IPML

IMPRECISE PROBABILISTIC MACHINE LEARNING

Lecture 2: Overview of Imprecise Probability

Krikamol Muandet

24 October 2025

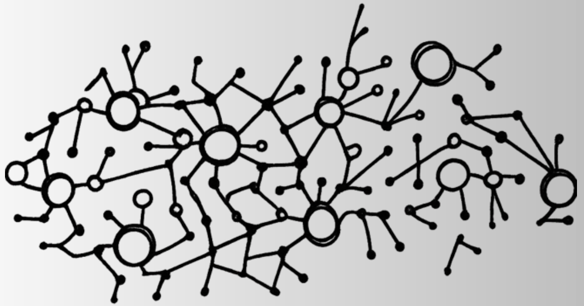
The Achilles' heel of machine learning

- False confidence in prediction
- OOD generalisation
- Algorithmic biases
- Adversarial robustness
- Trustworthiness
- AI safety and misalignment
- …

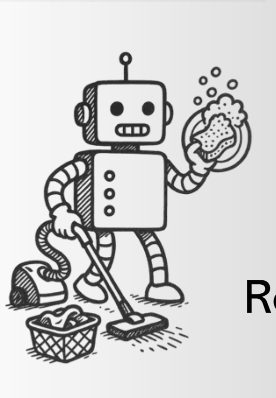
Known Unknowns

Discovering patterns from data under known uncertainties.

Large-scale Pattern Discovery



Learned Associations



Routine Tasks

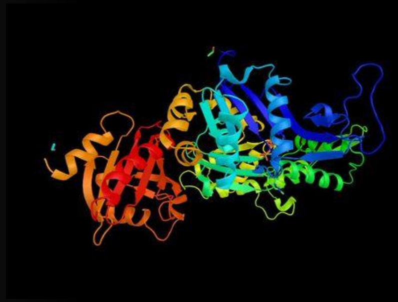
IPML

From Generalization to Robustness

“Bridging Systematicity and Safety”

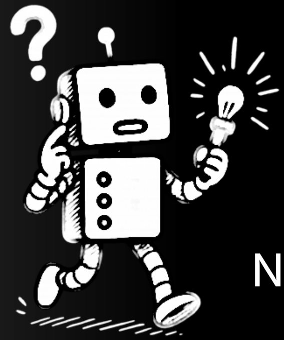
Unknown Unknowns

Generalizing to unforeseen situations under deep uncertainty.



