Causality: A Decision Theoretic Foundation

Pablo Schenone (2019)

Presented by Silvio Ravaioli

April 1, 2019

Overview

- Savage (1972) derives *probability beliefs* from preferences
- Similarly, this paper derives *causality beliefs* from preferences
  - Introduce the framework
  - Agent behavior *as if* the relation between two variables is causal
  - Axioms on preferences and representation
  - Existence and uniqueness of a DAG [directed acyclic graph]
- **Motivating Example**
  1. Consider a random US citizen **with** college degree: would you guess she earns more or less than $50K a year?
  2. Consider a random US citizen **without** college degree: would you guess she earns more or less than $50K a year?
  3. Would you encourage the US government to make college education compulsory?
- **Correlation does not mean causation**

Introduction: *Subjective* Causation

- Consider the relation between Ability, Education, and Income
- Savage: we can reveal subjective probability beliefs
  - Consider again a randomly drawn citizen with high school/college degree
  - Accept a safe bet (\$0 for sure) or a risky bet (+\$1 if income>\$50K, -\$1 otherwise)
  - Possible pattern: accept the risky bet conditional on college degree, accept the safe bet conditional on high school degree [correlation]
- Schenone: we can also reveal subjective causality beliefs
  - Accept a safe bet (\$0 for sure) or a risky "policy+bet" (college education becomes compulsory, +\$1 if income>\$50K, -\$1 otherwise)
  - Possible pattern: accept the risky bet conditional on college degree, accept the safe bet in the policy case [no causation]

Introduction: *Subjective* Causation

- Definition of causation: a variable subjectively causes another variable if, holding all other variables constant, policy interventions on the first variable affects subjective beliefs about the second
- We may believe that education policies are useless to affect income (holding fixed intellectual ability), i.e. believe that education levels do not cause income levels
- The definition includes correlation and policy recommendation
- **This paper: foundation for selecting models with which empirical researchers can estimate causal effects**

Preview of Results

- Decision problem similar to Savage's [set of states, set of acts mapping states into monetary amounts, the DM chooses among acts based on a preference relation]
- Additionally, we accommodate for the possibility of choosing policies that affect the states
- Axioms necessary and sufficient for a numerical representation of beliefs
- Bayesian networks as main tool to represent preferences
- Additional axiom to identify causal effects (useful for empirical applications) - not discussed today [section 5]

Model (DAG - Notation)

- Directed graph (V, E)
- V set of nodes and $E \subset V \times V$ set of edges
- A directed graph is acyclic [DAG] iff for all $v \in V$, v is not a descendant of v [i.e. there is no directed path from v to itself]
- $D(v)$ and $ND(v)$ are the sets of descendants and non-descendants of v
- $Pa(v)$ is the set of parents of v: $Pa(v) = \{v' \in V : (v', v) \in E\}$

Model (Acts and Policies)

- State space $S = \Pi_i^N X_i$, where each X_i is finite
- Each $i \in \mathcal{N}$ indicates a variable
- Set of Savage acts $\mathcal{A} = \mathbb{R}^S$
- Preferences $\succ$ over $\mathcal{A}$

- Set of intervention policies $\mathcal{P} = \Pi_i^N (X_i \cup \{\varnothing\})$
- Policies with $p_i = \varnothing$ leave i unaffected
- Policies with $p_j = x_j \in X_j$ forces j value to be x_j
- Each policy implies a collection of interventions over states
- Precise definition: no "reduced spaces", yes "impossible policy"

Model (Preferences over Policies)

- Given a policy $p \in \mathcal{P}$:
- $\mathcal{N}(p)$ is the set of variables that p leaves unaffected
- $\mathcal{A}(p)$ is the set of acts over variables that p leaves unaffected
- The primitive domain is the set $\{(p, a) : p \in \mathcal{P}, a \in \mathcal{A}\}$
- The DM has a preference relation $\succsim$ over the set above

Savage vs. Schenone (again)

