Gaussian Prepivoting for Finite Population Causal Inference

Peter Cohen

April 1, 2021

MIT, Operations Research Center
plcohen@mit.edu

Coauthors

Colin B. Fogarty
Assistant Professor of Operations Research and Statistics,
Massachusetts Institute of Technology, Cambridge, MA 02142
(*cfogarty@mit.edu*)

Set-Up

The Objective

Detect Treatment Effect
Between two groups (treated vs. control) determine if the treatment did anything.

The Objective

Detect Treatment Effect
Between two groups (treated vs. control) determine if the treatment did anything.

Mathematical Formalism
Design an algorithm which builds level-α hypothesis tests for treatment effect given some user-specified test statistic.

(A level-α test controls the Type I error rate below α.)

The Objective

Detect Treatment Effect
Between two groups (treated vs. control) determine if the treatment did anything.

Mathematical Formalism
Design an algorithm which builds level-α hypothesis tests for treatment effect given some user-specified test statistic.

(A level-α test controls the Type I error rate below α.)

Foreshadowing: One procedure with finite sample exactness under one null with asymptotic conservativeness under other null (unified inference).

Notation

N total units (n_1 treated & n_0 control).

Unit i has outcomes $\mathbf{y}_i(0), \mathbf{y}_i(1) \in \mathbb{R}^d$ and covariates $\mathbf{x}_i \in \mathbb{R}^k$.
(*For the sake of this talk, we ignore covariates.*)

$\mathbf{Z}_i$ is the indicator of treatment (1 if treated and 0 if control).

Treatment effect for i^{th} unit is $\boldsymbol{\tau}_i = \mathbf{y}_i(1) - \mathbf{y}_i(0)$.

Average treatment effect is $\bar{\boldsymbol{\tau}} = N^{-1} \sum_{i=1}^{N} \boldsymbol{\tau}_i$.

Drop subscripts to denote concatenation: e.g.,
$\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$.

Allowable treatment allocation set is Ω: e.g.,
$\Omega_{CRE} = \{\mathbf{z} \in \{0,1\}^N \text{ s.t. } \sum_i \mathbf{z}_i = n_1\}$.

Some Asymptotic Quantities

The proportion of treated units: $n_1/N \to p \in (0, 1)$.

Limiting variances & covariances are denoted with Σ_∞; e.g., $\Sigma_{y(1),\infty}$ is limit for treated potential outcomes.

A Foundational Question

Treatment Effect
Between two groups (treated vs. control) determine if the treatment **did anything**.

A Foundational Question

Treatment Effect
Between two groups (treated vs. control) determine if the treatment **did anything**.

Competing Definitions:

Fisher's Sharp Null (H_F) versus Neyman's Weak Null (H_N)

Fisher's Sharp Null

No effect means no effect (Rosenbaum)

Fisher's Sharp Null

No effect means no effect (Rosenbaum) $\Longrightarrow$

$H_F : \boldsymbol{\tau}_i = \mathbf{0} \;\; (i = 1, \ldots, N).$

Fisher's Sharp Null

No effect means no effect (Rosenbaum) $\implies$
$H_F : \boldsymbol{\tau}_i = \mathbf{0} \;\; (i = 1, \ldots, N)$.

Alternative hypothesis is $H_A : \exists\, i$ s.t. $\boldsymbol{\tau}_i \neq \mathbf{0}$.

Fisher's Sharp Null

No effect means no effect (Rosenbaum) $\Longrightarrow$
$H_F : \boldsymbol{\tau}_i = \mathbf{0} \;\; (i = 1, \ldots, N)$.

Alternative hypothesis is $H_A : \exists\, i$ s.t. $\boldsymbol{\tau}_i \neq \mathbf{0}$.

Pros

- Randomization inference provides exact tests (i.e., tests are guaranteed to be level-α for each $N \in \mathbb{N}$).

Cons

- Sometimes thought to be a very brittle null.
- Can be misinterpreted by users in practice.

Neyman's Weak Null

Treatment did nothing on average

Neyman's Weak Null

Treatment did nothing on average $\implies H_N : \bar{\boldsymbol{\tau}} = \mathbf{0}$.

Neyman's Weak Null

Treatment did nothing on average $\implies H_N : \bar{\boldsymbol{\tau}} = \mathbf{0}$.

Alternative hypothesis is $H_A : \bar{\boldsymbol{\tau}} \neq \mathbf{0}$.

Neyman's Weak Null

Treatment did nothing on average $\implies H_N : \bar{\boldsymbol{\tau}} = \mathbf{0}$.

Alternative hypothesis is $H_A : \bar{\boldsymbol{\tau}} \neq \mathbf{0}$.

Pros

- Practitioners rarely misinterpret.
- Less "brittle" than H_F.

Cons

- Randomization inference can be **anticonservative** when there is treatment effect heterogeneity (i.e., when $\boldsymbol{\tau}_i$ is not constant for all i).

Neyman's Weak Null

Treatment did nothing on average $\implies H_N : \bar{\boldsymbol{\tau}} = \mathbf{0}$.

Alternative hypothesis is $H_A : \bar{\boldsymbol{\tau}} \neq \mathbf{0}$.

Pros

- Practitioners rarely misinterpret.
- Less "brittle" than H_F.

Cons

- Randomization inference can be **anticonservative** when there is treatment effect heterogeneity (i.e., when $\boldsymbol{\tau}_i$ is not constant for all i). H_N does not constrain counterfactuals!

Refined Objective

1. Let a practitioner pick a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ which is asymptotically valid for H_N.

Refined Objective

1. Let a practitioner pick a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ which is asymptotically valid for H_N.
2. Automatically build a new statistic $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ based on $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$.

