Identifying Assumptions and Research Dynamics

Andrew Ellis & Ran Spiegler
LSE TAU & UCL

November 2024

INTRODUCTION

- How should research / learning be conducted?
  - **Bayesianism**: well-studied rational model of learning
- In research communities, acceptance of result requires identification, which requires an (untestable) assumption
- Assumptions **do not** have any special role in Bayesianism
- How does **assumption-based learning** differ from Bayesian learning?

- How should research / learning be conducted?
  - **Bayesianism**: well-studied rational model of learning
- In research communities, acceptance of result requires identification, which requires an (untestable) assumption
- Assumptions **do not** have any special role in Bayesianism
- How does **assumption-based learning** differ from Bayesian learning?

Assumption-based Learning

- Model where research conducted $\iff$ an "**identifying assumption**" is **sufficiently plausible** & beliefs updated as if assumption held
- Rationales:
  - Complexity of processing and communicating all uncertainties
  - Impracticality of strict Bayesianism
  - The need for consensus
- Two key **frictions** relative to Bayesianism:
  - Not all informative research conducted
  - Uncertainty about assumption not incorporated in update

Assumption-based Learning

- Model where research conducted $\iff$ an "**identifying assumption**" is **sufficiently plausible** & beliefs updated as if assumption held
- Rationales:
  - Complexity of processing and communicating all uncertainties
  - Impracticality of strict Bayesianism
  - The need for consensus
- Two key **frictions** relative to Bayesianism:
  - Not all informative research conducted
  - Uncertainty about assumption not incorporated in update

Assumption-based Learning

- Model where research conducted $\iff$ an "**identifying assumption**" is **sufficiently plausible** & beliefs updated as if assumption held
- Rationales:
  - Complexity of processing and communicating all uncertainties
  - Impracticality of strict Bayesianism
  - The need for consensus
- Two key **frictions** relative to Bayesianism:
  - Not all informative research conducted
  - Uncertainty about assumption not incorporated in update

Results

1. Application to stylized examples of research methodologies
2. Impossibility of (certain) research speed up
3. Sufficient condition for constant research speed
4. Characterization of limiting beliefs

Model

- Fixed parameter $\omega = (\omega_1, \ldots, \omega_n) \in \Omega \subset \mathbb{R}^n$ drawn once
- Representative researcher wants to answer a **research question**, represented by a subset Q of $\{1, \ldots, n\}$ that indicates which of the fixed parameters they are trying to learn
  - $\omega_Q \equiv (\omega_i)_{i \in Q}$ is the answer to the question
- Researcher has a prior belief μ over Ω
  - Independence across components of ω
  - Admits probability density function (also denoted μ)

Running example, a contaminated experiment:
Fixed parameters are a true effect ω_1 and a "friction" ω_2
Researcher wants to know true effect: $Q = \{1\}$

Model

- Fixed parameter $\omega = (\omega_1, \ldots, \omega_n) \in \Omega \subset \mathbb{R}^n$ drawn once
- Representative researcher wants to answer a **research question**, represented by a subset Q of $\{1, \ldots, n\}$ that indicates which of the fixed parameters they are trying to learn
  - $\omega_Q \equiv (\omega_i)_{i \in Q}$ is the answer to the question
- Researcher has a prior belief μ over Ω
  - Independence across components of ω
  - Admits probability density function (also denoted μ)

Running example, a contaminated experiment:
Fixed parameters are a true effect ω_1 and a "friction" ω_2
Researcher wants to know true effect: $Q = \{1\}$

Model

Time is discrete & infinite: $t = 1, 2, 3, ...$

In period t:

- beliefs entering period have been updated using history h^t
- **context** $\theta^t \in \Theta$ is drawn (iid) and observed by researcher
- **latent variable** $u^t \in U$ also drawn (iid) but **not** observed
- Researcher decides whether to conduct research
- If they conduct research, they observe **statistic** $s^t \in S$; otherwise move on to next period with same beliefs

Beliefs over $x^t = (s^t, u^t, \theta^t, \omega)$ have density

$$P(x^t) = \mu(\omega) \cdot p_\theta\left(\theta^t\right) \cdot p_u\left(u^t\right) \cdot p\left(s^t | u^t, \theta^t, \omega\right)$$

For results/definitions: S, U are finite, conditional distribution of s^t has full-support, & Ω, Θ are compact, convex

Model

Time is discrete & infinite: $t = 1, 2, 3, \ldots$

In period t:

- beliefs entering period have been updated using history h^t
- **context** $\theta^t \in \Theta$ is drawn (iid) and observed by researcher
- **latent variable** $u^t \in U$ also drawn (iid) but **not** observed
- Researcher decides whether to conduct research
- If they conduct research, they observe **statistic** $s^t \in S$; otherwise move on to next period with same beliefs

Beliefs over $x^t = (s^t, u^t, \theta^t, \omega)$ have density

$$P(x^t) = \mu(\omega) \cdot p_\theta\left(\theta^t\right) \cdot p_u\left(u^t\right) \cdot p\left(s^t | u^t, \theta^t, \omega\right)$$

For results/definitions: S, U are finite, conditional distribution of s^t has full-support, & Ω, Θ are compact, convex

Model

Time is discrete & infinite: $t = 1, 2, 3, ...$

In period t:

- beliefs entering period have been updated using history h^t
- **context** $\theta^t \in \Theta$ is drawn (iid) and observed by researcher
- **latent variable** $u^t \in U$ also drawn (iid) but **not** observed
- Researcher decides whether to conduct research
- If they conduct research, they observe **statistic** $s^t \in S$; otherwise move on to next period with same beliefs

Beliefs over $x^t = (s^t, u^t, \theta^t, \omega)$ have density

$$P(x^t) = \mu(\omega) \cdot p_\theta\left(\theta^t\right) \cdot p_u\left(u^t\right) \cdot p\left(s^t | u^t, \theta^t, \omega\right)$$

For results/definitions: S, U are finite, conditional distribution of s^t has full-support, & Ω, Θ are compact, convex

Model

Running example, continued:

- a^t is whether to conduct experiment
- $s^t \in \mathbb{R}$ is result of experiment
- No unobserved variables
- $\theta^t \in [0, 1]$ is quality of experiment
- Data-generating process:

$$s^t = \omega_1 + \theta^t \omega_2 + \varepsilon^t$$

  - Result of experiment is true effect plus friction plus noise
  - Bias from friction larger in lower quality (higher θ^t settings)

Model: Assumptions

- An **assumption** is a value θ^* of the context parameters

Definition

An **assumption** θ^* **is identifying** w.r.t Q if for every $\omega, \omega' \in \Omega$:

$$\omega_Q \neq \omega'_Q \implies p\left(s \middle| \omega', \theta^*\right) \neq p\left(s \middle| \omega, \theta^*\right) \text{ for some } s$$

- **Interpretation**: Under the assumption, repeated observation of the statistic gives a definitive answer to the question
  - **Running example:** $\theta^* = 0$ is identifying (no friction)
- There may be zero, one, or many identifying assumptions; we assume one throughout but this can be relaxed

Model: Assumptions

- An **assumption** is a value θ^* of the context parameters

Definition

An **assumption** θ^* **is identifying** w.r.t Q if for every $\omega, \omega' \in \Omega$:

$$\omega_Q \neq \omega'_Q \implies p\left(s | \omega', \theta^*\right) \neq p\left(s | \omega, \theta^*\right) \text{ for some } s$$

- **Interpretation**: Under the assumption, repeated observation of the statistic gives a definitive answer to the question
  - **Running example:** $\theta^* = 0$ is identifying (no friction)
- There may be zero, one, or many identifying assumptions; we assume one throughout but this can be relaxed

Model: Assumptions

- An **assumption** is a value θ^* of the context parameters

Definition

An **assumption** θ^* **is identifying** w.r.t Q if for every $\omega, \omega' \in \Omega$:

$$\omega_Q \neq \omega'_Q \implies p\left(s \middle| \omega', \theta^*\right) \neq p\left(s \middle| \omega, \theta^*\right) \text{ for some } s$$

- **Interpretation**: Under the assumption, repeated observation of the statistic gives a definitive answer to the question
  - **Running example:** $\theta^* = 0$ is identifying (no friction)
- There may be zero, one, or many identifying assumptions; we assume one throughout but this can be relaxed

Model: Decision

- Researcher makes the identifying assumption in period t **iff**

$$D\left(P\left(s,u \mid \theta^{t},h^{t}\right) \parallel P\left(s,u \mid \theta^{*},h^{t}\right)\right) \leq K$$

  for some constant $K > 0$
  - D is an f-divergence:

$$D(p(x) \parallel q(x)) = \mathbb{E}_q\left[f\left(\frac{p(x)}{q(x)}\right)\right]$$

    where f strictly convex & $f(1) = 0$.
  - In examples, $D = D_{KL}$ is Kullback-Leibler: $f(y) = y \ln y$
  - **NB:** between beliefs about **both** s **and** u
- $\Theta^R(\mu')$ – contexts where learning occurs given belief μ' over Ω

