High-dimensional Inference and Beyond

Runmin Wang

March 18, 2023

Motivation

A micro-array dataset measures gene expression.

"**Large** p, **small** n": number of genes p of order 10^3, number of samples n of order 10^2. n can be even smaller for rare diseases.

Possible questions:

- Identify gene-sets that are associated with clinical outcomes.
- Compare gene expressions for different groups.

Other examples:

- Network/tensor-valued time series
- Optimal portfolio construction

Classical statistical analysis requires $\boldsymbol{n} \gg \boldsymbol{p}$.

New Challenges

1. Traditional methods will not work anymore.
   - Examples:
     - $\|\bar{X}\|_2$ is not a consistent estimator of $\|\mu\|_2$.
     - Sample covariance matrix may not be invertible. Difficult to normalize a statistic.
     - Overfitting when the number of predictors is larger than the sample size.
2. New theoretical tools need to be developed to handle growing p.
3. Computational complexity grows in both n and p.

Possible Solutions

Assume that $X_1, ..., X_n \in \mathbb{R}^p$ are i.i.d. Denote $\boldsymbol{\mu} = \mathbb{E}[X_1]$. Both n and p can grow to infinity. Test: $\mathcal{H}_0 : \boldsymbol{\mu} = \mathbf{0}_p$ v.s. $\mathcal{H}_1 : \boldsymbol{\mu} \neq \mathbf{0}_p$.

1. Dimension Reduction: find a coefficient matrix $\boldsymbol{A} \in \mathbb{R}^{m \times p}$ where $m \geq 1$ is fixed, such that $A\boldsymbol{\mu} = \mathbf{0}_m$ if and only if $\boldsymbol{\mu} = \mathbf{0}_p$.

   This technique is useful, if we know $\boldsymbol{\mu}$ is a sparse vector, i.e. most components of $\boldsymbol{\mu}$ are zero.

2. Equivalent hypothesis: $\mathcal{H}_0' : \|\boldsymbol{\mu}\| = 0$ v.s. $\mathcal{H}_1' : \|\boldsymbol{\mu}\| \neq 0$, where $\| \cdot \|$ is some norm defined on $\mathbb{R}^p$.

   Popular choices: for $\boldsymbol{x} = (x_1, x_2, ..., x_p) \in \mathbb{R}^p$,
   - $\|\boldsymbol{x}\|_\infty = \max(|x_1|, |x_2|, ..., |x_p|)$ (ℓ_∞ norm)
   - $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}$ (ℓ_2 norm)

   No need to find $\boldsymbol{A}$. In fact looking for $\boldsymbol{A}$ is not easy as well, and it is often a separate problem.

Problem Setting

- One-sample testing
  - Problem: Given i.i.d. high-dimensional random vectors $X_1, \ldots, X_n \in \mathbb{R}^p$, and $\Theta = \Theta(X_1) = \{\theta_l : l \in \mathcal{L}\}$.
  - Goal: Test $\Theta = \mathbf{0}$ against $\Theta \neq \mathbf{0}$.
- Two-sample testing
  - Problem: Given i.i.d. high-dimensional random vectors $X_1, \ldots, X_n; Y_1, \ldots, Y_m \in \mathbb{R}^p$, and $\Theta = \Theta(X_1, Y_1) = \{\theta_l : l \in \mathcal{L}\}$.
  - Goal: Test $\Theta = \mathbf{0}$ against $\Theta \neq \mathbf{0}$.

Literature Review

Focus on i.i.d. data under high-dimensional setting.

ℓ_2-type statistics for **dense** alternatives:

- Mean testing: Bai and Saranadasa (1996); Chen and Qin (2010); Goeman et al. (2006); Gregory et al. (2015); Srivastava and Du (2008); Srivastava et al. (2016);
- Covariance testing: Bai et al. (2009); Chen et al. (2010); Ledoit and Wolf (2002); Li et al. (2012);
- Component-wise independence testing: Leung and Drton (2018);
- Simultaneous testing for the coefficients of linear model: Zhong and Chen (2011);
- ...

Literature Review

ℓ_∞-type statistics for **sparse** alternatives:

- Mean testing: Cai et al. (2014); Hall and Jin (2010);
- Covariance testing: Cai and Jiang (2011); Cai et al. (2013); Jiang (2004); Liu et al. (2008); Shao and Zhou (2014);
- Component-wise independence testing: Han et al. (2017); Drton et al. (2020);
- ...