Refined Objective

1. Let a practitioner pick a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ which is asymptotically valid for H_N.
2. Automatically build a new statistic $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ based on $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$.
3. Use **FRT** with $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ to test H_N asymptotically at level-α, but retain finite N guarantee of exactness under H_F.

Refined Objective

1. Let a practitioner pick a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ which is asymptotically valid for H_N.
2. Automatically build a new statistic $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ based on $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$.
3. Use **FRT** with $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ to test H_N asymptotically at level-α, but retain finite N guarantee of exactness under H_F.

Design a procedure to get finite sample exactness under H_F **for free** without sacrificing asymptotic inference for H_N.

Gaussian Prepivoting

Main idea: Use the Fisher Randomization Test, but let your test statistic be the p-value of a large-sample test for H_N using $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$.

Why "prepivoting"?
Beran (1980s) and Chung & Romano (2016) proposed transforming a test statistic T by an estimate of its distribution $\hat{F}$ to get a new statistic $\hat{F}(T)$.

- Beran's objective was asymptotic refinement in *iid* models.
- We use prepivoting to get asymptotically pivotal quantities (and stochastic dominance).

Inference Approaches and Sharp Dominance

The Observed Difference in Means

For any matrix $\mathbf{r} \in \mathbb{R}^{N \times \ell}$ and $\mathbf{w} \in \Omega$, we define the function

$$\hat{\tau}(\mathbf{r}, \mathbf{W}) = \frac{1}{n_1} \sum_{i=1}^{N} W_i \mathbf{r}_i - \frac{1}{n_0} \sum_{i=1}^{N} (1 - W_i) \mathbf{r}_i.$$

The Observed Difference in Means

For any matrix $\mathbf{r} \in \mathbb{R}^{N\times\ell}$ and $\mathbf{w} \in \Omega$, we define the function

$$\hat{\tau}(\mathbf{r}, \mathbf{W}) = \frac{1}{n_1}\sum_{i=1}^{N} W_i \mathbf{r}_i - \frac{1}{n_0}\sum_{i=1}^{N}(1 - W_i)\mathbf{r}_i.$$

Special Cases

- $\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is the treated-minus-controlled difference in outcome means.
- $\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W})$ is the treated-minus-controlled difference in outcome means *after relabeling treatment allocation by* $\mathbf{W}$.

Using the Difference in Means to Test H_F

For a statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ generate the reference distribution:

$$\hat{\mathscr{P}}_T(t) = \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T\left(\mathbf{y}(\mathbf{Z}), \mathbf{w}\right) \leq t\right\}.$$

Using the Difference in Means to Test H_F

For a statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ generate the reference distribution:

$$\hat{\mathscr{P}}_T(t) = \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T\left(\mathbf{y}(\mathbf{Z}), \mathbf{w}\right) \leq t\right\}.$$

The reference distribution $\hat{\mathscr{P}}_T$ is the conditional distribution of $T\left(\mathbf{y}(\mathbf{Z}), \mathbf{W}\right)$ given $\mathbf{Z}$ with $\mathbf{W} \sim \text{Unif}(\Omega)$.

Using the Difference in Means to Test H_F

Theorem (Informal)
Under H_F the test

$$\varphi_T(\alpha) = \mathbb{1}\left\{T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \geq \hat{\mathscr{P}}_T^{-1}(1-\alpha)\right\}.$$

is exact.

Using the Difference in Means to Test H_F

Theorem (Informal)
Under H_F the test

$$\varphi_T(\alpha) = \mathbb{1}\left\{T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \geq \hat{\mathscr{P}}_T^{-1}(1-\alpha)\right\}.$$

is exact.

Historically well known result. Driving reason for use of the Fisher Randomization Test (FRT) with H_F.

Why Does This Work?

In a CRE **true** (randomization) distribution of a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is

$$\begin{aligned}\mathscr{R}_T(t) &= \mathbb{P}\left(T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \leq t\right) \\ &= \sum_{\mathbf{w} \in \Omega} \mathbb{P}\left(\mathbf{Z} = \mathbf{w}\right) \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \leq t\right\} \\ &= \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \leq t\right\}.\end{aligned}$$

Why Does This Work?

In a CRE **true** (randomization) distribution of a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is

$$\begin{aligned} \mathscr{R}_T(t) &= \mathbb{P}\left(T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \leq t\right) \\ &= \sum_{\mathbf{w} \in \Omega} \mathbb{P}\left(\mathbf{Z} = \mathbf{w}\right) \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \leq t\right\} \\ &= \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \leq t\right\}. \end{aligned}$$

Under H_F the observations $\mathbf{y}(\mathbf{Z})$ exactly equals the potential outcomes $\mathbf{y}(0)$ and $\mathbf{y}(1) \implies \mathbf{y}(\mathbf{w}) = \mathbf{y}(\mathbf{Z})$ for all $\mathbf{w} \in \Omega$.

Why Does This Work?

In a CRE **true** (randomization) distribution of a test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is

$$\begin{aligned}\mathscr{R}_T(t) &= \mathbb{P}\left(T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \le t\right) \\ &= \sum_{\mathbf{w} \in \Omega} \mathbb{P}\left(\mathbf{Z} = \mathbf{w}\right) \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \le t\right\} \\ &= \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \le t\right\}.\end{aligned}$$

Under H_F the observations $\mathbf{y}(\mathbf{Z})$ exactly equals the potential outcomes $\mathbf{y}(0)$ and $\mathbf{y}(1) \implies \mathbf{y}(\mathbf{w}) = \mathbf{y}(\mathbf{Z})$ for all $\mathbf{w} \in \Omega$.

Consequence: under H_F the true randomization distribution $\mathscr{R}_T$ **exactly matches** the reference distribution $\hat{\mathscr{P}}_T$.

In general, $\mathscr{R}_T \neq \hat{\mathscr{P}}_T$ under H_N due to effect heterogeneity. Furthermore, $\hat{\mathscr{P}}_T$ is a random CDF under H_N.

