

The Influence of the Number of Voters on Ranking Data Sampled with the Mallows Model

Original

The Influence of the Number of Voters on Ranking Data Sampled with the Mallows Model / Villar, M., Rico, N., Baz, J., Díaz, I.. - In: INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE SYSTEMS. - ISSN 1875-6883. - (2026). [10.1007/s44196-026-01514-6]

Availability:

This version is available at: 11583/3015626 since: 2026-09-16T15:21:00Z

Publisher:

Springer Nature

Published

DOI:10.1007/s44196-026-01514-6

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

The Influence of the Number of Voters on Ranking Data Sampled with the Mallows Model

Received: 1 May 2026

Accepted: 27 July 2026

Published online: 26 August 2026

Cite this article as: Villar M., Rico N., Baz J. *et al.* The Influence of the Number of Voters on Ranking Data Sampled with the Mallows Model. *Int J Comput Intell Syst* (2026). <https://doi.org/10.1007/s44196-026-01514-6>

Mario Villar, Noelia Rico, Juan Baz & Irene Díaz

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

The Influence of the Number of Voters on Ranking Data Sampled with the Mallows Model

Mario Villar [0000-0003-1885-0452]^{1*}, Noelia Rico
[0000-0002-6169-4523]^{1†}, Juan Baz [0000-0003-1142-6077]^{2†},
Irene Díaz [0000-0002-3024-6605]^{1†}

^{1*}Department of Computer Science, University of Oviedo, San Francisco Street, 3, Oviedo, 33003, Asturias, Spain.

²Department of Mathematical Sciences “G.L. Lagrange”, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, 10129, Piedmont, Italy.

*Corresponding author(s). E-mail(s): villarmario@uniovi.es;
Contributing authors: noeliarico@uniovi.es; juan.bazgonzalez@polito.it;
sirene@uniovi.es;

[†]These authors contributed equally to this work.

Abstract

Ranking aggregation algorithms are typically evaluated using synthetic data generated by the Mallows model. Comparisons are often made across profiles with varying numbers of voters under the assumption that their structural properties remain consistent. However, previous empirical studies have shown that this assumption does not always hold, as the number of voters appears to affect profile characteristics such as the Condorcet properties. In this paper, we provide a theoretical explanation of these phenomena by introducing a probabilistic framework that characterises events arising from the Mallows sampling process, including the probability that a given alternative is the Condorcet winner. We demonstrate that the number of voters influences the distribution of sampled profiles and, consequently, their structural characteristics. Building on these theoretical results, we propose methodological guidelines for the fair evaluation of ranking aggregation algorithms, facilitating unbiased comparisons across benchmark instances generated with different parameter configurations.

Keywords: applied probability, ranking aggregation, Mallows model, Impartial Culture, decision making

1 Introduction

Aggregating the preferences of a group of voters is crucial across a wide range of fields, such as decision making in complex multi-criteria problems [1] or the aggregation of predictions in machine learning model ensembles [2], where clustering-based approaches and evaluation metrics such as the Fowlkes-Mallows index are frequently employed [3]. In most of these scenarios, different criteria can be used to rank multiple alternatives, for example, the class rankings produced by each model in an ensemble [4, 5]. In this context, *ranking aggregation methods* are essential for combining these diverse criteria into a *consensus ranking* [6] that facilitates fair decision-making.

One of the earliest ranking aggregation methods is the Condorcet criterion [7], which determines a winner by simple majorities in pairwise comparisons among all alternatives. However, this method exhibits an inherent issue that makes it impossible to derive a consensus ranking in some situations [8]. Numerous works have aimed to extend the Condorcet criterion to overcome this limitation [9]. Most of these techniques have solid theoretical foundations and satisfy multiple desirable properties, but they do not account for the practical problems that may arise from their complexity [10]. These problems have raised interest in the research field known as *Computational Social Choice* (COMSOC) [9], with applications spanning areas such as group decision-making [11] and machine learning [12].

One relevant area in COMSOC is the design of algorithms that efficiently implement ranking aggregation methods, facilitating their application to real-time problems such as aggregating recommendations in recommender systems [13]. The evaluation process of such implementations typically involves multiple profiles of rankings (sets of rankings) and the measurement of evaluation metrics, such as their execution time. Real-life datasets provide interesting analyses of real-world situations, but they usually have a limited number of alternatives and voters [14]. Therefore, synthetic data is usually sampled to provide voting scenarios with different numbers of voters and alternatives, as well as different levels of agreement among the voters. The probabilistic model proposed by Colin Mallows [15] is one of the most frequent sampling techniques [16].

To generate a profile with the Mallows model, rankings are randomly sampled from a probability distribution. It is usually assumed that its probabilistic nature justifies any choice of the model's parameters, which can be problematic. For instance, while the dependence of Condorcet-consistent outcomes on the number of voters has a long history in classical voting theory under uniform random distributions (Impartial Culture) [17, 18], it remains almost unexamined for parametric distributions such as the Mallows model. In particular, previous studies have empirically shown that increasing the number of voters reduces the appearance of the Condorcet paradox [19] and modifies structural properties of sampled data [20]. These observations suggest that the choice of the number of voters may bias the evaluation of ranking aggregation algorithms. However, the probabilistic mechanisms underlying these phenomena remain unexplained, and there are currently no theoretically grounded guidelines for selecting the number of voters when generating benchmark instances.

This paper addresses this gap by providing a theoretical analysis of the influence of the number of voters on the sampling process of profiles using the Mallows model.

Our contribution extends classical asymptotic results from Impartial Culture to the Mallows model, presents a finite-sample analysis of sampled data, and combines both perspectives in a methodological framework for designing experimental evaluation protocols. Specifically, we derive probability expressions related to structural properties of the sampled data that explain previously reported empirical observations, thereby identifying the probabilistic causes of these phenomena. Building on this analysis, we propose methodological guidelines for the fair evaluation of ranking aggregation algorithms using Mallows-generated profiles.

The remainder of this paper is organised as follows. Section 2 reviews current evaluation techniques, and Section 3 introduces the required theoretical concepts. Section 4 describes the Mallows model, and Section 5 illustrates the empirical phenomena motivating our study. Section 6 develops the probability framework explaining these observations. Section 7 presents a methodological procedure for the fair evaluation of ranking aggregation algorithms, and Section 8 concludes the paper.

2 Related Work

The Mallows model [15] is used to sample profiles of rankings from a probabilistic distribution and benchmark ranking aggregation algorithms. A recent literature review [16] places the Mallows model as the second most popular sampling technique, only after Impartial Culture (IC) [21]. However, the Mallows model can be seen as a generalisation of IC.

The Mallows model depends on three parameters: the number of alternatives, the number of voters, and the dispersion. However, the choice of parameters still heavily depends on the author: some choose to fix all parameters to create a single dataset of profiles [22, 23], while others experiment with a range of dispersion values [24, 25], fix the number of alternatives and vary the numbers of voters [26, 27] or vice versa [28, 29], or generate multiple profiles with some combinations of three parameters [30, 31].

The choice of parameters is often not systematically analysed, overlooking their impact on evaluation. In particular, a fair evaluation protocol requires that any of the performance differences across different parameter settings are solely attributable to the experimental factors under study, rather than unintended variations in the structural properties of the sampled profiles. To address this issue, the *map of elections* [32] visualises ranking profiles in a 2D space based on the position-wise distance. Using this approach, it has been shown that the Mallows model covers most real-world datasets [33], supporting its widespread use when parameters are properly selected. However, the model tends to concentrate on high-agreement profiles, limiting its flexibility and leading to uneven coverage between maximum and minimum agreement. A normalised variant has recently been proposed [34] to address this by adjusting dispersion to the number of alternatives.

On the other hand, prior work has empirically observed that varying the number of voters can affect the performance of some ranking aggregation algorithms [19]. In particular, profiles with more voters exhibit different behaviour regarding the existence of a Condorcet ranking or winner than those with fewer voters. Another empirical study analysed how the frequency of rankings in sampled data, including the modal

ranking, evolves as the number of voters increases [20]. It found that while the mean frequency remains constant, the standard deviation decreases as the number of voters grows, and the modal ranking more often matches the central ranking of the Mallows model. Overall, these results suggest that the intrinsic properties of sampled profiles change with the voter count.

