Universal aggregation of permutations

Ekhine Irurozki[1] and **Stephan Clemencon**[2]

Abstract. The estimation of the median and mean is a central problem in statistics. When it comes to permutation data the median estimation problem is called *aggregation problem*. The problem is challenging both computationally and statistically, since provable algorithms are known only for a hand-full of cases.

In this paper, we consider the permutation aggregation problem in a general probabilistic setting. This algorithm is given a data-set drawn form a distribution P and a choose for a distance function d and outputs the provably good estimator for the median of P in polynomial time. For the first time, one algorithm can be shown to work for different generating distributions.

1 Introduction

The permutation aggregation problem (PA) has been studied in different communities, such as Social Choice [2], Machine Learning [16] and Operation Research [1, 11], under a variety of names such as Rank Aggregation Problem, Linear Ordering Problem (LOP) or Kemeny ranking problem. A metric-based approach to PA looks for the permutation that minimizes the sum of the distances to a sample of permutations. For most choices of distances this problem is NP-hard [11, 12]. Statistical-based approaches characterize generating distribution for which algorithm converge towards correct estimators [16, 4]. In this paper, we take a statistical perspective to the permutation aggregation problem and propose an algorithm that is shown to recover the correct estimator with high probability in polynomial time when for several generating distributions of the given sample, thus the name *Universal* aggregation.

The *depth functions* offer theoretical foundation to extend statistical concepts of quantiles or robustness to multivariate data. Recently, the notion of depth has been generalized to permutation spaces and different applications have been studied. One of these applications is a rank aggregation algorithm that trims data points until a median trivially arises. In this paper, we improve this algorithm by (i) generalizing the distances and (ii) proposing a Markov Chain framework for the sake of analysis.

Problem statement: aggregating permutations Let P be a distribution for permutations centered at $\sigma_0 \in \mathfrak{S}_n$. The problem of aggregation consists on obtaining an estimator for σ_0 given sample S drawn from P.

Contribution We give an algorithm for the aggregation problem that returns an estimator for σ_0 with provable guarantees for different distributions P, in particular:

- We give the first provable aggregation algorithm for samples distributed according Plackett-Luce, Mallows and Generalized Mallows under Spearman's-ρ and Spearman's-footrule, Ulam and Cayley distances.
- We give a new provable algorithm for Mallows and Generalized Mallows under Kendall's-τ and Hamming distances.

2 Preliminaries

We start by recalling basic definition and previous results. Throughout the paper, permutations are denoted with Greek letters. The group operation is the composition $\sigma \cdot \pi$, denoted $\sigma\pi$, and the inverse of σ is denoted σ^{-1}. The group of permutations of n items is denoted Symmetric group, $\mathfrak{S}_n$.

2.1 Ranking models

Probability models for permutations [5] assign a probability value to each permutation $\sigma \in \mathfrak{S}_n$, $P : \mathfrak{S}_n \to \mathbb{R}_+$. The *universality* of the proposed algorithm comes for the fact that it can be shown to return the correct estimator with high probability for a number of distributions. In particular,we consider the distributions that satisfy the next property. Later, we give examples of distributions that satisfy it.

The next definition generalizes a property presented in [5] for distributions on permutations. Intuitively, it says that the probability P is non-increasing as it moves farther from the mode, along a certain type of path. The original definition considers transpositions, i.e., the Kendall's-τ distance, while the following definition extends it to Cayley and Hamming distances, i.e., swaps and insertions.

Definition 1 (STRONGLY-UNI-MODAL RANKING MODELS)**.** *Let P be a distribution for permutations with mode in σ_0. Let d be one of Kendall's-τ, Cayley or Ulam. Let τ be a permutation that moves from any permutation σ to another permutation $\sigma\tau$ at distance 1, i.e., $d(\sigma, \sigma\tau) = 1$ for all σ. Then P is strongly-uni-modal w.r.t. d if for every τ, σ, $d(\sigma_0, \sigma\tau) > d(\sigma_0, \sigma) \iff P(\sigma) \geq P(\sigma\tau)$.*

We now characterize the best know models under the extension of the strongly-uni-modal condition (summary in Table 1):

- The Plackett-Luce [18, 19], Mallows model based on Kendall's-τ, Spearman's-ρ and Spearman's-footrule and the Generalized Mallows model under the same distances (for a decreasing value of the dispersion parameters) [5, 9, 8, 12] are strongly-uni-modal w.r.t. the Kendall's-τ distance.
- The Mallows and Generalized Mallows (for a decreasing value of the dispersion parameters) under the Cayley [13, 12] and Hamming distances [14, 12] are strongly-uni-modal w.r.t. Cayley distance.
- The Mallows model under the Ulam [15, 12] distance is strongly-uni-modal w.r.t. Ulam distance.

