Multivariate Distribution-free Testing using Optimal Transport

Bodhisattva Sen[1]
Department of Statistics
Columbia University, New York

ICERM workshop on "Optimal Transport in Data Science"
Brown University
May 08, 2023

[1]Supported by NSF grant DMS-2015376

Multivariate two-sample testing

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

Multivariate two-sample testing

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$

- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

- When $d = 1$: Student (1908), Wilcoxon (1945), Cramér von-Mises (1928), Smirnov (1939), Wald and Wolfowitz (1940), Mann and Whitney (1947), Anderson (1962), ...

- When $d > 1$: Hotelling (1931), Weiss (1960), Bickel (1969), Friedman and Rafsky (1979), Schilling (1986), Henze (1988), Liu and Singh (1993), Székely (2003), Rosenbaum (2005), Gretton et al. (2012), Chen and Friedman (2017), Bhattacharya (2019), ...

Multivariate two-sample testing

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$

- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

- When $d = 1$: Student (1908), Wilcoxon (1945), Cramér von-Mises (1928), Smirnov (1939), Wald and Wolfowitz (1940), Mann and Whitney (1947), Anderson (1962), ...

- When $d > 1$: Hotelling (1931), Weiss (1960), Bickel (1969), Friedman and Rafsky (1979), Schilling (1986), Henze (1988), Liu and Singh (1993), Székely (2003), Rosenbaum (2005), Gretton et al. (2012), Chen and Friedman (2017), Bhattacharya (2019), ...

What is the distribution-free analogue of Hotelling's test when $d > 1$?

When $d = 1$

Two-sample t-test

- **Two-sample t-test**: Compares $\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$ & $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$

When $d = 1$

Two-sample t-test

- **Two-sample t-test**: Compares $\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$ & $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$
- Under H_0, the t-statistic has approximately t_{m+n-2} distribution
- Approximate (not valid for small sample sizes) level α test; requires moment assumptions; not robust to outliers

When $d = 1$

Two-sample t-test

- **Two-sample t-test**: Compares $\bar{X}_m = \frac{1}{m}\sum_{i=1}^m X_i$ & $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^n Y_j$
- Under H_0, the t-statistic has approximately t_{m+n-2} distribution
- Approximate (not valid for small sample sizes) level α test; requires moment assumptions; not robust to outliers

Question: Can we find a distribution-free test that is robust to outliers and heavy-tailed distributions and is also efficient?

When $d = 1$

Two-sample t-test

- **Two-sample t-test**: Compares $\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$ & $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$
- Under H_0, the t-statistic has approximately t_{m+n-2} distribution
- Approximate (not valid for small sample sizes) level α test; requires moment assumptions; not robust to outliers

Question: Can we find a distribution-free test that is robust to outliers and heavy-tailed distributions and is also efficient?

Answer: Wilcoxon rank-sum test (WRS) [Wilcoxon (1945)]

WRS test is **distribution-free**: null distribution is universal — does not depend on the distribution of the data (if it is a continuous dist.)

Wilcoxon rank-sum test (WRS)

- Pool $\{X_1, \ldots, X_m, Y_1, \ldots, Y_n\}$ to obtain ranks $\widehat{R}_{m,n}(X_i)$'s and $\widehat{R}_{m,n}(Y_j)$'s:

$$\mathrm{WRS} = \frac{1}{n}\sum_{j=1}^{n} \widehat{R}_{m,n}(Y_j) - \frac{1}{m}\sum_{i=1}^{m} \widehat{R}_{m,n}(X_i)$$

- WRS test is distribution-free and exact for all $P_1 = P_2$ continuous

Wilcoxon rank-sum test (WRS)

- Pool $\{X_1, \ldots, X_m, Y_1, \ldots, Y_n\}$ to obtain ranks $\widehat{R}_{m,n}(X_i)$'s and $\widehat{R}_{m,n}(Y_j)$'s:

$$\mathrm{WRS} = \frac{1}{n}\sum_{j=1}^{n} \widehat{R}_{m,n}(Y_j) - \frac{1}{m}\sum_{i=1}^{m} \widehat{R}_{m,n}(X_i)$$

- WRS test is distribution-free and exact for all $P_1 = P_2$ continuous
- Under H_0, $\left(\widehat{R}_{m,n}(X_1), \ldots, \widehat{R}_{m,n}(X_m), \widehat{R}_{m,n}(Y_1), \ldots, \widehat{R}_{m,n}(Y_n)\right)$ distributed uniformly over all $(m+n)!$ permutations of $\left\{\frac{1}{m+n}, \frac{2}{m+n}, \ldots, 1\right\}$

Wilcoxon rank-sum test (WRS)

- Pool $\{X_1, \ldots, X_m, Y_1, \ldots, Y_n\}$ to obtain ranks $\widehat{R}_{m,n}(X_i)$'s and $\widehat{R}_{m,n}(Y_j)$'s:

$$\mathrm{WRS} = \frac{1}{n}\sum_{j=1}^{n}\widehat{R}_{m,n}(Y_j) - \frac{1}{m}\sum_{i=1}^{m}\widehat{R}_{m,n}(X_i)$$

- WRS test is distribution-free and exact for all $P_1 = P_2$ continuous

- Under H_0, $\left(\widehat{R}_{m,n}(X_1), \ldots, \widehat{R}_{m,n}(X_m), \widehat{R}_{m,n}(Y_1), \ldots, \widehat{R}_{m,n}(Y_n)\right)$ distributed uniformly over all $(m+n)!$ permutations of $\left\{\frac{1}{m+n}, \frac{2}{m+n}, \ldots, 1\right\}$

- **WRS**: Exact test valid for all sample sizes

- Robust to outliers; does not need moment assumptions

- Based on univariate ranks — advent of classical nonparametrics

Efficiency of the WRS test

Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

Comparison of WRS test with t-test (under location shift alternatives)

- WRS test has 0.95 Pitman efficiency w.r.t t-test when P_1 is Gaussian

Efficiency of the WRS test

Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

Comparison of WRS test with t-test (under location shift alternatives)

- WRS test has 0.95 Pitman efficiency w.r.t t-test when P_1 is Gaussian
- Non-trivial efficiency lower bound of 0.864 w.r.t t-test [Hodges and Lehmann (1956)]; efficiency can be $+\infty$ (for heavy-tailed dist.)

Efficiency of the WRS test

Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

Comparison of WRS test with *t*-test (under location shift alternatives)

- WRS test has 0.95 Pitman efficiency w.r.t *t*-test when P_1 is Gaussian
- Non-trivial efficiency lower bound of 0.864 w.r.t *t*-test [Hodges and Lehmann (1956)]; efficiency can be $+\infty$ (for heavy-tailed dist.)
- Non-trivial efficiency lower bound of 1 w.r.t *t*-test [Chernoff and Savage (1958)] when the following revised statistic is used:

$$\frac{1}{n}\sum_{j=1}^{n} \Phi^{-1}(\widehat{R}_{m,n}(Y_j)) - \frac{1}{m}\sum_{i=1}^{m} \Phi^{-1}(\widehat{R}_{m,n}(X_i))$$

Generalize all these properties to multivariate data

Question: How to construct efficient distribution-free **multivariate** tests?

- When $d = 1$ tests based on "ranks" are distribution-free
- How to define multivariate ranks that lead to distribution-free tests?

Optimal transport!

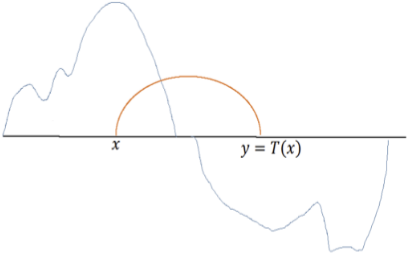
Outline

Outline

1. Optimal Transport: Monge's Problem
   - Introduction
   - Multivariate Ranks via Optimal Transport
2. Multivariate Two-sample Goodness-of-fit Testing
   - Distribution-free Testing — Hotelling T^2 and Rank Hotelling
   - Lower bounds on Asymptotic (Pitman) Relative Efficiency
3. Other Applications of Distribution-free Inference
   - Testing for Mutual Independence
   - Testing for Symmetry

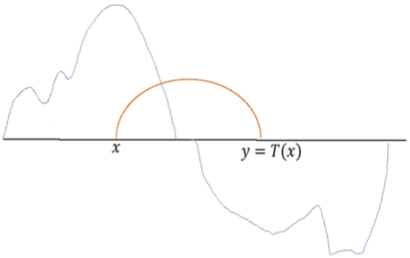
Optimal transport: Monge's problem

Gaspard Monge (1781): What is the cheapest way to transport a pile of sand to cover a sinkhole?



Optimal transport: Monge's problem

Gaspard Monge (1781): What is the cheapest way to transport a pile of sand to cover a sinkhole?



