Consistency of invariance-based randomization tests

Edgar Dobriban
University of Pennsylvania

July 2, 2022

Many thanks

- Conference organizing committe

Many thanks

- Conference organizing committe
- Award committe

Many thanks

- Conference organizing committe
- Award committe
- Audience

Many thanks

- Conference organizing committe
- Award committe
- Audience
- And congratulations to all other awardees!

My own involvement with ICSA

- I am very grateful to be involved in ICSA

My own involvement with ICSA

- I am very grateful to be involved in ICSA
- Have already benefited from this:

My own involvement with ICSA

- I am very grateful to be involved in ICSA
- Have already benefited from this:
  - Got invited to give talk at ICSA 2019 International Conference in Hangzhou (had to withdraw last minute); gave virtual presentation at ICSA 2021 Applied Statistics Symposium (thanks to Tracy Ke).

My own involvement with ICSA

- I am very grateful to be involved in ICSA
- Have already benefited from this:
  - Got invited to give talk at ICSA 2019 International Conference in Hangzhou (had to withdraw last minute); gave virtual presentation at ICSA 2021 Applied Statistics Symposium (thanks to Tracy Ke).
  - Many of my colleagues, mentors, and even some students are ICSA members.

My own involvement with ICSA

- I am very grateful to be involved in ICSA
- Have already benefited from this:
  - Got invited to give talk at ICSA 2019 International Conference in Hangzhou (had to withdraw last minute); gave virtual presentation at ICSA 2021 Applied Statistics Symposium (thanks to Tracy Ke).
  - Many of my colleagues, mentors, and even some students are ICSA members.
- Look forward to more involvement

The paper I am presenting

CONSISTENCY OF INVARIANCE-BASED RANDOMIZATION TESTS

BY EDGAR DOBRIBAN[1]

[1]*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, dobriban@wharton.upenn.edu*

Invariance-based randomization tests—such as permutation tests, rotation tests, or sign changes—are an important and widely used class of statistical methods. They allow drawing inferences under weak assumptions on the data distribution. Most work focuses on their type I error control properties, while their consistency properties are much less understood.

We develop a general framework and a set of results on the consistency of invariance-based randomization tests in signal-plus-noise models. Our framework is grounded in the deep mathematical area of representation theory. We allow the transforms to be general compact topological groups, such as rotation groups, acting by general linear group representations. We study test statistics with a generalized sub-additivity property.

We apply our framework to a number of fundamental and highly important problems in statistics, including sparse vector detection, testing for low-rank matrices in noise, sparse detection in linear regression, and two-sample testing. Comparing with minimax lower bounds we develop, we find perhaps surprisingly that in some cases, randomization tests detect signals at the minimax optimal rate.

Brief summary

- Randomization tests (e.g., permutation tests) are a widely used and important class of statistical methods

Brief summary

- Randomization tests (e.g., permutation tests) are a widely used and important class of statistical methods
- Popular: Require weak assumptions, broadly applicable

Brief summary

- Randomization tests (e.g., permutation tests) are a widely used and important class of statistical methods
- Popular: Require weak assumptions, broadly applicable
- Most work focuses studies them under the null hypothesis (type I error, false positive rate)

Brief summary

- Randomization tests (e.g., permutation tests) are a widely used and important class of statistical methods
- Popular: Require weak assumptions, broadly applicable
- Most work focuses studies them under the null hypothesis (type I error, false positive rate)
- This work develops a general framework to study them under the alternative hypothesis (consistency, power)

Randomization tests

- Data X

Randomization tests

- Data X
  - For instance, $X = (X_1, \ldots, X_n)$, where each $X_i \sim P$ iid.

Randomization tests

- Data X
  - For instance, $X = (X_1, \ldots, X_n)$, where each $X_i \sim P$ iid.
- Under null hypothesis, X has a distribution that is unchanged under a set of transformations $\mathcal{G}$, acting as $X \to gX$ for $g \in \mathcal{G}$

Randomization tests

- Data X
  - For instance, $X = (X_1, \ldots, X_n)$, where each $X_i \sim P$ iid.
- Under null hypothesis, X has a distribution that is unchanged under a set of transformations $\mathcal{G}$, acting as $X \to gX$ for $g \in \mathcal{G}$
  - For instance, when $n = 3$, (X_1, X_2, X_3) has the same distribution as permutation (X_2, X_1, X_3); because all are iid; order of observations does not matter.

