University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Stanisław Cichomski

Combinatorial methods in the analysis of coherent distributions and related inequalities

PhD *dissertation*
in Mathematics

Supervisor
prof. dr hab. Adam Osękowski
Institute of Mathematics
University of Warsaw

Warsaw, September 2024

Author's declaration.

I hereby declare that I have written this dissertation myself and all the contents of the dissertation have been obtained by legal means.

September 21, 2024 ....

mgr Stanisław Cichomski

Supervisors' declaration.

The dissertation is ready to be reviewed.

September 21, 2024 ....

prof. dr hab. Adam Osękowski

Abstract

The primary objective of this dissertation is to expand our current understanding of coherent distributions, their geometric properties and related maximal inequalities. The importance of this topic stems from its significant applications in probability theory, microeconomics and statistics: coherent vectors appear naturally when the odds of a given random event are estimated based on various sources of information. The thesis enhances the underlying theory of coherent distributions and introduces some efficient strategies tailored for the optimization context.

The work presented in this thesis is based on five published research articles and one paper accepted for publication. For sake of convenience, we have decided to conduct the analysis separately in two directions: for the bivariate and for the multivariate distributions.

In the first part, we focus on the convex structure of coherent distributions and the general characterization of corresponding extreme points. We analyze the supports of extreme coherent measures with finite number of atoms and provide an example of an extreme distribution with an uncountable support and with no atoms. Finally, we also confirm the Burdzy–Pitman conjecture about the maximal spread of coherent and independent distributions.

In the second part, we investigate some open questions connected with the maximal dispersion of multivariate coherent distributions. We show a useful symmetrization technique and verify its essential properties. As a primary application, we generalize the two-variate inequality of Burdzy and Pal, and in particular offer its alternative derivation. Furthermore, we explore the underlying interrelation between coherent distributions and classical martingale theory, and propose an appropriate version of Doob's maximal estimate for special "tree-shaped" filtrations.

The characteristic feature of the work is the highly combinatorial nature of the reasoning. We will repeatedly use the discretization argument and the technique of consecutive transformations and reductions. This approach enables us to benefit from the interplay between coherent distributions, graph theory, and combinatorial matrix theory.

Keywords: coherent distributions, expert opinions, feasible joint posterior beliefs, contradictory predictions, maximal spread, conditional probability, joint distribution of conditional expectations, stochastic inequalities.

AMS MSC 2020 classification: 60C05, 60E05, 60E15, 62E10.

Streszczenie

Głównym celem tej rozprawy jest poszerzenie dostępnej wiedzy dotyczącej rozkładów zgodnych, ich własności geometrycznych i powiązanych nierówności maksymalnych. Waga tego tematu wynika z jego istotnych zastosowań w teorii prawdopodobieństwa, mikroekonomii i statystyce: wektory zgodne pojawiają się w sytuacji, gdy szanse danego zdarzenia losowego liczone są równocześnie na podstawie różnych źródeł informacji. Niniejsza rozprawa rozwija podwaliny teorii rozkładów zgodnych oraz wprowadza szereg efektywnych narzędzi dostosowanych do kontekstu optymalizacji.

Wyniki przedstawione w rozprawie opierają się na pięciu opublikowanych artykułach oraz jednej pracy przyjętej do druku. Ich prezentację podzieliliśmy tematycznie na dwie rozłączne części: pierwsza przedstawia wyniki dla rozkładów dwuwymiarowych, część druga dotyczy rozkładów wielowymiarowych.

W części pierwszej skoncentrujemy się na strukturze wypukłej rozkładów zgodnych oraz ogólnej charakteryzacji odpowiadających jej punktów ekstremalnych. Później poddamy też badaniu nośniki miar ekstremalnych ze skończoną liczbą atomów oraz wskażemy przykład rozkładu ekstremalnego o nieprzeliczalnym nośniku i bez atomów. Na zakończenie udowodnimy hipotezę Burdzego–Pitmana o maksymalnym rozrzucie rozkładów zgodnych i niezależnych.

Druga część pracy dotyczy pewnych pytań otwartych związanych z maksymalnym rozrzutem wielowymiarowych rozkładów zgodnych. Zaproponujemy tam użyteczną technikę symetryzacji i omówimy jej podstawowe własności. Jako jej najważniejsze zastosowanie uogólnimy zaś dwuwymiarową nierówność Burdzego i Pala, a zatem w szczególności podamy jej alternatywne uzasadnienie. Zbadamy również związek pomiędzy rozkładami zgodnymi a klasyczną teorią martyngałów, oraz pokażemy nową wersję nierówności maksymalnej Dooba dla pewnych "drzewiastych" filtracji.

Dominującą cechą tej pracy jest wysoce kombinatoryczny charakter rozumowania. Będziemy wielokrotnie korzystać z argumentu dyskretyzacji oraz techniki kolejnych przekształceń i redukcji. Podejście to umożliwi nam skorzystanie z powiązań między rozkładami zgodnymi, teorią grafów i kombinatoryczną teorią macierzy.

Słowa kluczowe: rozkłady zgodne, opinie ekspertów, możliwe rozkłady uzasadnionych przekonań, przeciwstawne przewidywania, maksymalny rozrzut, prawdopodobieństwo warunkowe, wspólny rozkład warunkowych wartości oczekiwanych, nierówności losowe.

Klasyfikacja AMS MSC 2020: 60C05, 60E05, 60E15, 62E10.

Acknowledgements

I am deeply grateful to my advisor, Professor Adam Osękowski,
for guidance, invaluable help and inspiration.

I would like to express my gratitude to Professor John M. Noble,
for pointing me to this great research topic and encouragement.

Finally, I would like to thank Maja Szlenk,
for wide-ranging support and enthusiasm.

Contents

Chapter 1. Introduction 9
1.1. Definitions, Examples 9
1.2. Maximal spread of coherent vectors 14
1.3. Connections with combinatorics 16
1.4. Relations to other subfields 18
1.5. Organization, Contributions 19

Part 1. Coherent distributions on the square 21

Chapter 2. Extreme points and asymptotics 23
2.1. Coherent measures, Representations 26
2.2. Extreme points with a finite support 28
2.3. Asymptotic estimate 36
2.4. Concluding remarks 44

Chapter 3. Extreme distributions with no atoms 47
3.1. Main tools and outline of the proof 48
3.1.1. On the idea behind the special measure m_r 49
3.1.2. Some basic facts from the theory of dynamical systems 50
3.2. Proof of uniqueness 52
3.3. Proof of minimality 55

Chapter 4. Proof of the Burdzy–Pitman conjecture 57
4.1. Relation with the Erdős "spread bounding theorem" 58
4.2. Number of high degree differences in bipartite graphs 59
4.3. Proof of the Burdzy–Pitman conjecture 61

Part 2. Multivariate coherent distributions 65

Chapter 5. Maximal difference among expert's opinions 67
5.1. Symmetrization procedure 68
5.2. Preparatory reductions 70
5.3. Combinatorial optimization 73

Chapter 6. Contradictory predictions with multiple agents 83
6.1. Basic reductions, Symmetrization 84
6.2. Proof of Lemma 6.1 88
6.2.1. Tree structure 89
6.2.2. Properties of $\mathcal{T}_{(H,L)}$ 91
6.2.3. Dynamic programming 93

Chapter 7. Doob's estimate for coherent random variables 97
7.1. Proof of Theorem 7.4 99
7.2. L^p estimates 101
7.3. Proof of Theorem 7.2 105

Bibliography 107

CHAPTER 1

Introduction

1.1. Definitions, Examples

The primary purpose of this dissertation is to increase our current understanding of the so-called coherent distributions, a certain family of special probability measures on $[0,1]^n$. Let us start with the precise definition.

DEFINITION 1.1 (Dawid, DeGroot, Mortera, 1995). *Let μ be a probability measure on the n-dimensional cube $[0,1]^n$. We say that the measure μ is* **coherent**, *if it is the joint distribution of an n-variate random vector $(X_1, X_2, \ldots, X_n)$ defined on some arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that we have the almost sure identities*

$$X_i = \mathbb{P}(A|\mathcal{G}_i), \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

for some measurable event $A \in \mathcal{F}$ and some sequence $\mathcal{G}_1$, $\mathcal{G}_2$, $\ldots$, $\mathcal{G}_n$ of sub-σ-fields of $\mathcal{F}$. In such a case, a vector $(X_1, X_2, \ldots, X_n)$ is also called coherent.

The class of all coherent distributions on $[0,1]^n$ will be denoted by $\mathcal{C}_n$. For the sake of convenience (and with a slight abuse of notation), we will frequently write $X \in \mathcal{C}_n$ to indicate that the distribution of the random vector $X = (X_1, \ldots, X_n)$ is coherent. From now on, we will drop the subscript "2" and write $\mathcal{C}$ instead of $\mathcal{C}_2$ for the foundational class of all two-dimensional coherent distributions on $[0,1]^2$. We also write $\mathcal{C}_{n,\mathcal{I}}$ ($\mathcal{C}_{\mathcal{I}}$ when $n = 2$) for the family of coherent measures with product structure – distributions of all coherent vectors with jointly independent components.

A few remarks on the historical perspective are appropriate. The deliberate research on coherent distributions (under a different name) was started in the eighties, see the pioneering results [**28, 29**]. However, slightly less direct roots of this concept can be probably found in even earlier papers, see [**48, 57, 66**]. The term "coherent" was first introduced in 1995, in the landmark article by Dawid, DeGroot and Mortera – although focused rather on statistical applications, [**23**] contains also some initial insight into the structure of coherent distributions. Later on, coherent distributions appeared in disguise in many places in the literature; see Section 1.4 for more details.

Finally, during the last five years, the transition into a more theoretical investigation was very rapid; confront the works by Burdzy, Pal, Pitman and Zhu [**9, 10, 75**]. In particular, the second of these papers contains an up-to-date list of crucial references and formulates a number of major open problems. Since then, this active topic was successfully investigated by several authors [**3, 4, 15, 52**]. The main aim of the thesis is to present the further development of this line of research. We will introduce some novel methods and apply them to selected problems from [**10**].

There are two main types of problems we will be studying in this thesis.

• *Structural properties of coherent distributions.* There are many appealing problems concerning the structure of coherent distributions. Let us present a few questions in this direction, some of which will be studied in the thesis. For instance, given a probability measure μ on $[0,1]^n$, is there a comprehensible algebraic characterization of the inclusion $\mu \in \mathcal{C}_n$? Similarly, assuming that μ is finitely supported, is there an efficient algorithm to check whether $\mu \in \mathcal{C}_n$? Does the answer depend on dimension n? Another class of questions concerns the geometry of coherent distributions. What can be said about the geometry of the support? For a given Borel set $B \subseteq [0,1]^n$, does there exist $\mu \in \mathcal{C}_n$ such that $\text{supp}(\mu) = B$? The main topic we discuss is the following. The class $\mathcal{C}_n$ can be shown to be a convex and weak*-compact subset of the family of all probability measures on $[0,1]^n$. What can be said about the extremal points (characterization, type of distribution, properties of the support, etc.)? For example, are they always discrete? Can they be absolutely continuous with respect to the Lebesgue's measure?

• *Sharp inequalities for coherent random vectors.* The second topic concerns various estimates which arise naturally in the context of coherent distributions. In general, for a fixed function $\Phi : [0,1]^n \to \mathbb{R}_+$, the goal is to evaluate $\sup \mathbb{E}\Phi(X_1, X_2, \ldots, X_n)$, where the supremum is taken over all probability models satisfying condition (1.1) and other auxiliary restrictions (such as independence of σ-fields $\mathcal{G}_1, \ldots, \mathcal{G}_n$, or fixed probability of an event A). For example, one can ask about tight inequalities for the maximal and minimal functions of the sequence $X = (X_1, \ldots, X_n) \in \mathcal{C}_n$, i.e., the bounds of the type $\mathbb{E}\max_{1\le k\le n} X_k \le c_1(n, \mathbb{E}X_1)$ and $\mathbb{E}\min_{1\le k\le n} X_k \ge c_2(n, \mathbb{E}X_1)$, for some constants c_1, c_2 depending only on the parameters indicated. Of course, there are many classes of inequalities in this setting, but we will be mostly interested in good estimates for the spread of X, i.e., for the difference of the maximal and the minimal functions. We will elaborate more on this problem in the next section. Here, note that the study of extremal coherent distributions, mentioned in the previous research topic, remains very important in the context of optimization.

From a purely mathematical point of view, the definition of coherent distributions is very short and simple, and the problems formulated above are very natural. As we will see, the successful treatment of the topic will require developing various novel combinatorial and analytic arguments – the introduction of these methods can be regarded as another important contribution of the thesis. Interestingly, our study will also reveal some unexpected connections of coherent distributions with other areas of mathematics: theory of dynamical systems, extremal problems in graph theory, boundedness properties of Hardy-Littlewood maximal operator on tree measure spaces. It is also worth repeating here that the progress in this direction is not of theoretical value only: as already noted above, coherent distributions arise frequently in many applications, also outside mathematics. See Section 1.4 below.

Let us proceed with the discussion on three distinctive examples. First, coherent distributions appear commonly whenever some survey or experimental data is being organized according to two or more qualitative variables. See e.g. [**34, 48, 55**].

EXAMPLE 1.1. Consider a hypothetical study: 500 volunteers are divided into two cohorts (**A** and **B**, not necessarily of equal size) and consequently vaccinated with two different doses of medicine. Later on, all participants are tested to see whether they have developed a proper immune response – see Table 1.

	VACCINATED		IMMUNIZED	
	A	**B**	**A**	**B**
MEN	200	50	150	25
WOMEN	100	150	70	100

TABLE 1. Hypothetical data set.

Let us pick a random volunteer. From the viewpoint of applications, it is natural to study the interplay between the random variables $S =$ sex, $C =$ cohort and the event $E = \{\text{immunized}\}$. For instance, one can distinguish the variables $X_1 = \mathbb{P}(E), X_2 = \mathbb{P}(E|S), X_3 = \mathbb{P}(E|C)$ and $X_4 = \mathbb{P}(E|S,C)$ and ask about their joint behavior. Note that $\mathbb{P}(E) = \mathbb{P}(E|\{\emptyset, \Omega\})$, so X_1 is a conditional probability of event E as well, and hence $(X_1, X_2, X_3, X_4) \in \mathcal{C}_4$.

It is clear that this example can be modified and applied in numerous directions. The data set can be very large, it can be arranged in various ways, and any abstract result on the structure of coherent distributions can bring a lot of valuable insight and practical information. △

The next example is a bit longer: we first need to develop some helpful intuition behind the notions of "information" and "knowledge".

EXAMPLE 1.2. A point ω is placed at random on the interval $\Omega = [0,1]$. Let A be the event that $\omega \in [1/8, 3/8] \cup [6/8, 7/8]$. Suppose that we have no extra information about ω. Hence, our best estimate on the chances of A is simply $X_1 = \mathbb{P}(A) = 3/8$. Now, assume that we are also informed whether $\omega \le 1/2$ or $\omega > 1/2$. This additional knowledge can be formally described using the σ-field $\mathcal{G}'$ generated by a partition $\Omega = [0, 1/2] \cup (1/2, 1]$. Consequently, our updated belief is

$$X_2(\omega) = \mathbb{P}(A|\mathcal{G}')(\omega) = \begin{cases} 1/2 & \text{if } \omega \in [0, 1/2], \\ 1/4 & \text{if } \omega \in (1/2, 1]. \end{cases}$$

In a similar way, let $\mathcal{G}''$ be the σ-field related to the division of $[0,1]$ into 8 consecutive subintervals of equal length – see Figure 1. We can clearly write $X_3 = \mathbb{P}(A|\mathcal{G}'') = \mathbb{1}_A$, so this last partition already provides a complete information about the odds of A.

Unfortunately, this appealing interpretation of "knowledge" breaks down in the limit. Indeed, set $\mathcal{P} = \{\{x\} : x \in [0,1]\}$, $\mathcal{G}''' = \sigma(\mathcal{P})$ and put $X_4 = \mathbb{P}(A|\mathcal{G}''')$. Consistently, partition $\mathcal{P}$ contains full information on the location of point ω. Yet, the σ-field $\mathcal{G}'''$ is nearly trivial: for every $G \subseteq [0,1]$ we get $G \in \mathcal{G}'''$ if and only if G or $\Omega \setminus G$ is at most countable. Accordingly, we have $\mathbb{P}(G) = 0$ in the first case and $\mathbb{P}(G) = 1$ in the second. One can verify that $X_4 = X_1 = 3/8$ almost surely. Surprisingly, there is no new knowledge about the event A in the σ-field $\mathcal{G}'''$ (not contained in $\{\emptyset, \Omega\}$).

By the $\mathcal{G}'''$-measurability of X_4, we have $F_{X_4}(t) = \mathbb{P}(X_4 \le t) \in \{0, 1\}$ for all $t \in [0,1]$. Denote the number $t_0 = \inf\{t \in [0,1] : F_{X_4}(t) = 1\}$. The CDF function F_{X_4} is nondecreasing and right-continuous, and so we can write

$$\mathbb{P}(X_4 = t_0) = \mathbb{P}(X_4 \le t_0) - \mathbb{P}(X_4 < t_0) = 1 - 0 = 1.$$

Now, we get $t_0 = \mathbb{E}X_4 = \mathbb{E}\left[\mathbb{E}(\mathbb{1}_A|\mathcal{G}''')\right] = \mathbb{E}\mathbb{1}_A = 3/8$.

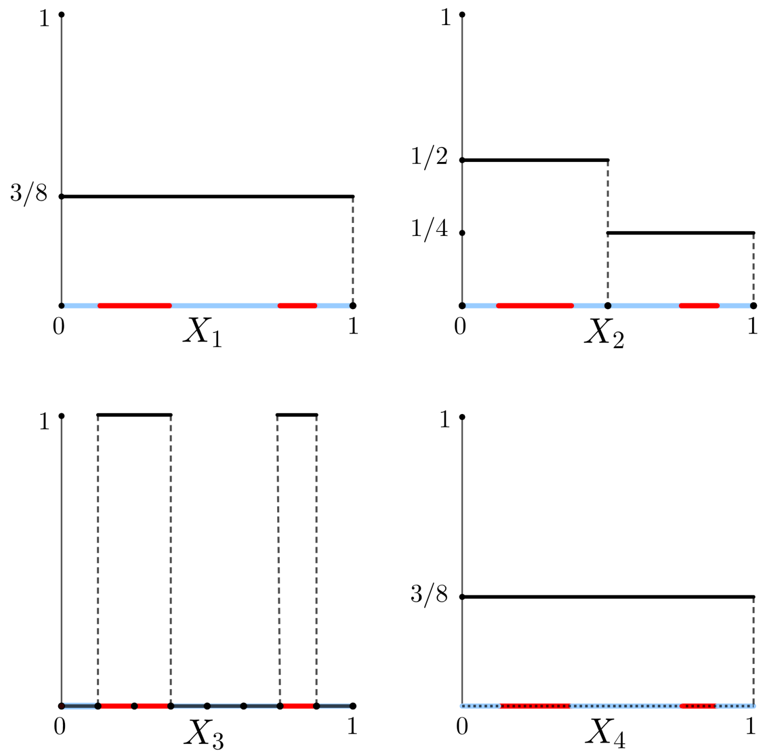

FIGURE 1. Nested partitions of $[0, 1]$ and varying likelihood of A (red).

As emphasized in [**6, 31**], the interpretation of σ-fields as "informational content" is sometimes misleading. Nevertheless, as indicated in [**45**], this useful heuristic is well justified for finite or countable partitions and remains crucial for most applications.

Now we come to the heart of the example. The above discussion clearly motivates a particularly productive perspective on coherent distributions. Namely, suppose that a group of n entities (experts, players, economic agents) are asked to provide their personal estimates on the likelihood of some uncertain event A. At first, they do agree upon some basic model $(\Omega, \mathcal{F}, \mathbb{P})$. Later on, their unified beliefs are updated with distinct sources of information – assume that the knowledge of j-th person is represented by the σ-algebra $\mathcal{G}_j$, $j = 1, 2, \ldots, n$. This gives rise to the coherent vector $X = (X_1, \ldots, X_n)$, whose coordinates (1.1) correspond to the opinions of the above entities. From this standpoint, the problem of estimating divergence of X becomes very important: any concentration-type bounds can be further used to extract some additional information about the event A. $\triangle$

More explicitly, let us briefly discuss below a very specific and elementary example.

EXAMPLE 1.3. ***A**lice*, ***B**ob*, ***C**indy*, and ***D**ave* are playing cards. The short 20-card deck (from 10s up) is used. Each player is dealt two cards, face down – see Table 2. The remaining 12 cards remain hidden. In what follows, all four players bet whether there are all four aces among their 8 cards.

ALICE	A♠	A♥	*plays fair*	$X_1 \approx 0.10$
BOB	A♦	K♠	*spying on Alice*	$X_2 = 0.25$
CINDY	J♣	10♣	*plays fair*	$X_3 < 0.01$
DAVE	A♣	Q♦	*marked the aces*	$X_4 = 1.00$

TABLE 2. Sample shuffle and players' private knowledge.

Now we assume that ***A**lice* and ***C**indy* play honestly – they only know their own cards, ***B**ob* is spying on ***A**lice* and knows her hand, and ***D**ave* is a crook – he marked all four aces with a secret sign. Formally, let the random variables a_i, b_i, c_i, d_i stand for the i-th cards ($i = 1, 2$) obtained by the corresponding players, and let $E = \big\{\{$**A♠A♥A♦A♣**$\} \subseteq \{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2\}\big\}$. The knowledge of the players is represented by the σ-algebras $\mathcal{G}_1 = \sigma(a_1, a_2)$, $\mathcal{G}_2 = \sigma(a_1, a_2, b_1, b_2)$, $\mathcal{G}_3 = \sigma(c_1, c_2)$ and a slightly more complicated algebra $\mathcal{G}_4$, which contains $\mathbb{1}_E$. We have $(X_1, X_2, X_3, X_4) \in \mathcal{C}_4$. △

In the remaining part of this section we discuss two simple facts. First, we present a well-known alternative characterization of coherent distributions – see e.g. [**10, 23**]. We sketch the elementary proof for the sake of completeness.

PROPOSITION 1.1. *Let μ be a probability measure on the n-dimensional cube $[0,1]^n$. The measure μ is a coherent distribution if and only if it is the joint distribution of an n-variate random vector $(X_1, X_2, \ldots, X_n)$ such that*

$$X_i = \mathbb{E}(Z|X_i) \quad \text{for all } i = 1, 2, \ldots, n, \quad \text{almost surely,} \tag{1.2}$$

for some random variable Z with $0 \le Z \le 1$.

PROOF. We check both implications separately.

The implication "⇒". In accordance with Definition 1.1, there exists a random vector $(X_1, X_2, \ldots, X_n) \sim \mu$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and satisfying

$$X_i = \mathbb{E}(\mathbb{1}_A|\mathcal{G}_i) \quad \text{for all } i = 1, 2, \ldots, n, \quad \text{almost surely,} \tag{1.3}$$

for some $A \in \mathcal{F}$ and sub-σ-fields $\mathcal{G}_1$, $\mathcal{G}_2$, $\ldots$, $\mathcal{G}_n \subseteq \mathcal{F}$. Clearly, we have $0 \le \mathbb{1}_A \le 1$ and it is enough to set $Z = \mathbb{1}_A$. Due to (1.3), random variables X_i are $\mathcal{G}_i$-measurable, so $\sigma(X_i) \subseteq \mathcal{G}_i$ for $1 \le i \le n$ and hence

$$X_i = \mathbb{E}(X_i|X_i) = \mathbb{E}[\mathbb{E}(Z|\mathcal{G}_i)|X_i] = \mathbb{E}(Z|X_i).$$

The implication "⇐". Suppose that $(X_1, X_2, \ldots, X_n) \sim \mu$ satisfies (1.2) for a certain $0 \le Z \le 1$. Therefore, enlarging the probability space if necessary, we can find a uniform random variable $U \sim \mathcal{U}([0,1])$ which is jointly independent of $(X_1, \ldots, X_n, Z)$. Then we can write

$$X_i = \mathbb{E}(Z|X_i) = \mathbb{E}\Big[\mathbb{E}(\mathbb{1}\{U \le Z\}|\sigma(Z, X_i))\Big|\mathcal{G}_i\Big] = \mathbb{E}(\mathbb{1}\{U \le Z\}|\mathcal{G}_i),$$

where $\mathcal{G}_i = \sigma(X_i)$, $i = 1, 2, \ldots, n$. □

In our considerations below, we will frequently make use of the following convenient discretization. This will allow us to benefit from the combinatorial nature of coherent distributions.

Proposition 1.2. *Let n, m be two positive integers. Then, for every:*

(i) $(X_1, X_2, \dots, X_n) \in \mathcal{C}_n$ *there exists* $(\widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_n) \in \mathcal{C}_n$,

(ii) $(X_1, X_2, \dots, X_n) \in \mathcal{C}_{n,\mathcal{I}}$ *there exists* $(\widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_n) \in \mathcal{C}_{n,\mathcal{I}}$,

such that each $\widetilde{X}_i$ takes at most m different values and

$$|X_i - \widetilde{X}_i| \le 1/m \quad \text{for all } i = 1, 2, \dots, n, \quad \text{almost surely.} \tag{1.4}$$

Proof. To prove (i), fix any $(X_1, X_2, \dots, X_n) \in \mathcal{C}_n$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a fitting probability space. As follows from the proof of Proposition 1.1, we can write

$$X_i = \mathbb{E}(\mathbb{1}_A | X_i) \quad \text{for all } i = 1, 2, \dots, n, \quad \text{almost surely,}$$

for some $A \in \mathcal{F}$. Next, let $\mathcal{G}_{i,m}$ be the σ-field generated by the partition

$$\mathcal{P}_{i,m} = \left\{ \left\{ X_i \in \left[0, \frac{1}{m}\right] \right\}, \left\{ X_i \in \left(\frac{1}{m}, \frac{2}{m}\right] \right\}, \dots, \left\{ X_i \in \left(\frac{m-1}{m}, 1\right] \right\} \right\},$$

and set $\widetilde{X}_i = \mathbb{E}(X_i | \mathcal{G}_{i,m}) = \mathbb{E}[\mathbb{E}(\mathbb{1}_A | X_i) | \mathcal{G}_{i,m}] = \mathbb{E}(\mathbb{1}_A | \mathcal{G}_{i,m})$, for $1 \le i \le n$. Therefore, we clearly have $(\widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_n) \in \mathcal{C}_n$. Since every $\mathcal{P}_{i,m}$ is an m-element partition, each variable $\widetilde{X}_i$ takes at most m different values. The condition (1.4) is just a direct consequence of this construction ($\widetilde{X}_i$ is constant on every element of $\mathcal{P}_{i,m}$).

To show (ii), we proceed in the same way. Note that the σ-fields $\mathcal{G}_{1,m}, \mathcal{G}_{2,m}, \dots, \mathcal{G}_{n,m}$ are now jointly independent. As a result, we get $(\widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_n) \in \mathcal{C}_{n,\mathcal{I}}$. □

1.2. Maximal spread of coherent vectors

As we have already mentioned above, one of the primary challenges addressed in this dissertation is to determine the optimal estimates associated with coherent random vectors. The following milestone result can be extracted from [**29**].

Theorem 1.1 (Dubins, Pitman, 1980). *For $n \ge 2$, let $(X_1, \dots, X_n) \in \mathcal{C}_n$ be a coherent vector with $\mathbb{E}X_1 = p$. Then we have*

$$\mathbb{E} \max_{1 \le i \le n} X_i \le \frac{p(n-p)}{1 + p(n-2)}. \tag{1.5}$$

It turns out that the constant on the right cannot be improved, as the following construction from [**28**] shows.

Example 1.4. For $n \ge 2$ and $p \in [0, 1]$, the (n, p)-daisy is a coherent random vector corresponding to the following partition of space Ω into $n + 1$ disjoint subsets:

· a single event A, *the center of the daisy*, with $\mathbb{P}(A) = p$,

· n consecutive events $B_1, B_2, \dots, B_n$, *the petals*, with

$$\mathbb{P}(B_i) = \frac{1-p}{n}, \quad \text{for } i = 1, 2, \dots, n.$$

See Figure 2. For $1 \le i \le n$, let $\mathcal{G}_i$ be the σ-field generated by $A \cup B_i$, and define $X_i = \mathbb{P}(A | \mathcal{G}_i)$. Then $(X_1, \dots, X_n) \in \mathcal{C}_n$ and we have the identity

$$X_i(\omega) = \begin{cases} 0 & \text{if } \omega \notin A \cup B_i, \\ \dfrac{np}{(n-1)p + 1} & \text{if } \omega \in A \cup B_i. \end{cases}$$

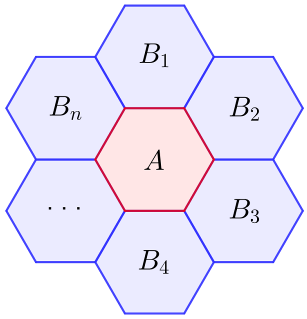

FIGURE 2. The (n,p)-daisy with equal (blue) petals and a (red) center A.

Now, let $(\widetilde{X}_1,\dots,\widetilde{X}_{n-1})$ be the $(n-1,p)$-daisy, and complement it with the variable $\widetilde{X}_n = \mathbb{1}_A = \mathbb{P}(A|\sigma(A))$. Then $(\widetilde{X}_1,\dots,\widetilde{X}_n) \in \mathcal{C}_n$ and so, by the formula for $X_i(\omega)$,

$$\begin{aligned}\mathbb{E}\max_{1\le i\le n}\widetilde{X}_i &= \sum_{j=1}^{n-1}\int_{B_j}\max_{1\le i\le n-1}\widetilde{X}_i\mathrm{d}\mathbb{P} + \mathbb{P}(A)\\ &= (n-1)\cdot\int_{B_1}\widetilde{X}_1\mathrm{d}\mathbb{P} + p = \frac{p(n-p)}{1+p(n-2)}.\end{aligned}$$

This shows that the estimate (1.5) is sharp. △

Notably, in the last chapter of this thesis, we will extend the estimate of the above example to a certain sharp Doob-type inequality in L^p. However, we will be mostly interested in the estimates for the maximal *spread* between the coherent opinions. More specifically, we will concentrate on the following functionals.

DEFINITION 1.2. *For* $n \in \{2,3,\dots\}$ *and* $\delta \in (1/2,1]$, *define*

$$\mathbf{P}(n,\delta) = \sup_{(X_1,X_2,\dots,X_n)\in\mathcal{C}_n}\mathbb{P}\Big(\max_{1\le i<j\le n}|X_i - X_j| \ge \delta\Big),$$

$$\mathbf{P}_{\mathcal{I}}(n,\delta) = \sup_{(X_1,X_2,\dots,X_n)\in\mathcal{C}_{n,\mathcal{I}}}\mathbb{P}\Big(\max_{1\le i<j\le n}|X_i - X_j| \ge \delta\Big).$$

Furthermore, for $n \in \{2,3,\dots\}$ *and* $\alpha > 0$, *let*

$$\mathbf{E}(n,\alpha) = \sup_{(X_1,X_2,\dots,X_n)\in\mathcal{C}_n}\mathbb{E}\max_{1\le i<j\le n}|X_i - X_j|^\alpha,$$

$$\mathbf{E}_{\mathcal{I}}(n,\alpha) = \sup_{(X_1,X_2,\dots,X_n)\in\mathcal{C}_{n,\mathcal{I}}}\mathbb{E}\max_{1\le i<j\le n}|X_i - X_j|^\alpha.$$

The above functions, in a mathematically precise way, determine what should be understood as the maximum discrepancy of multidimensional coherent distributions. Let us emphasize that all the above suprema are taken over all probability spaces $(\Omega,\mathcal{F},\mathbb{P})$, all measurable events $A \in \mathcal{F}$ and all sub-σ-fields $\mathcal{G}_1,\mathcal{G}_2,\dots,\mathcal{G}_n \subseteq \mathcal{F}$ such that (1.1) holds.

EXAMPLE 1.5. As verified in [**10**], we have the identity

$$\mathbf{E}(2,1) = \sup_{(X,Y)\in\mathcal{C}}\mathbb{E}|X-Y| = \frac{1}{2}. \tag{1.6}$$

Indeed, for any $(X,Y) \in \mathcal{C}$, we have $|X-Y| = 2\max\{X,Y\} - X - Y$. Set $\mathbb{E}X = p$. Accordingly, by Theorem 1.1 for $n=2$, we can write $\mathbb{E}|X-Y| \le 2p(1-p) \le 1/2$. Conversely, consider $X' = \mathbb{1}_A$ and $Y' = \mathbb{P}(A)$ for an arbitrary event A with $\mathbb{P}(A) = \frac{1}{2}$. In this case $(X',Y') \in \mathcal{C}_{\mathcal{I}}$ and $\mathbb{E}|X'-Y'| = 1/2$. This finishes the proof of (1.6). △

Surprisingly, since $(X',Y') \in \mathcal{C}_{\mathcal{I}}$, we also obtain $\mathbf{E}(2,1) = \mathbf{E}_{\mathcal{I}}(2,1)$. The subsequent, more general conclusion, was stated independently and concurrently in [**4**] and [**15**].

Example 1.6. For any $\alpha \in [0,2]$, we have

$$\mathbf{E}(2,\alpha) = \mathbf{E}_{\mathcal{I}}(2,\alpha) = 2^{-\alpha}. \tag{1.7}$$

Equalities (1.7) are easily proven using the L^2-norm and Hilbert space framework. Interestingly, the bound $2^{-\alpha}$ does not hold for exponents $\alpha \ge 3$. △

In our considerations below, we will obtain a number of sharp results for $\mathbf{P}(n,\delta)$ and $\mathbf{E}(n,\alpha)$. See Section 1.5 for the summary.

1.3. Connections with combinatorics

Coherent distributions and discrete mathematics are closely related. As an example, a very important class of coherent vectors is that associated with bipartite graphs.

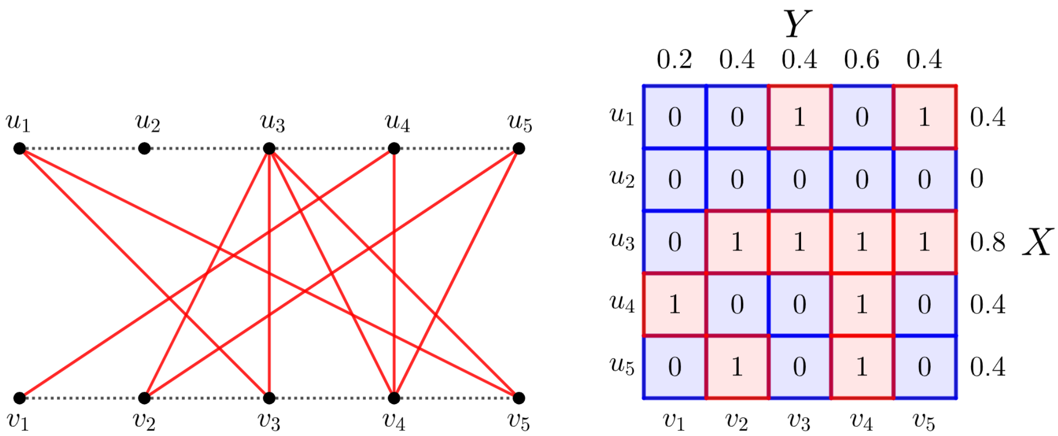

Figure 3. A bipartite graph G, a $(0,1)$-matrix A, an associated $(X,Y) \in \mathcal{C}_{\mathcal{I}}$.

Example 1.7. A bipartite graph $G = (U,V,E)$ consists of two disjoint parts

$$U = \{u_1, u_2, \ldots, u_n\}, \quad V = \{v_1, v_2, \ldots, v_m\}, \quad n, m \in \mathbb{Z}_+,$$

and a subset $E \subseteq U \times V$. Generic elements $u_i, v_j \in U \cup V$ are referred to as "vertices" of G and pairs $(u_i, v_j) \in E$ are interpreted as "edges" (or intervals) connecting them. Then, we can represent such a bipartite graph G with an $n \times m$ binary matrix A: set $a_{i,j} = \mathbb{1}\{(u_i, v_j) \in E\}$ for all $u_i \in U$, $v_j \in V$. Next, let $(\alpha_i)_{i=1}^n$ and $(\beta_j)_{j=1}^m$ denote the "degree sequences" of G:

$$\alpha_i = \deg(u_i) = \sum_{j=1}^{m} a_{ij}, \quad \text{for } 1 \leq i \leq n, \tag{1.8}$$

$$\beta_j = \deg(v_j) = \sum_{i=1}^{n} a_{ij}, \quad \text{for } 1 \leq j \leq m. \tag{1.9}$$

Pairs (α, β) obeying (1.8) and (1.9) for some bipartite graph G are called "bigraphic".

Consider the probability space $(U \times V, \mathcal{P}(U \times V), \mathbb{P}_{\#})$, where $\mathbb{P}_{\#}(A) = \frac{|A|}{n \cdot m}$ for all $A \in \mathcal{P}(U \times V)$ and $\mathcal{P}$ denotes the power set. Now, let the σ-fields $\mathcal{G}_1$ and $\mathcal{G}_2$ coincide with the parallel division of $(0,1)$-matrix A into rows and columns, correspondingly:

$$\mathcal{G}_1 = \sigma(\{u_1\} \times V, \ldots, \{u_n\} \times V), \quad \mathcal{G}_2 = \sigma(U \times \{v_1\}, \ldots, U \times \{v_m\}).$$

Accordingly, for random $\omega = (u_\omega, v_\omega) \in U \times V$ (distributed according to $\mathbb{P}_{\#}$), put

$$X(\omega) = \frac{\deg(u_\omega)}{m} \quad \text{and} \quad Y(\omega) = \frac{\deg(v_\omega)}{n}.$$

Stated differently, we have $(X, Y) = (\mathbb{P}(E|\mathcal{G}_1), \mathbb{P}(E|\mathcal{G}_2))$, so $(X, Y) \in \mathcal{C}_{\mathcal{I}}$, as the random variables X and Y are also independent – see Figure 3. $\triangle$

For a deeper comprehension of Example 1.7 (and its quite surprising generality), let us additionally compare the following two important results from distinct subfields.

THEOREM 1.2 (Burdzy, Pitman, 2020, [**10**]). *Let μ be a probability measure on $[0,1]^2$. The measure μ is a coherent distribution if and only if it is the joint distribution of a bivariate random vector (X, Y) such that $\mathbb{E}X = \mathbb{E}Y = p$ for some $0 \leq p \leq 1$, and*

$$\mathbb{E}[X \mathbb{1}\{X \in B\}] + \mathbb{E}[(1 - Y)\mathbb{1}\{Y \in C\}] \geq \mathbb{P}(X \in B, Y \in C),$$

for all Borel subsets $B, C \in \mathcal{B}([0,1])$.

THEOREM 1.3 (Ford, Fulkerson, 1962, [**35**]). *Let $(\alpha_i)_{i=1}^n, (\beta_j)_{j=1}^n$ be two sequences of natural numbers. The pair (α, β) is bigraphic if and only if $\sum_{i=1}^n \alpha_i = \sum_{j=1}^n \beta_j$ and*

$$\sum_{i \in I} \alpha_i + \sum_{j \in J} (n - \beta_j) \geq |I| \cdot |J|,$$

for all $I, J \subseteq \{1, 2, \ldots, n\}$.

Observe that Theorems 1.2 and 1.3 are remarkably similar – they express the same idea in different settings. In general, structural properties of $\mathcal{C}$ and $\mathcal{C}_{\mathcal{I}}$ are strongly connected with graph theory: the analysis of $\mathcal{C}_n$ and $\mathcal{C}_{n,\mathcal{I}}$ correlates with the study of n-partite hypergraphs. Another example of this phenomenon can be found in [**70**].

The interplay between coherent distributions and classical combinatorial analysis is significantly deeper. Without going into any details, we will just highlight some of the most important connections. In this context, the topics of threshold graphs [**26, 59**], graphic sequences [**32, 68**] and various variants of the Gale–Ryser theorem [**7, 57, 66**] are of particular interest. In a wider perspective, associations with combinatorial matrix theory [**8**], discrete tomography [**44**], geometry of transportation polytopes [**56**], are also very influential.

1.4. Relations to other subfields

A number of concurrent topics in probability theory, statistics and microeconomics provides another impulse for further exploration of coherent distributions.

Probability theory

• *Martingale theory* – a process $\{M_t\}_{t\in\mathcal{T}}$ indexed by a partially ordered set $\langle\mathcal{T},\preceq\rangle$ is called a martingale, if $\mathbb{E}|M_s| < \infty$ and $\mathbb{E}(M_s|\mathcal{F}_t) = M_t$ for all $t \preceq s$ (where $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ is a fixed filtration, i.e., a collection of sub-σ-fields such that $\mathcal{F}_t \subseteq \mathcal{F}_s$ when $t \preceq s$). Now, let $(X_1, X_2, \ldots, X_n) \in \mathcal{C}_n$ satisfy the equations $X_i = \mathbb{E}(\mathbb{1}_A|\mathcal{G}_i)$, $i = 1, 2, \ldots, n$, just as in (1.1). Set $\mathbb{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n\}$ and consider a pair $\langle\mathbb{G},\subseteq\rangle$, where the relation "$\subseteq$" denotes the standard inclusion order. Therefore, we clearly have $X_i = \mathbb{E}(X_j|\mathcal{G}_i)$ for all $\mathcal{G}_i \subseteq \mathcal{G}_j$ and so $\{X_i\}_{\mathcal{G}_i\in\mathbb{G}}$ is a set-indexed martingale. For that reason, research on coherent distributions may offer some fresh insights into a still vague framework of set-inexed martingales; see [**17, 28, 29, 49**].

• *Marginal problems* – there are many works devoted to the study of probabilistic measures on $[0, 1]^n$ with a given family of marginal distributions. Classical research concentrates on the the question of existence and structural properties of solutions – convex geometry methods are very common and supports of extreme points (within specific convex subsets of measures) are thoroughly characterised. From a broader context, this class of problems closely resembles the combinatorial matrix theory and expands its results. In our further inspection of $\mathcal{C}$ (coherent distributions on the square) we will heavily rely on similar ideas; see [**36, 40, 46, 69**].

