# An Adaptive Multivariate Point Null Test

Adam Elder

June $27^{\text{th}}$ 2021

Joint work with Alex Luedtke and Marco Carone
University of Washington

# Motivation

In this project, we consider data from the HVTN 505 clinical trial, a phase IIB preventative HIV vaccine efficacy trial[1].

- The primary analysis for the trial did not find the treatment to be effective.
- Secondary analyses studied the relationships between the vaccine immune response and risk of infection.

[1] Neidich et al. 2019.

Thank you to:

- The HVTN 505 study participants
- The HVTN 505 study team
- Brian Williamson, Fred Hutch
- Peter Gilbert, Fred Hutch
- Youyi Fong, Fred Hutch

## Motivation

One question from the secondary analysis was if there is any association between vaccine immune response and HIV infection.

- The vaccine immune response was measured using a large number of biomarkers, including antibodies, T cells, and $F_{c\gamma}$ receptors.
- The null hypothesis for this question is that none biomarkers are associated with infection.

# Notation

Let $O = (Y, X_1, \ldots, X_d)$ and let $O_1, O_2, \ldots, O_n$ be drawn independently from a common unknown distribution $P_0$ in the statistical model $\mathcal{M}$.

We are interested in whether the outcome $Y$ is associated with any of the covariates $X_1, \ldots, X_d$.

- Let the measure of association (for example correlation) between $Y$ and $X_i$ be denoted by $\psi_i$.
- Let $\hat{\psi}_i$ be the corresponding estimate of $\psi_i$ based on $O_1, O_2, \ldots, O_n$.
- We wish to test the multivariate point null:

$$H_0 : \psi_1 = \psi_2 = \ldots = \psi_d = 0$$

# Previous Work

Previous work in this area can be broken down into two categories

- Construct a test for each $\psi_i$, and correct for multiple testing. That is, define

$$H_{0i} : \psi_i = 0 \text{ v.s. } H_{1i} : \psi_i \neq 0,$$

  and generate p-values for each test that correct for multiple hypothesis testing. Then if any of the $d$ hypothesis tests rejects, reject the multivariate null.
- Directly test the multivariate point null.

# Downsides of multiple hypothesis testing

While it is possible to test a multivariate point null using multiple hypothesis testing procedures, there can be a loss of power when compared with methods that directly test the multivariate point null.

- Multiple hypothesis testing methods allow users to know which hypothesis was rejected.
- This knowledge can come at the cost of power when testing a multivariate point null.

# Previous work in multivariate point null tests

Some recent multivariate point null testing methods provide more power than other methods, but are often difficult to generalize to other parameters and data-generating mechanisms[2]. Here we propose a general-purpose testing procedure that aims to

- provide more power than the easily applicable multiple hypothesis correction methods, and
- is applicable for a wide variety of data-generating mechanisms and parameters of interest.

[2] Donoho and Jin 2004; McKeague and Qian 2015; Pan et al. 2014; Xu et al. 2016.

# Creating a test

Creating a test can usually be broken down into three steps:

- Picking a test statistic
- Determining the (limiting) distribution of the test statistic under $H_0$
- Determining which values of the test statistic would be extreme under $H_0$.

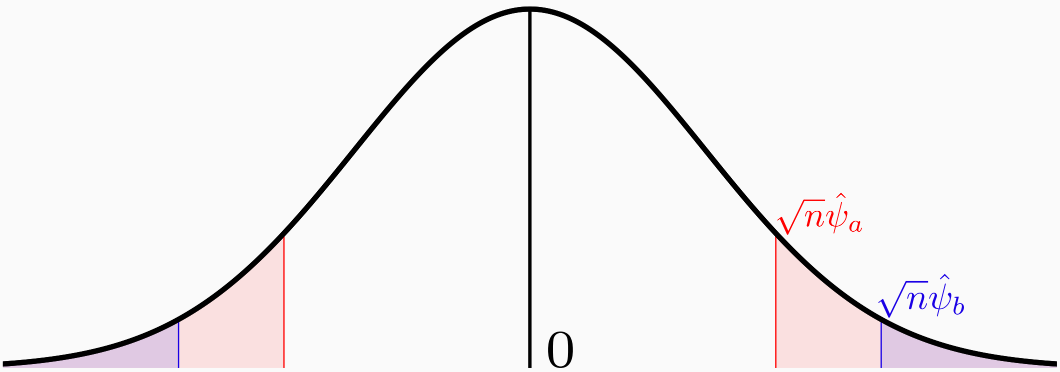
# Understanding the vector of parameter estimates

Define $\psi = (\psi_1, \ldots, \psi_d)$ and $\hat{\psi} = \left(\hat{\psi}_1, \ldots, \hat{\psi}_d\right)$

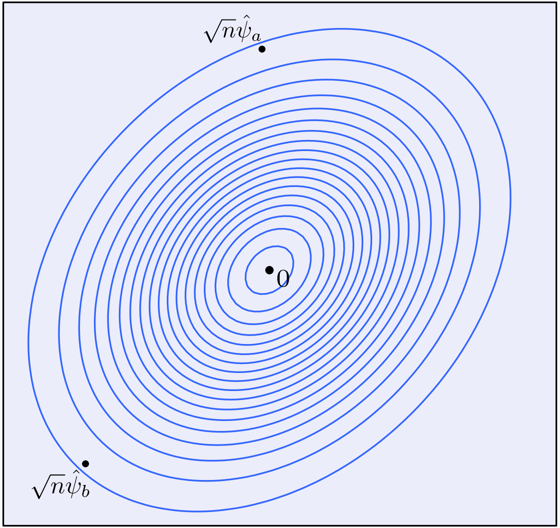
Suppose that for each $P \in \mathcal{M}$,

$$\sqrt{n}\left(\hat{\psi} - \psi\right) \xrightarrow{d} Z \sim N\left(0, \Sigma_P\right)$$

# Considering a test in $\mathbb{R}$



## Considering a test in $\mathbb{R}^2$



# Transformation to create a test statistic

Defining which observations are more extreme in higher dimensions is difficult. We will consider a simple example in which the test statistic is the $\ell_2$ or Euclidean norm of the vector of parameter estimators.

$$\hat{t} = \sqrt{\sum_{i=1}^{d} \hat{\psi}_i^2}$$

Next we consider the limiting distribution of $\sqrt{n}\hat{t}$ under $H_0$.

