What should an active causal learner value?

Neil R. Bramley¹, Jonathan D. Nelson², Maarten Speekenbrink¹, Vincenzo Crupi³, & David A. Lagnado¹
nrbramley@gmail.com, jonathan.d.nelson@gmail.com, m.speekenbrink@ucl.ac.uk, vincenzo.crupi@unito.it, d.lagnado@ucl.ac.uk
University College London ¹ Max Planck Institute for Human Development ² University of Turin ³

What is active causal learning?

- Intervening on a causal system provides important information about causal structure (Pearl, 2000; Spirtes et al., 1993; Steyvers, 2003; Bramley et al., 2014)
- Interventions render intervened-on variables independent of their normal causes: \( P(A|\text{Do}(B,C)) \neq P(A|B,C) \)
- Which interventions are useful depends on the prior and the hypothesis space
- Goal: Maximize probability of identifying the correct graph

Beyond Shannon: Generalized entropy landscapes

Entropy is expected surprise across possible outcomes. Shannon entropy uses the normal mean (e.g. \( \sum \text{Surprise}(x) \times p(x) \)) as expectation and \( \log \frac{1}{p(x)} \) for surprise. Different types of mean, and different types of surprise, can be used (Crupi and Tentori, 2014). Sharma-Mittal entropy (Nielson and Nock, 2011) provides a unifying framework:
- Sharma-Mittal entropy: \( H_{\alpha,\beta}(g) = \frac{1}{\alpha - 1} \left( \sum p(x) \left( \frac{p(x)}{g(x)} \right)^{\alpha - 1} \right) \), where \( \alpha > 0, \beta \neq 1, \) and \( \beta \neq 1 \)
- Renyi (1961) entropy if degree \( (\beta = 1) \)
- Tsallis (1988) entropy if degree \( (\beta = -\alpha) \)
- Error entropy (probability gain; Nelson et al. (2010)) if degree \( (\beta = 2, \) order \( (\alpha = \infty) \)

Simulating many strategies in many environments

- Learners based on 5 \( \times \) 5 entropies in Sharma-mittal space \( (\alpha, \beta) \in [1/10, 1/2, 1, 2, 10] \), plus our 5 weird and wonderful entropy functions
- Test cases: all 25 of the 3-variable causal graphs, through 5 sequentially chosen interventions. Repeated simulations 10 times each in 3 \( \times \) 3 environments:
  1. Spontaneous activation rates of \([2, 4, 6]\)
  2. Causal powers of \([1, 6, 9]\)

Results

- Shannon entropy achieves higher accuracy than probability gain
- many entropy functions achieve similar accuracy as Shannon entropy
- but the interventions chosen vary widely;

Weird and wonderful (bad) entropy functions

To further broaden the space of entropy functions considered, we first created a variety of atomic surprise functions. From these we derived corresponding entropy functions, always defining entropy as \( \sum \text{Surprise}(x) \times p(x) \). The resulting entropy functions vary in which entropy axioms (Csiszar, 2008) they satisfy.

Questions:

- in which instances do different strategies strongly contradict each other?
- which properties of entropy functions are important for particular situations?
- when are particular entropy functions cognitively plausible?

References