Goal: $$\inf_{T : T(X) \sim \mu} \mathbb{E}_\nu[c(X, T(X))] \qquad X \sim \nu$$

- ν ("data" dist.) and μ ("reference" dist.)
- $c(x, y) \geq 0$: cost of transporting x to y (e.g., $c(x, y) = \|x - y\|^2$)
- T transports ν to μ: $T\#\nu = \mu$ (i.e., $T(X) \sim \mu$ where $X \sim \nu$)

Rank function as the optimal transport (OT) map: when $d = 1$

- $X \sim \nu$ (continuous dist.) on $\mathbb{R}$, $F \equiv F_\nu$ c.d.f. of ν
- **Rank**: The population rank of $x \in \mathbb{R}$ is $F(x)$ (a.k.a. the c.d.f. at x)
- **Property**: $F(X) \sim \text{Uniform}([0,1]) \equiv \mu$; i.e., F transports ν to μ

Rank function as the optimal transport (OT) map: when $d = 1$

- $X \sim \nu$ (continuous dist.) on $\mathbb{R}$, $F \equiv F_\nu$ c.d.f. of ν
- **Rank**: The population rank of $x \in \mathbb{R}$ is $F(x)$ (a.k.a. the c.d.f. at x)
- **Property**: $F(X) \sim \text{Uniform}([0, 1]) \equiv \mu$; i.e., F transports ν to μ
- If $\mathbb{E}_\nu[X^2] < \infty$, the c.d.f F is the optimal transport (OT) map as

$$F = \underset{T:T\#\nu=\mu}{\arg\min}\ \mathbb{E}_\nu[(X - T(X))^2]$$

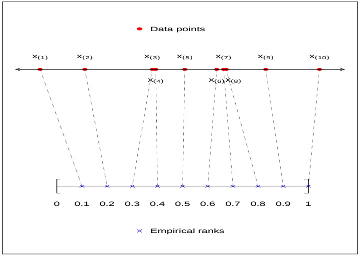
where we take

$$c(x, y) = (x - y)^2$$

Sample rank map: when $d = 1$

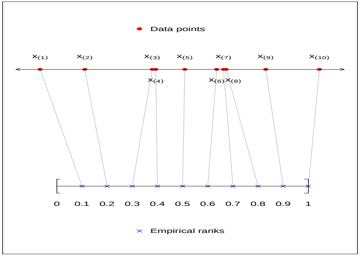
- **Data**: $X_1, \ldots, X_n$ iid ν (cont. distribution) on $\mathbb{R}$

- **Sample rank map**: $\hat{R}_n : \{X_1, X_2, \ldots, X_n\} \longrightarrow \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$



Sample rank map: when $d = 1$

- **Data**: $X_1, \ldots, X_n$ iid ν (cont. distribution) on $\mathbb{R}$
- **Sample rank map**: $\hat{R}_n : \{X_1, X_2, \ldots, X_n\} \longrightarrow \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$



Sample rank map $\hat{R}_n$ is the OT map that transports

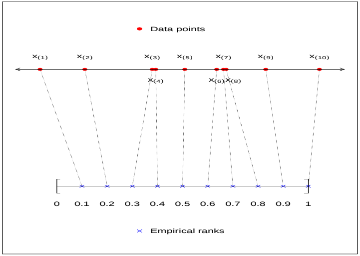
$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \qquad \text{to} \qquad \mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{\frac{j}{n}},$$

i.e., $$\hat{R}_n := \mathop{\arg\min}_{T:T\#\nu_n=\mu_n} \frac{1}{n}\sum_{i=1}^{n} |X_i - T(X_i)|^2$$

Sample rank map: when $d = 1$

- **Data**: $X_1, \ldots, X_n$ iid ν (cont. distribution) on $\mathbb{R}$

- **Sample rank map**: $\hat{R}_n : \{X_1, X_2, \ldots, X_n\} \longrightarrow \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$



Sample rank map $\hat{R}_n$ is the OT map that transports

$$\nu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \qquad \text{to} \qquad \mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{\frac{j}{n}},$$

i.e., $$\hat{R}_n := \underset{T:T\#\nu_n=\mu_n}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} |X_i - T(X_i)|^2 = \underset{T:T\#\nu_n=\mu_n}{\arg\max} \frac{1}{n}\sum_{i=1}^{n} X_{(i)} \cdot T(X_{(i)})$$

- $X \sim \nu$; ν is a probability measure (abs. cont.) in $\mathbb{R}^d$ $(d \geq 1)$
- **Reference dist.**: $U \sim \mu$ on $\mathcal{S} \subset \mathbb{R}^d$ $(\mu = \mathrm{Unif}([0,1]^d),\ N(0, I_d))$
- Find OT map T s.t. $T(X) \stackrel{d}{=} U \sim \mu$ (μ abs. cont.)

- $X \sim \nu$; ν is a probability measure (abs. cont.) in $\mathbb{R}^d$ ($d \geq 1$)
- **Reference dist.**: $U \sim \mu$ on $\mathcal{S} \subset \mathbb{R}^d$ ($\mu =$ Unif$([0,1]^d)$, $N(0, I_d)$)
- Find OT map T s.t. $T(X) \stackrel{d}{=} U \sim \mu$ (μ abs. cont.)

Population rank function [Hallin (2017), Chernozhukov et al. (2017)]

If $\mathbb{E}_\nu \|X\|^2 < \infty$, rank function $R : \mathbb{R}^d \to \mathcal{S}$ is the OT map s.t.

$$R := \operatorname*{arg\,min}_{T : T\#\nu = \mu} \mathbb{E}_\nu \|X - T(X)\|^2$$

- $X \sim \nu$; ν is a probability measure (abs. cont.) in $\mathbb{R}^d$ $(d \geq 1)$
- **Reference dist.**: $U \sim \mu$ on $\mathcal{S} \subset \mathbb{R}^d$ $(\mu = \text{Unif}([0,1]^d),\ N(0, I_d))$
- Find OT map T s.t. $T(X) \stackrel{d}{=} U \sim \mu$ (μ abs. cont.)

Population rank function [Hallin (2017), Chernozhukov et al. (2017)]

If $\mathbb{E}_\nu \|X\|^2 < \infty$, rank function $R : \mathbb{R}^d \to \mathcal{S}$ is the OT map s.t.

$$R := \underset{T: T\#\nu = \mu}{\arg\min}\ \mathbb{E}_\nu \|X - T(X)\|^2$$

Properties of population rank function [Brenier (1991), McCann (1995)]

- $R(\cdot)$ characterizes distribution: $R_1(x) = R_2(x)\ \forall\, x \in \mathbb{R}^d$ iff $P_1 = P_2$

- $X \sim \nu$; ν is a probability measure (abs. cont.) in $\mathbb{R}^d$ $(d \geq 1)$
- **Reference dist.**: $U \sim \mu$ on $\mathcal{S} \subset \mathbb{R}^d$ $(\mu = \text{Unif}([0,1]^d),\ N(0, I_d))$
- Find OT map T s.t. $T(X) \stackrel{d}{=} U \sim \mu$ (μ abs. cont.)

Population rank function [Hallin (2017), Chernozhukov et al. (2017)]

If $\mathbb{E}_\nu \|X\|^2 < \infty$, rank function $R : \mathbb{R}^d \to \mathcal{S}$ is the OT map s.t.

$$R := \underset{T: T\#\nu = \mu}{\arg\min}\ \mathbb{E}_\nu \|X - T(X)\|^2$$

Properties of population rank function [Brenier (1991), McCann (1995)]

- $R(\cdot)$ characterizes distribution: $R_1(x) = R_2(x)\ \forall\ x \in \mathbb{R}^d$ iff $P_1 = P_2$
- $R(\cdot)$ is invertible, i.e., there exists unique $Q(\cdot)$ s.t.

$$R \circ Q(u) = u \ \ (\mu\text{-a.e.}) \qquad \text{and} \qquad Q \circ R(x) = x \ \ (\nu\text{-a.e.})$$

- $X \sim \nu$; ν is a probability measure (abs. cont.) in $\mathbb{R}^d$ $(d \geq 1)$
- **Reference dist.**: $U \sim \mu$ on $\mathcal{S} \subset \mathbb{R}^d$ $(\mu = \text{Unif}([0,1]^d),\ N(0, I_d))$
- Find OT map T s.t. $T(X) \stackrel{d}{=} U \sim \mu$ (μ abs. cont.)

