3ab^2 && (2.19) \\ &= 3ab(2 - 3(a+b)) && (2.20) \\ &\geq 0 && (2.21) \end{aligned}$$

since a and b are the two smallest values among $\{a, b, p_3, \ldots, p_m\}$ so $a + b \leq \frac{2}{m} \leq \frac{2}{3}$ for $m \geq 3$. □

This shows that for every $(p_1, \ldots, p_m)$ with $m \geq 3$, merging the two smallest values among $\{p_1, \ldots, p_m\}$ does not decrease q_2.

Proof of Proposition 2.0.1. By de Finetti's theorem, it suffices to prove the inequality for i.i.d. sequences. Since

$$q_2(p_1, p_2, \ldots) = \sum_{i=1}^{\infty} 3p_i^2(1-p_i) = \lim_{m\to\infty} \sum_{i=1}^{m} 3p_i^2(1-p_i) = \lim_{m\to\infty} q_2(p_1, \ldots, p_m), \tag{2.22}$$

it is enough to establish the inequality $q_2(p_1, \ldots, p_m) \leq \frac{3}{4}$ for finite discrete distributions $(p_1, \ldots, p_m)$. If $m = 2$, then $q_2(p_1, p_2) = 3p_1^2(1-p_1) + 3p_2^2(1-p_2)$ which attains its maximum value of $\frac{3}{4}$ subject to $p_1, p_2 \geq 0$ and $p_1 + p_2 \leq 1$ at $p_1 = p_2 = \frac{1}{2}$. For $m \geq 3$, by Lemma 2.2.1 repeatedly merging the two smallest values until no more than two nonzero values remain gives $q_2(p_1, \ldots, p_m) \leq q_2(\frac{1}{2}, \frac{1}{2}) = \frac{3}{4}$. □

Consider the law of K_3 for an i.i.d. sequence $(X_1, X_2, \ldots)$ where each X_i has the uniform distribution $\boldsymbol{u}_N := (\frac{1}{N}, \ldots, \frac{1}{N})$. A probability distribution (q_1, q_2, q_3) of K_3 (on $\{1, 2, 3\}$) can be represented by any pair of its coordinates; here we shall work with $(q_1, q_3) := \big(\mathbb{P}(K_3 = 1), \mathbb{P}(K_3 = 3)\big)$. Then

$$q_1(\boldsymbol{u}_N) := \mathbb{P}(K_3(\boldsymbol{u}_N) = 1) = \frac{1}{N^2} \tag{2.23}$$

$$q_3(\boldsymbol{u}_N) := \mathbb{P}(K_3(\boldsymbol{u}_N) = 3) = \frac{(N-1)(N-2)}{N^2}. \tag{2.24}$$

The set of points $\{\boldsymbol{v}_N : N \in \mathbb{N}\} = \{(1,0), (\frac{1}{4}, 0), (\frac{1}{9}, \frac{2}{9}), (\frac{1}{16}, \frac{6}{16}), (\frac{1}{25}, \frac{12}{25}), (\frac{1}{36}, \frac{20}{36}), \ldots\}$ where

$$\boldsymbol{v}_N := (q_1(\boldsymbol{u}_N), q_3(\boldsymbol{u}_N)) = \left(\frac{1}{N^2}, \frac{(N-1)(N-2)}{N^2}\right) \tag{2.25}$$

are shown in Figures 2.1 and 2.2, with line segments connecting consecutive points.

The slope of the line connecting $\boldsymbol{v}_N = \big(\frac{1}{N^2}, \frac{(N-1)(N-2)}{N^2}\big)$ and $\boldsymbol{v}_{N+1} = \big(\frac{1}{(N+1)^2}, \frac{N(N-1)}{(N+1)^2}\big)$ is

$$\frac{\frac{N(N-1)}{(N+1)^2} - \frac{(N-1)(N-2)}{N^2}}{\frac{1}{(N+1)^2} - \frac{1}{N^2}} = -\frac{(N-1)(3N+2)}{2N+1}. \tag{2.26}$$

This is increasing in N which proves Theorem 2.0.2(i). The equation of the Nth line is given by

$$q_3 - \frac{(N-1)(N-2)}{N^2} = -\frac{(N-1)(3N+2)}{2N+1}\left(q_1 - \frac{1}{N^2}\right) \tag{2.27}$$

or after rearranging,

$$q_3 + \frac{(N-1)(3N+2)}{2N+1} q_1 = \frac{2N-2}{2N+1}. \tag{2.28}$$

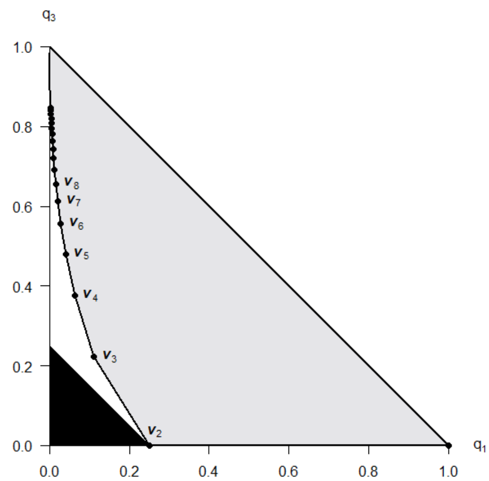


Figure 2.1: Probability distributions of K_3 represented as points $(q_1, q_3) = \big(\mathbb{P}(K_3 = 1), \mathbb{P}(K_3 = 3)\big)$ with q_1 horizontal and q_3 vertical. Shaded in black is the restricted region specified by Proposition 2.0.1. The gray region is the closed convex hull of $\{\boldsymbol{v}_N : N \in \mathbb{N}\}$ where $\boldsymbol{v}_N$ corresponds to the distribution of K_3 for i.i.d. sampling from a discrete uniform distribution on N elements, as defined in (2.25).

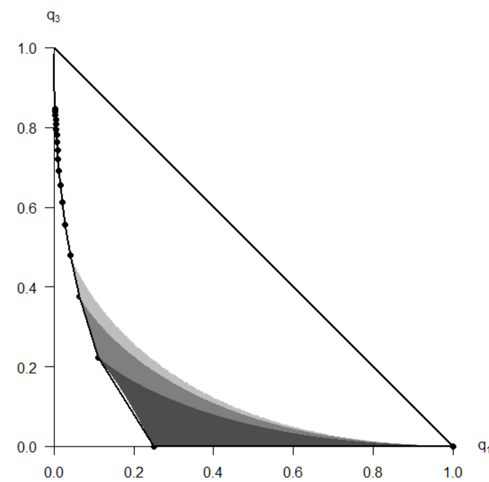


Figure 2.2: The shaded regions (nested) correspond to the images of $\{(p_1, \dots, p_m) : p_i \geq 0, \sum p_i = 1\}$ under the map $(p_1, \dots, p_m) \mapsto \big(q_1(p_1, \dots, p_m), q_3(p_1, \dots, p_m)\big)$, i.e. distributions of K_3 for i.i.d. sampling from a discrete distribution with at most m atoms for $m = 3$ (dark), $m = 4$ (dark and medium), and $m = 5$ (dark, medium, and light). The existence of the gap between the left boundary of the dark region and the line segment connecting $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ is a consequence of Lemma 2.2.6. The midpoint of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$, for example, does not correspond to i.i.d. sampling from any discrete distribution; however, it does correspond to the the exchangeable sequence with law given by i.i.d. sampling from $\boldsymbol{u}_2$ with probability $\frac{1}{2}$ and i.i.d. sampling from $\boldsymbol{u}_3$ with probability $\frac{1}{2}$.

For $\boldsymbol{p} = (p_1, \dots, p_m)$, define according to the left-hand side of (2.28) the functional

$$L_N(\boldsymbol{p}) := q_3(\boldsymbol{p}) + \frac{(N-1)(3N+2)}{2N+1} q_1(\boldsymbol{p}) \tag{2.29}$$

which may be reexpressed as

$$
\begin{aligned}
L_N(\boldsymbol{p}) &= 1 - \big(1 - L_N(\boldsymbol{p})\big) && (2.30)\\
&= 1 - \left(1 - q_3(\boldsymbol{p}) - q_1(\boldsymbol{p}) - \left[\frac{(N-1)(3N+2)}{2N+1} - 1\right] q_1(\boldsymbol{p})\right) && (2.31)\\
&= 1 - q_2(\boldsymbol{p}) + \frac{3(N^2-N-1)}{2N+1} q_1(\boldsymbol{p}) && (2.32)\\
&= 1 - \sum_{i=1}^{m} 3p_i^2(1-p_i) + \frac{3(N^2-N-1)}{2N+1}\sum_{i=1}^{m} p_i^3 && (2.33)\\
&= 1 - 3\sum_{i=1}^{m} p_i^2 + \frac{3N(N+1)}{2N+1}\sum_{i=1}^{m} p_i^3. && (2.34)
\end{aligned}
$$

Note that L_N is nonlinear as a function of discrete distributions $\boldsymbol{p}$. Define

$$f(N) := \frac{3N(N+1)}{2N+1} \tag{2.35}$$

so

$$L_N(\boldsymbol{p}) = 1 - 3\sum_{i=1}^{m} p_i^2 + f(N)\sum_{i=1}^{m} p_i^3. \tag{2.36}$$

To better understand the sequence of values $f(N)$, note that f is increasing and

$$N < \frac{2N+2}{2N+1}(N) = \frac{2}{3}\cdot \underbrace{\frac{3N(N+1)}{2N+1}}_{f(N)} = \frac{2N}{2N+1}(N+1) < N+1. \tag{2.37}$$

The first few values are $f(1) = 2$, $f(2) = \frac{18}{5}$, $f(3) = \frac{36}{7}$, $f(4) = \frac{60}{9}$.

Lemma 2.2.2. *For $N \geq 1$ and every $\boldsymbol{p} = (p_1, \ldots, p_m)$ with $p_1 \geq \ldots \geq p_m \geq 0$ and $\sum p_i \leq 1$,*

$$L_N(\boldsymbol{p}) \geq \frac{2N-2}{2N+1}. \tag{2.38}$$

Geometrically, Lemma 2.2.2 asserts that for every $\boldsymbol{p} = (p_1, \ldots, p_m)$, the point $\big(q_1(\boldsymbol{p}), q_3(\boldsymbol{p})\big)$ lies on or above each of the lines connecting $\boldsymbol{v}_N$ and $\boldsymbol{v}_{N+1}$ for $N \in \mathbb{N}$. It will be shown in the proof that for $N \geq 2$, $L_N(\boldsymbol{p}) = \frac{2N-2}{2N+1}$ if and only if $\boldsymbol{p} = \boldsymbol{u}_N$ or $\boldsymbol{p} = \boldsymbol{u}_{N+1}$; as for $N = 1$, $L_1(\boldsymbol{p}) = q_3(\boldsymbol{p}) = 0$ is attained if and only if $\boldsymbol{p} = (p_1, p_2)$ with $p_1 + p_2 = 1$.

The strategy for proving Lemma 2.2.2 is to show that L_N is minimized at precisely $\boldsymbol{v}_N$ and $\boldsymbol{v}_{N+1}$ by reducing the domain of minimization in stages, first to $(p_1, \ldots, p_m)$ with $\sum p_i = 1$, then to the uniform distributions, and finally to $\boldsymbol{u}_N$ and $\boldsymbol{u}_{N+1}$. The key to the proof is the following *merging* lemma, which generalizes Lemma 2.2.1.

Lemma 2.2.3. *For $N \geq 1$ and $(p_1, \ldots, p_m)$ with $m \geq 2$,*

$$L_N(p_1 + p_2, p_3, \ldots, p_m) - L_N(p_1, p_2, p_3, \ldots, p_m) = 3p_1p_2\big[(p_1 + p_2)f(N) - 2\big] \tag{2.39}$$

which is positive, negative, or zero according to the sign of $p_1 + p_2 - \frac{2}{f(N)}$.

Proof. Let $a = p_1$ and $b = p_2$. We have

$$L_N(a, b, p_3, \ldots, p_m) = 1 - 3a^2 - 3b^2 - 3\sum_{i=3}^{m} p_i^2 + f(N)(a^3 + b^3) + f(N)\sum_{i=3}^{m} p_i^3 \tag{2.40}$$

and

$$L_N(a + b, p_3, \ldots, p_m) = 1 - 3(a + b)^2 - 3\sum_{i=3}^{m} p_i^2 + f(N)(a + b)^3 + f(N)\sum_{i=3}^{m} p_i^3. \tag{2.41}$$

Then

$$L_N(a + b, p_3, \ldots, p_m) - L_N(a, b, p_3, \ldots, p_m) = -6ab + f(N)(3a^2b + 3ab^2) \tag{2.42}$$
$$= 3ab\big[(a + b)f(N) - 2\big]. \tag{2.43}$$

□

The proof of Lemma 2.2.2 is organized according to the following lemmas.

Lemma 2.2.4. *Let $\mathcal{P}$ denote the set of all finite ranked discrete distributions, and let $\mathcal{P}^1$ denote the set of finite ranked discrete distributions $(p_1, \ldots, p_m)$ with $\sum p_i = 1$. Then for every $N \geq 1$, we have the equality of sets*

$$\arg\min_{\boldsymbol{p}\in\mathcal{P}} L_N(\boldsymbol{p}) = \arg\min_{\boldsymbol{p}\in\mathcal{P}^1} L_N(\boldsymbol{p}) \tag{2.44}$$

Proof. Let $\boldsymbol{p}_0 = (p_1, \ldots, p_m) \in \mathcal{P}$ such that $\sum_{i=1}^{m} p_i < 1$. Let ε satisfy

$$0 < \varepsilon < \min\Big\{\frac{3}{f(N)}, 1 - \sum_{i=1}^{m} p_i\Big\}. \tag{2.45}$$

Then

$$L_N(\varepsilon, p_1, \ldots, p_m) = -3\varepsilon^2 + f(N)\varepsilon^3 + L_N(p_1, \ldots, p_m) \tag{2.46}$$
$$= \varepsilon^2(f(N)\varepsilon - 3) + L_N(p_1, \ldots, p_m) \tag{2.47}$$
$$< L_N(p_1, \ldots, p_m). \tag{2.48}$$

This shows that if $\boldsymbol{p}_0 \notin \mathcal{P}^1$, then $\boldsymbol{p}_0 \notin \arg\min_{\boldsymbol{p}\in\mathcal{P}} L_N(\boldsymbol{p})$. □

Lemma 2.2.5. *Let $\mathcal{P}^1$ denote the set of finite ranked discrete distributions $(p_1, \ldots, p_m)$ with $\sum p_i = 1$, and let $\mathcal{U} := \{\boldsymbol{u}_m : m \in \mathbb{N}\}$. Then for $N \geq 2$, we have the equality of sets*

$$\arg\min_{\boldsymbol{p} \in \mathcal{P}^1} L_N(\boldsymbol{p}) = \arg\min_{\boldsymbol{p} \in \mathcal{U}} L_N(\boldsymbol{p}) \tag{2.49}$$

Proof. Let $\boldsymbol{p}_0 = (p_1, \ldots, p_m)$, not necessarily ranked, such that $\sum_{i=1}^m p_i = 1$. Suppose $\boldsymbol{p}_0$ has a pair of distinct nonzero values, say $a = p_1$ and $b = p_2$ with $a, b > 0$ and $a \neq b$. Consider the three cases as designated in Lemma 2.2.3, noting that $\frac{2}{f(N)} < 1$ for $N \geq 2$.

(i) If $a + b < \frac{2}{f(N)}$, then $L_N(a + b, p_3, \ldots, p_m) < L_N(a, b, p_3, \ldots, p_m)$ by Lemma 2.2.3.

(ii) If $a + b > \frac{2}{f(N)}$, then

$$\begin{aligned} & L_N(\tfrac{a+b}{2}, \tfrac{a+b}{2}, p_3, \ldots, p_m) - L_N(a, b, p_3, \ldots, p_m) && (2.50)\\ &= \big(L_N(a + b, p_3, \ldots, p_m) - L_N(a, b, p_3, \ldots, p_m)\big) && (2.51)\\ &- \big(L_N(a + b, p_3, \ldots, p_m) - L_N(\tfrac{a+b}{2}, \tfrac{a+b}{2}, p_3, \ldots, p_m)\big) \\ &= 3ab\big((a + b)f(N) - 2\big) - 3(\tfrac{a+b}{2})^2\big((a + b)f(N) - 2\big) && (2.52)\\ &= 3\big(ab - (\tfrac{a+b}{2})^2\big)\big((a + b)f(N) - 2\big) && (2.53)\end{aligned}$$

which is negative since $ab - (\frac{a+b}{2})^2 < 0$ and $(a + b)f(N) - 2 > 0$.

(iii) If $a + b = \frac{2}{f(N)} < 1$, then there must exist a third nonzero value, say $p_3 = c > 0$. If $c = \frac{2}{f(N)}$, then $a \neq c$ and $a + c > \frac{2}{f(N)}$ so $L_N(\frac{a+c}{2}, \frac{a+c}{2}, b, p_4, \ldots, p_m) < L_N(a, b, c, p_4, \ldots, p_m)$ by case (ii). If $c \neq \frac{2}{f(N)}$, then merging a and b, which does not change L_N, followed by averaging $a + b$ and c gives $L_N(\frac{a+b+c}{2}, \frac{a+b+c}{2}, p_4, \ldots, p_m) < L_N(a, b, c, p_4, \ldots, p_m)$ by case (ii) again.

Since permuting values in any discrete distribution does not change L_N, the analysis above holds for all ranked discrete distributions and thus shows that among $\boldsymbol{p} \in \mathcal{P}^1$, L_N cannot be minimized at any $\boldsymbol{p}$ with a pair of distinct nonzero values, i.e. any non-uniform distribution. □

Remark. As mentioned previously, for $N = 1$,

$$\arg\min_{\boldsymbol{p} \in \mathcal{P}} L_1(\boldsymbol{p}) = \{(p_1, p_2) : p_1 \geq p_2 \geq 0, p_1 + p_2 = 1\}$$

which differs from the general case $N \geq 2$. The reason the proof of Lemma 2.2.5 fails for $N = 1$ is that $f(1) = 2$, so $\frac{2}{f(1)} = 1$ and case (iii) of the proof breaks down.

Lemma 2.2.6. *Let $\mathcal{U} := \{\boldsymbol{u}_m : m \in \mathbb{N}\}$. Then for $N \geq 1$,*

$$\arg\min_{\boldsymbol{p} \in \mathcal{U}} L_N(\boldsymbol{p}) = \{\boldsymbol{u}_N, \boldsymbol{u}_{N+1}\} \tag{2.54}$$

Proof. The claim is obvious based on Figure 2.1, which shows that the slopes between $\boldsymbol{v}_N$ and $\boldsymbol{v}_{N+1}$ for $N \in \mathbb{N}$ are decreasing in N. Indeed, the slope of the Nth line segment is computed in (2.26) as

$$-\frac{(N-1)(3N+2)}{2N+1} = -\frac{3N^2-N-2}{2N+1} = 2 - \frac{3N(N+1)}{2N+1} = 2 - f(N) \tag{2.55}$$

which is decreasing in N. □

Proof of Lemma 2.2.2. The claim holds trivially for $N = 1$. For $N \geq 2$, applying Lemmas 2.2.4, 2.2.5, and 2.2.6 yields

$$\arg\min_{\boldsymbol{p}\in\mathcal{P}} L_N(\boldsymbol{p}) = \arg\min_{\boldsymbol{p}\in\mathcal{P}^1} L_N(\boldsymbol{p}) = \arg\min_{\boldsymbol{p}\in\mathcal{U}} L_N(\boldsymbol{p}) = \{\boldsymbol{u}_N, \boldsymbol{u}_{N+1}\} \tag{2.56}$$

and therefore for every $\boldsymbol{p} = (p_1, \ldots, p_m)$ with $p_i \geq 0$ and $\sum p_i \leq 1$,

$$L_N(\boldsymbol{p}) \geq L_N(\boldsymbol{u}_N) = L_N(\boldsymbol{u}_{N+1}) = \frac{2N-2}{2N+1}. \tag{2.57}$$

□

Proof of Theorem 2.0.2. Part (i) was proven earlier by the slope computation (2.26) and illustrated in Figure 2.1. For part (ii), Lemma 2.2.2 asserts that $(q_1(\boldsymbol{p}), q_2(\boldsymbol{p}), q_3(\boldsymbol{p})) \in H_3 = \mathrm{conv}(V_3)$ for every finite ranked discrete distribution $\boldsymbol{p}$. Extension to infinite discrete distributions $(p_1, p_2, \ldots)$ follows because $\lim_{m\to\infty} q_i(p_1, \ldots, p_m) = q_i(p_1, p_2, \ldots)$, and then extension to exchangeable sequences holds by convexity. □

2.3 Higher dimensions

This section aims to extend some of the results in the previous section to K_n for larger n. Here $q_{n,i} := \mathbb{P}(K_n = i)$. We begin by generalizing Lemma 2.2.1 and Proposition 2.0.1.

Lemma 2.3.1. *For $n \geq 3$ and $(p_1, \ldots, p_m)$ with $m \geq 3$, $\sum_{i=1}^m p_i = 1$, $p_1 \leq \ldots \leq p_m$,*

$$q_{n,2}(p_1 + p_2, p_3, \ldots, p_m) \geq q_{n,2}(p_1, p_2, p_3, \ldots, p_m). \tag{2.58}$$

The proof requires the following inequality:

Lemma 2.3.2. *For $a, b > 0$ and $n \geq 2$,*

$$4\left(\tfrac{n-1}{n}\right)ab(a+b)^{n-2} \leq (a+b)^n - a^n - b^n \leq nab(a+b)^{n-2} \tag{2.59}$$

Proof. We have

$$(a+b)^n - a^n - b^n = \sum_{k=1}^{n-1} \binom{n}{k} a^k b^{n-k} = ab \sum_{k=0}^{n-2} \binom{n}{k+1} a^k b^{n-2-k}. \tag{2.60}$$

Observe that

$$\binom{n}{k+1} = \frac{n(n-1)(n-2)!}{(k+1)k!(n-k-1)(n-k-2)!} = \frac{n(n-1)}{(k+1)(n-k-1)} \binom{n-2}{k}; \tag{2.61}$$

the denominator $(k+1)(n-k-1)$ is no greater than $(n/2)^2$, and is minimized at $k=0$ and $k=n-2$, so

$$\binom{n}{k+1} \geq \frac{n(n-1)}{(n/2)^2} \binom{n-2}{k} = 4\frac{n-1}{n} \binom{n-2}{k} \tag{2.62}$$

and

$$\binom{n}{k+1} \leq n \binom{n-2}{k}. \tag{2.63}$$

The result follows by substituting inequalities (2.62) and (2.63) into (2.60) and appealing to the binomial theorem. □

Proof of Lemma 2.3.1. Let $a = p_1$ and $b = p_2$. We can compute

$$q_{n,2}(a, b, p_3, \ldots, p_m) = \mathbb{P}\big(K_n(a, b, p_3, \ldots, p_m) = 2\big)$$

by conditioning on the appearance of the first two values:

$$q_{n,2}(a, b, p_3, \ldots, p_m) = \sum_{k=1}^{n-1} \binom{n}{k} a^k b^{n-k} + \sum_{k=1}^{n-1} \binom{n}{k} a^k \sum_{i=3}^{m} p_i^{n-k} \tag{2.64}$$

$$+ \sum_{k=1}^{n-1} \binom{n}{k} b^k \sum_{i=3}^{m} p_i^{n-k} + \sum_{3 \leq i < j \leq m} \sum_{k=1}^{n-1} \binom{n}{k} p_i^k p_j^{n-k}. \tag{2.65}$$

Note that the first term, which is an expression for the probability that the first two values both appear and are the only ones to appear in the first n observations, is also equal to $(a+b)^n - a^n - b^n$. Similarly,

$$q_{n,2}(a+b, p_3, \ldots, p_m) = \sum_{k=1}^{n-1} \binom{n}{k} (a+b)^k \sum_{i=3}^{m} p_i^{n-k} + \sum_{3 \leq i < j \leq m} \sum_{k=1}^{n-1} \binom{n}{k} p_i^k p_j^{n-k}. \tag{2.66}$$

For $m \geq 3$, the difference after appropriate cancellations and applying Lemma 2.3.2 is

$$q_{n,2}(a+b, p_3, \ldots, p_m) - q_{n,2}(a, b, p_3, \ldots, p_m) \tag{2.67}$$

$$= \sum_{k=1}^{n-1} \binom{n}{k} \left[(a+b)^k - a^k - b^k\right] \sum_{i=3}^{m} p_i^{n-k} - \sum_{k=1}^{n-1} \binom{n}{k} a^k b^{n-k} \tag{2.68}$$

$$= \underbrace{\sum_{k=1}^{n-2} \binom{n}{k} \left[(a+b)^k - a^k - b^k\right] \sum_{i=3}^{m} p_i^{n-k}}_{\geq 0} + n \underbrace{\left[(a+b)^{n-1} - a^{n-1} - b^{n-1}\right]}_{\geq 4(\frac{n-2}{n-1})ab(a+b)^{n-3} \geq 2ab(a+b)^{n-3}} \sum_{i=3}^{m} p_i \tag{2.69}$$

$$- \underbrace{\left[(a+b)^n - a^n - b^n\right]}_{\leq nab(a+b)^{n-2}} \tag{2.70}$$

$$\geq nab(a+b)^{n-3}\Big[2\sum_{i=3}^{m} p_i - (a+b)\Big]. \tag{2.71}$$

Since $\sum_{i=1}^{m} p_i = 1$ and $a \leq b \leq p_3 \leq \ldots \leq p_m$, it follows that $\sum_{i=3}^{m} p_i \geq \frac{m-2}{m}$ and $a + b \leq \frac{2}{m}$, so

$$2\sum_{i=3}^{m} p_i - (a+b) \geq 2\Big(\frac{m-2}{m}\Big) - \frac{2}{m} = \frac{2(m-3)}{m} \geq 0 \tag{2.72}$$

and therefore merging the two smallest values among $\{p_1, \ldots, p_m\}$ does not decrease $q_{n,2}$ provided that there are at least 3 nonzero values. □

Lemma 2.3.3. *For every $(p_1, \ldots, p_m)$ and $n \geq 3$,*

$$q_{n,2}(p_1, \ldots, p_m, p_*) \geq q_{n,2}(p_1, \ldots, p_m) \tag{2.73}$$

where $p_* := 1 - \sum_{i=1}^{m} p_i$.

Proof. We have

$$q_{n,2}(p_1, \ldots, p_m) = \sum_{1 \leq i < j \leq m} \sum_{k=1}^{n-1} \binom{n}{k} p_i^k p_j^{n-k} + \sum_{i=1}^{m} n p_i^{n-1} p_* \tag{2.74}$$

and

$$q_{n,2}(p_1, \ldots, p_m, p_*) = \sum_{1 \leq i < j \leq m} \sum_{k=1}^{n-1} \binom{n}{k} p_i^k p_j^{n-k} + \sum_{i=1}^{m} \sum_{k=1}^{n-1} \binom{n}{k} p_i^k p_*^{n-k}, \tag{2.75}$$

so

$$q_{n,2}(p_1, \ldots, p_m, p_*) - q_{n,2}(p_1, \ldots, p_m) = \sum_{i=1}^{m} \sum_{k=1}^{n-2} p_i^k p_*^{n-k} \geq 0. \tag{2.76}$$

□

Theorem 2.3.4. *For every exchangeable sequence of random variables* $(X_1, X_2, \ldots)$ *and every* $n \geq 3$,

$$\mathbb{P}(K_n = 2) \leq 1 - 2^{-(n-1)}. \tag{2.77}$$

Proof. As in the proof of Proposition 2.0.1, it suffices to show that $q_{n,2}(p_1, \ldots, p_m) \leq 1 - 2^{-(n-1)}$ for any $(p_1, \ldots, p_m)$. If $m = 2$ and $p_1 + p_2 = 1$, then

$$q_{n,2}(p_1, p_2) = 1 - p_1^n - p_2^n \tag{2.78}$$

which attains its maximum of $1 - 2^{-(n-1)}$ at $p_1 = p_2 = \frac{1}{2}$. For $m \geq 3$, by Lemmas 2.3.1 and 2.3.3 we have

$$q_{n,2}(p_1, \ldots, p_m) \leq q_{n,2}(p_1, \ldots, p_m, p_*) \leq q_{n,2}\big(\tfrac{1}{2}, \tfrac{1}{2}\big) = 1 - 2^{-(n-1)}. \tag{2.79}$$

□

The difficulty in extending the proof of Theorem 2.0.2(ii) to the problem in higher dimensions is that there is no simple generalization of Lemma 2.2.3. Lemma 2.2.3 is essential because it asserts that whether merging two values in a discrete distribution increases, decreases, or preserves the functionals L_N is determined by only the sum of the two value to be merged. The corresponding functionals for the higher dimensional problem are more complicated and do not have the same convenient property.

Recently, Yakubovich [77] resolved the previously standing conjecture regarding the assertions in Theorem 2.0.2 for $n \geq 3$.

Recall some notation from the beginning of the chapter: for $n \geq 3$ and $m = 1, 2, \ldots, \infty$, denote by $\boldsymbol{v}_{n,m}$ the law of $K_n(X_{m,1}, \ldots, X_{m,n})$ where $X_{m,i}$ are i.i.d. with uniform distribution on m elements, i.e.

$$\boldsymbol{v}_{n,m} = \big(\mathbb{P}(K_{n,m} = k) : 1 \leq k \leq n\big) \tag{2.80}$$

and $\boldsymbol{v}_{n,\infty} = (0, \ldots, 0, 1)$. By a standard combinatorial argument, we have the formula

$$\boldsymbol{v}_{n,m}(k) = \frac{S(n,k)\binom{m}{k}k!}{m^n} \qquad (1 \leq k \leq n) \tag{2.81}$$

where the $S(n, k)$ are Stirling numbers of the second kind. Let

$$V_n := \{\boldsymbol{v}_{n,m} : m = 1, 2, \ldots, \infty\} \tag{2.82}$$

and let H_n denote the convex hull of V_n. Yakubovich proved the following:

Proposition 2.3.5. *For* $n \geq 3$, *the set of extreme points of* H_n *is* V_n.

This is a consequence of the following two lemmas, in which orthogonal vectors to the supporting hyperplanes of H_n are found, revealing its geometry.

Lemma 2.3.6. *Let $n \geq 3$ be odd. Let $\boldsymbol{\gamma}_{n,1} := \delta_n = (0, \dots, 0, 1) \in \mathbb{R}^n$, and for $r \geq 2$ define $\boldsymbol{\gamma}_{n,r} \in \mathbb{R}^n$ by*

$$\boldsymbol{\gamma}_{n,r}(k) := \frac{(-1)^{k-1}\binom{n-1}{k-1}}{S(n,k)\binom{n+r-3}{k-1}(k-1)!} \qquad \text{for } k = 1, \dots, n. \tag{2.83}$$

Then for $r \geq 1$ we have $\langle \boldsymbol{\gamma}_{n,r}, \boldsymbol{v}_{n,m} \rangle = 0$ for $m = r, r+1, \dots, r+n-2$ and $\langle \boldsymbol{\gamma}_{n,r}, \boldsymbol{v}_{n,m} \rangle > 0$ for $m < r$ or $m > r+n-2$.

Proof. The assertion for $r = 1$ is obvious because the probability of observing n distinct values in a n-sample from a uniform distribution on m elements is 0 for $m = 1, \dots, n-1$ and positive for $m \geq n$. For $r \geq 2$, observe that

$$\langle \boldsymbol{\gamma}_{n,r}, \boldsymbol{v}_{n,m} \rangle = \sum_{k=1}^{n} \frac{(-1)^{k-1}\binom{n-1}{k-1}}{S(n,k)\binom{n+r-3}{k-1}(k-1)!} \frac{S(n,k)\binom{m}{k}k!}{m^n} = \sum_{k=1}^{n} \frac{(-1)^{k-1}\binom{m-1}{k-1}\binom{n-k+r-2}{r-2}}{\binom{n+r-3}{r-2}m^{n-1}} \tag{2.84}$$

because $k\binom{m}{k} = m\binom{m-1}{k-1}$ and $\binom{n-1}{k-1}\binom{n+r-3}{k-1}^{-1} = \binom{n-k+r-2}{r-2}\binom{n+r-3}{r-2}^{-1}$. Note that the denominator in the summand does not depend on k. It can be shown using generating functions that the numerator evaluates to 0 for $r \leq m \leq r+n-2$ and is otherwise positive for all odd n (proof omitted). □

Lemma 2.3.7. *Let $n \geq 4$ be even. Let $\boldsymbol{\gamma}''_{n,2} := \delta_n = (0, \dots, 0, 1) \in \mathbb{R}^n$. For $r \geq 2$ define $\boldsymbol{\gamma}'_{n,r} \in \mathbb{R}^n$ by*

$$\boldsymbol{\gamma}'_{n,r}(k) := \frac{(-1)^{k-1}\binom{n-2}{k-1}}{S(n,k)\binom{n+r-4}{k-1}(k-1)!} \qquad \text{for } k = 1, \dots, n-1, \tag{2.85}$$

$$\boldsymbol{\gamma}'_{n,r}(n) := 0 \tag{2.86}$$

and for $r \geq 3$ define $\boldsymbol{\gamma}''_{n,r} \in \mathbb{R}^n$ by

$$\boldsymbol{\gamma}''_{n,r}(1) := 0, \tag{2.87}$$

$$\boldsymbol{\gamma}''_{n,r}(k) := \frac{(-1)^{k}\binom{n-2}{k-2}}{S(n,k)\binom{n+r-5}{k-2}(k-2)!} \qquad \text{for } k = 2, \dots, n. \tag{2.88}$$

Then for $r \geq 2$ we have $\langle \boldsymbol{\gamma}'_{n,r}, \boldsymbol{v}_{n,m} \rangle = 0$ for $m = \infty, r, r+1, \dots, r+n-3$ and $\langle \boldsymbol{\gamma}'_{n,r}, \boldsymbol{v}_{n,m} \rangle > 0$ for $m < r$ or $r+n-3 < m < \infty$, and we have $\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,m} \rangle = 0$ for $m = 1, r, r+1, \dots, r+n-3$ and $\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,m} \rangle > 0$ for $1 < m < r$ or $m > r+n-3$.