Statistics

• *Forecast calibration* – often, for a fixed event A (like "it will rain tomorrow"), we need to verify several predictions $F_1, F_2, \ldots, F_n$ (say "different weather forecasts"). Then, we exploit the empirical joint distributions of all pairs $(\mathbb{1}_A, F_i)$, $i = 1, 2, \ldots, n$. In this setup, the conditional expected value $\mathbb{E}(\mathbb{1}_A|F_i = f_i)$ is an average frequency of A given a precise value $f_i \in [0, 1]$ of the i-th forecast F_i. Put $\Delta_i = |\mathbb{E}(\mathbb{1}_A|F_i) - F_i|$, $1 \le i \le n$. For a credible prediction F_i we anticipate that Δ_i is small. Moreover, when $\Delta_i = 0$ almost surely, we say that the forecast F_i is "perfectly calibrated". In most theoretical models, we get $(F_1, F_2, \ldots, F_n) \in \mathcal{C}_n$; see [**22, 24, 60, 67**].

• *Optimal combining* – it is occasionally desired to integrate multiple individual forecasts $F_1, F_2, \ldots, F_n$ into a unified and reliable prediction F_0. We expect a formula in the form $F_0 = \Phi(F_1, F_2, \ldots, F_n)$ for some continuous function $\Phi : [0, 1]^n \to [0, 1]$. This problem is especially present when we are denied access to the "raw data" used by forecasters. The theoretical literature on this challenge reveals numerous notable answers – most of the solutions depend on various distributional assumptions and auxiliary restrictions on Φ. For a convenient idealization, it is usually assumed that $(F_1, F_2, \ldots, F_n) \in \mathcal{C}_n$ (all F_i are perfectly calibrated); see [**23, 25, 37, 64**].

Microeconomics

• *Posterior feasibility* – in a dynamic learning structure, every person from a group of n agents (nodes of a network) starts off with a common prior knowledge. Next, across $T \in \mathbb{N} \cup \{+\infty\}$ rounds, agents receive non-identical sets of random signals. For each node i and every round t, let $\mathcal{G}_t^i$ denote the σ-field generated by all data available to agent i at the end of round t – we have $\mathcal{G}_1^i \subseteq \mathcal{G}_2^i \subseteq \cdots \subseteq \mathcal{G}_T^i$. Then, for any fixed event E, a coherent vector $(\mathbb{P}(E|\mathcal{G}_T^i))_{i=1}^n$ encodes the collected group beliefs on chances of E. Investigation of all feasible outcomes of such a general learning process represents a significant theoretical problem; see [**2, 4, 41, 54**].

• *Bayesian persuasion* – in many situations, an informed sender (as public authorities or corporate management) has a complete knowledge about some random process and is free to arbitrarily reveal this information. Within a game theoretic approach, given a specific payoff function, we can simulate the sender's gain from exposing particular pieces of data to different entities – thus altering their perceptions and responses. We assume that all players are well aware about the joint distribution of the model, but it is entirely up to the sender to design a dynamic learning structure responsible for the signals generating mechanism; see [**1, 3, 43, 50**].

1.5. Organization, Contributions

The work presented in this dissertation is based on five published research articles and one paper accepted for publication. For convenience, we split the thesis into two separate parts: one presenting our results on the bivariate and one for the multivariate coherent distributions. The content of each part corresponds to exactly three papers. Let us briefly describe their scope and primary contributions.

Part 1: `Coherent distributions on the square`

Chapter 2. We examine the set $\mathrm{ext}(\mathcal{C})$ of extreme points of $\mathcal{C}$ and provide its general characterisation. Furthermore, we establish several structural properties of finitely-supported elements of $\mathrm{ext}(\mathcal{C})$. We apply these results to obtain the asymptotic sharp bound $\lim_{\alpha\to\infty} \alpha \cdot \mathbf{E}(2,\alpha) = 2/e$. The contents of the chapter are based on the work

• S. Cichomski, A. Osękowski, *Coherent distributions on the square – extreme points and asymptotics*, Journal of Applied Probability, **Online first**, 1–23, **2024**, [**18**].

Chapter 3. We prove the existence of an extreme coherent distribution with the uncountable support and with no atoms. Our argument is based on classical tools and ideas from the dynamical systems theory. This unexpected connection can be regarded as an independent contribution. The contents of the chapter are based on the paper

• S. Cichomski, A. Osękowski, *On the existence of extreme coherent distributions with no atoms*, arXiv:2311.08140, **math.PR**, 1–12, **2023**, [**20**], accepted in Journal of Theoretical Probability.

Chapter 4. We confirm the Burdzy–Pitman conjecture about the maximal spread of coherent and independent vectors: for threshold $\delta \in (1/2, 1]$ we prove that $\mathbf{P}_{\mathcal{I}}(2,\delta) = 2\delta(1-\delta)$. We investigate the graph theoretic counterpart of this problem. The contents of the chapter are based on the paper

• S. Cichomski, F. Petrov, *A combinatorial proof of the Burdzy–Pitman conjecture*, Electronic Communications in Probability, **28**, 1–7, **2023**, [**21**].

Part 2: `Multivariate coherent distributions`

Chapter 5. We prove a sharp bound for the expected spread upon opinions of multiple experts – for $n \geq 2$ we calculate the exact value of $\mathbf{E}(n,1)$ (the quantity $\mathbf{E}(2,1)$ was already known). For this purpose, we introduce a novel symmetrization procedure and apply several direct combinatorial reductions. The contents of the chapter are based on the paper

• S. Cichomski, A. Osękowski, *The maximal difference among expert's opinions*, Electronic Journal of Probability, **26**, 1–17, **2021**, [**16**].

Chapter 6. We present the proof of the estimate $\mathbf{P}(n,\delta) \leq \frac{n(1-\delta)}{2-\delta} \wedge 1$, where $n \geq 2$ and $\delta \in (1/2, 1]$. This result generalizes the two-variate inequality of Burdzy and Pal, see [**9**]. Our argument rests on the dynamic programming technique, combinatorial reductions and symmetrization procedure from the previous chapter. The contents of the chapter are based on the paper

• S. Cichomski, A. Osękowski, *Contradictory predictions with multiple agents*, ALEA: Latin American Journal of Probability and Mathematical Statistics, **21**, 369–383, **2024**, [**19**].

Chapter 7. Let ξ be an integrable random variable. Fix $k \in \mathbb{N}$ and let $\{\mathcal{G}_i^j\}_{1\leq i\leq n, 1\leq j\leq k}$ be a family of sub-σ-fields, such that $\{\mathcal{G}_i^j\}_{1\leq i\leq n}$ is a filtration for each $j \in \{1,2,\ldots,k\}$. We explain the relation between the maximal function of the (generalized) coherent vector and basic properties of the uncentered Hardy–Littlewood maximal operator. We establish a relevant version of the Doob's maximal estimate. The contents of the chapter are based on the paper

• S. Cichomski, A. Osękowski, *Doob's estimate for coherent random variables and maximal operators on trees*, Probability and Mathematical Statistics, **43**, 109–119, **2023**, [**17**].

Part 1

Coherent distributions on the square

CHAPTER 2

Extreme points and asymptotics

Preliminaries

The analysis of the expression

$$\mathbf{E}(2,\alpha) = \sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha \tag{2.1}$$

for exponents $\alpha > 2$ remains a major open problem and constitutes one of the main motivations for this chapter. Accordingly, we investigate the asymptotic behavior of (2.1) and derive an appropriate sharp estimate.

THEOREM 2.1. *We have*

$$\lim_{\alpha\to\infty} \alpha \cdot \mathbf{E}(2,\alpha) = \lim_{\alpha\to\infty} \alpha \cdot \Big(\sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha \Big) = \frac{2}{e}. \tag{2.2}$$

The proof of (2.2) that we present below rests on a novel, geometric-type approach. As verified in [**10**], the family of all two-dimensional coherent distributions is a convex, compact subset of the space of probability distributions on $[0,1]^2$ equipped with the usual weak topology. One of the main results of this chapter is to provide a helpful characterisation of the extremal points of $\mathcal{C}$, which is considered to be one of the major challenges of the topic [**10, 75**].

It is instructive to take a brief look at the corresponding problem arising in the theory of martingales, the solution to which is well-known. Namely (see [**30**]), fix a number $N \in \mathbb{N}$ and consider the class of all finite martingales $(M_1, M_2, \ldots, M_N)$ and their induced distributions on $\mathbb{R}^N$. The extremal distributions can be characterised as follows:

(i) M_1 is concentrated in one point,
(ii) for any $n = 2, 3, \ldots, N$, the conditional distribution of M_n given $(M_i)_{i=1}^{n-1}$ is concentrated on the set of cardinality at most two.

In particular, the support of a two-variate martingale with an extremal distribution cannot exceed two points. Surprisingly, the structure of $\mathrm{ext}(\mathcal{C})$ (the set of extreme points of $\mathcal{C}$) is much more complex, as there exist extremal coherent measures with arbitrary large or even countable infinite number of atoms (see [**4, 75**]).

Conversely, as proved in [**4**], elements of $\mathrm{ext}(\mathcal{C})$ are always supported on sets of Lebesgue measure zero. The existence of non-atomic extreme points represents a yet another captivating problem. It turns out that such distributions do exist and we will return to this question in Chapter 3.

For the further discussion, we need to introduce some additional background and notation. For a measure μ supported on $[0,1]^2$, we will write μ^x and μ^y for the marginal measures of μ on $[0,1]$, i.e. for the measures obtained by projecting μ on the first and the second coordinate, correspondingly.

DEFINITION 2.1. *Introduce the family $\mathcal{R}$, which consists of all ordered pairs (μ,ν) of nonnegative Borel measures on $[0,1]^2$ for which*

$$\int_A (1-x)\,d\mu^x = \int_A x\,d\nu^x,$$

and

$$\int_B (1-y)\,d\mu^y = \int_B y\,d\nu^y,$$

for any Borel subsets $A, B \in \mathcal{B}([0,1])$.

It turns out that the family $\mathcal{R}$ above is very closely related to the class of coherent distributions. We will prove the following statement (a slightly different formulation can be found in [**4**]).

PROPOSITION 2.1. *Let m be a probability measure on $[0,1]^2$. Then m is coherent if and only if there exists $(\mu,\nu)\in\mathcal{R}$ such that $m=\mu+\nu$.*

The above result motivates the following.

DEFINITION 2.2. *For a fixed $m\in\mathcal{C}$, consider the class*

$$\mathcal{R}(m) = \{(\mu,\nu)\in\mathcal{R}:\ m=\mu+\nu\}.$$

Any pair $(\mu,\nu)\in\mathcal{R}(m)$ will be called a representation *of a coherent distribution m.*

By the very definition, families $\mathcal{C}$ and $\mathcal{R}$, and hence also $\mathcal{R}(m)$, are convex sets. To proceed, let us distinguish the ordering in the class of measures, which will often be used in our considerations below. Namely, for two Borel measures μ_1,μ_2 supported on the unit square, we will write $\mu_1\le\mu_2$ if we have $\mu_1(A)\le\mu_2(A)$ for all $A\in\mathcal{B}([0,1]^2)$.

DEFINITION 2.3. *Let $m\in\mathcal{C}$. We say that the representation (μ,ν) of m is*
· unique, *if for every $(\tilde\mu,\tilde\nu)\in\mathcal{R}$ with $m=\tilde\mu+\tilde\nu$, we have $\tilde\mu=\mu$ and $\tilde\nu=\nu$;*
· minimal, *if for all $(\tilde\mu,\tilde\nu)\in\mathcal{R}$ with $\tilde\mu\le\mu$ and $\tilde\nu\le\nu$, there exists $\alpha\in[0,1]$ such that $(\tilde\mu,\tilde\nu)=\alpha\cdot(\mu,\nu)$.*

In practice, we are interested only in the minimality of those representations that have been previously verified to be unique. In such a case, the minimality of (μ,ν) is just an indecomposability condition for m: we are asking whether every "coherent subsystem" $(\tilde\mu,\tilde\nu)$ contained in m is necessarily just a smaller copy of m. To gain some intuition about the above concepts, let us briefly discuss the following example.

EXAMPLE 2.1. Consider an arbitrary probability distribution m supported on the diagonal. This distribution is coherent: to see this, let ξ be a random variable on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ such that $\xi\sim m^x$. Consider the product space

$$(\tilde\Omega,\tilde{\mathcal{F}},\tilde{\mathbb{P}}) = (\Omega\times[0,1],\mathcal{F}\otimes\mathcal{B}([0,1]),\mathbb{P}\otimes|\cdot|),$$

where $|\cdot|$ denotes the Lebesgue measure. Then ξ has the same distribution as $\tilde{\mathbb{P}}(E|\mathcal{G})$, where $\mathcal{G}$ is the sub-σ-algebra of $\tilde{\mathcal{F}}$ consisting of all sets of the form $A\times[0,1]$, with $A\in\mathcal{F}$, and the event $E\in\tilde{\mathcal{F}}$ is given by

$$E=\{(\omega,y)\in\Omega\times[0,1]:\ y\le\xi(\omega)\}.$$

Consequently, we have $(\tilde{\mathbb{P}}(E|\mathcal{G}), \tilde{\mathbb{P}}(E|\mathcal{G})) \sim (\xi, \xi) \sim m$ and hence m is coherent.

Next, let us identify the representation (μ, ν) of m. Since the measure is supported on the diagonal, both components μ and ν (if exist) must also have this property and hence, when checking the conditions in the definition of the family $\mathcal{R}$, it is enough to verify the first of them. But the first condition is equivalent to saying that

$$\mathrm{d}\mu^x = x\mathrm{d}m^x \quad \text{and} \quad \mathrm{d}\nu^x = (1-x)\mathrm{d}m^x.$$

This gives the existence and uniqueness of the representation.

Finally, let us discuss the minimality of the representation of m. If m is concentrated at a single point (δ, δ), then the same is true for μ and ν, and hence also for $\tilde{\mu}$ and $\tilde{\nu}$, where $(\tilde{\mu}, \tilde{\nu}) \in \mathcal{R}$ is a pair as in the definition of minimality. Now one easily verifies that $(\tilde{\mu}, \tilde{\nu})$ is proportional to (μ, ν), applying directly the equations in the definition of class $\mathcal{R}$ with $A = B = \{\delta\}$; thus, the representation is minimal.

It remains to study the case in which measure m is not concentrated at a single point. Then there is a measure $\tilde{m}$ satisfying $\tilde{m} \leq m$, which is not proportional to m. Repeating the above argumentation with m replaced by $\tilde{m}$, we see that $\tilde{m}$ can be decomposed as the sum $\tilde{\mu} + \tilde{\nu}$, where $(\tilde{\mu}, \tilde{\nu}) \in \mathcal{R}$ is a pair of measures supported on the diagonal uniquely determined by

$$\mathrm{d}\tilde{\mu}^x = x\mathrm{d}\tilde{m}^x \quad \text{and} \quad \mathrm{d}\tilde{\nu}^x = (1-x)\mathrm{d}\tilde{m}^x.$$

Since $\tilde{m} \leq m$, we also have $\tilde{\mu} \leq \mu$ and $\tilde{\nu} \leq \nu$. It remains to note that $(\tilde{\mu}, \tilde{\nu})$ is not proportional to (μ, ν), since the same is true for $\tilde{m}$ and m. This proves that the representation (μ, ν) is not minimal. △

With these notions at hand, we shall establish a general characterisation of $\mathrm{ext}(\mathcal{C})$.

THEOREM 2.2. *Let m be a coherent distribution on $[0,1]^2$. Then m is extremal if and only if the representation of m is unique and minimal.*

This statement will be established in the next section. Later on, in Section 2.2, we intend to concentrate on extremal coherent measures with a finite support. Let

$$\mathrm{ext}_f(\mathcal{C}) = \{\eta \in \mathrm{ext}(\mathcal{C}) : |\mathrm{supp}(\eta)| < \infty\}.$$

Theorem 2.2 will enable us to deduce several structural properties of $\mathrm{ext}_f(\mathcal{C})$; most importantly, as conjectured in [**75**], we show that support of $\eta \in \mathrm{ext}_f(\mathcal{C})$ cannot contain any axial cycles. Here is the definition.

DEFINITION 2.4. *The sequence $\big((x_i, y_i)\big)_{i=1}^{2n}$ with values in $[0,1]^2$ is called an* axial cycle, *if all points (x_i, y_i) are distinct, the endpoint coordinates x_1 and x_{2n} coincide, and we have*

$$x_{2i} = x_{2i+1} \quad \textit{and} \quad y_{2i-1} = y_{2i} \quad \textit{for all } i.$$

Remarkably, the same "no axial cycle" property also holds true for extremal doubly stochastic measures (permutons) – for the relevant discussion, see [**46**].

Next, in Section 2.3, we apply our previous results and obtain the following reduction towards Theorem 2.1. Namely, for all $\alpha \geq 1$, we have

$$\sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha = \sup_{\mathbf{z}} \sum_{i=1}^n z_i \left| \frac{z_i}{z_{i-1}+z_i} - \frac{z_i}{z_i+z_{i+1}} \right|^\alpha. \tag{2.3}$$

Here the supremum is taken over all n and sequences $\tilde{\mathbf{z}} = (z_0, z_1, \ldots, z_{n+1})$ such that

$$z_0 = z_{n+1} = 0, \quad z_i > 0 \quad \text{for all} \quad i = 1, 2, \ldots, n, \quad \text{and} \quad \textstyle\sum_{i=1}^n z_i = 1.$$

Finally, using several combinatorial arguments and additional reductions, we prove Theorem 2.1 by a direct analysis of the right-hand side of (2.3).

2.1. Coherent measures, Representations

Let $\mathcal{M}([0,1]^2)$ and $\mathcal{M}([0,1])$ stand for the space of nonnegative Borel measures on $[0,1]^2$ and $[0,1]$, respectively. For $\mu \in \mathcal{M}([0,1]^2)$, let $\mu^x, \mu^y \in \mathcal{M}([0,1])$ denote

$$\mu^x(A) = \mu(A \times [0,1]) \quad \text{and} \quad \mu^y(B) = \mu([0,1] \times B),$$

for all Borel subsets $A, B \in \mathcal{B}([0,1])$. We begin with a helpful characterisation of $\mathcal{C}$. Recall the definition of the class $\mathcal{R}$ formulated in the previous section. Let us study the connection between this class and the family of all coherent distributions.

Proof of Proposition 2.1. First, we show that the decomposition $m = \mu + \nu$ exists for all $m \in \mathcal{C}$. Indeed, by virtue of Proposition 1.1, we can find a random vector $(X, Y) \sim m$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that

$$X = \mathbb{E}(Z|X) \quad \text{and} \quad Y = \mathbb{E}(Z|Y)$$

for some random variable $Z \in [0,1]$. For a set $C \in \mathcal{B}([0,1]^2)$, we put

$$\mu(C) = \int_{\{(X,Y)\in C\}} Z \mathrm{d}\mathbb{P} \quad \text{and} \quad \nu(C) = \int_{\{(X,Y)\in C\}} (1-Z) \mathrm{d}\mathbb{P}. \tag{2.4}$$

Then the equality $m = \mu + \nu$ is evident. Besides, for a fixed $A \in \mathcal{B}([0,1])$, we have

$$\int_{\{X\in A\}} X \mathrm{d}\mathbb{P} = \int_{\{X\in A\}} Z \mathrm{d}\mathbb{P} = \int_A 1 \mathrm{d}\mu^x, \tag{2.5}$$

where the first equality is due to $X = \mathbb{E}(Z|X)$ and the second is a result of (2.4). Moreover, we may also write

$$\int_{\{X\in A\}} X \mathrm{d}\mathbb{P} = \int_{A\times[0,1]} x \mathrm{d}m = \int_A x \mathrm{d}\mu^x + \int_A x \mathrm{d}\nu^x. \tag{2.6}$$

Combining (2.5) and (2.6), we get

$$\int_A (1-x) \mathrm{d}\mu^x = \int_A x \mathrm{d}\nu^x,$$

for all $A \in \mathcal{B}([0,1])$. The symmetric condition (the second part of Definition 2.1) is shown analogously. This completes the first part of the proof.

Now, pick a probability measure m on $[0,1]^2$ such that $m = \mu + \nu$ for some $(\mu, \nu) \in \mathcal{R}$. We need to show that m is coherent. To this end, consider the probability space $([0,1]^2, \mathcal{B}([0,1]^2), m)$ and the random variables $X, Y : [0,1]^2 \to [0,1]$ defined by

$$X(x,y) = x \quad \text{and} \quad Y(x,y) = y, \quad x, y \in [0,1].$$

Additionally, let Z denote the Radon–Nikodym derivative of μ with respect to m: we have $0 \le Z \le 1$ m–almost surely and $\mu(C) = \int_C Z \mathrm{d}m$ for all $C \in \mathcal{B}([0,1]^2)$. Again by Proposition 1.1, it is sufficient to verify that

$$X = \mathbb{E}(Z|X) \quad \text{and} \quad Y = \mathbb{E}(Z|Y).$$

By symmetry, it is enough to show the first equality. Fix $A \in \mathcal{B}([0,1])$ and note that

$$\int_{\{X\in A\}} X \mathrm{d}m = \int_{A\times[0,1]} x \mathrm{d}m = \int_A x \mathrm{d}\mu^x + \int_A x \mathrm{d}\nu^x. \tag{2.7}$$

Similarly, we also have

$$\int_{\{X\in A\}} Z \mathrm{d}m = \int_{A\times[0,1]} Z \mathrm{d}m = \mu(A\times[0,1]) = \int_A 1 \mathrm{d}\mu^x. \tag{2.8}$$

Finally, note that by $(\mu,\nu) \in \mathcal{R}$, the right-hand sides of (2.7) and (2.8) are equal. Therefore we obtain the identity

$$\int_{\{X\in A\}} X \mathrm{d}m = \int_{\{X\in A\}} Z \mathrm{d}m$$

for arbitrary $A \in \mathcal{B}([0,1])$. This yields the claim. □

We turn our attention to the characterisation of $\mathrm{ext}(\mathcal{C})$ stated in the previous section.

Proof of Theorem 2.2, the implication "⇒". Let us assume that m is an extremal coherent measure and suppose, on contrary, that (μ_1,ν_1) and (μ_2,ν_2) are two different elements of $\mathcal{R}(m)$. We will prove that $m-\mu_1+\mu_2$ and $m-\mu_2+\mu_1$ are also coherent distributions. Because of

$$m = \frac{1}{2}(m-\mu_1+\mu_2) + \frac{1}{2}(m-\mu_2+\mu_1),$$

we will obtain the contradiction with the assumed extremality of m. By symmetry, it is enough to show that $(m-\mu_1+\mu_2) \in \mathcal{C}$. To this end, by virtue of Proposition 2.1, it suffices to check that $m-\mu_1+\mu_2$ is a probability measure and $(\mu_2, m-\mu_1) \in \mathcal{R}$.

First, note that $\nu_1 = m-\mu_1$ is nonnegative and fix an arbitrary $A \in \mathcal{B}([0,1])$. As (μ_1,ν_1) and (μ_2,ν_2) are representations of m, Definition 2.1 gives

$$\int_A 1 \mathrm{d}\mu_1^x = \int_A x(\mathrm{d}\nu_1^x + \mathrm{d}\mu_1^x) = \int_A x \mathrm{d}m^x,$$

and

$$\int_A 1 \mathrm{d}\mu_2^x = \int_A x(\mathrm{d}\nu_2^x + \mathrm{d}\mu_2^x) = \int_A x \mathrm{d}m^x, \tag{2.9}$$

so $\mu_1^x(A) = \mu_2^x(A)$. Similarly, we can conclude that $\mu_1^y = \mu_2^y$, which means that marginal distributions of μ_1 and μ_2 are equal. This, together with $m-\mu_1 \geq 0$, proves that $m-\mu_1+\mu_2$ is a probability measure.

Next, using (2.9) and $\mu_1^x = \mu_2^x$, we can also write

$$\int_A (1-x) \mathrm{d}\mu_2^x = \int_A x \mathrm{d}m^x - \int_A x \mathrm{d}\mu_1^x = \int_A x \mathrm{d}(m-\mu_1)^x. \tag{2.10}$$

In the same way we get

$$\int_B (1-y) \mathrm{d}\mu_2^y = \int_B y \mathrm{d}(m-\mu_1)^y, \tag{2.11}$$

for all $B \in \mathcal{B}([0,1])$. By (2.10) and (2.11), we obtain that $(\mu_2, m-\mu_1) \in \mathcal{R}$ and this completes the proof of the uniqueness.

To show the minimality, let us assume that m is an extremal coherent measure with the representation (μ, ν) (which is unique, as we have just proved). Consider any nonzero $(\tilde{\mu}, \tilde{\nu}) \in \mathcal{R}$ with $\tilde{\mu} \le \mu$ and $\tilde{\nu} \le \nu$. Then, by the very definition of $\mathcal{R}$, we have $(\mu - \tilde{\mu}, \nu - \tilde{\nu}) \in \mathcal{R}$. Therefore, by Proposition 2.1, we get

$$\alpha^{-1}(\tilde{\mu} + \tilde{\nu}), (1-\alpha)^{-1}(m - \tilde{\mu} - \tilde{\nu}) \in \mathcal{C},$$

where $\alpha = (\tilde{\mu} + \tilde{\nu})([0,1]^2) \in (0,1]$. We have the identity

$$m = \alpha \cdot \Big(\alpha^{-1}(\tilde{\mu} + \tilde{\nu})\Big) + (1-\alpha) \cdot \Big((1-\alpha)^{-1}(m - \tilde{\mu} - \tilde{\nu})\Big),$$

which combined with the extremality of m yields

$$m = \alpha^{-1}(\tilde{\mu} + \tilde{\nu}) = \alpha^{-1}\tilde{\mu} + \alpha^{-1}\tilde{\nu}.$$

But $(\alpha^{-1}\tilde{\mu}, \alpha^{-1}\tilde{\nu})$ belongs to $\mathcal{R}$, since $(\tilde{\mu}, \tilde{\nu})$ does, and so $(\alpha^{-1}\tilde{\mu}, \alpha^{-1}\tilde{\nu})$ is also a representation of m. By the uniqueness, we deduce that $(\tilde{\mu}, \tilde{\nu}) = \alpha \cdot (\mu, \nu)$. □

Proof of Theorem 2.2, the implication "⇐". Suppose that measure m is a coherent distribution with the unique and minimal representation (μ, ν). To show that m is extremal, consider the decomposition

$$m = \beta \cdot m_1 + (1-\beta) \cdot m_2$$

for some $m_1, m_2 \in \mathcal{C}$ and $\beta \in (0,1)$. Moreover, let

$$(\mu_1, \nu_1) \in \mathcal{R}(m_1) \quad \text{and} \quad (\mu_2, \nu_2) \in \mathcal{R}(m_2).$$

By the convexity of $\mathcal{R}$, we have

$$(\mu', \nu') := (\beta\mu_1 + (1-\beta)\mu_2, \beta\nu_1 + (1-\beta)\nu_2) \in \mathcal{R}(m) \tag{2.12}$$

and hence, by the uniqueness, we get $(\mu', \nu') = (\mu, \nu)$. Then, from (2.12), we have

$$\beta\mu_1 \le \mu \ \text{ and } \ \beta\nu_1 \le \nu. \tag{2.13}$$

Combining this with the minimality of (μ, ν), we get $(\beta\mu_1, \beta\nu_1) = \alpha(\mu, \nu)$ for some $\alpha \in [0,1]$. Since $m = \mu + \nu$ and $m_1 = \mu_1 + \nu_1$ are probability measures, this gives $\alpha = \beta$ and hence $(\mu_1, \nu_1) = (\mu, \nu)$. This implies that $m = m_1$ and ends the proof. □

2.2. Extreme points with a finite support

In this section we study the geometric structure of the supports of measures from

$$\text{ext}_f(\mathcal{C}) = \{\eta \in \text{ext}(\mathcal{C}) : \ |\text{supp}(\eta)| < \infty\}.$$

Our key result is presented in Theorem 2.3 – we prove that the support of an extremal coherent distribution cannot contain any axial cycles (recall Definition 2.4). Let us emphasize that this property has been originally conjectured in [**75**]. We start with a simple combinatorial observation: it is straightforward to check that certain special "alternating" cycles are forbidden.

Definition 2.5. *Let η be a coherent distribution with a fixed representation (μ, ν) and let $\big((x_i, y_i)\big)_{i=1}^{2n}$ be an axial cycle contained in* $\text{supp}(\eta)$. *Then $\big((x_i, y_i)\big)_{i=1}^{2n}$ is an alternating cycle if*

$$(x_{2i+1}, y_{2i+1}) \in \text{supp}(\mu) \quad \textit{and} \quad (x_{2i}, y_{2i}) \in \text{supp}(\nu),$$

for all $i = 1, 2, \ldots, n$ – *with the convention* $x_{2n+1} = x_1$, $y_{2n+1} = y_1$.

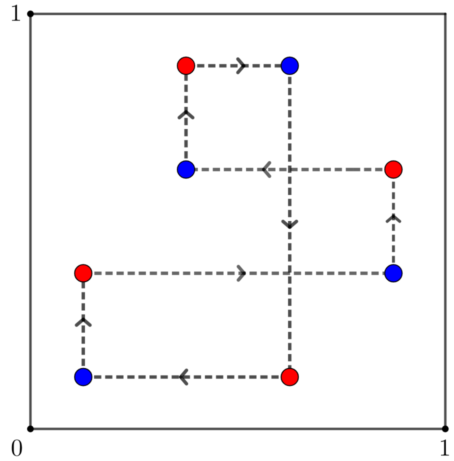

FIGURE 1. An example of an alternating cycle. Red points represent probability masses in $\mathrm{supp}(\mu)$, while blue points indicate probability masses in $\mathrm{supp}(\nu)$. Arrows outline a possible transformation of the representation (μ, ν).

PROPOSITION 2.2. *If distribution* $\eta \in \mathrm{ext}_f(\mathcal{C})$, *then* $\mathrm{supp}(\eta)$ *does not contain any alternating cycles.*

PROOF. Let η be a coherent distribution with a unique representation (μ, ν) and a finite support. Additionally, assume that $\big((x_i, y_i)\big)_{i=1}^{2n}$ is an alternating cycle contained in $\mathrm{supp}(\eta)$. Let δ be the smaller of the two numbers

$$\min_{0 \le i \le n-1} \mu(x_{2i+1}, y_{2i+1}) \quad \text{and} \quad \min_{1 \le i \le n} \nu(x_{2i}, y_{2i})$$

(for brevity, in what follows we will skip the parentheses and write $\mu(a, b)$, $\nu(a, b)$ instead of $\mu(\{a, b\})$, $\nu(\{a, b\})$, respectively). By Definition 2.5, we have $\delta > 0$. Now, consider the transformation $(\mu, \nu) \mapsto (\mu', \nu')$ described by the following requirements:

1. for $i = 0, 1, \ldots, n-1$, put
$$\mu'(x_{2i+1}, y_{2i+1}) := \mu(x_{2i+1}, y_{2i+1}) - \delta$$
$$\nu'(x_{2i+1}, y_{2i+1}) := \nu(x_{2i+1}, y_{2i+1}) + \delta,$$
2. for $i = 1, 2, \ldots, n$, put
$$\mu'(x_{2i}, y_{2i}) := \mu(x_{2i}, y_{2i}) + \delta$$
$$\nu'(x_{2i}, y_{2i}) := \nu(x_{2i}, y_{2i}) - \delta,$$
3. for $(x, y) \notin \{(x_i, y_i) : 1 \le i \le 2n\}$, set
$$\mu'(x, y) = \mu(x, y),$$
$$\nu'(x, y) = \nu(x, y).$$

Note that μ and μ', as well as ν and ν', have the same marginal distributions and hence $(\mu', \nu') \in \mathcal{R}$. We also have $\mu' + \nu' = \mu + \nu = \eta$ and thus $(\mu', \nu') \in \mathcal{R}(\eta)$. This contradicts the uniqueness of the representation (μ, ν) and shows that $\mathrm{supp}(\eta)$ cannot contain an alternating cycle. By Theorem 2.2, this ends the proof. □

Before the further combinatorial analysis, we need to introduce some useful notation.

DEFINITION 2.6. *For $\mu, \nu \in \mathcal{M}([0,1]^2)$ with $|\mathrm{supp}(\mu+\nu)| < \infty$, we define a quotient function $q_{(\mu,\nu)} : \mathrm{supp}(\mu+\nu) \to [0,1]$ by*

$$q_{(\mu,\nu)}(x,y) = \frac{\mu(x,y)}{\mu(x,y) + \nu(x,y)}.$$

In what follows, we will omit the subscripts and write q for $q_{(\mu,\nu)}$ whenever the choice for (μ, ν) is clear from the context.

PROPOSITION 2.3. *Let $\mu, \nu \in \mathcal{M}([0,1]^2)$ and $|\mathrm{supp}(\mu+\nu)| < \infty$.*
Then $(\mu, \nu) \in \mathcal{R}$ if and only if the following conditions hold simultaneously:

- *for every x satisfying $\mu(\{x\} \times [0,1]) + \nu(\{x\} \times [0,1]) > 0$, we have*

$$\sum_{\substack{y \in [0,1], \\ (x,y) \in \mathrm{supp}(\mu+\nu)}} q(x,y) \frac{\mu(x,y) + \nu(x,y)}{\mu(\{x\} \times [0,1]) + \nu(\{x\} \times [0,1])} = x, \tag{2.14}$$

- *for every y satisfying $\mu([0,1] \times \{y\}) + \nu([0,1] \times \{y\}) > 0$, we have*

$$\sum_{\substack{x \in [0,1], \\ (x,y) \in \mathrm{supp}(\mu+\nu)}} q(x,y) \frac{\mu(x,y) + \nu(x,y)}{\mu([0,1] \times \{y\}) + \nu([0,1] \times \{y\})} = y, \tag{2.15}$$

where sums in (2.14) and (2.15) are well defined – in both cases, there is only a finite number of nonzero summands.

PROOF. As $|\mathrm{supp}(\mu+\nu)| < \infty$, this is a simple consequence of Definition 2.1. □

Next, we will require an additional distinction between three different types of points.

DEFINITION 2.7. *Let $(\mu, \nu) \in \mathcal{R}$. A point $(x, y) \in \mathrm{supp}(\mu + \nu)$ is said to be*

a) a lower out point, if $q(x, y) < \min(x, y)$;
b) an upper out point, if $q(x, y) > \max(x, y)$;
c) a cut point, if it is not an out point, i.e.

$$x \le q(x,y) \le y \quad \text{or} \quad y \le q(x,y) \le x.$$

Finally, for the sake of completeness, we include a formal definition of an axial path.

DEFINITION 2.8. *The sequence $\big((x_i, y_i)\big)_{i=1}^{n}$ with terms in $[0,1]^2$ is called an axial path if: all points (x_i, y_i) are distinct; we have*

$$x_{2i} = x_{2i+1} \quad \text{and} \quad y_{2i-1} = y_{2i},$$

or

$$y_{2i} = y_{2i+1} \quad \text{and} \quad x_{2i-1} = x_{2i},$$

for all i.

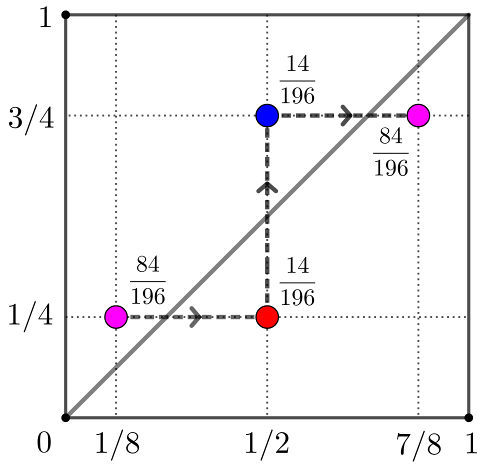

FIGURE 2. Support of a coherent distribution m. Purple points (endpoints of the path) are cut points. Red point represents a mass in supp(μ) and is an upper out point. Blue point indicates a mass in supp(ν) and it is a lower out point.

To develop some intuition, it is convenient to inspect the example given below.

EXAMPLE 2.2. Let m be a probability measure given by

$$m\Big(\frac{1}{8},\frac{1}{4}\Big)=\frac{84}{196},\quad m\Big(\frac{1}{2},\frac{1}{4}\Big)=\frac{14}{196},\quad m\Big(\frac{1}{2},\frac{3}{4}\Big)=\frac{14}{196},\quad m\Big(\frac{7}{8},\frac{3}{4}\Big)=\frac{84}{196}.$$

There are five observations, which will be discussed separately.

(i) Consider the decomposition $m=\mu+\nu$, where (μ,ν) is determined by the quotient

$$q\Big(\frac{1}{8},\frac{1}{4}\Big)=\frac{1}{8},\quad q\Big(\frac{1}{2},\frac{1}{4}\Big)=1,\quad q\Big(\frac{1}{2},\frac{3}{4}\Big)=0,\quad q\Big(\frac{7}{8},\frac{3}{4}\Big)=\frac{7}{8}.$$

Using Proposition 2.3, we can check that $(\mu,\nu)\in\mathcal{R}$. For instance, for $y=\frac{1}{4}$ we get

$$\frac{q(\frac{1}{8},\frac{1}{4})\cdot m(\frac{1}{8},\frac{1}{4})+q(\frac{1}{2},\frac{1}{4})\cdot m(\frac{1}{2},\frac{1}{4})}{m(\frac{1}{8},\frac{1}{4})+m(\frac{1}{2},\frac{1}{4})}=\frac{\frac{1}{8}\cdot\frac{84}{196}+1\cdot\frac{14}{196}}{\frac{84}{196}+\frac{14}{196}}=\frac{1}{4},\tag{2.16}$$

which agrees with (2.15). As a consequence, by Proposition 2.1, we have $m\in\mathcal{C}$.

(ii) Observe that $(\frac{1}{8},\frac{1}{4})$ and $(\frac{7}{8},\frac{3}{4})$ are cut points, $(\frac{1}{2},\frac{1}{4})$ is an upper and $(\frac{1}{2},\frac{3}{4})$ is a lower out point. Besides, supp(m) is an axial path without cycles – see Figure 2.

(iii) Notably, pair (μ,ν) is a unique representation of measure m. Indeed, $(\frac{1}{8},\frac{1}{4})$ is the only point in supp(m) with x-coordinate equal to $\frac{1}{8}$ and hence $q(\frac{1}{8},\frac{1}{4})=\frac{1}{8}$. Accordingly, $q(\frac{1}{2},\frac{1}{4})=1$ is now a consequence of (2.16). The derivation of $q(\frac{1}{2},\frac{3}{4})=0$ and $q(\frac{7}{8},\frac{3}{4})=\frac{7}{8}$ follows from an analogous computation.

(iv) Finally, the representation (μ,ν) is minimal; let $(\tilde{\mu},\tilde{\nu})\in\mathcal{R}$ satisfy $\tilde{\mu}\le\mu$ and $\tilde{\nu}\le\nu$. Suppose that $(\frac{1}{8},\frac{1}{4})\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu})$. Again, as $(\frac{1}{8},\frac{1}{4})$ is the only point in supp(m) with x-coordinate equal to $\frac{1}{8}$, we get $q_{(\tilde{\mu},\tilde{\nu})}(\frac{1}{8},\frac{1}{4})=\frac{1}{8}$. Next, assume that $(\frac{1}{2},\frac{1}{4})\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu})$. As $\tilde{\nu}(\frac{1}{2},\frac{1}{4})\le\nu(\frac{1}{2},\frac{1}{4})=0$, we have $q_{(\tilde{\mu},\tilde{\nu})}(\frac{1}{2},\frac{1}{4})=1$. Likewise, we can check that

$$q_{(\tilde{\mu},\tilde{\nu})}(x,y)=q_{(\mu,\nu)}(x,y)\quad\text{for all }(x,y)\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu}).\tag{2.17}$$

By Proposition 2.3 and the equation (2.17), we easily obtain that $\tilde{\mu}+\tilde{\nu}=0$ or $\operatorname{supp}(\tilde{\mu}+\tilde{\nu})=\operatorname{supp}(m)$. For example,

· if $(\frac{1}{2},\frac{1}{4})\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu})$, then (2.14) gives $(\frac{1}{2},\frac{3}{4})\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu})$;

· if $(\frac{1}{2},\frac{3}{4})\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu})$, then (2.15) yields $(\frac{7}{8},\frac{3}{4})\in\operatorname{supp}(\tilde{\mu}+\tilde{\nu})$.

Therefore, if $\tilde{\mu}+\tilde{\nu}\neq 0$, then the measure $\tilde{\mu}+\tilde{\nu}$ is supported on the same set as m and $q_{(\tilde{\mu},\tilde{\nu})}\equiv q_{(\mu,\nu)}$. For the same reason, i.e. using Proposition 2.3 and path structure of $\operatorname{supp}(m)$, it follows that $\tilde{\mu}+\tilde{\nu}=\alpha\cdot m$ for some $\alpha\in[0,1]$. For instance, by (2.15) for $y=\frac{1}{4}$, we get

$$\frac{\frac{1}{8}\cdot\tilde{m}(\frac{1}{8},\frac{1}{4})+1\cdot\tilde{m}(\frac{1}{2},\frac{1}{4})}{\tilde{m}(\frac{1}{8},\frac{1}{4})+\tilde{m}(\frac{1}{2},\frac{1}{4})}=\frac{1}{4},$$

where $\tilde{m}=\tilde{\mu}+\tilde{\nu}$. Hence $\tilde{m}(\frac{1}{8},\frac{1}{4})\tilde{m}(\frac{1}{2},\frac{1}{4})^{-1}=m(\frac{1}{8},\frac{1}{4})m(\frac{1}{2},\frac{1}{4})^{-1}=\frac{84}{14}$.

(v) By the above analysis and Theorem 2.2, we conclude that $m\in\operatorname{ext}_f(\mathcal{C})$. △

To clarify the main reasoning, we first record an evident geometric lemma.

LEMMA 2.1. *Let* $\big((x_i,y_i)\big)_{i=1}^{n}$ *be an axial path without cycles.*

A. If $x_{n-1}=x_n$ *(or* $y_{n-1}=y_n$*), then* $y_n\neq y_j$ *(or* $x_n\neq x_j$*) for all* $j<n$.
B. For every $x,y\in[0,1]$, *we have*

$$\max\Big\{\,|\{i:x_i=x\}|\,,|\{j:y_j=y\}|\,\Big\}<3.$$

PROOF. Part A can be verified by induction. Part B follows from A. □

We are now ready to demonstrate the central result of this section.

THEOREM 2.3. *If* $\eta\in\operatorname{ext}_f(\mathcal{C})$, *then* $\operatorname{supp}(\eta)$ *is an axial path without cycles.*

Let us briefly explain the main idea of the proof. For $\eta\in\operatorname{ext}_f(\mathcal{C})$, we inductively construct a special axial path contained in $\operatorname{supp}(\eta)$, which does not contain any cut points (apart from the endpoints). We show that axial path obtained in this process is acyclic and involves all points from $\operatorname{supp}(\eta)$.

