

Learning to Rank in High Dimensions and Applications

Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes

Myrto LIMNIOS

Bernoulli Instructor

Institute of Mathematics, EPFL, Switzerland

April 17, 2025

Joint work with: Stephan Cléménçon (Telecom Paris, Institut Polytechnique de Paris, FR) and
Nicolas Vayatis (ENS Paris-Saclay, Université Paris-Saclay, FR)



1. Theoretical analysis of rank processes and application to bipartite ranking

Cléménçon S., Limnios M., and Vayatis N. Concentration inequalities for two-sample rank processes with application to bipartite ranking. *Electronic Journal of Statistics*, 2021.

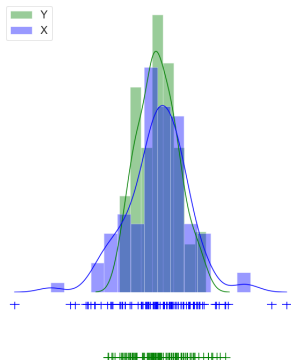
<https://hal.archives-ouvertes.fr/hal-03190532>

2. Application to the independence testing with theoretical analysis

Limnios M., Cléménçon S. On Ranking-based Tests of Independence. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, PMLR 238:577-585, 2024.

<https://proceedings.mlr.press/v238/limnios24a.html>

Two-sample rank statistics



- Student's grades of the **morning class**

Marie	8.2
Antoine	14
...	

- Student's grades of the **evening class**

Clémence	12
Arthur	9
...	
x	0.5

- Are the students of similar level in the morning compared to the evening?
- How **not** to penalize the evening group because of x?
- A ubiquitous hypothesis testing problem, *e.g.* clinical trials

⇒ **Spearman (1904)** introduced **rank statistics** to “reduce the “accidental errors”” and as a response to the Gaussian assumption

The univariate two-sample problem

Based on $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} G$, $Y_1, \dots, Y_m \stackrel{i.i.d.}{\sim} H$, independent, valued in \mathbb{R} , test at level $\alpha \in (0, 1)$

$$\mathcal{H}_0 : G = H \quad \text{versus} \quad \mathcal{H}_1 : G \neq H$$

- Define the ranks of the X_i 's in the **pooled sample**

$$\text{Rank}(X_i) = \sum_{k=1}^n \mathbb{I}\{X_k \leq X_i\} + \sum_{j=1}^m \mathbb{I}\{Y_j \leq X_i\}$$

A basic observation

Under \mathcal{H}_0 , the ranks of the X_i 's are **uniformly distributed** on $\{1, \dots, n + m\}$

- Typical example:** Mann-Whitney-Wilcoxon or ranksum statistic (Wilcoxon (1945); Mann and Whitney (1947))

$$W_{n,m} = \sum_{i=1}^n \text{Rank}(X_i)$$

If $G \geq_{sto} H$, then the X_i 's should be higher ranked with large probability

Two-sample (linear) rank statistics

- $\phi : [0, 1] \rightarrow \mathbb{R}$ *score-generating* function, nondecreasing

$$\widehat{W}_{n,m}^{\phi} = \sum_{i=1}^n \phi \left(\frac{\text{Rank}(X_i)}{n+m+1} \right)$$

- **Distribution-free** under $\mathcal{H}_0 \Rightarrow$ unbiased tests
- Competitive **power**, asymptotic efficiency for **contiguous** alternatives

ex: optimality of the MWW test when $G(x) = F(x - \theta)$ with $\theta \in \mathbb{R}$ (in this case, $\mathcal{H}_0 : \theta = 0$), see e.g. [Lehmann and Romano \(2005\)](#)

- Can be (non-)asymptotically analyzed by means of standard linearization tricks, c.f. **Hajék projection**

$$\widehat{W}_{n,m} - \mathbb{E}[\widehat{W}_{n,m}] = \underbrace{\frac{1}{n} \sum_{i=1}^n (H(X_i) - \mathbb{E}[H(X_i)]) - \frac{1}{m} \sum_{j=1}^m (G(Y_j) - \mathbb{E}[G(Y_j)])}_{\text{independent i.i.d. sums} \Rightarrow \text{classic theorems/bounds}} + \mathcal{O}_{\mathbb{P}} \left(\frac{1}{n} + \frac{1}{m} \right)$$

Extension to independence testing

Based on $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} G$, $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} H$, and the joint distribution of the pair (X, Y) to be F , valued in $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, test at level $\alpha \in (0, 1)$

$$\mathcal{H}_0 : F = G \otimes H \quad \text{versus} \quad \mathcal{H}_1 : F \neq G \otimes H$$

- Consider σ to be the permutation of the index set $\{1, \dots, n\}$, with $n = n_- + n_+$, and suppose the r.v. to be continuous
- Rank the pairs (X_i, Y_i) according to increasing values of the X_i 's: $(X_{\sigma(1)}, Y_{\sigma(1)}), \dots, (X_{\sigma(n_-)}, Y_{\sigma(n_-)})$, s.t. $X_{\sigma(1)} < \dots < X_{\sigma(n_-)}$ and analyzing the ranks of the $Y_{\sigma(i)}$'s through the rank correlation coefficient

Conditionally on the sample $\{X_{\sigma(1)}, \dots, X_{\sigma(n_-)}\}$, the rank of $Y_{\sigma(i)}$ is **stochastically increasing with i** under the alternative hypothesis of dependence

\Rightarrow We can extend rank statistics used for the two-sample problem to independence testing

Visualizing the testing performance in the ROC space...

- Consider a test statistic $T = T(Z)$ based on the observation Z , defining *critical regions* $\{T > t\}$ with $t \in \mathbb{R}$. For fixed and noncomposite hypotheses, find an acceptable trade-off between the two types of error by plotting the parametrized curve

$$\text{ROC} : t \in \mathbb{R} \mapsto \left(\underbrace{\mathbb{P}_{\mathcal{H}_0}\{T > t\}}_{\text{type-I error}}, \underbrace{\mathbb{P}_{\mathcal{H}_1}\{T > t\}}_{\text{power}} \right)$$

where ROC is the *Receiver Operator Characteristic* curve

- Neyman-Pearson's lemma:** the likelihood ratio test statistic is **Uniformly Most Powerful**
 \implies Its ROC curve dominates any other ROC curve of a test statistic everywhere, is concave.

... but also the deviations from \mathcal{H}_0 ...

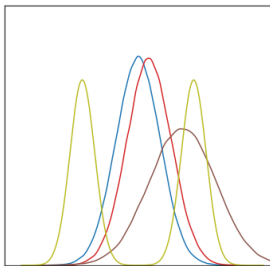
- Define the ROC as the Probability-Probability plot

$$t \mapsto (1 - H(t), 1 - G(t)) = (\mathbb{P}\{Y > t\}, \mathbb{P}\{X > t\})$$

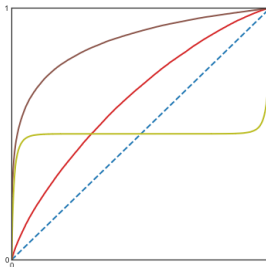
\Rightarrow the ROC curve can be seen as the graph of a continuous mapping

$$\alpha \in (0, 1) \mapsto \text{ROC}_{H,G}(\alpha) \quad (= 1 - G \circ H^{-1}(1 - \alpha) \text{ in absence of jumps})$$

\mathcal{H}_0 : The ROC curve coincides with the diagonal of the unit square



(a) Probability distributions



(b) ROC curves

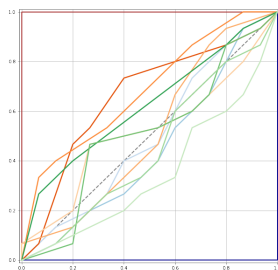
... and devise univariate two-sample testing procedures...

