Optimal Pre-Analysis Plans: Statistical Decisions Subject to Implementability

Maximilian Kasy   Jann Spiess

June 2024

Introduction

- Trial registration and pre-analysis plans (PAPs) have become a standard requirement for experimental research.
  - For clinical studies in medicine starting in the 1990s.
  - For experimental research in economics more recently.
- Standard justification: Guarantee validity of inference.
  - P-hacking, specification searching, and selective publication distort inference.
  - Tying researchers' hands prevents selective reporting.
  - Christensen and Miguel (2018); Miguel (2021).
- The widespread adoption of PAPs has not gone uncontested, however.
  - Coffman and Niederle (2015); Olken (2015); Duflo et al. (2020).

Open questions

1. Why do we need a commitment device?
   Standard decision theory has no time inconsistency!

2. How should the structure of PAPs look like?
   How can we derive optimal PAPs?

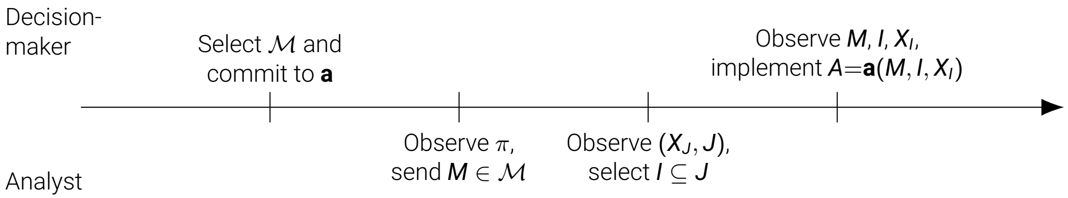
Key insight:

- Single-agent decision-theory cannot make sense of these debates.
- We need to consider multiple agents,
  conflicts of interest, and asymmetric information.

Setup: Notation

- Two parties, decision-maker and analyst.
- Message M ("pre-analysis plan") sent from analyst to decision-maker.
- Data $X = (X_1, \ldots, X_n) \sim \mathrm{P}_\theta$.
  - Unknown parameter $\theta \in \Theta$.
- Index sets:
  - $K = \{1, \ldots, n\}$ fixed, finite, commonly known.
  - $J \subset K$ subset of data available to the analyst, privately known.
  - $I \subset J$ subset of available data reported to the decision-maker.
- Decision $A \in \mathcal{A} \subseteq \mathbb{R}$.

Setup: Timeline



Discussion

- The analyst can withhold information,
  but they cannot lie.
- The components of X might represent different
  - hypothesis tests,
  - estimates,
  - subgroups,
  - outcome variables, etc.
- Possible model interpretations:
  1. Drug approval (pharma company vs. FDA).
  2. Hypothesis testing (researcher vs. reader).
  3. Publication decision (researcher vs. journal).

