Studentized Sensitivity Analysis in Paired Observational Studies

Colin Fogarty
Massachusetts Institute of Technology
Operations Research and Statistics Group

November 20, 2017

On "No Treatment Effect"

What does it mean for a treatment to have no effect?

- Fisher: the treatment does not have an effect for *any* individual.
  - Called "Fisher's sharp null" (H_F).
- Neyman: the treatment has no effect *on average* for the individuals in our study (yet might affect certain individuals).
  - Called "Neyman's weak null" (H_N).

Truth of Fisher's sharp null implies truth of Neyman's weak null

- Neyman's null is composite; Fisher's null is a particular element ($H_F \in H_N$).

Does the distinction have practical implications?

The Neyman-Fisher Controversy

Neyman, in a paper presented to the Royal Statistical Society in 1935, suggested that for certain experimental designs Fisher's tools for inference assuming no effect at all might be anti-conservative when there's only no effect on average.

> "Professor R. A. Fisher, in opening the discussion, said he had hoped that Dr. Neyman's paper would be on a subject with which the author was fully acquainted, and on which he could speak with authority ... Since seeing the paper, he had come to the conclusion that Dr. Neyman had been somewhat unwise in his choice."
>
> — Fisher's transcribed response to Neyman (1935)

Time has done little to temper the debate. See Caughey et al. (2017+), "Beyond the Sharp Null..." for a nice overview.

Constant Effects in Observational Studies

Oftentimes, the debate is without consequence in practice for the analysis of randomized experiments

- Paired experiment: the largest variance for the classical estimator of the average treatment effect occurs when assuming the treatment effect is constant.

In **observational studies**, we assess the robustness of our study's findings to unmeasured confounding through a **sensitivity analysis**

- We'll review a model for such an analysis
- Methods are well established when we assume the treatment effect is **constant** for all individuals

Essential Heterogeneity

Suppose individuals self-select into receiving the intervention they know will be most beneficial for them

- For individuals with the same observed X, the one who self selected into the treatment group would likely have the larger treatment effect
- Narrative implies that treatment effects are not constant!

What does this say about assessing robustness to hidden bias assuming constant effects?

- **Are conventional sensitivity analyses overly optimistic?**

What are the Ramifications for Practitioners?

We seek to assess

1. Whether one can conduct a sensitivity analysis while allowing for heterogeneous effects.
2. Whether the conventional sensitivity analysis assuming constant effects is valid for heterogeneous effects.
3. If the conventional approach is invalid, whether assuming constant effects can materially alter one's conclusions.

Outline

1. Background
   - Review of Paired Studies
   - Sensitivity Analysis with Constant Effects
2. Sensitivity Analysis with Heterogeneous Effects
   - A Large-Sample Procedure
   - Studentized Sensitivity Analysis
3. Comparison of Methods
   - Motivating the Studentized Procedure
   - Quantifying the Gap
   - An Illustration: Alcoholism and Genetic Damage

Outline

Notation for Paired Studies

- The i^{th} of I pairs has one treated individual, $Z_{ij} = 1$, and one control, $Z_{ij'} = 0$, such that $Z_{i1} + Z_{i2} = 1$.
- Each individual has observed covariates, x_{ij}, and unobserved covariates, u_{ij}.
- r_{Tij} and r_{Cij} are the potential outcomes under treatment and control for individual j in pair i
- $\tau_{ij} = r_{Tij} - r_{Cij}$ is the treatment effect for each individual
- $R_{ij} = Z_{ij} r_{Tij} + (1 - Z_{ij}) r_{Cij}$ is the observed response
- $Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2})$ is the treated-minus-control pair difference in pair i

Notation for Paired Studies

Another way to write Y_i...

- $\ell_{ij} = (r_{Tij} + r_{Cij})/2$ is the "level" of the potential outcomes
- $\eta_i = (\ell_{i1} - \ell_{i2})$ is the difference in levels between the potential outcomes
- $\Delta_i = (\tau_{i1} + \tau_{i2})/2$ is the average of the treatment effects
- $Y_i = \Delta_i + (Z_{i1} - Z_{i2})\eta_i$

Let $\Omega_I = \{z : z_{i1} + z_{i2} = 1\}$, $\mathcal{Z}_I = \{Z \in \Omega_I\}$, $\mathcal{F}_I = \{r_{Tij}, r_{Cij}, x_{ij}, u_{ij}\}$

- What's $\pi_i = \text{pr}(Z_{i1} = 1 \mid \mathcal{F}_I, \mathcal{Z}_I)$?

Paired Experiments

Let's say we're in control of the assignment mechanism....

- $\pi_i = \text{pr}(Z_{i1} = 1 \mid \mathcal{F}_I, \mathcal{Z}_I) = 1/2$ by design
- $\text{pr}(Z = z \mid \mathcal{F}_I, \mathcal{Z}_I) = 2^{-I} = |\Omega_I|^{-1}$
- $E(Y_i \mid \mathcal{F}_I, \mathcal{Z}_I) = \Delta_i + (2\pi_i - 1)\eta_i = \Delta_i$
- $E(\bar{Y} \mid \mathcal{F}_I, \mathcal{Z}_I) = I^{-1} \sum_{i=1}^{I} \Delta_i = \bar{\Delta}$
- We'll call $\bar{\Delta}$ the *sample average treatment effect*

In words, the treated-minus-control paired difference is an unbiased estimator for the average of the treatment effects for the $2I$ individuals being compared.

Paired Observational Studies

But what if we can't control assignment to treatment?

- Observational study: group membership is not decided by us. In our data, we already know who has $Z = 1$ and who has $Z = 0$
- Idea: seek to mimic an idealized paired experiment
- Take initial study, and find pairs with one treated individual and one control where $x_{i1} \approx x_{i2}$ in each pair.

Suppose $(r_T, r_C) \perp\!\!\!\perp Z \mid X, \quad 0 < \text{pr}(Z = 1 \mid X) < 1$

- If $x_{i1} = x_{i2}$, then $\pi_i = \text{pr}(Z_{i1} = 1 \mid \mathcal{F}_I, \mathcal{Z}_I) = 1/2$!

