

Optimal Pre-Analysis Plans: Statistical Decisions Subject to Implementability

Maximilian Kasy Jann Spiess

October 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.
 - 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.

Our approach

- Import insights from contract theory / mechanism design to statistics.
 - We consider (optimal) statistical decision rules subject to the constraint of implementability.
 - PAPs are generically necessary for implementation.
 - They allow to leverage researcher expertise while maintaining incentive compatibility of non-selective reporting.
- 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 working paper on the same topic.

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

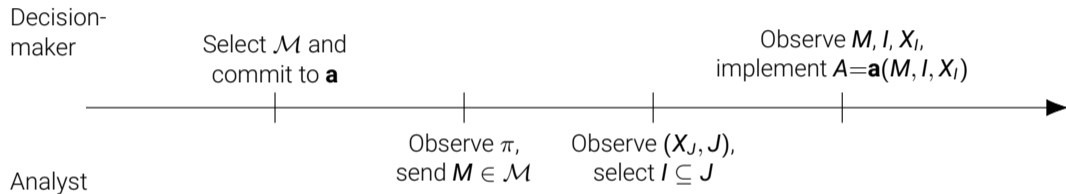
The PAP App

Conclusion and outlook

Setup: Notation

- Two parties, decision-maker and analyst.
- Message \mathbf{M} (“pre-analysis plan”) sent from analyst to decision-maker.
- Data $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n) \sim \mathbf{P}_\theta$.
 - Unknown parameter $\theta \in \Theta$.
- Index sets:
 - $\mathbf{K} = \{1, \dots, n\}$ fixed, finite, commonly known.
 - $\mathbf{J} \subset \mathbf{K}$ subset of data available to the analyst, privately known.
 - $\mathbf{I} \subset \mathbf{J}$ subset of available data reported to the decision-maker.
- Decision $\mathbf{A} \in \mathcal{A} \subseteq \mathbb{R}$.

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).

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

The PAP App

Conclusion and outlook

Motivating example: Normal testing

- $K = \{1, 2\}$.
- $X_1, X_2 \sim N(\theta, 1)$.
- Prior of the decision-maker : $(J_1, J_2) \sim \text{Ber}(\eta_1) \times \text{Ber}(\eta_2)$.
- 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 \mathbf{X}_t).
3. The optimal implementable test without a PAP.
4. The optimal implementable test with a PAP.

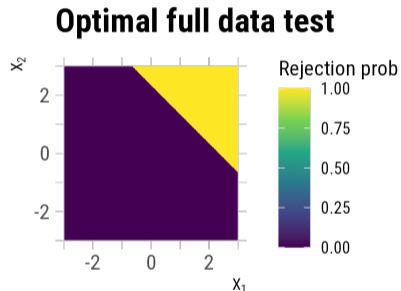
The optimal full data test

- 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

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

- Critical value:

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



The naive test

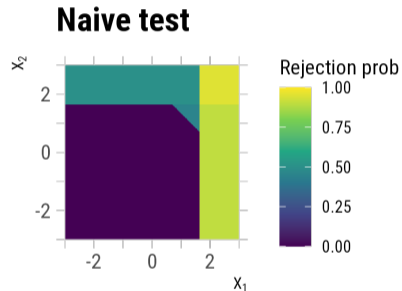
- Treat the reported data I as if there were no selective reporting.

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

- 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!

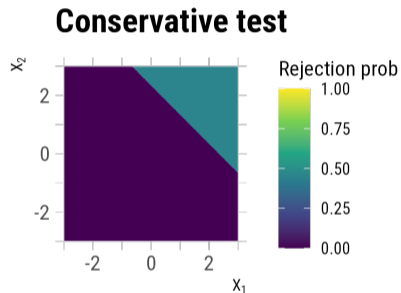


The conservative test

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

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

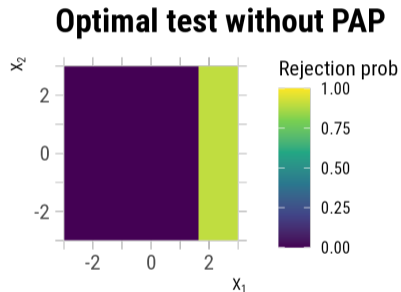
- This test controls size.
- But it has low power.