Proof. First, for $r \geq 2$ we have $\langle \boldsymbol{\gamma}'_{n,r}, \boldsymbol{v}_{n,\infty} \rangle = 0$ since $\boldsymbol{v}_{n,\infty} = \delta_n$, and for $1 \leq m < \infty$ we have

$$\langle \boldsymbol{\gamma}'_{n,r}, \boldsymbol{v}_{n,m} \rangle = \sum_{k=1}^{n-1} \frac{(-1)^{k-1}\binom{n-2}{k-1}}{S(n,k)\binom{n+r-4}{k-1}(k-1)!} \frac{S(n,k)\binom{m}{k}k!}{m^n} = \sum_{k=1}^{n-1} \frac{(-1)^{k-1}\binom{m-1}{k-1}\binom{n-k+r-3}{r-2}}{\binom{n+r-4}{r-2}m^{n-1}} \tag{2.89}$$

which is up to a factor of m the same as (2.84) with n replaced by $n-1$. Therefore it evaluates to 0 for $r \le m \le r+n-3$ and is positive for other integer values of m for all even n.

Next, the assertion about $\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,m} \rangle$ for $r=2$ holds by the same reasoning as in the case for n odd and $r=1$. For $r \ge 3$ we have $\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,\infty} \rangle > 0$ and for $1 \le m < \infty$ we have

$$\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,m} \rangle = \sum_{k=2}^{n} \frac{(-1)^k \binom{n-2}{k-2}}{S(n,k)\binom{n+r-5}{k-2}(k-2)!} \frac{S(n,k)\binom{m}{k}k!}{m^n} \tag{2.90}$$

$$= \sum_{k=2}^{n} \frac{(-1)^k (m-1)\binom{n-k+r-3}{r-3}\binom{m-2}{k-2}}{\binom{n+r-5}{r-3} m^{n-1}} \tag{2.91}$$

$$= \sum_{k=1}^{n-1} \frac{(-1)^{k-1} (m-1)\binom{n-k+r-4}{r-3}\binom{m-2}{k-1}}{\binom{n+r-5}{r-3} m^{n-1}} \tag{2.92}$$

because $k(k-1)\binom{m}{k} = m(m-1)\binom{m-2}{k-2}$ and $\binom{n-2}{k-2}\binom{n+r-5}{k-2}^{-1} = \binom{n-k+r-3}{r-3}\binom{n+r-5}{r-3}^{-1}$. We see that $\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,m} \rangle = 0$ for $m=1$. For $m>1$ by shifting the variables accordingly, specifically $r-1 \mapsto r$ and $m-1 \mapsto m$, we obtain (2.89) up to a positive factor and thus $\langle \boldsymbol{\gamma}''_{n,r}, \boldsymbol{v}_{n,m} \rangle = 0$ for $r \le m \le r+n-3$, and is positive for other integer values of $m>1$ and $m=\infty$ by definition of $\boldsymbol{\gamma}''_{n,r}$. □

The set $H_n \in \mathbb{R}^n$ is a $(n-1)$-dimensional *apeirotope*, or a generalized polytope which has infinitely many facets, lying in the $(n-1)$-dimensional affine subspace $\{(x_1, \ldots, x_n) : x_1 + \ldots + x_n = 1\}$ intersected with the positive orthant in $\mathbb{R}^n$. Lemmas 2.3.6 and 2.3.7 show that the geometry of H_n depends on the parity of n. Specifically,

- for odd $n \ge 3$, the facets of H_n are $(n-2)$-dimensional polytopes given by the vertices $\boldsymbol{v}_{n,1}, \boldsymbol{v}_{n,2}, \ldots, \boldsymbol{v}_{n,n-2}, \boldsymbol{v}_{n,\infty}$ and the vertices $\boldsymbol{v}_{n,r}, \boldsymbol{v}_{n,r+1}, \ldots, \boldsymbol{v}_{n,r+n-2}$ for $r = 1, 2, \ldots$;
- for even $n \ge 4$, the facets of H_n are $(n-2)$-dimensional polytopes given by the vertices $\boldsymbol{v}_{n,1}, \boldsymbol{v}_{n,2}, \ldots, \boldsymbol{v}_{n,n-2}, \boldsymbol{v}_{n,\infty}$, the vertices $\boldsymbol{v}_{n,1}, \boldsymbol{v}_{n,r}, \ldots, \boldsymbol{v}_{n,r+n-3}$ for $r = 2, 3, \ldots$, and the vertices $\boldsymbol{v}_{n,\infty}, \boldsymbol{v}_{n,r}, \ldots, \boldsymbol{v}_{n,r+n-3}$ for $r = 2, 3, \ldots$.

For some intuition regarding the structural difference between the two cases, see Figure 2.1 ($n=3$) and Figure 2.3 ($n=4$).

Yakubovich [77] also found the following counterexample to assertion (*ii*) in Theorem 2.0.2 for $n=7$. Consider the distribution of K_7 induced by i.i.d. sampling from the discrete distribution $\boldsymbol{p}^{(t)} := (\frac{1}{4+t}, \frac{1}{4+t}, \frac{1}{4+t}, \frac{1}{4+t}, \frac{t}{4+t}) = \frac{4-4t}{4+t}\boldsymbol{u}_4 + \frac{5t}{4+t}\boldsymbol{u}_5$ for some $t>0$. The corresponding distribution of K_7 can be computed according to

$$\boldsymbol{v}_7^{(t)}(k) := \mathbb{P}(K_7(\boldsymbol{p}^{(t)}) = k) = \sum \binom{7}{n_1, \ldots, n_5} \prod_{i=1}^{5} (\boldsymbol{p}^{(t)}(i))^{n_i} \tag{2.93}$$

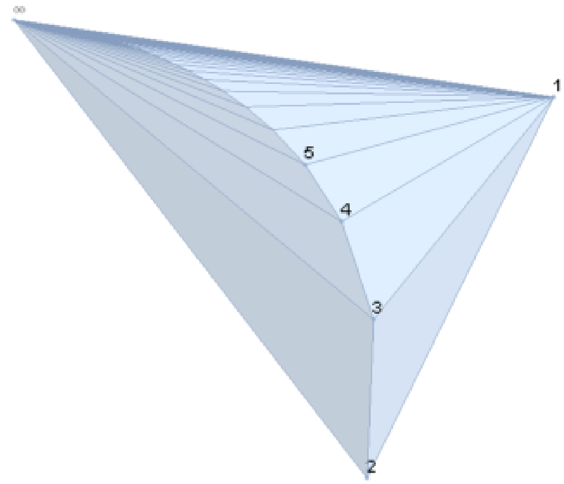


Figure 2.3: The set $H_4 \in \mathbb{R}^4$ projected onto the first, second, and fourth coordinates. Each point $v_{4,m}$ is labeled by m, for $m = 1, 2, 3, 4, 5, \infty$.

where the sum is taken over quintuples $(n_1, \ldots, n_5)$ with $\sum n_i = 7$ and $\#\{i : n_i > 0\} = k$. In particular, it can be verified numerically that for $t \in (0, 0.13)$, $\langle \boldsymbol{\gamma}_{7,2}, \boldsymbol{v}_7^{(t)} \rangle < 0$ and hence by Lemma 2.3.6, $\boldsymbol{v}_7^{(t)} \notin V_7$. This result is proved in [77] using a calculus argument, by showing that function $t \mapsto \langle \boldsymbol{\gamma}_{7,2}, \boldsymbol{v}_7^{(t)} \rangle$, which takes the value 0 at $t = 0$ by Lemma 2.3.6, has a negative one-sided derivative at $t = 0$, implying that $\langle \boldsymbol{\gamma}_{7,2}, \boldsymbol{v}_7^{(t)} \rangle < 0$ for small positive values of t.

A similar counterexample works for $n = 6$, with $\boldsymbol{p}^{(t)} := (\frac{1}{5+t}, \frac{1}{5+t}, \frac{1}{5+t}, \frac{1}{5+t}, \frac{1}{5+t}, \frac{t}{5+t})$ and the hyperplane inequality with $\boldsymbol{\gamma}'_{6,3}$ as in Lemma 2.3.7. It appears empirically that similar modifications can be made to produce counterexamples for larger values of n.

2.4 Finite exchangeable sequences

In this section, we consider the law of K_n for a *finite exchangeable sequence* $(X_1, \ldots, X_m)$ with $m \geq n$. Note the deviation from the original problem: the first m terms of an infinite exchangeable sequence always form a finite exchangeable sequence, but a finite exchangeable sequence need not have an embedding into an infinite one, nor one with more terms. See [17] for some nice geometric pictures of this property; [40] for an extension of de Finetti's theorem to finite exchangeable sequences in which such a sequence can be identified as a "mixture" of i.i.d. random variables, but allowing for a signed mixing measure; and [51] for conditions for the existence of an embedding of a finite exchangeable sequence in a longer one. The presence of negative signs in the mixture confirms that the laws of K_n in this setting are not simply derived from the i.i.d. case by convexity.

The set of possible laws of K_n for finite exchangeable sequences $(X_1, \ldots, X_m)$ form decreasing nested subsets for $m \geq n$, all of which contain that for infinite exchangeable sequences. To analyze this problem, we shift to the framework of *exchangeable random partitions*, for which we provide some background below.

A *partition* of $[m] := \{1, \ldots, m\}$ is an unordered collection of disjoint non-empty subsets $\{A_i\}$ of $[m]$ with $\bigcup_i A_i = [m]$. The A_i are called the *clusters* of the partition. The *restriction* of a partition $\{A_i\}$ of $[m]$ to $[n]$ where $n < m$ is the partition of $[n]$ whose clusters are the nonempty members of $\{A_i \cap [n]\}$.

Every infinite sequence of random variables $(X_1, X_2, \ldots)$ induces a random partition of $\mathbb{N}$ according to the relation $i \sim j$ if and only if $X_i = X_j$. More precisely, a random partition Π of $\mathbb{N}$ is a sequence (Π_m) where for each m, Π_m is a random partition of $[m]$, and for $n < m$, the restriction of Π_m to $[n]$ is Π_n. For the random partition Π of $\mathbb{N}$ induced by a sequence $(X_1, X_2, \ldots)$, the clusters of Π_m are the indices associated to each distinct value among $\{X_1, \ldots, X_m\}$. For example, if

$$(X_1(\omega), X_2(\omega), \ldots) = (7, 6, 7, 8, 8, 7 \ldots),$$

then

$$\Pi_1(\omega) = \{\{1\}\}, \qquad \Pi_2(\omega) = \{\{1\}, \{2\}\}, \qquad \Pi_3(\omega) = \{\{1, 3\}, \{2\}\},$$
$$\Pi_4(\omega) = \{\{1, 3\}, \{2\}, \{4\}\}, \quad \Pi_5(\omega) = \{\{1, 3\}, \{2\}, \{4, 5\}\}, \quad \Pi_6(\omega) = \{\{1, 3, 6\}, \{2\}, \{4, 5\}\}.$$

Observe that K_n as previously defined for a sequence $(X_1, X_2, \ldots)$ counts the number of clusters of Π_n for the associated partition Π. When $(X_1, X_2, \ldots)$ is exchangeable, it induces an *exchangeable random partition* Π of $\mathbb{N}$, meaning that for each m, the distribution of Π_m is invariant under every deterministic permutation of $[m]$. In this scenario, associated to Π is a function p defined for all finite sequences of positive integers such that for every m and every partition $\{A_1, \ldots, A_k\}$ of $[m]$,

$$\mathbb{P}(\Pi_m = \{A_1, \ldots, A_k\}) = p(|A_1|, \ldots, |A_k|). \tag{2.94}$$

Here p is called the *exchangeable partition probability function (EPPF)* associated to Π. A consequence of exchangeability is that the EPPF is a symmetric function of its arguments. The probability mass function for K_n can therefore be expressed in terms of the EPPF as

$$\mathbb{P}(K_n = k) = \sum_{\substack{n_1 + \ldots + n_k = n \\ n_1 \geq \ldots \geq n_k \geq 1}} C(n_1, \ldots, n_k) p(n_1, \ldots, n_k) \tag{2.95}$$

where

$$C(n_1, \ldots, n_k) := \frac{n!}{\prod_{j=1}^{n} (j!)^{s_j} s_j!}, \qquad s_j = s_j(n_1, \ldots, n_k) := \#\{i : n_i = j\} \tag{2.96}$$

counts the number of partitions of $[n]$ whose cluster sizes in descending order are given by $n_1, \ldots, n_k$. Furthermore, the EPPF p satisfies the following *consistency* relation:

$$p(n_1, \ldots, n_k) = p(n_1, \ldots, n_k, 1) + \sum_{i=1}^{k} p(n_1, \ldots, \ n_i + 1 \ , \ldots, n_k). \tag{2.97}$$

Reposed in this alternate framework, the goal of this section is to understand the possible distributions of $K_n = K_n(\Pi_m)$ for an exchangeable random partition Π_m of $[m]$ for $m \geq n$, meaning the number of clusters of the restriction $\Pi_{m\downarrow n}$ of Π_m to $[n]$. A consequence of the exchangeability of Π_m is that $\Pi_{m\downarrow n}$ is an exchangeable random partition of $[n]$, whose EPPF is determined recursively by the EPPF for Π_m and the consistency relations (2.97). Note that for $m = n$, $K_n(\Pi_n)$ can have any general probability distribution on $[n]$: for example, given such a probability distribution $(a_1, \ldots, a_n)$, define an EPPF according to

$$p(n-k+1, \underbrace{1, \ldots, 1}_{k-1 \text{ singletons}}) = \frac{a_k}{\binom{n}{k-1}}, \qquad k = 1, \ldots, n \tag{2.98}$$

where the rest of the values are either 0 or specified by symmetry. By construction, p corresponds to an exchangeable random partition of $[n]$ such that $\mathbb{P}(K_n = k) = a_n$ for $1 \leq k \leq n$. However, for $m > n$, the consistency relations (2.97) must be satisfied, so it is not immediately clear given n and $m > n$ what restrictions there are on the distribution of K_n, if any. The next proposition shows that there are indeed nontrivial restrictions on the law of K_n in this setting.

Proposition 2.4.1. *Let $n \geq 3$, and let Π_{n+1} be an exchangeable random partition of $[n+1]$. Then we have the sharp bound*

$$\mathbb{P}(K_n(\Pi_{n+1}) = n-1) \leq \frac{\max\{4, n-1\}}{n+1} \tag{2.99}$$

Proof. We have

$$\mathbb{P}(K_n = n-1) = \binom{n}{2} p(2, 1^{n-2}) = \binom{n}{2}\left[p(3, 1^{n-2}) + (n-2)p(2, 2, 1^{n-3}) + p(2, 1^{n-1})\right] \tag{2.100}$$

where 1^m is shorthand for m clusters of size 1. We consider the appearance of each of the three terms $p(3, 1^{n-2})$, $p(2, 2, 1^{n-3})$, and $p(2, 1^{n-1})$ in the expansion (2.97) of $p(n_1, \ldots, n_k)$ for $(n_1, \ldots, n_k)$ with $\sum_{i=1}^k n_i = n$ and $n_1 \geq \ldots \geq n_k \geq 1$.

- $p(3, 1^{n-2})$ appears in the expansion of only $p(2, 1^{n-2})$ with coefficient 1 and $p(3, 1^{n-3})$ with coefficient 1. $p(3, 1^{n-3})$ appears in the expansion of $\mathbb{P}(K_n = n-2)$ according to (2.95) with coefficient $C(3, 1^{n-3}) = \binom{n}{3}$.

- $p(2, 2, 1^{n-3})$ appears in the expansion of only $p(2, 1^{n-2})$ with coefficient $n-2$ and $p(2, 2, 1^{n-4})$ with coefficient 1. $p(2, 2, 1^{n-4})$ appears in the expansion of $\mathbb{P}(K_n = n-2)$ according to (2.95) with coefficient $C(2, 2, 1^{n-4}) = 3\binom{n}{4}$.

- $p(2, 1^{n-1})$ appears in the expansion of only $p(2, 1^{n-2})$ with coefficient 1 and $p(1^n)$ with coefficient n. $p(1^n)$ appears in the expansion of $\mathbb{P}(K_n = n)$ with coefficient $C(1^n) = 1$.

Hence the problem reduces to maximizing (2.100) subject to the linear constraints

$$\left[\binom{n}{2}+\binom{n}{3}\right]p(3,1^{n-2})+\left[\binom{n}{2}(n-2)+3\binom{n}{4}\right]p(2,2,1^{n-3})+\left[\binom{n}{2}+n\right]p(2,1^{n-1})\leq 1. \tag{2.101}$$

The maximum value of (2.100) is evidently equal to

$$\max\left\{\frac{\binom{n}{2}}{\binom{n}{2}+\binom{n}{3}},\frac{\binom{n}{2}(n-2)}{\binom{n}{2}(n-2)+3\binom{n}{4}},\frac{\binom{n}{2}}{\binom{n}{2}+n}\right\}, \tag{2.102}$$

with the first expression corresponding to Π_{n+1} having 1 cluster of size 3 and $n-2$ clusters of size 1 with probability 1; the second expression corresponding to Π_{n+1} having 2 clusters of size 2 and $n-3$ clusters of size 1 with probability 1; and the third expression corresponding to Π_{n+1} having 1 cluster of size 2 and $n-1$ clusters of size 1 with probability 1. Simplifying each of the three expressions yields

$$\max\left\{\frac{3}{n+1},\frac{4}{n+1},\frac{n-1}{n+1}\right\}=\frac{\max\{4,n-1\}}{n+1}. \tag{2.103}$$

□

It follows from Proposition 2.4.1 that for $n=3$, there are no restrictions on the distribution of $K_3(\Pi_4)$ on $\{1,2,3\}$. The same claim cannot be made for $n\geq 4$, as $\mathbb{P}(K_4(\Pi_5)=3)\leq\frac{4}{5}$ and $\mathbb{P}(K_n(\Pi_{n+1})=n-1)\leq\frac{n-1}{n+1}$ for $n\geq 5$.

The remainder of the section will focus on $K_3(\Pi_n)$ for $n\geq 3$. Intuitively, as $n\to\infty$, the set of probability distributions of $K_3(\Pi_n)$ should tend to the corresponding set for $K_3(\Pi)$ for exchangeable random partitions Π of $\mathbb{N}$, which was explicitly characterized in Section 2. Fix $n\geq 3$, and as before, consider the parameterization $q_1=\mathbb{P}(K_3(\Pi_n)=1)$ and $q_3=\mathbb{P}(K_3(\Pi_n)=3)$. By repeated application of (2.97), q_1 and q_3 may be written in terms of the EPPF as

$$q_1=p(3)=\sum_{\substack{1\leq k\leq n\\ n_1+\ldots+n_k=n\\ n_1\geq\ldots\geq n_k\geq 1}}A(n_1,\ldots,n_k)p(n_1,\ldots,n_k) \tag{2.104}$$

and

$$q_3=p(1,1,1)=\sum_{\substack{1\leq k\leq n\\ n_1+\ldots+n_k=n\\ n_1\geq\ldots\geq n_k\geq 1}}B(n_1,\ldots,n_k)p(n_1,\ldots,n_k) \tag{2.105}$$

for uniquely defined nonnegative integer coefficients $A(n_1,\ldots,n_k)$ and $B(n_1,\ldots,n_k)$.The problem is to describe the set of points (q_1,q_3) arising in this manner subject to

$$\sum_{\substack{1\leq k\leq n\\ n_1+\ldots+n_k=n\\ n_1\geq\ldots\geq n_k\geq 1}}C(n_1,\ldots,n_k)p(n_1,\ldots,n_k)=1 \tag{2.106}$$

where $C(n_1,\ldots,n_k)$ is as defined in (2.96). Observe that, in vector notation,

$$(q_1, q_3) = \Big(\sum A(n_1,\ldots,n_k)p(n_1,\ldots,n_k), \sum B(n_1,\ldots,n_k)p(n_1,\ldots,n_k)\Big) \tag{2.107}$$

$$= \sum C(n_1,\ldots,n_k)p(n_1,\ldots,n_k)\Big(\tfrac{A(n_1,\ldots,n_k)}{C(n_1,\ldots,n_k)}, \tfrac{B(n_1,\ldots,n_k)}{C(n_1,\ldots,n_k)}\Big) \tag{2.108}$$

This shows that every (q_1, q_3) is a convex combination of points of the form $\big(\frac{A(\boldsymbol{n})}{C(\boldsymbol{n})}, \frac{B(\boldsymbol{n})}{C(\boldsymbol{n})}\big)$, and thus the set of probability distributions of $K_3(\Pi_n)$ over all exchangeable random partitions Π_n of $[n]$, expressed in the parameterization (q_1, q_3), is the convex hull of the finite set of points

$$S_n := \Big\{\Big(\tfrac{A(n_1,\ldots,n_k)}{C(n_1,\ldots,n_k)}, \tfrac{B(n_1,\ldots,n_k)}{C(n_1,\ldots,n_k)}\Big) : 1 \le k \le n, n_1 + \ldots + n_k = 1, n_1 \ge \ldots \ge n_k \ge 1\Big\}. \tag{2.109}$$

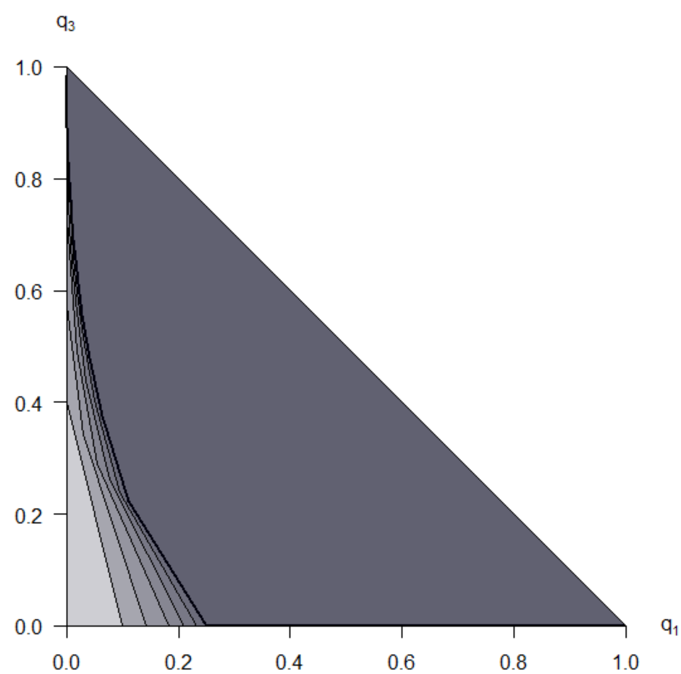


Figure 2.4: The nested regions are the possible probability distributions of $K_3(\Pi_n)$ for Π_n an exchangeable random partition of $[n]$ for $n = 4, 5, 7, 12, 19, 41$, which tend to the region corresponding to K_3 for infinite exchangeable sequences, as described in Theorem 2.0.2 and shown in Figure 2.1.

Listed below is the sequence (s_n) for the number of extreme points of the convex hull of S_n for $3 \le n \le 34$, computed using SciPy [74].

n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
s_n	3	3	4	4	5	5	6	6	7	6	8	7	8	8	9	8	10

n	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
s_n	9	10	10	11	9	12	11	11	11	13	11	13	12	13	13	14

2.5 The two-parameter family

It was shown in [64] that every pair of real parameters (α, θ) satisfying either of the conditions

$$\text{(i)}\ \ 0 \le \alpha < 1 \text{ and } \theta > -\alpha;\text{ or} \tag{2.110}$$

$$\text{(ii)}\ \ \alpha < 0 \text{ and } \theta = -m\alpha \text{ for some } m \in \mathbb{N} \tag{2.111}$$

corresponds to an exchangeable random partition $\Pi_{\alpha,\theta} = (\Pi_n)$ of $\mathbb{N}$ according to the following sequential construction known as the Chinese restaurant process: for each $n \in \mathbb{N}$, conditionally given $\Pi_n = \{C_1, \ldots, C_k\}$, Π_{n+1} is formed by having $n+1$

$$\begin{aligned} &\text{attach to cluster } C_i \text{ with probability } \frac{|C_i| - \alpha}{n+\theta}, \quad 1 \le i \le k\ ; \\ &\text{form a new cluster with probability } \frac{\theta + k\alpha}{n+\theta}. \end{aligned} \tag{2.112}$$

The corresponding EPPF is given by

$$p_{\alpha,\theta}(n_1, \ldots, n_k) = \frac{\prod_{i=0}^{k-1}(\theta + i\alpha) \prod_{j=1}^{k}(1-\alpha)_{n_j - 1}}{(\theta)_n} \tag{2.113}$$

where $n = n_1 + \ldots + n_k$ and

$$(x)_m := x(x+1)\cdots(x+m-1) = \frac{\Gamma(x+m)}{\Gamma(x)}. \tag{2.114}$$

Let $\mathbb{P}_{\alpha,\theta}$ denote the law of $\Pi_{\alpha,\theta}$. The distribution of K_3 for $\Pi_{\alpha,\theta}$ is given by

$$q_1(\alpha, \theta) = \frac{(1-\alpha)(2-\alpha)}{(1+\theta)(2+\theta)} \tag{2.115}$$

$$q_2(\alpha, \theta) = \frac{3(1-\alpha)(\theta+\alpha)}{(1+\theta)(2+\theta)} \tag{2.116}$$

$$q_3(\alpha, \theta) = \frac{(\theta+\alpha)(\theta+2\alpha)}{(1+\theta)(2+\theta)} \tag{2.117}$$

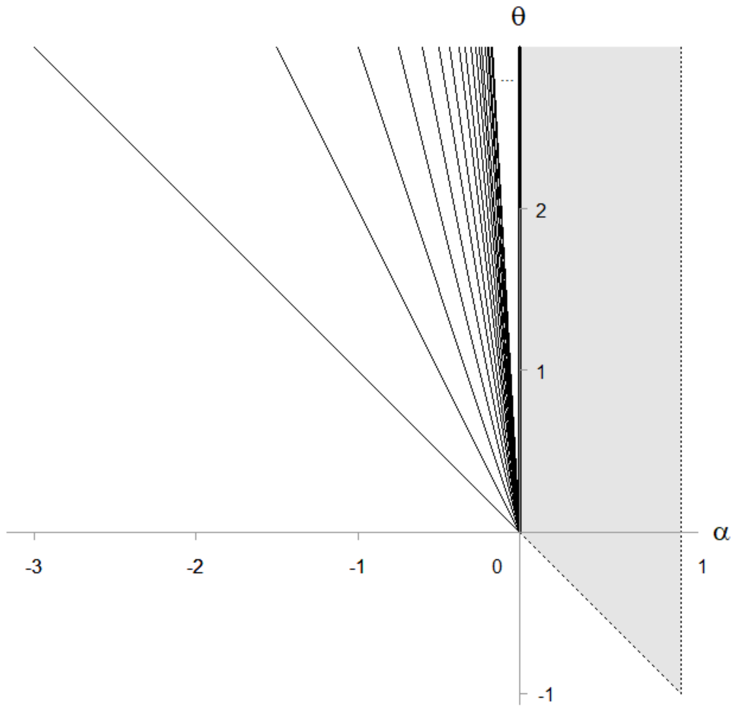


Figure 2.5: The (α, θ) parameter space.

where

$$q_i(\alpha, \theta) := \mathbb{P}_{\alpha,\theta}(K_3 = i). \tag{2.118}$$

For $m > 0$, let

$$A_m := \left\{(m + m\theta, \theta) : -\tfrac{m}{m+1} < \theta < \tfrac{1-m}{m}\right\} \subseteq \{(\alpha, \theta) : 0 \leq \alpha < 1, \theta > -\alpha\} \tag{2.119}$$

and let $A_0 := \{(0, \theta) : \theta > 0\}$, the parameter subspace corresponding to the well-known one-parameter Ewens sampling formula [27]. The line segments and one ray $\{A_m\}_{m \geq 0}$ with inverse slope m in the (α, θ) plane, each of which would pass through the point $(\alpha, \theta) = (0, -1)$ if extended, partition the parameter subspace $\{(\alpha, \theta) : 0 \leq \alpha < 1, \theta > -\alpha\}$. Hence the distribution of K_3 can be reparameterized in m and θ as

$$q_1^{(m)}(\theta) = \frac{(1 - m - m\theta)(2 - m - m\theta)}{(1 + \theta)(2 + \theta)} \tag{2.120}$$

$$q_2^{(m)}(\theta) = \frac{3(1 - m - m\theta)[m + (m + 1)\theta]}{(1 + \theta)(2 + \theta)} \tag{2.121}$$

$$q_3^{(m)}(\theta) = \frac{[m + (m + 1)\theta][2m + (2m + 1)\theta]}{(1 + \theta)(2 + \theta)} \tag{2.122}$$

It can be checked by calculus that for each fixed $m > 0$,

- the function $q_1^{(m)}(\theta)$ is strictly decreasing for $\theta \in (-\frac{m}{m+1}, \frac{1-m}{m})$ with $\lim_{\theta \to -\frac{m}{m+1}} q_1^{(m)}(\theta) = 1$ and $\lim_{\theta \to \frac{1-m}{m}} q_1^{(m)}(\theta) = 0$.

- the function $q_3^{(m)}(\theta)$ is strictly increasing for $\theta \in (-\frac{m}{m+1}, \frac{1-m}{m})$ with $\lim_{\theta \to -\frac{m}{m+1}} q_3^{(m)}(\theta) = 0$ and $\lim_{\theta \to \frac{1-m}{m}} q_3^{(m)}(\theta) = 1$.

- the function $q_2^{(m)}(\theta)$ is strictly increasing on $\big(-\frac{m}{m+1}, \tau(m)\big]$ and strictly decreasing on $\big[\tau(m), \frac{1-m}{m}\big)$, with a unique maximum value of $9 - 6\big(\sqrt{(m+1)(m+2)} - m\big)$ at $\theta = \tau(m) := \frac{-m^2-3m+\sqrt{(m+1)(m+2)}}{1+3m+m^2}$, which is also the unique value of θ in the domain at which $q_1^{(m)}(\theta) = q_3^{(m)}(\theta)$.

The properties above also hold for $m = 0$ after slight modification by replacing each instance of $\frac{1-m}{m}$ with $\lim_{m \to 0^+} \frac{1-m}{m} = \infty$, and this remark also applies to subsequent discussion.

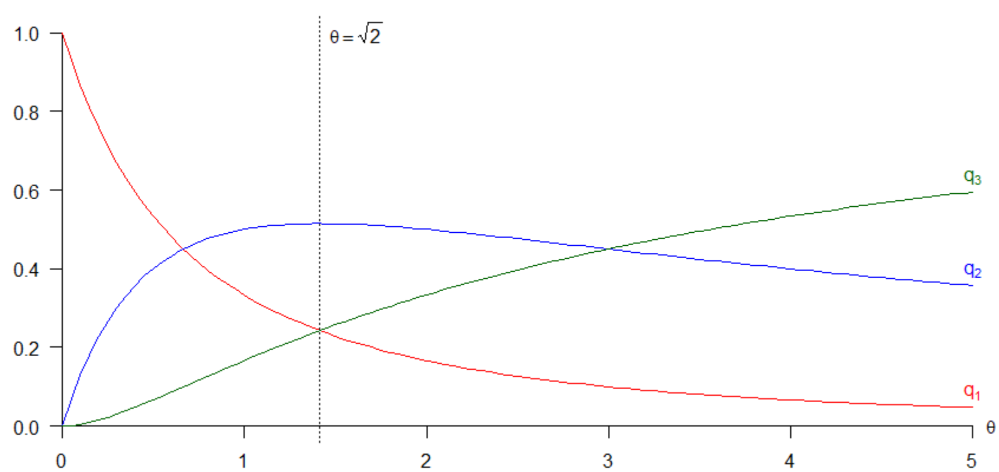


Figure 2.6: Graphs of $q_i^{(m)}(\theta)$ for $m = 0$ and $\theta \in [0, 5]$. Observe that q_1 and q_3 intersect at the same value of θ as where q_2 attains its maximum value. The corresponding graphs for every $m > 0$ also share this property.