[1] Telecom Paris, Institut Polytechnique de Paris, email: irurozki@telecom-paris.fr
[2] Telecom Paris, Institut Polytechnique de Paris, email: stephan.clemencon@telecom-paris.fr

	Kendall's-τ	Cayley	Ulam
MM Kendall GMM Kendall MM Spearman's-ρ MM Spearman's-footrule Plackett-Luce	X		
MM Cayley GMM Cayley MM Hamming GMM Hamming		X	
MM Ulam			X

Table 1. Summary of the pairs of distributions P and distances d such that P is strongly-uni-modal w.r.t. d.

The next definition describes a graph pn the permutations in $\mathfrak{S}_n$ where the adjacency matrix is parameterized by a distance, which again is one of Kendall's-τ, Cayley or Ulam. This graph will be the base for the Markov Chain.

Definition 2. *The Cayley-Graph* $\mathbf{\Gamma^d} = \Gamma(\mathfrak{S}_n, d)$ *is defined on* $\mathfrak{S}_n$*: each permutation* $\sigma \in \mathfrak{S}_n$ *is associated to a node; There is an edge in* $\sigma, \sigma' \iff d(\sigma, \sigma') = 1$. *It follows that the Cayley-Graph differs depending on the distance* d *considered.*
The three distances are editing distances *[6] w.r.t. three* generating sets *, i.e., inversions, swaps and insertions. These generating sets can be used in the definition of the Cayley-Graph to obtain equivalent definitions.*

2.2 Aggregation problem

Statistical approach The statistical framework for the aggregation problem assumes that there exists a generative distribution P for the data. We want to find the *median permutation*, defined as the σ^* that satisfies

$$L_P(\sigma^*) = \min_{\sigma \in \mathfrak{S}_n} L_P(\sigma), \tag{1}$$

where

$$L_P(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\sigma, \Sigma)] \tag{2}$$

and P is one of: Plackett-Luce, Mallows and Generalized Mallows under Spearman's-ρ and Spearman's-footrule, Kendall's-τ, Hamming, Ulam, Cayley distances. When P is strongly-uni-modal w.r.t. distance d its easy to see that the minimizer is σ_0, i.e., $\sigma^* = \sigma_0$. Full statistical problem analysis for Kendall's-τ is given in [16] and references therein. Moreover, its known that Borda algorithm [3] returns an unbiased estimator of the median permutation with low sample complexity and in quasi-linear time for data distributed according to MM under Kendall's-τ but it does not return an unbiased estimator when other distances are considered [4, 8]. The statistical problem under Hamming has also been considered in [14].

2.3 Depth functions

The concept of *statistical depth* was introduced to allow defining a center-outward ordering of points in the support of a probability distribution P on $\mathbb{R}^d$, so as to extend the notions of order and (signed) rank statistics, see *e.g.* and [17, 20]. This idea has recently been generalized to permutation spaces [10], and several depth properties (such as invariance, maximality at center and monotonicity) have been shown to be satisfied for distributions on permutations. A depth function for permutation data $D_P : \mathfrak{S}_n \to \mathbb{R}_+$ relative to $P : \mathfrak{S}_n \to \mathbb{R}_+$ should ideally assign the highest values $D_P(x)$ to points $x \in \mathfrak{S}_n$ near the "center" of the distribution. The next lines provide a formal description.

Algorithm 1: Universal permutation aggregation

Data: $\{\sigma_1, \ldots, \sigma_N\} \sim P$, a distance d
Result: $\widehat{\sigma}^*$

```
σ ← random permutation of n items;
while True do
    T ← {π : d(σ, π) = 1} ;
    if T is empty then return σ;
    ∀π ∈ T, with probability p_{σ,π} σ ← π ;
```

Definition 3. (METRIC-BASED RANKING DEPTH [10]) Let d be a distance and P a distribution on $\mathfrak{S}_n$. The ranking depth based on d is defined as: $D_P^{(d)}$: $\forall \sigma \in \mathfrak{S}_n$, $D_P^{(d)}(\sigma) = \mathbb{E}_P[||d||_\infty - d(\sigma, \Sigma)] = ||d||_\infty - L_P(\sigma)$, with $||d||_\infty = \max_{(\sigma,\sigma') \in \mathfrak{S}_n^2} d(\sigma, \sigma')$.