Paired Observational Studies

If $(r_T, r_C) \not\perp\!\!\!\perp Z \mid X$, then $\pi_i \neq 1/2$

- Individuals in the same matched set might differ intrinsically on unobserved covariates u_{ij}.
- These latent discrepancies might make individuals more likely to self-select into the treatment group, invalidating strong ignorability
- $E(Y_i \mid \mathcal{F}_I, \mathcal{Z}_I) = \Delta_i + (2\pi_i - 1)\eta_i \neq \Delta_i$
- $E(\bar{Y} \mid \mathcal{F}_I, \mathcal{Z}_I) \neq \bar{\Delta}$

Outline

A Simple Model for Hidden Bias

In general, the conditional probability for Z is

$$\text{pr}(Z = z \mid \mathcal{F}_I, \mathcal{Z}_I) = \prod_{i=1}^{I} \pi_i^{z_{i1}} (1 - \pi_i)^{z_{i2}}.$$

The model of Rosenbaum (1987, 2002) parameterizes departures from a paired experiment through a parameter $\Gamma \geq 1$

$$1/(1 + \Gamma) \leq \pi_i \leq \Gamma/(1 + \Gamma).$$

- $\Gamma = 1$ recovers a paired randomized experiment.
- $\Gamma > 1$ encodes a family of distributions for $\text{pr}(Z = z \mid \mathcal{F}_I, \mathcal{Z}_I)$.

Let $\Pi(\Gamma) = \{\pi : 1/(1 + \Gamma) \leq \pi_i \leq \Gamma/(1 + \Gamma), \quad i = 1, ..., I\}$

Sensitivity Analysis for the Difference-in-Means

Suppose we want to test Fisher's notion of no effect

$$H_F : r_{Tij} = r_{Cij} \Leftrightarrow \tau_{ij} = 0 \ \ \forall i, j.$$

Consider as a test statistic the classical difference-in-means $\bar{Y}$. (referred to as the permutational t-statistic).

In each pair under the null hypothesis...

- $R_{ij} = r_{Tij} = r_{Cij}$.
- $Y_i = (Z_{i1} - Z_{i2})(R_{i1} - R_{i2}) = (Z_{i1} - Z_{i2})(r_{Ci1} - r_{Ci2})$.
  - $(r_{Ci1} - r_{Ci2}) = Y_i/(Z_{i1} - Z_{i2})$ is known given the data.
- If we swapped treatment assignment, $(Z_{i1} - Z_{i2})$ flips sign, so we instead observe $-Y_i$.

Randomization Inference

The distribution of $\bar{Y}$ under H_F is then

$$\begin{aligned}
&\text{pr}(\bar{Y} \leq k \mid \mathcal{F}_I, \mathcal{Z}_I, H_F) \\
&= \sum_{z \in \Omega_I} \mathbb{1}\{I^{-1} \sum_{i=1}^{I} (z_{i1} - z_{i2})(r_{Ci1} - r_{Ci2}) \leq k\} \prod_{i=1}^{I} \pi_i^{z_{i1}} (1 - \pi_i)^{z_{i2}}
\end{aligned}$$

For a randomized experiment, $\prod_{i=1}^{I} \pi_i^{z_{i1}} (1 - \pi_i)^{z_{i2}} = 1/2^I = 1/|\Omega_I|$ for any $z \in \Omega_I$.

- Same reference distribution as permutation test for the difference-in-means a paired study (despite starting from a different model)

In an observational study, $\pi_i \neq 1/2$.

If We Knew π_i...

For any allocation of π_i, the tail probability is easy to compute.

- For each $z \in \Omega_I$
  1. Compute $\mathbb{1}\{I^{-1}\sum_{i=1}^{I}(z_{i1} - z_{i2})(r_{Ci1} - r_{Ci2}) \leq k\}$.
  2. Weight $\mathbb{1}\{I^{-1}\sum_{i=1}^{I}(z_{i1} - z_{i2})(r_{Ci1} - r_{Ci2}) \leq k\}$ by $\prod_{i=1}^{I}\pi_i^{z_{i1}}(1-\pi_i)^{z_{i2}}$.
- At the end, sum up the weighted indicators.

Sadly, we don't know the weighting function $\prod_{i=1}^{I}\pi_i^{z_{i1}}(1-\pi_i)^{z_{i2}}$ since we don't know π_i

If we restrict $\pi \in \Pi(\Gamma)$ for a fixed Γ, can we bound the distribution of $\bar{Y}$?

Defining a Valid Sensitivity Analysis

In a sensitivity analysis, we take an adversarial approach and conduct the worst-case inference for a given Γ.

- Let $\varphi(\alpha, \Gamma)$ be a testing procedure (1 if reject the null, 0 otherwise) for a hypothesis H.

Valid Sensitivity Analysis

We call $\varphi(\alpha, \Gamma)$ an exact level-α sensitivity analysis at Γ for H if, for any $\pi \in \Pi(\Gamma)$ and for all I,

$$E(\varphi(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H) \leq \alpha.$$

We'll call it an asymptotically valid sensitivity analysis if $\pi \in \Pi(\Gamma)$ implies

$$\lim_{I \to \infty} E(\varphi(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H) \leq \alpha.$$

A Bounding Random Variable

Suppose $\pi \in \Pi(\Gamma)$, i.e. $1/(1+\Gamma) \leq \pi_i \leq \Gamma/(1+\Gamma)$.

Under H_F, Y_i is stochastically dominated by the random variable

$$T_{i,\Gamma} = V_{i,\Gamma}|Y_i|,$$

where

$$V_{i,\Gamma} = \pm 1, \quad \text{pr}(V_{i,\Gamma} = 1) = \Gamma/(1+\Gamma).$$

- $T_{i,\Gamma}$ places largest allowed mass on $|Y_i|$ for each pair.

