



A weighted distance metric for assessing ranking dissimilarity and inter-group heterogeneity

Amalia Vanacore¹ · Maria Sole Pellegrino¹

Received: 30 September 2020 / Accepted: 15 January 2021 / Published online: 24 February 2021
© The Author(s) 2021

Abstract

In this paper, a weighted variant of the normalized pairwise angular distance metric is proposed. The inclusion of position weights aims at penalizing inversions in the top of the ranking more than inversions in the tail of the ranking. The performance of the proposed weighted distance metric for assessing ranking dissimilarity and its impact on a procedure for testing inter-group heterogeneity have been investigated via a Monte Carlo simulation study under several scenarios—differing for group size, number of ranked alternatives and system of hypotheses—and compared against those obtained for the unweighted variant.

Keywords Ranking dissimilarity · Weighted distance metric · Inter-group heterogeneity · Power analysis

1 Introduction

Analyzing consumer satisfaction by listening to their needs is fundamental in any private and public sector since it represents a useful tool for improving the efficacy of management and policy actions [5]; it becomes critical when the consumers cannot switch to other providers, deny or reduce the service [8].

Satisfaction data are usually collected through opinion survey questionnaires where each respondent is asked to value a set of attributes independently of one another and rate them accordingly, in such a way that no direct comparison is made between attributes. The widespread adoption of ordered categorical response scales is due to two factors: they are ease to manage and low time-consuming [1, 12, 16, 17]. Despite these practical advantages, this evaluation method requires a specific expertise that some respondents may lack producing two main criticisms: response biases like agreement response style (tendency to always agree with every statement irrespective of its content) and non-differentiation (tendency to not really differentiate between the statements irrespective of their content) [1, 12].

✉ Amalia Vanacore
amalia.vanacore@unina.it

¹ Department of Industrial Engineering, University of Naples “Federico II”, p.le Tecchio 80, 80125 Naples, Italy

A valid alternative solution stands in elicitation methods based on ranking according which each respondent is asked to sort a set of alternatives with respect to some criteria (e.g. liking, satisfaction, agreement, importance). Such elicitation methods give high quality and informative data, high test–retest and cross-sectional reliability and high discriminate and correlation validity [11,17,20]; nevertheless, their widespread adoption is limited by the higher availability of suitable statistical methods for handling ordered categorical response data.

When dealing with ranking data, two typical issues of interest are to assess the diversity between two subjects expressing their preferences over the same set of alternatives and compare two or more groups of subjects (e.g. men and women, young and elderly, people from different countries) in order to investigate concordance, agreement or homogeneity among them and identify the significant factors impacting on subject preferences and, possibly, cluster the subjects [15].

Different proximity and distance measures have been proposed in the literature over the years for assessing the similarity between two rankings. A widespread approach in ranking data analysis uses similarity measures based on correlation coefficients and derives distance measures via linear transformations. Among these methods, Kendall's τ rank correlation coefficient [9,10], Pearson correlation coefficient and Spearman's ρ are the most widely used. Other typical examples of distances are Kendall, Spearman and Cayley's distances. See Mallows [14], Critchlow et al. [2] and Diaconis [4] for more details; moreover, an extensive listing of such measures classified according to the main field of application or scientific area of interest is available in Deza and Deza [3].

A different approach to measure distances between rankings stems from multidimensional geometry, according which each ranking is a vector in the multidimensional Euclidean space and the distance between two of them is measured via the pairwise angular distance. Following the multidimensional geometry approach, in this paper the rankings are analyzed and interpreted using a simple distance metric based on cosine similarity, that is the normalized pairwise angular distance metric [21].

It is noteworthy that, as recently pointed out by Kumar and Vassilvitskii [13], the traditional similarity and distance measures between rankings do not care about two crucial concepts: (1) some alternatives could be more important than others, so that swapping equally important alternatives should be less penalizing than swapping not equally important alternatives; (2) swapping alternatives belonging to the top of the ranking could be more relevant than swapping alternatives belonging to the tail of the ranking. In order to overcome the two criticisms, Kumar and Vassilvitskii [13] defined the *element weights* and *position weights* for handling the first and second criticism, respectively.

The aim of this paper is to propose a variant of the normalized pairwise angular distance metric [21,23] suitable for position weighted rankings that penalizes inversions in the top of the ranking more than inversions in the tail of the ranking. The proposed weighted distance metric will be adopted for testing preference heterogeneity across groups of subjects via an inferential procedure based on the index of segregation power I_{SP} introduced by Gadrach et al. [6] for numerical and categorical variables and further specialized by Vanacore et al. [23] to complete rankings.