PROOF OF THEOREM 2.3. Fix measure $\eta\in\operatorname{ext}_f(\mathcal{C})$ and let (μ,ν) be the unique representation of η. By $\mathcal{L}(\eta)$ and $\mathcal{U}(\eta)$ denote the sets of lower and upper out points, correspondingly. Choose any $(x_0,y_0)\in\operatorname{supp}(\eta)$. We will consider two separate cases:

Case I: (x_0,y_0) is an out point. With no loss of generality, we can assume that $(x_0,y_0)\in\mathcal{L}(\eta)$. We then use the following inductive procedure.

1° Suppose we have successfully found $(x_n,y_n)\in\mathcal{L}(\eta)$ and it is the first time we have chosen a point with the x-coordinate equal to x_n. Since $(x_n,y_n)\in\mathcal{L}(\eta)$, we have $q(x_n,y_n)<x_n$. By (2.14), there must exist a point $(x_{n+1},y_{n+1})\in\operatorname{supp}(\eta)$ such that $x_{n+1}=x_n$ and $q(x_{n+1},y_{n+1})>x_n$. We pick one such point and add it at the end of the path. If (x_{n+1},y_{n+1}) is a cut point or an axial cycle was just created, we exit the loop. Otherwise, note that $(x_{n+1},y_{n+1})\in\mathcal{U}(\eta)$ and $y_{n+1}\neq y_j$ for all $j<n+1$ (by part A of Lemma 2.1). Go to 2°.

2° Assume we have successfully found $(x_n,y_n)\in\mathcal{U}(\eta)$ and it is the first time we have chosen a point with the y-coordinate equal to y_n. Since $(x_n,y_n)\in\mathcal{U}(\eta)$, we have

$q(x_n, y_n) > y_n$. By (2.15), there must exist a point $(x_{n+1}, y_{n+1}) \in \operatorname{supp}(\eta)$ such that $y_{n+1} = y_n$ and $q(x_{n+1}, y_{n+1}) < y_n$. We pick one such point and add it at the end of the path. If (x_{n+1}, y_{n+1}) is a cut point or an axial cycle was just created, we exit the loop. Otherwise, note that $(x_{n+1}, y_{n+1}) \in \mathcal{L}(\eta)$ and $x_{n+1} \neq x_j$ for all $j < n + 1$ (by part A of Lemma 2.1). Go to 1°.

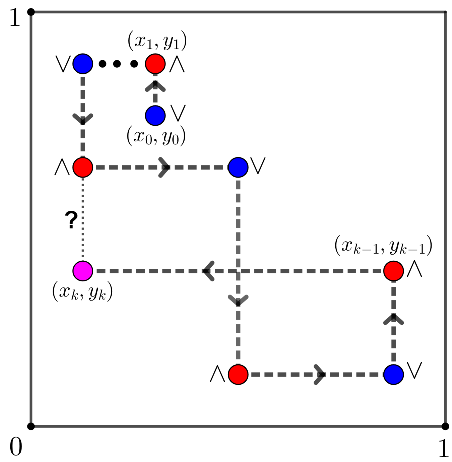

FIGURE 3. An example of an axial path constructed by the algorithm. Symbols $\vee, \wedge$ are placed next to lower ($\vee$) and upper ($\wedge$) out points. Purple point (x_k, y_k) is the endpoint of the path. Red points represent probability masses in $\operatorname{supp}(\mu)$, while blue points indicate probability masses in $\operatorname{supp}(\nu)$.

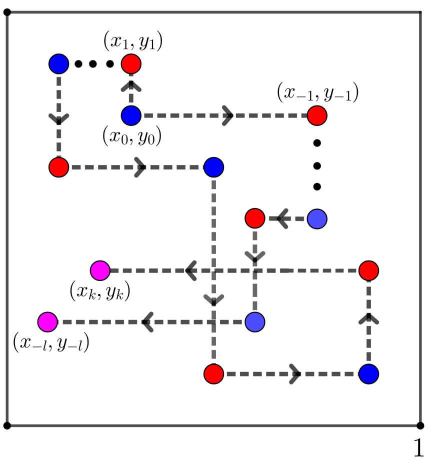

FIGURE 4. An example of an axial path Γ constructed after the second (reversed) run of the algorithm. Purple points (x_k, y_k) and (x_{-l}, y_{-l}) (endpoints of Γ) are cut points. Red points represent probability masses in $\operatorname{supp}(\mu)$, while blue points indicate masses in $\operatorname{supp}(\nu)$.

As $|\operatorname{supp}(\eta)| < \infty$, the above procedure terminates after a finite number of steps (let us denote it by k) and produces an axial path $\big((x_i, y_i)\big)_{i=0}^{k}$ contained in $\operatorname{supp}(\eta)$.

Notice that it is possible that (x_k, y_k) is a third point on some horizontal or vertical line – in such a case, by part B of Lemma 2.1, the sequence $\big((x_i, y_i)\big)_{i=0}^{k}$ contains an axial cycle. Now, by the construction of the loop, point (x_k, y_k) is either an endpoint of an axial cycle or a cut point.

Let us show that the first alternative is impossible. First, we clearly have

$$\mathcal{L}(\eta) \subseteq \operatorname{supp}(\nu) \quad \text{and} \quad \mathcal{U}(\eta) \subseteq \operatorname{supp}(\mu),$$

see Figure 3. Next, assume that $(x_{k-1}, y_{k-1}) \in \mathcal{U}(\eta)$. This means that (x_k, y_k) was found in step 2° and $q(x_k, y_k) < y_{k-1} \leq 1$. Therefore point $(x_k, y_k) \in \operatorname{supp}(\nu)$ and there exists an alternating cycle in $\operatorname{supp}(\eta)$. However, this is not possible because of Proposition 2.2. If $(x_{k-1}, y_{k-1}) \in \mathcal{L}(\eta)$, the argument is analogous.

We have shown that (x_k, y_k) is a cut point. Set $\Gamma_+ = \bigcup_{i=1}^{k}\{(x_i, y_i)\}$. Moving on, we can return to the starting point (x_0, y_0) and repeat the above construction in the reversed direction. By switching the roles of x and y-coordinates in steps 1° and 2°, we produce another axial path $(x_i, y_i)_{i=0}^{-l}$. Set $\Gamma_- = \bigcup_{i=-1}^{-l}\{(x_i, y_i)\}$ and

$$\Gamma = \Gamma_+ \cup \{(x_0, y_0)\} \cup \Gamma_-.$$

Repeating the same arguments as before, we show that (x_{-l}, y_{-l}) is a cut point and Γ is an axial path without cycles, see Figure 4.

It remains to verify that $\operatorname{supp}(\eta) = \Gamma$. This will be accomplished by showing that there exists $(\tilde{\mu}, \tilde{\nu}) \in \mathcal{R}$ with $\tilde{\mu} \leq \mu$, $\tilde{\nu} \leq \nu$ and $\operatorname{supp}(\tilde{\mu} + \tilde{\nu}) = \Gamma$. This will give the claim: by the minimality of the representation (μ, ν), we will deduce that $\tilde{\mu} + \tilde{\nu} = \alpha \cdot \eta$ for some $\alpha \in (0, 1]$, and hence $\operatorname{supp}(\tilde{\mu} + \tilde{\nu}) = \operatorname{supp}(\eta)$.

We begin with the endpoints of Γ. As (x_k, y_k) is a cut point, there exists $\gamma \in [0, 1]$ such that $q(x_k, y_k) = \gamma x_k + (1 - \gamma) y_k$. We can write

$$\eta(x_k, y_k) = \eta'(x_k, y_k) + \eta''(x_k, y_k), \tag{2.18}$$

where $\eta'(x_k, y_k) = \gamma \eta(x_k, y_k)$ and $\eta''(x_k, y_k) = (1 - \gamma)\eta(x_k, y_k)$. Set

$$\mu'(x_k, y_k) = x_k \eta'(x_k, y_k) \quad \text{and} \quad \mu''(x_k, y_k) = y_k \eta''(x_k, y_k). \tag{2.19}$$

By (2.18) and (2.19), we have

$$\mu'(x_k, y_k) + \mu''(x_k, y_k) = \Big(x_k \gamma + y_k(1 - \gamma)\Big)\eta(x_k, y_k) = \mu(x_k, y_k). \tag{2.20}$$

Equations (2.18) and (2.20) have a clear and convenient interpretation. Indeed, we can visualize it as "cutting" the point (x_k, y_k) into two separate points: $(x_k, y_k)'$ with mass $\eta'(x_k, y_k)$ and $(x_k, y_k)''$ with mass $\eta''(x_k, y_k)$. Moreover, calculating their quotient functions independently, we get

$$q'(x_k, y_k) = x_k \quad \text{and} \quad q''(x_k, y_k) = y_k.$$

Performing the same "cut" operation on (x_{-l}, y_{-l}) we can divide this point into $(x_{-l}, y_{-l})'$ and $(x_{-l}, y_{-l})''$ such that $q'(x_{-l}, y_{-l}) = x_{-l}$ and $q''(x_{-l}, y_{-l}) = y_{-l}$.

Observe that (x_k, y_k) and (x_{k-1}, y_{k-1}) have exactly one common coordinate, say $y_k = y_{k-1}$. Consequently, (x_k, y_k) is the only point in Γ with x-coordinate equal to x_k. Additionally, by (2.15) and $(x_{k-1}, y_{k-1}) \in \mathcal{U}(\eta)$, this means that $q(x_k, y_k) \neq y_k$ and $\gamma > 0$. Hence $\eta'(x_k, y_k) > 0$. Similarly, suppose that $y_{-l} = y_{-l+1}$ (as shown in Figure 4; for other configurations of endpoints, we proceed by analogy). Thus, (x_{-l}, y_{-l}) is the only point in Γ with x-coordinate equal to x_{-l}. By (2.15) and $(x_{-l+1}, y_{-l+1}) \in \mathcal{L}(\eta)$, we have $\eta'(x_{-l}, y_{-l}) > 0$.

Next, consider the following function $\tilde{q} : \Gamma \to [0,1]$ uniquely determined by the following requirements:

1. $\tilde{q}(x_k, y_k) = x_k$ (if $y_k = y_{k-1}$, as we have assumed) or $\tilde{q}(x_k, y_k) = y_k$ (in the case when $x_k = x_{k-1}$),
2. $\tilde{q}(x_{-l}, y_{-l}) = x_{-l}$ (if $y_{-l} = y_{-l+1}$, as we have assumed) or $\tilde{q}(x_{-l}, y_{-l}) = y_{-l}$ (in the case when $x_{-l} = x_{-l+1}$),
3. $\tilde{q}(x,y) = 0$ for all $(x,y) \in \Gamma \cap \mathcal{L}(\eta)$,
4. $\tilde{q}(x,y) = 1$ for all $(x,y) \in \Gamma \cap \mathcal{U}(\eta)$.

Set $\delta = \min(a,b,c,d)$, where

$$a = \eta'(x_k, y_k) \ \ (\text{if } y_k = y_{k-1}) \quad \text{or} \quad a = \eta''(x_k, y_k) \ \ (\text{if } x_k = x_{k-1}),$$

$$b = \eta'(x_{-l}, y_{-l}) \ \ (\text{if } y_{-l} = y_{-l+1}) \quad \text{or} \quad b = \eta''(x_{-l}, y_{-l}) \ \ (\text{if } x_{-l} = x_{-l+1}),$$

$$c = \min\nolimits_{(x,y)\in\Gamma\cap\mathcal{L}(\eta)} \nu(x,y), \quad d = \min\nolimits_{(x,y)\in\Gamma\cap\mathcal{U}(\eta)} \mu(x,y).$$

Then $\delta > 0$, which follows from the previous discussion. Finally, using the acyclic path structure of Γ and Proposition 2.3 (just as in Example 2.2), we are able to find a pair $(\tilde{\mu}, \tilde{\nu}) \in \mathcal{R}$ with $\operatorname{supp}(\tilde{\mu} + \tilde{\nu}) = \Gamma$ and a quotient function $q_{(\tilde{\mu},\tilde{\nu})} = \tilde{q}$. Letting

$$\beta = \delta \cdot \Big(\max_{(x,y)\in\Gamma} (\tilde{\mu} + \tilde{\nu})(x,y) \Big)^{-1},$$

we see that $\beta\tilde{\mu} \le \mu$ and $\beta\tilde{\nu} \le \nu$, as desired.

Case II: (x_0, y_0) is a cut point. Suppose that $x_0 = y_0$ and $q(x_0, x_0) = x_0$. Put

$$\tilde{\mu} = \mathbb{1}_{\{(x_0,x_0)\}} x_0 \eta(x_0, y_0) \ \text{ and } \ \tilde{\nu} = \mathbb{1}_{\{(x_0,x_0)\}} (1 - x_0) \eta(x_0, y_0).$$

We have $(\tilde{\mu}, \tilde{\nu}) \in \mathcal{R}$ and $\tilde{\mu} \le \mu$, $\tilde{\nu} \le \nu$. Hence $\operatorname{supp}(\eta) = \{(x_0, x_0)\}$. Next, assume that $x_0 \neq y_0$. In that case, $q(x_0, y_0)$ cannot be equal to both x_0 and y_0 at the same time and we can clearly apply the same recursive procedure as in Case I.

For example, let us assume that $q(x_0, y_0) < x_0$. Although $(x_0, y_0) \notin \mathcal{L}(\eta)$, by (2.14) there still must exist a point $(x_1, y_1) \in \operatorname{supp}(\eta)$ such that $x_1 = x_0$ and $q(x_1, y_1) > x_0$. If (x_1, y_1) is not a cut point, then $(x_1, y_1) \in \mathcal{U}(\eta)$ and we can go to step 2°. Now the procedure continues without any further changes. The details of the proof remain the same as in Case I. □

From the proof provided, we can deduce yet another conclusion.

Corollary 2.1. *If* $\eta \in \operatorname{ext}_f(\mathcal{C})$, *then*

$$q(x,y) = 0 \quad \textit{for all} \quad (x,y) \in \mathcal{L}(\eta)$$

and

$$q(x,y) = 1 \quad \textit{for all} \quad (x,y) \in \mathcal{U}(\eta).$$

Except for the endpoints of this axial path (which are cut points), $\operatorname{supp}(\eta)$ *consists of lower and upper out points, appearing alternately.*

PROOF. Note that $\mathcal{L}(\eta)$ and $\mathcal{U}(\eta)$ are well defined as the representation of η is unique. The statement follows directly from the proof of Theorem 2.3. □

2.3. Asymptotic estimate

Equipped with the machinery developed in the previous sections, we are ready to establish the asymptotic estimate (2.2). We need to clarify how the properties of $\mathrm{ext}_f(\mathcal{C})$ covered in the preceding part apply to this problem. Referring to the prior notation, we will write

$$(X,Y) \in \mathcal{C}_f \quad \text{or} \quad (X,Y) \in \mathrm{ext}_f(\mathcal{C}),$$

to indicate that the distribution of a random vector (X,Y) is a coherent (or an extremal coherent) measure with a finite support.

PROPOSITION 2.4. *For any $\alpha > 0$, we have*

$$\sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha = \sup_{(X,Y)\in\mathcal{C}_f} \mathbb{E}|X-Y|^\alpha.$$

PROOF. Fix any $(X,Y) \in \mathcal{C}$. Directly by Proposition 1.2, there exists a sequence $(X_n, Y_n) \in \mathcal{C}_f$ such that X_n, Y_n each take at most n different values and

$$\max\Big\{|X - X_n|, |Y - Y_n|\Big\} \le \frac{1}{n}, \quad \text{for all } n = 1, 2, \ldots \tag{2.21}$$

almost surely. Consequently, by dominated convergence and (2.21), we obtain

$$\mathbb{E}|X-Y|^\alpha = \lim_{n\to\infty} \mathbb{E}|X_n - Y_n|^\alpha,$$

and thus

$$\mathbb{E}|X-Y|^\alpha \le \sup_{n\in\mathbb{N}} \mathbb{E}|X_n - Y_n|^\alpha \le \sup_{(X,Y)\in\mathcal{C}_f} \mathbb{E}|X-Y|^\alpha.$$

This proves the "$\le$"-inequality, while in the reversed direction it is evident. □

Next, we will apply the celebrated Krein–Milman theorem, see [**53, 65**].

THEOREM 2.4 (Krein–Milman). *A compact convex subset of a Hausdorff locally convex topological vector space is equal to the closed convex hull of its extreme points.*

The above statement enables us to restrict the analysis of the estimate (2.2) to the simpler study of extremal measures. Precisely, we have the following reduction.

PROPOSITION 2.5. *For any $\alpha > 0$, we have*

$$\sup_{(X,Y)\in\mathcal{C}_f} \mathbb{E}|X-Y|^\alpha = \sup_{(X,Y)\in\mathrm{ext}_f(\mathcal{C})} \mathbb{E}|X-Y|^\alpha.$$

PROOF. Let $Z = C([0,1]^2, \mathbb{R})$; then Z^* is the space of all finite signed Borel measures with the total variation norm $\|\cdot\|_{\mathrm{TV}}$. Let us equip Z^* with the topology of weak* convergence. Under this topology, Z^* is a Hausdorff and a locally convex space. For a fixed $m \in \mathcal{C}_f$, let

$$\mathcal{C}_m = \{m' \in \mathcal{C}_f : \ \mathrm{supp}(m') \subseteq \mathrm{supp}(m)\}$$

denote the family of coherent distributions supported on the subsets of $\mathrm{supp}(m)$.

Firstly, observe that $\mathcal{C}_m$ is convex. Secondly, we can easily verify that

$$\mathrm{ext}(\mathcal{C}_m) = \mathcal{C}_m \cap \mathrm{ext}_f(\mathcal{C}).$$

Plainly, if $m' \in \mathcal{C}_m$ and $m' = \alpha \cdot m_1 + (1-\alpha) \cdot m_2$ for some $\alpha \in (0,1)$ and $m_1, m_2 \in \mathcal{C}$, then $\mathrm{supp}(m') = \mathrm{supp}(m_1) \cup \mathrm{supp}(m_2)$ and we must have $m_1, m_2 \in \mathcal{C}_m$. Hence $\mathrm{ext}(\mathcal{C}_m) \subseteq \mathrm{ext}_f(\mathcal{C})$, whereas $\mathrm{ext}_f(\mathcal{C}) \cap \mathcal{C}_m \subseteq \mathrm{ext}(\mathcal{C}_m)$ is obvious.

In addition, we claim that $\mathcal{C}_m$ is compact in the weak* topology. Indeed, by the Banach–Alaoglu theorem,

$$B_{Z^*} = \{\mu \in Z^* : \ \|\mu\|_{\mathrm{TV}} \leq 1\}$$

is weak* compact. As $\mathcal{C}_m \subseteq B_{Z^*}$, it remains to check that set $\mathcal{C}_m$ is weak* closed. We can write $\mathcal{C}_m = \mathcal{C} \cap \mathcal{P}_m$, where $\mathcal{P}_m$ stands for the set of all probability measures supported on the subsets of $\mathrm{supp}(m)$. Note that $\mathcal{P}_m$ is clearly weak* closed. Lastly, coherent distributions on $[0,1]^2$ are also weak* closed, as demonstrated in [**10**].

Thus, by Krein–Milman theorem, there exists a sequence $(m_n)_{n=1}^{\infty}$ with values in $\mathcal{C}_m$, satisfying

$$m_n = \beta_1^{(n)}\eta_1^{(n)} + \beta_2^{(n)}\eta_2^{(n)} + \cdots + \beta_{k_n}^{(n)}\eta_{k_n}^{(n)}, \tag{2.22}$$

where $\eta_1^{(n)}, \ldots, \eta_{k_n}^{(n)} \in \mathrm{ext}(\mathcal{C}_m)$ and $\beta_1^{(n)}, \ldots, \beta_{k_n}^{(n)}$ are positive numbers summing up to 1, such that

$$\int_{[0,1]^2} f \mathrm{d}m_n \longrightarrow \int_{[0,1]^2} f \mathrm{d}m, \tag{2.23}$$

for all bounded, continuous functions $f : [0,1]^2 \to \mathbb{R}$. Put $f(x,y) = |x-y|^\alpha$. By (2.23) and (2.22), we have

$$\begin{aligned}\int_{[0,1]^2} |x-y|^\alpha \mathrm{d}m &\leq \sup_{n\in\mathbb{N}} \int_{[0,1]^2} |x-y|^\alpha \mathrm{d}m_n \\ &\leq \sup_{\substack{n\in\mathbb{N},\\ 1\leq i\leq k_n}} \int_{[0,1]^2} |x-y|^\alpha \mathrm{d}\eta_i^{(n)} \\ &\leq \sup_{\eta\in\mathrm{ext}_f(\mathcal{C})} \int_{[0,1]^2} |x-y|^\alpha \mathrm{d}\eta,\end{aligned}$$

and hence

$$\sup_{(X,Y)\in\mathcal{C}_f} \mathbb{E}|X-Y|^\alpha \leq \sup_{(X,Y)\in\mathrm{ext}_f(\mathcal{C})} \mathbb{E}|X-Y|^\alpha.$$

The reverse inequality is clear. □

Now, we have the following significant reduction. Denote by $\mathcal{S}$ the family of all finite sequences $\mathbf{z} = (z_0, z_1, \ldots, z_{n+1})$, $n \in \mathbb{N}$, with $z_0 = z_{n+1} = 0$, $\sum_{i=1}^n z_i = 1$ and $z_i > 0$ for $i = 1, 2, \ldots, n$. We emphasize that $n = n(\mathbf{z})$, the length of $\mathbf{z}$, is also allowed to vary. In what follows, we will write n instead of $n(\mathbf{z})$; this should not lead to any confusion.

Proposition 2.6. *For any $\alpha \geq 1$, we have*

$$\sup_{(X,Y)\in\mathrm{ext}_f(\mathcal{C})} \mathbb{E}|X-Y|^\alpha = \sup_{z\in\mathcal{S}} \sum_{i=1}^{n} z_i \left| \frac{z_i}{z_{i-1}+z_i} - \frac{z_i}{z_i+z_{i+1}} \right|^\alpha. \tag{2.24}$$

PROOF. Consider an arbitrary $\eta \in \text{ext}_f(\mathcal{C})$ and let pair (μ, ν) be its unique representation. Recall, based on Theorem 2.3, that $\text{supp}(\eta)$ is an axial path without cycles. Set $\text{supp}(\eta) = \{(x_i, y_i)\}_{i=1}^n$ and let $q : \text{supp}(\eta) \to [0,1]$ be the quotient function associated with (μ, ν). In this setup, by (2.14) and (2.15), we can write

$$\int_{[0,1]^2} |x-y|^\alpha \, \mathrm{d}\eta = \sum_{i=1}^n z_i \left| \frac{q_{i-1}z_{i-1} + q_i z_i}{z_{i-1} + z_i} - \frac{q_i z_i + q_{i+1} z_{i+1}}{z_i + z_{i+1}} \right|^\alpha, \tag{2.25}$$

where $z_0 = z_{n+1} = 0$, $q_0 = q_{n+1} = 0$,

$$q_i = q(x_i, y_i) \quad \text{and} \quad z_i = \eta(x_i, y_i), \quad \text{for all} \;\; i = 1, 2, \ldots, n.$$

Note that if $n = 1$, then both sides of (2.25) are equal to zero; hence η does not bring any contribution to (2.24). Hence, from now on, we will assume that $n \geq 2$. Notice that by Corollary 2.1, the sequence $(q_1, q_2, \ldots, q_n)$ is given by

$$(q_1, 0, 1, 0, 1, \ldots, q_n) \quad \text{or} \quad (q_1, 1, 0, 1, 0, \ldots, q_n)$$

– except for q_1 and q_n, $(q_2, q_3, \ldots, q_{n-1})$ is simply an alternating binary sequence. Furthermore, the right-hand side of (2.25) is the sum of

$$P(q_1) := z_1 \left| q_1 - \frac{q_1 z_1 + q_2 z_2}{z_1 + z_2} \right|^\alpha + z_2 \left| \frac{q_1 z_1 + q_2 z_2}{z_1 + z_2} - \frac{q_2 z_2 + q_3 z_3}{z_2 + z_3} \right|^\alpha$$

and some other terms not involving q_1. Since $\alpha \geq 1$, P is a convex function on $[0,1]$ and hence it is maximized by some $q_1' \in \{0,1\}$; in the case of $P(0) = P(1)$, we choose q_1' arbitrarily. Depending on q_1', we shall now perform one of the following transformations $(q, z) \mapsto (\tilde{q}, \tilde{z})$:

a. If $q_1' \neq q_2$, we let $\tilde{n} = n$, $\tilde{q}_1 = q_1'$ and $\tilde{q}_i = q_i$ for $i \in \{0\} \cup \{2, 3, \ldots, n+1\}$, $\tilde{z}_i = z_i$ for $i \in \{0, 1, \ldots, n+1\}$. This operation only changes q_1 into q_1' – we increase the right-hand side of (2.25) by "correcting" the quotient function on the first atom.

b. If $q_1' = q_2$, we take $\tilde{n} = n - 1$, $\tilde{q}_0 = 0$, $\tilde{z}_0 = 0$ and

$$\tilde{q}_i = q_{i+1}, \qquad \tilde{z}_i = \frac{z_{i+1}}{z_2 + z_3 + \ldots + z_n} \qquad \text{for } i \in \{1, 2, \ldots, \tilde{n} + 1\}.$$

This modification removes the first atom and rescales the remaining ones. It is easy to see that for the transformed sequences $(\tilde{q}, \tilde{z})$, the right-hand side of (2.25) does not decrease.

Performing a similar transformation for the last summand in (2.25) (depending on q_n' and q_{n-1}) we obtain a pair of sequences $(\tilde{q}, \tilde{z})$, such that $(\tilde{q}_1, \ldots, \tilde{q}_{\tilde{n}})$ is an alternating binary sequence and

$$\begin{aligned}
\int_{[0,1]^2} |x-y|^\alpha \mathrm{d}\eta \;&\leq\; \sum_{i=1}^{\tilde{n}} \tilde{z}_i \left| \frac{\tilde{q}_{i-1}\tilde{z}_{i-1} + \tilde{q}_i \tilde{z}_i}{\tilde{z}_{i-1} + \tilde{z}_i} - \frac{\tilde{q}_i \tilde{z}_i + \tilde{q}_{i+1}\tilde{z}_{i+1}}{\tilde{z}_i + \tilde{z}_{i+1}} \right|^\alpha \\
&= \sum_{i=1}^{\tilde{n}} \tilde{z}_i \left| \frac{\tilde{z}_i}{\tilde{z}_{i-1} + \tilde{z}_i} - \frac{\tilde{z}_i}{\tilde{z}_i + \tilde{z}_{i+1}} \right|^\alpha \\
&\leq \sup_{\mathbf{z}} \sum_{i=1}^{n} z_i \left| \frac{z_i}{z_{i-1} + z_i} - \frac{z_i}{z_i + z_{i+1}} \right|^\alpha,
\end{aligned}$$

which proves the desired inequality "$\leq$" in (2.24). The reverse bound follows by a straightforward construction, involving measures with quotient functions equal to 0 or 1 (see (2.25)). $\square$

We require some further notation. Given $\alpha > 0$, let $\Phi_\alpha : \mathcal{S} \to [0,1]$ be defined by

$$\Phi_\alpha(z) = \sum_{i=1}^{n} z_i \Big| \frac{z_i}{z_{i-1}+z_i} - \frac{z_i}{z_i+z_{i+1}} \Big|^{\alpha}.$$

By the preceding discussion, for $\alpha \geq 1$ we have

$$\sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha = \sup_{z\in\mathcal{S}} \Phi_\alpha(z),$$

and our main problem amounts to the identification of

$$\limsup_{\alpha\to\infty} \Big[\alpha \cdot \sup_{z\in\mathcal{S}} \Phi_\alpha(z)\Big]. \tag{2.26}$$

It will later become clear that lim sup in (2.26) can be replaced by an ordinary limit. We begin by making some introductory observations.

Definition 2.9. *Fix $\alpha \geq 1$ and let $\boldsymbol{z} = (z_0, z_1, \ldots, z_{n+1})$ be a generic element of $\mathcal{S}$. For $1 \leq i \leq n$, we say that the term (component) z_i of $\boldsymbol{z}$ is* significant *if*

$$\sqrt{\alpha}\cdot z_{i-1} < z_i \quad \text{and} \quad \sqrt{\alpha}\cdot z_i < z_{i+1},$$

or

$$z_{i-1} > \sqrt{\alpha}\cdot z_i \quad \text{and} \quad z_i > \sqrt{\alpha}\cdot z_{i+1}.$$

The set of all significant components of z will be denoted by $\phi_\alpha(z)$.

Whenever a component z_i of $\boldsymbol{z}$ ($1 \leq i \leq n$) is not significant, we say that z_i is negligible. *The terms z_0 and z_{n+1} will be treated as neither significant nor negligible.*

Now we will show that the contribution of all negligible terms of z to the total sum $\Phi_\alpha(z)$ vanishes in the limit $\alpha \to \infty$. Precisely, we have the following.

Proposition 2.7. *For $\alpha \geq 1$ and $z \in \mathcal{S}$, we have*

$$\Phi_\alpha(z) \leq \Psi_\alpha(z) + \Big|1 - \frac{1}{1+\sqrt{\alpha}}\Big|^{\alpha},$$

where $\Psi_\alpha : \mathcal{S} \to [0,1]$ is defined by

$$\Psi_\alpha(z) = \sum_{z_i\in\phi_\alpha(z)} z_i \Big| \frac{z_i}{z_{i-1}+z_i} - \frac{z_i}{z_i+z_{i+1}} \Big|^{\alpha}.$$

Proof. Since $z_1 + z_2 + \cdots + z_n = 1$, it is sufficient to show that

$$\Big| \frac{z_i}{z_{i-1}+z_i} - \frac{z_i}{z_i+z_{i+1}} \Big| \leq \Big|1 - \frac{1}{1+\sqrt{\alpha}}\Big|, \tag{2.27}$$

for all negligible components z_i. Assume that (2.27) does not hold. Since the ratios $z_i/(z_{i-1}+z_i)$ and $z_i/(z_i+z_{i+1})$ take values in $[0,1]$, we must have

$$\min\Big\{\frac{z_i}{z_{i-1}+z_i}, \frac{z_i}{z_i+z_{i+1}}\Big\} < \frac{1}{1+\sqrt{\alpha}} \tag{2.28}$$

and

$$\max\left\{\frac{z_i}{z_{i-1}+z_i},\frac{z_i}{z_i+z_{i+1}}\right\} > \frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}. \tag{2.29}$$

It remains to note that component z_i fulfilling (2.28) and (2.29) is significant. □

It is also useful to consider some special arrangements consisting of three successive components (z_{i-1}, z_i, z_{i+1}) of the generic sequence $z \in \mathcal{S}$.

DEFINITION 2.10. *Let* $\boldsymbol{z} = (z_0, z_1, \dots, z_{n+1})$ *be an element of* $\mathcal{S}$. *For* $1 \le i \le n$, *we say that a subsequence* (z_{i-1}, z_i, z_{i+1}) *of* $\boldsymbol{z}$ *is*

- *a* split, *if* $z_{i-1} > z_i < z_{i+1}$,
- *a* peak, *if* $z_{i-1} < z_i > z_{i+1}$.

In what follows, let $\mathcal{S}'$ be the subset of all those $z \in \mathcal{S}$, which satisfy:

1. $z_{i-1} \neq z_i$ for all $i \in \{1, 2, \dots, n+1\}$,
2. there are no split subsequences in z,
3. there is exactly one peak in z,
4. there is exactly one negligible component z_{j_0} in z, and z_{j_0} is the center of the unique peak $(z_{j_0-1}, z_{j_0}, z_{j_0+1})$.

PROPOSITION 2.8. *For* $\alpha \ge 1$, *we have*

$$\sup_{z\in\mathcal{S}} \Psi_\alpha(z) \le \sup_{z\in\mathcal{S}'} \Psi_\alpha(z).$$

PROOF. Let us start by outlining the structure of the proof. Pick an arbitrary sequence $z \in \mathcal{S}$. We will gradually improve z by a series of subsequent combinatorial reductions

$$z \longrightarrow z^{(1)} \longrightarrow z^{(2)} \longrightarrow z^{(3)} \longrightarrow z^{(4)},$$

such that

$$\Psi_\alpha(z) \le \Psi_\alpha(z^{(i)}) \le \Psi_\alpha(z^{(j)}) \quad \text{for} \quad 1 \le i \le j \le 4,$$

and $z^{(i)}$ will satisfy the requirements from 1. to i. in the definition of $\mathcal{S}'$. This will give $\Psi_\alpha(z) \le \Psi_\alpha(z^{(4)})$ for some $z^{(4)} \in \mathcal{S}'$ and the claim will be proved.

1. $z \to z^{(1)}$. Put $z = (z_0, z_1, \dots, z_{n+1})$. If $z_{i-1} \neq z_i$ for all $i \in \{1, 2, \dots, n+1\}$, then we are done. Otherwise, let i_0 be the smallest index without this property. As $z_0 = 0$ and z_1 is strictly positive, we must have $i_0 > 1$. Analogously, we have $i_0 < n+1$. Consequently, observe that z_{i_0-1} and z_{i_0} are negligible. Examine the transformation $z \mapsto \tilde{z}$,

$$(\dots, z_{i_0-1}, z_{i_0}, z_{i_0+1}, \dots) \longrightarrow w^{-1} \cdot (\dots, z_{i_0-1}, z_{i_0+1}, \dots), \tag{2.30}$$
$$w = 1 - z_{i_0},$$

which removes z_{i_0} and rescales the remaining elements. If $z_{i_0+1} \in \phi_\alpha(z)$, then $w^{-1} z_{i_0+1}$ will remain a significant component of $\tilde{z}$. The contribution of z_{i_0+1} (and all the other significant components of z) to the overall sum will grow by a factor of $w^{-1} > 1$. The contribution of z_{i_0-1} to $\Psi_\alpha(z)$ is zero and it can only increase if z_{i_0-1} becomes significant. Therefore $\Psi_\alpha(z) \le \Psi_\alpha(\tilde{z})$. After a finite number of such operations, we obtain a sequence $z^{(1)}$ for which 1. holds.

2. $z^{(1)} \to z^{(2)}$. Set $z^{(1)} = (z_i^{(1)})_{i=0}^{n+1}$ and suppose that $(z_{i_0-1}^{(1)}, z_{i_0}^{(1)}, z_{i_0+1}^{(1)})$ is a split for some $i_0 \in \{2, 3, \ldots, n-1\}$ – by the definition of split configuration, i_0 must be greater than 1 and smaller than n. Accordingly, note that $z_{i_0}^{(1)}$ is negligible and consider the preliminary modification $z^{(1)} \mapsto \hat{z}^{(1)}$ given by

$$(\ldots, z_{i_0-1}^{(1)}, z_{i_0}^{(1)}, z_{i_0+1}^{(1)}, \ldots) \longrightarrow (\ldots, z_{i_0-1}^{(1)}, 0, z_{i_0+1}^{(1)}, \ldots),$$

which changes $z_{i_0}^{(1)}$ into 0 (so $\hat{z}^{(1)} \notin \mathcal{S}$: we will handle this later). As $z_{i_0-1}^{(1)} > z_{i_0}^{(1)}$, we have

$$\left| \frac{z_{i_0-1}^{(1)}}{z_{i_0-2}^{(1)} + z_{i_0-1}^{(1)}} - \frac{z_{i_0-1}^{(1)}}{z_{i_0-1}^{(1)} + z_{i_0}^{(1)}} \right| < \left| \frac{z_{i_0-1}^{(1)}}{z_{i_0-2}^{(1)} + z_{i_0-1}^{(1)}} - 1 \right|, \tag{2.31}$$

if only $z_{i_0-1}^{(1)} \in \phi_\alpha(z^{(1)})$. Similarly, as $z_{i_0}^{(1)} < z_{i_0+1}^{(1)}$, we get

$$\left| \frac{z_{i_0+1}^{(1)}}{z_{i_0}^{(1)} + z_{i_0+1}^{(1)}} - \frac{z_{i_0+1}^{(1)}}{z_{i_0+1}^{(1)} + z_{i_0+2}^{(1)}} \right| < \left| 1 - \frac{z_{i_0+1}^{(1)}}{z_{i_0+1}^{(1)} + z_{i_0+2}^{(1)}} \right|, \tag{2.32}$$

as long as $z_{i_0+1}^{(1)} \in \phi_\alpha(z^{(1)})$. By (2.31) and (2.32), with a slight abuse of notation (the domain of Ψ_α formally does not contain $\hat{z}^{(1)}$, but we may extend the definition for $\Psi_\alpha(\hat{z}^{(1)})$ in a straightforward way), we can write $\Psi_\alpha(z^{(1)}) \le \Psi_\alpha(\hat{z}^{(1)})$. Now, let us denote

$$\hat{z}^{(1,\leftarrow)} = (0, \hat{z}_1^{(1)}, \ldots, \hat{z}_{i_0-1}^{(1)}, 0)$$

and

$$\hat{z}^{(1,\rightarrow)} = (0, \hat{z}_{i_0+1}^{(1)}, \ldots, \hat{z}_n^{(1)}, 0).$$

In other words, sequences $\hat{z}^{(1,\leftarrow)}$ and $\hat{z}^{(1,\rightarrow)}$ are two consecutive parts of $\hat{z}^{(1)}$ and we can restore $\hat{z}^{(1)}$ by glueing their corresponding zeroes together. Moreover, after normalizing them by the weights

$$w^{(1,\leftarrow)} = \textstyle\sum_{i=1}^{i_0-1} \hat{z}_i^{(1)} \quad \text{and} \quad w^{(1,\rightarrow)} = \textstyle\sum_{i=i_0+1}^{n} \hat{z}_i^{(1)},$$

we get $(w^{(1,\leftarrow)})^{-1}\hat{z}^{(1,\leftarrow)}, (w^{(1,\rightarrow)})^{-1}\hat{z}^{(1,\rightarrow)} \in \mathcal{S}$. Next, in this setup, we are left with

$$\begin{aligned} \Psi_\alpha(\hat{z}^{(1)}) \; &= w^{(1,\leftarrow)} \cdot \Psi_\alpha\left(\frac{\hat{z}^{(1,\leftarrow)}}{w^{(1,\leftarrow)}}\right) \\ &+ w^{(1,\rightarrow)} \cdot \Psi_\alpha\left(\frac{\hat{z}^{(1,\rightarrow)}}{w^{(1,\rightarrow)}}\right) \\ &\le \max\left\{\Psi_\alpha\left(\frac{\hat{z}^{(1,\leftarrow)}}{w^{(1,\leftarrow)}}\right), \Psi_\alpha\left(\frac{\hat{z}^{(1,\rightarrow)}}{w^{(1,\rightarrow)}}\right)\right\}, \end{aligned}$$

where we have used $w^{(1,\leftarrow)} + w^{(1,\rightarrow)} \le 1$. Let

$$\tilde{z}^{(1)} = \arg\max\left\{\Psi_\alpha(z) : \; z \in \left\{\frac{\hat{z}^{(1,\leftarrow)}}{w^{(1,\leftarrow)}}, \frac{\hat{z}^{(1,\rightarrow)}}{w^{(1,\rightarrow)}}\right\}\right\}.$$

By the construction, we have $\Psi_\alpha(z^{(1)}) \le \Psi_\alpha(\tilde{z}^{(1)})$, the new sequence $\tilde{z}^{(1)}$ is shorter than $z^{(1)}$ and $\tilde{z}^{(1)}$ contains less split configurations than $z^{(1)}$. After repeating this procedure ($z^{(1)} \mapsto \tilde{z}^{(1)}$) multiple times, we acquire a new sequence $z^{(2)}$ obeying 1. and 2.

3. $z^{(2)} \to z^{(3)}$. Surprisingly, it is enough to put $z^{(3)} = z^{(2)}$. Indeed, we can show that sequence $z^{(2)}$ already satisfies the third condition. First, suppose that $(z^{(2)}_{j_0-1}, z^{(2)}_{j_0}, z^{(2)}_{j_0+1})$ and $(z^{(2)}_{j_1-1}, z^{(2)}_{j_1}, z^{(2)}_{j_1+1})$ are two different peaks with indices $j_0 < j_1$. Hence, as $z^{(2)}_{j_0} > z^{(2)}_{j_0+1}$ and $z^{(2)}_{j_1-1} < z^{(2)}_{j_1}$, there is at least one $i_0 \in \{j_0+1, \ldots, j_1-1\}$ at which we are forced to "flip" the direction of the previous inequality sign:

$$z^{(2)}_{j_0-1} < z^{(2)}_{j_0} > z^{(2)}_{j_0+1} > \cdots > z^{(2)}_{i_0} < \cdots < z^{(2)}_{j_1-1} < z^{(2)}_{j_1} > z^{(2)}_{j_1+1}.$$

Equivalently, this means that triple $(z^{(2)}_{i_0-1}, z^{(2)}_{i_0}, z^{(2)}_{i_0+1})$ is a split configuration. This contradicts our initial assumptions about $z^{(2)}$ (the requirement 2. is not met) and proves that there is at most one peak in $z^{(2)}$. Second, we have

$$0 = z^{(2)}_0 < z^{(2)}_1 \quad \text{and} \quad z^{(2)}_n > z^{(2)}_{n+1} = 0,$$

so there exists a point j_0 at which the direction of the inequalities must be changed from "$<$" to "$>$". Thus, there is at least one peak in $z^{(2)}$.

4. $z^{(3)} \to z^{(4)}$. Let $z^{(3)} = (z^{(3)}_i)_{i=0}^{n+1}$ and assume that $(z^{(3)}_{j_0-1}, z^{(3)}_{j_0}, z^{(3)}_{j_0+1})$ is the unique peak of $z^{(3)}$:

$$0 < z^{(3)}_1 < \cdots < z^{(3)}_{j_0-1} < z^{(3)}_{j_0} > z^{(3)}_{j_0+1} > \cdots > z^{(3)}_n > 0. \tag{2.33}$$

Further reasoning is similar to the previous ones (from points 1. and 2.), so we will just sketch it. Appropriately, if the requirement 4. is not satisfied, pick a negligible component $z^{(3)}_{i_0}$ with $i_0 \neq j_0$. Next, apply the transformation $z^{(3)} \mapsto \tilde{z}^{(3)}$ defined by (2.30), i.e. remove $z^{(3)}_{i_0}$ and rescale the remaining components. Thanks to the "single peak structure" (2.33), all the significant components of $z^{(3)}$ stay significant for $\tilde{z}^{(3)}$. The terms associated with components $z^{(3)}_i \in \phi_\alpha(z^{(3)}) \setminus \{z^{(3)}_{i_0-1}, z^{(3)}_{i_0+1}\}$ are not changed (and their contribution grows after the rescaling). Observe that the summands corresponding to $z^{(3)}_{i_0-1}$ and $z^{(3)}_{i_0+1}$ can only increase, just as in (2.31) and (2.32). Therefore $\Psi_\alpha(z^{(3)}) \le \Psi_\alpha(\tilde{z}^{(3)})$. After several repetitions and discarding of all unnecessary negligible components (beyond the central z_{j_0}), we finally obtain the desired sequence $z^{(4)} \in \mathcal{S}'$. □

We proceed to the proof of our main result.