# Obtaining the limiting distribution under the null

Under the null:

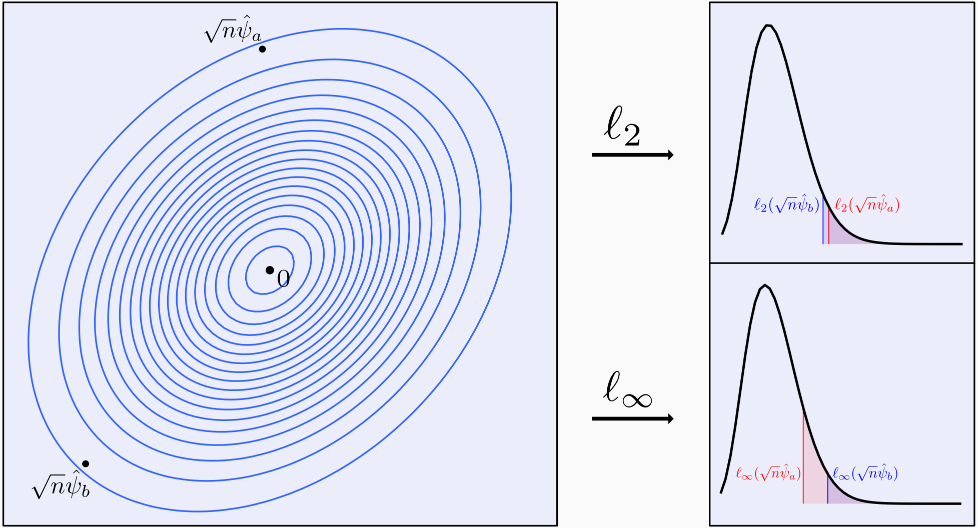
$$\sqrt{n}\hat{\psi} \xrightarrow{d} Z \sim N\left(0, \Sigma_P\right).$$

The continuous mapping theorem tells us that under the null

$$\|\sqrt{n}\hat{\psi}\|_2 \xrightarrow{d} \|Z\|_2, \quad Z \sim N\left(0, \Sigma_P\right).$$

While knowing the exact distribution of $\|Z\|_2$ is difficult, we may sample from $\|Z\|_2$ to carry out a test.

# Comparing between norms



# Choosing a norm

While it may be possible to know a priori which norm will perform optimally, this will not generally be the case.

We have developed a method that adaptively chooses a norm that optimized the estimated local power while controlling type-1 error. This method requires measuring the “performance” of each norm for any potential alternative $\psi$.

This idea was inspired in part by the work of Ian McKeague and a commentary on this work from Yichi Zhang and Eric Laber[3].

[3] McKeague and Qian 2015; Zhang and Laber 2015.

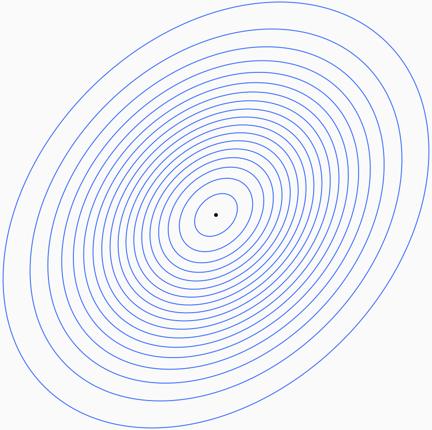
# Defining a performance metric

Here, we denote by $\Gamma$ a performance metric that maps from $\mathbb{R}^d \times \mathbb{M}_{d\times d}$ into $\mathbb{R}$ where $\mathbb{M}_{d\times d}$ contains all positive definite $d \times d$ matrices.

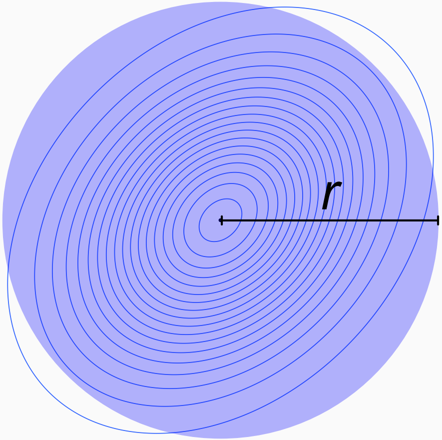
An example of such a performance metric is the acceptance rate performance metric:

$$\Gamma_{\|\cdot\|}(\omega, \Sigma) = \Pr(\|Z + \omega\| \leq c_\alpha(\Sigma)) \text{ where}$$
$$c_\alpha(\Sigma) \equiv \inf\{c : \Pr(\|Z\| < c) \geq \alpha\} \text{ and } Z \sim N(0, \Sigma).$$

# Estimating local acceptance rate



# Estimating local acceptance rate



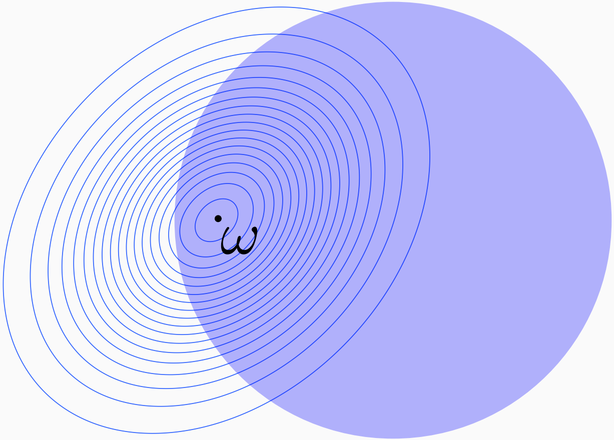
## Estimating local acceptance rate

Thinking back to the definition of our performance metric:

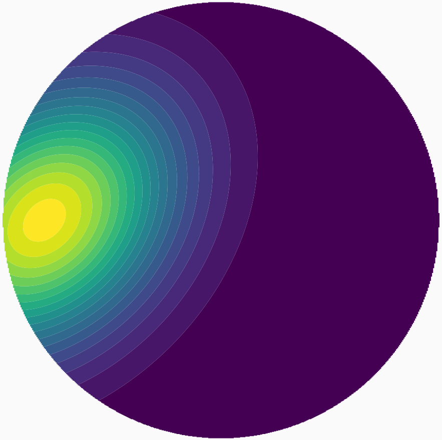
$$\Gamma_{\|\cdot\|}(\omega, \Sigma) = \Pr(\|Z + \omega\| \leq c_\alpha(\Sigma)) \text{ where}$$
$$c_\alpha(\Sigma) \equiv \inf\{c : \Pr(\|Z\| < c) \geq \alpha\} \text{ and } Z \sim N(0, \Sigma),$$

Note that $c_\alpha(\Sigma) = r$.

