The game of mind and dice

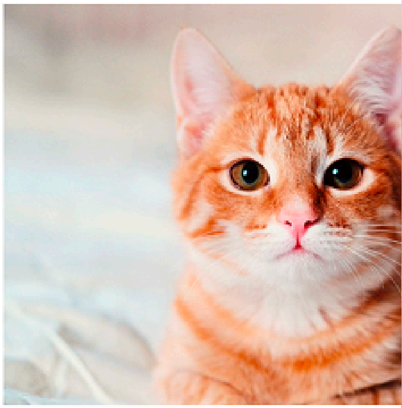
Uncertainty Quantification

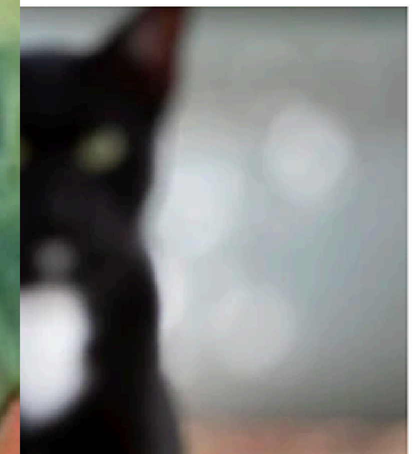
Veronika Lohmanova
The University of Manchester

Plan

1. **Why mind? Why dice?** **(Rationale)**
2. **IPs are nice** **(Toolbox)**
3. **What we can do with it?** **(Application to ML)**

Cat vs Dog





ale **mic**

(irreducib n-specifity)

total

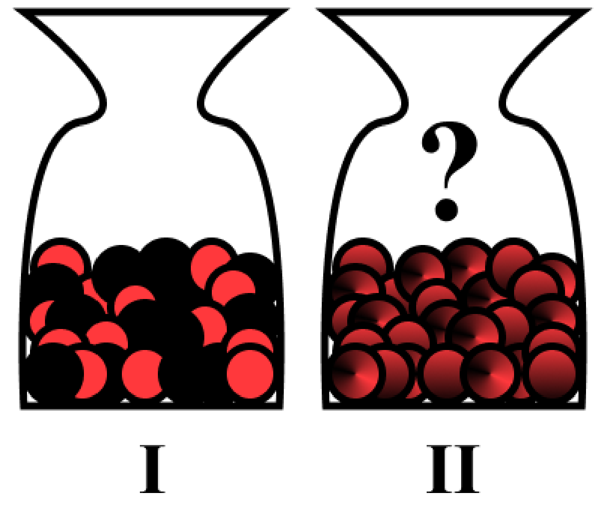
$TU = AU + EU$

Source: KYM

Classical probabilities

$$\begin{aligned}\mathrm{Var}(\hat{y} \mid x) = \mathbb{E}_\theta\big[\mathrm{Var}(\hat{y} \mid x, \theta)\big] \\ + \mathrm{Var}_\theta(\mathbb{E}_\theta\big[\hat{y} \mid x, \theta\big])\end{aligned}$$

(Waegeman 2026)