- The preference relation is defined over pairs of *policies* and *acts* over unaffected variables
- Given $\bar{\succ}$, each p induces an intervention preference on $\mathcal{A}(p)$
- $f, g \in \mathcal{A}(p)$, we say $f \succ_p g \iff (p, f) \bar{\succ} (p, g)$

- In Savage model the DM chooses an act, and Nature select a value for the state
- In this model Nature is replaced by the DM active choice of p
- Savage conditional preferences correspond to a special case where $p = (\emptyset, \ldots, \emptyset)$

Causal Effect

- A variable j causes a variable i when policies that **intervene** j at different levels have a **ceteris paribus** effect on the DM beliefs about i
- If the choices of acts over i is insensitive to whether j was intervened, j does not cause i
- Based on this definition, subjective causation is a correlation that is sufficient for policy recommendations

- We separate direct and indirect effects by defining *intervention independence*
- Consider a set of variables $\mathcal{K}$, and two variables $i, j \notin \mathcal{K}$
- i is $\mathcal{K}$-independent of j if, after eliminating mediation effect through $\mathcal{K}$, the choice of acts over i is insensitive to interventions of j

$\mathcal{K}$-Independence (Formal definition)

- i is $\mathcal{K}$-independent of j if the following holds for every f, g, x_j, x_j'

$$f \succ_{x_j, x_{\mathcal{K}}} g \iff f \succ_{x_j', x_{\mathcal{K}}} g$$

$$f \succ_{x_j, x_{\mathcal{K}}} g \iff \mathbb{1}_{X_j} f \succ_{x_{\mathcal{K}}} \mathbb{1}_{X_j} g$$

- Line 1: conditional on $\mathcal{K}$ intervention, *intervening* on j has no effect on choices of acts
- Line 2: conditional on $\mathcal{K}$ intervention, the *ability* to intervene on j has no effect on choices
- The second line implies the first (so we use the second as formal definition)

Directed and Indirected Causality

- Variable j **causes** variable i if i is not $\{i, j\}^C$-independent of j [complementary set of variables]
- $Ca(i)$ is the causal set of i
- Variable j is an **indirect cause** of i if there is a sequence $j_0, \ldots, j_T$ such that for all t we have j_t causes j_{t+1}, with $j_0 = j$ and $j_T = i$
- Variable i is an **exogenous primitive** if $Ca(i) = \varnothing$

- Definition of **causal graph**
- Let $\bar{\succ}$ be a preference and $\{Ca(i) : i \in \mathcal{N}\}$ be the collection of causal sets derived from it. Define $G(\bar{\succ}) = (V, E)$ with $V = \mathcal{N}$ and $E = \{(j, i) : j \in Ca(i)\}$

Representation Theorem (Preview)

Theorem 1. Let $\bar{\succ}$ satisfy Axiom 1, and let μ_p be the intervention beliefs elicited from $\succ_p$. The following statements are equivalent:

- Axioms 2 and 3 hold
- Exists a unique G such that G is a DAG and G represents $\bar{\succ}$ in the following sense:
  - Unconditional beliefs are equal to the product of conditional beliefs (conditional on non-descendant variables)
  - The set of parents nodes is the smallest set with the above property
  - If i causes j, then $\mu_{x_{\{j\}^C}}(x_j) = \mu_{x_{\{i,j\}^C}}(x_j|x_i)$

Axioms

- **Axiom 1.** For each $\mathcal{J} \subset \mathcal{N}$ the followings are true
  - For each p, $\succ_p$ satisfy Gul '95 axioms [completeness, transitivity, independence, monotonicity, continuity]
  - If $j \in Ca(i)$, then $f \succ_{x_{\{j\}^C}} g \iff \mathbb{1}_{x_j} f \succ_{x_{\{i,j\}^C}} \mathbb{1}_{x_j} g$
  - There are no null states: $\mathbb{1}_x \succ \mathbb{1}_X 0$ for all $x \in X$
  - Policies do not affect preferences (Bernoulli utility indexes)
- **Axiom 2.** For all $i \in \mathcal{N}$, i is not an indirect cause of i
- **Axiom 3.** $Ca(i) \subset \mathcal{I}$ iff the set $\mathcal{I} \subset \{i\}^C$ satisfies