The optimal implementable test without PAP

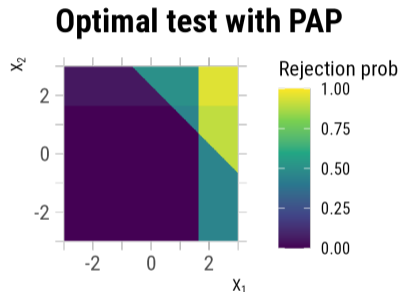
- Requirements:
 1. Size control.
 2. Incentive compatibility.
 3. Maximizes expected power.
- Solution without a PAP:
 1. Pick a full-data test,
 2. make worst-case assumptions about unreported components.
- Choose the full-data test to maximize expected power.
- Here:

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

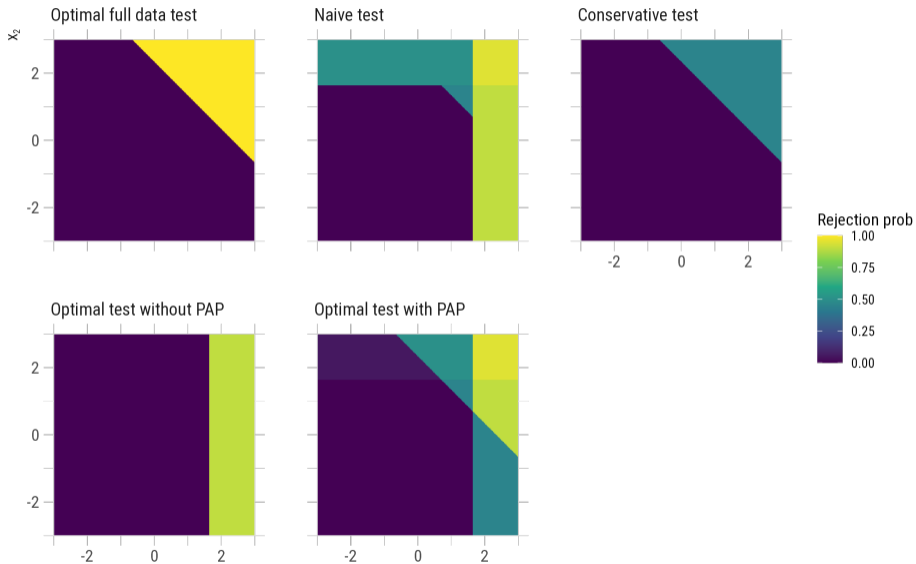


The optimal implementable test with PAP

- 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.

