Optimal Pre-Analysis Plans: Statistical Decisions Subject to Implementability

Maximilian Kasy   Jann Spiess

May 2023

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.
- Counter-argument:
    - Interesting findings are unexpected and flexibility is necessary.

Open questions

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

2. Under what conditions are PAPs more or less useful?
   How do we trade off the benefits and costs of PAPs?

3. How should the structure of PAPs look like?

Key insight:

- Single-agent decision-theory cannot make sense of these debates.

- We need to consider multiple agents,
  conflicts of interest, and asymmetric information.

Our approach

- Import insights from contract theory / mechanism design to statistics.
  - PAPs can be rationalized with multiple parties,
    conflicts of interest, and asymmetric information.
  - We consider (optimal) statistical decision rules
    subject to the constraint of implementability.
- Our model:
  1. A decision-maker commits to a decision rule,
  2. then an analyst communicates a PAP,
  3. then observes the data,
     reports selected (!) statistics to the decision-maker,
  4. who then applies the decision rule.

Note: The model presented in this talk is different from that discussed in an earlier workingpaper on the same topic.

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

1. The decision-maker commits to a decision function $\mathbf{a}(\cdot)$.
2. The analyst:
   a) Observes a private signal π.
   b) Sends a message M to the decision-maker.
   c) Observes (X_J, J).
   d) Reports (X_I, I) with $I \subseteq J$.
3. The decision-maker implements the decision $A = \mathbf{a}(M, X_I, I)$.

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

- $K = \{1, 2\}$.
- $X_1, X_2 \sim N(\theta, 1)$.
- 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 (only available if $I = J = \{1, 2\}$).
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.

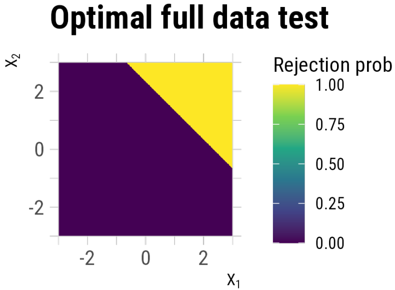
The optimal full data test

- Suppose availability and selective reporting were no concern.
- Then $X_1 + X_2$ is a sufficient statistic.
- By Neyman-Pearson, the uniformly most powerful test is given by

$$\mathbf{1}\left(X_1 + X_2 > \sqrt{2} \cdot z\right).$$

- Critical value:

$$z = \Phi^{-1}(1 - \alpha).$$



The naive test

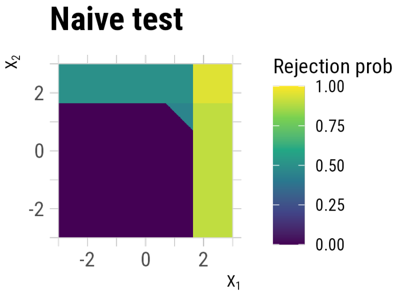
- Treat the reported data I as if there were no selective reporting.

$$\mathbf{a}_1(X_I, I) = \mathbf{1}\left(\sum_{\iota \in I} X_\iota > z \cdot \sqrt{|I|}\right).$$

- The analyst chooses $I \subset J$ to maximize rejection,

$$\bar{\mathbf{a}}_1(X_J, J) = \max_{I \subset J} \mathbf{a}(X_I, I).$$

- Such p-hacking violates size control!

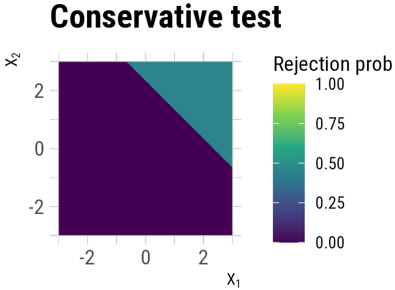


The conservative test

- Possible remedy:
  Worst-case assumptions about unreported components.

$$\mathbf{a}_2(X_I, I) = \mathbf{1}\left(X_1 + X_2 > \sqrt{2} \cdot z \text{ and } I = K\right).$$

- This test controls size.
- But it has low power.



The optimal implementable test without PAP

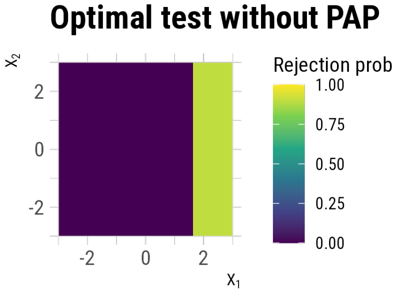
- Requirements:
  1. Size control.
  2. Incentive compatibility.
  3. Maximizes expected power.

