Electron. Commun. Probab. **25** (2020), no. 14, 1–12.
https://doi.org/10.1214/20-ECP296
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Bounds on the probability of radically different opinions*

Krzysztof Burdzy† Jim Pitman‡

Abstract

We establish bounds on the probability that two different agents, who share an initial opinion expressed as a probability distribution on an abstract probability space, given two different sources of information, may come to radically different opinions regarding the conditional probability of the same event.

Keywords: conditional probability; opinion; maximal inequality; joint distribution of conditional expectations.
AMS MSC 2010: 60E15.
Submitted to ECP on March 19, 2019, final version accepted on January 31, 2020.
Supersedes arXiv:1903.07773.

1 Introduction

Let $A \in \mathcal{F}$ be an event in some probability space $(\Omega, \mathcal{F}, P)$, and let

$$X = P(A \,|\, \mathcal{G}) \qquad \text{and} \qquad Y = P(A \,|\, \mathcal{H}) \tag{1.1}$$

for two sub-σ-fields $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$. Equivalently, X and Y are random variables with

$$0 \le X, Y \le 1 \text{ and } X = P(A \,|\, X) \text{ and } Y = P(A \,|\, Y), \text{ hence } EX = EY = P(A) = p \tag{1.2}$$

for some $p \in [0, 1]$ and $A \in \mathcal{F}$. Following [7], we interpret X and Y as the opinions of two experts about the probability of A given different sources of information $\mathcal{G}$ and $\mathcal{H}$, assuming the experts agree on some initial assignment of probability P to events in $\mathcal{F}$.

There is a body of literature on related topics, some of them inspired by modern uses of technology. Consider N experts represented by sub-σ-fields who are all trying to predict the probability of a common event. A natural question is if there is a way to combine their predictions to come up with a better forecast. Introduced this way in the mid 80's onwards, see [16, 10, 7], such combinations typically take the form of weighted averages ([9]). The field has found a renewed interest in the current age of social networks (see [25, 15]). In particular, [31, 17] recommend both linear and non-linear combinations, [32] develops a mathematical framework to combine predictions when experts use "partially overlapping information sources", and [8] uses it for the case of $N = 2$ experts in prediction markets who take turn in updating their beliefs. Also see [27, 6, 20] for applications to economics, [23] for applications to banking and

*Research of KB was supported in part by Simons Foundation Grant 506732.
†Department of Mathematics, University of Washington, Seattle, WA 98195. E-mail: burdzy@uw.edu
‡Departments of Statistics and Mathematics, University of California, Berkeley, CA 94720.
E-mail: pitman@berkeley.edu

finance, [26] for applications to meteorology, [34] for applications to maintenance of wind turbines, and [14] for philosophical implications. The problem is also related to modeling insider trading in finance [21] where the insider has more information that the rest of the traders, i.e., $\mathcal{G} \subseteq \mathcal{H}$, although the general non-containment scenario makes sense for two different insiders.

We will use the term *coherent*, as in [7], for (X,Y) as in (1.1) or (1.2), or for the joint distribution of such (X,Y) on $[0,1]^2$. Note the obvious *reflection symmetry* that

$$\text{if } (X,Y) \text{ is coherent then so are } (Y,X), (1-X, 1-Y), \text{ and } (1-Y, 1-X). \tag{1.3}$$

Elementary examples in [7, §4.1] show that for any prescribed value of $EX = EY = P(A) \in (0,1)$, the correlation between coherent opinions X and Y about A can take any value in $(-1,1]$. Consider for instance, for $\delta \in (0,1)$, the distribution of (X,Y) concentrated on the three points $(1-\delta, 1-\delta)$ and $(0, 1-\delta)$ and $(1-\delta, 0)$, with

$$P(X=Y) = P(1-\delta, 1-\delta) = \frac{1-\delta}{1+\delta} \qquad \text{and} \qquad P(0,1-\delta) = P(1-\delta, 0) = \frac{\delta}{1+\delta}. \tag{1.4}$$

This example from [12] gives a pair of coherent opinions (X,Y) about the event $A = (X=Y)$, with correlation $\rho(X,Y) = -\delta$ which can be any value in $(-1,0)$.

The idea expressed above, that coherent opinions X and Y should not be too radically different, leads to the following precise problem, posed in [3, Sect. 14.4, p. 242] and [30]: for $0 \le \delta \le 1$, evaluate

$$\varepsilon(\delta) := \sup_{\text{coherent } (X,Y)} P(|X-Y| \ge 1-\delta) = \sup_{\text{coherent } (X,Y)} P(1-|X-Y| \le \delta). \tag{1.5}$$

For $m,n = 1,2,3,\ldots$ consider also $\varepsilon_{m\times n}(\delta) = \varepsilon_{n \times m}(\delta)$ defined by restricting the above supremum to $m \times n$ coherent (X,Y), meaning that X takes at most m and Y at most n possible values. Let $\varepsilon_{\text{finite}}(\delta) := \sup_{m,n} \varepsilon_{m\times n}(\delta)$, which is the supremum in (1.5) restricted to (X,Y) with a finite number of possible values. Each of these functions of δ is evidently non-decreasing and bounded above by 1. Then for all $\delta \in [0,1]$

$$\frac{2\delta}{1+\delta} \le \varepsilon_{2\times 2}(\delta) \le \varepsilon_{\text{finite}}(\delta) \le \varepsilon(\delta) \le \lim_{a \downarrow \delta} \varepsilon_{\text{finite}}(a). \tag{1.6}$$

The first inequality is due to the example (1.4). The second and third are obvious, and the last is by elementary construction of $n \times n$ coherent (X_n, Y_n) with $|X_n - X| + |Y_n - Y| \le 2/n$ for any coherent (X,Y) (see [5, Lemma 2.2]). We use the notation $x \wedge y := \min(x,y)$ and $x \vee y := \max(x,y)$, and either $\mathbb{1}_A$ or $\mathbb{1}(A)$ for an indicator function whose value is 1 if A and 0 else.