- **Method 1:** By building critical regions, away from the diagonal, in the ROC space and comparing the empirical ROC curve based on

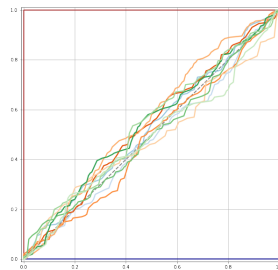
$$\hat{H}_m(t) = \frac{1}{m} \sum_{j=1}^m \mathbb{I}\{Y_j \leq t\} \text{ and } \hat{G}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq t\}$$

and plot the curve, pivotal under \mathcal{H}_0

$$\widehat{\text{ROC}}_{\text{H,G}} = \text{ROC}_{\hat{H}_m, \hat{G}_n}$$



a. $n = m = 15$



b. $n, m = 200, 150$

Figure: Examples of empirical ROC curves simulations under the null hypothesis.

... and devise univariate two-sample testing procedures...

The empirical ROC curve is fully determined by the $\text{Rank}(X_i)$'s

- The breakpoints of the piecewise linear ROC curve necessarily belong to the set of gridpoints

$$\{(j/m, i/n) : j \in \{1, \dots, m-1\}, i \in \{1, \dots, n-1\}\}$$

- Denote by $X_{(i)}$ the order statistics related to the sample $\{X_1, \dots, X_n\}$ satisfy

$$\text{Rank}(X_{(n)}) > \dots > \text{Rank}(X_{(1)})$$

We obtain the ROC curve by connecting with a continuous broken line the jump points of the step curve

$$\alpha \in [0, 1] \mapsto \sum_{j=1}^m \hat{\gamma}_j \times \mathbb{I}\{\alpha \in [(j-1)/m, j/m)\}$$

with

$$\hat{\gamma}_j = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{j \geq N - \text{Rank}(X_{(n-i+1)}) - i + 2\}$$

\Rightarrow Only interesting if the number of all possible empirical ROC curves of size $\binom{N}{n}$ is not too large...

... and devise univariate two-sample testing procedures...

Method 2: By considering statistics summarizing the ROC curve

Example: Area Under the ROC Curve (AUC)

$$\text{AUC}_{H,G} := \int_0^1 \text{ROC}_{H,G}(\alpha) d\alpha = \mathbb{P}\{Y < X\} + \frac{1}{2} \mathbb{P}\{X = Y\}$$

$$\mathcal{H}_0 : \text{AUC} := \int_0^1 \text{ROC}_{H,G} = \frac{1}{2} \quad \text{vs.} \quad \mathcal{H}_1 : \text{AUC} > \frac{1}{2}$$

Remark: The empirical version is an affine transformation of the MWW statistic

$$\widehat{W}_{n,m} = nm \text{AUC}_{\widehat{H}_m, \widehat{G}_n} + \frac{n(n+1)}{2}$$

... and end up with two-sample linear rank tests!

$\widehat{W}_{n,m}^\phi$ are **flexible summaries of the empirical ROC curve** depending on ϕ

- **Classic two-sample setting:** fix $p \in (0, 1)$ and for $N \geq 1/p$, set

$$n = \lfloor pN \rfloor \text{ and } m = \lceil (1 - p)N \rceil = N - n$$

- Define $F = pG + (1 - p)H$
- As $N \rightarrow \infty$, the asymptotic mean of the two-sample rank statistic $\widehat{W}_{n,m}^\phi/n$ is

$$W_\phi = \mathbb{E}[\phi \circ F(X)]$$

$$\mathcal{H}_0 : W_\phi = \int_0^1 \phi(u) du \quad \text{vs.} \quad \mathcal{H}_1 : W_\phi > \int_0^1 \phi(u) du$$

What happens in multivariate settings?

How to define rank statistics?

- Component-wise ranks (Lung-Yut-Fong et al. (2015))
- Data-depth ranks with *e.g.* quantile functions (Chaudhuri (1996); Oja (1983))
- Spatial ranks (Möttönen and Oja (1995); Möttönen et al. (1997); Möttönen et al. (2005))
- Distance-based ranks (Hallin and Paindaveine (2002a,b, 2008))
- Optimal transport-based multivariate ranks (center-outward distributions Hallin et al. (2021); Shi et al. (2022), rank maps Deb and Sen (2021))

How to test for independence?

Reject \mathcal{H}_0 for “large $\mathcal{D}(F, G \otimes H)$ ” with \mathcal{D} a pseudo-distance.

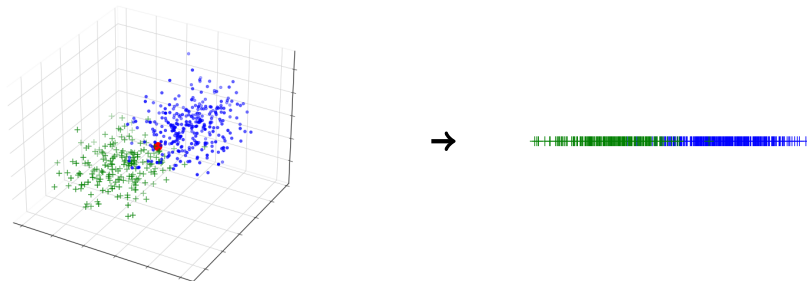
- Covariance-based distances, generalized to metric spaces in Lyons (2013); Jakobsen (2017)
- Kernel-based extensions relying on the *Hilbert-Schmidt Independence Criterion* (HSIC) Gretton et al. (2005a,b, 2007)
- Mutual information Berrett and Samworth (2019); Gonzalez et al. (2021)
- Partitioning techniques Gretton and Györfi (2010); Heller et al. (2016)
- Maximal information criterion Reshef et al. (2011, 2016, 2018)
- Extension of Rank statistics based on optimal transport distances Ramdas et al. (2015); Deb and Sen (2021)

Main limitations

- **Statistics dependent on the model specification** (local representation of the data, definition of the statistic): only one rank definition is distribution-free under mild assumptions on the distributions (based on measure transportation e.g. [Hallin \(2017\)](#)) but no practical formula \Rightarrow **data-driven testing threshold** \Rightarrow *requires additional computations and approximations*
- **Asymptotic guarantees**: usually for the consistency, and distributions under the null and alternative statistical hypothesis, and local power/efficiency analysis
- **Curse of dimensionality**: the estimators suffer from settings where dimension d is ‘large’ \Rightarrow misspecification of the asymptotic distribution ([Huang et al. \(2023\)](#)) + difficulty (even impossibility) to recover parametric rates $\mathcal{O}_{\mathbb{P}}(1/\sqrt{N})$ ([Arias-Castro et al. \(2018\)](#))
- Few testing methods are adapted to **general structures** of data under nonparametric assumptions (non-Euclidean, e.g. graphs, time-series, etc.)