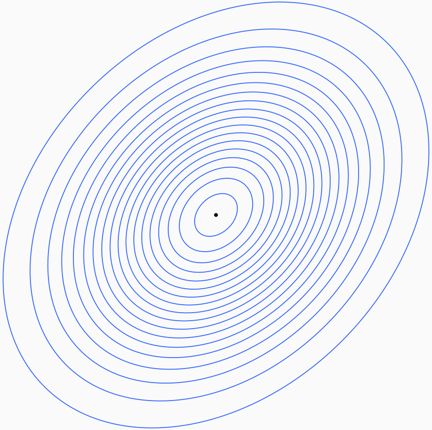
# Estimating local acceptance rate



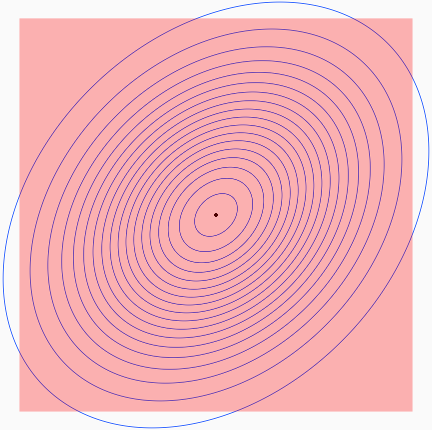
# Estimating local acceptance rate



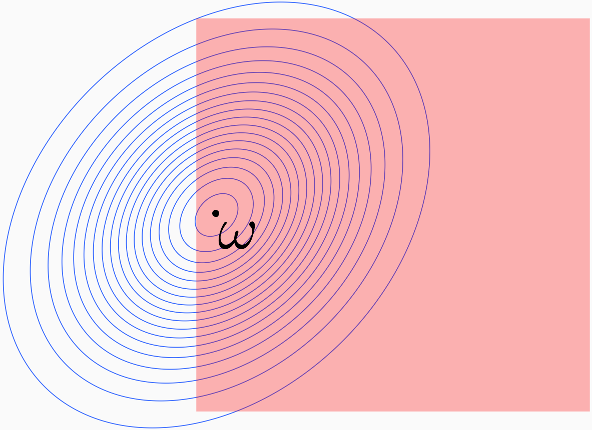
# Estimating local acceptance rate



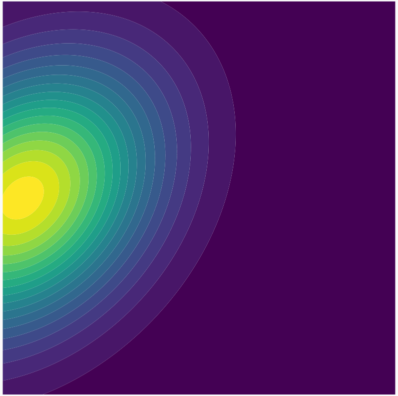
# Estimating local acceptance rate



# Estimating local acceptance rate



# Estimating local acceptance rate



# Defining an adaptive test statistic

Now, that we have a method for determining which norm will provide better power at a given alternative, we wish to define an adaptive test statistic in order to have an adaptive test.

A first thought for such a test statistic could be:

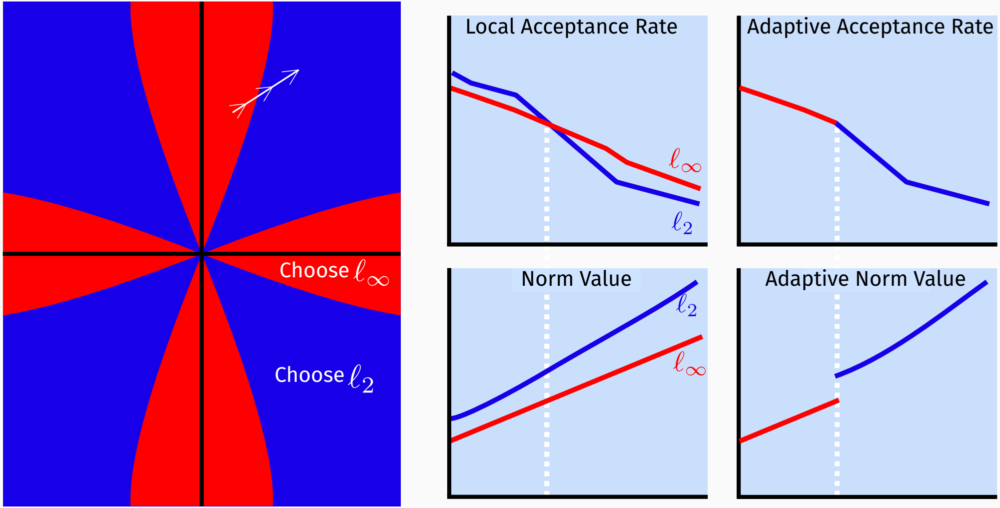
$$\left\|\sqrt{n}\hat{\psi}\right\|_{k_n} \text{ where } k_n = \operatorname{argmin}\left\{\Gamma_{\|\cdot\|_k}(\sqrt{n}\hat{\psi}, \hat{\Sigma}) : k \in \{1, 2, \dots\}\right\}.$$

Unfortunately, using this test statistic has some important downsides, including the discontinuity of:

$$\|\psi\|_{k(\psi)} \text{ where } k(\psi) = \operatorname{argmin}\left\{\Gamma_{\|\cdot\|_k}(\psi, \Sigma) : k \in \{1, 2, \dots\}\right\},$$

with respect to $\omega$.

# Defining an adaptive test statistic

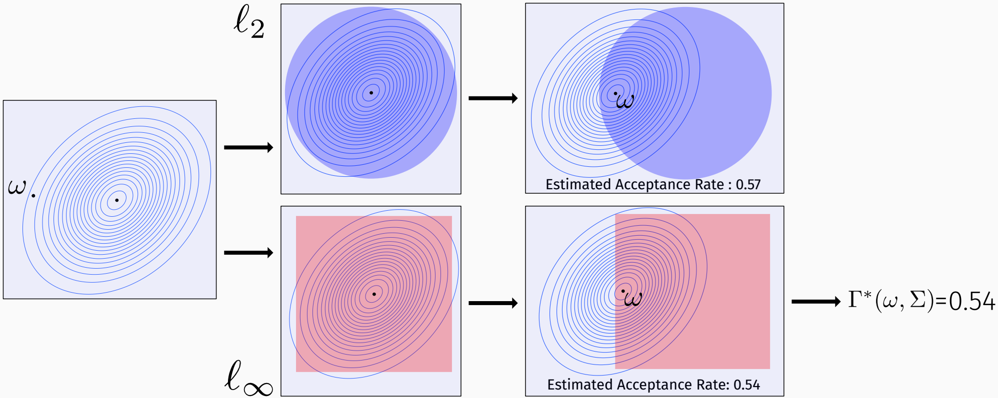


# Defining an adaptive test statistic

For a set of potential norms $\|\cdot\|_1, \|\cdot\|_2, \ldots, \|\cdot\|_k$, the adaptive performance metric is

$$\Gamma^*(\omega, \Sigma) = \min\left(\Gamma_{\|\cdot\|_1}(\omega, \Sigma), \Gamma_{\|\cdot\|_2}(\omega, \Sigma), \ldots, \Gamma_{\|\cdot\|_k}(\omega, \Sigma)\right).$$

# An adaptive performance metric



# What has changed

While both the use of a performance metric and the adaptive selection of a norm add complexity to our method, the underlying testing procedure remains largely unchanged.