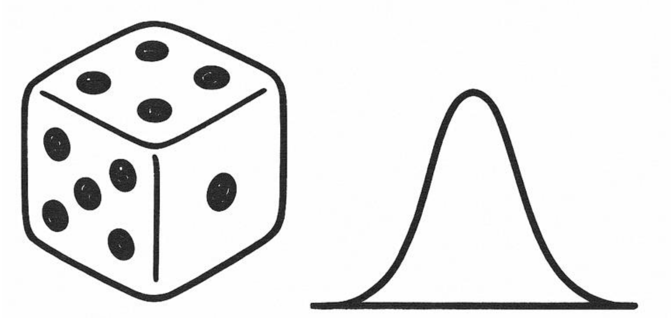
Exploration & Scientific Discovery

Adversarial Attacks

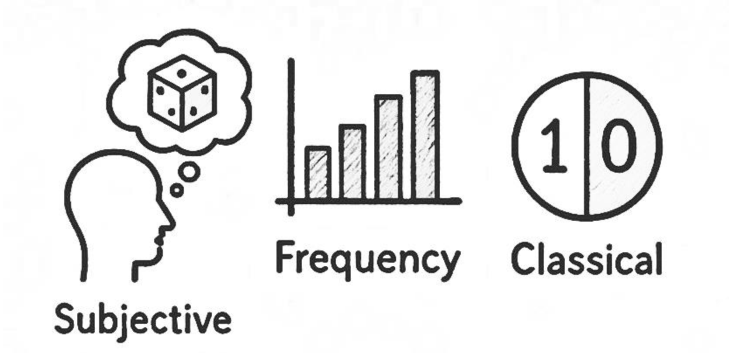


Novel Tasks

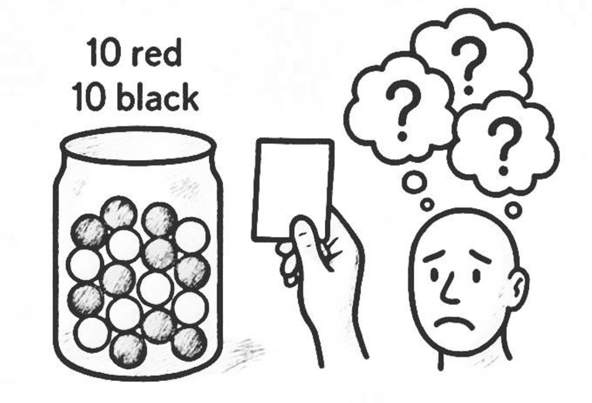
Outline



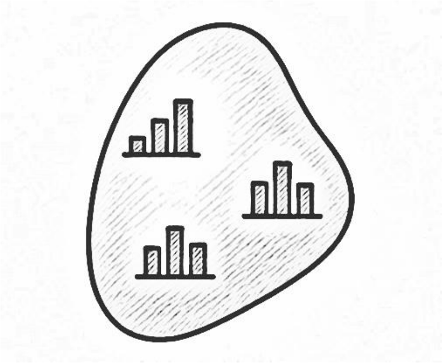
1. Uncertainty and Probability



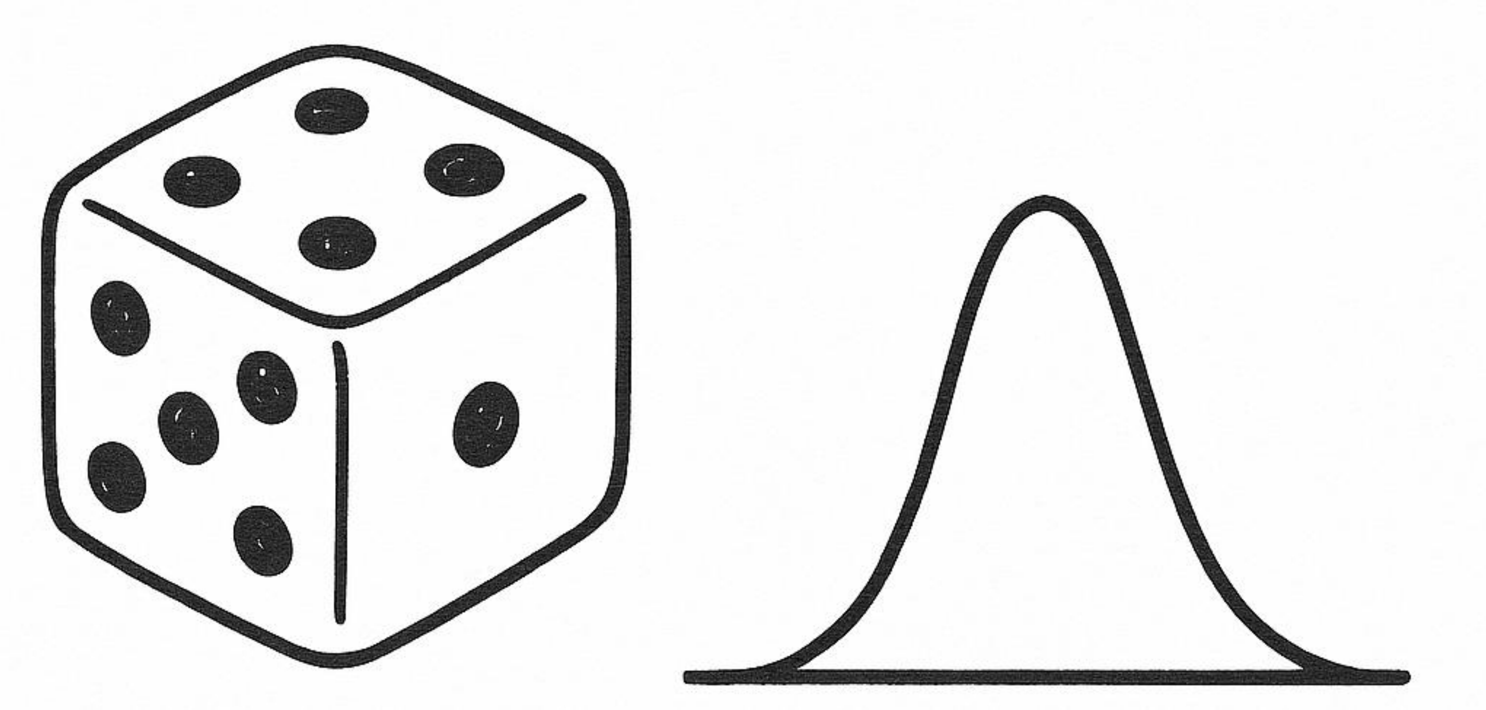
2. Interpretations of Probability



3. Pitfalls of Precise Probability



4. Imprecise Probability



Uncertainty and Probability

Uncertainty in Machine Learning

Data D

- Sampling uncertainty
- Label noise / annotation error
- Measurement noise
- Missing data / censoring
- Latent confounding
- Class imbalance
- Rare events

Aleatoric (irreducible)

Models M

- Model misspecification
- Parameter uncertainty
- Hyperparameter uncertainty
- Approximation error
- Training stochasticity
- Overfitting / underfitting
- Representation uncertainty

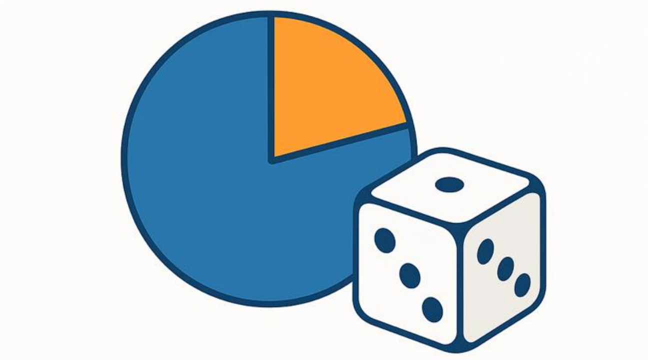
Epistemic (reducible)

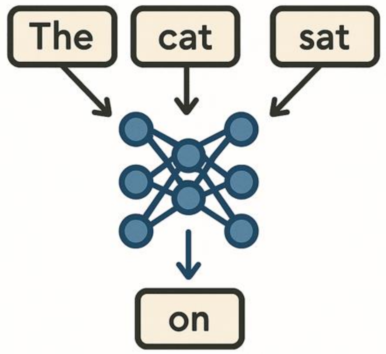
Environments E

- Distribution shift
- Concept drifts
- Intervention / policy shift
- Adversarial perturbations
- Hardware / system noise
- Social / contextual uncertainty
- Task redefinition

Out-of-distribution / Structural

Probability Quantifies Uncertainty





Probability is a measures of how likely an event is to occur.

- *The probability of a tossed coin landing on heads.*
- *The probability that it will rain tomorrow.*
- *Given a sentence, "Broccoli is …", the probability of "healthy" as the next word in the sentence.*
- *Given an image, the probability that it contains animals.*

Definition of (Additive) Probability

A **probability space** $(\Omega, \mathcal{E}, P)$ consists of a **sample space** Ω and **event space** $\mathcal{E} \subseteq 2^{\Omega}$ and a **probability function** $P: \mathcal{E} \to [0,1]$:

1. $P(\emptyset) = 0$, $P(\Omega) = 1$
2. $A \subseteq B \implies P(A) \leq P(B)$
3. $A \cap B = \emptyset \implies P(A \cup \mathrm{B}) = P(A) + P(B)$

Exercise: Show that probability is self-conjugate, i.e., for any $A, A^c \in \mathcal{E}, P(A) \ = \ 1 - P(A^c)$.