The performance of the testing procedure based on I_{SP} with the weighted normalized angular distance metric is investigated via a Monte Carlo simulation study under different scenarios, differing for group size, number of ranked alternatives and system of hypotheses. The results of the Monte Carlo simulation study are compared against those obtained for the normalized angular distance metric discussed in Vanacore and Pellegrino [22].

The paper is organized as follows: the weighted version of the normalized pairwise angular distance metric together with the testing procedure are introduced in Sect. 2; the design and the main results of the Monte Carlo simulation study are described in Sect. 3; finally, conclusions are summarized in Sect. 4.

2 Weighted normalized pairwise angular distance metric

Let M subjects rank a finite set $A = \{A_1, A_2, \dots, A_n\}$ of n alternatives. Each ranking can be represented as an antisymmetric square alternative-to-alternative matrix $(a_{lm})_{n \times n}$ of order n . The generic element a_{lm} of the (l, m) cell represents the preference relation between the ranked alternatives A_l and A_m and thus it assumes the following values:

$$a_{lm} = \begin{cases} -1 & \text{if the alternative } A_l \text{ precedes the alternative } A_m \\ 0 & \text{if } A_l \text{ and } A_m \text{ are indistinguishable alternatives} \\ 1 & \text{if the alternative } A_l \text{ follows the alternative } A_m \end{cases}$$

A suitable distance measure between two rankings comes from multidimensional geometry, according which each ranking, and thus matrix $(a_{lm})_{n \times n}$, can be considered as a vector in a multidimensional Euclidean space. A measure of dissimilarity between two rankings is the normalized pairwise angular distance metric [21]. Specifically, let \mathbf{a} and \mathbf{b} be two such vectors, the normalized pairwise angular distance metric $L_{\mathbf{ab}}$ is expressed as follows:

$$L_{\mathbf{ab}} = \frac{1}{\pi} \cdot \arccos \left[\frac{\sum_{l=1}^{n-1} \sum_{m=l+1}^n a_{lm} b_{lm}}{\sqrt{\sum_{l=1}^{n-1} \sum_{m=l+1}^n a_{lm}^2} \sqrt{\sum_{l=1}^{n-1} \sum_{m=l+1}^n b_{lm}^2}} \right] = \frac{\hat{\theta}_{\mathbf{a}, \mathbf{b}}}{\pi} \tag{1}$$

where π is the straight angle against which the angular distance $\hat{\theta}_{\mathbf{a}, \mathbf{b}}$ is normalized.

Position weighted distances are built to enhance the reactivity to the role of the position of each alternative, that is swaps at the top of the ranking are costlier than those at the tail of the ranking. The weighted variant of the normalized pairwise angular distance metric can be formulated as follows:

$$L_{\mathbf{ab}_w} = \frac{1}{\pi} \cdot \arccos \left[\frac{\sum_{l=1}^{n-1} \sum_{m=l+1}^n a_{lm_w} b_{lm_w}}{\sqrt{\sum_{l=1}^{n-1} \sum_{m=l+1}^n a_{lm_w}^2} \sqrt{\sum_{l=1}^{n-1} \sum_{m=l+1}^n b_{lm_w}^2}} \right] = \frac{\hat{\theta}_{\mathbf{a}_w, \mathbf{b}_w}}{\pi} \tag{2}$$

where a_{lm_w} (with $l = 1, \dots, n - 1; m = l + 1, \dots, n$) is the position weighted version of the generic element a_{lm} of the antisymmetric square alternative-to-alternative matrix $(a_{lm})_{n \times n}$. Specifically, let $w = (w_1, w_2, \dots, w_n)$ be the vector of non-increasing weights for each ranking position, the generic elements a_{lm_w} and b_{lm_w} are given by:

$$a_{lm_w} = a_{lm} \cdot w_l; \quad b_{lm_w} = b_{lm} \cdot w_l \tag{3}$$

It is worth to pinpoint that the square alternative-to-alternative matrix $(a_{lm})_{n \times n}$ is antisymmetric so that, in order to reduce the computational burden, Eqs. 1 and 2 are applied only on the elements of the strictly upper triangular matrix. Anyway, the weighted square alternative-to-alternative matrix can be easily obtained by filling in the empty cells with their antisymmetric value.