The phenomenon whereby the probability of observing Condorcet-consistent outcomes varies with the number of voters is well-established under IC. Multiple works have thoroughly characterised the exact and asymptotic probabilities of the existence of Condorcet winners, together with the Condorcet efficiency of different voting rules [17, 18, 35]. While these classical contributions focus primarily on theoretical bounds under uniform cultures, evaluating algorithmic performance requires shifting the focus towards non-uniform parametric data models, such as the Mallows model.

However, existing studies on the influence of the number of voters in data sampled from the Mallows model are purely empirical and do not offer a theoretical account of the observed variations. This work addresses that gap by connecting classical asymptotic analysis with practical benchmarking, deriving probability expressions that characterise how the number of voters affects outcomes under the Mallows model. Accordingly, this paper makes both a theoretical contribution, by explaining the probabilistic mechanisms underlying previously observed empirical phenomena, and a methodological contribution, by translating these results into practical guidelines for the fair evaluation of ranking aggregation algorithms.

3 Preliminaries

Consider a set of m alternatives $\mathcal{A} = \{a_1, a_2, \dots, a_m\}$. A *strict ranking* v is a linear order over $\mathcal{A}$ such that, for any $a_i, a_j \in \mathcal{A} : i \neq j$, either $a_i \succ_v a_j$ (a_i is preferred over a_j) or $a_j \succ_v a_i$ [36]. Note that a strict ranking establishes a complete, transitive and antisymmetric relation on $\mathcal{A}$. We denote the set of all strict rankings of $\mathcal{A}$ by $\mathcal{L}(\mathcal{A})$, and the alternative ranked in the i -th position of a ranking v by $v(i)$. On the contrary, we denote the mapping that returns the position in v of any alternative $a \in \mathcal{A}$ by $v^{-1} : \mathcal{A} \rightarrow \{1, \dots, m\}$. To facilitate understanding, we refer to strict rankings simply as rankings in the remainder of this work.

A partial ranking is a linear order over a subset of alternatives. As a consequence, some alternatives of $\mathcal{A}$ may remain unranked or incomparable to others following the partial ranking. The set of all partial rankings of l alternatives is defined as the set of l -permutations over $\mathcal{A}$, i.e., $\mathcal{L}_l(\mathcal{A}) = \{(a_1, \dots, a_l) : a_i \in \mathcal{A}, a_i \neq a_j \text{ for } i \neq j\}$ with $1 \leq l \leq m$. The interpretation of each element $(a_1, \dots, a_l) \in \mathcal{L}_l(\mathcal{A})$ as a partial ranking over $\mathcal{A}$ is such that $a_1 \succ \dots \succ a_l$ (a strict order over the l alternatives).

A prefix $\rho \in \mathcal{L}_l(\mathcal{A})$ of length l is a partial ranking that gathers the l alternatives ranked on top of some ranking $v \in \mathcal{L}(\mathcal{A})$, that is, $\rho = v(1) \succ v(2) \succ \dots \succ v(l)$. In such a case, we say that v starts with the prefix ρ . On the other hand, a suffix $\eta \in \mathcal{L}_h(\mathcal{A})$ is a partial ranking that contains the h last ranked alternatives of a ranking $u \in \mathcal{L}(\mathcal{A})$, i.e., $\eta = u(m - h + 1) \succ \dots \succ u(m)$. We say that the ranking u ends with the suffix

η . The number of prefixes (and of suffixes) of length l is given by the expression

$$\frac{m!}{(m-l)!}.$$

A suffix η of length $m-l$ is compatible with a prefix ρ of length l if their concatenation, $\rho \cdot \eta$, gives a complete strict ranking of the m alternatives, i.e., if $v(1) \succ v(2) \succ \dots \succ v(l) \succ u(l+1) \succ u(l+2) \succ \dots \succ u(m)$ is a ranking over $\mathcal{A}$. We denote the set of all suffixes which are compatible with prefix ρ by $\mathcal{H}(\rho)$.

Consider n independent voters, each providing a ranking over $\mathcal{A}$. A *list of votes* $\mathcal{V} = (v_1, \dots, v_n)$ is the ordered set of these rankings. Under anonymity, where voter order is irrelevant, a *profile of rankings* π is obtained by grouping identical rankings and counting their frequencies.

The number of rankings is $m!$. With n voters, the number of possible lists of votes is $(m!)^n$, since each voter selects one of the $m!$ rankings. The number of profiles corresponds to the number of multisets of size n from these rankings, that is,

$$\binom{\binom{m!}{n}}{n} = \binom{m! + n - 1}{n}.$$

Thus, the number of profiles is at most the number of lists of votes.

When a collective decision is required, preferences in a profile must be combined into a single outcome. *Ranking aggregation methods* derive a consensus ranking, aggregating individual preferences into a collective order. Many such methods have been proposed [6], with the Condorcet criterion being particularly influential and forming the basis for numerous extensions [9].

3.1 The Condorcet criterion

The Condorcet criterion is a ranking aggregation technique based on simple majorities: an alternative is ranked above another if a majority of voters prefers it. The outranking matrix $\mathbf{O}$ compactly represents a profile [37], where each entry $o_{h,k}$ counts voters preferring a_h over a_k . It is an $m \times m$ matrix with $o_{h,k} + o_{k,h} = n$ for all pairs, and zeros on the diagonal.

The *Condorcet ranking* (CR) is obtained by placing an alternative a_h above another a_k whenever $o_{h,k} > \frac{n}{2}$. The CR does not exist when cycles between alternatives emerge (Condorcet paradox), that is, when $\exists a_h, a_k, a_t \in \mathcal{A}$ such that $o_{h,k} > \frac{n}{2}$, $o_{k,t} > \frac{n}{2}$, $o_{t,h} > \frac{n}{2}$. In addition, the CR does not exist in cases with even numbers of voters when ties appear in pairwise comparisons. The *Condorcet winner* (CW) is the alternative preferred to all others, i.e., $a_c \in \mathcal{A}$ such that $o_{c,k} > \frac{n}{2} \forall a_k \in \mathcal{A} \setminus \{a_c\}$. The CW does not exist if such a condition is not met.

The absence of a CW implies no CR, but not vice versa: a CW may exist even if no CR exists if an alternative is still preferred above all others. Thus, three cases arise: (i) a CR exists, (ii) no CR but a CW exists, and (iii) neither exists (NC).

The Condorcet ranking can be obtained in polynomial time, and has laid the foundation for multiple extensions that aim to overcome its limitations. However,

some of these, such as the Kemeny method [38], are NP-hard and cannot be solved in polynomial time when the CR does not exist.

4 The Mallows Model for Sampling Rankings

Evaluating ranking aggregation methods is crucial, especially for those with high computational complexity. For example, the Kemeny method [38] is theoretically validated, but its practical application in problems such as feature selection in machine learning models [39] depends on the execution time of its implementation. Rigorous evaluation, therefore, requires benchmarking implementations across diverse, controlled scenarios, varying the numbers of alternatives, voters, and levels of agreement.

The *Mallows model* [15] is commonly used to sample profiles used in the evaluation process. It is defined by a central ranking σ and a dispersion parameter $\phi \in [0, 1]$, where a higher ϕ implies greater divergence from σ . Together, they define an instance $\mathcal{M}_{\phi, \sigma}$. As mentioned in Section 2, Impartial Culture (IC) [21] is obtained when $\phi = 1$, assigning equal probability to all rankings.