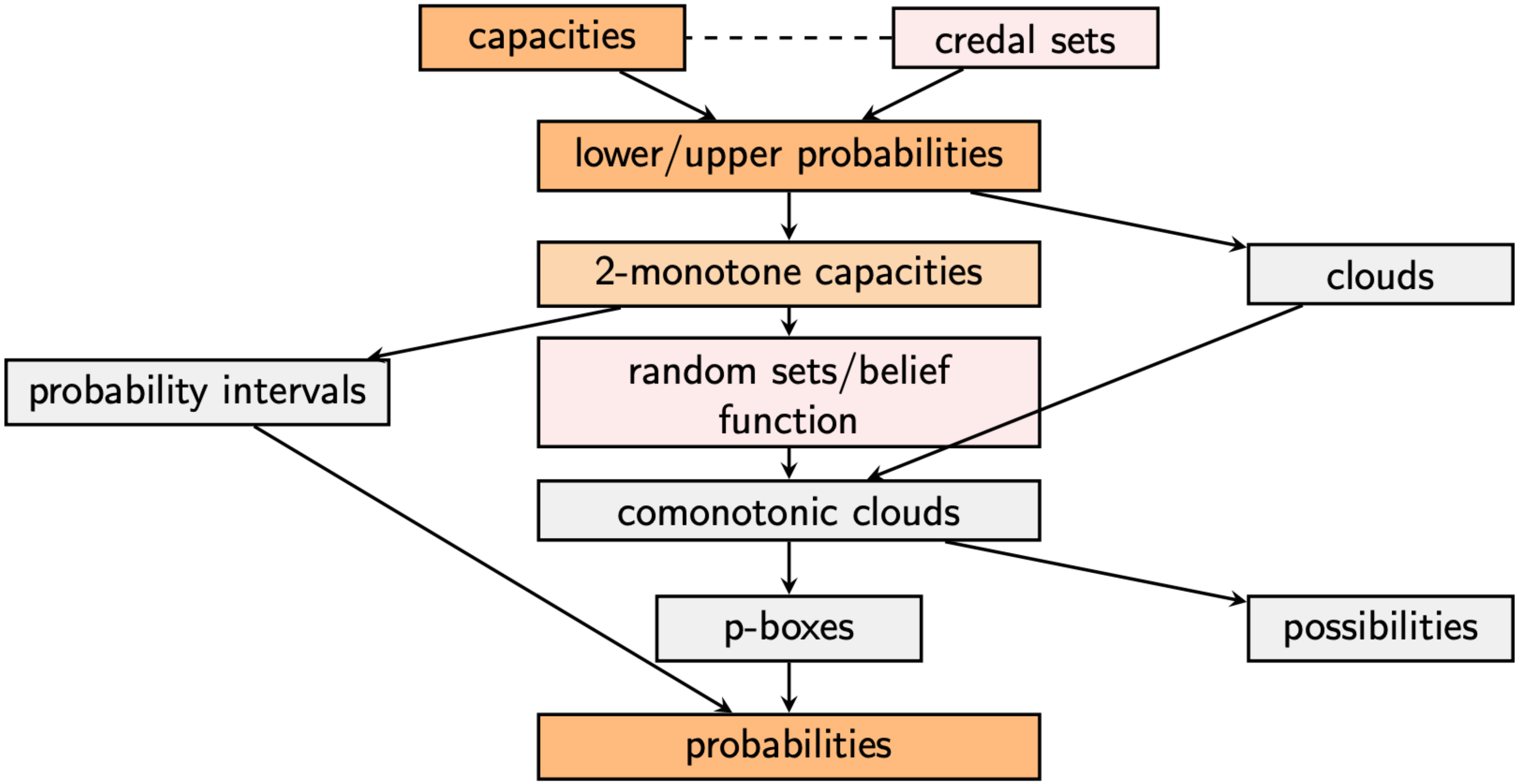


Ellsberg paradox (Ellsberg 1961)

> If I do not know anything about event A, I want to encode $\mathbb{P}[A] = 0$, but that implies that I know everything about A^C since now $\mathbb{P}[A^C] = 1$ (Chau 2026).

Toolbox

Imprecise Probabilities



Source: (Chau 2026). A skeleton demonstrating the connection between various uncertainty calculi. $A \rightarrow B$ means A generalises B, meaning that B is a specific instance of A.