The position weights, modeling the cost of swapping two alternatives according to their position in the ranking, can be arbitrary defined, constrained to $\sum_l w_l = 1$. The position weighting scheme here adopted, derived from the radical weighting scheme adopted in

Table 1 Estimates of total ranking variation and its components

Variation	Formula	<i>df</i>
\hat{V}_k	$\frac{1}{M^2} \cdot \sum_{i,j=1}^M L_{i_k j_k}$	$M - 1$
\hat{V}_{TOT}	$\frac{1}{(MK)^2} \cdot \sum_{i,j=1}^M \sum_{h,k=1}^K L_{i_h j_k}$	$MK - 1$
\hat{V}_{WG}	$\frac{1}{K} \cdot \sum_{k=1}^K \hat{V}_k$	$K(M - 1)$
\hat{V}_{BG}	$\hat{V}_{TOT} - \hat{V}_{WG}$	$M(K - 1)$

agreement studies [7], is proportional to the distance between positions and is formulated as follows:

$$w_l = 1 - \sqrt{\frac{l - 1}{n - 1}}. \tag{4}$$

In order to compare several groups of subjects for testing ranking heterogeneity, the total variation among the rankings provided by all subjects (i.e. K groups of M subjects) \hat{V}_{TOT} , formulated as proposed in Vanacore et al. [23], is split—according to Rao’s apportionment of diversity [19]—into within \hat{V}_{WG} and between group \hat{V}_{BG} components. The inter-group ranking heterogeneity is measured through the index of segregation power I_{SP} [6], defined as the quotient of normalized between-group variation (\hat{V}_{BG}) to normalized total variation (\hat{V}_{TOT}):

$$I_{SP} = \frac{\hat{V}_{BG}/df_{BG}}{\hat{V}_{TOT}/df_{TOT}}. \tag{5}$$

Let $L_{i_h j_k}$ be the distance between the rankings provided by i^{th} subject belonging to group h and j^{th} subject belonging to group k , the total ranking variation and its components are formulated as shown in Table 1. The readers can refer to [22] for more computational details.

It is worth noting that although $\hat{V}_{BG}/\hat{V}_{TOT}$ is always less than 1, the I_{SP} value could exceed 1 since the ratio $df_{TOT}/df_{BG} = (MK - 1)/(MK - M) \geq 1$ with $M \geq 2$.

3 Monte Carlo simulation

The Monte Carlo simulation study aims at investigating: (1) the performance of the weighted normalized angular distance metric in discriminating dissimilarity between rankings; (2) the impact of the position weights on the performance of the procedure for testing preference heterogeneity. The results are compared against those obtained for the normalized angular distance metric without position weights. The algorithm for Monte Carlo simulations have been implemented using Mathematica (Version 11.0, Wolfram Research, Inc., Champaign, IL, USA).

3.1 Discriminating performance of the weighted normalized angular distance metric

Let us consider two subjects i and j expressing their preferences about $n = 5, 6, 7, 8, 9, 10$ alternatives; the $n! \times n!$ normalized pairwise angular distances L_{ij} and L_{ij_w} are represented in the box-plots of Fig. 1. It is evident that the introduction of position weights produces a wider range of distance values, allowing to better discriminate dissimilarity between rankings.