The probabilistic distribution of the Mallows model is based on the Kendall-tau distance. The *Kendall-tau distance* [40], denoted by $\kappa(u, v)$, between two rankings $u, v \in \mathcal{L}(\mathcal{A})$ is the number of swaps of consecutive alternatives required to transform u into v . The probability of sampling a ranking $v \in \mathcal{L}(\mathcal{A})$ using $\mathcal{M}_{\phi, \sigma}$ is

$$P(v) = \frac{\phi^{\kappa(v, \sigma)}}{Z(\phi, m)},$$

where $Z(\phi, m)$ is a constant introduced to account for the $m!$ possible rankings so the sum of their sampling probabilities equals 1 [15]. It is defined as

$$Z(\phi, m) = \sum_{v \in \mathcal{L}(\mathcal{A})} \phi^{\kappa(v, \sigma)} = \prod_{i=1}^{m-1} \left(1 + \sum_{j=1}^i \phi^j \right).$$

Rankings can be efficiently sampled from the Mallows model using the *Repeated Insertion Model* (RIM) [41, 42]. It constructs rankings by iteratively inserting the alternatives from the central ranking. Starting with $\sigma(1)$, each alternative a is inserted into position j with probability $P_{\text{RIM}}(a, j)$. At step i , $\sigma(i)$ is inserted into one of the i available positions, continuing until $\sigma(m)$ is placed.

Thus, the i -th alternative is inserted into a partial ranking that already contains the $i - 1$ preceding alternatives in σ , with its position chosen according to predefined probabilities. These probabilities are independent of prior insertions, implying independence between insertions. The following theorem establishes the link between the Mallows model and the RIM.

Theorem 1 [41]. An instance $\mathcal{M}_{\phi,\sigma}$ of the Mallows model produces the same identical distribution as a RIM with insertion probabilities following

$$P_{\text{RIM}}(a, j) = \frac{\phi^{\sigma^{-1}(a)-j}}{\sum_{k=1}^{\sigma^{-1}(a)} \phi^{\sigma^{-1}(a)-k}} \quad (1)$$

with $a \in \mathcal{A}$, $1 \leq j \leq \sigma^{-1}(a)$ and $\sigma^{-1}(a)$ being the position of a in the central ranking.

4.1 Probability of Sampling a Ranking

In this paper, we analyse the probability of sampling rankings with specific properties under the Mallows model. Before presenting our results, we briefly review the probability distribution over rankings induced by the RIM, included here only to provide essential background.

Sampling a ranking v occurs with probability $P(v)$, obtained by iteratively inserting the alternatives of σ into an initially empty prefix. The insertion position of each alternative $v(i)$ is determined by the number of alternatives ranked above $v(i)$ in both σ and v , plus one [41]. Thus, the probability of sampling v is:

$$P(v) = \prod_{i=1}^m P_{\text{RIM}}(v(i), \text{Above}(v, i) + 1), \quad (2)$$

where $\text{Above}(v, i)$ is the number of alternatives ranked above $v(i)$ both in v and in σ , that is

$$\text{Above}(v, i) = \left| \{a \in \mathcal{A} : v^{-1}(a) < i \text{ and } \sigma^{-1}(a) < \sigma^{-1}(v(i))\} \right|.$$

A ranking is sampled from $\mathcal{M}_{\phi,\sigma}$ using the RIM with the insertion probabilities of Theorem 1, as illustrated in Example 1. When $\phi = 1$, the model reduces to IC, and all rankings occur with equal probability. The central ranking has the highest sampling probability. For any alternative $a \in \mathcal{A}$, the denominator of Equation 1 is independent of its insertion position, while the numerator is maximised when a is inserted at its position in the central ranking, i.e., when $j = \sigma^{-1}(a)$. Thus, inserting each alternative in its position in σ yields the central ranking with the highest probability. Conversely, the inverse ranking of σ results from inserting every alternative in the first position, minimising the numerator of Equation 1 and giving it the lowest probability.

Example 1 (Sampling probability of a ranking) Consider the Mallows model $\mathcal{M}_{\phi,\sigma}$, given the set $\mathcal{A} = \{A, B, C\}$ of three alternatives, the central ranking $\sigma = A \succ B \succ C$ and a dispersion $\phi = 0.95$. By the Theorem 1, the RIM equivalent to $\mathcal{M}_{\phi,\sigma}$ has the insertion probabilities shown in Table 1.

Let u be the ranking $B \succ C \succ A$. This ranking is constructed by inserting A in first position (the only available at the initial step), B in first position and, then, C in second position. By Equation 2, the probability of sampling this ranking is

$$P(u) \approx 1 \cdot 0.487 \cdot 0.333 \approx 0.162$$

The central ranking is sampled by inserting every alternative in the last available position, i.e. A , in the first position, B in the second and C in the third position. Thus, its sampling probability is

$$P(\sigma) \approx 1 \cdot 0.513 \cdot 0.351 \approx 0.180$$

Alternative	Probability of position 1	Probability of position 2	Probability of position 3
A	1		
B	0.487	0.513	
C	0.316	0.333	0.351

Table 1: Insertion probabilities, approximated to three decimal places, of each alternative in each position using the RIM equivalent to the Mallows model considered in Example 1.

A list of votes for n voters can be drawn by independently repeating the sampling process n times. When voters are assumed to be anonymous, and all votes carry equal weight, the result is a **profile of rankings**.

4.2 The Normalised Mallows Model

As described in Section 2, the normalised Mallows model [34] is a variation of the original model that tackles an issue in the model by adapting the dispersion with the number of alternatives employed. Under this approach, a new parameter $\text{norm-}\phi$ (*normalised dispersion*) is used, and its value is internally transformed to the standard dispersion value ϕ . The relationship between both parameters is given by

$$\text{norm-}\phi = \frac{4 \cdot \mathbb{E}_{\phi, m}[\kappa(v, \sigma)]}{m \cdot (m - 1)},$$

where $\mathbb{E}_{\phi, m}[\kappa(v, \sigma)]$ is the expected Kendall-tau distance between the central ranking and a ranking of m alternatives sampled using a Mallows model with a dispersion parameter ϕ . The resulting value of ϕ is then used in the RIM, as described in the previous section. We adopt this normalisation throughout the remainder of the paper: unless otherwise stated, all reported dispersion values refer to $\text{norm-}\phi$ and are internally converted to the model's dispersion for sampling rankings.

5 Identifying the Problem

This section presents an empirical example of the influence of the voter count on the Mallows model when evaluating ranking aggregation algorithms. We use the Condorcet criterion, since some of its extensions, such as the Kemeny method [38], benefit from the presence of a CR [43]. Consequently, comparing execution times across sets of profiles with different frequencies of existence of CRs would be unfair.

We have used a publicly available dataset of profiles sampled with the Mallows model [44]. This dataset was sampled using the RIM and using the implementation of the normalisation procedure published in the *PrefLib* library [45]. The dataset covers a wide spectrum of profiles, and we have considered a subset of it. Table 2 lists the parameter values considered. The chosen values for the number of alternatives are intended to cover profiles that can be used to evaluate the current state-of-the-art implementations for the Kemeny method [43, 46], which cannot handle instances with

15 or more alternatives. The values for the number of voters and dispersion were chosen to cover a wide range of scenarios, and all normalised dispersion values were mapped numerically to the corresponding Mallows dispersion following Subsection 4.2. We excluded the case $\text{norm-}\phi = 0$, as the central ranking is always returned in this case, which is trivially a CR. We used the lexicographic order as the central ranking, though the analysis generalises to any choice of σ . For each parameter combination, we used the 1000 profiles included in the dataset.

Number of alternatives	$\{3, 5, 10, 20, 30\}$
Number of voters	$\{10\} \cup \{x \in \mathbb{N} \mid 25 \leq x \leq 1000 \text{ and } x \equiv 0 \pmod{25}\}$
$\text{norm-}\phi$	$\{0.4, 0.6, 0.7, 0.8, 0.9, 0.95, 1\}$

Table 2: Values considered for the parameters of the Mallows model to sample the dataset. Note that the reported values of $\text{norm-}\phi$ were internally converted to the final dispersion of the Mallows used to sample the rankings.

Figure 1 shows the behaviour of the Condorcet criterion across profiles with three alternatives and different numbers of voters and levels of dispersion¹. Each facet corresponds to a fixed number of voters, and each bar represents the percentage of sampled profiles with a given Condorcet property. Profiles with a CR are shown in light yellow, profiles without a CR but with a CW in light green, and profiles which have neither (NC) in dark green. For any given dispersion value, moving from left to right facets illustrates how these Condorcet properties evolve as the number of voters increases, while the number of alternatives and the dispersion remain fixed. The standard errors of the proportions of CR, CW and NC for every parameter combination are reported in the Supplementary Material due to their extent.