Proposition 1.1. *There are the following evaluations and bounds: for $\delta \in [0,1]$ and $n \ge 2$,*

$$\varepsilon_{1\times n}(\delta) = \delta \quad \text{if } \delta \in [0, \tfrac{1}{2}) \text{ and } 1 \text{ if } \delta \in [\tfrac{1}{2}, 1], \tag{1.7}$$

$$\varepsilon_{2\times 2}(\delta) = \frac{2\delta}{1+\delta} \text{ if } \delta \in [0, \tfrac{1}{2}) \text{ and } 1 \text{ if } \delta \in [\tfrac{1}{2}, 1], \tag{1.8}$$

$$\varepsilon_{2\times 2}(\delta) \le \varepsilon(\delta) \le (2\delta) \wedge 1. \tag{1.9}$$

The bounds (1.6) and (1.9) were given in [3, Theorem 14.1, p. 243], [30] and [4, Theorem 18.1, p. 389], while (1.7) and (1.8) are new. Our renewed interest in these results is prompted by

Theorem 1.2 ([5])**.** $\varepsilon_{2\times 2}(\delta) = \varepsilon_{\text{finite}}(\delta) = \varepsilon(\delta)$ *for all* $\delta \in [0,1]$.

To see that this identity holds with all values 1 for $\delta \in [\frac{1}{2}, 1]$, consider the coherent 1×2 distribution of (X, Y) with equal probability $\frac{1}{2}$ at the points $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, 1) \in [0, 1]^2$. That is

$$X = E(Y) = \tfrac{1}{2} \text{ for } Y = B_{1/2} \tag{1.10}$$

where B_p for $0 \le p \le 1$ denotes a random variable with the *Bernoulli*(p) *distribution*

$$P(B_p = 1) = p \text{ and } P(B_p = 0) = 1 - p. \tag{1.11}$$

For $\delta \in (0, \frac{1}{2})$, Theorem 1.2 is that equality holds in all the inequalities (1.6). The first of these equalities is proved here as (1.8). Equality in the second inequality of (1.6) for $\delta \in (0, \frac{1}{2})$ is much less obvious. The proof of this in [5] is quite long and difficult, by recursive reduction of m and n for $m \times n$ coherent (X, Y), until the problem is reduced to the 2×2 case treated here by (1.8). We hope this exposition of the easier evaluations in Proposition 1.1 might provoke someone to find a simpler proof of Theorem 1.2.

Note from (1.7), (1.8) and Theorem 1.2 that each of the functions $\varepsilon_{1\times n}(\delta)$ and $\varepsilon_{2\times 2}(\delta) = \varepsilon(\delta)$ is continuous on each of the intervals $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1]$, but has an upward jump to 1 at $\delta = \frac{1}{2}$.

The rest of this article is organized as follows. Section 2 recalls some background related to Proposition 1.1, which is proved in Section 3. Section 4 recalls some known characterizations of coherent distributions of (X, Y). For reasons we do not understand well, these general characterizations seem to be of little help in establishing the evaluations of $\varepsilon(\delta)$ discussed above, or in settling a number of related problems about coherent distributions, which we present in Section 5. So much is left to be understood about the limitations on coherent opinions.

2 Background

Let $(X_i, i \in I)$ be a finite collection of random variables defined on some common probability space $(\Omega, \mathcal{F}, P)$, and suppose that each X_i is the conditional expectation of some integrable random variable X_* given some sub-σ-field $\mathcal{F}_i$ of $\mathcal{F}$:

$$X_i = E(X_* \,|\, \mathcal{F}_i) \qquad (i \in I). \tag{2.1}$$

Doob's well known bounds for tail probabilities and moments of the distributions of $\max_{i\in I} X_i$ and $\max_{i\in I} |X_i|$, for either an increasing or decreasing family of σ-fields, and extensions of these inequalities to families of σ-fields indexed by a directed set I, with suitable conditional independence conditions, play a central role in the theory of martingale convergence. See for instance [22, 19] and [29] for recent refinements of Doob's inequalities, and further references. For the *diameter* of a martingale

$$\max_{i,j\in I} |X_i - X_j| = \big(\max_{i\in I} X_i\big) + \big(-\max_{i\in I}(-X_i)\big) \le 2 \max_{i\in I} |X_i| \tag{2.2}$$

there is no difficulty in bounding tail probabilities and moments, with an additional factor of 2 to a suitable power. But finer results with best constants for the diameter have also been obtained in [11, 28].

Much less is known about limitations on the distributions of such maximal variables for finite collections of σ-fields $(\mathcal{F}_i, i \in I)$ without conditions of nesting or conditional independence. We focus here on joint distributions of $X_i = E(X_* \,|\, \mathcal{F}_i)$ for X_* with $0 \le X_* \le 1$, and no restrictions except $\mathcal{F}_i \subseteq \mathcal{F}$ in a probability space $(\Omega, \mathcal{F}, P)$. Setting $X_J := E[X_* \,|\, \sigma(\cup_{i\in J}\mathcal{F}_i)]$ makes $((X_J, \mathcal{F}_J), J \subseteq I)$ a martingale indexed by subsets of J of I, with $(X_i, i \in I)$ the random vector of values of this martingale on singleton subsets of I. Assuming the basic probability space is sufficiently rich, there is a random variable U

with uniform distribution on $[0,1]$, with U independent of X_* and $\mathcal{F}_I$. Then X_* can be be replaced by the indicator random variable $\mathbb{1}(U \le X_*)$. So there is no loss of generality in supposing $X_* = \mathbb{1}(A)$ is the indicator of some event A with $P(A) = p \in [0,1]$. It follows that each X_i is the conditional probability of A given $\mathcal{F}_i$:

$$X_i = P(A \mid \mathcal{F}_i) \text{ implying } EX_i \equiv p := P(A) \qquad (i \in I). \tag{2.3}$$

