Making confident decisions with model ensembles

Joe Roussos

Roman Frigg

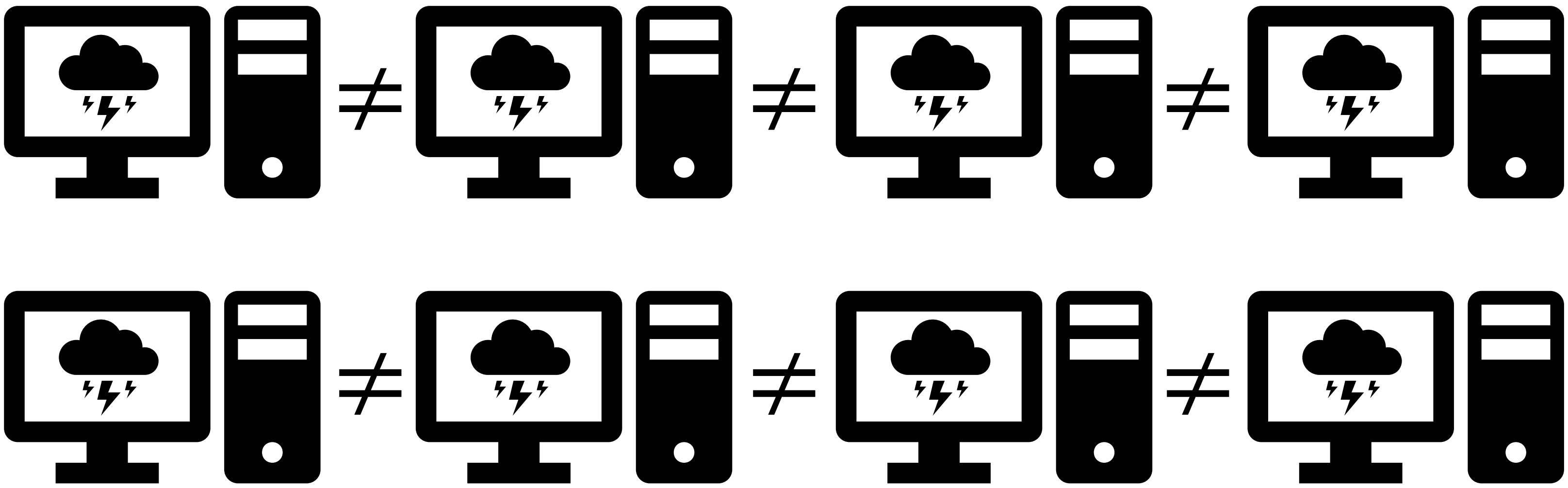
11 June 2018

Models to Decisions, Cambridge

London School of Economics

The Problem

How do we use model ensembles to inform decision-making, in a way that reflects and makes use of scientific uncertainty?



Our Decision Problem

How much does it cost to offer insurance against natural catastrophes?

Basics of the Decision Problem

The price of an insurance contract is a function of the **probability of the event** insured against

Two nice features of this:

1. Probabilities of natural catastrophes like hurricanes are provided by **scientific models**, which combine statistical analysis and physical understanding
2. Probabilistic model outputs made it natural for us to approach the problem with techniques from statistics/decision theory using **imprecise probabilities**

Toy Example

Simplified problem:

- You are a new insurer
- You want to sell a *single* insurance contract on house damage due to hurricanes.
- Event *E*: “a hurricane strikes Fort Lauderdale in 2019”.

What should you charge for this contract?

→ Need to know *p(E)*

The Dream Answer

Take perfect model and calculate $p(E)$, and plug it into a pricing model.

But …

- There is no perfect model. Of necessity models omit factors (known and unknown) and make idealisations.
- Many models and impossible to decide between them on the basis of available evidence.
  - *Florida Commission on Hurricane Loss Prevention* 2007 assessment: ensemble of 972 models.

The Real Answer

- Buy an ensemble of predictive models from a commercial modelling company
  - The ensemble members are chosen such that they reflect scientific disagreement
  - There are known inadequacies with all models

→ How does this ensemble inform pricing?

Toy example "model outputs"

You consult a modelling firm

"There are 10 models, and they disagree!"