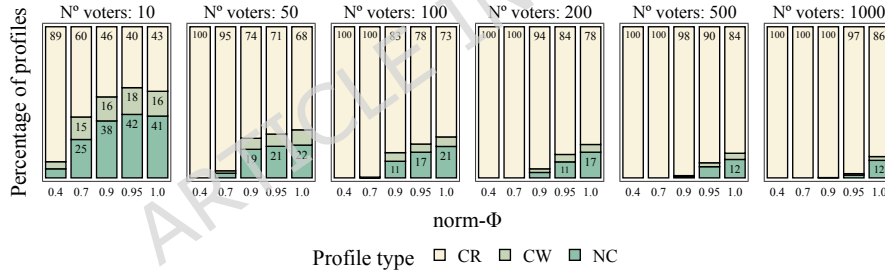


Fig. 1: Condorcet properties of profiles with three alternatives, disaggregated by dispersion level and number of voters. Each bar reports the percentage of profiles within a given parameter configuration exhibiting a specific Condorcet property, restricted to profiles where a Condorcet ranking (CR) exists, those that exhibit a Condorcet winner (CW), and those with neither (NC).

¹Figure A1 of Appendix A reproduces this figure for elections with five, ten, twenty and thirty alternatives, and for a wider range of dispersion values.

This figure shows that the percentage of profiles with a CR increases as the number of voters grows, for fixed dispersion and number of alternatives. Although the proportion of profiles with only a CW appears to decrease, profiles with a CR also contain a CW, so the total number of profiles with a CW also increases. As a result, the performance of ranking aggregation algorithms that benefit from the existence of a CR is artificially boosted in such cases. A similar pattern occurs for other numbers of alternatives, as shown in Figure A1.

This effect can introduce a bias when comparing implementations across elections with different voter counts, similar to the issue discussed in Subsection 4.2 when varying the number of alternatives [34]. For example, the state-of-the-art implementation for the Kemeny method [46] has factorial complexity with respect to the number of alternatives when no CR exists, but polynomial when it does. Hence, the results shown in Figure 1 indicate that the execution times of this algorithm are not comparable across profiles with fixed alternatives and dispersion but different voter counts, as execution times are likely shorter when this parameter increases due to the higher prevalence of CRs.

To examine this effect in more detail, we introduce Figures 2 and 3. Figure 2 visualises the mean and standard deviation (STD) of the positions of the alternatives in the CRs, computed only for profiles where a CR exists. The shaded region indicates one standard deviation above and below the mean. The example is restricted to profiles with three alternatives and dispersion 0.95, though the same pattern appears for all other parameter combinations². The mean positions of the alternatives ranked first and last in the central ranking vary significantly, and the CR becomes closer to the central ranking as the number of voters increases, even with fixed dispersion. Increasing the number of voters appears to shift the CR toward the central ranking, with dispersion mainly affecting the rate of convergence. The STD of the position of each alternative also decreases with the number of voters. Thus, the average CR positions move toward those in the central ranking, and the variability around these positions shrinks as the number of voters grows. Figure 3 visualises the number of profiles with a CR for each number of voters considered in Figure 2. It shows that the sample of CR profiles is sufficiently large to justify the use of conditional averages in the previous figure³.

Additionally, Figure 4 visualises the evolution of the percentage of profiles in which each alternative was the CW as the number of voters is increased, for profiles with three alternatives and normalised dispersions of 0.95 and 1 (IC). A clear distinction between the two dispersion values can be observed. For 0.95 dispersion, the alternative ranked first in the central ranking is the CW more frequently than the others, and this trend is amplified as the number of voters increases (going from $\approx 30\%$ of profiles with 10 voters to almost 80% with 1000 voters). On the other hand, all alternatives emerge as the CW on a similar ratio of profiles in IC, and this behaviour is maintained even when the number of voters increases.

Figure 4 also shows an alternating pattern for odd and even numbers, as fewer CWs tend to appear for even numbers of voters. This phenomenon is related to ties

²Additional examples with different dispersion values are included in Figure A2 of Appendix A.

³Additional examples with different dispersion values are included in Figure A3 of Appendix A.

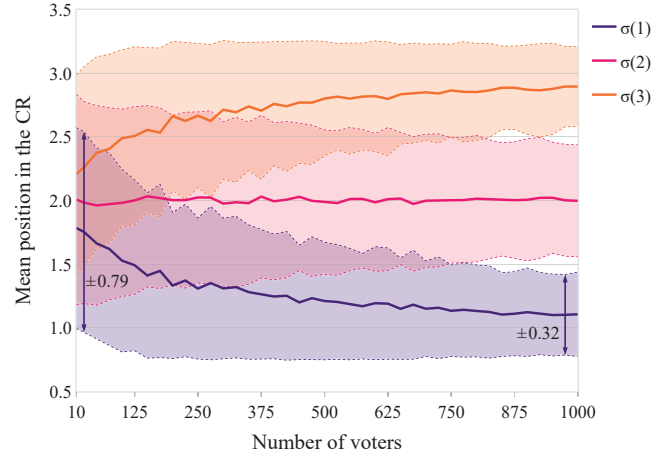


Fig. 2: Evolution of the mean position of alternatives within Condorcet rankings (CRs) as the number of voters increases. The analysis is restricted to profiles with three alternatives and normalised dispersion of 0.95, and only to profiles where a CR exists. For each alternative, the curve shows the average position in the CR, while the shaded band indicates one standard deviation around the mean.

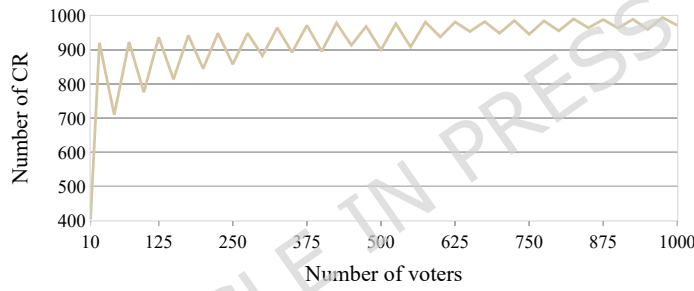


Fig. 3: Evolution of the number of profiles where the CR exists as the number of voters increases, for profiles with three alternatives and normalised dispersion of 0.95. Out of the 1000 profiles sampled for each value of the number of voters, only the ones where a CR exists are displayed in this chart.

in pairwise comparisons that can emerge in scenarios where exactly half of the voters prefer one alternative over another, while the other half have opposite preferences.

The preceding examples illustrate some of the issues that can arise when evaluating algorithms that implement Condorcet extensions (ranking aggregation methods that ensure the CW is ranked first whenever it exists) on data sampled using the Mallows model with multiple numbers of voters. In the next section, we examine the probability distribution induced by the Mallows model to clarify the underlying causes of the structural variations observed in these profiles. The remainder of the paper aims to

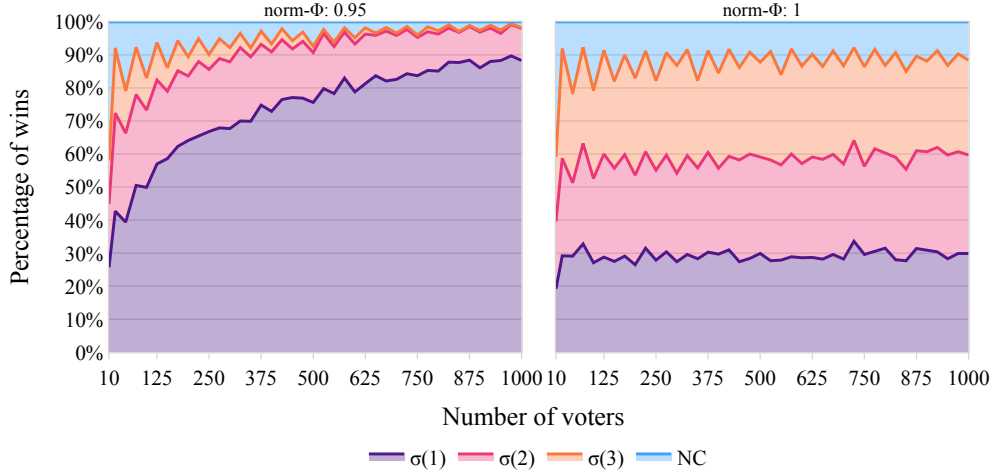


Fig. 4: Evolution of the Condorcet winners (CWs) as the number of voters increases. This chart gathers all profiles with three alternatives and normalised dispersions of 0.95 and 1 (IC), as profiles without a CW are shown in light blue (NC). The percentage of profiles in which each alternative of the central ranking was the CW is shown in the y-axis.

assess the influence of the number of voters on the comparability across profiles with fixed dispersion and number of alternatives. This issue is particularly relevant for the fair evaluation of algorithmic implementations of ranking aggregation methods, whose reliability is important in decision-support applications, including practical ranking problems that may benefit from efficient implementations [47].