- Solution without a PAP:
  1. Pick a full-data test,
  2. make worst-case assumptions about unreported components.

- Choose the full-data test to maximize expected power.

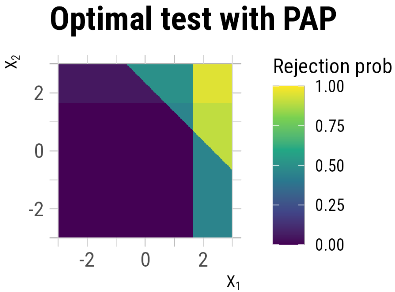
- Here:

$$\mathbf{a}_3(X_I, I) = \mathbf{1}\left(X_1 > z \text{ and } 1 \in I\right).$$

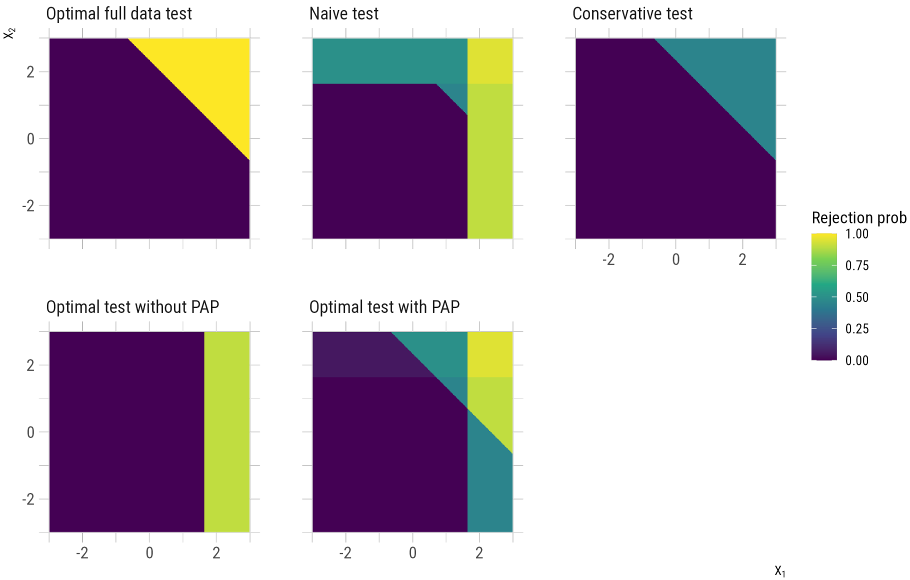


The optimal implementable test with PAP

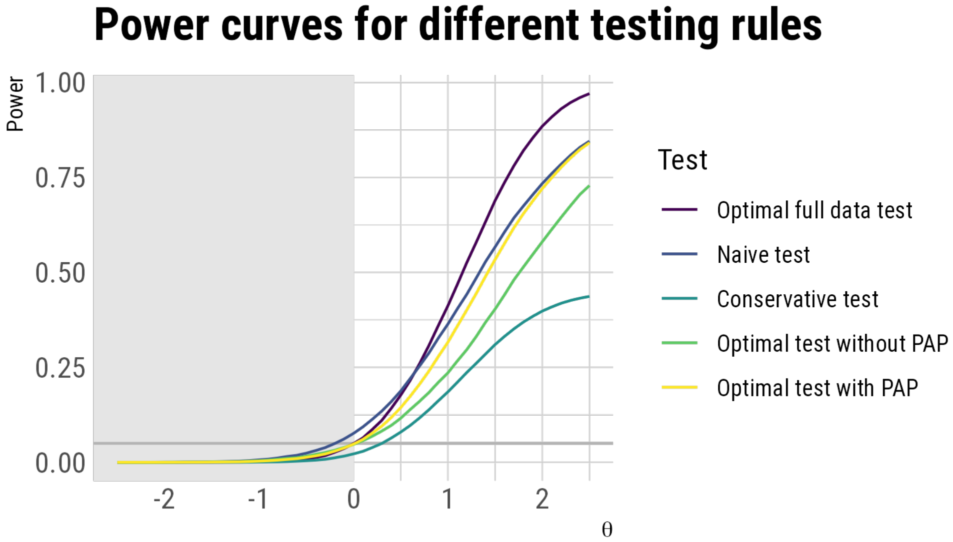
- Allow an analyst message before seeing data.
- Solution *with* a PAP :
  1. Let the *analyst* pick a full-data test,
  2. make worst-case assumptions about unreported components.
- The analyst knows J when choosing the full-data test.