Our approach based on bipartite ranking

How to *rank* a new observation Z given the pooled sample $\mathbf{X} \sim G$, $\mathbf{Y} \sim H$, in any multivariate feature space \mathcal{Z} ?



Goal: Learn the optimal **scoring function** $s : \mathcal{Z} \rightarrow (-\infty, \infty]$ such that the \mathbf{X} 's are ranked statistically higher than the \mathbf{Y} 's.

\Rightarrow Minimize the loss function = number of misranked pairs mapped using the function $s(z)$

A gold standard criterion in bipartite ranking: the ROC curve

- Consider a scoring function candidate s : $s(\mathbf{X}) \sim G_s$ and $s(\mathbf{Y}) \sim H_s$
- **Bipartite ranking expected loss and AUC**

$$\begin{aligned} L(s) &= \mathbb{E}[\mathbb{I}\{s(\mathbf{Y}) > s(\mathbf{X})\}] + \frac{1}{2} \mathbb{P}\{s(\mathbf{Y}) = s(\mathbf{X})\} \\ &= 1 - \int_0^1 \text{ROC}_{H_s, G_s} = 1 - \text{AUC}(s) \end{aligned}$$

- **Optimal elements**¹: \mathcal{S}^* = increasing transforms of the likelihood ratio $\Psi(\mathbf{z}) = (dG/dH)(\mathbf{z})$
- **Neyman-Pearson**: for any s and for all $s^* \in \mathcal{S}^*$,

$$\text{ROC}_s(\alpha) \leq \text{ROC}_{s^*}(\alpha) = \text{ROC}_\Psi(\alpha) := \text{ROC}^*(\alpha)$$

\mathcal{H}_0 : The ROC^{*} curve coincides with the diagonal of the unit square

But ROC^{*} is **unknown** \implies We need to solve the bipartite ranking problem first!

¹Proposition 2, Cl  men  on S. and Vayatis, N. Tree-structured ranking rules and approximation of the optimal ROC curve. In *ALT '08: Proceedings of the 2008 conference on Algorithmic Learning Theory*, 2008. ◀ ≡ ▶ ◀ ≡ ▶ ≡|≡ ↻ 🔍

Bipartite Ranking: state of the art algorithms

'Scalar' view:

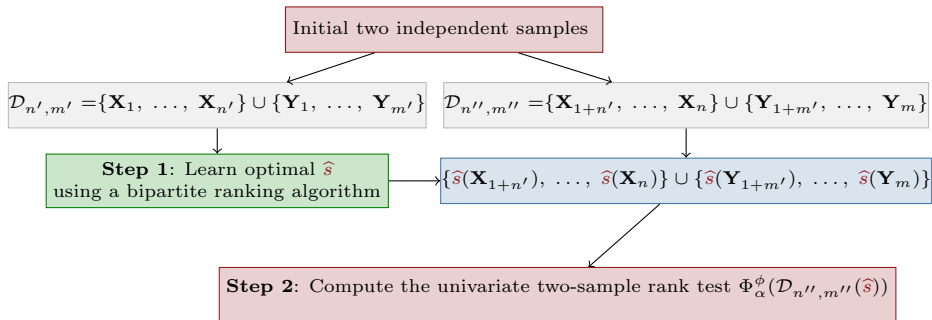
Optimize an empirical counterpart of a scalar criterion $W(s)$ summarizing the ROC curve

$$\max_{s \in \mathcal{S}_0} W(s)$$

- AUC, see [Cl  men  on et al. \(2008\)](#)
- local AUC, see [Cl  men  on and Vayatis \(2007\)](#)
($\phi(u) = u\mathbb{I}\{u \geq 1 - u_0\}$ for the fraction $u_0 \in (0, 1)$ of the 'best instances')
- pairwise exponential loss, see [Freund et al. \(2003\)](#)
- q -norm push, see [Rudin et al. \(2005\)](#)
- $W_\phi(s) = W_\phi(H_s, G_s)$, see [Cl  men  on et al. \(2021\)](#)

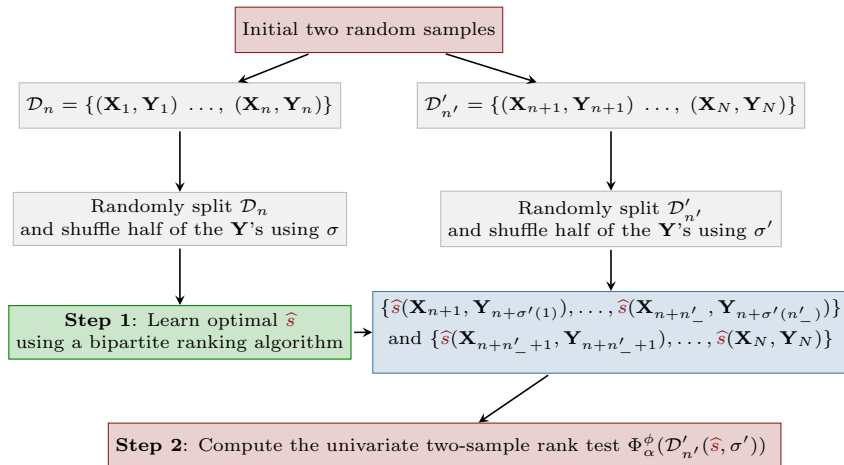
Additional references: Theoretical contributions [Agarwal et al. \(2005\)](#); [Cl  men  on et al. \(2008\)](#); [Menon and Williamson \(2016\)](#), learning algorithms [Burges et al. \(2007, 2005\)](#); [Joachims \(2006\)](#); [Rudin \(2006\)](#); [Cl  men  on and Vayatis \(2009\)](#)

The generic procedure for homogeneity testing¹



¹Cléménçon S., **Limnios M.**, and Vayatis N. A bipartite ranking approach to the two-sample problem. *Submitted, arXiv:2302.03592*, 2023.

And for independence testing²



⇒ **Flexibility** through the choices of bipartite ranking algorithm, and ϕ recovering classic univariate rank-tests and tailoring the ranks depending on the applied study

How to define the test statistic Φ_α^ϕ and what are the related properties?

²Limnios M., Cléménçon S. On Ranking-based Tests of Independence. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, PMLR 238:577-585, 2024.

We now focus on the situation where bipartite ranking is formulated as
maximization of two-sample linear rank statistics

Statistical learning framework

- $p \in (0, 1)$, $N \in \mathbb{N}^*$ such that $N \geq 1/p$
- $n = \lfloor pN \rfloor$ and $m = \lceil (1 - p)N \rceil = N - n$
- $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} G$, $\mathbf{Y}_1, \dots, \mathbf{Y}_m \stackrel{i.i.d.}{\sim} H$, independent
- G, H unknown (**nonparametric**) and continuous, defined on \mathcal{Z} **high-dimensional** and measurable
- \mathcal{S} class of measurable *scoring functions* $s : \mathcal{Z} \rightarrow (-\infty, \infty]$

Generalization of two-sample linear R -statistics¹

- $\phi : [0, 1] \rightarrow \mathbb{R}$ *score-generating* function, nondecreasing
- For $s \in \mathcal{S}$, we consider the univariate samples $\{s(\mathbf{X}_1), \dots, s(\mathbf{X}_n)\}, \{s(\mathbf{Y}_1), \dots, s(\mathbf{Y}_m)\}$

$$\text{Rank}(t) = \sum_{k=1}^n \mathbb{I}\{s(\mathbf{X}_k) \leq t\} + \sum_{j=1}^m \mathbb{I}\{s(\mathbf{Y}_j) \leq t\}$$

Definition

$$\widehat{W}_{n,m}^{\phi}(s) = \sum_{i=1}^n \phi\left(\frac{\text{Rank}(s(\mathbf{X}_i))}{N+1}\right)$$

¹Cléménçon S., **Limnios M.**, and Vayatis N. Concentration inequalities for two-sample rank processes with application to bipartite ranking. *Electronic Journal of Statistics*, 2021.