Model	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
p(E)	0.0070	0.0083	0.0071	0.0074	0.0091	0.0076	0.0061	0.0092	0.0068	0.0086
Weight	0.2368	0.0729	0.2071	0.1575	0.0158	0.0317	0.1157	0.0173	0.1148	0.0300

"But we've averaged them using scoring rule R"

"The answer is *p(E)=0.0072*"

We think we can do better…

The Confidence Approach

We want to make explicit use of the following:

- What is at stake in the decision
- Uncertainty attitude of the decision maker
- The nature and spread of evidence available

Point of departure: There is something wrong with the “linear” process of decision-making just described. Issues of model uncertainty and their effect on the “answer” cannot be separated from the decision problem we want to solve.

Overview of the Approach

1. Assess **how important** the decision is to the agent
2. Using this, settle on **how confident** the agent wants to be to make this decision (using simplified levels of confidence: Low, Medium, High)
3. Use the **scientific evidence** (model outputs) to construct an answer which trades off specificity and robustness according to the confidence level required
   - Confidence is generated by **weight of evidence** = f(quantity, quality, diversity)
   - We will consider a set of nested claims—representing less specific but more reliable potential answers—and then classify them into the levels of confidence that they licence: Low, Medium, High
4. Finally, we select the claim which best fits the confidence required

Step 1: What is at stake?

Stakes of the decision: the agent’s assessment of how important it is.

e.g., What’s the worst that could happen?

Convention: number on a 0-to-1 scale.

0: You don’t care (e.g. £1 bet)

1: Highly significant (e.g. you’re shot if you lose)

Toy example: Stakes

- This contract will constitute your whole business and so the risk of ruin is very high.
- Still, no one's life is at stake and there is no impact on anything outside of the realm of this decision.

Conclusion: The stake is moderately high

Let us use s=0.75

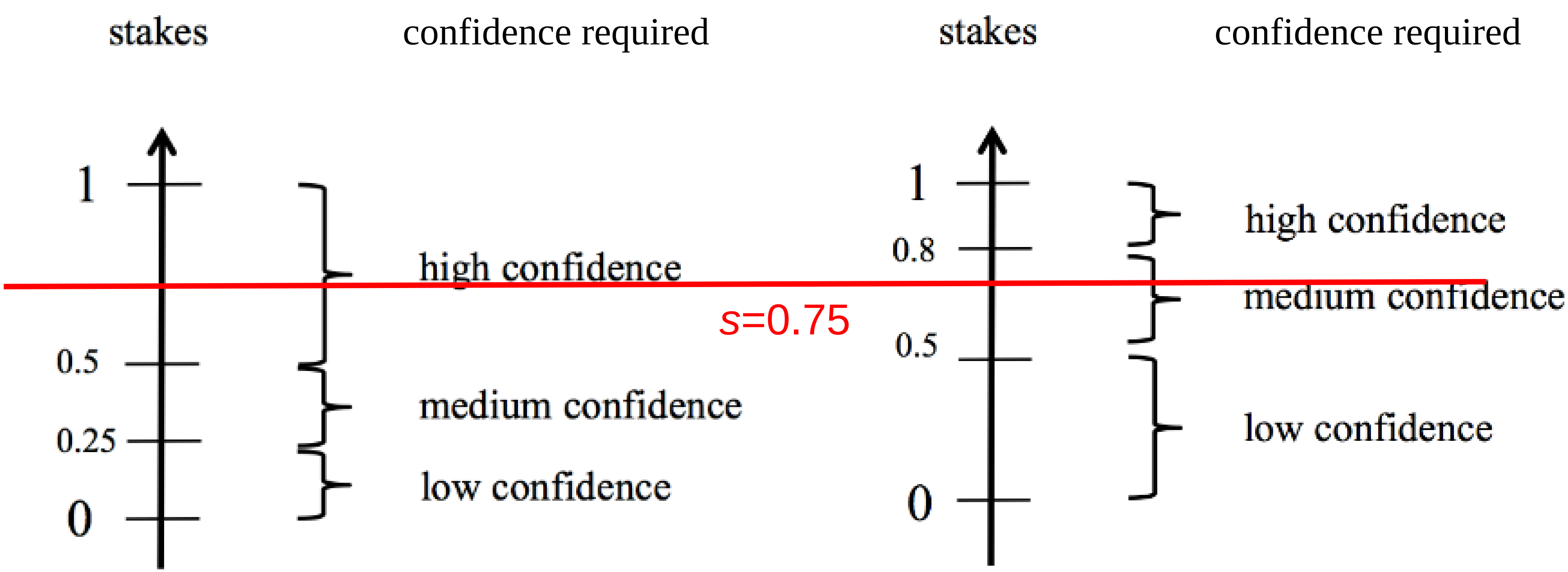
Step 2: Cautiousness

Given the importance of the decision, how confident do you want to be in order to act?