Denote the randomization distribution for $\bar{T}_\Gamma$ by

$$\hat{F}_\Gamma(k) = \sum_{z \in \Omega_I} \mathbb{1}\{I^{-1}\sum_{i=1}^{I}(z_{i1} - z_{i2})|Y_i| \leq k\} \prod_{i=1}^{I} \left(\frac{\Gamma}{1+\Gamma}\right)^{z_{i1}} \left(\frac{1}{1+\Gamma}\right)^{z_{i2}}.$$

Exact Sensitivity Analysis under H_F

Define a candidate level$-\alpha$ sensitivity analysis for H_F by

$$\varphi_F(\alpha, \Gamma) = \mathbb{1}\{\bar{Y} \geq \hat{F}_\Gamma^{-1}(1-\alpha)\},$$

where $\hat{F}_\Gamma^{-1}(1-\alpha) = \inf\{k : \hat{F}_\Gamma(k) \geq 1-\alpha\}$.

Rosenbaum (2007)

For all I, if $\pi \in \Pi(\Gamma)$,

$$E(\varphi_F(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H_F) \leq \alpha.$$

How is this used in practice?

- Fixing α, find the largest Γ such that $\varphi_F(\alpha, \Gamma) = 1$.
- This changepoint Γ attests to how robust the rejection of H_F is to unmeasured biases.

Outline

The Sample Average Treatment Effect

Our developments relied upon constant effects! What about the sample average treatment effect?

$$\bar{\Delta} = I^{-1} \sum_{i=1}^{I} \Delta_i = (2I)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{2} \tau_{ij}$$

Suppose we wanted to test the null that the average of the treatment effects equaled zero,

$$H_N : \bar{\Delta} = 0.$$

What follows readily extends to the null $\bar{\Delta} = \bar{\Delta}_0$ for $\bar{\Delta}_0 \neq 0$.

Facilitating Inference

Define I new random variables by

$$D_{i,\Gamma} = Y_i - \left(\frac{\Gamma - 1}{1 + \Gamma}\right) |Y_i|,$$

Motivating the form...

- $T_{i,\Gamma} = V_{i,\Gamma}|Y_i|$ has expectation $\{(\Gamma - 1)/(1 + \Gamma)\}|Y_i|$ given Y_i.
- $|Y_i|$ was fixed across randomizations under H_F, but is random under H_N.

We'll further investigate the distribution of $\bar{D}_\Gamma$ given $\mathcal{F}_I$, $\mathcal{Z}_I$.

- At $\Gamma = 1$ (a paired experiment), $\bar{D}_1 = \bar{Y}$.

A Familiar Standard Error

As a means of assessing the uncertainty in $\bar{D}_\Gamma$, consider the variance estimator

$$S^2_{D_\Gamma} = \frac{1}{I(I-1)} \sum_{i=1}^{I} (D_{i,\Gamma} - \bar{D}_\Gamma)^2.$$

$\Gamma = 1 \Rightarrow D_{i,1} = Y_i$, so $S^2_{D_1}$ is the variance of the paired differences divided by I, the classical standard error for a paired study.

A Test for $\bar{\Delta} = 0$

Define a candidate level-α test valid at unmeasured confounding level Γ with a greater-than alternative by

$$\varphi_N(\alpha, \Gamma) = \mathbb{1}\{\bar{D}_\Gamma / S_{D_\Gamma} \geq \Phi^{-1}(1-\alpha)\},$$

where $\Phi(\cdot)$ is the standard normal CDF.

Proposition

Under mild regularity conditions, if $\pi \in \Pi(\Gamma)$

$$\lim_{I \to \infty} E(\varphi_N(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H_N) \leq \alpha.$$

That is, $\varphi_N(\alpha, \Gamma)$ asymptotically provides a valid level-α sensitivity analysis while accommodating effect heterogeneity.

Sketch of the Proof

1. Assuming $\pi \in \Pi(\Gamma)$, construct a stochastically bounding random variable $\bar{U}_\Gamma$ for $\bar{D}_\Gamma$.
   - Can't use the randomization distribution of $\bar{U}_\Gamma$ directly since it depends on the missing potential outcomes.
2. $E(\bar{U}_\Gamma \mid \mathcal{F}_I, \mathcal{Z}_I) \leq 0$ when $\bar{\Delta} = 0$ and $\pi \in \Pi(\Gamma)$.
3. $\text{var}(\bar{U}_\Gamma \mid \mathcal{F}_I, \mathcal{Z}_I) \leq E(S^2_{D_\Gamma} \mid \mathcal{F}_I, \mathcal{Z}_I)$ when $\pi \in \Pi(\Gamma)$.
4. $\bar{U}_\Gamma$ is asymptotically normal.
5. Replacing $E(\bar{U}_\Gamma \mid \mathcal{F}_I, \mathcal{Z}_I)$ with zero and $\text{var}(\bar{U}_\Gamma \mid \mathcal{F}_I, \mathcal{Z}_I)$ with $S^2_{D_\Gamma}$ doesn't corrupt the asymptotic level of the procedure.

Outline

Improving Finite-Sample Performance

Let's take another look at $\varphi_N(\cdot)$

$$\varphi_N(\alpha, \Gamma) = \mathbb{1}\{\bar{D}_\Gamma / S_{D_\Gamma} \geq \Phi^{-1}(1-\alpha)\},$$

Proposition would also hold if we replace $\Phi^{-1}(1-\alpha)$ with $\hat{G}_I^{-1}(1-\alpha)$ for any sequence of distribution functions such that, for all points t,

$$\hat{G}_I(t) \xrightarrow{p} \Phi(t) \text{ as } I \to \infty$$

Can we form a sequence of distribution functions better reflecting the finite-sample behavior of $\bar{D}_\Gamma / S_{D_\Gamma}$?

Studentization

Studentization is the division of a first-degree statistic derived from a sample by a sample-based estimate of its population standard deviation.

- Most famous example is in Student's t-test: we divide the sample average by its estimated standard error.

We utilize studentization in two ways

1. Defining our test statistic $\bar{D}_\Gamma / S_{D_\Gamma}$ (classical).
2. Constructing a randomization distribution (recent)

Studentized Permutation Tests

Two-sample permutation tests yield exact tests for

$$H_D : F_1 = F_2,$$

i.e. the null of equality in distribution. In general, they are not valid for testing that for some functional $\theta(\cdot)$ of the distributions

$$H_{\theta(D)} : \theta(F_1) = \theta(F_2).$$

- Example: $\theta(\cdot)$ could be the expectation.

Chung and Romano (2013): permutation tests based on **studentized** test statistics can be asymptotically valid for the null $H_{\theta(D)}$.