As the distribution P of interest is generally unknown in practice, its analysis relies on the observation of $N \geq 1$ independent realizations $X_1, \ \ldots, \ X_N$ of P. A statistical version of $D_P(x)$ can be built by replacing P with its empirical counterpart $\widehat{P}_N = (1/N)\sum_{i=1}^N \delta_{X_i}$, yielding the *empirical depth function* $D_{\widehat{P}_N}(x)$. Its consistency and asymptotic normality have been studied for various notions of depth, refer to *e.g.* [7, 21],

3 Contribution: Aggregation as a random process based on depths

In this section, we detail the main contribution of this paper, an algorithm for permutation aggregation with theoretical guarantees. The pseudo-code of our proposal is given in Algorithm 1. It takes as input (i) the data-set S i.i.d. with regard to P and (ii) a distance function d. This algorithm is guaranteed to output the median ranking of the data-set with high probability if P is strongly-uni-modal w.r.t. d.

Algorithm 1 is a random process defined by a Markov chain over the Cayley-Graph defined by d, Γ^d. The transition probability matrix assigns equal probability to all the neighbouring nodes that have a larger depth $\widehat{D}(\pi)$, i.e.,

$$p_{\sigma,\pi} = \frac{\mathbb{1}[\widehat{D}_N(\pi) - \widehat{D}_N(\sigma) > 0]}{|\sum_{\pi'} \mathbb{1}[\widehat{D}_N(\pi') - \widehat{D}_N(\sigma) > 0]|} \tag{3}$$

Esentially, the algorithm works as follows: It starts from a random permutation σ. It evaluates the neighbours permutations σ', i.e., the permutations at distance 1 from σ. It moves to the first neighbour with greater empirical depth $D(\sigma) < D(\sigma')$.

The next result summarizes the main contribution of this paper, the quality guarantees for Algorithm 1 and the characterization of the proper choice for the distance for convergence.

Theorem 4. (PROVABLE GUARANTEES FOR ALGORITHM 1) *Let* P *be strongly-uni-modal w.r.t. distance* d *and centered in* σ_0 *and let* $\{\sigma_1, \ldots, \sigma_N\} \sim P$*. Algorithm 1 based on distance* d *returns the median permutation* σ_0 *with high probability if* P *is strongly-uni-modal w.r.t.* d *(see summary in of these pairs in Table 1). Its time complexity is characterized in Lemma 6.*

To show this result, we first need to proof the following one. Theorem 5 shows that if we move far away from the mode with a particular type of move then the depth decreases. In particular, it characterizes the moves and the generating distribution for which this holds.

Theorem 5. (MONOTONICITY ON DEPTHS) *Let P be strongly-unimodal w.r.t. d and centered at σ_0. Let $\sigma, \sigma\tau \in \mathfrak{S}_n^2$ be two permutations at distance 1 from each other, $d(\sigma_0, \sigma) + 1 = d(\sigma_0, \sigma\tau)$. Let the depth function $D_P^{(d)}$ be based on the same distance. Then*

$$D_P^{(d)}(\sigma) > D_P^{(d)}(\sigma\tau).$$

Proof. This result was given in [10] for Kendall's-τ, here we extend it for Cayley, Hamming and Ulam. By definition, we have

$$D_P^{(d)}(\sigma) > D_P^{(d)}(\sigma\tau) \iff L_P(\sigma) < L_P(\sigma\tau) \tag{4}$$

Where the left hand side can be written as follows

$$L_P(\sigma) = \sum_{\pi \in \mathfrak{S}_n} p(\pi) d(\sigma, \pi) \tag{5}$$

and the right hand side as follows

$$\begin{aligned} L_P(\sigma\tau) &= \sum_{\pi \in \mathfrak{S}_n} p(\pi) d(\sigma\tau, \pi) \\ &= \sum_{\pi \in \mathfrak{S}_n^+} p(\pi)(d(\sigma\tau, \pi) + 1) + \sum_{\pi \in \mathfrak{S}_n^+} p(\pi)(d(\sigma\tau, \pi) - 1) \\ &= \sum_{\pi \in \mathfrak{S}_n} p(\pi) d(\sigma\tau, \pi) + \sum_{\pi \in \mathfrak{S}_n^+} p(\pi) - \sum_{\pi \in \mathfrak{S}_n^-} p(\pi) \end{aligned} \tag{6}$$