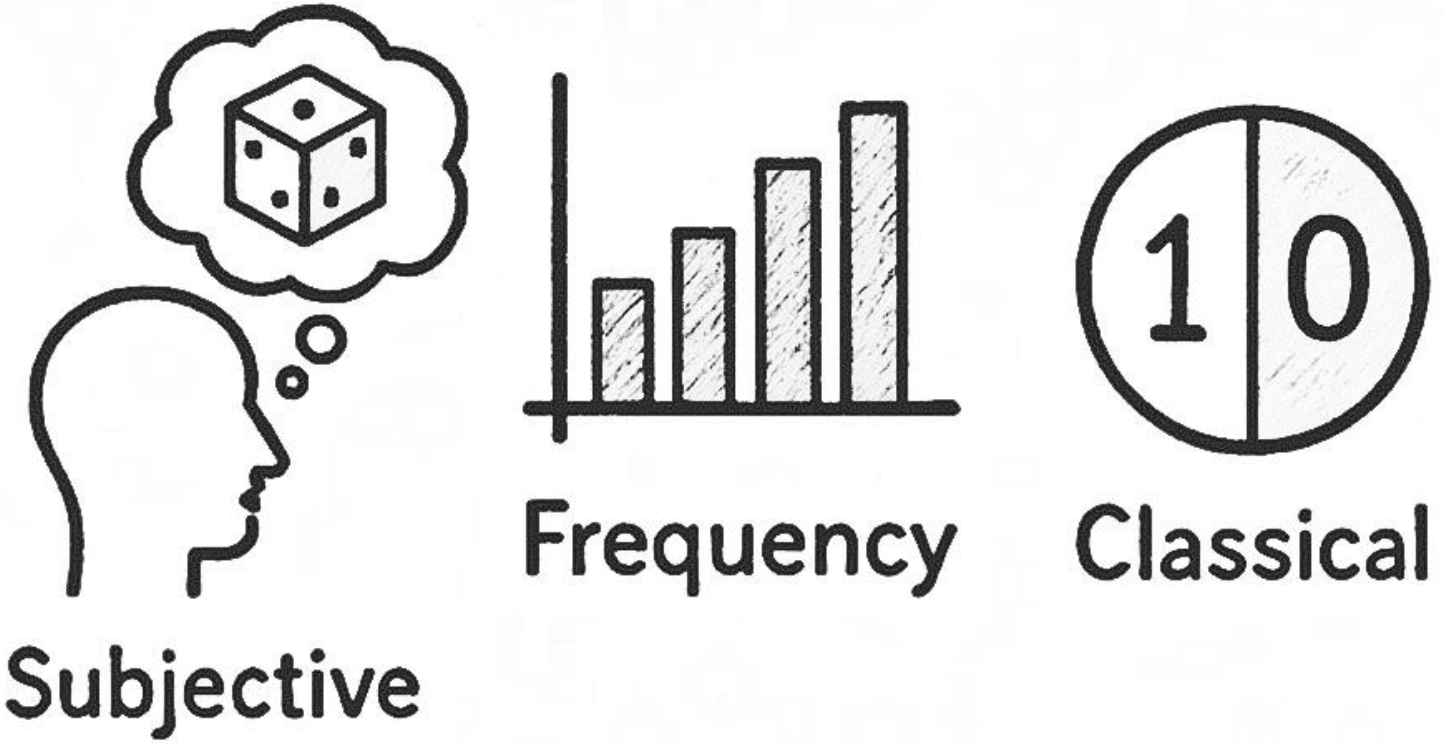
Illustration

- Consider the observed genetic variations: $\{X, Y, X, X, XY, YY, Y, YY\}$
- Treatments are applied to individuals with different genetic variations

	X	Y	XY	YY
f_1	0.1	0.4	0.7	0.8
f_2	0.3	0.7	0.1	0.3

p_X	p_Y	p_{XY}	p_{YY}
3/8	2/8	1/8	2/8

- **Reasoning**:
  - Probability: $P(\{X, Y\}) = P(X) + P(Y) =$ 3/8 + 2/8 = 5/8
  - Expectation: $E_p[f_1] =$ $p_X \cdot f_1(X) + p_Y \cdot f_1(Y) + p_{XY} \cdot f_1(XY) + p_{YY} \cdot f_1(YY) = 0.425$
  - What is the outcome with maximal probability?
  - Which treatment is more effective?



Interpretations of Probability

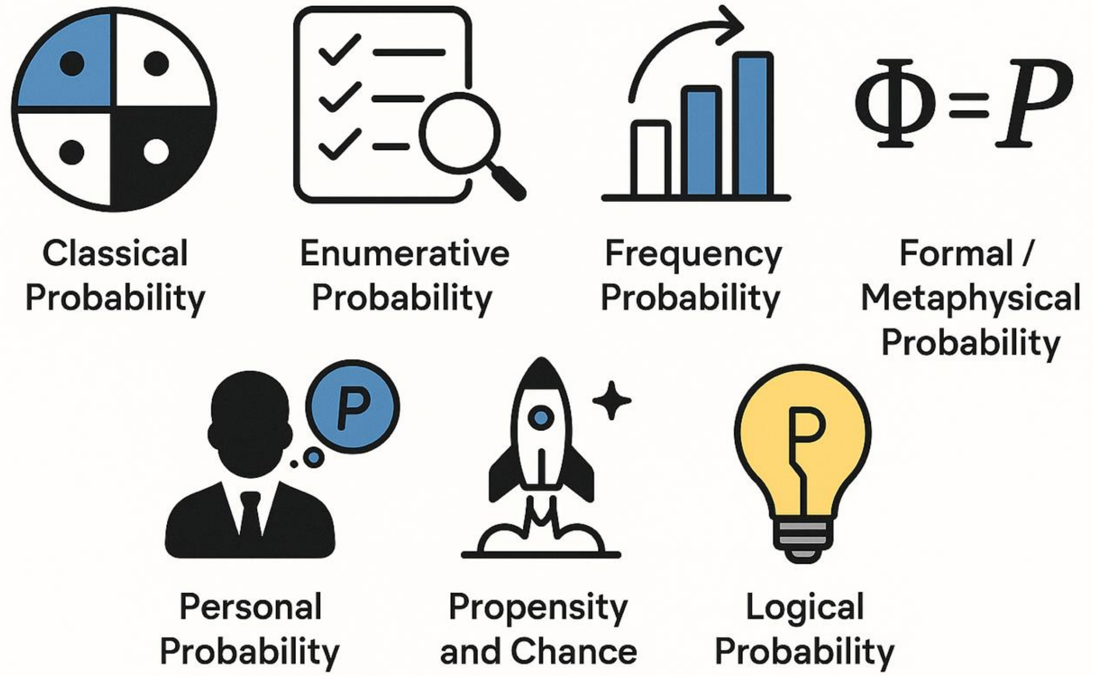
There is a 75% probability that it will rain tomorrow.

How should we interpret this probability?

There is 60% risk that an individual with features x will default on the loan.

How should we interpret this probability?

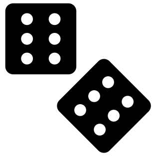
Interpretations of Probability



Recommended Reading:
- Dawid (2018), *On Individual Risk (Chapter 3)*.
- Hájek (2023), *Interpretations of Probability*, The Stanford Encyclopedia of Philosophy

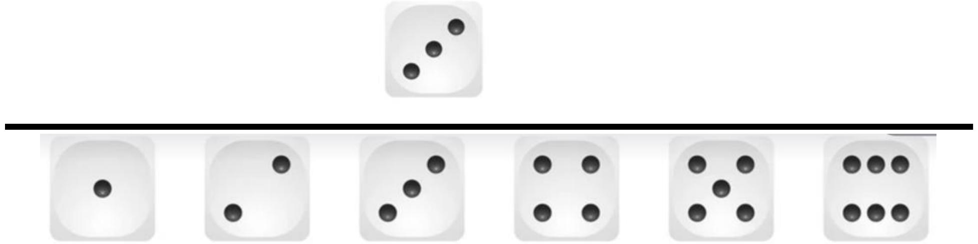
Classical Probability

Behavior of unbiased coins, packs of cards, roulette wheels (i.e., casino)



An experiment has several possible ***elementary outcomes****, but only one will occur when it is performed.*

$\boldsymbol{P(dice = 3)} =$



$= \mathbf{1/6}$

Some criticisms:

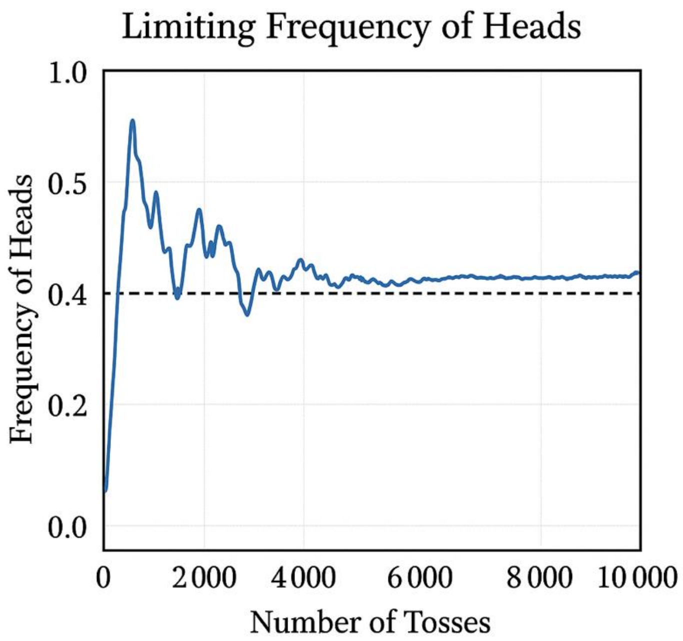
- Presuppose that each elementary outcome is **equally likely** to occur.
- There might be **more than one** natural way to define the elementary events.
  __Tossing two coins__: we can define either three elementary events: “0 heads”, “1 head”, and “2 heads” or four elementary events, “tail tail”, “tail head”, “head tail”, and “head head”.
- It’s problematic to cope with an infinite number of events

Enumerative Probability

Consider a finite collection of individuals (of any nature), on which we can measure one or more pre-existing attributes.

$$P(\text{smoker}) = \frac{\text{Number of } I_k \text{ who smokes}}{N}$$

Frequency Probability



An enumerative probability in an **infinite set**!