6 A Probabilistic Approach

In this section, we derive the probabilities of sampling a profile, a prefix, and an alternative in a given position under the Mallows model. We first compute the probability of a profile given its rankings. We then analyse the probability of sampling rankings with a given prefix and those placing an alternative in a specific position. Finally, we derive the probability of profiles where a fraction of rankings share a prefix or rank an alternative in a given position. Note that the development of this section is not limited to the sampling probabilities of the Mallows model, as it applies to any RIM model.

6.1 Sampling a Profile

A profile of rankings is a compact representation of a list of votes that preserves voter anonymity and summarises the corresponding individual rankings. Consequently, a profile is sampled whenever any list of votes that aggregates into it is sampled.

A list of votes is sampled by independently generating n rankings using the RIM. Therefore, given the list of votes $\mathcal{V} = (v_1, v_2, \dots, v_n)$, the events *sampling the ranking* v_k are independent for all $k = 1, \dots, n$. The event *sampling the list of votes* $\mathcal{V}$ occurs

when its rankings are generated in the exact order of the list. Its probability, by independence, is

$$P(\mathcal{V}) = P((v_1, \dots, v_n)) = \prod_{k=1}^n P(v_k). \quad (3)$$

Example 2 (Sampling probability of a list of votes) Consider the Mallows model with $\phi = 0.95$, the RIM and the rankings σ, u from Example 1. Let the number of voters be $n = 3$, and let $\mathcal{V} = (\sigma, \sigma, u)$ be a list of votes. The probability of sampling $\mathcal{V}$, given by Equation 3, is

$$P(\mathcal{V}) = P(\sigma) \cdot P(\sigma) \cdot P(u) \approx 0.005$$

We say that two or more lists of votes are equivalent if they are summarised by the same profile, i.e., they contain the same rankings, and each ranking occurs the same number of times, regardless of the order in which they were sampled. Equivalent lists of votes have the same probability, as their rankings are simply permuted. We denote the set of all lists of votes equivalent to $\mathcal{V}$ by $\mathcal{E}(\mathcal{V})$.

The event *sampling the profile* π is the union of the events *sampling a list of votes* over all lists in $\mathcal{E}(\mathcal{V})$. Its probability is the sum of their probabilities and can be computed as the probability of sampling $\mathcal{V}$ multiplied by the number of equivalent lists. Suppose that the profile π consists of τ distinct rankings $v_1, v_2, \dots, v_\tau$, repeated $n_1, n_2, \dots, n_\tau$ times, respectively, with $\sum_{k=1}^\tau n_k = n$. Any list of votes summarised by π is a permutation of the list of votes

$$\mathcal{V} = \left(\underbrace{v_1, \dots, v_1}_{n_1}, \underbrace{v_2, \dots, v_2}_{n_2}, \dots, \underbrace{v_\tau, \dots, v_\tau}_{n_\tau} \right).$$

The number of lists of votes equivalent to $\mathcal{V}$ is the number of permutations of τ elements, each one repeated $n_1, n_2, \dots, n_\tau$ times, which is given by

$$|\mathcal{E}(\mathcal{V})| = \binom{n}{n_1, n_2, \dots, n_\tau} = \frac{n!}{\prod_{k=1}^\tau n_k!}. \quad (4)$$

Therefore, the probability of *sampling the profile* π is

$$P(\pi) = P(\cup_{V \in \mathcal{E}(\mathcal{V})} V) = \sum_{V \in \mathcal{E}(\mathcal{V})} P(V) = |\mathcal{E}(\mathcal{V})| \cdot P(\mathcal{V}) = \frac{n!}{\prod_{r=1}^\tau n_r!} \prod_{r=1}^\tau P(v_r)^{n_r}, \quad (5)$$

which is the same as the probability associated to a multinomial distribution with parameters $n_1, \dots, n_r$ and event probabilities $P(v_1), \dots, P(v_r)$.

Example 3 (Sampling probability of a profile) Consider the Mallows model with $\phi = 0.95$, the RIM and the list of votes $\mathcal{V}$ for three voters from Example 2. Let π be the profile that

summarises this list, that is

Ranking	Votes
$\pi : A \succ B \succ C$	2
$B \succ C \succ A$	1

The number of lists equivalent to $\mathcal{V}$, given by Equation 4, is

$$\frac{n!}{\prod_{k=1}^r n_k!} = \frac{3!}{2! \cdot 1!} = 3$$

Indeed, these three lists are (u, σ, σ) , (σ, u, σ) and $\mathcal{V}$ itself. The sampling probability of π is the sum of the sampling probabilities of the three lists:

$$P(\pi) = P(u, \sigma, \sigma) + P(\sigma, u, \sigma) + P(\mathcal{V}) = 3 \cdot P(\mathcal{V}) \approx 0.016$$

The number of equivalent lists of votes, given by Equation 4, lies in the range $[1, \dots, n!]$. Therefore, the sampling probability of any profile is at least that of the list of votes it summarises and at most $n!$ times this value.

Several conclusions follow from these probabilistic expressions. As noted in Subsection 4.1, the central ranking has the highest sampling probability among all rankings. Hence, the list of votes in which all voters select the central ranking, denoted by $\mathcal{V}\sigma$, has the highest sampling probability. Since the probability of a list of votes is the product of the probabilities of its rankings, the probability of $\mathcal{V}\sigma$ is the n -th power of the probability of σ , making it maximal. More generally, the sampling probability of a list of votes increases with the number of occurrences of σ .

Secondly, the probability of sampling a profile that summarises a list of votes $\mathcal{V}$ depends on two factors: the sampling probability of $\mathcal{V}$ and the number of lists of votes equivalent to it. The former is maximised for $\mathcal{V}\sigma$, but the latter is minimised in this case, as there is only one list of votes in which σ appears n times. Conversely, a list of votes with no repeated rankings has a lower sampling probability than $\mathcal{V}\sigma$, but the number of equivalent lists is maximised at $n!$. Hence, there is a trade-off between sampling fewer distinct rankings with higher probabilities and sampling more diverse profiles with lower-probability rankings. Moreover, the influence of the second factor increases with the number of voters, as the difference between low-diversity and high-diversity profiles becomes more pronounced (see Example 4).

Example 4 (Trade-off in profile probabilities) Consider the Mallows model with $\phi = 0.95$, the RIM, and the rankings σ and u from Example 1. Let π_{10} and π'_{10} be profiles with 10 voters, and π_{20} and π'_{20} profiles with 20 voters, as follows

Ranking	Votes	Ranking	Votes
$\pi_n : A \succ B \succ C$	n	$\pi'_n : A \succ B \succ C$	$0.9n$
		$B \succ C \succ A$	$0.1n$

Sampling a list of votes corresponding to π_{10} is 5% more likely than one corresponding to π'_{10} . This difference increases for 20 voters, where the probability for π_{20} is 11% higher than for π'_{20} . However, when considering profile probabilities, which account for the number of equivalent lists, the situation reverses. Sampling π_{10} is approximately 10 times less likely

than π'_{10} , and π_{20} is over 172 times less likely than π'_{20} . Numerically,

Profile	$P(\mathcal{V})$	$P(\pi)$
π_{10}	$3.53 \cdot 10^{-8}$	$3.53 \cdot 10^{-8}$
π'_{10}	$3.35 \cdot 10^{-8}$	$3.35 \cdot 10^{-7}$
π_{20}	$1.24 \cdot 10^{-15}$	$1.24 \cdot 10^{-15}$
π'_{20}	$1.12 \cdot 10^{-15}$	$2.13 \cdot 10^{-13}$

Although the relative frequencies of rankings are identical across $n = 10$ and $n = 20$, the difference in the sampling probabilities of the profiles accentuates when increasing the number of voters.