- Example: permute $(\bar{x}_1 - \bar{x}_2)/(s_1^2/n_1 + s_2^2/n_2)^{1/2}$ instead of $\bar{x} - \bar{y}$.

What's the Connection?

Sharp null hypotheses are closely tied to the nonparametric tests of equality in distribution.

- Fisher's sharp null, $H_F : r_{Tij} = r_{Cij}$, implies equality of empirical marginal distributions for potential outcomes under treatment and control.

Neyman's null is closer to a parametric hypothesis.

- Neyman's null, $H_N : \bar{\Delta} = 0$, implies equality of the **expectations** of the empirical marginals.

The connections to studentized permutation tests are clear in the case of unconfoundedness ($\Gamma = 1$). Do they extend to $\Gamma > 1$ in a sensitivity analysis?

A Candidate Randomization Distribution

Let's consider I new random variables $A_{i,\Gamma} = A_{i,\Gamma}(V_\Gamma, Y)$, defined by

$$\begin{aligned} A_{i,\Gamma} &= V_{i,\Gamma}|Y_i| - \left(\frac{\Gamma - 1}{1 + \Gamma}\right)|Y_i| \\ &= T_{i,\Gamma} - E(T_{i,\Gamma} \mid Y, \mathcal{F}_I, \mathcal{Z}_I) \end{aligned}$$

$T_{i,\Gamma} = V_{i,\Gamma}|Y_i|$ was the stochastically dominating variable for Y_i under Fisher's sharp null H_F. Compare $A_{i,\Gamma}$ to

$$D_{i,\Gamma} = Y_i - \left(\frac{\Gamma - 1}{1 + \Gamma}\right)|Y_i|$$

Define the variance estimate for $\bar{A}_\Gamma$ by

$$S^2_{A_\Gamma}(V_\Gamma, Y) = \frac{1}{I(I-1)} \sum_{i=1}^{I} (A_{i,\Gamma} - \bar{A}_\Gamma)^2$$

Studentized Randomization Distribution

Consider $\hat{G}_\Gamma(t)$ defined by

$$\hat{G}_\Gamma(t) = \sum_{z \in \Omega_I} \mathbb{1}\left\{ \frac{\bar{A}_\Gamma(z_1 - z_2, Y)}{S_{A_\Gamma}(z_1 - z_2, Y)} \leq t \right\} \prod_{i=1}^{I} \left(\frac{\Gamma}{1+\Gamma} \right)^{z_{i1}} \left(\frac{1}{1+\Gamma} \right)^{z_{i2}}.$$

$\hat{G}_\Gamma(t)$ studentizes $\bar{A}_\Gamma(\cdot)$ by $S_{A_\Gamma}(\cdot)$, using the same worst-case distribution for treatment assignments as the bounding distribution under constant effects, $\hat{F}_\Gamma(k)$.

- $\bar{A}_\Gamma(z_1 - z_2, Y)$ and $S_{A_\Gamma}(z_1 - z_2, Y)$ vary over $z \in \Omega_I$

Studentized Sensitivity Analysis

Proposition

Under mild regularity conditions, for all points t

$$\hat{G}_\Gamma(t) \xrightarrow{p} \Phi(t)$$

So, define the studentized sensitivity analysis by

$$\varphi_S(\alpha, \Gamma) = \mathbb{1}\left\{\bar{D}_\Gamma / S_{D_\Gamma} \geq \hat{G}_\Gamma^{-1}(1-\alpha)\right\}.$$

Corollary

If $\pi \in \Pi(\Gamma)$,

$$\lim_{I \to \infty} E(\varphi_S(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H_N) \leq \alpha.$$

Outline

Comparing $\varphi_N(\cdot)$ to $\varphi_S(\cdot)$

Our two asymptotically valid sensitivity analyses for testing H_N, $\bar{\Delta} = 0$, are

$$\varphi_N(\alpha, \Gamma) = \mathbb{1}\{\bar{D}_\Gamma / S_{D_\Gamma} \geq \Phi^{-1}(1-\alpha)\},$$

and

$$\varphi_S(\alpha, \Gamma) = \mathbb{1}\left\{\bar{D}_\Gamma / S_{D_\Gamma} \geq \hat{G}_\Gamma^{-1}(1-\alpha)\right\}.$$

In reality, *neither* of these tests are based on the true bounding distribution for $\bar{D}_\Gamma / S_{D_\Gamma}$. Why prefer one over the other?

What Does $\hat{G}_\Gamma(t)$ Buy Us?

Let's look at the following allocation of treatment effects, differences in levels, and probabilities with $I = 100$ pairs.

