# Conformal Prediction Regions are Imprecise Highest Density Regions

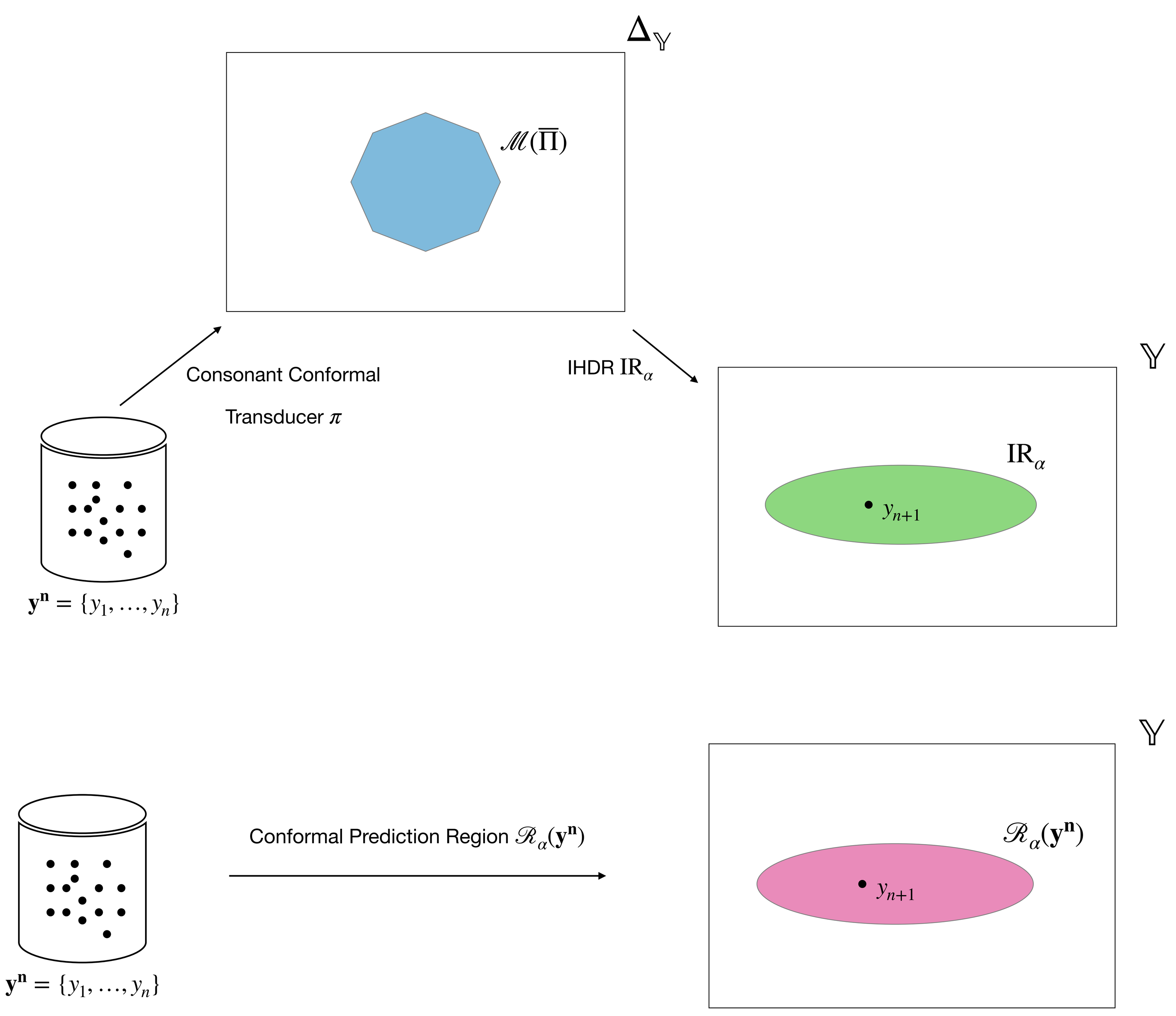

**Michele Caprio** **Yusuf Sale** **Eyke Hüllermeier**

The University of Manchester & Manchester Centre for AI Fundamentals, Manchester, UK
Ludwig-Maximilians-Universität München & Munich Center for Machine Learning, Munich, Germany

michele.caprio@manchester.ac.uk
{yusuf.sale,eyke}@ifi.lmu.de


## 1. Motivation

CONTINUE the endeavor started by Cella and Martin [5] to relate Conformal Prediction (CP, [1]) and Imprecise Probabilities (IPs, [2, 10, 13]).