Credal sets

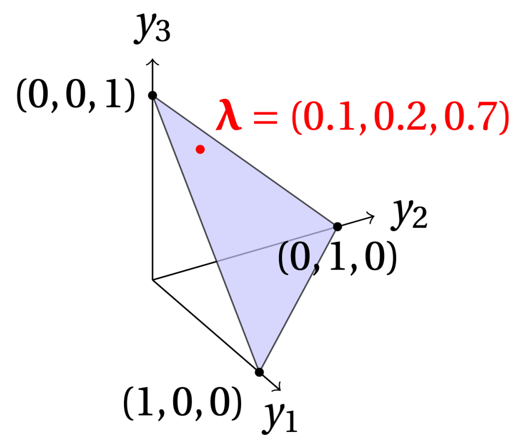
Credal sets $\mathcal{P}$ (Walley 1991):

- Closed and convex sets of probability distributions $\forall P_1, P_2 \in \mathcal{P}$

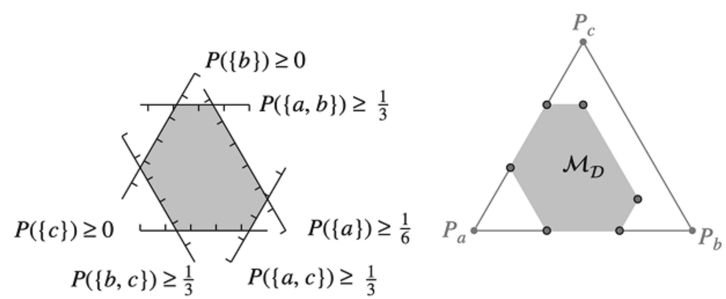
$$\{\lambda P_1 + (1-\lambda)P_2 \mid \lambda \in [0,1]\} \subseteq \mathcal{P}$$

- For finite space, represented as simplex $\Delta^{|\Omega|-1} := \{\boldsymbol{v} \in \mathbb{R}_{\geq 0}^{|\Omega|} \mid \sum_{i=1}^{|\Omega|} v_i = 1\}$

$$x + y + z = 1$$



Source: (Lohmanova Under review)

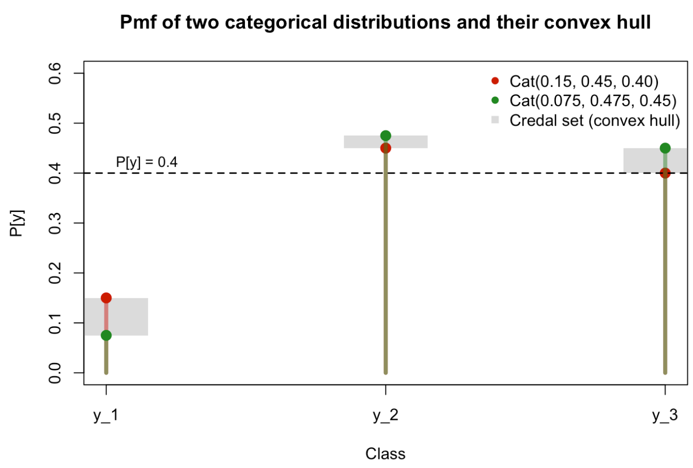


Source: (Augustin et al. 2014).

Lower and upper probabilities

Credal sets $\mathcal{P}$ (Walley 1991):

- Closed and convex sets of probability distributions



Source: (Lohmanova Under review)

Lower/upper probabilities

$$\bar{\mathbb{P}}[A] = \sup_{P \in \mathcal{P}} P(A)$$

$$\underline{\mathbb{P}}[A] = \inf_{P \in \mathcal{P}} P(A)$$

$\forall A \in \sigma(\Omega)$ (Caprio, Sale, and Hüllermeier 2025)

MANCHESTER 1824 The University of Manchester

Uncertainty decomposition

- Shannon entropy (Shannon 1948): $H(P) = -\mathbb{E}_P\left[\log_2 P(X)\right]$
- **Lower** and **upper** Shannon entropies (Abellán, Klir, and Moral 2006):

