

---

# Conformal prediction regions are imprecise highest density regions

---

Michele Caprio<sup>1,2</sup>

Yusuf Sale<sup>3,4</sup>

Eyke Hüllermeier<sup>3,4</sup>

<sup>1</sup>Department of Computer Science, University of Manchester, United Kingdom

<sup>2</sup>Manchester Centre for AI Fundamentals (MCAIF), United Kingdom

<sup>3</sup>Institute of Informatics, LMU Munich, Germany

<sup>4</sup>Munich Center for Machine Learning, Germany

## ABSTRACT

Recently, Cella and Martin proved how, under an assumption called *consonance*, a credal set (i.e. a closed and convex set of probabilities) can be derived from the conformal transducer associated with transductive conformal prediction. We show that the Imprecise Highest Density Region (IHDR) associated with such a credal set corresponds to the classical Conformal Prediction Region. In proving this result, we establish a new relationship between Conformal Prediction and Imprecise Probability (IP) theories, via the IP concept of a cloud. A byproduct of our presentation is the discovery that consonant plausibility functions are monoid homomorphisms, a new algebraic property of an IP tool.

**Keywords.** conformal prediction, credal machine learning, prediction sets, consonance

## 1. INTRODUCTION

Conformal prediction (CP) is a methodology introduced by Vovk, Gammerman, and Shafer [68] whose main goal is to output a prediction for the next observation's value, given a collection of data points, which takes into account the uncertainty faced by the agent. Loosely, CP accounts for the latter by producing predictions in the form of *conformal prediction regions* (CPRs), subsets of the output space that are likely to cover the target of prediction – the true outcome – with high probability, for all possible exchangeable probability distributions on the output space. CP is a model-free approach: it does not require any distributional assumptions on the underlying data-generating process, making it a versatile and widely applicable tool e.g. for statistical inference.

A key part of (transductive) CP is to derive the conformal transducer  $\pi$ , a “score” between 0 and 1 that measures how “conformal” an element  $y$  of the output space  $\mathbb{Y}$  is to the data previously available to the user. The “more conformal” element  $y$  is, the closer its “score” is to 1. In their work [19, 20], Cella and Martin show that, once  $\pi$  is available, an upper probability  $\bar{\Pi}$  (that can be

thought of as the upper envelope of a closed and convex set of probabilities, i.e. a credal set) can be derived from it. This is possible under a given assumption called *consonance*, which, roughly, tells us that there is at least an element  $y \in \mathbb{Y}$  such that  $\pi(y) = 1$ .<sup>1</sup> Under consonance, upper probability  $\bar{\Pi}$  is defined as  $\bar{\Pi}(A) = \sup_{y \in A} \pi(y)$ , for all  $A \subseteq \mathbb{Y}$ , and the corresponding credal set is given by  $\mathcal{M}(\bar{\Pi}) = \{P : P(A) \leq \bar{\Pi}(A), \forall A \subseteq \mathbb{Y}\}$ . That is,  $\mathcal{M}(\bar{\Pi})$  contains all the probabilities on  $\mathbb{Y}$  that are set-wise dominated by  $\bar{\Pi}$ .

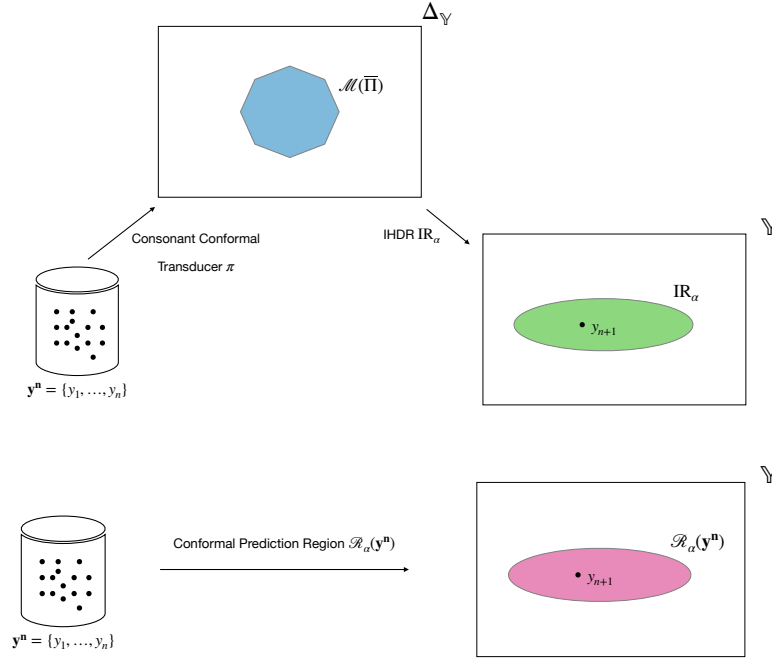
In this work, we further the study of the relationship between CP and Imprecise Probabilities (IPs). We show that the Imprecise Highest Density Region (IHDR) associated with  $\mathcal{M}(\bar{\Pi})$  is equivalent to the classical Conformal Prediction Region (CPR), and it retains the same (uniform) probabilistic guarantee. Loosely, an IHDR [26] is the IP counterpart of a Highest Density Region, which in turn can be thought of as a Bayesian version of a Confidence Interval. It is a collection of elements of the output space  $\mathbb{Y}$ , that *all* the elements of credal set  $\mathcal{M}(\bar{\Pi})$  indicate as being correct with high probability  $1 - \alpha$ , with significance level  $\alpha$  chosen by the user. A visual representation of our proposed method is given in Figure 1. In proving this result, we also relate Conformal Prediction to the concept of cloud [57], and discover a very interesting algebraic property of consonant plausibility functions like the upper probability  $\bar{\Pi}$ .

We conclude with a discussion on the open problems and unanswered questions on the nature of CP, and on the relationship between CP and IP theory.

The paper is organized as follows. Section 2.1 recalls the basics of (transductive) Conformal Prediction. Section 2.2 introduces the concepts from Imprecise Probability (IP) theory that are needed in our work. Section 3 presents the relationship between conformal prediction and the IP notion of plausibility functions noted by Cella and Martin [19, 20]. Section 4 introduces Imprecise Highest Density Regions (IHDRs). In Section 5,

---

<sup>1</sup>Consonance can for instance be obtained “artificially” by putting to 1 the value of  $\pi$  for  $\arg \max_y \pi(y)$  [19, Section 7].



**Figure 1.** Top: Our proposed, “indirect” methodology to derive a prediction region. We first use the consonant conformal transducer  $\pi$  to derive credal set  $\mathcal{M}(\overline{\Pi})$ , and then extract from the latter the IHDR  $\text{IR}_{\alpha}$ . Bottom: Classical CP methodology, in which the Conformal Prediction Region is obtained as in (1).

we derive the IHDR of our conformally-built credal set, that corresponds to the classical conformal prediction region. Section 6 concludes our work and discusses further interesting open problems.

## 2. PRELIMINARIES

**2.1. Conformal prediction.** This section is based on Cella and Martin [20, Section 4.1], who summarize work by Shafer and Vovk [64] and Vovk, Gammerman, and Shafer [68] on transductive conformal prediction.

Suppose that there is an exchangeable process  $Y_1, Y_2, \dots$  with distribution  $\mathbf{P}$ , where each  $Y_i$  is a random element taking values in  $\mathbb{Y}$ . Throughout the paper, we denote by  $\mathbf{P}$  the true data generating process, and by  $P$  a generic element of  $\Delta_{\mathbb{Y}}$ , the space of all (countably additive) probability measures on  $\mathbb{Y}$ . Recall that a sequence is exchangeable if, for any  $k \in \mathbb{N}$  and any permutation  $\tau$ , the two random vectors  $(Y_1, \dots, Y_k)^{\top}$  and  $(Y_{\tau(1)}, \dots, Y_{\tau(k)})^{\top}$  have the same joint distribution. This implies that the marginal distributions of the  $Y_i$ ’s are the same.<sup>2</sup>

We want to solve the following statistical problem. Suppose we observe the first  $n$  terms of the process, that is,  $\mathbf{Y}^n = (Y_1, \dots, Y_n)^{\top}$ . With this data, and the assumption of exchangeability, the goal is to predict  $Y_{n+1}$  using a method that is valid or reliable in a certain sense.

<sup>2</sup>Exchangeability does not imply independence, so the standard i.i.d. setup is in a sense more restrictive. Note also that we are not assuming any parametric form for the distribution  $\mathbf{P}$ .

Let  $\mathbf{Y}^{n+1} = (\mathbf{Y}^n, Y_{n+1})^{\top}$  be an  $(n+1)$ -dimensional vector consisting of the observable  $\mathbf{Y}^n$  and the yet-to-be-observed value  $Y_{n+1}$ . Consider the transform

$$\mathbf{Y}^{n+1} \rightarrow \mathbf{T}^{n+1} = (T_1, \dots, T_{n+1})^{\top}$$

defined by the rule

$$T_i := \psi_i(\mathbf{Y}^{n+1}) \equiv \Psi(\mathbf{y}_{-i}^{n+1}, y_i)$$

for all  $i \in \{1, \dots, n+1\}$ , where  $\mathbf{y}_{-i}^{n+1} = \mathbf{y}^{n+1} \setminus \{y_i\}$  and  $\Psi : \mathbb{Y}^n \times \mathbb{Y} \rightarrow \mathbb{R}$  is a fixed function that is invariant to permutations in its first vector argument. Function  $\Psi$ , which is called a *non-conformity measure*, is constructed in such a way that  $\psi_i(\mathbf{y}^{n+1})$  is small if and only if  $y_i$  agrees with – i.e. is “close to” – a prediction based on the data  $\mathbf{y}_{-i}^{n+1}$ . Or, the other way around, large values  $\psi_i(\mathbf{y}^{n+1})$  suggest that the observation  $y_i$  is “strange” and does not conform to the rest of the data  $\mathbf{y}_{-i}^{n+1}$ . The key idea is to define  $\psi_i(\mathbf{y}^{n+1})$  in a way that allows one to compare  $y_i$  to a suitable summary of  $\mathbf{y}_{-i}^{n+1}$ , e.g.  $\psi_i(\mathbf{y}^{n+1}) = |\text{mean}(\mathbf{y}_{-i}^{n+1}) - y_i|$ , for all  $i \in \{1, \dots, n+1\}$ . Notice that transformation  $\mathbf{Y}^{n+1} \rightarrow \mathbf{T}^{n+1}$  preserves exchangeability.