Using the Difference in Means to Test H_N

Theorem (Informal (Li & Ding))
Under H_N and mild conditions on the potential outcomes

$$\sqrt{N}(\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) - \overline{\boldsymbol{\tau}}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\Sigma_{y(1),\infty}}{p} + \frac{\Sigma_{y(0),\infty}}{1-p} - \Sigma_{\tau,\infty}\right)$$

Using the Difference in Means to Test H_N

Theorem (Informal (Li & Ding))
Under H_N and mild conditions on the potential outcomes

$$\sqrt{N}(\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) - \overline{\boldsymbol{\tau}}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\Sigma_{y(1),\infty}}{p} + \frac{\Sigma_{y(0),\infty}}{1-p} - \Sigma_{\tau,\infty}\right)$$

Because we can't observe counterfactuals $\Sigma_{\tau,\infty}$ cannot be consistently estimated from observed data.

So just use c_α as the critical value from $\mathcal{N}\left(\mathbf{0}, \frac{\Sigma_{y(1),\infty}}{p} + \frac{\Sigma_{y(0),\infty}}{1-p}\right)$
Yields an **asymptotically** conservative test for H_N.

Fundamental Problem

The critical value from $\mathcal{N}\left(\mathbf{0}, \frac{\Sigma_{y(1),\infty}}{p} + \frac{\Sigma_{y(0),\infty}}{1-p}\right)$ only gives asymptotically conservative test; there is no way to get exactness from that because we used asymptotic distributional properties of $\sqrt{N}(\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) - \overline{\boldsymbol{\tau}})$ to derive the test.

What if we use Fisher Randomization Test to test H_N?
Problem: The test can be anticonservative (Type I error rate substantially exceeds α).

Classical Solution

Studentize! Estimate variance of $\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ with $\hat{V}$ and use

$$T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \frac{\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})}{\sqrt{\hat{V}}}.$$

Wu & Ding (2020) show that the FRT is exact under H_F for the studentized statistic **and also is asymptotically conservative under** H_N!

Classical Solution

Studentize! Estimate variance of $\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ with $\hat{V}$ and use

$$T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \frac{\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})}{\sqrt{\hat{V}}}.$$

Wu & Ding (2020) show that the FRT is exact under H_F for the studentized statistic **and also is asymptotically conservative under** H_N!

Problem solved?

Problem Solved? Not Quite.

Does studentizing fix your favorite statistic for H_N?

Problem Solved? Not Quite.

Does studentizing fix your favorite statistic for H_N?

- You need to analyse each statistic and figure out the right studentizing factor separately (Wu & Ding have lots of examples).

Problem Solved? Not Quite.

Does studentizing fix your favorite statistic for H_N?

- You need to analyse each statistic and figure out the right studentizing factor separately (Wu & Ding have lots of examples).
- Some statistics – even after studentization – aren't amenable to the Fisher Randomization Test for H_N.

Example: The max absolute t-statistic for multivariate data

$$T_{|max|}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \max_{1 \leq j \leq d} \frac{\sqrt{N}|\hat{\tau}_j|}{\sqrt{\hat{V}_{jj}}}.$$

A General Framework Through Stochastic Dominance

Suppose

- $\mathscr{R}_T$ converges weakly to $\mathscr{R}_{T,\infty}$

A General Framework Through Stochastic Dominance

Suppose

- $\mathscr{R}_T$ converges weakly to $\mathscr{R}_{T,\infty}$
- $\hat{\mathscr{P}}_T$ converges weakly in probability to $\mathscr{P}_{T,\infty}$
  - For all continuity points of $\mathscr{P}_{T,\infty}$, $\hat{\mathscr{P}}_T(t) \xrightarrow{p} \mathscr{P}_{T,\infty}(t)$

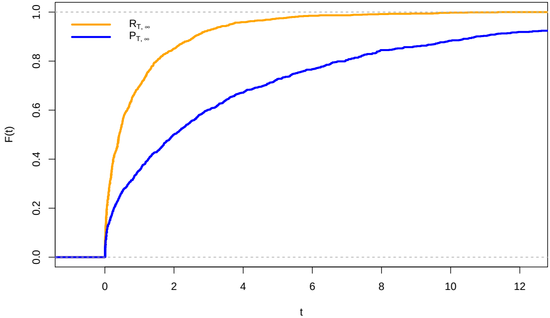
A General Framework Through Stochastic Dominance

Suppose

- $\mathscr{R}_T$ converges weakly to $\mathscr{R}_{T,\infty}$
- $\hat{\mathscr{P}}_T$ converges weakly in probability to $\mathscr{P}_{T,\infty}$
  - For all continuity points of $\mathscr{P}_{T,\infty}$, $\hat{\mathscr{P}}_T(t) \xrightarrow{p} \mathscr{P}_{T,\infty}(t)$

Asymptotic Sharp Dominance
$T(\cdot,\cdot)$ is called **asymptotically sharp-dominant** if, for all t,

$$\mathscr{P}_{T,\infty}(t) \leq \mathscr{R}_{T,\infty}(t).$$



Why is this useful?

Suppose that $T(\cdot, \cdot)$ is asymptotically sharp-dominant under H_N, so

$$\mathscr{P}_{T,\infty}(t) \leq \mathscr{R}_{T,\infty}(t) \quad \forall\, t \in \mathbb{R}.$$

Then T's limiting upper tail probabilities may be upper bounded by those of $\mathscr{P}_{T,\infty}$:

$$\underbrace{1 - \mathscr{P}_{T,\infty}(t)}_{p\text{-value under reference}} \geq \underbrace{1 - \mathscr{R}_{T,\infty}(t)}_{\text{true } p\text{-value}} \quad \forall\, t \in \mathbb{R}.$$

Why is this useful?