Proof of Theorem 2.1. We will start with the lower estimate, for which the argument is simpler. By Proposition 2.6 and reformulation (2.26), for $\alpha > 2$ we have

$$\begin{aligned}\alpha \cdot \sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha &= \alpha \cdot \sup_{z\in\mathcal{S}} \Phi_\alpha(z) \\ &\ge \alpha \cdot \Phi_\alpha\left(0, \frac{1}{\alpha}, \frac{\alpha-2}{\alpha}, \frac{1}{\alpha}, 0\right) \\ &= \alpha \cdot \frac{2}{\alpha}\left|1 - \frac{1}{\alpha-1}\right|^\alpha \xrightarrow{\alpha\to\infty} \frac{2}{e}.\end{aligned}$$

We turn our attention to the upper estimate. By Propositions 2.7 and 2.8, we get

$$\alpha \cdot \sup_{(X,Y)\in\mathcal{C}} \mathbb{E}|X-Y|^\alpha \le \alpha \cdot \left(\left|1 - \frac{1}{1+\sqrt{\alpha}}\right|^\alpha + \sup_{z\in\mathcal{S}'} \Psi_\alpha(z)\right).$$

Next, because of

$$\lim_{\alpha\to\infty} \alpha \cdot \left|1 - \frac{1}{1+\sqrt{\alpha}}\right|^{\alpha} = 0,$$

it is enough to provide an asymptotic estimate for $\alpha \cdot \sup_{z\in\mathcal{S}'} \Psi_\alpha(z)$. Fix an arbitrary sequence $z = (z_0, z_1, \ldots, z_{n+1}) \in \mathcal{S}'$ and let z_{j_0} be the center of the unique peak contained in z:

$$0 < z_1 < \cdots < z_{j_0-1} < z_{j_0} > z_{j_0+1} > \cdots > z_n > 0.$$

As z_{j_0} is the only negligible component contained in z, we have

$$\sqrt{\alpha} \cdot z_i < z_{i+1} \quad \text{for} \quad 1 \le i \le j_0 - 1,$$

and

$$z_{i-1} > \sqrt{\alpha} \cdot z_i \quad \text{for} \quad j_0 + 1 \le i \le n.$$

In particular, we get $0 \le z_{j_0-1}, z_{j_0+1} < 1/\sqrt{\alpha}$. Consequently, we can write $\Psi_\alpha(z) = A + B + C$, where

$$A = \sum_{|i-j_0|>2} z_i \left| \frac{z_i}{z_{i-1}+z_i} - \frac{z_i}{z_i + z_{i+1}} \right|^{\alpha},$$

$$B = z_{j_0-2} \left| \frac{z_{j_0-2}}{z_{j_0-3}+z_{j_0-2}} - \frac{z_{j_0-2}}{z_{j_0-2}+z_{j_0-1}} \right|^{\alpha} + z_{j_0+2} \left| \frac{z_{j_0+2}}{z_{j_0+1}+z_{j_0+2}} - \frac{z_{j_0+2}}{z_{j_0+2}+z_{j_0+3}} \right|^{\alpha}$$

and

$$C = z_{j_0-1} \left| \frac{z_{j_0-1}}{z_{j_0-2}+z_{j_0-1}} - \frac{z_{j_0-1}}{z_{j_0-1}+z_{j_0}} \right|^{\alpha} + z_{j_0+1} \left| \frac{z_{j_0+1}}{z_{j_0}+z_{j_0+1}} - \frac{z_{j_0+1}}{z_{j_0+1}+z_{j_0+2}} \right|^{\alpha}.$$

We will examine these three parts separately.

The term A. Since $z_i/(z_{i-1} + z_i)$ and $z_i/(z_i + z_{i+1})$ belong to $[0, 1]$, we may write

$$\begin{aligned} A \le& \sum_{i=1}^{j_0-3} z_i + \sum_{i=j_0+3}^{n} z_i \\ <& z_{j_0-3} \cdot \sum_{i=0}^{j_0-4} \left(\frac{1}{\sqrt{\alpha}}\right)^i + z_{j_0+3} \cdot \sum_{i=0}^{n-j_0-3} \left(\frac{1}{\sqrt{\alpha}}\right)^i \\ <& (z_{j_0-1} + z_{j_0+1}) \cdot \frac{1}{\alpha} \cdot \sum_{i=0}^{\infty} \left(\frac{1}{\sqrt{\alpha}}\right)^i \\ <& \frac{2}{\alpha\sqrt{\alpha}} \cdot \sum_{i=0}^{\infty} \left(\frac{1}{\sqrt{\alpha}}\right)^i = \frac{2}{\alpha(\sqrt{\alpha}-1)} \end{aligned}$$

and hence

$$\alpha \cdot A < \frac{2}{\sqrt{\alpha}-1} \xrightarrow{\alpha\to\infty} 0.$$

The term B. We have

$$\begin{aligned} B &\leq z_{j_0-2}\left|1-\frac{z_{j_0-2}}{z_{j_0-2}+z_{j_0-1}}\right|^\alpha + z_{j_0+2}\left|\frac{z_{j_0+2}}{z_{j_0+1}+z_{j_0+2}}-1\right|^\alpha \\ &< z_{j_0-2}\left|1-\frac{z_{j_0-2}}{z_{j_0-2}+\frac{1}{\sqrt{\alpha}}}\right|^\alpha + z_{j_0+2}\left|\frac{z_{j_0+2}}{\frac{1}{\sqrt{\alpha}}+z_{j_0+2}}-1\right|^\alpha \\ &\leq 2\cdot \sup_{x\in[0,1]} x\left|1-\frac{x}{x+\frac{1}{\sqrt{\alpha}}}\right|^\alpha = \frac{2}{\sqrt{\alpha}(\alpha-1)}\cdot\left(1-\frac{1}{\alpha}\right)^\alpha. \end{aligned}$$

This yields

$$\alpha\cdot B < \frac{2\sqrt{\alpha}}{\alpha-1}\cdot\left(1-\frac{1}{\alpha}\right)^\alpha \xrightarrow{\alpha\to\infty} 0.$$

The term C. Finally, we observe that

$$\begin{aligned} C &\leq z_{j_0-1}\left|1-\frac{z_{j_0-1}}{z_{j_0-1}+z_{j_0}}\right|^\alpha + z_{j_0+1}\left|\frac{z_{j_0+1}}{z_{j_0}+z_{j_0+1}}-1\right|^\alpha \\ &\leq z_{j_0-1}\left|1-z_{j_0-1}\right|^\alpha + z_{j_0+1}\left|z_{j_0+1}-1\right|^\alpha \\ &\leq 2\cdot \sup_{x\in[0,1]} x\left|1-x\right|^\alpha = \frac{2}{\alpha+1}\cdot\left(1-\frac{1}{\alpha+1}\right)^\alpha. \end{aligned}$$

Consequently, we obtain

$$\alpha\cdot C \leq \frac{2\alpha}{\alpha+1}\cdot\left(1-\frac{1}{\alpha+1}\right)^\alpha \xrightarrow{\alpha\to\infty} \frac{2}{e}.$$

The estimates for A, B and C give the desired upper bound. The proof is done. ☐

2.4. Concluding remarks

The proof of Theorem 2.1 presented in this chapter is just an example of a novel, geometric-type approach in the exploration of coherent distributions and related inequalities. We strongly believe that our study of extreme coherent distributions will turn out useful in other applications. While some of the obtained results can be easily extended to a wider context, others appear to be more difficult to generalize. Let us include a short discussion.

A. Definition 2.1, Proposition 2.1 and Theorem 2.2 extend naturally to higher dimensions – the omitted proofs remain analogous. For arbitrary number $n \geq 2$, extreme coherent distributions on $[0,1]^n$ are exactly those, whose representations are unique and minimal.

B. It is unclear whether Theorem 2.3 enjoys any comparable counterpart for $n \geq 3$. Without resolving this open question, all plausible applications of our geometric approach are rather limited to the two-variate setup.

C. A stronger result follows at once from the proof of Proposition 2.5. We have

$$\sup_{(X,Y)\in\mathcal{C}} \mathbb{E}\tilde{f}(X,Y) = \sup_{(X,Y)\in\mathrm{ext}_f(\mathcal{C})} \mathbb{E}\tilde{f}(X,Y),$$

for every continuous function $\tilde{f} : [0,1]^2 \to [0,1]$.

D. The converse of the Theorem 2.3 does not hold true. For example, let m be a probability distribution given by
$$m\Big(\frac{1}{4}, 0\Big) = \frac{1}{3}, \quad m\Big(\frac{1}{4}, \frac{3}{4}\Big) = \frac{1}{3}, \quad m\Big(1, \frac{3}{4}\Big) = \frac{1}{3}.$$
The support of m is just a three point axial path without cycles. Next, the unique representation of m is now clearly determined by the quotient function
$$q\Big(\frac{1}{4}, 0\Big) = 0, \quad q\Big(\frac{1}{4}, \frac{3}{4}\Big) = \frac{1}{2}, \quad q\Big(1, \frac{3}{4}\Big) = 1.$$
Note that $(\frac{1}{4}, \frac{3}{4})$ is a cut point, even though it is not an endpoint of the axial path. Thus, by Corollary 2.1, coherent measure m is not an extreme point. Indeed, we have $m = \frac{m_1+m_2}{2}$, where $m_1, m_2 \in \mathcal{C}$ are given by
$$m_1\Big(\frac{1}{4}, 0\Big) = \frac{2}{3}, \quad m_1\Big(\frac{1}{4}, \frac{3}{4}\Big) = \frac{1}{3}, \quad m_2\Big(\frac{1}{4}, \frac{3}{4}\Big) = \frac{1}{3}, \quad m_2\Big(1, \frac{3}{4}\Big) = \frac{2}{3}.$$
In practice, to prove the extremality of a discrete coherent distribution (whose support is an axial path without cycles), we need to compute its quotient function.

E. In the proof of Proposition 2.6, we have applied Corollary 2.1 to justify the reduction (2.3). A similar argumentation might enable or simplify the evaluation of the quantity
$$\sup_{(X,Y)\in \mathrm{ext}_f(\mathcal{C})} \mathbb{E}\tilde{f}(X,Y),$$
for multiple (not necessarily convex) continuous functions $\tilde{f} : [0,1]^2 \to [0,1]$.

CHAPTER 3

Extreme distributions with no atoms

Preliminaries

Now, we bring back the question that appeared in the previous chapter – our present goal is to investigate the natural and intriguing problem formulated in [**4, 18, 75**]: does there exist any non-atomic extreme point of $\mathcal{C}$? We will answer this question in the affirmative.

As we have already mentioned (see [**4, 75**]), there exist extreme coherent measures with arbitrary large or even countable infinite number of atoms. For an insightful comparison, recall that martingales (M_1, M_2), whose distributions are extremal, are always supported on one or two points, see [**30**]. Next, as verified in [**4**], elements of $\text{ext}(\mathcal{C})$ are always supported on sets of Lebesgue measure zero.

Through a sophisticated use of Theorem 2.2, i.e. our novel characterization of $\text{ext}(\mathcal{C})$, we shall bridge the remaining gap between those two results. More precisely, we provide a relatively simple construction of extreme coherent distributions with an uncountable support and with no atoms. Our argument is based on classical tools and ideas from the dynamical systems theory. This unexpected connection can be regarded as an independent contribution of this chapter.

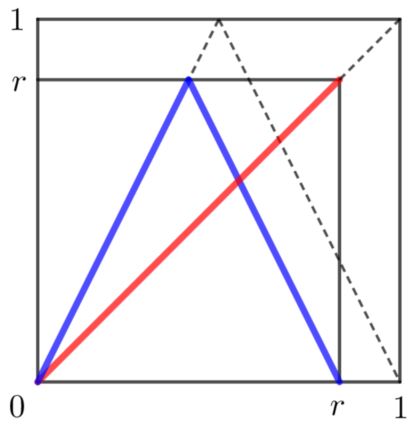

FIGURE 1. Red color stands for Δ_r, the support of measure μ_r. Blue color represents Γ_r, the support of measure ν_r. The union $\Delta_r \cup \Gamma_r$ is the support of an extreme coherent distribution m_r.

Beforehand, we need to develop some auxiliary notation. For a parameter $r \in (0, 1)$, introduce a function $t_r : [0, r] \to [0, r]$, described by $t_r(x) = 2\min\{x, r - x\}$, and distinguish the two special sets

$$\Delta_r = \{(x, x) : x \in [0, r]\} \quad \text{and} \quad \Gamma_r = \{(x, t_r(x)) : x \in [0, r]\}.$$

Next, consider a pair of nonnegative Borel measures μ_r and ν_r, supported on Δ_r and Γ_r, respectively, which are defined by

$$\mu_r\left(\{(s,s) : 0 \leq s \leq x\}\right) = c_r^{-1} \int_0^x \frac{s}{1-s} \mathrm{d}s, \tag{3.1}$$

$$\nu_r\left(\{(s,t_r(s)) : 0 \leq s \leq x\}\right) = c_r^{-1} \int_0^x 1 \mathrm{d}s, \tag{3.2}$$

for $x \in [0, r]$, where

$$c_r = \int_0^r \left(\frac{s}{1-s} + 1\right) \mathrm{d}s = -\log(1-r) < \infty. \tag{3.3}$$

See Figure 1. Finally, let $m_r = \mu_r + \nu_r$. Our main result is the following.

THEOREM 3.1. *For every $r \in (0, 1)$, the measure m_r is an extreme point of $\mathcal{C}$.*

For notational simplicity, fix any $r \in (0, 1)$. Although the choice of r is arbitrary, we shall not highlight it any further. The main tools and short outline of the proof are provided in the next section.

3.1. Main tools and outline of the proof

We begin with the following fact.

PROPOSITION 3.1. *The pair (μ_r, ν_r) is a representation of a coherent measure m_r.*

PROOF. By Proposition 2.1, we only need to check that $(\mu_r, \nu_r) \in \mathcal{R}$. Firstly, let us note that $\mu_r^x = \mu_r^y$ since the measure μ_r is concentrated on the diagonal Δ_r. Secondly, we can easily check that $\nu_r^x = \nu_r^y$ and this common measure is proportional to $\mathcal{U}(r)$, the uniform distribution on $[0, r]$:

$$(c_r/r) \cdot \nu_r^x = \mathcal{U}(r), \tag{3.4}$$

where c_r is defined as in (3.3). Hence, in accordance with the Definition 2.1, we will be done if we verify that

$$\int_A (1-x) \mathrm{d}\mu_r^x = \int_A x \mathrm{d}\nu_r^x,$$

for any Borel subset $A \in \mathcal{B}([0, r])$. However, using (3.1) and (3.2), we get

$$\int_A (1-x) \mathrm{d}\mu_r^x = \frac{1}{c_r} \int_A (1-x) \frac{x}{1-x} \mathrm{d}x,$$

and

$$\int_A x \mathrm{d}\nu_r^x = \frac{1}{c_r} \int_A x \mathrm{d}x,$$

which completes the proof. □

As previously stated, to demonstrate the primary finding of this chapter we are going to use Theorem 2.2. Therefore, in order to prove Theorem 3.1 we just need to verify that the representation (μ_r, ν_r) of m_r is unique and minimal. These two properties will be proved in Section 3.2 and Section 3.3 below. Still, a few more preliminary results and definitions will be needed in order to continue the discussion. For the sake of clarity, we split this material into two separate parts of the remaining section.

3.1.1. On the idea behind the special measure m_r. For a fixed $\eta \in [0,2]$, let $T_\eta : [0,1] \to [0,1]$ denote the celebrated "tent map" function, defined by $T_\eta(x) = \eta \cdot \min\{x, 1-x\}$. For our purposes, only the case $\eta = 2$ will be needed; the corresponding function T_2 is sometimes referred to as a "full tent map". After picking a starting point $x_0 \in [0,1]$, map T_2 defines a discrete-time dynamical system via the recurrence relation

$$x_{n+1} = T_2(x_n) = \begin{cases} 2x_n & \text{if } x_n \leq \frac{1}{2}, \\ 2(1-x_n) & \text{if } x_n > \frac{1}{2}, \end{cases}$$

for $n = 0, 1, 2, \ldots$ Now, if x_0 is irrational, then the resulting sequence $(x_n)_{n=0}^{\infty}$ becomes injective. To describe the dynamics, one often makes use of the so-called cobweb plot (or Verhulst diagram): put points successively and alternately on the main diagonal and on the graph of function T_2, according to the following algorithm:

0. place a starting point (x_0, x_0) on the diagonal,

then, for $n = 1, 2, \ldots$, repeat

1. add a point $(x_{n-1}, T_2(x_{n-1}))$ on the graph of function T_2,
2. place a point (x_n, x_n) on the diagonal and increase n by 1.

Let us consider a simple extension of this procedure: apart from drawing points on the diagram, we will also assign specific weights to each new location. Namely, for a fixed huge integer N, place (red) mass $x_n/(1-x_n)$ at every point (x_n, x_n) on the diagonal, $n = 1, 2, \ldots, N-1$, and the (blue) mass 1 for each point on the graph of function T_2. Finally, assign weight zero (and a purple color) to the initial and terminal points (x_0, x_0), (x_{N-1}, x_N). See Figure 2.

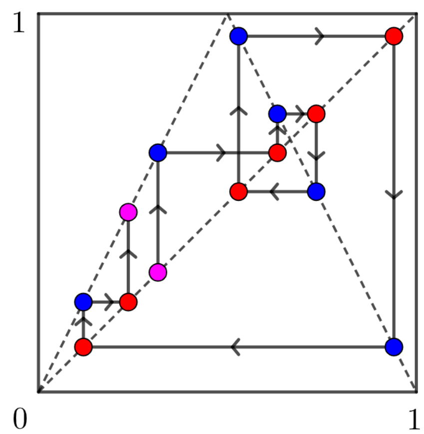

FIGURE 2. An example of a cobweb plot showing the first iterations of a full tent map system. Purple points are endpoints of an "axial path" obtained in this process. Red points represent masses placed on the diagonal, while blue points indicate masses added on the graph of function T_2.

We are ready to discuss (informally) the intuition which has led us to the measure m_r. Take the initial subsequence $(x_0, x_0), (x_0, x_1), (x_1, x_1), \ldots, (x_{N-1}, x_N)$, for some big number N. Let $\widetilde{m}^N$ be a probability distribution supported on those points, determined by their normalized weights. Then, choose a random point P on the square (distributed according to $\widetilde{m}^N$) and let vector (X, Y) be its coordinates. Let E denote the event that P belongs to the diagonal. Then, for any $1 \le i, j \le N-2$, we may write

$$\mathbb{P}(E|X = x_i) = \frac{\frac{x_i}{1-x_i}}{\frac{x_i}{1-x_i} + 1} = x_i \tag{3.5}$$

and

$$\mathbb{P}(E|Y = x_j) = \frac{\frac{x_j}{1-x_j}}{\frac{x_j}{1-x_j} + 1} = x_j, \tag{3.6}$$

where the middle fraction is the ratio of the red mass to the total mass in the line. Intuitively, equations (3.5) and (3.6) show that $\widetilde{m}^N$ is "almost coherent": the representation of this measure is the pair of its red and blue component. Now it is natural to try to carry out a limiting argument: as N goes to infinity, red points uniformly cover the entire diagonal and blue points fill up the graph of T_2. It seems plausible that $\widetilde{m}^N$ converges to a certain limiting measure $\widetilde{m} \in \operatorname{ext}(\mathcal{C})$ with no atoms. Unfortunately, this is not the case: as $\frac{x}{1-x} \to \infty$ for $x \to 1$, the red mass explodes uncontrollably near the corner $(1, 1)$.

Nevertheless, there is an easy way to overcome this obstacle. Rather than working with the transform T_2, we will simply employ its rescaled version $t_r : [0, r] \to [0, r]$,

$$t_r(x) = 2\min\{x, r - x\}.$$

Observe that images of transform T_2 and its shrunken version t_r are homothetic. Thus, if we just repeat the former reasoning with t_r acting on $[0, r]$, instead of T_2 acting on $[0, 1]$, then any point mass added to the diagonal will not exceed $\frac{r}{1-r} < \infty$.

This leads us to the consideration of a new sequence $\widetilde{m}_r^N$, defined analogously to $\widetilde{m}^N$, of "almost coherent" and "resembling extreme points" measures on $[0, r]^2$. Letting $N \to \infty$, looking at the red and blue components of $\widetilde{m}_r^N$ separately, we see that red mass approaches to μ_r and blue mass tends to ν_r. This motivates our claim: $m_r = \mu_r + \nu_r$ is an extreme coherent distribution.

3.1.2. Some basic facts from the theory of dynamical systems. To carry out a formally strict justification, we will need the following classical definition of a measure-preserving and ergodic transformation. By symbol $B \triangle C$ we will denote the symmetric difference $(B \setminus C) \cup (C \setminus B)$ of sets B and C.

DEFINITION 3.1. *Fix a constant $s > 0$ and let π be a Borel probability measure on $[0, s]$. Let $T : [0, s] \to [0, s]$ be a Borel measurable function. We say that the transformation T is*

- measure-preserving *(or equivalently, that π is T-*invariant*), if*
$$\pi(T^{-1}(A)) = \pi(A),$$
for all Borel subsets $A \in \mathcal{B}([0, s])$;
- ergodic *(or equivalently, that π is an* ergodic measure*), if T is measure-preserving and $\pi(A) \in \{0, 1\}$ for all Borel subsets $A \in \mathcal{B}([0, s])$ with*
$$\pi\left(T^{-1}(A) \triangle A\right) = 0.$$

Therefore, if a measure-preserving transformation T is not ergodic, then we can find some $A \in \mathcal{B}([0,s])$ with $T^{-1}(A) = A$ and $\pi(A) \in (0,1)$. In principle, this means that restrictions $T|_A : A \to A$ and $T|_{A^c} : A^c \to A^c$ define a non-trivial decomposition of the map T into two simpler subsystems. Worthy of note, there is a subtle similarity between the above ergodicity of transformations and minimality of representations as introduced in Definition 2.3.

Eventually, we need to make more precise the statement: "full tent map T_2 covers $[0,1]$ uniformly". The following classical result will be required, see [**11, 71**].

THEOREM 3.2 (Ergodicity of map T_2). *$\mathcal{U}(1)$, the uniform distribution on $[0,1]$, is T_2-invariant and ergodic.*

More precisely, we will employ a simple reformulation of this statement for $r < 1$.

COROLLARY 3.1. *$\mathcal{U}(r)$, the uniform distribution on $[0,r]$, is t_r-invariant and ergodic.*

PROOF. For $x \in [0,r]$, we have

$$t_r^{-1}(x) = r \cdot T_2^{-1}(x/r). \tag{3.7}$$

To check that $\mathcal{U}(r)$ is t_r-invariant, we need to show that

$$|t_r^{-1}(A)| \cdot \frac{1}{r} = |A| \cdot \frac{1}{r}, \tag{3.8}$$

for all $A \in \mathcal{B}([0,r])$, where $|\cdot|$ stands for the Lebesgue measure on the line. Yet, by (3.7) and Theorem 3.2 (as $\mathcal{U}(1)$ is T_2-invariant), we can write

$$|t_r^{-1}(A)| \cdot r^{-1} = |T_2^{-1}(A/r)| = |A/r| = |A| \cdot r^{-1},$$

which proves (3.8). Next, if t_r was not ergodic, then we could find a set $B \in \mathcal{B}([0,r])$ with $t_r^{-1}(B) = B$ and $|B| \in (0,r)$. Consequently, by (3.7) we would obtain

$$T_2^{-1}(B/r) = \frac{1}{r} \cdot t_r^{-1}(B) = B/r,$$

while $B/r \in \mathcal{B}([0,1])$ and $|B/r| \in (0,1)$. This would indicate that the map T_2 is not ergodic, contradicting Theorem 3.2. □

One more (folklore) lemma will be vital – we provide its proof for completeness.

LEMMA 3.1. *For $s > 0$, let $T : [0,s] \to [0,s]$ be a Borel measurable transformation. Let π and σ be two T-invariant probability measures on $\mathcal{B}([0,s])$ and suppose that π is additionally ergodic. If $\sigma \ll \pi$ (i.e. if σ is absolutely continuous with respect to π), then $\sigma = \pi$.*

PROOF. Let ρ denote the Radon–Nikodym derivative of σ with respect to π: for any $C \in \mathcal{B}([0,s])$ we have $\sigma(C) = \int_C \rho \, \mathrm{d}\pi$. Let us distinguish the set $A = \{\rho < 1\}$. Because π and σ are T-invariant, we may write

$$\pi\left(A \setminus T^{-1}(A)\right) = \pi\left(T^{-1}(A) \setminus A\right), \tag{3.9}$$

and

$$\sigma\left(A \setminus T^{-1}(A)\right) = \sigma\left(T^{-1}(A) \setminus A\right). \tag{3.10}$$

Hence, if $\pi(A \setminus T^{-1}(A)) > 0$, then by the very definition of A we get

$$\pi(A \setminus T^{-1}(A)) > \int_{A\setminus T^{-1}(A)} \rho \mathrm{d}\pi = \sigma(A \setminus T^{-1}(A)). \tag{3.11}$$

Analogously, we show that

$$\sigma\left(T^{-1}(A) \setminus A\right) = \int_{T^{-1}(A)\setminus A} \rho \, \mathrm{d}\pi \geq \pi\left(T^{-1}(A) \setminus A\right). \tag{3.12}$$

Combining (3.10), (3.11) and (3.12), we obtain

$$\pi\left(A \setminus T^{-1}(A)\right) > \pi\left(T^{-1}(A) \setminus A\right),$$

which contradicts (3.9) and proves that $\pi(A \triangle T^{-1}(A)) = 0$. Since measure π is ergodic, we conclude that $\pi(A) \in \{0, 1\}$. But case of $\pi(\{\rho < 1\}) = 1$ is impossible as measures π and σ are both probabilistic. Therefore we must have $\pi(\{\rho \geq 1\}) = 1$. Again, since π and σ are probabilistic, we deduce that $\rho = 1$ π-almost everywhere and thus $\sigma = \pi$. □

3.2. Proof of uniqueness

Now we will show that (μ_r, ν_r) is the only representation of measure m_r. To this end, assume that $m_r = \widetilde{\mu}_r + \widetilde{\nu}_r$ for some representation $(\widetilde{\mu}_r, \widetilde{\nu}_r) \in \mathcal{R}$. We will demonstrate that $\widetilde{\mu}_r = \mu_r$ and $\widetilde{\nu}_r = \nu_r$. From now on, to simplify the notation, we will skip the index r and write μ, ν, $\widetilde{\mu}$, $\widetilde{\nu}$, m instead of μ_r, ν_r, $\widetilde{\mu}_r$, $\widetilde{\nu}_r$ and m_r, respectively. First, fix an arbitrary $A \in \mathcal{B}([0, r])$. By Definition 2.1, we have

$$\int_A 1\mathrm{d}\mu^x = \int_A x(\mathrm{d}\nu^x + \mathrm{d}\mu^x) = \int_A x\mathrm{d}m^x,$$

and

$$\int_A 1\mathrm{d}\widetilde{\mu}^x = \int_A x(\mathrm{d}\widetilde{\nu}^x + \mathrm{d}\widetilde{\mu}^x) = \int_A x\mathrm{d}m^x,$$

so $\mu^x(A) = \widetilde{\mu}^x(A)$. Similarly, we can deduce that $\mu^y = \widetilde{\mu}^y$, and hence the marginal distributions of μ and $\widetilde{\mu}$ are equal. Because $\mu + \nu = \widetilde{\mu} + \widetilde{\nu}$, we also get

$$\nu^x = \widetilde{\nu}^x \quad \text{and} \quad \nu^y = \widetilde{\nu}^y. \tag{3.13}$$

However, directly from (3.2), we infer that $\nu^x = \nu^y$ and both these measures are proportional to the uniform distribution $\mathcal{U}(r)$. Let us inspect the finite Borel signed measure $\delta = \widetilde{\nu} - \nu$. By the Jordan decomposition theorem, we can find two nonnegative measures δ_+ and δ_-, which are mutually singular and such that

$$\delta = \delta_+ - \delta_-. \tag{3.14}$$

As a consequence of (3.13), the marginal measures δ^x and δ^y are identically equal to zero. Therefore, using (3.14) we obtain

$$\delta_+^x = \delta_-^x \quad \text{and} \quad \delta_+^y = \delta_-^y. \tag{3.15}$$

The following observation is elementary, nonetheless very helpful.

Proposition 3.2. *We have $\delta_- \leq \nu$ and $\delta_+ \leq \mu$.*

PROOF. Since δ_+ and δ_- are mutually singular, there exist two disjoint Borel sets S_+, S_- with $S_+ \cup S_- = [0,r]^2$, such that

$$\delta_+(A) = \delta_+(A \cap S_+), \tag{3.16}$$

$$\delta_-(A) = \delta_-(A \cap S_-), \tag{3.17}$$

for all Borel subsets $A \in \mathcal{B}([0,r]^2)$. To show that $\delta_- \leq \nu$, it is sufficient to check that $\delta_-(B \cap S_-) \leq \nu(B \cap S_-)$, for every choice of $B \in \mathcal{B}([0,r]^2)$. However, by (3.14) and (3.16), we can write

$$-\delta_-(B \cap S_-) = \delta(B \cap S_-) = \widetilde{\nu}(B \cap S_-) - \nu(B \cap S_-),$$

which yields

$$\nu(B \cap S_-) = \delta_-(B \cap S_-) + \widetilde{\nu}(B \cap S_-) \geq \delta_-(B \cap S_-).$$

In the same way, by (3.14) and (3.17), we get

$$\begin{aligned}\delta_+(B \cap S_+) + (\widetilde{\mu} - \mu)(B \cap S_+) &= \delta(B \cap S_+) + (\widetilde{\mu} - \mu)(B \cap S_+) \\ &= (\widetilde{\nu} + \widetilde{\mu})(B \cap S_+) - (\nu + \mu)(B \cap S_+) = 0,\end{aligned}$$

which results in

$$\mu(B \cap S_+) = \delta_+(B \cap S_+) + \widetilde{\mu}(B \cap S_+) \geq \delta_+(B \cap S_+),$$

and proves that $\delta_+ \leq \mu$. □

A central component of our argument is the next conclusion.

PROPOSITION 3.3. *Assume that δ_- is not identically equal to zero and let*

$$\alpha_- = \frac{1}{\delta_-([0,r]^2)}$$

denote its norming constant. Then $\alpha_- \cdot \delta_-^x$ is a t_r-invariant measure.

PROOF. Since δ_- is not a zero measure and we have

$$\delta_-([0,r]^2) = \delta_-^x([0,r]),$$

then $\alpha_- \cdot \delta_-^x$ is a probability measure. Now, by Proposition 3.2 we have

$$\delta_- \leq \nu \quad \text{and} \quad \delta_+ \leq \mu,$$

which leads to

$$\operatorname{supp}(\delta_-) \subseteq \operatorname{supp}(\nu) = \Gamma_r \tag{3.18}$$

and

$$\operatorname{supp}(\delta_+) \subseteq \operatorname{supp}(\mu) = \Delta_r, \tag{3.19}$$

respectively. Directly from (3.19), we see that $\operatorname{supp}(\delta_+)$ is a subset of the main diagonal, and therefore $\delta_+^x = \delta_+^y$. Consequently, due to (3.15), we get the identities

$$\delta_-^x = \delta_+^x = \delta_+^y = \delta_-^y. \tag{3.20}$$

Finally, take an arbitrary Borel set $A \in \mathcal{B}([0,r])$. From (3.18) we know that $\operatorname{supp}(\delta_-)$ is contained in Γ_r, i.e. in the graph of function t_r, and so

$$\begin{aligned}\delta_-^y(A) = \delta_-([0,r] \times A) &= \delta_-\Big[\big([0,r] \times A\big) \cap \Gamma_r\Big] \\ &= \delta_-^x\big(\{x \in [0,r] : \ t_r(x) \in A\}\big) = \delta_-^x\left(t_r^{-1}(A)\right),\end{aligned} \tag{3.21}$$

see Figure 3. Combining (3.20) and (3.21), we conclude that

$$\alpha_- \cdot \delta_-^x(A) = \alpha_- \cdot \delta_-^y(A) = \alpha_- \cdot \delta_-^x\left(t_r^{-1}(A)\right),$$

which is the claim. □

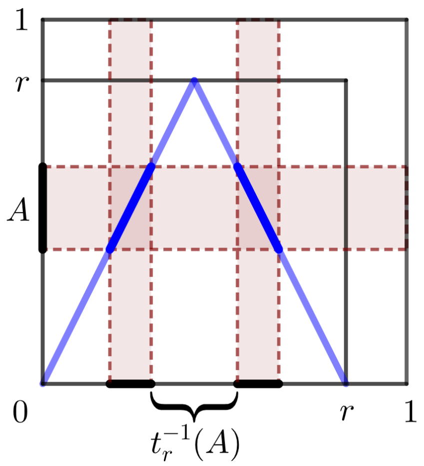

FIGURE 3. The preimage $t_r^{-1}(A)$ of a Borel set A is made up of two non-overlapping subsets with twice as small Lebesgue measure.

We proceed to the the primary objective of this section.

PROOF OF UNIQUENESS. Let us assume that δ_- is not identically equal to zero. Then by Proposition 3.2 and equality (3.4), we know that

$$\delta_-^x \le \nu^x \quad \text{and} \quad \nu^x = \frac{r}{c_r} \cdot \mathcal{U}(r),$$

and hence $\alpha_- \cdot \delta_-^x \ll \mathcal{U}(r)$. Moreover, by Proposition 3.3 and Corollary 3.1, we note that $\alpha_- \cdot \delta_-^x$ and $\mathcal{U}(r)$ are both t_r-invariant. But $\mathcal{U}(r)$ is also ergodic, so Lemma 3.1 gives $\alpha_- \cdot \delta_-^x = \mathcal{U}(r)$. Combining this with (3.15) and Proposition 3.2, we get

$$\alpha_-^{-1} \cdot \mathcal{U}(r) = \delta_-^x = \delta_+^x \le \mu^x. \tag{3.22}$$

However, directly from (3.1), for $s \in (0, s_0)$ we have

$$\mu^x([0,s]) = c_r^{-1} \int_0^s \frac{z}{1-z} \mathrm{d}z < \alpha_-^{-1} \cdot \frac{s}{r},$$

provided s_0 is sufficiently small. This contradicts (3.22) and shows that δ_- is a zero measure. Consequently, we get $\delta = \delta_+$: δ is a nonnegative measure. But (3.13) gives $\delta^x = 0$, so we must have $\delta = 0$ and $\widetilde{\nu} = \nu$. Finally, we write down the identity

$$\widetilde{\mu} = m - \widetilde{\nu} = m - \nu = \mu,$$

and get $(\widetilde{\mu}, \widetilde{\nu}) = (\mu, \nu)$. This proves that (μ, ν) is a unique representation of m. □

3.3. Proof of minimality

It remains to verify the minimality of the representation (μ, ν). Suppose that

$$\widetilde{\mu} \leq \mu \quad \text{and} \quad \widetilde{\nu} \leq \nu,$$

for some $(\widetilde{\mu}, \widetilde{\nu}) \in \mathcal{R}$. We need to show that $(\widetilde{\mu}, \widetilde{\nu})$ is proportional to (μ, ν).

Proof of minimality. Thanks to $\widetilde{\mu} \leq \mu$, we have

$$\operatorname{supp}(\widetilde{\mu}) \subseteq \operatorname{supp}(\mu) = \Delta_r. \tag{3.23}$$

In particular, $\operatorname{supp}(\widetilde{\mu})$ is a subset of the main diagonal, which implies that

$$\widetilde{\mu}^x = \widetilde{\mu}^y. \tag{3.24}$$

Next, using (3.24) and Definition 2.1, we get

$$\begin{aligned} \int_A x \mathrm{d}\widetilde{\nu}^x &= \int_A (1-x) \mathrm{d}\widetilde{\mu}^x \\ &= \int_A (1-x) \mathrm{d}\widetilde{\mu}^y = \int_A x \mathrm{d}\widetilde{\nu}^y, \end{aligned} \tag{3.25}$$

for any Borel subset $A \in \mathcal{B}([0,r])$. Due to $\widetilde{\nu} \leq \nu$, we have

$$\widetilde{\nu}^x \ll \nu^x \quad \text{and} \quad \widetilde{\nu}^y \ll \nu^y.$$

Furthermore, based on (3.4), recall that $\nu^x = \nu^y$ and ν^x is proportional to $\mathcal{U}(r)$. Applying the above observations, let φ_x, φ_y denote the Radon–Nikodym derivatives of $\widetilde{\nu}^x$ and $\widetilde{\nu}^y$ with respect to ν^x. Put

$$A_+ = \{\varphi_x > \varphi_y\} \quad \text{and} \quad A_- = \{\varphi_x < \varphi_y\}.$$

Inserting sets A_+, A_- into (3.25), we obtain

$$\int_{A_\pm} x\left(\varphi_x - \varphi_y\right) \mathrm{d}\nu^x = \int_{A_\pm} x \mathrm{d}\widetilde{\nu}^x - \int_{A_\pm} x \mathrm{d}\widetilde{\nu}^y = 0, \tag{3.26}$$

which shows that $A_+ \cup A_-$ is a set of ν^x-measure zero and hence $\widetilde{\nu}^x = \widetilde{\nu}^y$. On the other hand, we have

$$\operatorname{supp}(\widetilde{\nu}) \subseteq \operatorname{supp}(\nu) = \Gamma_r, \tag{3.27}$$

so $\operatorname{supp}(\widetilde{\nu})$ is a subset of the graph of the function t_r. Thus, we can simply repeat the reasoning from the proof of Proposition 3.3 – rewriting the equation (3.21) with δ_- replaced by $\widetilde{\nu}$, we get that $\widetilde{\nu}^x$ is a t_r-invariant measure (up to proportionality). Then, again by Lemma 3.1,

$$\widetilde{\nu}^x = \alpha \cdot \mathcal{U}(r) = \alpha' \cdot \nu^x, \tag{3.28}$$

for some factors $\alpha, \alpha' \geq 0$. Combining (3.28) with Definition 2.1, we get

$$\begin{aligned} \int_A (1-x) \mathrm{d}\widetilde{\mu}^x &= \int_A x \mathrm{d}\widetilde{\nu}^x \\ &= \alpha' \cdot \int_A x \mathrm{d}\nu^x = \alpha' \cdot \int_A (1-x) \mathrm{d}\mu^x, \end{aligned} \tag{3.29}$$

for any Borel subset $A \in \mathcal{B}([0,r])$. Just as in (3.26), plugging the Radon–Nikodym derivative of $\widetilde{\mu}^x$ with respect to μ^x into (3.29), we check that $\widetilde{\mu}^x = \alpha' \cdot \mu^x$. Altogether with (3.28), (3.23) and (3.27), this yields $(\widetilde{\mu}, \widetilde{\nu}) = \alpha' \cdot (\mu, \nu)$. Hence, the proof of the minimality is now complete. □

CHAPTER 4

Proof of the Burdzy–Pitman conjecture

Preliminaries

In this chapter we confirm the following claim stated as a conjecture in [**10**].

THEOREM 4.1. *For any threshold* $\delta \in (1/2, 1]$*, we have*

$$\mathbf{P}_{\mathcal{I}}(2, \delta) = \sup_{(X,Y)\in\mathcal{C}_{\mathcal{I}}} \mathbb{P}(|X - Y| \geq \delta) = 2\delta(1 - \delta). \tag{4.1}$$

Theorem 4.1 provides a sharp upper bound on the maximal spread of coherent opinions in the special case of two experts with access to independent sources of information. Let us point out that restricting δ to $(1/2, 1]$ does not weaken the generality of the result. Consider $X' = \mathbb{1}_A$ and $Y' = \mathbb{P}(A)$ for an arbitrary event A with $\mathbb{P}(A) = \frac{1}{2}$. It is easy to see that $(X', Y') \in \mathcal{C}_{\mathcal{I}}$. In this case, $\mathbb{P}(|X'-Y'| \geq \frac{1}{2}) = 1$. Hence, for all $\delta \in [0, 1/2]$ the problem is trivial.

Let us briefly describe our approach and the organization of the chapter. Our main intention is to use the remarkable similarity between the analysis of two-dimensional coherent vectors and closely related theory of degree sequences of bipartite graphs. In order to take advantage of the combinatorial nature of the claim made in Theorem 4.1, we start by discussing its appropriate graph-theoretic version. More precisely, we prove the following theorem.

THEOREM 4.2. *Let* $G = (U, V, E)$ *be a bipartite graph with an equal bipartition,*

$$U = \{u_1, u_2, \ldots, u_n\}, \quad V = \{v_1, v_2, \ldots, v_n\},$$

for some $n \in \mathbb{Z}_+$*. For* $n \geq k > \frac{n}{2}$ *we have*

$$\sum_{1\leq i,j\leq n} \mathbb{1}\Big\{|\deg(u_i) - \deg(v_j)| \geq k\Big\} \leq 2k(n - k). \tag{4.2}$$

Note that the trivial upper bound n^2 is the best possible upper bound in the case $k \leq \frac{n}{2}$. The proof of Theorem 4.2, given in Section 4.2, is based on an idea similar to the "spread bounding theorem" of Erdős, Chen, Rousseau and Schelp – Section 4.1 is solely devoted to the conscious explanation of this relationship. Later in Section 4.2 we provide an elementary example showing that the bound (4.2) is sharp.

In Section 4.3 we show how to transform the initial Theorem 4.1 to Theorem 4.2. To this end, we make use of an appropriate sampling construction, similar in spirit to [**58**]. The key idea is to approximate a fixed coherent distribution with a randomly generated sequence of graphs. We then apply Theorem 4.2 to each of the graphs in the sequence and obtain (4.1) by passing to the limit.