Randomization tests

- Data X
  - For instance, $X = (X_1, \ldots, X_n)$, where each $X_i \sim P$ iid.
- Under null hypothesis, X has a distribution that is unchanged under a set of transformations $\mathcal{G}$, acting as $X \to gX$ for $g \in \mathcal{G}$
  - For instance, when $n = 3$, (X_1, X_2, X_3) has the same distribution as permutation (X_2, X_1, X_3); because all are iid; order of observations does not matter.
- Randomization test: compare $T(X)$ with values of $T(gX)$, where $g \in \mathcal{G}$, reject null if $T(X)$ is large enough

Randomization tests

- Data X
  - For instance, $X = (X_1, \ldots, X_n)$, where each $X_i \sim P$ iid.
- Under null hypothesis, X has a distribution that is unchanged under a set of transformations $\mathcal{G}$, acting as $X \to gX$ for $g \in \mathcal{G}$
  - For instance, when $n = 3$, (X_1, X_2, X_3) has the same distribution as permutation (X_2, X_1, X_3); because all are iid; order of observations does not matter.
- Randomization test: compare $T(X)$ with values of $T(gX)$, where $g \in \mathcal{G}$, reject null if $T(X)$ is large enough
  - For instance, compare $T(X_1, \ldots, X_n)$ with $T(X_{\pi(1)}, \ldots, X_{\pi(n)})$ for permutations π, reject if $T(X)$ large enough—Permutation test

Randomization tests, continued

- Set of transformations $\mathcal{G}$ may be very large, so in practice often sample random transforms $G_1, \ldots, G_K$ uniformly from $\mathcal{G}$

Randomization tests, continued

- Set of transformations $\mathcal{G}$ may be very large, so in practice often sample random transforms $G_1, \ldots, G_K$ uniformly from $\mathcal{G}$
  - For instance, there are $n!$ permutations; so sample random permutations in permutation testing

Randomization tests, continued

- Set of transformations $\mathcal{G}$ may be very large, so in practice often sample random transforms $G_1, \ldots, G_K$ uniformly from $\mathcal{G}$
  - For instance, there are $n!$ permutations; so sample random permutations in permutation testing
- How large should $T(X)$ be? Large enough to ensure that this happens with probability at most $\alpha \implies$ larger than $1-\alpha$-th quantile of $T(X), T(G_1X), \ldots, T(G_KX)$

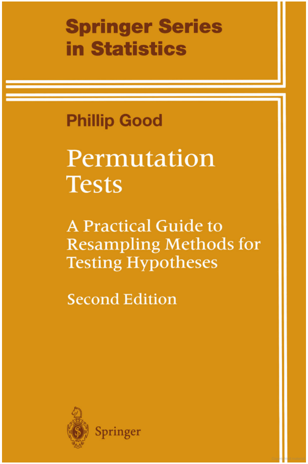
Randomization tests, continued

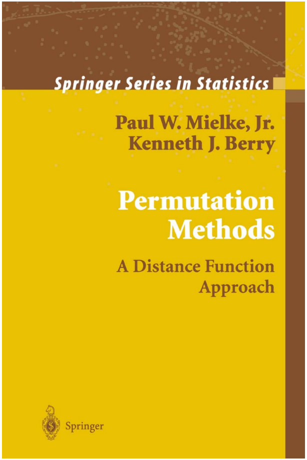
- Set of transformations $\mathcal{G}$ may be very large, so in practice often sample random transforms $G_1, \ldots, G_K$ uniformly from $\mathcal{G}$
  - For instance, there are $n!$ permutations; so sample random permutations in permutation testing
- How large should $T(X)$ be? Large enough to ensure that this happens with probability at most $\alpha \implies$ larger than $1 - \alpha$-th quantile of $T(X), T(G_1X), \ldots, T(G_KX)$
  - For instance, $T(X) > \max(T(G_1X), \ldots, T(G_{19}X))$ happens on average at most 1 out of 20 times (each of the 20 equally likely to be the maximum), so $\alpha = 0.05$ ok.