Population rank function [Hallin (2017), Chernozhukov et al. (2017)]

If $\mathbb{E}_\nu \|X\|^2 < \infty$, rank function $R : \mathbb{R}^d \to \mathcal{S}$ is the OT map s.t.

$$R := \operatorname*{arg\,min}_{T:T\#\nu=\mu} \mathbb{E}_\nu \|X - T(X)\|^2$$

Properties of population rank function [Brenier (1991), McCann (1995)]

- $R(\cdot)$ characterizes distribution: $R_1(x) = R_2(x)\ \forall\, x \in \mathbb{R}^d$ iff $P_1 = P_2$
- $R(\cdot)$ is invertible, i.e., there exists unique $Q(\cdot)$ s.t.

$$R \circ Q(u) = u \ \ (\mu\text{-a.e.}) \qquad \text{and} \qquad Q \circ R(x) = x \ \ (\nu\text{-a.e.})$$

- Both $R(\cdot)$ and $Q(\cdot)$ and gradients of convex functions

- If $\mathbb{E}_\nu\|X\|^2 < \infty$, the population rank function $R(\cdot)$ is defined as

$$R := \operatorname*{arg\,min}_{T: T\#\nu=\mu} \mathbb{E}_\nu\|X - T(X)\|^2 \tag{1}$$

- Even when $\mathbb{E}_\nu\|X\|^2 = +\infty$, $R(\cdot)$ can still be defined

- If $\mathbb{E}_\nu \|X\|^2 < \infty$, the population rank function $R(\cdot)$ is defined as

$$R := \underset{T: T\#\nu=\mu}{\arg\min}\ \mathbb{E}_\nu \|X - T(X)\|^2 \tag{1}$$

- Even when $\mathbb{E}_\nu \|X\|^2 = +\infty$, $R(\cdot)$ can still be defined

Characterization of the population rank function [McCann (1995)]

Suppose $X \sim \nu$ abs. cont. on $\mathbb{R}^d$. Then $\exists$ ν-a.e. unique meas. mapping $R : \mathbb{R}^d \to \mathcal{S}$, transporting ν to μ (i.e., $R\#\nu = \mu$), of the form

$$R(x) = \nabla\varphi(x), \qquad \text{for } \nu\text{-a.e. } x, \tag{2}$$

where $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a convex function (cf. when $d = 1$).

- If $\mathbb{E}_\nu \|X\|^2 < \infty$, the population rank function $R(\cdot)$ is defined as

$$R := \underset{T:T\#\nu=\mu}{\arg\min}\ \mathbb{E}_\nu \|X - T(X)\|^2 \tag{1}$$

- Even when $\mathbb{E}_\nu \|X\|^2 = +\infty$, $R(\cdot)$ can still be defined

Characterization of the population rank function [McCann (1995)]

Suppose $X \sim \nu$ abs. cont. on $\mathbb{R}^d$. Then $\exists$ ν-a.e. unique meas. mapping $R : \mathbb{R}^d \to \mathcal{S}$, transporting ν to μ (i.e., $R\#\nu = \mu$), of the form

$$R(x) = \nabla\varphi(x), \qquad \text{for } \nu\text{-a.e. } x, \tag{2}$$

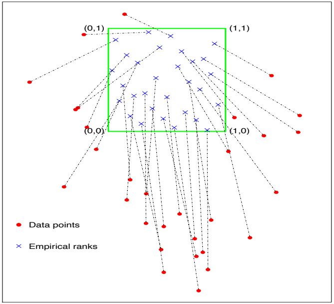
where $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a convex function (cf. when $d = 1$).

Moreover, when $\mathbb{E}_\nu \|X\|^2 < \infty$, $R(\cdot)$ as defined in (2) also satisfies (1).

- $X_1, \ldots, X_n \overset{iid}{\sim} \nu$ on $\mathbb{R}^d$ (abs. cont.); $\mu \sim \text{Unif}([0,1]^d)$
- Empirical rank map $\hat{R}_n$: $\{X_1, \ldots, X_n\} \to \{c_1, \ldots, c_n\} \subset [0,1]^d$ — sequence of "uniform-like" points (or quasi-Monte Carlo sequence)

- $X_1, \ldots, X_n \overset{iid}{\sim} \nu$ on $\mathbb{R}^d$ (abs. cont.); $\mu \sim \text{Unif}([0,1]^d)$

- Empirical rank map $\hat{R}_n$: $\{X_1, \ldots, X_n\} \to \{c_1, \ldots, c_n\} \subset [0,1]^d$ — sequence of "uniform-like" points (or quasi-Monte Carlo sequence)



- Sample multivariate rank map $\hat{R}_n$ is defined as the OT map s.t.

$$\hat{R}_n := \underset{T : T\#\nu_n = \mu_n}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \|X_i - T(X_i)\|^2$$

where T transports $\nu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ to $\mu_n := \frac{1}{n} \sum_{j=1}^{n} \delta_{c_j}$

- Assignment problem (can be reduced to a linear program — $O(n^3)$)

Distribution-free property [Hallin (2017), Deb and S. (2019)]

Suppose that $X_1, \ldots, X_n$ iid on $\mathbb{R}^d$ with abs. cont. distribution. Then,

$$(\hat{R}_n(X_1), \ldots, \hat{R}_n(X_n))$$

is uniformly distributed over the $n!$ permutations of $\{c_1, \ldots, c_n\}$.

Distribution-free property [Hallin (2017), Deb and S. (2019)]

Suppose that $X_1, \ldots, X_n$ iid on $\mathbb{R}^d$ with abs. cont. distribution. Then,

$$(\hat{R}_n(X_1), \ldots, \hat{R}_n(X_n))$$

is uniformly distributed over the $n!$ permutations of $\{c_1, \ldots, c_n\}$.

The first step to obtaining distribution-free tests [Hallin et al. (2021)]

Distribution-free property [Hallin (2017), Deb and S. (2019)]

Suppose that $X_1, \ldots, X_n$ iid on $\mathbb{R}^d$ with abs. cont. distribution. Then,

$$(\hat{R}_n(X_1), \ldots, \hat{R}_n(X_n))$$

is uniformly distributed over the $n!$ permutations of $\{c_1, \ldots, c_n\}$.

The first step to obtaining distribution-free tests [Hallin et al. (2021)]

Consistency [Deb and S. (2019), Deb, Bhattacharya and S. (2021)]

$X_1, \ldots, X_n$ iid ν (abs. cont.). If $\mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{c_j} \xrightarrow{d} \mu$ (abs. cont.), then

$$\frac{1}{n}\sum_{i=1}^{n} \|\hat{R}_n(X_i) - R(X_i)\|^2 \xrightarrow{p} 0$$

Regularity to the empirical multivariate rank/OT map

Question: What is the **rate of convergence** of $\hat{R}_n$?

- **Recall**: $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $\mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j}$
- **OT maps**: $R\#\nu = \mu$, $\hat{R}_n\#\nu_n = \mu_n$
- Assume $\int \|x\|^2 \, d\nu(x) < \infty$, $\int \|y\|^2 \, d\mu(y) < \infty$

Rate of convergence [Deb, Ghosal and S. (2021)] Proof of this result

Suppose the population rank map $R(\cdot)$ is Lipschitz. Then, under appropriate conditions on μ_n,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\hat{R}_n(X_i) - R(X_i)\|^2\right] \lesssim \begin{cases} n^{-1/2} & d = 2, 3, \\ n^{-1/2}\log n & d = 4, \\ n^{-2/d} & d > 4. \end{cases}$$

Question: What is the **rate of convergence** of $\hat{R}_n$?

- **Recall**: $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $\mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j}$
- **OT maps**: $R\#\nu = \mu$, $\hat{R}_n\#\nu_n = \mu_n$
- Assume $\int \|x\|^2\, d\nu(x) < \infty$, $\int \|y\|^2\, d\mu(y) < \infty$

Rate of convergence [Deb, Ghosal and S. (2021)] Proof of this result

Suppose the population rank map $R(\cdot)$ is Lipschitz. Then, under appropriate conditions on μ_n,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\hat{R}_n(X_i) - R(X_i)\|^2\right] \lesssim \begin{cases} n^{-1/2} & d = 2, 3, \\ n^{-1/2}\log n & d = 4, \\ n^{-2/d} & d > 4. \end{cases}$$

This is the optimal rate for $d \geq 4$ [Hütter & Rigollet (2021)]

Question: What is the **rate of convergence** of $\hat{R}_n$?

- **Recall**: $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $\mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j}$
- **OT maps**: $R\#\nu = \mu$, $\hat{R}_n\#\nu_n = \mu_n$
- Assume $\int \|x\|^2\, d\nu(x) < \infty$, $\int \|y\|^2\, d\mu(y) < \infty$

Rate of convergence [Deb, Ghosal and S. (2021)] Proof of this result

Suppose the population rank map $R(\cdot)$ is Lipschitz. Then, under appropriate conditions on μ_n,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\hat{R}_n(X_i) - R(X_i)\|^2\right] \lesssim \begin{cases} n^{-1/2} & d = 2, 3, \\ n^{-1/2}\log n & d = 4, \\ n^{-2/d} & d > 4. \end{cases}$$

This is the optimal rate for $d \geq 4$ [Hütter & Rigollet (2021)]

Connection to estimation of the OT map R ($R\#\nu = \mu$)

References: Hütter & Rigollet (2021), Ghosal and S. (2019), Manole et al. (2021), Pooladian and Niles-Weed (2022), Gunsilius (2022), ...