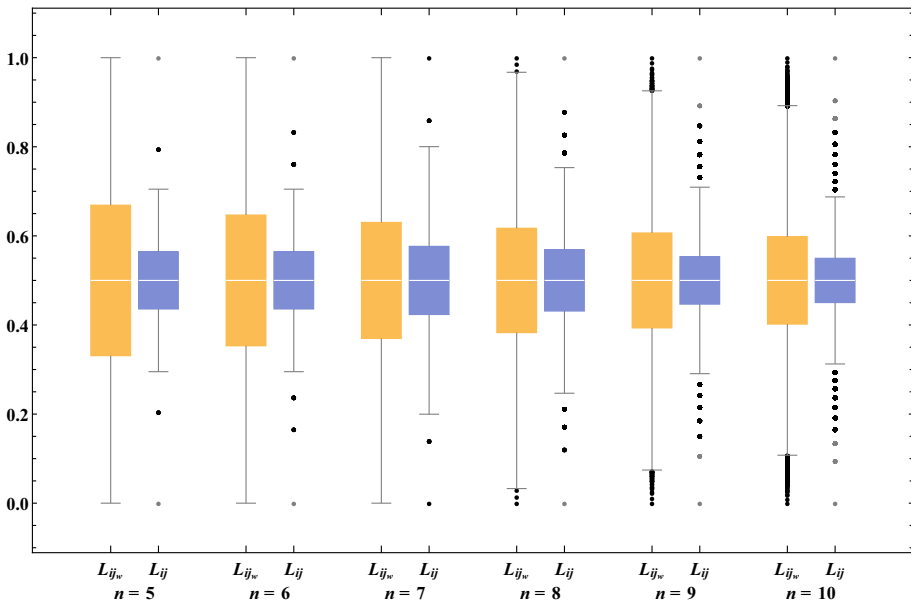


Fig. 1 Box-plot of weighted (in orange) and unweighted (in blue) normalized pairwise angular distance metric for n alternatives

The discriminating performance of L_{ij_w} has been investigated and compared against that of L_{ij} focusing on rankings differing each other for only one swap. Specifically, two factors have been considered: swap position with two levels (i.e. at the top and at the tail of the ranking) and swap dimension with four levels (i.e. between adjacent, 1-step apart, 2-steps apart and 3-steps apart positions), for a total of $2 \times 4 = 8$ simulated scenarios for $n = 5, 6, 7, 8, 9, 10$ alternatives. The distance values L_{ij} and L_{ij_w} are plotted in Fig. 2 for all the investigated scenarios.

The plotted curves show that both L_{ij} and L_{ij_w} decrease with the number of ranked alternatives and increase with the swap dimension; however, it is worth to note that only the weighted normalized angular distance metric takes into account the position of the swap. Indeed, for any swap dimension, L_{ij_w} is about three times higher if the swap is at the top preferences rather than if it is at the tail; vice-versa, L_{ij} does not depend on swap position, being its value the same either the swap is at the top or at the tail of the ranking.

3.2 Statistical power of the inter-group preference heterogeneity testing procedure

Let us consider M subjects expressing their preferences about n alternatives. In order to simulate different levels of inter-group heterogeneity, preferences have been generated using the distance-based model developed by Diaconis [4], which assumes that the probability of each ranking depends on its distance to a chosen modal ranking, to which it is expected most of rankings are close.

Different samples of rankings have been simulated by varying either the modal ranking or the dispersion parameter λ_k for the k^{th} group of subjects; specifically, λ_k controls the probability of each ranking and higher differences among λ_k produce higher heterogeneity among the rankings provided by different groups of subjects.

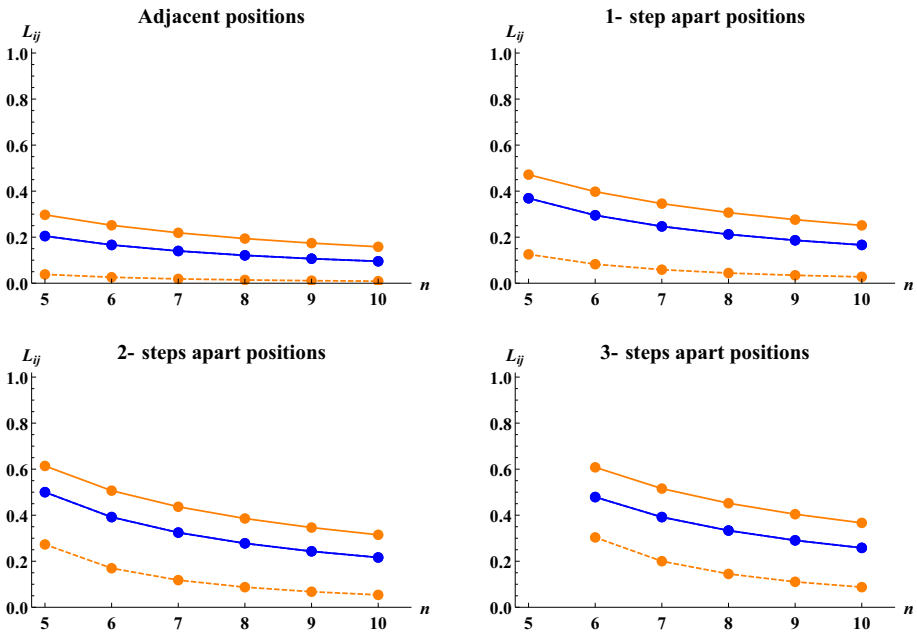


Fig. 2 Weighted (in orange) and unweighted (in blue) normalized pairwise angular distance metric for 1 swap of adjacent, 1-step apart, 2-steps apart and 3-steps apart positions at the top (solid lines) or at the tail (dashed lines) of the ranking of n alternatives

The case of $K = 3$ groups of M subjects has been considered and 108 scenarios have been simulated differing for group size, number of alternatives to rank and system of hypotheses. The group sizes simulated for this study are $M = 10, 20, 50$ subjects and the number of alternatives to rank are $n = 5, 10$; these values have been selected in order to represent a range of values that might be seen in practice in a survey [18].