4.1. Relation with the Erdős "spread bounding theorem"

Let pair $G = (V, E)$ be a simple graph: G consists of vertex set $V = \{v_1, v_2, \ldots, v_n\}$ and edge set E (given collection of two-element subsets of V). This section discusses a simplified version of Theorem 4.2 and its connections with classical graph theory.

THEOREM 4.3. *Let G be a simple graph with $n \geq 2$ vertices. For $n > k > \frac{n}{2}$ we have*

$$\sum_{1 \leq i < j \leq n} \mathbb{1}\Big\{|\deg(v_i) - \deg(v_j)| \geq k\Big\} \leq (k+1)(n-k-1). \tag{4.3}$$

In this setup, let us start with some auxiliary terminology.

DEFINITION 4.1. *For a vertex set $B \subseteq V$, we define the "spread" of B as*

$$\operatorname{sp}(B) = \max_{u,v \in B} |\deg(u) - \deg(v)|.$$

Then, for a fixed integer $k \geq 0$, the "k-th spread" of a simple graph G is

$$\operatorname{sp}(G, k) = \max\{|B| : B \subseteq V \wedge \operatorname{sp}(B) \leq k\}.$$

EXAMPLE 4.1. In each simple graph G with $|V| \geq 2$, there are at least two vertices $u, v \in V$ such that $\deg(u) = \deg(v)$. Indeed, otherwise we must have

$$\{\deg(v) : v \in V\} = \{0, 1, \ldots, |V| - 1\}.$$

This yields a clear contradiction: there is an isolated (degree 0) and dominant vertex (degree $|V|-1$) at the same time. We can reformulate this result as $\operatorname{sp}(G, 0) \geq 2$. △

The next theorem reveals an important extension of this uncomplicated example. For the proof, see [**33, 13**].

THEOREM 4.4 (Erdős, Chen, Rousseau, Schelp). *Let $G = (V, E)$ be a simple graph with $|V| \geq k+2$ vertices. Then, we have* $\operatorname{sp}(G, k) \geq k + 2$.

PROOF OF THEOREM 4.3. By Theorem 4.4, there is a subset $S \subseteq V$ such that $\operatorname{sp}(S) \leq k - 1$ and $|S| = k + j$ for some $j \in \{1, 2, \ldots, n - k\}$. Moreover, we further assume that S is the maximal set with those properties. Let m and M denote the minimum and maximum degree of vertices in S. We introduce the sets

$$A = \{v \in V : \deg(v) < m\}, \quad B = \{v \in V : \deg(v) > M\},$$

and put $a = |A|$, $b = |B|$. Therefore, from $A \cup B = V \setminus S$ and $A \cap B = \emptyset$, we have $a + b = n - k - j$. From the maximality of S, we can also write

$$\deg(u_A) \leq n/2 \leq \deg(u_B) \quad \text{for all } u_A \in A, u_B \in B,$$

and hence $\max\{\operatorname{sp}(A), \operatorname{sp}(B)\} < k$. Next, let us denote

$$S_A = \{v \in S : \ \deg(v) - \deg(u) \geq k \ \text{ for some } u \in A\},$$

$$S_B = \{v \in S : \ \deg(u) - \deg(v) \geq k \ \text{ for some } u \in B\}.$$

Define $s_a = |S_A|$, $s_b = |S_B|$. As a consequence of $k > n/2$, we get $S_A \cap S_B = \emptyset$ and $s_a + s_b \leq k + j$. Combining all the above observations, we obtain the inequality

$$\sum_{1 \leq i < j \leq n} \mathbb{1}\Big\{|\deg(v_i) - \deg(v_j)| \geq k\Big\} \leq ab + as_a + bs_b,$$

where the right handside can be further estimated by

$$\begin{aligned}&\le \min\{a,b\}(a+b)+\max\{a,b\}(s_a+s_b)\\&\le \min\{a,b\}(n-k-j)+\max\{a,b\}(k+j)\\&\le (a+b)\max\{n-k-j,k+j\}\\&=(n-k-j)(k+j)\\&\le (n-k-1)(k+1),\end{aligned}$$

which ends the proof. □

4.2. Number of high degree differences in bipartite graphs

Let $G=(U,V,E)$ be a bipartite graph with an equal bipartition, that is a triplet

$$U=\{u_1,u_2,\dots,u_n\},\quad V=\{v_1,v_2,\dots,v_n\},\quad \text{and}\quad E\subseteq U\times V,$$

for some fixed $n\in\mathbb{Z}_+$. Let us fix a natural number k satisfying $n\ge k>\frac{n}{2}$. Hereinafter, we denote the degree sequences of G as $(\alpha_i)_{i=1}^n$ and $(\beta_j)_{j=1}^n$, i.e.

$$\alpha_i=\deg(u_i)\quad\text{and}\quad \beta_j=\deg(v_j),$$

for all $1\le i,j\le n$. Without loss of generality we can also assume that

$$\begin{aligned}\alpha_1\ge\alpha_2\ge\cdots\ge\alpha_n,\\\beta_1\ge\beta_2\ge\cdots\ge\beta_n.\end{aligned}$$

We start with an observation similar to the spread bounding theorem of Erdős et al.

LEMMA 4.1. *There exist $s,t\in\{1,2,\dots,n-k+1\}$ such that $\alpha_s\le\beta_{s+k-1}+k-1$ and $\beta_t\le\alpha_{t+k-1}+k-1$.*

PROOF. We will prove only the existence of s, as the case of t is analogous. Assume for the sake of contradiction that such a number s does not exists. Therefore, the total number of edges incident to $u_1,u_2,\dots,u_{n-k+1}$ is at least

$$\beta_k+\beta_{k+1}+\cdots+\beta_n+k(n-k+1).$$

Observe that at least $k(n-k+1)$ of these edges go to vertices $v_1,v_2,\dots,v_{k-1}$. Let us denote

$$\tilde{E}:=E\cap\Big(\{u_1,u_2,\dots,u_{n-k+1}\}\times\{v_1,v_2,\dots,v_{k-1}\}\Big).$$

We have just shown that $|\tilde{E}|\ge k(n-k+1)$. On the other hand, we clearly have $|\tilde{E}|\le(k-1)(n-k+1)$, which is a contradiction. □

PROOF OF THEOREM 4.2. For $1\le i,j\le n$, let us call (i,j)

- an $\mathcal{A}$-pair if $\alpha_i\ge\beta_j+k$,
- a $\mathcal{B}$-pair if $\beta_j\ge\alpha_i+k$.

Since $k>\frac{n}{2}$, we have $\alpha_i>\frac{n}{2}$ for all $\mathcal{A}$-pairs (i,j) and $\alpha_i<\frac{n}{2}$ for all $\mathcal{B}$-pairs (i,j). As a consequence, there exists an $i_0\in\{1,2,\dots,n+1\}$ such that:

1. $i\le i_0-1$ for any $\mathcal{A}$-pair (i,j),
2. $i\ge i_0$ for any $\mathcal{B}$-pair (i,j).

Analogously, there exists $j_0\in\{1,2,\dots,n+1\}$ such that:

3. $j \leq j_0 - 1$ for any $\mathcal{B}$-pair (i,j),
4. $j \geq j_0$ for any $\mathcal{A}$-pair (i,j).

Observe that by Lemma 4.1,

5. for any $\mathcal{A}$-pair (i,j) either $i < s$ or $j > s+k-1$,
6. for any $\mathcal{B}$-pair pair (i,j) either $j < t$ or $i > t+k-1$.

We will now show that conditions 1–6 imply that the total number of $\mathcal{A}$-pairs and $\mathcal{B}$-pairs is at most $2k(n-k)$. Let us fix $i_0, j_0 \in \{1,2,\ldots,n+1\}$. First, we will show that it is sufficient to consider only s and t such that $s,t \in \{1, n-k+1\}$ because these values of s and t are optimal in the sense that they maximize the total number of pairs (i,j) fulfilling all conditions 1–6.

Note that the variable s appears only in the 5-th condition and thus the value of s is not relevant for bounding the number of $\mathcal{B}$-pairs. Moreover, observe that if $i_0 \leq n-k+1$, then for $s = n-k+1$ condition 5 is automatically fulfilled and thus $s = n-k+1$ is an optimal value. Similarly, if $j_0 \geq k+1$, then for $s = 1$ condition 5 is also automatically fulfilled and $s = 1$ is an optimal value. Finally, let us assume that $i_0 \geq n-k+2$ and $j_0 \leq k$. In this case, the restrictions imposed by condition 5 remove exactly $(i_0 - s)(s+k-j_0)$ additional pairs. Therefore, as the last expression is a concave function of $s \in [1, n-k+1]$, it is minimized in one of the endpoints. Hence we may assume that $s = 1$ or $s = n-k+1$, as desired. Analogously, we show that $t = 1$ or $t = n-k+1$ is optimal. There are four possible cases now:

a. $s = 1,\ t = n-k+1$. We have $j \geq k+1$ for all $\mathcal{A}$-pairs and $j \leq n-k$ for all $\mathcal{B}$-pairs (i,j). Thus any i participates in at most $n-k$ of $\mathcal{A}$-pairs and in at most $n-k$ of $\mathcal{B}$-pairs. Therefore, since a fixed vertex can not participate in both types of pairs, every i participates overall in at most $n-k$ pairs. As a consequence, the total number of pairs does not exceed $n(n-k) < 2k(n-k)$.

b. $s = n-k+1,\ t = 1$. This case is symmetric to the previous one.

c. $s = 1,\ t = 1$. We have $j \geq k+1$ for all $\mathcal{A}$-pairs and $i \geq k+1$ for all $\mathcal{B}$-pairs (i,j). Let us denote $a := \max(k+1, j_0)$ and $b := \max(k+1, i_0)$. Then the total number of $\mathcal{A}$-pairs is bounded by $(n-a+1)(b-1)$, while the total number of $\mathcal{B}$-pairs is at most $(n-b+1)(a-1)$. Notice, that for $a, b \in [k+1, n+1]$ the sum

$$S := (n-a+1)(b-1) + (n-b+1)(a-1),$$

is bilinear and it is maximized at one of four corners. For $a = b = k+1$, we get $S = 2k(n-k)$. For, say $a = n+1$, we get

$$S = n(n-b+1) \leqslant n(n-k) < 2k(n-k).$$

d. $s = n-k+1,\ t = n-k+1$. This case is analogous to c.

Hence we have shown that Theorem 4.2 holds in all cases. This ends the proof. □

We end this section with an example showing that the upper bound $2k(n-k)$ in (4.2) cannot be improved. Note that a straightforward modification of this example shows that $2\delta(1-\delta)$ in (4.1) is also sharp.

EXAMPLE 4.2. Consider $n, k \in \mathbb{Z}_+$, with $n \geq k > \frac{n}{2}$. Let $G_{n,k} = (U, V, E)$, where $U = V = \{1, 2, \ldots, n\}$ and $E = \{(u, v) \in U \times V : \max(u, v) \leq k\}$. We clearly have

$$\sum_{1 \leq i,j \leq n} \mathbb{1}\Big\{|\deg(u_i) - \deg(v_j)| \geq k\Big\} = 2k(n-k). \quad \triangle$$

Moreover, one can check that inequality (4.2) becomes an equality exactly for those graphs G that are isomorphic to $G_{n,k}$ or to its complement $\overline{G}_{n,k}$. This follows easily from the proof of Theorem 4.2 and we leave the details to interested reader.

4.3. Proof of the Burdzy–Pitman conjecture

By $\mathcal{C}_\mathcal{I}(m)$ we denote the set of $(X, Y) \in \mathcal{C}_\mathcal{I}$ such that both X and Y take at most m different values. In what follows, fix any $\delta \in (1/2, 1]$.

PROPOSITION 4.1. *To prove Theorem 4.1 it is enough to verify it for all vectors $(X, Y) \in \mathcal{C}_\mathcal{I}(m)$, $m \geq 1$.*

PROOF. Fix an arbitrary coherent vector $(X, Y) \in \mathcal{C}_\mathcal{I}$ and choose (X_m, Y_m) as in Proposition 1.2. By the triangle inequality we get

$$\mathbb{P}(|X - Y| \geq \delta) \leq \mathbb{P}(|X_m - Y_m| \geq \delta - 2/m).$$

Thus, assuming that Theorem 4.1 is true for all $(X, Y) \in \cup_{m=1}^{\infty} \mathcal{C}_\mathcal{I}(m)$, for m large enough so that $\delta - 2/m > 1/2$, we obtain

$$\mathbb{P}(|X - Y| \geq \delta) \leq 2(\delta - 2/m)(1 - \delta + 2/m).$$

Letting $m \to \infty$ completes the proof. $\square$

We are now able to prove our main result.

PROOF OF THEOREM 4.1. Fix $(X, Y) \in \bigcup_{m=1}^{\infty} \mathcal{C}_\mathcal{I}(m)$. Hence, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, independent sub σ-fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ and an event $A \in \mathcal{F}$, such that $X = \mathbb{E}(\mathbb{1}_A | \mathcal{G})$ and $Y = \mathbb{E}(\mathbb{1}_A | \mathcal{H})$. Furthermore, for some $N, M \in \mathbb{Z}_+$, we may suppose that X and Y takes values

$$x_1, x_2, \ldots, x_N \quad \text{on sets} \quad G_1, G_2, \ldots, G_N,$$

$$y_1, y_2, \ldots, y_M \quad \text{on sets} \quad H_1, H_2, \ldots, H_M,$$

respectively. We can also assume, without loss of generality, that

$$\mathcal{G} = \sigma\Big(G_1, G_2, \ldots, G_N\Big) \quad \text{and} \quad \mathcal{H} = \sigma\Big(H_1, H_2, \ldots, H_M\Big).$$

For $1 \leq i \leq N$ and $1 \leq j \leq M$, let $p_i = \mathbb{P}(G_i), q_j = \mathbb{P}(H_j)$ and

$$\rho_{i,j} = \frac{\mathbb{P}(G_i \cap H_j \cap A)}{\mathbb{P}(G_i \cap H_j)}.$$

Then, by independence we have $\mathbb{P}(G_i \cap H_j) = p_i q_j$ and

$$x_i = \sum_{j=1}^{M} q_j \rho_{i,j}, \quad 1 \leq i \leq N, \tag{4.4}$$

$$y_j = \sum_{i=1}^{N} p_i \rho_{i,j}, \quad 1 \le j \le M. \tag{4.5}$$

First, we show how to construct a sequence of bipartite graphs $G_n = (U_n, V_n, E_n)$ with $|U_n| = |V_n| = n$, such that:

(C1) there are $p_i n + O(n^{3/4})$ vertices in U_n of degree $x_i n + O(n^{3/4})$, $i = 1, 2, \ldots, N$,

(C2) there are $q_j n + O(n^{3/4})$ vertices in V_n of degree $y_j n + O(n^{3/4})$, $j = 1, 2, \ldots, M$,

where by $O(n^{3/4})$ we denote any quantity bounded in magnitude by $Cn^{3/4}$ for some constant $C < \infty$ independent of n, N, M, i and j.

Fix $n \ge 1$ and choose n independent points $u_1, u_2, \ldots, u_n$ in the initial space Ω (distributed according to $\mathbb{P}$) and for $1 \le i \le n$ denote $\alpha_i = s$ if $u_i \in G_s$. In other words, $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is an i.i.d. sample from the set $\{1, 2, \ldots, N\}$ with weights $p_1, p_2, \ldots, p_N$, respectively.

We can think about this sample as a randomly generated sequence of labels. Let $A_s = \sum_{i=1}^{n} \mathbb{1}_{\{\alpha_i = s\}}$ be the number of labels equal to s, $1 \le s \le N$. Observe that A_s is the sum of n independent Bernoulli random variables. Next, we will apply the celebrated Hoeffding's inequality, see [**47, 73**].

THEOREM 4.5 (Hoeffding). *Let $X_1, \ldots, X_n$ be independent random variables with $a_i \le X_i \le b_i$ almost surely, $i = 1, 2, \ldots, n$. Put $S_n = X_1 + X_2 + \cdots + X_n$. We have*

$$\mathbb{P}(|S_n - \mathbb{E}S_n| \ge nr) \le 2 \cdot \exp\left(-\frac{2n^2 r^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \quad \textit{for all } r > 0.$$

Hence, by an application of Hoeffding's inequality, we get

$$\mathbb{P}(|A_s - np_s| \ge nr) \le 2 \cdot e^{-2nr^2} \quad \text{for all } r > 0.$$

Consequently, for $r = n^{-1/4}$ we have $\mathbb{P}(|A_s - np_s| \ge n^{3/4}) \le 2 \cdot e^{-2\sqrt{n}}$. Thus, for large n, with high probability we get $|A_s - np_s| < n^{3/4}$ simultaneously for all $1 \le s \le N$.

Analogously, we choose n points $v_1, v_2, \ldots, v_n$ in ground space Ω and generate an i.i.d. sample $(\beta_1, \beta_2, \ldots, \beta_n)$ from the set $\{1, 2, \ldots, M\}$ with weights $q_1, q_2, \ldots, q_M$. If $B_t = \sum_{j=1}^{n} \mathbb{1}_{\{\beta_j = t\}}$ for $1 \le t \le M$, then for large n, with high probability we have $|B_t - nq_t| < n^{3/4}$ for all t simultaneously.

Given the points $(u_i)_{i=1}^n$ and $(v_j)_{j=1}^n$ and the corresponding labels $(\alpha_i)_{i=1}^n$ and $(\beta_j)_{j=1}^n$, we will generate independently edges of a random bipartite graph (U_n, V_n, E_n), where

$$U_n = \{u_1, u_2, \ldots, u_n\} \quad \text{and} \quad V_n = \{v_1, v_2, \ldots, v_n\}.$$

The subscripts on $\mathbb{P}_{\alpha,\beta}$ and $\mathbb{E}_{\alpha,\beta}$ will denote conditioning on $(\alpha_i)_{i=1}^n$ and $(\beta_j)_{j=1}^n$.

(1) Generate independent indicator variables $Z_{i,j}$, for $1 \le i, j \le n$ satisfying

$$\mathbb{P}_{\alpha,\beta}(Z_{i,j} = 1) = 1 - \mathbb{P}_{\alpha,\beta}(Z_{i,j} = 0) = \rho_{\alpha_i, \beta_j},$$

(2) for $1 \le i, j \le n$, set $(u_i, v_j) \in E_n$ iff $Z_{i,j} = 1$.

Hence, $Z_{i,j} = \mathbb{1}_{\{(u_i,v_j)\in E_n\}}$. For $1 \le i \le n$, we have

$$\mathbb{E}_{\alpha,\beta}\mathrm{deg}(u_i) = \mathbb{E}_{\alpha,\beta}\left(\sum_{j=1}^{n} Z_{i,j}\right) = \sum_{t=1}^{M} B_t \rho_{\alpha_i,t} = \sum_{t=1}^{M}\Big(nq_t + O(n^{3/4})\Big)\rho_{\alpha_i,t},$$

and hence, by (4.4),

$$\mathbb{E}_{\alpha,\beta}\mathrm{deg}(u_i) = nx_{\alpha_i} + O(n^{3/4}). \tag{4.6}$$

Similarly, for $1 \le j \le n$, by (4.5) we get

$$\mathbb{E}_{\alpha,\beta}\mathrm{deg}(v_j) = ny_{\beta_j} + O(n^{3/4}). \tag{4.7}$$

We apply Hoeffdings's inequality again to obtain

$$\mathbb{P}_{\alpha,\beta}\Big(|\mathrm{deg}(u_i) - \mathbb{E}_{\alpha,\beta}\mathrm{deg}(u_i)| \ge n^{3/4}\Big) \le 2\cdot e^{-2\sqrt{n}}, \tag{4.8}$$

and

$$\mathbb{P}_{\alpha,\beta}\Big(|\mathrm{deg}(v_j) - \mathbb{E}_{\alpha,\beta}\mathrm{deg}(v_j)| \ge n^{3/4}\Big) \le 2\cdot e^{-2\sqrt{n}}, \tag{4.9}$$

for all $i,j \in \{1,2,\dots,n\}$.

Note that the concentration rates (4.8) and (4.9) are exponential in $\sqrt{n}$. Thus, since n is large, with high probability all these inequalities hold simultaneously. Then, by (4.6) and (4.7), we have

$$\mathrm{deg}(u_i) = nx_{\alpha_i} + O(n^{3/4}) \quad \text{and} \quad \mathrm{deg}(v_j) = ny_{\beta_j} + O(n^{3/4})$$

for all $i,j \in \{1,2,\dots,n\}$ with high probability. This, together with bounds on $(A_s)_{s=1}^N$ and $(B_t)_{t=1}^M$, proves that (deterministic graphs) G_n's satisfying conditions (C1)-(C2) exist for large n.

In what follows, we add additional subscripts and write $u_i^{(n)}$ and $v_j^{(n)}$ for generic elements of U_n and V_n, respectively. We can now write

$$\mathbb{P}(|X-Y| \ge \delta) = \sum_{\substack{1\le i\le N\\ 1\le j\le M}} \mathbb{1}_{\{|x_i - y_j| \ge \delta\}}\cdot p_i q_j$$

$$= \lim_{n\to\infty}\frac{1}{n^2}\sum_{\substack{1\le i\le N\\ 1\le j\le M}} \mathbb{1}_{\{|nx_i - ny_j| \ge n\delta\}}\cdot\Big(p_i n + O(n^{3/4})\Big)\Big(q_j n + O(n^{3/4})\Big). \tag{4.10}$$

By the triangle inequality

$$|nx_{\alpha_i} - ny_{\beta_j}| \le |\mathrm{deg}(u_i^{(n)}) - \mathrm{deg}(v_j^{(n)})| + 2\cdot O(n^{3/4}),$$

for all $i,j \in \{1,2,\dots,n\}$. This and (C1)-(C2) imply that we can bound the right hand side of (4.10) by

$$\le \limsup_{n\to\infty}\frac{1}{n^2}\sum_{1\le i,j\le n}\mathbb{1}\Big\{|\mathrm{deg}(u_i^{(n)}) - \mathrm{deg}(v_j^{(n)})| \ge n\delta - 2O(n^{3/4})\Big\}.$$

Finally, applying Theorem 4.2 to bipartite graphs G_n, we obtain

$$\le \limsup_{n\to\infty}\frac{1}{n^2}\cdot 2\Big(n\delta - 2O(n^{3/4})\Big)\Big(n - n\delta + 2O(n^{3/4})\Big) = 2\delta(1-\delta),$$

which ends the proof. □

Part 2

Multivariate coherent distributions

CHAPTER 5

Maximal difference among expert's opinions

Preliminaries

Now, for a fixed $\alpha \in \mathbb{R}_+$ and $n \geq 2$, consider the number

$$\mathbf{E}(n,\alpha) = \sup_{(X_1,X_2,\ldots,X_n)\in\mathcal{C}_n} \mathbb{E} \max_{1\leq i<j\leq n} |X_i - X_j|^\alpha.$$

The question about the explicit formula for this supremum can be regarded as the precise mathematical reformulation of the commonsense and captivating problem: how large is the maximal spread of multivariate coherent opinions? Significantly, this question also constitutes a natural generalization of the martingale diameter problem, see e.g. [**27, 62**]. The primary goal of this chapter is to introduce a relevant symmetrization technique and carry out a complete analysis in the basic case of $\alpha = 1$. Our main result is the following.

Theorem 5.1. *We have*

$$\mathbf{E}(n,1) = \sup_{(X_1,X_2,\ldots,X_n)\in\mathcal{C}_n} \mathbb{E} \max_{1\leq i<j\leq n} |X_i - X_j| = \begin{cases} \frac{1}{2} & \text{if } n=2,\\ 2-\sqrt{2} & \text{if } n=3,\\ \frac{7}{2}-2\sqrt{2} & \text{if } n=4,\\ \dfrac{n-2}{n-1} & \text{if } n\geq 5. \end{cases} \tag{5.1}$$

Note that the above supremum behaves in a quite surprising manner: we have a few "irregular" terms corresponding to small values of n, and then, for $n \geq 5$, it is given by a simple and compact expression. We would like to point out that the result in the special case of $n = 2$ has already been known in the literature, see [**10**].

Let us quickly discuss our approach and the organization of the chapter. In the next section we present our novel symmetrization procedure and establish its essential properties – the results of this section will be equally important in the next chapter. Consequently, in Section 5.2, we apply this technique to the left-hand side of (5.1) and reduce the remaining evaluation to the analysis of $\sup \mathbb{E} \max_{1\leq i\leq n} X_i$, where the supremum is taken over all coherent vectors satisfying certain additional constraints.

Then, in Section 5.3, using various combinatorial arguments, we gradually simplify the context. We show that the optimal coherent vectors (those for which the simpler supremum is attained) can be assumed to satisfy more and more useful properties. After several steps, this allows us to express the supremum as the extremal value of a certain function of one variable, which in turn can be computed explicitly.

Deserving of note, our approach was greatly inspired by the influential paper [**9**]: in that article, a related problem for coherent vectors was also studied with the use of a certain discretization and subsequent combinatorial reductions.

5.1. Symmetrization procedure

We have the following (relatively simple but very helpful) observation.

LEMMA 5.1. *Assume that* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space and let* $\{G_1, G_2, \ldots, G_m\}$ *be a finite partition of* Ω. *For an arbitrary event* $A \in \mathcal{F}$, *we put*

$$Y = \mathbb{E}(\mathbb{1}_A | \sigma(G_1, G_2, \ldots, G_m)).$$

Then for any $y \in (0, 1]$ *such that* $\mathbb{P}(Y = y) > 0$, *we have*

$$\mathbb{P}(\{Y = y\} \cap A^c) = \mathbb{P}(\{Y = y\} \cap A) \cdot \frac{1-y}{y}.$$

PROOF. For any $G \in \sigma(G_1, G_2, \ldots, G_m)$ such that $\mathbb{E}(\mathbb{1}_A | G) = y$, we write

$$y = \frac{\mathbb{P}(A \cap G)}{\mathbb{P}(G)} = \frac{\mathbb{P}(A \cap G)}{\mathbb{P}(A \cap G) + \mathbb{P}(A^c \cap G)}.$$

This is equivalent to

$$y \cdot \Big(\mathbb{P}(A \cap G) + \mathbb{P}(A^c \cap G)\Big) = \mathbb{P}(A \cap G),$$

or

$$\mathbb{P}(A^c \cap G) = \frac{1-y}{y} \cdot \mathbb{P}(A \cap G).$$

It remains to take $G = \{Y = y\}$. Indeed, we clearly have $G \in \sigma(G_1, G_2, \ldots, G_m)$, since Y is measurable with respect to the latter σ-algebra. $\square$

Next we will describe a useful symmetrization procedure, which will later allow us to replace the left-hand side of (5.1) with a more regular expression. To achieve this goal, we need to introduce some additional notation.

DEFINITION 5.1. *For a positive integer* m, *let* $\mathcal{C}(n, m)$ *be the family of all those vectors* $X = (X_1, X_2, \ldots, X_n) \in \mathcal{C}_n$ *such that each* X_j *takes at most* m *different values,* $j = 1, 2, \ldots, n$.

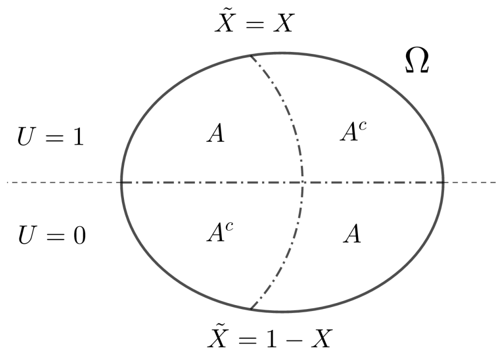

FIGURE 1. The modified sequence $\widetilde{X}$.

Fix a positive integer m and suppose that $X \in \mathcal{C}(n, m)$ is a coherent vector with

$$X_i = \mathbb{E}(\mathbb{1}_A | \mathcal{G}_i), \quad i = 1, 2, \ldots, n.$$

Furthermore, let U be a random variable independent of $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ and A, having the two-point distribution $\mathbb{P}(U=0)=\mathbb{P}(U=1)=1/2$. Then modification $\widetilde{X}$, the mixture of vectors X and $1-X$, is given by

$$(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n) = U \cdot (X_1, X_2, \ldots, X_n) + (1-U)\cdot(1-X_1, 1-X_2, \ldots, 1-X_n),$$

see Figure 1. Furthermore, we define the mixture $\widetilde{A}$ of A and A^c by the requirement $\mathbb{1}_{\widetilde{A}} = \widetilde{\mathbb{1}_A}$, or more explicitly, $\widetilde{A} = (A\cap\{U=1\})\cup(A^c\cap\{U=0\})$. Let us distinguish the σ-algebras $\widetilde{\mathcal{G}_i} = \sigma(\mathcal{G}_i, U)$, $i=1,2,\ldots,n$. The key properties of these objects are summarized in a supporting lemma below.

LEMMA 5.2. *With the above notation, the following holds true.*

(i) We have $\mathbb{P}(\widetilde{A}) = \frac{1}{2}$, $\widetilde{X} \in \mathcal{C}(n, 2m)$ *and* $\widetilde{X}_i = \mathbb{E}(\mathbb{1}_{\widetilde{A}}|\widetilde{\mathcal{G}_i})$ *for all* i.

(ii) For any sequence $(x_i)_{i=1}^n \subseteq [0,1]$, *we have the identity*

$$\mathbb{P}\Big(\bigcap_{i=1}^n \{\widetilde{X}_i = x_i\} \cap \widetilde{A}\Big) = \mathbb{P}\Big(\bigcap_{i=1}^n \{\widetilde{X}_i = 1-x_i\} \cap \widetilde{A}^c\Big). \tag{5.2}$$

(iii) For any $x \in (0,1]$,

$$\frac{1-x}{x}\cdot\sum_{i=1}^n \mathbb{P}\Big(\{\widetilde{X}_i = x\}\cap\widetilde{A}\Big) = \sum_{i=1}^n \mathbb{P}\Big(\{\widetilde{X}_i = 1-x\}\cap\widetilde{A}\Big). \tag{5.3}$$

PROOF. Since U is measurable with respect to $\widetilde{\mathcal{G}_i}$ and independent of A, we get

$$\mathbb{E}\big(\mathbb{1}_{\widetilde{A}}|\widetilde{\mathcal{G}_i}\big) = \mathbb{1}_{\{U=1\}}\mathbb{E}\big(\mathbb{1}_A|\widetilde{\mathcal{G}_i}\big) + \mathbb{1}_{\{U=0\}}\mathbb{E}\big(\mathbb{1}_{A^c}|\widetilde{\mathcal{G}_i}\big) = UX_i + (1-U)(1-X_i)$$

and

$$\mathbb{P}(\widetilde{A}) = \mathbb{E}\Big[\mathbb{E}\big(\mathbb{1}_{\widetilde{A}}|\widetilde{\mathcal{G}_i}\big)\Big] = \mathbb{E}\big[UX_i + (1-U)(1-X_i)\big] = \frac{1}{2},$$

for all i. It remains to note that since $U \in \{0,1\}$, the set of all values attained by $\widetilde{X}_i$ has at most $2m$ elements; this gives (i).

To show (ii), observe that $\mathbb{P}\Big(\bigcap_{i=1}^n\{\widetilde{X}_i = x_i\}\cap\widetilde{A}\Big) = \ldots$

$$\begin{aligned}
\ldots &= \mathbb{P}\Big(\bigcap_{i=1}^n\{\widetilde{X}_i = x_i\}\cap\widetilde{A}\cap\{U=0\}\Big) + \mathbb{P}\Big(\bigcap_{i=1}^n\{\widetilde{X}_i = x_i\}\cap\widetilde{A}\cap\{U=1\}\Big)\\
&= \mathbb{P}\Big(\bigcap_{i=1}^n\{X_i = 1-x_i\}\cap A^c\cap\{U=0\}\Big) + \mathbb{P}\Big(\bigcap_{i=1}^n\{X_i = x_i\}\cap A\cap\{U=1\}\Big).
\end{aligned}$$

Since U is independent of X_i's and A, and satisfies $\mathbb{P}(U=0)=\mathbb{P}(U=1)=1/2$, the above expression is equal to

$$\begin{aligned}
&\mathbb{P}\Big(\bigcap_{i=1}^n\{X_i = 1-x_i\}\cap A^c\cap\{U=1\}\Big) + \mathbb{P}\Big(\bigcap_{i=1}^n\{X_i = x_i\}\cap A\cap\{U=0\}\Big)\\
&= \mathbb{P}\Big(\bigcap_{i=1}^n\{\widetilde{X}_i = 1-x_i\}\cap\widetilde{A}^c\cap\{U=1\}\Big) + \mathbb{P}\Big(\bigcap_{i=1}^n\{\widetilde{X}_i = 1-x_i\}\cap\widetilde{A}^c\cap\{U=0\}\Big)\\
&= \mathbb{P}\Big(\bigcap_{i=1}^n\{\widetilde{X}_i = 1-x_i\}\cap\widetilde{A}^c\Big),
\end{aligned}$$

so (ii) is established.

To prove the third part, fix $x \in (0,1]$ and write

$$\frac{1-x}{x} \cdot \sum_{i=1}^{n} \mathbb{P}\Big(\{\widetilde{X}_i = x\} \cap \widetilde{A}\Big) = \sum_{i=1}^{n} \mathbb{P}\Big(\{\widetilde{X}_i = x\} \cap \widetilde{A}^c\Big) = \sum_{i=1}^{n} \mathbb{P}\Big(\{\widetilde{X}_i = 1-x\} \cap \widetilde{A}\Big),$$

where the first equality is due to the Lemma 5.1 and the second is a result of (ii). □

In closing, we include an extra lemma – it will be needed in Chapter 6.

LEMMA 5.3. *Using the same setup, for any $\delta \in [0,1]$, we have the equality*

$$\mathbb{P}\Big(\max_{1\le i<j\le n} |X_i - X_j| \ge \delta\Big) = 2 \cdot \mathbb{P}\Big(\Big\{\max_{1\le i<j\le n} |\widetilde{X}_i - \widetilde{X}_j| \ge \delta\Big\} \cap \widetilde{A}\Big).$$

PROOF. Fix $m \in \{1,2,\dots\}$ and an arbitrary vector $X \in \mathcal{C}(n,m)$. Notice that

$$\max_{1\le i<j\le n} |X_i - X_j| = \max_{1\le i<j\le n} |(1-X_i) - (1-X_j)| \tag{5.4}$$

almost surely. Hence, by (5.4) and part (ii) of Lemma 5.2, we deduce

$$\mathbb{P}\Big(\max_{1\le i<j\le n} |X_i - X_j| \ge \delta\Big) = \mathbb{P}\Big(\max_{1\le i<j\le n} |\widetilde{X}_i - \widetilde{X}_j| \ge \delta\Big) = 2 \cdot \mathbb{P}\Big(\Big\{\max_{1\le i<j\le n} |\widetilde{X}_i - \widetilde{X}_j| \ge \delta\Big\} \cap \widetilde{A}\Big),$$

as desired. □

5.2. Preparatory reductions

Our starting point is the following discretization, which enables us to restrict our argument to random variables taking values in a finite set. For convenience, we will consequently use the shorter notation and write X instead of $(X_1, X_2, \dots, X_n)$.

PROPOSITION 5.1. *We have the identity*

$$\sup_{X\in\mathcal{C}_n} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j| = \sup_{\substack{m\in\{1,2,3,\dots\},\\ X\in\mathcal{C}(n,m)}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j|.$$

PROOF. By Proposition 1.2, for every $X \in \mathcal{C}_n$ and ever $m \in \{1,2,3,\dots\}$, there exists $\widetilde{X} \in \mathcal{C}(n,m)$ such that $\max\{|X_i - \widetilde{X}_i| : 1 \le i \le n\} \le 1/m$ almost surely. Now, the statement follows from the direct application of the triangle inequality. □

We now have to establish the following reduction.

PROPOSITION 5.2. *We have the identity*

$$\sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}(n,m)}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j| = \sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}'(n,m)}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j|, \tag{5.5}$$

where $\mathcal{C}'(n,m)$ is the subset of all those $X \in \mathcal{C}(n,m)$ which satisfy

$$\mathbb{P}\left(\Big\{\max_{1\le i\le n} X_i = 1\Big\} \cup \Big\{\min_{1\le j\le n} X_j = 0\Big\}\right) = 1. \tag{5.6}$$

Proof. Fix $X \in \mathcal{C}(n,m)$. Then there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an event $A \in \mathcal{F}$ and σ-algebras $\mathcal{G}_1$, $\mathcal{G}_2$, …, $\mathcal{G}_n$ such that

$$X_i = \mathbb{E}(\mathbb{1}_A|\mathcal{G}_i), \qquad i = 1, 2, \ldots, n.$$

With no loss of generality, we may assume that the probability space is non-atomic. Now we will perform a sequence of transformations of the variables X_i (or rather of the corresponding σ-algebras $\mathcal{G}_i$), after which

· the maximum $\max_{1\le i\le n} X_i$ will increase to 1 on A;

· the minimum $\min_{1\le i\le n} X_i$ will decrease to 0 on A^c;

· the expectation $\mathbb{E}\max_{1\le i<j\le n} |X_i - X_j|$ will increase or stay unchanged.

This will yield the claim. For a given $i \in \{1, 2, \ldots, n\}$, the transformation of X_i can be described as follows. Split the set $\{\max_{1\le j\le n} X_j = X_i\} \cap A$ into the events

$$A_{i,x} = \Big\{\max_{1\le j\le n} X_j = X_i\Big\} \cap A \cap \big\{X_i = x\big\}, \qquad x \in [0,1].$$

Fix $x \in [0,1)$ such that $\mathbb{P}(A_{i,x}) > 0$; there is a finite number of such parameters, since X_i takes values in a finite set. Furthermore, we have $x > 0$, since $\max_{1\le j\le n} X_j > 0$ almost surely on A. By Lemma 5.1, there exists an event $\widehat{A}_{i,x} \subseteq \{X_i = x\} \cap A^c$ satisfying $\mathbb{P}(\widehat{A}_{i,x}) = \mathbb{P}(A_{i,x}) \cdot \frac{1-x}{x}$. We introduce the modification $\widehat{X}_i$ of X_i, given by

$$\widehat{X}_i(\omega) = \begin{cases} 1, & \text{for } \ \omega \in A_{i,x} \\ 0, & \text{for } \ \omega \in \widehat{A}_{i,x} \\ X_i(\omega), & \text{otherwise,} \end{cases}$$

or equivalently, $\widehat{X}_i = \mathbb{E}(\mathbb{1}_A|\sigma(\mathcal{G}_i, A_{i,x}, \widehat{A}_{i,x}))$. Setting $\widehat{X}_j \equiv X_j$ for $j \neq i$, we see that

$$\begin{aligned} &\mathbb{E}\max_{1\le k<l\le n} |\widehat{X}_k - \widehat{X}_l| - \mathbb{E}\max_{1\le k<l\le n} |X_k - X_l| \\ &\ge \mathbb{P}(A_{i,x}) \cdot (1-x) - \mathbb{P}(\widehat{A}_{i,x}) \cdot x \\ &= \mathbb{P}(A_{i,x}) \cdot \left[(1-x) - \frac{1-x}{x} \cdot x\right] = 0, \end{aligned}$$

and hence the alteration $X \mapsto \widehat{X}$ does not decrease the maximized expectation. Furthermore, note that quantity $\max_{1\le j\le n} \widehat{X}_j$ has increased to 1 on $A_{i,x}$. The desired transformation of X_i is obtained by applying the above modification for all $x \in (0,1)$ with $\mathbb{P}(A_{i,x}) > 0$.

Now, performing the above transformations of $X_1, X_2, \ldots, X_n$, we obtain a new vector X for which

$$\mathbb{P}\left(\Big\{\max_{1\le i\le n} X_i = 1\Big\} \cap A\right) = \mathbb{P}(A).$$

Furthermore, after applying the above procedure to the coherent sequence $(1-X_1, 1-X_2, \dots, 1-X_n)$ (corresponding to the event A^c), we may also guarantee that

$$\mathbb{P}\left(\left\{\min_{1\le i\le n} X_i = 0\right\} \cap A^c\right) = \mathbb{P}(A^c),$$

which completes the proof. □

Remark 5.1. It follows easily from the above argument that if $X \in \mathcal{C}'(n,m)$, then the corresponding event A satisfies

$$A = \{\max_{1\le i\le n} X_i = 1\} \quad \text{and} \quad A^c = \{\min_{1\le i\le n} X_i = 0\},$$

up to sets of probability zero.

Our next step is to simplify the maxima in (5.5).

Proposition 5.3. *For any $n \ge 2$, we have the equality*

$$\sup_{X\in\mathcal{C}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j| = \sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j|,$$

where $\mathcal{C}''(n,m)$ is the subset of all those $X \in \mathcal{C}'(n,m)$ that satisfy (5.2) and (5.3).

Proof. By Proposition 5.2, it is enough to apply the symmetrization procedure $X \mapsto \widetilde{X}$ (as described in the previous section), for every $X \in \mathcal{C}'(n,m)$, $m = 1,2,\dots$. Note that (5.6) is clearly preserved by this modification. Points (5.2) and (5.3) are now a direct consequence of parts (ii) and (iii) of Lemma 5.2. □

Remark 5.2. By (5.2), all the coordinates of an arbitrary vector $X \in C''(n,m)$ have expectation 1/2. This in particular implies that the event A which "generates" X satisfies $\mathbb{P}(A) = 1/2$ – confront part (i) of Lemma 5.2. Actually, (5.2) yields the stronger symmetry property of X around 1/2: we have the equality of distributions

$$\left(X_i - \frac{1}{2}\right)_{i=1}^n \overset{\mathcal{D}}{=} \left(\frac{1}{2} - X_i\right)_{i=1}^n. \tag{5.7}$$

The above remark enables the following further reduction.