Scalar performance measure¹

- $\mathbf{X} \sim G$, $\mathbf{Y} \sim H$, independent, valued in \mathcal{Z}
- $G_s(t) = \mathbb{P}\{s(\mathbf{X}) \leq t\}$, $H_s(t) = \mathbb{P}\{s(\mathbf{Y}) \leq t\}$
- $p \in (0, 1)$ ‘theoretical’ proportion of \mathbf{X} s among the pooled sample
- $F_s = pG_s + (1 - p)H_s$

Definition

The W_ϕ -ranking performance criterion based on F_s is

$$W_\phi(s) = \mathbb{E}[(\phi \circ F_s)(s(\mathbf{X}))]$$

What are the *best* scoring functions s^* ? The oracle class is composed of the nondecreasing transforms of the likelihood ratio $\Psi(\mathbf{z}) = dG/dH(\mathbf{z})$ AND maximize W_ϕ (cf. Proposition 6¹)

¹Cléménçon S., **Limnios M.**, and Vayatis N. Concentration inequalities for two-sample rank processes with application to bipartite ranking. *Electronic Journal of Statistics*, 2021.

Problem statement

Based on $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} G$, $\mathbf{Y}_1, \dots, \mathbf{Y}_m \stackrel{i.i.d.}{\sim} H$, independent, $S_0 \subset S$, find

$$\hat{s} \in \arg \max_{s \in S_0} \widehat{W}_{n,m}^\phi(s)$$

Statistical learning guarantees: decompose the generalization error

$$W_\phi(s^*) - W_\phi(\hat{s}) \leq \underbrace{2 \sup_{s \in S_0} \left| \frac{1}{n} \widehat{W}_{n,m}^\phi(s) - W_\phi(s) \right|}_{\text{estimation error}} + \underbrace{W_\phi(s^*) - \sup_{s \in S_0} W_\phi(s)}_{\text{approximation error/bias}}$$

Main result: generalization error bound

- (A1) Let $M > 0$. For all $s \in \mathcal{S}_0$, the random variables $s(\mathbf{X})$ and $s(\mathbf{Y})$ are continuous, with c.d.f. that are twice differentiable and have Sobolev $\mathcal{W}^{2,\infty}$ -norms bounded by $M < +\infty$
- (A2) The score-generating function $\phi : [0, 1] \mapsto \mathbb{R}$, is nondecreasing and twice continuously differentiable
- (A3) The class of scoring functions \mathcal{S}_0 is a VC class of finite VC dimension $\mathcal{V} < +\infty$

Theorem (Corollary 7¹)

Suppose (A1-3) fulfilled. For any $\delta \in (0, 1)$, with probability at least $1 - \delta$

$$W_\phi(s^*) - W_\phi(\hat{s}) \leq 2C_3 \sqrt{\frac{\log(C_2/\delta)}{pN}} + \left(W_\phi(s^*) - \sup_{s \in \mathcal{S}_0} W_\phi(s) \right)$$

for sufficiently large N , where the constants C_i , $i \leq 4$ depend only on ϕ , \mathcal{V}

¹Cléménçon S., Limnios M., and Vayatis N. Concentration inequalities for two-sample rank processes with application to bipartite ranking. *Electronic Journal of Statistics*, 2021.

Ingredients for the proof

Main tool: linearization of R -processes with uniform control of the remainder w.p. $1 - \delta$ over \mathcal{S}_0 ³

$$\widehat{W}_{n,m}^\phi(s) = \underbrace{n\widehat{W}_\phi(s)}_{\text{central statistic}} + \underbrace{\left(\widehat{V}_n^X(s) - \mathbb{E}\left[\widehat{V}_n^X(s)\right]\right) + \left(\widehat{V}_m^Y(s) - \mathbb{E}\left[\widehat{V}_m^Y(s)\right]\right)}_{\text{empirical processes}} + \underbrace{\mathcal{R}_{n,m}(s)}_{\text{remainder}}$$

And...

- (i) Permanence results to control the complexity of classes of functions⁴
- (ii) Linearization methods applied to one-/two-samples asymmetric U -processes using⁵:
 - (a) Hájek projection method
 - (b) Hoeffding decomposition results for generalized U -statistics
- (iii) **Uniform concentration bounds for one-/two-samples asymmetric (degenerate) U -processes**⁶ (*symmetrization, chaining, ...*⁷)
- (iv) Maximal inequalities⁸

³Proposition 5, Cléménçon, Limnios and Vayatis (2021)

⁴Lemma 19-20, Cléménçon, Limnios and Vayatis (2021)

⁵Hájek (1968); Serfling (1980)

⁶Theorem 2 Major (2006), Lemma 16, Cléménçon, Limnios and Vayatis (2021)

⁷van de Geer (2000); van der Vaart (1998); De la Pena and Giné (1999)

⁸Lemma 2.4 Neumeyer (2004), Theorem 6 Nolan and Pollard (1987), Cléménçon, Limnios and Vayatis (2021)

Main consequences and additional contributions

- Generalization of R -statistics to **nonparametric** distributions and **high-dimensional** spaces

$$\widehat{W}_{n,m}^{\phi}(s) = \sum_{i=1}^n \phi\left(\frac{\text{Rank}(s(\mathbf{X}_i))}{N+1}\right)$$

- Series of **nonasymptotic uniform results of order $\mathcal{O}_{\mathbb{P}}(N^{-1/2})$** \implies Just like classic Empirical Risk Minimization theory
- The obtained W_{ϕ} is a **scalar performance measure for ranking procedures**
 - ▶ generic thanks to ϕ and the ranking algorithm operating over \mathcal{S}_0
 - ▶ robust w.r.t sampling bias: **bipartite ranking** (*e.g. recommendation systems, computational biology*), **learning to rank anomalies** (*e.g. fraud detection*)
 - ▶ tailors the summaries of the *Receiver Operating Characteristic* (ROC) curve (*e.g. Area Under the ROC Curve (AUC)*)
- **Numerical experiments:** Simple algorithm based on gradient ascent proposed to learn the optimal \widehat{s} , open-access online at https://github.com/MyrtoLimmios/grad_2sample

\implies **How this allows to prove finite sample guarantees for new hypothesis testing procedures to compare populations?**

Independence testing with R -statistics

$$\mathcal{H}_0 : F = G \otimes H \quad \text{versus} \quad \mathcal{H}_1 : F \neq G \otimes H$$

From two-sample R -statistics...