→ Cautiousness: function from stakes to “levels of confidence”

- Cautiousness represents uncertainty *attitude*.
- It will be subjective and will need to be elicited.
- Can be different for different agents.

Simple Examples: Cautiousness



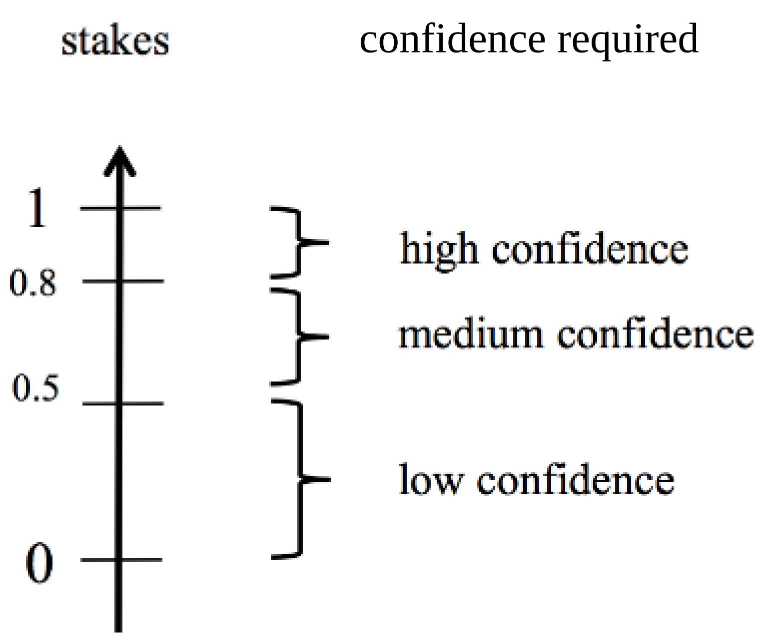
Cautious Agent

Bold Agent

Toy Example

Insurance of natural catastrophes involves significant uncertainty, so you can’t be *overly* uncertainty averse.

Let’s use the “bold” attitude:



Progress

We now know

(1) How important the decision-maker thinks this decision is: ***s=0.75***

(2) Given that, how confident they want to be in order to decide: ***Medium***

In real situations, (1) and (2) will both be informed by other decisions they make, and the nature of their field

Step 3: Nested Intervals

We now turn to the evidence base: in this case, outputs from models

We use these to construct a series of nested claims

The probability of a hurricane striking Fort Lauderdale in 2019 is…

- = 0.007
- between 0.007 and 0.0072
- between 0.0068 and 0.0072
- …

Where these values are **model outputs**

Assumption

Best model assumption: there is a best model.