Then either $(X_i, i \in I)$ or its joint distribution on $[0,1]^I$ will be called *coherent*. Besides $EX = EY$, another necessary condition for a pair (X,Y) to be coherent is provided by the following simplification and extension of [7, Theorem 5.2]. See also Proposition 4.1 for some conditions that are both necessary and sufficient for (X,Y) to be coherent.

Proposition 2.1. *Consider a pair of real-valued random variables (X,Y) and assume that there exist disjoint intervals G and H and Borel sets $G' \subseteq G$ and $H' \subseteq H$ such that the events $(X \in G')$ and $(Y \in H')$ are almost surely identical, with $P(X \in G') > 0$.*

(i) There is no integrable Z with $X = E(Z \mid X)$ and $Y = E(Z \mid Y)$.
(ii) If (X,Y) takes values in $[0,1]^2$ then (X,Y) is not coherent.
(iii) Suppose (X,Y) takes values in $[0,1]^2$. If $X - a$ and $Y - b$ are sure to be of opposite sign for some $0 \le a \le b \le 1$:

$$P((X-a)(Y-b) < 0) = 1, \tag{2.4}$$

and $P(Y > b) > 0$, then the distribution of (X,Y) is not coherent.

Proof. Suppose that $G' \subseteq G$ and $H' \subseteq H$. If $X = E(Z \mid X)$ and $Y = E(Z \mid Y)$ for some integrable Z then it is easily seen that

$$G \ni E(Z \mid X \in G') = E(Z \mid Y \in H') \in H, \tag{2.5}$$

where $E(Z \mid B)$ denotes $E(Z\mathbb{1}_B)/P(B)$ for any B with $P(B) > 0$. Since $G \cap H = \emptyset$, we obtain (i). Part (ii) follows from (i) and (1.2). Part (iii) follows by applying (ii) to $G' = G = [0,a)$ and $H' = H = (b,1]$. $\square$

Proposition 2.1 (iii) corrects the claim above [7, Theorem 5.2] that (2.4) alone makes (X,Y) not coherent. (This is false if $P(Y > b) = 0$; take $a = \frac{1}{4}, b = \frac{3}{4}$ and $X = Y = \frac{1}{2}$).

The following construction of a coherent distribution of n variables $(X_1, \dots, X_n)$ was used in [12] to build counterexamples in the theory of almost sure convergence of martingales relative to directed sets.

Example 2.2 (The (n,p)-daisy, with n petals and a Bernoulli(p) center [12])**.** Let $A, A_1, \dots, A_n$ be a measurable partition of Ω with

$$P(A) = p \text{ and } P(A_i) = \frac{1-p}{n} \text{ for } 1 \le i \le n.$$

For $1 \le i \le n$ let $\mathcal{F}_i$ be the σ-field generated by $A \cup A_i$. Then set

$$X_i := P(A \mid \mathcal{F}_i) = p_n \mathbb{1}(A \cup A_i) \text{ with } p_n := \frac{np}{np - p + 1}. \tag{2.6}$$

To explain the daisy mnemonic, imagine Ω is the union of $n+1$ parts of a daisy flower, with center A of area p, surrounded by n petals A_i of equal areas, with total petal area $1-p$. For each petal A_i, an ith *petal observer* learns whether or not a point picked at random from the daisy area has fallen in (the center A or their petal A_i), or in some other petal. Each petal observer's conditional probability X_i of A is then as in (2.6). The sequence of n variables $(X_1, \dots, X_n)$ is both coherent and exchangeable, with constant expectation p:

- given A the sequence $(X_1, \dots, X_n)$ is identically equal to the constant p_n;
- given the complement A^c, the sequence $(X_1, \dots, X_n)$ is p_n times an indicator sequence with a single 1 at a uniformly distributed index in $\{1, \dots, n\}$.

The (n, p)-daisy example was designed to make $\max_{1 \le i \le n} X_i = p_n$, a constant, as large as possible with $EX_i \equiv p$. As observed in [13, p. 224], this p_n is the largest possible essential infimum of values of $\max_i X_i$ for any coherent distribution of $(X_1, \dots, X_n)$ with $EX_i \equiv p$. This special property involves the n-petal daisy in the solution in various extremal problems for coherent opinions. For instance, $(X, Y) = (X_1, X_2)$ derived from the $(2, p)$ daisy with $p = (1 - \delta)/(1 + \delta)$, so $p_2 = 1 - \delta$, is the coherent pair in (1.4). This provides the lower bound for $\varepsilon_{2\times 2}(\delta)$ in (1.6), which according to (1.8) is attained with equality for $\delta \in [0, \frac{1}{2})$. Also:

Proposition 2.3. *(i)* [13] *For every coherent distribution of $(X_i, 1 \le i \le n)$ with $EX_i \equiv p$,*

$$E \max_{1 \le i \le n} X_i \le \frac{p(n-p)}{1 + p(n-2)}. \tag{2.7}$$

Moreover, this bound is attained by taking $(X_1, \dots, X_{n-1})$ to be the $(n-1, p)$-daisy sequence, and $X_n = \mathbb{1}_A$, the Bernoulli(p) indicator of the daisy center.