- $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} G$, $\mathbf{Y}_1, \dots, \mathbf{Y}_n \stackrel{i.i.d.}{\sim} H$, valued in resp. \mathcal{X} and \mathcal{Y}
- F joint distribution of the $(\mathbf{X}_i, \mathbf{Y}_i)$'s valued in $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$
- σ a permutation of the index set of the \mathbf{Y}' s on $[n_-]$, $n = n_- + n_+$, and form two independent samples

$$\mathcal{D}_n(\sigma) = \{(\mathbf{X}_1, \mathbf{Y}_{\sigma(1)}), \dots, (\mathbf{X}_{n_-}, \mathbf{Y}_{\sigma(n_-)})\} \cup \{(\mathbf{X}_{n_-+1}, \mathbf{Y}_{n_-+1}), \dots, (\mathbf{X}_n, \mathbf{Y}_n)\}$$

$$\text{Rank}_\sigma(t) = \sum_{i=1}^{n_-} \mathbb{I}\{s(\mathbf{X}_i, \mathbf{Y}_{\sigma(i)}) \leq t\} + \sum_{i=1+n_-}^n \mathbb{I}\{s(\mathbf{X}_i, \mathbf{Y}_i) \leq t\}$$

Definition

$$\widehat{W}_{n_-, n_+}^\phi(s, \sigma) = \sum_{i=1+n_-}^n \phi\left(\frac{\text{Rank}_\sigma(s(\mathbf{X}_i, \mathbf{Y}_i))}{n+1}\right)$$

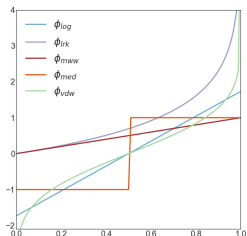
And similar theoretical treatment applies and obtained results with $(F, G \otimes H) \dots$

... to the design of independence testing

Based on $\mathbf{X} \sim G$, $\mathbf{Y} \sim H$, valued in resp. \mathcal{X} and \mathcal{Y} , test at level $\alpha \in (0, 1)$

$$\mathcal{H}_0 : W_\phi(s^*) = \int_0^1 \phi(u) du \quad \text{versus} \quad \mathcal{H}_1 : W_\phi(s^*) - \int_0^1 \phi(u) du > 0$$

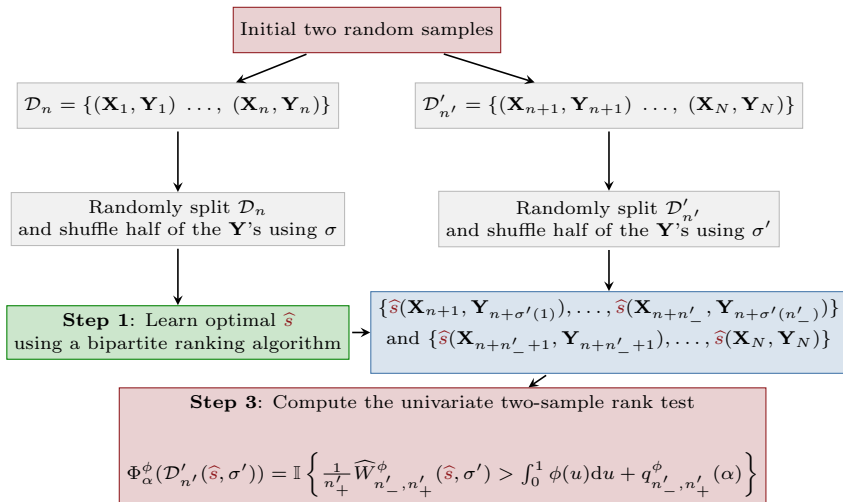
where $s^* : \mathcal{X} \times \mathcal{Y} \rightarrow (-\infty, \infty]$ a non-decreasing transform of $\Psi = dF/d(G \otimes H)$



Statistic	ϕ
MWW (green)	$\phi(u) = u$
Logistic (blue)	$\phi_{log}(u) = 2\sqrt{3}(u - 1/2)$
Logrank (orange)	$\phi_{lrk}(u) = -\log(1 - u)$
Median (red)	$\phi_{med}(u) = \text{sgn}(u - 1/2)$
Van der Waerden (purple)	$\phi_{vdw}(u) = \Delta^{-1}(u)$

Figure: Graphs of typical choices of score-generating function ϕ

The generic procedure for independence testing



Ranking-based rank test statistic¹

- Let \widehat{s} optimal element minimizing the bipartite ranking loss (*Step 1*) over $\mathcal{S}_0 \subset \mathcal{S}$
- We compare

$$\widehat{W}_{n'_-, n'_+}^\phi(\widehat{s}, \sigma') = \sum_{i=1+n+n'_-}^N \phi\left(\frac{\text{Rank}_{\sigma'}(\widehat{s}(\mathbf{X}_i, \mathbf{Y}_i))}{n' + 1}\right)$$

to its null $(1 - \alpha)$ -quantile for Φ_α^ϕ after centering

Theorem (Type-I error bound)

Let $\phi(u)$, $1 \leq n'_- < n'$ and $1 \leq n'_+ < n'$, and fix $\alpha \in (0, 1)$. Under the null hypothesis \mathcal{H}_0 , the type-I error of the test is less than α for all pair of distributions $(H \otimes G, F)$

$$\mathbb{P}_{\mathcal{H}_0} \left\{ \Phi_\alpha^\phi(\mathcal{D}'_{n'}(\widehat{s})) = +1 \right\} \leq \alpha$$

¹**Limnios M.**, Cl  men  on S. On Ranking-based Tests of Independence. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, PMLR 238:577-585, 2024.

Under the alternative hypothesis of dependence

- The **bias of the model** is bounded by $\delta > 0$ s.t. $\mathcal{B}(\delta)$ is the set of all pairs $(H \otimes G, F)$ of probability distributions on $\mathcal{X} \times \mathcal{Y}$ satisfying

$$W_{\phi}^* - \sup_{s \in \mathcal{S}_0} W_{\phi}(s) \leq \delta$$

- The **deviation from the null hypothesis** \mathcal{H}_0 is greater than $\varepsilon > 0$

$$W_{\phi}^* - \int_0^1 \phi(u) du \geq \varepsilon$$

Uniform concentration bound for the statistical type-II error¹

Theorem (Type-II error bound)

Let ϕ and $\varepsilon > \delta > 0$. Let $p \in (0, 1)$ such that $n \wedge n' \geq 1/p$. Set $n_+ = \lfloor pn \rfloor$ and $n_- = \lceil (1-p)n \rceil = n - n_+$, as well as $n'_+ = \lfloor pn' \rfloor$ and $n'_- = \lceil (1-p)n' \rceil = n' - n'_+$. Fix $\alpha \in (0, 1)$. Suppose (A1-3) are fulfilled. Then, there exist constants C_1 and $C_2 \geq 24$, such that the type-II error of the test is uniformly bounded

$$\sup_{(G \otimes H, F) \in \mathcal{H}_1(\varepsilon) \cap \mathcal{B}(\delta)} \mathbb{P}_{\mathcal{H}_1} \left\{ \Phi_\alpha^\phi = 0 \right\} \leq \underbrace{18 \exp \left(-\frac{C n' (\varepsilon - \delta)^2}{16} \right)}_{\text{Step 2 (type-II error bound of a univariate rank test)}} + \underbrace{C_2 \exp \left(-\frac{n}{8C_2} p(p \wedge (1-p)) (\varepsilon - \delta) \log \left(1 + \frac{\varepsilon - \delta}{32C_1(p \wedge (1-p))} \right) \right)}_{\text{Step 1 (ranking)}}$$

for sufficiently large n' , where the constants C (resp. C_1, C_2) depend only on p, ϕ (resp. ϕ, \mathcal{V})

¹Limnios M., Cl  men  on S. On Ranking-based Tests of Independence. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, PMLR 238:577-585, 2024.