Model: Decision

- Researcher makes the identifying assumption in period t **iff**

$$D\left(P\left(s,u \mid \theta^t, h^t\right) \parallel P\left(s,u \mid \theta^*, h^t\right)\right) \leq K$$

  for some constant $K > 0$
  - D is an f-divergence:

$$D(p(x) \parallel q(x)) = \mathbb{E}_q\left[f\left(\frac{p(x)}{q(x)}\right)\right]$$

    where f strictly convex & $f(1) = 0$.
  - In examples, $D = D_{KL}$ is Kullback-Leibler: $f(y) = y \ln y$
  - **NB:** between beliefs about **both** s **and** u
- $\Theta^R(\mu')$ – contexts where learning occurs given belief μ' over Ω

Model: Decision

- Researcher makes the identifying assumption in period t **iff**

$$D\left(P\left(s,u \mid \theta^t, h^t\right) \parallel P\left(s,u \mid \theta^*, h^t\right)\right) \leq K$$

  for some constant $K > 0$
  - D is an f-divergence:

  $$D(p(x) \parallel q(x)) = \mathbb{E}_q\left[f\left(\frac{p(x)}{q(x)}\right)\right]$$

    where f strictly convex & $f(1) = 0$.
  - In examples, $D = D_{KL}$ is Kullback-Leibler: $f(y) = y \ln y$
  - **NB:** between beliefs about **both s and u**
- $\Theta^R(\mu')$ – contexts where learning occurs given belief μ' over Ω

Model: Updating

- If they conduct research, then they observe s^t and update via Bayes' rule **given** θ^*:

$$\mu\left(\omega|h^{t+1}\right) = \frac{p\left(s^t|\theta^*, \omega\right)\mu\left(\omega|h^t\right)}{p\left(s^t|\theta^*, h^t\right)}$$

for (almost) every ω

(even if $\theta^t \neq \theta^*$!)

- If they don't, then they pass over the opportunity to learn

$$\mu\left(\cdot|h^{t+1}\right) = \mu\left(\cdot|h^t\right)$$

Contaminated Experiment

- $\omega = (\omega_1, \omega_2)$ & $\omega_Q = \omega_1$
- $s^t = \omega_1 + \theta^t \omega_2 + \varepsilon^t$
- ε^t is noise: zero mean, unit variance, & indep. of all other vars
- $\theta^t \in [0, 1]$ so $\theta^* = 0$ is an identifying assumption

- If $\omega_1, \omega_2, \varepsilon^t$ are indep. Normals w/ means $m_1, m_2, 0$, then
- $\theta^* = 0$ is the **unique** identifying assumption

Assumes the friction can be **neglected**

NB: under θ^*, s^t reveals nothing about ω_2

Contaminated Experiment

- $\omega = (\omega_1, \omega_2)$ & $\omega_Q = \omega_1$
- $s^t = \omega_1 + \theta^t \omega_2 + \varepsilon^t$
- ε^t is noise: zero mean, unit variance, & indep. of all other vars
- $\theta^t \in [0,1]$ so $\theta^* = 0$ is an identifying assumption

- If $\omega_1, \omega_2, \varepsilon^t$ are indep. Normals w/ means $m_1, m_2, 0$, then
- $\theta^* = 0$ is the **unique** identifying assumption

  Assumes the friction can be **neglected**

  NB: under θ^*, s^t reveals nothing about ω_2

Contaminated Experiment

- In period t given h^t:
  $s^t|\theta^t \sim N\left(m_1(h^t) + \theta^t m_2, (\sigma_1(h_t))^2 + (\theta^t)^2\sigma_2^2 + 1\right)$
  $s^t|\theta^* \sim N\left(m_1(h^t), (\sigma_1(h_t))^2 + 1\right)$
- KL Divergence of assumption equals

$$\frac{1}{2}\left[\left(\theta^t\right)^2 \frac{\sigma_2^2 + m_2^2}{1 + (\sigma_1(h_t))^2} - \ln\left(1 + \frac{\left(\theta^t\right)^2\sigma_2^2}{1 + (\sigma_1(h_t))^2}\right)\right]$$