The groups have been firstly simulated having all the same dispersion parameter λ_k , chosen in the range $[1 \div 50]$, but different modal ranking with one swap at the top of the ranking (i.e. first simulation setting). Specifically, 7 different λ_k values have been considered, that is $\lambda_k = 1, 3, 5, 10, 20, 35, 50$, so as to represent as many alternative hypotheses of growing levels of inter-group ranking heterogeneity.

Then, the groups have been simulated each with a different dispersion parameter λ_k chosen in the range $[1 \div 30]$ but having all the same modal ranking (i.e. second simulation setting). Specifically, 11 combinations of different values of λ_k have been considered so as to represent as many alternative hypotheses H_1 (see Table 2) of growing levels of inter-group ranking heterogeneity, obtained by choosing combinations of λ_k values characterized by increasing values of maximum difference in ranking dispersion ($\Delta\lambda_k$).

For each scenario, $R = 2000$ data sets with K groups of M rankings have been generated, the distances between rankings have been measured via the weighted and unweighted normalized pairwise angular distance metrics [21] and the I_{SP} has been assessed.

The simulation procedure has been developed through the following steps:

1. set a modal ranking and the dispersion parameters λ_k for each of the K groups of size M ($M = 10, 20, 50$);

Table 2 Values of λ_k identifying the nine alternative hypotheses H_1 of the second simulation setting

$(\lambda_1, \lambda_2, \lambda_3)$	(1,2,4)	(1,5,6)	(1,4,8)	(1,5,10)	(1,5,12)	(5,10,18)	(5,10,20)	(1,5,18)	(1,5,20)	(5,10,30)	(1,5,30)
$\Delta\lambda_k$	3	5	7	9	11	13	15	17	19	25	29

Table 3 Critical values I_{cr} of the indicator I_{SP} for the significance level $1 - \alpha = 0.95$, computed with and without position weights considering $K = 3$ groups of M subjects ranking n alternatives

$\lambda_1 = \lambda_2 = \lambda_3 = 1$	M = 10	M = 20	M = 50	M = 10	M = 20	M = 50
$n = 5$	1.63	1.67	1.70	1.53	1.53	1.53
$n = 10$	1.34	1.36	1.36	1.28	1.28	1.26

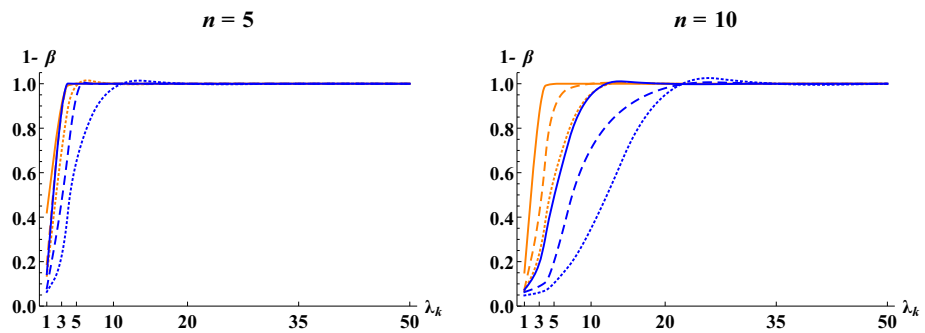


Fig. 3 Statistical power curves when testing alternative hypotheses H_1 with increasing dispersion parameter λ_k against the null hypothesis $H_0: \lambda_1 = \lambda_2 = \lambda_3 = 1$, for $n = 5$ and $n = 10$ alternatives, $M = 10$ (short-dashed lines), $M = 20$ (long-dashed lines) and $M = 50$ (solid lines) subjects, with (in orange) and without (in blue) position weights