As the value  $Y_{n+1}$  has not yet been observed, and is actually the target of prediction, the above calculations cannot be carried out exactly. Nevertheless, the exchangeability-preserving properties of the transformations described above provide a procedure to rank candidate values  $\tilde{y}$  of  $Y_{n+1}$  based on the observed  $\mathbf{Y}^n = \mathbf{y}^n$ , as shown in Algorithm 1.

**Algorithm 1** Conformal prediction (CP)

---

```

Initialize: data  $\mathbf{y}^n$ , non-conformity measure  $\Psi$ , grid of
 $\tilde{y}$  values
for each  $\tilde{y}$  value in the grid do
    set  $y_{n+1} = \tilde{y}$  and write  $\mathbf{y}^{n+1} = \mathbf{y}^n \cup \{y_{n+1}\}$ ;
    define  $T_i = \psi_i(\mathbf{y}^{n+1})$ , for all  $i \in \{1, \dots, n+1\}$ ;
    evaluate  $\pi(\tilde{y}, \mathbf{y}^n) = (n+1)^{-1} \sum_{i=1}^{n+1} \mathbb{1}[T_i \geq T_{n+1}]$ ;
end for
return  $\pi(\tilde{y}, \mathbf{y}^n)$  for each  $\tilde{y}$  on the grid.

```

---

The output of Algorithm 1 is a data-dependent function  $\tilde{y} \mapsto \pi(\tilde{y}, \mathbf{y}^n)$  that can be interpreted as a measure of plausibility of the assertion that  $Y_{n+1} = \tilde{y}$ , given data  $\mathbf{y}^n$ . Vovk, Gammerman, and Shafer [68] refer to the function  $\pi$  as *conformal transducer*. Conformal transducer  $\pi$  plays a key role in the construction of *conformal prediction regions* (CPRs). For any  $\alpha \in [0, 1]$ , the  $\alpha$ -level CPR is defined as [67, Equation (2)]

$$\mathcal{R}_\alpha(\mathbf{y}^n) := \{y_{n+1} \in \mathbb{Y} : \pi(y_{n+1}, \mathbf{y}^n) > \alpha\}, \quad (1)$$

and it satisfies

$$P[Y_{n+1} \in \mathcal{R}_\alpha(\mathbf{y}^n)] \geq 1 - \alpha, \quad (2)$$

uniformly in  $n$  and in  $P$  [68]. That is, (2) is satisfied for all  $n \in \mathbb{N}$  and all exchangeable distributions  $P$  on  $\mathbb{Y}$ . Note that (2) holds regardless of the choice of the non-conformity measure  $\Psi$ . However, the choice of this function is crucial in terms of the *efficiency* of conformal prediction, that is, the *size* of the prediction regions.

Transductive conformal prediction (TCP) is not the only way of carrying out conformal prediction. There exists another type called inductive (or split) conformal prediction (ICP). It assumes exchangeability and it is used to build the same region  $\mathcal{R}_\alpha(\mathbf{y}^n)$  in (1) having the same guarantee (2), while being computationally less expensive than TCP. Since ICP produces the same region as TCP, which is what we need in light of the goal of our paper, we do not present ICP in detail. We refer the interested reader to [3, 58].

**2.2. Imprecise probabilities.** Let  $(\mathbb{Y}, \Sigma_{\mathbb{Y}})$  be a measurable (prediction) space, where  $\mathbb{Y}$  is a nonempty set and  $\Sigma_{\mathbb{Y}}$  is a  $\sigma$ -algebra on  $\mathbb{Y}$ . A set function  $\nu : \Sigma_{\mathbb{Y}} \rightarrow [0, 1]$  is called (*Choquet*) *capacity* if  $\nu(\emptyset) = 0$ ,  $\nu(\mathbb{Y}) = 1$ , and  $\nu(A) \leq \nu(B)$  for all  $A, B \in \Sigma_{\mathbb{Y}}$  such that  $A \subseteq B$  [20, 22, 25]. When  $\mathbb{Y}$  is not finite,  $\nu$  needs to be continuous from above and below [5, Chapter 6].

Capacity  $\nu$  is called a *lower probability* (LP) if it is superadditive, that is, if  $\nu(A \cup B) \geq \nu(A) + \nu(B)$ , for all disjoint sets  $A, B \in \Sigma_{\mathbb{Y}}$ . In the remainder of the paper, we denote a generic LP by  $\underline{P}$  in view of the following property. A lower probability  $\underline{P}$  can be obtained as the *lower envelope* of a set  $\mathcal{P}$  of probability measures on  $(\mathbb{Y}, \Sigma_{\mathbb{Y}})$ ,

that is,

$$\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A),$$

for all  $A \in \Sigma_{\mathbb{Y}}$  [20, 22, 69]. If  $\mathcal{P}$  is convex and closed,<sup>3</sup> we call it *credal set*; a credal set having a finite number of extreme elements is called a *finitely generated credal set* (FGCS).<sup>4</sup>

Let  $\bar{P}$  denote the dual of an LP  $\underline{P}$ , that is,  $\bar{P}(A) = 1 - \underline{P}(A^c)$ , for all  $A \in \Sigma_{\mathbb{Y}}$ . We call  $\bar{P}$  an *upper probability* (UP). It is subadditive, that is,  $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B)$ , for all disjoint sets  $A, B \in \Sigma_{\mathbb{Y}}$ . In addition, it can be obtained as the *upper envelope* of a set  $\mathcal{P}$  of probability measures on  $(\mathbb{Y}, \Sigma_{\mathbb{Y}})$ , that is,  $\bar{P}(A) = \sup_{P \in \mathcal{P}} P(A)$ , for all  $A \in \Sigma_{\mathbb{Y}}$ .

The duality of  $\underline{P}$  and  $\bar{P}$  means that knowing one of them is sufficient to retrieve the other. We may refer to either of them individually with the understanding of their one-to-one relationship [42].

Capacity  $\nu$  is said to be *k-monotone* if for every collection  $\{A, A_1, \dots, A_k\} \subseteq \Sigma_{\mathbb{Y}}$  such that  $A_i \subseteq A$ , for all  $i \in \{1, \dots, k\}$ , we have

$$\nu(A) \geq \sum_{\emptyset \neq \mathcal{J} \subseteq \{1, \dots, k\}} (-1)^{|\mathcal{J}|-1} \nu\left(\bigcap_{i \in \mathcal{J}} A_i\right). \quad (3)$$

If a capacity is  $(k+1)$ -monotone, it is *k-monotone* as well. The smaller the  $k$ , the broader the class [42]. Capacity  $\nu$  is said to be *k-alternating* if for every collection  $\{A, A_1, \dots, A_k\} \subseteq \Sigma_{\mathbb{Y}}$  such that  $A_i \subseteq A$ , for all  $i \in \{1, \dots, k\}$ ,

$$\nu(A) \leq \sum_{\emptyset \neq \mathcal{J} \subseteq \{1, \dots, k\}} (-1)^{|\mathcal{J}|-1} \nu(\cup_{i \in \mathcal{J}} A_i). \quad (4)$$

If an LP  $\underline{P}$  is a *k-monotone* capacity, then its dual UP  $\bar{P}$  will be *k-alternating*. Similarly, if an UP  $\bar{P}$  is a *k-alternating* capacity, then its dual LP  $\underline{P}$  will be *k-monotone*.<sup>5</sup>

Lower probability  $\underline{P}$  is said to be *convex* if it is a 2-monotone capacity, that is, if  $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B) - \underline{P}(A \cap B)$ , for all  $A, B \in \Sigma_{\mathbb{Y}}$ . Similarly, UP  $\bar{P}$  is said to be *concave* if it is a 2-alternating capacity, that is, if  $\bar{P}(A \cup B) \leq \bar{P}(A) + \bar{P}(B) - \bar{P}(A \cap B)$ , for all  $A, B \in \Sigma_{\mathbb{Y}}$  [15].

Lower probability  $\underline{P}$  is called a *belief function* if it is an  $\infty$ -monotone capacity, i.e., if (3) holds for every  $k$ ; we denote it by *bel*. Upper probability  $\bar{P}$  is called a *plausibility function* if it is an  $\infty$ -alternating capacity, i.e., if (4) holds for every  $k$ ; we denote it by *pl*. Belief and plausibility functions are dual [63]. A belief function induces

<sup>3</sup>Here and in the rest of the paper, closed has to be understood with respect to the topology endowed to the space  $\Delta_{\mathbb{Y}}$  of all probability measures on  $(\mathbb{Y}, \Sigma_{\mathbb{Y}})$ , typically the weak or weak\* topologies.

<sup>4</sup>The extreme elements of an FGCS are those that cannot be written as convex combinations of one another.

<sup>5</sup>Notice that an LP cannot be *k-alternating* because it is superadditive, and similarly a UP cannot be *k-monotone* because it is subadditive.

a set-valued random variable on the power set of  $\mathbb{Y}$  [42]: if  $bel$  is a belief function, its associated *mass function* is the non-negative set function

$$m : \Sigma_{\mathbb{Y}} \rightarrow [0, 1], \quad A \mapsto m(A) := \sum_{B \subseteq A} (-1)^{|A-B|} bel(B), \quad (5)$$

where  $A - B \equiv A \cap B^c$ , and the subsets  $B$  of  $A$  have to belong to  $\Sigma_{\mathbb{Y}}$  as well. Mass function  $m$  has the following properties:

- (a)  $m(\emptyset) = 0$ ;
- (b)  $\sum_{B \subseteq \mathbb{Y}} m(B) = 1$ ;
- (c)  $bel(A) = \sum_{B \subseteq A} m(B)$ , and this representation is unique to  $bel$ .

Equation (5) is called the Möbius transform of  $bel$  [71]. A mass function  $m$  induces a probability measure on  $\Sigma_{\mathbb{Y}}$ , namely as the distribution of a random set. These concepts are further studied by Gong and Meng [42].