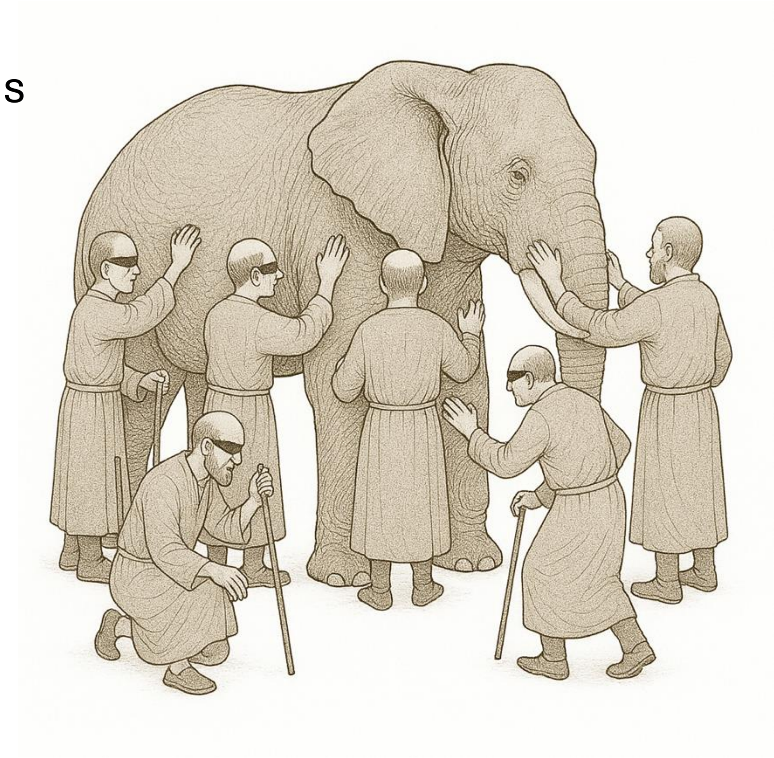
Consider **a sequence of coin tosses** from toss 1 up to toss N:

- f_N: The relative frequency of heads in this finite set (i.e., enumerative prob).
- As we increase N, f_N may approach some mathematical limit p, which we call the "**limit relative frequency**" of heads.
- The limiting value p may be termed the "**frequency probability**" of heads.

Some criticisms: The existence of p is an assumption and even if it exists, we may not have sufficient data to determine its value precisely.

Formal (Metaphysical) Probability

Probability is a *mathematical formalism* — a consistent calculus for reasoning about uncertainty, **without committing to what probabilities "mean" in the real world**.



Probability reflects something **real in the world**, not just our knowledge or conventions.
It's a statement about how the world itself *is* — about objective tendencies, propensities, or chances.

Personal (Subjective) Probability

A probability value p is associated with

1. An **individual event** E, e.g., the tomorrow's weather
2. The **individual**, say "You", who make the assignments, e.g., the forecaster
3. The **information** H available to You when making the assignment, e.g., historical weather data

*The probability p corresponds to the **odds at which You would be willing to bet** on the event E, e.g., rain tomorrow.*

Propensity and Chance

- A particular proposed coin-toss has a certain (typically unknown) “**propensity**” to yields heads, if it were to be conducted.
  - Championed by Popper (1959), and is still much discussed by philosophers—though hardly at all by statisticians.

- The “**Principal Principle**” of Lewis (1980):
  *if You learn that the **chance** of an event A is (say) 0.6, and nothing else, then Your personal probability of A should be updated to be 0.6.*

Logical Probability

It's about **how strongly one statement follows from another**, based purely on *logic and information*.

"All balls in this urn are either red or blue."

Without further information, what's the probability that the first ball you pick is red?

- You don't know the frequencies or propensities.
- But ***logically***, given symmetry, the propositions "it's red" and "it's blue" are equally supported.
- So you assign ½ to each — not because of repeated trials, but because that's the most ***rationally neutral*** assignment.

So, **logical probability** expresses how much *the premises imply the conclusion*, when viewed as a matter of rational consistency.

There is a 75% probability that it will rain tomorrow.

There is 60% risk that an individual with features x will default on the loan.

Avoiding Sure Loss

The fundamental **rationality** condition introduced by de Finetti:

> *"You should not be vulnerable to a combination of bets (based on your stated probabilities) that would guarantee a loss no matter what outcome occurs."*

Otherwise, a bookmaker could construct a "**Dutch book**" against you.

Precise Probability: Coherence $\Leftrightarrow$ No Dutch Book $\Leftrightarrow$ Avoiding sure loss
Imprecise Probability: Coherence $\Longrightarrow$ Avoiding sure loss, but not conversely

Why **Additive** Probability?

De Finetti's subjective probability represents the belief of a "**rational**" agent.

- Consider a **gamble** 1_A whose payoff is 1 unit of utility when the event A occurs, and 0 otherwise.
- A **fair price** associated to this gamble is denoted by $\boldsymbol{P}: \rightarrow [0,1]$
- The pricing scheme P is **coherent** if there is no finite collection of transaction (buy or sell) that guarantees a **sure loss:** For any events $A_1, A_2 \ldots, A_K$, there doesn't exist $\mathrm{c}_1, \mathrm{c}_2 \ldots, \mathrm{c}_K \in R$ such that

$$\sum_{k=1}^{K} c_k [1_{A_k} - \boldsymbol{P}(A_k)] < 0$$

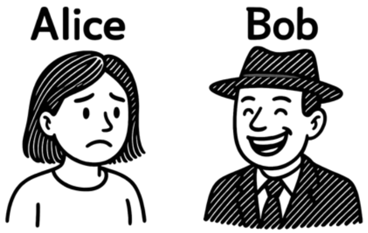
The Overconfident Weather Gambler

Alice believe she's good at predicting the weather, so she offers to buy and sell bets on whether **it will rain tomorrow (R)**.

- Alice's belief: $P(R) = 0.7,\ P(\neg R) = 0.7$

Bet 1: Pays €1 if it rains. Alice is willing to buy it for €0.7 (since she thinks rain is 70% likely).

Bet 2: Pays €1 if it doesn't rain. Alice is willing to buy it for €0.7 as well.



