

Contradictory predictions with multiple agents

Stanisław Cichomski and Adam Osękowski

Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

E-mail address: S.Cichomski@uw.edu.pl

Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

E-mail address: A.Osekowski@mimuw.edu.pl

Abstract. Let X_1, X_2, \dots, X_n be a sequence of coherent random variables, i.e., satisfying the equalities

$$X_j = \mathbb{P}(A|\mathcal{G}_j), \quad j = 1, 2, \dots, n,$$

almost surely for some event A . This paper contains the proof of the estimate

$$\mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) \leq \frac{n(1-\delta)}{2-\delta} \wedge 1,$$

where $\delta \in (\frac{1}{2}, 1]$ is a given parameter. The inequality is sharp: for any δ , the constant on the right cannot be replaced by any smaller number. The argument rests on several novel combinatorial and symmetrization arguments, combined with dynamic programming. Our result generalizes the two-variate inequality of K. Burdzy and S. Pal and in particular provides an alternative derivation.

1. Introduction

To begin, let us recall the following celebrated definition due to Dawid et al. (1995). Henceforth, let n be a positive integer.

Definition 1.1. Let μ be a probability measure on $[0, 1]^n$. We say that μ is *coherent*, if it is the joint distribution of an n -variate random vector (X_1, X_2, \dots, X_n) defined on some arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that

$$X_i = \mathbb{P}(A|\mathcal{G}_i) \quad \text{for all } i = 1, 2, \dots, n, \quad \text{almost surely,} \quad (1.1)$$

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

The class of all coherent distributions on $[0, 1]^n$ will be denoted as \mathcal{C}_n . For the sake of convenience (and with a slight abuse of notation), we will also write $(X_1, X_2, \dots, X_n) \in \mathcal{C}_n$ to indicate that

Received by the editors September 18th, 2023; accepted January 18th, 2024.

2010 *Mathematics Subject Classification.* 60E15.

Key words and phrases. conditional probability; joint distribution of conditional expectations; opinion; best constant.

the distribution of the random vector (X_1, X_2, \dots, X_n) is coherent. Finally, we will use one more helpful convention: in case a vector $(X_1, X_2, \dots, X_n) \in \mathcal{C}_n$ is given and fixed, the corresponding measurable event and the sequence of sub- σ -fields satisfying (1.1) will automatically be denoted with the same letters A and $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n$.

The above setup has a nice and transparent interpretation, which is important for many applications. Namely, suppose that a group of n experts provides their personal estimates on the likelihood of some random event A , and assume that the knowledge of j -th expert is represented by the σ -algebra \mathcal{G}_j , $j = 1, 2, \dots, n$. Of course, there are many possible mechanisms leading to more or less effective predictions; one of the simplest approaches is to use the conditional expectation - then the predictions X_1, X_2, \dots, X_n of the experts are precisely given by (1.1).

In general, there are three basic categories of problems which are studied in the above context, stemming from applications in statistics, decision theory, economics and game theory as well as probability and information theory.

- *Optimal combining* – depending on auxiliary distributional assumptions of the model, to find an optimal procedure that combines multiple coherent opinions in order to produce a better forecast; see Dawid et al. (1995); DeGroot (1988); Ranjan and Gneiting (2010); Satopää et al. (2016).

- *Bayesian persuasion* – to compute, given a specific payoff function, how much one of the agents can benefit by selectively revealing parts of his information to other players, thus changing their beliefs and reactions; see Arieli and Babichenko (2024); Arieli et al. (2022, 2021); He et al. (2021).

- *Maximal discrepancy* – to provide sharp bounds on the maximal possible spread of coherent opinions. For instance, for a fixed function $\Phi : [0, 1]^n \rightarrow \mathbb{R}_+$, evaluate

$$\sup \mathbb{E}\Phi(X_1, X_2, \dots, X_n), \quad (1.2)$$

where the supremum is taken over all probability models as described above; see Burdzy and Pal (2021); Burdzy and Pitman (2020); Cichomski and Osekowski (2021); Cichomski and Petrov (2023).

The contribution of this paper concerns the last category. Our motivation comes from the following foundational result of Burdzy and Pal (2021).

Theorem 1.2. *For any threshold $\delta \in (\frac{1}{2}, 1]$, we have*

$$\sup_{(X,Y) \in \mathcal{C}_2} \mathbb{P}(|X - Y| \geq \delta) = \frac{2(1 - \delta)}{2 - \delta}. \quad (1.3)$$

In the language of applications, Theorem 1.2 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 (1.3) is remarkably complex and rather difficult: an explicit optimizer is obtained by a series of consecutive reductions and simplifications. As pointed out in Burdzy and Pitman (2020), finding a simpler proof of this result would be highly desirable. Another natural and important question concerns the extension of the threshold bound (1.3) to $n > 2$ coherent opinions. Our main result in this paper is as follows, we use the notation $a \wedge b$ for the minimum of the numbers a and b .

Theorem 1.3. *For any threshold $\delta \in (\frac{1}{2}, 1]$ and every integer $n \geq 2$, we have*

$$\sup_{(X_1, X_2, \dots, X_n) \in \mathcal{C}_n} \mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) = \frac{n(1 - \delta)}{2 - \delta} \wedge 1. \quad (1.4)$$

Correspondingly, Theorem 1.3 expands the range of applications from two experts to multiple agents scenario. Quite unexpectedly (at least to the authors), the threshold bound (1.4) reveals an almost linear dependence between the examined quantities and the number of coherent random variables. The proof of (1.4) that we present below is completely independent from the reasoning in Burdzy and Pal (2021) and hence can be regarded as an alternative demonstration of (1.3). Moreover, our approach does not refer in any significant way to the particular choice of integer n .