2. sample the rankings provided by the M subjects of the K groups over n alternatives according to the framework of distance-based models;
3. compute the weighted normalized distance L_{ijw} for each pair of subjects (Eq. 2);
4. assess the total ranking variation and its within and between components (see Table 1);
5. compute the I_{SP} index (Eq. 5);
6. repeat R times steps 1–5;
7. for a significance level α , define the critical value I_{cr} as the $(1 - \alpha)$ percentile of the empirical sampling distribution of I_{SP} built under the assumption of homogeneity;
8. compute the statistical power for each alternative hypothesis as:

$$1 - \beta = \frac{1}{R} \sum_{r=1}^R I[I_{SP_r} > I_{cr}|H_1]; \tag{6}$$

9. compute the unweighted normalized angular distance metric L_{ij} for each pair of subjects (Eq. 1) with the outputs of step 2;
10. repeat steps 4–8 adopting the weighted distance values L_{ijw} .

For $\alpha = 0.05$, the values of I_{cr} for the null hypothesis $H_0: \lambda_1 = \lambda_2 = \lambda_3 = 1$ are reported in Table 3; whereas the values of statistical power for the two testing procedures for each tested hypothesis H_1 are reported in Tables 4 and 5, respectively.

Table 4 Statistical power when testing alternative hypotheses with increasing dispersion parameter λ_k for different n and M values with (on the left) and without (on the right) position weights against the null hypothesis $H_0: \lambda_1 = \lambda_2 = \lambda_3 = 1$

	λ_k	With position weights			Without position weights		
		M = 10	M = 20	M = 50	M = 10	M = 20	M = 50
$n = 5$	1	0.136	0.189	0.421	0.064	0.080	0.144
	3	0.697	0.956	1.000	0.225	0.472	0.969
	5	0.991	1.000	1.000	0.650	0.942	1.000
	10	1.000	1.000	1.000	1.000	1.000	1.000
	20	1.000	1.000	1.000	1.000	1.000	1.000
	35	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000
$n = 10$	1	0.067	0.076	0.151	0.048	0.063	0.073
	3	0.235	0.416	0.865	0.061	0.089	0.192
	5	0.575	0.885	0.999	0.103	0.171	0.491
	10	0.994	1.000	1.000	0.349	0.708	1.000
	20	1.000	1.000	1.000	0.978	1.000	1.000
	35	1.000	1.000	1.000	1.000	1.000	1.000
	50	1.000	1.000	1.000	1.000	1.000	1.000

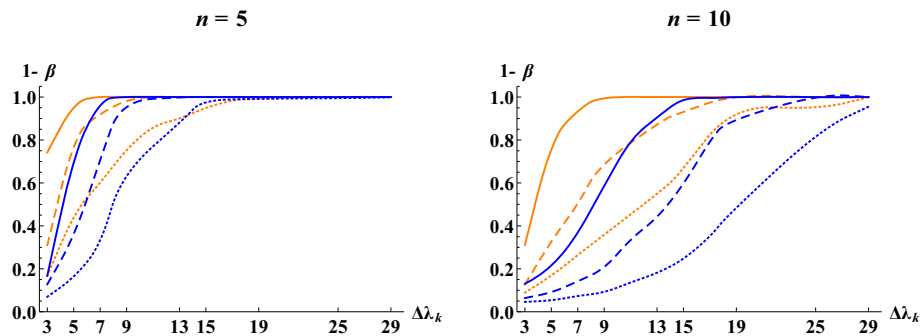


Fig. 4 Statistical power curves when testing alternative hypotheses H_1 with increasing difference in ranking dispersion (i.e. $\Delta\lambda_k$), for $n = 5$ and $n = 10$ alternatives, $M = 10$ (short-dashed lines), $M = 20$ (long-dashed lines) and $M = 50$ (solid lines) subjects, with (in orange) and without (in blue) position weights

The results highlight that the adoption of position weights for assessing the dissimilarity between rankings makes the statistical power of the testing procedure for inter-group heterogeneity increase. The improving power rate is more evident when the modal ranking changes across groups (i.e. first simulation setting) rather than with groups having the same modal ranking but different dispersion parameters (i.e. second simulation setting).