Numerical applications

Experimental settings for homogeneity testing

- **Bipartite ranking algorithms:** RankNN (RNN, [Burges et al. \(2005\)](#)), linear RankSVM (see [Joachims \(2002\)](#)) with L_2 loss (rSVM2), RankBoost (rBoost, [Freund et al. \(2003\)](#)), Ranking Forest (rForest, [Cléménçon et al. \(2013\)](#))
- **Score-generating functions:** $\phi_{MWW}(u) = u$ (MWW, [Wilcoxon \(1945\)](#))
- **Two-sample statistics:**

MMD: Maximum Mean Discrepancy [Gretton et al. \(2007, 2012\)](#), \mathcal{F} unit ball of a Reproducing Kernel Hilbert Space (RKHS)

$$\text{MMD}(G, H) = \sup_{f \in \mathcal{F}} |\mathbb{E}[f(\mathbf{X})] - \mathbb{E}[f(\mathbf{Y})]| \quad (1)$$

Energy: metric-based Energy test [Székely and Rizzo \(2013\)](#), $\|\cdot\|$ Euclidean norm in \mathbb{R}^d

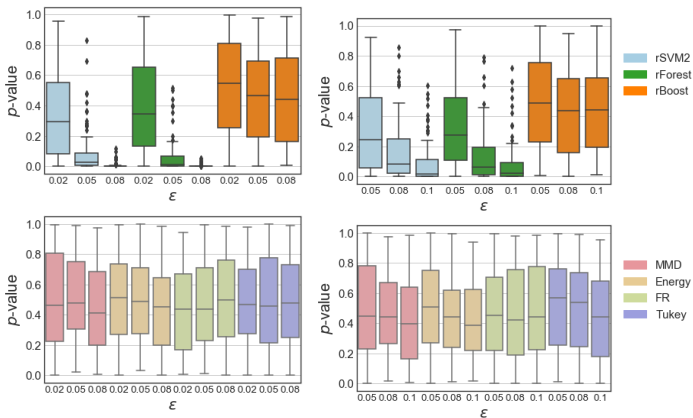
$$\mathcal{E}_{n,m} = \frac{mn}{m+n} \left(\frac{2}{nm} \sum_{i,j \leq n,m} \|\mathbf{X}_i - \mathbf{Y}_j\| - \frac{1}{n^2} \sum_{i,j \leq n} \|\mathbf{X}_i - \mathbf{X}_j\| - \frac{1}{m^2} \sum_{i,j \leq m} \|\mathbf{Y}_i - \mathbf{Y}_j\| \right) \quad (2)$$

FR: generalization of graph-based Wald-Wolfowitz Runs test [Friedman and Rafsky \(1979\)](#)

Tukey: statistical depth-based generalization of ranks [Tukey \(1975\)](#) with method for testing of [Liu and Singh \(1993\)](#)

Gaussian location model

$$\mathbf{X} \sim \mathcal{N}_d((\varepsilon/\sqrt{d}) \times \mathbf{1}_d, \Sigma), \mathbf{Y} \sim \mathcal{N}_d(0_d, \Sigma)$$



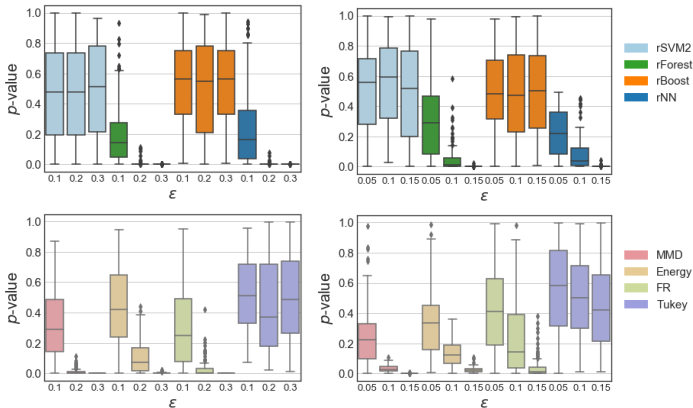
(a) Low dependence
 $\text{Cov}(X_i, Y_j) < 0, \forall j \leq m,$
 $\text{Cov}(X_i, Y_j) = 0$ else

(b) Strong dependence
 $\text{Cov}(X_i, Y_j) > 0, \forall i \leq n, j \leq m$

Numerical parameters: $n = m = 1000, n' = m' = 4N/5, d = 6$

Gaussian scale model

$\mathbf{X} \sim \mathcal{N}_d(0_d, \Sigma_X)$ and $\mathbf{Y} \sim \mathcal{N}_d(0_d, \Sigma_Y)$



(a) Toeplitz covariance matrix

$$\Sigma_{X,i,j} = (\beta + \epsilon)^{|i-j|}$$

$$\Sigma_{Y,i,j} = \beta^{|i-j|}$$

(b) Spiked identity model

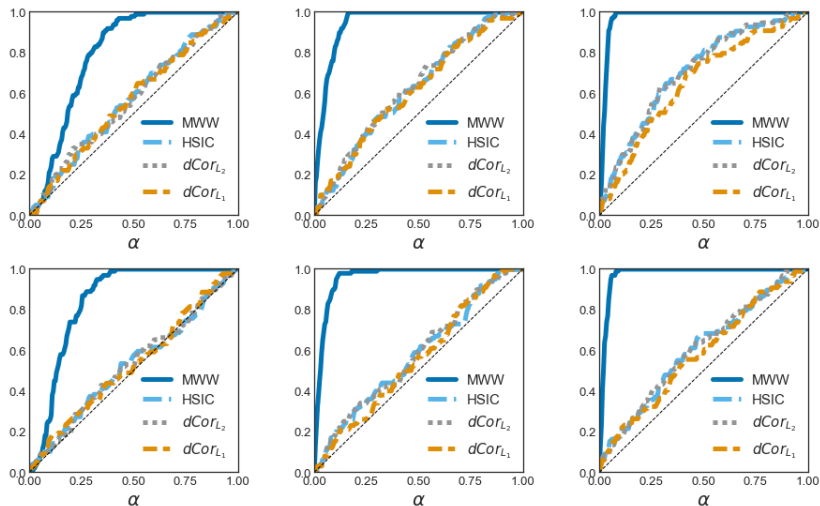
$$\Sigma_X = (1 - (\beta + \epsilon))\mathbb{I}_d + (\beta + \epsilon)\mathbf{1}_d\mathbf{1}_d^T$$

$$\Sigma_Y = (1 - \beta)\mathbb{I}_d + \beta\mathbf{1}_d\mathbf{1}_d^T$$

Numerical parameters: $n = m = 1000$, $n' = m' = 4N/5$, $d = 20$ (a) and $d = 30$ (b)