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

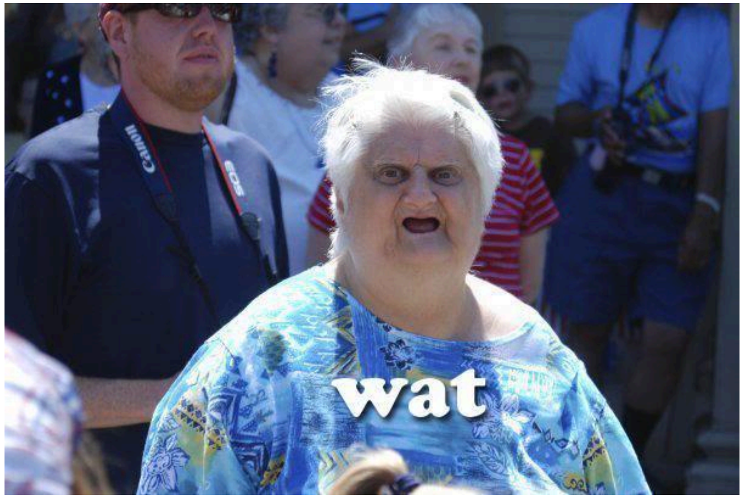
$$\bar{H}(\mathcal{P}) = \sup_{P \in \mathcal{P}} H(P)$$

$$TU(\mathcal{P}) = AU(\mathcal{P}) + EU(\mathcal{P})$$

$$\bar{H}(\mathcal{P}) = \underline{H}(\mathcal{P}) + (\bar{H}(\mathcal{P}) - \underline{H}(\mathcal{P}))$$

Progress check

- ☑ We know how to create our abstract model - Credal sets
- ☑ We know how to draw inference on the credal set - Lower and upper probabilities
- ☑ We know how it relates to uncertainty decomposition - Lower and upper Shannon entropies



Source: KYM

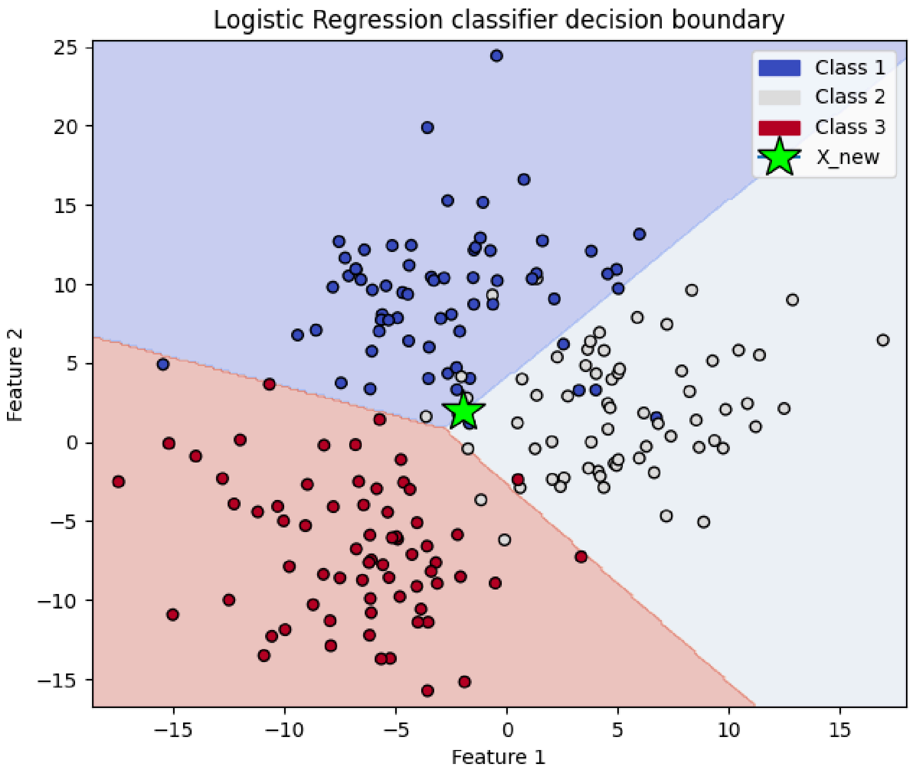
Conformal Prediction (CP)