Plausibility function  $pl$  is said to be *consonant* if there exists a function  $\pi : \mathbb{Y} \rightarrow [0, 1]$  such that

- (i)  $\sup_{y \in \mathbb{Y}} \pi(y) = 1$ ;
- (ii)  $pl(A) = \sup_{y \in A} \pi(y)$ ,  $A \in \Sigma_{\mathbb{Y}}$ .

Function  $\pi$  corresponds to the probability density or mass function (pdf/pmf) of  $m$ , and is called a *plausibility contour*. We now show an interesting algebraic property of consonant plausibility functions.

Recall that a *semigroup* is a generic set  $S$  together with a binary operation  $\dagger$  (that is, a function  $\dagger : S \times S \rightarrow S$ ) that satisfies the associative property, i.e.  $(a \dagger b) \dagger c = a \dagger (b \dagger c)$ , for all  $a, b, c \in S$ . Semigroup  $S$  is called a *monoid* if it has an identity element  $e$ , that is, an element such that  $a \dagger e = e \dagger a = a$ , for all  $a \in S$ . Recall also that a *monoid homomorphism* is a map between monoids that preserves the monoid operation and maps the identity element of the first monoid to that of the second monoid. We have the following.

**Lemma 2.1.** *A consonant plausibility function  $pl$  is a monoid homomorphism between the monoids  $(\Sigma_{\mathbb{Y}}, \cup)$  and  $([0, 1], \oplus)$ , where  $\cup$  is the set union operation and  $\oplus$  is the tropical addition on  $[0, 1]$*

*Proof.* Consider the semigroups  $(\Sigma_{\mathbb{Y}}, \cup)$  and  $([0, 1], \oplus)$ , where  $\cup$  is the usual set union operation, and  $\oplus$  is the tropical addition, i.e.  $a \oplus b = \max\{a, b\}$ , for all  $a, b \in [0, 1] \subset \mathbb{R}$ . They are both monoids with identity elements  $\emptyset$  and  $0$ , respectively.

Then, a consonant plausibility function  $pl$  is a map  $pl : \Sigma_{\mathbb{Y}} \rightarrow [0, 1]$  such that  $pl(\emptyset) = 0$ , and  $pl(A \cup B) = pl(A) \oplus pl(B)$ . This shows that  $pl$  is a monoid homomorphism, as desired.  $\square$

While at the moment Lemma 2.1 seems like a result of independent interest, unrelated to our main result

in Section 5, we point out how in the future it may be useful to relate Imprecise Probabilities to an algebraic version of Conformal Prediction, in the spirit of the field of Algebraic Statistics [65].

Note that  $\pi(y)$  is the upper probability assigned to the singleton set  $\{y\}$ , which means that the entire lower and upper probability pair  $(P, \bar{P})$  is determined by the plausibility assigned to singletons. This is a unique feature of the consonance model [20, 36]. If we denote by  $\mathcal{F}_m$  the set of all focal elements, i.e.,  $A \in \Sigma_{\mathbb{Y}}$  such that  $m(A) > 0$ , then consonance corresponds to the case of having nested focal elements. Specifically, for any two elements  $A, B \in \Sigma_{\mathbb{Y}}$  we have either  $A \subseteq B$  or  $B \subseteq A$ , indicating a hierarchical structure [63]. We mention in passing that in classical possibility theory, “consonance” refers to the nestedness property we just described, while condition  $\sup_{y \in \mathbb{Y}} \pi(y) = 1$  is usually referred to as “normalization”. We do not make a distinction between the two to be consistent with the terminology in Cella and Martin’s works, and because normalization combined with the condition  $pl(A) = \sup_{y \in A} \pi(y)$ ,  $A \in \Sigma_{\mathbb{Y}}$ , implies nestedness.

We conclude by pointing out that Imprecise Probabilistic concepts like the ones we presented in this section are routinely implemented in many applied fields, such as Machine Learning and Artificial Intelligence [17, 32, 34, 37, 48, 61, 62, 70, 72], Statistics [9–14, 24, 60], Engineering [4, Chapter 4] – including aerospace engineering [66, Chapter 5] – and Economics [23, 30, 31].

### 3. RELATIONSHIP BETWEEN BELIEF FUNCTIONS AND TCP

This section is based on Cella and Martin [20, Section 4.2], who highlight the relationship between conformal transducers and plausibility contours. Further works that inspect such a relationship are [16, 44–46, 54].

One of the main concerns in the conformal literature is to identify function  $\pi$ , taking values in  $[0, 1]$ , such that  $\pi(Y_{n+1}, \mathbf{Y}^n)$  is stochastically no smaller than  $\text{Unif}(0, 1)$  under any exchangeable  $P \in \Delta_{\mathbb{Y}}$ . Also compare this to classical hypothesis testing in statistics, where the p-values under a true null hypothesis are distributed uniformly. Vovk, Gammerman, and Shafer [68, Theorem 11.1] show that the conformal transducer returned by Algorithm 1 satisfies this requirement. Suppose now that function  $\pi$  also satisfies

$$\sup_{\tilde{y} \in \mathbb{Y}} \pi(\tilde{y}, \mathbf{y}^n) = 1, \quad \text{for all } \mathbf{y}^n \in \mathbb{Y}^n. \quad (6)$$

This property holds quite generally for conformal prediction in continuous-data problems: if  $\tilde{y}$  is a point at which the minimum of  $\tilde{y} \mapsto \Psi(\mathbf{y}^n, \tilde{y})$  is achieved, then  $\pi(\tilde{y}, \mathbf{y}^n) = 1$  and (6) holds. For discrete-data problems, see Cella and Martin [19]. Notice that consonance can also be satisfied with a transformation of



the conformal transducer  $\pi$ . For example, in Cella and Martin [19, Section 7], the authors suggest the following two adjusted conformal transducers:  $\pi'(y, \mathbf{y}^n) = \pi(y, \mathbf{y}^n) / \sup_y \pi(y, \mathbf{y}^n)$ , for all  $y \in \mathbb{Y}$ , and

$$\pi''(y, \mathbf{y}^n) = \begin{cases} 1 & \text{if } y \in \operatorname{argsup}_{y \in \mathbb{Y}} \pi(y, \mathbf{y}^n) \\ \pi(y, \mathbf{y}^n) & \text{otherwise} \end{cases},$$

with the latter being more efficient. From (6), we can define a (predictive) upper probability as follows<sup>6</sup>

$$\bar{\Pi}_{\mathbf{y}^n}(A) = \sup_{\tilde{y} \in A} \pi(\tilde{y}, \mathbf{y}^n), \quad A \in \Sigma_{\mathbb{Y}}. \quad (7)$$

It is immediate to see how the upper probability in (7) is a consonant plausibility function, as introduced in Section 2.2. Conformal transducer  $\pi$  corresponds to a plausibility contour, and it fully determines the upper and lower probabilities of the elements of  $\Sigma_{\mathbb{Y}}$  via (7) and the duality property of lower probabilities.

We now discuss some interesting properties of upper probability  $\bar{\Pi}_{\mathbf{y}^n}$ , namely that it is coherent, supremum preserving, and tropically finitely additive.

Recall that the tropical semiring is the 3-tuple  $(\mathbb{R} \cup \{\infty\}, \oplus, \otimes)$ , where, for all  $a, b \in \mathbb{R} \cup \{\infty\}$ ,  $a \oplus b = \max\{a, b\}$ , and  $a \otimes b = a + b$  [49]. As a consequence, we can write that, for all  $A, B \in \Sigma_{\mathbb{Y}}$ ,  $\bar{\Pi}_{\mathbf{y}^n}(A) \oplus \bar{\Pi}_{\mathbf{y}^n}(B) = \max\{\bar{\Pi}_{\mathbf{y}^n}(A), \bar{\Pi}_{\mathbf{y}^n}(B)\}$  and  $\bar{\Pi}_{\mathbf{y}^n}(A) \otimes \bar{\Pi}_{\mathbf{y}^n}(B) = \bar{\Pi}_{\mathbf{y}^n}(A) + \bar{\Pi}_{\mathbf{y}^n}(B)$ .

Recall also that a generic upper probability  $\bar{P}$  is supremum preserving if  $\bar{P}(\cup_{A \in \mathcal{A}} A) = \sup_{A \in \mathcal{A}} \bar{P}(A)$ , for any possible collection  $\mathcal{A} \subseteq \Sigma_{\mathbb{Y}}$  [29].

Finally, we say that a generic upper probability  $\bar{P}$  is *coherent à la Walley* [69, Section 2.5] if its dual lower probability  $\underline{P}$  is such that, given an arbitrary collection  $\mathcal{K}$  of bounded random variables on  $\mathbb{Y}$ , we have that  $\sup[\sum_{j=1}^n (X_j - \underline{P}(X_j)) - m(X_0 - \underline{P}(X_0))] \geq 0$ , whenever  $m, n \in \mathbb{Z}_+$ , and  $X_0, X_1, \dots, X_n$  (not necessarily distinct) are in  $\mathcal{K}$ .

**Lemma 3.1** (Properties of a Consonant Upper Probability). *Let  $\mathcal{A} \subseteq \Sigma_{\mathbb{Y}}$  be a generic collection of subsets of  $\mathbb{Y}$ . Upper probability  $\bar{\Pi}_{\mathbf{y}^n}$  defined in (7) is supremum preserving and coherent à la Walley. In addition,  $\bar{\Pi}_{\mathbf{y}^n}$  is tropically finitely additive, that is,*

$$|\mathcal{A}| < \infty \implies \bar{\Pi}_{\mathbf{y}^n} \left( \bigcup_{A \in \mathcal{A}} A \right) = \bigoplus_{A \in \mathcal{A}} \bar{\Pi}_{\mathbf{y}^n}(A).$$

*Proof.* Upper probability  $\bar{\Pi}_{\mathbf{y}^n}$  is supremum preserving and coherent à la Walley by de Cooman and Aeyels [29]

<sup>6</sup>Let us note in passing that, when defining upper probability on the basis of the plausibility contour, we turn ordinal (order) information into cardinal information.

and Augustin, Coolen, De Cooman, and Troffaes [5, Section 4.6.1]. Tropical finite additivity follows from  $\bar{\Pi}_{\mathbf{y}^n}$  being supremum preserving, having assumed  $|\mathcal{A}| < \infty$ , and the definition of additivity in the tropical semiring. It can also be derived immediately from Lemma 2.1, or the “maxitivity” of possibility theory [35].  $\square$

Unlike classical finite additivity, tropical finite additivity does not require the events in  $\mathcal{A}$  to be disjoint. The practical implications of Lemmas 2.1 and 3.1 – especially the algebraic properties and the tropical additivity of consonant plausibility functions – will be the subject of future work.