6.2 Sampling a Prefix

A prefix is sampled whenever a ranking that begins with it is generated. The event *sampling a prefix* is the union of the events *sampling a ranking* over all rankings that begin with that prefix, obtained by concatenating a compatible suffix.

Let $\rho \in \mathcal{L}_l(\mathcal{A})$ be a prefix of length l . The probability of sampling a ranking that begins with ρ and ends with a suffix $\eta \in \mathcal{H}(\rho)$ is, following Equation 2, the product of the cumulative weight of the alternatives in ρ and that of the alternatives in η .

$$P(\rho \cdot \eta) = \left(\prod_{i=1}^l P_{\text{RIM}}(\rho(i), \text{Above}(\rho, i) + 1) \right) \cdot \left(\prod_{i=1}^{m-l} P_{\text{RIM}}(\eta(i), \text{Above}(\rho \cdot \eta, l + i) + 1) \right) \quad (6)$$

By concatenating ρ to each suffix in $\mathcal{H}(\rho)$, we obtain all the rankings that start with ρ . Thus, the probability of *sampling the prefix* ρ is the sum of the prior probability when considering all the compatible suffixes:

$$\begin{aligned} P(\rho) &= P(\cup_{\eta \in \mathcal{H}(\rho)} (\rho \cdot \eta)) = \sum_{\eta \in \mathcal{H}(\rho)} P(\rho \cdot \eta) \\ &= \sum_{\eta \in \mathcal{H}(\rho)} \left(\prod_{i=1}^l P_{\text{RIM}}(\rho(i), \text{Above}(\rho, i) + 1) \right) \\ &\quad \cdot \left(\prod_{i=1}^{m-l} P_{\text{RIM}}(\eta(i), \text{Above}(\rho \cdot \eta, l + i) + 1) \right). \end{aligned} \quad (7)$$

Note that, by Equation 7, the prefix of length l with maximum probability corresponds to the prefix that gathers the l alternatives ranked on top of the central ranking in that exact order, i.e., the prefix $\sigma(1) \succ \dots \succ \sigma(l)$.

Example 5 (Sampling probability of a prefix) Consider the Mallows model with $\phi = 0.95$, the RIM and the ranking σ from Example 1. Consider the prefix $\rho = B$ of length 1. This prefix

is sampled whenever a ranking starting with it is sampled; that is, whenever $B \succ A \succ C$ or $B \succ C \succ A$ are sampled. Thus, its compatible suffixes are $\eta_1 = A \succ C$ and $\eta_2 = C \succ A$.

The probability of sampling the prefix ρ , given by Equation 7, is

$$P(\rho) \approx 0.487 \cdot (0.351 + 0.333) \approx 0.332$$

The direct evaluation of Equation (7) involves summing over the set $\mathcal{H}(\rho)$ of all $(m-l)!$ suffixes compatible with the prefix ρ . A naive computation would therefore require evaluating an exponential number of full rankings, each involving a product of $(m-l)$ sequential RIM insertion probabilities (the first sequence product only needs to be computed once, and then reused). This leads to a worst-case complexity of $\mathcal{O}((m-l)! \cdot (m-l))$.

6.3 Sampling an Alternative in a Position

An alternative $a \in \mathcal{A}$ is sampled in position j whenever a ranking is generated in which the j -th alternative is a or, equivalently, when a prefix of length j ending with a is sampled. Formally, this occurs when a ranking $v \in \mathcal{L}(\mathcal{A})$ with $v(j) = a$, or a prefix $\rho \in \mathcal{L}_j(\mathcal{A})$ with $\rho(j) = a$, is sampled.

Let $\mathcal{L}_j^a(\mathcal{A})$ denote the set of prefixes of length j that end with a . Then, the probability of *sampling a in position j* is the sum of the sampling probabilities of all prefixes in $\mathcal{L}_j^a(\mathcal{A})$.

$$P(v(j) = a) = P(\cup_{\rho \in \mathcal{L}_j^a(\mathcal{A})} \rho) = \sum_{\rho \in \mathcal{L}_j^a(\mathcal{A})} P(\rho). \quad (8)$$

Example 6 (Sampling probability of an alternative in a position) Consider the Mallows model with $\phi = 0.95$, the RIM and the ranking σ from Example 1. The alternative A is sampled in second position whenever a prefix of length 2 that ends in A is sampled, that is whenever $\rho_1 = B \succ A$ or $\rho_2 = C \succ A$ are sampled.

The probability of sampling A in second position, given by Equation 8, is the sum of the sampling probabilities of these prefixes:

$$P(v(2) = A) = P(\rho_1) + P(\rho_2) \approx 0.333$$

The direct evaluation of Equation (8) requires summing over all prefixes of length j that end in a , each of which has to be assigned a probability computed through Equation (7), which has efficiency $\mathcal{O}((m-j)! \cdot (m-j))$. The number of prefixes of length j that end in a is $\frac{(m-1)!}{(m-j)!}$, which leads to an increased overall complexity of $\mathcal{O}((m-1)! \cdot (m-j))$ for Equation (8).

6.4 Sampling a Profile with a Specific Frequency of a Prefix

In the previous sections, we detailed the probabilities of events related to sampling individual rankings, as well as lists of votes and profiles given their rankings. We now examine the sampling probabilities of profiles with specific characteristics without fixing their rankings.

Let $\rho \in \mathcal{L}_l(\mathcal{A})$ be a prefix of length l . A profile in which ρ appears ω times contains exactly ω rankings that begin with ρ , where $0 \leq \omega \leq n$. We denote the set of such profiles by π_ω and refer to ω as the frequency of ρ . The sampling probability of π_ω equals the probability that a binomial random variable with parameters n and $p = P(\rho)$ takes the value ω , expressed as [48].

$$\begin{aligned} P(\pi_\omega) &= P(\cup_{V \in \mathcal{E}(\mathcal{V}_\omega)} V) = |\mathcal{E}(\mathcal{V}_\omega)| \cdot P(\mathcal{V}_\omega) \\ &= \binom{n}{\omega} \cdot (P(\rho))^\omega \cdot (1 - P(\rho))^{n-\omega}, \end{aligned} \quad (9)$$

where $\mathcal{E}(\mathcal{V}_\omega)$ is the set of equivalent lists of votes to $\mathcal{V}_\omega$.

Example 7 (Probability of a profile given the frequency of a prefix) Consider the Mallows model with $\phi = 0.95$, the RIM and the rankings σ, u from Example 1. Consider the prefix $\rho = B$ (its probability was calculated in Example 5) and 10 voters.

The probability of sampling the prefix ρ in five of the ten possible rankings of a profile can be computed through Equation 9. Let π_ω be the class of profiles in which the number of rankings that start with ρ is $\omega = 5$. The probability of sampling π_ω is

$$P(\pi_\omega) \approx \frac{10 \cdot \dots \cdot (5+1)}{(10-5)!} \cdot 0.332^5 \cdot (1 - 0.332)^{10-5} \approx 0.137$$

The computational efficiency of Equation 9 is $\mathcal{O}(n \cdot (m-l)! \cdot (m-l))$, and it is mainly dominated by the computation of the probability of the prefix.