(ii) For every coherent distribution of (X, Y) on $[0, 1]^2$ with $EX = EY = p$,

$$E|X - Y| \le 2p(1-p) \le \tfrac{1}{2} \tag{2.8}$$

with equality in the first inequality if $X = p$ and $Y \stackrel{d}{=} B_p$ as in (1.11).

Proof. See the cited paper for the proof of (i). For (ii), take $n = 2$ in (2.7) and use $|X - Y| = 2(X \vee Y) - X - Y$. □

3 Proof of Proposition 1.1

The evaluation (1.7) in Proposition 1.1 is implied by Lemma 3.1 for $\delta \in [0, \frac{1}{2})$ and by example (1.10) for $\delta \in [\frac{1}{2}, 1]$.

Lemma 3.1. *If $X = E(Y \mid X)$ and $0 \le Y \le 1$ then $P(|Y - X| \ge 1 - \delta) \le \delta$ for $\delta \in [0, \frac{1}{2})$, with equality if $X = \delta$ and $Y = B_\delta$.*

Proof. Suppose $X = p$ is constant and $Y = Y_p \in [0, 1]$ has $EY_p = p$. By consideration of $Y_{1-p} = 1 - Y_p$ it can be supposed that $p \in [0, \frac{1}{2}]$. But then for $\delta \in [0, \frac{1}{2})$

$$|Y_p - p| \ge 1 - \delta \text{ iff } Y_p \ge 1 - \delta + p,$$

so Markov's inequality gives

$$P(|Y_p - p| \ge 1 - \delta) \le \frac{p\, \mathbb{1}(p \le \delta)}{1 - \delta + p} \le \delta \text{ for } 0 \le p \le \tfrac{1}{2} \text{ and } 0 \le \delta < \tfrac{1}{2}. \tag{3.1}$$

The more general assertion of the lemma follows by conditioning on X. □

Turning to consideration of (1.8), we start with a lemma of independent interest, which controls the variability of $P(A \mid G)$ as a function of G with $P(G) > 0$ by a bound that does not depend on A. We work here with the elementary conditional probability which is the number $P(A \mid G) := P(A \cap G)/P(G)$ rather than a random variable. Let $G \triangle H := (G \cap H^c) \cup (G^c \cap H)$ denote the symmetric difference of G and H.

Lemma 3.2. *For events A, G and H with $P(G) > 0$ and $P(H) > 0$,*

$$|P(A\,|\,G) - P(A\,|\,H)| \leq P(G \triangle H\,|\,G \cup H) = 1 - \frac{P(G \cap H)}{P(G) + P(H) - P(G \cap H)}. \tag{3.2}$$

Consequently, for each $0 \leq \delta \leq 1$,

$$|P(A\,|\,G) - P(A\,|\,H)| \geq 1 - \delta \implies P(G \cap H) \leq \frac{\delta}{(1+\delta)}(P(G) + P(H)). \tag{3.3}$$

Proof. Let $p = P(G \cap H^c)$, $q = P(G \cap H)$, $r = P(G^c \cap H)$ and $a = P(A\,|\,G \cap H^c)$, $b = P(A\,|\,G \cap H)$, $c = P(A\,|\,G^c \cap H)$, with the convention that $a = 0$ if $P(G \cap H^c) = 0$, and a similar convention for b and c. Then

$$P(A\,|\,G) - P(A\,|\,H) = \frac{pa + qb}{p+q} - \frac{qb + rc}{q+r} \leq \frac{p+r}{p+q+r} \tag{3.4}$$

from which (3.2)–(3.3) follow easily. To check the inequality in (3.4), observe that for fixed p, q, r the difference of fractions in the middle is obviously maximized by taking $a = 1, c = 0$. That done, the difference is a linear function of b, whose maximum over $0 \leq b \leq 1$ is attained either at $b = 0$ or at $b = 1$, when the inequality is obvious. □

It is easily checked that for p, q, r as above, with $p + q > 0$ and $q + r > 0$, there is equality in (3.4) iff one of the following three conditions holds, where in each case the condition on G, H, and A should be understood modulo events of probability 0:

- either $p > 0, q = 0, r > 0, a = 1, b = c = 0$, meaning $G \cap H = \emptyset$ and $A = G$;
- or $p = 0, q > 0, r > 0, a = 0, b = 1, c = 0$, meaning $G \subseteq H$ and $A = G$;
- or $p > 0, q > 0, r = 0, a = 1, b = c = 0$, meaning $H \subseteq G$ and $A = G \cap H^c$.

Consequently, there is equality in (3.2) iff one of these three conditions holds, either exactly as above or with G and H switched.

Lemma 3.3. *Suppose that $X = P(A\,|\,X)$ and $Y = P(A\,|\,Y)$ have discrete distributions. Fix $0 < \delta < 1/2$, and suppose that for each pair of possible values (x, y) of (X, Y) with $|y - x| \geq 1 - \delta$ there is no other such pair (x', y') with either $x' = x$ or $y' = y$. Then*

$$P(|Y - X| \geq 1 - \delta) \leq \frac{2\delta}{1+\delta} \qquad (0 < \delta < 1/2). \tag{3.5}$$

Proof. Application of (3.3) gives for each pair (x, y) with $|y - x| \geq 1 - \delta$

$$P(X = x, Y = y) \leq \frac{\delta}{1+\delta}\left(P(X = x) + P(Y = y)\right). \tag{3.6}$$

The assumption is that as (x, y) ranges over pairs (x, y) with $|y - x| \geq 1 - \delta$, the events $(X = x)$ are disjoint, and so are the events $(Y = y)$. So (3.5) follows by summation of (3.6) over such (x, y). □