$$\mathbb{1}_{x_\mathcal{J}} \mathbb{1}_{x_{\mathcal{I} \setminus \mathcal{K}}} f \succ_{x_\mathcal{K}} \mathbb{1}_{x_\mathcal{J}} \mathbb{1}_{x_{\mathcal{I} \setminus \mathcal{K}}} g \iff \mathbb{1}_{x_{\mathcal{I} \setminus \mathcal{K}}} f \succ_{x_\mathcal{K}} \mathbb{1}_{x_{\mathcal{I} \setminus \mathcal{K}}} g$$

Representation (DAG - 1)

- Definition of a DAG representing a probability distribution (from Lauritzen et al. 1990)
- Let $p \in \Delta(\Pi_i X_i)$. A DAG represents p iff

$$p(x) = \Pi_i p(x_i | Pa(i))$$

$$p(x) = \Pi_i p(x_i | \mathcal{T}_i) \Rightarrow \mathcal{T}_i = Pa(i) \quad \forall i \in \mathcal{N}$$

- Line 1: the DAG summarizes the conditional independence properties of p
- Line 2: the set of parents is the smallest possible set (minimality requirement)
- New preferences $\bar{\succ}$ are associated with a *collection* of beliefs (one for each induced $\succ_p$), so we need to define how a DAG represents a *collection* of probability distributions

Representation (DAG - 2)

- Let G be a DAG and $W \subset V$
- **Truncation of a DAG**: the W-truncated DAG G_W is the DAG obtained by eliminating all nodes in W and their arrows in both directions
- A truncated DAG is a representation of intervention beliefs: after intervening in all the variables in W, they are removed from the statistical model
- **G represents $\bar{\succ}$ if these two conditions hold:**
  - $G_{\mathcal{K}}$ represents $\mu_{x_{\mathcal{K}}}$ $\quad \forall \mathcal{K} \subset \mathcal{N}$ [correlation]
  - If $(i,j) \in E$, then $\mu_{x_{\{j\}^C}}(x_j) = \mu_{x_{\{i,j\}^C}}(x_j|x_i)$ [direction of causality]

Representation (DAG - 3)

- **Proposition 1.** Let $\succsim$ be a DM's preferences, and let $G(\succsim) = (\mathcal{N}, E)$ be the directed graph defined by setting $Pa(i) = Ca(i)\ \forall i$. If $\succsim$ satisties Axiom 1, then the following are true:
  - If $G = (\mathcal{N}, F)$ is a directed graph that represents $\succsim$, then $(j, i) \in F \Rightarrow j \in Ca(i)$
  - If $G = (\mathcal{N}, F)$ is a directed graph that represents $\succsim$, then $j \in Ca(i) \Rightarrow (j, i) \in F$ or $i \in Ca(j)$
- **No 2-cycles.** A graph has no 2-cycles if, for each pair i, j, $(i, j) \in E \Rightarrow (j, i) \notin E$. Axiom 2 implies G has no 2-cycles.
- **Corollary 1.** Under the assumptions of Proposition 1, if $G(\succsim)$ has no 2-cycles, then $G(\succsim) = G$.

Representation Theorem (again)

Theorem 1. Let $\bar{\succ}$ satisfy Axiom 1, and let μ_p be the intervention beliefs elicited from $\succ_p$. The following statements are equivalent:

- Axioms 2 and 3 hold
- Exists a unique G such that G is a DAG and G represents $\bar{\succ}$ in the following sense:
    - Unconditional beliefs are equal to the product of conditional beliefs (conditional on non-descendant variables)
    - The set of parents nodes is the smallest set with the above property
    - If i causes j, then $\mu_{x_{\{j\}^C}}(x_j) = \mu_{x_{\{i,j\}^C}}(x_j|x_i)$

Causality: A Decision Theoretic Foundation

Pablo Schenone (2019)

Presented by Silvio Ravaioli

April 1, 2019