Determine the norm

- Practically, which norm should we use? Any difference?
- It usually depends on the **sparsity** of Θ (under the alternative).
- When Θ is **sparse and strong**, the test is more powerful when using a larger q (ℓ_∞ is most powerful).
- When Θ is **dense and weak**, the test is more powerful when using a smaller q.
- The test using a wrong norm could have no power at all.

  Example:
  - $\boldsymbol{\mu} = (p^{-1/2}, p^{-1/2}, ..., p^{-1/2})$: $\|\boldsymbol{\mu}\|_2 = 1$ and $\|\boldsymbol{\mu}\|_\infty = p^{-1/2} \to 0$.
  - $\boldsymbol{\mu} = (1, 0, ..., 0)$: Although $\|\boldsymbol{\mu}\|_2 = \|\boldsymbol{\mu}\|_\infty = 1$, using ℓ_2 norm needs to aggregate all component sequences in the data therefore it has a much larger noise comparing to ℓ_∞ norm based method.

Literature Review

He et al. (2021, AoS): focus on mean (covariance) testing.

- Construct ℓ_q norm based tests for even q and $q = \infty$.
- Combine tests for different q's to achieve adaptive testing, i.e. the test is powerful against both sparse and dense alternatives.

Our method:

- ℓ_q norm based statistics for a general high-dimensional parameter.
- Powerful against both dense and sparse alternative.
- Asymptotically normal and independent under the null and alternative.
- Dynamic programming method to speed up computation.

U-statistic

A U-statistic is an unbiased estimator of the parameter in interest θ, which is defined as an average (across all combinatorial selections of the given size from the full set of observations) of the basic estimator applied to the sub-samples.

Example: Assume that $X_1, ..., X_n \in \mathbb{R}$ are i.i.d..

- $\theta = \mathbb{E}[X_1]$. Kernel function: $h(x) = x$.

$$U_n(X_1, ..., X_n) = \binom{n}{1}^{-1} \sum_{i=1}^{n} h(X_i).$$

- $\theta = \mathbb{E}^2[X_1]$. Kernel function: $h(x_1, x_2) = x_1 x_2$.

$$U_n(X_1, ..., X_n) = \binom{n}{2}^{-1} \sum_{i,j=1, i \neq j}^{n} h(X_i, X_j).$$

U-Statistic Construction (even q)

Consider $\|\Theta\|_q^q \triangleq \sum_{l=1}^{|\mathcal{L}|} \theta_l^q$ (for one-sample testing).

- Start with symmetric (core) kernel functions $h = (h_1, ..., h_{|\mathcal{L}|})$ (of order r) s.t.

$$\mathbb{E}[h_l(X_1, \ldots, X_r)] = \theta_l.$$

  for any $l = 1, 2, ..., |\mathcal{L}|$.

- Then we have

$$\mathbb{E}[h_l(X_1, \ldots, X_r) \cdots h_l(X_{(q-1)r+1}, \ldots, X_{qr})] = \theta_l^q.$$

- Derive an unbiased U-statistic for $\|\Theta\|_q^q$ (order qr):

$$U_{n,q} = \sum_{l \in \mathcal{L}} (P_{qr}^n)^{-1} \sum_{1 \le i_1, \ldots, i_{qr} \le n}^{*} \prod_{c=1}^{q} h_l(X_{i_{(c-1)r+1}}, \ldots, X_{i_{cr}}),$$

  where $\sum^*$ is over all distinct indexes.

Examples

- Test against $\mathcal{H}_0 : \boldsymbol{\mu} := \mathbb{E}[X_i] \equiv \mathbf{0}$. Consider $h_l(X_i) = x_{i,l}$, so

$$U_{n,q} = \sum_{l=1}^{p} (P_q^n)^{-1} \sum_{1 \le i_1, \ldots, i_q \le n}^{*} \prod_{c=1}^{q} x_{i_c,l}.$$

- Test against $\mathcal{H}_0 : Var(X_i)$ is a diagonal matrix (assuming $\boldsymbol{\mu} = 0$). Consider $h_l(X_i) = x_{ip_1} x_{ip_2}$, $p_1 < p_2$.

Examples

- Spatial sign based testing (Wang et al., 2015, JASA),
  $\mathcal{H}_0 : \mathbb{E}[X_i \,/\, \|X_i\|\,] = \mathbf{0}$.
  $\mathcal{L} = [p]$, and $h_l(X_i) = x_{i,l} \,/\, \|X_i\|$ with $r = 1$.