Suppose that $T(\cdot,\cdot)$ is asymptotically sharp-dominant under H_N, so

$$\mathscr{P}_{T,\infty}(t) \leq \mathscr{R}_{T,\infty}(t) \quad \forall\, t \in \mathbb{R}.$$

Then T's limiting upper tail probabilities may be upper bounded by those of $\mathscr{P}_{T,\infty}$:

$$\underbrace{1 - \mathscr{P}_{T,\infty}(t)}_{p\text{-value under reference}} \geq \underbrace{1 - \mathscr{R}_{T,\infty}(t)}_{\text{true } p\text{-value}} \quad \forall\, t \in \mathbb{R}.$$

If T is asymptotically sharp dominant then the FRT is exact under H_F and asymptotically conservative under H_N.

How Do We Make Sharp Dominant T?

Certainly not all test statistics are asymptotically sharp dominant (e.g., unstudentized difference in means, max absolute t-statistic, ...).

How Do We Make Sharp Dominant T?

Certainly not all test statistics are asymptotically sharp dominant (e.g., unstudentized difference in means, max absolute t-statistic, ...).

Question: Is there a way to take a test statistic which is not asymptotically sharp dominant and make a new statistic that is?

How Do We Make Sharp Dominant T?

Certainly not all test statistics are asymptotically sharp dominant (e.g., unstudentized difference in means, max absolute t-statistic, ...).

Question: Is there a way to take a test statistic which is not asymptotically sharp dominant and make a new statistic that is?

Answer: Prepivoting!

Prepivoting

A Quick Reminder

Beran's prepivoting:

1. Take as input a test statistic T,
2. Form an estimate of T's distribution, $\hat{F}$,
3. Transform T by $\hat{F}$ yielding $G = \hat{F}(T)$.

A Quick Reminder

Beran's prepivoting:

1. Take as input a test statistic T,
2. Form an estimate of T's distribution, $\hat{F}$,
3. Transform T by $\hat{F}$ yielding $G = \hat{F}(T)$.

Two Main Ingredients:
Test statistic T and distributional estimator $\hat{F}$.

How Do We Make Sharp Dominant T?

The base statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ must be of the form

$$T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau}),$$

where $\hat{\xi}$ and f_η satisfy the following conditions over some set Ξ.

How Do We Make Sharp Dominant T?

The base statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ must be of the form

$$T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau}),$$

where $\hat{\xi}$ and f_η satisfy the following conditions over some set Ξ.

Conditions on f_η
For any $\eta \in \Xi$, $f_\eta(t) : \mathbb{R}^d \mapsto \mathbb{R}_+$ is jointly continuous in η and t, quasi-convex, and nonnegative with $f_\eta(t) = f_\eta(-t)$ for all $t \in \mathbb{R}^d$.

How Do We Make Sharp Dominant T?

The base statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ must be of the form

$$T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau}),$$

where $\hat{\xi}$ and f_η satisfy the following conditions over some set Ξ.

Conditions on f_η
For any $\eta \in \Xi$, $f_\eta(t) : \mathbb{R}^d \mapsto \mathbb{R}_+$ is jointly continuous in η and t, quasi-convex, and nonnegative with $f_\eta(t) = f_\eta(-t)$ for all $t \in \mathbb{R}^d$.

Conditions on $\hat{\xi}$
For $W, Z \overset{iid}{\sim} \text{Unif}(\Omega)$ and for some $\xi, \tilde{\xi} \in \Xi$,

$$\hat{\xi}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} \xi; \quad \hat{\xi}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{\xi}$$

Most common statistics for H_N are of this form!

Finite Population CLT

In a completely randomized design, $\sqrt{N}(\hat{\tau} - \bar{\tau}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, V_{\tau\tau})$ with

$$V_{\tau\tau} = p^{-1}\Sigma_{y(1),\infty} + (1-p)^{-1}\Sigma_{y(0),\infty} - \Sigma_{\tau,\infty}.$$

Wu and Ding (2020)
Under H_N, the conditional distribution of $\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \mid \mathbf{Z}$ in a CRE converges weakly in probability to that of $\mathcal{N}\left(\mathbf{0}, \tilde{V}_{\tau\tau}\right)$ with

$$\tilde{V}_{\tau\tau} = (1-p)^{-1}\Sigma_{y(1),\infty} + p^{-1}\Sigma_{y(0),\infty}.$$

Finite Population CLT

In a completely randomized design, $\sqrt{N}(\hat{\tau} - \bar{\tau}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, V_{\tau\tau})$ with

$$V_{\tau\tau} = p^{-1}\Sigma_{y(1),\infty} + (1-p)^{-1}\Sigma_{y(0),\infty} - \Sigma_{\tau,\infty}.$$

Wu and Ding (2020)
Under H_N, the conditional distribution of $\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \mid \mathbf{Z}$ in a CRE converges weakly in probability to that of $\mathcal{N}\left(\mathbf{0}, \tilde{V}_{\tau\tau}\right)$ with

$$\tilde{V}_{\tau\tau} = (1-p)^{-1}\Sigma_{y(1),\infty} + p^{-1}\Sigma_{y(0),\infty}.$$

Generally, $\tilde{V}_{\tau\tau} \neq V_{\tau\tau}$, so the reference distribution does not align with the actual limiting distribution (unless H_F holds).