- **only** non-constant terms are $(\sigma_1(h_t))^2$ & θ^t
  - D_{KL} **decreases** with $\sigma_1^2(h^t)$
    more precise beliefs $\Longrightarrow$ more change from θ
  - D_{KL} **increases** with θ^t, so $\Theta^R(\mu(\cdot|h^t)) = [0, \overline{\theta}(h^t)]$
    more different contexts $\Longrightarrow$ less similar distributions

Contaminated Experiment

- $(\sigma_1(h_t))^2$ shrinks deterministically each time research occurs

  $\Longrightarrow D_{\mathsf{KL}}$ **increases** over time for **any given** θ

  $\Longrightarrow \overline{\theta}(h^t)$ **decreases** over time (to some $\bar{\theta}^* > 0$)

- Therefore research **slows down** over time
  (but never entirely stops)

- If ω^* is true parameter, then beliefs converge a.s. to

$$\omega_1^* + \mathbb{E}_0\left[\theta^t | \theta^t \leq \bar{\theta}^*\right] \omega_2^*$$

Result I: Impossiblity of accelerating research

Proposition

Suppose that $D\left(P\left(s,u|\theta^{t},h^{t}\right)||P\left(s,u|\theta^{*},h^{t}\right)\right)$ is always quasi-convex in θ.
If $\Theta^{R}\left(\mu\left(h^{t+1}\right)\right)\setminus\Theta^{R}\left(\mu\left(h^{t}\right)\right)\neq\emptyset$ with positive probability then $\Theta^{R}\left(\mu\left(h^{t}\right)\right)\setminus\Theta^{R}\left(\mu\left(h^{t+1}\right)\right)\neq\emptyset$ with positive probability.

- If the propensity to research goes up from period t to period $t+1$, then it might have gone down
- Θ^{R} might contract for sure but it can never expand for sure
- The proof is based on the convexity of f-divergences
- Similar but weaker result without quasi-convexity assumption

Proof

- Let $q(\theta, h) = P(\cdot|\theta, h)$ and suppose that $\exists \theta^1, h^t, s$ so that

$$D\left(q\left(\theta^1, h^t\right)||q\left(\theta^*, h^t\right)\right) > K > D\left(q\left(\theta^1, h^t, s\right)||q\left(\theta^*, h^t, s\right)\right)$$

- There exists $\theta = \beta\theta^1 + (1-\beta)\theta^* \in \Theta^R(\mu(.|h^t))$ so that

$$D\left(q\left(\theta, h^t\right)||q\left(\theta^*, h^t\right)\right) = K,$$

- Quasi-convexity gives that $D\left(q\left(\theta, h^t, s\right)||q\left(\theta^*, h^t, s\right)\right) < K$
- Also: $\sum_{s^t \in S} q\left(\theta, h^t, s^t\right)(s^{t+1}) p\left(s^t|\theta^*, h^t\right) = q\left(\theta^*, h^t\right)(s^{t+1})$
- But D is convex, so

$$\begin{aligned} K &= D\left(q\left(\theta, h^t\right)||q\left(\theta^*, h^t\right)\right) \\ &\leq \sum_{s^t} D\left(q\left(\theta, h^t, s^t\right)||q\left(\theta^*, h^t, s^t\right)\right) p(s^t|\theta^*, h^t) \end{aligned}$$

- Hence $\theta \in \Theta^R(\mu(.|h^t)) \setminus \Theta^R(\mu(.|h^t, s'))$ for some s'

Causal Inference

- $\omega = (\omega_1, \omega_2) = (\beta, \sigma) \in \mathbb{R}^2$:
  Causal effect of x on y & variance of potential confounder
- Researcher wants to know causal effect:

$$\omega_Q = \beta$$

- $s^t = (x^t, y^t) \in \mathbb{R}^2$: observed cause & effect
- $u^t \in \mathbb{R}$: unobserved confounder
- $\theta^t \in [0, 1]$: strength of confounding
  $\theta^* = 0$ is identifying assumption

Data-generating process:

$$\begin{aligned} x^t =& \theta^t \sigma u^t + \varepsilon_x^t \\ y^t =& \beta x^t + u^t + \varepsilon_y^t \end{aligned}$$

Causal Inference

$$x^t = \theta^t \sigma u^t + \varepsilon_x^t \qquad y^t = \beta x^t + u^t + \varepsilon_y^t$$