# Taking draws from the limiting distribution

As the test has been described, p-values are calculated by taking draws $\underline{z}_1, \ldots, \underline{z}_B$ from the normal limiting distribution $Z$, then computing

$$\frac{1}{B}\sum_{i=1}^{B} I\left\{\Gamma^*(\sqrt{n}\hat{\psi}, \Sigma) > \Gamma^*(z_i, \Sigma)\right\}$$

In practice, we will not know $\Sigma$, and will instead use a consistent estimator $\hat{\Sigma}$ of $\Sigma$ and take draws from $\hat{Z} \sim N(0, \hat{\Sigma})$.

## Estimating local acceptance rate

Our framework allows for the construction of tests with useful theoretical guarantees for most data generating mechanism and most parameters of interest.

- Asymptotic Type-1 error control
- Power approaching 1 under any fixed alternative
- Power that is asymptotically greater than $\alpha$ under all local alternatives

# Numerical Examples

While not shown here, we have found via numerical examples that tests created using our framework have similar power to other modern methods in settings where they exist.

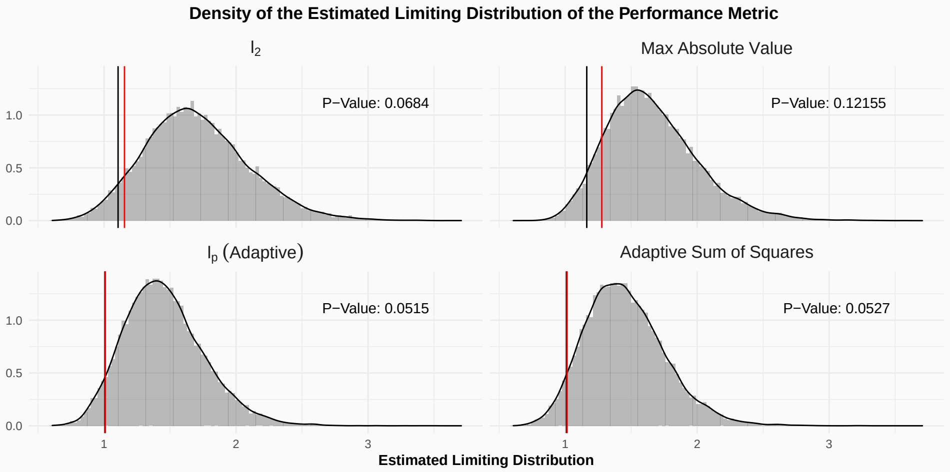
# Application of method

We applied the testing framework to data from the HVTN 505 clinical trial, a phase IIB preventative HIV vaccine efficacy trial[4].

- One set of data measured the immune response among 25 cases and 125 randomly sampled frequency-matched vaccine controls.
- We conducted a test for a biomarker set and test if any biomarkers in this set are associated with HIV infection.

[4] Neidich et al. 2019.

# Application of method



## Thank you!

# Thank you for listening!

# References i

## References


Donoho, David and Jiashun Jin (June 2004). “Higher criticism for detecting sparse heterogeneous mixtures”. en. In: *The Annals of Statistics* 32.3, pp. 962–994. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/009053604000000265. URL: https://projecteuclid.org/euclid.aos/1085408492 (visited on 04/18/2019).

Liu, Yaowu and Jun Xie (Jan. 2, 2020). “Cauchy Combination Test: A Powerful Test With Analytic p-Value Calculation Under Arbitrary Dependency Structures”. In: *Journal of the American Statistical Association* 115.529. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2018.1554485, pp. 393–402. ISSN: 0162-1459. DOI: 10.1080/01621459.2018.1554485. URL: https://doi.org/10.1080/01621459.2018.1554485 (visited on 05/10/2021).

McKeague, Ian W. and Min Qian (Oct. 2015). “An Adaptive Resampling Test for Detecting the Presence of Significant Predictors”. In: *Journal of the American Statistical Association* 110.512, pp. 1422–1433. ISSN: 0162-1459. DOI: 10.1080/01621459.2015.1095099. URL: https://amstat.tandfonline.com/doi/abs/10.1080/01621459.2015.1095099 (visited on 04/18/2019).

Neidich, Scott D. et al. (Nov. 2019). “Antibody Fc effector functions and IgG3 associate with decreased HIV-1 risk”. en. In: *The Journal of Clinical Investigation* 129.11, pp. 4838–4849. ISSN: 0021-9738. DOI: 10.1172/JCI126391. URL: https://www.jci.org/articles/view/126391 (visited on 08/04/2020).

# References ii


Pan, Wei et al. (Aug. 2014). "A Powerful and Adaptive Association Test for Rare Variants". en. In: *Genetics* 197.4, pp. 1081–1095. ISSN: 0016-6731, 1943-2631. DOI: 10.1534/genetics.114.165035. URL: https://www.genetics.org/content/197/4/1081 (visited on 07/12/2019).

Xu, Gongjun et al. (Sept. 2016). "An adaptive two-sample test for high-dimensional means". en. In: *Biometrika* 103.3, pp. 609–624. ISSN: 0006-3444. DOI: 10.1093/biomet/asw029. URL: https://academic.oup.com/biomet/article/103/3/609/1744173 (visited on 07/12/2019).

Zhang, Yichi and Eric B. Laber (Oct. 2015). "Comment". In: *Journal of the American Statistical Association* 110.512, pp. 1451–1454. ISSN: 0162-1459. DOI: 10.1080/01621459.2015.1106403. URL: https://amstat.tandfonline.com/doi/full/10.1080/01621459.2015.1106403 (visited on 04/18/2019).

# Conditions on $\Gamma$

1. The performance metric $\Gamma$ is continuous and non-negative on $\mathbb{R}^d \times \mathbb{M}_{d \times d}$ where $\mathbb{M}_{d \times d}$ contains all positive definite $d \times d$ matrices.

   This means the performance metric is smooth with respect to both the vector and matrix arguments.

2. $\Pr(\Gamma(Z, \Sigma) = t) = 0$ where $Z \sim N(0, \Sigma)$ for every $t$ and positive definite $\Sigma$.

   The random variable defined by $\Gamma(Z, \Sigma)$ has a continuous cumulative distribution function.

# Conditions on $\Gamma$

3. $\Gamma(\sqrt{n}\hat{\psi}, \widehat{\Sigma}) \xrightarrow{p} 0$ under sampling from any fixed alternative.

   Under fixed alternatives, as $n$ gets larger $\sqrt{n}\hat{\psi}$ will move further away from the origin. This condition requires that as this happens, the performance metric converges in probability to zero.