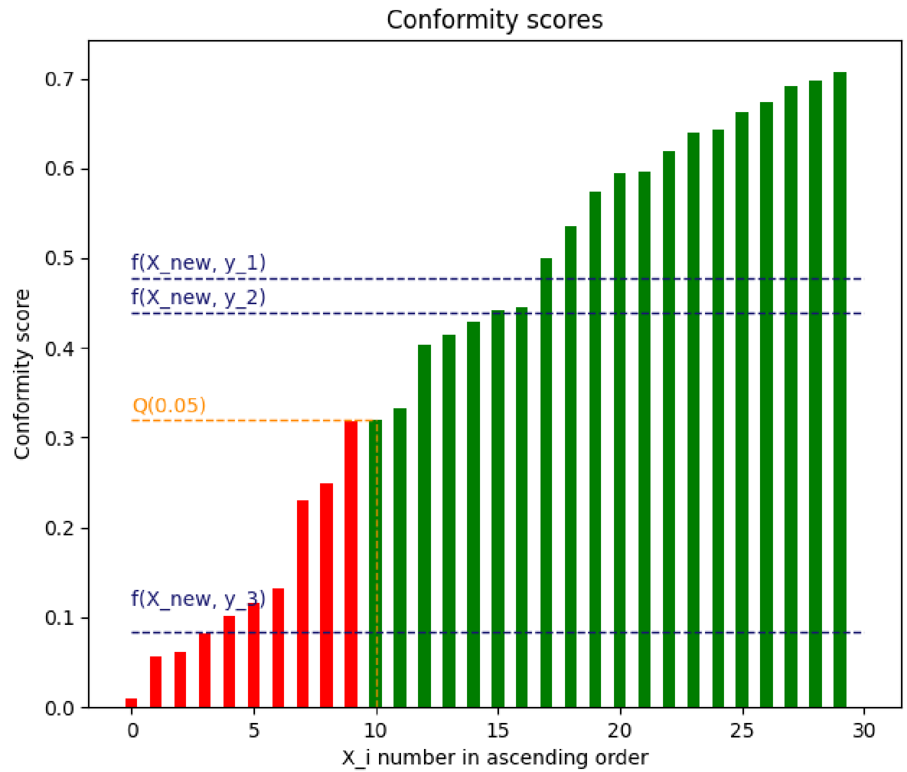
- Introduced by Vovk, Gammerman, and Shafer (2005)
- $\mathcal{D}_{\text{cal}} = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\} \subseteq \mathcal{X} \times \mathcal{Y}$
- **Key assumption:** $\mathcal{D}_{\text{cal}} \cup (X_{n+1}, Y_{n+1})$ is **exchangeable**

- $s : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ - (non)conformity function
- α - user specified miscoverage level
- Form a prediction set $\mathcal{C}(X_{n+1}) = \{y \in \mathcal{Y} : s(X_{n+1}, y) \geq \tau\}$ where $\tau = q(\{s(x, y) \mid (x, y) \in \mathcal{D}_{\text{cal}}\}, \alpha)$

Conformal Prediction (CP)

- Form a prediction set $\mathcal{C}(X_{n+1}) = \{y \in \mathcal{Y} : s(X_{n+1}, y) \geq \tau\}$ where $\tau = q(\{s(x,y) \mid (x,y) \in \mathcal{D}_{\text{cal}}\}, \alpha)$