where $\mathfrak{S}_n^+$ is the subsets of $\mathfrak{S}_n$ closest to σ than to $\sigma\tau$ (respectively $\mathfrak{S}_n^-$ the farthest). Noting that for every $\pi \in \mathfrak{S}_n^+$ we can construct $\pi\tau \in \mathfrak{S}_n^-$ (and the other way around) and that $p(\pi) > p(\pi\tau)$ concludes that $L_P(\sigma\tau) > L_P(\sigma)$ and conclude. □

Theorem 5 ensures that for the specific choices of P and d in Table 1, convergence to the optimum is guaranteed, i.e., the moves on the Cayley-Graph that decrease the distance to σ_0 will increase the depth. In practice we use $\widehat{D}_N(\sigma)$ as an approximation of $D_P(\sigma)$, and it can be shown that $\sup_{\sigma \in \mathfrak{S}_n} |\widehat{D}_N(\sigma) - D_P(\sigma)|$ gets arbitrarily small. [10].

We can now provide the proof for Theorem 4.

Proof. (for THEOREM 4) Algorithm 1 visits a sequence of permutations with increasing depth $D_P^{(d)}$. By Theorem 5 this means that the same sequence is decreasing in distance d to the the median permutation iff the population is distributed according to P. We end this by referring to [10] for the learning rates $\sup_{\sigma \in \mathfrak{S}_n} |\widehat{D}_N(\sigma) - D_P(\sigma)|$ get arbitrarily small. □

Regarding the time complexity, we have the following result.

Lemma 6 (Time complexity)**.** *The naive time complexity of Algorithm 1 is $O(a \cdot b \cdot c)$ where a is the $\mathbb{E}_{\Sigma \in P}[d(\Sigma, \sigma)]$, b is the number of EOs for d, i.e., the number of permutations at distance 1, and c the cost of computing the depth under a d. Slight modifications over the trivial version yield a $O(n^3)$ for Kendall's-τ and $O(n^2 m)$ for the Cayley.*

4 Experiments

We illustrate the performance of the proposed Algorithm 1 for permutation aggregation via simulated experiments.

Setting 1 We draw a sample of $m = 50$ permutations of $n = \{30, 50\}$ from distribution P (Mallows under Cayley, Kendall's-τ, Hamming and Ulam). We run Algorithm 1 with distance d (one of Cayley, Kendall's-τ, Hamming and Ulam).

Reading the figures The results of this experiment is shown in Figure 1 for P being a Mallows model based on Kendall's-τ and a depth based on Kendall's-τ, Figure 2 for P being a Mallows model based on Cayley and a depth based on Cayley, Figure 3 for P being a Mallows model based on Ulam and a depth based on Ulam, and Figure 4 for P being a Mallows model based on Hamming and a depth based on Cayley. The figures represent the permutations in the dataset as blue points and the sequence of permutations visited by the algorithm as a red line.

Each permutation in the sample is plot in the point specified by its empirical depth (X-axis) and its distance to the median (Y-axis). We avoid plotting in the same figure every $\sigma \in \mathfrak{S}_n$ due to the large cardinality of $\mathfrak{S}_n$, which is $n!$. However, we claim based upon previous results that the empirical depth is close enough to the real one even with such small sample size $m << n!$, [10].

The permutation aggregation problem consists on obtaining the minimizer in

$$\arg\min_{\sigma \in \mathfrak{S}_n} \frac{1}{N} \sum_{\Sigma \in S} [d(\sigma, \Sigma)] \tag{7}$$

given a sample drawn from P as an approximant of the loss minimizer Equation (2). Since this in a synthetic experiment, σ_0 is known and so is $d(\sigma, \sigma_0)$ for all σ in the sample but, in general, its not known. In this paper, we present an algorithm for estimating σ_0 using the empirical depth $\widehat{D}_P^{(d)}$, which is always known provided a sample.

A baseline method for the aggregation probem is to get the deepest permutation in the sample S, given in Equation (**??**). The deepest permutation in each plot are the top-most permutations. This is justified by Theorem 5.