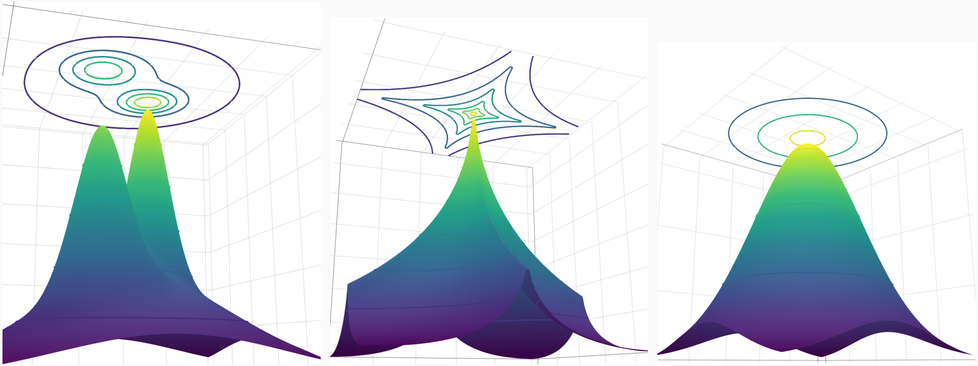
# Conditions on $\Gamma$

4. $\Gamma(\cdot, \Sigma)$ is quasi-concave for every positive definite $\Sigma$ (for every $k$, the set $\{\omega : \Gamma(\omega, \Sigma) \geq k\}$ is convex).

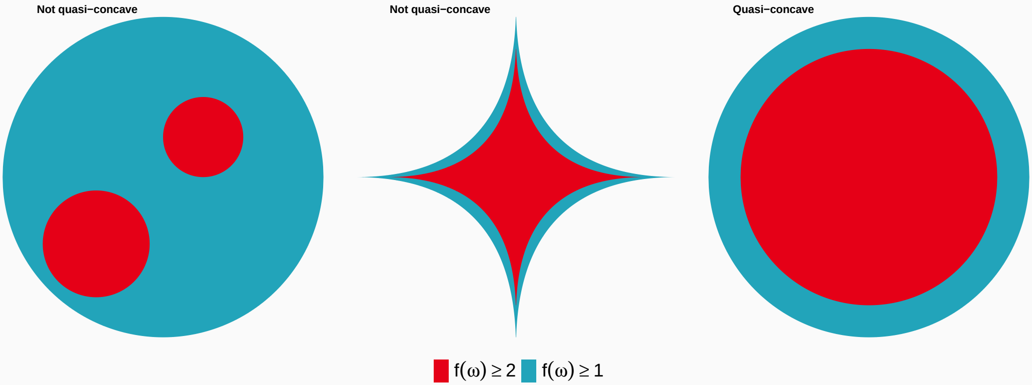
   We will cover this item in the next slides.

5. $\Gamma(\cdot, \Sigma)$ is centrally symmetric for every positive definite $\Sigma$, i.e., $\Gamma(-\omega, \Sigma) = \Gamma(\omega, \Sigma)$.

# Visualizing quasi-concavity



# Visualizing quasi-concavity



# Examples

We now turn to the examples in which we studied our test.

For these examples, we consider two different sets of norms. The first set of norms is the $\ell_p$ norm we have considered up to this point:

$$\ell_p(\omega) = \sqrt[p]{\sum_{i=1}^{d} |\omega_i|^p}.$$

The other norm considered is referred to the sum of squares norm and is defined by:

$$\jmath_k(\omega) = \sqrt{\sum_{i=1}^{k} \omega^2_{(d-i+1)}}$$

where $\omega^2_{(1)}, \ldots, \omega^2_{(d)}$ are the order statistics of $\omega^2_1, \ldots, \omega^2_d$.

# Example 1: Correlation

In our first and simplest example, we consider the data unit $X = (W_1, W_2, \ldots, W_d, Y)$, where $W_1, W_2, \ldots, W_d$ represent real-valued covariates and $Y$ is some outcome of interest, and take the parameter of interest

$$\psi_j := corr(W_j, Y)$$

to be the marginal correlation between $W_j$ and $Y$.

We compare our method to two other methods[5].

[5] Zhang and Laber 2015; Liu and Xie 2020.

# Example 1: Correlation

Data are generated from a linear model:

$$Y = W_1\beta_1 + \ldots + W_d\beta_d + \varepsilon,$$

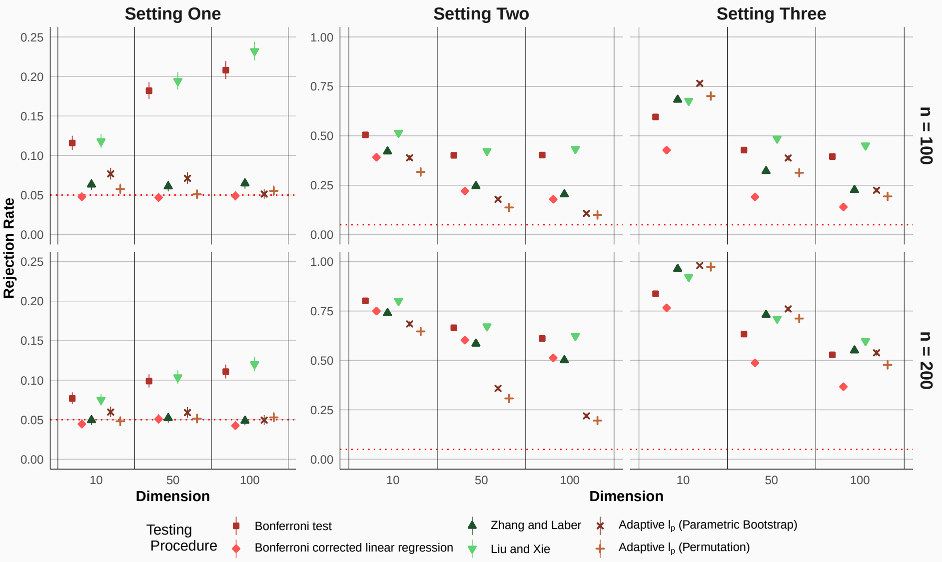
where

- $\varepsilon$ is independent of $(W_1, \ldots, W_d)$
- the between $W$ correlation is 0, 0.5, or 0.8 for all $W$.

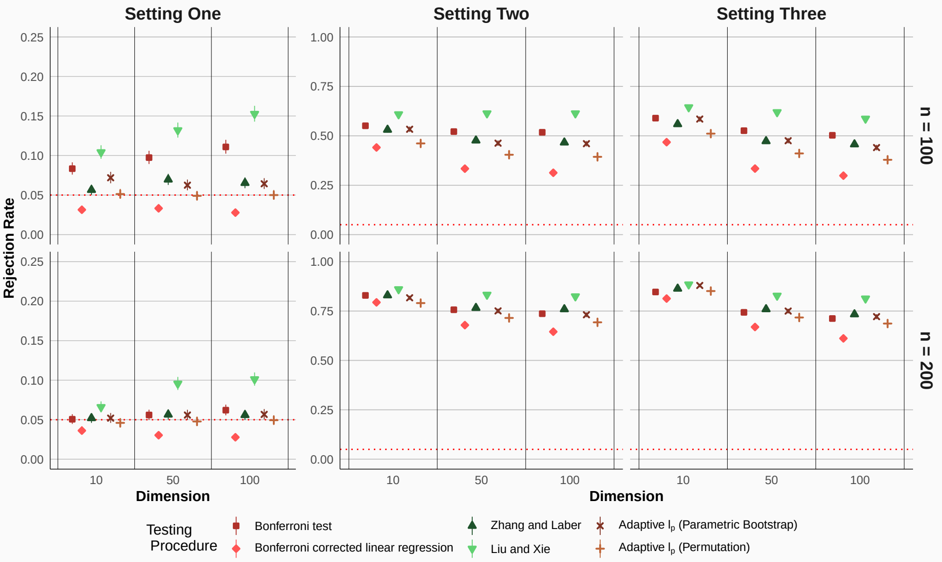
We consider three different settings defined by different sets of $\beta$ values.