**Research Question.** Do Imprecise and Conformal Prediction Regions **coincide**? Do we get some bonus intuition on CP and IPs (and their relations) when investigating this?

## 2. Full 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 $(\mathbb{Y}, \Sigma_{\mathbb{Y}})$. We observe the first $n$ terms of the process, $\mathbf{Y^n} = (Y_1, \dots, Y_n)^\top$.

**Goal.** Predict $Y_{n+1}$ using a method that is **valid**.

Let $\mathbf{Y^{n+1}} = (\mathbf{Y^n}, Y_{n+1})^\top$. Consider the transform

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

defined by the rule

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

where $\mathbf{y^{n+1}_{-i}} = \mathbf{y^{n+1}} \backslash \{y_i\}$, and the *non-conformity measure* $\Psi : \mathbb{Y}^n \times \mathbb{Y} \to \mathbb{R}$ is invariant to permutations in its first vector argument. Large values of $\psi_i(\mathbf{y^{n+1}})$ suggest that the observation $y_i$ is "strange" and does not conform to the rest of the data $\mathbf{y^{n+1}_{-i}}$.

$Y_{n+1}$ has not yet been observed: The above calculations cannot be carried out exactly. Nevertheless, the exchangeability-preserving properties of the transformations $\psi_i$ provide a procedure to rank candidate values $\tilde{y}$ of $Y_{n+1}$ based on the observed $\mathbf{Y^n} = \mathbf{y^n}$.

**Algorithm 1** Full 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 *conformal transducer* $\pi(\cdot, \mathbf{y^n})$ [12]: A measure of plausibility of the assertion that $Y_{n+1} = \tilde{y}$, given data $\mathbf{y^n}$.

$\pi(\cdot, \mathbf{y^n})$ is used to derive the *Conformal Prediction Regions* (CPRs) [11, Equation (2)],

$$\forall \alpha \in [0,1], \quad \mathscr{R}_\alpha(\mathbf{y^n}) := \{\tilde{y} \in \mathbb{Y} : \pi(\tilde{y}, \mathbf{y^n}) > \alpha\},$$

which satisfy the following uniformly in $n$ and in $P$ [12],

$$P\left[Y_{n+1} \in \mathscr{R}_\alpha(\mathbf{y^n})\right] \geq 1 - \alpha.$$

## 3. Imprecise Probabilistic Background Notions

**Definition: (Consonant) Plausibility Function [2]**

An upper probability $\overline{P}$ is $k$*-alternating* if for every collection $\{A, A_1, \dots, A_k\} \subseteq \Sigma_{\mathbb{Y}}$ such that $A_i \subseteq A$, for all $i \in \{1, \dots, k\}$,

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

$\overline{P}$ is called a **plausibility function** $pl$ if it is an $\infty$-alternating capacity, i.e., if (1) holds $\forall k$. A plausibility function $pl$ is said to be **consonant** if there exists a *plausibility contour* $\pi : \mathbb{Y} \to [0,1]$ such that (i) $\sup_{y \in \mathbb{Y}} \pi(y) = 1$; (ii) $pl(A) = \sup_{y \in A} \pi(y)$, $A \in \Sigma_{\mathbb{Y}}$.

**Lemma: Algebraic Properties of Consonant Plausibility Function**

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

**Definition: Imprecise Highest Density Region (IHDR) [7]**

Let $\underline{P}$ be a lower probability, and fix any $\alpha \in [0,1]$. An **Imprecise Highest Density Region** is a set $\text{IR}_\alpha \subset \Sigma_{\mathbb{Y}}$ such that (i) $\underline{P}[\text{IR}_\alpha] = 1 - \alpha$; (ii) $\int_{\text{IR}_\alpha} \mathrm{d}y$ is a minimum.

## 4. Relations Between Conformal Prediction and Imprecise Probabilities

**Definition: Under Consonance CP is Associated with an Upper Probability [5]**

Suppose that the conformal transducer satisfies consonance, i.e.

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

Then, we can define a **(predictive) upper probability** as $\overline{\Pi}_{\mathbf{y^n}}(A) = \sup_{\tilde{y} \in A} \pi(\tilde{y}, \mathbf{y^n})$, $A \in \Sigma_{\mathbb{Y}}$.

Consonance (that can also be satisfied with a transformation of the conformal transducer $\pi$) holds quite generally e.g. for conformal prediction in continuous-data problems [4].
With $\overline{\Pi}_{\mathbf{y^n}}$, **we obtain "for free" a credal set** $\mathcal{M}(\overline{\Pi}_{\mathbf{y^n}}) := \{P : P(A) \leq \overline{\Pi}_{\mathbf{y^n}}(A), \forall A \in \Sigma_{\mathbb{Y}}\}$.