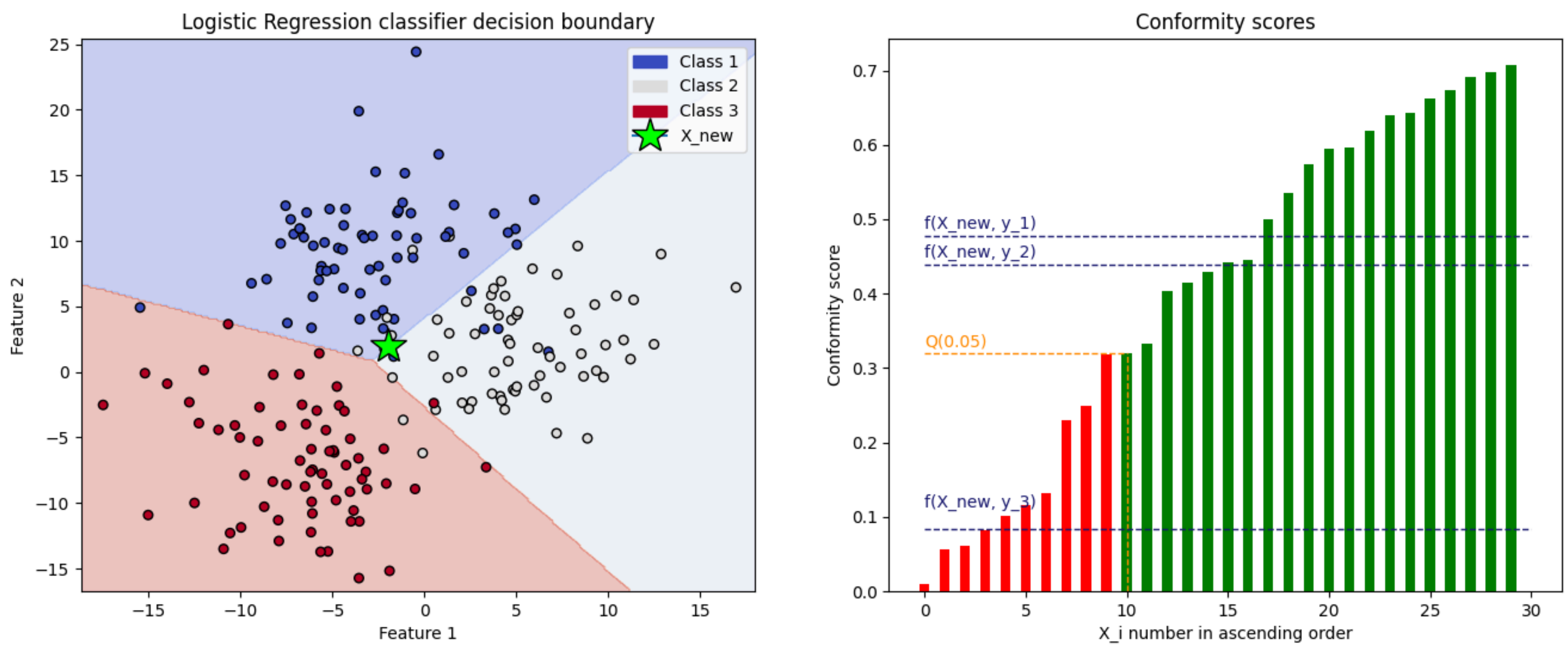


Source: (Lohmanova Under review). Example of conformal prediction at significance level $\alpha = 0.05$ applied to a logistic regression classifier with three classes. The model's predicted label is *Class 1*, but the conformal predictor yields a prediction set $C(x_{\text{new}}) = \{\text{Class } 1, \text{Class } 2\}$.

Conformal Prediction (CP)

Statistical guarantee:

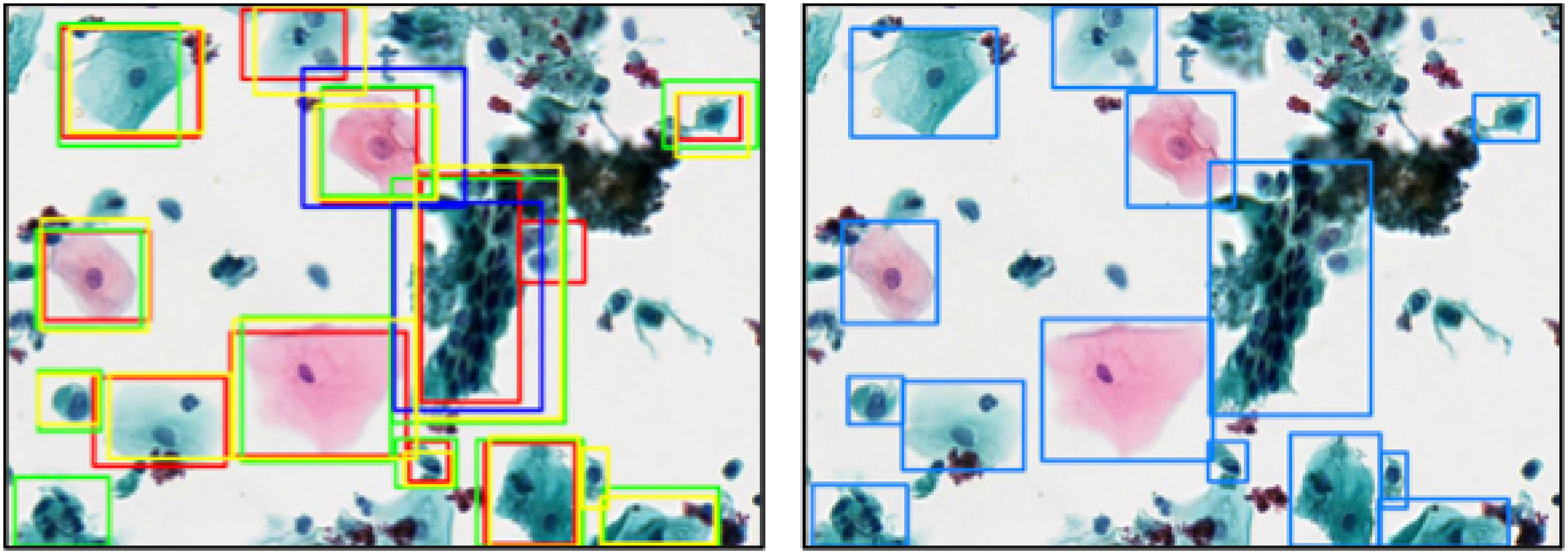
$$\mathbb{P}\left[Y_{n+1} \in \mathcal{C}\left(X_{n+1}\right)\right] \geq 1-\alpha$$



Source: (Lohmanova Under review). Example of conformal prediction at significance level $\alpha = 0.05$ applied to a logistic regression classifier with three classes. The model's predicted label is *Class 1*, but the conformal predictor yields a prediction set $C(x_{\text{new}}) = \{\text{Class 1}, \text{Class 2}\}$.

Imprecise Probabilistic ML

Application: medical imaging



(a) Raw Annotations (b) Ground Truth Annotations

Source: (Si et al. 2026). Raw expert annotations (left) vs. the final goldstandard ground truth (right) on a sample image from CytoCrowd.