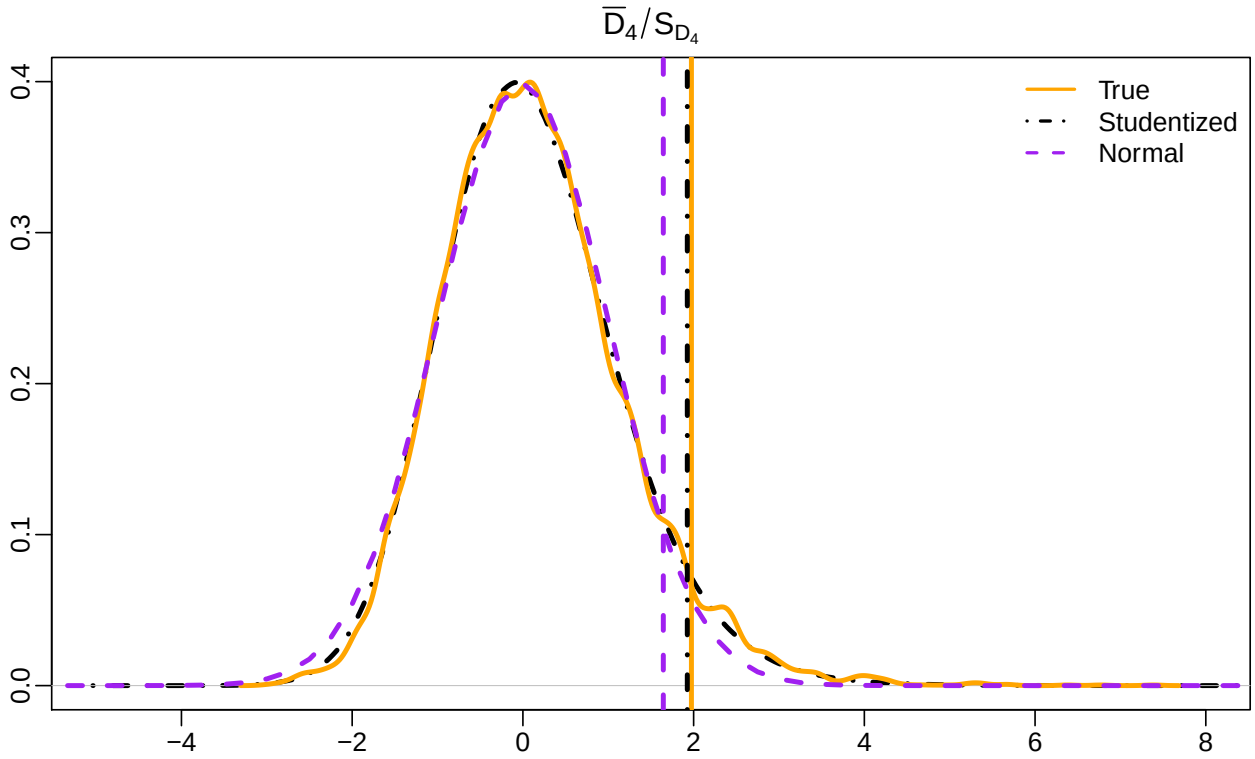
$$\{\Delta_i, \eta_i, \pi_i\} = \begin{cases} \{2.5, 5, 4/5\} & i = 1, ..., I/2 \\ \{-2.5, 20, 4/5\} & i = I/2 + 1, ..., I \end{cases}$$

- $\bar{\Delta} = 0$; $E(\bar{Y} \mid \mathcal{F}_I, \mathcal{Z}_I) = 7.5$
- $\pi \in \Pi(4)$

First, we'll compare procedures $\varphi_N(0.05, 4)$ and $\varphi_S(0.05, 4)$.

To provide a valid sensitivity analysis, the procedures need to furnish a null distribution that bounds the true randomization distribution of $\bar{D}_4 / S_{D_4}$.

Improving Finite-Sample Performance



Estimated Type I Error Rate for $\varphi_N(0.05, 4)$: 0.0798

Estimated Type I Error Rate for $\varphi_S(0.05, 4)$: 0.0530

Q: Was the Studentization Necessary?

Why not just decide to reject H_N at Γ based on the randomization distribution $\hat{F}_\Gamma(k)$ (i.e, the randomization distribution under Fisher's sharp null)?

$$\varphi_F(\alpha, \Gamma) = \mathbb{1}\{\bar{Y} \geq \hat{F}_\Gamma^{-1}(1-\alpha)\},$$
$$= \mathbb{1}\left\{\bar{D}_\Gamma \geq \hat{F}_\Gamma^{-1}(1-\alpha) - I^{-1}\sum_{i=1}^{I}\left(\frac{\Gamma-1}{1+\Gamma}\right)|Y_i|\right\}$$

We know that $E(\varphi_F(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H_F) \leq \alpha$ when the model holds at Γ for any I. Maybe it's also valid (at least asymptotically) under H_N...

- $\hat{G}_\Gamma(t)$ and $\hat{F}_\Gamma(k)$ use the same biased treatment assignment distribution.
- $\hat{F}_\Gamma(k)$ does *not* studentize within the permutation distribution.

A: Yes, it Was.

If the sensitivity model holds at $\Gamma = 1$, and we test at $\Gamma = 1$

$$\lim_{I \to \infty} E(\varphi_F(\alpha, 1) \mid \mathcal{F}_I, \mathcal{Z}_I, H_N) \leq \alpha.$$

That is, there's no issue in paired experiments. Does the insight carry over to observational studies?