The above statement should also be compared to its version concerning the maximal spread of expectations. For any pair (X, Y) of coherent random variables we have the sharp estimate

$$\mathbb{E}|X - Y| \leq \frac{1}{2}$$

(see e.g. [Burdzy and Pitman \(2020\)](#)). The paper by [Cichomski and Osekowski \(2021\)](#) contains the following extension of this result to the case of an arbitrary number of variables.

Theorem 1.4. *Under the above notation, we have*

$$\sup_{(X_1, X_2, \dots, X_n) \in \mathcal{C}_n} \mathbb{E} \max_{1 \leq i < j \leq n} |X_i - X_j| = \begin{cases} \frac{1}{2} & \text{if } n = 2, \\ 2 - \sqrt{2} & \text{if } n = 3, \\ \frac{7}{2} - 2\sqrt{2} & \text{if } n = 4, \\ \frac{n-2}{n-1} & \text{if } n \geq 5. \end{cases} \tag{1.5}$$

It might be a little unexpected that the above upper bound involves four different formulas depending on the value of n , while (1.4) is formulated with a single expression. We would also like to mention that the linearity of the expectation makes the proof of (1.5) a bit simpler; the successful treatment of the tail inequality will require a broader class of arguments.

Let us say a few words about our approach and the organization of the remaining part of the paper. In the next section we apply a special symmetrization technique, which reduces the problem of calculating the left-hand side of (1.4) to the combinatorial optimization over specific objects of geometrical nature. This approach seems to be especially beneficial due to certain convenient symmetry constraints it enforces. Then, in Section 3, using dynamic programming arguments, we solve the previously obtained optimization problem. This boils down to the derivation of a suitable Bellman function and some further reductions of the problem. This appearance of dynamic programming is not surprising: as evidenced in numerous papers, the Bellman function method is a powerful tool used widely in martingale theory and harmonic analysis to obtain sharp inequalities – see e.g. [Osekowski \(2012\)](#); [Pinelis et al. \(2017\)](#); [Vasyunin and Volberg \(2020\)](#) and consult the references therein. We strongly believe that our argumentation can be pushed further and successfully applied in the subsequent study of problem (1.2) for other multidimensional functionals. For example, the sharp L^p -extension of (1.5), $1 < p < \infty$, seems to be a natural and a challenging question. One could also study the weighted versions of (1.4) and (1.5), in which the expression $\max_{1 \leq i < j \leq n} |X_i - X_j|$ is replaced by $\max_{1 \leq i < j \leq n} \alpha_{ij} |X_i - X_j|$, for some given family $(\alpha_{ij})_{1 \leq i < j \leq n}$ of nonnegative coefficients.

2. Basic reductions and symmetrizations

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. Let $\mathcal{C}(n, m)$ be the family of all vectors $X = (X_1, X_2, \dots, X_n) \in \mathcal{C}_n$ such that each X_j takes at most m different values, $j = 1, 2, \dots, n$.

Proposition 2.1. *To prove the threshold bound (1.4), it is enough to verify that*

$$\sup_{\substack{m \in \{1, 2, \dots\} \\ X \in \mathcal{C}(n, m)}} \mathbb{P} \left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta \right) = \frac{n(1 - \delta)}{2 - \delta} \wedge 1. \tag{2.1}$$

Proof: Assume that (2.1) holds and fix any n -variate vector $X \in \mathcal{C}_n$. Let m be a positive integer with $\delta > \frac{2}{m} + \frac{1}{2}$. As shown in [Burdzy and Pal \(2021\)](#); [Cichomski \(2020\)](#), 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

$$\mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) \leq \mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i^{(m)} - X_j^{(m)}| \geq \delta - \frac{2}{m}\right) \leq \frac{n\left(1 - \left(\delta - \frac{2}{m}\right)\right)}{2 - \left(\delta - \frac{2}{m}\right)},$$

where the second inequality follows from (2.1). Taking $m \rightarrow \infty$ ends the proof. □

Later on, we will need the following structural fact.

Lemma 2.2. *Assume that $\{G_1, G_2, \dots, G_m\}$ is a finite partition of Ω , let $A \in \mathcal{F}$ be an arbitrary event and put $Y = \mathbb{E}(\mathbb{1}_A | \sigma(G_1, G_2, \dots, G_m))$. 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: This is straightforward. For any $G \in \sigma(G_1, G_2, \dots, G_m)$ 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 $y \cdot (\mathbb{P}(A \cap G) + \mathbb{P}(A^c \cap G)) = \mathbb{P}(A \cap G)$, or $\mathbb{P}(A^c \cap G) = \frac{1-y}{y} \cdot \mathbb{P}(A \cap G)$. It remains to take $G = \{Y = y\}$; we have $G \in \sigma(G_1, G_2, \dots, G_m)$, since Y is measurable with respect to the latter σ -algebra. □

Now we will describe a useful symmetrization procedure, which allow us to replace the left-hand side of (2.1) with a more regular expression (see Corollary 2.4 below). We need some additional notation. Fix a positive integer m and let $X \in \mathcal{C}(n, m)$ be a coherent vector with $X_i = \mathbb{E}(\mathbb{1}_A | \mathcal{G}_i)$, $i = 1, 2, \dots, n$. Let U be a random variable independent of $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n$ and A , having the two-point distribution $\mathbb{P}(U = 0) = \mathbb{P}(U = 1) = 1/2$. Then \tilde{X} , the mixture of vectors X and $1 - X$, is given by

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

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

Proposition 2.3. *With the above notation, the following holds true.*

- (i) *We have $\mathbb{P}(\tilde{A}) = \frac{1}{2}$, $\tilde{X} \in \mathcal{C}(n, 2m)$ and $\tilde{X}_i = \mathbb{E}(\mathbb{1}_{\tilde{A}} | \tilde{\mathcal{G}}_i)$ for all i .*
- (ii) *For any sequence $(x_i)_{i=1}^n \subset [0, 1]$ we have the identity*

$$\mathbb{P}\left(\bigcap_{i=1}^n \{\tilde{X}_i = x_i\} \cap \tilde{A}\right) = \mathbb{P}\left(\bigcap_{i=1}^n \{\tilde{X}_i = 1 - x_i\} \cap \tilde{A}^c\right).$$