Conservative Covariance Estimators

We consider covariance estimators $\hat{V}$ such that

Conditions on $\hat{V}$
For $W, Z \overset{iid}{\sim} \text{Unif}(\Omega)$

$$\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} \bar{V} = V_{\tau\tau} + \Delta; \;\; \Delta \succeq 0$$

$$\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{V}_{\tau\tau}$$

Example: $\hat{V}_{Neyman}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) := N \left(\frac{\hat{\Sigma}_1(\mathbf{y}(\mathbf{Z}),\mathbf{Z})}{n_1} + \frac{\hat{\Sigma}_0(\mathbf{y}(\mathbf{Z}),\mathbf{Z})}{n_0} \right)$

Gaussian Prepivoting (Finally)

1. Given a base statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau})$.
2. Compute a conservative covariance estimate $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$.
3. Form the **prepivoted statistic**

$$G(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \gamma^{(d)}_{\mathbf{0},\hat{V}} \left\{ a : f_{\hat{\xi}}(a) \leq T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \right\}.$$

$\gamma^{(d)}_{\mathbf{0},\hat{V}}$ denotes the Gaussian measure centered at **0** with covariance $\hat{V}$.

Interpreting $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$

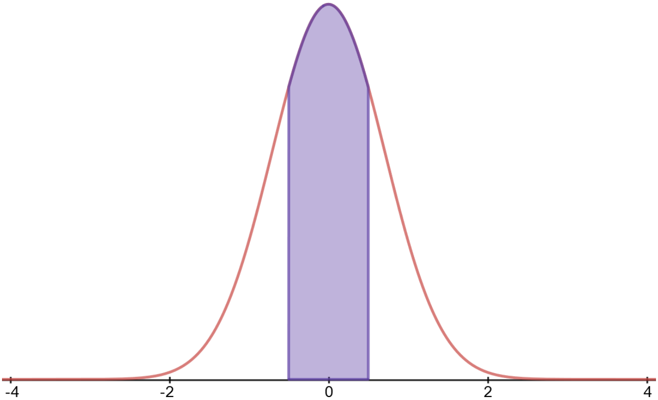
Say that $a \sim \mathcal{N}\left(\mathbf{0}, \hat{V}\right)$.

$G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is the probability that $f_{\hat{\xi}}(a)$ lies in $(-\infty, T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})]$.

Equivalently stated, $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is the measure of the set $(-\infty, T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})]$ under the $f_{\hat{\xi}}$-pushforward of the Gaussian measure $\gamma^{(d)}_{\mathbf{0},\hat{V}}$.

$G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is just the complement of the p-value for the large-sample test of H_N where you used $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ as the statistic.

Interpreting $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$



$$G(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \gamma^{(d)}_{\mathbf{0},\hat{V}} \left\{ a : f_{\hat{\xi}}(a) \leq T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \right\}.$$

Examples

- Absolute difference in means: $\sqrt{N}||\hat{\tau}||$.
- Studentized: $\left(\sqrt{N}\hat{\tau}\right)^T \hat{V}_{Neyman}^{-1} \left(\sqrt{N}\hat{\tau}\right)$
- **Incorrectly** Studentized: $\left(\sqrt{N}\hat{\tau}\right)^T \hat{V}_{pool}^{-1} \left(\sqrt{N}\hat{\tau}\right)$ where $\hat{V}_{Pool} = \left(\frac{N}{n_0} + \frac{N}{n_1}\right) \left(\frac{(n_1-1)\hat{\Sigma}_{y(1)}+(n_0-1)\hat{\Sigma}_{y(0)}}{n_1+n_0-2}\right)$.
- Max absolute t-stat: $\max_{1 \leq j \leq d} \frac{\sqrt{N}|\hat{\tau}_j|}{\sqrt{\hat{V}_{Neyman,jj}}}$

A Multivariate Simulation of Type I Error Rates

Set-up: Outcomes in $\mathbb{R}^{25}$, CRE ($n_1 = .2N$), $\alpha = 0.25$.

	Hotelling, Pooled			Max t-stat		
	No Pre.	Pre.	LS	No Pre.	Pre.	LS
Sharp, $N = 300$	0.251	0.249	0.365	0.254	0.252	0.300
Sharp, $N = 5000$	0.248	0.243	0.257	0.251	0.247	0.255
Weak, $N = 300$	0.996	0.361	0.433	0.321	0.071	0.082
Weak, $N = 5000$	0.990	0.064	0.067	0.308	0.060	0.064

"No Pre." = FRT without prepivoting.

"Pre." = FRT after prepivoting.

"LS" = Large Sample test valid for H_N (benchmark).

A Multivariate Simulation of Type I Error Rates

Set-up: Outcomes in $\mathbb{R}^{25}$, CRE ($n_1 = .2N$), $\alpha = 0.25$.

	Hotelling, Pooled			Max t-stat		
	No Pre.	Pre.	LS	No Pre.	Pre.	LS
Sharp, $N = 300$	0.251	0.249	0.365	0.254	0.252	0.300
Sharp, $N = 5000$	0.248	0.243	0.257	0.251	0.247	0.255
Weak, $N = 300$	0.996	0.361	0.433	0.321	0.071	0.082
Weak, $N = 5000$	0.990	0.064	0.067	0.308	0.060	0.064

"No Pre." = FRT without prepivoting.

"Pre." = FRT after prepivoting.

"LS" = Large Sample test valid for H_N (benchmark).

A Multivariate Simulation of Type I Error Rates

Set-up: Outcomes in $\mathbb{R}^{25}$, CRE ($n_1 = .2N$), $\alpha = 0.25$.

	Hotelling, Pooled			Max t-stat		
	No Pre.	Pre.	LS	No Pre.	Pre.	LS
Sharp, $N = 300$	0.251	0.249	0.365	0.254	0.252	0.300
Sharp, $N = 5000$	0.248	0.243	0.257	0.251	0.247	0.255
Weak, $N = 300$	0.996	0.361	0.433	0.321	0.071	0.082
Weak, $N = 5000$	0.990	0.064	0.067	0.308	0.060	0.064

"No Pre." = FRT without prepivoting.

"Pre." = FRT after prepivoting.

"LS" = Large Sample test valid for H_N (benchmark).