In our example: m_1 – *chosen by scoring rule R*

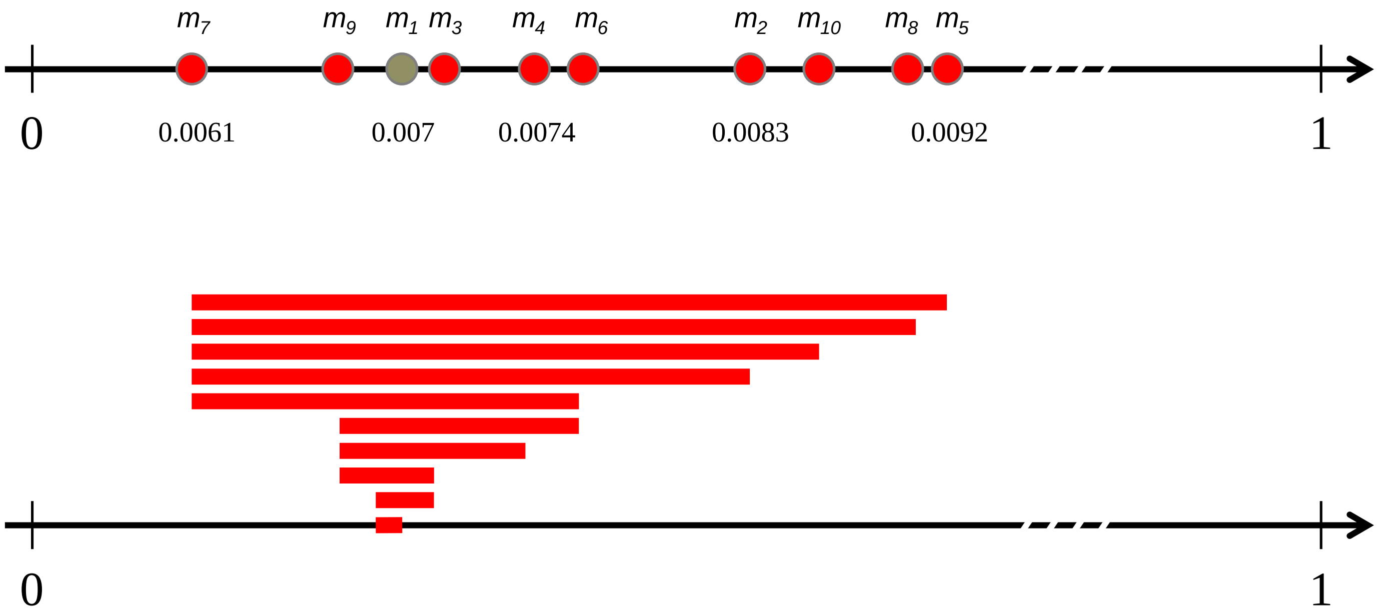
Our lowest level, most specific claim is that the probability of the event just is 0.007.

We can form wider intervals by including the predictions in the order of their distance from the best model.

Let’s discuss in the Q&A

Step 3: Nested Intervals

From model outputs to nested intervals



Step 4: Confidence Grading

Confidence is generated by examining the weight of evidence supporting a claim.

Aim: attach to each interval a confidence level.

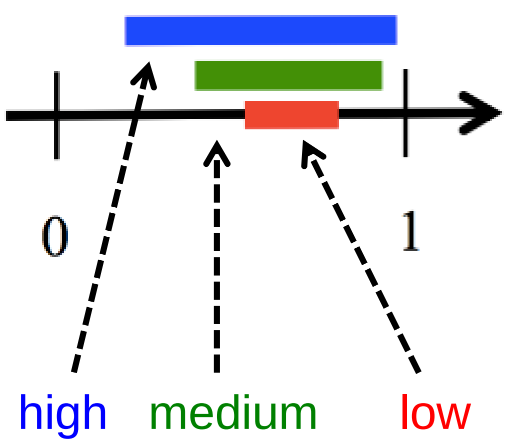
Important: increased confidence does *not* change the probabilities; it makes us more confident that probabilities we have are right. (Analogy: QM.)

Step 4: Confidence Grading

Convention:

- Three “levels” of confidence
- Low, Medium, High

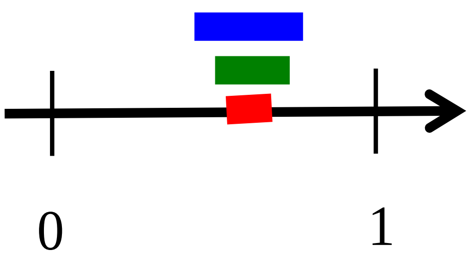
Logic dictates: less confidence in more precise claims.



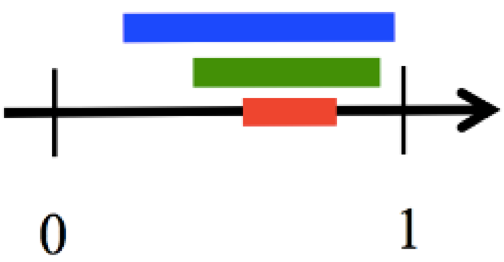
Step 4: Confidence Grading

Confidence grading reflects the state of scientific understanding

- Width of intervals reflects confidence-precision trade-off on some claim
- Wide/narrow intervals show that weight of evidence for projections is low/high.