- Testing for linear model coefficients (Zhong and Chen, 2011, JASA),
  $Y = X\boldsymbol{\beta} + \varepsilon$, $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 (\Leftrightarrow \Theta = \Sigma_X(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = 0)$.
  $\mathcal{L} = [p]$, and

  $$h_l\big((X_1, Y_1), (X_2, Y_2)\big) = [\,(X_1 - X_2)(Y_1 - Y_2 - (X_1 - X_2)^T\boldsymbol{\beta}_0)\,]_l/2$$

  with $r = 2$.

Null Distribution

Regularity conditions:

- Guarantee the dominance of the leading term;
- Guarantee the weak cross-sectional dependence.

Theorem 1 (Limiting null distribution)

Suppose $h = (h_1, ..., h_{|\mathcal{L}|})$ is a kernel with order r, under some regularity conditions.
Then we have under the null,

$$[(qs)!]^{-1/2} \binom{r}{s}^{-q} n^{qs/2} \widetilde{\Sigma}_s^{-1/2}(q) U_{n,q} \xrightarrow{\mathcal{D}} N(0,1), \tag{1}$$

where s is the order of degeneracy of h, and $\tilde{\Sigma}_s^{-1/2}(q)$ is a technical quantity which needs to be estimated later on.
Furthermore, for any finite set $I \subset \boldsymbol{Z}_+$, $(U_{n,q})_{q \in I}$ are **asymptotically jointly independent**.

Sketch of the proof

1. Decompose $U_{n,q}$ and find the leading term: Hajék projection and Hoeffding decomposition.

$$U_{n,q} = \binom{r}{s}^q U_{n,q}^{(qs)}[1 + o_p(1)].$$

2. $U_{n,q}^{(qs)}$ can be further written as a martingale: Martingale central limit theorem.

An Asymmetric U-statistic with Dynamic Programming

Define

$$D_{q,l}^{M}(m) = \sum_{1 \leq i_1 < \cdots < i_{qr} \leq m} \prod_{c=1}^{q} h_l(X_{i_{(c-1)r+1}}, \ldots, X_{i_{cr}}),$$

and

$$U_{n,q}^{M} = \binom{n}{qr}^{-1} \sum_{l \in \mathcal{L}} D_{q,l}^{M}(n).$$

We may calculate $D_{q,l}^{M}$ recursively by

$$D_{c,l}^{M}(m) = D_{c,l}^{M}(m-1) + D_{c-1,l}^{M}(m-1) h_l(X_m), \quad m \geq c,$$

with $D_{c,l}^{M}(m) = 0$ for $1 \leq m < c$.
Reduce computation from $O(q n^{qr} |\mathcal{L}|)$ to $O(q n^{r} |\mathcal{L}|)$.

Variance Estimator

We consider two approaches for estimating the variance.

- Plug-in method (mainly used for $U_{n,q}$ with $r = 1$): Construct the **consistent** estimator of $\tilde{\Sigma}_s(q)$, which has the form of proposed statistic associated with some kernels derived by h_l.
- **Permutation** based variance estimator (for both $U_{n,q}$ and $U^M_{n,q}$): Find the variance of the statistics computed on permuted data.

Asymptotic Distribution under Alternative

Define

$$\gamma_{n,q} = n^{qs/2}\widetilde{\Sigma}_s^{-1/2}(q)\|\Theta\|_q^q.$$

Theorem 2 (Alternative)

Under the same assumption as null, we have

- *Suppose* $\gamma_{n.q} \to \infty$. *Then* $n^{qs/2}\widetilde{\Sigma}_s^{-1/2}(q)U_{n,q} \xrightarrow{\mathcal{P}} \infty$ *and the power goes to 1.*
- *Suppose* $\gamma_{n.q} \to 0$. *Then* $U_{n,q}$ *has the same asymptotic distribution as null and the power converges to* α.
- *Suppose* $\gamma_{n,q} \to \gamma \in (0, \infty)$ *(***local*** alternative). We have*

$$n^{qs/2}\widetilde{\Sigma}_s^{-1/2}(q)U_{n,q} \xrightarrow{\mathcal{D}} N\left(\gamma, [(qs)!]^{1/2}\binom{r}{s}^q\right).$$

Adaptive Testing

1. Conduct tests for a set of q's, i.e. $q_1, q_2, ..., q_I$;

2. Obtain p-values from each test: p_{q_1},...,p_{q_I};

3. Adaptive test statistic: $p_{adp} = \min\{p_{q_1}, ..., p_{q_I}\}$;

4. For a level α test, reject if $p_{adp} < 1 - (1-\alpha)^{1/I}$.

Remark: In real applications we recommend to combine two tests with different q's to obtain an adaptive test.