**Remark 3.1.** Let us pause here to discuss a slight notational abuse that we perpetrate throughout the paper. When we say that “conformal transducer  $\pi$  corresponds to a plausibility contour”, we mean that conformal transducer  $\pi(\cdot, \mathbf{y}^n)$  corresponds to a plausibility contour  $\pi(\cdot)$ , for all  $\mathbf{y}^n \in \mathbb{Y}^n$ .

From a practical point of view, adding consonance to conformal prediction (that is, requiring (6) and (7) to hold) creates no new computational challenges. The standard use of Algorithm 1’s output is to extract the prediction region  $\mathcal{R}_\alpha(\mathbf{y}^n)$  which is just the collection of all  $\tilde{y}$ ’s such that  $\pi(\tilde{y}, \mathbf{y}^n)$  exceeds  $\alpha$ .

With the addition of consonance, Cella and Martin [20] recommend two additional summaries. First, at least in low-dimensional problems, a plot of  $\tilde{y} \mapsto \pi(\tilde{y}, \mathbf{y}^n)$  to give a visual assessment of the information available in the data  $\mathbf{y}^n$  regarding  $Y_{n+1}$ , similar to the Bayesian posterior predictive density function [20, Section 6]. Second, for any  $A \in \Sigma_{\mathbb{Y}}$ , the (prediction) upper probability at  $A$  can be approximated as

$$\bar{\Pi}_{\mathbf{y}^n}(A) \approx \max_{\tilde{y} \text{ on the grid and in } A} \pi(\tilde{y}, \mathbf{y}^n).$$

Consonance also induces several practically relevant properties that are inspected in Cella and Martin [20, Section 4.2] and in Augustin, Coolen, De Cooman, and Troffaes [5, Chapter 4].

#### 4. IMPRECISE HIGHEST DENSITY REGIONS

This section introduces the concept of Imprecise Highest Density Region (IHDR), a subset of  $\mathbb{Y}$  obtained from a methodology that employs techniques from the imprecise probability literature [26]. Consider again a process  $Y_1, Y_2, \dots$  of random elements taking values in  $\mathbb{Y}$ .

Like in the case of conformal prediction, the goal is to predict  $Y_{n+1}$  using a method that is valid. To this end, an IHDR proceeds from a credal set  $\mathcal{P}_{\text{pred}}$  of candidate distributions on  $\mathbb{Y}$ , which is supposed to be reliable in the sense that it contains the true distribution.<sup>7</sup>

<sup>7</sup>Subscript “pred” denotes the fact that, in the present paper, we are interested in a credal set of predictive distributions derived from the data  $\mathbf{y}^n$  at hand via some procedure, e.g. generalized Bayes’ or geometric updating [42].

**Definition 4.1** (Imprecise Highest Density Region [26]). Let  $\mathcal{P}_{\text{pred}}$  be a credal set of distributions on  $\mathbb{Y}$ , and  $\underline{P}$  be its lower probability. Call  $Y_{n+1}$  a random quantity that takes values in  $\mathbb{Y}$ , and  $\alpha$  any value in  $[0, 1]$ . Then, set  $\text{IR}_\alpha \subseteq \mathbb{Y}$  is called a  $(1 - \alpha)$ -*Imprecise Highest Density Region* (IHDR) if

$$\underline{P}[Y_{n+1} \in \text{IR}_\alpha] = 1 - \alpha$$

and  $\int_{\text{IR}_\alpha} dy$  is a minimum. If  $\mathbb{Y}$  is at most countable, we replace  $\int_{\text{IR}_\alpha} dy$  with  $|\text{IR}_\alpha|$ .

As a consequence of Definition 4.1, we have that  $\text{IR}_\alpha$  is the smallest subset of  $\mathbb{Y}$  such that

$$P[Y_{n+1} \in \text{IR}_\alpha] \geq 1 - \alpha, \quad \forall P \in \mathcal{P}_{\text{pred}}. \quad (8)$$

Here lies the appeal of the IHDR concept. Contrary to the guarantee in (2) though – which holds for all the possible exchangeable distributions  $P$  on  $\mathbb{Y}$  – the one in (8) holds only for the distributions that are plausible for the analysis at hand, that is, for all those in  $\mathcal{P}_{\text{pred}}$ .

**Example.** Let us now give an illustrating example of how to derive an IHDR. Credal set  $\mathcal{P}_{\text{pred}}$  may originate from different sources. One instance is robust Bayesian inference [8, 18, 33], where  $\mathcal{P}_{\text{pred}}$  corresponds to the (posterior) predictive credal set induced by a prior credal set. Suppose that we observe the first  $n$  elements of the process  $\{Y_k\}_{k \in \mathbb{N}}$ . A traditional Bayesian learner would first specify a likelihood for the observations, and then elicit a prior on the parameter of the likelihood distribution. Imagine that the  $Y_k$ 's are distributed according to a Poisson distribution  $\text{Pois}(\lambda)$ , so that the conjugate prior for the rate parameter  $\lambda$  is a Gamma distribution. Suppose then that prior information is available but comes from different sources or experiments. For instance, assuming  $\lambda \sim \text{Gamma}(a_j, b_j)$  may align with experiments on one group  $j \in \{1, \dots, J\}$ . Call  $P_j$  the probability measure whose pdf is  $\text{Gamma}(a_j, b_j)$ ,  $j \in \{1, \dots, J\}$ . In this case, a prior finitely generated credal set (FGCS)

$$\mathcal{P}_{\text{prior}} = \left\{ Q : Q = \sum_{j=1}^J \beta_j \cdot P_j, \beta_j \geq 0, \sum_{j=1}^J \beta_j = 1 \right\}$$

is a natural choice. This corresponds to the Bayesian sensitivity analysis (BSA) approach to inference [7, 38, 41], [69, Section 5.9]. The extreme elements of  $\mathcal{P}_{\text{prior}}$  are  $\text{ex}\mathcal{P}_{\text{prior}} = \{P_1, \dots, P_J\}$ . By computing their posterior, we obtain the extreme elements of the posterior FGCS,  $\text{ex}\mathcal{P}_{\text{post}} \subseteq \{P_1^{\text{post}}, \dots, P_J^{\text{post}}\}$  (continuing our example, the elements of  $\text{ex}\mathcal{P}_{\text{post}}$  are Gamma distributions with updated shape and rate parameters).

The goal now is to predict  $Y_{n+1}$  using a valid or reliable method. To do so, first the agent derives the extrema of the predictive FGCS,  $\text{ex}\mathcal{P}_{\text{pred}} \subseteq \{P_1^{\text{pred}}, \dots, P_J^{\text{pred}}\}$ . These are the probability measures whose pdf's are

computed as  $p_j^{\text{pred}}(y_{n+1}|y_1, \dots, y_n) = \int_0^\infty \ell(y_{n+1}|\lambda) \cdot p_j^{\text{post}}(\lambda|y_1, \dots, y_n) d\lambda$ ,  $j \in \{1, \dots, J\}$ , where  $\ell(y_{n+1}|\lambda)$  is our Poisson likelihood, and  $p_j^{\text{post}}$  is the pdf of the posterior measure  $P_j^{\text{post}}$ ,  $j \in \{1, \dots, J\}$ . By well-known results in Bayesian inference [43], each such a predictive distribution is a Negative Binomial. The predictive FGCS is given by  $\mathcal{P}_{\text{pred}} = \text{Conv}(\text{ex}\mathcal{P}_{\text{pred}}) \subseteq \Delta_{\mathbb{Y}}$ , where  $\text{Conv}(\cdot)$  denotes the convex hull operator. By Caprio et al. [18, Proposition 3], we know that  $\underline{P}(A) = \inf_{P \in \text{ex}\mathcal{P}_{\text{pred}}} P(A) = \inf_{P \in \mathcal{P}_{\text{pred}}} P(A)$ , for all  $A \in \Sigma_{\mathbb{Y}}$ . Hence, to find our desired IHDR, we only have to focus on searching for the smallest region  $\text{IR}_\alpha$  such that (8) is satisfied for all  $P \in \text{ex}\mathcal{P}_{\text{pred}}$ . Notice that the complexity of computation is linear in  $|\text{ex}\mathcal{P}_{\text{pred}}|$ , that is, in the number of extreme elements of  $\mathcal{P}_{\text{pred}}$  [5, Section 9.2.1].  $\blacktriangle$

Oftentimes, in Imprecise Probability theory, it is posited that

$$\mathbf{P} \in \mathcal{P}_{\text{pred}}, \quad (9)$$

where – with an abuse of notation –  $\mathbf{P}$  has to be understood as the marginal probability measure on  $Y_{n+1}$  derived from the true joint distribution of  $Y^\infty$ .<sup>8</sup> This assumption is not overly strong, as it may appear at first glance. The credal set  $\mathcal{P}_{\text{pred}}$  will generally be wider, the higher the uncertainty around the true  $\mathbf{P}$ . There is also an active literature that studies statistical tests which check whether credal sets are calibrated à la (9) [2, 24, 39, 47, 56].

Assumption (9) is not needed for  $\mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$ . Indeed, Martin [51] shows that  $\overline{\Pi}_{\mathbf{y}^n}$  is the minimal outer consonant approximation of the true data generating process  $\mathbf{P}$ . This means that  $\overline{\Pi}_{\mathbf{y}^n}$  is the narrowest upper bound for the true distribution, that also satisfies the consonance assumption. We can then conclude – although it is not explicitly shown in their work – that credal region  $\mathcal{M}(\overline{\Pi})$  contains the true data generating process.