- $\varepsilon_x^t, \varepsilon_y^t$ are independent Normals & the support of Ω and the variance of $\varepsilon_x^t, \varepsilon_y^t$ are chosen so that $x^t, y^t \sim N(0,1)$
- **only relevant observable** is their correlation, ρ_{12}
- But ρ_{12} reflects both β & confounding by u
- Overall strength of confounding unknown (because of σ), so no $\theta > 0$ can identify β
- $\theta^* = 0$ is the **unique** identifying assumption

  Assumes the confounding effect can be **neglected**

  NB: under θ^*, s reveals nothing about σ

Causal Inference

$$x^t = \theta^t \sigma u^t + \varepsilon_x^t \qquad y^t = \beta x^t + u^t + \varepsilon_y^t$$

- $\varepsilon_x^t, \varepsilon_y^t$ are independent Normals & the support of Ω and the variance of $\varepsilon_x^t, \varepsilon_y^t$ are chosen so that $x^t, y^t \sim N(0,1)$

- **only relevant observable** is their correlation, ρ_{12}
- But ρ_{12} reflects both β & confounding by u
- Overall strength of confounding unknown (because of σ), so no $\theta > 0$ can identify β

- $\theta^* = 0$ is the **unique** identifying assumption

  Assumes the confounding effect can be **neglected**

  NB: under θ^*, s reveals nothing about σ

Causal Inference

$$x^t = \theta^t \sigma u^t + \varepsilon_x^t \qquad y^t = \beta x^t + u^t + \varepsilon_y^t$$

- $\varepsilon_x^t, \varepsilon_y^t$ are independent Normals & the support of Ω and the variance of $\varepsilon_x^t, \varepsilon_y^t$ are chosen so that $x^t, y^t \sim N(0, 1)$
- **only relevant observable** is their correlation, ρ_{12}
- But ρ_{12} reflects both β & confounding by u
- Overall strength of confounding unknown (because of σ), so no $\theta > 0$ can identify β
- $\theta^* = 0$ is the **unique** identifying assumption

  Assumes the confounding effect can be **neglected**

  NB: under θ^*, s reveals nothing about σ

Causal Inference

$$D_{\mathsf{KL}}\left(P\left(s,u \mid \theta,h^t\right) \parallel P\left(s,u \mid \theta^*,h^t\right)\right)$$
$$= \int \ln \frac{\int_\sigma \int_\beta P(u)P(x|u,\theta,\sigma)P(y|x,u,\beta)d\mu\left(\beta|h^t\right)d\mu\left(\sigma|h^t\right)}{\int_\sigma \int_\beta P(u)P(x|u,\theta^*,\sigma)P(y|x,u,\beta)d\mu(\beta|h^t)d\mu(\sigma|h^t)} dP\left(s,u \mid \theta,h^t\right)$$

$$= \int \ln \frac{P(u)\int_\sigma P(x|u,\theta,\sigma)d\mu\left(\sigma|h^t\right)\int_\beta P(y|x,u,\beta)d\mu\left(\beta|h^t\right)}{P(u)\int_\sigma P(x|u,\theta^*,\sigma)d\mu\left(\sigma|h^t\right)\int_\beta P(y|x,u,\beta)d\mu\left(\beta|h^t\right)} dP\left(s,u \mid \theta,h^t\right)$$

$$= \int \ln \frac{P\left(x|u,\theta^t\right)}{P(x|u,\theta^*)} dP\left(x,u\right)$$

- $P\left(x|u,\theta\right)$ depends **only on the constant** beliefs about σ

Causal Inference

$$D_{\mathsf{KL}}\left(P\left(s,u \mid \theta,h^t\right) \parallel P\left(s,u \mid \theta^*,h^t\right)\right)$$

$$= \int \ln \frac{\int_\sigma \int_\beta P(u)P(x|u,\theta,\sigma)P(y|x,u,\beta)d\mu\left(\beta|h^t\right)d\mu\left(\sigma|h^t\right)}{\int_\sigma \int_\beta P(u)P(x|u,\theta^*,\sigma)P(y|x,u,\beta)d\mu(\beta|h^t)d\mu(\sigma|h^t)} dP\left(s,u \mid \theta,h^t\right)$$

$$= \int \ln \frac{\cancel{P(u)}\int_\sigma P(x|u,\theta,\sigma)d\mu\left(\sigma|\cancel{h^t}\right)\cancel{\int_\beta P(y|x,u,\beta)d\mu\left(\beta|h^t\right)}}{\cancel{P(u)}\int_\sigma P(x|u,\theta^*,\sigma)d\mu\left(\sigma|\cancel{h^t}\right)\cancel{\int_\beta P(y|x,u,\beta)d\mu\left(\beta|h^t\right)}} dP\left(s,u \mid \theta\cancel{,h^t}\right)$$

$$= \int \ln \frac{P\left(x|u,\theta^t\right)}{P(x|u,\theta^*)} dP\left(x,u\right)$$