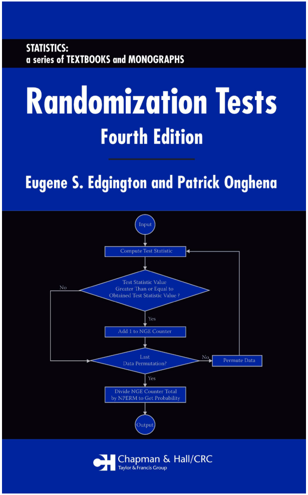
Flowchart

$$
\begin{tikzcd}
& \text{Test statistic } T(X) \arrow[ddr] & \\
& \text{Test statistic after transform } T(G_1X) \arrow[dr] & \\
\text{Data } X \arrow[uur] \arrow[ur] \arrow[ddr] & & T(X) > q_{1-\alpha}\left(T(X), T(G_1X), \ldots, T(G_KX)\right) \\
& \vdots & \\
& \text{Test statistic after transform } T(G_KX) \arrow[uur] &
\end{tikzcd}
$$

Books on permutation and randomization tests







Randomization tests are important: Neuroscience

Proceedings of the National Academy of Sciences of the United States of America

> Current Issue > vol. 113 no. 28 > Anders Eklund, 7900–7905, doi: 10.1073/pnas.1602413113

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

Randomization tests are important: Genomics

Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

Aravind Subramanian[a,b], Pablo Tamayo[a,b], Vamsi K. Mootha[a,c], Sayan Mukherjee[d], Benjamin L. Ebert[a,e], Michael A. Gillette[a,f], Amanda Paulovich[g], Scott L. Pomeroy[h], Todd R. Golub[a,e], Eric S. Lander[a,c,i,j,k], and Jill P. Mesirov[a,k]

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]
  - Other cases [Pesarin, Salmaso, Kim, Balakrishnan, Wasserman]

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]
  - Other cases [Pesarin, Salmaso, Kim, Balakrishnan, Wasserman]
- Limitations: need special structure of parameter space (e.g., contiguity), tests (e.g., establish limit distribution), or transforms (e.g., permutations)

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]
  - Other cases [Pesarin, Salmaso, Kim, Balakrishnan, Wasserman]
- Limitations: need special structure of parameter space (e.g., contiguity), tests (e.g., establish limit distribution), or transforms (e.g., permutations)
- Our contribution: significantly broaden these conditions; using a new approach.

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]
  - Other cases [Pesarin, Salmaso, Kim, Balakrishnan, Wasserman]
- Limitations: need special structure of parameter space (e.g., contiguity), tests (e.g., establish limit distribution), or transforms (e.g., permutations)
- Our contribution: significantly broaden these conditions; using a new approach.
  - Enables studying high-dimensional problems; not accessible before

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]
  - Other cases [Pesarin, Salmaso, Kim, Balakrishnan, Wasserman]
- Limitations: need special structure of parameter space (e.g., contiguity), tests (e.g., establish limit distribution), or transforms (e.g., permutations)
- Our contribution: significantly broaden these conditions; using a new approach.
  - Enables studying high-dimensional problems; not accessible before
  - In some cases, prove matching minimax lower bounds

Null vs alternative: No signal vs signal

- Under the null hypothesis, test controls type I error at level α
- How about under the alternative hypothesis? When does it have power?
  - Uniformly most powerful tests [Lehmann and Stein, 1949]
  - Tests whose limiting distributions can be controlled [Hoeffding, 1952]
  - Special cases, local alternatives [Bickel, van Zwet, 1970's]
  - Other cases [Pesarin, Salmaso, Kim, Balakrishnan, Wasserman]
- Limitations: need special structure of parameter space (e.g., contiguity), tests (e.g., establish limit distribution), or transforms (e.g., permutations)
- Our contribution: significantly broaden these conditions; using a new approach.
  - Enables studying high-dimensional problems; not accessible before
  - In some cases, prove matching minimax lower bounds
  - Main surprise: Randomization tests work almost as well as parametric tests, with much fewer assumptions

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$
    - Under null $S = 0$

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$
  - Under null $S = 0$
  - Under alternative, $S \in \Theta$, for some parameter space Θ

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$
  - Under null $S = 0$
  - Under alternative, $S \in \Theta$, for some parameter space Θ
  - Distribution of noise N unchanged under transforms $\mathcal{G}$

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$
  - Under null $S = 0$
  - Under alternative, $S \in \Theta$, for some parameter space Θ
  - Distribution of noise N unchanged under transforms $\mathcal{G}$
- Examples:
  - Location models (One-sample and two-sample testing)

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$
  - Under null $S = 0$
  - Under alternative, $S \in \Theta$, for some parameter space Θ
  - Distribution of noise N unchanged under transforms $\mathcal{G}$
- Examples:
  - Location models (One-sample and two-sample testing)
  - (Sparse) signal/ low-rank matrix detection