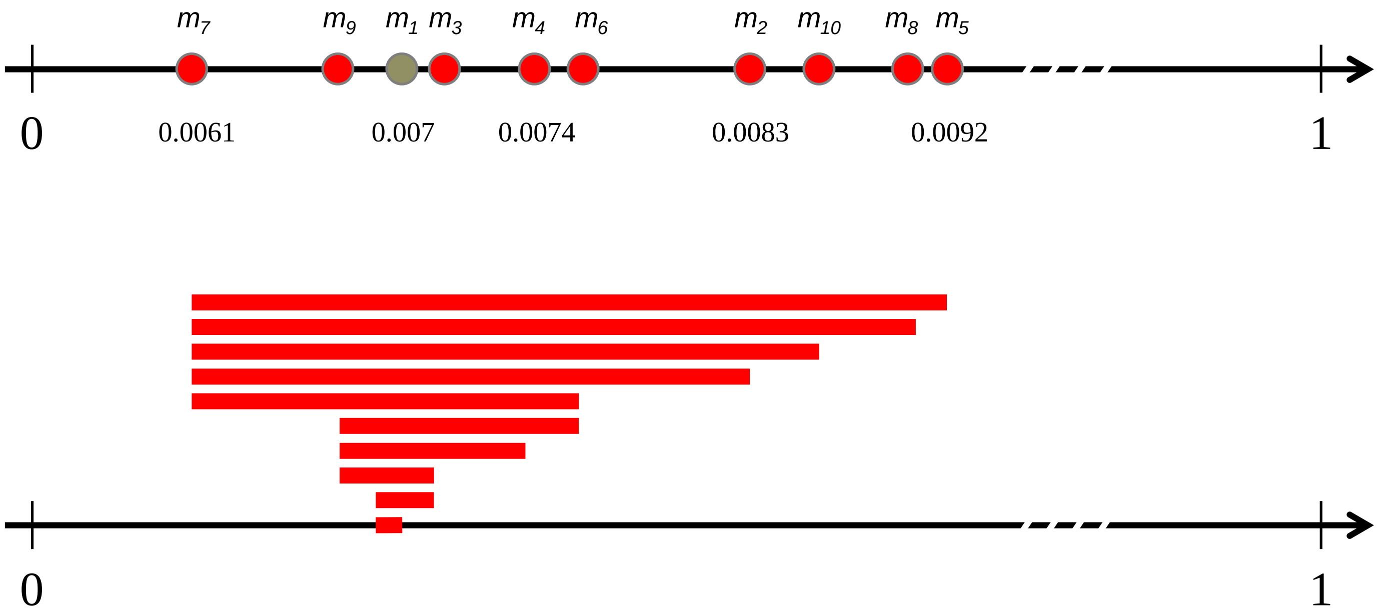
Good understanding



Poor understanding

Toy Example

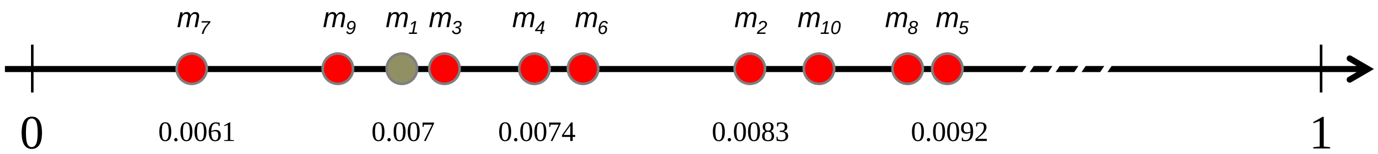
Recall our nested interval structure

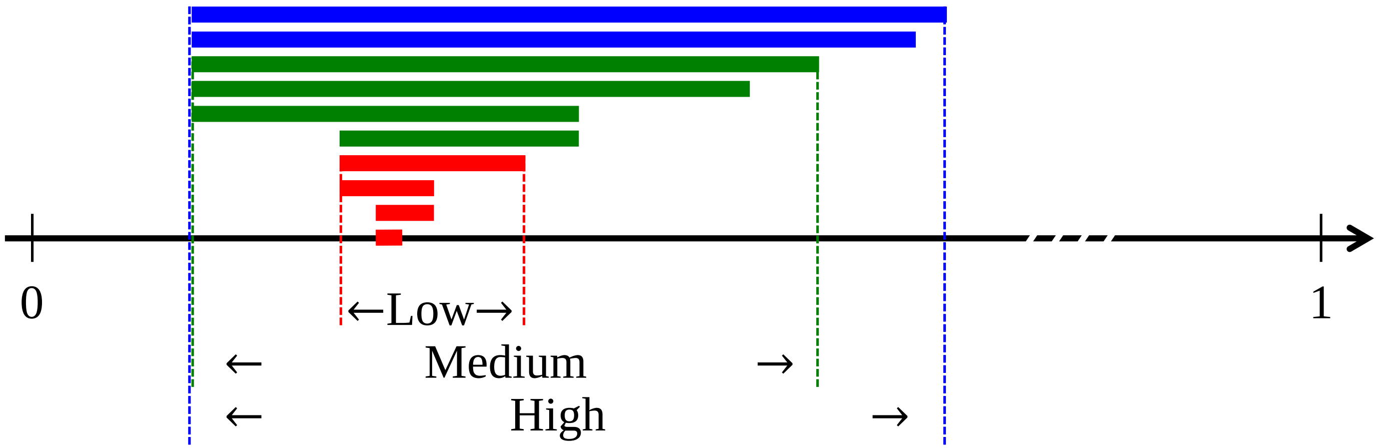


The only evidence I've described is the collection of model outputs → they are all we can use to assign confidence