Outline

Multivariate two-sample goodness-of-fit test

Testing for equality of two multivariate distributions

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

Multivariate two-sample goodness-of-fit test

Testing for equality of two multivariate distributions

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

- **Hotelling T^2 statistic** [Hotelling (1931)]: The multivariate analogue of Student's t-statistic, given by

$$\mathrm{T}^2_{m,n} := \frac{mn}{m+n} \left(\bar{X} - \bar{Y}\right)^\top S_{m,n}^{-1} \left(\bar{X} - \bar{Y}\right);$$

where $S_{m,n}$ is pooled covariance matrix

Multivariate two-sample goodness-of-fit test

Testing for equality of two multivariate distributions

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

- **Hotelling T^2 statistic** [Hotelling (1931)]: The multivariate analogue of Student's t-statistic, given by

$$\mathrm{T}^2_{m,n} := \frac{mn}{m+n} \left(\bar{X} - \bar{Y}\right)^\top S_{m,n}^{-1} \left(\bar{X} - \bar{Y}\right);$$

where $S_{m,n}$ is pooled covariance matrix

- Reject H_0 iff $\mathrm{T}^2_{m,n} > c_\alpha$ [asymp. cut-off c_α: $(1-\alpha)$ quantile of χ^2_d]
- **Consistency:** $\mathbb{P}(\mathrm{T}^2_{m,n} > c_\alpha) \to 1$ when $\mathbb{E}[X_1] \neq \mathbb{E}[Y_1]$

Multivariate two-sample goodness-of-fit test

Testing for equality of two multivariate distributions

- **Data**: $\{X_i\}_{i=1}^m$ iid P_1 on $\mathbb{R}^d$; $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$

- Test if the two samples came from the same distribution, i.e.,

$$\mathrm{H}_0 : P_1 = P_2 \qquad \text{versus} \qquad \mathrm{H}_1 : P_1 \neq P_2$$

- **Hotelling T^2 statistic** [Hotelling (1931)]: The multivariate analogue of Student's t-statistic, given by

$$\mathrm{T}^2_{m,n} := \frac{mn}{m+n} \left(\bar{X} - \bar{Y}\right)^\top S_{m,n}^{-1} \left(\bar{X} - \bar{Y}\right);$$

  where $S_{m,n}$ is pooled covariance matrix

- Reject H_0 iff $\mathrm{T}^2_{m,n} > c_\alpha$ [asymp. cut-off c_α: $(1-\alpha)$ quantile of χ^2_d]
- **Consistency:** $\mathbb{P}(\mathrm{T}^2_{m,n} > c_\alpha) \to 1$ when $\mathbb{E}[X_1] \neq \mathbb{E}[Y_1]$

Question: What is the distribution-free analogue of Hotelling's T^2?

Data: $\{X_i\}_{i=1}^m$ iid P_1 (abs. cont.), $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
Reference dist.: μ on $\mathcal{S} \subset \mathbb{R}^d$ (abs. cont.; $\mu = \text{Unif}([0,1]^d)$ or $N(0, I_d)$)

Proposed test [Deb, Bhattacharya and S. (2021)]

- Joint rank map: The sample ranks of the pooled observations:

$$\hat{R}_{m,n} : \{X_1, \ldots, X_m, Y_1, \ldots, Y_n\} \to \{c_1, \ldots, c_{m+n}\} \subset \mathcal{S}$$

- **Rank Hotelling**: $\mathrm{RT}^2_{m,n} := \mathrm{T}^2_{m,n}\left(\{\hat{R}_{m,n}(X_i)\}, \{\hat{R}_{m,n}(Y_j)\}\right)$

Data: $\{X_i\}_{i=1}^m$ iid P_1 (abs. cont.), $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
Reference dist.: μ on $\mathcal{S} \subset \mathbb{R}^d$ (abs. cont.; $\mu = \text{Unif}([0,1]^d)$ or $N(0, I_d)$)

Proposed test [Deb, Bhattacharya and S. (2021)]

- Joint rank map: The sample ranks of the pooled observations:

$$\hat{R}_{m,n} : \{X_1, \dots, X_m, Y_1, \dots, Y_n\} \to \{c_1, \dots, c_{m+n}\} \subset \mathcal{S}$$

- **Rank Hotelling**: $\mathrm{RT}^2_{m,n} := \mathrm{T}^2_{m,n}\left(\{\hat{R}_{m,n}(X_i)\}, \{\hat{R}_{m,n}(Y_j)\}\right)$

This yields the Wilcoxon rank-sum test when applied to the t-test

General principle [Deb and S. (2019)]

Start with any "good" test & replace X_i's & Y_j's with pooled OT ranks

Data: $\{X_i\}_{i=1}^m$ iid P_1 (abs. cont.), $\{Y_j\}_{j=1}^n$ iid P_2 on $\mathbb{R}^d$, $d \geq 1$
Reference dist.: μ on $\mathcal{S} \subset \mathbb{R}^d$ (abs. cont.; $\mu = \text{Unif}([0,1]^d)$ or $N(0, I_d)$)

Proposed test [Deb, Bhattacharya and S. (2021)]

- Joint rank map: The sample ranks of the pooled observations:

$$\hat{R}_{m,n} : \{X_1, \ldots, X_m, Y_1, \ldots, Y_n\} \to \{c_1, \ldots, c_{m+n}\} \subset \mathcal{S}$$

- **Rank Hotelling**: $\mathrm{RT}^2_{m,n} := \mathrm{T}^2_{m,n}\left(\{\hat{R}_{m,n}(X_i)\}, \{\hat{R}_{m,n}(Y_j)\}\right)$

This yields the Wilcoxon rank-sum test when applied to the t-test

General principle [Deb and S. (2019)]

Start with any "good" test & replace X_i's & Y_j's with pooled OT ranks

Distribution-freeness: Under H_0, the dist. of $\mathrm{RT}^2_{m,n}$ is free of $P_1 \equiv P_2$

The only known efficient, computationally feasible, distribution-free analogue of Hotelling's T^2; cf. Puri & Sen (1971), Hallin et al. (2020), ...

Rank Hotelling test

Rank Hotelling test: $\phi_{m,n} \equiv \mathbf{1}\{\mathrm{RT}^2_{m,n} > \kappa_\alpha^{(m,n)}\}$ — distribution-free

$\kappa_\alpha^{(m,n)}$ depends on c_j's, m, n and d

Rank Hotelling test

Rank Hotelling test: $\phi_{m,n} \equiv \mathbf{1}\{\mathrm{RT}^2_{m,n} > \kappa_\alpha^{(m,n)}\}$ — distribution-free

$\kappa_\alpha^{(m,n)}$ depends on c_j's, m, n and d

Asymptotic null distribution [Deb, Bhattacharya, and S. (2021)]

Under H_0, if $\mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{c_j} \xrightarrow{d} \mu$, then,

$$\mathrm{RT}^2_{m,n} \xrightarrow{d} \chi^2_d \qquad \text{as} \quad \min\{m,n\} \to \infty.$$

The choice of the c_j's have no effect for large m, n

Rank Hotelling test

Rank Hotelling test: $\phi_{m,n} \equiv \mathbf{1}\{\mathrm{RT}^2_{m,n} > \kappa_\alpha^{(m,n)}\}$ — distribution-free

$\kappa_\alpha^{(m,n)}$ depends on c_j's, m, n and d

Asymptotic null distribution [Deb, Bhattacharya, and S. (2021)]

Under H_0, if $\mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{c_j} \xrightarrow{d} \mu$, then,

$$\mathrm{RT}^2_{m,n} \xrightarrow{d} \chi^2_d \qquad \text{as} \quad \min\{m,n\} \to \infty.$$

The choice of the c_j's have no effect for large m, n

Consistency [Deb, Bhattacharya, and S. (2021)]

Under location shift alternatives ($P_1 \neq P_2$), if (i) $\mu_n \xrightarrow{d} \mu$, and (ii) $\frac{m}{m+n} \to \lambda \in (0,1)$, then,

$$\lim_{m,n\to\infty} \mathbb{E}_{\mathrm{H}_1}[\phi_{m,n}] = 1.$$

Rank Hotelling test

Rank Hotelling test: $\phi_{m,n} \equiv \mathbf{1}\{\mathrm{RT}^2_{m,n} > \kappa_\alpha^{(m,n)}\}$ — distribution-free

$\kappa_\alpha^{(m,n)}$ depends on c_j's, m, n and d

Asymptotic null distribution [Deb, Bhattacharya, and S. (2021)]

Under H_0, if $\mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j} \xrightarrow{d} \mu$, then,

$$\mathrm{RT}^2_{m,n} \xrightarrow{d} \chi^2_d \qquad \text{as} \quad \min\{m,n\} \to \infty.$$

The choice of the c_j's have no effect for large m, n

Consistency [Deb, Bhattacharya, and S. (2021)]

Under location shift alternatives ($P_1 \neq P_2$), if (i) $\mu_n \xrightarrow{d} \mu$, and (ii) $\frac{m}{m+n} \to \lambda \in (0,1)$, then,

$$\lim_{m,n\to\infty} \mathbb{E}_{\mathrm{H}_1}[\phi_{m,n}] = 1.$$

Question: How does the efficiency of $\mathrm{RT}^2_{m,n}$ compare with $\mathrm{T}^2_{m,n}$?