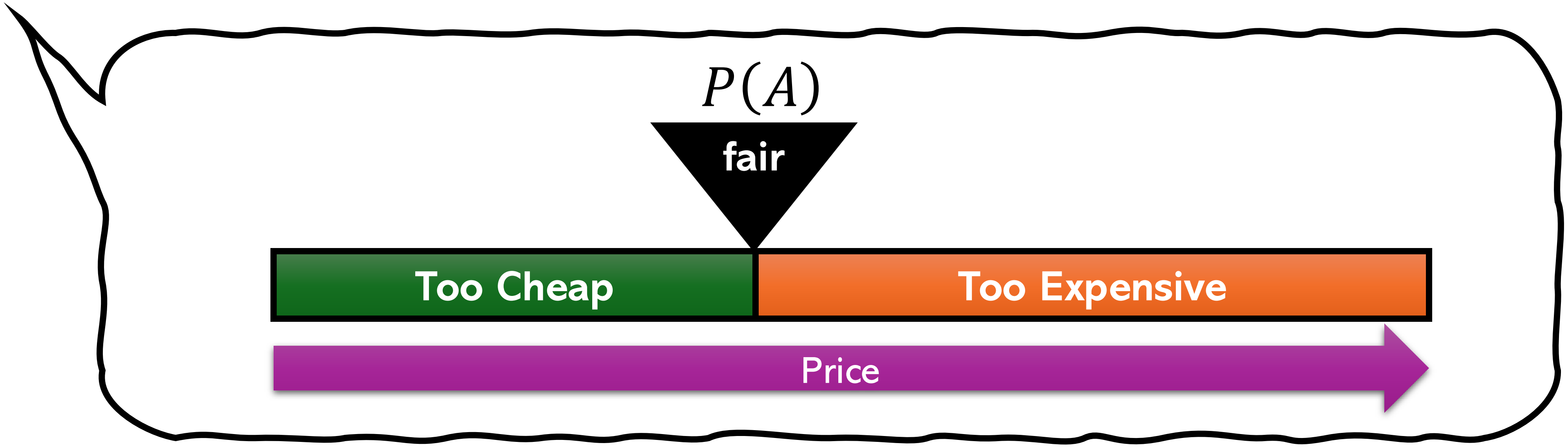
Bob, a clever bookmaker, **sells both bets** to Alice at her own stated prices.

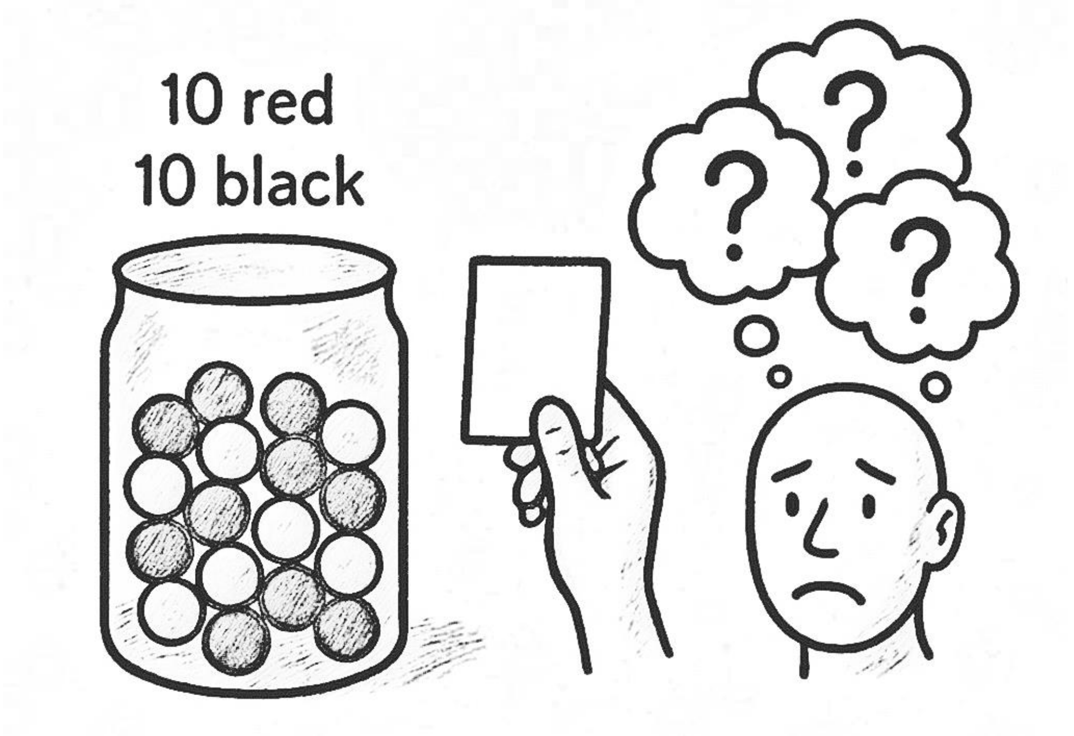
- Alice pays Bob: 0.7 + 0.7 = 1.4 euros.
- Tomorrow, exactly one event happens, either rain or no rain.
- Alice **receives only €1** back in any case.

Why **Additive** Probability?

How much would you pay for the gamble 1_A? What is a fair price?

Theorem (Coherence): *A pricing scheme P is* ***coherent*** *if and only if it is a (finitely additive) probability.*

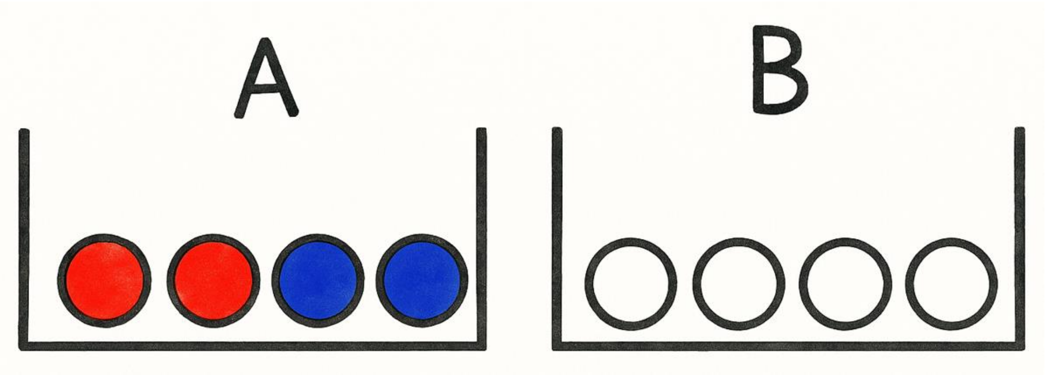




Pitfalls of Precise Probability

Ellsberg Paradox

Ambiguity Aversion



A person is shown two urns, A and B. In urn A, there are 50 red balls and 50 blue balls. There are red and blue balls in urn B with unknown proportion.

One ball is drawn at random from each urn:

1. Bet on **Ar** or **Ab** (indifferent)
2. Bet on **Br** or **Bb** (indifferent)
3. Bet on **Ar** or **Br** (**Ar** $\succ$ **Br** $\Longrightarrow p_{\mathrm{Br}} < p_{\mathrm{Ar}}$)
4. Bet on **Ab** or **Bb** (**Ab** $\succ$ **Bb** $\Longrightarrow p_{\mathrm{Bb}} < p_{\mathrm{Ab}} \Leftrightarrow 1 - p_{\mathrm{Br}} < 1 - p_{\mathrm{Ar}} \Leftrightarrow p_{\mathrm{Br}} > p_{\mathrm{Ar}}$)

Contradiction!