Toy Example

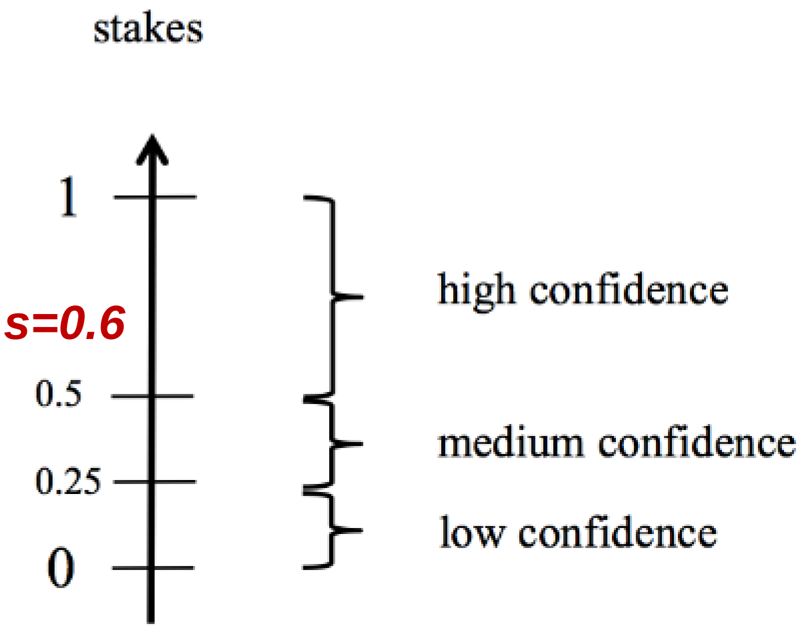
Step 4: Assign confidence levels

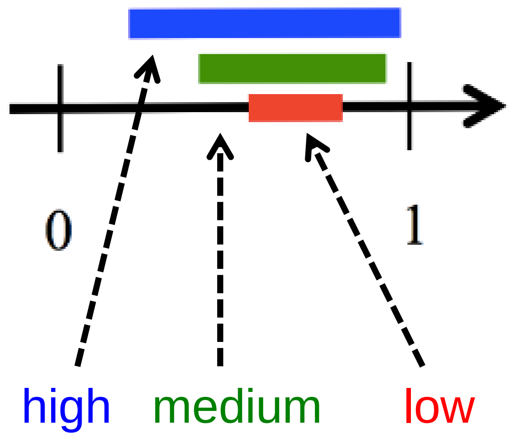




Step 5: Select "your" interval

Apply your cautiousness function to your assessed stakes

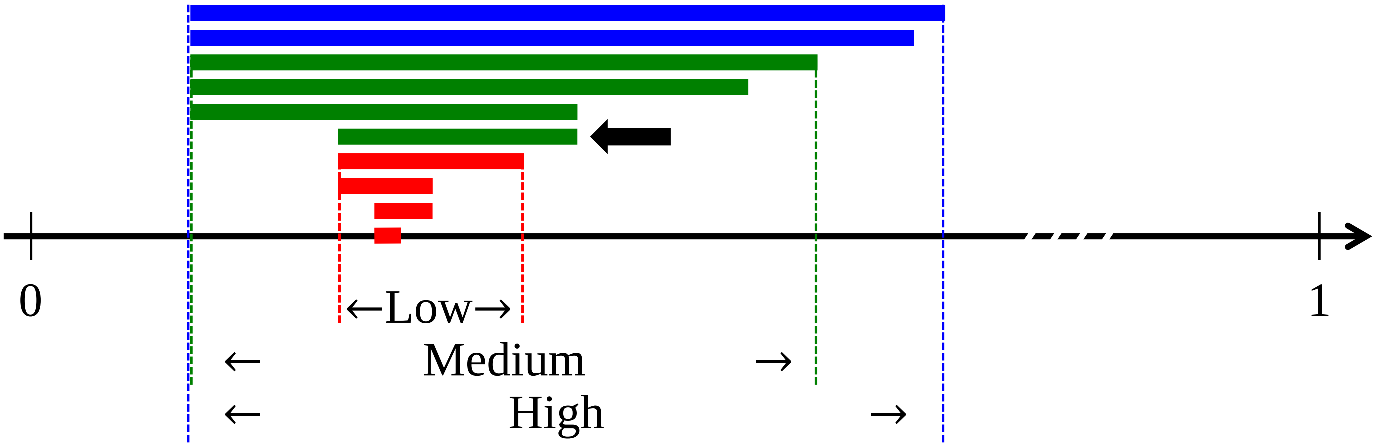




You choose the blue interval.

Toy Example

Given s=0.75, **medium** confidence is required.



The narrowest interval in the medium range is I_5=[0.0068,0.0076].

Choose this interval for the decision!

Progress

We have settled on a particular interval of probabilities, I_5=[0.0068,0.0076].

It represents a particular trade-off between

- **specificity** (narrowness of the interval), which is valuable in distinguishing between courses of action, and