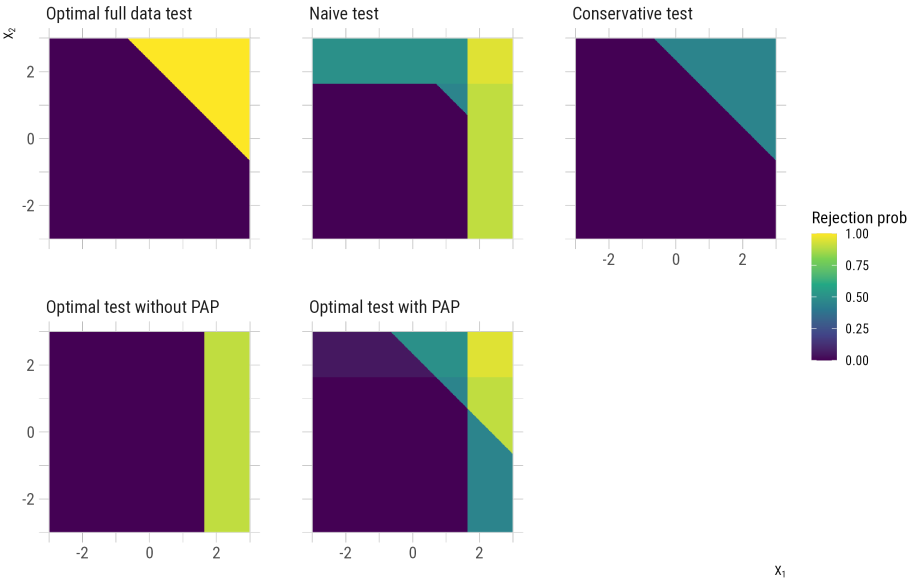
Motivating example: Normal testing

- $X_1, X_2 \sim N(\theta, 1)$.
- Prior of the decision-maker : $(J_1, J_2) \sim Ber(\eta_1) \times Ber(\eta_2)$.
- The analyst knows J.
- Null hypothesis $H_0 : \theta \leq 0$.
- The analyst selectively reports, to get a rejection of the null.

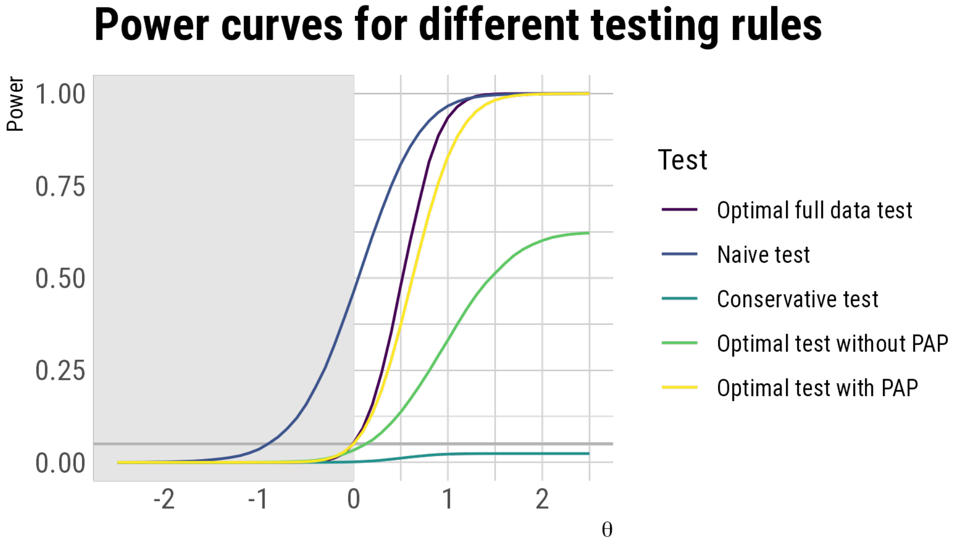
Compare 5 testing rules

0. The optimal full data test (infeasible).
1. The naive test (ignores selective reporting).
2. The conservative test (worst-case assumptions about unreported X_ι).
3. The optimal implementable test without a PAP.
4. The optimal implementable test with a PAP.

Rejection probabilities for different testing rules



Degrees of freedom $n = 10$



Implementable decision functions

- A **reduced form decision function** maps the full data into a decision $\mathbf{a}$:

$$\bar{\mathbf{a}}(\pi, X_J, J)$$

- A reduced form decision function $\bar{\mathbf{a}}$ is **implementable**
  - if there exist a decision function $\mathbf{a}$
  - with best responses $\mathbf{m}^*, \mathbf{i}^*$
  - such that

$$\bar{\mathbf{a}}(\pi, X_J, J) = \mathbf{a}(M^*, X_{I^*}, I^*).$$

- **Assumption**:
  The analyst is an expected utility maximizer with utility

$$v(A)$$

  for a (strictly) monotonically increasing function v.