A Multivariate Simulation of Type I Error Rates

Set-up: Outcomes in $\mathbb{R}^{25}$, CRE ($n_1 = .2N$), $\alpha = 0.25$.

	Hotelling, Pooled			Max t-stat		
	No Pre.	Pre.	LS	No Pre.	Pre.	LS
Sharp, $N = 300$	0.251	0.249	0.365	0.254	0.252	0.300
Sharp, $N = 5000$	0.248	0.243	0.257	0.251	0.247	0.255
Weak, $N = 300$	0.996	0.361	0.433	0.321	0.071	0.082
Weak, $N = 5000$	0.990	0.064	0.067	0.308	0.060	0.064

"No Pre." = FRT without prepivoting.

"Pre." = FRT after prepivoting.

"LS" = Large Sample test valid for H_N (benchmark).

The Main Theorem

Theorem (C., Fogarty)
Suppose that the potential outcomes and covariates are sufficiently regular (e.g., limiting finite population means and covariances exist, bounded "fourth moment"). In a completely randomized experiment with $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau})$ *and* $\hat{V}$ *a conservative covariance estimator the prepivoted statistic* $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ *is asymptotically sharp dominant under* H_N*.*

The Main Theorem

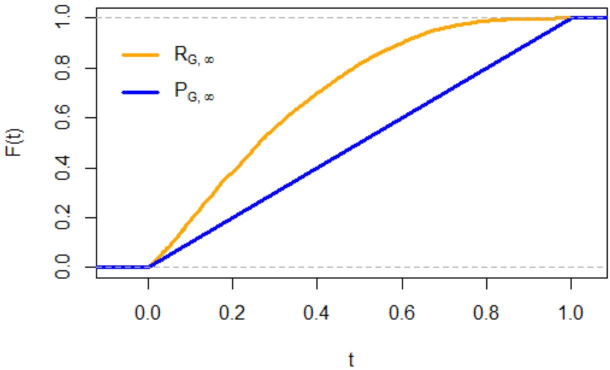
Theorem (C., Fogarty)
Suppose that the potential outcomes and covariates are sufficiently regular (e.g., limiting finite population means and covariances exist, bounded "fourth moment"). In a completely randomized experiment with $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau})$ *and* $\hat{V}$ *a conservative covariance estimator the prepivoted statistic* $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ *is asymptotically sharp dominant under* H_N.

Practical Implication: The FRT using $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is asymptotically conservative under H_N and is exact under H_F for all significance levels $\alpha \in (0, 1)$.

Unpacking Why This Works

To prove asymptotic sharp dominance, we need to understand:

- $\mathscr{P}_{G,\infty}$, the limit of the reference distributions $\hat{\mathscr{P}}_G$
  - Show that $\mathscr{P}_{G,\infty}$ is the uniform distribution on $(0, 1)$.
- $\mathscr{R}_{G,\infty}$, the limit of the true distributions $\mathscr{R}_G$.
  - Show that $\mathscr{R}_{G,\infty}$ is *dominated by* the uniform distribution on $(0, 1)$.



Analyzing $\mathscr{P}_{G,\infty}$

Simple concrete case:

- $f_\eta(\cdot) = ||\cdot||_2^2$ (the general case behaves similarly)
- Ignore covariates for now

Analyzing $\mathscr{P}_{G,\infty}$

Simple concrete case:

- $f_\eta(\cdot) = ||\cdot||_2^2$ (the general case behaves similarly)
- Ignore covariates for now

Under H_N, the conditional distribution of $\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \mid \mathbf{Z}$ in a CRE converges weakly in probability to that of $\mathcal{N}\left(\mathbf{0}, \tilde{V}_{\tau\tau}\right)$ with

$$\tilde{V}_{\tau\tau} = (1-p)^{-1}\Sigma_{y(1),\infty} + p^{-1}\Sigma_{y(0),\infty}.$$

Analyzing $\mathscr{P}_{G,\infty}$

Simple concrete case:

- $f_\eta(\cdot) = ||\cdot||_2^2$ (the general case behaves similarly)
- Ignore covariates for now

Under H_N, the conditional distribution of $\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \mid \mathbf{Z}$ in a CRE converges weakly in probability to that of $\mathcal{N}\left(\mathbf{0}, \tilde{V}_{\tau\tau}\right)$ with

$$\tilde{V}_{\tau\tau} = (1-p)^{-1}\Sigma_{y(1),\infty} + p^{-1}\Sigma_{y(0),\infty}.$$

$\implies f_{\hat{\xi}}(\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W})) \mid \mathbf{Z}$ has asymptotic behavior like $||A||_2^2$ with $A \sim \mathcal{N}\left(\mathbf{0}, \tilde{V}_{\tau\tau}\right)$.

Analyzing $\mathscr{P}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{V}_{\tau\tau}$

Analyzing $\mathscr{P}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{V}_{\tau\tau} \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S)$ for any Borel set S.

Analyzing $\mathscr{P}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{V}_{\tau\tau} \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S)$ for any Borel set S.

Define

$$S = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W})||_2^2\right\}$$

$$S_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||A||_2^2\right\}$$

Analyzing $\mathscr{P}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{V}_{\tau\tau} \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S)$ for any Borel set S.

Define

$$S = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W})||_2^2\right\}$$

$$S_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||A||_2^2\right\}$$

Asymptotically speaking, $G(\mathbf{y}(\mathbf{Z}), \mathbf{W}) = \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ given $\mathbf{Z}$ behaves like $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S_\infty)$.

Analyzing $\mathscr{P}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{W}) \xrightarrow{p} \tilde{V}_{\tau\tau} \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S)$ for any Borel set S.