Aleatoric and Epistemic Uncertainties

$P(Y = \text{camel} \mid X = x) = 0.6$

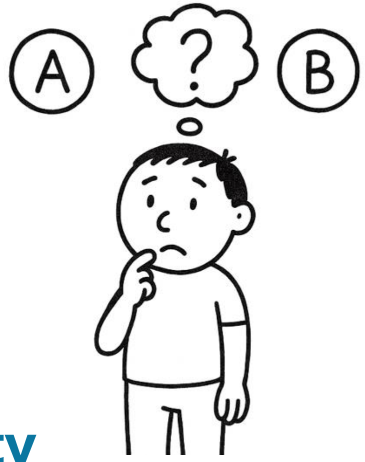
$$\boldsymbol{P_{\theta}(Y|X = x)}$$

A **single** probabilistic model cannot capture the **epistemic uncertainty**

Incompleteness and Incomparability

Precise probabilism assumes that some relation or other is **complete**: either $A \succ B$ or $B \succ A$ for some options A and B

Indifference
A determination that A and B are equally preferable ($A \sim B$)



Incomparability
Lack of such determination
(A and B are not comparable)

Weight and Balance of Evidence

$P(head) = 0.5$

*A probability is obtained through **experimentation** by repeatedly observing an event under controlled conditions and recording how often it occurs.*

A coin is tossed 100 times:

Guess the bias of the new coin:

$P(head) = 0.5$

*Due to lack of information, a probability is obtained by invoking the **principle of symmetry**, i.e., If there is no reason to believe one outcome is more likely than another, treat them as equally likely.*

Suspension of Judgement (Indecision)

- Sometimes, little or no information on which to base our conclusions
- In classical theory, **rational agents** model uncertainty using a **single probability measure** and choose between alternatives by maximizing expected utility:

$$a^* = \operatorname{argmax}_{a \in \mathcal{A}} E_{x \sim P}[\boldsymbol{u}(x, a)]$$

- For precise probability P, we have $I(X) = 1 - (P(X) + P(\neg X)) = 0$
- Some might argue $P(X) = 0.5$ ***is*** suspending judgement (maximum entropy), but it cannot signal the difference between **suspension of judgement** and **strong evidence of probability half**.

Collective Belief

- How to represent belief and uncertainty of a group of agents like committees, governments, and companies in which **conflicts** may arise?

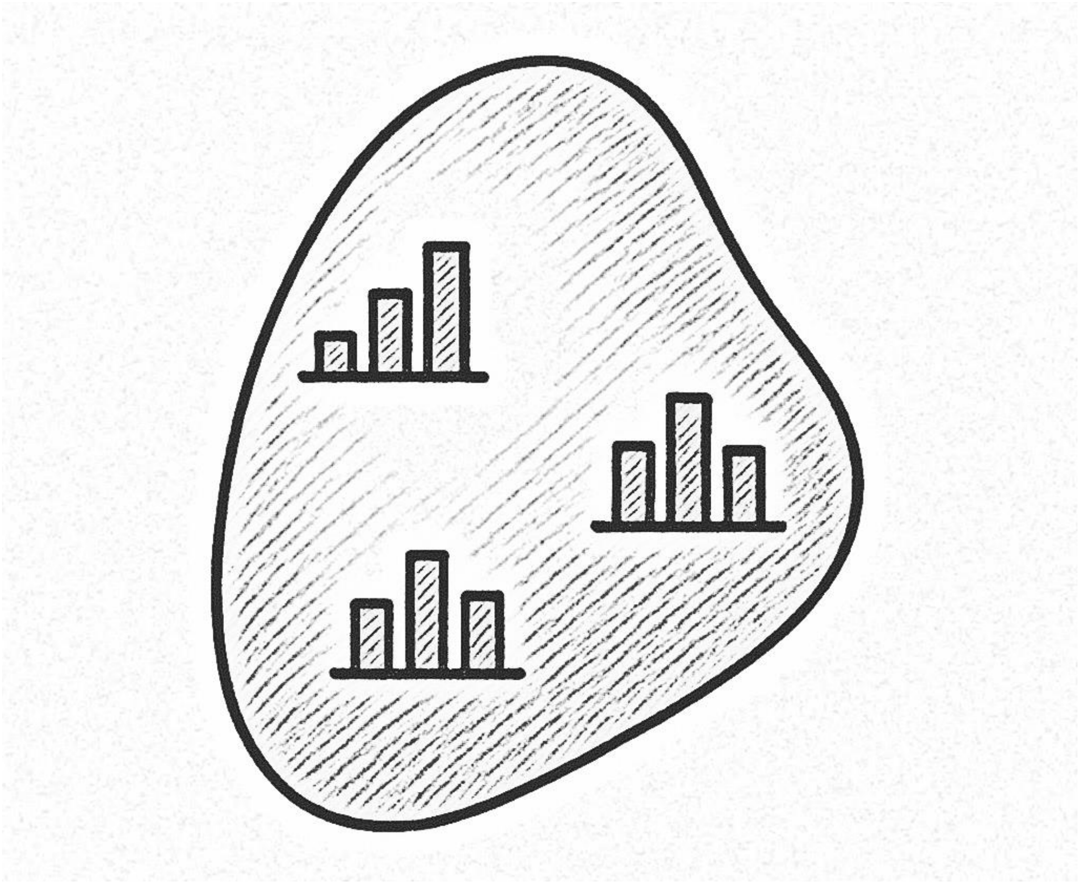
Precise Model of Group Agents:

For two probability P, Q and a **linear pooling** $R = P/2 + Q/2$:

$$P(X) = P(Y) = 1/3, \qquad \boldsymbol{P(X \mid Y) = P(X)}$$
$$Q(X) = Q(Y) = 2/3, \qquad \boldsymbol{Q(X \mid Y) = Q(X)}$$