**Possible criticism to a generic IHDR.** It requires the agent to come up with  $\mathcal{P}_{\text{pred}}$ . To counter this objection, we point out how modeling choices are crucial to model-based inference. A priori, we cannot say that model-based approaches are to be preferred to model-free ones, or vice versa. The choice will depend on the available information. Let us also remark that conformal prediction is not a completely model-free approach either, since the user is required to specify a non-conformity measure  $\Psi$ , which has a strong influence on the size of the conformal prediction region.

<sup>8</sup>Throughout the paper, we denote by  $\mathbf{P}$  both the true data generating process on  $Y^\infty$  and its marginal on  $Y_{n+1}$ . This abuse of notation allows us to better communicate our results, and does not induce confusion, as it is obvious from the context whether we are referring to the joint or the marginal distributions.

## 5. BUILDING THE EQUIVALENT PREDICTION REGION

In this section, we present the main contribution of our work. We show that – for a given nonconformity measure  $\Psi$  used to derive the conformal transducer  $\pi$  – credal set  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  can be used to obtain a prediction region that is equivalent to the CPR, and that retains the same uniform guarantees.

Suppose (i) that the process  $Y_1, Y_2, \dots$  is exchangeable and governed by a unique distribution  $\mathbf{P}$  (so that distribution shifts are not allowed), and (ii) that the conformal transducer  $\pi$  is consonant, that is,  $\sup_{\tilde{y} \in \mathbb{Y}} \pi(\tilde{y}, \mathbf{y}^n) = 1$ .

Let  $\bar{\Pi}_{\mathbf{y}^n}(A) = \sup_{\tilde{y} \in A} \pi(\tilde{y}, \mathbf{y}^n)$ , for all  $A \in \Sigma_{\mathbb{Y}}$ , be the upper probability induced by  $\pi$ . Let also  $\underline{\Pi}_{\mathbf{y}^n}(A) = 1 - \bar{\Pi}_{\mathbf{y}^n}(A^c)$ , for all  $A \in \Sigma_{\mathbb{Y}}$ , be the lower probability dual to  $\bar{\Pi}_{\mathbf{y}^n}$ . Let  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  be the credal set induced by  $\bar{\Pi}_{\mathbf{y}^n}$ , that is,

$$\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n}) := \{P : P(A) \leq \bar{\Pi}_{\mathbf{y}^n}(A), \forall A \in \Sigma_{\mathbb{Y}}\}.$$

To see that  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  is indeed a credal set, we refer the reader to Marinacci and Montrucchio [50] and Walley [69]. We denote by  $\text{IR}_{\alpha}^{\mathcal{M}}, \alpha \in [0, 1]$ , the IHDR associated with the credal set  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  induced by the conformal procedure, that is,  $\text{IR}_{\alpha}^{\mathcal{M}} \subseteq \mathbb{Y}$  such that  $\underline{\Pi}_{\mathbf{y}^n}(\text{IR}_{\alpha}^{\mathcal{M}}) = 1 - \alpha$  and  $\int_{\text{IR}_{\alpha}^{\mathcal{M}}} dy$  is minimal. Then, we have the following.

**Proposition 5.1** (IHDR Corresponds to the Conformal Prediction Region). *For any  $\alpha \in [0, 1]$  and any  $n \in \mathbb{N}$ , the following is true*

$$\text{IR}_{\alpha}^{\mathcal{M}} = \mathcal{R}_{\alpha}(\mathbf{y}^n).$$

*Proof.* Consider the function  $\gamma : \mathbb{Y} \rightarrow [0, 1]$ ,

$$y \mapsto \gamma(y) = \begin{cases} \pi(y) & \text{if } \pi(y) \leq 0.5 \\ 1 - \pi(y) & \text{if } \pi(y) > 0.5 \end{cases}.$$

It is easy to see that  $\gamma(y) \leq \pi(y)$ , for all  $y \in \mathbb{Y}$ . In addition, by the consonance property of  $\pi$ , there exists  $\tilde{y} \in \mathbb{Y}$  such that  $\gamma(\tilde{y}) = 0$ . In turn, we have that  $[\gamma, \pi]$  is a cloud [5, Definition 4.6]. Neumaier’s probabilistic constraint on clouds [57], [5, Equation (4.9)] gives us the following

$$\begin{aligned} P[Y_{n+1} \in \mathcal{R}_{\alpha}(\mathbf{y}^n)] &= P[Y_{n+1} \in \{y \in \mathbb{Y} : \pi(y, \mathbf{y}^n) > \alpha\}] \\ &\geq 1 - \alpha = \underline{\Pi}_{\mathbf{y}^n}(Y_{n+1} \in \text{IR}_{\alpha}^{\mathcal{M}}), \end{aligned}$$

for all  $P \in \mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$ . In turn, we have

$$\begin{aligned} \underline{\Pi}_{\mathbf{y}^n}[Y_{n+1} \in \mathcal{R}_{\alpha}(\mathbf{y}^n)] &= \underline{\Pi}_{\mathbf{y}^n}[Y_{n+1} \in \{y \in \mathbb{Y} : \pi(y, \mathbf{y}^n) > \alpha\}] \\ &\geq 1 - \alpha = \underline{\Pi}_{\mathbf{y}^n}(Y_{n+1} \in \text{IR}_{\alpha}^{\mathcal{M}}). \end{aligned} \tag{10}$$

By the monotonicity of lower probabilities (see Section 2.2), then, we can conclude that  $\text{IR}_{\alpha}^{\mathcal{M}} \subseteq \mathcal{R}_{\alpha}(\mathbf{y}^n)$ . But, by the definition of Conformal Prediction Region,  $\text{IR}_{\alpha}^{\mathcal{M}}$  cannot be strictly included in  $\mathcal{R}_{\alpha}(\mathbf{y}^n)$ . In turn, this implies that  $\text{IR}_{\alpha}^{\mathcal{M}} = \mathcal{R}_{\alpha}(\mathbf{y}^n)$ , as desired.  $\square$

Proposition 5.1 ensures us of the following. Once we derive the conformal transducer  $\pi$ , if we take the imprecise probabilistic route based on (predictive) credal set  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$ , we are able to obtain a prediction region (an IHDR)  $\text{IR}_{\alpha}^{\mathcal{M}}$ . The latter corresponds to the Conformal Prediction Region  $\mathcal{R}_{\alpha}(\mathbf{y}^n)$ , for all possible choices of significance level  $\alpha \in [0, 1]$ . We also point out how the proof of Proposition 5.1 links Conformal Prediction to the IP concept of a cloud, another profound connection between the two literatures.

We note in passing that another proof of Proposition 5.1 exists, and is based on Martin [51, Page 25] as follows. Thanks to the consonance of  $\pi$ , we have that

$$\begin{aligned} \mathcal{R}_{\alpha}(\mathbf{y}^n) &= \{y \in \mathbb{Y} : \pi(y, \mathbf{y}^n) > \alpha\} \\ &= \bigcap \left\{ A \in \Sigma_{\mathbb{Y}} : \underline{\Pi}_{\mathbf{y}^n}(A) \geq 1 - \alpha \right\} \\ &= \text{IR}_{\alpha}^{\mathcal{M}}, \end{aligned} \tag{11}$$

for all  $\alpha \in [0, 1]$ .

A natural question at this point is whether  $\text{IR}_{\alpha}^{\mathcal{M}}$  enjoys the same probabilistic guarantees as  $\mathcal{R}_{\alpha}(\mathbf{y}^n)$ . The following proposition and corollary provide us with a positive answer.

**Proposition 5.2** (Elements of the IHDR). *For all  $\alpha \in (0, 1]$ ,  $P \in \mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  if and only if  $P[Y_{n+1} \in \{y \in \mathbb{Y} : \pi(y, \mathbf{y}^n) > \alpha\}] \geq 1 - \alpha$ .*<sup>9</sup>

*Proof.* Immediate from Augustin, Coolen, De Cooman, and Troffaes [5, Proposition 4.1].  $\square$

**Corollary 5.1** (Probabilistic Guarantees of the IHDR). *For all  $\alpha \in [0, 1]$ , the following holds uniformly in  $P$  and in  $n$ ,*<sup>10</sup>

$$P[Y_{n+1} \in \text{IR}_{\alpha}^{\mathcal{M}}] \geq 1 - \alpha.$$

*Proof.* Pick any  $\alpha \in (0, 1]$  and any  $n \in \mathbb{N}$ . Then, by Vovk, Gammerman, and Shafer [68, Theorem 11.1] and Cella and Martin [20, Section 4.2], we have that

$$P[Y_{n+1} \in \{y \in \mathbb{Y} : \pi(y, \mathbf{y}^n) > \alpha\}] \geq 1 - \alpha, \tag{12}$$

uniformly in  $P$ , that is, for all exchangeable distribution  $P$  on  $\mathbb{Y}$ . By Proposition 5.2, then, all the (exchangeable) probabilities  $P \in \Delta_{\mathbb{Y}}$  satisfying (12) belong to

<sup>9</sup>The set  $\{y \in \mathbb{Y} : \pi(y, \mathbf{y}^n) > \alpha\}$  is called a *strong  $\alpha$ -cut* of  $\pi$ . If we replace the strong inequality with a weak one, the set is called a *regular  $\alpha$ -cut* of  $\pi$  [5, Section 4.6].

<sup>10</sup>Here too “uniformly in  $P$ ” has to be understood as “for all exchangeable  $P$  on  $\mathbb{Y}$ ”.

the credal set  $\mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$ . We also have that, by definition,  $\underline{\Pi}_{\mathbf{y}^n}[Y_{n+1} \in \text{IR}_\alpha^{\mathcal{M}}] = 1 - \alpha$ , which implies that  $P[Y_{n+1} \in \text{IR}_\alpha^{\mathcal{M}}] \geq 1 - \alpha$ , for all  $P \in \mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$ .