- In setting one (the null setting), all $\beta_i = 0$
- In setting two, $\beta_1 = 1/4$ and all other $\beta_i = 0$
- In setting three, $\beta_1 \ldots \beta_5 = 0.15$, $\beta_6 \ldots \beta_{10} = -0.1$, and all other $\beta_i = 0$

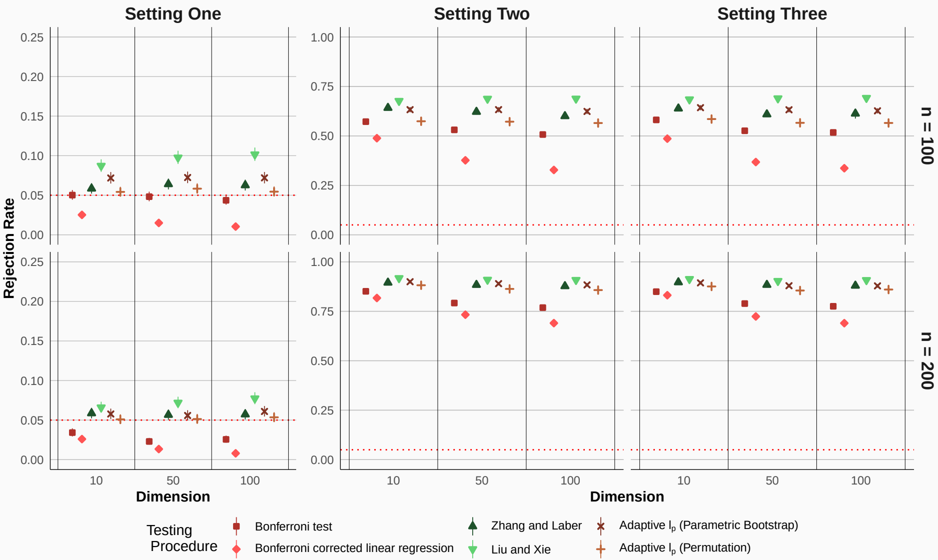
# Example 1 (Between X correlation is 0)



# Example 1 (Between W correlation is 0.5)



# Example 1 (Between X correlation is 0.8)



# Example 2: Working log-linear regression model under missingness

In our second example, the data unit is $(W_1, W_2, \ldots, W_d, U, \Delta)$, where

- $W_1, W_2, \ldots, W_d$ represent real-valued covariates,
- $\Delta$ is an indicator that the binary outcome $Y$ is observed, and
- $U := \Delta Y$ equals $Y$ if $\Delta = 1$ and is set to zero otherwise.

In other words, this data unit is similar to that defined in the first example, but with potential missingness of the outcome value in some observations.

## Example 2: Working log-linear regression model under missingness

For the second example, outcome data are simulated using a binomial model defined by

$$\log(\Pr(Y = 1|W)) = \beta_0 + W_1\beta_1 + \ldots + W_d\beta_d.$$

In all settings, the probability of missingness is given by

$$\text{logit}(\Pr(\Delta = 1|W)) = -0.25 + 1W_{d-1} - 1.5W_d,$$

and when $\Delta = 0$, $Y$ is missing.

# Example 2: Working log-linear regression model under missingness

We focus here on coefficients of the working log-linear regression model

$$\log\left[\Pr(Y = 1 \mid W_1 = w_1, \ldots, W_d = w_d)\right] = \alpha + \beta_1 w_1 + \ldots + \beta_d w_d \ .$$

Assuming that $Y$ is missing at random given $W$, that is, that $Y$ and $\Delta$ are independent conditionally upon $W$, the parameter

$$\psi_j := \frac{cov\{W_j, \log E[\,\Pr(Y = 1 \mid \Delta = 1, W) \mid W_j\,]\}}{var(W_j)} \tag{1}$$

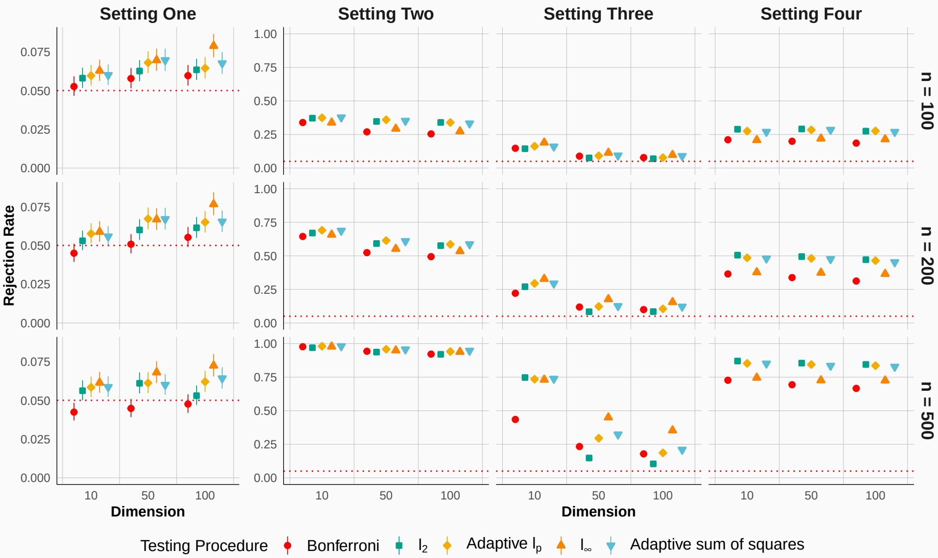
identifies the coefficient associated to $W_j$ in this working model, and simplifies to $\beta_j$ when the working log-linear model holds.

# Example 2: Working log-linear regression model under missingness

In each setting, the vector of covariates $W = (W_1, \ldots W_d)$ is drawn from a multivariate normal with mean zero and covariance matrix $\Sigma$, where $\Sigma_{i,j} = 1$ for $i = j$ and $\Sigma_{i,j} = 0.5$ for $i \neq j$ in all four settings.