We describe now the random-chain of permutations visited by Algorithm 1 and displayed as red lines in the plots. For each sample we run the Algorithm 1 3 times and therefore, there are 3 red lines in each plot. Each algorithm run starts on a random permutation σ (Line 1 in the algorithm). It follows that this point is plotted as the bottom-most point of the red line, in $(d(\sigma, \sigma_0), \widehat{D}_P^{(d)}(\sigma))$. The sequence of permutations accepted by the algorithm (Line 5) is a sequence with increasing depth and (with high probability) with smallest distance to the median permutation σ^*. The sequence of permutations is plot in red in the same figures, starting in the bottom-right corner and increasing in depth. The algorithm halts when there is no permutation at distance 1 from the current one with larger depth, this point is displayed with a red cross.

Results Algorithm 1 returns the top-most permutation of the sequence (red cross). This algorithm returns the ground truth σ^* iff the sequence finishes at X-axis value equals 0. In all the cases the results improve on the baseline method that consists on returning the deepest permutation in the sample.

Setting 2: Is the choice of the distance really relevant in Algorithm 1? The next experimental setting is designed to show that even though the minimizer of all distance functions is the same in all cases it is still very relevant which distance you choose in Algorithm 1. This point has been already answered in a theoretical way in Theorem 5 and now we will provide empirical evidence.

Figures 1 corresponds to the case in which the P is strongly-uni-modal w.r.t. distance d_1 and the depth in Algorithm 1 is a different distance d_2. In all the figures, we see, as before, the sequence of visited permutations in a red line. However, we now see that a increase in depth does not necessarily imply a decrease in the distance to the ground truth σ^*, i.e., the red line is not now diagonally increasing from right to left, but instead it increases vertically.

Moreover, we can see that in all the 3 cases the deepest permutation (provided we choose the correct distance) is better than the algorithm returned by Algorithm 1 when the distance chosen in not strongly-uni-modal w.r.t. the distribution that generated the permutation.

The takeaway for this experiment is that the choose of the distance in Algorithm 1 is not arbitrary, and we have this point shown both theoretically and empirically.

5 Conclusions and future work

In this paper, we have proposed the first permutation aggregation algorithm in a universal setting, i.e., for a wide number of generating distributions. For some distributions, such as the Plackett-Luce and Mallows under Hamming and Ulam this is the first algorithm with guarantees.

A natural improvement is to analyse a randomized version of the algorithm, i.e., to characterize the number of times one has to run the algorithm in order to hit the optimum with arbitrary large probability.

The problem of the robustness of the estimators is intrinsic to any estimation algorithm. Depths are a classical tool for analyzing and improving robustness. Indeed, the idea of trimming data points based on their depth has already been studied for permutation problems [10]. We plan to extend our analysis to a fully stochastic approach in which we use a re-sampling scheme of the sample with probability proportional to this empirical depths to increase robustness.

In a broader perspective, this paper opens a new line in the analysis of combinatorial optimization problems. In the absence of a well-established notion of convergence in functions for permutation spaces, depths are a tool to analyze performance of combinatorial algorithm. Indeed, the algorithm seen here can be seen as Local search algorithm for the Linear Ordering Problem (LOP) or Kemeny problem. Under this perspective, the results in this paper ensure that there exists optimality conditions that depend on the instance and the neighbourhood system that can be easily identified.

REFERENCES

[1] Alnur Ali and Marina Meila, 'Experiments with kemeny ranking: What works when?', *Mathematical Social Sciences*, **64**, 28–40, (2012). Computational Foundations of Social Choice.

[2] J. J. Bartholdi, C. A. Tovey, and M. A. Trick, 'The computational difficulty of manipulating an election', *Social Choice and Welfare*, (1989).

[3] J. Borda, *Memoire sur les Elections au Scrutin.*, Histoire de l'Academie Royal des Sciences., 1781.

[4] I. Caragiannis, A. D. Procaccia, and N. Shah, 'When do noisy votes reveal the truth?', in *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, pp. 143–160, New York, (2013). ACM.

[5] D. E. Critchlow, M. A. Fligner, and J. S. Verducci, 'Probability models on rankings', *Journal of Mathematical Psychology*, **35**(3), 294–318, (1991).

[6] Michel Marie Deza and Elena Deza, *Encyclopedia of Distances*, Springer-Verlag, 2009.

[7] D. L. Donoho and M. Gasko, 'Breakdown properties of location estimates based on halfspace depth and projected outlyingness', *The Annals of Statistics*, **20**, 1803–1827, (1992).