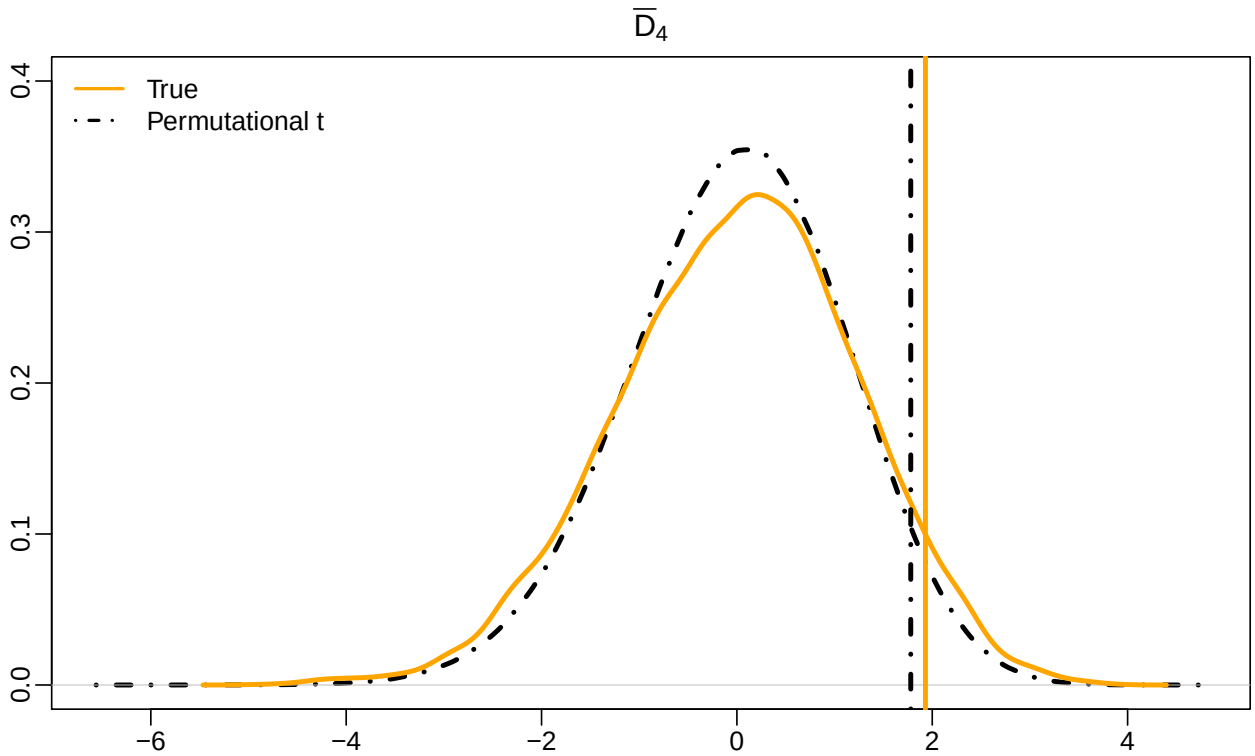
Proposition

For $\Gamma > 1$, there exist allocations of probabilities $\pi \in \Pi(\Gamma)$ and patterns of heterogeneous effects within H_N such that

$$\lim_{I \to \infty} E(\varphi_F(\alpha, \Gamma) \mid \mathcal{F}_I, \mathcal{Z}_I, H_N) > \alpha.$$

Failure of the Permutational t

In the same simulation setting as before, we also assess the performance $\varphi_F(0.05, 4)$ as a test for $\bar{\Delta} = 0$.



Estimated Type I Error Rate for $\varphi_F(0.05, 4)$: 0.0702

Outline

How Wrong Can the Permutational-t Be?

$\varphi_F(\alpha, \Gamma)$ is the standard way to conduct a sensitivity analysis for the difference-in-means. How misleading can inference be under heterogeneous effects?

Proposition

Suppose $\pi \in \Pi(\Gamma)$, but we instead test at $\Gamma + \epsilon$ for any $\epsilon > 0$. Under mild conditions,

$$\lim_{I \to \infty} E(\varphi_F(\alpha, \Gamma + \epsilon) \mid \mathcal{F}_I, \mathcal{Z}_I, H_N) = 0$$

You'd be worried if the above limit could also be greater than α at $\Gamma + \epsilon$. Rest easy.

Insight

Why is this the case?

- Permutational t-based sensitivity analysis actually correctly bounds the expectation of $\bar{Y}$ if $\pi \in \Pi(\Gamma)$.
- Unfortunately, it can have too small a variance (this isn't possible at $\Gamma = 1$, where constant effects yield the worst-case variance).
- Increasing Γ by ϵ, under weak assumptions, yields a positive gap between the true expectation of $\bar{D}_{\Gamma+\epsilon}$ and worst-case expectation over $\pi \in \Pi(\Gamma + \epsilon)$
- Discrepancy in variance ceases to matter once there's a gap

In the presence of unmeasured confounding, **bias trumps variance**.