Implementability without PAPs

Lemma

If no pre-analysis messages $\boldsymbol{M}$ are allowed,
a reduced-form decision function $\bar{\mathbf{a}}(\pi, X_J, J)$ is implementable iff

1. $\bar{\mathbf{a}}$ *does not depend on* π*, and*

2. $\bar{\mathbf{a}}$ *is* ***monotonic*** *in* J*,*

$$\bar{\mathbf{a}}(X_I, I) \leq \bar{\mathbf{a}}(X_J, J)$$

   for almost all $\boldsymbol{X}, \boldsymbol{J}$ *and all* $\boldsymbol{I} \subseteq \boldsymbol{J}$*.*

Implementability with PAPs

Theorem

A reduced-form decision function $\bar{\mathbf{a}}$ *is implementable iff both of the following conditions hold:*

1. ***Truthful PAP***
   For almost all π *and all* π',
   $$\mathrm{E}[v(\bar{\mathbf{a}}(\pi', X_J, J))|\pi] \leq \mathrm{E}[v(\bar{\mathbf{a}}(\pi, X_J, J))|\pi].$$
2. ***Monotonicity***
   For almost all π, X, J, *and all* $I \subseteq J$
   $$\bar{\mathbf{a}}(\pi, X_I, I) \leq \bar{\mathbf{a}}(\pi, X_J, J)$$

Revelation and delegation

Lemma

A reduced-form decision rule $\bar{\mathbf{a}}$ *can be implemented iff:*

1. ***Implementation by truthful revelation***
   It can be implemented with a decision rule $\mathbf{a}$ *for which*

$$\mathbf{a}(\pi, X_J, J) = \bar{\mathbf{a}}(\pi, X_J, J).$$

2. ***Implementation by delegation***
   It can be implemented with a decision rule $\mathbf{a}$ *for which*

$$\mathbf{a}(b, X_J, J) = b(X_J, J),$$

where $\mathbf{b}$ *is restricted to lie in some set* $\mathcal{B}$.

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

Conclusion and outlook

Hypothesis testing

- Null hypothesis $\theta \in \Theta_0$.
- Rejection probability $\mathbf{a} \in [0, 1]$.

$\Rightarrow$ w.l.o.g. $v(\mathbf{a}) = \mathbf{a}$.

- Size control at level $\alpha \in (0, 1)$:

$$\sup_{\theta \in \Theta_0} \sup_{\pi} \mathrm{E}[\bar{\mathbf{a}}(\pi, X_J, J) | \theta, \pi] \leq \alpha.$$

- Expected power:

$$\mathrm{E}[\bar{\mathbf{a}}(\pi, X_J, J)].$$

Implementing the optimal test by delegation

Theorem

- *The test with maximal expected power*
- *subject to implementability and size control*
- *can be implemented by requiring the analyst to communicate a full-data test* **t** *which satisfies, for all $\theta \in \Theta_0$,*

$$\mathrm{E}[t(X)|\theta] \leq \alpha$$

- *and then implementing the test*

$$b(X_I, I) = \min_{X';\, X'_I = X_I} t(X').$$

The analyst's problem as a linear program

$$
\begin{aligned}
\max_b \int b(X_J, J)\, d\,\mathrm{P}_\pi(X, J), & & & \text{(Interim expected power)} \\
\text{s.t.} \int b(X, K)\, d\,\mathrm{P}_{\theta_0}(X) \leq \alpha, & & & \text{(Size control)} \\
b(X_J, J) \in [0, 1] & \quad \forall\, J, X, & & \text{(Support)} \\
b(X_J, J) \leq b(X, K) & \quad \forall\, J, X. & & \text{(Monotonicity)}
\end{aligned}
$$

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

Conclusion and outlook

Discussion

- Conflicts of interest, private information.
  $\Rightarrow$ Not all decision rules are implementable.

- Mechanism design: Optimal implementable rules.

- Statistical reporting: Partial verifiability.
  1. No lying about reported statistics.
  2. Private information about which statistics were available.

- Pre-analysis plans:
  - No role in single-agent decision-theory.
  - But increase the set of implementable rules in multi-agent settings.

- We characterize
  1. implementable rules,
  2. optimal implementable hypothesis tests,

Thank you!