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

$$\frac{1 - x}{x} \cdot \sum_{i=1}^n \mathbb{P}\left(\{\tilde{X}_i = x\} \cap \tilde{A}\right) = \sum_{i=1}^n \mathbb{P}\left(\{\tilde{X}_i = 1 - x\} \cap \tilde{A}\right).$$

- (iv) *We have the equality*

$$\mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) = 2 \cdot \mathbb{P}\left(\left\{\max_{1 \leq i < j \leq n} |\tilde{X}_i - \tilde{X}_j| \geq \delta\right\} \cap \tilde{A}\right).$$

Proof: Since U is measurable with respect to $\tilde{\mathcal{G}}_i$ and independent of A , we obtain

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

and

$$\mathbb{P}(\tilde{A}) = \mathbb{E}\left[\mathbb{E}(\mathbb{1}_{\tilde{A}}|\tilde{\mathcal{G}}_i)\right] = \mathbb{E}[UX_i + (1 - U)(1 - X_i)] = \frac{1}{2}$$

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

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

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

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

so (ii) is established. To prove the third part, fix $x \in (0, 1]$ and write

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

where the first equality is due to the Lemma 2.2 and the second is a consequence of (ii). Finally, fix $m \in \{1, 2, \dots\}$, $X \in \mathcal{C}(n, m)$ and notice that

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

almost surely. Hence we deduce (iv) from

$$\mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) = \mathbb{P}\left(\max_{1 \leq i < j \leq n} |\tilde{X}_i - \tilde{X}_j| \geq \delta\right) = 2 \cdot \mathbb{P}\left(\left\{\max_{1 \leq i < j \leq n} |\tilde{X}_i - \tilde{X}_j| \geq \delta\right\} \cap \tilde{A}\right),$$

as desired. □

As a direct consequence, we have the following crucial reduction.

Corollary 2.4. *We have the inequality*

$$\sup_{\substack{m \in \{1, 2, \dots\} \\ X \in \mathcal{C}(n, m)}} \mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) \leq 2 \cdot \sup_{\substack{m \in \{1, 2, \dots\} \\ X \in \mathcal{C}'(n, m)}} \mathbb{P}\left(\left\{\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right\} \cap A\right), \quad (2.2)$$

where $\mathcal{C}'(n, m)$ is the subset of all those $X \in \mathcal{C}(n, m)$ that satisfy $\mathbb{P}(A) = \frac{1}{2}$ and

$$\frac{1-x}{x} \cdot \sum_{i=1}^n \mathbb{P}\left(\{X_i = x\} \cap A\right) = \sum_{i=1}^n \mathbb{P}\left(\{X_i = 1-x\} \cap A\right) \quad \text{for all } x \in (0, 1]. \quad (2.3)$$

Proof: By Proposition 2.3 (iii) and (iv), the left-hand side of (2.2) does not exceed the right-hand side. □

It will later become clear that (2.2) is in fact an equality. As for now, the above argumentation allows us to reduce our main problem to the identification of

$$\sup_{\substack{m \in \{1, 2, \dots\} \\ X \in \mathcal{C}'(n, m)}} \mathbb{P} \left(\left\{ \max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta \right\} \cap A \right). \tag{2.4}$$

The advantage over the original formulation (appearing on the left-hand side of (2.2)) lies in the fact that we study the behavior of X restricted to the set A . As we will see, the analysis of this expression can be performed in a purely analytic setup, with the use of combinatorial arguments. Accordingly, consider the measure space $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda)$, where λ stands for the Lebesgue measure.

Definition 2.5. For $k \in \mathbb{N} \setminus \{0\}$, let $\Lambda(k)$ be the family of all those functions $(H, L) : \mathbb{R}_+ \rightarrow [0, 1]^2$, which satisfy the following four requirements:

- (A.1) $L(x) \leq \frac{1}{2} \leq H(x)$ for all $x \in \mathbb{R}_+$,
- (A.2) H and L are right-continuous step functions with a finite number of steps,
- (A.3) $\lambda(\{H > \frac{1}{2}\}) + \lambda(\{L < \frac{1}{2}\}) \leq \frac{k}{2}$,
- (A.4) for any $y \in (0, 1]$ we have

$$\frac{1-y}{y} \cdot \left(\lambda(\{H = y\}) + \lambda(\{L = y\}) \right) = \lambda(\{H = 1 - y\}) + \lambda(\{L = 1 - y\}).$$

Here is a key statement, which links the above probabilistic considerations with the analytic context we have just introduced.

Proposition 2.6. *The value of (2.4) is not bigger than*

$$\sup_{(H, L) \in \Lambda(n)} \lambda(\{H \geq L + \delta\}). \tag{2.5}$$

Proof: Fix $m \in \{1, 2, \dots\}$, $X \in \mathcal{C}'(n, m)$ and the corresponding event A . We will construct $(H_X, L_X) \in \Lambda(n)$ such that

$$\mathbb{P} \left(\left\{ \max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta \right\} \cap A \right) = \lambda(\{H_X \geq L_X + \delta\}). \tag{2.6}$$

As $X \in \mathcal{C}'(n, m)$, there exists a natural number $l \leq m^n$ 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 $1 \leq k \leq l$, we set

$$p_k = \begin{cases} n\mathbb{P}(A_k) & \text{if } \max_{1 \leq i < j \leq n} |X_i - X_j| < \delta \text{ on } A_k, \\ (n-1)\mathbb{P}(A_k) & \text{if } \max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta \text{ on } A_k, \end{cases}$$

and introduce a disjoint partition $\mathcal{I} = \{I_k\}_{k=1}^{l+1}$ of \mathbb{R}_+ by

$$I_k = \begin{cases} [p_1 + \dots + p_{k-1}, p_1 + \dots + p_k) & \text{for } 1 \leq k \leq l, \\ [p_1 + \dots + p_l, \infty) & \text{for } k = l + 1. \end{cases}$$

Now we are ready to define (H_X, L_X) , setting its values on each element of \mathcal{I} separately. Assume that $k \in \{1, 2, \dots, l + 1\}$ and distinguish three major cases.