Define

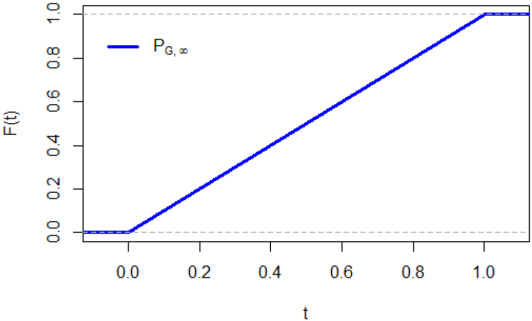
$$S = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{W})||_2^2\right\}$$
$$S_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||A||_2^2\right\}$$

Asymptotically speaking, $G(\mathbf{y}(\mathbf{Z}), \mathbf{W}) = \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ given $\mathbf{Z}$ behaves like $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S_\infty)$.

Punchline: $\gamma^{(d)}_{\mathbf{0},\tilde{V}_{\tau\tau}}(S_\infty)$ is a fancy way of writing the probability integral transform ($F_X(X)$ is uniform for continuous X!)

Theorem (C., Fogarty)
Under H_N and mild regularity conditions. In a completely randomized experiment with $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau})$ and $\hat{V}$ a conservative covariance estimator the reference distribution of $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ limits to the uniform distribution on $[0, 1]$, i.e.,

$$\mathscr{P}_G(t) \xrightarrow{p} t \quad \forall\, t \in [0, 1].$$



Under H_N, $\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}),\mathbf{Z})$ in a CRE converges in distribution to $\mathcal{N}\left(\mathbf{0}, V_{\tau\tau}\right)$ with

$$V_{\tau\tau} = p^{-1}\Sigma_{y(1),\infty} + (1-p)^{-1}\Sigma_{y(0),\infty} - \Sigma_{\tau,\infty}.$$

Analyzing $\mathscr{R}_{G,\infty}$

Under H_N, $\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ in a CRE converges in distribution to $\mathcal{N}(\mathbf{0}, V_{\tau\tau})$ with

$$V_{\tau\tau} = p^{-1}\Sigma_{y(1),\infty} + (1-p)^{-1}\Sigma_{y(0),\infty} - \Sigma_{\tau,\infty}.$$

$\implies f_{\hat{\xi}}(\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}))$ has asymptotic behavior like $||B||_2^2$ with $B \sim \mathcal{N}(\mathbf{0}, V_{\tau\tau})$.

Analyzing $\mathscr{R}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} V_{\tau\tau} + \Delta$

Analyzing $\mathscr{R}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} V_{\tau\tau} + \Delta \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},V_{\tau\tau}+\Delta}(S)$ for any Borel set S.

Analyzing $\mathscr{R}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} V_{\tau\tau} + \Delta \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},V_{\tau\tau}+\Delta}(S)$ for any Borel set S.

Define

$$\mathcal{S} = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})||_2^2\right\}$$

$$\mathcal{S}_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||B||_2^2\right\}$$

Analyzing $\mathscr{R}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} V_{\tau\tau} + \Delta \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},V_{\tau\tau}+\Delta}(S)$ for any Borel set S.

Define

$$\mathcal{S} = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})||_2^2\right\}$$
$$\mathcal{S}_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||B||_2^2\right\}$$

Asymptotically speaking, $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \gamma^{(d)}_{\mathbf{0},\hat{V}}(\mathcal{S})$ behaves like $\gamma^{(d)}_{\mathbf{0},V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$.

Analyzing $\mathscr{R}_{G,\infty}$

Recall $\hat{V}(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \xrightarrow{p} V_{\tau\tau} + \Delta \implies \gamma^{(d)}_{\mathbf{0},\hat{V}}(S)$ limits to $\gamma^{(d)}_{\mathbf{0},V_{\tau\tau}+\Delta}(S)$ for any Borel set S.

Define

$$\mathcal{S} = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq ||\sqrt{N}\hat{\tau}(\mathbf{y}(\mathbf{Z}), \mathbf{Z})||_2^2\right\}$$

$$\mathcal{S}_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq \underbrace{||B||_2^2}_{V_{\tau\tau}}\right\}$$

Asymptotically speaking, $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \gamma^{(d)}_{\mathbf{0},\hat{V}}(\mathcal{S})$ behaves like $\gamma^{(d)}_{\mathbf{0},V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$.

Punchline: Similar to the probability integral transform, but notice the **conservative covariance**!

Analyzing $\mathscr{R}_{G,\infty}$

$$\mathcal{S}_\infty = \left\{\mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq \underbrace{||B||_2^2}_{V_{\tau\tau}}\right\} \quad \text{versus} \quad \gamma^{(d)}_{\mathbf{0}, V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$$

If we had matched covariances, then we would be back to the probability integral transform (so $\mathscr{R}_{G,\infty}$ would be the standard uniform distribution).

Analyzing $\mathscr{R}_{G,\infty}$

$$\mathcal{S}_\infty = \left\{ \mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq \underbrace{||B||_2^2}_{V_{\tau\tau}} \right\} \quad \text{versus} \quad \gamma^{(d)}_{\mathbf{0}, V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$$

If we had matched covariances, then we would be back to the probability integral transform (so $\mathscr{R}_{G,\infty}$ would be the standard uniform distribution).

By using a overestimate of covariance we overestimate the probability of S_∞ with $\gamma^{(d)}_{\mathbf{0}, V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$

Analyzing $\mathscr{R}_{G,\infty}$

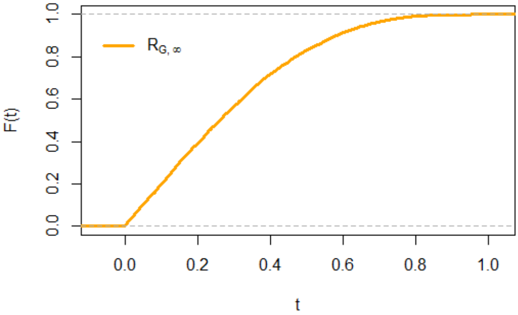
$$\mathcal{S}_\infty = \left\{ \mathbf{a} \in \mathbb{R}^d \text{ s.t. } ||\mathbf{a}||_2^2 \leq \underbrace{||B||_2^2}_{V_{\tau\tau}} \right\} \quad \text{versus} \quad \gamma^{(d)}_{\mathbf{0}, V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$$

If we had matched covariances, then we would be back to the probability integral transform (so $\mathscr{R}_{G,\infty}$ would be the standard uniform distribution).