Proof of (1.8)*.* The example given in (1.10) proves (1.8) for $\delta \in [\frac{1}{2}, 1]$. The claim (1.7) has been proved at the beginning of this section. Hence, we can limit our attention to the case when each X and Y takes two values. For (X, Y) in (1.4), $P(|X - Y| \geq 1 - \delta) = 2\delta/(1 + \delta)$ so the lower bound in (1.8) is proved. It is now enough to establish (3.5) for 2×2 coherent (X, Y) whose possible values are contained in the 4 corners of a rectangle $R := [x_1, x_2] \times [y_1, y_2] \subseteq [0, 1]^2$ with $x_1 < x_2$ and $y_1 < y_2$. Fix $0 < \delta < \frac{1}{2}$. Then $\{(x, y) : |y - x| \geq 1 - \delta\} = T \cup T'$ for right triangles T and T' in the upper left and lower right corners of $[0, 1]^2$. If neither T nor T' contains two corners on the same side of R, then (3.5) holds by Lemma 3.3. Otherwise, by the reflection symmetries (1.3), it is enough to discuss the case when T contains the two left corners of R. If T contains at least three corners of R then $EX \leq \delta < 1/2 < 1-\delta \leq EY$. This is not possible because $EX = EY$. Finally, suppose that the two left corners of R are in T and the two right corners not in T and, therefore, not in $T \cup T'$. Let $Y' \equiv y_3 = EY$. Note that $y_1 \leq y_3 \leq y_2$ so $(x_1, y_3) \in T$ and $(x_2, y_3) \notin T \cup T'$. Hence, by (1.7) applied to (X, Y'),

$$\begin{aligned} P(|X-Y| \geq 1-\delta) = P((X,Y) \in T) = P((X,Y') \in T) = P(|X-Y'| \geq 1-\delta) \\ \leq \delta \leq \frac{2\delta}{1+\delta}. \end{aligned}$$

□

Proof of (1.9). This argument from [30] was presented in [4, Theorem 18.1, p. 389], but is included here for the reader's convenience. The lower bound in (1.9) is obvious from (1.6). For the upper bound, it is enough to discuss the case $\delta \in [0, \frac{1}{2})$. Observe that

$$(|X-Y| \geq 1-\delta) \subseteq (X \leq \delta, Y \geq 1-\delta) \cup (Y \leq \delta, X \geq 1-\delta). \tag{3.7}$$

But since $X = P(A \,|\, X)$ and $1 - Y = P(A^c \,|\, Y)$,

$$\begin{aligned} P(X \leq \delta, Y \geq 1-\delta, A) &\leq P(X \leq \delta, A) = E\mathbb{1}(X \leq \delta)X \leq \delta P(X \leq \delta), \\ P(X \leq \delta, Y \geq 1-\delta, A^c) &\leq P(Y \geq 1-\delta, A^c) = E\mathbb{1}(1-Y \leq \delta)(1-Y)) \leq \delta P(Y \geq 1-\delta). \end{aligned}$$

It follows that

$$P(X \leq \delta, Y \geq 1-\delta) \quad \leq \quad \delta[P(X \leq \delta) + P(Y \geq 1-\delta)], \tag{3.8}$$

$$P(Y \leq \delta, X \geq 1-\delta) \quad \leq \quad \delta[P(Y \leq \delta) + P(X \geq 1-\delta)]. \tag{3.9}$$

For $\delta < 1/2$ the events $(X \leq \delta)$ and $(X \geq 1-\delta)$ are disjoint, so $P(X \leq \delta) + P(X \geq 1-\delta) \leq 1$, and the same for Y. Add (3.8) and (3.9) and use (3.7) to obtain the upper bound in (1.9). □

4 Coherent distributions

The following proposition summarizes a number of known characterizations of the set of coherent distributions of (X, Y), due to [13], [18] and [7].

Proposition 4.1. *Let (X, Y) be a pair of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, on which there is also defined a random variable U with uniform distribution, independent of (X, Y). Then the following conditions are equivalent:*

(i) The joint law of (X, Y) is coherent.

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

$$E[Zg(X)] = E[Xg(X)] \quad \text{and} \quad E[Zg(Y)] = E[Yg(Y)] \tag{4.1}$$

either for all bounded measurable functions g with domain $[0,1]$, or for all bounded continuous functions g.

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

$$E[\phi(X,Y)g(X)] = E[Xg(X)] \quad \text{and} \quad E[\phi(X,Y)g(Y)] = E[Yg(Y)] \tag{4.2}$$

either for all bounded measurable g, or for all bounded continuous g.

(iv) $EX = EY = p$ for some $0 \leq p \leq 1$, and

$$E\big[X\mathbb{1}(X \in B)\big] + E\big[Y\mathbb{1}(Y \in C)\big] \leq p + P(X \in B, Y \in C) \tag{4.3}$$

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

Proof. Condition (i) is just (ii) for Z an indicator variable, while (ii) for $0 \le Z \le 1$ implies (iii) for $\phi(X,Y) = E(Z \mid X,Y)$. Assuming (iii), (ii) holds with $Z = \mathbb{1}(U \le \phi(X,Y))$ for the uniform $[0,1]$ variable U independent of (X,Y). So (i), (ii) and (iii) are equivalent. The equivalence of (iii) and (iv) is an instance of [33, Theorem 6], according to which for any finite measure m on $[0,1]^2$, a pair of probability distributions Q and R on $[0,1]$ are the marginals of the measure $\phi(x,y)m(dx\,dy)$ on $[0,1]^2$, for ϕ a product measurable function with $0 \le \phi \le 1$, iff

$$Q(B) + R(C) \le 1 + m(B \times C)$$

for all Borel sets B and C. This is equivalent to the same condition for all finite unions of intervals, by elementary measure theory. After dismissing the trivial case $p = 0$, this result is applied here to $m(\cdot) = P((X,Y) \in \cdot)/p$ for X and Y with mean p, with $Q(B) := E\left[X\mathbb{1}(X \in B)\right]/p$ and $R(C) := E\left[Y\mathbb{1}(Y \in C)\right]/p$. □

The characterizations (ii) and (iii) above extend easily to a coherent family $(X_i, i \in I)$, while (iv) does not [7, p. 288].