Duality. The last observation implies that for $m \geq 0$ and every real number p such that

$0 < p < 9 - 6(\sqrt{(m+1)(m+2)} - m)$, there are exactly two values $\theta_{\pm}^{(m)}(p)$ with

$$-\frac{m}{m+1} < \theta_{-}^{(m)}(p) < \tau(m) < \theta_{+}^{(m)}(p) < \frac{1-m}{m}. \tag{2.123}$$

satisfying

$$q_2^{(m)}(\theta_{-}^{(m)}(p)) = q_2^{(m)}(\theta_{+}^{(m)}(p)). \tag{2.124}$$

For $p = 9 - 6(\sqrt{(m+1)(m+2)} - 2)$, define $\theta_{-}^{(m)}(p) = \theta_{+}^{(m)}(p) = \varphi(m)$. As $\theta_{\pm}^{(m)}(p)$ are defined as the solutions to the equation

$$\frac{3(1-m-m\theta)[m+(m+1)\theta]}{(1+\theta)(2+\theta)} = p \tag{2.125}$$

or equivalently the quadratic equation

$$p(1+\theta)(2+\theta) - 3(1-m-m\theta)[m+(m+1)\theta] = 0, \tag{2.126}$$

we have the polynomial identity

$$(\theta - \theta_{+}^{(m)}(p))(\theta - \theta_{-}^{(m)}(p)) = \theta^2 + \frac{3p - 3 + 6m^2}{p + 3m + 3m^2}\theta + \frac{2p - 3m + 3m^2}{p + 3m + 3m^2} \tag{2.127}$$

after rearranging (2.126). It follows that

$$\theta_{+}^{(m)}\theta_{-}^{(m)} = \frac{2p - 3m + 3m^2}{p + 3m + 3m^2}. \tag{2.128}$$

For $-\frac{m}{m+1} < \theta < \frac{1-m}{m}$, define the *m-dual* $\theta_{*}^{(m)}$ of θ according to (2.123). Rearranging (2.128) and simplifying gives the explicit formula

$$\theta_{*}^{(m)} = \frac{2 - m(3+m)(1+\theta)}{\theta + m(3+m)(1+\theta)}. \tag{2.129}$$

Theorem 2.5.1. *For $m \geq 0$ and $-\frac{m}{m+1} < \theta < \frac{1-m}{m}$, we have*

$$q_1^{(m)}(\theta_{*}^{(m)}) = q_3^{(m)}(\theta) \qquad \text{and} \qquad q_3^{(m)}(\theta_{*}^{(m)}) = q_1^{(m)}(\theta). \tag{2.130}$$

Proof. It suffices to verify the first of the two identities since (2.129) is constructed as an involution. Let $D(m,\theta)$ be the denominator in (2.129). Substituting and simplifying yields

$$1 + \theta_{*}^{(m)} = \frac{2+\theta}{D(m,\theta)}\ ; \tag{2.131}$$

$$2 + \theta_{*}^{(m)} = \frac{(1+\theta)(1+m)(2+m)}{D(m,\theta)}\ ; \tag{2.132}$$

$$1 - m - m\theta_{*}^{(m)} = \frac{(1+m)[m+(m+1)\theta]}{D(m,\theta)}\ ; \tag{2.133}$$

$$2 - m - m\theta_{*}^{(m)} = \frac{(2+m)[2m+(2m+1)\theta]}{D(m,\theta)}. \tag{2.134}$$

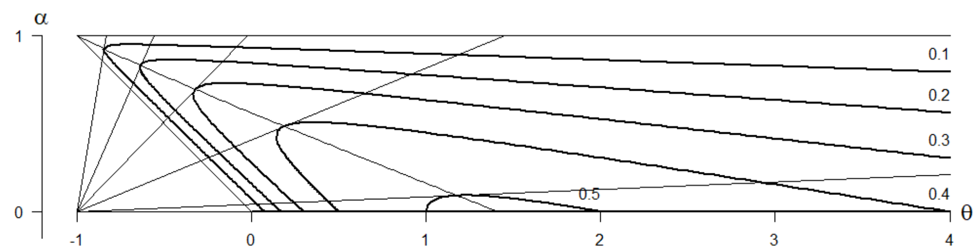


Figure 2.7: Contour plot of $q_2(\alpha, \theta)$. The level curves for $q_2(\alpha, \theta) \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ are shown, along with their tangent lines where they meet the curve $q_1(\alpha, \theta) = q_3(\alpha, \theta)$. Observe that each tangent line passes through the point $(\alpha, \theta) = (0, -1)$. Note that here α is plotted on the vertical axis, for convenience of display.

Hence we have

$$q_1^{(m)}(\theta_*^{(m)}) = \frac{(1 - m - m\theta_*^{(m)})(2 - m - m\theta_*^{(m)})}{(1 + \theta_*^{(m)})(2 + \theta_*^{(m)})} \tag{2.135}$$

$$= \frac{[m + (m+1)\theta][2m + (2m+1)\theta]}{(1+\theta)(2+\theta)} \tag{2.136}$$

$$= q_3^{(m)}(\theta) \tag{2.137}$$

as desired. □

Symmetry. A consequence of Theorem 2.5.1 is a surprising symmetry in the set of laws of K_3 arising from the two-parameter model. To make this observation explicit, for any $m \geq 0$ we solve for $q_3 = q_3^{(m)}$ in terms of $q_1 = q_1^{(m)}$ as defined in (2.120) and (2.122) to obtain the formula

$$q_3 = \varphi_m(q_1) := 1 + \frac{3}{4}m + \frac{5}{4}q_1 - \frac{3}{4}\sqrt{m^2 + 6q_1 m + q_1(8 + q_1)}. \tag{2.138}$$

Rearranging to eliminate the radical yields the relation

$$(4 + 3m)(q_1 + q_3) + 5q_1q_3 - 2(q_1^2 + q_3^2) - 2 - 3m = 0 \tag{2.139}$$

which verifies the symmetry. For $m = 0$ the identity reduces to

$$h(q_1, q_3) := 4(q_1 + q_3) + 5q_1q_3 - 2(q_1^2 + q_3^2) - 2 = 0. \tag{2.140}$$

This appears to be an exclusive property of the case $n = 3$, as no similar symmetry appears to manifest for larger n.

Theorem 2.5.2. *The mapping $(\alpha, \theta) \mapsto (q_1, q_3)$ defined by (2.115) and (2.117) is a bijection between the regions*

$$\{(\alpha, \theta) : 0 \leq \alpha < 1,\ \theta > -\alpha\} \qquad \text{and} \qquad \{(q_1, q_3) : h(q_1, q_3) \geq 0,\ q_1 + q_3 < 1\} \tag{2.141}$$

where $h(q_1, q_3)$ is defined as in (2.140).

Proof. Consider $\varphi(m, q_1) := \varphi_m(q_1)$ as in (2.138). To show the desired bijection, it suffices to show that for every fixed $0 < q_1 < 1$ that (i) $\varphi(m, q_1)$ is increasing in m, and (ii) $\lim_{m\to\infty} \varphi(m, q_1) = 1 - q_1$.

(i)

$$\frac{\partial}{\partial m}\varphi(m, q_1) = \frac{3}{4}(1 - \frac{2m + 6q_1}{2\sqrt{m^2 + 6q_1 m + q_1(8 + q_1)}}) > \frac{3}{4}(1 - \frac{2m + 6q_1}{2\sqrt{m^2 + 6q_1 m + 9q_1^2}}) = 0 \tag{2.142}$$

(ii)

$$\lim_{m\to\infty} \varphi(m, q_1) = \lim_{m\to\infty} 1 + \frac{5}{4}q_1 + \frac{3}{4}\left(\frac{m^2 - (m^2 + 6q_1 m + q_1(8 + q_1))}{m + \sqrt{m^2 + 6q_1 m + q_1(8 + q_1)}}\right) \tag{2.143}$$

$$= \lim_{m\to\infty} 1 + \frac{5}{4}q_1 + \frac{3}{4}\left(\frac{-6q_1 - \frac{q_1(8+q_1)}{m}}{1 + \sqrt{1 + \frac{6q_1}{m} + \frac{q_1(8+q_1)}{m^2}}}\right) \tag{2.144}$$

$$= 1 - q_1 \tag{2.145}$$

□

Explicit inverse. Define the ratios

$$r(\alpha, \theta) := \frac{q_1(\alpha, \theta)}{q_2(\alpha, \theta)} = \frac{2 - \alpha}{3(\theta + \alpha)}, \qquad s(\alpha, \theta) := \frac{q_2(\alpha, \theta)}{q_3(\alpha, \theta)} = \frac{3(1 - \alpha)}{(\theta + 2\alpha)} \tag{2.146}$$

These ratios uniquely define the law of K_3 for the corresponding (α, θ). The map $(\theta, \alpha) \mapsto (r, s)$ can be explicitly inverted as

$$\alpha(r, s) = \frac{9r - 2s}{9r - s + 3rs}, \qquad \theta(r, s) = \frac{3 - 9r + 4s}{9r - s + 3rs} \tag{2.147}$$

Expressed in terms of q_1 and q_3, this gives the inversion formulas

$$\alpha(q_1, q_3) = \frac{4q_1 + 4q_3 + 5q_1q_3 - 2q_1^2 - 2q_3^2 - 2}{5q_1 + 2q_3 + 4q_1q_3 - 4q_1^2 - q_3^2 - 1}, \tag{2.148}$$

$$\theta(q_1, q_3) = -\frac{8q_1 + 5q_3 + 4q_1q_3 - 4q_1^2 - q_3^2 - 4}{5q_1 + 2q_3 + 4q_1q_3 - 4q_1^2 - q_3^2 - 1}. \tag{2.149}$$

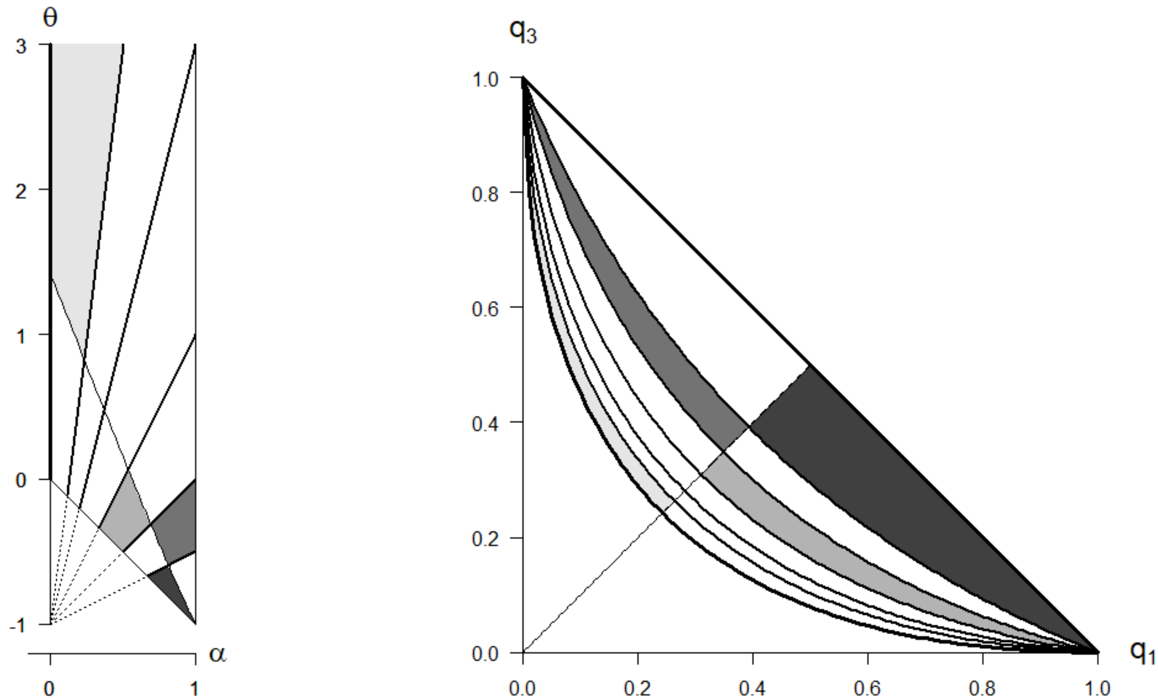


Figure 2.8: The bijection of Theorem 2.5.2. The regions colored in different shades of gray reveal the geometry of the bijection.

Note that the numerator in the formula for $\alpha(q_1, q_3)$ is equal to $h(q_1, q_3)$ as defined in (2.140). It is easy to verify that these formulas give an algebraic inverse. Observe that the denominator which is the same in both formulas is nonvanishing on the region $\{(q_1, q_3) : h(q_1, q_3) \geq 0,\ q_1 + q_3 < 1\}$, since

$$2(5q_1 + 2q_3 + 4q_1q_3 - 4q_1^2 - q_3^2 - 1) = h(q_1, q_3) + 6q_1 - 6q_1^2 + 3q_1q_3 > 0. \tag{2.150}$$

Corollary 2.5.3. *For any parameters* (α, θ) *with* $0 \leq \alpha < 1$ *and* $\theta > -\alpha$*, there exists a unique pair* (α_*, θ_*) *with* $0 \leq \alpha_* < 1$ *and* $\theta_* > -\alpha_*$ *such that*

$$q_{2\pm 1}(\alpha, \theta) = q_{2\mp 1}(\alpha_*, \theta_*). \tag{2.151}$$

Explicit formulas for α_* and θ_* in terms of α and θ can be computed as

$$\alpha^* = \frac{(2 - 3\alpha)(1 + \theta) - \alpha^2}{(\theta + 3\alpha)(1 + \theta) + \alpha^2} \tag{2.152}$$

$$\theta^* = \frac{\alpha(2 + \theta)}{(\theta + 3\alpha)(1 + \theta) + \alpha^2}. \tag{2.153}$$

Exceptional parameters. $\alpha < 0$, $\theta = -m\alpha$ for some $m \in \mathbb{N}$

It is well-known that in this case, the exchangeable random partition (Π_n) of $\mathbb{N}$ generated according to the Chinese restaurant construction is distributed as if by sampling from a symmetric Dirichlet distribution with m parameters equal to $-\alpha$ [63]. Hence for fixed $m \in \mathbb{N}$, as $\alpha \downarrow -\infty$ the exchangeable random partition of $\mathbb{N}$ corresponding to the parameter pair $(\alpha, \theta) = (\alpha, -m\alpha)$ converges in distribution to that obtained by sampling from the discrete uniform distribution on m elements. For K_3, the (α, θ) to (q_1, q_3) correspondence can be seen in Figure 2.9.

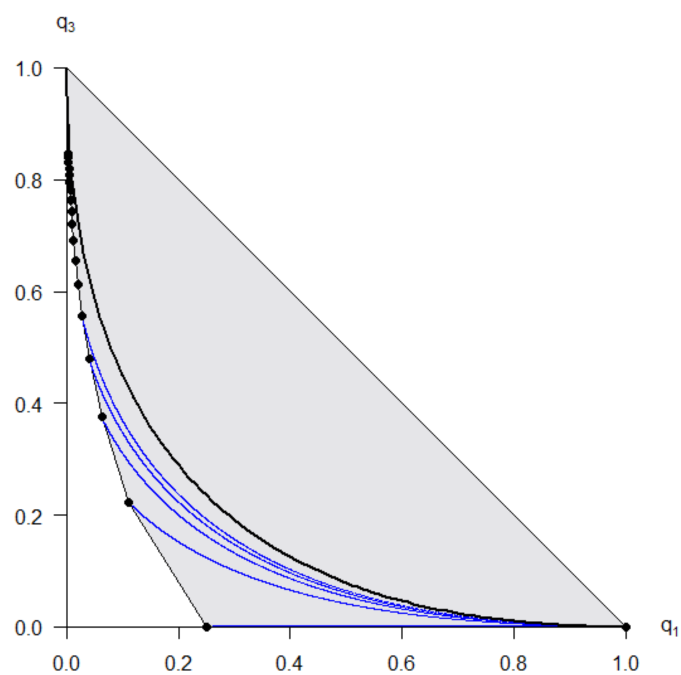


Figure 2.9: The blue curves correspond to the images of $(\alpha, \theta) = (\alpha, -m\alpha)$ for $\alpha \in (-\infty, 0)$ and fixed m under the $(\alpha, \theta) \mapsto (q_1, q_3)$ map, for $m = 2, 3, 4, 5, 6$. The curve defined by (2.140) is included in black.

2.6 Complements

In this section, we point out an interesting convexity property for the the law of K_3. With notation as in Section 2.2, for $\boldsymbol{p} \in \nabla_\infty$, let

$$\boldsymbol{Q}(\boldsymbol{p}) := \big(q_1(\boldsymbol{p}), q_3(\boldsymbol{p})\big) \tag{2.154}$$

be the mapping from a ranked discrete distribution to its corresponding law of K_3 obtained by i.i.d. sampling. In Section 2.2 we established that the range of $\boldsymbol{Q}$ is a (strict) subset of the closed convex hull of the set of points $\{\boldsymbol{Q}(\boldsymbol{u}_N) : N \in \mathbb{N}\}$. Note that the range of $\boldsymbol{Q}$ includes only distributions of K_3 which arise from i.i.d. sampling. Here are some preliminary efforts to better understand the geometry of this mapping.

Proposition 2.6.1. *For every* $0 \le \lambda \le 1$ *and* $N \ge 1$,

$$\boldsymbol{Q}(\lambda \boldsymbol{u}_N + (1-\lambda)\boldsymbol{u}_{2N}) = \lambda^2 \boldsymbol{Q}(\boldsymbol{u}_N) + (1-\lambda^2)\boldsymbol{Q}(\boldsymbol{u}_{2N}) \tag{2.155}$$

Proof. We have

$$\lambda \boldsymbol{u}_N + (1-\lambda)\boldsymbol{u}_{2N} = \big(\underbrace{\tfrac{1+\lambda}{2N}, \dots, \tfrac{1+\lambda}{2N}}_{N \text{ times}}, \underbrace{\tfrac{1-\lambda}{2N}, \dots, \tfrac{1-\lambda}{2N}}_{N \text{ times}}\big). \tag{2.156}$$

Hence

$$q_1(\lambda \boldsymbol{u}_N + (1-\lambda)\boldsymbol{u}_{2N}) = N\Big(\frac{1+\lambda}{2N}\Big)^3 + N\Big(\frac{1-\lambda}{2N}\Big)^3 = \frac{1+3\lambda^2}{4N^2} \tag{2.157}$$

and

$$q_3(\lambda \boldsymbol{u}_N + (1-\lambda)\boldsymbol{u}_{2N}) = \binom{N}{3}\Big(\frac{1+\lambda}{2N}\Big)^3 + \binom{N}{2}N\Big(\frac{1+\lambda}{2N}\Big)^2\Big(\frac{1-\lambda}{2N}\Big) \tag{2.158}$$

$$+ N\binom{N}{2}\Big(\frac{1+\lambda}{2N}\Big)\Big(\frac{1-\lambda}{2N}\Big)^2 + \binom{N}{3}\Big(\frac{1-\lambda}{2N}\Big)^3 \tag{2.159}$$

$$= \binom{N}{3}\frac{1+3\lambda^2}{4N^3} + N\binom{N}{2}\frac{1-\lambda^2}{4N^3} \tag{2.160}$$

$$= \frac{N-1}{3}\Big(\frac{2N-1-3\lambda^2}{4N^2}\Big). \tag{2.161}$$

On the other side,

$$\lambda^2 q_1(\boldsymbol{u}_N) + (1-\lambda^2)q_1(\boldsymbol{u}_{2N}) = \frac{\lambda^2}{N^2} + \frac{1-\lambda^2}{4N^2} = \frac{1+3\lambda^2}{4N^2} \tag{2.162}$$

and

$$\lambda^2 q_3(\boldsymbol{u}_N) + (1-\lambda^2)q_3(\boldsymbol{u}_{2N}) = \lambda^2\binom{N}{3}\frac{1}{N^3} + (1-\lambda^2)\binom{2N}{3}\frac{1}{8N^3} \tag{2.163}$$

$$= \frac{N(N-1)(N-2)}{6}\cdot\frac{\lambda^2}{N^3} + \frac{2N(2N-1)(2N-2)}{6}\cdot\frac{1-\lambda^2}{8N^3} \tag{2.164}$$

$$= \frac{N-1}{3}\Big(\frac{2N-1-3\lambda^2}{4N^2}\Big).$$

□

Chapter 3

Extreme coherent distributions

A probability measure μ on $[0,1]^2$ is called *coherent* if it is the joint distribution of a pair of random variables (X, Y) on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying

$$X = \mathbb{P}(A \mid \mathcal{G}) \qquad \text{and} \qquad Y = \mathbb{P}(A \mid \mathcal{H}) \tag{3.1}$$

for some event $A \in \mathcal{F}$ and two sub-σ-fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$. Following [15], X and Y as in (3.1) may be interpreted as the opinions of two *experts* about the probability of an event A, given different sources of information $\mathcal{G}$ and $\mathcal{H}$. In this setup, we assume that the experts agree on some initial assignment $\mathbb{P}$ of probabilities to events in $\mathcal{F}$. We adopt the terminology of [15] and use the term *coherent* to describe (X, Y) as in (3.1) as well as the joint distribution of such.

The definition of coherence is easily extended to a family of random variables $(X_i, i \in I)$. That is, $(X_i, i \in I)$ is coherent if there exists an event $A \in \mathcal{F}$ and sub-σ-fields $\mathcal{G}_i \subseteq \mathcal{F}$ such that $X_i = \mathbb{P}(A \mid \mathcal{G}_i)$ for all $i \in I$. Then a coherent family based on increasing information, in the sense that the sequence of σ-fields is increasing, forms a martingale. Other important examples of coherent families include reversed martingales and martingales relative to directed index sets [19], [49].

Dawid, DeGroot, and Mortera [15] pioneered the notion of coherent expert opinions from the perspective of Bayesian statistics. However, well beforehand Dubins and Pitman [20] studied the same probabilistic structure from the vantage of martingale theory, specifically with interest in various maximal inequalities. Recently there has been a resurgence of interest in the subject. The notion of coherence is studied from the perspective of Bayesian economics and game theory in [2] and [37], and the problem of maximizing various linear functionals over coherent distributions is explored in [9], [8], [12], and [13]. However, the known fact that the set of coherent distributions is a compact convex set of measures is not exploited in any of this analysis, largely due to the lack of understanding of both the geometry and the extreme points of this set. This chapter documents our contributions on this front.

The chapter is organized as follows. Section 3.1 provides some background and intuition for coherent distributions. Section 3.2 covers a couple of classical examples on extreme points of certain sets of probability measures. Section 3.3 presents some inequalities for coherent distributions. In Section 3.4, we characterize the subset of extreme coherent distributions which are supported on the corners of a rectangle, solving a problem posed in [9]. In Section 3.5, we demonstrate an application of this characterization to a maximization problem in [9]. In Section 3.6, we construct a large class of extreme coherent distributions which are uniquely determined by a support set satisfying certain properties. In particular, we prove the existence of coherent distributions with countably infinite support, a result which was also independently obtained in [2]. In Section 3.7, we consider an interesting example of a coherent distribution with uniform marginals.

3.1 Background

Some basic consequences of the definition of coherence are

- $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{P}(A) = p$ for some $p \in [0,1]$;
- If (X,Y) is coherent, then (Y,X), $(1-X,1-Y)$, and $(1-Y,1-X)$ are also coherent (reflection symmetry).

Note also that in the definition of coherence, by law of iterated expectation (3.1) can be reformulated as $X = \mathbb{P}(A \mid X)$ and $Y = \mathbb{P}(A \mid Y)$.

It is not immediately obvious that coherence is a stronger condition than just having equal marginal expectations. To see that it is, let X be a random variable with $\mathbb{E}X = \frac{1}{2}$ and suppose that the pair $(X, 1-X)$ is coherent. Then we must have

$$X = \mathbb{P}(A \mid X) = \mathbb{P}(A \mid 1-X) = 1-X \text{ a.s.} \tag{3.2}$$

Hence $(X, 1-X)$ with $\mathbb{E}X = \frac{1}{2}$ is coherent if and only if $X = \frac{1}{2}$ a.s.

In the case that X and Y are discrete random variables, we also have the following condition.

Lemma 3.1.1. *Suppose (X,Y) is coherent and that (x,y) is a possible pair, meaning $\mathbb{P}((X,Y)=(x,y)) > 0$. Then either $x = y$, or there exists a different possible pair (x',y') with either $x' = x$ or $y' = y$.*

Proof. If there does not exist a possible pair $(x',y') \neq (x,y)$ with either $x' = x$ or $y' = y$, then the events $\{X = x\}$ and $\{Y = y\}$ are almost surely equal, in which case $x = \mathbb{P}(A \mid X = x) = \mathbb{P}(A \mid Y = y) = y$. □

The following proposition gives some known characterizations of coherent pairs of random variables.

Proposition 3.1.2 ([9])**.** *Let (X,Y) be a pair of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, on which there is also a random variable U with uniform distribution on $[0,1]$, independent of (X,Y). Then the following conditions are equivalent:*

(i) (X, Y) *is coherent.*

(ii) *There exists a random variable* Z *defined on* $(\Omega, \mathcal{F}, \mathbb{P})$*, with* $0 \le Z \le 1$*, such that both*

$$\mathbb{E}[Zg(X)] = \mathbb{E}[Xg(X)] \qquad \text{and} \qquad \mathbb{E}[Zg(Y)] = \mathbb{E}[Yg(Y)] \tag{3.3}$$

either for all bounded measurable functions g *defined on* $[0, 1]$*, or for all bounded continuous functions* g*.*

(iii) *There exists a measurable function* $\phi : [0, 1]^2 \to [0, 1]$ *such that*

$$\mathbb{E}[\phi(X, Y)g(X)] = \mathbb{E}[Xg(X)] \qquad \text{and} \qquad \mathbb{E}[\phi(X, Y)g(Y)] = \mathbb{E}[Yg(Y)] \tag{3.4}$$

either for all bounded measurable g *defined on* $[0, 1]$*, or for all bounded continuous* g*.*

(iv) $\mathbb{E}X = \mathbb{E}Y = p$ *for some* $0 \le p \le 1$*, and*

$$\mathbb{E}[X\mathbb{1}(X \in B)] + \mathbb{E}[Y\mathbb{1}(Y \in C)] \le p + \mathbb{P}(X \in B, Y \in C) \tag{3.5}$$

for all $B, C \in \mathcal{B}$*, where* $\mathcal{B}$ *may be either the collection of all Borel subsets of* $[0, 1]$*, or the collection of all finite unions of intervals contained in* $[0, 1]$*.*

Condition (iv) is derived from a classical result due to Strassen [72] on the existence of probability measures with given marginals. Observe that the inequality (3.5) is trivial if either B or C is equal to $\emptyset$ or $[0, 1]$. In the case that X and Y each take only finitely many distinct values, say m and n, respectively, this in general gives us a system of $(2^m - 2)(2^n - 2)$ nontrivial inequalities, which we shall refer to as the *Strassen inequalities*.

The characterizations (ii) and (iii) in Proposition 3.1.2 extend easily to a coherent family $(X_i, i \in I)$ of random variables, while (iv) does not [9].

Proposition 3.1.3 ([9])**.** *For a finite index set* I*, the set of coherent distributions of* $(X_i, i \in I)$ *is a convex, compact subset of probability distributions on* $[0, 1]^I$ *with the usual weak topology.*

3.2 Some examples of extremal probability measures

In this section, we review some classical results on extremal probability measures. It is well-known that for a compact subset $X \subseteq \mathbb{R}^n$, the set $\mathcal{M}(X)$ of Borel probability measures on X is convex and compact in the weak topology. (More generally, X can be taken to be any compact Hausdorff space and $\mathcal{M}(X)$ the set of *regular* Borel probability measures on X; see e.g. [31].) Recall that the *support* of a measure μ on X, denoted $\mathrm{supp}(\mu)$, is defined as the smallest closed set E such that $\mu(X \setminus E) = 0$; equivalently, $\mathrm{supp}(\mu)$ is the set of all $x \in X$ for which $\mu(U) > 0$ for every open set U containing x.

Theorem 3.2.1. *[70] Let X be a compact subset of $\mathbb{R}^n$. The set of extreme points of $\mathcal{M}(X)$ is $\{\delta_x : x \in X\}$ where δ_x is the Dirac measure*

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \tag{3.6}$$

Proof. First, we show that an extreme probability measure μ must have the property that $\mu(B) \in \{0, 1\}$ for every measurable set $B \subseteq X$. Indeed, if $0 < \mu(B) < 1$, define the probability measures

$$\mu_B(A) := \frac{\mu(A \cap B)}{\mu(B)}, \qquad \mu_{B^c}(A) := \frac{\mu(A \cap B^c)}{\mu(B^c)}. \tag{3.7}$$

Then

$$\mu = \alpha\mu_B + (1 - \alpha)\mu_{B^c} \tag{3.8}$$

where $\alpha := \mu(B)$, hence μ is not extreme.

Now suppose μ is extreme. If $x, y \in \operatorname{supp}(\mu)$ and $x \neq y$, then there exist disjoint open sets U and V with $x \in U$ and $y \in V$; then $\mu(U) = \mu(V) = 1$ which is a contradiction. Hence $\operatorname{supp}(\mu)$ must contain exactly 1 element, so $\mu = \delta_x$ for some $x \in X$. To see that each δ_x is indeed an extreme point, note that if $\delta_x = \frac{1}{2}\nu + \frac{1}{2}\rho$, then $\operatorname{supp}(\nu) \subseteq \{x\}$ and therefore $\nu = \delta_x$, and likewise $\rho = \delta_x$. $\square$

For our next example, for a subset X of $\mathbb{R}^n$ define $\mathcal{M}_{\boldsymbol{a}}(X)$ to be the set of probability measures on X with mean vector $\boldsymbol{a} = (a_1, \ldots, a_n)$. Note that $\mathcal{M}_{\boldsymbol{a}}(X)$ is a convex and closed subset of $M(X)$ in the weak topology, hence it is compact if X is compact. Let f_i be the projection map $f_i(x_1, \ldots, x_n) = x_i$ for $1 \leq i \leq n$, and let $f_0 \equiv 1$. So for $\mu \in \mathcal{M}_{\boldsymbol{a}}(X)$,

$$\left(\int_X f_i \, d\mu \ : 0 \leq i \leq n \right) = (1, a_1, \ldots, a_n). \tag{3.9}$$

Let U be the linear span of $\{f_i : 0 \leq i \leq n\}$.

Theorem 3.2.2. *Let X be a compact subset of $\mathbb{R}^n$ and let $\mu \in \mathcal{M}_{\boldsymbol{a}}(X)$. The following are equivalent:*

(a) μ is an extreme point of $\mathcal{M}_{\boldsymbol{a}}(X)$;

(b) $L^1(\mu) = \operatorname{Span}\{f_i : 0 \leq i \leq n\}$;

(c) $\{(1, \boldsymbol{x}) : \boldsymbol{x} \in \operatorname{supp}(\mu)\}$ is a linearly independent set of vectors in $\mathbb{R}^{n+1}$.

Proof.
(a) $\Rightarrow$ *(b)*. Let $U := \operatorname{Span}\{f_i : 0 \leq i \leq n\}$, and suppose $U \neq L^1(\mu)$, so U is a proper closed

subspace of $L^1(\mu)$. It follows from the Hahn-Banach theorem and standard theory of L^p spaces that there exists a nonzero function $g \in L^\infty(\mu)$ such that

$$\int_X fg \; d\mu = 0 \qquad \text{for } f \in U. \tag{3.10}$$

Define a signed measure ν according to

$$\nu(E) := \frac{1}{\|g\|_\infty} \int_E g \; d\mu. \tag{3.11}$$

Note that $\nu(X) = 0$ since $f_0 \equiv 1 \in U$, and $\nu \geq -1$ by construction, so $\mu \pm \nu$ is a probability measure, and

$$\int_X f_i \; d(\rho \pm \nu) = \int_X f_i \; d\mu \pm \int_X f_i \; d\nu = m_i + \frac{1}{\|g\|_\infty} \int_X f_i g \; d\mu = m_i \tag{3.12}$$

so $\mu \pm \nu \in \mathcal{M}_{\boldsymbol{a}}$. Since $\mu = \frac{1}{2}(\mu + \nu) + \frac{1}{2}(\mu - \nu)$, μ is not an extreme point of $\mathcal{M}_{\boldsymbol{a}}$.

