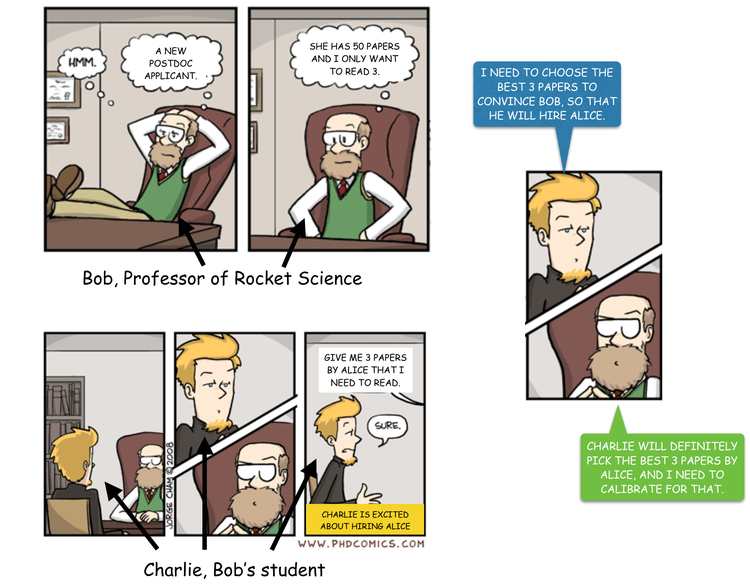
When Samples Are Strategically Selected

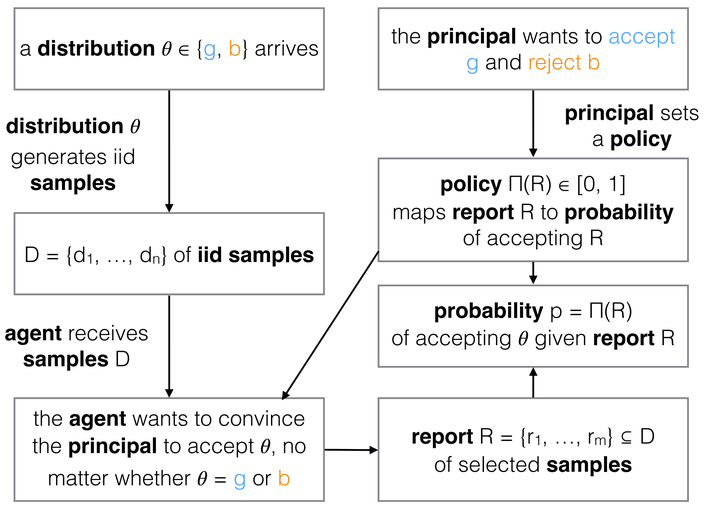
Hanrui Zhang Yu Cheng Vincent Conitzer

Duke University

PhD Comics: Academia in 20 Years



The Problem



Implications of Strategic Samples

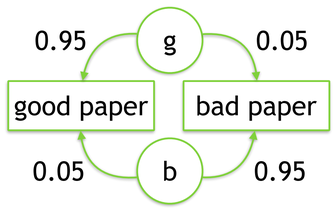
- How does **strategic selection** affect the principal's policy?
- Is classification easier with **iid samples** or **strategic samples**?
- Does **diversity** help the agent / the principal?

Example.

- Each postdoc applicant has n = 50 papers; the professor wants to read only m = 1
- Each paper can be good or bad; good applicants write good papers with a higher probability

A reasonable policy: **accept** iff the reported paper is **good**

An "easy" world:



strategic samples:

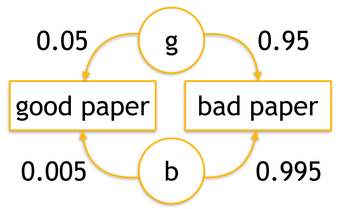
$p_g = 1 - (1 - 0.95)^{50} \approx 1$
$p_b = 1 - (1 - 0.05)^{50} \approx 0.92$

iid samples:

$p_g = 0.95$, $p_b = 0.05$

Strategic selection hurts!

A "hard" world:



strategic samples:

$p_g = 1 - (1 - 0.05)^{50} \approx 0.92$
$p_b = 1 - (1 - 0.005)^{50} \approx 0.22$

iid samples:

$p_g = 0.05$, $p_b = 0.005$

Strategic selection helps!

The Cure: Irrelevant Information

A modified policy: **accept** iff the paper is **good** and the **42nd letter is an "A"** (which happens, say, independently w.p. 0.05)

Now in the easy world:

$$p_g \approx 1 - (1 - 0.05)^{50} \approx 0.92,\ p_b \approx 1 - (1 - 0.003)^{50} \approx 0.14$$

Strategic selection doesn't hurt so much!

One Sample: Case Solved

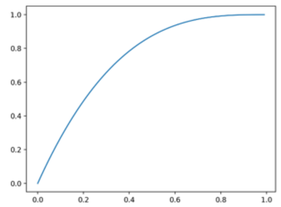
Theorem.

When m = 1, any **Pareto optimal** deterministic policy:

- **orders the sample space** according to the **likelihood ratio** g(x) / b(x), and
- the limiting acceptance probabilities satisfy:

$$p_g + (1 - p_b)^r = 1$$

where r = max (g(x) / b(x)) is the maximum likelihood ratio



The ROC curve (or the Pareto frontier) when r = 3

Multiple Samples: Exponentially Diminishing Error

Theorem.

With m samples, there is a deterministic policy

- which **orders the sample space**, and
- whose limiting error rate is at most $\exp(-(1 - r^{-1/2})^2 m / 2)$

– Does **every** pareto optimal policy **order the sample space**?

– Answer: No, **Pareto optimality sometimes requires diversity**!

– Okay, so what exactly do optimal policies look like?

– Answer: In general (1 < m < n), we don't know

Generalizations: Multiple Good / Bad Distributions

- What to expect: locality in special cases, NP-hardness of reporting, poly-time algorithm to achieve target probabilities
- Where to find them: see our paper!