Proposition 5.4. *For any $n \ge 2$, we have*

$$\sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j| + 1 = 2 \cdot \sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E} \max_{1\le i\le n} X_i.$$

Proof. We may write

$$\max_{1\le i<j\le n} |X_i - X_j| = \max_{1\le i,j\le n} (X_i - X_j) = \max_{1\le i\le n} \left(X_i - \frac{1}{2}\right) + \max_{1\le j\le n} \left\{-\left(X_j - \frac{1}{2}\right)\right\},$$

and hence by (5.7) we get

$$\sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E} \max_{1\le i<j\le n} |X_i - X_j| = 2 \cdot \sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E} \max_{1\le i\le n} \left(X_i - \frac{1}{2}\right),$$

which completes the proof. □

By the above reductions, we see that it is enough to handle the expression

$$\sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E}\max_{1\le i\le n} X_i = \sup_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \left(\mathbb{E}\left(\mathbb{1}_A \max_{1\le i\le n} X_i\right) + \mathbb{E}\left(\mathbb{1}_{A^c}\max_{1\le i\le n} X_i\right)\right). \tag{5.8}$$

Observe that for a fixed m and an $X\in\mathcal{C}''(n,m)$, we have $\mathbb{1}_A\cdot\max_{1\le i\le n} X_i = \mathbb{1}_A$ almost surely, due to Remark 5.1. In addition, we have the identity

$$\mathbb{P}\left(\left\{\max_{1\le j\le n} X_j = X_i\right\}\cap\left\{X_i = 1-x\right\}\cap A^c\right) = \mathbb{P}\left(\left\{\min_{1\le j\le n} X_j = X_i\right\}\cap\left\{X_i = x\right\}\cap A\right),$$

for $1\le i\le n$ and $x\in[0,1]$, which follows from the condition (5.2). In other words, the distributions of $\mathbb{1}_{A^c}\cdot\max_{1\le i\le n} X_i$ and $\mathbb{1}_A\,(1-\min_{1\le i\le n} X_i)$ coincide, and we get the following.

PROPOSITION 5.5. *For any $n\ge 2$, a sequence X is a maximizer of (5.8) if and only if it minimizes the quantity*

$$\inf_{\substack{m\in\{1,2,\dots\},\\ X\in\mathcal{C}''(n,m)}} \mathbb{E}\left(\mathbb{1}_A\cdot\min_{1\le i\le n} X_i\right). \tag{5.9}$$

Thanks to Proposition 5.5, we can limit the remaining analysis to the study of coherent vector X on the smaller set A.

5.3. Combinatorial optimization

Now we are going to present a combinatorial analysis of (5.9), which will be done by some geometrical considerations in a slightly different setup. Let us start with introducing some auxiliary notation. Throughout, we assume that $n\ge 2$ is a fixed integer. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a non-atomic probability space and let A be a fixed event satisfying $\mathbb{P}(A)=1/2$. For an event $S\subseteq A$ and any $x\in(0,1]$, we set

$$S^x := S\times\{x\}\subseteq A\times(0,1].$$

In our considerations below, it will be convenient to interpret S^x as horizontal line segments at level x, or finite unions of sets of this type. We assign to each S^x the corresponding function $\mathcal{S}^x : A\to[0,1]$, given by

$$\mathcal{S}^x(\omega) = \begin{cases} x, & \text{for } \omega\in S\\ 1, & \text{for } \omega\in A\setminus S.\end{cases}$$

For $k\in\mathbb{N}\setminus\{0\}$, we denote by $\Lambda(n,k)$ the family of all sequences $(S_1^{x_1}, S_2^{x_2},\dots,S_k^{x_k})$ which satisfy the conditions

$$\sum_{i=1}^k \mathbb{P}(S_i) = \frac{n-1}{2}, \tag{5.10}$$

and

$$\sum_{\{1\le i\le k:\ x_i=x\}} \mathbb{P}(S_i) = \frac{x}{1-x}\cdot\sum_{\{1\le j\le k:\ x_j=1-x\}} \mathbb{P}(S_j), \quad \forall x\in(0,1). \tag{5.11}$$

In this setup, we consider the following, new optimization problem:

$$\inf_{\substack{(\Omega,\mathcal{F},\mathbb{P}),\\ A\in\mathcal{F},\ \mathbb{P}(A)=\frac{1}{2}}} \inf_{\substack{k\in\{1,2,\dots\},\\ (S_1^{x_1},S_2^{x_2},\dots,S_k^{x_k})\in\Lambda(n,k)}} \int_A \min(\mathcal{S}_1^{x_1},\mathcal{S}_2^{x_2},\dots,\mathcal{S}_k^{x_k})\mathrm{d}\mathbb{P}.$$

By a straightforward measure-preserving transformation argument, we see that the infimum inside does not depend on the choice of the particular probability space or the underlying event A. In other words, the above quantity is equal to

$$\inf_{\substack{k\in\{1,2,\dots\},\\ (S_1^{x_1},S_2^{x_2},\dots,S_k^{x_k})\in\Lambda(n,k)}} \int_A \min(\mathcal{S}_1^{x_1},\mathcal{S}_2^{x_2},\dots,\mathcal{S}_k^{x_k})\mathrm{d}\mathbb{P}. \tag{5.12}$$

PROPOSITION 5.6. *The value of (5.9) is not smaller than the value of (5.12).*

PROOF. Fix $m\in\{1,2,\dots\}$, $X\in\mathcal{C}''(n,m)$ and the corresponding "generating" event A. As we have seen in Remark 5.2, we have $\mathbb{P}(A)=1/2$. For a given $1\leq i\leq n$, let $\{x_1^i,x_2^i,\dots,x_m^i\}$ denote the set of values attained by X_i, $1\leq i\leq n$ (if X_i takes less than m different values, we add some extra, superfluous elements to the set). Introduce the events $T_{i,j}=\{\omega\in A:\ X_i(\omega)=x_j^i\}$, for $1\leq i\leq n$ and $1\leq j\leq m$. Since $\mathbb{P}(A)=1/2$, we have the straightforward equality $\sum_{\substack{1\leq i\leq n,\\ 1\leq j\leq m}}\mathbb{P}(T_{i,j})=n\cdot 1/2$.

Next, by the property (5.3) of class $\mathcal{C}''(n,m)$, for a fixed $x\in(0,1)$ we may write

$$\sum_{i=1}^n \sum_{\{1\leq j\leq m:\ x_j^i=x\}} \mathbb{P}(T_{i,j}) = \frac{x}{1-x}\cdot\sum_{i=1}^n \sum_{\{1\leq j\leq m:\ x_j^i=1-x\}} \mathbb{P}(T_{i,j}).$$

This yields the condition (5.11) for the family $(T_{i,j}^{x_j^i})_{i,j}$. Next, let $(U_i)_{i=1}^n$ be a partition of A such that $U_i\in\mathcal{F}$ and $U_i\subseteq\{X_i=1\}$ for all $1\leq i\leq n$, up to a set of measure zero. The existence of such a partition is an obvious consequence of Remark 5.1. Define a modification $(S_{i,j}^{x_j^i})_{i,j}$ of $(T_{i,j}^{x_j^i})_{i,j}$ by

$$S_{i,j}=\begin{cases} T_{i,j}\setminus U_i, & \text{if } x_j^i=1\\ T_{i,j}, & \text{if } x_j^i\neq 1.\end{cases}$$

Since $\mathbb{P}(A)=1/2$, we get $\sum_{\substack{1\leq i\leq n,\\ 1\leq j\leq m}}\mathbb{P}(S_{i,j})=(n-1)\cdot\frac{1}{2}$, and the condition (5.11) is still satisfied for $(S_{i,j}^{x_j^i})_{i,j}$ (the sets $T_{i,j}$ are modified only if $x_j^i=1$, and these values are not considered in (5.11). Consequently, we have $(S_{i,j}^{x_j^i})_{i,j}\in\bigcup_{k\geq 1}\Lambda(n,k)$ and so

$$\mathbb{E}\left(\mathbb{1}_A\cdot\min_{1\leq i\leq n}X_i\right)=\int_A \min_{\substack{1\leq i\leq n,\\ 1\leq j\leq m}}\mathcal{S}_{i,j}^{x_j^i}\mathrm{d}\mathbb{P}$$

is not smaller than (5.12). Since m and X were arbitrary, the proof is complete. □

Our plan is to solve the problem (5.12), by performing a sequence of combinatorial and geometrical reductions. We start with some simple observations. First, note that by (5.11), a sequence $(S_1^{x_1},S_2^{x_2},\dots,S_k^{x_k})\in\Lambda(n,k)$ enjoys a sort of skew-symmetry around $1/2$ – in particular, if a level x belongs to $\{x_1,x_2,\dots,x_k\}$, then so does $1-x$. Second, obviously, the integral in (5.12) does not depend on the order of the sets $S_1^{x_1}$, $S_2^{x_2}$, ..., $S_k^{x_k}$, so we may permute them arbitrarily. Furthermore, note that if we split $S_k^{x_k}$ into two sets $\widetilde{S}_k^{x_k}$ and $\widetilde{S}_{k+1}^{x_k}$ (of course, the level x_k needs to be preserved) and replace

$$(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k}) \quad \text{with} \quad (S_1^{x_1}, S_2^{x_2}, \dots, S_{k-1}^{x_{k-1}}, \widetilde{S}_k^{x_k}, \widetilde{S}_{k+1}^{x_k}),$$

then the integral will not change either. Similar phenomenon occurs if we splice two disjoint sets lying at the same level. In other words, given a vector $(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k})$, we may cut some of the sets into pieces or merge some of them, with no effect on the expression optimized in (5.12).

The next step is the following.

PROPOSITION 5.7. *The problem (5.12) can be rewritten as*

$$\inf_{\substack{k\in\{1,2,\dots\},\\ (S_1^{x_1},S_2^{x_2},\dots,S_k^{x_k})\in\Lambda^\star(n,k)}} \int_A \min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \dots, \mathcal{S}_k^{x_k})\, d\mathbb{P}, \tag{5.13}$$

where $\Lambda^\star(n,k)$ *is the subset of all those* $(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k}) \in \Lambda(n,k)$*, which satisfy*

$$\mathbb{P}(S_i \cap S_j) = 0 \quad \textit{whenever} \quad x_i, x_j \le 1/2. \tag{5.14}$$

PROOF. Fix $k \in \{1, 2, \dots\}$, a vector $(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k}) \in \Lambda(n,k)$ and assume that the condition (5.14) is not satisfied for some $0 < x_i \le x_j \le \frac{1}{2}$. Of course, the claim will follow if we find a vector $(\widetilde{S}_1^{y_1}, \widetilde{S}_2^{y_2}, \dots, \widetilde{S}_l^{y_l}) \in \Lambda^\star(n,l)$, $l \in \mathbb{Z}_+$, such that

$$\int_A \min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \dots, \mathcal{S}_k^{x_k}) \mathrm{d}\mathbb{P} \ge \int_A \min(\widetilde{\mathcal{S}}_1^{y_1}, \widetilde{\mathcal{S}}_2^{y_2}, \dots, \widetilde{\mathcal{S}}_l^{y_l}) \mathrm{d}\mathbb{P}. \tag{5.15}$$

There are two possible scenarios: we either have

$$\mathbb{P}\left(\left\{\min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \dots, \mathcal{S}_k^{x_k}) > \frac{1}{2}\right\} \cap A\right) > 0, \tag{5.16}$$

or

$$\mathbb{P}\left(\left\{\min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \dots, \mathcal{S}_k^{x_k}) > \frac{1}{2}\right\} \cap A\right) = 0. \tag{5.17}$$

In the case of (5.16), we can simply cut off a part of S_j which is already covered by a smaller value x_i and transfer as much of it as possible to the set

$$\left\{\min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \dots, \mathcal{S}_k^{x_k}) > \frac{1}{2}\right\}.$$

See Figure 2. Observe that this modification satisfies condition (5.15). After a finite number of such transformations, we obtain a sequence $(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k})$ for which either (5.14) holds, or we have (5.17).

Now, suppose that (5.14) does not hold for some $0 < x_i \le x_j \le \frac{1}{2}$, but we also have the equality (5.17). First, set

$$c = \mathbb{P}(S_i \cap S_j), \quad a = \frac{1 - x_i}{x_i} \cdot c, \quad b = \frac{1 - x_j}{x_j} \cdot c.$$

Let us pick all those $S_\ell^{x_\ell}$ from the sequence $(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k})$, for which $x_\ell = 1 - x_i$ or $x_\ell = 1 - x_j$, and arrange them into two lists

$$(S_{i,1}^{1-x_i}, \dots, S_{i,n_i}^{1-x_i}) \quad \text{and} \quad (S_{j,1}^{1-x_j}, \dots, S_{j,n_j}^{1-x_j}).$$

Next, we choose two sequences $(T_{i,m})_{m=1}^{n_i}, (T_{j,m})_{m=1}^{n_j} \subseteq \mathcal{F}$ such that:

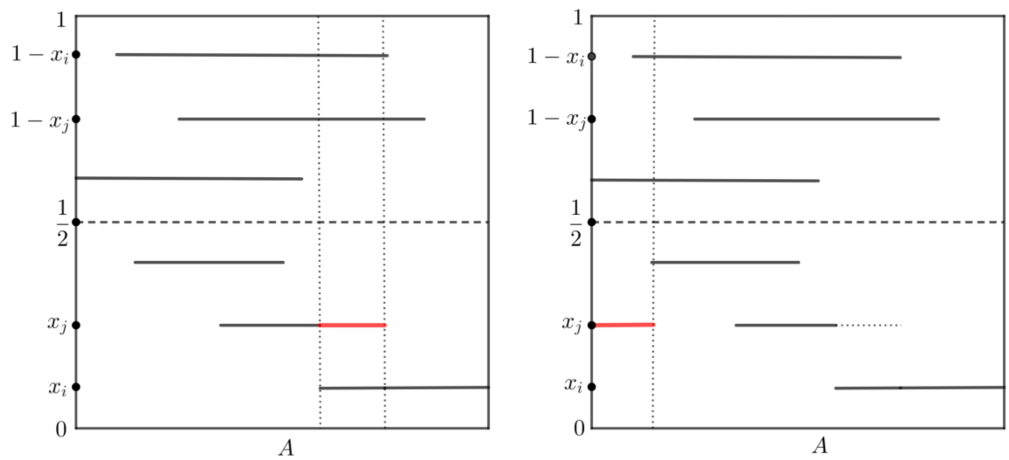

FIGURE 2. The horizontal line segments are (fragments of) graphs of functions $\mathcal{S}_1^{x_1}$, $\mathcal{S}_2^{x_2}$, …, $\mathcal{S}_n^{x_n}$. In the figure above, the entire problematic intersection (in red) has been moved to the area in which all the other line segments lie above $1/2$. However, if the set $\{\min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \ldots, \mathcal{S}_k^{x_k}) > \frac{1}{2}\}$ is too small, it might be impossible to remove the whole troublesome intersection – this leads to case (5.17).

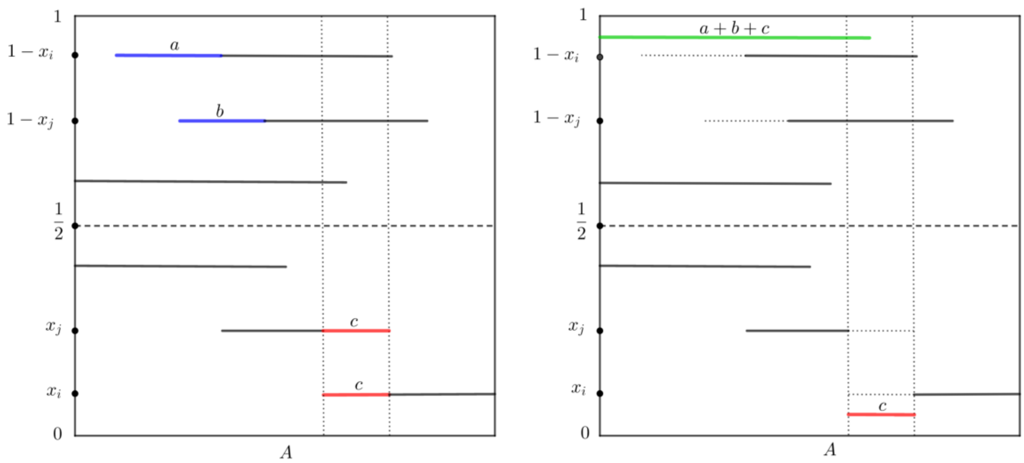

FIGURE 3. The transformation in the case of (5.17) – the integral $\int_A \min(\mathcal{S}_i^{x_i})\mathrm{d}\mathbb{P}$ can only decrease. Indeed, the new line segment of length c on the right picture lies below x_i. Though the line segment of length $a+b+c$ might lie on a higher level, the condition (5.17) guarantees that there must be a "layer" of line segments lying beneath.

· $T_{i,m} \subseteq S_{i,m}$ for $1 \le m \le n_i$; $T_{j,m} \subseteq S_{j,m}$ for $1 \le m \le n_j$,

· $\sum_{m=1}^{n_i} \mathbb{P}(T_{i,m}) = a$ and $\sum_{m=1}^{n_j} \mathbb{P}(T_{j,m}) = b$.

By (5.11), such $(T_{i,m})_m, (T_{j,m})_m$ exist (since the probability space is nonatomic). Assume further that x_0 is a number satisfying

$$\frac{1-x_0}{x_0}\cdot c = a+b+c,$$

and observe that $x_0 < x_i$. Indeed, we have $a = (1-x_i)c/x_i < a+b+c$.

We shall now perform the following transformation:

(1) remove $(T_{i,m}^{1-x_i})_{m=1}^{n_i}$, $(T_{j,m}^{1-x_j})_{m=1}^{n_j}$, $(S_i\cap S_j)^{x_i}$ and $(S_i\cap S_j)^{x_j}$,
(2) add $(T_{i,m}^{1-x_0})_{m=1}^{n_i}$, $(T_{j,m}^{1-x_0})_{m=1}^{n_j}$, $(S_i\cap S_j)^{1-x_0}$ and $(S_i\cap S_j)^{x_0}$,

– see Figure 3. It is straightforward to check that the new, modified sequence $(S_\ell^{x_\ell})$ satisfies points (5.10), (5.11) and (5.15). After a finite number of such changes, we guarantee the validity of (5.14). □

Hence, from now on, we may restrict our analysis of (5.12) to the class $\bigcup_{k\geq 1}\Lambda^\star(n,k)$. To proceed, consider a vector $(S_1^{x_1}, S_2^{x_2}, \ldots, S_k^{x_k}) \in \Lambda^\star(n,k)$. We shall introduce a partition of A into the following four basic components. Namely, we first put

$$A_1 = \left\{\omega\in A:\ \min_{1\leq i\leq k}\mathcal{S}_i^{x_i} = \frac{1}{2}\right\}\quad\text{and}\quad A_2 = \left\{\omega\in A:\ \min_{1\leq i\leq k}\mathcal{S}_i^{x_i} > \frac{1}{2}\right\}.$$

Let $y_1, y_2, \ldots, y_m$ be the collection of all distinct values taken by $\min_{1\leq i\leq k}\mathcal{S}_i^{x_i}$ on the set A_2. For $1\leq j\leq m$, we set $M_{y_j} = A_2\cap\{\min_{1\leq i\leq k}\mathcal{S}_i^{x_i} = y_j\}$. By (5.11), we can find events $(N_{1-y_j})_{j=1}^m$ satisfying

$$N_{1-y_j}\subseteq A\cap\left\{\min_{1\leq i\leq k}\mathcal{S}_i^{x_i} = 1-y_j\right\}\quad\text{and}\quad \mathbb{P}(N_{1-y_j}) = \frac{1-y_j}{y_j}\cdot\mathbb{P}(M_{y_j}).$$

By condition (5.14), elements of $(N_{1-y_j})_{j=1}^m$ are disjoint and well-defined. Let us set

$$A_3 = \bigcup_{1\leq j\leq m} N_{1-y_j}\quad\text{and}\quad A_4 = \left\{\omega\in A:\ \min_{1\leq i\leq k}\mathcal{S}_i^{x_i} < \frac{1}{2}\right\}\setminus A_3.$$

– See Figure 4.a. We can think about the set A_3 as "generated" or "induced" by A_2.

Proposition 5.8. *The problem (5.13) can be rewritten as*

$$\inf_{\substack{k\in\{1,2,\ldots\},\\ (S_1^{x_1},S_2^{x_2},\ldots,S_k^{x_k})\in\Lambda^{\star\star}(n,k)}} \int_A \min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \ldots, \mathcal{S}_k^{x_k})\,d\mathbb{P}, \tag{5.18}$$

where $\Lambda^{\star\star}(n,k)$ is the subset of all those $(S_1^{x_1}, S_2^{x_2}, \ldots, S_k^{x_k}) \in \Lambda^\star(n,k)$, which satisfy

$$\mathbb{P}\left(A_2\cap\bigcup_{1\leq i<j\leq k}(S_i\cap S_j)\right) = 0. \tag{5.19}$$

Proof. Fix $k\in\{1,2,\ldots\}$, a vector $(S_1^{x_1}, S_2^{x_2}, \ldots, S_k^{x_k}) \in \Lambda^\star(n,k)$ and assume that the condition (5.19) is not satisfied. This in particular implies that $\mathbb{P}(A_2) > 0$ and $\mathbb{P}(A_3\cup A_4) > 0$ (see the skew-symmetry of $(S_1^{x_1}, S_2^{x_2}, \ldots, S_k^{x_k})$, mentioned above

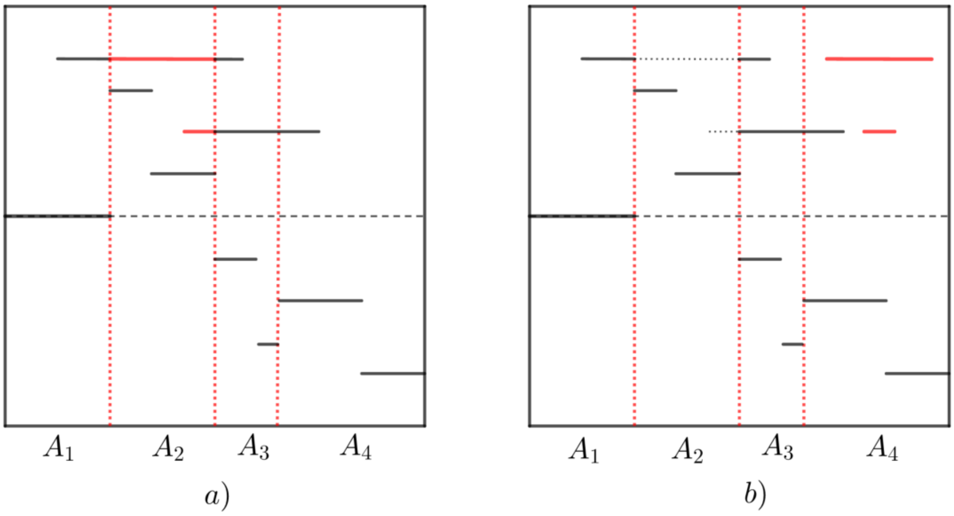

FIGURE 4. Basic components of A: in a) before, in b) after the rearrangement.

Proposition 5.7). Now, let us fix an index $i \in \{1, 2, \dots, k\}$ with $x_i > 1/2$. Then, for any event $T \subseteq S_i \cap A_2$, such that

$$T \subseteq \left\{ \min_{1 \le j \le k} \mathcal{S}_j^{x_j} < x_i \right\} \quad \text{or} \quad T \subseteq \left(S_i \cap \bigcup_{\substack{1 \le j \le k, \\ j \ne i,\ x_j = x_i}} S_j \right),$$

we can perform the following rearrangement:

(1) remove T^{x_i},
(2) add new events $(T_m^{x_i})_{m=1}^N$, such that

$$T_m \subseteq A_3 \cup A_4 \quad \text{for} \quad 1 \le m \le N,$$

and $\sum_{m=1}^N \mathbb{P}(T_m) = \mathbb{P}(T)$,

– see Figure 4b. The existence of $(T_m^{x_i})_{m=1}^N$ follows trivially from $\mathbb{P}(A_3 \cup A_4) > 0$, as we allow the overlapping of the sets. The obtained modified sequence belongs to $\Lambda^\star(n, \ell)$ for some ℓ and the minimum of the corresponding functions is unchanged almost surely on set A (in comparison to the initial minimum $\min \mathcal{S}_\ell^{x_\ell}$). It remains to observe, that we may guarantee the validity of (5.19), just by performing sufficiently many such transformations. □

PROPOSITION 5.9. *In the problem (5.18), we can further restrict ourselves to exactly all those sequences* $(S_1^{x_1}, \dots, S_k^{x_k}) \in \Lambda^{\star\star}(n, k)$, *which additionally satisfy*

$$\mathbb{P}(A_2 \cup A_3) = 0, \tag{5.20}$$

and

$$|\{1 \le i \le k : \ x_i > 1/2\}| \le 1. \tag{5.21}$$

PROOF. The argument follows the above pattern. Namely, we fix $k \in \{1, 2, \dots\}$, a sequence $(S_1^{x_1}, \dots, S_k^{x_k}) \in \Lambda^{\star\star}(n, k)$ and assume that the condition (5.20) is not satisfied. Let the number $x > 1/2$ be such that $\mathbb{P}(M_x) > 0$, where (as defined before) $M_x = A_2 \cap \{\min_{1 \le i \le k} \mathcal{S}_i^{x_i} = x\}$. Moreover, recall that N_{1-x} is an event satisfying

$$N_{1-x} \subseteq A_3 \cap \left\{ \min_{1 \le i \le k} \mathcal{S}_i^{x_i} = 1 - x \right\} \quad \text{and} \quad \mathbb{P}(N_{1-x}) = \frac{1-x}{x} \cdot \mathbb{P}(M_x).$$

Note that we have

$$\begin{aligned} &\int_{M_x \cup N_{1-x}} \min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \dots, \mathcal{S}_k^{x_k}) \mathrm{d}\mathbb{P} = x \cdot \mathbb{P}(M_x) + (1-x) \cdot \mathbb{P}(N_{1-x}) \\ &= \mathbb{P}(M_x \cup N_{1-x}) \cdot \left(x^2 + (1-x)^2 \right) > \frac{\mathbb{P}(M_x \cup N_{1-x})}{2}. \end{aligned} \tag{5.22}$$

Consider the following operation:

(1) remove M_x^x and N_{1-x}^{1-x},

(2) add $M_x^{\frac{1}{2}}$ and $N_{1-x}^{\frac{1}{2}}$,

– see Figure 5.a. By (5.22), such a transformation yields a sequence from the class $\bigcup_{\ell \in \{1,2,\dots\}} \Lambda^{\star\star}(n, \ell)$, for which the integral in (5.18) is decreased. After finitely many such operations, we will remove A_2 and thus enforce the validity of (5.20).

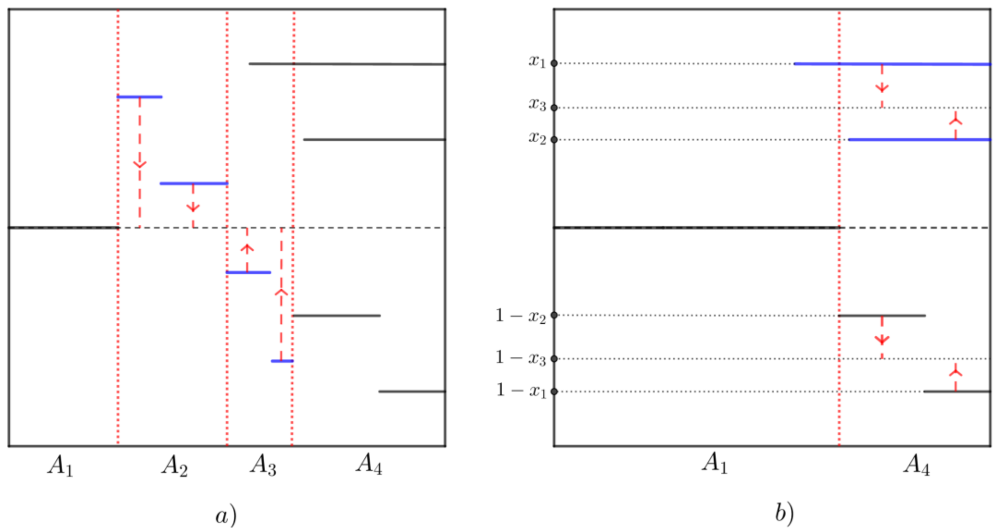

FIGURE 5. a: removing $A_2 \cup A_3$. b: leveling $x_1, x_2 > \frac{1}{2}, x_1 \neq x_2$.

To guarantee the second condition, we proceed as previously. We start with an arbitrary sequence $(S_1^{x_1}, S_2^{x_2}, \dots, S_k^{x_k}) \in \Lambda^{\star\star}(n, k)$, for which (5.20) is satisfied, but (5.21) is not. We may assume, performing a permutation of the indices if necessary, that $x_1, x_2 > 1/2$, $x_1 \neq x_2$ and $s_1 := \mathbb{P}(S_1) > 0$, $s_2 := \mathbb{P}(S_2) > 0$. By (5.11), we can find two events $T_{1-x_1}, T_{1-x_2} \in \mathcal{F}$, such that

$$T_{1-x_1} \subseteq \left\{ \min_{1 \le j \le k} \mathcal{S}_j^{x_j} = 1 - x_1 \right\}, \quad T_{1-x_2} \subseteq \left\{ \min_{1 \le j \le k} \mathcal{S}_j^{x_j} = 1 - x_2 \right\},$$

and

$$t_1 := \mathbb{P}(T_{1-x_1}) = \frac{1-x_1}{x_1} \cdot s_1, \qquad t_2 := \mathbb{P}(T_{1-x_2}) = \frac{1-x_2}{x_2} \cdot s_2.$$

Consider the auxiliary equation

$$t_1 \cdot \frac{x_1}{1-x_1} + t_2 \cdot \frac{x_2}{1-x_2} = (t_1+t_2) \cdot \frac{x_0}{1-x_0}, \tag{5.23}$$

or, after transforming equivalently,

$$1 - x_0 = \frac{t_1+t_2}{\frac{t_1}{1-x_1} + \frac{t_2}{1-x_2}}.$$

If $\max\{x_1, x_2\} = 1$, we understand this equation as $x_0 = \min\{x_1, x_2\}$. We will show

$$(t_1+t_2) \cdot (1-x_0) \le t_1 \cdot (1-x_1) + t_2 \cdot (1-x_2). \tag{5.24}$$

This holds true for $\max\{x_1, x_2\} = 1$ (we have t_1 or t_2 equal to 0). We omit some further details of this simple case. For the general case, we substitute the previous identity and rewrite the estimate in the form

$$\frac{(t_1+t_2)^2}{\frac{t_1}{1-x_1} + \frac{t_2}{1-x_2}} \le t_1 \cdot (1-x_1) + t_2 \cdot (1-x_2),$$

or

$$2 \le \frac{1-x_1}{1-x_2} + \frac{1-x_2}{1-x_1},$$

which is evident. Having that in mind, let us consider the transformation:

(1) remove $S_1^{x_1}, S_2^{x_2}$ and $T_{1-x_1}^{1-x_1}, T_{1-x_2}^{1-x_2}$,
(2) add $S_1^{x_0}, S_2^{x_0}$, and $T_{1-x_1}^{1-x_0}, T_{1-x_2}^{1-x_0}$.

By (5.23), the obtained new vector still belongs to $\bigcup_{\ell \ge 1} \Lambda^{\star\star}(n, \ell)$ and enjoys (5.20). Furthermore, by (5.24), the appropriate minimized integral over A does not increase. It remains to note that after a finite number of the above transformations, the condition (5.21) will eventually become true. □

We are almost ready to prove our main result. Let us denote the auxiliary constant

$$s(n) := 1 - 2 \cdot \inf_{\substack{k \in \{1,2,\ldots\}, \\ (S_1^{x_1}, S_2^{x_2}, \ldots, S_k^{x_k}) \in \Lambda(n,k)}} \int_A \min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \ldots, \mathcal{S}_k^{x_k}) \mathrm{d}\mathbb{P}.$$

Lemma 5.4. *For any $n \ge 2$, the number $s(n)$ is equal to the right-hand side of* (5.1).

Proof. By the above reductions, in the definition of quantity $s(n)$ we may restrict ourselves to those $(S_1^{x_1}, S_2^{x_2}, \ldots, S_k^{x_k}) \in \Lambda^{\star\star}(n, k)$, which additionally satisfy (5.20) and (5.21). This is a very simple context: there are at most three different levels of the sets S_j. See Figure 6. Put $p := \mathbb{P}(A_1)$ and suppose that the maximal level is equal to x_1. Note that the random variable $\mathbb{1}_A \cdot \min(\mathcal{S}_1^{x_1}, \mathcal{S}_2^{x_2}, \ldots, \mathcal{S}_k^{x_k})$ is equal to $1/2$ on A_1 and to $1 - x_1$ on A_4, so we have

$$s(n) = 1 - 2 \cdot \inf_{p \in [0, \frac{1}{2}]} \left[p \cdot \frac{1}{2} + \left(\frac{1}{2} - p \right) \cdot (1 - x_1) \right].$$

To express x_1 in terms of p, we apply (5.10) and (5.11) with $x = x_1$. We obtain

$$\left[\frac{n-1}{2} - p - \left(\frac{1}{2} - p \right) \right] \cdot \frac{1-x_1}{x_1} = \frac{1}{2} - p. \tag{5.25}$$

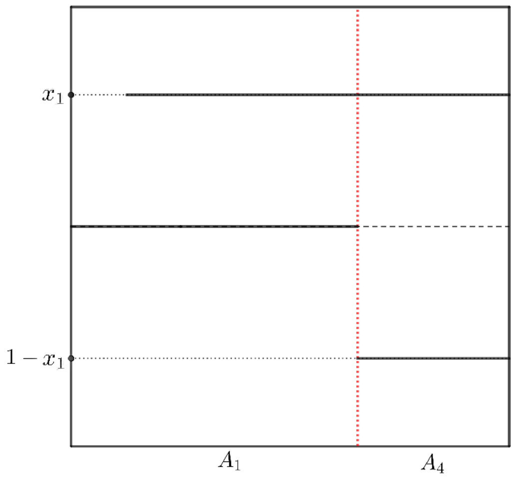

FIGURE 6. There are at most three different values in the set $\{x_1, x_2, \ldots, x_k\}$. We may assume that x_1 is the largest of them.

If $n = 2$, we get $p = 1/2$ and $s(n) = 1/2$. For $n \geq 3$, the equation (5.25) yields

$$x_1 = x_1(p) = \frac{n-2}{n-1-2p}, \tag{5.26}$$

and hence we obtain

$$s(n) = 1 - 2 \cdot \inf_{p \in [0, \frac{1}{2}]} \left[p \cdot \frac{1}{2} + \left(\frac{1}{2} - p \right) \cdot \frac{1-2p}{n-1-2p} \right]. \tag{5.27}$$

Now, we consider the cases of $n \in \{3, 4\}$ and $n \geq 5$ separately. In the first case, we make some elementary calculations to obtain

$$\operatorname*{argmin}_{p \in [0, \frac{1}{2}]} \left[p \cdot \frac{1}{2} + \left(\frac{1}{2} - p \right) \cdot \frac{1-2p}{n-1-2p} \right] = \begin{cases} 1 - \frac{1}{\sqrt{2}}, & \text{for } \ n = 3, \\ \frac{3}{2} - \sqrt{2}, & \text{for } \ n = 4. \end{cases}$$

This gives $s(3) = 2 - \sqrt{2}$ and $s(4) = \frac{7}{2} - 2\sqrt{2}$, as announced above. For $n \geq 5$, it is straightforward to check that

$$t(n, p) := \left[p \cdot \frac{1}{2} + \left(\frac{1}{2} - p \right) \cdot \frac{1-2p}{n-1-2p} \right],$$

satisfies $\partial t(n, p)/\partial p > 0$ for $p \in (0, 1/2)$. Consequently, we get

$$s(n) = 1 - 2t(n, 0) = \frac{n-2}{n-1}, \quad \text{for } n \geq 5,$$

and the proof is complete. □

We turn our attention to the main result.

PROOF OF THEOREM 5.1. We need to show that

$$\sup_{(X_1, X_2, \ldots, X_n) \in \mathcal{C}} \mathbb{E} \max_{1 \leq i < j \leq n} |X_i - X_j| = s(n).$$

By Propositions 5.4, 5.5 and 5.6, we get

$$\sup_{(X_1,\ldots,X_n)\in\mathcal{C}} \mathbb{E}\max_{1\leq i<j\leq n}|X_i-X_j|\leq s(n),$$

and it is enough to check that those bounds cannot be improved. To this end, we construct an appropriate coherent vector $(X_1, X_2, \ldots, X_n)$ on $([0,1],\mathcal{B}([0,1]),|\cdot|)$, where $\mathcal{B}([0,1])$ is the σ-field of Borel subsets of $[0,1]$ and $|\cdot|$ is the Lebesgue measure. Fix $p\in(0,1/2]$ and put $A=[0,1/2]$, $B=(1/2,1]$. Then, for $1\leq i\leq n-1$, consider the following families of intervals

$$A_1^i=\left(\frac{p(i-1)}{n-1},\frac{pi}{n-1}\right],\quad B_1^i=\left(\frac{1}{2}+\frac{p(i-1)}{n-1},\frac{1}{2}+\frac{pi}{n-1}\right],$$

$$A_2^i=\left(p+\frac{(\frac{1}{2}-p)(i-1)}{n-1},p+\frac{(\frac{1}{2}-p)i}{n-1}\right],$$

and

$$B_2^i=\left(\frac{1}{2}+p+\frac{(\frac{1}{2}-p)(i-1)}{n-1},\frac{1}{2}+p+\frac{(\frac{1}{2}-p)i}{n-1}\right].$$

One can check that all these intervals are pairwise disjoint. We also have $A_1^i, A_2^i\subseteq A$ and $B_1^i, B_2^i\subseteq B$ for all i. Therefore, if we take

$$\mathcal{G}_i=\sigma\left(A_1^i\cup B_1^i, A_2^i\cup(B\setminus(B_1^i\cup B_2^i)),(A\setminus(A_1^i\cup A_2^i))\cup B_2^i\right)$$

and $X_i=\mathbb{E}(\mathbb{1}_A|\mathcal{G}_i)$, then some simple calculations show that

$$X_i=\begin{cases}1/2 & \text{on } A_1^i\cup B_1^i,\\ 1-x_1(p) & \text{on } A_2^i\cup(B\setminus(B_1^i\cup B_2^i)),\\ x_1(p) & \text{on } (A\setminus(A_1^i\cup A_2^i))\cup B_2^i,\end{cases}\tag{5.28}$$

for all $1\leq i\leq n-1$, where $x_1(p)$ is given by (5.26). In the end, we add a variable $X_n=\mathbb{1}_A$ (it corresponds to taking $\mathcal{G}_n=\sigma(A)$). We get $(X_1,X_2,\ldots,X_n)\in\mathcal{C}_n$. For any $\omega\in[0,p]$, we have $X_n(\omega)=1$. By (5.28), there is an index i with $X_i(\omega)=1/2$. Consequently, we get

$$\mathbb{E}\max_{1\leq i<j\leq n}|X_i-X_j|\mathbb{1}_{[0,p]}\geq\frac{p}{2},$$

while similarly, we have

$$\mathbb{E}\max_{1\leq i<j\leq n}|X_i-X_j|\mathbb{1}_{(p,1/2]}\geq\left(\frac{1}{2}-p\right)\cdot x_1(p),$$

$$\mathbb{E}\max_{1\leq i<j\leq n}|X_i-X_j|\mathbb{1}_{(1/2,1/2+p]}\geq\frac{p}{2},$$

$$\mathbb{E}\max_{1\leq i<j\leq n}|X_i-X_j|\mathbb{1}_{(1/2+p,1]}\geq\left(\frac{1}{2}-p\right)\cdot x_1(p).$$

Hence, summing the above four inequalities, we get

$$\mathbb{E}\max_{1\leq i<j\leq n}|X_i-X_j|\geq 1-p-(1-2p)\cdot(1-x_1(p)).$$

Now, the supremum of the right-hand side over $p\in(0,1/2]$ is precisely $s(n)$ – see equation (5.27). This completes the proof. □

CHAPTER 6

Contradictory predictions with multiple agents

Preliminaries

The starting point of this chapter is the following foundational result of [**9**].

THEOREM 6.1. *For any threshold* $\delta \in (\frac{1}{2}, 1]$*, we have*

$$\mathbf{P}(2,\delta) = \sup_{(X,Y)\in\mathcal{C}} \mathbb{P}(|X-Y| \geq \delta) = \frac{2(1-\delta)}{2-\delta}. \tag{6.1}$$

In the language of applications, Theorem 6.1 establishes a sharp upper bound for the probability that two experts, with access to different information sources, will deliver highly incongruent or contradictory opinions. The original proof of equality (6.1) is remarkably complex and somewhat difficult: an explicit optimizer is obtained by a series of consecutive reductions and simplifications.

As pointed out in [**10**], finding a simpler proof of this result would be highly desirable. Another natural and important question concerns the extension of the threshold bound (6.1) to he case of $n > 2$ coherent opinions. Our main result in this chapter is as follows; we use the notation $a \wedge b$ for the minimum of the numbers a and b.

THEOREM 6.2. *For any threshold* $\delta \in (\frac{1}{2}, 1]$ *and every integer* $n \geq 2$*, we have*

$$\mathbf{P}(n,\delta) = \sup_{(X_1,X_2,\ldots,X_n)\in\mathcal{C}_n} \mathbb{P}\Big(\max_{1\leq i<j\leq n} |X_i - X_j| \geq \delta\Big) = \frac{n(1-\delta)}{2-\delta} \wedge 1. \tag{6.2}$$

Correspondingly, Theorem 6.2 expands the range of applications from two experts to multiple agents scenario. Quite unexpectedly (at least to the authors), the threshold bound (6.2) reveals an almost linear dependence between the examined quantities and the number of coherent random variables. The proof of the estimate (6.2) that we present below is completely independent from the reasoning in [**9**] and hence can be regarded as an alternative demonstration of (6.1). Moreover, our approach does not refer in any significant way to the particular choice of integer n.

Let us say a few words about our approach and the organization of the remaining part of the chapter. In the next section we apply our special symmetrization technique, just as introduced in Chapter 5. In this way, we will be able to reduce the problem of calculating the left-hand side of (6.2) to the combinatorial optimization over specific objects of geometrical nature. This approach seems to be especially beneficial due to certain convenient skew-symmetric constraints it enforces.

Then, in Section 6.2, using dynamic programming arguments, we solve the previously obtained optimization problem. In essence, this involves determining an appropriate Bellman function and some further combinatorial reductions of the problem. This appearance of dynamic programming is not surprising: as evidenced in numerous articles, the Bellman function method is a powerful tool used widely in martingale theory and harmonic analysis to obtain sharp inequalities – see e.g. [**61, 63, 72**] and consult the references therein.

6.1. Basic reductions, Symmetrization

Throughout, we assume that $n \geq 2$ is a fixed integer and $\delta \in (\frac{1}{2}, 1]$ is a given threshold. We begin with the standard discretization, which will later allow us to pass to various combinatorial and optimization arguments. We consequently use the notation $\mathcal{C}(n,m)$, as previously introduced in Definition 5.1.