Remark: The equivalence of conditions (a) and (b) is a version of Douglas's theorem. See [18] for the direct proof of equivalence, independent of condition (c).

(b) $\Rightarrow$ *(c)*. As previously remarked, condition (b) implies $\#(\text{supp}(\mu)) \leq n + 1$, so let $\text{supp}(\mu) = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ with $k \leq n + 1$. Consider the $(n + 1) \times k$ matrix

$$A = \begin{bmatrix} 1 & \cdots & 1 \\ x_1^{(1)} & \cdots & x_1^{(k)} \\ \vdots & \ddots & \vdots \\ x_n^{(1)} & \cdots & x_n^{(k)} \end{bmatrix} \tag{3.13}$$

Since $\dim U = \dim L^1(\mu) = k$, the span of the rows has dimension k and it follows from the fundamental theorem of linear algebra that the span of the columns also has dimension k, so the k columns must be linearly independent.

(c) $\Rightarrow$ *(a)*. Since the columns of A are linearly independent, $\boldsymbol{\mu} := (\mu(\{\boldsymbol{x}^{(1)}\}), \ldots, \mu(\{\boldsymbol{x}^{(k)}\}))$ is the unique vector $\boldsymbol{y} = (y_1, \ldots, y_k) \in \mathbb{R}^k$ satisfying

$$\begin{bmatrix} 1 & \cdots & 1 \\ x_1^{(1)} & \cdots & x_1^{(k)} \\ \vdots & \ddots & \vdots \\ x_n^{(1)} & \cdots & x_n^{(k)} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} = \begin{bmatrix} 1 \\ m_1 \\ \vdots \\ m_n \end{bmatrix} \tag{3.14}$$

Therefore μ is the only probability measure in $\mathcal{M}_{\boldsymbol{a}}$ that is supported on $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}\}$, so μ cannot be a convex combination of any two other elements $\mu_1, \mu_2 \in \mathcal{M}_{\boldsymbol{a}}$ for otherwise μ_1 and μ_2 would each be supported on $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}\}$. $\square$

Corollary 3.2.3. *Let X be a compact subset of $\mathbb{R}^n$. If μ is an extreme point of $\mathcal{M}_{\boldsymbol{a}}(X)$, then*

$$\#\mathrm{supp}(\mu) \leq n+1. \tag{3.15}$$

Proof. Conditions (b) and (c) in Theorem 3.2.2 each imply that μ is supported on at most $n+1$ points. For (b), $\dim L^1(\mu) = \dim U \leq n+1$ implies the claim, and from (c) it is obvious. □

Condition (c) ensures that the points in the support of μ do not lie on the same hyperplane in $\mathbb{R}^n$. For example, if μ on $X \subseteq \mathbb{R}^2$ has support on exactly 3 points which lie on the same line in $\mathbb{R}^2$, then μ is not extreme because it is a mixture of two 2 point distributions.

More general versions of Theorem 3.2.2 can be found in [46] and [76]. The proof presented above is based on the proof in [46], with some insightful modifications. In particular, [46] uses condition (b) to prove the implication $(a) \Rightarrow (c)$ and does not note the equivalence between the three conditions.

3.3 Some inequalities

In this section, we present some new inequalities for the law of a coherent pair of random variables (X, Y). Recall Lemma 3.1.1; intuitively, if (x, y) is a possible pair for (X, Y) with no other possible pair lying on either the horizontal or vertical lines containing (x, y), so (X, Y) is not coherent, then one should not expect to make the distribution coherent just by adding some small mass to one or both lines. The amount of mass necessary should be a function of x, y, and $p_{x,y} := \mathbb{P}(X = x, Y = y)$. Indeed,

Proposition 3.3.1. *If (X, Y) is coherent and $\mathbb{P}(X = x, Y = y) > 0$ for some pair $x, y \in [0, 1]$ with $x \neq y$, then*

$$\mathbb{P}\big(\{X = x\} \triangle \{Y = y\}\big) \geq \frac{|x-y|}{1 + x \vee y}\mathbb{P}(X = x, Y = y) \tag{3.16}$$

Here $\triangle$ denotes symmetric difference.

More generally, let $A_{x_1,x_2} := \{x_1 \leq X \leq x_2\}$ and $B_{y_1,y_2} := \{y_1 \leq Y \leq y_2\}$.

Proposition 3.3.2. *If (X, Y) is coherent and $\mathbb{P}(A_{x_1,x_2} B_{y_1,y_2}) > 0$, then*

$$\mathbb{P}\big(A_{x_1,x_2} \triangle B_{y_1,y_2}\big) \geq \frac{\max\{x_1 - y_2, y_1 - x_2, 0\}}{1 + \min\{x_2, y_2\}}\mathbb{P}(A_{x_1,x_2} B_{y_1,y_2}) \tag{3.17}$$

Proof. Assume $0 \leq x_1 \leq x_2 < y_1 \leq y_2 \leq 1$, so the (possibly degenerate) rectangle $[x_1, x_2] \times [y_1, y_2]$ lies completely above the diagonal $y = x$. Let $p = \mathbb{E}X = \mathbb{E}Y$ and let $A = A_{x_1,x_2}$ and

$B = B_{y_1,y_2}$. By Proposition 3.1.2,

$$\mathbb{E}X\mathbb{1}_A + \mathbb{E}Y\mathbb{1}_{B^c} \le p + \mathbb{P}(AB^c) \tag{3.18}$$
$$\mathbb{E}X\mathbb{1}_A + p - \mathbb{E}Y\mathbb{1}_B \le p + \mathbb{P}(AB^c) \tag{3.19}$$
$$\mathbb{E}X\mathbb{1}_A - \mathbb{E}Y\mathbb{1}_B \le \mathbb{P}(AB^c). \tag{3.20}$$

Likewise,

$$\mathbb{E}Y\mathbb{1}_B - \mathbb{E}X\mathbb{1}_A \le \mathbb{P}(A^cB). \tag{3.21}$$

Hence

$$\mathbb{P}(AB^c) + \mathbb{P}(A^cB) \ge \mathbb{E}Y\mathbb{1}_B - \mathbb{E}X\mathbb{1}_A \tag{3.22}$$
$$\ge y_1\mathbb{P}(B) - x_2\mathbb{P}(A) \tag{3.23}$$
$$= (y_1 - x_2)\mathbb{P}(AB) + y_1\mathbb{P}(A^cB) - x_2\mathbb{P}(AB^c) \tag{3.24}$$

so

$$(1 + x_2)\mathbb{P}(AB^c) + (1 - y_1)\mathbb{P}(A^cB) \ge (y_1 - x_2)\mathbb{P}(AB). \tag{3.25}$$

Then $\mathbb{P}(A\triangle B) = \mathbb{P}(AB^c) + \mathbb{P}(A^cB)$ is minimized subject to the constraint (3.25) when $\mathbb{P}(A^cB) = 0$ and $\mathbb{P}(AB^c) = \frac{y_1-x_2}{1+x_2}\mathbb{P}(AB)$, so

$$\mathbb{P}(A\triangle B) \ge \frac{y_1 - x_2}{1 + x_2}\mathbb{P}(AB). \tag{3.26}$$

The general form (3.17) is obtained by consideration of the case $y_1 \le y_2 < x_1 \le x_2$ where the rectangle lies completely under the diagonal, and the case that the rectangle intersects the diagonal, in which case $x_1 \le y_2$ and $y_1 \le x_2$ and the numerator in (3.17) is 0 as should be. □

Note that in the inequality (3.22), there is no guarantee that the r.h.s. $\mathbb{E}Y\mathbb{1}_B - \mathbb{E}X\mathbb{1}_A$ is nonnegative; if it is negative, then (3.22) can be replaced by the stronger inequality $\mathbb{P}(AB^c) + \mathbb{P}(A^cB) \ge \mathbb{E}X\mathbb{1}_A - \mathbb{E}Y\mathbb{1}_B > 0$, but this does not seem to lead to anything better under the assumption that $x_2 < y_1$.

Corollary 3.3.3. *If (X, Y) is coherent, then*

$$\mathbb{P}(A_{x_1,x_2}B_{y_1,y_2}) \le \frac{1 + \min\{x_2, y_2\}}{1 + \max\{x_1, y_1, \min\{x_2, y_2\}\}} \tag{3.27}$$

Proof.

$$1 \ge \mathbb{P}(A\triangle B) + \mathbb{P}(AB) \ge \frac{\max\{x_1 - y_2, y_1 - x_2, 0\}}{1 + \min\{x_2, y_2\}}\mathbb{P}(AB) + \mathbb{P}(AB) \tag{3.28}$$
$$= \frac{1 + \max\{x_1, y_1, \min\{x_2, y_2\}\}}{1 + \min\{x_2, y_2\}}\mathbb{P}(AB) \tag{3.29}$$

□

The inequality in Corollary 3.3.3 may be easier to understand by unraveling the formula according to the three cases as before, where the rectangular region is either completely above the diagonal, completely below the diagonal, or on the diagonal:

$$\mathbb{P}(AB) \leq \begin{cases} \dfrac{1+x_2}{1+y_1} & \text{if } x_1 \leq x_2 < y_1 \leq y_2 \\ \dfrac{1+y_2}{1+x_1} & \text{if } y_1 \leq y_2 < x_1 \leq x_2 \\ 1 & \text{otherwise} \end{cases} \tag{3.30}$$

When restricted to a single point, i.e. $x_1 = x_2 = x$ and $y_1 = y_2 = y$, the bound reduces to

$$\mathbb{P}(X = x, Y = y) \leq \frac{1 + \min\{x, y\}}{1 + \max\{x, y\}} \tag{3.31}$$

Note that this is strictly better than the following sharp bound for joint distributions of (X, Y) with equal mean.

Proposition 3.3.4. *If X and Y are random variables on $[0, 1]$ with $\mathbb{E}X = \mathbb{E}Y$, then*

$$\mathbb{P}(X = x, Y = y) \leq \frac{1}{1 + |x - y|}. \tag{3.32}$$

Proof. If $x = y$, the inequality is trivial. Without loss of generality, assume $x < y$. By Theorem 3.2.2 each extreme point of $\mathcal{M}_{(a,a)}([0, 1]^2)$ for $0 \leq a \leq 1$ is supported on at most 3 points in $[0, 1]^2$ and is uniquely defined by those points. It can be shown by analysis of these extreme points that the unique extreme probably measure which maximizes the linear functional $\mathbb{P}(X = x, Y = y) = \mathbb{E}1(X = x, Y = y)$ subject to the constraint of equal means is the one supported on the two points (x, y) and $(1, 0)$, which has $\mathbb{P}(X = x, Y = y) = 1/(1 + y - x)$. Combining this with the case for $x > y$ yields the desired inequality.

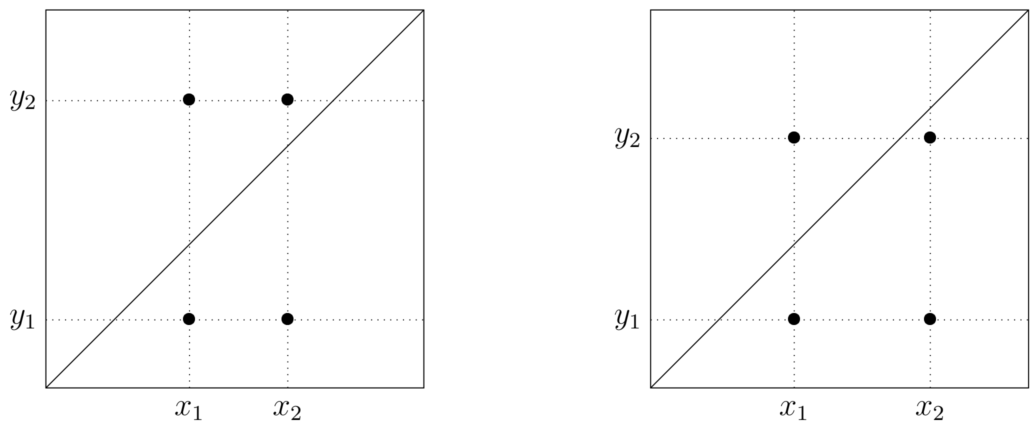
We omit the complete argument. □

3.4 2×2 extreme coherent laws

Given $0 \leq x_1 < x_2 \leq 1$ and $0 \leq y_1 < y_2 \leq 1$, we aim to describe all extreme coherent laws supported on $\{x_1, x_2\} \times \{y_1, y_2\}$, or in words, the corners of a rectangle. The problem is trivial if $x_1 \geq y_2$ or $y_1 \geq x_2$, since in either case the rectangle lies completely on one side of the diagonal, so there are no coherent laws with that support. For our analysis, let us assume that

$$y_1 \leq x_1 < y_2 \tag{3.33}$$

so the diagonal intersects left side of rectangle, including the bottom left corner, as in either of the figures below:



Results for the other case (diagonal intersects bottom side of rectangle) can then be obtained by the reflection symmetry. By Proposition 3.1.2 (iv), we have the inequalities

$$x_i p_{i\bullet} + y_j p_{\bullet j} \le p + p_{ij} \tag{3.34}$$

where

$$p_{i\bullet} := \sum_j p_{ij} \quad \text{and} \quad p_{\bullet j} := \sum_i p_{ij}. \tag{3.35}$$

The inequalities for $(i,j) = (1,1)$ and $(i,j) = (2,1)$ are trivial. The inequality for $(i,j) = (1,2)$ can be rewritten in either of the two forms

$$(x_1 - y_1)p_{11} \le y_1 p_{21} + (1 - x_1)p_{12} \tag{3.36}$$

$$(y_2 - x_2)p_{22} \le x_2 p_{21} + (1 - y_2)p_{12} \tag{3.37}$$

and the inequality for $(i,j) = (2,2)$ can be rewritten in either of the two forms

$$(x_2 - y_1)p_{21} \le y_1 p_{11} + (1 - x_2)p_{22} \tag{3.38}$$

$$(y_2 - x_1)p_{12} \le x_1 p_{11} + (1 - y_2)p_{22} \tag{3.39}$$

Proposition 3.4.1. *If $p_{ij} > 0$ for all i, j, then (p_{ij}) is not an extreme coherent law.*

Proof. First suppose (3.36) is a strict inequality, and let

$$\varepsilon := y_1 p_{21} + (1 - x_1)p_{12} - (x_1 - y_1)p_{11} > 0. \tag{3.40}$$

The homogeneous system of equations

$$z_{11} + z_{12} + z_{21} + z_{22} = 0 \tag{3.41}$$

$$(x_1 - y_1)z_{11} - (y_2 - x_1)z_{12} + (x_2 - y_1)z_{21} + (x_2 - y_2)z_{22} = 0 \tag{3.42}$$

$$y_1 z_{11} - (x_2 - y_1)z_{21} + (1 - x_2)z_{22} = 0 \tag{3.43}$$

has an at least one dimensional space of solutions in $\mathbb{R}^4$, so there exists $(\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}) \neq (0,0,0,0)$ satisfying the system of equations above such that

$$0 \leq p_{ij} \pm \delta_{ij} \leq 1 \quad \text{for all } i, j \tag{3.44}$$

and

$$|y_1\delta_{21} + (1-x_1)\delta_{12} - (x_1 - y_1)\delta_{11}| \leq \varepsilon. \tag{3.45}$$

Then $(p_{ij} \pm \delta_{ij})$ is a probability distribution by (3.41) and (3.44), whose marginal distributions have equal means by (3.42), and still satisfies (3.36) and (3.38) by (3.45) and (3.43), respectively. Therefore $(p_{ij} \pm \delta_{ij})$ are coherent and

$$p_{ij} = \frac{1}{2}(p_{ij} + \delta_{ij}) + \frac{1}{2}(p_{ij} - \delta_{ij}) \tag{3.46}$$

so (p_{ij}) is not an extreme coherent law.

Similarly, if (3.38) is a strict inquality, an analogous argument shows that (p_{ij}) is not an extreme coherent law. This leaves only the case that both (3.36) and (3.38) are equalities. A coherent law with equality in both (3.36) and (3.38) must satisfy the system of equations

$$p_{11} + p_{12} + p_{21} + p_{22} = 1 \tag{3.47}$$

$$(x_1 - y_1)p_{11} - (y_2 - x_1)p_{12} + (x_2 - y_1)p_{21} - (y_2 - x_2)p_{22} = 0 \tag{3.48}$$

$$(1 - y_2)p_{12} + x_2 p_{21} - (y_2 - x_2)p_{22} = 0 \tag{3.49}$$

$$x_1 p_{11} - (y_2 - x_1)p_{12} + (1 - y_2)p_{22} = 0 \tag{3.50}$$

Equation (3.48) can be replaced with (3.50) minus (3.48), and after rearranging and expressing in matrix notation:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1-y_2 & x_2 & -(y_2-x_2) \\ y_1 & 0 & -(x_2-y_1) & 1-x_2 \\ x_1 & -(y_2-x_1) & 0 & 1-y_2 \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \\ p_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.51}$$

Adding the second and third equations above gives

$$y_1 p_{11} + (1-y_2)p_{12} + y_1 p_{21} + (1-y_2)p_{22} = 0 \tag{3.52}$$

- If $y_1 \neq 0$ and $y_2 \neq 1$, then there is no way to assign positive probabilities p_{ij} so that this equation holds.

- If $y_1 = 0$ and $y_2 \neq 1$, then the third equation reduces to $x_2 p_{21} = (1-x_2)p_{22}$. Substituting this into the second equation gives $(1-y_2)p_{12} + (1-y_2)p_{22} = 0$. There is no way to assign positive probabilities p_{ij} to satisfy this equation.

- If $y_2 = 1$ and $y_1 \neq 0$, then the second equation reduces to $x_2 p_{21} = (1-x_2)p_{22}$. Substituting this into the third equation gives $y_1 p_{11} + y_1 p_{21} = 0$. There is no way to assign positive probabilities p_{ij} to satisfy this equation.

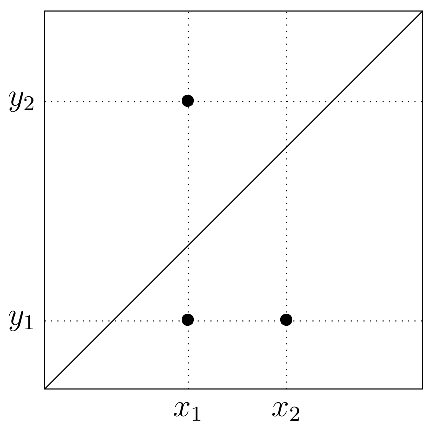
- If $y_1 = 0$ and $y_2 = 1$, then the system of equations reduces to

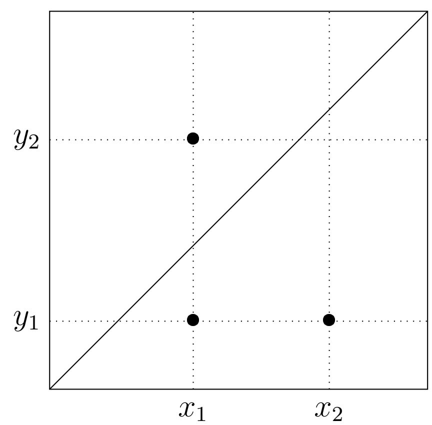
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & -(1-x_1) & 0 & 0 \\ 0 & 0 & x_2 & -(1-x_2) \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \\ p_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{3.53}$$

  This has solutions with p_{ij} all positive, but they are obviously all mixtures of the two-point laws concentrated on the lines $x = x_1$ and $x = x_2$.

□

Proposition 3.4.2. *If $y_1 \le x_1 < y_2$, there exists an extreme coherent law whose support is $\{(x_1, y_1), (x_1, y_2), (x_2, y_1)\}$ if and only if $y_1 \neq 0$, in which case it is the unique extreme coherent law with this support.*





Proof. Suppose $p_{22} = 0$. Then (3.37) holds trivially, and (3.38) reduces to

$$(x_2 - y_1)p_{21} \le y_1 p_{11}. \tag{3.54}$$

By the same argument as in Proposition 3.4.1, if (3.54) is a strict inequality then (p_{ij}) is not an extreme coherent law. Then if (3.54) is an equality, (p_{ij}) must satisfy the system of equations

$$p_{11} + p_{12} + p_{21} = 1 \tag{3.55}$$

$$y_1 p_{11} - (x_2 - y_1)p_{21} = 0 \tag{3.56}$$

$$(x_1 - y_1)p_{11} - (y_2 - x_1)p_{12} + (x_2 - y_1)p_{21} = 0 \tag{3.57}$$

which has the unique solution

$$\begin{bmatrix} p_{12} & p_{22} \\ p_{11} & p_{21} \end{bmatrix} = \begin{bmatrix} \frac{x_1(x_2-y_1)}{(y_2-x_1)(x_2-y_1)+x_1(x_2-y_1)+y_1(y_2-x_1)} & 0 \\ \frac{(y_2-x_1)(x_2-y_1)}{(y_2-x_1)(x_2-y_1)+x_1(x_2-y_1)+y_1(y_2-x_1)} & \frac{y_1(y_2-x_1)}{(y_2-x_1)(x_2-y_1)+x_1(x_2-y_1)+y_1(y_2-x_1)} \end{bmatrix} \tag{3.58}$$

with positive p_{11}, p_{12}, p_{21} if and only if $y_1 \neq 0$.

Here, the auxiliary function ϕ from Proposition 3.1.2 (iii) with $\mathbb{E}(\phi(X,Y)|X) = X$ and $\mathbb{E}(\phi(X,Y)|Y) = Y$ is given by

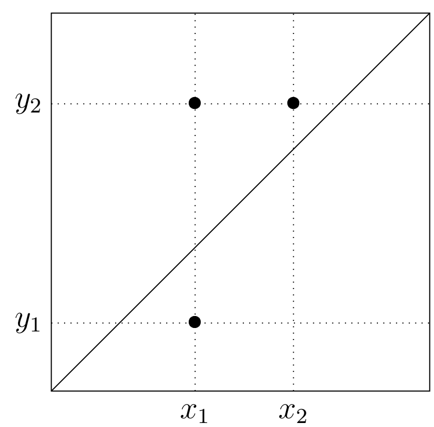
$$\phi(x_1, y_2) = \mathbb{E}(\phi(X,Y)|Y = y_2) = y_2, \tag{3.59}$$

$$\phi(x_2, y_1) = \mathbb{E}(\phi(X,Y)|X = x_2) = x_2, \tag{3.60}$$

$$\phi(x_1, y_1)\frac{y_2 - x_1}{y_2} + \underbrace{\phi(x_1, y_2)}_{y_2}\frac{x_1}{y_2} = \mathbb{E}(\phi(X,Y)|X = x_1) = x_1 \implies \phi(x_1, y_1) = 0. \tag{3.61}$$

□

Proposition 3.4.3. *If $y_1 \leq x_1 < y_2$, there exists an extreme coherent law whose support is $\{(x_1, y_1), (x_1, y_2), (x_2, y_2)\}$ if and only if $y_1 < x_1$ and $x_2 < y_2$, and $y_2 \neq 1$, in which case it is the unique extreme coherent law with this support.*



Proof. Suppose $p_{21} = 0$. Then (3.38) holds trivially, and (3.36) reduces to

$$(x_1 - y_1)p_{11} \leq (1 - x_1)p_{12}. \tag{3.62}$$

As before, if (3.62) is a strict inequality then (p_{ij}) is not an extreme coherent law. Then if (3.62) is an equality, (p_{ij}) must satisfy the system of equations

$$p_{11} + p_{12} + p_{22} = 1 \tag{3.63}$$

$$(x_1 - y_1)p_{11} - (1 - x_1)p_{12} = 0 \tag{3.64}$$

$$(x_1 - y_1)p_{11} - (y_2 - x_1)p_{12} - (y_2 - x_2)p_{22} = 0 \tag{3.65}$$

which has the unique solution

$$\begin{bmatrix} p_{12} & p_{22} \\ p_{11} & p_{21} \end{bmatrix} = \begin{bmatrix} \frac{(y_2-x_2)(x_1-y_1)}{(y_2-x_2)(x_1-y_1)+(1-y_2)(x_1-y_1)+(1-x_1)(y_2-x_2)} & \frac{(1-y_2)(x_1-y_1)}{(y_2-x_2)(x_1-y_1)+(1-y_2)(x_1-y_1)+(1-x_1)(y_2-x_2)} \\ \frac{(1-x_1)(y_2-x_2)}{(y_2-x_2)(x_1-y_1)+(1-y_2)(x_1-y_1)+(1-x_1)(y_2-x_2)} & 0 \end{bmatrix} \tag{3.66}$$

with positive p_{11}, p_{12}, p_{22} if and only if $x_1 > y_1$, $y_2 > x_2$, and $y_2 < 1$.

Here, the auxiliary function ϕ is given by

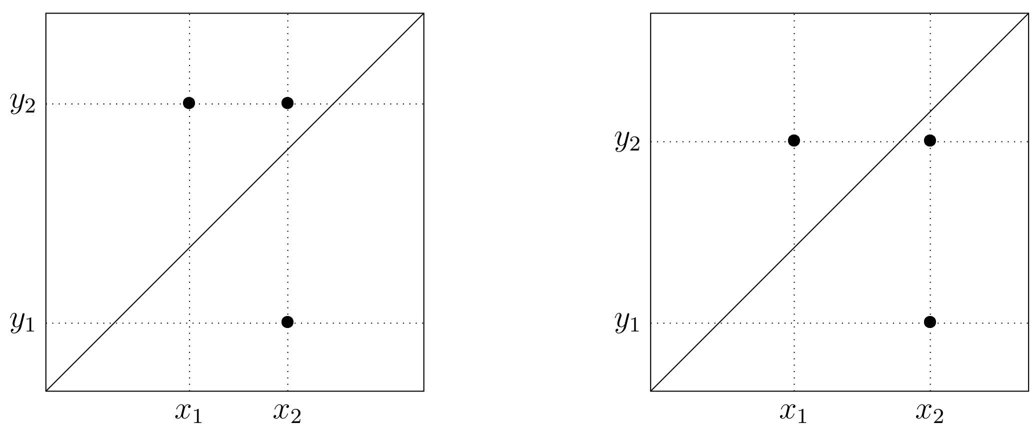
$$\phi(x_1, y_1) = \mathbb{E}(\phi(X, Y)|Y = y_1) = y_1, \tag{3.67}$$

$$\phi(x_2, y_2) = \mathbb{E}(\phi(X, Y)|X = x_2) = x_2, \tag{3.68}$$

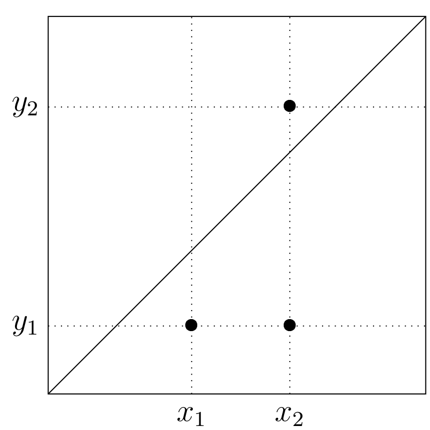
$$\underbrace{\phi(x_1, y_1)}_{y_1}\frac{1 - x_1}{1 - y_1} + \phi(x_1, y_2)\frac{x_1 - y_1}{1 - y_1} = \mathbb{E}(\phi(X, Y)|X = x_1) = x_1 \implies \phi(x_1, y_2) = 1. \tag{3.69}$$

□

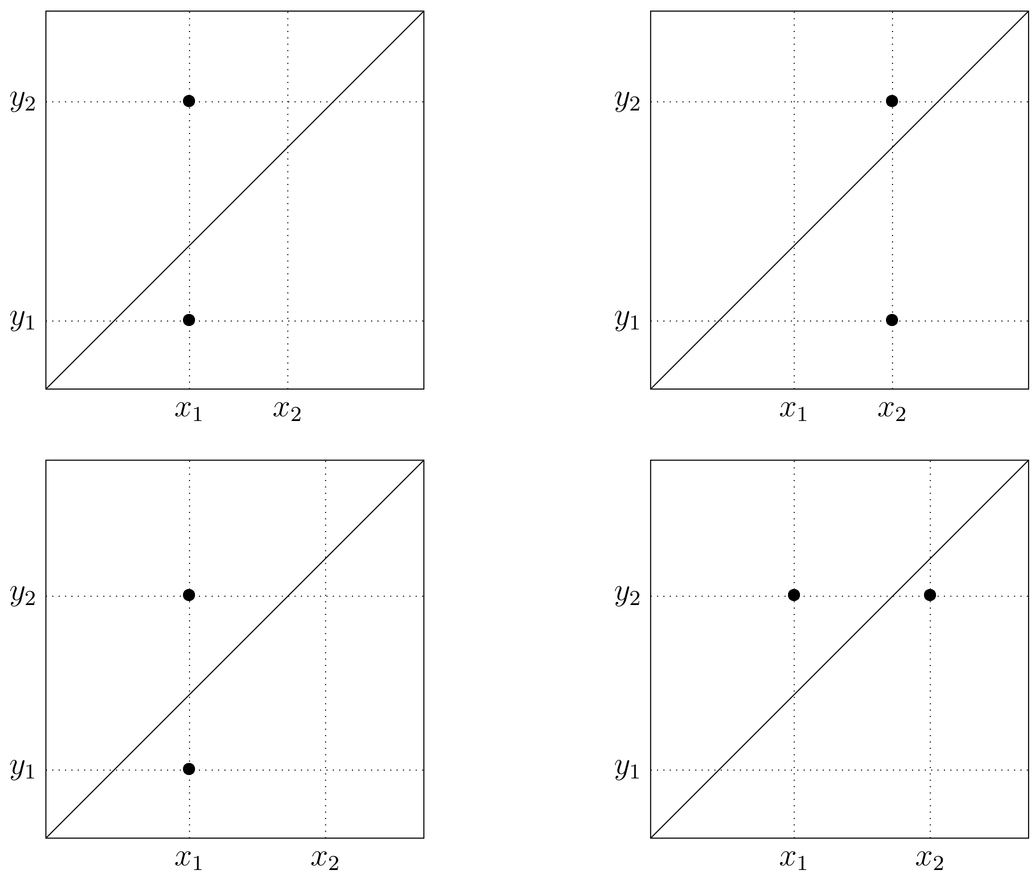
Corollary 3.4.4. *If $y_1 \leq x_1 < y_2$, there exists an extreme coherent law whose support is $\{(x_1, y_2), (x_2, y_1), (x_2, y_2)\}$ if and only if $x_2 \neq 1$, in which case it is the unique extreme coherent law with this support.*



Corollary 3.4.5. *If $y_1 \leq x_1 < y_2$, there exists an extreme coherent law whose support is $\{(x_1, y_1), (x_2, y_1), (x_2, y_2)\}$ if and only if $y_1 < x_1$ and $x_2 < y_2$, and $y_1 \neq 0$, in which case it is the unique extreme coherent law with this support.*



Proposition 3.4.6. *If $y_1 \leq x_1 < y_2$, there are exactly two distinct extreme coherent laws that are supported on at most two points of $\{x_1, x_2\} \times \{y_1, y_2\}$.*



Corollary 3.4.7. *If $y_1 \leq x_1 < y_2$, the number of extreme coherent laws supported on $\{x_1, x_2\} \times \{y_1, y_2\}$ is exactly*

(i) *2 if $y_1 = 0$ and $y_2 = 1$ OR $y_1 = 0$ and $x_2 = 1$;*

(ii) *3 if $y_2 \leq x_2 < 1$ and $y_1 = 0$ OR if $x_1 = y_1 > 0$ and $y_2 = 1$ OR if $y_1 > 0$ and $x_2 = 1$;*

(iii) *4 if $0 < y_1 \leq x_1 < y_2 \leq x_2 < 1$ OR $0 = y_1 < x_1 < x_2 < y_2 < 1$ OR $0 < y_1 < x_1 < x_2 < y_2 = 1$;*

(iv) *6 if $0 < y_1 < x_1 < x_2 < y_2 < 1$.*

3.5 Application: an optimization problem

In [9] and [8], the authors consider for $0 \leq \delta \leq 1$ the problem of evaluating

$$\varepsilon(\delta) := \sup_{\text{coherent } (X,Y)} \mathbb{P}(|X - Y| \geq 1 - \delta) \tag{3.70}$$

and $\varepsilon_{2\times 2}(\delta)$ defined like (3.70) but with the supremum taken over the restricted set of 2×2 coherent laws. The quantity

$$\mathbb{P}(|X - Y| \geq 1 - \delta) = \int_{[0,1]^2} 1(|x - y| \geq 1 - \delta) d\mu(x, y), \tag{3.71}$$

where μ is the law of (X, Y), is a linear functional over the set of probability measures on $[0, 1]$. Since the set of coherent distributions is a compact convex subset, the objective function must attain its maximum at an extreme point of the set, and similarly over 2×2 coherent distributions the functional must attain its maximum at one of the 2×2 extreme coherent distributions. We present an alternative proof to the following result from [9] using extreme point analysis, based on the results in the previous section.