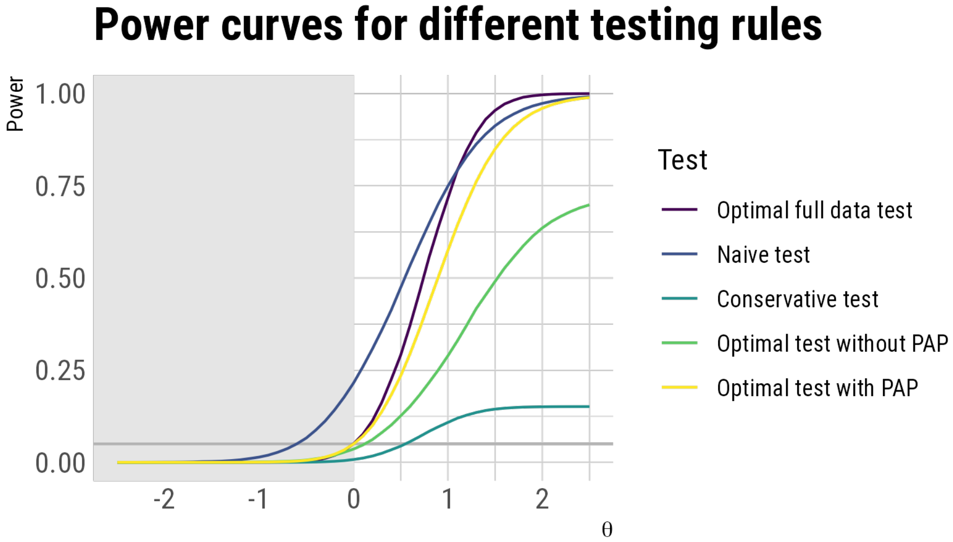
Rejection probabilities for different testing rules



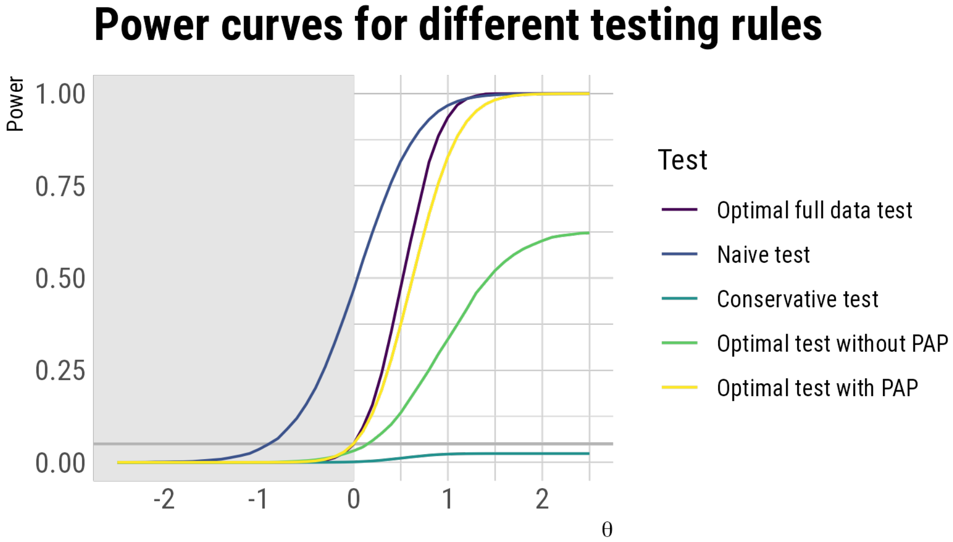
Degrees of freedom $n = 2$



Degrees of freedom $n = 5$



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.