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- $X_1, \dots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \dots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$; $\frac{m}{N} \approx \lambda \in (0, 1)$
- $\{P_\theta\}_{\theta \in \Theta \subset \mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- $X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$; $\frac{m}{N} \approx \lambda \in (0, 1)$
- $\{P_\theta\}_{\theta \in \Theta \subset \mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family

- **Test** $\mathrm{H}_0 : \theta_2 = \theta_1 \qquad$ vs. $\qquad \mathrm{H}_1 : \theta_2 = \theta_1 + \Delta; \quad \Delta \to 0$

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- $X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$; $\;\frac{m}{N} \approx \lambda \in (0,1)$
- $\{P_\theta\}_{\theta \in \Theta \subset \mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family

- **Test** $\mathrm{H}_0 : \theta_2 = \theta_1 \quad$ vs. $\quad \mathrm{H}_1 : \theta_2 = \theta_1 + \Delta; \quad \Delta \to 0$
- Fix $\alpha \in (0,1)$ (level) and $\beta \in (\alpha, 1)$ (power)

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- $X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$; $\frac{m}{N} \approx \lambda \in (0,1)$
- $\{P_\theta\}_{\theta \in \Theta \subset \mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family

- **Test** $\mathrm{H}_0 : \theta_2 = \theta_1 \qquad$ vs. $\qquad \mathrm{H}_1 : \theta_2 = \theta_1 + \Delta; \quad \Delta \to 0$
- Fix $\alpha \in (0,1)$ (level) and $\beta \in (\alpha, 1)$ (power)
- Let $N_\Delta(T.) \equiv N_\Delta$ denote the minimum number of samples s.t:

$$\mathbb{E}_{\mathrm{H}_0}[T_{N_\Delta}] = \alpha \qquad \text{and} \qquad \mathbb{E}_{\mathrm{H}_1}[T_{N_\Delta}] \geq \beta$$

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- $X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$; $\frac{m}{N} \approx \lambda \in (0,1)$
- $\{P_\theta\}_{\theta \in \Theta \subset \mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family

- **Test** $\mathrm{H}_0 : \theta_2 = \theta_1 \quad$ vs. $\quad \mathrm{H}_1 : \theta_2 = \theta_1 + \Delta; \quad \Delta \to 0$
- Fix $\alpha \in (0,1)$ (level) and $\beta \in (\alpha, 1)$ (power)
- Let $N_\Delta(T_\cdot) \equiv N_\Delta$ denote the minimum number of samples s.t:

$$\mathbb{E}_{\mathrm{H}_0}[T_{N_\Delta}] = \alpha \qquad \text{and} \qquad \mathbb{E}_{\mathrm{H}_1}[T_{N_\Delta}] \geq \beta$$

- The asymptotic (Pitman) efficiency of S_N w.r.t T_N is given by

$$\mathrm{ARE}\,(S_N, T_N) := \lim_{\Delta \to 0} \frac{N_\Delta(T_\cdot)}{N_\Delta(S_\cdot)}$$

- **Question**: How to compare two consistent tests S_N and T_N?
- **Asymptotic relative (Pitman) efficiency** (ARE) [Pitman (1948), Serfling (1980), Lehmann & Romano (2005), van der Vaart (1998)]

- $X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$; $\frac{m}{N} \approx \lambda \in (0,1)$
- $\{P_\theta\}_{\theta \in \Theta \subset \mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family

- **Test** $\mathrm{H}_0 : \theta_2 = \theta_1 \qquad$ vs. $\qquad \mathrm{H}_1 : \theta_2 = \theta_1 + \Delta; \quad \Delta \to 0$
- Fix $\alpha \in (0,1)$ (level) and $\beta \in (\alpha, 1)$ (power)
- Let $N_\Delta(T_.) \equiv N_\Delta$ denote the minimum number of samples s.t:

$$\mathbb{E}_{\mathrm{H}_0}[T_{N_\Delta}] = \alpha \qquad \text{and} \qquad \mathbb{E}_{\mathrm{H}_1}[T_{N_\Delta}] \geq \beta$$

- The asymptotic (Pitman) efficiency of S_N w.r.t T_N is given by

$$\mathrm{ARE}\,(S_N, T_N) := \lim_{\Delta \to 0} \frac{N_\Delta(T_.)}{N_\Delta(S_.)}$$

$\mathrm{ARE}\,(S_N, T_N)$ can depend on α and β, but in some cases it doesn't!

Hotelling T^2: $\quad \mathrm{T}^2_{m,n}(\{X_i\},\{Y_j\}) = \frac{mn}{m+n}\left(\bar{X}-\bar{Y}\right)^\top S^{-1}_{m,n}\left(\bar{X}-\bar{Y}\right)$

Rank Hotelling: $\quad \mathrm{RT}^2_{m,n} = \mathrm{T}^2_{m,n}\left(\{\hat{R}_{m,n}(X_i)\},\{\hat{R}_{m,n}(Y_j)\}\right)$

- $X_1,\ldots,X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1,\ldots,Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m+n$
- $\{P_\theta\}_{\theta\in\Theta\subset\mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family
- Consider $\mathrm{H}_0 : \theta_2 = \theta_1$ vs. $\mathrm{H}_1 : \theta_2 = \theta_1 + hN^{-1/2}$; $\quad h \neq 0 \in \mathbb{R}^p$

$\mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right)$ can be derived under the above alternatives

Hotelling T^2: $\quad \mathrm{T}^2_{m,n}(\{X_i\},\{Y_j\}) = \frac{mn}{m+n}\left(\bar{X}-\bar{Y}\right)^\top S^{-1}_{m,n}\left(\bar{X}-\bar{Y}\right)$

Rank Hotelling: $\quad \mathrm{RT}^2_{m,n} = \mathrm{T}^2_{m,n}\left(\{\hat{R}_{m,n}(X_i)\},\{\hat{R}_{m,n}(Y_j)\}\right)$

- $X_1,\ldots,X_m \stackrel{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1,\ldots,Y_n \stackrel{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m+n$
- $\{P_\theta\}_{\theta\in\Theta\subset\mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family
- Consider $\mathrm{H}_0 : \theta_2 = \theta_1$ vs. $\mathrm{H}_1 : \theta_2 = \theta_1 + hN^{-1/2}$; $\quad h \neq 0 \in \mathbb{R}^p$

$\mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right)$ can be derived under the above alternatives

Some observations

- Expression of $\mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right)$ does not depend on α and β
- Asymp. dist. of $\mathrm{RT}^2_{m,n}$ can depend on choice of μ (reference dist.)

Hotelling T^2: $\mathrm{T}^2_{m,n}(\{X_i\},\{Y_j\}) = \frac{mn}{m+n}\left(\bar{X}-\bar{Y}\right)^\top S^{-1}_{m,n}\left(\bar{X}-\bar{Y}\right)$

Rank Hotelling: $\mathrm{RT}^2_{m,n} = \mathrm{T}^2_{m,n}\left(\{\hat{R}_{m,n}(X_i)\},\{\hat{R}_{m,n}(Y_j)\}\right)$

- $X_1,\ldots,X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1,\ldots,Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $N = m+n$
- $\{P_\theta\}_{\theta\in\Theta\subset\mathbb{R}^p}$: "smooth" (satisfies DQM) parametric family
- Consider $\mathrm{H}_0 : \theta_2 = \theta_1$ vs. $\mathrm{H}_1 : \theta_2 = \theta_1 + hN^{-1/2}$; $h \neq 0 \in \mathbb{R}^p$

$\mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right)$ can be derived under the above alternatives

Some observations

- Expression of $\mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right)$ does not depend on α and β
- Asymp. dist. of $\mathrm{RT}^2_{m,n}$ can depend on choice of μ (reference dist.)