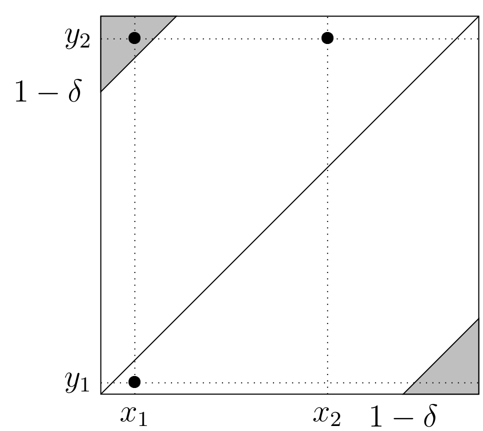
Proposition 3.5.1. *[9] For* $\delta \in [0, \frac{1}{2})$,

$$\varepsilon_{2\times 2}(\delta) = \frac{2\delta}{1 + \delta}. \tag{3.72}$$

Proof. By symmetries, it suffices to consider the extreme coherent laws with support either $\{(x_1, y_1), (x_1, y_2), (x_2, y_2)\}$ with $y_1 \leq x_1 < y_2 < 1$, or $\{(x_1, y_1), (x_1, y_2), (x_2, y_1)\}$ with $0 < y_1 \leq x_1 < y_2$.

First, consider the extreme coherent law with support $\{(x_1, y_1), (x_1, y_2), (x_2, y_2)\}$.

Case 1. $y_2 - x_1 \geq 1 - \delta$ and $y_2 - x_2 < 1 - \delta$.



Consider

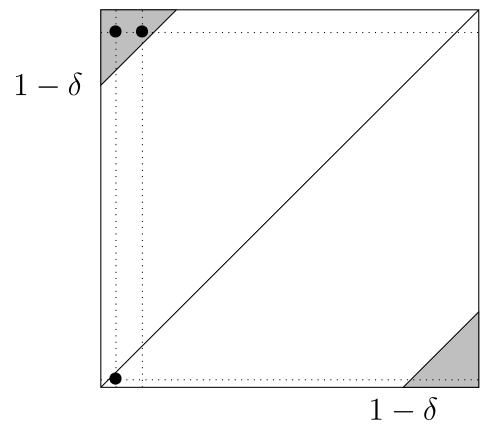
$$\mathbb{P}(|X - Y| \geq 1 - \delta) = p_{12} = \frac{(y_2 - x_2)(x_1 - y_1)}{(y_2 - x_2)(x_1 - y_1) + (1 - y_2)(x_1 - y_1) + (1 - x_1)(y_2 - x_2)}. \tag{3.73}$$

Any (x_1, x_2, y_1, y_2) satisfying the specified constraints maximizes (3.73) if and only if it minimizes

$$\frac{1-y_2}{y_2-x_2}+\frac{1-x_1}{x_1-y_1}. \tag{3.74}$$

But no matter the choice of x_2 and y_2 satisfying $y_2 - x_2 < 1-\delta$, x_2 can always be adjusted to make (3.74) smaller, and therefore (3.73) cannot be maximized subject to these constraints.

Case 2. $y_2 - x_2 \geq 1-\delta$.



In this case,

$$\mathbb{P}(|X-Y| \geq 1-\delta) = p_{12}+p_{22} = \frac{(1-x_2)(x_1-y_1)}{(y_2-x_2)(x_1-y_1)+(1-y_2)(x_1-y_1)+(1-x_1)(y_2-x_2)}. \tag{3.75}$$

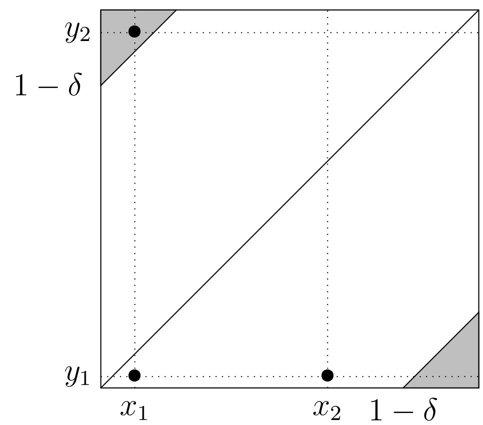
Any (x_1, x_2, y_1, y_2) satisfying the specified constraints maximizes (3.75) if and only if it minimizes

$$\frac{(1-x_1)(y_2-x_2)}{(1-x_2)(x_1-y_1)} \tag{3.76}$$

But for a $x_1 < x_2$, x_1 can always be increased to make (3.76) smaller, and therefore (3.75) cannot be maximized subject to these constraints.

Now consider the extreme coherent law with support $\{(x_1, y_1), (x_1, y_2), (x_2, y_1)\}$.

Case 1. $y_2 - x_1 \geq 1-\delta$ and $x_2 - y_1 < 1-\delta$.



Here

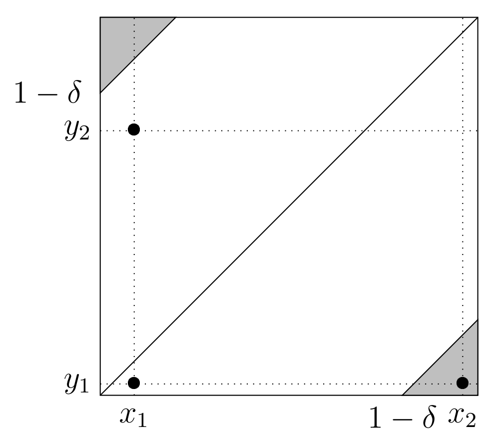
$$\mathbb{P}(|X-Y| \geq 1-\delta) = p_{12} = \frac{x_1(x_2-y_1)}{(y_2-x_1)(x_2-y_1)+x_1(x_2-y_1)+y_1(y_2-x_1)} \tag{3.77}$$

which is maximized if and only if

$$\frac{x_2(y_2-x_1)}{x_1(x_2-y_1)} \tag{3.78}$$

is minimized. But for every choice of x_2 and y_1 satisfying $x_2 - y_1 < 1-\delta$, y_1 can always be decreased to make (3.78) smaller, so (3.77) cannot be maximized subject to these constraints.

Case 2. $y_2 - x_1 < 1-\delta$ and $x_2 - y_1 \geq 1-\delta$.



Here

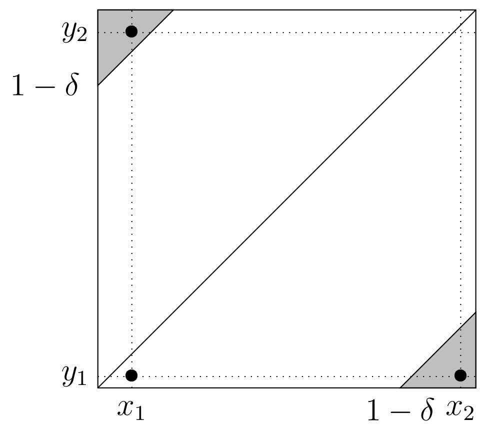
$$\mathbb{P}(|X-Y| \geq 1-\delta) = p_{21} = \frac{y_1(y_2-x_1)}{(y_2-x_1)(x_2-y_1)+x_1(x_2-y_1)+y_1(y_2-x_1)} \tag{3.79}$$

which is maximized if and only if

$$\frac{y_2(x_2-y_1)}{y_1(y_2-x_1)} \tag{3.80}$$

is minimized. But for every choice of x_1 and y_2 with $y_2 - x_1 < 1 - \delta$, x_1 can be decreased to make (3.80) smaller, so (3.79) cannot be maximized subject to these constraints.

Case 3. $y_2 - x_1 \geq 1 - \delta$ and $x_2 - y_1 \geq 1 - \delta$.



Then

$$\mathbb{P}(|X - Y| \geq 1 - \delta) = p_{12} + p_{21} = \frac{x_1(x_2 - y_1) + y_1(y_2 - x_1)}{(y_2 - x_1)(x_2 - y_1) + x_1(x_2 - y_1) + y_1(y_2 - x_1)} \tag{3.81}$$

which is maximized if and only if

$$\frac{x_1}{y_2 - x_1} + \frac{y_1}{x_2 - y_1} \tag{3.82}$$

is maximized. For fixed x_1 and y_1 with $x_1 \geq y_1$, (3.82) is maximized when $y_2 - x_1 = 1 - \delta$ and $x_2 - y_1 = 1 - \delta$. Therefore (3.82) is maximized when $x_1 = y_1 = \delta$ and $x_2 = y_2 = 1$, in which case (3.81) subject to the specified constraints attains its maximum value of

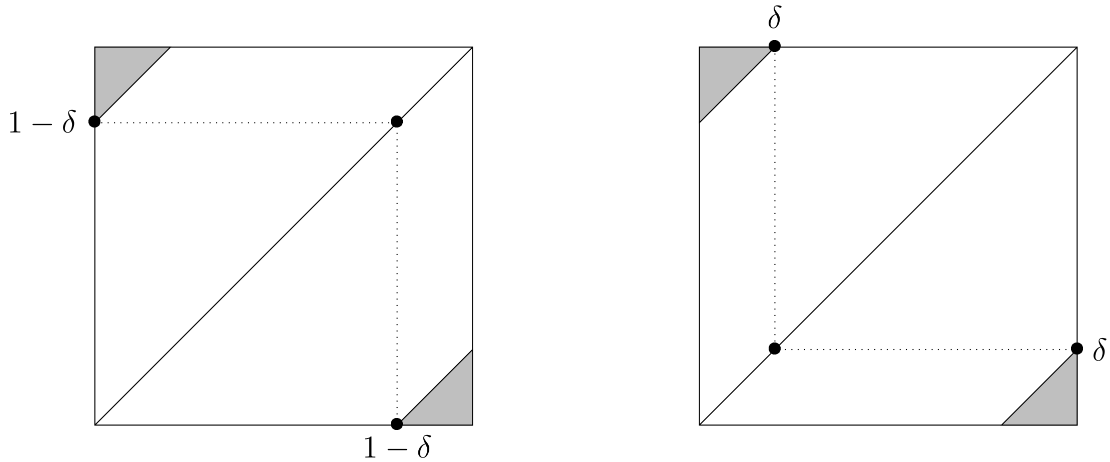
$$\frac{2\delta}{1 + \delta}. \tag{3.83}$$

Combining all the different cases analyzed here along with the easy result that $\varepsilon_{2\times 1}(\delta) = \delta < \frac{2\delta}{1+\delta}$, it follows that

$$\varepsilon_{2\times 2}(\delta) = \frac{2\delta}{1 + \delta}. \tag{3.84}$$

□

By consideration of symmetries, we see that $\mathbb{P}(|X - Y| \geq 1 - \delta)$ is maximized at exactly two of the 2×2 extreme coherent laws: the one with support $\{(\delta, \delta), (\delta, 1), (1, \delta)\}$, and the one with support $\{(1 - \delta, 1 - \delta), (0, 1 - \delta), (1 - \delta, 0)\}$.



3.6 More extreme coherent laws

We first establish the following lemma.

Lemma 3.6.1. *Let (X_1, Y_1) and (X_2, Y_2) be pairs of coherent random variables, independent of each other, with corresponding auxiliary functions ϕ_i, i.e.*

$$\sum_y \phi_i(x,y) r_i(x,y) = x \sum_y r_i(x,y) \tag{3.85}$$

$$\sum_x \phi_i(x,y) r_i(x,y) = y \sum_x r_i(x,y) \tag{3.86}$$

where

$$r_i(x,y) := \mathbb{P}(X_i = x, Y_i = y). \tag{3.87}$$

Let B_p, $0 \le p \le 1$ be a Bernoulli(p) random variable independent of X_i and Y_i for $i = 1, 2$. Define

$$X' = B_p X_1 + (1 - B_p) X_2, \qquad Y' = B_p Y_1 + (1 - B_p) Y_2. \tag{3.88}$$

Then (X', Y') is coherent with corresponding auxiliary function

$$\phi'(x,y) := \frac{p r_1(x,y)}{p r_1(x,y) + (1-p) r_2(x,y)} \phi_1(x,y) + \frac{(1-p) r_2(x,y)}{p r_1(x,y) + (1-p) r_2(x,y)} \phi_2(x,y) \tag{3.89}$$

Proof. Let ϕ' be defined as above. Note that

$$r'(x,y) := \mathbb{P}(X' = x, Y' = y) = p r_1(x,y) + (1-p) r_2(x,y). \tag{3.90}$$

Then

$$x\sum_y r'(x,y) = x\sum_y \big[pr_1(x,y) + (1-p)r_2(x,y)\big] \tag{3.91}$$

$$= p\Big[x\sum_y r_1(x,y)\Big] + (1-p)\Big[x\sum_y r_2(x,y)\Big] \tag{3.92}$$

$$= p\sum_y \phi_1(x,y)r_1(x,y) + (1-p)\sum_y \phi_2(x,y)r_2(x,y) \tag{3.93}$$

$$= \sum_y \phi'(x,y)r'(x,y) \tag{3.94}$$

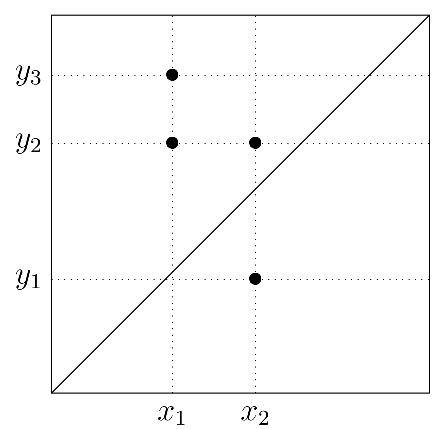
and likewise

$$y\sum_x r'(x,y) = \sum_x \phi'(x,y)r'(x,y). \tag{3.95}$$

□

To better understand the properties of extreme coherent laws, consider the following 2×3 examples.

Example 1. Let $0 < x_1 < x_2$ and $y_1 < y_2 < y_3$ and $y_1 < x_2 < y_2$. Let $(p_{12}, p_{13}, p_{21}, p_{22})$ be a probability distribution on $S = \{(x_1, y_2), (x_1, y_3), (x_2, y_1), (x_2, y_2)\}$ with all nonzero probabilities.



It is coherent if and only if there is a corresponding auxiliary function $\phi : S \to [0, 1]$ satisfying the constraints for coherence: ($\phi_{ij} := \phi(x_i, y_j)$)

$$y_3 = \phi_{13} \tag{3.96}$$

$$x_1 = \frac{p_{12}}{p_{12}+p_{13}}\phi_{12} + \frac{p_{13}}{p_{12}+p_{13}}\phi_{13} \tag{3.97}$$

$$y_2 = \frac{p_{12}}{p_{12}+p_{22}}\phi_{12} + \frac{p_{22}}{p_{12}+p_{22}}\phi_{22} \tag{3.98}$$

$$x_2 = \frac{p_{21}}{p_{21}+p_{22}}\phi_{21} + \frac{p_{22}}{p_{21}+p_{22}}\phi_{22} \tag{3.99}$$

$$y_1 = \phi_{21} \tag{3.100}$$

Substituting and rearranging the middle three equations yields

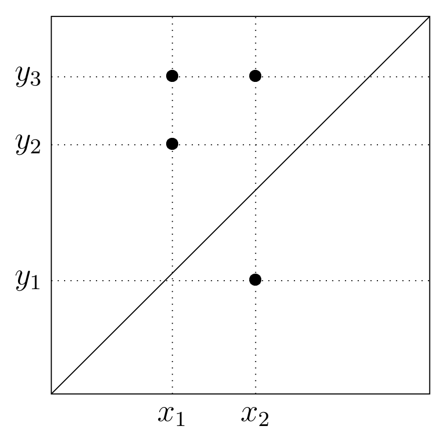
$$p_{12}(x_1 - \phi_{12}) = p_{13}(y_3 - x_1) \tag{3.101}$$

$$p_{12}(y_2 - \phi_{12}) = p_{22}(\phi_{22} - y_2) \tag{3.102}$$

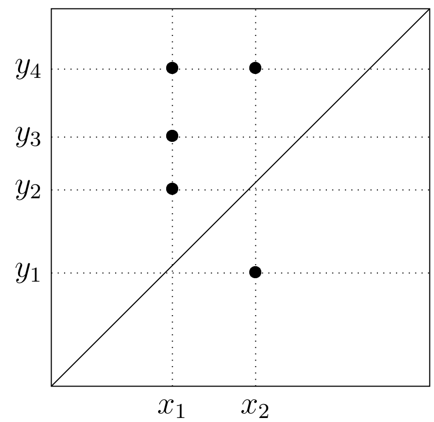
$$p_{21}(x_2 - y_1) = p_{22}(\phi_{22} - x_2) \tag{3.103}$$

Observe that ϕ exists and satisfies $\phi_{13} = y_3$, $\phi_{21} = y_1$, $\phi_{12} < x_1$, and $\phi_{22} > y_2$ if and only if (p_{ij}) is coherent with all nonzero probabilities, and if ϕ satisfies these conditions, then (p_{ij}) is uniquely determined. Suppose $(\phi_{12}, \phi_{13}, \phi_{21}, \phi_{22})$ satisfies these conditions, and suppose $0 < \phi_{12} < x_1$. Then $0 \le \phi_{12} - \varepsilon < \phi_{12} < \phi_{12} + \varepsilon < x_1$ for some $\varepsilon > 0$. Then by Lemma 3.6.1, the law corresponding to $(\phi_{12}, \phi_{13}, \phi_{21}, \phi_{22})$ is a probabilistic mixture of the laws corresponding to $(\phi_{12} - \varepsilon, \phi_{13}, \phi_{21}, \phi_{22})$ and $(\phi_{12} + \varepsilon, \phi_{13}, \phi_{21}, \phi_{22})$, so it is not an extreme coherent law. Likewise if $y_2 < \phi_{22} < 1$; therefore the only law with support S which could be extreme is $(\phi_{12}, \phi_{13}, \phi_{21}, \phi_{22}) = (0, y_3, y_1, 1)$, and indeed it must be because there are no other extreme laws supported on S for which (x_1, y_3) has nonzero probability.

Example 2. Let $0 < x_1 < x_2$ and $y_1 < y_2 < y_3 < 1$ and $y_1 < x_2 < y_3$ and $x_1 < y_2$. By similar reasoning as in Example 1, there is a unique extreme coherent law with support $\{(x_1, y_2), (x_1, y_3), (x_2, y_1), (x_2, y_3)\}$ corresponding to $(\phi_{12}, \phi_{13}, \phi_{21}, \phi_{23}) = (y_2, 0, y_1, 1)$.



Example 3. (Nonexample) Let $0 < x_1 < x_2$ and $y_1 < y_2 < y_3 < y_4 < 1$. and $y_1 < x_2 < y_4$ and $x_1 < y_2$. Let $(p_{12}, p_{13}, p_{14}, p_{21}, p_{24})$ be a probability distribution on $S = \{(x_1, y_2), (x_1, y_3), (x_1, y_4), (x_2, y_1), (x_2, y_4)\}$.



Then

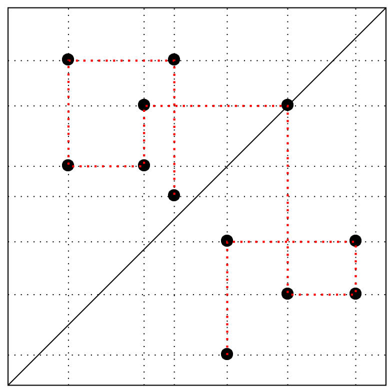
$$y_2 = \phi_{12} \tag{3.104}$$
$$y_3 = \phi_{13} \tag{3.105}$$
$$x_1 = \frac{p_{12}}{p_{12}+p_{13}+p_{14}}\phi_{12} + \frac{p_{13}}{p_{12}+p_{13}+p_{14}}\phi_{13} + \frac{p_{14}}{p_{12}+p_{13}+p_{14}}\phi_{14} \tag{3.106}$$
$$y_4 = \frac{p_{14}}{p_{14}+p_{24}}\phi_{14} + \frac{p_{24}}{p_{14}+p_{24}}\phi_{24} \tag{3.107}$$
$$x_2 = \frac{p_{21}}{p_{21}+p_{24}}\phi_{21} + \frac{p_{24}}{p_{21}+p_{24}}\phi_{24} \tag{3.108}$$
$$y_1 = \phi_{21} \tag{3.109}$$

Substituting and rearranging the unsolved equations yields

$$p_{14}(x_1 - \phi_{14}) = p_{12}(y_2 - x_1) + p_{13}(y_3 - x_1) \tag{3.110}$$
$$p_{14}(y_4 - \phi_{14}) = p_{24}(\phi_{24} - y_4) \tag{3.111}$$
$$p_{21}(x_2 - y_1) = p_{24}(\phi_{24} - x_2) \tag{3.112}$$

It follows that if (p_{ij}) is coherent, then $\phi_{14} < x_1$ and $\phi_{24} > y_4$. On the other hand, given ϕ satisfying $\phi_{12} = y_2$, $\phi_{13} = y_3$, $\phi_{21} = y_1$, $\phi_{14} < x_1$, and $\phi_{24} > y_4$, for $r > 0$, there is a unique probability distribution (p_{ij}) satisfying the requisite equations such that $p_{13}/p_{12} = r$. But by Example 2, there is a coherent law with support $\{(x_1, y_2), (x_1, y_4), (x_2, y_1), (x_2, y_4)\}$ corresponding to ϕ' with $(\phi'_{12}, \phi'_{14}, \phi'_{21}, \phi'_{24}) = (\phi_{12}, \phi_{14}, \phi_{21}, \phi_{24})$, and a coherent law with support $\{(x_1, y_3), (x_1, y_4), (x_2, y_1), (x_2, y_4)\}$ corresponding to ϕ'' with $(\phi''_{13}, \phi''_{14}, \phi''_{21}, \phi''_{24}) = (\phi_{13}, \phi_{14}, \phi_{21}, \phi_{24})$. Then there is a unique mixture of these two laws for which the ratio of probabilities of (x_1, y_3) to (x_1, y_2) is equal to r, with corresponding ϕ as desired. Therefore, there are no extreme coherent laws with support S.

Extrapolation of the arguments in the preceding examples reveals a structure which is a sufficient condition for a set to be the support of a unique extreme coherent law. For example, the configuration of points in the figure below:



Let us formalize the properties of this type of set and our claim.

A large class of extreme coherent laws with finite support.

Call a finite set of points $S \subseteq [0,1]^2$ *traceable* if for every pair of distinct points $(x, y), (x', y') \in S$, there exists a unique sequence of distinct points $(x_0, y_0), \ldots, (x_k, y_k)$ in S with $(x_0, y_0) = (x, y)$ and $(x_k, y_k) = (x', y')$ such that for $1 \le i \le k$, either $x_i = x_{i-1}$ or $y_i = y_{i-1}$. In particular, if S is traceable, then

- no horizontal or vertical line contains more than two points in S;
- there are no "loops" in S;
- there exists an ordering of S which satisfies the condition in the definition of traceable, i.e. the points in S can be labeled so that $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ with either $x_i = x_{i-1}$ or $y_i = y_{i-1}$ for $i = 2, \ldots, n$.

Given a traceable set S, let

$$\overline{S} := \{(x, y) : \exists\, x' \le x \le x'' \text{ with } (x', y), (x'', y) \in S \text{ or } \exists\, y' \le y \le y'' \text{ with } (x, y'), (x, y'') \in S\} \tag{3.113}$$

In words, $\overline{S}$ is the union of all horizontal and vertical line segments connecting points in S.

Theorem 3.6.2. *Let S be a finite subset of $[0,1]^2$ with at least 2 points. Then there exists an extreme coherent law with support S if* *(i)* S *is traceable,*

(ii) $S \cap \{(x, y) : x < y\} \neq \emptyset$*,* $S \cap \{(x, y) : x > y\} \neq \emptyset$*, and* $\left|\overline{S} \cap \{(x, y) : x = y\}\right| = 1$*, and*

(iii) $\left|\overline{S} \cap \left\{(x, y) : x \in \{0, 1\} \text{ or } y \in \{0, 1\}\right\}\right| \leq 2$*,*

in which case the extreme coherent law is unique.

Proof. Suppose S satisfies (i), (ii), and (iii) and $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where $y_1 > x_1$ and $x_n > y_n$, and for each $2 \leq i \leq n$, either $x_i = x_{i-1}$ or $y_i = y_{i-1}$. Let $c := \min\{k : x_k \geq y_k\}$. There are several cases to consider: $x_1 = x_2$ or $y_1 = y_2$, n even or odd, and c even or odd. Let us assume that $x_1 = x_2$, n is odd (hence $y_{n-1} = y_n$), and c is even; the proofs of the other cases are nearly the same. Suppose $(p_1, \ldots, p_n)$ is a coherent law on S with all nonzero probabilities, and let $\phi = (\phi_1, \ldots, \phi_n)$ where $\phi_i = \phi(x_i, y_i)$ be the corresponding auxiliary function. It is neccessary that $\phi_1 = y_1$ and $\phi_n = x_n$. Furthermore,

$$\frac{p_i}{p_i + p_{i+1}}\phi_i + \frac{p_{i+1}}{p_i + p_{i+1}}\phi_{i+1} = x_i = x_{i+1} \qquad \text{for } i \text{ odd} \tag{3.114}$$

$$\frac{p_i}{p_i + p_{i+1}}\phi_i + \frac{p_{i+1}}{p_i + p_{i+1}}\phi_{i+1} = y_i = y_{i+1} \qquad \text{for } i \text{ even} \tag{3.115}$$

$$\tag{3.116}$$

or equivalently

$$p_i(\phi_i - x_i) = p_{i+1}(x_{i+1} - \phi_{i+1}) \qquad \text{for } i \text{ odd} \tag{3.117}$$

$$p_i(\phi_i - y_i) = p_{i+1}(y_{i+1} - \phi_{i+1}) \qquad \text{for } i \text{ even.} \tag{3.118}$$

Note that condition (ii) guarantees that $0 < x_i, y_i < 1$ for $2 \leq i \leq n - 1$. By consideration of the edge cases $\phi_1 = y_1$ and $\phi_n = x_n$ and sign analysis,

$$\phi_i < x_i \qquad \text{for } i \text{ even, } 2 \leq i \leq c - 1 \tag{3.119}$$

$$\phi_i > y_i \qquad \text{for } i \text{ odd, } 2 \leq i \leq c - 1 \tag{3.120}$$

$$\phi_i < y_i \qquad \text{for } i \text{ even, } c \leq i \leq n - 1 \tag{3.121}$$

$$\phi_i > x_i \qquad \text{for } i \text{ odd, } c \leq i \leq n - 1 \tag{3.122}$$

which simplifies to

$$0 \leq \phi_i < \min(x_i, y_i) \qquad \text{for } i \text{ even, } 2 \leq i \leq n - 1 \tag{3.123}$$

$$1 \geq \phi_i > \max(x_i, y_i) \qquad \text{for } i \text{ odd, } 2 \leq i \leq n - 1 \tag{3.124}$$

Conversely, given $\phi = (\phi_1, \ldots, \phi_n)$ satisfying $\phi_1 = y_1$, $\phi_n = x_n$, and (3.123) and (3.124), there is a unique law $(p_1, \ldots, p_n)$ corresponding to ϕ which is uniquely recovered according to (3.117) and (3.118). Hence there is a bijection between coherent laws with support S and auxiliary functions ϕ satisfying the specified conditions.

Now suppose $\phi = (\phi_1, \ldots, \phi_n)$ satisfies $\phi_1 = y_1$, $\phi_n = x_n$, and (3.123) and (3.124), and that $0 < \phi_k < 1$ for some $2 \leq k \leq n-1$. Let $\varepsilon > 0$ be such that

$$\phi' := (\phi_1, \ldots, \phi_k - \varepsilon, \ldots, \phi_n) \qquad \text{and} \qquad \phi'' := (\phi_1, \ldots, \phi_k + \varepsilon, \ldots, \phi_n) \tag{3.125}$$

satisfy (3.123) and (3.124). Then by Lemma 3.6.1, the coherent law corresponding to ϕ is a mixture of the coherent laws corresponding to ϕ' and ϕ'' and hence is not extreme. It follows that the coherent law corresponding to $\phi = (y_1, 0, 1, 0, \ldots, 1, 0, x_n)$ (alternating 0s and 1s) is extreme and is therefore the unique extreme coherent law with support S. □

The validity of the converse to Theorem 3.6.2 remains an open problem:

Conjecture 3.6.3. *Let S be a finite subset of $[0,1]^2$ with at least 2 points. Then there exists an extreme coherent law with support S if and only if*

(i) S is traceable,

(ii) $S \cap \{(x,y) : x < y\} \neq \emptyset$, $S \cap \{(x,y) : x > y\} \neq \emptyset$, and $\left|\overline{S} \cap \{(x,y) : x = y\}\right| = 1$, and

(iii) $\left|\overline{S} \cap \{(x,y) : x \in \{0,1\} \text{ or } y \in \{0,1\}\}\right| \leq 2$,

in which case the extreme coherent law is unique.

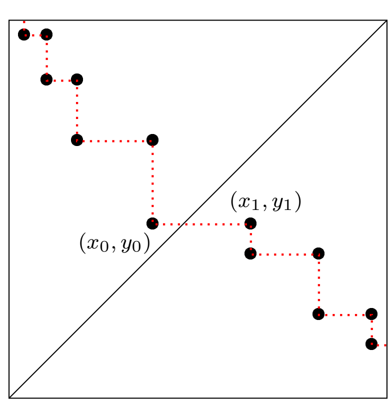
Extreme coherent laws with infinite support

The ideas from the previous discussion of extreme coherent laws with support on traceable sets can, to some extent, be pushed to infinite sets. Series convergence becomes a concern for general "infinite traceable sets," but can be guaranteed if restricted properly. Consider an infinite collection of points $S = \{(x_i, y_i)\}_{i \in \mathbb{Z}}$ in $[0,1]^2$ satisfying

(i) $\inf\{k : x_k > y_k\} = 1$

(ii) $x_{i+1} = x_i$ and $y_{i+1} < y_i$ for i odd

(iii) $y_{i+1} = y_i$ and $x_{i+1} > x_i$ for i even



Consider the prescribed auxiliary function ϕ where $\phi_i := \phi(x_i, y_i) = 1$ for i odd and 0 for i even. Then a coherent law $(p_i)_{i\in\mathbb{Z}}$ corresponding to ϕ satisfies (3.117) and (3.118) which can be simplified to

$$p_{2k} = p_0 \prod_{i=1}^{k} \frac{y_{2i-1}}{x_{2i-1}} \cdot \frac{1-x_{2i-1}}{1-y_{2i-1}} \tag{3.126}$$

$$p_{2k+1} = p_0 \frac{y_0}{1-y_0} \prod_{i=1}^{k} \frac{y_{2i}}{x_{2i}} \cdot \frac{1-x_2}{1-y_2} \tag{3.127}$$

$$p_{-2k} = p_0 \prod_{i=1}^{k} \frac{x_{-2i+1}}{y_{-2i+1}} \cdot \frac{1-y_{-2i+1}}{1-x_{-2i+1}} \tag{3.128}$$

$$p_{-2k-1} = p_0 \frac{x_0}{1-x_0} \prod_{i=1}^{k} \frac{x_{-2i}}{y_{-2i}} \cdot \frac{1-y_{-2i}}{1-x_{-2i}} \tag{3.129}$$

for $k = 0, 1, 2, \ldots$. Let $R_+ = \frac{y_1}{x_1} \cdot \frac{1-x_1}{1-y_1}$ and $R_- = \frac{x_{-1}}{y_{-1}} \cdot \frac{1-y_{-1}}{1-x_{-1}}$ and note that $0 < R_\pm < 1$ and

$$p_{2k} \le p_0 R_+^k \tag{3.130}$$

$$p_{2k+1} \le p_0 \frac{y_0}{1-y_0} R_+^k \tag{3.131}$$

$$p_{-2k} \le p_0 R_-^k \tag{3.132}$$

$$p_{-2k-1} \le p_0 \frac{x_0}{1-x_0} R_-^k \tag{3.133}$$

so the series

$$\begin{aligned}\sigma := \sum_{k=0}^{\infty} \Bigg[&\prod_{i=1}^{k} \frac{y_{2i-1}}{x_{2i-1}} \cdot \frac{1-x_{2i-1}}{1-y_{2i-1}} + \frac{y_0}{1-y_0} \prod_{i=1}^{k} \frac{y_{2i}}{x_{2i}} \cdot \frac{1-x_2}{1-y_2} + \\ &\prod_{i=1}^{k} \frac{x_{-2i+1}}{y_{-2i+1}} \cdot \frac{1-y_{-2i+1}}{1-x_{-2i+1}} + \frac{x_0}{1-x_0} \prod_{i=1}^{k} \frac{x_{-2i}}{y_{-2i}} \cdot \frac{1-y_{-2i}}{1-x_{-2i}} \Bigg]\end{aligned} \tag{3.134}$$

converges, and $p_0 = 1/(\sigma - 1)$. Hence this is the unique coherent law with support S and ϕ as defined. Furthermore, this must be an extreme coherent law: if $(\rho_i^{(1)})_{i\in\mathbb{Z}}$ and $(\rho_i^{(2)})_{i\in\mathbb{Z}}$ are coherent laws with supports $T^{(1)}$ and $T^{(2)}$ and auxiliary functions $\phi^{(1)}$ and $\phi^{(2)}$, respectively, and (p_i) is a nondegenerate convex combination of $(\rho_i^{(1)})$ and $(\rho_i^{(2)})$, then $T^{(1)}, T^{(2)} \subseteq S$ and $(x_0, y_0) \in T^{(1)} \cap T^{(2)}$. Then by Lemma 3.6.1, $\phi_0^{(1)} = \phi_0^{(2)} = 0$ which forces $(x_{-1}, y_{-1}), (x_1, y_1) \in T^{(1)} \cap T^{(2)}$. Applying this argument inductively shows that $T^{(1)} = T^{(2)} = S$ and hence $(\rho_i^{(1)}) = (\rho_i^{(2)}) = (p_i)$.