By using a overestimate of covariance we overestimate the probability of S_∞ with $\gamma^{(d)}_{\mathbf{0}, V_{\tau\tau}+\Delta}(\mathcal{S}_\infty)$

$\implies \mathscr{R}_{G,\infty}(t) \geq t \quad \forall\, t \in [0, 1]$. (Anderson's Theorem - 1955)

Theorem (C., Fogarty)
Under H_N and mild regularity conditions. In a completely randomized experiment with $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = f_{\hat{\xi}}(\sqrt{N}\hat{\tau})$ and $\hat{V}$ a conservative covariance estimator the true distribution of $G(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is asymptotically dominated by the uniform distribution on $[0, 1]$, i.e., $\mathscr{R}_{G,\infty}(t) \geq t \quad \forall\, t \in [0, 1]$.



Generalizations

Generalizations

- Easy to include covariates:
  - Rerandomized designs: Instead of performing a CRE, randomly select treatment allocation which **preserves covariate balance**.
  - Asymptotically linear statistics (so regression adjustment can be included).
- Multiple treatment arms: Gaussian prepivoting can be applied for experiments with any number $A \in \mathbb{N}$ of treatment arms ($A \geq 2$).
- Extension to bootstrapping methods!

https://arxiv.org/abs/2002.06654

Questions?

The Fisher Randomization Test i

Suppose that H_F holds, then $\mathbf{y}(\mathbf{Z}) = \mathbf{y}(0) = \mathbf{y}(1)$ no matter what value $\mathbf{Z}$ takes. For inference, we need the cumulative distribution function of the test statistic $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$.

$$
\begin{aligned}
\mathscr{R}_T(t) &= \mathbb{P}\left(T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \leq t\right) \\
&= \sum_{\mathbf{w} \in \Omega} \mathbb{P}\left(\mathbf{Z} = \mathbf{w}\right) \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \leq t\right\} \\
&= \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T(\mathbf{y}(\mathbf{w}), \mathbf{w}) \leq t\right\} \\
&= \frac{1}{|\Omega|} \sum_{\mathbf{w} \in \Omega} \mathbb{1}\left\{T(\mathbf{y}(\mathbf{Z}), \mathbf{w}) \leq t\right\} = \mathscr{P}_T(t).
\end{aligned}
$$

The Fisher Randomization Test ii

We want to reject the null when $T(\mathbf{y}(\mathbf{Z}), \mathbf{Z})$ is larger than some critical threshold:

- Under H_F we don't want to improperly reject the null with probability greater than α,
- We want the threshold to be as low as possible so that we have good detection power.

So c_α is determined by $\inf \left\{ c \in \mathbb{R} \text{ s.t. } \mathbb{P}\left(T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \geq c \right) \leq \alpha \right\}$.

This is exactly solved by taking $c_\alpha = \hat{\mathscr{P}}^{-1}(1 - \alpha)$.

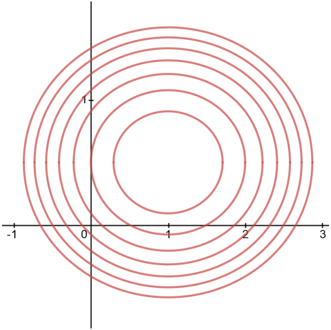
Go back.

Quasi-Convexity

A function $f : \mathbb{R}^d \to \mathbb{R}$ is quasi-convex if $\forall\, x, y \in \mathbb{R}^d$

$$f(\lambda x + (1-\lambda)y) \le \max\{f(x), f(y)\} \quad \forall\, \lambda \in [0,1].$$

Equivalently, the *sublevel-sets* of f must be convex.



Go back.

Anderson's Theorem

Anderson's Theorem (1995) implies:

Theorem (Tong, Thm 4.2.5)
For non-degenerate $X \sim \mathcal{N}(\mathbf{0}, A)$ *and* $Y \sim \mathcal{N}(\mathbf{0}, B)$ *with* $A \succeq B$

$$\mathbb{P}(Y \in S) \geq \mathbb{P}(X \in S)$$

for all measurable convex S *that are mirror-symmetric about the origin.*

This is the multivariate generalization of saying "The variance of univariate centered Gaussians controlls their concentration near the origin." Go back.

A ***balance criterion*** $\phi : \mathbb{R}^k \mapsto \{0,1\}$ is an indicator function such that the set $M = \{\mathbf{b} : \phi(\mathbf{b}) = 1\}$ is closed, convex, mirror-symmetric about the origin (i.e. $\mathbf{b} \in M \Leftrightarrow -\mathbf{b} \in M$) with non-empty interior.

Then define the prepivoted statistic as

$$G(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) = \frac{\gamma^{(d+k)}_{\mathbf{0},\hat{V}} \left\{(\mathbf{a}, \mathbf{b})^T : f_{\hat{\xi}}(\mathbf{a}) \leq T(\mathbf{y}(\mathbf{Z}), \mathbf{Z}) \ \wedge \ \phi(\mathbf{b}) = 1\right\}}{\gamma^{(k)}_{\mathbf{0},\hat{V}_{\delta\delta}} \{\mathbf{b} : \phi(\mathbf{b}) = 1\}}.$$