- In the first (null) setting, $\beta_1, \ldots, \beta_d = 0$.
- In the second setting, $\beta_1 = 0.6$ and $\beta_2, \ldots, \beta_d = 0$.
- In the third setting, $\beta_1, \ldots, \beta_5 = 0.32$, $\beta_6, \ldots, \beta_{10} = -0.32$ and $\beta_{11}, \ldots, \beta_d = 0$.
- In the last setting, $\beta_1, \ldots, \beta_5 = 0.03375$, $\beta_6, \ldots, \beta_{10} = 0.0675$ and $\beta_{11}, \ldots, \beta_d = 0$.

# Example 2: Working log-linear regression model under missingness



# Coefficients of a working effect modification model for randomized trials

We wish to consider the interaction coefficient of the least-squares projection of the true conditional counterfactual success probability onto the logit-linear regression model

$$\operatorname{logit} \mathrm{pr}(Y(a) = 1 \mid W_j = w) = \alpha_j + \beta_j a + \gamma_j w + \delta_j wa \ .$$

This coefficient provides a measure of the degree to which $W_j$ modifies the effect of $A$ on $Y$. For simplicity, we assume that treatment allocation is randomized, so that the counterfactual outcome $Y(a)$ corresponding to treatment level $a$ is independent of $A$ for each $a \in \{0, 1\}$. In this case, the parameter

$$\Psi_j(P) := \operatorname*{argmin}_{\delta} \min_{(\alpha,\beta,\gamma)} E_P \left\{ \operatorname{logit} E_P[\, P(Y = 1 \mid W, A) \mid W_j \,] - (\alpha + \beta A + \gamma W_j + \delta W_j A) \right\}^2$$

identifies the interaction coefficient in this working model, and again, simplifies to $\delta_j$ when the working logit-linear structural model above.

## Coefficients of a working effect modification model for randomized trials

In all settings for this example, the working model described in section is used to take draws from $Y$ and the vector of covariates $W$ is draw from a multivariate normal with mean zero and covariance matrix $\Sigma$ where $\Sigma_{i,j} = 1$ for $i = j$ and $\Sigma_{i,j} = 0.5$ for $i \neq j$.

The random variable $A$ is drawn from a binomial distribution independently of $W$.

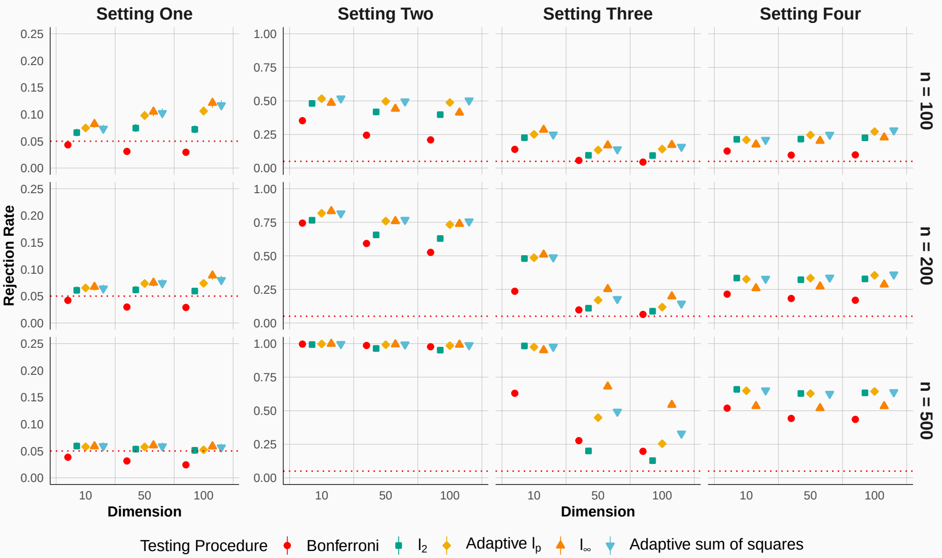
For all three settings,

$$\operatorname{logit}\left(Pr(Y=1|w,a)\right)=\alpha a+\sum_{i=1}^{d}\beta_i w_i+\sum_{j=1}^{d}\gamma_j w_j a.$$

Additionally, $\alpha=0.2$, $\beta_1,\ldots,\beta_{d/2}=2/\sqrt{d}$, and $\beta_{d/2},\ldots,\beta_d=0$ in each setting.

- In setting one: $\gamma_1,\ldots,\gamma_d=0$.
- In setting two: $\gamma_1=3$ and $\gamma_2,\ldots,\gamma_d=0$.
- In setting three: $\gamma_1,\ldots,\gamma_5=0.6$, $\gamma_6,\ldots,\gamma_{10}=-0.6$, and $\gamma_{11},\ldots,\gamma_d=0$.
- In setting four: $\gamma_1,\ldots,\gamma_5=0.09$, $\gamma_6,\ldots,\gamma_{10}=0.18$, and $\gamma_{11},\ldots,\gamma_d=0$.

# Example 3



Setting One
Setting Two
Setting Three
Setting Four
n = 100
n = 200
n = 500
Rejection Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.00
0.25
0.50
0.75
1.00
10
50
100
Dimension
Testing Procedure
Bonferroni
$l_2$
Adaptive $l_p$
$l_\infty$
Adaptive sum of squares

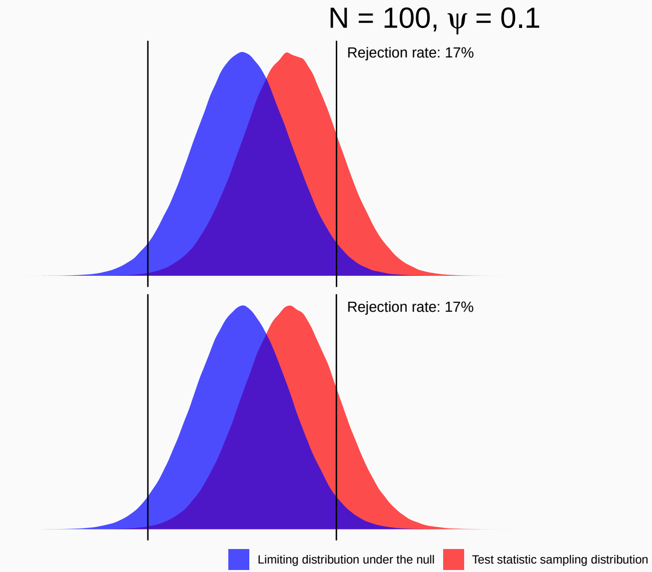
# Considering two tests

Consider two different tests. While both tests have an estimate of $\psi$ that is consistent, the second test has a standard error that shrinks towards zero at a rate slower than $n^{1/2}$.