- $P\left(x|u,\theta\right)$ depends **only on the constant** beliefs about σ

Causal Inference

$$D_{\mathsf{KL}}\left(P\left(s,u \mid \theta,h^t\right) \parallel P\left(s,u \mid \theta^*,h^t\right)\right)$$
$$= \int \ln \frac{\int_\sigma \int_\beta P(u)P(x|u,\theta,\sigma)P(y|x,u,\beta)d\mu\left(\beta|h^t\right)d\mu\left(\sigma|h^t\right)}{\int_\sigma \int_\beta P(u)P(x|u,\theta^*,\sigma)P(y|x,u,\beta)d\mu(\beta|h^t)d\mu(\sigma|h^t)} dP\left(s,u \mid \theta,h^t\right)$$
$$= \int \ln \frac{\cancel{P(u)} \int_\sigma P(x|u,\theta,\sigma)d\mu\left(\sigma|\cancel{h^t}\right)\cancel{\int_\beta P(y|x,u,\beta)d\mu\left(\beta|h^t\right)}}{\cancel{P(u)} \int_\sigma P(x|u,\theta^*,\sigma)d\mu\left(\sigma|\cancel{h^t}\right)\cancel{\int_\beta P(y|x,u,\beta)d\mu\left(\beta|h^t\right)}} dP\left(s,u \mid \theta\cancel{,h^t}\right)$$
$$= \int \ln \frac{P\left(x\middle|u,\theta^t\right)}{P(x|u,\theta^*)} dP\left(x,u\right)$$

- $P\left(x|u,\theta\right)$ depends **only on the constant** beliefs about σ

Causal Inference

- D_{KL} for assumption is **independent of history**, so propensity to conduct research is time-invariant and positive
- Research whenever θ^t in the interval $[0, \bar{\theta}]$

- Given true (β^*, σ^*), beliefs converge a.s. to

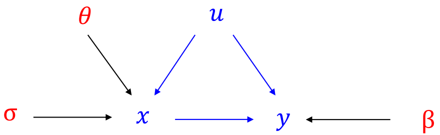
$$\beta = \mathbb{E}_0 \left[(\theta^t)^2 \sigma^{*2} \beta^* + \theta^t \sigma^* \sigma_U | \theta^t \leq \bar{\theta} \right] + \beta^*$$

Result II: Constant Research

- Let $x^t = (s^t, u^t, \theta^t, \omega) \in \mathbb{R}^m$ and N^s s.t. $x^t_{N^s} = s^t$
- Say that **data-generating process has recursive structure** $G = (\{1, \ldots, m\}, R)$ if G is a directed acyclic graph (DAG) with no edge into j for any $j \notin N^s$ and x^t has density

$$\mu(\omega)\, p_\theta\left(\theta^t\right) p_u\left(u^t\right) \prod_{i \in N^s} p\left(s^t_i | x^t_{R(i)}\right)$$

  where $R(i)$ is all nodes pointing into i
- Every p has **some** recursive structure (not unique)

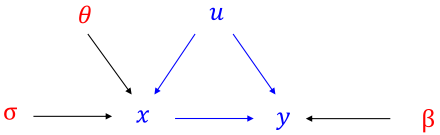


Result II: Constant Research

- In both examples, $R(i)$ equals the variables on the RHS of the equation for s_i

- DAG G satisfies a conditional independence property if all data-generating processes having structure G satisfy it
- DAG lit gives graphical characterization of these properties

- In CI Example, G satisfies $\beta \perp x|(y,u)$ but not $\beta \perp y|(x,u)$:



Result II: Constant Research

- The set of **active parameters** Q^* is the smallest set of indexes of ω that affect the distribution of s under θ^*
  - In CI Example, $\omega_{Q^*} = \beta = \omega_Q$
  - In CE Example, $\omega_{Q^*} = \omega_1 = \omega_Q$

- Say that θ and ω_{Q^*} are G**-separable** if for every i, G satisfies $s_i \perp \omega_{Q^*}$ whenever it satisfies $s_i \not\perp \theta^t | (s_{-i}, u)$
  - any **statistic** that is **context-dependent** (conditional on the other variables) is **unaffected by the active parameters**
  - context and active parameters have separate observable effects
  - CI Example satisfies but CE Example does not

Result II: Constant Research

Proposition

If data-generating process has a recursive structure G for which θ and ω_{Q^*} are **G-separable**, then **Θ^R is constant**.