For  $\alpha = 0$ , the statement follows from (11) and Vovk, Gammerman, and Shafer [68, Theorem 11.1]. Notice how we could have directly used Vovk, Gammerman, and Shafer [68, Theorem 11.1] to prove the whole result (without splitting in the  $\alpha \in (0, 1]$  and  $\alpha = 0$  cases), but we preferred to state it as a corollary to Proposition 5.2 to better study the relationship between conformal prediction and imprecise probabilities.  $\square$

By combining Proposition 5.1 and Corollary 5.1 together, we can conclude that an imprecise probabilistic approach delivers the same prediction region as classical Transductive Conformal Prediction, and it retains the same uniform probabilistic guarantees.

Let us also point out that Proposition 5.1 is an immediate consequence of Couso, Montes, and Gil [27, Theorem 2], where it is also shown that Corollary 5.1 holds for all  $P \in \mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$ . Our result generalizes it slightly, by showing that the statement actually holds for all exchangeable  $P$  on  $\mathbb{Y}$ .

**5.1. Varying the nonconformity measure  $\Psi$ .** Proposition 5.1 and Corollary 5.1 show that, *once we fix the non-conformity measure  $\Psi$* , an imprecise-probabilistic approach gives us a prediction region that is equivalent to the classical CPR. In the next paragraphs, we will study what happens when we change  $\Psi$ .

Notice that, for any  $n \in \mathbb{N}$  representing the cardinality of the (training) dataset, Conformal Prediction (CP) can be seen as a set-valued function

$$CP : [0, 1] \times \mathbb{Y}^n \times \mathcal{F} \rightarrow \Sigma_{\mathbb{Y}}, \quad (\alpha, \mathbf{y}^n, \Psi) \mapsto CP(\alpha, \mathbf{y}^n, \Psi) = \mathcal{R}_\alpha^\Psi(\mathbf{y}^n). \quad (13)$$

Here  $\mathcal{F} \subseteq \mathbb{R}^{\mathbb{Y}^{n+1}}$  denotes the set of all possible non-conformity measures,

$$\mathcal{F} := \{\Psi : \mathbb{Y}^n \times \mathbb{Y} \rightarrow \mathbb{R} : (\mathbf{y}_{-i}^{n+1}, y_i) \mapsto \psi_i(\mathbf{y}^{n+1}) \in \mathbb{R}\}, \quad (14)$$

and  $\mathcal{R}_\alpha^\Psi(\mathbf{y}^n) \equiv \mathcal{R}_\alpha(\mathbf{y}^n) \in \Sigma_{\mathbb{Y}}$  denotes the Conformal Prediction Region (in previous results we omitted the explicit reference to the choice of non-conformity measure  $\Psi$  for notational convenience).<sup>11</sup>

**Proposition 5.3** (Refining a CPR). *Pick any  $\alpha \in [0, 1]$  and any  $\Psi \in \mathcal{F}$ . Then, there exists  $\Psi' \neq \Psi$  such that  $\text{IR}_\alpha^{\mathcal{M}} = \mathcal{R}_\alpha^\Psi(\mathbf{y}^n) \supseteq \mathcal{R}_\alpha^{\Psi'}(\mathbf{y}^n)$ , and the inclusion is strict for some value of  $\alpha$ . In addition,  $P[Y_{n+1} \in \mathcal{R}_\alpha^{\Psi'}(\mathbf{y}^n)] \geq 1 - \alpha$ , uniformly in  $P$  and in  $n$ .*<sup>12</sup>

<sup>11</sup>We make the implicit assumption that  $\mathcal{R}_\alpha^\Psi(\mathbf{y}^n) \in \Sigma_{\mathbb{Y}}$ . To relax it, simply substitute  $\Sigma_{\mathbb{Y}}$  with  $2^{\mathbb{Y}}$ .

<sup>12</sup>Here too “uniformly” has to be understood in the sense of “for all the exchangeable probability measures on  $\mathbb{Y}$ ”.

*Proof.* The first part is an immediate consequence of Vovk, Gammerman, and Shafer [68, Theorem 2.10]. There, the authors provide a proof by construction, that can be used to derive  $\mathcal{R}_\alpha^{\Psi'}(\mathbf{y}^n)$ , and in turn the new non-conformity measure  $\Psi'$ . The uniform probabilistic guarantee follows from Vovk, Gammerman, and Shafer [68, Theorem 1.11].  $\square$

Proposition 5.3 tells us that, while it is true that the CPR  $\mathcal{R}_\alpha^\Psi(\mathbf{y}^n)$  corresponds to the IHDR  $\text{IR}_\alpha^{\mathcal{M}}$  associated with the conformally-built credal set  $\mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$  for a fixed non-conformity measure  $\Psi$ , we can always find a new non-conformity measure  $\Psi'$  that further improves on  $\text{IR}_\alpha^{\mathcal{M}} = \mathcal{R}_\alpha^\Psi(\mathbf{y}^n)$ . In principle, different nonconformity measures may reduce the size of the credal set; in addition, further shrinking the set may eventually require more data.

**5.2. Discussion.** Before concluding this section, let us add a discussion of the results we presented so far. As pointed out by Ryan Martin and coauthors in their recent works (see e.g. [21, 51, 52], and especially [19]), there seems to be a sense in which the conformal prediction procedure assumes a vacuous prior, which captures the concept of maximal (prior) ambiguity. That is, conformal prediction appears to be akin to a Bayesian Sensitivity Analysis (BSA) procedure [7] whose prior credal set is the whole space of parameter probabilities. In BSA, though, the posterior parameter probabilities (and hence, also the predictive probabilities on  $\mathbb{Y}$ ) obtained from this prior class are again vacuous: no finite sample is enough to annihilate a sufficiently extreme prior belief [59], [18, Appendix B]. On the contrary, conformal prediction is able to derive a non-vacuous predictive credal set  $\mathcal{M}(\overline{\Pi}_{\mathbf{y}^n}) \subsetneq \Delta_{\mathbb{Y}}$ .

This seems to suggest that there is something peculiar about conformal prediction’s (possibilistic) quantification of uncertainty, that allows it to handle the vacuous prior case more efficiently than other imprecise probability frameworks like BSA. We conjecture that the main reason for this is that CP enjoys *half-coherence* [51, Section 5.2.2], while departing from the classical imprecise probabilistic notion of (*full*) *coherence* [69, Section 2.5]. The latter entails extreme conservatism in the vacuous prior case: as we pointed out before, a vacuous prior set produces a vacuous predictive set. The gap between conformal prediction, a powerful tool that efficiently handles vacuous prior information, and the theoretically well-developed imprecise-probabilistic literature – together with how to strike a balance between these two paradigms – remains a largely unexplored research area, and will be the object of future work.

Another issue worth discussing is that the consonance assumption induces a “distortion” in the shape of the credal set  $\mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$  induced by the conformal transducer  $\pi$ . With this, we mean that  $\mathcal{M}(\overline{\Pi}_{\mathbf{y}^n})$  is “pushed” towards



the boundaries of the space of probabilities on  $\mathbb{Y}$ , and that the true data generating process is not its centroid. To see this, consider the following example.

Suppose that  $\mathbb{Y} = \{A, B, C\}$ , and that we observe 20  $A$ 's, 30  $B$ 's and 50  $C$ 's. As a consequence of these observations, we have that vector  $p^{\text{emp}} = (0.2, 0.3, 0.5)^\top$  represents the empirical pmf. Suppose that the learner chooses non-conformity measure  $\psi_i(\mathbf{y}^{n+1}) = 1 - p^{\text{emp}}(y_i = k)$ , for all  $i \in \{1, \dots, n+1\}$ . Put  $T_i = \psi_i(\mathbf{y}^{n+1})$ , for all  $i \in \{1, \dots, n+1\}$ . We have that  $\psi_i(\mathbf{y}^{n+1}) = 0.8$  for the  $y_i$ 's that are equal to  $A$ ,  $\psi_i(\mathbf{y}^{n+1}) = 0.7$  for the  $y_i$ 's that are equal to  $B$ , and  $\psi_i(\mathbf{y}^{n+1}) = 0.5$  for the  $y_i$ 's that are equal to  $C$ . In turn, we derive the conformal transducer  $\pi$  as  $\pi(\tilde{y} = k, \mathbf{y}^n) = \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{1}[T_i \geq T_{n+1}]$ , for all  $k \in \mathbb{Y}$ , so that

$$\pi(\tilde{y}, \mathbf{y}^n) = \begin{cases} \frac{21}{101} & \tilde{y} = A \\ \frac{51}{101} & \tilde{y} = B \\ 1 & \tilde{y} = C \end{cases}.$$

As we can see, our choice on non-conformity measure  $\psi_i(\mathbf{y}^{n+1})$ ,  $i \in \{1, \dots, n+1\}$ , ensures consonance. The values of the upper probabilities of the elements of  $\Sigma_{\mathbb{Y}} = 2^{\mathbb{Y}}$  can be found in Table 1. To derive them, we simply computed  $\bar{\Pi}_{\mathbf{y}^n}(A) = \sup_{\tilde{y} \in A} \pi(\tilde{y}, \mathbf{y}^n)$  and  $\underline{\Pi}_{\mathbf{y}^n}(A) = 1 - \bar{\Pi}_{\mathbf{y}^n}(A^c)$ , for all  $A \in 2^{\mathbb{Y}}$ . Of course,  $\bar{\Pi}_{\mathbf{y}^n}(\emptyset) = \underline{\Pi}_{\mathbf{y}^n}(\emptyset) = 0$ , and  $\bar{\Pi}_{\mathbf{y}^n}(\mathbb{Y}) = \underline{\Pi}_{\mathbf{y}^n}(\mathbb{Y}) = 1$ .

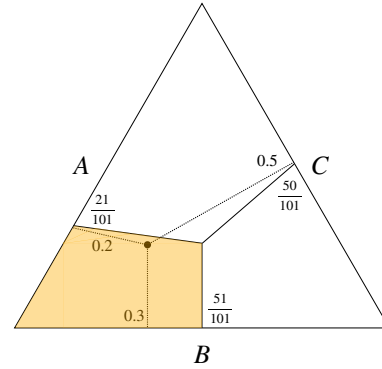
	$\underline{\Pi}_{\mathbf{y}^n}$	$\bar{\Pi}_{\mathbf{y}^n}$
$\{A\}$	0	21/101
$\{B\}$	0	51/101
$\{C\}$	50/101	1
$\{A, B\}$	0	51/101
$\{B, C\}$	80/101	1
$\{A, C\}$	50/101	1

**Table 1.** Values that the lower and upper probabilities assign to the elements of  $\Sigma_{\mathbb{Y}} = 2^{\mathbb{Y}}$ .