In the first simulation setting (see Fig. 3 and Table 4), for $n = 5$ a small difference in power rates is observed between weighted and unweighted distances; in both cases the statistical power is higher than 80% with small and homogeneous groups of subjects (i.e. starting from a λ_k value equal to 3 for $10 < M < 20$). The difference in power rates between weighted and unweighted distances increases with the number of alternatives. For $n = 10$, when the

Table 5 Statistical power when testing alternative hypotheses with increasing difference in ranking dispersion ($\Delta\lambda_k$) against the null hypothesis $H_0: \lambda_1 = \lambda_2 = \lambda_3 = 1$ for different values of n and M with (on the left) and without (on the right) position weights

	$\Delta\lambda_k$	With position weights			Without position weights		
		$M = 10$	$M = 20$	$M = 50$	$M = 10$	$M = 20$	$M = 50$
$n = 5$	3	0.167	0.310	0.742	0.070	0.128	0.168
	5	0.439	0.766	0.995	0.165	0.363	0.695
	7	0.572	0.919	1.000	0.334	0.706	0.986
	9	0.752	0.981	1.000	0.631	0.953	1.000
	11	0.857	1.000	1.000	0.881	0.998	1.000
	13	0.901	1.000	1.000	0.975	1.000	1.000
	15	0.949	1.000	1.000	0.992	1.000	1.000
	17	0.983	1.000	1.000	0.995	1.000	1.000
	19	0.991	1.000	1.000	0.999	1.000	1.000
	25	0.997	1.000	1.000	1.000	1.000	1.000
	29	1.000	1.000	1.000	1.000	1.000	1.000
$n = 10$	3	0.089	0.127	0.312	0.046	0.064	0.130
	5	0.169	0.326	0.746	0.054	0.092	0.215
	7	0.264	0.505	0.933	0.073	0.142	0.370
	9	0.399	0.715	0.994	0.093	0.209	0.586
	11	0.530	0.874	0.999	0.136	0.334	0.790
	13	0.443	0.842	1.000	0.182	0.443	0.935
	15	0.555	0.931	1.000	0.248	0.598	0.986
	17	0.861	0.998	1.000	0.353	0.792	0.995
	19	0.919	0.999	1.000	0.484	0.893	1.000
	25	0.953	1.000	1.000	0.815	0.997	1.000
	29	0.999	1.000	1.000	0.955	1.000	1.000

position weights are included the statistical power is higher than 80% with homogeneous groups of subjects (i.e. starting from $\lambda_k = 3$ for $M = 50$ or $\lambda_k = 5$ for $10 < M < 20$); whereas, when the weights are not included, the statistical power reaches 80% only with more heterogeneous groups of subjects (i.e. $\lambda_k \geq 10$).

The results obtained under the second simulation setting (see Fig. 4 and Table 5) reveal that in the case of $n = 5$ alternatives the increase in statistical power due to the inclusion of position weights is more evident with groups of $M > 10$ subjects; whereas in the case of $n = 10$ alternatives the increase in statistical power is evident for any group size.

For $n = 5$ alternatives, the adoption of position weights makes the testing procedure adequately powered in detecting inter-group ranking heterogeneity even for small group size (i.e. $M = 10$) with $\Delta\lambda_k \approx 10$. In scenarios with a fairly large set of alternatives, like $n = 10$, the statistical power of the testing procedure slightly worsens and with small groups of $M = 10$ subjects it reaches 80% when $\Delta\lambda_k \approx 17$. Nevertheless, the statistical power can be improved by increasing the group dimension, reaching 80% with groups of $M = 50$ subjects with $\Delta\lambda_k \approx 4$ for $n = 5$ or with $\Delta\lambda_k \approx 6$ for $n = 10$.

4 Conclusions

In this paper a position weighted variant of the normalized pairwise angular distance metric has been proposed.

The performance of the proposed metric in assessing ranking dissimilarity and inter-group heterogeneity has been investigated by an extensive Monte Carlo simulation study. The critical values of the I_{SP} index under the null hypothesis of inter-group ranking homogeneity have been obtained for different scenarios and the statistical power has been computed for increasing level of heterogeneity by varying either the modal ranking or the dispersion parameter.

The simulation results have been compared against those obtained for the unweighted distance metric. L_{ij} and L_{ijw} decrease with the number of alternatives to rank and are both positively related to swap dimension; only L_{ijw} accounts for the position of the swap taking on higher values if the swaps are at the top of the ranking. Whatever the adopted metric, the statistical power of the testing procedure improves for increasing group size and decreasing number of alternatives to rank and it reaches 80% even with as few as $M = 10$ subjects ranking $n = 5$ alternatives. The inclusion of position weights makes the performance of the inter-group heterogeneity testing procedure improve; the difference in statistical power is more evident when the groups have different modal ranking and the same dispersion parameter rather than when the groups have the same modal ranking but differ for the dispersion parameter.