Our approach 1: model

- Consider a signal-plus-noise model $X = S + N$
  - Under null $S = 0$
  - Under alternative, $S \in \Theta$, for some parameter space Θ
  - Distribution of noise N unchanged under transforms $\mathcal{G}$
- Examples:
  - Location models (One-sample and two-sample testing)
  - (Sparse) signal/ low-rank matrix detection
  - (Sparse) linear regression: $S = X\beta$

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"
- We assume that for some $\psi > 0$, T is ψ-subadditive:

$$\psi \cdot T(a+b) \leqslant T(a) + T(b).$$

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"
- We assume that for some $\psi > 0$, T is ψ-subadditive:

$$\psi \cdot T(a+b) \leqslant T(a) + T(b).$$

- Examples:
  - Linear functions

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"
- We assume that for some $\psi > 0$, T is ψ-subadditive:

$$\psi \cdot T(a+b) \leqslant T(a) + T(b).$$

- Examples:
  - Linear functions
  - Norms

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"
- We assume that for some $\psi > 0$, T is ψ-subadditive:

$$\psi \cdot T(a+b) \leqslant T(a) + T(b).$$

- Examples:
  - Linear functions
  - Norms
  - Convex functions of polynomial growth

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"
- We assume that for some $\psi > 0$, T is ψ-subadditive:

$$\psi \cdot T(a+b) \leqslant T(a) + T(b).$$

- Examples:
  - Linear functions
  - Norms
  - Convex functions of polynomial growth
  - Given subadditive functions: Finite convex combinations, maxima

Our approach 2: test statistic

- Want $T(X)$ to be larger under alternative, so "$T(S+N) \gg T(N)$"
- We assume that for some $\psi > 0$, T is ψ-subadditive:

$$\psi \cdot T(a+b) \leqslant T(a) + T(b).$$

- Examples:
  - Linear functions
  - Norms
  - Convex functions of polynomial growth
  - Given subadditive functions: Finite convex combinations, maxima
- Non-examples: functions of fast growth, e.g., $x \mapsto \exp(x)$

Main result

Theorem (informal). Consider sequence of signal-plus-noise problems with above conditions. Assume there is $t > 0$, such that the following hold:

1. **Strong signal.** For any signal s, there is $\tilde{t}(s) > 0$ such that

$$T(s) > \psi^{-2}\tilde{t}(s) + \psi^{-1}(\psi^{-1} + 1)t. \tag{1}$$

Main result

Theorem (informal). Consider sequence of signal-plus-noise problems with above conditions. Assume there is $t > 0$, such that the following hold:

1. **Strong signal.** For any signal s, there is $\tilde{t}(s) > 0$ such that

$$T(s) > \psi^{-2}\tilde{t}(s) + \psi^{-1}(\psi^{-1} + 1)t. \tag{1}$$

2. **Noise control.** $P(T(N) \leqslant t) \to 1$ and $P(T(-N) \leqslant t) \to 1$.

Main result

Theorem (informal). Consider sequence of signal-plus-noise problems with above conditions. Assume there is $t > 0$, such that the following hold:

1. **Strong signal.** For any signal s, there is $\tilde{t}(s) > 0$ such that
$$T(s) > \psi^{-2}\tilde{t}(s) + \psi^{-1}(\psi^{-1} + 1)t. \tag{1}$$
2. **Noise control.** $P(T(N) \leqslant t) \to 1$ and $P(T(-N) \leqslant t) \to 1$.
3. **Bound on randomized statistic**.
$$P_G\left(T(Gs) \leqslant \tilde{t}(s)\right) \to 1.$$

The randomization test is consistent, i.e.,

$$P_{G_1,\ldots,G_K,X}\left(T(X) > q_{1-\alpha}\left[T(X), T(G_1X), \ldots, T(G_KX)\right]\right) \to 1.$$

Discussion of main result

- If signal is sufficiently strong, and randomized signal vanishes, randomization test works.