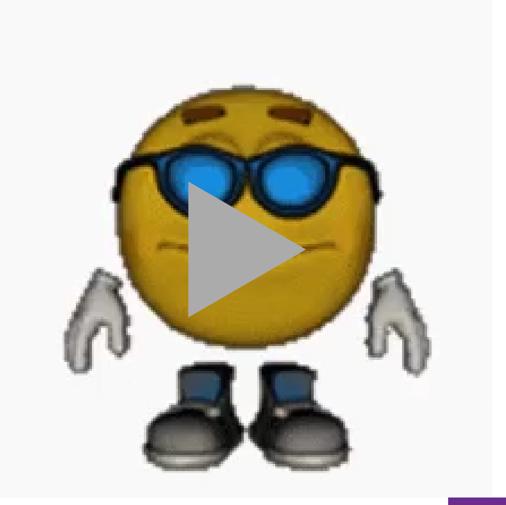
Imprecise Probabilistic CP

- Caprio et al. (2025)
- $\mathcal{D}_{\text{cal}} = \{(X_1, \boldsymbol{\Lambda_1}), (X_2, \boldsymbol{\Lambda_2}), \ldots, (X_n, \boldsymbol{\Lambda_n})\} \subseteq \mathcal{X} \times \Delta^{K-1}$
- $\implies \mathcal{C}(X_{n+1}) = \{\lambda \in \Delta^{K-1} : S(X_{n+1}, \lambda) \geq \tau\}$ plausibility region
- $S(x, \boldsymbol{\lambda}) = \sum_{i=1}^{K} \lambda_i \boldsymbol{s}(x, k_i)$
- $\mathcal{P} = \{\mathrm{Cat}(\lambda) : \lambda \in \mathcal{C}(X_{n+1})\}$ credal region

> Statistical guarantee:
>
> $$\mathbb{P}\left[\mathrm{Cat}\left(\Lambda_{n+1}\right) \in \mathcal{P}\right] \geq 1 - \alpha$$

Key takeaways

- Probability theory and statistics are incredibly interesting phenomena
- **WEWTYIW** principle - “Why Everything We Teach You Is Wrong”
- Kolmogorovian axiomatisation is actually not enough when reasoning about uncertainty $\Longrightarrow$ we need new models (IPs)
- Uncertainty can be aleatoric and epistemic
- Conformal Prediction is actually magical ~~black~~ box
- We need to teach AI about uncertainty



References

Abellán, J., G. J. Klir, and S. Moral. 2006. ‘Disaggregated Total Uncertainty Measure for Credal Sets’. *International Journal of General Systems* 35 (February): 29–44. https://doi.org/10.1080/03081070500473490.

Augustin, Thomas, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, eds. 2014. *Introduction to Imprecise Probabilities*. Wiley. https://doi.org/10.1002/9781118763117.

Bordini, Vitor Martin, Sébastien Destercke, and Benjamin Quost. 2026. ‘Cautious Self-Learning via Labelwise Uncertainty Quantification’. " WUML 2026 (https://sites.google.com/view/wuml2026/program?authuser=0)".

Caprio, Michele, Yusuf Sale, and Eyke Hüllermeier. 2025. ‘Conformal Prediction Regions Are Imprecise Highest Density Regions’, April.

Caprio, Michele, David Stutz, Shuo Li, and Arnaud Doucet. 2025. ‘Conformalized Credal Regions for Classification with Ambiguous Ground Truth’. https://arxiv.org/abs/2411.04852.

Chau, Siu Lun. 2026. ‘Recent Advances in Imprecise Probabilistic Machine Learning’. https://chau999.github.io/assets/research_presentation/2026-IIPM.pdf.

Ellsberg, Daniel. 1961. ‘Risk, Ambiguity, and the Savage Axioms’. *The Quarterly Journal of Economics* 75 (November): 643. https://doi.org/10.2307/1884324.

Lohmanova, Veronika. Under review. ‘Foundational Aspects of Imprecise Probabilistic Conformal Prediction (IPCP)’.

Shannon, C. E. 1948. ‘A Mathematical Theory of Communication’. *Bell System Technical Journal* 27 (July): 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Si, Yonghao, Xingyuan Zeng, Zhao Chen, Libin Zheng, Caleb Cao, Lei Chen, and Jian Yin. 2026. ‘CytoCrowd: A Multi-Annotator Benchmark Dataset for Cytology Image Analysis’, February. https://doi.org/10.48550/arXiv.2602.06674.

Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer. 2005. *Algorithmic Learning in a Random World*. Springer-Verlag. https://doi.org/10.1007/b106715.

Waegeman, Willem. 2026. ‘"Disentangling Aleatoric and Epistemic Uncertainty in ML: Opportunities and Challenges"’. " WUML 2026 (https://youtu.be/Q64FEk5wfQw?si=aXUv1xAcO8bZ-_Kc)".

Walley, Peter. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall.

Thank you for attention!

Any questions?