TECHNISCHE UNIVERSITÄT DRESDEN

VOTING FOR BINS

INTEGRATING IMPRECISE PROBABILISTIC BELIEFS INTO THE CONDORCET JURY THEOREM

Jonas Karge
Computational Logic Group, TU Dresden

KoDis 2023, Rhodes, Greece, September 03, 2023

Introduction

Scenario: Multiple experts assess the likelihood of an event such as:

Example: Global sea level will rise at least 1,5 meters until the year 2100 above the level of 2000.

Two fundamental questions:

(1) How can we appropriately represent the probabilistic beliefs of experts?

(2) What constitutes a reasonable method for aggregation?

Outline

(i) Aggregation Method: Voting in a jury theorem setting
(ii) Representation: Imprecise probabilistic beliefs
(iii) Voting with this Representation: Supervaluationism
(iv) Embedding

The Condorcet Jury Theorem

The Condorcet Jury Theorem (CJT)

Marie Jean Antoine Nicolas Caritat Marquis de Condorcet

Theorem: For odd-numbered **homogenous** groups of **independent** and **reliable** agents in a **dichotomic** voting setting, the probability that majority voting identifies the correct alternative

- increases monotonically with the number of agents and (non-asymptotic part)
- converges to 1 as the number of agents goes to infinity. (asymptotic part)

Voting

Define **approval voting** and obtain simpler voting mechanisms as special cases.

Given: finite set of n agents $\mathcal{A} = \{a_1, \dots, a_n\}$
finite set of m choices $\mathcal{W} = \{\omega_1, \dots, \omega_m\}$

- **approval voting (instance)**: relation $V \subseteq \mathcal{A} \times \mathcal{W}$
  $(a_i, \omega_j) \in V$ means agent a_i approves choice ω_j
- given $\omega \in \mathcal{W}$, obtain **score** $\#_V\, \omega$ as overall number of votes that ω receives, i.e.,

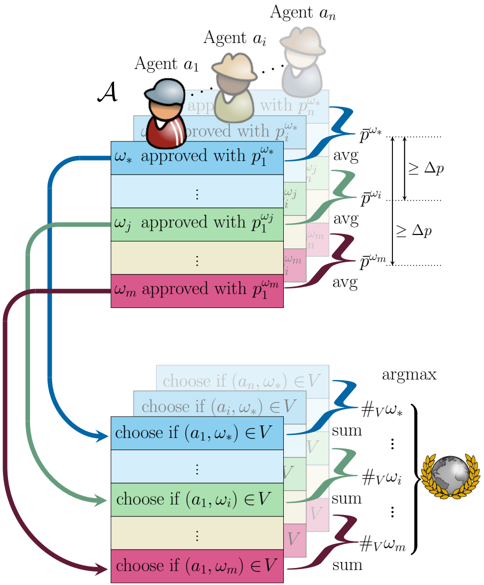
$$\#_V\, \omega = |\{a_i \in \mathcal{A}_n \mid (a_i, \omega) \in V\}|$$

- ω **wins approval vote** V if it receives strictly more votes than any other choice:

$$\#_V\, \omega > \max_{\omega' \in \mathcal{W} \setminus \{\omega\}} \#_V\, \omega'$$

The Voting Scenario

$\mathcal{W}$



Success:

CJT under approval voting

Asymptotic result:

Theorem: For odd-numbered **heterogenous** groups of **independent** and **Δ-reliable** agents in a voting setting with a finite number of alternatives, the probability that approval voting identifies the correct alternative

- converges to 1 as the number of agents goes to infinity.

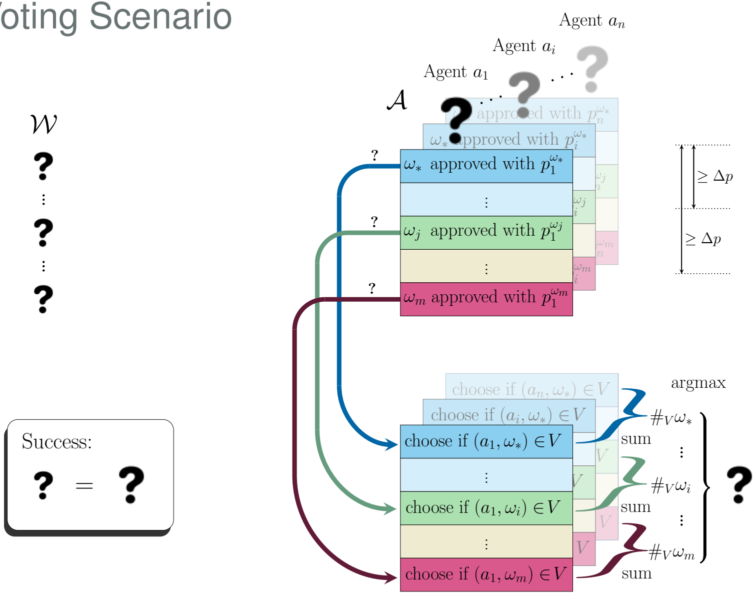
Beyond the convergence behavior in the infinite:

Theorem: In a Δp-group reliable setting with m choices, the worst case approval vote success probability is at least P_{min} whenever the number of agents is equal or higher than

$$\min\left(\tfrac{2}{\Delta p^2}\ln Q, 1+(\tfrac{1}{\Delta p^2}-1)Q\right), \qquad (1)$$

where $Q = 2\frac{m-1}{1-P_{\text{min}}}$ is the twofold ratio between the number of incorrect alternatives and the admissible error probability.

The Voting Scenario



Probabilistic Beliefs

Precise Probabilities

Starting point:

Definition: A probability function $\mathbb{P}$ is a function $\mathbb{P} : 2^{\Omega} \rightarrow \mathbb{R}$, satisfying the probability axioms.

$\Rightarrow$ Output of function reflects the agent's **degree of belief** in that proposition.

Reconsider:

Example: Global sea level will rise at least 1,5 meters until the year 2100 above the level of 2000.

Problem:

What probability is the expert supposed to assign to A?

Imprecise Probabilities

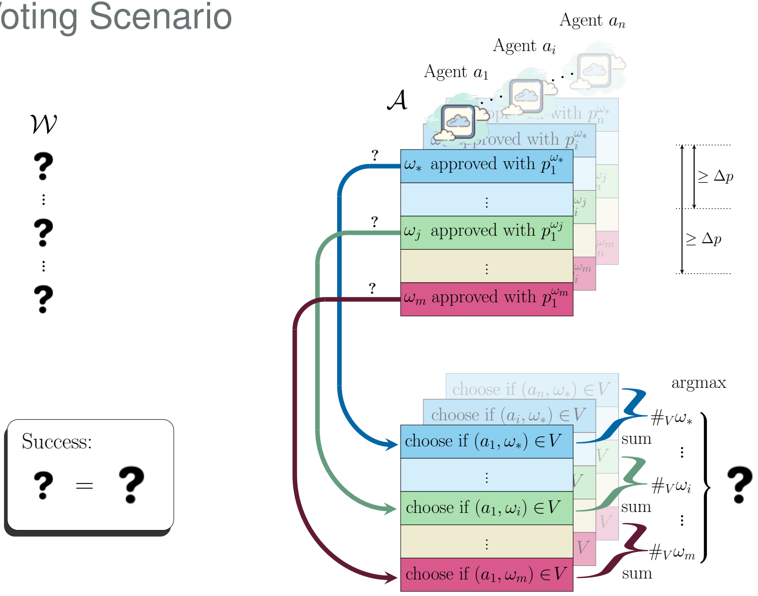
Definition:
Imprecise probabilities are sets of probability functions.

We refer to a specific set of probability functions as the agent's representor, denoted by $\mathcal{P}$.

Definition: An agent's imprecise degree of belief in a proposition H is represented by a function, $\mathcal{P}(H)$, with $\mathcal{P}(H) = \{\mathbb{P}(H) : \mathbb{P} \in \mathcal{P}\}$.

Example: Assume, the agent's representor consists of three probability functions that assign event A values from the set $\{0.4, 0.6, 0.8\}$. Assuming **convexity**, we may represent the agent's imprecise degree of belief with $\mathcal{P}(A) = [0.4, 0.8]$. Thus, our agent is $40 - 80\%$ confident that event A will occur, i.e., that proposition A is true.

The Voting Scenario



Supervaluationism and Voting

Standard Supervaluationism

Consider a **vague predicate** such as **tall**
⇒ can be made more precise by introducing cutoff points (i.e. 300cm, 180cm, 20cm).

Each cutoff point represents a **precisification** of that predicate.

Truth value of vague predicates:

- **Determinate truth** (true according to all admissible precisifications, person who is 400cm tall);
- **Determinate falsehood** (false according to all admissible precisifications, 10cm);
- **Indeterminate truth** (true and false according to some admissible precisifications, 190cm).

Modified Supervaluationism

Definition: A proposition is predominantly true if it is true according to a relative majority of admissible precisifications.

Definition: Given two propositions, A and B, an agent is considered to be predominantly more confident in proposition A than in proposition B if a greater proportion of elements within the agent's imprecise degree of belief satisfy the condition $Pr(A) > Pr(B)$.

Problem: we need to measure the proportion of possibly infinitely many elements.