We depict the credal set  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  associated with the values of upper probability  $\bar{\Pi}_{\mathbf{y}^n}$  in Figure 2. As we can see, the consonance assumption “pushes”  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  towards the boundary of the unit simplex, and – in this particular case – towards the vertex associated with event  $C$ . Notice that this is not a mere consequence of the non-conformity measure that we chose in this example, but rather a characteristic ensuing from enforcing consonance. Whether this deformation has an effect on the size of the IHDR is an open question. What is apparent, though, is that it makes it difficult to gain insights around the true data generating process, which is not the centroid of  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  [55].

In addition, if we measure the aleatoric uncertainty  $\text{AU}[\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})]$  – the type of uncertainty that is inherent to the data generating process, and as such, irreducible – associated with  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  via the lower entropy  $\underline{H}(P) = \inf_{P \in \mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})} H(P)$  [1], we have that  $\text{AU}[\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})] = \underline{H}(P) = 0$ , because  $\delta_{\{C\}} \in \mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$ .<sup>13</sup> While this is a feature of the measure chosen for the AU, it still seems to highlight a shortcoming of the consonance assumption.

Finally, we point out how proving or disproving a fundamental incompatibility between consonance and minimality (i.e. whether consonance might prevent the credal set from being minimal in size) is a highly non-trivial question, which we plan to inspect in future research.



**Figure 2.** Visual representation of  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  in our example. As we can see, it is “pushed” towards the boundary of the unit simplex. We also depicted  $p^{\text{emp}} = (0.2, 0.3, 0.5)^\top$  as a black dot.

## 6. CONCLUSION

In the present work we study the conformal construction of credal sets. We show that the IHDR generated by the credal set induced by the conformal transducer  $\pi$  is equivalent to the classical CPR, and it retains the same uniform probabilistic guarantee. Our results are devised in the general effort of framing conformal prediction as a model-free imprecise probabilistic method. This continues the undertaking of Cella and Martin [20], who show that CP is related to credal sets and inferential models [53]; we prove that it is also linked to clouds.

Notice that throughout the paper we tacitly operated under the *closed world assumption* [15], that is, we assumed that the support of the true data generating process does not change, and it is equal to  $\mathbb{Y}$ . In the future, we plan to forego this requirement.

We also intend to extend the conformal construction of credal sets to the generalized CP frameworks studied in Barber, Candès, Ramdas, and Tibshirani [6] and Gibbs, Cherian, and Candès [40].

<sup>13</sup>Here  $\delta_{\cdot}$  denotes the Dirac measure.

Furthermore, we plan to (i) study in greater detail the empirical implications of our results, and (ii) inspect whether set-function  $CP$  in (13) is continuous (in some sense) in  $\Psi$ . If that were true, it would be easier to find  $\Psi^*$  that minimizes  $\int_{\mathcal{R}_\alpha^\Psi(\mathbf{y}^n)} dy$ , that is, the non-conformity measure  $\Psi^*$  that induces the smallest possible CPR which still satisfies the  $1 - \alpha$  uniform probabilistic guarantee.

In addition – building on our considerations in Section 5.2 – we plan to study the geometry of the conformally-built credal set  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$ , e.g. in the spirit of Cuzzolin [28]. We are particularly interested in (a) its position within the simplex, and (b) the distance between the true data generating process  $\mathbf{P}$  and any notion of the centroid of  $\mathcal{M}(\bar{\Pi}_{\mathbf{y}^n})$  proposed by Miranda and Montes [55].

Finally, we would like to further our investigation on the shortcomings of the consonance assumption. This is because [16] showed that there is at least a case of conformalized credal sets for classification problems in which such an assumption can be foregone.

#### ADDITIONAL AUTHOR INFORMATION

**Acknowledgements.** Michele Caprio was partially funded by the Army Research Office (ARO MURI W911NF2010080). He wishes to express his gratitude to Ryan Martin and Marianne Johnson for insightful discussions. Yusuf Sale is supported by the DAAD program Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research. All the authors wish to express their gratitude to four anonymous reviewers for their insightful comments.

#### REFERENCES

- [1] Joaquín Abellán, George Jiří Klir, and Serafín Moral. “Disaggregated total uncertainty measure for credal sets”. In: *International Journal of General Systems* 1.35 (2006), pp. 29–44.
- [2] Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. “Optimal testing for properties of distributions”. In: *Advances in Neural Information Processing Systems* 28 (2015).
- [3] Anastasios N. Angelopoulos, Rina Foygel Barber, and Stephen Bates. *Theoretical Foundations of Conformal Prediction*. 2024. arXiv: [2411.11824 \[math.ST\]](#).
- [4] Louis J. M. Aslett, Frank P. A. Coolen, and Jasper De Bock, eds. *Uncertainty in Engineering: Introduction to Methods and Applications*. Vol. 7. Springer-Briefs in Statistics. Cham, Switzerland: Springer, 2022.
- [5] Thomas Augustin, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes. *Introduction to imprecise probabilities*. Vol. 591. John Wiley & Sons, 2014.
- [6] Rina Foygel Barber, Emmanuel J Candès, Aaditya Ramdas, and Ryan J Tibshirani. “Conformal prediction beyond exchangeability”. In: *The Annals of Statistics* 51.2 (2023), pp. 816–845.
- [7] James Berger. *The robust Bayesian viewpoint (with discussion)*. *Robustness of Bayesian Analysis*, (J. Kadane, Ed.) 1984.
- [8] James O. Berger. “Robust Bayesian analysis: sensitivity to the prior”. In: *Journal of Statistical Planning and Inference* 25 (1990), pp. 303–328.
- [9] Michele Caprio. “Imprecise Markov Semigroups and their Ergodicity”. 2024. arXiv: [2405.00081](#).
- [10] Michele Caprio. *Optimal Transport for  $\epsilon$ -Contaminated Credal Sets*. 2025. arXiv: [2410.03267 \[stat.ML\]](#).
- [11] Michele Caprio and Ruobin Gong. “Dynamic precise and imprecise probability kinematics”. In: *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*. Ed. by Enrique Miranda, Ignacio Montes, Erik Quaeghebeur, and Barbara Vantaggi. Vol. 215. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 72–83.
- [12] Michele Caprio and Sayan Mukherjee. “Extended probabilities and their application to statistical inference”. In: *Understanding Information and Its Role as a Tool: In Memory of Mark Burgin — PART II: Information Realm: Information as a Tool for Exploring Reality*. World Scientific, 2023. Chap. 2.
- [13] Michele Caprio and Sayan Mukherjee. “Ergodic theorems for dynamic imprecise probability kinematics”. In: *International Journal of Approximate Reasoning* 152 (2023), pp. 325–343.
- [14] Michele Caprio, Yusuf Sale, Eyke Hüllermeier, and Insup Lee. “A Novel Bayes’ Theorem for Upper Probabilities”. In: *Epistemic Uncertainty in Artificial Intelligence*. Ed. by Fabio Cuzzolin and Maryam Sultana. Cham: Springer Nature Switzerland, 2024, pp. 1–12.
- [15] Michele Caprio and Teddy Seidenfeld. “Constriction for sets of probabilities”. In: *International Symposium on Imprecise Probability: Theories and Applications*. PMLR. 2023, pp. 84–95.
- [16] Michele Caprio, David Stutz, Shuo Li, and Arnaud Doucet. “Conformalized Credal Regions for Classification with Ambiguous Ground Truth”. In: *Transactions on Machine Learning Research* (2025).