PROPOSITION 6.1. *To prove the threshold bound (6.2), it is enough to verify that*

$$\sup_{\substack{m\in\{1,2,\dots\}\\ X\in\mathcal{C}(n,m)}} \mathbb{P}\Big(\max_{1\leq i<j\leq n}|X_i-X_j|\geq\delta\Big)=\frac{n(1-\delta)}{2-\delta}\wedge 1. \tag{6.3}$$

PROOF. Assume that (6.3) holds and fix any n-variate vector $X\in\mathcal{C}_n$. Let m be a positive integer with $\delta>\frac{2}{m}+\frac{1}{2}$. According to Proposition 1.2, there exists a vector $X^{(m)}\in\mathcal{C}(n,m)$ such that $|X_j-X_j^{(m)}|\leq\frac{1}{m}$ almost surely for all indices $j=1,2,\dots,n$. Thus, by the triangle inequality, we have

$$\mathbb{P}\Big(\max_{1\leq i<j\leq n}|X_i-X_j|\geq\delta\Big)\leq\mathbb{P}\Big(\max_{1\leq i<j\leq n}|X_i^{(m)}-X_j^{(m)}|\geq\delta-\frac{2}{m}\Big)\leq\frac{n\Big(1-(\delta-\frac{2}{m})\Big)}{2-(\delta-\frac{2}{m})},$$

where the second inequality follows from (6.3). Taking $m\to\infty$ ends the proof. □

Applying the symmetrization technique developed in the previous chapter, we obtain the following crucial reduction.

COROLLARY 6.1. *We have the inequality*

$$\begin{aligned}&\sup_{\substack{m\in\{1,2,\dots\}\\ X\in\mathcal{C}(n,m)}} \mathbb{P}\Big(\max_{1\leq i<j\leq n}|X_i-X_j|\geq\delta\Big)\\ &\qquad\leq 2\cdot\sup_{\substack{m\in\{1,2,\dots\}\\ X\in\mathcal{C}'(n,m)}}\mathbb{P}\Big(\Big\{\max_{1\leq i<j\leq n}|X_i-X_j|\geq\delta\Big\}\cap A\Big),\end{aligned} \tag{6.4}$$

where $\mathcal{C}'(n,m)$ is the subset of all those $X\in C(n,m)$ that satisfy $\mathbb{P}(A)=1/2$ and

$$\frac{1-x}{x}\cdot\sum_{i=1}^n\mathbb{P}\Big(\{X_i=x\}\cap A\Big)=\sum_{i=1}^n\mathbb{P}\Big(\{X_i=1-x\}\cap A\Big),\quad\forall x\in(0,1]. \tag{6.5}$$

PROOF. By Lemma 5.2 point (iii) and Lemma 5.3, the left-hand side of (6.4) does not exceed the right-hand side. □

It will later become clear that (6.4) is in fact an equality. For the time being, the above argumentation allows us to reduce our main problem to the identification of

$$\sup_{\substack{m\in\{1,2,\dots\}\\ X\in\mathcal{C}'(n,m)}}\mathbb{P}\Big(\Big\{\max_{1\leq i<j\leq n}|X_i-X_j|\geq\delta\Big\}\cap A\Big). \tag{6.6}$$

The advantage of this new formulation over the original one (seen on the left-hand side of (6.4)) comes from the fact that we can study the behavior of X restricted to the set A. As we will demonstrate, the analysis of this expression can be performed in a purely analytic setup, with the use of combinatorial arguments. Suitably, consider the measure space $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda)$, where λ stands for the Lebesgue measure.

DEFINITION 6.1. *For $k \in \mathbb{N} \setminus \{0\}$, let $\Lambda(k)$ be the family of all those functions $(H, L) : \mathbb{R}_+ \to [0, 1]^2$, which satisfy the following four requirements:*

(Λ.1) $L(x) \le \frac{1}{2} \le H(x)$ *for all* $x \in \mathbb{R}_+$,
(Λ.2) *H and L are right-continuous step functions with a finite number of steps,*
(Λ.3) $\lambda(\{H > \frac{1}{2}\}) + \lambda(\{L < \frac{1}{2}\}) \le \frac{k}{2}$,
(Λ.4) *for any $y \in (0, 1]$ we have*

$$\frac{1-y}{y} \cdot [\lambda(\{H = y\}) + \lambda(\{L = y\})] = \lambda(\{H = 1-y\}) + \lambda(\{L = 1-y\}).$$

Here is a key statement, which links the above probabilistic considerations with the analytic context we have just introduced.

PROPOSITION 6.2. *The value of (6.6) is not bigger than*

$$\sup_{(H,L)\in\Lambda(n)} \lambda(\{H \ge L + \delta\}). \tag{6.7}$$

PROOF. Fix $m \in \{1, 2, \dots\}$, $X \in \mathcal{C}'(n, m)$ and the associated event A. We will construct $(H_X, L_X) \in \Lambda(n)$ such that

$$\mathbb{P}\Big(\Big\{\max_{1\le i<j\le n} |X_i - X_j| \ge \delta\Big\} \cap A\Big) = \lambda(\{H_X \ge L_X + \delta\}). \tag{6.8}$$

As $X \in \mathcal{C}'(n, m)$, there exists a natural number $l \le m^n$ such that X takes exactly l different values. It follows that A can be partitioned into disjoint family $\{A_k\}_{k=1}^l$ of events of positive probability, so that X is constant on every element of this partition: let $(X_1, \dots, X_n) \equiv (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$ on A_k. For $1 \le k \le l$, we set

$$p_k = \begin{cases} n\mathbb{P}(A_k) & \text{if } \max_{1\le i<j\le n} |X_i - X_j| < \delta \text{ on } A_k, \\ (n-1)\mathbb{P}(A_k) & \text{if } \max_{1\le i<j\le n} |X_i - X_j| \ge \delta \text{ on } A_k, \end{cases}$$

and introduce a disjoint partition $\mathcal{I} = \{I_k\}_{k=1}^{l+1}$ of $\mathbb{R}_+$ by

$$I_k = \begin{cases} [p_1 + \cdots + p_{k-1}, p_1 + \cdots + p_k) & \text{for } 1 \le k \le l, \\ [p_1 + \cdots + p_l, \infty) & \text{for } k = l+1. \end{cases}$$

Now we are ready to define (H_X, L_X), setting its values on each element of family $\mathcal{I}$ separately. Assume that $k \in \{1, 2, \dots, l+1\}$ and distinguish three major cases.

• $k = l + 1$. We put $H_X(x) = L_X(x) = \frac{1}{2}$ for all $x \in I_{l+1}$.

• $k \le l$, $p_k = n\mathbb{P}(A_k)$. We split I_k into n consecutive intervals $\{I_{k,s}\}_{s=1}^n$ (left-closed, right-open) of equal length and set

$$H_X(x) = \max\left(x_s^{(k)}, 1/2\right), \qquad L_X(x) = \min\left(x_s^{(k)}, 1/2\right) \tag{6.9}$$

whenever $x \in I_{k,s}$ and $s = 1, 2, \dots, n$.

• $k \le l$, $p_k = (n-1)\mathbb{P}(A_k)$. Then there are two indices $1 \le i_1 \ne i_2 \le n$ such that $X_{i_1} \ge X_{i_2} + \delta$ on A_k. The choice of i_1, i_2 may not be unique (in such a case, we pick any pair with this property). We divide I_k into $n-1$ consecutive intervals $I_{k,s}$ of equal length, $s \in \{1, 2, \ldots, n\} \setminus \{i_2\}$, and put

$$H_X(x) = x_{i_1}^{(k)}, \qquad L_X(x) = x_{i_2}^{(k)} \qquad \text{if } x \in I_{k,i_1},$$

while for $x \in I_{k,s}$ and $s \in \{1, 2, \ldots, n\} \setminus \{i_1, i_2\}$, we use (6.9). Put differently, we proceed as in the previous case, but the intervals I_{k,i_1}, I_{k,i_2} are now "glued" into one.

Let us check that the function (H, L) we have just obtained does belong to $\Lambda(n)$, i.e., it satisfies the four requirements (Λ.1)-(Λ.4). The first two conditions hold directly by the construction. To verify the point (Λ.3), we inspect carefully the three cases considered above. Note that $H = L = 1/2$ on I_{l+1}, so

$$\lambda(\{H > 1/2\} \cap I_{l+1}) + \lambda(\{L < 1/2\} \cap I_{l+1}) = 0.$$

If $k \le l$ and $p_k = n\mathbb{P}(A_k)$, then the restrictions of H and L to I_k are given by (6.9). Directly by this formula, we see that the sets $\{H > 1/2\} \cap I_k$ and $\{L < 1/2\} \cap I_k$ are disjoint and hence

$$\lambda(\{H > 1/2\} \cap I_k) + \lambda(\{L < 1/2\} \cap I_k) \le \lambda(I_k) = p_k = n\mathbb{P}(A_k).$$

Finally, if $k \le l$ and $p_k = (n-1)\mathbb{P}(A_k)$, then the above construction implies that the intersection of $\{H > 1/2\} \cap I_k$ and $\{L < 1/2\} \cap I_k$ is precisely the interval I_{k,i_1}. Consequently,

$$\lambda(\{H > 1/2\} \cap I_k) + \lambda(\{L < 1/2\} \cap I_k) \le \lambda(I_{k,i_1}) + \lambda(I_k) = n\mathbb{P}(A_k).$$

Summing the above inequalities/equalities over k and noting that

$$\mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots + \mathbb{P}(A_l) = \mathbb{P}(A) = 1/2,$$

we obtain (Λ.3). The last property is currently a direct consequence of (6.5). $\square$

The next step is the following reduction.

PROPOSITION 6.3. *The quantity (6.7) can be rewritten as*

$$\sup_{(H,L)\in\Lambda^\delta(n)} \lambda(\{H \ge L + \delta\}),$$

where $\Lambda^\delta(n)$ is the subset of all $(H, L) \in \Lambda(n)$ satisfying

$$\{H \in (1/2, \delta)\} \cup \{L \in (1-\delta, 1/2)\} = \emptyset \tag{6.10}$$

and

$$\{H \ge L + \delta\} = \{L < 1/2\}. \tag{6.11}$$

PROOF. Fix $(H, L) \in \Lambda(n)$ and assume that the condition (6.10) or (6.11) is not satisfied. If (6.10) fails, we modify H and/or L on the "bad" sets, changing their values to $1/2$ there. After this modification, the points (Λ.1)-(Λ.4) are still satisfied and the value of $\lambda(\{H \ge L+\delta\})$ remains unchanged. Now suppose that (6.11) does not hold. Because of the trivial inclusion $\{H \ge L+\delta\} \subseteq \{L < 1/2\}$ and the equality $\{L < 1/2\} = \{L \le 1-\delta\}$ we have just guaranteed, there must exist $0 \le a < b$ and $0 < \gamma \le 1-\delta$ such that $L = \gamma$ and $H < \gamma + \delta$ on $[a, b)$. By the point (Λ.4), we can find pairwise disjoint intervals $[a_j, b_j)$, $j = 1, \ldots, m$, satisfying

$$\bigcup_{j=1}^m [a_j, b_j) \subset \{H = 1-\gamma\} \quad \text{and} \quad \sum_{j=1}^m (b_j - a_j) = \frac{1-\gamma}{\gamma} \cdot (b-a).$$

Therefore, we can perform the following rearrangement:

(1) change L on $[a, b)$ from γ to $\frac{1}{2}$,
(2) change H on $\bigcup_{j=1}^m [a_j, b_j)$ from $1-\gamma$ to 1.

This "corrects" the behavior of (H, L) on the troublesome interval $[a, b)$. Note that the obtained function belongs to $\Lambda(n)$ and the value of $\lambda(\{H \geq L+\delta\})$ is not decreased. It remains to observe that we may guarantee the validity of (6.11), by performing sufficiently many such transformations. □

The central part of our reasoning is the following estimate.

LEMMA 6.1. *We have the identity*

$$\phi := \sup \frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}\}) - \lambda(\{L < \frac{1}{2}\})} = \frac{1-\delta}{\delta},$$

where the supremum is taken over all $k \in \mathbb{N} \setminus \{0\}$ *and all* $(H, L) \in \Lambda^\delta(k)$ *satisfying* $\lambda(\{H > 1/2\}) > 0$.

We postpone a careful verification of this lemma to the next section and proceed with the proof of our main result.

PROOF OF THEOREM 6.2. Eventually, exploiting Proposition 6.1, Lemma 5.3, Propositions 6.2 and 6.3, we can write

$$\begin{aligned} \sup_{(X_1,X_2,\ldots,X_n)\in\mathcal{C}_n} \mathbb{P}\Big(\max_{1\leq i<j\leq n} |X_i - X_j| \geq \delta\Big) &\leq 2 \cdot \sup_{(H,L)\in\Lambda(n)} \lambda(\{H \geq L+\delta\}) \\ &= 2 \cdot \sup_{(H,L)\in\Lambda^\delta(n)} \lambda\left(\{L < 1/2\}\right). \end{aligned}$$

Fix $(H, L) \in \Lambda^\delta(n)$. By Lemma 6.1, we have $\lambda(\{L < 1/2\}) \leq (1-\delta)\lambda(\{H > 1/2\})$, while the point ($\Lambda$.3) gives $\lambda(\{L < 1/2\}) + \lambda(\{H > 1/2\}) \leq n/2$. Combining these two estimates, we immediately obtain

$$\mathbb{P}\Big(\max_{1\leq i<j\leq n} |X_i - X_j| \geq \delta\Big) \leq \frac{n(1-\delta)}{2-\delta}.$$

It remains to prove the sharpness of (6.2). Observe that the function $\delta \mapsto \frac{1-\delta}{2-\delta}$ is decreasing on $[0, 1]$, so the claim will follow if we construct an appropriate coherent vector $(Z_i)_{i=1}^n$ for every δ with $n(1-\delta)/(2-\delta) \leq 1$. To this end, let $\{A_0, A_1, \ldots, A_n\} \cup \{B_0, B_1, \ldots, B_n\}$ be a measurable partition of Ω satisfying

$$\mathbb{P}(A_0) = \mathbb{P}(B_0) = \frac{1}{2} \cdot \left(1 - \frac{n(1-\delta)}{2-\delta}\right)$$

and

$$\mathbb{P}(A_i) = \mathbb{P}(B_i) = \frac{1}{2} \cdot \frac{1-\delta}{2-\delta} \quad \text{for } 1 \leq i \leq n.$$

Put $A = \bigcup_{i=0}^n A_i$, $B = \bigcup_{i=0}^n B_i$ and consider the σ-algebras

$$\mathcal{F}_i = \sigma\Big(A_i, B_i, (A \cup B_{i+1}) \setminus A_{i+1}, (B \cup A_{i+1}) \setminus B_{i+1}\Big), \quad i = 1, 2, \ldots, n$$

(with the cyclic convention $A_{n+1} = A_1$, $B_{n+1} = B_1$). It is straightforward to check that the variables $Z_i = \mathbb{E}(\mathbb{1}_A|\mathcal{F}_i)$, $i = 1, 2, \ldots, n$, satisfy

$$Z_i = \begin{cases} 1 & \text{on } A_i, \\ 0 & \text{on } B_i, \\ \delta & \text{on } (A \cup B_{i+1}) \setminus (A_i \cup A_{i+1}), \\ 1-\delta & \text{on } (B \cup A_{i+1}) \setminus (B_i \cup B_{i+1}). \end{cases}$$

Consequently, we have the bound $\max_{1\leq i<j\leq n} |Z_i - Z_j| \geq \delta$ on each A_k and each B_k. This proves the estimate

$$\mathbb{P}\Big(\max_{1\leq i<j\leq n} |Z_i - Z_j| \geq \delta \Big) \geq \mathbb{P}(\Omega \setminus (A_0 \cup B_0)) = \frac{n(1-\delta)}{2-\delta},$$

which is the desired lower bound. □

6.2. Proof of Lemma 6.1

For the sake of clarity, let us first describe the rough idea behind the proof. We have

$$\phi = \sup \frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}\}) - \lambda(\{L < \frac{1}{2}\})} = \sup \frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}, L = \frac{1}{2}\})},$$

where both suprema are as in the statement of Lemma 6.1. Our approach rests on expressing the set $\{H > \frac{1}{2}, L = \frac{1}{2}\}$ as the finite union of appropriate pairwise disjoint intervals $I_1^{(1)}$, $I_2^{(1)}$, …, $I_{k_1}^{(1)}$, and splitting $\{L < \frac{1}{2}\}$ into the union of k_1 families $\mathcal{A}_1$, $\mathcal{A}_2$, …, $\mathcal{A}_{k_1}$, each of which consists of pairwise disjoint intervals. Then we may write

$$\frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}, L = \frac{1}{2}\})} = \frac{\sum_{j=1}^{k_1} \sum_{J\in\mathcal{A}_j} \lambda(J)}{\sum_{j=1}^{k_1} \lambda(I_j^{(1)})}. \tag{6.12}$$

While the choice of $I_j^{(1)}$ is simple (the intervals correspond to different values taken by the function H), the construction of $\mathcal{A}_j$ is more complex. For each j, we will use a recursive procedure, starting from $I_j^{(1)}$, which will arrange the elements of $\mathcal{A}_j \cup \{I_j^{(1)}\}$ into a directed tree structure. Now, the obvious inequality

$$\frac{\sum_{j=1}^{k_1} \sum_{J\in\mathcal{A}_j} \lambda(J)}{\sum_{j=1}^{k_1} \lambda(I_j)} \leq \max_{1\leq j\leq k_1} \frac{\sum_{J\in\mathcal{A}_j} \lambda(J)}{\lambda(I_j)}$$

will enable us to restrict the analysis of (6.12) to the case $k_1 = 1$, i.e., to the case of a single tree. Actually, this inequality leads to a much more significant reduction: as we will see below, it allows us to assume that the tree is a path. This very special case will be successfully treated by arguments coming from dynamic programming.

We turn our attention to the rigorous verification of the above plan. It is convenient to split the remaining part of this section into three separate parts, devoted to the construction of the tree structure, the verification of its properties and the dynamic programming argument.

6.2.1. Tree structure. We will use some basic terminology from the theory of graphs. Recall that a simple (directed) graph G is an ordered pair (V_G, E_G), where V_G is the set of vertices and $E_G \subseteq V_G \times V_G$ is the collection of all edges. A simple graph is called a tree, if any two vertices are connected by exactly one path; a forest is a disjoint union of trees.

From now on, we will use a shorter notation and write $\Lambda^\delta(\mathbb{N})$ instead of $\bigcup_{k=1}^\infty \Lambda^\delta(k)$. We start with an arbitrary $(H, L) \in \Lambda^\delta(\mathbb{N})$ satisfying $\lambda(\{H > 1/2\}) > 0$ and describe how such a function gives rise to a (directed) forest graph $\mathcal{T}_{(H,L)} = (\mathcal{V}_{(H,L)}, \mathcal{E}_{(H,L)})$. We will proceed by induction, the intervals under consideration will always be left-closed and right-open:

(1) Induction base. By $(\Lambda.4)$, we have $\lambda(\{L < 1/2\}) < \lambda(\{H > 1/2\})$ and hence $\lambda(\{H > 1/2, L = 1/2\}) > 0$. Therefore, we can find a finite family $\mathcal{V}_1 = \{I_1^{(1)}, I_2^{(1)}, \ldots, I_{k_1}^{(1)}\}$ of disjoint intervals, such that

$$\bigcup_{j=1}^{k_1} I_j^{(1)} = \{H > 1/2,\ L = 1/2\}$$

and such that H is constant on each interval, say, $H = x_j^{(1)}$ on $I_j^{(1)}$ for $1 \le j \le k_1$. Set $\mathcal{E}_1 = \emptyset$.

(2) Induction step. Suppose that we have successfully constructed $\mathcal{V}_j$ and $\mathcal{E}_j$ for $j \le i-1$. Moreover, assume that $\mathcal{V}_{i-1} = \{I_1^{(i-1)}, I_2^{(i-1)}, \ldots, I_{k_{i-1}}^{(i-1)}\}$ and $H = x_j^{(i-1)}$ on $I_j^{(i-1)}$ for $1 \le j \le k_{i-1}$. By the point $(\Lambda.4)$, there exists a finite family $\bigcup_{j=1}^{k_{i-1}} \{J_1^j, J_2^j, \ldots, J_{m_j}^j\}$ of disjoint intervals, such that

$$\bigcup_{l=1}^{m_j} J_l^j \subseteq \{L = 1 - x_j^{(i-1)}\} \setminus \bigcup_{n=1}^{i-1} \bigcup \mathcal{V}_n, \quad \sum_{l=1}^{m_j} \lambda(J_l^j) = \frac{1 - x_j^{(i-1)}}{x_j^{(i-1)}} \cdot \lambda(I_j^{(i-1)})$$

for $j = 1, 2, \ldots, k_{i-1}$, and such that H is constant on each J_l^j. Set

$$\mathcal{V}_i = \bigcup_{j=1}^{k_{i-1}} \{J_1^j, J_2^j, \ldots, J_{m_j}^j\} \ \text{ and } \ \mathcal{E}_i = \mathcal{E}_{i-1} \cup \bigcup_{j=1}^{k_{i-1}} \{I_j^{(i-1)}\} \times \{J_1^j, J_2^j, \ldots, J_{m_j}^j\},$$

and put $\mathcal{V}_{(H,L)} = \bigcup_{i=1}^\infty \mathcal{V}_i$, $\mathcal{E}_{(H,L)} = \bigcup_{i=1}^\infty \mathcal{E}_i$.

Let us perform an explicit calculation to acquire some stronger intuition regarding the above construction.

Example 6.1. Let $\delta = 0.7$ and consider a pair (H, L) given by

$$H = \chi_{[1,3)} + \frac{7}{8}\left(\chi_{[0,1)} + \chi_{[3,5)} + \chi_{[8,12)}\right) + \frac{3}{4}\left(\chi_{[5,8)} + \chi_{[12,15)}\right) + \frac{1}{2}\chi_{[15,\infty)},$$
$$L = \frac{1}{8}\chi_{[0,1)} + \frac{1}{4}\chi_{[1,3)} + \frac{1}{2}\chi_{[3,\infty)}.$$

It is not difficult to check that $(H, L) \in \Lambda^\delta(36)$ – see Figure 1.

Let us briefly explain the construction of the forest $\mathcal{T}_{(H,L)}$. The starting point is to look at the set $\{H > 1/2, L = 1/2\} = [3, 15)$. In our case, this set splits into

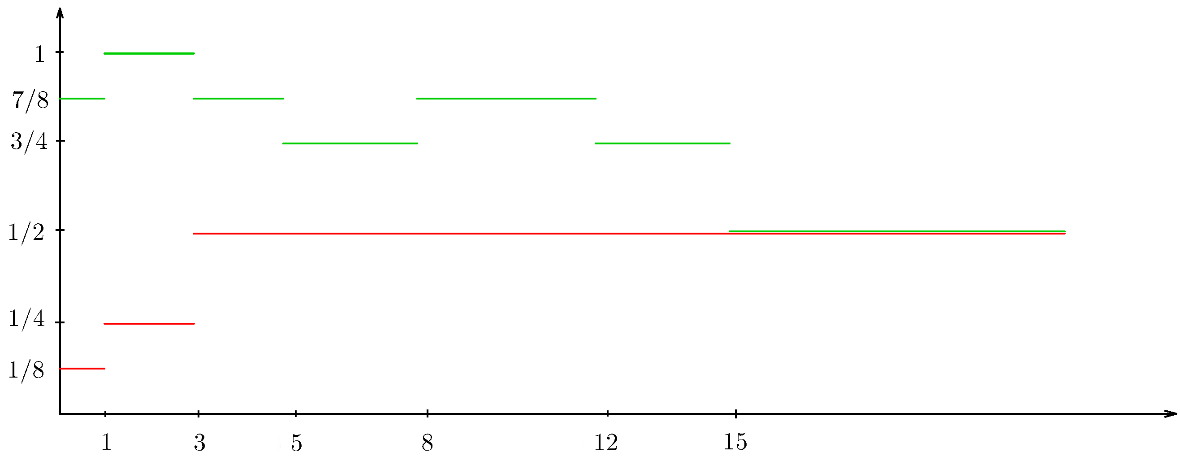

FIGURE 1. The graphs of the function H (green) and L (red).

four intervals on which H is constant: $[3, 5)$, $[5, 8)$, $[8, 12)$ and $[12, 15)$. These four intervals are the roots of four trees which will form the forest $\mathcal{T}_{(H,L)}$.

Next, for each root we describe its descendants. It is best to explain the procedure on a given root, say, $[5, 8)$. The length of the interval is equal to 3 and the function H is equal to 3/4 there. The application of the property (Λ.4) with $y = 3/4$ gives

$$\lambda(\{H = 3/4\}) = 3\lambda(\{L = 1/4\}), \tag{6.13}$$

i.e., the set $\{L = 1/4\}$ is three times smaller than $\{H = 3/4\}$. The children of $[5, 8)$ are the pairwise disjoint subintervals $J_1^1, J_2^1, \ldots, J_{m_1}^1$ of $\{L = 1/4\}$ for which the measure constraint (6.13) is preserved:

$$\lambda([5, 8)) = 3\lambda\Big(\bigcup_{j=1}^{m_1} J_j^1\Big),$$

and such that H is constant on each J_j^1. There is a lot of ambiguity with the choice of J^j's, one may actually take a single child $J_1^1 = [1, 2)$. We repeat a similar procedure with each root, making sure that all the children obtained in the process are pairwise disjoint. For example, at the end we may obtain the following (partial) forest.

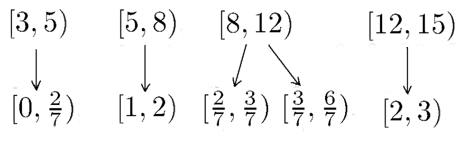

FIGURE 2. Partial forest – roots and their children.

Note that the intervals $[2/7, 3/7)$ and $[3/7, 6/7)$ could as well be merged into one $[2/7, 6/7)$: then the root $[8, 12)$ would have just one descendant. Next, we continue the procedure, but now the role of the roots is played by the children of the first generation which have been just constructed. It is clear that the procedure is well-defined – by property (Λ.4), at each step there are no problems with the existence of intervals satisfying appropriate measure and disjointness requirements. We would just like to mention that any interval, on which H is equal to 1, does not have any descendants (the tree is cut at such a vertex). $\triangle$

6.2.2. Properties of $\mathcal{T}_{(H,L)}$. We start with the following important fact.

PROPOSITION 6.4. *The family $\mathcal{V}_{(H,L)}$ is disjoint and*

$$\bigcup \mathcal{V}_{(H,L)} = \{H > 1/2\},$$

up to a set of measure zero.

PROOF. The first part follows from the very construction. To prove the second part, we will first show inductively that

$$H > 1/2 \quad \text{for every } J \in \mathcal{V}_{(H,L)}. \tag{6.14}$$

Indeed, we obviously have $H > 1/2$ on $\bigcup \mathcal{V}_1$. So, fix $i \in \{2, 3, \dots\}$ and assume that $H > 1/2$ on $\bigcup_{n=1}^{i-1} \bigcup \mathcal{V}_n$. Let $J \in \mathcal{V}_i$ be an arbitrary interval and let $I \in \mathcal{V}_{i-1}$ be the father of J (relative to the structure of the tree $\mathcal{T}_{(H,L)}$). Then there exists $x > \frac{1}{2}$ such that $H \equiv x$ on I and $L \equiv 1 - x < \frac{1}{2}$ on J. By the definition of $\Lambda^\delta(\mathbb{N})$, we have $\{L < 1/2\} = \{H \geq L + \delta\}$ and hence $H \geq 1 - x + \delta \geq \delta > \frac{1}{2}$ on J. This completes the proof of (6.14). To show the reverse inclusion (up to a set of measure zero), put $U = \{H > 1/2\} \setminus \bigcup \mathcal{V}_{(H,L)}$ and assume that $\lambda(U) > 0$. Recall that, again by the definition of $\Lambda^\delta(\mathbb{N})$, we have

$$\{H \in (1/2, \delta)\} = \{L \in (1 - \delta, 1/2)\} = \emptyset.$$

Fix $y \in [\delta, 1]$ and note that by the construction of the sets $\mathcal{V}_i$ above, we may write

$$\frac{1-y}{y} \cdot \sum_{n=1}^{i-1} \sum_{\{I \in \mathcal{V}_n:\ H \equiv y \text{ on } I\}} \lambda(I) = \sum_{n=2}^{i} \sum_{\{J \in \mathcal{V}_n:\ L \equiv 1-y \text{ on } J\}} \lambda(J),$$

for all $i = 2, 3, \dots$ Hence, passing with i to infinity yields

$$\frac{1-y}{y} \cdot \sum_{\{I \in \mathcal{V}_{(H,L)}:\ H \equiv y \text{ on } I\}} \lambda(I) = \sum_{\{J \in \mathcal{V}_{(H,L)}:\ L \equiv 1-y \text{ on } J\}} \lambda(J). \tag{6.15}$$

On the other hand, just by the property $(\Lambda.4)$, we have

$$\frac{1-y}{y} \cdot \lambda(\{H = y\}) = \lambda(\{L = 1 - y\}). \tag{6.16}$$

Subtracting (6.15) from (6.16), we get

$$\frac{1-y}{y} \cdot \lambda\left(\{H = y\} \cap U\right) = \lambda\left(\{L = 1 - y\} \cap U\right), \quad \forall y \in [\delta, 1]. \tag{6.17}$$

Next, by point $(\Lambda.2)$, there exists a finite sequence $y_1, y_2, \dots, y_k \in [\delta, 1]$, satisfying

$$\lambda\left(\{H \notin \{y_1, y_2, \dots, y_k\}\} \cap U\right) = \lambda\left(\{L \notin \{1 - y_1, 1 - y_2, \dots, 1 - y_k\}\} \cap U\right) = 0.$$

Therefore, summing (6.17) for $y_1, y_2, \ldots, y_k$, we obtain the inequality

$$\begin{aligned}\sum_{i=1}^{k} \lambda\left(\{L = 1 - y_i\} \cap U\right) &= \sum_{i=1}^{k} \frac{1-y_i}{y_i} \cdot \lambda\left(\{H = y_i\} \cap U\right) \\ &\le \frac{1-\delta}{\delta} \cdot \sum_{i=1}^{k} \lambda\left(\{H = y_i\} \cap U\right),\end{aligned}$$

and hence

$$\lambda\left(\{L < 1/2\} \cap U\right) < \lambda\left(\{H > 1/2\} \cap U\right) \tag{6.18}$$

if only the right-hand side of (6.18) is positive. At the same time, we have

$$U \cap \{H > 1/2, L = 1/2\} = \emptyset,$$

since the set $\{H > 1/2, L = 1/2\}$ has been already covered by $\mathcal{V}_1$. Thus, we get

$$\{H > 1/2\} \cap U = \{L < 1/2\} \cap U = U$$

and thus

$$\lambda\left(\{L < 1/2\} \cap U\right) = \lambda\left(\{H > 1/2\} \cap U\right),$$

which contradicts (6.18). □

As we explained in the beginning of this section, the above graph structure can be exploited in the study of Lemma 6.1. Under the notation we have just introduced, the expression for ϕ can be rewritten in the form

$$\phi = \sup_{(H,L) \in \Lambda^\delta(\mathbb{N})} \frac{\sum_{J \in \mathcal{V}_{(H,L)} \setminus \mathcal{V}_1} \lambda(J)}{\sum_{I \in \mathcal{V}_1} \lambda(I)}. \tag{6.19}$$

We split the forest $\mathcal{T}_{(H,L)}$ into the disjoint trees: for $1 \le j \le k_1$, let $\mathcal{T}^j_{(H,L)}$ denote the directed tree with root $I^{(1)}_j$. Then we have

$$\begin{aligned}\frac{\sum_{J \in \mathcal{V}_{(H,L)} \setminus \mathcal{V}_1} \lambda(J)}{\sum_{I \in \mathcal{V}_1} \lambda(I)} &= \frac{\sum_{j=1}^{k_1} \left[\lambda\left(\bigcup \mathcal{T}^j_{(H,L)}\right) - \lambda\left(I^{(1)}_j\right)\right]}{\sum_{j=1}^{k_1} \lambda(I^{(1)}_j)} \\ &\le \max_{1 \le j \le k_1} \frac{\lambda\left(\bigcup \mathcal{T}^j_{(H,L)}\right) - \lambda\left(I^{(1)}_j\right)}{\lambda(I^{(1)}_j)}.\end{aligned}$$

Hence, as we have already mentioned above, in the problem (6.19) it is enough to consider (H, L) with $\mathcal{T}_{(H,L)} = \mathcal{T}^1_{(H,L)}$, i.e. in the context when the underlying forest structure consists of a single tree.

Let us discuss some further simplifications. With no loss of generality, we may assume that $\lambda(I^{(1)}_1) = 1$. Indeed, scaling $I^{(1)}_1$ by $c > 0$ results in scaling all intervals in $\mathcal{V}_2$ by the same factor, which, in turn, leads to the same scaling of all intervals generated by $\mathcal{V}_2$ (i.e., $\mathcal{V}_3$), and so on. Summarizing, we have obtained

$$\phi = \sup_{\Xi^\delta(\mathbb{N})} \left[\lambda\left(\bigcup \mathcal{V}_2\right) + \lambda\left(\bigcup \mathcal{V}_3\right) + \ldots\right], \tag{6.20}$$

where supremum is taken over

$$\Xi^\delta(\mathbb{N}) := \Big\{(H,L) \in \Lambda^\delta(\mathbb{N}) : \ \mathcal{T}_{(H,L)} = \mathcal{T}^1_{(H,L)} \ \text{ and } \ \lambda(I_1^{(1)}) = 1\Big\}.$$

Note that the series under supremum in (6.20) is uniformly convergent: by the construction, we have

$$\lambda\left(\bigcup \mathcal{V}_{m+1}\right) \leq \frac{1-\delta}{\delta} \cdot \lambda\left(\bigcup \mathcal{V}_m\right),$$

for all $m = 1, 2, \dots$ Therefore, we can reformulate (6.20) as

$$\phi = \lim_{m\to\infty} \left(\sup_{\Xi^\delta(\mathbb{N})} \sum_{j=2}^{m} \lambda\left(\bigcup \mathcal{V}_j\right)\right). \tag{6.21}$$

6.2.3. Dynamic programming. To compute the supremum discussed above, it is convenient to apply dynamic programming techniques. Let $\Phi : [\delta, 1] \to \mathbb{R}_+$ satisfy

$$\Phi(x) = \sup_{\mathbf{x}} \sum_{n=0}^{\infty} \prod_{i=0}^{n} \frac{1-x_i}{x_i}, \tag{6.22}$$

where the supremum is taken over all sequences $\mathbf{x} = (x_0, x_1, x_2, \dots)$ such that

$$x_0 = x, \quad x_n \in [\delta, 1] \quad \text{and} \quad x_{n+1} \geq 1 - x_n + \delta \quad \text{ for } n = 0, 1, 2, \dots$$

We may call Φ the Bellman function associated with (6.21). Its connection to the problem is described in the following statement.

PROPOSITION 6.5. *We have the identity*

$$\phi = \sup_{x\in[\delta,1]} \Phi(x).$$

PROOF. Fix $m \in \{1, 2, \dots\}$. Analogously to the reduction $\mathcal{T}_{(H,L)} = \mathcal{T}^1_{(H,L)}$, we easily verify that it is enough to handle (H, L) with $\mathcal{V}_j = \{I_1^{(j)}\}$ for $1 \leq j \leq m$. Taking $m \to \infty$, just as in (6.21), we get

$$\phi = \sup_{\Xi^\delta(\mathbb{N})} \sum_{j=2}^{\infty} \lambda(I_1^{(j)}), \tag{6.23}$$

where $I_1^{(n+1)}$ is generated by $I_1^{(n)}$ for each $n \geq 1$. Recall from construction that H is constant on such intervals: denote $H = x_{n-1}$ on $I_1^{(n)}$, $n = 1, 2, \dots$ Note that inequality $x_{n+1} \geq 1 - x_n + \delta$ is a straightforward consequence of $(H, L) \in \Lambda^\delta(\mathbb{N})$. Lastly, let $\Phi(x)$ denote the right-hand side of (6.23) with an additional restriction to $x_0 = x$. This yields the claim. □

We turn our attention to the identification of the formula for Φ. We start with a structural property of the Bellman function.

PROPOSITION 6.6. *For any $x \in [\delta, 1]$ we have the recurrence relation*

$$\Phi(x) = \frac{1-x}{x}\Big(1 + \sup_{y\geq 1-x+\delta} \Phi(y)\Big). \tag{6.24}$$

PROOF. The argument depends on the so-called Bellman's optimality principle. By (6.22), we simply have

$$\Phi(x) = \frac{1-x}{x} \cdot \Big(1 + \sup_{\widetilde{\mathbf{x}}} \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{1-x_i}{x_i}\Big),$$

where supremum is taken over all sequences $\widetilde{\mathbf{x}} = (x_1, x_2, \dots)$ such that

$$x_1 \geq 1 - x + \delta, \quad x_n \in [\delta, 1] \text{ and } x_{n+1} \geq 1 - x_n + \delta \quad \text{for } n = 1, 2, 3, \dots. \quad \square$$

Now we will make use of the following procedure, which is often successful in the treatment of various problems in dynamic programming. Namely, based on some experimentation, we will "guess" for which choice of $\mathbf{x}$ the supremum defining $\Phi(x)$ is attained, thus obtaining "a candidate" Ψ for the Bellman function. By the very definition, this candidate must satisfy $\Psi \leq \Phi$. The reverse estimate will be obtained by the verification that the candidate also satisfies the structural requirement (6.24), and exploiting this condition appropriately.

We proceed to the choice of $\mathbf{x}$. A little thought and a closer inspection suggests that problem (6.22) should be maximized by an alternating sequence

$$\hat{\mathbf{x}} = (x, 1 - x + \delta, x, 1 - x + \delta, x, \dots).$$

Indeed, this is quite a natural guess – we come up with $\hat{\mathbf{x}}$ simply by assuming equalities in the contraints for the coordinates x_0, x_1, x_2, ... Plugging this sequence into (6.22), we compute the corresponding candidate for $\Phi(x)$, obtaining

$$\Psi(x) := \frac{1-x}{x} + \frac{1-x}{x} \cdot \frac{x-\delta}{1-x+\delta} + \frac{1-x}{x} \cdot \frac{x-\delta}{1-x+\delta} \cdot \frac{1-x}{x} + \dots = \frac{1-x}{\delta},$$

for all $x \in [\delta, 1]$. Then $\Psi \leq \Phi$, as we have already commented above, so the proof will be complete if we manage to check that $\Psi \geq \Phi$.

PROOF OF LEMMA 6.1. First, we show that Ψ fulfills the recurrence (6.24). Indeed, for $x \in [\delta, 1]$, we have

$$\begin{aligned} \frac{1-x}{x}\Big(1 + \sup_{y \geq 1-x+\delta} \Psi(y)\Big) &= \frac{1-x}{x}\Big(1 + \Psi(1-x+\delta)\Big) \\ &= \frac{1-x}{x}\Big(1 + \frac{x-\delta}{\delta}\Big) = \Psi(x). \end{aligned}$$

Pick any $x \in [\delta, 1]$ and $\varepsilon > 0$. By (6.22), we can choose an admissible sequence $\mathbf{x}_\varepsilon = (x_0, x_1, \dots)$ (satisfying $x_0 = x$ and $x_{n+1} \geq 1 - x_n + \delta$, $n = 0, 1, 2, \dots$) such that

$$\Phi(x) \leq \varepsilon + \sum_{n=0}^{\infty} \prod_{i=0}^{n} \frac{1-x_i}{x_i}.$$

Since $\frac{1-x_i}{x_i} \leq \frac{1-\delta}{\delta}$, $i = 1, 2, \dots$, there is a natural number m for which

$$\Phi(x) \leq 2\varepsilon + \sum_{n=0}^{m} \prod_{i=0}^{n} \frac{1-x_i}{x_i}. \tag{6.25}$$

On the other hand, by recurrence relation (6.24), we can write

$$\begin{aligned}\Psi(x) = \Psi(x_0) &\geq \frac{1-x_0}{x_0}\Big(1+\Psi(x_1)\Big)\\ &= \frac{1-x_0}{x_0} + \frac{1-x_0}{x_0}\Psi(x_1)\\ &\geq \frac{1-x_0}{x_0} + \frac{1-x_0}{x_0}\frac{1-x_1}{x_1}\Big(1+\Psi(x_2)\Big)\\ &= \frac{1-x_0}{x_0} + \frac{1-x_0}{x_0}\frac{1-x_1}{x_1} + \frac{1-x_0}{x_0}\frac{1-x_1}{x_1}\Psi(x_2)\end{aligned}$$

and so on. After m steps, we obtain

$$\Psi(x) \geq \sum_{n=0}^{m}\prod_{i=0}^{n}\frac{1-x_i}{x_i} + \left(\prod_{i=0}^{m}\frac{1-x_i}{x_i}\right)\Psi(x_{m+1}) \geq \sum_{n=0}^{m}\prod_{i=0}^{n}\frac{1-x_i}{x_i}.$$

Hence, by (6.25), we get $\Psi(x) + 2\varepsilon \geq \Phi(x)$, and since $\varepsilon > 0$ was chosen arbitrarily, the reverse bound $\Psi \geq \Phi$ follows. The formula $\phi = \sup_{x\in[\delta,1]}\Phi(x) = (1-\delta)/\delta$ is thus established. This proves our claim and completes the proof of (6.2). $\square$

CHAPTER 7

Doob's estimate for coherent random variables

Preliminaries

To introduce the necessary notions, suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is an arbitrary nonatomic probability space. For a predefined integrable random variable ξ and a sequence $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_n)$ of sub-σ-algebras of $\mathcal{F}$, let vector $X = (X_1, \ldots, X_n)$ be given by

$$X_k = \mathbb{E}(\xi|\mathcal{G}_k) \quad \text{for all} \quad k = 1, 2, \ldots, n.$$

Notice that we no longer assume that random vector X is strictly coherent – it will be instructive to drop the assumption $\mathbb{P}(\xi \in \{0, 1\}) = 1$.

In this chapter, we will be interested in the universal sharp norm comparison of ξ and the associated maximal function of X, $M_{\mathcal{G}}\xi = \sup_j |\mathbb{E}(\xi|\mathcal{G}_j)|$. The starting point is the classical result of Doob, which asserts that

$$\left\| M_{\mathcal{G}}\xi \right\|_p \leq \frac{p}{p-1} \|\xi\|_p, \qquad 1 < p \leq \infty, \tag{7.1}$$

in the case when $\mathcal{G}$ is a filtration, i.e., we have the nesting condition

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \ldots \subseteq \mathcal{G}_n.$$

Furthermore, for each p the number $p/(p-1)$ is the best universal constant (i.e., not depending on the length of $\mathcal{G}$) allowed in the estimate.