Impact of the dimension d on the power: independence testing

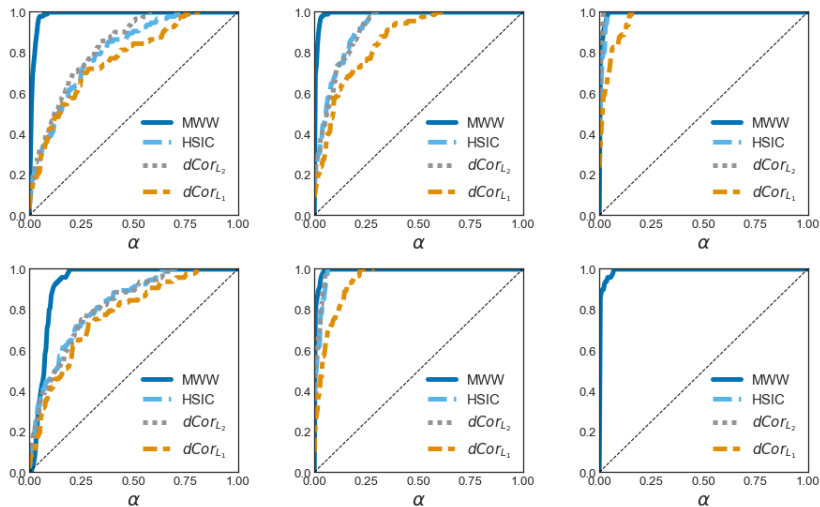
$(\mathbf{X}, \mathbf{Y}) \sim \mathcal{N}_d(0_d, \Sigma_\rho)$, $\text{Cov}(X^1, Y^k) = \rho$, for all $k \leq l$ and $\Sigma_{\rho, i, j} = \delta_{ij}$ otherwise



Numerical parameters: $N = 1000$, $d = 4$ top, $d = 10$ bottom, number of permutations for rForest_{MWW} $K_p = 10$, number of permutations for SoA's $K_0 = 200$, $B = 100$, $\rho \in \{0.2, 0.3, 0.4\}$ (top), $\rho \in \{0.10, 0.15, 0.20\}$ (bottom)

Sparse (top) vs. dense (bottom)

$X^u = \rho \cos \Theta + \omega_1/4$, $Y^v = \rho \sin \Theta + \omega_2/4$, with $\rho \in \{1, 2, 3\}$, $\omega_i \sim \mathcal{N}(0, 1)$, $i \in \{1, 2\}$, and $\Theta \sim \mathcal{U}([0, 2\pi])$



Numerical parameters: $N = 500$, $d = 50$, number of permutations for rForest_{MWW} $K_p = 10$, number of permutations for SoA's $K_0 = 200$, $B = 100$, $\rho \in \{0.3, 0.4, 0.5\}$ (top), $\rho \in \{0.2, 0.3, 0.4\}$ (bottom)

Guarantees and main contributions

Theoretical guarantees¹

- **Exact control of type-I error** i.e. less than α for all sample sizes and all α , with probability 1 (Theorem 7)
⇒ **Exact calibration of the test for any algorithm** that only relies on the possible optimization of its hyperparameters
- **Finite-sample uniform bound for the type-II error** over the class of alternative distributions (deviation from \mathcal{H}_0 + bias from bipartite ranking algorithm) with high probability under some conditions (Theorem 11)

Empirical guarantees¹

- **Competitive power** for small deviations from the null and high dimensions compared to state-of-the-art methods (*Maximum Mean Discrepancy* [Gretton et al. \(2007, 2012\)](#), *graph-based Wald-Wolfowitz Runs test* [Friedman and Rafsky \(1979\)](#), *statistical depth-based generalization of ranks* [Tukey \(1975\)](#), *metric-based Energy test* [Székely and Rizzo \(2013\)](#))
- **Numerical experiments:** Algorithmic procedure open-access online at https://github.com/MyrtoLimnios/independence_ranktest

Remark: We obtain similar results for homogeneity testing under nonparametric alternatives²

¹**Limnios M.**, Cléménçon S. On Ranking-based Tests of Independence. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, PMLR 238:577-585, 2024.

²Cléménçon S., **Limnios M.**, and Vayatis N. A bipartite ranking approach to the two-sample problem. *Under review, arXiv:2302.03592*, 2023.

Conclusions

- **Theory.** Generic ranking-based method of performance **analyzed under nonparametric assumptions**
 - ▶ **Nonasymptotic guarantees** for the solution output by the chosen learning algorithm
 - ▶ The guarantees depend on the dimension of the feature space only through the choice of learning algorithm
 - ▶ The rank test is chosen according to the nature of these guarantees
- **Practice.** If well-implemented, the ranking-based two-sample testing procedure is **highly competitive**
 - ▶ It **resists to the increasing dimensions** compared to state-of-the-art methods
 - ▶ It **inherits the advantages of the univariate rank tests**: c.f. capacity to detect ‘small’ departures from the homogeneity assumption (i.e. alternatives contiguous to \mathcal{H}_0)
 - ▶ It is **not that sensitive to the choice of the bipartite ranking algorithm**
- **Ongoing projects:** Minimax results for hypothesis testing, uniform concentration inequalities for ROC curves, extension to multi-sample hypothesis testing

Further readings..

1. **Theoretical analysis of R -processes and application to bipartite ranking**¹
Cléménçon S., Limnios M., and Vayatis N. Concentration inequalities for two-sample rank processes with application to bipartite ranking. *Electronic Journal of Statistics*, 2021. <https://hal.archives-ouvertes.fr/hal-03190532>
2. **Application to independence testing**
Limnios M., Cléménçon S. On Ranking-based Tests of Independence. *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, PMLR 238:577-585, 2024.
<https://proceedings.mlr.press/v238/limnios24a.html>
3. **Application to learning to rank anomalies (+ 3bis. journal long version in Machine Learning, Springer)**
Limnios M., Noiry N., and Cléménçon S. Learning to rank anomalies: Scalar performance criteria and maximization of two-sample rank statistics. *Proceedings of Machine Learning Research*, 2021.
<https://proceedings.mlr.press/v154/limnios21a.html>
4. **Application to the nonparametric two-sample problem with posturographic data**
Bargiotas I., Kalogeratos A., Limnios M., Vidal P-P., Ricard D., and Vayatis N. Revealing posturographic profile of patients with Parkinsonian syndromes through a novel hypothesis testing framework based on machine learning. *PLOS ONE*, 2021. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0246790>
5. **Two-sample problem with theoretical analysis (under review)**¹
Cléménçon S., Limnios M., and Vayatis N. A bipartite ranking approach to the two-sample problem. *arXiv:2302.03592*. 2023. <https://arxiv.org/abs/2302.03592>

Merci !

¹Alphabetical order that does not represent the contributions of the author.