Thus there exist extreme coherent laws with countably infinite support. Note that this fact was also independently discovered in [2], using a slightly different but still "well-behaved" configuration of points than the one in our example. Whether or not there are any non-atomic extreme coherent laws remains an open problem. However, the following result was proved in [2]:

Theorem 3.6.4. *[2] Let $n \geq 2$. Every extreme coherent distribution μ on $[0,1]^n$ is singular with respect to Lebesgue measure on $[0,1]^n$, i.e. μ is supported on a Lebesgue null set.*

Remarks. For other configurations of infinitely many points in $[0,1]^2$ other than the particularly friendly one specified for the preceding example, it may potentially be the case that the series in (3.134) diverges, in which case there is no extreme coherent law on that support with the prescribed ϕ. Also, it may be tempting to think that by constructing a sequence of extreme coherent laws with finite nested supports, that in the limit one would automatically get an extreme coherent law with infinite support. However, it is not true in general that the set of extreme points of a compact convex set is closed. Consider the following example from [70]: in $\mathbb{R}^3$, $(0,1,0)$ is a limit of extreme points but is not an extreme point of the closed convex hull of $\{(x,y,0) : x^2+y^2=1\} \cup \{(0,1,\pm 1)\}$.

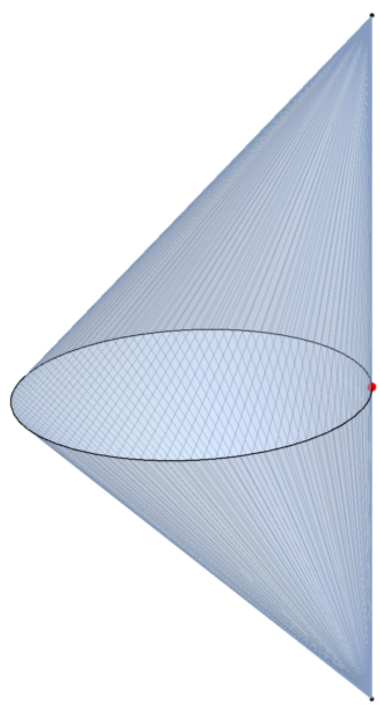


Figure 3.1: Example from [70]. The extreme points of this convex body are drawn in black. The red point is not an extreme point but is a limit point of the set of extreme points.

3.7 Coherent copulas

An n-dimensional *copula* is a probability measure on $[0,1]^n$ whose univariate marginals are uniform on $[0,1]$. The theory of copulas has applications in quantitative finance, statistics, engineering, and many other disciplines [7], [57], [52]. It is well-known and easy to see that the set of n-dimensional copulas is convex and weakly compact; see e.g. [21]. Thus it is natural to consider the problem of describing the extreme points of this set. This topic was recently explored by Ghosh and Bhandari [32] who gave some sufficient conditions for copulas to be extreme. For the case $n = 2$, two-dimensional copulas are also known as *doubly stochastic measures*, which have been extensively studied in the literature. Douglas [18] and Lindenstrauss [53] give a characterization of the extreme points of the set of doubly stochastic measures, but as remarked in [69] and [11], this characterization has limited value in identifying concrete examples.

In this section, we discuss a couple of examples of doubly stochastic measures in connection to our study of coherent distributions. Of course, the intersection of these two compact convex sets is itself a compact convex set. The simplest example of a doubly stochastic measure is uniform (Lebesgue) measure λ on $[0,1]^2$, corresponding to independent (X,Y). In this case, as remarked in [2], (X,Y) is coherent; taking

$$A := \{Y > 1 - X\}, \tag{3.135}$$

it is readily checked that

$$\mathbb{P}(A \mid X = x) = x, \qquad \mathbb{P}(A \mid Y = y) = y. \tag{3.136}$$

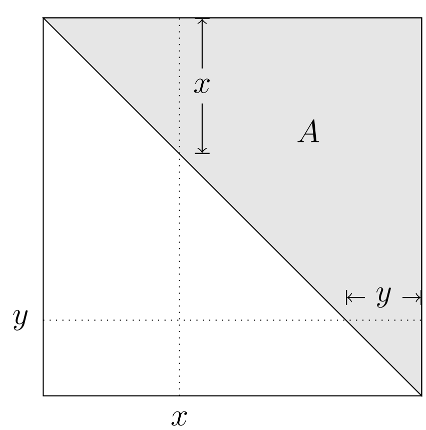


Figure 3.2: Uniform measure on $[0,1]^2$. Taking A to be the event $\{Y > 1-X\}$, it is readily seen from the figure that the definition of coherence is satisfied.

Lindenstrauss [53] proved that every extreme doubly stochastic measure is singular with respect to Lebesgue measure, and Theorem 3.6.4, which is derived from Lindenstrauss's result, asserts the same for extreme coherent distributions. Thus λ is not an extreme point of either set. However, it still appears to be an interesting problem to formally, or even informally, express λ as a mixture of extreme coherent distributions, or doubly stochastic measures.

The next example is derived from a general result in [20] concerning the existence of a stochastically minimal minimum $M := \min(X_1, \ldots, X_n)$ over coherent families $(X_1, \ldots, X_n)$ with given marginals. Define $f_0, f_1 : [0,1] \to [0,1]$ by

$$f_0(x) = 1 - \sqrt{1-(1-x)^2}, \tag{3.137}$$

$$f_1(x) = \sqrt{1-x^2}. \tag{3.138}$$

Let Γ_0 and Γ_1 denote the graphs of f_0 and f_1, respectively; that is,

$$\Gamma_i := \{(x, f_i(x)) : 0 \le x \le 1\} \qquad i = 0, 1. \tag{3.139}$$

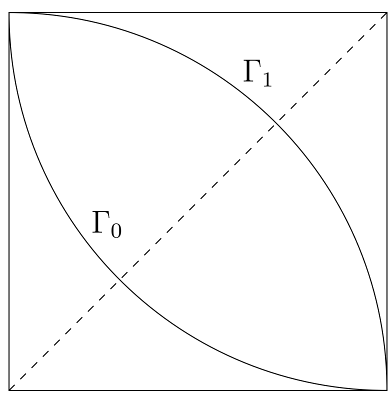


Figure 3.3: The graphs of f_0 and f_1. The union of these two circular arcs are the support of the doubly stochastic measure μ given by (3.140) and (3.141).

Consider the probability measure μ with support $\Gamma_0 \cup \Gamma_1$, defined by

$$\mu(\{(t, f_0(t)) : 0 \le t \le x\}) := \int_0^x (1-t)\ dt = x - \frac{1}{2}x^2 \tag{3.140}$$

$$\mu(\{(t, f_1(t)) : 0 \le t \le x\}) := \int_0^x t\ dt = \frac{1}{2}x^2. \tag{3.141}$$

It is immediate from (3.140) and (3.141) that the first marginal is uniform. Since f_0 and f_1 are involutions, we have

$$\mu(\{(f_0(s), s) : 0 \leq s \leq y\}) = \mu(\{(s, f_0(s)) : f_0(y) \leq s \leq 1\}) = \int_{f_0(y)}^{1} (1-s)\; ds = y - \frac{1}{2}y^2, \tag{3.142}$$

$$\mu(\{(f_1(s), s) : 0 \leq s \leq y\}) = \mu(\{(s, f_1(s)) : f_1(y) \leq s \leq 1\}) = \int_{f_1(y)}^{1} s\; ds = \frac{1}{2}y^2., \tag{3.143}$$

so we have the symmetry in distribution

$$(X, Y) \overset{d}{=} (Y, X). \tag{3.144}$$

Hence μ is a doubly stochastic measure which is supported on the graphs of two functions. The following theorem from [68] implies that μ is an extreme point of the set of doubly stochastic measures.

Theorem 3.7.1. *[68] Let $f, g : [0,1] \to [0,1]$ such that $f \leq g$. If f or g is one-to-one, then there exists at most one doubly stochastic measure supported on the union of the graphs of f and g.*

Next, for the same example, suppose (X, Y) has law μ, and let $A := \{(X, Y) \in \Gamma_1\}$. Then by (3.140) and (3.141),

$$\mathbb{P}(A \mid X = x) = \frac{x}{(1-x) + x} = x \tag{3.145}$$

and similarly $\mathbb{P}(A | Y = y) = y$, so μ is coherent. The natural question then is whether or not μ, which has uncountable support, is an extreme point of the set of coherent distributions. In principle, it appears that it may be possible to represent μ as a mixture of the "zigzag" extreme coherent laws from Section 3.6 with countable support indexed by the integers; but we do not have a definitive answer.

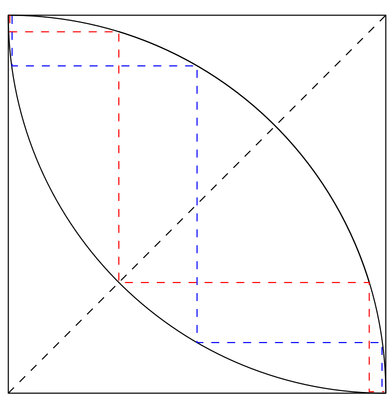


Figure 3.4: The "zigzag" extreme coherent laws supported on the union of the two circular arcs. The blue and red dotted lines identify two infinite traceable sets which are each the support of an extreme coherent law.

Chapter 4

Polynomial probability densities

A *polynomial probability density* on the unit interval $[0, 1]$ is a polynomial

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n \tag{4.1}$$

satisfying

(i) $p(x) \geq 0$ for $0 \leq x \leq 1$,

(ii) $\int_0^1 p(x)dx = 1$.

The study of polynomials that are positive or nonnegative on an interval fall into a larger framework of sum-of-squares (SOS) representation theorems for polynomials satisfying various positivity conditions. Such results are known as *positivstellensatz* and are of great interest in algebraic geometry with important applications in theoretical computer science and convex optimization; see e.g. [58]. Karlin and Shapley [44] studied the set of polynomial probability densities from a geometric perspective in connection to moment spaces. They proved a SOS-type representation theorem which refines a classical result due to Fekete on nonnegative polynomials on an interval. See [65] which contains an excellent survey with many references.

From a more combinatorial and probabilistic perspective, it is natural to consider the *Bernstein basis polynomials*, which are inherently nonnegative on $[0, 1]$ and when normalized are densities of the beta family of probability distributions, which has many applications in probability and statistics. For example, the beta distribution is essential in Bayesian inference as a conjugate prior for many discrete distributions [47]. In density estimation problems and for smoothing it is natural to consider models of polynomial densities, in particular positive mixtures of beta densities, which has been studied by various authors; see e.g. the works of Petrone [59], Guan [35],[34],[36], and Vitale [75].

4.1 Preliminaries

For $n \geq 0$, let $\mathcal{P}_n$ denote the set of real univariate polynomials of degree at most n, and let $\mathcal{P} = \bigcup_{n=0}^{\infty} \mathcal{P}_n$. Define

$$\mathcal{P}_n^+ = \mathcal{P}_n^+([0,1]) := \{p(x) \in \mathcal{P}_n : p(x) \geq 0 \text{ for } x \in [0,1]\}, \qquad \mathcal{P}^+ := \bigcup_{n=0}^{\infty} \mathcal{P}_n^+ \tag{4.2}$$

$$\mathcal{P}_n^1 = \mathcal{P}_n^1([0,1]) := \{p(x) \in \mathcal{P}_n : \int_0^1 p(x)dx = 1\}, \qquad \mathcal{P}^1 := \bigcup_{n=0}^{\infty} \mathcal{P}_n^1. \tag{4.3}$$

Note that geometrically, $\mathcal{P}_n^+$ is a convex cone in $\mathcal{P}_n$, and $\mathcal{P}_n^1$ is the affine hyperplane in $\mathcal{P}_n$ of polynomials $p(x) = a_0 + a_1 x + \ldots + a_n x^n$ satisfying the integral condition

$$\sum_{i=0}^{n} \frac{a_i}{i+1} = 1. \tag{4.4}$$

Let

$$\mathcal{D}_n := \mathcal{P}_n^+ \cap \mathcal{P}_n^1, \qquad \mathcal{D} := \bigcup_{n=0}^{\infty} \mathcal{D}_n = \mathcal{P}^+ \cap \mathcal{P}^1 \tag{4.5}$$

denote sets of polynomials which we shall refer to without ambiguity as *polynomial probability densities.*

Proposition 4.1.1. *The set $\mathcal{D}_n$, identified as a subset of $\mathbb{R}^{n+1}$ according to*

$$a_0 + a_1 x + \ldots + a_n x^n \longleftrightarrow (a_0, a_1 \ldots, a_n) \in \mathbb{R}^{n+1}, \tag{4.6}$$

is a compact convex set in $\mathbb{R}^{n+1}$.

Proof. Here $\mathcal{D}_n$ corresponds to the set

$$S := \left\{(a_0, a_1, \ldots, a_n) : \sum_{i=0}^{n} a_i x^i \geq 0 \text{ for all } x \in [0,1] \text{ and } \sum_{i=0}^{n} \frac{a_i}{i+1} = 1\right\} \subseteq \mathbb{R}^{n+1}. \tag{4.7}$$

It is readily checked that S is convex and closed. For compactness, let $0 \leq x_0 < \ldots < x_n \leq 1$. Then S is a closed subset of

$$\left\{(a_0, a_1, \ldots, a_n) : \sum_{i=0}^{n} a_i x_j^i \geq 0 \text{ for } j = 0, 1, \ldots, n \text{ and } \sum_{i=0}^{n} \frac{a_i}{i+1} = 1\right\} \tag{4.8}$$

which is an n-dimensional polytope in the affine hyperplane $\sum_{i=0}^{n} \frac{a_i}{i+1} = 1$ whose facets are defined by the $n+1$ supporting hyperplanes through the origin with linearly independent normal vectors due to the Vandermonde system (see e.g. [56].) □

Note that since compactness and convexity are invariant set properties under linear maps, Proposition 4.1.1 implies that $\mathcal{D}_n$ is a compact convex set when embedded in $\mathbb{R}^{n+1}$ according to any basis of $\mathcal{P}_n$. In this setting, it is natural and convenient to work with the *Bernstein basis polynomials.*

The rest of the chapter is organized as follows. Section 4.2 recalls a number of known results regarding the Bernstein basis polynomials, including the definition of the Lorentz degree of a polynomial. Section 4.3 discusses the extreme points of $\mathcal{D}_n$ and presents a representation theorem due to Karlin and Shapley [44]. In Section 4.4 we map out the geometry of $\mathcal{D}_2$, specifically in terms of the level sets defined by the Lorentz degree. In Section 4.5 we consider uniform sampling from $\mathcal{D}_n$ and prove some properties of the Lorentz degree of a random polynomial probability density sampled from $\mathcal{D}_2$. In Section 4.6 we derive a formula for the upper envelope of $\mathcal{D}_n$ using orthogonal polynomials.

4.2 Bernstein basis polynomials

Let

$$U_{(1)} < U_{(2)} < \ldots < U_{(n+1)} \tag{4.9}$$

denote the order statistics obtained by ranking $n+1$ i.i.d. uniform $[0,1]$ random variables $U_1, \ldots, U_{n+1}$. It is well-known that for $0 \le k \le n$, $U_{(k+1)}$ has the beta$(k+1,\, n-k+1)$ distribution with probability density function

$$\frac{d}{dx}\mathbb{P}(U_{(k+1)} \le x) = (n+1)\binom{n}{k}x^k(1-x)^{n-k}. \tag{4.10}$$

We refer to the polynomials

$$b_{n,k}(x) := \binom{n}{k}x^k(1-x)^{n-k} \tag{4.11}$$

as the *Bernstein basis polynomials* of degree n, and

$$\tilde{b}_{n,k}(x) := (n+1)b_{n,k}(x) \tag{4.12}$$

in (4.10) as normalized Bernstein basis polynomials. It is well-known and easy to see that $\{b_{n,k}(x) : 0 \le k \le n\}$ forms a basis for the real vector space $\mathcal{P}_n$ for each $n \ge 0$. Indeed, we can expand any monomial x^j for $0 \leq j \leq n$ in the Bernstein basis for $\mathcal{P}_n$ as

$$\begin{aligned} x^j &= x^j(x + (1-x))^{n-j} && (4.13)\\ &= x^j \sum_{k=0}^{n-j} \binom{n-j}{k} x^k (1-x)^{n-j-k} && (4.14)\\ &= \sum_{k=0}^{n-j} x^{j+k}(1-x)^{n-j-k} && (4.15)\\ &= \sum_{k=j}^{n} \binom{n-j}{k-j} x^k (1-x)^{n-k} && (4.16)\\ &= \sum_{k=j}^{n} \frac{\binom{n-j}{k-j}}{\binom{n}{k}} b_{n,k}(x). && (4.17) \end{aligned}$$

Then for a polynomial

$$f(x) = a_0 + a_1 x + \ldots + a_d x^d \tag{4.18}$$

with degree $d \leq n$, we have

$$\begin{aligned} f(x) &= \sum_{j=0}^{d} a_j \sum_{k=j}^{n} \frac{\binom{n-j}{k-j}}{\binom{n}{k}} b_{n,k}(x) && (4.19)\\ &= \sum_{k=0}^{n} \left[\sum_{j=0}^{k \wedge d} a_j \frac{\binom{n-j}{k-j}}{\binom{n}{k}} \right] b_{n,k}(x). && (4.20) \end{aligned}$$

In particular, taking the constant polynomial $f(x) = 1$, this gives the well-known identity

$$\sum_{k=0}^{n} b_{n,k}(x) = 1. \tag{4.21}$$

Recall from (4.12) that $\tilde{b}_{n,k}(x)$ are the normalized Bernstein basis polynomials.

Proposition 4.2.1. *For $n \geq 0$, every $p(x) \in \mathcal{D}_n$ has a unique representation*

$$p(x) = \sum_{k=0}^{n} c_{n,k} \tilde{b}_{n,k}(x) \tag{4.22}$$

with real coefficients $c_{n,k}$, $0 \leq k \leq n$, subject to the constraint

$$\sum_{k=0}^{n} c_{n,k} = 1. \tag{4.23}$$

If each $c_{n,k}$ in (4.22) is nonnegative, then $p(x)$ is the probability density on $[0,1]$ of a random order statistic of $n+1$ i.i.d. uniform $[0,1]$ random variables $U_1,\dots,U_{n+1}$. Explicitly, $p(x)$ corresponds to the distribution of the I^{th} order statistic where I is a random variable independent of $U_1,\dots,U_{n+1}$ with

$$\mathbb{P}(I=i)=c_{n,i-1} \qquad (1\le i\le n+1). \tag{4.24}$$

In general, the coefficients in (4.22) need not all be nonnegative. For example, for $n=2$, $12(x-\frac{1}{2})^2\in\mathcal{D}_2$ has the representation

$$12(x-\tfrac{1}{2})^2=3(1-x)^2-6x(1-x)+3x^2=\tilde{b}_{2,0}-\tilde{b}_{2,1}+\tilde{b}_{2,2}. \tag{4.25}$$

In fact, a polynomial probability density which takes the value 0 somewhere in the open interval $(0,1)$ cannot admit the representation (4.22) with all coefficients nonnegative, since each $\tilde{b}_{n,k}(x)$ is positive on $(0,1)$.

We can approximate the coefficients

$$c_{n,k}:=\sum_{j=0}^{k\wedge d}a_j\frac{\binom{n-j}{k-j}}{\binom{n}{k}}=\sum_{j=0}^{d}a_j\frac{\binom{n-j}{k-j}}{\binom{n}{k}} \tag{4.26}$$

from (4.20) using

$$a_{n,k}\approx f(k/n)=a_0+a_1\Big(\frac{k}{n}\Big)+a_2\Big(\frac{k}{n}\Big)^2+\ldots+a_d\Big(\frac{k}{n}\Big)^d. \tag{4.27}$$

following Bernstein's approximation theorem:

Theorem 4.2.2 (Bernstein)**.** *[5] Let g be a continuous function on the interval $[0,1]$. For $n\ge 1$, define the polynomial*

$$B_n(g)(x)=\sum_{k=0}^{n}g(k/n)b_{n,k}(x). \tag{4.28}$$

Then $B_n(g)\to g$ uniformly on $[0,1]$.

For the approximation (4.27), we have

$$|a_{n,k}-f(k/n)|=\left|\sum_{j=2}^{d}a_j\left[\Big(\frac{k}{n}\Big)^j-\frac{k(k-1)\cdots(k-j+1)}{n(n-1)\cdots(n-j+1)}\right]\right| \tag{4.29}$$

$$\le\sum_{j=2}^{d}|a_j|\left[\Big(\frac{k}{n}\Big)^j-\frac{k(k-1)\cdots(k-j+1)}{n(n-1)\cdots(n-j+1)}\right]. \tag{4.30}$$

It is an elementary exercise to show for $2 \leq j \leq d$ that

$$\sup_{0 \leq k \leq n} \left[\left(\frac{k}{n}\right)^j - \frac{k(k-1)\cdots(k-j+1)}{n(n-1)\cdots(n-j+1)} \right] \to 0 \qquad \text{as } n \to \infty. \tag{4.31}$$

Note that the expression over which the supremum is taken in (4.31) has the probabilistic interpretation as the difference between drawing with and without replacement for the probability of drawing only "good" objects in j draws from a box containing k good objects among n total objects. See [67] for explicit bounds on this expression. Consequently,

$$\sup_{0 \leq k \leq n} |a_{n,k} - f(k/n)| \to 0 \qquad \text{as } n \to \infty. \tag{4.32}$$

Proposition 4.2.3. *[5] Let $p(x) \in \mathcal{D}$. Then $p(x) > 0$ for all $0 < x < 1$ if and only if there exists a nonnegative integer n such that $p(x)$ has the expansion*

$$p(x) = \sum_{k=0}^{n} c_{n,k} \tilde{b}_{n,k}(x) \tag{4.33}$$

with $c_{n,k} \geq 0$ for all $0 \leq k \leq n$.

Proof. The "if" direction is obvious. For the other direction, if $p(x) > 0$ for $0 < x < 1$, then $p(x) = x^i(1-x)^j q(x)$ for some $i, j \geq 0$ and a polynomial $q(x)$ which is positive for $0 \leq x \leq 1$. Let $u := \min_{0 \leq x \leq 1} q(x) > 0$. By (4.32), there exists m such that $q(x) = \sum_{k=0}^{m} a_{m,k} b_{m,k}(x)$ with

$$\sup_{0 \leq k \leq m} |a_{m,k} - q(k/m)| < u. \tag{4.34}$$

Since $q(x) \geq u$ for all $0 \leq x \leq 1$, it follows that $a_{m,k} \geq 0$ for $0 \leq k \leq m$. Then for $n = m + i + j$, $p(x) = x^i(1-x)^j q(x)$ has the representation

$$\sum_{k=0}^{n} c_{n,k} \tilde{b}_{n,k}(x) \tag{4.35}$$

with $c_{n,k} \geq 0$ for $0 \leq k \leq n$. □

Note that if $p(x)$ has the representation (4.33) for a nonnegative integer n, it does for all $n' > n$ as well; indeed, we have the identity

$$\begin{aligned}
\tilde{b}_{n,k}(x) &= \tilde{b}_{n,k}(x)(x + (1-x)) &(4.36)\\
&= (n+1)\binom{n}{k} x^{k+1}(1-x)^{n-k} + (n+1)\binom{n}{k} x^k (1-x)^{n-k+1} &(4.37)\\
&= (k+1)\binom{n+1}{k+1} x^{k+1}(1-x)^{n-k} + (n+1-k)\binom{n+1}{k} x^k (1-x)^{n-k+1} &(4.38)\\
&= \tfrac{k+1}{n+2}\tilde{b}_{n+1,k+1}(x) + \tfrac{n+1-k}{n+2}\tilde{b}_{n+1,k}(x) &(4.39)
\end{aligned}$$

and subsequently

$$
\begin{aligned}
p(x) &= \sum_{k=0}^{n} c_{n,k}\tilde{b}_{n,k}(x) && (4.40)\\
&= \sum_{k=0}^{n} c_{n,k}\left[\tfrac{k+1}{n+2}\tilde{b}_{n+1,k+1}(x) + \tfrac{n+1-k}{n+2}\tilde{b}_{n+1,k}(x)\right] && (4.41)\\
&= \tfrac{n+1}{n+2}c_{n,0}\tilde{b}_{n+1,0}(x) + \sum_{k=1}^{n}\left[\tfrac{k}{n+2}c_{n,k+1} + \tfrac{n+1-k}{n+2}c_{n,k}\right]\tilde{b}_{n+1,k}(x) + \tfrac{n+1}{n+2}c_{n,n}\tilde{b}_{n+1,n+1}(x) && (4.42)\\
&= \sum_{k=0}^{n+1} c_{n+1,k}\tilde{b}_{n+1,k}(x) && (4.43)
\end{aligned}
$$

where the coefficients $c_{n+1,k}$ are all nonnegative if the $c_{n,k}$ are all nonnegative. The claim follows by induction.

The smallest nonnegative integer n for which a polynomial probability density or more generally $p(x) \in \mathcal{P}^+$ has the representation (4.33) is known as the *Lorentz degree* of $p(x)$, which we denote by $\delta(p(x))$ [55],[25]. If no such n exists, i.e. $p(x) = 0$ for some $0 < x < 1$, we define $\delta(p(x)) = \infty$. It is clear that for $p(x) \in \mathcal{P}^+$ that

$$\delta(p(x)) \geq \deg p(x) \tag{4.44}$$

but in general $\delta(p(x))$ can be arbitrarily large over polynomial probability densities with any fixed ordinary degree of at least 2; see Section 4.4 for this discussion for the quadratic case. See e.g. [24],[26],[25] for more on the Lorentz degree of polynomials. Let

$$\mathcal{L}_{d,n} := \{p(x) \in \mathcal{D}_d : \delta(p(x)) \leq n\}, \tag{4.45}$$

i.e. the set of polynomial densities with ordinary degree at most d and Lorentz degree at most n.

4.3 Extreme points

In this section, we consider the set of extreme points of the compact convex set $\mathcal{D}_n$ for $n \geq 1$. An immediate observation is that every extreme point of this set must have a zero in $[0, 1]$. Indeed, suppose $p(x) \in \mathcal{D}_n$ with $p(x) > 0$ for $0 \leq x \leq 1$ and let $\varepsilon := \min_{0\leq x\leq 1} p(x) > 0$ by compactness of $[0, 1]$. Then

$$p(x) = \tfrac{1}{2}[p(x) + \varepsilon(2x - 1)] + \tfrac{1}{2}[p(x) - \varepsilon(2x - 1)] \tag{4.46}$$

is the midpoint of two distinct elements of $\mathcal{D}_n$, where the perturbation $\varepsilon(2x - 1)$ is chosen so that $p(x) \pm \varepsilon(2x - 1)$ remains nonnegative with integral 1. Hence $p(x)$ is not extreme.

For $n = 1$, it is obvious that $\mathcal{D}_1$ has exactly two extreme points: $2x$ and $2(1-x)$. For $n = 2$, if $p(x) \in \mathcal{D}_2$ has a zero $r \in (0,1)$ then $p(x)$ must be extreme, because it is the unique polynomial in $\mathcal{D}_2$ with a zero at r, given by

$$C(x-r)^2 \tag{4.47}$$

where C is the appropriate normalizing constant, hence it cannot be a convex combination of two distinct polynomials in $\mathcal{D}_2$. If $p(0) = 0$ then $p(x)$ has either the form $Cx(x-a)$ for $r \leq 0$ or $Cx(a-x)$ for $a \geq 1$.

- If $p(x) = Cx(x-a)$ with $a < 0$, then $p(x)$ is not extreme; for instance, we have

  $$Cx(x-a) = \tfrac{C-3}{D-3}Dx(x-2a) + \tfrac{D-C}{D-3}3x^2 \tag{4.48}$$

  where $C = (\frac{1}{3} - \frac{a}{2})^{-1}$ and $D = (\frac{1}{3} - a)^{-1}$.

- If $p(x) = Cx(a-x)$ with $a > 1$, then $p(x)$ is not extreme; for instance, we have

  $$Cx(a-x) = \tfrac{C-6}{D-6}Dx(2a-x) + \tfrac{D-C}{D-6}6x(1-x) \tag{4.49}$$

  where $C = (\frac{a}{2} - \frac{1}{3})^{-1}$ and $D = (a - \frac{1}{3})^{-1}$.

- The polynomial $p(x) = 3x^2$ is extreme, because it is the unique polynomial in $\mathcal{D}_2$ satisfying $p(0) = p'(0) = 0$, and every other polynomial $q(x) \in \mathcal{D}_2$ with $q(0) = 0$ must have $q'(0) > 0$, so $p(x)$ cannot be a convex combination of two distinct polynomials in $\mathcal{D}_2$.

- The polynomial $p(x) = 6x(1-x)$ is extreme, because it is the unique polynomial in $\mathcal{D}_2$ satisfying $p(0) = p(1) = 0$, so it cannot be a convex combination of two distinct polynomials in $\mathcal{D}_2$.

By symmetry, the polynomial $3(1-x)^2$ is also extreme in $\mathcal{P}_2$. This completes the list of extreme points of $\mathcal{D}_2$: $C(r)(x-r)^2$ for $0 \leq r \leq 1$ where $C(r) := (\int_0^1 (x-r)^2\, dx)^{-1}$, and $6x(1-x)$.

The problem of classifying the extreme points of $\mathcal{D}_n$ for general n was solved by Karlin and Shapley [44] who gave a precise representation theorem for polynomial densities.

Theorem 4.3.1. *[44] For $q(x) \in \mathcal{P}^+$ and $m \geq 0$, define the set*

$$T_m(q(x)) := \left\{p(x) \in \mathcal{D} : p(x) = Cq(x)\prod_{i=1}^{m}(x-r_i)^2,\ 0 \leq r_1 \leq \cdots \leq r_m \leq 1\right\}. \tag{4.50}$$

Let $n \geq 1$. Then

$$\operatorname{ext}(\mathcal{D}_n) = \begin{cases} T_m(1) \cup T_{m-1}(x(1-x)) & \text{if } n = 2m, \\ T_m(x) \cup T_m(1-x) & \text{if } n = 2m+1. \end{cases} \tag{4.51}$$

Furthermore, every nonextreme $p(x) \in \mathcal{D}_n$ has a unique representation as a convex combination of a pair of extreme polynomials with interlacing roots; more precisely, $p(x)$ has a unique representation of the form

- *for $n = 2m$,*

$$p(x) = \alpha C_1 \prod_{i=1}^{m} (x - r_{2i-1})^2 + (1-\alpha) C_2 x(1-x) \prod_{i=1}^{m-1} (x - r_{2i})^2 \tag{4.52}$$

  where $0 \leq r_1 \leq \cdots \leq r_{2m-1} \leq 1$;

- *for $n = 2m+1$,*

$$p(x) = \alpha C_1 x \prod_{i=1}^{m} (x - r_{2i})^2 + (1-\alpha) C_2 (1-x) \prod_{i=1}^{m} (x - r_{2i-1})^2 \tag{4.53}$$

  where $0 \leq r_1 \leq \cdots \leq r_{2m} \leq 1$.

Moreover, $p(x)$ is an interior point of $\mathcal{D}_n$ if and only if all of the inequalities for the interlacing roots are strict.

