Rational Confidence and Unknown Unknowns

Isaac Swift

Hong Kong Baptist University

19 April 2025

Confidence-Knowledge Graph

What do you expect to see when you graph confidence against knowledge?

Confidence-Knowledge Graph

What do you expect to see when you graph confidence against knowledge?

- Charles Darwin noticed it and remarked, “Ignorance more frequently begets confidence than does knowledge.”
- Mark Twain made a similar observation, “When I was a boy of 14, my father was so ignorant I could hardly stand to have the old man around. But when I got to be 21, I was astonished at how much the old man had learned in seven years.”
- In *As You Like It*, William Shakespeare put it clearly, “the fool doth think he is wise, but the wise man knows himself to be a fool.”

Questions

- What confidence-knowledge graphs are rationalizable?

Are you above average?

Many studies with similar results

- Lake Wobegon Effect
- 88% of drivers think they are above average
- Svenson 1981
- Buunk and Van Yperen

The opposite effect also exists

Questions

We now have two big questions

- What confidence-knowledge graphs are rationalizable?
- How many rational people can think they're above average?

Unknown unknowns

"We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know."
-United States Secretary of Defense Donald Rumsfeld

You're reading papers in a new field you want to study.

- You don't know how many paper there are on the subject
- You may not find every paper that has been written
- You may not have time to carefully study every paper you find

Set up

You will stop reading papers for one of three reasons

- You stop if you've read all the papers
  - Suppose there are 0, 1, 2, or 3 papers
  - Each is equally likely
- You stop if you don't find any more papers
  - After reading each paper (0, 1, or 2), if there is another paper you find it with probability 2/3
- You stop if you run out of budget
  - You have a budget of 2 papers

Confidence-knowledge graph

Conditional on the number of papers you've read, what is the probability that you've read everything available?

Consider first someone who's read zero papers

- It's possible that there aren't any papers (ex ante 25%)
- It's possible that there are papers they didn't find ($\frac{1}{3}$ conditional on existence)

$$prob = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{3}{4}\frac{1}{3}} = \frac{1}{2} = 50\%$$

Confidence-knowledge graph

We can compute the confidence level of someone who has read one paper

$$prob = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{2}{3}\frac{1}{3}} = \frac{3}{5} = 60\%$$

Confidence-knowledge graph

We can compute the confidence level of someone who has read two papers

- Say don't have the budget to read a third paper, so they don't search for one
  - There are equally likely to be two or three papers, so confidence is 50%
- Say they do still search for a third even though they won't read it
  - If they find a third paper, confidence is 0%
  - If they don't find a third, confidence is 75%
  - On average, the confidence level will be 50% or 0% (or 25%)

Confidence-knowledge graph

Result

- Not the 45 degree line
- Can slope downward
- Hump-shaped

Comparisons

Are you above average?

- Mass of independent learners
- Each person guesses whether they are (strictly) above average or not
- Payoff of 1 if guess is correct 0 if incorrect
- What fraction of the population thinks they are above average?
- It depends on the realization of the number of papers

Comparisons

Suppose there aren't any papers on the subject

- Everyone has read zero papers
- 100% of the population guesses (correctly) they are not above average

Comparisons

Suppose there aren't any papers on the subject

- Everyone has read zero papers
- 100% of the population guesses (correctly) they are not above average

Suppose there is one paper on the subject

- People who have read one paper think they're above average (correctly)
  - There is a 60% chance there is only one paper
  - Two-thirds of the population will have read one paper

Comparisons

Suppose there aren't any papers on the subject

- Everyone has read zero papers
- 100% of the population guesses (correctly) they are not above average

Suppose there is one paper on the subject

- People who have read one paper think they're above average (correctly)
  - There is a 60% chance there is only one paper
  - Two-thirds of the population will have read one paper

Suppose there are two or three papers on the subject

- People who have read two papers know they're above average
- 67% of the population thinks they are above average while only 44% actually are

Conclusions

Confidence-knowledge graphs

- Can be downward sloping or hump-shaped
- Can we bound all possible graphs?

Comparisons

- Can be above or below 50%
- People can be wrong
- Below when realization is low, above when realization is high
- Statistical property

Model

You start learning at time zero until you stop for one of two reasons

- $t^I \sim \mathcal{I}(t)$ is the amount of information available
- $t^S \sim \mathcal{S}(t)$ is the amount of information you would conditionally find

You stop learning at the minimum of those two times.
$\hat{t} = \min\{t^I, t^S\}$

The budget from the example isn't needed for any results, but doesn't hinder any results either.

Definitions

Definition

Given some distributions, $\mathcal{I}(t)$ and $\mathcal{S}(t)$, the **confidence-knowledge curve**, $f : \mathbb{R}_+ \to [0, 1]$, is the average posterior probability of having all information conditional on stopping at time t.