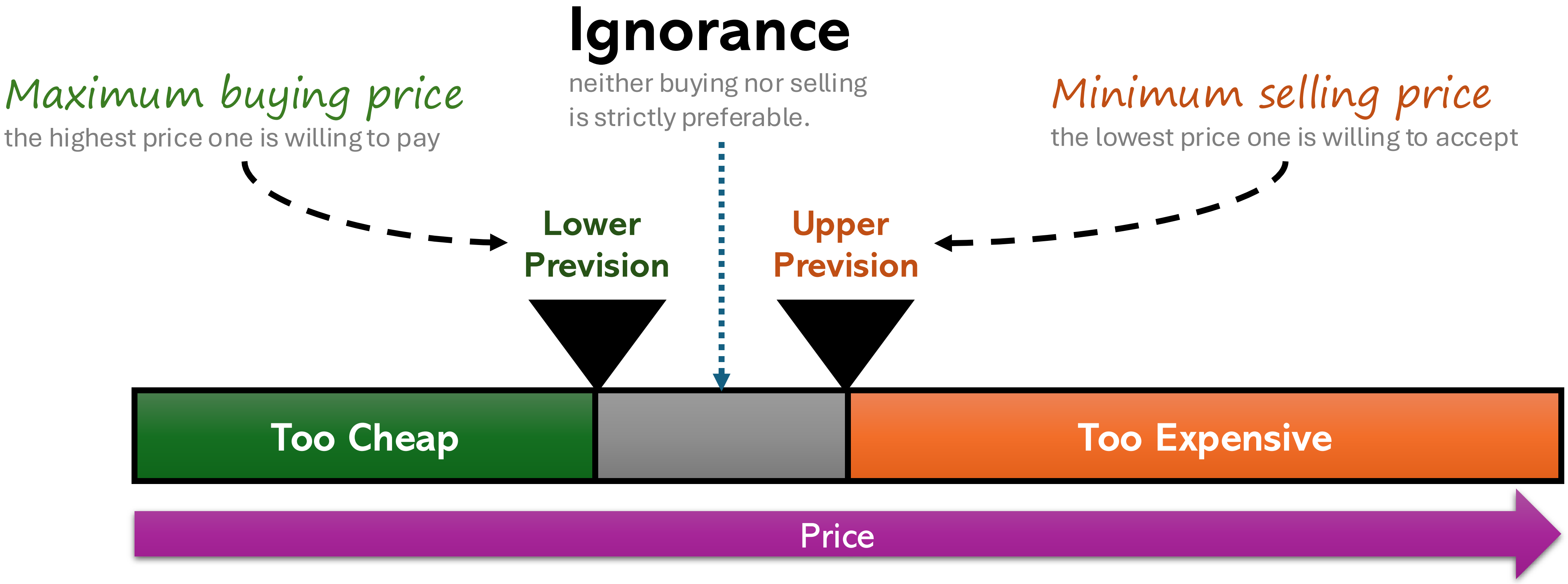
After pooling, we have $R(X \cap Y) = 5/18$ while $R(X)R(Y) = 1/4$.

The aggregation does not preserve the **statistical independence**



Imprecise Probability

Betting Perspective



Illustration

- Consider the *partial* observations:

$$\{\{X, \mathrm{Y}\}, \{\mathrm{X}, Y\}, \{X, \mathrm{Y}\}, \{X, \mathrm{Y}\}, XY, YY, \{\mathrm{X}, Y\}, YY\}$$

- Treatments are applied to individuals with different genetic variations

	X	Y	XY	YY
f_1	0.1	0.4	0.7	0.8
f_2	0.3	0.7	0.1	0.3

p_X	p_Y	p_{XY}	p_{YY}
[0,5/8]	**[0,5/8]**	1/8	2/8

- How can we perform **reasoning** with these probabilities?

Probability Intervals

- Let X be a random variable taking values in a *finite set* $\mathcal{X}$
- A **probability interval** $(\underline{p}, \overline{p})$ is a pair of lower and upper probability mass functions satisfying
  1. $0 \leq \underline{p}_x \leq \overline{p}_x \leq 1$ for all outcome $x \in \mathcal{X}$ (**bounded**)
  2. $\sum_{x \in \mathcal{X}} \underline{p}_x \leq 1 \leq \sum_{x \in \mathcal{X}} \overline{p}_x$ (**proper**)
  3. $\underline{p}_x \geq 1 - \sum_{z \neq x} \overline{p}_z$ and $\overline{p}_x \geq 1 - \sum_{z \neq x} \underline{p}_z$ for all $x \in \mathcal{X}$ (**reachable**)

- A set of compatible pmfs is called the **credal set**:

$$\mathcal{C} = \{p : \underline{p}_x \leq p_x \leq \overline{p}_x\}$$

Lower and Upper Probabilities

The lower and upper probabilities can be defined as

$$\underline{P}(S) := \max\left\{\sum_{x\in S}\underline{p}_x, 1-\sum_{x\in S^c}\overline{p}_x\right\}, \qquad \overline{P}(S) := \min\left\{\sum_{x\in S}\overline{p}_x, 1-\sum_{x\in S^c}\underline{p}_x\right\}$$

Exercise: Consider the following lower and upper pmf:

	X	Y	XY	YY
$\overline{p}$	3/8	3/8	5/8	5/8
$\underline{p}$	1/8	1/8	3/8	0/8

Then, calculate:

- $\overline{P}(\{Y, XY\}) = ?$
- $\underline{P}(\{X, Y\}) = ?$

Lower and Upper Previsions

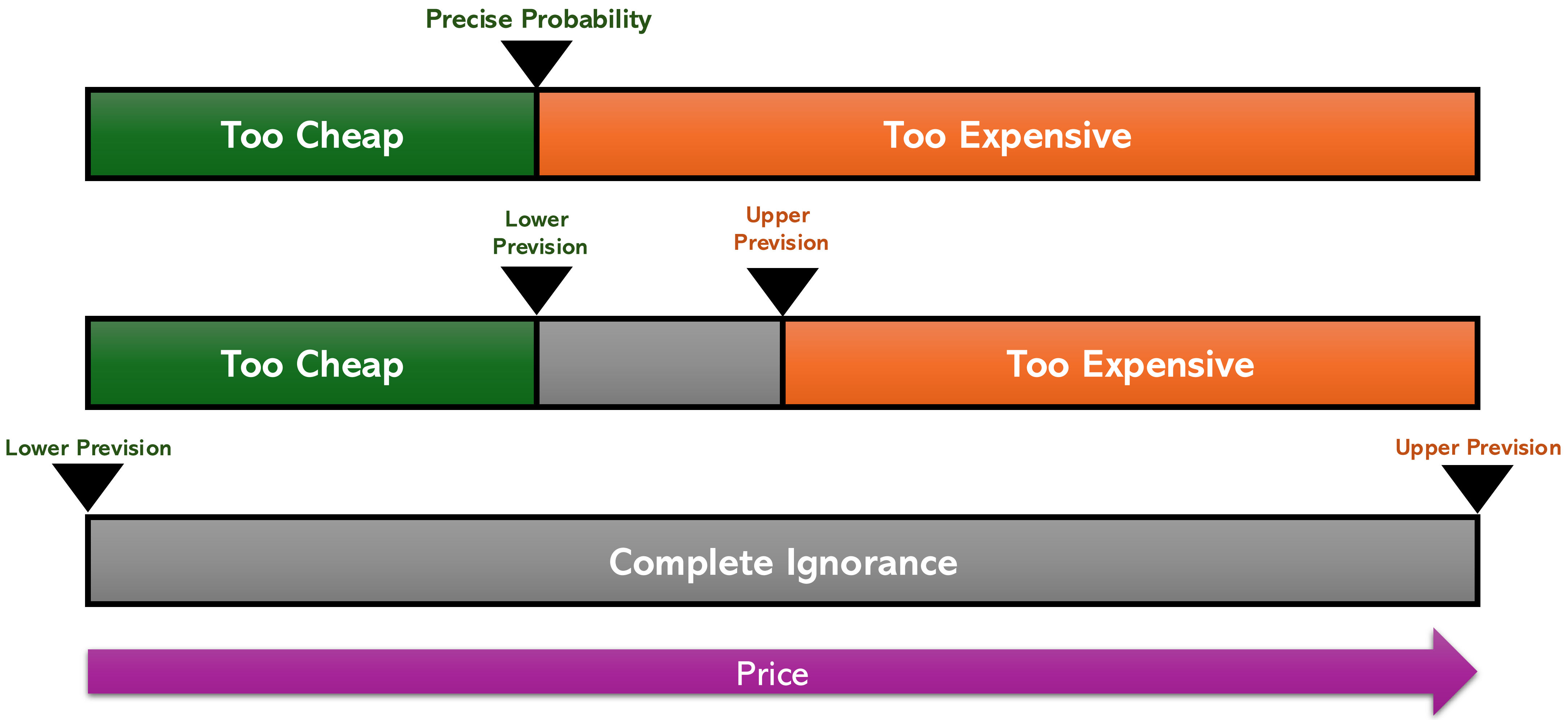
- Let $\mathcal{C}$ be a non-empty subset of a set of all probabilities, i.e., $\mathcal{C} \subseteq \mathcal{P}$. The **lower and upper previsions** can be defined as