- If $k = l + 1$, we put $H_X(x) = L_X(x) = \frac{1}{2}$ for all $x \in I_{l+1}$.
- If $k \leq l$ and $p_k = n\mathbb{P}(A_k)$, we split I_k into n consecutive intervals $\{I_{k,s}\}_{s=1}^n$ (left-closed, right-open) of equal length and set

$$H_X(x) = \max \left(x_s^{(k)}, \frac{1}{2} \right), \quad L_X(x) = \min \left(x_s^{(k)}, \frac{1}{2} \right) \tag{2.7}$$

whenever $x \in I_{k,s}$ and $s = 1, 2, \dots, n$.

· Finally, suppose that $k \leq l$ and $p_k = (n - 1)\mathbb{P}(A_k)$. Then there are two indices $1 \leq i_1 \neq i_2 \leq n$ such that $X_{i_1} \geq X_{i_2} + \delta$ 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, $s \in \{1, 2, \dots, n\} \setminus \{i_2\}$, and put

$$H_X(x) = x_{i_1}^{(k)}, \quad L_X(x) = x_{i_2}^{(k)} \quad \text{if } x \in I_{k,i_1},$$

while for $x \in I_{k,s}$ and $s \in \{1, 2, \dots, n\} \setminus \{i_1, i_2\}$, we use (2.7). In other words, we proceed as in the previous case, but the intervals I_{k,i_1} and 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 $(\Lambda.1)$ - $(\Lambda.4)$. The first two conditions hold directly by the construction. To verify the point $(\Lambda.3)$, 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 $k \leq l$ and $p_k = n\mathbb{P}(A_k)$, then the restrictions of H and L to I_k are given by (2.7); 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

$$\lambda(\{H > 1/2\} \cap I_k) + \lambda(\{L < 1/2\} \cap I_k) \leq \lambda(I_k) = p_k = n\mathbb{P}(A_k).$$

Finally, if $k \leq l$ 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,

$$\lambda(\{H > 1/2\} \cap I_k) + \lambda(\{L < 1/2\} \cap I_k) \leq \lambda(I_{k,i_1}) + \lambda(I_k) = n\mathbb{P}(A_k).$$

Summing the above inequalities/equalities over k and noting that $\mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_l) = \mathbb{P}(A) = 1/2$, we obtain $(\Lambda.3)$. It remains to note that the last property is a direct consequence of (2.3). □

The next step is the following reduction.

Proposition 2.7. *The quantity (2.5) can be rewritten as*

$$\sup_{(H,L) \in \Lambda^\delta(n)} \lambda(\{H \geq L + \delta\}), \tag{2.8}$$

where $\Lambda^\delta(n)$ is the subset of all $(H, L) \in \Lambda(n)$ satisfying

$$\left\{H \in \left(\frac{1}{2}, \delta\right)\right\} \cup \left\{L \in \left(1 - \delta, \frac{1}{2}\right)\right\} = \emptyset \tag{2.9}$$

and

$$\{H \geq L + \delta\} = \left\{L < \frac{1}{2}\right\}. \tag{2.10}$$

Proof: Fix $(H, L) \in \Lambda(n)$ and assume that the condition (2.9) or (2.10) is not satisfied. If (2.9) fails, we modify H and/or L on the “bad” sets, changing their values to $\frac{1}{2}$ there. After this modification, the points $(\Lambda.1)$ - $(\Lambda.4)$ are still satisfied and the value of $\lambda(\{H \geq L + \delta\})$ remains unchanged. Now suppose that (2.10) does not hold. Because of the trivial inclusion $\{H \geq L + \delta\} \subseteq \{L < 1/2\}$ and the equality $\{L < 1/2\} = \{L \leq 1 - \delta\}$ we have just guaranteed, there must exist $0 \leq a < b$ and $0 < \gamma \leq 1 - \delta$ such that $L = \gamma$ and $H < \gamma + \delta$ on $[a, b)$. By the point $(\Lambda.4)$, we can find pairwise disjoint intervals $[a_j, b_j)$, $j = 1, \dots, m$, 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 (2.10), by performing sufficiently many such transformations. \square

The central part of the proof is the following estimate.

Lemma 2.8. *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 $k \in \{1, 2, \dots\}$ and all $(H, L) \in \Lambda^\delta(k)$ satisfying $\lambda(\{H > 1/2\}) > 0$.

We postpone the proof of this lemma to the next section, and proceed with our main result.