Corollary 4.2 ([13])**.** *For any finite I, the set of coherent distributions of $(X_i, i \in I)$ is a convex, compact subset of probability distributions on $[0,1]^I$ with the usual weak topology.*

Proof. To check convexity, suppose that $(X_i, i \in I)$ is subject to the extension of (4.1). That is for some additional index $* \notin I$ and $X_* = Z \in [0,1]$,

$$E[X_* g(X_i)] = E[X_i g(X_i)] \text{ for all bounded continuous } g \text{ and } i \in I, \tag{4.4}$$

and the same for $Y = (Y_i, i \in I_*)$ instead of X, with $I_* := I \cup \{*\}$. Construct these random vectors X and Y on a common probability space with a Bernoulli(p) variable B_p, with X, Y and B_p independent. Let $W := B_p X + (1 - B_p)Y$, so the law of W is the mixture of laws of X and Y with weights p and $1 - p$. Then (4.4) for X and Y implies (4.4) for W.

The proof of compactness is similar. Suppose X is the limit in distribution of some sequence of random vectors $X_n := (X_{n,i}, i \in I)$. Then the sequence of random vectors $X_n := (X_{n,i}, i \in I_*)$ subject to (4.4) has a subsequence which converges in distribution to some $(X_i, i \in I_*)$, and deduce (4.4) for $(X_i, i \in I_*)$ using bounded convergence. □

Corollary 4.3. *Let $\mathcal{C}$ be a non-empty set of distributions of $X = (X_i, i \in I)$ on $\mathbb{R}^I$ that is compact in the topology of weak convergence, such as coherent distributions of X on $[0,1]^I$. Let $G(x) := \sup_{\mathcal{C}} P(g(X) \le x)$ for some particular continuous function g, and $x \in \mathbb{R}$, where the $\sup_{\mathcal{C}}$ is over X with a distribution in $\mathcal{C}$. Then*

- *(i) for each fixed $x \in \mathbb{R}$ there exists a distribution of X in $\mathcal{C}$ with $G(x) = P(g(X) \le x)$;*
- *(ii) $G(x) = P(\gamma \le x)$ is the cumulative distribution function of a random variable γ which is stochastically smaller than $g(X)$ for every distribution of X in $\mathcal{C}$: $P(\gamma > x) \le P(g(X) > x)$ for all real x.*

Proof. By definition of $G(x)$, for each fixed x there exists a sequence of random vectors X_n with distributions in $\mathcal{C}$ such that $F_n(x) := P(g(X_n) \le x) \uparrow G(x)$. By compactness of $\mathcal{C}$, it may be supposed that $X_n \xrightarrow{d} X$, meaning the distribution of X_n converges to that of some $X \in \mathcal{C}$. That implies $g(X_n) \xrightarrow{d} g(X)$. Let $F(x) := P(g(X) \le x)$. Since $F_n(x)$ and $F(x)$ are the probabilities assigned by the laws of $g(X_n)$ and $g(X)$ to the closed set $(-\infty, x]$, [2, Theorem 29.1] gives

$$G(x) \ge F(x) \ge \limsup_n F_n(x) = G(x).$$

For (ii), the only property of a cumulative distribution function that is not an obvious property of G is right continuity. To see this, take $x_n \downarrow x$ and X_n with $P(g(X_n) \le x) = F_n(x)$ such that $F_n(x_n) = G(x_n)$, and $X_n \xrightarrow{d} X$ with distribution in $\mathcal{C}$. Let $F(x) := P(g(X) \le x)$. Then for each fixed m, by the same result of [2],

$$F(x_m) \ge \limsup_n F_n(x_m) \ge \limsup_n F_n(x_n) = \limsup_n G(x_n) = G(x+).$$

Finally, letting $m \to \infty$ gives $G(x) \ge F(x) = F(x+) \ge G(x+) \ge G(x)$. □

Returning to discussion of just a pair random variables (X, Y) with values in $[0,1]^2$, as in Proposition 4.1, suppose further that X and Y are independent, with $EX = EY = p$. Then the inequality (4.3) becomes

$$EX\mathbb{1}(X \in B) + EY\mathbb{1}(Y \in C) \le p + P(X \in B)P(Y \in C). \tag{4.5}$$

It was shown in [18, Theorem 4] that this condition, just for $B = (s, 1]$ and $C = (t, 1]$ for $0 \le s, t \le 1$, characterizes all possible pairs of marginal distributions on $[0, 1]$ of independent X and Y with mean p such that (X, Y) is coherent. See also [24, Proposition 3].

5 Open problems

Conjecture 5.1. *If (X, Y) is coherent, and X and Y are independent, then*

$$P(|X - Y| \ge 1 - \delta) \le 2\delta(1 - \delta) \qquad \textit{for } \delta \in [0, \tfrac{1}{2}). \tag{5.1}$$

Equality is attained in (5.1) for independent X and Y with

$$X \stackrel{d}{=} Y \stackrel{d}{=} (1 - \delta)B_{1-\delta} \text{ and } A = (X = Y = 1 - \delta). \tag{5.2}$$

One can prove (5.1) for 2×2 laws of (X, Y) in a manner similar to the proof of (1.8); we leave the proof to the reader. But like Theorem 1.2, the extension of (5.1) to general distributions of X and Y seems quite challenging.