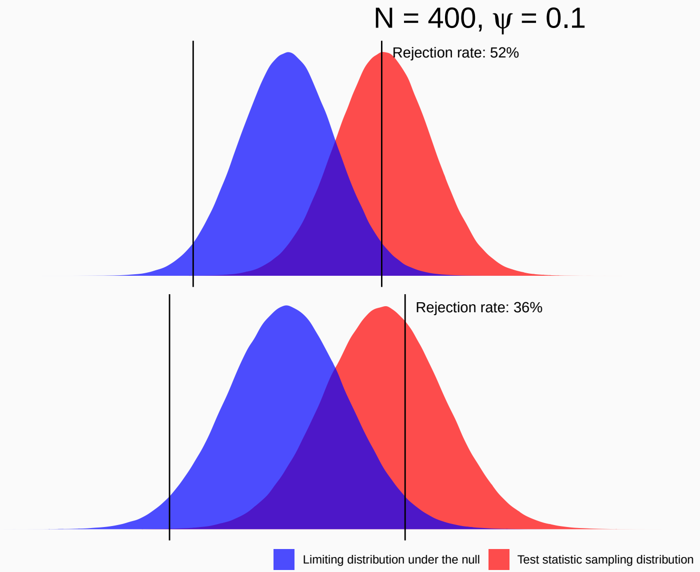
- It is still possible for both tests to achieve type-1 error control and consistency.
- This can happen as long as the standard error of the estimator does not shrink too slowly.

In the following slides, we compare the sampling distribution of test statistic $\sqrt{n}\hat{\psi}$ to the estimated limiting distribution for $\sqrt{n}\hat{\psi}$ under the null.

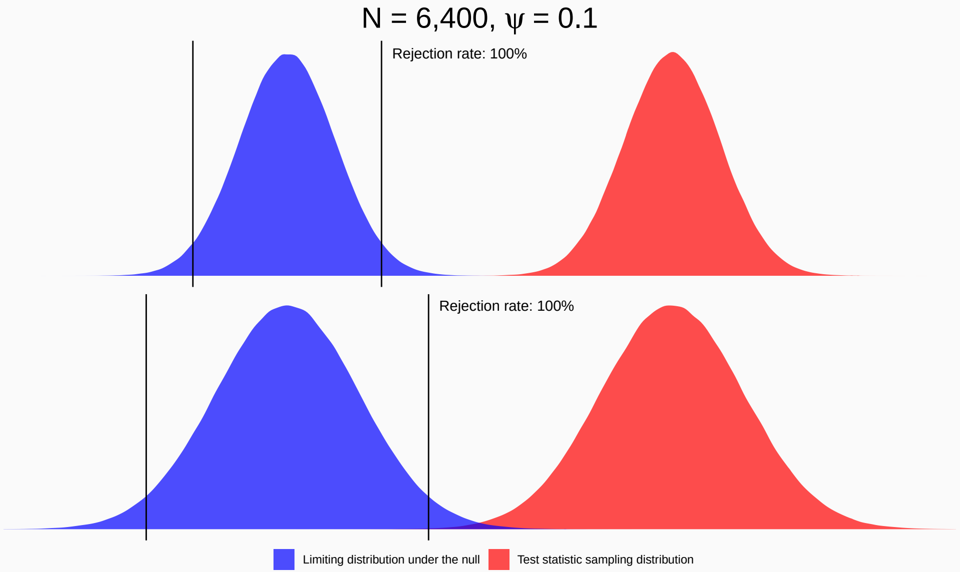
# Considering two tests



# Considering two tests



# Considering two tests

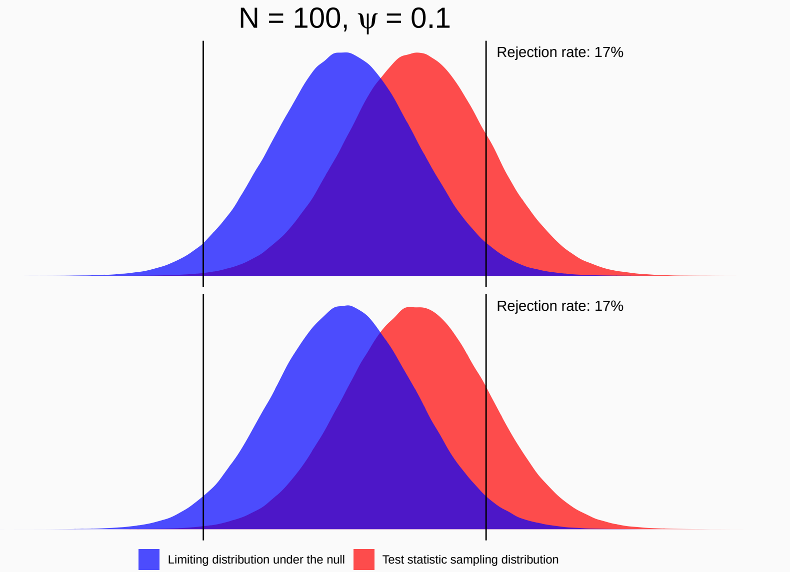


## Considering two tests

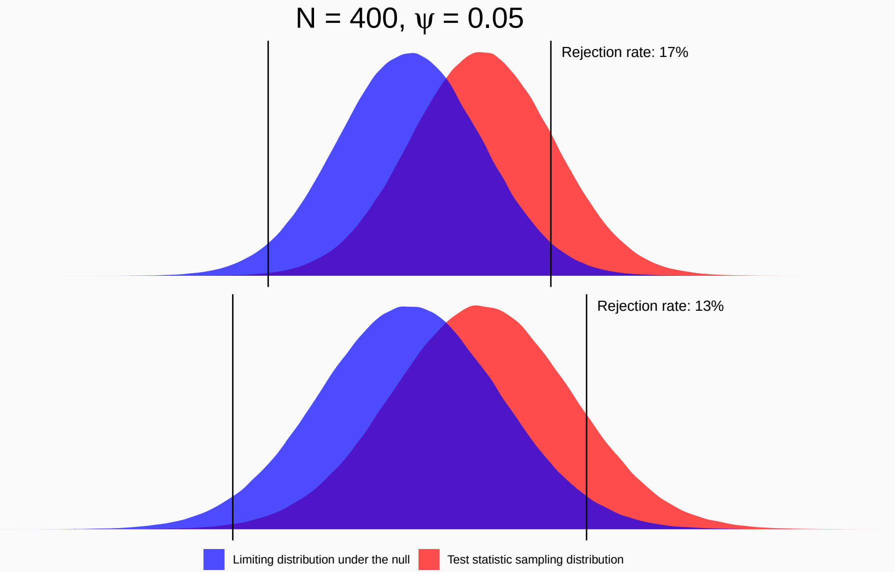
We now consider the above tests under a sequence of local alternatives.

- Under this sequence of local alternatives as sample size grows, the true value of $\psi$ shrinks towards the null ($\psi = 0$ at a rate of $1/sqrtn$).
- Under this sequence of alternatives, a test based on an inefficient estimator will have a rejection rate approaching $\alpha$.

# Considering local alternatives



# Considering local alternatives



# Considering local alternatives