Proof of Theorem 1.3: By Propositions 2.1, 2.3 (iv), 2.6 and 2.7, we can write

$$\begin{aligned} \sup_{(X_1, X_2, \dots, X_n) \in \mathcal{C}_n} \mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) &\leq 2 \cdot \sup_{(H, L) \in \Lambda(n)} \lambda(\{H \geq L + \delta\}) \\ &= 2 \cdot \sup_{(H, L) \in \Lambda^\delta(n)} \lambda\left(\left\{L < \frac{1}{2}\right\}\right). \end{aligned}$$

Fix $(H, L) \in \Lambda^\delta(n)$. By Lemma 2.8, we have $\lambda(\{L < 1/2\}) \leq (1 - \delta)\lambda(\{H > 1/2\})$, while the point (A.3) gives $\lambda(\{L < 1/2\}) + \lambda(\{H > 1/2\}) \leq n/2$. Combining these two estimates, we immediately obtain

$$\mathbb{P}\left(\max_{1 \leq i < j \leq n} |X_i - X_j| \geq \delta\right) \leq \frac{n(1 - \delta)}{2 - \delta}.$$

It remains to prove the sharpness of (1.4). Observe that the function $\delta \mapsto (1 - \delta)/(2 - \delta)$ 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 $\{A_0, A_1, \dots, A_n\} \cup \{B_0, B_1, \dots, B_n\}$ 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

$$\mathcal{F}_i = \sigma\left(A_i, B_i, (A \cup B_{i+1}) \setminus A_{i+1}, (B \cup A_{i+1}) \setminus B_{i+1}\right), \quad i = 1, 2, \dots, n$$

(with the cyclic convention $A_{n+1} = A_1$, $B_{n+1} = B_1$). It is straightforward to check that the variables $Z_i = \mathbb{E}(\mathbb{1}_A | \mathcal{F}_i)$, $i = 1, 2, \dots, n$, 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 $\max_{1 \leq i < j \leq n} |Z_i - Z_j| \geq \delta$ on each A_k and each B_k ; this proves the estimate

$$\mathbb{P}\left(\max_{1 \leq i < j \leq n} |Z_i - Z_j| \geq \delta\right) \geq \mathbb{P}(\Omega \setminus (A_0 \cup B_0)) = \frac{n(1 - \delta)}{2 - \delta},$$

which is the desired lower bound. \square

3. Proof of Lemma 2.8

For the sake of clarity, let us first describe the rough idea behind the proof. We have the identity

$$\phi = \sup \frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}\}) - \lambda(\{L < \frac{1}{2}\})} = \sup \frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}, L = \frac{1}{2}\})},$$

where both suprema are as in the statement of Lemma 2.8. Our approach rests on expressing the set $\{H > \frac{1}{2}, L = \frac{1}{2}\}$ as the finite union of appropriate pairwise disjoint intervals $I_1^{(1)}, I_2^{(1)}, \dots, I_{k_1}^{(1)}$, and splitting $\{L < \frac{1}{2}\}$ into the union of k_1 families $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_{k_1}$, each of which consists of pairwise disjoint intervals. Then we may write

$$\frac{\lambda(\{L < \frac{1}{2}\})}{\lambda(\{H > \frac{1}{2}, L = \frac{1}{2}\})} = \frac{\sum_{j=1}^{k_1} \sum_{J \in \mathcal{A}_j} \lambda(J)}{\sum_{j=1}^{k_1} \lambda(I_j^{(1)})}. \tag{3.1}$$

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 (3.1) 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.

3.1. *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 $E_G \subset V_G \times V_G$ 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 (A.4), we have $\lambda(\{L < 1/2\}) < \lambda(\{H > 1/2\})$ and hence $\lambda(\{H > \frac{1}{2}, L = \frac{1}{2}\}) > 0$. Therefore, we can find a finite family $\mathcal{V}_1 = \{I_1^{(1)}, I_2^{(1)}, \dots, I_{k_1}^{(1)}\}$ of disjoint intervals, such that

$$\bigcup_{j=1}^{k_1} I_j^{(1)} = \left\{ H > \frac{1}{2}, L = \frac{1}{2} \right\}$$

and such that H is constant on each interval, say, $H = x_j^{(1)}$ on $I_j^{(1)}$ for $1 \leq j \leq k_1$. Set $\mathcal{E}_1 = \emptyset$.

- (2) Induction step. Suppose that we have successfully constructed \mathcal{V}_j and \mathcal{E}_j for $j \leq i - 1$. Moreover, assume that $\mathcal{V}_{i-1} = \{I_1^{(i-1)}, I_2^{(i-1)}, \dots, I_{k_{i-1}}^{(i-1)}\}$ and $H = x_j^{(i-1)}$ on $I_j^{(i-1)}$ for $1 \leq$

$j \leq k_{i-1}$. By the point (A.4), there exists a finite family $\bigcup_{j=1}^{k_{i-1}} \{J_1^j, J_2^j, \dots, J_{m_j}^j\}$ of disjoint intervals, such that

$$\bigcup_{l=1}^{m_j} J_l^j \subset \{L = 1 - x_j^{(i-1)}\} \setminus \bigcup_{n=1}^{i-1} \mathcal{V}_n, \quad \sum_{l=1}^{m_j} \lambda(J_l^j) = \frac{1 - x_j^{(i-1)}}{x_j^{(i-1)}} \cdot \lambda(I_j^{(i-1)})$$

for $j = 1, 2, \dots, k_{i-1}$, and such that H is constant on each J_l^j . Set

$$\mathcal{V}_i = \bigcup_{j=1}^{k_{i-1}} \{J_1^j, J_2^j, \dots, J_{m_j}^j\} \quad \text{and} \quad \mathcal{E}_i = \mathcal{E}_{i-1} \cup \bigcup_{j=1}^{k_{i-1}} \{I_j^{(i-1)}\} \times \{J_1^j, J_2^j, \dots, J_{m_j}^j\},$$

and put $\mathcal{V}_{(H,L)} = \bigcup_{i=1}^{\infty} \mathcal{V}_i$, $\mathcal{E}_{(H,L)} = \bigcup_{i=1}^{\infty} \mathcal{E}_i$.

To gain more intuition about the above construction, let us carry out an explicit calculation.