⇒ For any closed, $[a, b]$, open, (a, b), or half open, $(a, b]$ or $[a, b)$, interval it holds that its **Lebesque measure** is of length $l = b - a$;

⇒ determine the proportion of elements in favor of a proposition by measuring the length of the corresponding interval.

Modified Supervaluationism and Voting

Example: Consider proposition A and its complement B, i.e. global sea level will **not** rise at least 1,5 meters until the year 2100 above the level of 2000. Suppose we have $\mathcal{P}(A) = [0.4, 1]$ as our agent's imprecise degree of belief. For those elements represented by $(0.5, 1]$ it holds true that $Pr(A) > Pr(B)$. For those represented by $[0.4, 0.5)$ we have $Pr(B) > Pr(A)$. Taking their Lebesque measure, we obtain $l(A) = 0.5$ as well as $l(B) = 0.1$. Thus, the agent is predominantly more confident in proposition A.

Predominant confidence and voting:

Definition: Given a set of alternatives $\mathcal{W} = \{\omega_1, \ldots, \omega_m\}$ and set of agents $\mathcal{A} = \{a_1, \ldots, a_n\}$, agent a_i approves alternative ω_j if the agent is predominantly more confident in that alternative than in its competitors.

Embedding

What is an Alternative?

Recall: We are given a **finite** set of alternatives $\omega_1, ..., \omega_m$ and one $\omega_k \in \mathcal{W}$ represents the **correct probability** for an event to occur.

Suppose, proposition A has a probability of 40.1862345% to occur.

First idea: Each alternative represents a precise probability value.
$\Rightarrow$ Similar problems as on the belief level.

Second idea: Each alternative represents an **interval of probability values** of the form: $[X_{min}, X_{max}]$.

Simplest (theoretically excluded) case: A **single alternative** with $[X_{min}, X_{max}] = [0, 1]$.

More Alternatives?

To obtain more alternatives, $[X_{\text{min}}, X_{\text{max}}]$ is further divided into **subintervals**.

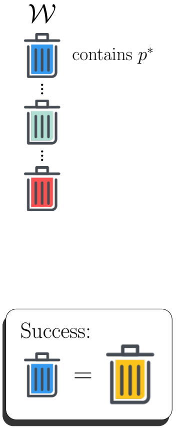
$\Rightarrow$ Partition the unit interval $[0, 1]$ into $2k$ subintervals of equal Lebesgue measure, denoted as $[X_{j-1}, X_j)$, where $j = 1, 2, ..., 2k$.

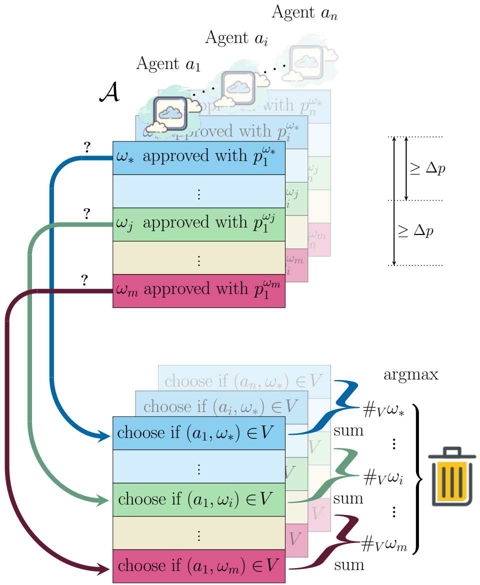
Note: The value of k depends on the desired precision.

Example: If a small group of agents is highly reliable in their estimates, a moderate precision of 10% may be adequate, achieved by setting $k = 5$ (yielding $2k = 10$) and providing 10 subintervals of equal Lebesgue measure.

Note: Each subinterval is referred to as a bin.

The Voting Scenario





Confidence in Bins

Problem: **predominant confidence** compares the beliefs in two propositions.

⇒ compare confidence in two probabilistic assessments for the same proposition.

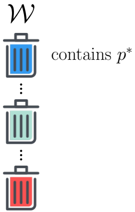
Definition: Let A be a proposition, $\mathcal{P}(A) = [a, b]$ be an agent's imprecise degree of belief in A, and let $[X_{j-1}, X_j)$, $j = 1, 2, ..., 2k$ be $2k$ bins defined on the unit interval reflecting probability values for A to occur. Given two bins B_1 and B_2, we say that an agent is predominantly more confident in B_1 if the intersection of $\mathcal{P}(A)$ and B_1 is of greater Lebesque measure than the one of $\mathcal{P}(A)$ and B_2. That is, $l(\mathcal{P}(A) \cap B_1) \geq l(\mathcal{P}(A) \cap B_2)$.