The main objective of this chapter is to consider (7.1) for more general families of σ-algebras: we will assume that $\mathcal{G}$ can be decomposed into the union of filtrations. Specifically, we let $\mathcal{G}$ be of the form

$$\mathcal{G} := \left\{ \mathcal{G}_i^j \right\}_{\substack{1 \leq i \leq n, \\ 1 \leq j \leq k}},$$

and require the inclusions $\mathcal{G}_1^j \subseteq \mathcal{G}_2^j \subseteq \ldots \subseteq \mathcal{G}_n^j$ for every j, $1 \leq j \leq k$. No relation between σ-algebras $\mathcal{G}_i^j$ with different j is imposed.

Thus, our investigation can be seen as a natural halfway state between the study of general coherent distributions and classical martingales. Furthermore, this subject enters into the still vague framework of martingales indexed by partially ordered sets. For a general introduction to this emerging theory see [**49**], for related Doob's type inequalities see [**12, 14, 42, 74**].

Our approach will reveal an unexpected relation between the study of $\max_{i,j} |\mathbb{E}(\xi|\mathcal{G}_i^j)|$ and basic combinatorial properties of the uncentered Hardy–Littlewood maximal operator on tree-shaped domains. Due to this interdependence, we will be able to extend the classical approach introduced in [**38, 39**] and derive an appropriate sharp version of (7.1).

THEOREM 7.1. *Let* $1 < p < \infty$ *be a given parameter and assume that* $\mathcal{G} = \left\{\mathcal{G}_i^j\right\}_{\substack{1\leq i\leq n,\\ 1\leq j\leq k}}$, *is the union of filtrations as above. Then for any random variable* $\xi \in L^p$ *we have*

$$\|M_{\mathcal{G}}\xi\|_p \leq C_{p,k}\|\xi\|_p, \tag{7.2}$$

where $C_{p,k}$ *is the unique root of the equation*

$$(p-1)C_{p,k}^p - pC_{p,k}^{p-1} - (k-1) = 0. \tag{7.3}$$

For fixed $1 < p < \infty$ *and* $k \geq 1$, *the constant* $C_{p,k}$ *is the best possible: given* $\varepsilon > 0$, *there is* $n \in \mathbb{N}$, *a family* $\mathcal{G}$ *as above and a positive random variable* $\xi \in L^p$ *for which*

$$\|M_{\mathcal{G}}\xi\|_p > (C_{p,k} - \varepsilon)\|\xi\|_p. \tag{7.4}$$

That is, the constant $C_{p,k}$ is the best universal constant allowed in (7.2), where the universality is the non-dependence on n, the length of the filtrations building $\mathcal{G}$.

We would like to point out that the constant $C_{p,k}$ is still optimal if we restrict ourselves to random variables ξ taking values in $[0,1]$. This follows by an effortless approximation argument: given a positive almost extremal variable ξ (i.e., satisfying (7.4)), we replace it with $\min\{\xi, L\}$, where L is a positive constant. If L is sufficiently large, then this new variable still satisfies (7.4), and hence so does $\min\{\xi, L\}/L$, by homogeneity. It remains to note that the latter variable takes values in $[0,1]$.

Interestingly, in the case $\xi \in \{0,1\}$, which originates in the coherent context, the optimal constant is smaller: here is the precise formulation.

THEOREM 7.2. *Let* $\mathcal{G} = \left\{\mathcal{G}_i^j\right\}_{1\leq i\leq n, 1\leq j\leq k}$ *be the union of filtrations as above and let* $1 < p < \infty$. *Then for any random variable* ξ *with values in* $\{0,1\}$ *we have*

$$\|M_{\mathcal{G}}\xi\|_p \leq \left(1 + \frac{k}{p-1}\right)^{1/p}\|\xi\|_p. \tag{7.5}$$

The constant is the best possible for each k *and each* p.

We turn our attention to the analytic contents of the chapter. Let k be a fixed positive integer. Consider the set $\mathcal{R}_k = \bigcup_{j=1}^k H_j$, where H_j is the line segment on the complex plane, with endpoints 0 and $e^{2\pi ij/k}$, $j = 1, 2, \ldots, k$. That is, $\mathcal{R}_k$ is a tree-shaped domain being the union of k rays H_1, H_2, $\ldots$, H_k, each having length one. We equip $\mathcal{R}_k$ with the standard British railway metric and the normalized one-dimensional Lebesgue measure λ_k.

Then we can introduce the concept of the decreasing rearrangement on $\mathcal{R}_k$. Namely, for an arbitrary random variable ξ on $(\Omega, \mathcal{F}, \mathbb{P})$, we at first define its distribution function $d_\xi : [0,\infty) \to [0,1]$ by $d_\xi(s) = \mathbb{P}(|\xi| > s)$. Then the associated k-decreasing rearrangement $\xi^*_{(k)} : \mathcal{R}_k \to [0,\infty)$ is given by

$$\xi^*_{(k)}(e^{2\pi ij/k}t) = \inf\{s > 0 : d_\xi(s) \leq t\}, \qquad j = 1, 2, \ldots, k.$$

Equivalently, $\xi^*_{(k)}$ can be defined by taking the standard decreasing rearrangement ξ^* on $[0,1]$ and copying it on each ray H_j, in accordance with the natural order induced by the distance from 0. Thus, we immediately see that $|\xi|$ and $\xi^*_{(k)}$ have the same distributions (as random variables on Ω and $\mathcal{R}_k$, respectively). Furthermore, $\xi^*_{(k)}$ is radially decreasing, i.e., $\xi^*_{(k)}(x) = \xi^*_{(k)}(|x|)$ decreases as $|x|$ grows.

Finally, we introduce the uncentered Hardy–Littlewood maximal function $\mathcal{M}_{(k)}$ in the above setup. This operator acts on integrable functions f on $\mathcal{R}_k$ by the usual formula

$$\mathcal{M}_{(k)}f(x) = \sup \frac{1}{\lambda_k(B)} \int_B |f| \mathrm{d}\lambda_k, \qquad x \in \mathcal{R}_k,$$

where the supremum is taken over all open balls $B \subseteq \mathcal{R}_k$ which contain x. We will identify the L^p norm of this object.

Theorem 7.3. *For any $1 < p < \infty$ and any $k \geq 2$ we have $\|\mathcal{M}_{(k)}\|_{L^p \to L^p} = C_{p,k}$, where $C_{p,k}$ is given in* (7.3).

The core case $k = 2$ was established by Grafakos and Montgomery-Smith [**39**]. Our contribution is the analysis for $k \geq 3$. Furthermore, we will link the context of coherent distributions with the analytic setup above, intertwining the contents of Theorems 7.1 and 7.3.

Theorem 7.4. *Let k, $n \geq 1$ be fixed integers. Suppose further that ξ is an integrable random variable and assume that $\mathcal{G} = \left\{\mathcal{G}_i^j\right\}_{\substack{1 \leq i \leq n,\\ 1 \leq j \leq k}}$, is a union of filtrations as above. Then the maximal function $M_{\mathcal{G}}\xi$ satisfies*

$$(M_{\mathcal{G}}\xi)^*_{(k)} \leq \mathcal{M}_{(k)}(\xi^*_{(k)}) \qquad \lambda_k\text{-almost everywhere on } \mathcal{R}_k. \tag{7.6}$$

The remaining part of the chapter is split into three sections. In Section 7.1 we prove Theorem 7.4. Later on, in Section 7.2 we establish the L^p bound $\|\mathcal{M}_{(k)}\|_{L^p \to L^p} \leq C_{p,k}$, which allows us to deduce (7.2) immediately. Furthermore, we show there the sharpness of the latter inequality, thus completing the proofs of all associated results. Finally, in the last part of the chapter, we verify Theorem 7.2.

From now on, the parameter k will be kept fixed; to simplify the notation, we will skip the index and write ξ^*, $\mathcal{M}$ instead of $\xi^*_{(k)}$ and $\mathcal{M}_{(k)}$, respectively.

7.1. Proof of Theorem 7.4

We will need the following property of the Hardy–Littlewood maximal operator.

Lemma 7.1. *Suppose that ξ is an integrable random variable. Then for any $s > 0$ such that $\lambda_k(\mathcal{M}\xi^* > s) < 1$ we have*

$$s\Big((k-1)\lambda_k(\xi^* > s) + \lambda_k(\mathcal{M}\xi^* > s)\Big)$$
$$= (k-1)\int_{\{\xi^* > s\}} \xi^* \, d\lambda_k + \int_{\{\mathcal{M}\xi^* > s\}} \xi^* \, d\lambda_k.$$

Proof. If $s \geq \|\xi\|_\infty$, then the assertion is evident (both sides are zero), so from now on we assume that $s < \|\xi\|_\infty$. The function $\mathcal{M}\xi^*$ is radially decreasing along the rays of $\mathcal{R}_k$. Furthermore, it is continuous, which follows directly from Lebesgue's dominated convergence theorem. Thus there exists a point $u \in \mathcal{R}_k$, lying on the first ray H_1, for which $s = \mathcal{M}\xi^*(u)$. It is easy to identify the ball B for which the supremum defining $\mathcal{M}\xi^*(u)$ is attained: u must be one of its boundary points, and the intersection $B \cap H_j$ for $j \neq 1$ must be the part of H_j on which we have $f > s$. It remains to note that the equality

$$s = \mathcal{M}\xi^*(u) = \frac{1}{\lambda_k(B)} \int_B \xi^* \mathrm{d}\lambda_k$$

is equivalent to the claim. Indeed, we have

$$\lambda_k(B) = \tfrac{k-1}{k}\lambda_k(\xi^* > s) + \tfrac{1}{k}\lambda_k(\mathcal{M}\xi^* > s),$$

with a similar identity for $\int_B \xi^* \mathrm{d}\lambda_k$. □

Now we proceed to the proof of Theorem 7.4.

Proof of Theorem 7.4. It is enough to show the tail inequality

$$\mathbb{P}(M_{\mathcal{G}}\xi > s) \leq \lambda_k(\mathcal{M}\xi^* > s) \tag{7.7}$$

for all s. Now we consider two separate steps.

Step 1. Reductions. Let us first exclude the trivial cases: from now on, we will assume that $\lambda_k(\mathcal{M}\xi^* > s) < 1$ and $s < \|\xi\|_\infty$. Indeed, if $\lambda_k(\mathcal{M}\xi^* > s) = 1$, then there is nothing to prove, while for $s \geq \|\xi\|_\infty$ both sides of (7.7) are zero. Adding the full σ-algebras $\mathcal{G}^j_{n+1} = \mathcal{F}$, $j = 1, 2, \ldots, k$ to the collection $\mathcal{G}$ if necessary, we may and do assume that

$$\max_i |\mathbb{E}(\xi|\mathcal{G}_i^j)| \geq |\xi| \qquad \text{almost surely for all } j. \tag{7.8}$$

In particular, this gives $M_{\mathcal{G}}\xi \geq |\xi|$ with probability 1.

Step 2. Proof of theorem. Fix an arbitrary $s > 0$ and write

$$\mathbb{P}(M_{\mathcal{G}}\xi > s) = \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_k),$$

where $A_j = \{\max_i |\mathbb{E}(\xi|\mathcal{G}_i^j)| > s\}$, $j = 1, 2, \ldots, k$. Let us distinguish the additional event $A_0 = \{|\xi| > s\}$ and observe that $A_0 \subseteq A_j$ for each j, in the light of (7.8). Note that if $\tilde{A}_j$ is an arbitrary event satisfying $A_0 \subseteq \tilde{A}_j \subseteq A_j$, then we have

$$s\mathbb{P}(\tilde{A}_j) - \int_{\tilde{A}_j} |\xi| \mathrm{d}\mathbb{P} = \int_{\tilde{A}_j} (s - |\xi|)\mathrm{d}\mathbb{P} \leq \int_{A_j} (s - |\xi|)\mathrm{d}\mathbb{P} \leq 0, \tag{7.9}$$

where the latter bound follows from Doob's weak-type bound for martingale maximal function. Next, we write

$$\begin{aligned} &\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_k) \\ &= \mathbb{P}(A_0 \cup A_1 \cup A_2 \cup \ldots \cup A_k) \\ &= \mathbb{P}(A_0) + \mathbb{P}(A_1 \setminus A_0) + \mathbb{P}(A_2 \setminus (A_1 \cup A_0)) \\ &\quad + \ldots + \mathbb{P}(A_n \setminus (A_{n-1} \cup A_{n-2} \cup \ldots \cup A_0)). \end{aligned}$$

Set $\tilde{A}_j = A_0 \cup \big(A_j \setminus (A_{j-1} \cup A_{j-2} \cup \ldots \cup A_0)\big)$, apply (7.9) and add the estimates over j. Combining the result with the formula for $\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_k)$, we obtain

$$s\Big[\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_k) + (k-1)\mathbb{P}(A_0)\Big] = s\sum_{j=1}^k \mathbb{P}(\tilde{A}_j) \leq \sum_{j=1}^k \int_{\tilde{A}_j} |\xi| \mathrm{d}\mathbb{P},$$

or equivalently,

$$s\big[\mathbb{P}(M_{\mathcal{G}}\xi > s) + (k-1)\mathbb{P}(A_0)\big] \leq \int_{\{M_{\mathcal{G}}\xi>s\}} |\xi| \mathrm{d}\mathbb{P} + (k-1)\int_{A_0} |\xi| \mathrm{d}\mathbb{P}.$$

Since $|\xi|$ and ξ^* are equidistributed, we have $\mathbb{P}(A_0) = \lambda_k(\xi^* > s)$ and

$$\int_{A_0} |\xi| \mathrm{d}\mathbb{P} = \int_{\{\xi^*>s\}} \xi^* \mathrm{d}\lambda_k.$$

Plugging this above and applying Lemma 7.1, we get

$$\int_{\{M_{\mathcal{G}}\xi>s\}} (s-|\xi|)\mathrm{d}\mathbb{P} \leq \int_{\{M\xi^*>s\}} (s-\xi^*)\mathrm{d}\lambda_k,$$

or, subtracting the equality $\int_{\{|\xi|>s\}}(s-|\xi|)\mathrm{d}\mathbb{P} = \int_{\{\xi^*>s\}}(s-\xi^*)\mathrm{d}\lambda_k$,

$$\begin{aligned}\int_{\{M_{\mathcal{G}}\xi>s\}} (s-|\xi|)_+\mathrm{d}\mathbb{P} &\leq \int_{\{\mathcal{M}\xi^*>s\}} (s-\xi^*)_+\mathrm{d}\lambda_k \\ &\leq \int_{\mathcal{R}_k} \chi_{\{\mathcal{M}\xi^*>s\}}(s-\xi^*)_+\mathrm{d}\lambda_k.\end{aligned}$$

However, notice that the nonnegative functions $\chi_{\{\mathcal{M}\xi^*>s\}}$ and $(s-\xi^*)_+$ have the reversed monotonicity along the rays: the first of them is non-increasing, while the second is non-decreasing. Since $(s-|\xi|)_+$ and $(s-\xi^*)_+$ have the same distribution, inequality (7.7) follows. ☐

7.2. L^p estimates

We turn our attention to Theorems 7.1 and 7.3. Let us start with the L^p bound for the uncentered maximal operator; the key ingredient of the proof is the following weak-type estimate.

Proposition 7.1. *For every integrable function f on $\mathcal{R}_k$ and any $s>0$ we have*

$$\begin{aligned}&s\lambda_k(\mathcal{M}f > s) + s(k-1)\lambda_k(|f|>s) \\ &\qquad \leq \int_{\{\mathcal{M}f>s\}} |f| d\lambda_k + (k-1)\int_{\{|f|>s\}} |f| d\lambda_k. \qquad (7.10)\end{aligned}$$

Proof. It is convenient to split the reasoning into two steps.

Step 1. Special balls in $\mathcal{R}_k$. Let us consider the level set

$$E = \{x \in \mathcal{R}_k : \mathcal{M}f > s\}.$$

Then for each $x \in E$ there is an open ball $B_x \subseteq \mathcal{R}_k$ which contains x and satisfies

$$\lambda_k(B_x)^{-1}\int_{B_x} |f| \mathrm{d}\lambda_k > s.$$

This inequality implies that $B_x \subseteq E$ and hence $\bigcup_{x\in E} B_x = E$. Next, we will apply the classical Lindelöf's theorem, see [**5, 51**].

Theorem 7.5 (Lindelöf). *There is a countable subcover of each open cover of a subset of a space whose topology has a countable base.*

By the Lindelöf's theorem, we may pick a countable subcollection $(B_{x_n})_{n=1}^\infty$ such that $\bigcup_{n=1}^\infty B_{x_n} = E$. With no loss of generality, we may assume that B_{x_i} is not a subset of B_{x_j} for $i \neq j$. We fix an integer N and restrict ourselves to the finite family $\mathcal{B} = (B_{x_n})_{n=1}^N$.

The idea is to pick a subcollection $\mathcal{B}'$ of $\mathcal{B}$ which does not overlap too much. To this end, we will choose appropriate balls from each separate ray of $\mathcal{R}_k$, exploiting the natural order induced by the distance from 0. For simplicity, we will only describe the procedure for the k-th ray (i.e., for the interval $[0,1]$), the argument for other rays is the same, up to rotation. First, we pick a ball from $\mathcal{B}$ which contains zero and call it J_0 (if no ball in $\mathcal{B}$ contains zero, we let $J_0 = \emptyset$; if there are several balls with this property, we take the ball whose intersection with $[0,1]$ has the biggest measure). Next we apply the following inductive procedure.

1° Suppose that we have successfully defined J_n. Consider the family of all intervals $J \in \mathcal{B}$ which intersect J_n and satisfy $\sup J > \sup J_n$. If this family is nonempty, choose the interval with largest left-endpoint (if this object is not unique, pick the one with the biggest measure) and denote it by J_{n+1}.

2° If the family in 1° is empty, then consider all intervals $J \in \mathcal{B}$ with $\inf J \geq \sup J_n$. If this family is nonempty, choose an element with the smallest left-endpoint (if this object is not unique, pick the one with the biggest measure) and denote it by J_{n+1}.

3° Go to 1°.

Since the family $\mathcal{B}$ is finite, the procedure stops after a number of sets (in 1° and 2°, there are no balls to choose from) and returns a family J_0^j, J_1^j, J_2^j, $\ldots$, $J_{m_j}^j$ of balls. Observe that by the very construction, J_0^j, J_2^j, J_4^j, $\ldots$ are pairwise disjoint and the same is true for J_1^j, J_3^j, J_5^j, $\ldots$. Letting

$$\mathcal{B}' = \Big\{ J_\ell^j : 1 \leq \ell \leq m_j,\ j = 1, 2, \ldots, k \Big\},$$

we easily check that

$$\bigcup_{B \in \mathcal{B}} B = \bigcup_{B \in \mathcal{B}'} B. \tag{7.11}$$

Next, by the disjointness properties of the sequences J_i^j, note that a family $\mathcal{B}'$ has the following property: each point $x \in \mathcal{R}_k$ belongs to at most $k+1$ elements of $\mathcal{B}'$. Moreover, we can actually improve this last bound by 1. Now, say that there is a point $x_0 \in \mathcal{R}_k$ which belongs to exactly $k+1$ elements of $\mathcal{B}'$ and let us assume that x_0 belongs to the the k-th ray H_k. By the extremality of J_0^k we must have $(J_0^i \cap [0,1]) \subset (J_0^k \cap [0,1])$ for all $i = 1, 2, \ldots, k-1$, and hence

$$x_0 \in \bigcap_{j=1}^k J_0^j \cap J_1^k.$$

Thus, we simply remove J_0^k from the family $\mathcal{B}'$. Such a modification does not affect the validity of (7.11) and proves our assertion.

Step 2. Calculation. Since $\mathcal{B}' \subseteq \mathcal{B}$, each element B of $\mathcal{B}'$ satisfies

$$s\lambda_k(B) \leq \int_B |f| \mathrm{d}\lambda_k.$$

Summing over all $B \in \mathcal{B}'$, we thus obtain

$$s\left[\lambda\left(\bigcup_{B\in\mathcal{B}'} B\right) + \sum_{j=2}^{k} \lambda_k(A_j)\right] \leq \int_{\bigcup_{B\in\mathcal{B}'} B} |f| \mathrm{d}\lambda_k + \sum_{j=2}^{k} \int_{A_j} |f| \mathrm{d}\lambda_k,$$

where A_j is the collection of all $x \in \mathcal{R}_k$ which belong to exactly j elements of $\mathcal{B}'$. This is equivalent to

$$\begin{aligned} s\lambda\left(\bigcup_{B\in\mathcal{B}} B\right) &\leq \int_{\bigcup_{B\in\mathcal{B}} B} |f| \mathrm{d}\lambda_k + \sum_{j=2}^{k} \int_{A_j} (|f| - s) \mathrm{d}\lambda_k \\ &\leq \int_{\bigcup_{B\in\mathcal{B}} B} |f| \mathrm{d}\lambda_k + \sum_{j=2}^{k} \int_{A_j} (|f| - s)_+ \mathrm{d}\lambda_k \\ &\leq \int_{\bigcup_{B\in\mathcal{B}} B} |f| \mathrm{d}\lambda_k + (k-1) \int_{\bigcup_{j=2}^{k} A_j} (|f| - s)_+ \mathrm{d}\lambda_k \\ &\leq \int_{\bigcup_{B\in\mathcal{B}} B} |f| \mathrm{d}\lambda_k + (k-1) \int_{\mathcal{R}_k} (|f| - s)_+ \mathrm{d}\lambda_k. \end{aligned}$$

Now recall that the family $\mathcal{B}$ depended on N. Letting this parameter to infinity and using Lebesgue's monotone convergence theorem, we obtain

$$s\lambda(E) \leq \int_E |f| \mathrm{d}\lambda_k + (k-1) \int_{\mathcal{R}_k} (|f| - s)_+ \mathrm{d}\lambda_k.$$

This is precisely the claim. □

Now, using the standard integration argument, we obtain the L^p estimate for the uncentered maximal operator on $\mathcal{R}_k$.

Proof of (7.2). By Fubini's theorem, we have

$$\begin{aligned} &\int_{\mathcal{R}_k} (\mathcal{M}f)^p \mathrm{d}\lambda_k + (k-1) \int_{\mathcal{R}_k} |f|^p \mathrm{d}\lambda_k \\ &= p \int_0^\infty s^{p-1} \left[\lambda_k(\mathcal{M}f > s) + (k-1)\lambda_k(|f| > s)\right] \mathrm{d}s, \end{aligned}$$

which by (7.10) does not exceed

$$\begin{aligned} &p \int_0^\infty s^{p-2} \left[\int_{\{\mathcal{M}f > s\}} |f| \mathrm{d}\lambda_k + (k-1) \int_{\{|f| > s\}} |f| \mathrm{d}\lambda_k\right] \mathrm{d}s \\ &= \frac{p}{p-1} \int_{\mathcal{R}_k} \Big((\mathcal{M}f)^{p-1} |f| + (k-1)|f|^p\Big) \mathrm{d}\lambda_k. \end{aligned}$$

Here in the last passage we have used Fubini's theorem again. This gives the bound

$$\int_{\mathcal{R}_k} (\mathcal{M}f)^p \mathrm{d}\lambda_k \leq \frac{p}{p-1} \int_{\mathcal{R}_k} (\mathcal{M}f)^{p-1} |f| \mathrm{d}\lambda_k + \frac{k-1}{p-1} \int_{\mathcal{R}_k} |f|^p \mathrm{d}\lambda_k.$$

However, by Hölder's inequality, we have

$$\int_{\mathcal{R}_k} (\mathcal{M}f)^{p-1} |f| \mathrm{d}\lambda_k \leq \left(\int_{\mathcal{R}_k} (\mathcal{M}f)^p \mathrm{d}\lambda_k\right)^{(p-1)/p} \left(\int_{\mathcal{R}_k} |f|^p \mathrm{d}\lambda_k\right)^{1/p},$$

which combined with the previous estimate yields

$$(p-1)\left(\frac{\|\mathcal{M}f\|_{L^p(\mathcal{R}_k)}}{\|f\|_{L^p(\mathcal{R}_k)}}\right)^p - p\left(\frac{\|\mathcal{M}f\|_{L^p(\mathcal{R}_k)}}{\|f\|_{L^p(\mathcal{R}_k)}}\right)^{p-1} - (k-1) \leq 0.$$

It remains to note that the function

$$s \mapsto (p-1)s^p - ps^{p-1} - (k-1)$$

is increasing on $[1,\infty)$ and $C_{p,k}$ is its unique root. This establishes the desired L^p bound $\|\mathcal{M}f\|_{L^p(\mathcal{R}_k)} \leq C_{p,k}\|f\|_{L^p(\mathcal{R}_k)}$. □

Combining the L^p estimate we have just proved with the majorization inequality (7.6), we immediately obtain (7.2), Doob's inequality for the generalized coherent random variables. It remains to prove the optimality of the constant $C_{p,k}$ in the latter estimate. Having proved this sharpness, we immediately deduce the optimality of the constant for the uncentered maximal operator.

Proof of sharpness of $C_{p,k}$. Let $1 < p < \infty$ and $k \in \{1, 2, \ldots\}$ be fixed. Consider the probability space $\mathcal{R}_k$ with its Borel σ-algebra and normalized one-dimensional Lebesgue's measure λ_k. Fix an auxiliary constant $r \in (0, p^{-1})$ and consider the random variable $\xi(x) = |x|^{-r}$: then the estimate $r < p^{-1}$ guarantees that this variable belongs to L^p.

To define the filtrations, let $\lambda_{r,k}$ be the unique root of the equation

$$\lambda(1-r) - (k-1)r\lambda^{(r-1)/r} - 1 = 0, \qquad 1 \leq \lambda < \infty. \tag{7.12}$$

The existence and uniqueness of $\lambda_{r,k}$ is direct consequence of the fact that the left-hand side, considered as a function of λ, is strictly increasing, negative at $\lambda = 1$ and positive for large λ. Now, for any $j \in \{1, 2, \ldots, k\}$, introduce the closed ball B_j which has the center $e^{2\pi ij/k}(1 - \lambda_{r,k}^{-1/r})/2$ and radius $(1 + \lambda_{r,k}^{-1/r})/2$. This ball covers the whole ray H_j and some portion of the remaining rays: $|B_j \cap H_i| = \lambda_{r,k}^{-1/r}$ for $i \neq j$. Therefore if x lies on the j-th ray of $\mathcal{R}_k$, then the rescaled ball

$$|x|B_j = \{|x|y \in \mathcal{R}_k : \ y \in B_j\}$$

satisfies

$$\frac{1}{\lambda_k(|x|B_j)} \int_{|x|B_j} \xi \mathrm{d}\lambda_k = \frac{\int_0^{|x|} \omega^{-r}\mathrm{d}\omega + (k-1)\int_0^{\lambda_{r,k}^{-1/r}|x|} \omega^{-r}\mathrm{d}\omega}{|x| + (k-1)\lambda_{r,k}^{-1/r}|x|} = \lambda_{r,k} \cdot \xi(x),$$

by (7.12). Since both sides are homogeneous of order $-r$ (as a function of x), one can actually show a bit more: for any $\varepsilon > 0$ there is $\delta \in (0,1)$ such that if $y \in H_j$ satisfies $\delta < |y/x| \leq 1$, then

$$\frac{1}{\lambda_k(|x|B_j)} \int_{|x|B_j} \xi \mathrm{d}\lambda_k \geq (\lambda_{r,k} - \varepsilon) \cdot \xi(y). \tag{7.13}$$

Fix ε, δ with the above property and pick a large integer N. For any $n = 0, 1, \ldots, N$, let $\mathcal{G}_n^j$ be the σ-algebra generated by the balls $B_j, \delta B_j, \delta^2 B_j, \ldots, \delta^{n-1}B_j$. It follows directly from (7.13) that

$$M_{\mathcal{G}}\xi \geq (\lambda_{r,k} - \varepsilon)\xi \qquad \text{almost surely on } \mathcal{R}_k \setminus \delta^N B_j.$$

But $\xi \in L^p$, as we have already discussed above. Since ε and N were taken arbitrarily, the best constant allowed in the estimate (7.2) is at least $\lambda_{r,k}$. It remains to note that if we let $r \to p^{-1}$, then $\lambda_{r,k}$ converges to the constant $C_{p,k}$: in the limit, the equation (7.12) becomes (7.3). This proves the desired sharpness. □

7.3. Proof of Theorem 7.2

Lastly, we prove the sharp version of Doob's estimate in the classical coherent setting.

Proof of Theorem 7.2. Put $\mathbb{P}(\xi = 1) = q$. Then for $t \in [0,1]$ we have the identity $\xi^*(e^{2\pi ij/k}t) = \mathbb{1}(t \le q)$ and therefore

$$\mathcal{M}\xi^*(e^{2\pi ij/k}t) = \begin{cases} 1 & \text{if } t \le q, \\ \dfrac{kq}{(k-1)q+t} & \text{if } t > q, \end{cases}$$

for all $j = 1, 2, \ldots, k$. By Theorem 7.4, we can write

$$\begin{aligned} \frac{\|M_{\mathcal{G}}\xi\|_p^p}{\|\xi\|_p^p} \le \frac{\|\mathcal{M}\xi^*\|_p^p}{\|\xi\|_p^p} &= \left[q + \int_q^1 \left(\frac{kq}{(k-1)q+t}\right)^p \mathrm{d}t\right]\frac{1}{q} \\ &= 1 + \int_1^{1/q} \left(\frac{k}{k-1+s}\right)^p \mathrm{d}s \\ &\le 1 + \int_1^{\infty} \left(\frac{k}{k-1+s}\right)^p \mathrm{d}s = 1 + \frac{k}{p-1}, \end{aligned}$$

which gives the desired bound.

To see that the estimate is sharp, we construct a simple example for which all the inequalities above become almost-equalities. Precisely, consider the probability space $\mathcal{R}_k$ with its Borel subsets and normalized one-dimensional Lebesgue's measure λ_k and fix an arbitrary $\varepsilon > 0$. Introduce the random variable $\xi(x) = \mathbb{1}(|x| < q)$, where $q \in (0,1)$ satisfies

$$\int_1^{1/q} \left(\frac{k}{k-1+s}\right)^p \mathrm{d}s + \varepsilon = \int_1^{\infty} \left(\frac{k}{k-1+s}\right)^p \mathrm{d}s.$$

For fixed $1 \le j \le k$ and $0 \le n \le N$, distinguish the point $x_n = (N-n)/(2N)$ and let B_n^j be the ball centered at $e^{2\pi ij/k}x_n$ and of radius $x_n + q$.

Finally, consider the filtration $(\mathcal{G}_n^j)_{0 \le n \le N} = (\sigma(B_0^j, B_1^j, B_2^j, \ldots, B_n^j))_{0 \le n \le N}$. Arguing as above, one easily checks that the maximal function $M_{\mathcal{G}}\xi$ can be made arbitrarily close, in L^∞ norm, to $\mathcal{M}\xi^*$, by picking N sufficiently large. Thus we can guarantee that

$$\|M_{\mathcal{G}}\xi\|_p^p/\|\xi\|_p^p + \varepsilon > \|\mathcal{M}\xi^*\|_p^p/\|\xi\|_p^p,$$

and hence we obtain

$$\frac{\|M_{\mathcal{G}}\xi\|_p^p}{\|\xi\|_p^p} > 1 + \frac{k}{p-1} - 2\varepsilon.$$

Since ε was chosen arbitrarily, the sharpness follows. □

Bibliography

[1] I. Arieli and Y. Babichenko, *Private Bayesian persuasion*, Journal of Economic Theory, 182 (2019), pp. 185–217.

[2] ——, *A population's feasible posterior beliefs*, Journal of Economic Theory, 215 (2024). Paper No. 105764.

[3] I. Arieli, Y. Babichenko, and F. Sandomirskiy, *Persuasion as transportation*, arXiv:2307.07672, Preprint (2023), pp. 1–42.

[4] I. Arieli, Y. Babichenko, F. Sandomirskiy, and O. Tamuz, *Feasible joint posterior beliefs*, Journal of Political Economy, 129 (2021), pp. 2546–2594.

[5] M. A. Armstrong, *Basic topology*, Springer, 1983.

[6] P. Billingsley, *Probability and measure – anniversary edition*, Wiley, 2012.

[7] S. Boza, M. Křepela, and J. Soria, *Lorentz and Gale–Ryser theorems on general measure spaces*, Proceedings of the Royal Society of Edinburgh, Section A: Mathematics, 152 (2022), pp. 857–878.

[8] R. A. Brualdi, *Combinatorial matrix classes*, Cambridge University Press, 2006.

[9] K. Burdzy and S. Pal, *Can coherent predictions be contradictory?*, Advances in Applied Probability, 53 (2021), pp. 133–161.

[10] K. Burdzy and J. Pitman, *Bounds on the probability of radically different opinions*, Electronic Communications in Probability, 25 (2020), pp. 1–12.

[11] J. Buzzi, *Chaos and ergodic theory*, Springer, 2009, pp. 953–978. Encyclopedia of complexity and systems science.

[12] R. Cairoli and J. B. Walsh, *Stochastic integrals in the plane*, Acta Mathematica, 134 (1975), pp. 111–183.

[13] Y. Caro, J. Lauri, and C. Zarb, *Notes on spreads of degrees in graphs*, Bulletin of the Institute of Combinatorics and its Applications, 85 (2018), pp. 79–91.

[14] T. C. Christofides and R. J. Serfling, *Maximal inequalities for multidimensionally indexed submartingale arrays*, The Annals of Probability, 18 (1990), pp. 630–641.

[15] S. Cichomski, *Maximal spread of coherent distributions: a geometric and combinatorial perspective*, Master Thesis, University of Warsaw, 2020.

[16] S. Cichomski and A. Osękowski, *The maximal difference among expert's opinions*, Electronic Journal of Probability, 26 (2021), pp. 1–17.

[17] ——, *Doob's estimate for coherent random variables and maximal operators on trees*, Probability and Mathematical Statistics, 43 (2023), pp. 109–119.

[18] ——, *Coherent distributions on the square – extreme points and asymptotics*, Journal of Applied Probability, Online first (2024), pp. 1–23.

[19] ——, *Contradictory predictions with multiple agents*, ALEA: Latin American Journal of Probability and Mathematical Statistics, 21 (2024), pp. 369–383.

[20] ——, *On the existence of extreme coherent distributions with no atoms*, arXiv:2311.08140, Preprint (2024), pp. 1–12.

[21] S. Cichomski and F. Petrov, *A combinatorial proof of the Burdzy–Pitman conjecture*, Electronic Communications in Probability, 28 (2023), pp. 1–7.

[22] A. P. Dawid, *The well-calibrated Bayesian*, Journal of the American Statistical Association, 77 (1982), pp. 605–610.

[23] A. P. Dawid, M. H. DeGroot, and J. Mortera, *Coherent combination of experts' opinions*, Test, 4 (1995), pp. 263–313.

[24] M. H. DeGroot, *A Bayesian view of assessing uncertainty and comparing expert opinion*, Journal of Statistical Planning and Inference, 20 (1988), pp. 295–306.

[25] M. H. DeGroot and J. Mortera, *Optimal linear opinion pools*, Management Science, 37 (1991), pp. 546–558.

[26] P. Diaconis, S. Holmes, and S. Janson, *Threshold graph limits and random threshold graphs*, Internet Mathematics, 5 (2008), pp. 267–320.

[27] L. E. Dubins, D. Gilat, and I. Meilijson, *On the expected diameter of an L_2-bounded martingale*, The Annals of Probability, 37 (2009), pp. 393–402.

[28] L. E. Dubins and J. Pitman, *A divergent, two-parameter, bounded martingale*, Proceedings of the American Mathematical Society, 78 (1980), pp. 414–416.

[29] ——, *A maximal inequality for skew fields*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 52 (1980), pp. 219–227.

[30] L. E. Dubins and G. Schwarz, *On extremal martingale distributions*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 2 (1967), pp. 295–299.

[31] J. Dubra and F. Echenique, *Information is not about measurability*, Mathematical Social Sciences, 47 (2004), pp. 177–185.

[32] P. Erdös and T. Gallai, *Graphs with prescribed degrees of vertices*, Matematikai Lapok, 11 (1960), pp. 264–274.

[33] P. Erdös, G. Chen, C. C. Rousseau, and R. H. Schelp, *Ramsey problems involving degrees in edge-colored complete graphs of vertices belonging to monochromatic subgraphs*, European Journal of Combinatorics, 14 (1993), pp. 183–189.

[34] B. S. Everitt, *The analysis of contingency tables*, CRC Press, 1992.

[35] L. R. Ford and D. R. Fulkerson, *Flows in networks*, Princeton University Press, 1962.

[36] N. Gaffke and L. Rüschendorf, *On the existence of probability measures with given marginals*, Statistics & Risk Modeling, 2 (1984), pp. 163–174.

[37] T. Gneiting and R. Ranjan, *Combining predictive distributions*, Electronic Journal of Statistics, 7 (2013), pp. 1747–1782.

[38] L. Grafakos and J. Kinnunen, *Sharp inequalities for maximal functions associated with general measures*, Proceedings of the Royal Society of Edinburgh, Section A: Mathematics, 128 (1998), pp. 717–723.

[39] L. Grafakos and S. J. Montgomery-Smith, *Best constants for uncentred maximal functions*, Bulletin of the London Mathematical Society, 29 (1997), pp. 60–64.

[40] S. Gutmann, J. H. B. Kemperman, J. A. Reeds, and L. A. Shepp, *Existence of probability measures with given marginals*, The Annals of Probability, 19 (1991), pp. 1781–1797.

[41] J. C. Harsanyi, *Games with incomplete information played by "Bayesian" players, i-iii, part i, the basic model*, Management Science, 14 (1967), pp. 159–182.

[42] T. He and Y. Hou, *Some inequalities for tree martingales*, Acta Mathematicae Applicatae Sinica, 21 (2005), pp. 671–682.

[43] P. J. J. Herings, D. Karos, and T. T. Kerman, *Belief inducibility and informativeness*, Theory and Decision, Online first (2024), pp. 1–37.

[44] G. T. Herman and A. Kuba, *Discrete tomography: foundations, algorithms, and applications*, Springer, 1999.

[45] C. Hervés-Beloso and P. K. Monteiro, *Information and σ-algebras*, Economic Theory, 54 (2013), pp. 405–418.

[46] K. Hestir and S. C. Williams, *Supports of doubly stochastic measures*, Bernoulli, 1 (1995), pp. 217–243.

[47] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, 58 (1963), pp. 13–30.

[48] C. T. Ireland and S. Kullback, *Contingency tables with given marginals*, Biometrika, 55 (1968), pp. 179–188.

[49] B. G. Ivanoff and E. Merzbach, *Set–indexed martingales*, Chapman and Hall, 1999.

[50] E. Kamenica and M. Gentzkow, *Bayesian persuasion*, American Economic Review, 101 (2011), pp. 2590–2615.

[51] J. L. Kelley, *General topology*, van Nostrand, 1955.
[52] E. Kravchenko, *Coherent distributions: Hilbert space approach and duality*, arXiv:2405.04375, Preprint (2024), pp. 1–30.
[53] M. Krein and D. Milman, *On extreme points of regular convex sets*, Studia Mathematica, 9 (1940), pp. 133–138.
[54] X. Lang and J. Li, *Feasible joint posterior beliefs with binary signals*, Economics Letters, 227 (2023). Paper No. 111126.
[55] J. Lee, *Sampling contingency tables given sets of marginals and/or conditionals in the context of statistical disclosure limitation*, Doctoral Thesis, The Pennsylvania State University, 2009.
[56] J. A. Loera and E. D. Kim, *Combinatorics and geometry of transportation polytopes: an update*, Discrete Geometry and Algebraic Combinatorics, 625 (2013), pp. 37–76.
[57] G. G. Lorentz, *A problem of plane measure*, American Journal of Mathematics, 71 (1949), pp. 417–426.
[58] L. Lovász, *Large networks and graph limits*, American Mathematical Society, 2012.
[59] N. V. R. Mahadev and U. N. Peled, *Threshold graphs and related topics*, Elsevier, 1995.
[60] A. H. Murphy and R. L. Winkler, *A general framework for forecast verification*, Monthly Weather Review, 115 (1987), pp. 1330–1338.
[61] A. Osękowski, *Sharp martingale and semimartingale inequalities*, Birkhäuser, 2012.
[62] A. Osękowski, *Estimates for the diameter of a martingale*, Stochastics, 87 (2015), pp. 235–256.
[63] I. Pinelis, V. H. Peña, R. Ibragimov, A. Osękowski, and I. Shevtsova, *Inequalities and extremal problems in probability and statistics: selected topics*, Academic Press, 2017.
[64] R. Ranjan and T. Gneiting, *Combining probability forecasts*, Journal of the Royal Statistical Society, Series B: Statistical Methodology, 72 (2010), pp. 71–91.
[65] W. Rudin, *Functional analysis*, McGraw-Hill, 1991.
[66] H. J. Ryser, *Combinatorial properties of matrices of zeros and ones*, Canadian Journal of Mathematics, 9 (1957), pp. 371–377.
[67] V. A. Satopää, R. Pemantle, and L. H. Ungar, *Modeling probability forecasts via information diversity*, Journal of the American Statistical Association, 111 (2016), pp. 1623–1633.
[68] G. Sierksma and H. Hoogeveen, *Seven criteria for integer sequences being graphic*, Journal of Graph Theory, 15 (1991), pp. 223–231.
[69] V. Strassen, *The existence of probability measures with given marginals*, The Annals of Mathematical Statistics, 36 (1965), pp. 423–439.
[70] T. Tao, *Szemerédi's regularity lemma revisited*, Contributions to Discrete Mathematics, 1 (2006), pp. 8–28.
[71] M. Urbański, M. Roy, and S. Munday, *Ergodic theory – finite and infinite, thermodynamic formalism, symbolic dynamics and distance expanding maps*, De Gruyter, 2022. Volume 1.
[72] V. Vasyunin and A. Volberg, *The Bellman function technique in harmonic analysis*, Cambridge University Press, 2020.
[73] R. Vershynin, *High-dimensional probability: an introduction with applications in data science*, Cambridge University Press, 2018.
[74] E. Wong and M. Zakai, *Weak martingales and stochastic integrals in the plane*, The Annals of Probability, 4 (1976), pp. 570–586.
[75] T. Zhu, *Some problems on the convex geometry of probability measures*, Doctoral Thesis, University of California, Berkeley, 2022.