So, it's just the probability that $\hat{t} = t^I$ on average.

Definition

A function $g : \mathbb{R}_+ \to [0, 1]$ is **rationalizable** if there exists some distributions $\mathcal{I}(t)$ and $\mathcal{S}(t)$ such that $g(t)$ is the confidence-knowledge curve induced by those distributions.

Proposition 1

What confidence-knowledge curves are rationalizable?

Proposition

Every function $g : \mathbb{R}_+ \to [0, 1]$ is rationalizable.

Intuition

Consider the following distributions

$$\mathcal{I}(t) = 1 - e^{-2t}, \qquad \mathcal{S}(t) = 1 - e^{-t}$$

At the point where you stopped, it's twice as likely that you were stopped by $\mathcal{I}(t)$ than by $\mathcal{S}(t)$.

- Any flat confidence-knowledge curve can be obtained by scaling the arrival rate for $\mathcal{I}(t)$ up or down.

Intuition

Generally, the average confidence level is just the ratio of hazard rates.

$$p = \frac{\sigma^{\mathcal{I}}}{\sigma^{\mathcal{I}} + \sigma^{\mathcal{S}}}$$

So, the confidence-knowledge graph will be increasing (or decreasing) whenever the hazard rate for $\mathcal{I}(t)$ is increasing (or decreasing) relative to the hazard rate of $\mathcal{S}(t)$.

Proof

Since hazard rates are practically unrestrained, any function is rationalizable.

Let $g(t)$ be some function mapping into $[0, 1]$.

- Remove the budget because it isn't needed
- Set $\mathcal{S}(t) = 1 - e^{-t}$
- You will obtain $g(t)$ for the confidence-knowledge graph if the hazard rate of $\mathcal{I}(t)$ is equal to $\frac{g(t)}{1-g(t)}$.

The differential equation $\frac{\mathcal{I}'(t)}{1-\mathcal{I}(t)} = \frac{g(t)}{1-g(t)}$ has a solution.

$$\mathcal{I}(t) = 1 - e^{-\int_0^t \frac{g(\tau)}{1-g(\tau)} d\tau}$$

Comparisons

How many people can think they're above average?

- Consider the example from earlier.

$$\mathcal{I}(t) = 1 - e^{-2t}, \qquad \mathcal{S}(t) = 1 - e^{-t}$$

Everyone can believe they are above average.

- Greater likelihood of being above average, not above average of averages

Lake Wobegon

Proposition

For any value $p \in [0, 1]$, *there exist distributions of* $\mathcal{I}(t)$ *and* $\mathcal{S}(t)$ *such that* p *fraction of the population believes they are above average.*

Proof

A simple example

- Let $\mathcal{I}(t) = 1 - e^{-2t}$
- Define $\mathcal{S}(t)$ in two parts
  - Probability mass of $1 - p$ at $t = 0$
  - Density $\mathcal{S}(t) = p - pe^{-t}$ for all $t > 0$

Everybody Can be Wrong

Proposition

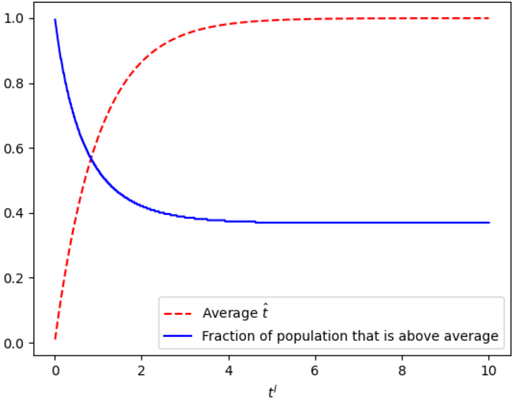
For any $\epsilon > 0$, *there exist distributions* $\mathcal{I}(t)$ *and* $\mathcal{S}(t)$ *and a realization* t^I *such that* $1 - \epsilon$ *fraction of the population believes they are above average but are actually below.*

With the same distributions, there is another realization of t^I where more than $1 - \epsilon$ believe they are below average but they are actually above average.

Complexity

Proposition

The fraction of the population that wrongly believes they are above average is weakly increasing in the realization t^I.



Implications

What does this mean for studies of overconfidence?

- Any confidence-knowledge curve is rationalizable
- Any comparison to the average is rationalizable
- Studies where you see an objective outcome aren't immune
- What if the agents communicate?

Literature

There are a million studies documenting overconfidence

- Kruger and Dunning (1999), Buunk and Van Yperen (1991), Svenson (1981), Malmendier and Tate (2005)

There are several answers to the 88% are above average fact

- Benoît and Dubra (2011), Zábojník (2004), Brocas and Carillo (2007), Köszegi (2006), Moore and Healy (2008)

- Modica and Rustichini (1999)