4.4 Geometry of quadratic densities

The normalized Bernstein basis polynomials of degree 2 are

$$\tilde{b}_{2,0} = 3(1-x)^2, \tag{4.54}$$

$$\tilde{b}_{2,1} = 6x(1-x), \tag{4.55}$$

$$\tilde{b}_{2,2} = 3x^2. \tag{4.56}$$

Consider the set $\mathcal{D}_2$ represented as a subset $S \subseteq \mathbb{R}^3$ according to the correspondence

$$p(x) = a(3(1-x)^2) + b(6x(1-x)) + c(3x^2) \longleftrightarrow (a, b, c) \in \mathbb{R}^3, \tag{4.57}$$

i.e. the coordinate representation under the normalized Bernstein polynomial basis of degree 2. It is natural to work in this basis because the $\tilde{b}_{2,k}$ for $k = 0, 1, 2$ are elements of $\mathcal{D}_2$ and also because of the reversal symmetry

$$(a, b, c) \in S \iff (c, b, a) \in S. \tag{4.58}$$

Following the discussion in Section 4.3, consider polynomial densities of the form

$$C(r)(x-r)^2 = C(r)[(1-r)x - r(1-x)]^2 \tag{4.59}$$

$$= C(r)\left[\tfrac{r^2}{3} 3(1-x)^2 - \tfrac{r(1-r)}{3} 6x(1-x) + \tfrac{(1-r)^2}{3} 3x^2\right] \tag{4.60}$$

for $r \in \mathbb{R}$ The coefficients $(a, b, c) := (\frac{C(r)r^2}{3}, -\frac{C(r)r(1-r)}{3}, \frac{C(r)(1-r)^2}{3})$ satisfy the conditions $b \leq 0$ and

$$b^2 = ac, \tag{4.61}$$

which is equivalent to the discriminant condition in the standard monomial basis for quadratic polynomials with a double root. This, in conjunction with the integral condition

$$a + b + c = 1, \tag{4.62}$$

reveals that this class of points in S lie on the sphere

$$(a-1)^2 + b^2 + (c-1)^2 = 1. \tag{4.63}$$

These points lie on the circle given by the intersection of the sphere (4.63) and the affine hyperplane (4.62), and the extreme points of S corresponding to polynomial densities of the form $C(r)(x-r)^2$ for $0 \leq r \leq 1$ lie on the intersection of this circle with the half-space $b \leq 0$. Since the part of the circle not corresponding to the extreme points lie in the simplex $\{(a, b, c) : a, b, c \geq 0, a + b + c \leq 1\}$, S is the convex hull of the point $(0, 1, 0)$ and the circle defined by (4.62) and (4.63).

It is convenient here to view S as a subset $\tilde{S} \subseteq \mathbb{R}^2$ via the projection

$$(a, b, c) \mapsto (a, c) =: (x, y) \in \mathbb{R}^2. \tag{4.64}$$

Then the circle in $\mathbb{R}^3$ defined by (4.62) and (4.63) projects to the ellipse satisfying the equation

$$(x-1)^2 + (1-x-y)^2 + (y-1)^2 = 1 \tag{4.65}$$

which can be rearranged as

$$\frac{(x+y-\frac{4}{3})^2}{\frac{4}{9}} + \frac{(x-y)^2}{\frac{4}{3}} = 1. \tag{4.66}$$

The ellipse is centered at $(\frac{2}{3}, \frac{2}{3})$, corresponding to the polynomial probability density $6(x - \frac{1}{2})^2 + \frac{1}{2}$; passes through the points $(1, 0)$ and $(0, 1)$, corresponding to the polynomial densities $3(1-x)^2$ and $3x^2$; has major axis of length $\frac{2\sqrt{6}}{3}$ along the line $x + y = \frac{4}{3}$; and has minor axis connecting $(\frac{1}{3}, \frac{1}{3})$ and $(1, 1)$, corresponding to the polynomial densities 1 and $12(x - \frac{1}{2})^2$. See Figure 4.1.

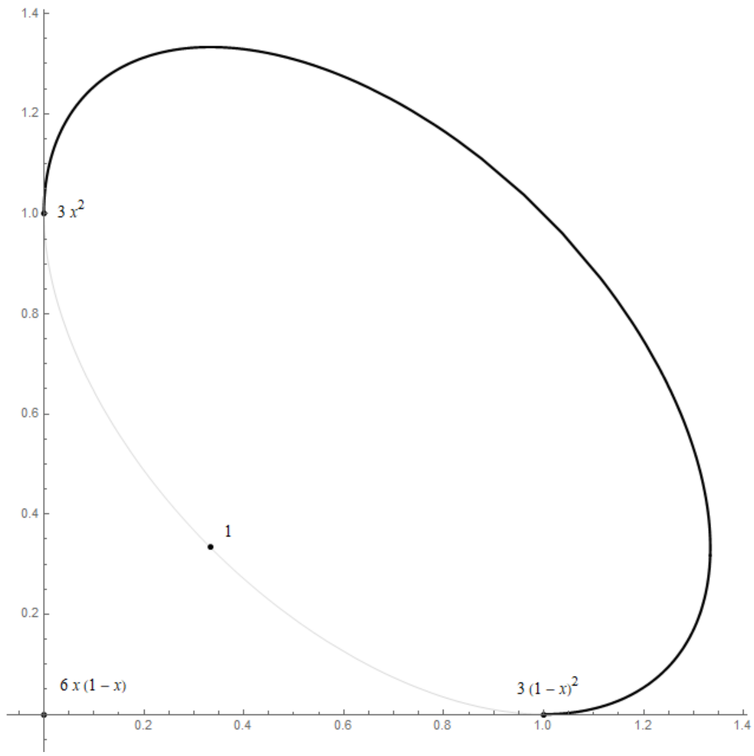


Figure 4.1: Planar representation of the extreme points of $\mathcal{D}_2$. This is a coordinate transformation of a graph presented in [44] from the standard monomial basis to the normalized Bernstein basis which displays a natural symmetry.

Proposition 4.4.1. *Let $n \geq 2$. The image of the set*

$$Q_n := \left\{ C_n(r)\left((x-r)^2 + \tfrac{r(1-r)}{n}\right) : r \in \mathbb{R} \right\} \subseteq \mathcal{P}_2^1 \tag{4.67}$$

where $C_n(r) := \left(\int_0^1 ((x-r)^2 + \frac{r(1-r)}{n}) dx \right)^{-1}$ under the map $\pi : \mathcal{P}_2 \to \mathbb{R}^2$ given by

$$p(x) = a3(1-x)^2 + b6x(1-x) + c3x^2 \mapsto (a, c) \tag{4.68}$$

is $E_n \setminus \{(\frac{1}{3}, \frac{1}{3})\}$ where E_n is the ellipse with equation

$$\frac{(a + c - \frac{4n+6}{3n+9})^2}{\frac{4n^2}{9(n+3)^2}} + \frac{(a-c)^2}{\frac{4n^2}{3(n+3)(n-1)}} = 1. \tag{4.69}$$

Note that

- The polynomial $(x-r)^2 + \frac{r(1-r)}{n}$ is nonnegative on $[0, 1]$ if and only if $0 \leq r \leq 1$, $r < -\frac{1}{n-1}$, or $r > 1 + \frac{1}{n-1}$. Hence $C_n(r)((x-r)^2 + \frac{r(1-r)}{n})$ is not a polynomial probability density for $-\frac{1}{n-1} < r < 0$ or $1 < r < 1 + \frac{1}{n-1}$.

- For every $r \in \mathbb{R}$, $\int_0^1 ((x-r)^2 + \frac{r(1-r)}{n})dx = (r-\frac{1}{2})^2 + \frac{1}{3}\frac{n}{n-1} - \frac{1}{4} > (r-\frac{1}{2})^2 + \frac{1}{3(n-1)} > 0.$

Proof. For every $r \in \mathbb{R}$, we have

$$(x-r)^2 + \tfrac{r(1-r)}{n} = (r^2 + \tfrac{r(1-r)}{n})(1-x)^2 + \underbrace{\left(-r(1-r) + \tfrac{r(1-r)}{n}\right)}_{-\frac{n-1}{n}r(1-r)} 2x(1-x) + ((1-r)^2 + \tfrac{r(1-r)}{n})x^2 \tag{4.70}$$

so

$$C_n(r)\big((x-r)^2 + \tfrac{r(1-r)}{n}\big) = a3(1-x)^2 + b6x(1-x) + c3x^2 \tag{4.71}$$

where

$$b = -\tfrac{C_n(r)}{3}\tfrac{n-1}{n}r(1-r) \tag{4.72}$$
$$a = \tfrac{C_n(r)}{3}r^2 - \tfrac{1}{n-1}b \tag{4.73}$$
$$c = \tfrac{C_n(r)}{3}(1-r)^2 - \tfrac{1}{n-1}b \tag{4.74}$$

satisfy

$$\big(\tfrac{n-1}{n}b\big)^2 = \big(a + \tfrac{1}{n-1}b\big)\big(c + \tfrac{1}{n-1}b\big). \tag{4.75}$$

Simplifying and substituting $b = 1-a-c$ yields

$$\tfrac{n+1}{n-1}(1-a-c)^2 - \tfrac{1}{n-1}(a+c)(1-a-c) - ac = 0. \tag{4.76}$$

Using the substitution

$$s = a+c, \qquad t = a-c, \tag{4.77}$$

equation (4.76) becomes

$$\tfrac{n+1}{n-1}(1-s)^2 - \tfrac{1}{n-1}s(1-s) - \tfrac{s+t}{2}\tfrac{s-t}{2} = 0 \tag{4.78}$$
$$\tfrac{n+1}{n-1}(1-2s+s^2) - \tfrac{1}{n-1}(s-s^2) - \tfrac{1}{4}s^2 + \tfrac{1}{4}t^2 = 0 \tag{4.79}$$
$$\tfrac{3n+9}{4(n-1)}s^2 - \tfrac{2n+3}{n-1}s + \tfrac{1}{4}t^2 + \tfrac{n+1}{n-1} = 0 \tag{4.80}$$
$$\tfrac{3(n+3)}{4(n-1)}\big(s - \tfrac{2(2n+3)}{3(n+3)}\big)^2 + \tfrac{1}{4}t^2 = \tfrac{n^2}{3(n+3)(n-1)} \tag{4.81}$$
$$\tfrac{9(n+3)^2}{4n^2}\big(s - \tfrac{4n+6}{3n+9}\big)^2 + \tfrac{3(n+3)(n-1)}{4n^2}t^2 = 1 \tag{4.82}$$

as desired. The map π restricted to Q_n is continuous and injective with

$$\pi\Big(C_n(r)\big((x-r)^2 + \tfrac{r(1-r)}{n}\big)\Big) = \left(\frac{r^2 + \frac{r(1-r)}{n}}{1 - 3(1-\frac{1}{n})r + 3(1-\frac{1}{n})r^2}, \frac{(1-r)^2 + \frac{r(1-r)}{n}}{1 - 3(1-\frac{1}{n})r + 3(1-\frac{1}{n})r^2}\right), \tag{4.83}$$

so

$$\lim_{r\to\pm\infty} \pi\big((x-r)^2 + \tfrac{r(1-r)}{n}\big) = \big(\tfrac{1}{3}, \tfrac{1}{3}\big) \tag{4.84}$$

which implies that the image of Q_n under π is $E_n \setminus \{(\frac{1}{3}, \frac{1}{3})\}$. □

It turns out that there is a direct connection between the class of polynomial densities in Proposition 4.4.1 and the set $\mathcal{L}_{2,n}$ of polynomial densities of ordinary degree at most 2 and Lorentz degree at most n for $n \geq 2$.

Theorem 4.4.2. *Let $n \geq 2$. The extreme points of $\mathcal{L}_{2,n}$ are the n polynomial densities of the form*

$$p_{n,k}(x) := C_n(\tfrac{k}{n-1})\big((x - \tfrac{k}{n-1})^2 + \tfrac{1}{n}\tfrac{k}{n-1}(1 - \tfrac{k}{n-1})\big), \qquad 0 \leq k \leq n-1 \tag{4.85}$$

and $6x(1-x)$.

Proof. For

$$p(x) = c_{2,0}\tilde{b}_{2,0}(x) + c_{2,1}\tilde{b}_{2,1}(x) + c_{2,2}\tilde{b}_{2,2}(x) \in \mathcal{P}_2^1$$

we have

$$\begin{aligned}
p(x) &= [c_{2,0}3(1-x)^2 + c_{2,1}6x(1-x) + c_{2,2}3x^2](x+1-x)^{n-2} && (4.86)\\
&= [c_{2,0}3(1-x)^2 + c_{2,1}6x(1-x) + c_{2,2}3x^2]\sum_{k=0}^{n-2}\binom{n-2}{k}x^k(1-x)^{n-2-k} && (4.87)\\
&= \sum_{k=0}^{n-2}\frac{3}{n+1}c_{2,0}\frac{(n-k)(n-k-1)}{n(n-1)}\tilde{b}_{n,k}(x) + \sum_{k=1}^{n-1}\frac{6}{n+1}c_{2,1}\frac{k(n-k)}{n(n-1)}\tilde{b}_{n,k}(x) \\
&\quad + \sum_{k=2}^{n}\frac{3}{n+1}c_{2,2}\frac{k(k-1)}{n(n-1)}\tilde{b}_{n,k}(x) && (4.88)\\
&= \frac{3}{(n+1)n(n-1)}\sum_{k=0}^{n}\big[(n-k)(n-k-1)c_{2,0} + 2k(n-k)c_{2,1} + k(k-1)c_{2,2}\big]\tilde{b}_{n,k}(x), && (4.89)
\end{aligned}$$

so $p(x) \in \mathcal{L}_{2,n}$ if and only if

$$(n-k)(n-k-1)c_{2,0} + 2k(n-k)c_{2,1} + k(k-1)c_{2,2} \geq 0 \qquad \text{for } 0 \leq k \leq n. \tag{4.90}$$

Now substituting $c_{2,1} = 1 - c_{2,0} - c_{2,2}$ and simplifying, we obtain

$$(n-k)(3k-n+1)c_{2,0} + k(3(n-k)-n+1)c_{2,2} \leq 2k(n-k) \qquad \text{for } 0 \leq k \leq n. \tag{4.91}$$

These $n+1$ inequalities define an $(n+1)$-gon in $\mathbb{R}^2$ which lies inside $\tilde{S}$. The vertices $(x_{n,k}, y_{n,k})$, $k = 0, 1, \ldots, n$ are computed as

$$x_{n,k} = \frac{k(k+1)}{k(k+1) - k(n-k-1) + (n-k)(n-k-1)}, \tag{4.92}$$

$$y_{n,k} = \frac{(n-k)(n-k-1)}{k(k+1) - k(n-k-1) + (n-k)(n-k-1)} \tag{4.93}$$

for $k = 0, 1, \ldots, n-1$ where $(x_{n,k}, y_{n,k})$ is computed by intersecting the lines corresponding to (4.4) for k and $k+1$, and $(x_{n,n}, y_{n,n}) = (0, 0)$ is the intersection of the lines for $k = 0$ and $k = n$. Then for $0 \le k \le n-1$, with $r_{n,k} := \frac{k}{n-1}$, we have

$$x_{n,k} = \frac{r_{n,k}(r_{n,k} + \frac{1}{n-1})}{r_{n,k}(r_{n,k} + \frac{1}{n-1}) - r_{n,k}(1 - r_{n,k}) + (1 - r_{n,k} + \frac{1}{n-1})(1 - r_{n,k})} \tag{4.94}$$

$$= \frac{\frac{n}{n-1} r_{n,k}^2 + \frac{1}{n-1} r_{n,k}(1 - r_{n,k})}{\frac{n}{n-1} - 3r_{n,k} + 3r_{n,k}^2} \tag{4.95}$$

$$= \frac{r_{n,k}^2 - \frac{r_{n,k}(1-r_{n,k})}{n}}{1 - 3(\frac{n-1}{n})r_{n,k} + 3(\frac{n-1}{n})r_{n,k}^2} \tag{4.96}$$

and similarly

$$y_{n,k} = \frac{(1 - r_{n,k})^2 + \frac{r_{n,k}(1-r_{n,k})}{n}}{1 - 3(\frac{n-1}{n})r_{n,k} + 3(\frac{n-1}{n})r_{n,k}^2}. \tag{4.97}$$

By (4.83), we see that the point $(x_{n,k}, y_{n,k})$ corresponds to the polynomial probability density

$$p_{n,k}(x) = C_n(r_{n,k})\big((x - r_{n,k})^2 + \tfrac{1}{n} r_{n,k}(1 - r_{n,k})\big). \tag{4.98}$$

□

It can be seen with a bit of effort that $(x_{n,k}, y_{n,k})$ as in (4.92) and (4.93) satisfies equality in (4.91) with $(n, k) \mapsto (n+1, k+1)$. This implies that the extreme points of $\mathcal{L}_{2,n}$ lie on the boundary of $\mathcal{L}_{2,n+1}$, which is an unexpected and nontrivial feature of the geometry of these sets; see Figure 4.2. Specifically,

$$p_{n,k}(x) = \alpha_{n,k} p_{n+1,k}(x) + (1 - \alpha_{n,k}) p_{n+1,k+1}(x) \tag{4.99}$$

where

$$\alpha_{n,k} := \frac{(n-k-1)C_{n+1}(\frac{k+1}{n})}{kC_{n+1}(\frac{k}{n}) + (n-k-1)C_{n+1}(\frac{k+1}{n})}. \tag{4.100}$$

Also, the lines in $\tilde{S}$ that pass through the point $(\frac{1}{3}, \frac{1}{3})$ (which corresponds to the constant polynomial 1) have a natural correspondence in terms of polynomial densities. It is easy to check that the line through $(\frac{1}{3}, \frac{1}{3})$ with slope -1 corresponds to $\mathcal{D}_1$, the polynomial densities of degree at most 1. If $p(x), q(x), 1 \in \mathcal{D}_2$ are collinear and $p(x)$ has the form $s(x-r)^2 + t$, then without loss of generality either

$$1 = \alpha p(x) + (1-\alpha) q(x) \qquad \text{or} \qquad q(x) = \alpha p(x) + (1-\alpha)1, \tag{4.101}$$

for some $\alpha \in [0, 1]$. In either case, $q(x)$ has the form $s'(x-r)^2 + t'$. In other words, the lines in $\tilde{S}$ passing through $(\frac{1}{3}, \frac{1}{3})$ with slope not equal to -1 can be parameterized by $r \in \mathbb{R}$,

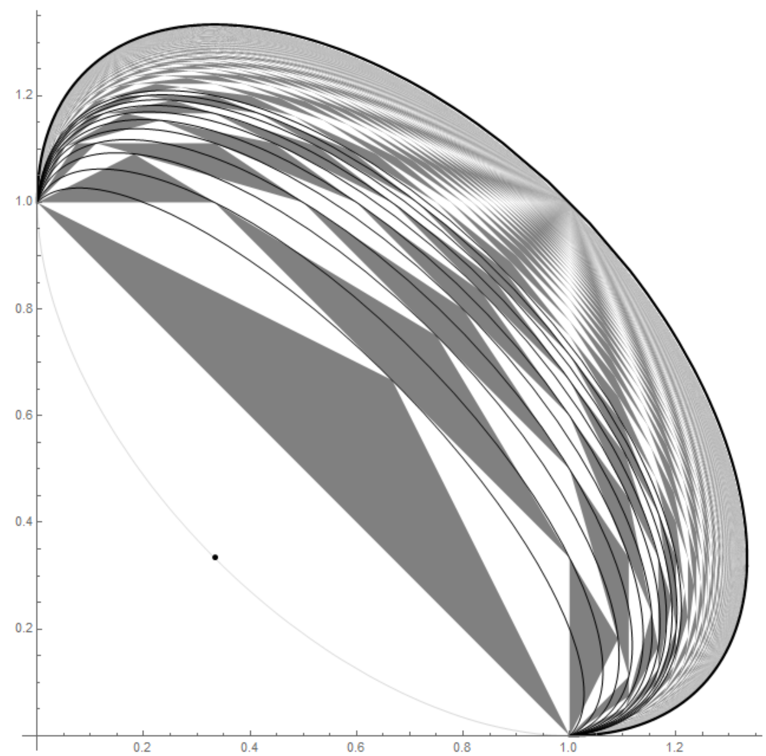


Figure 4.2: The geometry of the nested sets $(\mathcal{L}_{2,n})_{n\geq 2}$ in $\mathcal{D}_2$. For each n, $\mathcal{L}_{2,n}$ corresponds to an $(n+1)$-gon, n of whose vertices lie on the ellipse E_n defined in Proposition 4.4.1. Notice that the vertices of $\mathcal{L}_{2,n}$ lie on the boundary of $\mathcal{L}_{2,n+1}$.

corresponding to polynomial probability densities with functional critical point r, i.e. those of the form $s(x-r)^2+t$ where s and t may vary, and of course the polynomial probability density 1 which can be seen as the limiting probability density corresponding to $s(x-r)^2+t$ as $s \to 0$. In particular, observe that

$$\alpha 3(1-x)^2 + (1-\alpha)3x^2 = 3(x-\alpha)^2 + 3\alpha(1-\alpha) \tag{4.102}$$

which in $\tilde{S}$ corresponds to

$$\alpha(1,0) + (1-\alpha)(0,1) = (\alpha, 1-\alpha). \tag{4.103}$$

This leads to the following corollary.

Corollary 4.4.3. *The image of $\mathcal{L}_{2,n}$ under π is a $(n+1)$-gon whose vertices are $(0,0)$ and n vertices on the ellipse E_n with equation*

$$\frac{(x+y-\frac{4n+6}{3n+9})^2}{\frac{4n^2}{9(n+3)^2}} + \frac{(x-y)^2}{\frac{4n^2}{3(n+3)(n-1)}} = 1. \tag{4.104}$$

Geometrically, the n vertices on E_n are given by the intersection of the line through $(\frac{1}{3}, \frac{1}{3})$ and $(\frac{k}{n-1}, \frac{n-1-k}{n-1})$ and the elliptic arc $E_n \cap \{(x,y) : x+y \geq 1\}$, for $k = 0, 1, \ldots, n-1$.

4.5 A model for random polynomials

Consider sampling a polynomial probability density uniformly at random from $\mathcal{D}_n$. That is, let $F(x)$ be a random polynomial with law

$$\mathbb{P}(F(x) \in Q) := \frac{\mathrm{vol}_n(Q)}{\mathrm{vol}_n(\mathcal{D}_n)} \tag{4.105}$$

for every measurable $Q \subseteq \mathcal{D}_n$ in the usual sense, where the n-dimensional volume ratio can be taken in any suitable coordinate representation of $\mathcal{D}_n$.

Random polynomials have been studied extensively since the 1930s. Classical models pioneered by Block and Polya [6], Littlewood and Offord [54], and Kac [42],[43] consider random polynomials of the form $\sum A_k x^k$ where the coefficients A_k are real or complex random variables according to some prescribed distribution, with particular interest various properties of the zeros and asymptotic behavior. See e.g. [3] for a recent survey and [23] for a geometric interpretation of some of the classical results. This model can be extended to coefficients in other polynomial bases, including trigonometric polynomials and orthogonal polynomials; see works of Farahmand [28] and Das [14]. In particular, random polynomials in the Bernstein basis are studied in [4], and Petrone [60] considers a model for random densities which are convex combinations of beta densities with coefficients following a Dirichlet distribution, with applications in Bayesian statistics [59], [61]. However, such models do not faithfully sample from $\mathcal{D}_n$, since they only consider nonnegative combinations of beta densities, which are supported on the Bernstein simplex.

Our results in Section 4.4 give us an elementary approach to understanding the case $n = 2$, where we consider a random polynomial probability density $F(x)$ of degree at most 2 with law

$$\mathbb{P}(F(x) \in Q) = \frac{\mathrm{area}(\pi(Q))}{\mathrm{area}(\tilde{S})} \tag{4.106}$$

Let R_n denote the region in $\tilde{S}$ bounded by $x \geq 0$, $y \geq 0$, and the arc $E_n \cap \{(x, y) : x + y \geq 1\}$. Note that $\pi(\mathcal{L}_{2,n}) \subseteq R_n$. We have the following basic geometry formula:

Lemma 4.5.1. *Let $0 < s < b$ and $a > 0$ and let E be the ellipse with equation*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{4.107}$$

Then the area of the sector of the ellipse above and defined by the line $y = -s$ is

$$ab\Big(\frac{\pi}{2} + \arcsin(s/b)\Big). \tag{4.108}$$

Proof. See Figure 4.3. The idea here is to scale the ellipse along its major axis to a circle, apply basic trigonometry, and then rescale back to original. We leave the calculation as an exercise. □

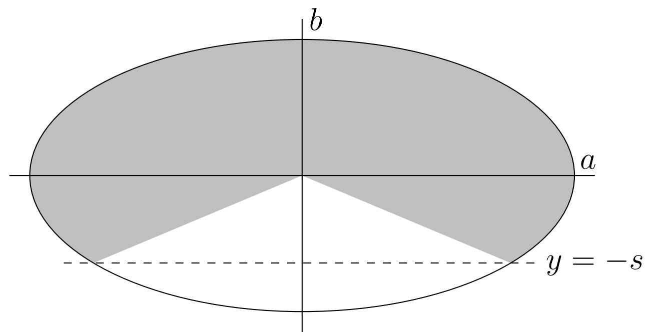


Figure 4.3: The sector of the ellipse (shaded) whose area is computed in Lemma 4.5.

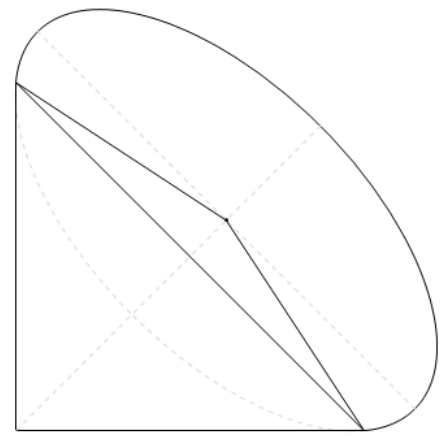


Figure 4.4: The region R_n, which is composed of two triangles and a sector of an ellipse. The area of R_n is computed in Proposition 4.5.2.

Proposition 4.5.2. *For $n \geq 3$, the area of R_n is*

$$\frac{n^2(\pi + 2\theta_n)}{3(n+3)\sqrt{3(n+3)(n-1)}} + \frac{2n+3}{3n+9}. \tag{4.109}$$

where

$$\theta_n := \arcsin\left(\tfrac{1}{2} - \tfrac{3}{2n}\right). \tag{4.110}$$

Proof. See Figure 4.4. The region R_n is composed of two triangles and a sector of the ellipse E_n. The triangle corresponding to the Bernstein simplex has area 1/2. The other triangle has base $\sqrt{2}$ and height

$$d\Big(\Big(\frac{1}{2}, \frac{1}{2}\Big), \Big(\frac{2n+3}{3n+9}, \frac{2n+3}{3n+9}\Big)\Big) = \sqrt{2}\frac{n-3}{6(n+3)}, \tag{4.111}$$

so it has an area of $\frac{n-3}{6(n+3)}$. The area of the sector of E_n can be computed using Lemma 4.5.1 with

$$a = a(n) := \frac{1}{\sqrt{2}}\frac{2n}{\sqrt{3(n+3)(n-1)}}, \tag{4.112}$$

$$b = b(n) := \frac{1}{\sqrt{2}}\frac{2n}{3(n+3)}, \tag{4.113}$$

$$s = s(n) := \sqrt{2}\frac{n-3}{6(n+3)}. \tag{4.114}$$

Adding the areas of the three components gives the desired formula. □

In particular, the probability that a random quadratic polynomial probability density has all coefficients nonnegative in the Bernstein basis $\{3(1-x)^2, 6x(1-x), 3x^2\}$ is

$$\mathbb{P}(\mathcal{L}_{2,2}) = \frac{\frac{1}{2}}{\frac{4\pi}{9\sqrt{3}} + \frac{2}{3}} \approx 0.33949. \tag{4.115}$$

Theorem 4.5.3. *Let $F(x)$ be a random polynomial sampled uniformly from $\mathcal{D}_2$. Then for $n \geq 3$,*

$$\mathbb{P}\big(\delta(F(x)) > n\big) > \left(\tfrac{4\pi}{9\sqrt{3}} + \tfrac{2}{3}\right)^{-1} \frac{1}{n+3}. \tag{4.116}$$

Proof. Observe that $0 \leq \frac{1}{2} - \frac{3}{2n} < \frac{1}{2}$ for $n \geq 3$, so $\theta_n < \frac{\pi}{6}$ as defined in (4.110) and

$$\text{area}(\pi(\mathcal{L}_{2,n})) < \text{area}(R_n) < \frac{n^2(\frac{4\pi}{3})}{3(n+3)\sqrt{3(n+3)(n-1)}} + \frac{2n+3}{3n+9} < \frac{4\pi}{9\sqrt{3}} + \frac{2n+3}{3n+9}. \tag{4.117}$$

By Proposition 4.5.2, for $n \geq 3$ we have

$$\mathbb{P}(\delta(F(x)) > n) = 1 - \mathbb{P}(F(x) \in \mathcal{L}_{2,n}) > 1 - \frac{\frac{4\pi}{9\sqrt{3}} + \frac{2n+3}{3n+9}}{\frac{4\pi}{9\sqrt{3}} + \frac{2}{3}} = \left(\tfrac{4\pi}{9\sqrt{3}} + \tfrac{2}{3}\right)^{-1} \frac{1}{n+3}. \tag{4.118}$$

□

Corollary 4.5.4. *Let $F(x)$ be a random polynomial sampled uniformly from $\mathcal{D}_2$. Then*

$$\mathbb{E}\delta(F(x)) = \infty. \tag{4.119}$$

4.6 Upper envelopes

In this section, we consider the problem of describing for each n the upper envelope of $\mathcal{D}_n$; that is, for each $u \in [0,1]$,

$$\overline{f}_n(u) := \max_{f \in \mathcal{D}_n} f(u) = \max_{f \in \text{ext}(\mathcal{D}_n)} f(u). \tag{4.120}$$

Note that finiteness of $\overline{f}_n(u)$ and the second equality follow from linearity of the functional $f \mapsto f(u)$ and compactness of $\mathcal{D}_n$. It is an easy calculus exercise to show that for $n = 2$,

$$\overline{f}_2(u) = \max(4 - 12u(1-u), 6u(1-u)) \tag{4.121}$$

$$= \begin{cases} 4 - 12u(1-u) & \text{if } 0 \leq u \leq \frac{1}{3} \text{ or } \frac{2}{3} \leq u \leq 1, \\ 6u(1-u) & \text{if } \frac{1}{3} < u < \frac{2}{3}. \end{cases} \tag{4.122}$$

We first establish some basic definitions from the theory of *orthogonal polynomials*; see e.g. [73].

An *orthonormal system* of polynomials on $[0,1]$ with respect to a *weight function* $w : [0,1] \to [0,\infty)$ is a sequence of polynomials $(p_n(x) : n \geq 0)$ such that $\deg(p_n(x)) = n$ and

$$\int_0^1 p_m(x)p_n(x)w(x)dx = 1(m = n) \tag{4.123}$$

for all $m,n \geq 0$. The normalization $\int_0^1 p_n(x)^2 w(x)dx = 1$ uniquely determines the sequence of polynomials. Note that for every $n \geq 0$, $\{p_k(x) : 0 \leq k \leq n\}$ forms a basis for $\mathcal{P}_n$. If $w \in \mathcal{D}$, then there is a natural probabilistic reformulation of (4.123): if X is a random variable with probability density w on $[0,1]$, then

$$\mathbb{E}p_m(X)p_n(X) = 1(m = n) \tag{4.124}$$

for all $m,n \geq 0$. In particular, the beta densities

$$w(x) = \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta+1)} x^\beta (1-x)^\alpha \tag{4.125}$$

correspond, up to a scaling factor, to the classical orthonormal system known as the (shifted) *Jacobi polynomials* $p_n^{(\alpha,\beta)}(x)$. We shall need the following three special cases of Jacobi polynomials.

- $(\alpha,\beta) = (0,0)$ corresponding to $w(x) \equiv 1$. This gives us the (shifted) *Legendre polynomials.*
- $(\alpha,\beta) = (0,1)$ corresponding to $w(x) = 2x$.
- $(\alpha,\beta) = (1,1)$ corresponding to $w(x) = 6x(1-x)$.

Table 4.1 displays the first few polynomials in each of these three orthonormal systems.