Discussion of main result

- If signal is sufficiently strong, and randomized signal vanishes, randomization test works.
- Proof approach: deterministic separating sequence (between $T(X)$ and randomized $T(GX)$)

Discussion of main result

- If signal is sufficiently strong, and randomized signal vanishes, randomization test works.
- Proof approach: deterministic separating sequence (between $T(X)$ and randomized $T(GX)$)
- Examples
  - Detecting sparse vectors
  - Detecting low-rank matrices
  - Sparse detection in linear regression

Follow-up work inspired by ours

FASTER EXACT PERMUTATION TESTING: USING A REPRESENTATIVE SUBGROUP

BY NICK W. KONING[1], JESSE HEMERIK[2]

[1]*Econometric Institute, Erasmus University Rotterdam, n.w.koning@ese.eur.nl*

[2]*Biometris, Wageningen University & Research, jesse.hemerik@wur.nl*

Abstract

Non-parametric tests based on permutation, rotation or sign-flipping are examples of so-called group-invariance tests. These tests rely on invariance of the null distribution under a set of transformations that has a group structure, in the algebraic sense. Such groups are often huge, which makes it computationally infeasible to use the entire group. Hence, it is standard practice to test using a randomly sampled set of transformations from the group. This random sample still needs to be substantial to obtain good power and replicability. We improve upon the standard practice by using a well-designed subgroup of transformations instead of a random sample. We show this can yield a more powerful and fully replicable test with the same number of transformations. For a normal location model and a particular design of the subgroup, we show that the power improvement is equivalent to the power difference between a Monte Carlo Z-test and Monte Carlo t-test. In our simulations, we find that we can obtain the same power as a test based on sampling with just half the number of transformations, or equivalently, more power for the same computation time. These benefits come entirely 'for free', as our methodology relies on an assumption of invariance under the subgroup, which is implied by invariance under the entire group.

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \ldots, n$

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \ldots, n$
- Interested to detect "sparse" vectors s; having large ℓ_∞ norm $\|s\|_\infty$.

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \ldots, n$
- Interested to detect "sparse" vectors s; having large ℓ_∞ norm $\|s\|_\infty$.
- Use the test statistic $f(X) = \|\sum_{i=1}^n X_i\|_\infty$.

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \ldots, n$
- Interested to detect "sparse" vectors s; having large ℓ_∞ norm $\|s\|_\infty$.
- Use the test statistic $f(X) = \|\sum_{i=1}^n X_i\|_\infty$.
- Assume each noise vector N_i has a sign-symmetric distribution: N_i has same distribution as $-N_i$

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \ldots, n$
- Interested to detect "sparse" vectors s; having large ℓ_∞ norm $\|s\|_\infty$.
- Use the test statistic $f(X) = \| \sum_{i=1}^n X_i \|_\infty$.
- Assume each noise vector N_i has a sign-symmetric distribution: N_i has same distribution as $-N_i$
- Use randomization test with transforms $(X_1, \ldots, X_n) \to (\pm X_1, \ldots, \pm X_n)$

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \ldots, n$
- Interested to detect "sparse" vectors s; having large ℓ_∞ norm $\|s\|_\infty$.
- Use the test statistic $f(X) = \|\sum_{i=1}^{n} X_i\|_\infty$.
- Assume each noise vector N_i has a sign-symmetric distribution: N_i has same distribution as $-N_i$
- Use randomization test with transforms $(X_1, \ldots, X_n) \to (\pm X_1, \ldots, \pm X_n)$
  - Valid under weak non-parametric assumption of symmetry; typically not true of tests with critical values chosen via parametric models

Detecting sparse vectors

- Observe independent vectors $X_i = s + N_i$, $i = 1, \dots, n$
- Interested to detect "sparse" vectors s; having large ℓ_∞ norm $\|s\|_\infty$.
- Use the test statistic $f(X) = \|\sum_{i=1}^n X_i\|_\infty$.
- Assume each noise vector N_i has a sign-symmetric distribution: N_i has same distribution as $-N_i$
- Use randomization test with transforms $(X_1, \dots, X_n) \to (\pm X_1, \dots, \pm X_n)$
  - Valid under weak non-parametric assumption of symmetry; typically not true of tests with critical values chosen via parametric models
- Result: consistent when $\|s\|_\infty > 2(1+\varepsilon) \cdot \|n^{-1} \sum_{i=1}^n N_i\|_\infty$ with probability $\to 1$; for any $\varepsilon > 0$.