Example 3.1. Let $\delta = 0.7$ and consider a pair (H, L) given by

$$H = \chi_{[1,3)} + \frac{7}{8}(\chi_{[0,1)} + \chi_{[3,5)} + \chi_{[8,12)}) + \frac{3}{4}(\chi_{[5,8)} + \chi_{[12,15)}) + \frac{1}{2}\chi_{[15,\infty)},$$

$$L = \frac{1}{8}\chi_{[0,1)} + \frac{1}{4}\chi_{[1,3)} + \frac{1}{2}\chi_{[3,\infty)}.$$

It is not difficult to check that $(H, L) \in \Lambda^\delta(36)$. Let us now explain the construction of the forest

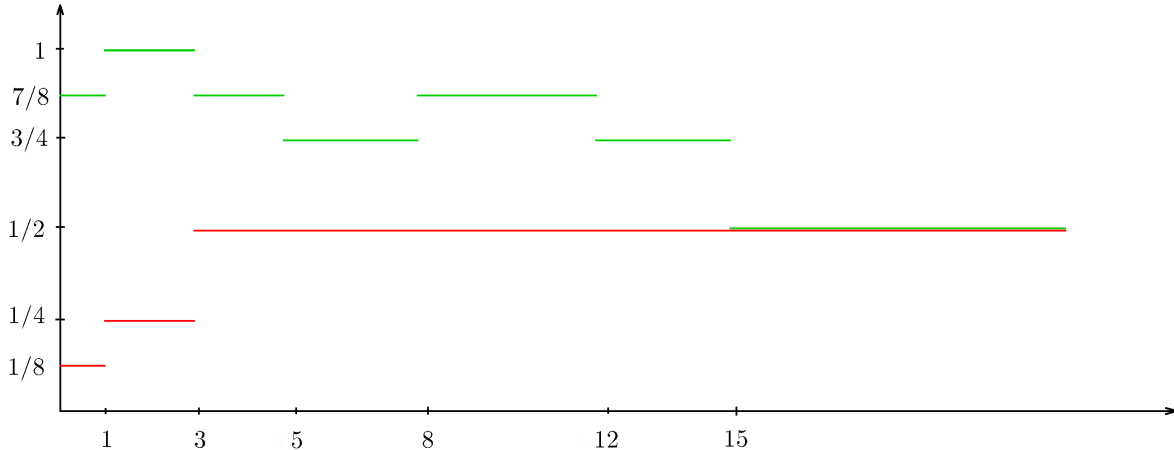


FIGURE 3.1. The graphs of the function H (green) and L (red).

$\mathcal{T}_{(H,L)}$. The starting point is to look at the set $\{H > 1/2, L = 1/2\} = [3, 15)$. In our case, this set splits into 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 (A.4) with $y = 3/4$ gives

$$\lambda(\{H = 3/4\}) = 3\lambda(\{L = 1/4\}), \tag{3.2}$$

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 $J_1^1, J_2^1, \dots, J_{m_1}^1$ of $\{L = 1/4\}$ for which the measure constraint (3.2) is preserved:

$$\lambda([5, 8)) = 3\lambda\left(\bigcup_{j=1}^{m_1} J_j^1\right),$$

and such that H is constant on each J_j^1 . There is a lot of ambiguity with the choice of J^j 's, we may actually take a single child $J_1^1 = [1, 2)$. We carry out 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:

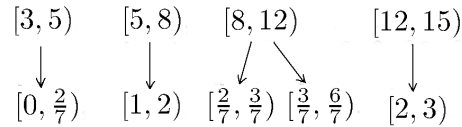


FIGURE 3.2. Partial forest: roots and their children.

Note that the intervals $[\frac{2}{7}, \frac{3}{7})$ and $[\frac{3}{7}, \frac{6}{7})$ could as well be merged into one $[\frac{2}{7}, \frac{6}{7})$: 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 (A.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).

3.2. *Properties of $\mathcal{T}_{(H,L)}$.* We start with the following important fact.

Proposition 3.2. *The family $\mathcal{V}_{(H,L)}$ is disjoint and*

$$\bigcup \mathcal{V}_{(H,L)} = \left\{ H > \frac{1}{2} \right\},$$

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

$$H > \frac{1}{2} \quad \text{for every } J \in \mathcal{V}_{(H,L)}. \tag{3.3}$$

Indeed, we obviously have $H > \frac{1}{2}$ on $\bigcup \mathcal{V}_1$. So, fix $i \in \{2, 3, \dots\}$ and assume that $H > \frac{1}{2}$ on $\bigcup_{n=1}^{i-1} \mathcal{V}_n$. 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 (3.3). 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

$$\left\{ H \in \left(\frac{1}{2}, \delta \right) \right\} = \left\{ L \in \left(1 - \delta, \frac{1}{2} \right) \right\} = \emptyset.$$

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

$$\frac{1-y}{y} \cdot \sum_{\{I \in \mathcal{V}_{(H,L)}: H \equiv y \text{ on } I\}} \lambda(I) = \sum_{\{J \in \mathcal{V}_{(H,L)}: L \equiv 1-y \text{ on } J\}} \lambda(J). \tag{3.4}$$

On the other hand, just by the property (A.4), we have

$$\frac{1-y}{y} \cdot \lambda(\{H = y\}) = \lambda(\{L = 1 - y\}). \tag{3.5}$$

Subtracting (3.4) from (3.5), we get

$$\frac{1-y}{y} \cdot \lambda(\{H = y\} \cap U) = \lambda(\{L = 1 - y\} \cap U), \tag{3.6}$$

for $y \in [\delta, 1]$. Next, by point (A.2), there exists a finite sequence $y_1, y_2, \dots, y_k \in [\delta, 1]$, satisfying

$$\lambda(\{H \notin \{y_1, y_2, \dots, y_k\}\} \cap U) = \lambda(\{L \notin \{1 - y_1, 1 - y_2, \dots, 1 - y_k\}\} \cap U) = 0.$$