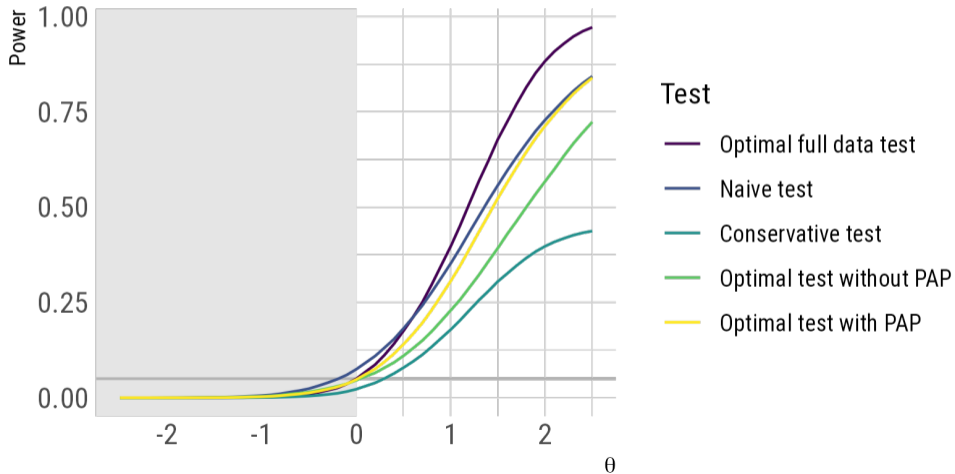


Rejection probabilities for different testing rules



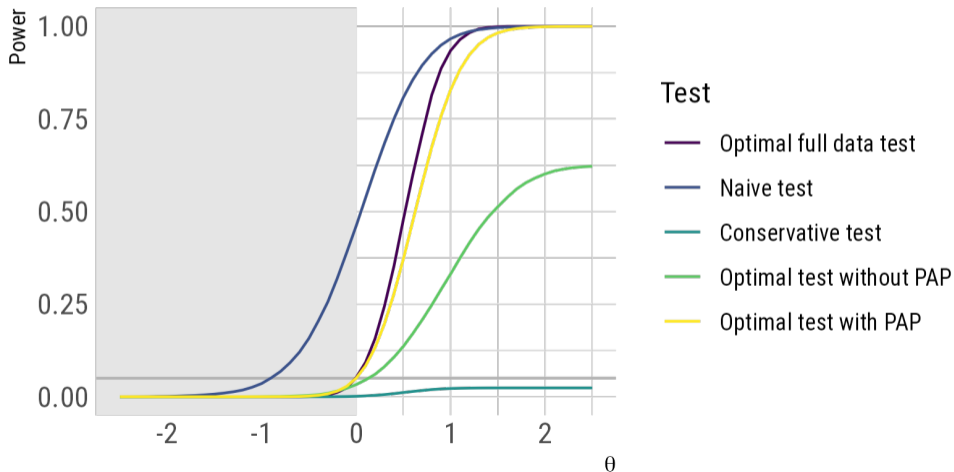
Degrees of freedom $n = 2$

Power curves for different testing rules



Degrees of freedom $n = 10$

Power curves for different testing rules



Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

The PAP App

Conclusion and outlook

Implementable decision functions

- A **reduced-form decision function** maps the full data into a decision **a**:

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

- A reduced-form decision function $\bar{\mathbf{a}}$ is **implementable**
 - if there exist a decision function **a**
 - with best responses M^*, 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^* of the analyst satisfies

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

- The optimal message M^* satisfies

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

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.
⇒ Can leverage analyst expertise!
- 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.
- For binary actions, the set of implementable decision functions is a **convex polytope**.
- Truthful revelation is closely related to **proper scoring**.

Implementability without PAPs

Proposition

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

1. $\bar{\mathbf{a}}$ does not depend on π , and
2. $\bar{\mathbf{a}}$ is **monotonic** in \mathbf{J} ,

$$\bar{\mathbf{a}}(\mathbf{X}_I, I) \leq \bar{\mathbf{a}}(\mathbf{X}_J, \mathbf{J})$$

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

Proof

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

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.



Revelation and delegation

Proposition

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

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

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 $\mathbf{a}(M, X_I, I)$, M^* , I^* .
 - Define $\tilde{\mathbf{a}}(b, X_J, J) = b(X_J, J)$ for $b \in \mathcal{B}$, where

$$\mathcal{B} = \{b(\cdot) : b(X_I, I) = \mathbf{a}(M, X_I, I), \text{ for some } M\}.$$

- It follows that $b(\cdot) = \mathbf{a}(M^*(\pi), X_I, I)$ is a best response to $\tilde{\mathbf{a}}$.
- Therefore $\tilde{\mathbf{a}}$ implements $\bar{\mathbf{a}}$.



The convex polytope of implementable rules

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

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

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

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

Proper scoring

- Define

$$S(\pi', \pi) = E[v(\bar{\mathbf{a}}(\pi', \mathbf{X}_J, \mathbf{J}) | \pi)].$$

- 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' .

Proposition

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

$$S(\pi', \pi) = G(P_{\pi'}) + \langle \nabla G(P_{\pi'}), P_\pi - P_{\pi'} \rangle.$$

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

The PAP App

Conclusion and outlook

Hypothesis testing

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

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

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

$$\sup_{\pi, \theta \in \Theta_0, \mathcal{J} \subseteq \{1, \dots, n\}} \mathbf{E}[\bar{\mathbf{a}}(\pi, \mathbf{X}_{\mathcal{J}}, \mathcal{J}) | \theta, \pi, \mathcal{J}] \leq \alpha.$$

- Expected power:

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

Preview of optimal implementable tests

- Implementable tests are monotonic, 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

$$\max_{\mathbf{a}, t} \int \mathbf{a}(\pi, \mathbf{X}_J, J) dP(\pi, \mathbf{X}_J, J) \quad \text{(Expected power)}$$

$$\text{s.t.} \quad \int t(\pi, \mathbf{X}) dP_{\theta}(\mathbf{X}) \leq \alpha \quad \forall \pi, \theta \in \Theta_0, \quad \text{(Size control)}$$

$$\mathbf{a}(\pi, \mathbf{X}_J, J), t(\pi, \mathbf{X}) \in [0, 1] \quad \forall \pi, \mathbf{J}, \mathbf{X}, \quad \text{(Support)}$$

$$\mathbf{a}(\pi, \mathbf{X}_J, J) \leq t(\pi, \mathbf{X}) \quad \forall \pi, \mathbf{J}, \mathbf{X}, \quad \text{(Monotonicity)}$$

$$\int (\bar{\mathbf{a}}(\pi', \mathbf{X}_J, J) - \bar{\mathbf{a}}(\pi, \mathbf{X}_J, J)) dP_{\pi}(\mathbf{X}_J, J) \leq 0 \quad \forall \pi', \pi. \quad \text{(Truthful PAP)}$$

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 $\theta \in \Theta_0$,*

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

- *and then implementing the test*

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

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

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

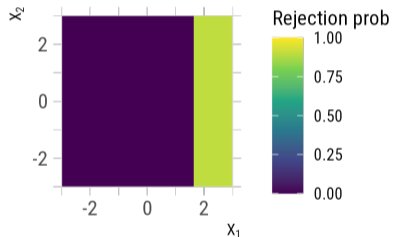


The analyst's problem as a (simpler) linear program

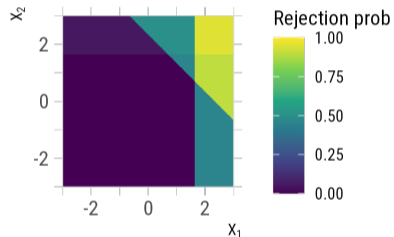
$$\begin{aligned} \max_b \int b(X_J, J) dP_\pi(X, J), & \quad \text{(Interim expected power)} \\ \text{s.t. } \int b(X, K) dP_{\theta_0}(X) \leq \alpha, & \quad \text{(Size control)} \\ b(X_J, J) \in [0, 1] \quad \forall J, X, & \quad \text{(Support)} \\ b(X_J, J) \leq b(X, K) \quad \forall J, X. & \quad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

Example revisited

Optimal test without PAP



Optimal test with PAP



Potentially optimal tests: Extremal points of \mathcal{B}

- Consider the “optimal test by delegation” mechanism,
- where \mathcal{B} is the set of tests available to the analyst.

Proposition

- *There exists a full-data test \mathbf{t} which is a best response of the analyst such that*

$$\mathbf{b}(X_J, J) = \inf_{X': X'_J = X_J} \mathbf{t}(X')$$

is extremal in \mathcal{B} .

- *Such a \mathbf{b} is extremal iff*
 1. $\mathbf{t}(X) \in \{0, \mathbf{q}, 1\}$ for all X , for some $0 < \mathbf{q} < 1$.
 2. *If there exists X such that $\mathbf{t}(X) = \mathbf{q}$, then $P_{\theta_0}(\mathbf{t}(X) = \mathbf{q}) > 0$.*
 3. *For any $X \neq X'$ such that $\mathbf{t}(X) = \mathbf{t}(X') = \mathbf{q}$, there exists a value J such that $X_J = X'_J$ and $\mathbf{b}(X_J, J) = \mathbf{b}(X'_J, J) = \mathbf{q}$.*