6.5 Sampling a Profile where an Alternative is Ranked in a Specific Position in a Number of Rankings

In a profile where an alternative $a \in \mathcal{A}$ appears λ times in position j , exactly λ voters select a ranking in which a occupies position j , with $1 \leq j \leq m$ and $0 \leq \lambda \leq n$. The probability of sampling any profile of the set π_λ of profiles that fulfil such a condition is obtained analogously to the previous case, as it follows a binomial distribution:

$$\begin{aligned} P(\pi_\lambda) &= P(\cup_{V \in \mathcal{E}(\mathcal{V}_\lambda)} V) = |\mathcal{E}(\mathcal{V}_\lambda)| \cdot P(\mathcal{V}_\lambda) = \\ &= \binom{n}{\lambda} \cdot (P(v(j) = a))^\lambda \cdot (1 - P(v(j) = a))^{n-\lambda}, \end{aligned} \quad (10)$$

where $\mathcal{V}_\lambda$ is a list of votes where a appears λ times in position j .

Note that the number of voters in the profiles is a significant factor in Equation 10. To compare probability distributions over profiles with different numbers of voters, we define the probability in terms of the proportion of rankings in which a appears in position j , rather than the absolute number λ . Let $\alpha \in \{t \in [0, 1] : tn \in \mathbb{N}\}$ denote this proportion, and let π_α be the set of profiles satisfying this condition. Then, the probability of sampling π_α can be written as

$$P(\pi_\alpha) = \binom{n}{\alpha n} \cdot (P(v(j) = a))^{\alpha n} \cdot (1 - P(v(j) = a))^{(1-\alpha)n}. \quad (11)$$

The computational efficiency of both Equation 10 and Equation 11 is $\mathcal{O}(n \cdot (m-1)! \cdot (m-j))$, mainly dominated by the computation of the probability of the alternative a being placed in the position j .

6.5.1 Asymptotic Behaviour of the Probability of a Profile where an Alternative is Ranked in a Specific Position in a Proportion of Rankings

We now turn to the effect of the number of voters on the probability expressions introduced above. In particular, to characterise the behaviour of Equation 11 as the number of voters increases, consider a fixed alternative $a \in \mathcal{A}$ and a fixed position $j \in 1, \dots, m$. Define, for each voter $r \in 1, \dots, n$, the indicator random variable

$$X_r^{a,j} = \mathbf{1}\{v_r(j) = a\},$$

where v_r denotes the ranking sampled for voter r . Since the rankings of the voters are independently and identically distributed according to the underlying ranking distribution, the variables $X_1^{a,j}, \dots, X_n^{a,j}$ are independent Bernoulli random variables with parameter $p_{a,j} = \mathbb{P}(v(j) = a)$ [48]. The empirical proportion of voters placing alternative a in position j is, therefore,

$$\alpha_{a,j} = \frac{1}{n} \sum_{r=1}^n X_r^{a,j}.$$

Observe that the event defining the set of profiles π_α corresponds precisely to the event $\alpha_{a,j} = \alpha$. Consequently, Equation 11 is the probability mass function of the empirical proportion $\alpha_{a,j}$. By the Strong Law of Large Numbers [48],

$$\alpha_{a,j} \xrightarrow[n \rightarrow \infty]{a.s.} p_{a,j},$$

which implies that the proportion of voters ranking a in position j converges almost surely, i.e. with probability one, to the underlying probability of observing a in that position. Therefore, for sufficiently large voter counts, the observed ratio becomes increasingly concentrated around $p_{a,j}$.

Since $\alpha_{a,j}$ is the average of n independent Bernoulli variables, its expectation and variance are given by $\mathbb{E}[\alpha_{a,j}] = p_{a,j}$ and $\text{Var}(\alpha_{a,j}) = p_{a,j}(1 - p_{a,j})/n$, thus the dispersion of the empirical proportion decreases at the classical rate $n^{-1/2}$. Consequently, larger numbers of voters yield increasingly stable estimates of the positional probability $p_{a,j}$. Furthermore, the Central Limit Theorem [48] yields the asymptotic normality of the empirical proportion:

$$\sqrt{n} \frac{\alpha_{a,j} - p_{a,j}}{\sqrt{p_{a,j}(1 - p_{a,j})}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

Equivalently,

$$\alpha_{a,j} = p_{a,j} + \mathcal{O}_p\left(n^{-1/2}\right),$$

which establishes that the (random) fluctuations around the limiting probability decrease at rate $n^{-1/2}$.

6.6 Probability of a Condorcet Winner

Section 5 showed that varying the number of voters can affect the behaviour of certain ranking aggregation algorithms, particularly those based on the Condorcet criterion. For Condorcet extensions, a central question is therefore how likely it is that a given alternative $a \in \mathcal{A}$ emerges as the Condorcet winner in profiles sampled from the Mallows model. We denote this event by $\text{CW}(a)$.

For a to be the Condorcet winner, the number of voters preferring a over any other alternative $b \in \mathcal{A}$ must strictly exceed half of the total number of voters. Formally, if $\mathbf{1}\{a \succ_r b\}$ is an indicator function that returns 1 if voter r ranks a strictly above b , and 0 otherwise, the probability of a being the Condorcet winner is defined as:

$$P(\text{CW}(a)) = P\left(\sum_{r=1}^n \mathbf{1}\{a \succ_r b\} > \frac{n}{2} \forall b \neq a\right). \quad (12)$$

To compute this exact probability under the Mallows model, we rely on the independence of the n voters. Let $\mathcal{V} = (v_1, \dots, v_n)$ be a random list of votes, where each ranking v_r is sampled independently from the distribution induced by the RIM. Because the condition for $\text{CW}(a)$ requires that $\sum_{r=1}^n \mathbf{1}\{a \succ_{v_r} b\} > \frac{n}{2}$ holds true for all $b \neq a$, we can express the exact probability by summing the joint probabilities of all possible lists of votes $\mathcal{V} \in \mathcal{L}(\mathcal{A})^n$ that satisfy this strict majority criterion:

$$P(\text{CW}(a)) = \sum_{(v_1, \dots, v_n) \in \mathcal{L}(\mathcal{A})^n} \mathbf{1}\left\{\sum_{r=1}^n \mathbf{1}\{a \succ_{v_r} b\} > \frac{n}{2} \forall b \neq a\right\} \prod_{r=1}^n P(v_r). \quad (13)$$

We can exploit the anonymity of voters to significantly reduce the state space by transitioning to profiles of rankings. For a profile π with τ different rankings, let n_r denote the number of voters who cast the specific ranking $v_r \in \pi$. Recalling from Subsection 6.1, the sampling probability of a class of profile that fulfil a specific condition follows a multinomial distribution with parameters n_r and event probabilities $P(v_r)$.

Rewriting the pairwise victory condition in terms of ranking frequencies, alternative a is the Condorcet winner if, for every $b \neq a$, the sum of counts n_r for all rankings v where $a \succ_{v_r} b$ strictly exceeds $n/2$. This yields the equivalent exact expression:

$$P(\text{CW}(a)) = \sum_{\substack{n_1, \dots, n_m \in \mathbb{N} \\ \sum_{r=1}^m n_r = n}} \frac{n!}{\prod_{r=1}^m n_r!} \prod_{r=1}^m P(v_r)^{n_r} \cdot \mathbf{1}\left\{\sum_{a \succ_{v_r} b} n_r > \frac{n}{2} \forall b \neq a\right\}, \quad (14)$$

where $(n_r)_{v_r \in \mathcal{L}(\mathcal{A})}$ is the vector of counts of each ranking for all possible profiles.

The outer summation requires iterating over every unique profile configuration. For each individual profile term generated, a naive evaluation must compute the multinomial coefficient and the underlying Mallows sampling probabilities using the Repeated

Insertion Model, incurring a cost of $\mathcal{O}(m^2 \cdot m!)$ per iteration. Consequently, the time complexity scales as

$$\mathcal{O}\left(\binom{m! + n - 1}{n} \cdot m^2 \cdot m!\right).$$

A procedure using dynamic programming to efficiently compute the probability of an alternative being the Condorcet winner is provided as part of the Supplementary Material.

6.6.1 Asymptotic Behaviour of the Condorcet Winner Probability

The exact computation of $P(\text{CW}(a))$ offers a precise characterisation of the probability that alternative a is selected as the Condorcet winner. However, evaluating Equation 14 quickly becomes computationally intractable as the number of voters grows. This motivates an asymptotic analysis, which allows us to examine the fundamental behaviour of the Mallows model in large-scale profiles. In what follows, we study the behaviour of $P(\text{CW}(a))$ as the number of voters tends to infinity, i.e., $n \rightarrow \infty$.