Simulation Studies

Consider testing for linear model $Y_i = X_i\beta + \varepsilon_i$.

$$\mathcal{H}_0 : \boldsymbol{\beta} = \mathbf{0} \quad v.s. \quad \mathcal{H}_a : \boldsymbol{\beta} \neq \mathbf{0}.$$

Simulation setting:

- $X_i \overset{i.i.d.}{\sim} N(0, I_p)$, independent of $\varepsilon \overset{i.i.d.}{\sim} N(0,1)$.
- $\boldsymbol{\beta} = \delta(\mathbf{1}_r, \mathbf{0}_{p-r})$.

(n,p)	δ, r	$q=2$	$q=4$	$q=6$	$q=2,4$	$q=2,6$
(100,50)	0,NA	5.7	6.4	3.1	6.3	4.6
	0.4,2	50.4	72.0	56.2	74.0	67.4
	0.05, p	70.4	19.8	12.6	65.2	63.8
(200,100)	0,NA	5.5	4.9	3.5	5.1	5.6
	0.4,2	76.0	98.0	95.2	98.2	96.4
	0.05, p	98.8	36.8	20.4	98.2	98.2

Table 1: Size and power in % for linear model coefficient testing

Summary

- ℓ_q-norm based U-statistic for high dimensional testing.
- Asymptotically normal and independent statistics; adaptive test with high power against both dense and sparse alternatives.
- No explicit constraints on p with encouraging finite sample performance.

Future Work

- Study the asymptotic independence of ℓ_∞-based statistic.
- Generalize to non-i.i.d. data.

Dimension-agnostic Inference

- Motivation:
  - The calibration of a test depends on the assumption of how p scales with n, which is usually pre-decided but unverifiable.
  - An illustrative example:

<table>
<tr><th>Data</th><th>Possible Scales</th><th>Calibration of a Test</th></tr>
<tr><td rowspan="4">$n = 100$
$p = 20$</td><td>$p = 20$ fixed</td><td>low-dimensional method
(fix p while let $n \to \infty$)</td></tr>
<tr><td>$p/n = 0.2$ fixed</td><td rowspan="3">high-dimensional method
(let $p \to \infty$ as $n \to \infty$)</td></tr>
<tr><td>$p/n^2 = 0.002$ fixed</td></tr>
<tr><td>$p/\sqrt{n} = 2$ fixed</td></tr>
</table>

  - **Is there a test that works under all dimensional settings?**
- Dimension-agnostic (Kim and Ramdas (2020)):

  $\cdots$ *the goal of dimension-agnostic inference: developing methods whose validity does not depend on any assumption on p versus n.*
- Our method: Dimension-agnostic change point detection

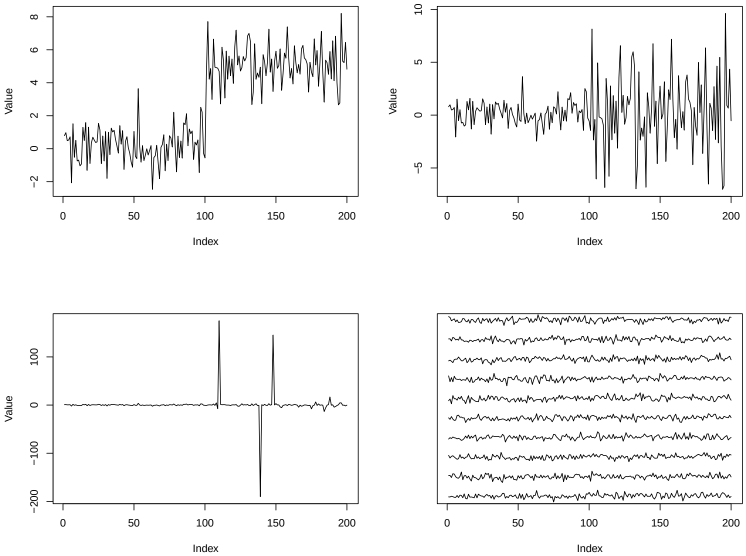
Literature Review

- Existing work:

Low/fixed-dimensional	High-dimensional
Page (1954)	Horváth and Hušková (2012)
Page (1955)	Jirak (2015), Cho (2016)
Shao and Zhang (2010)	Wang and Samworth (2018)
(Review) Aue and Horváth (2013)	Enikeeva and Harchaoui (2019)
(Review) Casini and Perron (2019), etc	Wang et al. (2022)
	Yu and Chen (2022), etc

- Limitations:
  - Low/fixed-dimensional setup (p is fixed and small)
    - theory is justified specifically for small/fixed p.
    - not applicable when $p > n$.
    - serious size distortion when p is moderate.
  - High-dimensional setup (p is high and is comparable to or exceeds n)
    - the approximation accuracy may highly rely on the central limit effect from the high dimension.
    - serious size distortion for data of low or moderate dimension.
    - different methods may require different growth rate of p, different sparsity, etc..

Change Point Analysis



Problem Setup

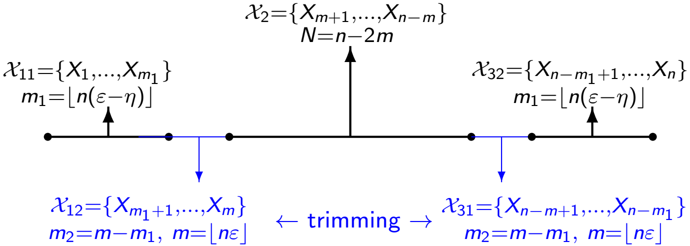
- Data: Given the observations $\{X_t\}_{t=1}^n \in \mathbb{R}^p$ with both temporal and cross-sectional dependence. We denote $\mu_t = \mathbb{E}\big[X_t\big]$ for $t = 1, \cdots, n$.
- Single change point testing:
  $H_0: \ \mu_1 = \cdots = \mu_n$ v.s.

$$H_1: \ \mu_1 = \cdots = \mu_{k_0} \neq \mu_{k_0+1} = \cdots = \mu_n$$

  where $k_0 = n\varepsilon_0$ with $\varepsilon_0 \in (0,1)$ is an unknown location.
- Applications:
  - Finance: stock return volatility change
  - Neuroscience: functional magnetic resonance imaging (fMRI) study
  - Credit card fraud detection/monitoring

Methodology

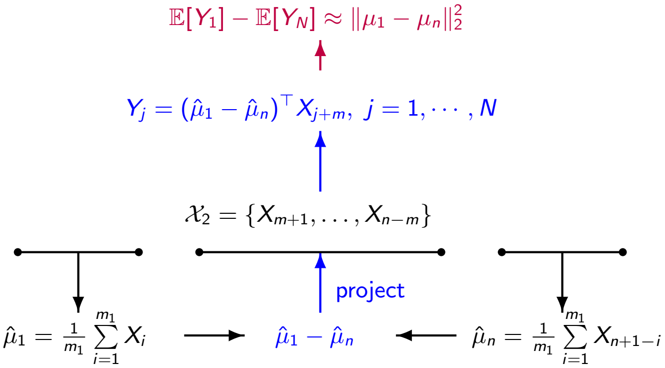
- **Step 1: Splitting (and trimming if temporal dependence exists)**



  - $\varepsilon \in (0, 1)$: the splitting ratio, satisfying that $\varepsilon_0 \in (\varepsilon, 1 - \varepsilon)$.
  - $\eta \in [0, \varepsilon)$: the trimming ratio.

Methodology

- **Step 2: Projection**



Proposed Test Statistic

- **Test statistic for the single change point** (Shao and Zhang (2010)):
  - For $k = 1, 2, \ldots, N-1$, define the CUMSUM statistic as

$$T_n(k) = N^{-1/2} \sum_{t=1}^{k} \left(Y_t - \bar{Y}_N\right) \tag{2}$$

    where $\bar{Y}_N = \frac{1}{N} \sum_{j=1}^{N} Y_j$.
  - For $k = 1, 2, \ldots, N-1$, define the self-normalizer as

$$V_n(k) = N^{-2} \left(\sum_{t=1}^{k} \left(S_{1,t} - \frac{t}{k} S_{1,k} \right)^2 + \sum_{t=k+1}^{N} \left(S_{t,N} - \frac{N-t+1}{N-k} S_{+1,N} \right)^2 \right), \tag{3}$$

    where $S_{a,b} = \sum_{j=a}^{b} Y_j$ denotes the cumulative sum.
  - The proposed test statistic is defined as

$$G_n = \sup_{k=1,\ldots,N-1} T_n(k) V_n^{-1/2}(k). \tag{4}$$

Theoretical Results

- Three data generating processes:

Dimensionality	Data Generating Process	Temporal Dependence	Cross-sectional Dependence
Fixed p	**stationary sequence (DGP1)** Long-run variance Ω positive definite	weak	arbitrary
Diverging p	**linear process (DGP2)** $X_t = \mu_t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, $\{\varepsilon_t\}_{t\in\mathbf{0}} \overset{iid}{\sim} (0, \Gamma)$	weak	weak
	static factor model (DGP3) $X_t = \mu_t + \Lambda F_t + Z_t$, $\Lambda \in \mathbb{R}^{p\times s}, s \ll p$ $\{F_t\}_{t=1}^{n} \in \mathbb{R}^s \sim (0, \Omega)$ $\{Z_t\}_{t=1}^{n} \in \mathbb{R}^p \sim (0, \Sigma)$ linear process $\{F_t\}_{t=1}^{n} \perp\!\!\!\perp \{Z_t\}_{t=1}^{n}$	weak	strong

Theoretical Results

- Limiting null distribution

p	$\{X_t\}_{t=1}^n$	Key assumptions	Limiting Null Distribution
Fixed p	Stationary (DGP1)	Functional CLT	$G_n \to^d G$
$p \to \infty$	Linear process (DGP2)	$\|a_j\| \lesssim \rho^j$ for $\rho \in (0,1)$ and $j \geq 0$	
		$\rho^{m_2/4} \|\Gamma\|_F = o\left(\frac{n}{\log(n)}\right)$	
	Factor model (DGP3)	Functional CLT for $\{F_t\}_{t=1}^n$	
		$\rho^{m_2/4} \|\Gamma\|_F = o\left(\frac{n}{\log(n)}\right)$	

where $G := \sup_{r \in [0,1]} \left(B(r) - rB(1)\right) V^{-1/2}(r)$ with $\{B(r)\}_{r \in [0,1]}$ denoting the standard Brownian motion, and $V(r)$ is given by

$$V(r) = \int_0^r \left(B(s) - \frac{s}{r}B(r)\right)^2 ds + \int_r^1 \left(B(1) - B(s) - \frac{1-s}{1-r}\left(B(1) - B(r)\right)\right)^2 ds.$$

Theoretical Results

- Theoretical asymptotic power against the local alternative

	Stationary (DGP1)	Linear process (DGP2) (with $A^{(0)} = \sum_{\ell=0}^{\infty} a_\ell$)	Factor model (DGP3)
$\mathbb{P}(G_n > G_{1-\alpha}) \to \alpha$	$\sqrt{n}\|\delta\|_2 \to 0$	$\frac{\sqrt{n}\|\delta\|_2}{\|A^{(0)}\Gamma(A^{(0)})^\top\|_F^{1/2}} \to 0$	$\frac{\sqrt{n}\|\delta\|_2}{\max\{\|\Lambda\|,\|\Gamma\|_F^{1/2}\}} \to 0$
$\mathbb{P}(G_n > G_{1-\alpha}) \to \beta$ $\beta \in (\alpha, 1)$	$\sqrt{n}\|\delta\|_2 \to c_1$ $c_1 \in (0, \infty)$	$\frac{\sqrt{n}\|\delta\|_2}{\|A^{(0)}\Gamma(A^{(0)})^\top\|_F^{1/2}} \to c_2$ $c_2 \in (0, \infty)$	$\frac{\sqrt{n}\|\delta\|_2}{\max\{\|\Lambda\|,\|\Gamma\|_F^{1/2}\}} \to c_3$ $c_3 \in (0, \infty)$
$\mathbb{P}(G_n > G_{1-\alpha}) \to 1$	$\sqrt{n}\|\delta\|_2 \to \infty$	$\frac{\sqrt{n}\|\delta\|_2}{\|A^{(0)}\Gamma(A^{(0)})^\top\|_F^{1/2}} \to \infty$	$\frac{\sqrt{n}\|\delta\|_2}{\max\{\|\Lambda\|,\|\Gamma\|_F^{1/2}\}} \to \infty$