- [17] Michele Caprio, Maryam Sultana, Eleni Elia, and Fabio Cuzzolin. “Credal Learning Theory”. In: *NeurIPS 2024* (2024).
- [18] Michele Caprio et al. “Credal Bayesian Deep Learning”. In: *Transactions on Machine Learning Research* (2024).
- [19] Leonardo Cella and Ryan Martin. “Valid inferential models for prediction in supervised learning problems”. In: *International Symposium on Imprecise Probability: Theories and Applications*. PMLR, 2021, pp. 72–82.
- [20] Leonardo Cella and Ryan Martin. “Validity, consonant plausibility measures, and conformal prediction”. In: *International Journal of Approximate Reasoning* 141 (2022), pp. 110–130.
- [21] Leonardo Cella and Ryan Martin. “Possibility-theoretic statistical inference offers performance and probativeness assurances”. In: *International Journal of Approximate Reasoning* 163 (2023), p. 109060.
- [22] Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci. “Ergodic theorems for lower probabilities”. In: *Proceedings of the American Mathematical Society* 144.8 (2016), pp. 3381–3396.
- [23] Simone Cerreia-Vioglio, Roberto Corrao, and Giacomo Lanzani. “Dynamic Opinion Aggregation: Long-Run Stability and Disagreement”. In: *The Review of Economic Studies* (2023), rdad072.
- [24] Siu Lun Chau, Antonin Schrab, Arthur Gretton, Dino Sejdinovic, and Krikamol Muandet. *Credal Two-Sample Tests of Epistemic Ignorance*. 2024. arXiv: [2410.12921](https://arxiv.org/abs/2410.12921) [stat.ML].
- [25] Gustave Choquet. “Theory of capacities”. In: *Annales de l’institut Fourier*. Vol. 5. 1954, pp. 131–295.
- [26] Franciscus Petrus Antonius Coolen. *Imprecise highest density regions related to intervals of measures*. 1992.
- [27] Inés Couso, Susana Montes, and Pedro Gil. “The Necessity Of The Strong  $\alpha$ -Cuts Of A Fuzzy Set”. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 09.02 (2001), pp. 249–262. DOI: [10 . 1142 / S0218488501000788](https://doi.org/10.1142/S0218488501000788).
- [28] Fabio Cuzzolin. *The geometry of uncertainty: the geometry of imprecise probabilities*. Springer Nature, 2020.
- [29] Gert de Cooman and Dirk Aeyels. “Supremum preserving upper probabilities”. In: *Information Sciences* 118.1 (1999), pp. 173–212.
- [30] Robert Denk, Michael Kupper, and Max Nendel. “Kolmogorov-type and general extension results for nonlinear expectations”. In: *Banach Journal of Mathematical Analysis* 12.3 (2018), pp. 515–540.
- [31] Robert Denk, Michael Kupper, and Max Nendel. “A semigroup approach to nonlinear Lévy processes”. In: *Stochastic Processes and their Applications* 130.3 (2020), pp. 1616–1642.
- [32] Thierry Denoeux. “A neural network classifier based on Dempster-Shafer theory”. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 30.2 (2000), pp. 131–150.
- [33] Lorraine DeRoberts and J. A. Hartigan. “Bayesian Inference Using Intervals of Measures”. In: *The Annals of Statistics* 9.2 (1981), pp. 235–244.
- [34] S. Destercke, D. Dubois, and E. Chojnacki. “Unifying practical uncertainty representations – I: Generalized p-boxes”. In: *International Journal of Approximate Reasoning* 49.3 (2008), pp. 649–663. DOI: [10.1016/j.ijar.2008.07.003](https://doi.org/10.1016/j.ijar.2008.07.003).
- [35] Didier Dubois, Hung T. Nguyen, and Henri Prade. “Possibility Theory, Probability and Fuzzy Sets Misunderstandings, Bridges and Gaps”. In: *Fundamentals of Fuzzy Sets*. Ed. by Didier Dubois and Henri Prade. Boston, MA: Springer US, 2000, pp. 343–438. DOI: [10.1007/978-1-4615-4429-6\\_8](https://doi.org/10.1007/978-1-4615-4429-6_8).
- [36] Didier Dubois and Henri Prade. “Consonant approximations of belief functions”. In: *International Journal of Approximate Reasoning* 4.5-6 (1990), pp. 419–449.
- [37] Souradeep Dutta et al. “Distributionally Robust Statistical Verification with Imprecise Neural Networks”. In: *Accepted to HSCC25* (2025).
- [38] Daniel Ellsberg. “Risk, ambiguity, and the Savage axioms”. In: *The quarterly journal of economics* 75.4 (1961), pp. 643–669.
- [39] Rui Gao, Liyan Xie, Yao Xie, and Huan Xu. “Robust hypothesis testing using Wasserstein uncertainty sets”. In: *Advances in Neural Information Processing Systems* 31 (2018).
- [40] Isaac Gibbs, John J Cherian, and Emmanuel J Candès. “Conformal prediction with conditional guarantees”. 2023. arXiv: [2305.12616](https://arxiv.org/abs/2305.12616).
- [41] Itzhak Gilboa and Massimo Marinacci. “Ambiguity and the Bayesian paradigm”. In: *Readings in formal epistemology: Sourcebook*. Springer, 2016, pp. 385–439.

- [42] Ruobin Gong and Xiao-Li Meng. “Judicious judgment meets unsettling updating: dilation, sure loss, and Simpson’s paradox”. In: *Statistical Science* 36.2 (2021), pp. 169–190.
- [43] Peter D. Hoff. *A First Course in Bayesian Statistical Methods*. New York : Springer, 2009.
- [44] Alireza Javanmardi, Yusuf Sale, Paul Hofman, and Eyke Hüllermeier. “Conformal Prediction with Partially Labeled Data”. In: *Proceedings of the Twelfth Symposium on Conformal and Probabilistic Prediction with Applications*. Ed. by Harris Papadopoulos, Khuong An Nguyen, Henrik Boström, and Lars Carlsson. Vol. 204. Proceedings of Machine Learning Research. PMLR, Sept. 2023, pp. 251–266. URL: <https://proceedings.mlr.press/v204/javanmardi23a.html>.
- [45] Alireza Javanmardi, David Stutz, and Eyke Hüllermeier. “Conformalized Credal Set Predictors”. In: *Proceedings of the 38th International Conference on Neural Information Processing Systems*. NeurIPS. 2024.
- [46] Julian Lienen, Caglar Demir, and Eyke Hüllermeier. “Conformal Credal Self-Supervised Learning”. 2023. arXiv: [2205.15239](https://arxiv.org/abs/2205.15239).
- [47] Xing Liu and François-Xavier Briol. “On the Robustness of Kernel Goodness-of-Fit Tests”. 2024. arXiv: [2408.05854](https://arxiv.org/abs/2408.05854).
- [48] Pengyuan Lu, Michele Caprio, Eric Eaton, and Insup Lee. “IBCL: Zero-shot Model Generation for Task Trade-offs in Continual Learning”. 2024. arXiv: [2305.14782](https://arxiv.org/abs/2305.14782).
- [49] Diane Maclagan and Bernd Sturmfels. *Introduction to Tropical Geometry*. Vol. 161. Graduate Studies in Mathematics. American Mathematical Society, Providence, 2015.
- [50] Massimo Marinacci and Luigi Montrucchio. “Introduction to the mathematics of ambiguity”. In: *Uncertainty in economic theory*. Routledge, 2004, pp. 46–107.
- [51] Ryan Martin. “Valid and efficient imprecise-probabilistic inference with partial priors, II. General framework”. 2022. arXiv: [2211.14567](https://arxiv.org/abs/2211.14567).
- [52] Ryan Martin. “Fisher’s underworld and the behavioral-statistical reliability balance in scientific inference”. 2023. arXiv: [2312.14912](https://arxiv.org/abs/2312.14912).
- [53] Ryan Martin and Chuanhai Liu. *Inferential models: reasoning with uncertainty*. CRC Press, 2015.
- [54] Vitor Martin Bordini, Sébastien Destercke, and Benjamin Quost. “Learning calibrated belief functions from conformal predictions”. In: *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*. Ed. by Enrique Miranda, Ignacio Montes, Erik Quaeghebeur, and Barbara Vantaggi. Vol. 215. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 311–320. URL: <https://proceedings.mlr.press/v215/martin-bordini23a.html>.
- [55] Enrique Miranda and Ignacio Montes. “Centroids of the core of exact capacities: a comparative study”. In: *Annals of Operations Research* 321.1 (2023), pp. 409–449.
- [56] Thomas Mortier, Viktor Bengs, Eyke Hüllermeier, Stijn Luca, and Willem Waegeman. “On the Calibration of Probabilistic Classifier Sets”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2023, pp. 8857–8870.
- [57] Arnold Neumaier. “Clouds, Fuzzy Sets and Probability Intervals”. In: *Reliable Computing* 10 (Nov. 2002), pp. 249–272. DOI: [10.1023/B:REOM.0000032114.08705.cd](https://doi.org/10.1023/B:REOM.0000032114.08705.cd).
- [58] Harris Papadopoulos. “Inductive conformal prediction: Theory and application to neural networks”. In: *Tools in artificial intelligence*. Citeseer, 2008.
- [59] Luis Raul Pericchi. “Sets of Prior Probabilities and Bayesian Robustness”. In: *Imprecise Probability Project* (<http://ippserv.rug.ac.be/home/ipp.html>), <http://ippserv.rug.ac.be/documentation/robust/robust.html> (1998).
- [60] Julian Rodemann and Thomas Augustin. “Imprecise Bayesian optimization”. In: *Knowledge-Based Systems* 300 (2024), p. 112186. DOI: [10.1016/j.knosys.2024.112186](https://doi.org/10.1016/j.knosys.2024.112186).
- [61] Yusuf Sale, Viktor Bengs, Michele Caprio, and Eyke Hüllermeier. “Second-Order Uncertainty Quantification: A Distance-Based Approach”. In: *Proceedings of the 41st International Conference on Machine Learning*. Ed. by Ruslan Salakhutdinov et al. Vol. 235. Proceedings of Machine Learning Research. PMLR, July 2024, pp. 43060–43076. URL: <https://proceedings.mlr.press/v235/sale24a.html>.
- [62] Yusuf Sale, Michele Caprio, and Eyke Hüllermeier. “Is the volume of a credal set a good measure for epistemic uncertainty?” In: *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*. Ed. by Robin J. Evans and Ilya Shpitser. Vol. 216. Proceedings of Machine Learning Research. PMLR, Aug. 2023, pp. 1795–1804.



URL: <https://proceedings.mlr.press/v216/sale23a.html>.

- [63] Glenn Shafer. *A mathematical theory of evidence*. Vol. 42. Princeton, NJ: Princeton university press, 1976.
- [64] Glenn Shafer and Vladimir Vovk. “A tutorial on conformal prediction.” In: *Journal of Machine Learning Research* 9.3 (2008).
- [65] Seth Sullivant. *Algebraic Statistics*. American Mathematical Society, 2018.
- [66] Massimiliano Vasile, ed. *Optimization Under Uncertainty with Applications to Aerospace Engineering*. Vol. 6. Physics and Astronomy. Cham, Switzerland: Springer, 2021.
- [67] Vladimir Vovk. “Transductive conformal predictors”. In: *Artificial Intelligence Applications and Innovations*. Ed. by Harris Papadopoulos, Andreas S. Andreou, Lazaros Iliadis, and Ilias Maglogiannis. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 348–360.
- [68] Vladimir Vovk, Alexander Gammernan, and Glenn Shafer. *Algorithmic learning in a random world*. Vol. 29. Springer, 2005.
- [69] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Vol. 42. Monographs on Statistics and Applied Probability. London: Chapman and Hall, 1991.
- [70] Lisa Wimmer, Yusuf Sale, Paul Hofman, Bernd Bischl, and Eyke Hüllermeier. “Quantifying aleatoric and epistemic uncertainty in machine learning: Are conditional entropy and mutual information appropriate measures?” In: *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*. Ed. by Robin J. Evans and Ilya Shpitser. Vol. 216. Proceedings of Machine Learning Research. PMLR, Aug. 2023, pp. 2282–2292. URL: <https://proceedings.mlr.press/v216/wimmer23a.html>.
- [71] Ronald R Yager and Liping Liu. *Classic works of the Dempster-Shafer theory of belief functions*. Vol. 219. Springer, 2008.
- [72] Marco Zaffalon. “The naive credal classifier”. In: *Journal of Statistical Planning and Inference* 105.1 (2002). Imprecise Probability Models and their Applications, pp. 5–21.