[8] M. A. Fligner and J. S. Verducci, 'Multistage ranking models', *Journal of the American Statistical Association*, **83**(403), 892–901, (1988).

[9] Michael A Fligner and Joseph S Verducci, 'Distance based ranking models', *Journal of the Royal Statistical Society*, **48**(3), 359–369, (1986).

[10] Morgane Goibert, Stéphan Clémençon, Ekhine Irurozki, and Pavlo Mozharovskyi, 'Statistical depth functions for ranking distributions: Definitions, statistical learning and applications', in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, (2022).

[11] O. Hudry, 'NP-hardness results for the aggregation of linear orders into median orders', *Annals of Operations Research*, **163**, 63–88, (2008).

[12] E. Irurozki, B. Calvo, and J. A. Lozano, 'PerMallows: An R package for mallows and generalized mallows models', *Journal of Statistical Software*, **71**, (2019).

[13] E. Irurozki, B. Calvo, and J.A. Lozano, 'Sampling and learning mallows and generalized mallows models under the cayley distance', *Methodology and Computing in Applied Probability*, (2016).

[14] E. Irurozki, B. Calvo, and J.A. Lozano, 'Mallows and generalized Mallows model for matchings', *Bernoulli*, **25**(2), (2019).

[15] Ekhine Irurozki, Borja Calvo, and Jose A. Lozano, 'Mallows model under the ulam distance: a feasible combinatorial approach', (2014).

[16] Anna Korba, Stephan Clémençon, and Eric Sibony, 'A learning theory of ranking aggregation', (2017).

[17] K. Mosler, 'Depth statistics', in *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*, eds., C. Becker, R. Fried, and S. Kuhnt, 17–34, Springer, (2013).

[18] R L Plackett, 'The analysis of permutations', *Journal of the Royal Statistical Society*, **24**, 193–202, (1975).

[19] Duncan Luce R., *Individual Choice Behavior*, John Wiley and Sons, 1959.

[20] J. W. Tukey, 'Mathematics and the picturing of data', in *Proceedings of the International Congress of Mathematicians*, ed., R. D. James, volume 2, pp. 523–531. Canadian Mathematical Congress, (1975).

[21] Y. Zuo and R. Serfling, 'Structural properties and convergence results for contours of sample statistical depth functions', *The Annals of Statistics*, **28**(2), 483–499, (2000).

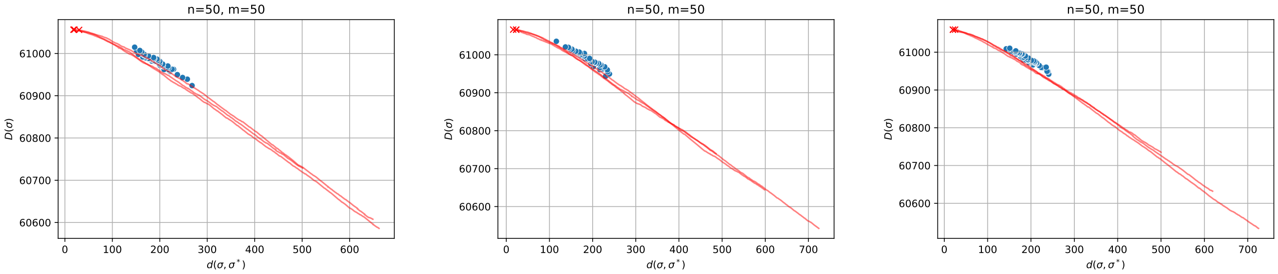

Figure 1. Distribution with strongly-uni-modal w.r.t. the Kendall's-τ distance and depth based on Kendall's-τ.

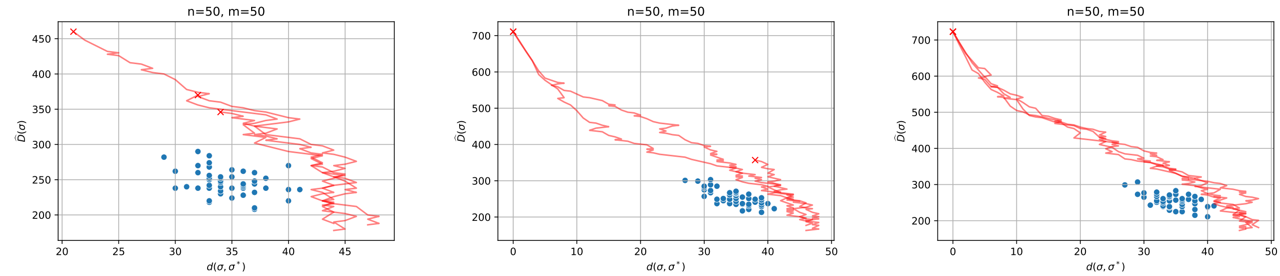

Figure 2. Distribution with strongly-uni-modal w.r.t. the Cayley distance and depth based on Cayley.