- Hypothesis on the structure of the distribution, not the distribution itself
- Proof uses DAG techniques to show that under G-separability, the only aspects of ω that determine how s_i varies with θ are associated with a **time-invariant** belief
- **Application:** Causal inference via **IV** where the identifying assumption is that instrument is independent of confounder

Result II: Constant Research

Proposition

If data-generating process has a recursive structure G for which θ and ω_{Q^*} are **G-separable**, then Θ^R **is constant**.

- Hypothesis on the structure of the distribution, not the distribution itself
- Proof uses DAG techniques to show that
  under G-separability,
  the only aspects of ω that determine how s_i varies with θ
  are associated with a **time-invariant** belief
- **Application:** Causal inference via **IV** where the identifying assumption is that instrument is independent of confounder

Result II: Constant Research

Proposition

If data-generating process has a recursive structure G for which θ and ω_{Q^*} are **G-separable**, then Θ^R **is constant**.

- Hypothesis on the structure of the distribution, not the distribution itself
- Proof uses DAG techniques to show that
  under G-separability,
  the only aspects of ω that determine how s_i varies with θ
  are associated with a **time-invariant** belief
- **Application:** Causal inference via **IV** where the identifying assumption is that instrument is independent of confounder

Result III: Stable beliefs

Definition

A belief $\mu^* \in \Delta(\Omega)$ is **stable** given ω^* if $\Pr\left(\mu\left(\cdot|h^t\right) \to^{w*} \mu^*|\omega^*\right) > 0$.

Proposition

If μ^* is stable given ω^* and Θ^R is continuous at μ^*, then $\mu^*(O) = 1$ for any open O s.t.

$$O \supset \text{argmin}_\omega \, D_{\text{KL}}\left(P\left(s|\,\omega^*,\ \theta \in \Theta^R\left(\mu^*\right)\right) \parallel P\left(s|\,\omega,\ \theta^*\right)\right)$$

Result III: Stable beliefs

Definition

A belief $\mu^* \in \Delta(\Omega)$ is **stable** given ω^* if
$\Pr\left(\mu\left(\cdot|h^t\right) \to^{w*} \mu^*|\omega^*\right) > 0$.

Proposition

If μ^* is stable given ω^* and Θ^R is continuous at μ^*, then
$\mu^*(O) = 1$ for any open O s.t.

$$O \supset \text{argmin}_\omega \, D_{\text{KL}}\left(P\left(s|\,\omega^*,\ \theta \in \Theta^R\left(\mu^*\right)\right) \parallel P\left(s|\,\omega,\ \theta^*\right)\right)$$

Result III: Stable beliefs

if μ^* stable then it rules out parameters that do not minimize

$$D_{\mathsf{KL}}\Big(\underbrace{P\left(s|\,\omega^*,\ \theta \in \Theta^R\left(\mu^*\right)\right)}_{\text{Actual distribution of } s} \,\|\, \underbrace{P\left(s|\,\omega,\ \theta^*\right)}_{\text{Predicted distribution of } s \text{ given } \omega \,\&\, \theta^*}\Big)$$

- Here, KL divergence is a result not assumption
- Distinct divergence than for plausibility
- Self-referential equation / equilibrium condition
- Related to Berk-Nash equilibrium (Esponda-Pouzo 2016)
- Stable belief biased in most generalizations of our examples

Wrap-up

In paper:

- Possibility of multiple stable beliefs
- Stable beliefs far from truth even with small K
- Extensions
  - Choosing between structural assumptions / setting the value of a fixed parameter; learning by “calibration”
  - Choosing between research-design and structural assumptions; “natural experiment” vs. “Heckman correction” identification strategies for selective samples

Wishlist:

- Choosing between a strong assumption (to answer an ambitious question) and a weak assumption (to answer a modest question)
- Hierarchy of identifying assumptions