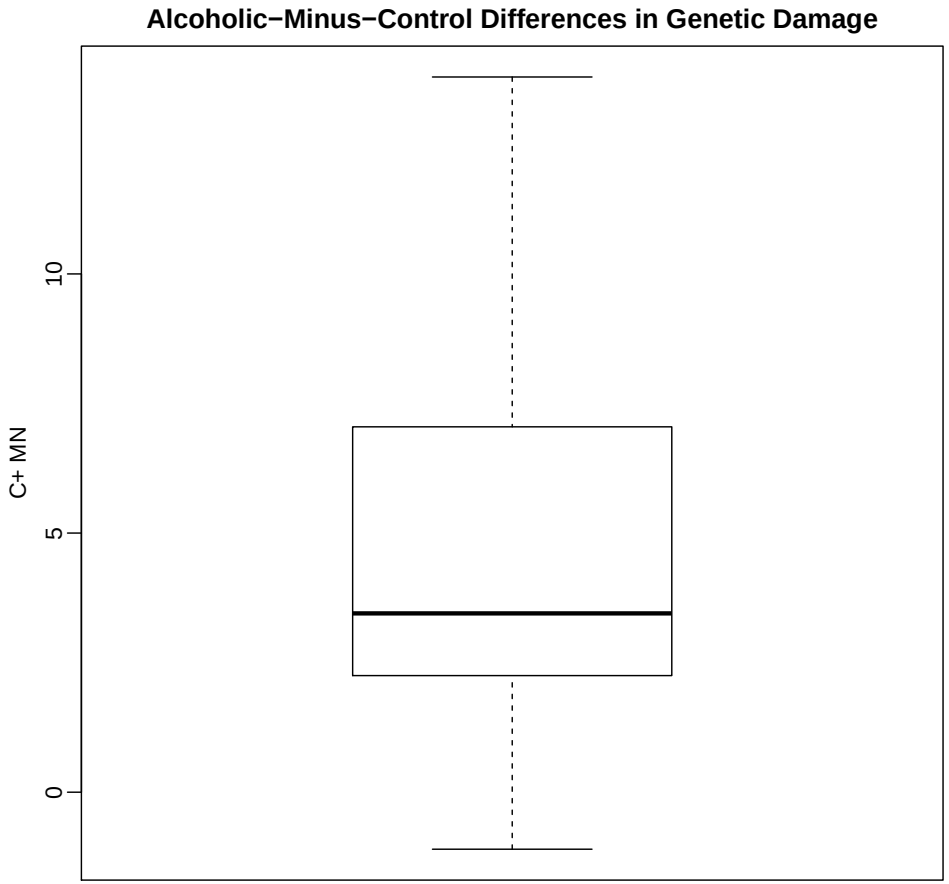
Outline

A Data Example

Maffei et al. (2001) investigated the impact of heavy drinking on genetic damage

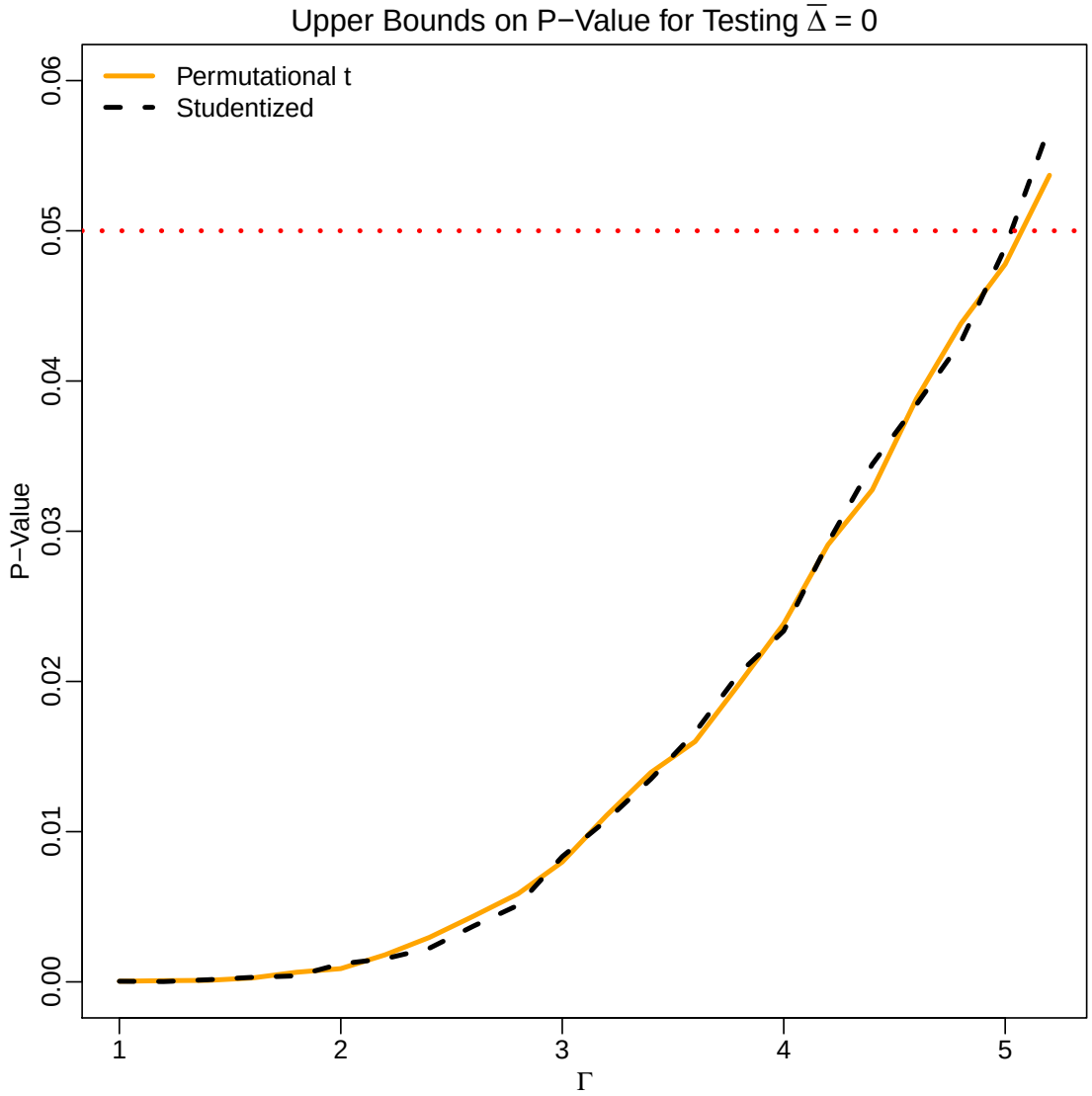
- Paired 20 alcoholics with controls based on age, smoking, gender
- Genetic damage: C+ MN (positive - more damage)
- Boxplot is right-skewed, indicating non-constant effects

Is there evidence of a positive average treatment effect?



Sensitivity Analyses

- $\varphi_F(0.05, \Gamma)$ rejects until $\Gamma = 5.04$
- $\varphi_S(0.05, \Gamma)$ rejects until $\Gamma = 5.02$
- Virtually identical perceptions of robustness to unmeasured confounding



Concluding Remarks: The Practitioner's Takeaway

1. Can one conduct a sensitivity analysis while allowing for heterogeneous effects?
   - **Yes**. Use the studentized sensitivity analysis.
2. Is the conventional sensitivity analysis based on the permutational t-test valid for heterogeneous effects?
   - **No**. It can be anti-conservative.
3. Can assuming constant effects through use of the permutational t-test materially alter your conclusions?
   - **Unlikely**. $\varphi_F(\alpha, \Gamma + \epsilon)$ is conservative asymptotically, limiting the extent to which it can mislead.
   - The changepoint Γ (where we go from rejecting to failing to reject) shouldn't be too far off.

Thanks!

- Fogarty, C.B. Studentized sensitivity analysis for the sample average treatment effect in paired observational studies. *arXiv*.

Bonus Slides: Design Sensitivity and the Power of a Sensitivity Analysis

How Much Have We Sacrificed?

Under heterogeneous effects

- $\varphi_F(\cdot)$ need not yield a valid sensitivity analysis
- $\varphi_S(\cdot)$ does asymptotically.

In scenarios where $\varphi_F(\cdot)$ would have been valid, how much is lost by using $\varphi_S(\cdot)$ instead?

- Ideally, the gap would be minimal in a quantifiable way
- Large gap could indicate
  1. Strength of constant effects assumption
  2. Conservativeness of $\varphi_S(\cdot)$.

A Favorable Situation...

Consider the following scenario:

- $Y_i \overset{iid}{\sim} \Upsilon(\cdot)$, $\Upsilon(\cdot)$ symmetric.
- $E(Y_i) = \mu$
- $\text{var}(Y_i) = \sigma^2 < \infty$
- $\Gamma = 1$

$\Rightarrow E[\varphi_S(\alpha, 1) \mid \mu = 0], E[\varphi_F(\alpha, 1) \mid \mu = 0] \leq \alpha$

...Unknown to the Practitioner

Suppose we're conducting an observational study where, unknown to us, we're in the scenario previously described with $\mu > 0$

- No hidden bias
- We use $\varphi_F(\cdot)$ and $\varphi_S(\cdot)$ to test for the existence of a treatment effect

Since we don't know that we're in this favorable scenario, we still want to assess the robustness of our study's finding of existence of a treatment effect

- Conduct a sensitivity analysis using $\varphi_S(\alpha, \Gamma)$ and $\varphi_F(\alpha, \Gamma)$
- Record the largest Γ such that we reject the null of no treatment effect