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

The PAP App

Conclusion and outlook

The PAP App

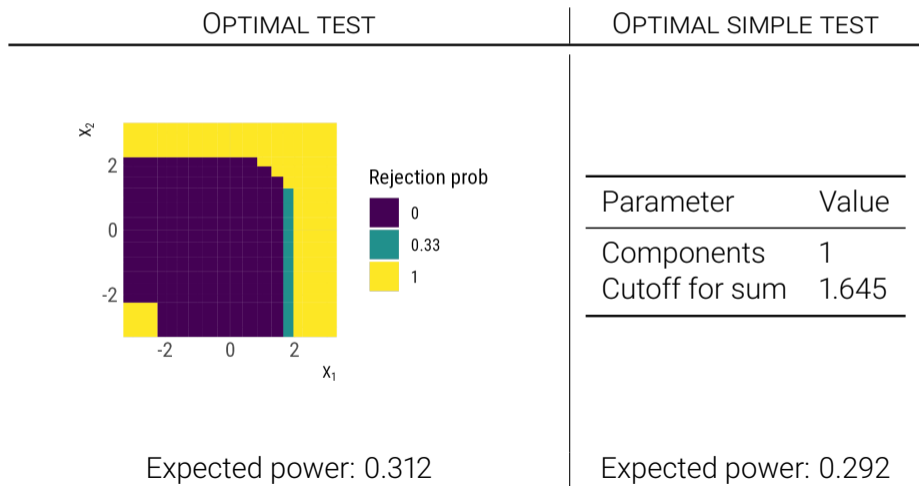
- The analyst needs to solve an LP problem, to find the optimal PAP.
- We provide an app which allows them to enter
 1. Prior parameters,
 2. null hypothesis and size,
 3. for normal and for binary data.
- The app provides
 1. The optimal full-data test.
 2. The optimal *simple* full-data test.
 3. The expected power of either.
- Suggestions for improvement are appreciated!

https://maxkasy.shinyapps.io/The_PAP_App/

Example 1 (normal data)

$$\eta = (0.9, 0.5).$$

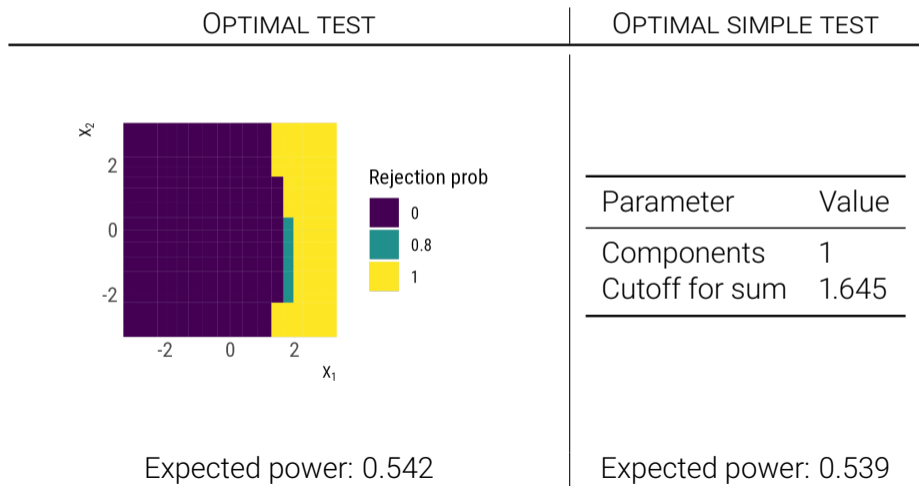
The interim prior is that X has a mean vector of $(1, 1)$, and a variance of $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.



Example 2 (normal data)

$$\eta = (0.9, 0.9).$$

The interim prior is that X has a mean vector of $(2, 0.5)$, and a variance of $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.



Example 3 (binary data)

$$\eta = (0.9, 0.5).$$

OPTIMAL TEST			OPTIMAL SIMPLE TEST	
X1	X2	t	Parameter	Value
1	0	0.44	Components	1,2
1	1	1.00	Cutoff for sum	1
			Rejection prob at the margin	0.19
Expected power: 0.283			Expected power: 0.228	

Example 4 (binary data)

$$\eta = (0.9, 0.5, 0.1).$$

OPTIMAL TEST				OPTIMAL SIMPLE TEST	
X1	X2	X3	t	Parameter	Value
1	0	0	0.44	Components	1,2
1	0	1	0.44	Cutoff for sum	1
1	1	0	1.00	Rejection prob at the margin	0.19
1	1	1	1.00		
Expected power: 0.283				Expected power: 0.228	

Introduction

Setup

Motivating example: Normal testing

Implementable decision functions

Hypothesis testing

The PAP App

Conclusion and outlook

Discussion

- Conflicts of interest, private information.
⇒ 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).

Thank you!