The problems solved by (1.8) for $t(X, Y) = 1(|X - Y| \ge 1 - \delta)$ and by the case $n = 2$ of (2.7) for $t(X, Y) = X \vee Y$, are instances of the following more general problem, with further variants as above, assuming X and Y are independent.

Problem 5.2 ([13, p. 224])**.** Given some target function $t(X, Y)$ defined on $[0, 1]^2$, evaluate $\sup_{\mathcal{C}} Et(X, Y)$, the supremum of $Et(X, Y)$ as the law of (X, Y) ranges over the set $\mathcal{C}$ of coherent laws on $[0, 1]^2$. Or the same for $\mathcal{C}(p)$, coherent laws of (X, Y) with $EX = EY = p$.

This problem seems to be open even for XY, or $|X - Y|^r$ for $r \ne 1$. Another instance of this problem is to evaluate

$$\varepsilon(\delta, p) := \sup_{\mathcal{C}(p)} P(|X - Y| \ge 1 - \delta). \tag{5.3}$$

For each $\delta \in (0, 1)$, examples of coherent (X, Y) with

$$P(|X - Y| \ge 1 - \delta) = p(\delta) := 2\delta/(1 + \delta) \tag{5.4}$$

are the 2×2 example (1.4), say (X_δ, Y_δ), its reflection $(1 - X_\delta, 1 - Y_\delta)$, and any mixture of these two laws, which is a 4×4 law in $\mathcal{C}(p)$ for p between $p(\delta)$ and $1 - p(\delta)$. So

$$p(\delta) \le \varepsilon(\delta, p) \le \varepsilon(\delta) \text{ for } p \text{ between } p(\delta) \text{ and } 1 - p(\delta). \tag{5.5}$$

It follows from Theorem 1.2 that both inequalities are equalities for $\delta \in (0, \frac{1}{2}]$. But that leaves open:

Problem 5.3. Find $\varepsilon(\delta, p)$ for $\delta \in (0, \frac{1}{2}]$, and p not covered by (5.5).

Problem 5.2 is related to some concepts in the optimal transport theory. For example, the square of the L^2-Wasserstein distance between the distributions of X and Y is the minimum of $Et(X', Y')$ for $t(x, y) = (x - y)^2$, over all (X', Y') with the marginal distributions the same as those of X and Y (see [35, Ch. 6]).

For a bounded upper semicontinuous t, such as the indicator of a closed set, the $\sup_{\mathcal{C}} Et(X, Y)$ will be attained at a distribution of (X, Y) in ext($\mathcal{C}$), the set of extreme points of the compact, convex set $\mathcal{C}$ of coherent distributions [1]. This leads to:

Problem 5.4 ([13, p. 224] [7, p. 273])**.** Characterize ext($\mathcal{C}$).

For the particular target functions t involved in (2.7) and in Theorem 1.2, the $\sup_{\mathcal{C}} Et(X, Y)$ is attained by 2×2 distributions of (X, Y).

It has been recently proved in [36] that there are extreme coherent laws of (X, Y) with an arbitrarily large finite number of atoms.

The following proposition is easily proved using (4.3):

Proposition 5.5. *For each a rectangle $R = [x_1, x_2] \times [y_1, y_2] \subseteq [0, 1]^2$, let $\mathcal{C}_{2\times 2}(R)$ denote the set of coherent laws of (X, Y) on the corners of R. Then*

- *$\mathcal{C}_{2\times 2}(R)$ is non-empty iff R intersects the diagonal $\{(p, p), 0 \leq p \leq 1\}$, that is iff $x_1 \vee y_1 \leq x_2 \wedge y_2$.*
- *If $x_1 \vee y_1 = x_2 \wedge y_2 = p$, then (p, p) is a corner of R, and the unique law in $\mathcal{C}_{2\times 2}(R)$ is degenerate with $X = Y = p$.*
- *If $x_1 \vee y_1 < x_2 \wedge y_2$, the set of laws $\operatorname{ext} \mathcal{C}_{2\times 2}(R)$ forms a convex polygon in a 2-dimensional affine subspace of the set of probability distributions on those corners, with at least 2 and at most 8 vertices.*

It has been proved in [36] that the number of vertices must be 2, 3, 4 or 6, and examples show that each of these cases holds for some distribution.

Problem 5.6. Provide an accounting of the extreme 2×2 coherent laws of (X, Y) which is adequate to recover (1.8) and (2.8), and to find the extrema of $Et(X, Y)$ over 2×2 coherent laws for other functions t, such as $t(X, Y) = XY$ or $|X - Y|^r$ for $r > 0$.

Problem 5.7. Extensions of above problems to $n > 2$ coherent opinions.

References

[1] Viktor Beneš and Josef Štěpán, *Extremal solutions in the marginal problem*, Advances in probability distributions with given marginals (Rome, 1990), Math. Appl., vol. 67, Kluwer Acad. Publ., Dordrecht, 1991, pp. 189–206. MR-1215952

[2] Patrick Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995, A Wiley-Interscience Publication. MR-1324786

[3] Krzysztof Burdzy, *The search for certainty*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2009, On the clash of science and philosophy of probability. MR-2510150

[4] Krzysztof Burdzy, *Resonance—from probability to epistemology and back*, Imperial College Press, London, 2016. MR-3468703

[5] Krzysztof Burdzy and Soumik Pal, *Contradictory predictions*, (2019), preprint, arXiv:1912.00126.