$$\underline{P}(S) \coloneqq \inf_{p \in \mathcal{C}} P_p(S), \qquad \overline{P}(S) \coloneqq \sup_{p \in \mathcal{C}} P_p(S)$$

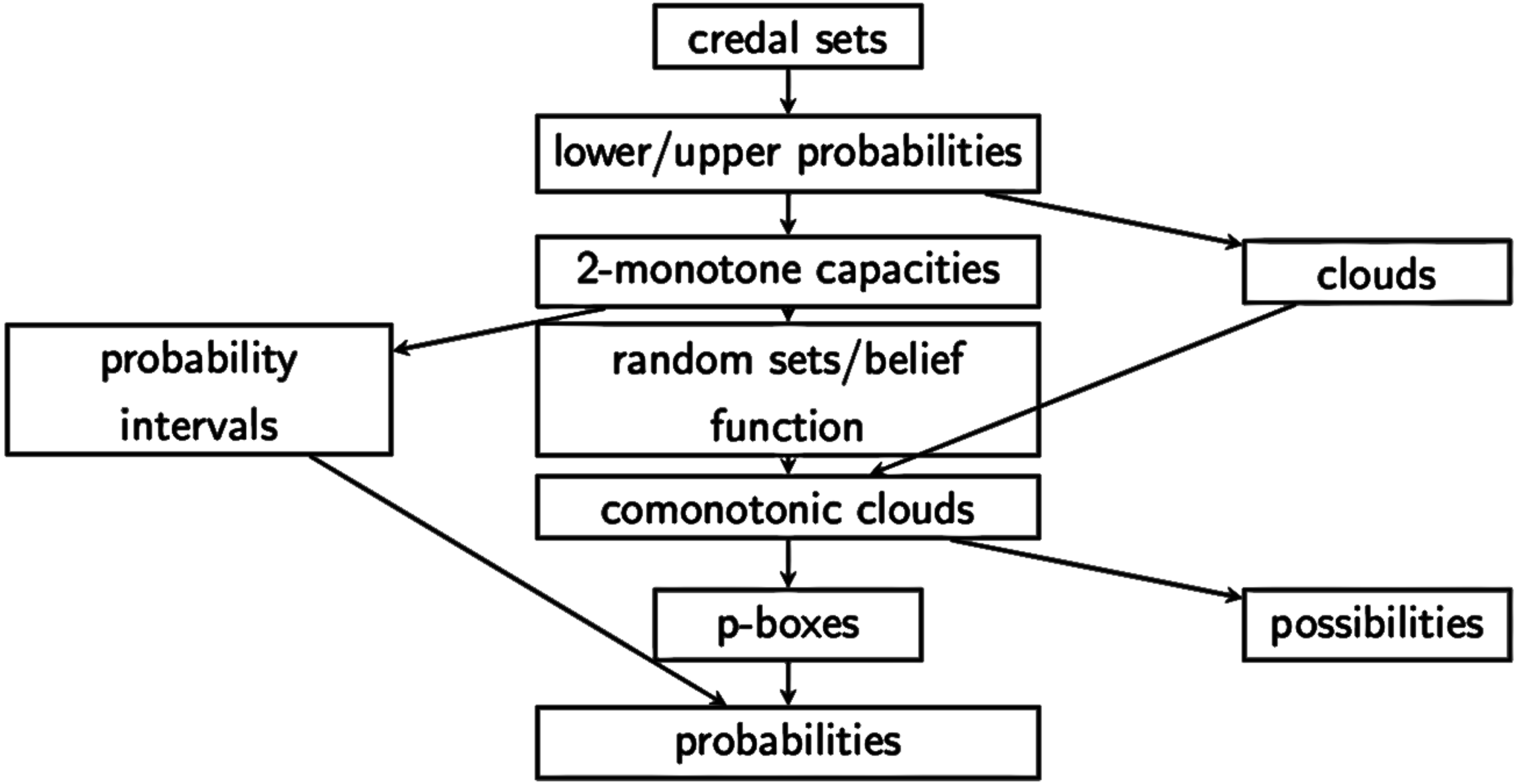
$$\underline{E}(f) \coloneqq \inf_{p \in \mathcal{C}} E_p(f), \qquad \overline{E}(f) \coloneqq \sup_{p \in \mathcal{C}} E_p(f)$$

- Examples of credal sets:
  - The vacuous credal set $\mathcal{P}^{\mathrm{S}} \coloneqq \{ p \in \mathcal{P} : p(S) = 1\}$ for some $S \in \mathcal{E}$.
  - The vacuous credal set $\mathcal{C} \coloneqq \mathcal{P}$ **(complete ignorance)**
  - The singleton credal set $\mathcal{C} \coloneqq \{p\}$ **(precise)**
  - and many more.

Degree of Imprecision



Summary of IP Models



Exercise

A coherent pricing scheme P is a (finitely additive) probability

Exercise 1: Prove that a *pricing scheme that is not finitely additive is not coherent.*

Solution: For any disjoint events A and B, we assume without loss of generality that $P(A \cup B) \leq P(A) + P(B)$. Then, the bookkeeper can buy the gamble $\mathbb{1}_{A\cup B}$ from You, and then sell You the gambles $\mathbb{1}_A$ and $\mathbb{1}_B$. Your earning is

$$P(A \cup B) - \big(P(A) + P(B)\big) - (\mathbb{1}_{A\cup B} - \mathbb{1}_A - \mathbb{1}_B) < 0$$

This guarantees **sure loss.** Hence, P is not coherent.

A coherent pricing scheme P is a (finitely additive) probability

Exercise 2: Prove that a *pricing scheme that is finitely additive is coherent.*

Solution: Consider any finite disjoint events $A_1, A_2, \dots, A_K$ and any buy or sell actions $c_1, c_2, \dots, c_K \in [-1, +1]$. Then, Your earning can be represented by a random variable:

$$W = \sum_{k=1}^{K} c_k \left[\mathbb{1}_{A_k} - P(A_k)\right]$$

which has zero expectation. Hence, there cannot be finite transactions that result in $W < 0$ all the time. Consequently, a pricing scheme that is finitely additive is coherent.