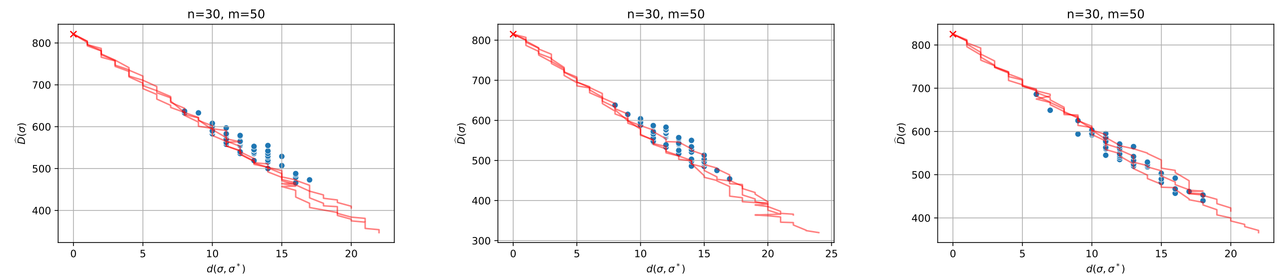

Figure 3. Distribution with strongly-uni-modal w.r.t. the Ulam distance and depth based on Ulam.

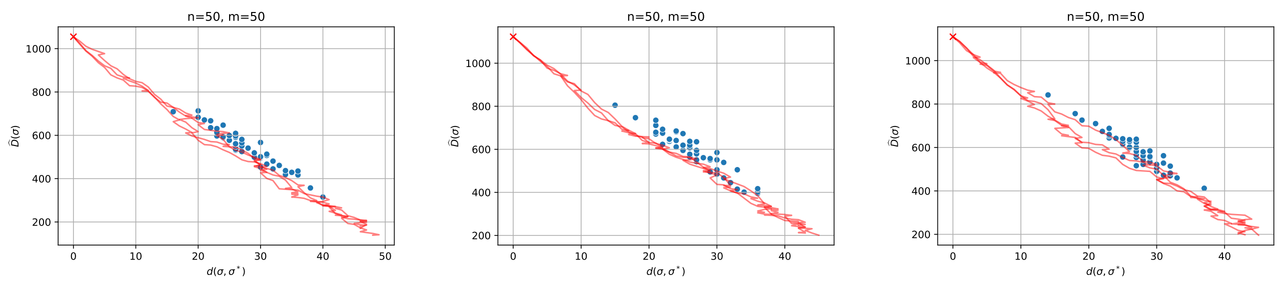

Figure 4. Distribution with strongly-uni-modal w.r.t. the Hamming distance and depth based on Cayley.

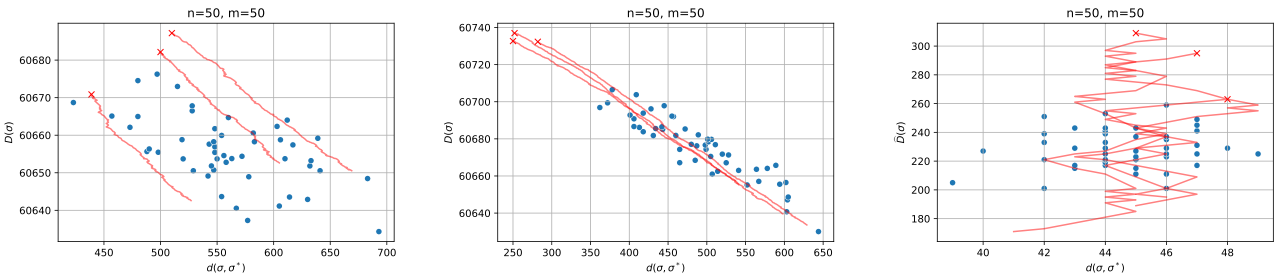

Figure 5. Distribution with strongly-uni-modal w.r.t. the a distance different from the one in the depth.