Generalization

- Single **dense** alternative $\rightarrow$ Single **sparse** alternative
  - use a sparse direction for projection
  - theory is wide open
- **Single** change point $\rightarrow$ **Multiple** change points
  - use scanning-based tests by Zhang and Lavitas (2018)
  - done with methodology, theory and simulations

Simulation Studies

- **DGP:** We generate the data from a p-dimensional AR(1) process

$$X_t - \mu_t = \kappa(X_{t-1} - \mu_{t-1}) + \epsilon_t \in \mathbb{R}^p, \qquad 1 \leq t \leq n,$$

  where $\kappa = 0.7$, $\{\epsilon_t\} \overset{iid}{\sim} \mathcal{N}_p(0, \Sigma)$ and Σ takes the following forms:
  - AR ($\Sigma_{i,j} = 0.8^{|i-j|}$)
  - CS ($\Sigma_{i,j} = 0.5 + 0.5\mathbf{1}\{i = j\}$)
  - ID ($\Sigma_{i,j} = \mathbf{1}\{i = j\}$)
- **Proposed Method:**
  - the test statistic targeting at dense alternatives: $G_{n,2}$
  - the test statistic targeting at sparse alternatives: $G_{n,\infty}$
  - the Bonferroni test based on $G_{n,2}$ and $G_{n,\infty}$: `Bonf`
  - the splitting ratio $\varepsilon = 0.1$, the trimming ratio $\eta = 0.04$
- **Comparison Methods:**
  - Wang et al. (2022), denoted by $T(\eta_0)$ where η_0 is a trimming parameter selected from $\{0, 0.01, 0.02, 0.05, 0.1\}$
- Significance level $\alpha = 0.05$
- 5000 Monte-Carlo replicates

Simulation Studies

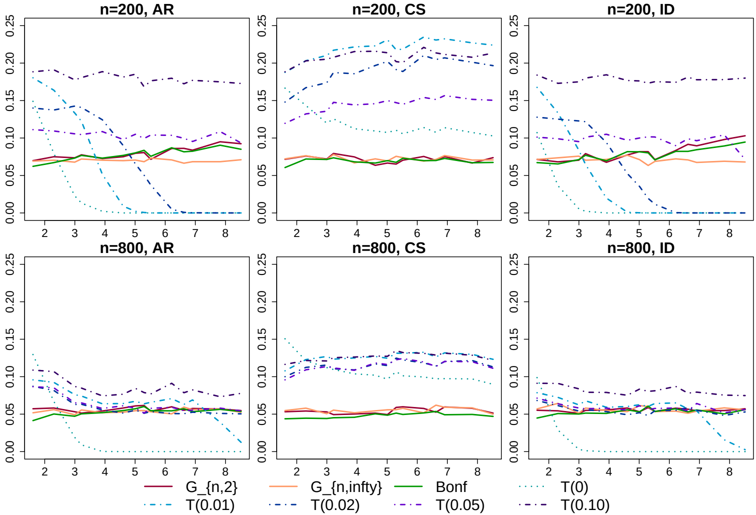


Figure 1: Empirical size curves versus the logarithm of p against the single change point alternative

Simulation Studies

- **Dense change point:**
  - $n = 200$, $k = \lfloor n/2 \rfloor + 1$ and $\mu_t = \begin{cases} (0, \ldots, 0)^\top, & 1 \leq t < k \\ c(1, \ldots, 1)^\top/\sqrt{p}, & k \leq t \leq n \end{cases}$
- **Sparse change point:**
  - $n = 200$, $k = \lfloor n/2 \rfloor + 1$ and $\mu_t = \begin{cases} (0, \ldots, 0)^\top, & 1 \leq t < k \\ c(0, 0, 1, 0, \ldots, 0)^\top, & k \leq t \leq n \end{cases}$
- **Parameters:**
  - The splitting ratio $\varepsilon = 0.1$, the trimming ratio $\eta = 0.04$.