Numerical example

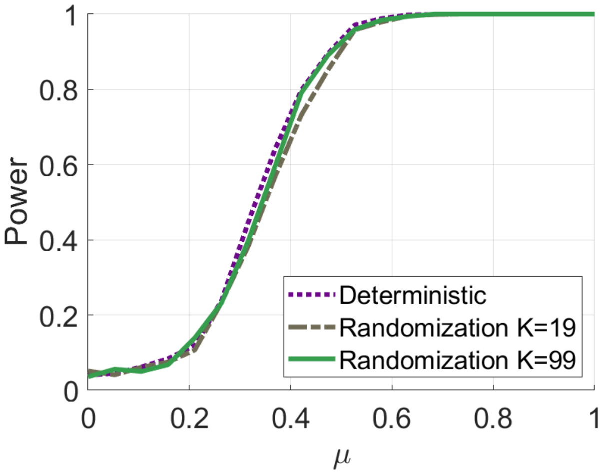


Figure: Evaluating the power of a randomization test in comparison with the deterministic test as a function of signal strength in sparse vector detection: standard normal noise, $n = p = 100$, $s = (\mu, 0, \ldots, 0)^\top$, $\alpha = 0.05$, 1000 repetitions.

Detecting sparse vectors: concrete examples

- For concrete results, need to control $\|n^{-1}\sum_{i=1}^{n} N_i\|_\infty$; can use *any concentration inequalities*

Detecting sparse vectors: concrete examples

- For concrete results, need to control $\|n^{-1}\sum_{i=1}^{n} N_i\|_\infty$; can use *any concentration inequalities*
- Sub-exponential noise with iid coordinates: $s \sim \sqrt{\log p/n}$

Detecting sparse vectors: concrete examples

- For concrete results, need to control $\|n^{-1}\sum_{i=1}^{n} N_i\|_\infty$; can use *any concentration inequalities*
- Sub-exponential noise with iid coordinates: $s \sim \sqrt{\log p/n}$
  - **Theorem.** This is the minimax optimal detection rate for sub-exponential noise (previously only known for Gaussian)

Detecting sparse vectors: concrete examples

- For concrete results, need to control $\|n^{-1}\sum_{i=1}^{n} N_i\|_\infty$; can use *any concentration inequalities*
- Sub-exponential noise with iid coordinates: $s \sim \sqrt{\log p/n}$
  - **Theorem.** This is the minimax optimal detection rate for sub-exponential noise (previously only known for Gaussian)
  - Randomization test is optimal

Detecting sparse vectors: concrete examples

- For concrete results, need to control $\|n^{-1}\sum_{i=1}^{n} N_i\|_\infty$; can use *any concentration inequalities*
- Sub-exponential noise with iid coordinates: $s \sim \sqrt{\log p/n}$
  - **Theorem.** This is the minimax optimal detection rate for sub-exponential noise (previously only known for Gaussian)
  - Randomization test is optimal
- Noise with finite Ψ-Orlicz norm $\Psi(x) = \exp(\ln[x+1]^\kappa) - 1$ (tail decay $\sim \exp(-\ln[x+1]^\kappa)$), $s \sim \exp[(\log p)^{1/\kappa}]/\sqrt{n}$.

Intriguing numerical example: two-sample t-test

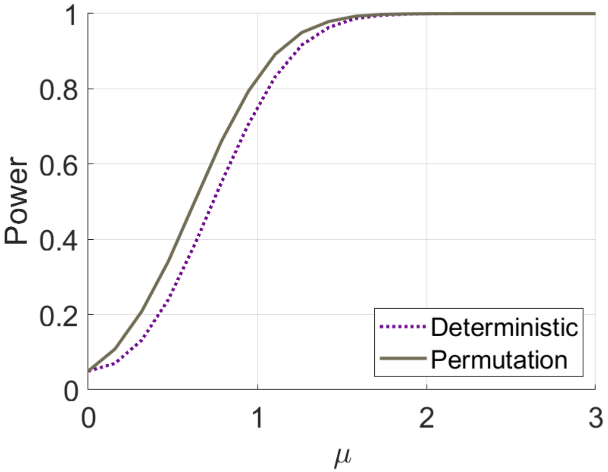


Figure: Evaluating the power of the permutation test in comparison with the deterministic test as a function of signal strength in two-sample testing, using t-test: standard normal noise, $n_1 = n_2 = 15$, $p = 1$, $s_1 - s_2 = (\mu, 0, \ldots, 0)^\top$, $\alpha = 0.05$, $K = 99$, 100,000 repetitions.

Summary

- Established consistency of randomization tests in modern settings: they can work almost as if they know the parametric model.

Summary

- Established consistency of randomization tests in modern settings: they can work almost as if they know the parametric model.
- Thank you!