Funding Open access funding provided by Università degli Studi di Napoli Federico II.

Compliance with ethical standards

Conflict of interest The authors declare that they have no potential conflict of interest in relation to the study in this paper.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication The article has not been previously published elsewhere and it is not under consideration for publication in another journal; both authors consent to the publication of the manuscript in Metron.

Availability of data and material All data generated during this study are available from the corresponding author on reasonable request.

Code availability The code is available from the corresponding author on reasonable request.

Authors' contributions AV conceived this research; both authors participated in the design and interpretation of the data, performed experiments and analysis, wrote the paper and participated in the revisions of it, read and approved the final manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alwin, D.F., Krosnick, J.A.: The measurement of values in surveys: a comparison of ratings and rankings. *Public Opin. Q* **49**(4), 535–552 (1985)
2. Critchlow, D.E., Fligner, M.A., Verducci, J.S.: Probability models on rankings. *J. Math. Psychol.* **35**(3), 294–318 (1991)
3. Deza, M.M., Deza, E.: Encyclopedia of distances. In: *Encyclopedia of Distances*, pp. 1–583. Springer (2009)
4. Diaconis, P.: Group representations in probability and statistics. *Lecture notes-monograph series* **11**, i–192 (1988)
5. Ferrari, P.A., Manzi, G.: Citizens evaluate public services: a critical overview of statistical methods for analysing user satisfaction. *J. Econ. Policy Reform* **17**(3), 236–252 (2014)
6. Gadrich, T., Bashkansky, E., Zitkikis, R.: Assessing variation: a unifying approach for all scales of measurement. *Quality & Quantity* **49**(3), 1145–1167 (2015)
7. Gwet, K.L.: Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters. *Advanced Analytics, LLC* (2014)
8. Hirschman, A.O.: *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, vol. 25. Harvard University Press, Harvard (1970)
9. Kendall, M.G.: A new measure of rank correlation. *Biometrika* **30**(1/2), 81–93 (1938)
10. Kendall, M.G.: Rank correlation methods. (1948)
11. Krosnick, J.A.: The threat of satisficing in surveys: the shortcuts respondents take in answering questions. *Surv. Methods Newslett.* **20**(1), 4–8 (2000)
12. Krosnick, J.A., Alwin, D.F.: A test of the form-resistant correlation hypothesis: ratings, rankings, and the measurement of values. *Public Opin. Quart.* **52**(4), 526–538 (1988)
13. Kumar, R., Vassilvitskii, S.: Generalized distances between rankings. In: *Proceedings of the 19th international conference on World wide web*, pp. 571–580 (2010)
14. Mallows, C.L.: Non-null ranking models. i. *Biometrika* **44**(1/2), 114–130 (1957)
15. Marmor, Y.N., Bashkansky, E.: Processing new types of quality data. *Qual. Reliab. Eng. Int.*
16. McCarty, J.A., Shrum, L.J.: The measurement of personal values in survey research: a test of alternative rating procedures. *Public Opin. Quart.* **64**(3), 271–298 (2000)
17. Munson, J.M., McIntyre, S.H.: Developing practical procedures for the measurement of personal values in cross-cultural marketing. *J. Mark. Res.* **16**(1), 48–52 (1979)
18. Ock, M., Yi, N., Ahn, J., Jo, M.W.: How many alternatives can be ranked? A comparison of the paired comparison and ranking methods. *Value Health* **19**(5), 655–660 (2016)
19. Rao, C.R.: Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol.* **21**(1), 24–43 (1982)
20. Reynolds, T.J., Jolly, J.P.: Measuring personal values: an evaluation of alternative methods. *J. Mark. Res.* **17**(4), 531–536 (1980)
21. Vanacore, A., Marmor, Y.N., Bashkansky, E.: Some metrological aspects of preferences expressed by prioritization of alternatives. *Measurement* **135**, 520–526 (2019)
22. Vanacore, A., Pellegrino, M.S.: Testing inter-group ranking heterogeneity: do patient characteristics matter for prioritization of quality improvements in healthcare service?, p. 100921. *Socio-Economic Planning Sciences* (2020)
23. Vanacore, A., Pellegrino, M.S., Marmor, Y.N., Bashkansky, E.: Analysis of consumer preferences expressed by prioritization chains. *Qual. Reliab. Eng. Int.* **35**(5), 1424–1435 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.