Simulation Studies

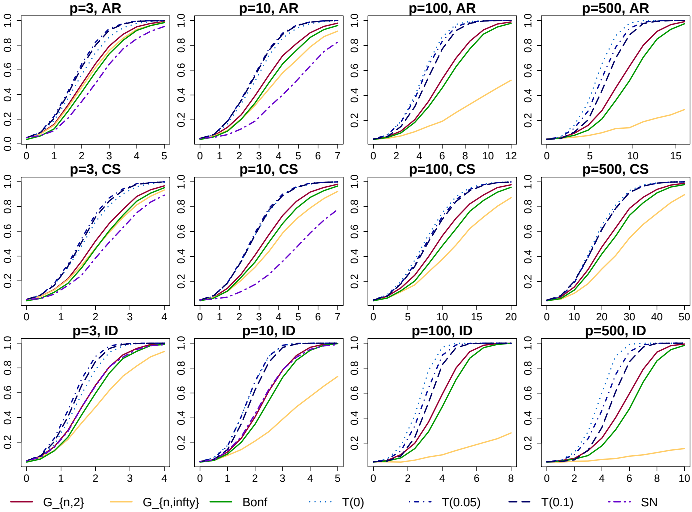


Figure 2: Power curves (size-adjusted) with a single dense change point

Simulation Studies

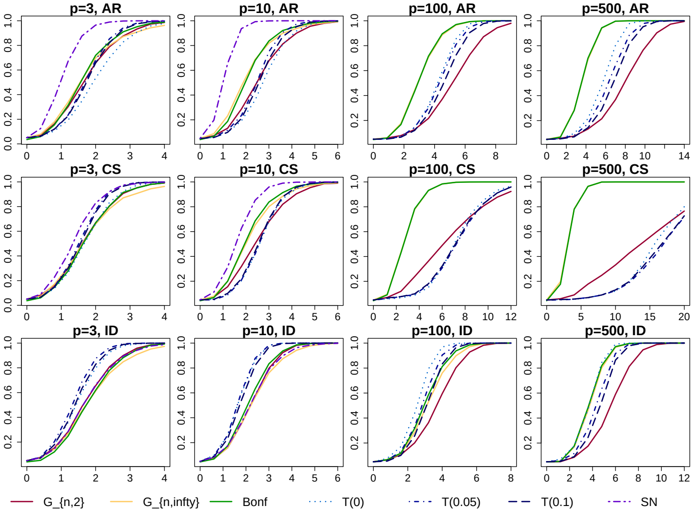


Figure 3: Power curves (size-adjusted) with a single sparse change point

Summary and Future Work

- **Summary**:
  - Dimension-agnostic testing for (weakly dependent) time series is feasible and natural since the sample split is unique for given proportions.
  - Agnostic to both the dimensionality and the magnitude of cross-sectional dependence.
- **Future work**:
  - Dimension-agnostic segmentation.
  - Dimension-agnostic testing for other parameters, e.g., covariance matrix, distribution, etc.

Thank you!

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *The Annals of Statistics*, pages 3822–3840.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6(2):311–329.

Cai, T., Liu, W., and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501):265–277.

Cai, T. T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *The Annals of Statistics*, 39(3):1496–1525.

Cai, T. T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 349–372.

Casini, A. and Perron, P. (2019). Structural breaks in time series. *Oxford Research Encyclopedia of Economics and Finance*.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, 10(2):2000–2038.

Drton, M., Han, F., and Shi, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *The Annals of Statistics*, 48(6):3206–3227.

Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, 47(4):2051–2079.

Goeman, J. J., Van De Geer, S. A., and Van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V., and Lahiri, S. N. (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association*, 110(510):837–849.

Hall, P. and Jin, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics*, 38(3):1686–1732.

Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, 104(4):813–828.

Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865–880.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics*, 43(6):2451–2483.

Kim, I. and Ramdas, A. (2020). Dimension-agnostic inference. *arXiv preprint arXiv:2011.05068*.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102.

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics*, 46(1):280–307.

Li, J., Chen, S. X., et al. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, 40(2):908–940.

Liu, W.-D., Lin, Z., and Shao, Q.-M. (2008). The asymptotic distribution and berry–esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *The Annals of Applied Probability*, 18(6):2337–2366.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527.

Shao, Q.-M. and Zhou, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *The Annals of Probability*, 42(2):623–648.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.

Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402.

Srivastava, R., Li, P., and Ruppert, D. (2016). Raptt: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics*, 25(3):954–970.

Wang, R., Zhu, C., Volgushev, S., and Shao, X. (2022). Inference for change points in high-dimensional data via selfnormalization. *The Annals of Statistics*, 50(2):781–806.

Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83.

Yu, M. and Chen, X. (2022). A robust bootstrap change point test for high-dimensional location parameter. *Electronic Journal of Statistics*, 16(1):1096–1152.

Zhang, T. and Lavitas, L. (2018). Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association*, 113:637–648.

Zhong, P.-S. and Chen, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association*, 106(493):260–274.