Design Sensitivity

In the scenario just described, as $I \to \infty$ there's a value $\tilde{\Gamma}$ called the **design sensitivity** such that

$$\lim_{I \to \infty} E[\varphi(\alpha, \Gamma)] \to \begin{cases} 1 & \Gamma < \tilde{\Gamma} \\ 0 & \Gamma > \tilde{\Gamma} \end{cases}$$

We prefer procedures with larger $\tilde{\Gamma}$ (akin the breakdown point of an estimator in robust statistics)

Proposition

$\varphi_F(\cdot)$ and $\varphi_S(\cdot)$ have the same design sensitivity, given by

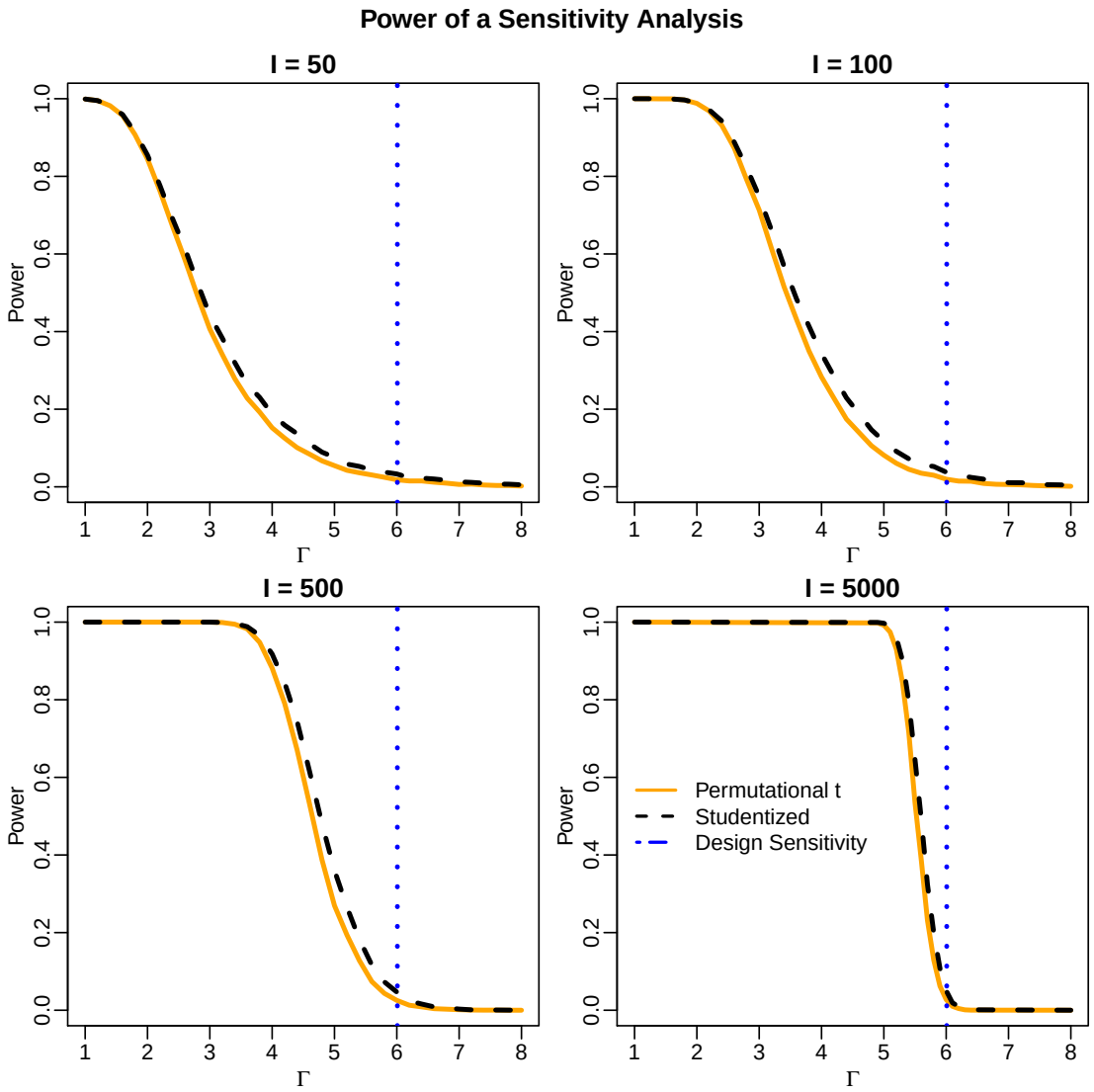
$$\tilde{\Gamma} = \frac{E|Y_i| + |\mu|}{E|Y_i| - |\mu|}$$

Power of a Sensitivity Analysis

We can also investigate the finite-sample power of the sensitivity analysis for various values of Γ and number of pairs I.

- $I = 50, 100, 500, 5000$
- $Y_i \overset{iid}{\sim} \mathcal{N}(1/2, 1/2) \Rightarrow \tilde{\Gamma} \approx 6$
- Test using $\varphi_F(0.05, \Gamma)$ and $\varphi_S(0.05, \Gamma)$ for a range of Γ

Power as a Function of Γ



Not a mistake. $\varphi_S(0.05, \Gamma)$ consistently more powerful

Explaining a "Paradox"

$\varphi_F(\cdot)$ tests Fisher's null, while $\varphi_S(\cdot)$ tests the weaker Neyman null.

- How could it reject less frequently?

$\varphi_F(\cdot)$ uses variance formed **under the null**.

- At $\Gamma = 1$: $I^{-2} \sum_{i=1}^{I} (Y_i - 0)^2$

$\varphi_S(\cdot)$ uses a variance estimator centered by the **observed** difference-in-means

- At $\Gamma = 1$: $(I(I-1))^{-1} \sum_{i=1}^{I} (Y_i - \bar{Y})^2$

If the null is false and $\mu >> 0$...

- Variance estimator used by $\varphi_F(\cdot)$ unduly inflated.
- $\varphi_S(\cdot)$ exploits local information even while testing $\mu = 0$.

Ding (2017) describes this issue in randomized experiments.