References I

- S. Agarwal, T. Graepel, R. Herbrich, S. Har-Peled, and D. Roth. Generalization bounds for the area under the ROC curve. *Journal of Machine Learning Research*, 6:393–425, 2005.
- E. Arias-Castro, B. Pelletier, and V. Saligrama. Remember the curse of dimensionality: the case of goodness-of-fit testing in arbitrary dimension. *Journal of Nonparametric Statistics*, 30(2):448–471, 2018.
- T. B. Berrett and R. J. Samworth. Nonparametric independence testing via mutual information. *Biometrika*, 106(3):547–566, 2019.
- C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, page 89–96, 2005.
- C. Burges, R. Ragno, and Q. Le. Learning to rank with nonsmooth cost functions. In *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2007.
- P. Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996.
- S. Cléménçon and N. Vayatis. Ranking the best instances. *Journal of Machine Learning Research*, 8: 2671–2699, 2007.
- S. Cléménçon and N. Vayatis. Tree-based ranking methods. *IEEE Transactions on Information Theory*, 55 (9):4316–4336, 2009.
- S. Cléménçon, G. Lugosi, and N. Vayatis. Ranking and empirical risk minimization of U-statistics. *The Annals of Statistics*, 36(2):844–874, 2008.
- S. Cléménçon, M. Depecker, and N. Vayatis. Ranking Forests. *Journal of Machine Learning Research*, 14: 39–73, 2013.
- S. Cléménçon, M. Limnios, and N. Vayatis. Concentration inequalities for two-sample rank processes with application to bipartite ranking. *Electronic Journal of Statistics*, 15(2):4659 – 4717, 2021.
- V. De la Pena and E. Giné. *Decoupling: from dependence to independence*. Springer Science and Business Media, 1999.
- N. Deb and B. Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Journal of the American Statistical Association*, pages 1–16, 2021.

References II

- Y. Freund, R. D. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.
- J. H. Friedman and L. C. Rafsky. Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests. *The Annals of Statistics*, 7(4):697 – 717, 1979.
- M. E. Gonzalez, J. F. Silva, M. Videla, and M. E. Orchard. Data-driven representations for testing independence: Modeling, analysis and connection with mutual information estimation. *IEEE Transactions on Signal Processing*, 70:158–173, 2021.
- A. Gretton and L. Györfi. Consistent nonparametric tests of independence. *Journal of Machine Learning Research*, 11(46):1391–1423, 2010.
- A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *Algorithmic Learning Theory: 16th International Conference, ALT 2005*, volume 3734, pages 63–78, 2005a.
- A. Gretton, A. Smola, O. Bousquet, R. Herbrich, A. Belitski, M. Augath, Y. Murayama, J. Pauls, B. Schölkopf, and N. Logothetis. Kernel constrained covariance for dependence measurement. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, volume R5, pages 112–119. Proceedings of Machine Learning Research, 2005b.
- A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholkopf, and A. Smola. A kernel method for the two-sample problem. In *Advances in Neural Information Processing Systems 19*. MIT Press, Cambridge, MA, 2007.
- A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholkopf, and A. Smola. A kernel two-sample problem. *Journal of Machine Learning Research*, 13:723–773, 2012.
- J. Hájek. Asymptotic normality of simple linear rank statistics under alternatives. *The Annals of Mathematical Statistics*, 39:325–346, 1968.
- M. Hallin. On Distribution and Quantile Functions, Ranks and Signs in R_d . *Working Papers ECARES*, 2017.
- M. Hallin and D. Paindaveine. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *The Annals of Statistics*, 30(4):1103 – 1133, 2002a.
- M. Hallin and D. Paindaveine. Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli*, 8(6):787 – 815, 2002b.

References III

- M. Hallin and D. Paindaveine. Optimal rank-based tests for homogeneity of scatter. *The Annals of Statistics*, 36(3):1261 – 1298, 2008.
- M. Hallin, E. del Barrio, J. Cuesta-Albertos, and C. Matrán. Distribution and quantile functions, ranks and signs in dimension d : A measure transportation approach. *The Annals of Statistics*, 49(2):1139 – 1165, 2021.
- R. Heller, Y. Heller, S. Kaufman, B. Brill, and M. Gorfine. Consistent distribution-free k -sample and independence tests for univariate random variables. *Journal of Machine Learning Research*, 17:1–54, 2016.
- K. H. Huang, X. Liu, A. B. Duncan, and A. Gandy. A high-dimensional convergence theorem for u -statistics with applications to kernel-based testing. In *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 3827–3918. Proceedings of Machine Learning Research, 2023.
- M. E. Jakobsen. Distance covariance in metric spaces: Non-parametric independence testing in metric spaces (master’s thesis), 2017.
- T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’02, page 133–142. Association for Computing Machinery, 2002.
- T. Joachims. Training linear svms in linear time. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’06, page 217–226. Association for Computing Machinery, 2006.
- E. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer, 2005.
- R. Y. Liu and K. Singh. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260, 1993.
- A. Lung-Yut-Fong, C. Lévy-Leduc, and O. Cappé. Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la société française de statistique*, 156(4):133–162, 2015.
- R. Lyons. Distance covariance in metric spaces. *The Annals of Probability*, 41(5):3284 – 3305, 2013.
- P. Major. An estimate on the supremum of a nice class of stochastic integrals and u -statistics. *Probability Theory and Related Fields*, 134(3):489–537, 2006.

References IV

- H. Mann and D. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18:50–60, 1947.
- A. K. Menon and R. C. Williamson. Bipartite ranking: A risk theoretic perspective. *Journal of Machine Learning Research*, 7:1–102, 2016.
- J. Möttönen and H. Oja. Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, 5(2):201–213, 1995.
- J. Möttönen, H. Oja, and J. Tienari. On the efficiency of multivariate spatial sign and rank tests. *The Annals of Statistics*, 25(2):542–552, 1997.
- J. Möttönen, H. Oja, and R. Serfling. Multivariate generalized spatial signed-rank methods. *Journal of Statistical Research*, 39(1):19–42, 2005.
- N. Neumeyer. A central limit theorem for two-sample u-processes. *Statistics and Probability Letters*, 67(1):73 – 85, 2004.
- D. Nolan and D. Pollard. U-Processes: Rates of Convergence. *The Annals of Statistics*, 15(2):780 – 799, 1987.
- H. Oja. Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1(6):327–332, 1983.
- A. Ramdas, N. Garcia, and M. Cuturi. On wasserstein two sample testing and related families of nonparametric tests, 2015.
- D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. Detecting novel associations in large data sets. *Science*, 334(6062): 1518–1524, 2011.
- D. N. Reshef, Y. A. Reshef, P. C. Sabeti, and M. Mitzenmacher. An empirical study of the maximal and total information coefficients and leading measures of dependence. *The Annals of Applied Statistics*, 12(1): 123 – 155, 2018.
- Y. A. Reshef, D. N. Reshef, H. K. Finucane, P. C. Sabeti, and M. Mitzenmacher. Measuring dependence powerfully and equitably. *Journal of Machine Learning Research*, 17(211):1–63, 2016.
- C. Rudin. Ranking with a P-Norm Push. In *Proceedings of COLT 2006*, volume 4005 of *Lecture Notes in Computer Science*, pages 589–604, 2006.

References V

- C. Rudin, C. Cortes, M. Mohri, and R. E. Schapire. Margin-based ranking and boosting meet in the middle. In *Proceedings of COLT 2005*, volume 3559 of *Lecture Notes in Computer Science*, pages 63–78. Springer, 2005.
- R. J. Serfling. *Approximation theorems of mathematical statistics*. John Wiley and Sons, 1980.
- H. Shi, M. Hallin, M. Drton, and F. Han. On universally consistent and fully distribution-free rank tests of vector independence. *The Annals of Statistics*, 50(4):1933 – 1959, 2022.
- C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.
- G. J. Székely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.
- J. W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, pages 523–531, 1975.
- S. van de Geer. *Empirical Processes in M-Estimation*. Cambridge University Press, 2000.
- A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1:80–83, 1945.