[6] R. Casarin, G. Mantoan, and F. Ravazzolo, *Bayesian calibration of generalized pools of predictive distributions*, Econometrics **4** (2016), no. 1.

[7] A. P. Dawid, M. H. DeGroot, and J. Mortera, *Coherent combination of experts' opinions*, Test **4** (1995), no. 2, 263–313. MR-1379793

[8] A. P. Dawid and J. Mortera, *A note on prediction markets*, Available at arxiv[math.ST] arXiv:1702.02502, 2018.

[9] M. H. DeGroot and J. Mortera, *Optimal linear opinion pools*, Management Science **37** (1991), no. 5, 546–558.

[10] Morris H. DeGroot, *A Bayesian view of assessing uncertainty and comparing expert opinion*, J. Statist. Plann. Inference **20** (1988), no. 3, 295–306. MR-976182

[11] Lester E. Dubins, David Gilat, and Isaac Meilijson, *On the expected diameter of an L_2-bounded martingale*, Ann. Probab. **37** (2009), no. 1, 393–402. MR-2489169

[12] Lester E. Dubins and Jim Pitman, *A divergent, two-parameter, bounded martingale*, Proc. Amer. Math. Soc. **78** (1980), no. 3, 414–416. MR-553386

[13] Lester E. Dubins and Jim Pitman, *A maximal inequality for skew fields*, Z. Wahrsch. Verw. Gebiete **52** (1980), no. 3, 219–227. MR-576883

[14] Kenny Easwaran, Luke Fenton-Glynn, Christopher Hitchcock, and Joel D. Velasco, *Updating on the credences of others: Disagreement, agreement, and synergy*, Philosophers' Imprint **16** (2016), 1–39.

[15] Simon French, *Aggregating expert judgement*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM **105** (2011), no. 1, 181–206. MR-2783806

[16] Christian Genest, *A characterization theorem for externally Bayesian groups*, Ann. Statist. **12** (1984), no. 3, 1100–1105. MR-751297

[17] Tilmann Gneiting and Roopesh Ranjan, *Combining predictive distributions*, Electron. J. Stat. **7** (2013), 1747–1782. MR-3080409

[18] Sam Gutmann, J. H. B. Kemperman, J. A. Reeds, and L. A. Shepp, *Existence of probability measures with given marginals*, Ann. Probab. **19** (1991), no. 4, 1781–1797. MR-1127728

[19] Pierre Henry-Labordère, Jan Obłój, Peter Spoida, and Nizar Touzi, *The maximum maximum of a martingale with given n marginals*, Ann. Appl. Probab. **26** (2016), no. 1, 1–44. MR-3449312

[20] G. Kapetanios, J. Mitchell, S. Price, and N. Fawcett, *Generalised density forecast combinations*, J. Econometrics **188** (2015), no. 1, 150–165. MR-3371665

[21] Younes Kchia and Philip Protter, *Progressive filtration expansions via a process, with applications to insider trading*, Int. J. Theor. Appl. Finance **18** (2015), no. 4, 1550027, 48. MR-3358108

[22] Davar Khoshnevisan, *Multiparameter processes*, Springer Monographs in Mathematics, Springer-Verlag, New York, 2002, An introduction to random fields. MR-1914748

[23] Fabian Krüger and Ingmar Nolte, *Disagreement versus uncertainty: Evidence from distribution forecasts*, Journal of Banking & Finance **72** (2016), S172–S186, IFABS 2014: Bank business models, regulation, and the role of financial market participants in the global financial crisis.

[24] Steffen L. Lauritzen, *Rasch models with exchangeable rows and columns*, Bayesian statistics, 7 (Tenerife, 2002), Oxford Univ. Press, New York, 2003, pp. 215–232. MR-2003175

[25] Jan Lorenz, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing, *How social influence can undermine the wisdom of crowd effect*, Proceedings of the National Academy of Sciences **108** (2011), no. 22, 9020–9025.

[26] Annette Möller and Jörgen Groß, *Probabilistic temperature forecasting based on an ensemble autoregressive modification*, Quarterly Journal of the Royal Meteorological Society **142** (2016), no. 696, 1385–1394.

[27] Enrique Moral-Benito, *Model averaging in economics: An overview*, Journal of Economic Surveys **29** (2015), no. 1, 46–75.

[28] Adam Osękowski, *Estimates for the diameter of a martingale*, Stochastics **87** (2015), no. 2, 235–256. MR-3316810

[29] Adam Osękowski, *Method of moments and sharp inequalities for martingales*, Inequalities and extremal problems in probability and statistics, Academic Press, London, 2017, pp. 1–27. MR-3702299

[30] Jim Pitman, *Bounds on the probability of radically different opinions*, Unpublished, 2014.

[31] Roopesh Ranjan and Tilmann Gneiting, *Combining probability forecasts*, J. R. Stat. Soc. Ser. B Stat. Methodol. **72** (2010), no. 1, 71–91. MR-2751244

[32] Ville A. Satopää, Robin Pemantle, and Lyle H. Ungar, *Modeling probability forecasts via information diversity*, J. Amer. Statist. Assoc. **111** (2016), no. 516, 1623–1633. MR-3601722

[33] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439. MR-177430

[34] James W. Taylor and Jooyoung Jeon, *Probabilistic forecasting of wave height for offshore wind turbine maintenance*, European J. Oper. Res. **267** (2018), no. 3, 877–890. MR-3760810

[35] Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR-2459454

[36] Theodore Zhu, *On coherent distributions*, (forthcoming), 2020.

Acknowledgments. We are grateful to David Aldous, Soumik Pal and Teddy Zhu for very helpful advice.