**Lemma: Properties of $\overline{\Pi}_{\mathbf{y^n}}$**

Let $\mathcal{A} \subseteq \Sigma_{\mathbb{Y}}$ be a generic collection of subsets of $\mathbb{Y}$. Upper probability $\overline{\Pi}_{\mathbf{y^n}}$ is (i) **supremum preserving**; (ii) **coherent à la Walley** [13]; and (iii) **tropically finitely additive**.

**Propositions: CPRs are IHDRs**

Let $\text{IR}^{\mathcal{M}}_\alpha$ be the IHDR for $\mathcal{M}(\overline{\Pi}_{\mathbf{y^n}})$. For any $\alpha \in [0,1]$ and any $n \in \mathbb{N}$, the following is true,

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

In addition, $P[Y_{n+1} \in \text{IR}^{\mathcal{M}}_\alpha] \geq 1 - \alpha$ holds uniformly in $P$ and in $n$.

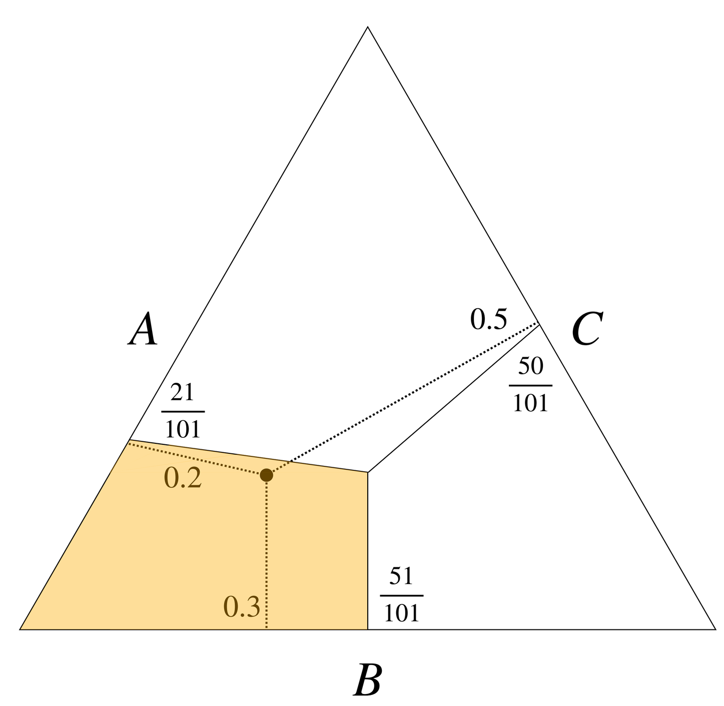
## 5. Discussion

THERE seems to be a sense in which CP assumes a vacuous prior [4, 6, 8, 9]. That is, CP appears akin to a Bayesian Sensitivity Analysis (BSA) procedure [3] whose prior credal set is the whole space of parameter probabilities. Contrary to BSA, though, CP is able to derive a non-vacuous predictive credal set $\mathcal{M}(\overline{\Pi}_{\mathbf{y^n}}) \subsetneq \Delta_{\mathbb{Y}}$.

**Conjecture.** The main reason for this, is that **CP enjoys half-coherence** [8, Section 5.2.2].

Consonance induces a **"distortion" in the shape** of $\mathcal{M}(\overline{\Pi}_{\mathbf{y^n}})$. The true data generating process need not be its centroid.

**Example.** Suppose $\mathbb{Y} = \{A, B, C\}$, and observe $20$ $A$'s, $30$ $B$'s and $50$ $C$'s. As a consequence, empirical pmf: $p^{\text{emp}} = (0.2, 0.3, 0.5)^\top$. Choose non-conformity measure $\psi_i(\mathbf{y^{n+1}}) = 1 - p^{\text{emp}}(y_i = k)$, for all $i \in \{1, \dots, n+1\}$.




**Open Question 1.** Why this happens?

**Open Question 2.** What are the consequences on the measure of AU?

**Open Question 2.** Is $\mathcal{M}(\overline{\Pi}_{\mathbf{y^n}})$ minimal (i.e. the smallest possible credal set associated with CP)?

**References.**


**ISIPTA 2025**, 14th International Symposium on Imprecise Probabilities: Theories and Applications. 15-18 July 2025, Bielefeld, Germany