Therefore, summing (3.6) for y_1, y_2, \dots, y_k , we obtain the inequality

$$\sum_{i=1}^k \lambda(\{L = 1 - y_i\} \cap U) = \sum_{i=1}^k \frac{1 - y_i}{y_i} \cdot \lambda(\{H = y_i\} \cap U) \leq \frac{1 - \delta}{\delta} \cdot \sum_{i=1}^k \lambda(\{H = y_i\} \cap U),$$

and hence

$$\lambda(\{L < \frac{1}{2}\} \cap U) < \lambda(\{H > \frac{1}{2}\} \cap U) \tag{3.7}$$

if only the right-hand side of (3.7) is positive. At the same time, we have

$$U \cap \{H > \frac{1}{2}, L = \frac{1}{2}\} = \emptyset,$$

since the set $\{H > \frac{1}{2}, L = \frac{1}{2}\}$ has been already covered by \mathcal{V}_1 . Consequently, we get

$$\{H > \frac{1}{2}\} \cap U = \{L < \frac{1}{2}\} \cap U = U$$

and thus

$$\lambda(\{L < \frac{1}{2}\} \cap U) = \lambda(\{H > \frac{1}{2}\} \cap U),$$

which contradicts (3.7). □

As we explained in the beginning of this section, the above graph structure can be exploited in the study of Lemma 2.8. Under the notation we have just introduced, the expression for ϕ can be rewritten in the form

$$\phi = \sup_{(H,L) \in \Lambda^\delta(\mathbb{N})} \frac{\sum_{J \in \mathcal{V}_{(H,L)} \setminus \mathcal{V}_1} \lambda(J)}{\sum_{I \in \mathcal{V}_1} \lambda(I)}. \tag{3.8}$$

We split the forest $\mathcal{T}_{(H,L)}$ into the disjoint trees: for $1 \leq j \leq k_1$, let $\mathcal{T}_{(H,L)}^j$ denote the directed tree with root $I_j^{(1)}$. Then we have

$$\frac{\sum_{J \in \mathcal{V}_{(H,L)} \setminus \mathcal{V}_1} \lambda(J)}{\sum_{I \in \mathcal{V}_1} \lambda(I)} = \frac{\sum_{j=1}^{k_1} [\lambda(\bigcup \mathcal{T}_{(H,L)}^j) - \lambda(I_j^{(1)})]}{\sum_{j=1}^{k_1} \lambda(I_j^{(1)})} \leq \max_{1 \leq j \leq k_1} \frac{\lambda(\bigcup \mathcal{T}_{(H,L)}^j) - \lambda(I_j^{(1)})}{\lambda(I_j^{(1)})}.$$

Hence, as we have already mentioned above, in the problem (3.8) it is enough to consider (H, L) with $\mathcal{T}_{(H,L)} = \mathcal{T}_{(H,L)}^1$, 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

$$\phi = \sup_{\Xi^\delta(\mathbb{N})} \left[\lambda(\bigcup \mathcal{V}_2) + \lambda(\bigcup \mathcal{V}_3) + \dots \right], \tag{3.9}$$

where supremum is taken over

$$\Xi^\delta(\mathbb{N}) := \left\{ (H, L) \in \Lambda^\delta(\mathbb{N}) : \mathcal{T}_{(H,L)} = \mathcal{T}_{(H,L)}^1 \text{ and } \lambda(I_1^{(1)}) = 1 \right\}.$$

Note that the series under supremum in (3.9) 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 (3.9) as

$$\phi = \lim_{m \rightarrow \infty} \left(\sup_{\Xi^\delta(\mathbb{N})} \sum_{j=2}^m \lambda\left(\bigcup \mathcal{V}_j\right) \right). \tag{3.10}$$

3.3. Dynamic programming. To compute the above supremum, it is convenient to apply dynamic programming techniques. Let $\Phi : [\delta, 1] \rightarrow \mathbb{R}_+$ be given by

$$\Phi(x) = \sup_{\mathbf{x}} \sum_{n=0}^\infty \prod_{i=0}^n \frac{1-x_i}{x_i}, \tag{3.11}$$

where the supremum is taken over all sequences $\mathbf{x} = (x_0, x_1, x_2, \dots)$ satisfying

$$x_0 = x, \quad x_n \in [\delta, 1] \quad \text{and} \quad x_{n+1} \geq 1 - x_n + \delta \quad \text{for } n = 0, 1, 2, \dots$$

We may call Φ the Bellman function associated with (3.10). Its connection to the problem is described in the following statement.

Proposition 3.3. *We have the identity*

$$\phi = \sup_{x \in [\delta, 1]} \Phi(x).$$

Proof: Fix $m \in \{1, 2, \dots\}$. Analogously to the reduction $\mathcal{T}_{(H,L)} = \mathcal{T}_{(H,L)}^1$, 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 $m \rightarrow \infty$, just as in (3.10), we get

$$\phi = \sup_{\Xi^\delta(\mathbb{N})} \sum_{j=2}^\infty \lambda(I_1^{(j)}), \tag{3.12}$$

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 (3.12) with an additional restriction to $x_0 = x$. This yields the claim. \square

We turn our attention to the identification of the formula for Φ . We start with a structural property of the Bellman function.

Proposition 3.4. *For any $x \in [\delta, 1]$ we have the recurrence relation*

$$\Phi(x) = \frac{1-x}{x} \left(1 + \sup_{y \geq 1-x+\delta} \Phi(y) \right). \tag{3.13}$$

Proof: The argument rests on the so-called optimality principle. By (3.11), we simply have

$$\Phi(x) = \frac{1-x}{x} \cdot \left(1 + \sup_{\tilde{\mathbf{x}}} \sum_{n=1}^\infty \prod_{i=1}^n \frac{1-x_i}{x_i} \right),$$

where supremum is taken over all sequences $\tilde{\mathbf{x}} = (x_1, x_2, \dots)$ such that

$$x_1 \geq 1 - x + \delta, \quad x_n \in [\delta, 1] \quad \text{and} \quad x_{n+1} \geq 1 - x_n + \delta \quad \text{for } n = 1, 2, 3, \dots \quad \square$$

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 (3.13), and exploiting this condition appropriately.