Analyst best responses

- The optimal report $I^* = \mathbf{i}^*(M, X_J, J)$ of the analyst satisfies

$$I^* \in \operatorname*{argmax}_{I \subseteq J} \; \mathbf{a}(M, X_I, I).$$

- The optimal message $M^* = \mathbf{m}^*(\pi)$ satisfies

$$M^* \in \operatorname*{argmax}_{m} \; \mathrm{E}[v(\mathbf{a}(m, I^*, X_{I^*}))|\pi].$$

Preview of implementability results

- Without PAPs, implementability is equivalent to **monotonicity** in J: Reporting more can only increase the decision.
- With PAPs, implementability only requires monotonicity in J **conditional** on the analyst signal.
- Implementation can use different approaches:
  1. Truthful **revelation** of the analyst signal.
  2. **Delegation** to the analyst, letting them choose a decision function from a constrained set.
- Truthful revelation is closely related to **proper scoring**.
- For binary actions, the set of implementable decision functions is a **convex polytope**.

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* $\boldsymbol{I} \subseteq \boldsymbol{J}$*.*

Proof

1. Suppose that both conditions hold.
   - Set $\mathbf{a}(X_I, I) = \bar{\mathbf{a}}(X_I, I)$.
   - Incentive compatibility of $\mathbf{i}^*(X_J, J) = J$ follows.
2. Consider a decision function $\bar{\mathbf{a}}$ that is implementable by $\mathbf{a}$.
   - Since $\mathbf{i}^*$ is an analyst best-response to this decision function $\mathbf{a}$,

$$\bar{\mathbf{a}}(\pi, X_J, J) = \max_{I \subseteq J} \mathbf{a}(X_I, I).$$

   - The maximum over subsets of J (weakly) increases in J. $\square$

Note: The revelation principle does not directly apply here, due to partial verifiability!

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

Sketch of proof

1. This is the revelation principle.
2. This follows by the same argument as before. $\square$

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* $\mathcal{B}$.

Sketch of proof

1. Immediate from previous result / revelation principle.
2. Suppose that $\bar{\mathbf{a}}$ is implemented by $\mathbf{a}(M, X_I, I)$, $\mathbf{m}^*, \mathbf{i}^*$.
   - Define $\tilde{\mathbf{a}}(b, X_J, J) = b(X_J, J)$ for $b \in \mathscr{B}$, where
     $$\mathscr{B} = \{b(\cdot) : \ b(X_I, I) = \mathbf{a}(M, X_I, I), \ \text{for some } M\}.$$
   - It follows that $b(\cdot) = \mathbf{a}(\mathbf{m}^*(\pi), X_I, I)$ is a best response to $\tilde{\mathbf{a}}$.
   - Therfore $\tilde{\mathbf{a}}$ implements $\bar{\mathbf{a}}$.

$\square$

Proper scoring

- Define
$$S(\pi', \pi) = E[v(\bar{\mathbf{a}}(\pi', X_J, J))|\pi].$$
- The condition for truthful revelation of π can be written as
$$S(\pi', \pi) \leq S(\pi, \pi).$$
for almost all π and all π'.

Lemma

The condition for truthful revelation of π holds iff there exists a convex function $G(P_\pi) = S(\pi, \pi)$ with sub-gradient $\nabla G(P_\pi)$ such that
$$S(\pi', \pi) = G(P_{\pi'}) + \langle \nabla G(P_{\pi'}), P_\pi - P_{\pi'} \rangle.$$

The convex polytope of implementable rules

- Assume that
  1. The action space $\mathscr{A} \subset \mathbb{R}$ is convex, and
  2. analyst utility is linear, $v(A) = A$.
- Leading example:
  Binary decision $A \in \{0, 1\}$, randomized rule $\bar{\mathbf{a}} \in \mathscr{A} = [0, 1]$.
- Then the set of implementable rules is given by a convex polytope:

$$\bar{\mathbf{a}}(\pi, X_J, J) \in \mathscr{A}, \qquad \text{(Support)}$$

$$\bar{\mathbf{a}}(\pi, X_I, I) - \bar{\mathbf{a}}(\pi, X_J, J) \leq 0 \qquad \forall\, \pi, X_J, J, I \subset J, \qquad \text{(Monotonicity)}$$

$$\sum_{X_J, J} \left(\bar{\mathbf{a}}(\pi', X_J, J) - \bar{\mathbf{a}}(\pi, X_J, J)\right) \cdot P_\pi(X_J, J) \leq 0 \qquad \forall\, \pi', \pi. \qquad \text{(Truthful PAP)}$$

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 $\alpha \in (0, 1)$:

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

- Expected power:

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

Size control for implementable tests

Lemma

1. *If* $\bar{\mathbf{a}}$ *satisfies the monotonicity restriction, and*
2. *the support of* π *includes a value such that* $\mathsf{P}_\pi(J = K) = 1$,

then size control holds iff

$$\mathsf{E}[\bar{\mathbf{a}}(\pi, X, K)|\theta] \leq \alpha.$$

for all $\theta \in \Theta_0$.