Can we lower bound ARE for sub-classes of multivariate dists., i.e.,

$$\min_{\mathcal{F}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = ??$$

$X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$

Independent coordinates case (location shift family)

$\mathcal{F}_{\text{ind}} = \{P_\theta\}_{\theta \in \Theta}$ has density $p_\theta(z_1, \ldots, z_d) = \prod_{i=1}^d f_i(z_i - \theta_i)$, $\theta \in \mathbb{R}^d$

Theorem [Deb, Bhattacharya, and S. (2021)]

Suppose $\frac{m}{N} \to \lambda \in (0,1)$. If $\mu_N := \frac{1}{N} \sum_{j=1}^N \delta_{c_j} \xrightarrow{d} \mathrm{Unif}([0,1]^d) \equiv \mu$, then

$$\min_{\mathcal{F}_{\text{ind}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 0.864.$$

$X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$

Independent coordinates case (location shift family)

$\mathcal{F}_{\mathrm{ind}} = \{P_\theta\}_{\theta \in \Theta}$ has density $p_\theta(z_1, \ldots, z_d) = \prod_{i=1}^d f_i(z_i - \theta_i)$, $\theta \in \mathbb{R}^d$

Theorem [Deb, Bhattacharya, and S. (2021)]

Suppose $\frac{m}{N} \to \lambda \in (0,1)$. If $\mu_N := \frac{1}{N} \sum_{j=1}^N \delta_{c_j} \xrightarrow{d} \mathrm{Unif}([0,1]^d) \equiv \mu$, then

$$\min_{\mathcal{F}_{\mathrm{ind}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 0.864.$$

If $\mu_N \xrightarrow{d} N(0, I_d) \equiv \mu$, then

$$\min_{\mathcal{F}_{\mathrm{ind}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 1.$$

$X_1, \ldots, X_m \overset{iid}{\sim} \mathrm{P}_{\theta_1}$ & $Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{P}_{\theta_2}$; $\quad N = m + n$

Independent coordinates case (location shift family)

$\mathcal{F}_{\mathrm{ind}} = \{P_\theta\}_{\theta \in \Theta}$ has density $p_\theta(z_1, \ldots, z_d) = \prod_{i=1}^d f_i(z_i - \theta_i)$, $\theta \in \mathbb{R}^d$

Theorem [Deb, Bhattacharya, and S. (2021)]

Suppose $\frac{m}{N} \to \lambda \in (0,1)$. If $\mu_N := \frac{1}{N}\sum_{j=1}^N \delta_{c_j} \xrightarrow{d} \mathrm{Unif}([0,1]^d) \equiv \mu$, then

$$\min_{\mathcal{F}_{\mathrm{ind}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 0.864.$$

If $\mu_N \xrightarrow{d} N(0, I_d) \equiv \mu$, then

$$\min_{\mathcal{F}_{\mathrm{ind}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 1.$$

- Generalizes Hodges & Lehmann (1956), Chernoff & Savage (1958)
- ARE can be arbitrarily large (can tend to $+\infty$) for heavy tailed dists.

Elliptically symmetric distributions

$\mathcal{F}_{\mathrm{ell}} = \{P_\theta\}_{\theta\in\Theta}$ is class of elliptically symmetric distributions on $\mathbb{R}^d$, i.e.,

$$p_\theta(x) \propto (\det(\Sigma))^{-\frac{1}{2}} \underline{f}\left((x-\theta)^\top \Sigma^{-1}(x-\theta)\right), \quad \text{for all } x \in \mathbb{R}^d$$

Elliptically symmetric distributions

$\mathcal{F}_{\mathrm{ell}} = \{P_\theta\}_{\theta\in\Theta}$ is class of elliptically symmetric distributions on $\mathbb{R}^d$, i.e.,

$$p_\theta(x) \propto (\det(\Sigma))^{-\frac{1}{2}} \underline{f}\left((x-\theta)^\top \Sigma^{-1}(x-\theta)\right), \quad \text{for all } x \in \mathbb{R}^d$$

Theorem [Deb, Bhattacharya, and S. (2021)]

Suppose: (i) $\mu_N \xrightarrow{d} N(0, I_d) \equiv \mu$, (ii) $\frac{m}{N} \to \lambda \in (0,1)$. Then,

$$\min_{\mathcal{F}_{\mathrm{ell}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 1.$$

- This generalizes the famous result of Chernoff and Savage (1958)

Elliptically symmetric distributions

$\mathcal{F}_{\mathrm{ell}} = \{P_\theta\}_{\theta\in\Theta}$ is class of elliptically symmetric distributions on $\mathbb{R}^d$, i.e.,

$$p_\theta(x) \propto (\det(\Sigma))^{-\frac{1}{2}} \underline{f}\left((x-\theta)^\top \Sigma^{-1}(x-\theta)\right), \quad \text{for all } x \in \mathbb{R}^d$$

Theorem [Deb, Bhattacharya, and S. (2021)]

Suppose: (i) $\mu_N \xrightarrow{d} N(0, I_d) \equiv \mu$, (ii) $\frac{m}{N} \to \lambda \in (0,1)$. Then,

$$\min_{\mathcal{F}_{\mathrm{ell}}} \mathrm{ARE}\left(\mathrm{RT}^2_{m,n}, \mathrm{T}^2_{m,n}\right) = 1.$$

- This generalizes the famous result of Chernoff and Savage (1958)
- Lower bounds can also be obtained for other sub-classes of multivariate distributions (e.g., the model for ICA)

Outline

Testing for Mutual Independence

- $(X, Y) \sim P$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$; $\qquad d_1, d_2 \geq 1$
- **Data**: n iid observations $\{(X_i, Y_i)\}_{i=1}^n$ from P
- Test if X is independent of Y, i.e.,

$$\mathrm{H}_0 : X \perp\!\!\!\perp Y \qquad \text{versus} \qquad \mathrm{H}_1 : X \not\perp\!\!\!\perp Y$$

Testing for Mutual Independence

- $(X, Y) \sim P$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$; $\qquad d_1, d_2 \geq 1$
- **Data**: n iid observations $\{(X_i, Y_i)\}_{i=1}^n$ from P
- Test if X is independent of Y, i.e.,

$$\mathrm{H}_0 : X \perp\!\!\!\perp Y \qquad \text{versus} \qquad \mathrm{H}_1 : X \not\perp\!\!\!\perp Y$$

- When $d_1 = d_2 = 1$: Pearson (1904), Spearman (1904), Kendall (1938), Hoeffding (1948), Blomqvist (1950), Blum et al. (1961), Rosenblatt (1975), Feuerverger (1993), ...
- When $d_1 > 1$ or $d_2 > 1$: Friedman and Rafsky (1979), Székely et al. (2007), Gretton et al. (2008), Oja (2010), Heller et al. (2013), Biswas et al. (2016), Berrett and Samworth (2019), ...

- $(X, Y) \sim P$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $\quad X \sim P_X$, $\quad Y \sim P_Y$, $\quad d_1, d_2 \geq 1$
- **Data**: $\{(X_i, Y_i) : 1 \leq i \leq n\}$ iid P
- **Test**: $\mathrm{H}_0 : X \perp\!\!\!\perp Y$ vs. $\mathrm{H}_1 : X \not\!\perp\!\!\!\perp Y$

- $(X, Y) \sim P$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $\quad X \sim P_X$, $\quad Y \sim P_Y$, $\quad d_1, d_2 \geq 1$
- **Data**: $\{(X_i, Y_i) : 1 \leq i \leq n\}$ iid P
- **Test**: $\quad \mathrm{H}_0 : X \perp\!\!\!\perp Y \quad$ vs. $\quad \mathrm{H}_1 : X \not\perp\!\!\!\perp Y$

Distance Covariance [Szekely et al. (2007, 2009), Feuerverger (1993)]

- Characterizes independence: $\mathrm{dCov}(X, Y) = 0 \quad$ iff $\quad X \perp\!\!\!\perp Y$

- $(X, Y) \sim P$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $\quad X \sim P_X$, $\quad Y \sim P_Y$, $\quad d_1, d_2 \geq 1$
- **Data**: $\{(X_i, Y_i) : 1 \leq i \leq n\}$ iid P
- **Test**: $\quad \mathrm{H}_0 : X \perp\!\!\!\perp Y \quad$ vs. $\quad \mathrm{H}_1 : X \not\perp\!\!\!\perp Y$

Distance Covariance [Szekely et al. (2007, 2009), Feuerverger (1993)]

- Characterizes independence: $\mathrm{dCov}(X, Y) = 0 \quad$ iff $\quad X \perp\!\!\!\perp Y$
- **Sample distance covariance**: dCov_n (V-statistic)
- **Distance covariance test**: Reject H_0 if

$$\mathrm{dCov}_n(\{(X_i, Y_i)\}_{i=1}^n) > c_\alpha$$

- Critical value c_α depends on n, P_X, P_Y! (can use permutation test)

- $(X, Y) \sim P$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $\quad X \sim P_X$, $\quad Y \sim P_Y$, $\quad d_1, d_2 \geq 1$
- **Data**: $\{(X_i, Y_i) : 1 \leq i \leq n\}$ iid P
- **Test**: $\quad \mathrm{H}_0 : X \perp\!\!\!\perp Y \quad$ vs. $\quad \mathrm{H}_1 : X \not\perp\!\!\!\perp Y$

Distance Covariance [Szekely et al. (2007, 2009), Feuerverger (1993)]

- Characterizes independence: $\mathrm{dCov}(X, Y) = 0 \quad$ iff $\quad X \perp\!\!\!\perp Y$
- **Sample distance covariance**: dCov_n (V-statistic)
- **Distance covariance test**: Reject H_0 if

$$\mathrm{dCov}_n(\{(X_i, Y_i)\}_{i=1}^n) > c_\alpha$$

- Critical value c_α depends on n, P_X, P_Y! (can use permutation test)

Question: What is the distribution-free analogue of distance covariance?