We proceed to the choice of \mathbf{x} . A little thought and a closer inspection suggests that problem (3.11) 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 constraints for the coordinates x_0, x_1, x_2, \dots . Plugging this sequence into (3.11), 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 2.8: First, we show that Ψ fulfills the recurrence (3.13). Indeed, for $x \in [\delta, 1]$, we have

$$\frac{1-x}{x} \left(1 + \sup_{y \geq 1-x+\delta} \Psi(y) \right) = \frac{1-x}{x} \left(1 + \Psi(1-x+\delta) \right) = \frac{1-x}{x} \left(1 + \frac{x-\delta}{\delta} \right) = \Psi(x).$$

Pick any $x \in [\delta, 1]$ and $\varepsilon > 0$. By (3.11), we can choose an admissible sequence $\mathbf{x}_\varepsilon = (x_0, x_1, \dots)$ (i.e., 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

$$\Phi(x) \leq 2\varepsilon + \sum_{n=0}^m \prod_{i=0}^n \frac{1-x_i}{x_i}. \quad (3.14)$$

On the other hand, by recurrence relation (3.13), we can write

$$\begin{aligned} \Psi(x) = \Psi(x_0) &\geq \frac{1-x_0}{x_0} \left(1 + \Psi(x_1) \right) = \frac{1-x_0}{x_0} + \frac{1-x_0}{x_0} \Psi(x_1) \\ &\geq \frac{1-x_0}{x_0} + \frac{1-x_0}{x_0} \frac{1-x_1}{x_1} \left(1 + \Psi(x_2) \right) \\ &= \frac{1-x_0}{x_0} + \frac{1-x_0}{x_0} \frac{1-x_1}{x_1} + \frac{1-x_0}{x_0} \frac{1-x_1}{x_1} \Psi(x_2) \end{aligned}$$

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 (3.14), we get $\Psi(x) + 2\varepsilon \geq \Phi(x)$, and since $\varepsilon > 0$ was chosen arbitrarily, the reverse bound $\Psi \geq \Phi$ follows. This proves the claim and completes the proof of (1.4): $\phi = \sup_{x \in [\delta, 1]} \Phi(x) = (1 - \delta)/\delta$. \square

Acknowledgments

The authors would like to thank an anonymous Referee for the very careful reading of the first version of the paper and many helpful suggestions. The second-named author was supported by University of Warsaw, Inicjatywa Doskonałości – Uczelnia Badawcza (IDUB), Nowe Idee 3B, grant no. 501-D110-20-3004310.

References

- Arieli, I. and Babichenko, Y. A population's feasible posterior beliefs. *J. Econom. Theory*, **215**, Paper No. 105764, 13 (2024). [MR4671090](#).
- Arieli, I., Babichenko, Y., and Sandomirskiy, F. Persuasion as transportation. *ArXiv Mathematics e-prints* (2022). [arXiv: 2307.07672](#).
- Arieli, I., Babichenko, Y., Sandomirskiy, F., and Tamuz, O. Feasible Joint Posterior Beliefs. *J. Pol. Econ.*, **129** (9), 2546–2594 (2021). [DOI: 10.1086/714993](#).
- Burdzy, K. and Pal, S. Can coherent predictions be contradictory? *Adv. in Appl. Probab.*, **53** (1), 133–161 (2021). [MR4232752](#).
- Burdzy, K. and Pitman, J. Bounds on the probability of radically different opinions. *Electron. Commun. Probab.*, **25**, Paper No. 14, 12 (2020). [MR4069734](#).
- Cichomski, S. *Maximal Spread of Coherent Distributions: a Geometric and Combinatorial Perspective*. Master's thesis, University of Warsaw (2020). Available at [arXiv: 2007.08022](#).
- Cichomski, S. and Osękowski, A. The maximal difference among expert's opinions. *Electron. J. Probab.*, **26**, Paper No. 105, 17 (2021). [MR4290507](#).
- Cichomski, S. and Petrov, F. A combinatorial proof of the Burdzy-Pitman conjecture. *Electron. Commun. Probab.*, **28**, Paper No. 3, 7 (2023). [MR4543968](#).
- Dawid, A. P., DeGroot, M. H., and Mortera, J. Coherent combination of experts' opinions. *Test*, **4** (2), 263–313 (1995). [MR1379793](#).
- DeGroot, M. H. A Bayesian view of assessing uncertainty and comparing expert opinion. *J. Statist. Plann. Inference*, **20** (3), 295–306 (1988). [MR976182](#).
- He, K., Sandomirskiy, F., and Tamuz, O. Private private information. *ArXiv Mathematics e-prints* (2021). [arXiv: 2112.14356](#).
- Osękowski, A. *Sharp martingale and semimartingale inequalities*, volume 72 of *Instytut Matematyczny Polskiej Akademii Nauk. Monografie Matematyczne (New Series)*. Birkhäuser/Springer Basel AG, Basel (2012). ISBN 978-3-0348-0369-4. [MR2964297](#).
- Pinelis, I., de la Peña, V. H., Ibragimov, R., Osękowski, A., and Shevtsova, I. *Inequalities and extremal problems in probability and statistics. Selected topics*. Academic Press, London (2017). ISBN 978-0-12-809818-9. [MR3727090](#).
- Ranjan, R. and Gneiting, T. Combining probability forecasts. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **72** (1), 71–91 (2010). [MR2751244](#).
- Satopää, V. A., Pemantle, R., and Ungar, L. H. Modeling probability forecasts via information diversity. *J. Amer. Statist. Assoc.*, **111** (516), 1623–1633 (2016). [MR3601722](#).
- Vasyunin, V. and Volberg, A. *The Bellman function technique in harmonic analysis*, volume 186 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (2020). ISBN 978-1-108-48689-7. [MR4411371](#).