Preview of optimal implementable tests

- Implementable tests are montonic,
  so that size control only binds for the full data.
- The optimal test
  - maximizes expected power,
  - subject to size control
  - and implementability.
- This test can be implemented as follows:
  - Ask the analyst to choose a full-data test that controls size.
  - For any report, assume the worst about the unreported components.
- The analyst problem of choosing the optimal full data test
  is a (simple) linear program.

The optimal test as solution to a linear program

$$
\begin{aligned}
\max_{\mathbf{a},t} \sum_{\pi,X,J} \mathbf{a}(\pi, X_J, J) \cdot P(\pi, X_J, J) & & & \text{(Expected power)} \\
\text{s.t.} \quad \sum_{X} t(\pi, X) \cdot \mathrm{P}_\theta(X) \leq \alpha & & \forall\, \pi, \theta \in \Theta_0, & \text{(Size control)} \\
a(\pi, X_J, J), t(\pi, X) \in [0,1] & & \forall\, \pi, J, X, & \text{(Support)} \\
\mathbf{a}(\pi, X_J, J) \leq t(\pi, X) & & \forall\, \pi, J, X, & \text{(Monotonicity)} \\
\sum_{X_J,J} \left(\bar{\mathbf{a}}(\pi', X_J, J) - \bar{\mathbf{a}}(\pi, X_J, J)\right) \cdot P_\pi(X_J, J) \leq 0 & & \forall\, \pi', \pi. & \text{(Truthful PAP)}
\end{aligned}
$$

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* $\mathbf{t}$ *which satisfies, for all* $\theta \in \Theta_0$,

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

- *and then implementing*

$$\mathbf{a}(t, X_J, J) = \min_{X';\, X'_J = X_J} t(X').$$

Sketch of proof

- Anything that can be implemented can be implemented by delegation.
- Implementable rules are monotonic.
- Monotonic rules satisfy size control iff they satisfy full-data size control.
- Subject to this constraint, analyst and decision-maker are aligned.
- Expected power given full-data size control and monotonicity is maximized by

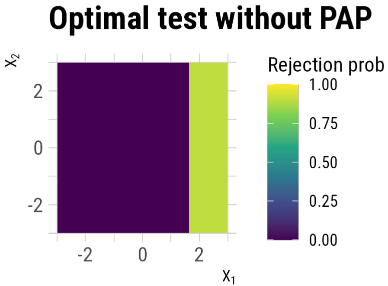
$$\mathbf{a}(t, X_J, J) = \min_{X';\, X'_J = X_J} t(X').$$

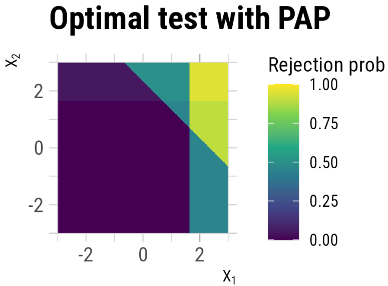
$\square$

The analyst's problem as a (simpler) linear program

$$
\begin{aligned}
\max_{\mathbf{a},t} \sum_{X_J,J} \mathbf{a}(X_J,J)\cdot P_\pi(X_J,J) & & & \text{(Expected power)}\\
\text{s.t.} \quad \sum_{X} t(X)\cdot \mathrm{P}_\theta(X) \le \alpha & & \forall\, \theta \in \Theta_0, & \quad \text{(Size control)}\\
\mathbf{a}(X_J,J), t(X) \in [0,1] & & \forall\, J, X, & \quad \text{(Support)}\\
\mathbf{a}(X_J,J) \le t(X) & & \forall\, J, X. & \quad \text{(Monotonicity)}
\end{aligned}
$$

Example revisited





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,
  3. optimal implementable unbiased estimators (not in these slides).

Next steps

1. Optimal implementable rules for Bayesian decision problems.
2. Explicit solutions for specific estimation and testing problems.
3. Intuitive characterizations of solutions
   to linear programming problems.

Thank you!