- **Test**: $\mathrm{H}_0 : X \perp\!\!\!\perp Y$ vs. $\mathrm{H}_1 : X \not\perp\!\!\!\perp Y$
- Take $\mu_1 = \text{Uniform}([0,1]^{d_1})$ and $\mu_2 = \text{Uniform}([0,1]^{d_2})$

Rank distance covariance [Deb and S. (2019)]

- OT rank of X_i: $\hat{R}_n^X : \{X_1, \ldots, X_n\} \to \{c_1^{(1)}, \ldots, c_n^{(1)}\} \subset [0,1]^{d_1}$
- OT rank of Y_i: $\hat{R}_n^Y : \{Y_1, \ldots, Y_n\} \to \{c_1^{(2)}, \ldots, c_n^{(2)}\} \subset [0,1]^{d_2}$

- **Test**: $\mathrm{H}_0 : X \perp\!\!\!\perp Y$ vs. $\mathrm{H}_1 : X \not\perp\!\!\!\perp Y$
- Take $\mu_1 = \text{Uniform}([0,1]^{d_1})$ and $\mu_2 = \text{Uniform}([0,1]^{d_2})$

Rank distance covariance [Deb and S. (2019)]

- OT rank of X_i: $\hat{R}_n^X : \{X_1, \ldots, X_n\} \to \{c_1^{(1)}, \ldots, c_n^{(1)}\} \subset [0,1]^{d_1}$
- OT rank of Y_i: $\hat{R}_n^Y : \{Y_1, \ldots, Y_n\} \to \{c_1^{(2)}, \ldots, c_n^{(2)}\} \subset [0,1]^{d_2}$
- **Rank distance cov.**: $\mathrm{RdCov}_n = \mathrm{dCov}_n \left(\left\{ (\hat{R}_n^X(X_i), \hat{R}_n^Y(Y_i)) \right\}_{i=1}^n \right)$

Distribution-freeness [Deb and S. (2019)]

X and Y abs. cont. Under H_0, the dist. of RdCov_n is free of P_X and P_Y.

- **Test**: $\mathrm{H}_0 : X \perp\!\!\!\perp Y$ vs. $\mathrm{H}_1 : X \not\perp\!\!\!\perp Y$
- Take $\mu_1 = \text{Uniform}([0,1]^{d_1})$ and $\mu_2 = \text{Uniform}([0,1]^{d_2})$

Rank distance covariance [Deb and S. (2019)]

- OT rank of X_i: $\hat{R}_n^X : \{X_1, \ldots, X_n\} \to \{c_1^{(1)}, \ldots, c_n^{(1)}\} \subset [0,1]^{d_1}$
- OT rank of Y_i: $\hat{R}_n^Y : \{Y_1, \ldots, Y_n\} \to \{c_1^{(2)}, \ldots, c_n^{(2)}\} \subset [0,1]^{d_2}$
- **Rank distance cov.**: $\mathrm{RdCov}_n = \mathrm{dCov}_n \left(\left\{ (\hat{R}_n^X(X_i), \hat{R}_n^Y(Y_i)) \right\}_{i=1}^n \right)$

Distribution-freeness [Deb and S. (2019)]

X and Y abs. cont. Under H_0, the dist. of RdCov_n is free of P_X and P_Y.

Leads to a test that is also: (i) consistent (against all fixed alternatives), (ii) computationally feasible, and (iii) has non-trivial efficiency

- **Test**: $\mathrm{H}_0 : X \perp\!\!\!\perp Y$ vs. $\mathrm{H}_1 : X \not\!\perp\!\!\!\perp Y$
- Take $\mu_1 = \text{Uniform}([0,1]^{d_1})$ and $\mu_2 = \text{Uniform}([0,1]^{d_2})$

Rank distance covariance [Deb and S. (2019)]

- OT rank of X_i: $\hat{R}_n^X : \{X_1, \ldots, X_n\} \to \{c_1^{(1)}, \ldots, c_n^{(1)}\} \subset [0,1]^{d_1}$
- OT rank of Y_i: $\hat{R}_n^Y : \{Y_1, \ldots, Y_n\} \to \{c_1^{(2)}, \ldots, c_n^{(2)}\} \subset [0,1]^{d_2}$
- **Rank distance cov.**: $\mathrm{RdCov}_n = \mathrm{dCov}_n \left(\left\{ (\hat{R}_n^X(X_i), \hat{R}_n^Y(Y_i)) \right\}_{i=1}^n \right)$

Distribution-freeness [Deb and S. (2019)]

X and Y abs. cont. Under H_0, the dist. of RdCov_n is free of P_X and P_Y.

Leads to a test that is also: (i) consistent (against all fixed alternatives), (ii) computationally feasible, and (iii) has non-trivial efficiency

Our general principle could have been used with any other procedure for mutual independence testing, e.g., the HSIC statistic [Gretton et al. (2005)] which uses ideas from the theory of RKHS, ...

Testing for Symmetry

Data: $\{X_i\}_{i=1}^n$ iid $X \sim P$ (abs. cont.) on $\mathbb{R}$

Test the hypothesis of symmetry, i.e.,

$$\mathrm{H}_0 : X \stackrel{d}{=} -X \qquad \text{versus} \qquad \mathrm{H}_1 : \text{not } \mathrm{H}_0 \qquad (\star)$$

Testing for Symmetry

Data: $\{X_i\}_{i=1}^n$ iid $X \sim P$ (abs. cont.) on $\mathbb{R}$

Test the hypothesis of symmetry, i.e.,

$$\mathrm{H}_0 : X \stackrel{d}{=} -X \qquad \text{versus} \qquad \mathrm{H}_1 : \text{not } \mathrm{H}_0 \qquad (\star)$$

Distribution-free testing for symmetry

- Sign test [Arbuthnott (1710)]: "... the first use of significance tests ..." (first nonparametric test)
- Wilcoxon signed-rank (WSR) test [Wilcoxon (1945)]: Created the field of (classical) nonparametrics
- Arises with paired (matched) data; when normality can be violated

Testing for Symmetry

Data: $\{X_i\}_{i=1}^n$ iid $X \sim P$ (abs. cont.) on $\mathbb{R}$

Test the hypothesis of symmetry, i.e.,

$$\mathrm{H}_0 : X \stackrel{d}{=} -X \qquad \text{versus} \qquad \mathrm{H}_1 : \text{not } \mathrm{H}_0 \qquad (\star)$$

Distribution-free testing for symmetry

- Sign test [Arbuthnott (1710)]: "... the first use of significance tests ..." (first nonparametric test)
- Wilcoxon signed-rank (WSR) test [Wilcoxon (1945)]: Created the field of (classical) nonparametrics
- Arises with paired (matched) data; when normality can be violated
- **Result** [van der Vaart (1998)]: Under H_0, the (i) signs are iid Bernoulli($\frac{1}{2}$), and (ii) signs and signed-ranks are independent

Goal: Develop multivariate distribution-free testing procedures for $(\star)$

Testing Multivariate Symmetry

There are many notions of symmetry in $\mathbb{R}^d$, for $d \geq 2$:

- **Central symmetry**: $\mathrm{H}_0 : X \stackrel{d}{=} -X$
- **Sign symmetry**: $\mathrm{H}_0 : X \stackrel{d}{=} DX$, $\quad D = \mathrm{diag}(\pm 1, \ldots, \pm 1)$
- **Spherical symmetry**: $\mathrm{H}_0 : X \stackrel{d}{=} QX$, $\quad Q$ orthogonal matrix

Testing Multivariate Symmetry

There are many notions of symmetry in $\mathbb{R}^d$, for $d \geq 2$:

- **Central symmetry**: $\mathrm{H}_0 : X \stackrel{d}{=} -X$
- **Sign symmetry**: $\mathrm{H}_0 : X \stackrel{d}{=} DX$, $D = \mathrm{diag}(\pm 1, \ldots, \pm 1)$
- **Spherical symmetry**: $\mathrm{H}_0 : X \stackrel{d}{=} QX$, Q orthogonal matrix

Our framework

- $\mathrm{O}(d)$: group of all orthogonal matrices on $\mathbb{R}^{d \times d}$
- $\mathcal{G}$: compact subgroup of $\mathrm{O}(d)$

- **Test**: $\mathrm{H}_0 : X \stackrel{d}{=} QX \quad \forall\, Q \in \mathcal{G}$, versus H_1 : not H_0