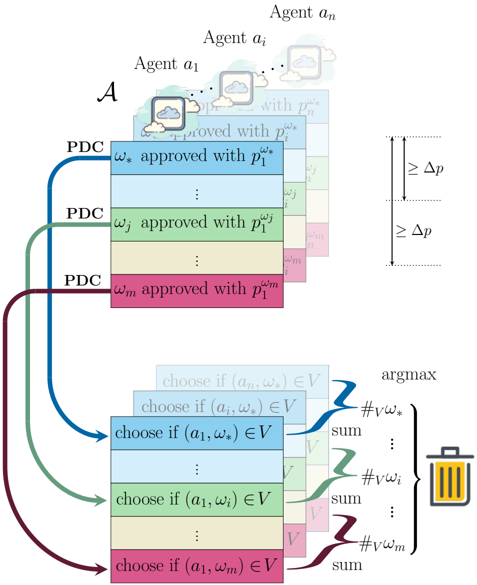
Voting for Bins

Example: Suppose there are only two bins for proposition A with $B_1 = [0, 0.5)$ and $B_2 = [0.5, 1]$ and let $\mathcal{P}(A) = [0.3, 0.9]$. We then have $\mathcal{P}(A) \cap B_1 = [0.3, 0.5)$ and $\mathcal{P}(A) \cap B_2 = [0.5, 0.9]$. This results in $l(\mathcal{P}(A) \cap B_1) = 0.2$ as well as $l(\mathcal{P}(A) \cap B_2) = 0.4$. Thus, the agent is predominantly more confident in the second bin.

Definition: Let $m = \omega_1, ..., \omega_m$ be a set of alternatives where each ω_i represents a bin of the form $[X_{j-1}, X_j)$. Moreover, let $a_1, ..., a_n$ be a set of agents and let V represent a single election. We say that an agent a_i votes for an alternative ω_j if she is predominantly confident in that alternative. That is, if $(l(\mathcal{P}(A) \cap \omega_j) \geq l(\mathcal{P}(A) \cap \omega_k))$ for all $j \neq k$ then $(a_i, \omega_j) \in V$.

The Voting Scenario





Estimate Permitted Precision

- In typical application scenarios, the number of voters on the expert board is known beforehand;
- i.e. studies on forecasting capabilities involving 42 climate scientists or 140 experts on COVID-19 outbreaks;

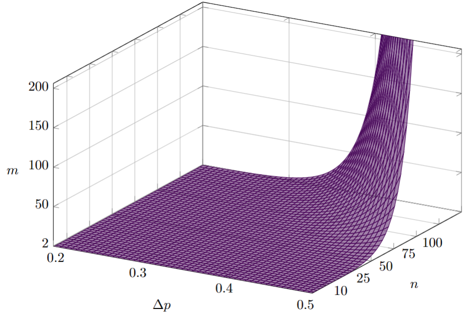
$\Rightarrow$ determine the maximal precision that can be allowed.

Theorem: In a Δp-group reliable setting where $\Delta p \in (0, 1)$ with n agents, the worst case approval vote success probability is at least $P_{\min}$ whenever the number of alternatives is equal or lower than

$$max(\frac{(1-p_{\min})}{(2e^{-\frac{1}{2}n\Delta p^2})} + 1, \frac{(1-p_{\min})(1+(n-1)\Delta p^2)}{2(1-\Delta p^2)} + 1). \tag{2}$$

$\Rightarrow$ Directly translates into the maximal allowed precision in percentage (C) with $C = \frac{100}{m}$.

Illustration



Selection of data points			
Number of Experts	Δp	Number of Bins	Precision
50	0.3	< 2	N.A.
50	0.4	4	25%
75	0.3	2	50%
75	0.4	21	4.8%
100	0.3	6	16%
100	0.4	150	$0.\overline{6}\%$
150	0.3	44	2.2%
150	0.4	8139	0.01%
200	0.3	406	0.25%
200	0.4	444307	0.0002%

Figure: Maximal number of permitted bins for $P_{min} = 0.9$ and varying Δp and n (left) as well as a selection of data points (right).

Summary and Future Work

Summary

We embedded imprecise probabilistic beliefs into a generalization of the **Condorcet Jury Theorem**:

- combined the epistemological account of **imprecise degrees of belief** as well as their interpretation in the **supervaluationinstic theory of vagueness** with a voting setting;
- each alternative represents an interval of **probability assessments (a bin)** for the same proposition as the imprecise degree of belief;
- established a direct correspondence between the number of **bins** in the voting process and the **maximal permitted precision** during aggregation;
- gave **estimate** for the allowed precision in the aggregation procedure.

Future work

- Compare the performance of **voting-based aggregation** of imprecise probabilistic beliefs with traditional methods of **probabilistic opinion pooling**;
- investigate applications at the intersection of ensemble learning in **ML and social choice**;
- consider **rough set theory** instead of **supervaluationism.**