Lemma 4.6.1. *[1][39] Let $(p_n(x) : n \geq 0)$ be an orthonormal system on $[0,1]$ with respect to weight function w. Suppose*

$$f(x) = \sum_{k=0}^m a_k p_k(x) \in \mathcal{P}_m \tag{4.126}$$

satisfies $\int_0^1 f(x)^2 w(x)dx = 1$. Then for every $0 \leq u \leq 1$,

$$f(u)^2 \leq \sum_{k=0}^m p_k(u)^2 \tag{4.127}$$

with equality if and only if

$$f(x) = \pm \frac{\sum_{k=0}^m p_k(u)p_k(x)}{\sqrt{\sum_{k=0}^m p_k(u)^2}}. \tag{4.128}$$

Proof. By orthonormality,

$$1 = \int_0^1 f(x)^2 w(x)dx = \int_0^1 \Big(\sum_{k=0}^m a_k p_k(x)\Big)^2 w(x)dx = \sum_{k=0}^m a_k^2. \tag{4.129}$$

Then by the Cauchy-Schwarz inequality,

$$f(u)^2 \le \sum_{k=0}^m a_k^2 \sum_{k=0}^m p_k(u)^2 = \sum_{k=0}^m p_k(u)^2 \tag{4.130}$$

with equality if and only if $a_k = ap_k(u)$ for all k, for some constant a. Hence by (4.129) $a = \pm(\sum p_k(u)^2)^{-1/2}$ and equality is attained by

$$f(x) = \pm\sum_{k=0}^m ap_k(u)p_k(x) \tag{4.131}$$

as desired. □

Table 4.1: Some Jacobi polynomials, defined according to (4.123) and the convention (4.125).

n	$p_n^{(0,0)}(x)$	$p_n^{(0,1)}(x)$
0	1	1
1	$\sqrt{3}(2x-1)$	$\sqrt{2}(3x-2)$
2	$\sqrt{5}(6x^2-6x+1)$	$\sqrt{3}(10x^2-12x+3)$
3	$\sqrt{7}(20x^3-30x^2+12x-1)$	$\sqrt{4}(35x^3-60x^2+30x-4)$
4	$\sqrt{9}(70x^4-140x^3+90x^2-20x+1)$	$\sqrt{5}(126x^4-280x^3+210x^2-60x+5)$

n	$p_n^{(1,1)}(x)$
0	1
1	$\sqrt{5}(2x-1)$
2	$\sqrt{14}(5x^2-5x+1)$
3	$\sqrt{30}(14x^3-21x^2+9x-1)$
4	$\sqrt{55}(42x^4-84x^3+56x^2-14x+1)$

Theorem 4.6.2. *Let $0 \leq u \leq 1$. Then*

(i) for even $n = 2m > 0$,

$$\overline{f}_n(u) = \max\left(\sum_{k=0}^{m} p_k^{(0,0)}(u)^2,\ 6u(1-u)\sum_{k=0}^{m-1} p_k^{(1,1)}(u)^2\right); \tag{4.132}$$

(ii) for odd $n = 2m+1$,

$$\overline{f}_n(u) = \max\left(2u\sum_{k=0}^{m} p_k^{(0,1)}(u)^2,\ 2(1-u)\sum_{k=0}^{m} p_k^{(1,0)}(u)^2\right). \tag{4.133}$$

Proof. The result follows from Lemma 4.6.1 and Theorem 4.3.1. Note that by symmetry, $p_k^{(1,0)}(x) = p_k^{(0,1)}(1-x)$. □

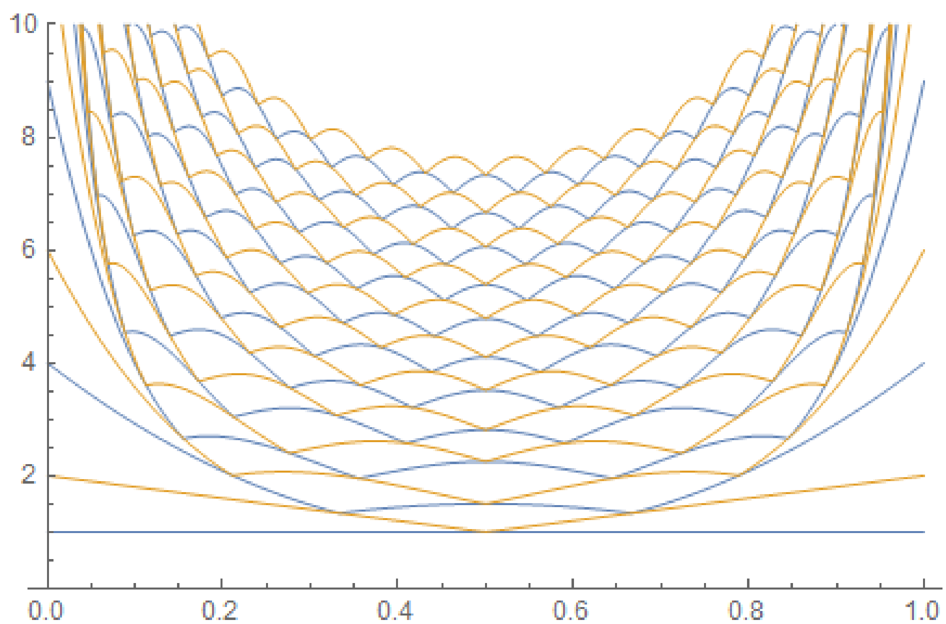


Figure 4.5: The upper envelopes $\overline{f}_n$ of $\mathcal{D}_n$ for $0 \leq n \leq 21$, colored blue for even n and yellow for odd n. The formulas for $\overline{f}_n$ are given in terms of Jacobi polynomials in Theorem 4.6.2.

Corollary 4.6.3. *We have*

$$\overline{f}_n(0) = \overline{f}_n(1) = \left\lfloor\frac{n+2}{2}\right\rfloor\left\lceil\frac{n+2}{2}\right\rceil = \begin{cases}(m+1)^2 & \textit{for even } n = 2m,\\ (m+1)(m+2) & \textit{for odd } n = 2m+1.\end{cases} \tag{4.134}$$

Proof. The result follows from Theorem 4.6.2 and the following well-known identities involving the Jacobi polynomials (see [66]) defined according to (4.125) which can be seen in Table 4.1 for low degrees:

- $p_k^{(0,0)}(1) = \sqrt{2k+1}$
- $p_k^{(0,1)}(1) = \sqrt{k+1}$

Then for $n = 2m$,

$$\overline{f}_n(1) = \sum_{k=0}^{m}(2k+1) = (m+1)^2 \tag{4.135}$$

and for $n = 2m+1$,

$$\overline{f}_n(1) = 2\sum_{k=0}^{m}(k+1) = (m+1)(m+2). \tag{4.136}$$

□

The sequence defined in (4.134) is the sequence of positive quarter-squares; see OEIS entry A002620.

Next, consider

$$M_n := \max_{f\in\mathcal{D}_n}\max_{0\le u\le 1} f(u) = \max_{0\le u\le 1}\overline{f}_n(u). \tag{4.137}$$

Following intuition, it is natural to suspect that $\overline{f}_n$ is maximized at the endpoints of the interval $[0,1]$. Indeed this is correct, but the proof is nontrivial. Let

$$\mathcal{I}_n := \{\text{non-negative, increasing polynomials on } [0,1] \text{ of degree at most } n\}. \tag{4.138}$$

Lemma 4.6.4. *[66] For every nonnegative, increasing polynomial function f on $[0,1]$, define*

$$\lambda(f) := \frac{\max_{u\in[0,1]} f'(u)}{f(1)}. \tag{4.139}$$

Then for $n \ge 1$,

$$\lambda_n := \sup_{f\in\mathcal{I}_n}\lambda(f) = \sup_{f\in\mathcal{I}_n}\frac{f'(1)}{f(1)} \tag{4.140}$$

Proof. Suppose there does not exist $f\in\mathcal{I}_n$ satisfying $\frac{f'(1)}{f(1)} = \lambda_n$. It follows by considering the function $f(1)-f(1-x)$ that also no $f\in\mathcal{I}_n$ satisfies $\frac{f'(0)}{f(1)} = \lambda_n$. Let $f\in\mathcal{I}_n$ satisfy $\lambda(f)=\lambda_n$. Then $\frac{f'(c)}{f(1)} = \lambda_n$ for some $c\in(0,1)$. Define $g,h\in\mathcal{I}_n$ by

$$g(x) = f(cx), \tag{4.141}$$

$$h(x) = f(c+(1-c)x) - f(c). \tag{4.142}$$

Then

$$\lambda_n > \frac{g'(1)}{g(1)} = \frac{cf'(c)}{f(c)} \implies f(c) > \frac{cf'(c)}{\lambda_n}, \tag{4.143}$$

$$\lambda_n > \frac{h'(0)}{h(1)} = \frac{(1-c)f'(c)}{f(1)-f(c)} \implies f(1) - f(c) > \frac{(1-c)f'(c)}{\lambda_n}. \tag{4.144}$$

Adding the two inequalities yields $f(1) > \frac{f'(c)}{\lambda_n}$ or $\lambda_n > \frac{f'(c)}{f(1)}$, which is a contradiction. □

Corollary 4.6.5. *For $n \geq 0$,*

$$M_n := \max_{0\leq u\leq 1} \overline{f}_n(u) = \begin{cases} (m+1)^2 & \text{if } n = 2m, \\ (m+1)(m+2) & \text{if } n = 2m+1. \end{cases} \tag{4.145}$$

Proof. Note that $\frac{f'(1)}{f(1)-f(0)} > \frac{f'(1)}{f(1)}$ whenever $f(0) > 0$, which implies that every $f \in \mathcal{I}_n$ which maximizes $\lambda(f)$ must have $f(0) = 0$. Thus Lemma 4.6.4 can be reformulated using calculus as

$$M_n = \sup_{f\in\mathcal{P}_n^+} \frac{\max_{0\leq u\leq 1} f(u)}{\int_0^1 f(x)dx} = \sup_{f\in\mathcal{P}_n^+} \frac{f(1)}{\int_0^1 f(x)dx} = \overline{f}_n(1). \tag{4.146}$$

□

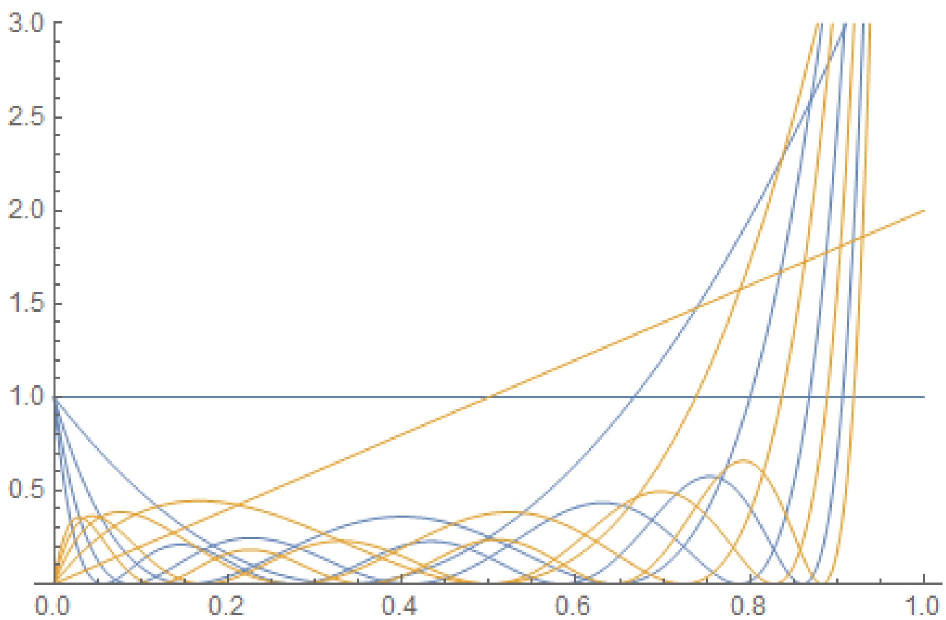


Figure 4.6: Graphs of the unique polynomials f which maximize $f(1)$ over $f \in \mathcal{D}_n$ for $0 \leq n \leq 9$, colored blue for even n and yellow for odd n. For each n, the degree n polynomial is an extreme point of $\mathcal{D}_n$. See Table 4.2 for exact formulas for $0 \leq n \leq 9$.

Table 4.2: The unique polynomials which maximize $f(1)$ over $f \in \mathcal{D}_n$ for $0 \leq n \leq 9$. Their graphs are shown in Figure 4.6. The general formulas for every n are given in Lemma 4.6.1 in terms of Jacobi polynomials. Observe that for odd n the polynomials are the normalized squared Jacobi polynomials of the form $p_n^{(1,1)}(x)$ (cf. Table 4.1).

n	$\arg\max_{f \in \mathcal{D}_n} f(1)$
0	1
1	$2x$
2	$(3x-1)^2$
3	$6x(2x-1)^2$
4	$(10x^2-8x+1)^2$
5	$12x(5x^2-5x+1)^2$
6	$(35x^3-45x^2+15x-1)^2$
7	$20x(14x^3-21x^2+9x-1)$
8	$(126x^4-224x^3+126x^2-24x+1)^2$
9	$30x(42x^4-84x^3+56x^2-14x+1)^2$

Bibliography

[1] George E. Andrews, Richard Askey, and Ranjan Roy. *Special functions*. Vol. 71. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1999, pp. xvi+664. ISBN: 0-521-62321-9; 0-521-78988-5. DOI: 10.1017/CBO9781107325937. URL: https://doi.org/10.1017/CBO9781107325937.

[2] Itai Arieli, Yakov Babichenko, Fedor Sandomirskiy, and Omer Tamuz. "Feasible Joint Posterior Beliefs". In: *Journal of Political Economy* 129.9 (2021), pp. 2546–2594. DOI: 10.1086/714993. URL: https://ideas.repec.org/a/ucp/jpolec/doi10.1086-714993.html.

[3] Diego Armentano. "A review of some recent results on random polynomials over ℝ and over ℂ". In: *Publ. Mat. Urug.* 12 (2011), pp. 1–14. ISSN: 0797-1443.

[4] Diego Armentano and Jean-Pierre Dedieu. "A note about the average number of real roots of a Bernstein polynomial system". In: *J. Complexity* 25.4 (2009), pp. 339–342. ISSN: 0885-064X. DOI: 10.1016/j.jco.2009.03.001. URL: https://doi.org/10.1016/j.jco.2009.03.001.

[5] S. N. Bernšteın. *Sobranie sočinenii. Tom I. Konstruktivnaya teoriya funkciı [1905–1930]*. Izdat. Akad. Nauk SSSR, Moscow, 1952, 581 pp. (1 plate).

[6] A. Bloch and G. Pólya. "On the Roots of Certain Algebraic Equations". In: *Proc. London Math. Soc. (2)* 33.2 (1931), pp. 102–114. ISSN: 0024-6115. DOI: 10.1112/plms/s2-33.1.102. URL: https://doi.org/10.1112/plms/s2-33.1.102.

[7] Eric Bouyé, Valdo Durlleman, Ashkan Nikeghbali, Gaël Riboulet, and Thierry Roncalli. "Copulas for Finance". In: (Jan. 2000). DOI: https://dx.doi.org/10.2139/ssrn.1032533.

[8] Krzysztof Burdzy and Soumik Pal. "Can coherent predictions be contradictory?" In: *Adv. in Appl. Probab.* 53.1 (2021), pp. 133–161. ISSN: 0001-8678. DOI: 10.1017/apr.2020.51. URL: https://doi.org/10.1017/apr.2020.51.

[9] Krzysztof Burdzy and Jim Pitman. *Bounds on the probability of radically different opinions*. 2019. eprint: arXiv:1903.07773.

[10] Constantin Carathéodory. "Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen". In: *Rendiconti Del Circolo Matematico di Palermo (1884-1940)* 32.1 (1911), pp. 193–217.

[11] Holly Carley. "Examples of doubly stochastic measures supported on the graphs of two functions". In: *Comm. Statist. Theory Methods* 46.17 (2017), pp. 8628–8630. ISSN: 0361-0926. DOI: 10.1080/03610926.2016.1186189. URL: https://doi.org/10.1080/03610926.2016.1186189.

[12] Stanisław Cichomski and Adam Os ekowski. "The maximal difference among expert's opinions". In: *Electron. J. Probab.* 26 (2021), Paper No. 105, 17. DOI: 10.1214/21-ejp675. URL: https://doi.org/10.1214/21-ejp675.

[13] Stanisław Cichomski and Fedor Petrov. *A combinatorial proof of the Burdzy-Pitman conjecture*. 2022. DOI: 10.48550/ARXIV.2204.07219. URL: https://arxiv.org/abs/2204.07219.

[14] Minaketan Das. "Real zeros of a random sum of orthogonal polynomials". In: *Proc. Amer. Math. Soc.* 27 (1971), pp. 147–153. ISSN: 0002-9939. DOI: 10.2307/2037279. URL: https://doi.org/10.2307/2037279.

[15] A. P. Dawid, M. H. DeGroot, and J. Mortera. "Coherent combination of experts' opinions". In: *Test* 4.2 (1995). With discussion and a reply by Dawid and Mortera, pp. 263–313. ISSN: 1133-0686. DOI: 10.1007/BF02562628. URL: https://doi.org/10.1007/BF02562628.

[16] Sudhakar Dharmadhikari and Kumar Joag-Dev. *Unimodality, convexity, and applications*. Probability and Mathematical Statistics. Academic Press, Inc., Boston, MA, 1988, pp. xiv+278. ISBN: 0-12-214690-5.

[17] Persi Diaconis. "Finite forms of de Finetti's theorem on exchangeability". In: *Synthese* 36.2 (1977). Foundations of probability and statistics, II, pp. 271–281. ISSN: 0039-7857. DOI: 10.1007/BF00486116. URL: https://doi.org/10.1007/BF00486116.

[18] R. G. Douglas. "On extremal measures and subspace density". In: *Michigan Math. J.* 11 (1964), pp. 243–246. ISSN: 0026-2285. URL: http://projecteuclid.org/euclid.mmj/1028999137.

[19] Lester E. Dubins and Jim Pitman. "A divergent, two-parameter, bounded martingale". In: *Proc. Amer. Math. Soc.* 78.3 (1980), pp. 414–416. ISSN: 0002-9939. DOI: 10.2307/2042335. URL: https://doi.org/10.2307/2042335.

[20] Lester E. Dubins and Jim Pitman. "A maximal inequality for skew fields". In: *Z. Wahrsch. Verw. Gebiete* 52.3 (1980), pp. 219–227. ISSN: 0044-3719. DOI: 10.1007/BF00538887. URL: https://doi.org/10.1007/BF00538887.

[21] Fabrizio Durante, Juan Fernández-Sánchez, and Carlo Sempi. "A topological proof of Sklar's theorem". In: *Applied Mathematics Letters* 26.9 (2013), pp. 945–948. ISSN: 0893-9659. DOI: https://doi.org/10.1016/j.aml.2013.04.005. URL: https://www.sciencedirect.com/science/article/pii/S0893965913001183.

[22] Rick Durrett. *Probability: theory and examples*. Fourth. Vol. 31. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010, pp. x+428. ISBN: 978-0-521-76539-8. DOI: 10.1017/CBO9780511779398. URL: https://doi.org/10.1017/CBO9780511779398.

[23] Alan Edelman and Eric Kostlan. "How many zeros of a random polynomial are real?" In: *Bull. Amer. Math. Soc. (N.S.)* 32.1 (1995), pp. 1–37. ISSN: 0273-0979. DOI: 10.1090/S0273-0979-1995-00571-9. URL: https://doi.org/10.1090/S0273-0979-1995-00571-9.

[24] T. Erdélyi. "George Lorentz and inequalities in approximation". In: *Algebra i Analiz* 21.3 (2009), pp. 1–57. ISSN: 0234-0852. DOI: 10.1090/S1061-0022-10-01099-X. URL: https://doi.org/10.1090/S1061-0022-10-01099-X.

[25] T. Erdélyi and J. Szabados. "On polynomials with positive coefficients". In: *J. Approx. Theory* 54.1 (1988), pp. 107–122. ISSN: 0021-9045. DOI: 10.1016/0021-9045(88)90119-0. URL: https://doi.org/10.1016/0021-9045(88)90119-0.

[26] Tamás Erdélyi. "Estimates for the Lorentz degree of polynomials". In: *J. Approx. Theory* 67.2 (1991), pp. 187–198. ISSN: 0021-9045. DOI: 10.1016/0021-9045(91)90017-5. URL: https://doi.org/10.1016/0021-9045(91)90017-5.

[27] W. J. Ewens. "The sampling theory of selectively neutral alleles". In: *Theoret. Population Biol.* 3 (1972). ISSN: 0040-5809. DOI: 10.1016/0040-5809(72)90035-4. URL: https://doi.org/10.1016/0040-5809(72)90035-4.

[28] Kambiz Farahmand. *Topics in random polynomials*. Vol. 393. Pitman Research Notes in Mathematics Series. Longman, Harlow, 1998, pp. x+163. ISBN: 0-582-35622-9.

[29] William Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968, pp. xviii+509.

[30] Bruno de Finetti. "La Prévision: Ses Lois Logiques, Ses Sources Subjectives". In: *Annales de l'Institut Henri Poincaré* 17 (1937), pp. 1–68.

[31] Gerald B. Folland. *Real analysis*. Second. Pure and Applied Mathematics (New York). Modern techniques and their applications, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, pp. xvi+386. ISBN: 0-471-31716-0.

[32] Partha Ghosh and Subir Bhandari. "Characterization of Extreme Copulas". In: (Sept. 2017).

[33] Alexander Gnedin, Ben Hansen, and Jim Pitman. "Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws". In: *Probab. Surv.* 4 (2007), pp. 146–171. DOI: 10.1214/07-PS092. URL: https://doi.org/10.1214/07-PS092.

[34] Zhong Guan. "Bernstein polynomial model for grouped continuous data". In: *J. Nonparametr. Stat.* 29.4 (2017), pp. 831–848. ISSN: 1048-5252. DOI: 10.1080/10485252.2017.1374384. URL: https://doi.org/10.1080/10485252.2017.1374384.

[35] Zhong Guan. "Efficient and robust density estimation using Bernstein type polynomials". In: *J. Nonparametr. Stat.* 28.2 (2016), pp. 250–271. ISSN: 1048-5252. DOI: 10.1080/10485252.2016.1163349. URL: https://doi-org.libproxy.berkeley.edu/10.1080/10485252.2016.1163349.

[36] Zhong Guan. "Maximum approximate Bernstein likelihood estimation in proportional hazard model for interval-censored data". In: *Stat. Med.* 40.3 (2021), pp. 758–778. ISSN: 0277-6715. DOI: 10.1002/sim.8801. URL: https://doi.org/10.1002/sim.8801.

[37] Kevin He, Fedor Sandomirskiy, and Omer Tamuz. "Private Private Information". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22. Boulder, CO, USA: Association for Computing Machinery, 2022, p. 1145. ISBN: 9781450391504. DOI: 10.1145/3490486.3538348. URL: https://doi.org/10.1145/3490486.3538348.

[38] Edwin Hewitt and Leonard J. Savage. "Symmetric measures on Cartesian products". In: *Trans. Amer. Math. Soc.* 80 (1955), pp. 470–501. ISSN: 0002-9947. DOI: 10.2307/1992999. URL: https://doi.org/10.2307/1992999.

[39] Mourad E. H. Ismail. *Classical and quantum orthogonal polynomials in one variable*. Vol. 98. Encyclopedia of Mathematics and its Applications. With two chapters by Walter Van Assche, With a foreword by Richard A. Askey. Cambridge University Press, Cambridge, 2005, pp. xviii+706. ISBN: 978-0-521-78201-2; 0-521-78201-5. DOI: 10.1017/CBO9781107325982. URL: https://doi.org/10.1017/CBO9781107325982.

[40] Svante Janson, Takis Konstantopoulos, and Linglong Yuan. "On a representation theorem for finitely exchangeable random vectors". In: *J. Math. Anal. Appl.* 442.2 (2016), pp. 703–714. ISSN: 0022-247X. DOI: 10.1016/j.jmaa.2016.04.070. URL: https://doi.org/10.1016/j.jmaa.2016.04.070.

[41] S. Johansen. "An application of extreme point methods to the representation of infinitely divisible distributions". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 5 (1966), pp. 304–316. DOI: 10.1007/BF00535361. URL: https://doi.org/10.1007/BF00535361.

[42] M. Kac. "On the average number of real roots of a random algebraic equation". In: *Bull. Amer. Math. Soc.* 49 (1943), pp. 314–320. ISSN: 0002-9904. DOI: 10.1090/S0002-9904-1943-07912-8. URL: https://doi.org/10.1090/S0002-9904-1943-07912-8.

[43] M. Kac. "On the average number of real roots of a random algebraic equation. II". In: *Proc. London Math. Soc. (2)* 50 (1949), pp. 390–408. ISSN: 0024-6115. DOI: 10.1112/plms/s2-50.5.390. URL: https://doi.org/10.1112/plms/s2-50.5.390.

[44] S. Karlin and L. S. Shapley. "Geometry of moment spaces". In: *Mem. Amer. Math. Soc.* 12 (1953), p. 93. ISSN: 0065-9266.

[45] Samuel Karlin. "Central limit theorems for certain infinite urn schemes". In: *J. Math. Mech.* 17 (1967), pp. 373–401. DOI: 10.1512/iumj.1968.17.17020. URL: https://doi.org/10.1512/iumj.1968.17.17020.

[46] Alan F. Karr. "Extreme points of certain sets of probability measures, with applications". In: *Math. Oper. Res.* 8.1 (1983), pp. 74–85. ISSN: 0364-765X. DOI: 10.1287/moor.8.1.74. URL: https://doi.org/10.1287/moor.8.1.74.

[47] Robert W. Keener. *Theoretical statistics*. Springer Texts in Statistics. Topics for a core course. Springer, New York, 2010, pp. xviii+538. ISBN: 978-0-387-93838-7. DOI: 10.1007/978-0-387-93839-4. URL: https://doi.org/10.1007/978-0-387-93839-4.

[48] A Ya Khintchine. "On unimodal distributions". In: *Izvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki* 2.2 (1938), pp. 1–7.

[49] Davar Khoshnevisan. *Multiparameter processes*. Springer Monographs in Mathematics. An introduction to random fields. Springer-Verlag, New York, 2002, pp. xx+584. ISBN: 0-387-95459-7. DOI: 10.1007/b97363. URL: https://doi.org/10.1007/b97363.

[50] J. F. C. Kingman. "The representation of partition structures". In: *J. London Math. Soc. (2)* 18.2 (1978), pp. 374–380. ISSN: 0024-6107. DOI: 10.1112/jlms/s2-18.2.374. URL: https://doi.org/10.1112/jlms/s2-18.2.374.

[51] Takis Konstantopoulos and Linglong Yuan. "On the extendibility of finitely exchangeable probability measures". In: *Trans. Amer. Math. Soc.* 371.10 (2019), pp. 7067–7092. ISSN: 0002-9947. DOI: 10.1090/tran/7701. URL: https://doi.org/10.1090/tran/7701.

[52] Pranesh Kumar. "Copula Functions and Applications in Engineering: Performance and Safety Management". In: Jan. 2019, pp. 195–209. ISBN: 978-981-13-0871-0. DOI: 10.1007/978-981-13-0872-7_15.

[53] J. Lindenstrauss. "A Remark on Extreme Doubly Stochastic Measures". In: 1965.

[54] J. E. Littlewood and A. C. Offord. "On the Number of Real Roots of a Random Algebraic Equation". In: *J. London Math. Soc.* 13.4 (1938), pp. 288–295. ISSN: 0024-6107. DOI: 10.1112/jlms/s1-13.4.288. URL: https://doi.org/10.1112/jlms/s1-13.4.288.

[55] G. G. Lorentz. "The degree of approximation by polynomials with positive coefficients". In: *Math. Ann.* 151 (1963), pp. 239–251. ISSN: 0025-5831. DOI: 10.1007/BF01398235. URL: https://doi.org/10.1007/BF01398235.

[56] Marvin Marcus and Henryk Minc. *A survey of matrix theory and matrix inequalities*. Reprint of the 1969 edition. Dover Publications, Inc., New York, 1992, pp. xii+180. ISBN: 0-486-67102-X.

[57] Roger Nelsen. "Properties and applications of copulas: A brief survey". In: *Proc. 1st Braz. Conf. on Stat. Modeling in Insurance and Finance* (Jan. 2003).

[58] Pablo A. Parrilo and Rekha R. Thomas, eds. *Sum of squares: theory and applications*. Vol. 77. Proceedings of Symposia in Applied Mathematics. AMS Short Course, Sum of Squares: Theory and Applications, January 14–15, 2019, Baltimore, MD. American Mathematical Society, Providence, RI, [2020] ©2020, pp. vii+142. ISBN: 978-1-4704-5025-0.

[59] Sonia Petrone. "Bayesian density estimation using Bernstein polynomials". In: *Canad. J. Statist.* 27.1 (1999), pp. 105–126. ISSN: 0319-5724. DOI: 10.2307/3315494. URL: https://doi-org.libproxy.berkeley.edu/10.2307/3315494.

[60] Sonia Petrone. "Random Bernstein polynomials". In: *Scand. J. Statist.* 26.3 (1999), pp. 373–393. ISSN: 0303-6898. DOI: 10.1111/1467-9469.00155. URL: https://doi-org.libproxy.berkeley.edu/10.1111/1467-9469.00155.

[61] Sonia Petrone and Larry Wasserman. "Consistency of Bernstein polynomial posteriors". In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64.1 (2002), pp. 79–100. ISSN: 1369-7412. DOI: 10.1111/1467-9868.00326. URL: https://doi-org.libproxy.berkeley.edu/10.1111/1467-9868.00326.

[62] L. A. Petrov. "A two-parameter family of infinite-dimensional diffusions on the Kingman simplex". In: *Funktsional. Anal. i Prilozhen.* 43.4 (2009), pp. 45–66. ISSN: 0374-1990. DOI: 10.1007/s10688-009-0036-8. URL: https://doi.org/10.1007/s10688-009-0036-8.

[63] J. Pitman. *Combinatorial stochastic processes*. Vol. 1875. Lecture Notes in Mathematics. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. Springer-Verlag, Berlin, 2006, pp. x+256. ISBN: 978-3-540-30990-1; 3-540-30990-X.

[64] Jim Pitman. "Exchangeable and partially exchangeable random partitions". In: *Probab. Theory Related Fields* 102.2 (1995), pp. 145–158. ISSN: 0178-8051. DOI: 10.1007/BF01213386. URL: https://doi.org/10.1007/BF01213386.

[65] Victoria Powers and Bruce Reznick. "Polynomials that are positive on an interval". In: *Transactions of the American Mathematical Society* 352 (June 2000), pp. 4677–4692. DOI: 10.1090/S0002-9947-00-02595-2.

[66] Q. I. Rahman and G. Schmeisser. *Analytic theory of polynomials*. Vol. 26. London Mathematical Society Monographs. New Series. The Clarendon Press, Oxford University Press, Oxford, 2002, pp. xiv+742. ISBN: 0-19-853493-0.

[67] T. J. Rivlin. "Bounds on a polynomial". In: *J. Res. Nat. Bur. Standards Sect. B* 74B (1970), pp. 47–54. ISSN: 0022-4340.

[68] T. L. Seethoff and R. C. Shiflett. "Doubly stochastic measures with prescribed support". In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 41.4 (1977/78), pp. 283–288. DOI: 10.1007/BF00533599. URL: https://doi.org/10.1007/BF00533599.

[69] H. Sherwood and M. D. Taylor. "Doubly stochastic measures with hairpin support". In: *Probab. Theory Related Fields* 78.4 (1988), pp. 617–626. ISSN: 0178-8051. DOI: 10.1007/BF00353879. URL: https://doi.org/10.1007/BF00353879.

[70] Barry Simon. *Convexity*. Vol. 187. Cambridge Tracts in Mathematics. An analytic viewpoint. Cambridge University Press, Cambridge, 2011, pp. x+345. ISBN: 978-1-107-00731-4. DOI: 10.1017/CBO9780511910135. URL: https://doi.org/10.1017/CBO9780511910135.

[71] Frank Spitzer. *Principles of random walk*. Second. Graduate Texts in Mathematics, Vol. 34. Springer-Verlag, New York-Heidelberg, 1976, pp. xiii+408.

[72] V. Strassen. "The existence of probability measures with given marginals". In: *Ann. Math. Statist.* 36 (1965), pp. 423–439. ISSN: 0003-4851. DOI: 10.1214/aoms/1177700153. URL: https://doi.org/10.1214/aoms/1177700153.

[73] Gábor Szegő. *Orthogonal polynomials*. Fourth. American Mathematical Society Colloquium Publications, Vol. XXIII. American Mathematical Society, Providence, R.I., 1975, pp. xiii+432.

[74] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[75] Richard A. Vitale. "Bernstein polynomial approach to density function estimation". In: *Statistical inference and related topics (Proc. Summer Res. Inst. Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., 1974, Vol. 2; dedicated to Z. W. Birnbaum)*. 1975, pp. 87–99.

[76] Gerhard Winkler. "Extreme points of moment sets". In: *Math. Oper. Res.* 13.4 (1988), pp. 581–587. ISSN: 0364-765X. DOI: 10.1287/moor.13.4.581. URL: https://doi.org/10.1287/moor.13.4.581.

[77] Yuri Yakubovich. "On the distribution of the number of distinct values in a finite sample". In: *Unpublished* (2021).

[78] Theodore Zhu. "The distribution of the number of distinct values in a finite exchangeable sequence". In: *Electronic Journal of Probability* 27 (2022), pp. 1–25. DOI: 10.1214/22-EJP815. URL: https://doi.org/10.1214/22-EJP815.