Testing Multivariate Symmetry

There are many notions of symmetry in $\mathbb{R}^d$, for $d \geq 2$:

- **Central symmetry**: $\mathrm{H}_0 : X \overset{d}{=} -X$
- **Sign symmetry**: $\mathrm{H}_0 : X \overset{d}{=} DX$, $D = \mathrm{diag}(\pm 1, \ldots, \pm 1)$
- **Spherical symmetry**: $\mathrm{H}_0 : X \overset{d}{=} QX$, Q orthogonal matrix

Our framework

- $\mathrm{O}(d)$: group of all orthogonal matrices on $\mathbb{R}^{d \times d}$
- $\mathcal{G}$: compact subgroup of $\mathrm{O}(d)$
- **Test**: $\mathrm{H}_0 : X \overset{d}{=} QX \quad \forall\, Q \in \mathcal{G}$, versus H_1 : not H_0

Results [Huang and S. (2023+)]

- Construct distribution-free analogues of signs and signed-ranks
- Generalized Wilcoxon signed-rank test for $\mathcal{G}$-symmetry
- Derive consistency, efficiency lower bounds, etc.
- Distribution-free confidence sets for the center of $\mathcal{G}$-symmetry

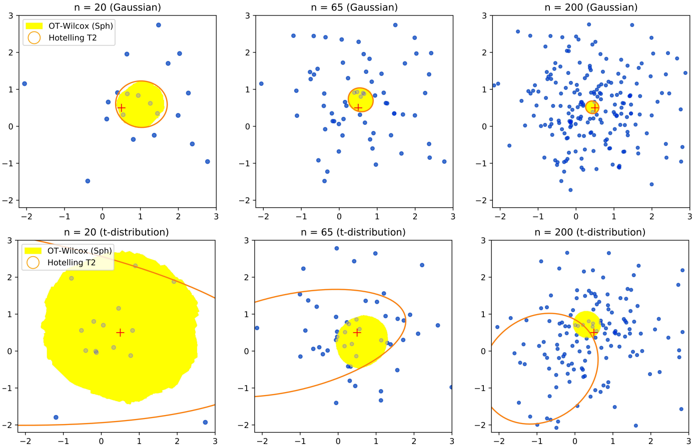


Figure: Confidence sets for center of $\mathrm{O}(d)$-symmetry (spherical symmetry) as the sample size n varies, obtained from (i) normal data (first row) and (ii) data from multivariate t-distribution with 1 degree of freedom (second row).

Summary

- Proposed a multivariate analogue of the Wilcoxon rank-sum test
- Studied its distribution-freeness and efficiency properties

Summary

- Proposed a multivariate analogue of the Wilcoxon rank-sum test
- Studied its distribution-freeness and efficiency properties

- Proposed a general framework — multivariate distribution-free testing procedures based on optimal transport; other examples may include testing for symmetry, testing the equality of K-distributions, independence testing of K-vectors, ...

- The proposed tests are: (i) distribution-free and have good efficiency, (ii) computationally feasible, (iii) more powerful for distributions with heavy tails, and (iv) robust to outliers & contamination

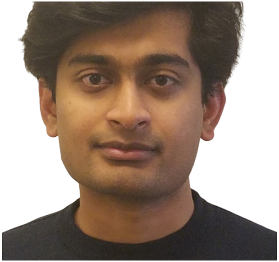
Ghosal and S. (2019). https://arxiv.org/abs/1905.05340 (AoS)
Deb and S. (2019). https://arxiv.org/pdf/1909.08733 (JASA)
Deb, Ghosal and S. (2021). https://arxiv.org/pdf/2107.01718. NeurIPS
Deb, Bhattacharya and S. (2021). https://arxiv.org/abs/2104.01986
Huang and S. (2023). https://arxiv.org/abs/2305.01839
Deb, Bhattacharya and S. (2023+). (working paper)

Thank you very much!

Questions?

Bhaswar Bhattacharya (UPenn)

Promit Ghosal (MIT)



Nabarun Deb (Columbia)

Zhen Huang (Columbia)

- $\nu_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}, \quad \mu_n := \frac{1}{n}\sum_{j=1}^{n}\delta_{c_j}$
- **OT maps**: $R\#\nu = \mu, \qquad \hat{R}_n\#\nu_n = \mu_n$

- $\nu_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}, \quad \mu_n := \frac{1}{n}\sum_{j=1}^{n}\delta_{c_j}$
- **OT maps**: $R\#\nu = \mu, \qquad \hat{R}_n\#\nu_n = \mu_n$
- Suppose $R = \nabla\varphi$, where $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex
- **Legendre-Fenchel dual of** φ: $\varphi^*(y) := \sup_{x\in\mathbb{R}^d}[x^\top y - \varphi(x)]$
- **Fact 1**: R is $\frac{1}{\lambda}$-Lipschitz iff φ^* is λ-strongly convex
- φ^* is λ-strongly convex if, for all $x, y \in \mathrm{Dom}(\varphi^*)$,

$$\varphi^*(y) \geq \varphi^*(x) + \nabla\varphi^*(x)^\top(y-x) + \frac{\lambda}{2}\|y-x\|^2$$

- $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $\quad \mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j}$

- **OT maps**: $R\#\nu = \mu$, $\qquad \hat{R}_n\#\nu_n = \mu_n$

- Suppose $R = \nabla\varphi$, $\quad$ where $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex

- **Legendre-Fenchel dual of** φ: $\varphi^*(y) := \sup_{x\in\mathbb{R}^d}[x^\top y - \varphi(x)]$

- **Fact 1**: R is $\frac{1}{\lambda}$-Lipschitz $\quad$ iff $\quad \varphi^*$ is λ-strongly convex

- φ^* is λ-strongly convex if, for all $x, y \in \mathrm{Dom}(\varphi^*)$,
$$\varphi^*(y) \geq \varphi^*(x) + \nabla\varphi^*(x)^\top(y-x) + \frac{\lambda}{2}\|y-x\|^2.$$

- **Fact 2**: $\nabla\varphi^*(R(x)) = x$ a.e.

- The 2-Wasserstein distance (squared) between ν and μ is defined as:
$$W_2^2(\nu,\mu) := \min_{\pi\in\Pi(\nu,\mu)} \int \|x-y\|^2\, d\pi(x,y),$$
where $\Pi(\nu,\mu) := \{$distributions on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals ν & $\mu\}$

Estimation of OT map [Deb, Ghosal and S. (2021)] Rate of convergence

If the population rank map $R(\cdot)$ is $\frac{1}{\lambda}$-Lipschitz, then

$$\lambda \int \|\hat{R}_n(x) - R(x)\|^2 \, d\nu_n(x) \leq W_2^2(\nu_n, \tilde{\mu}_n) - W_2^2(\nu_n, \mu_n) + 2 \int g \, d(\mu_n - \tilde{\mu}_n)$$

where $\tilde{\mu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{R(X_i)}$ and $g(y) := \varphi^*(y) - \frac{1}{2}\|y\|^2$.

Estimation of OT map [Deb, Ghosal and S. (2021)]

Rate of convergence

If the population rank map $R(\cdot)$ is $\frac{1}{\lambda}$-Lipschitz, then

$$\lambda \int \|\hat{R}_n(x) - R(x)\|^2 \, d\nu_n(x) \le W_2^2(\nu_n, \tilde{\mu}_n) - W_2^2(\nu_n, \mu_n) + 2 \int g \, d(\mu_n - \tilde{\mu}_n)$$

where $\tilde{\mu}_n := \frac{1}{n}\sum_{i=1}^n \delta_{R(X_i)}$ and $g(y) := \varphi^*(y) - \frac{1}{2}\|y\|^2$.

- Then, recalling $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and $\mu_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j}$,

$$\begin{aligned}
D_1 &:= \int \varphi^* d\mu_n - \int \varphi^* d\tilde{\mu}_n \\
&= \int [\varphi^*(\hat{R}_n(x)) - \varphi^*(R(x))] d\nu_n(x) \qquad \text{(as } \hat{R}_n \# \nu_n = \mu_n\text{)} \\
&\overset{(a)}{\ge} \int \left\{ \nabla\varphi^*(R(x))^\top (\hat{R}_n(x) - R(x)) + \frac{\lambda}{2}\|\hat{R}_n(x) - R(x)\|^2 \right\} d\nu_n(x) \\
&\overset{(b)}{=} \underbrace{\int x^\top (\hat{R}_n(x) - R(x)) d\nu_n(x)}_{D_2} + \frac{\lambda}{2} \int \|\hat{R}_n(x) - R(x)\|^2 d\nu_n(x)
\end{aligned}$$

- **Fact 3**: $2D_2 = W_2^2(\nu_n, \tilde{\mu}_n) - W_2^2(\nu_n, \mu_n) + \int \|y\|^2 \, d(\mu_n - \tilde{\mu}_n)(y)$