- **robustness** (breadth), which ensures we are confident enough in our decision given the uncertainty

Step 6: Make a decision

What we have now is a **set of probabilities**. We need a decision-rule that works with these. There are many candidates; we will use:

Maximin Expected Utility: *A* is preferred to *B* iff the minimum expected utility of *A* is greater than the minimum expected utility of *B*.

Colloquially: Choose the option that has the best outcome if things turn out to be as bad as they can be.

Toy Example

Our interval is I_5=[0.0068,0.0076].

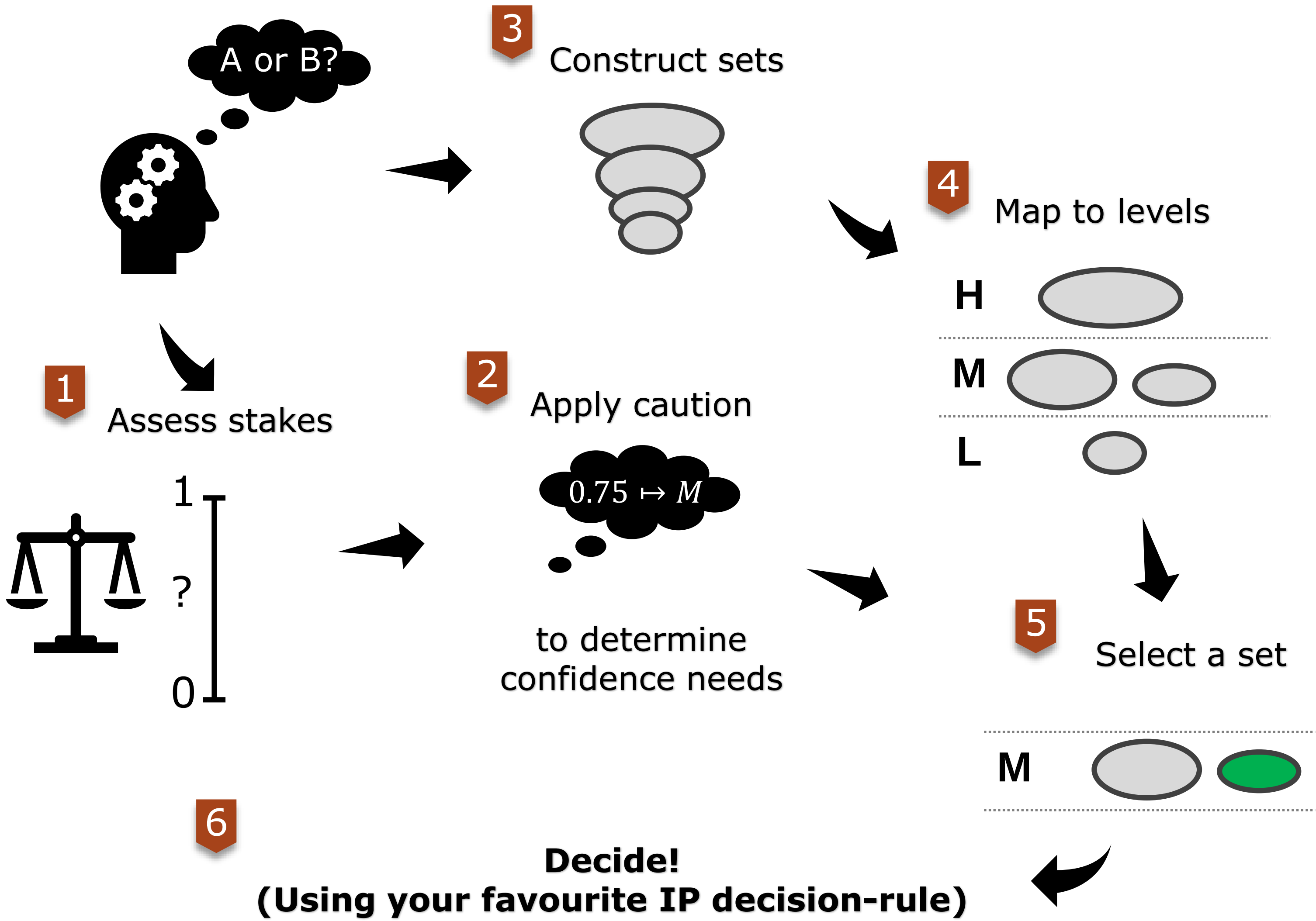
Let's assume for simplicity that things go badly when the probability is highest

Therefore work with $p(E) = 0.0076$

Compare: averaging $p(E)=0.0072$

5% higher → significant change in pricing! Many contracts shouldn't have been sold if we are right

Summary of the Procedure



Conclusion

- Main benefit of our approach is the structure it provides for managing uncertainty
- We build in DM caution, while avoiding *ad hoc* “ambiguity premiums”
- We better reflect hard cases: if there is no best model, no ranking, then no nesting – use the full range
- Open question: how best to construct nested intervals

Thanks

References

Bradley, Richard. 2017. *Decision Theory with a Human Face*. Cambridge University Press. https://doi.org/10.1017/9780511760105

Bradley, Richard, Casey Helgeson, and Brian Hill. 2017. "Climate Change Assessments : Confidence, Probability, and Decision." *Philosophy of Science* 84 (3): 500–522. https://doi.org/10.1086/692145.

Frigg, Roman, Seamus Bradley, Hailiang Du, and Leonard A. Smith. 2014. "Laplace's Demon and the Adventures of His Apprentices." *Philosophy of Science* 81 (1): 31–59. https://doi.org/10.1086/674416.

Frigg, Roman, Leonard A. Smith, and David A. Stainforth. 2015. "An Assessment of the Foundational Assumptions in High-Resolution Climate Projections: The Case of UKCP09." *Synthese* 192 (12): 3979–4008. https://doi.org/10.1007/s11229-015-0739-8.

Hill, Brian. 2013. "Confidence and Decision." *Games and Economic Behavior* 82: 675–92. https://doi.org/10.1016/j.geb.2013.09.009.

Stainforth, David A., M.R. Allen, E.R. Tredger, and Leonard Smith. 2007. "Confidence, Uncertainty and Decision-Support Relevance in Climate Predictions." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (June): 2145–61. https://doi.org/10.1098/rsta.2007.2074.

Thompson, Erica, Roman Frigg, and Casey Helgeson. 2016. "Expert Judgment for Climate Change Adaptation." *Philosophy of Science* 83 (5): 1110–21. https://doi.org/10.1086/687942.

Wüthrich, Nicolas. "Conceptualizing Uncertainty: An Assessment of the Uncertainty Framework of the Intergovernmental Panel on Climate Change." in *EPSA15 Selected Papers: The 5th conference of the European Philosophy of Science Association in Düsseldorf*. Springer, 2016. https://doi.org/10.1007/978-3-319-53730-6_9

A draft of the accompanying paper to this presentation is also available upon request.