Since voters are assumed to be independent and identically distributed, the limiting behaviour depends critically on the dispersion parameter $\phi \in (0, 1]$ and the central ranking σ . Let $p_{ab} = P(a \succ b)$ denote the marginal probability that a randomly sampled voter strictly prefers alternative a to alternative b . The asymptotic behaviour naturally divides into two cases: models exhibiting a central tendency ($\phi < 1$) and the limiting case corresponding to Impartial Culture ($\phi = 1$).

Case 1: Models with Central Tendency ($\phi < 1$)

When $\phi \in [0, 1)$, the probability distribution over rankings is centred on the central ranking σ . By the Weak Law of Large Numbers [48], the empirical proportion of voters who prefer a to b converges in probability to the marginal probability p_{ab} . Formally, for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{r=1}^n \mathbf{1}\{a \succ_r b\} - p_{ab}\right| < \epsilon\right) = 1.$$

Since $\text{CW}(a)$ requires this empirical proportion to exceed $1/2$ for every $b \neq a$, the asymptotic probability is entirely determined by the position of a in σ :

- **If a is the alternative ranked first in σ :** Consider another alternative b and a ranking v such that $v(a) < v(b)$ with Kendall-tau distance $\kappa(v, \sigma)$ to the central ranking σ . Denote as $d = v(b) - v(a)$ the distance between a and b in v . Now consider v' as the same ranking as v but swapping the positions of a and b . Since the Kendall-tau distance can be defined as the number of discordant pairs between both rankings, moving a , which is the first alternative in σ , d positions upwards will reduce, necessarily, the Kendall-tau distance in d units. At the same time, moving the alternative b d positions downwards will increase, at most, the Kendall-tau distance in $d - 1$ units (since the order of the pair (a, b) is the same in v and σ but different in v'). Therefore, we have that $\kappa(v, \sigma) > \kappa(v', \sigma)$. By the sampling probability formula of the Mallows model we reach that $P(v) < P(v')$. Define the

following two sets of rankings

$$\mathbb{V} = \{v \in \mathcal{L}(\mathcal{A}) \mid v(a) < v(b)\}, \quad \mathbb{V}' = \{v \in \mathcal{L}(\mathcal{A}) \mid v(a) > v(b)\}$$

Using that the function that takes a ranking $v \in \mathbb{V}$ and returns a ranking $v' \in \mathbb{V}'$ which is the one obtained by swapping the alternatives a and b is a bijection between $v \in \mathbb{V}$ and $v' \in \mathbb{V}'$, we reach

$$p_{ba} = \sum_{v \in \mathbb{V}} P(v) < \sum_{v' \in \mathbb{V}'} P(v') = p_{ab}$$

so $p_{ab} > 1/2$ for all $b \in \mathcal{A} \setminus a$. Consequently, as n approaches infinity, the proportion of voters preferring a to every other alternative converges to a value strictly greater than $1/2$, implying

$$\lim_{n \rightarrow \infty} P(\text{CW}(a)) = 1.$$

- **If a is not the alternative ranked first in σ :** Then another alternative b is the one ranked first in σ , so

$$\lim_{n \rightarrow \infty} P(\text{CW}(b)) = 1,$$

which leads, by the uniqueness of the Condorcet winner, to

$$\lim_{n \rightarrow \infty} P(\text{CW}(a)) = 0.$$

Hence, in standard Mallows models with $\phi < 1$, the effect of individual-level noise vanishes asymptotically, and the central ranking uniquely determines the Condorcet winner.

Case 2: The Impartial Culture Limit ($\phi = 1$)

For maximum dispersion, the Mallows model reproduces the Impartial Culture (IC) model. In this setting, all rankings in $\mathcal{L}(\mathcal{A})$ share the same sampling probability, yielding $p_{ab} = 1/2$ for every pair of alternatives.

For a voter r and two alternatives a and b , define the pairwise score variable $X_{ab}^{(r)} = \mathbf{1}\{a \succ_r b\} - \mathbf{1}\{b \succ_r a\}$, which takes values in $\{-1, 1\}$ with equal probability. Its expectation and variance are given by $\mathbb{E}[X_{ab}^{(r)}] = 0$ and $\text{Var}(X_{ab}^{(r)}) = 1$. Moreover, for another alternative c , we have that, from the 6 possible equiprobable orderings of a , b and c , only when a is ranked in the middle position we have that $X_{ab}^{(r)} X_{ac}^{(r)} = -1$, while in the rest of the cases $X_{ab}^{(r)} X_{ac}^{(r)} = 1$. Therefore,

$$\text{Cov}(X_{ab}^{(r)}, X_{ac}^{(r)}) = \mathbb{E}[X_{ab}^{(r)} X_{ac}^{(r)}] - \mathbb{E}[X_{ab}^{(r)}] \mathbb{E}[X_{ac}^{(r)}] = \frac{4-2}{6} - 0 = \frac{1}{3}.$$

To evaluate the Condorcet criterion for alternative a , consider the random vector

$$\mathbf{M}_n = \sum_{r=1}^n \mathbf{X}^{(r)},$$

where $\mathbf{X}^{(r)}$ contains the variables $X_{ab}^{(r)}$ for all $b \neq a$, so $\mathbf{M}_n$ is an $m - 1$ -dimensional random vector. Considering the independence between the voters, m_n is a sum of independent and identically distributed random vectors with finite covariance matrix, see Example 2.18 in [49], we can apply the multivariate central limit theorem, so

$$\frac{1}{\sqrt{n}}\mathbf{M}_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \Sigma),$$

where Σ is an $(m - 1) \times (m - 1)$ covariance matrix with diagonal entries equal to 1 and off-diagonal entries equal to $1/3$.

Alternative a is a Condorcet winner if and only if all components of $\mathbf{M}_n$ are strictly positive. Therefore, the asymptotic probability corresponds to the probability contained in the positive orthant of the limiting multivariate normal distribution:

$$\lim_{n \rightarrow \infty} P(\text{CW}(a)) = P(Z_1 > 0, Z_2 > 0, \dots, Z_{m-1} > 0),$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. This limit is a strictly positive constant that depends only on the number of alternatives m . For instance, in the case of three alternatives ($m = 3$), the orthant probability for two standard normal variables with correlation $1/3$ has a closed expression, (see, for instance [50]), so

$$\lim_{n \rightarrow \infty} P(\text{CW}(a)) = \frac{1}{4} + \frac{\arcsin(1/3)}{2\pi} \approx 0.304.$$

However, it is relevant to note that the computation of such probabilities becomes harder when moving to dimensions higher than 5, see [51].

6.7 Analysis of the Probability Distributions

Following Equation 11, we analytically computed the probability of sampling any profile of the class of profiles with three and five alternatives in which a specific fraction, α , of the voters rank an alternative in a particular position, for all $\alpha \in \{t \in [0, 1] : tn \in \mathbb{N}\}$.

Figure 5 illustrates the evolution of these probabilities as the number of voters increases, in elections of three alternatives and using a Mallows model with 0.9 dispersion⁴. Each cell corresponds to a specific number of voters, an alternative $a \in \mathcal{A}$, a position $j \in \{1, \dots, m\}$, and a range of percentages $[x, y]$ such that $x, y \in \{0, 100\}$. Each percentage range represents an interval of proportions of rankings satisfying the condition, i.e., the range $[x, y]$ represents all profiles with $\alpha \in \{t \in [\frac{x}{100}, \frac{y}{100}] : tn \in \mathbb{N}\}$. Thus, the cell shows the cumulative probability of sampling a profile where a occupies position j in a proportion of rankings for all $\alpha \in \{t \in [\frac{x}{100}, \frac{y}{100}] : tn \in \mathbb{N}\}$. For example, the top cell of the left column in the top-left facet shows the probability (0.07) of sampling a profile with 50 voters where $\sigma(1)$ is in the first position in 0% to 29% of rankings.

⁴Figure A4 of Appendix A reproduces this chart for elections with five alternatives and using a Mallows model with 0.9 dispersion.

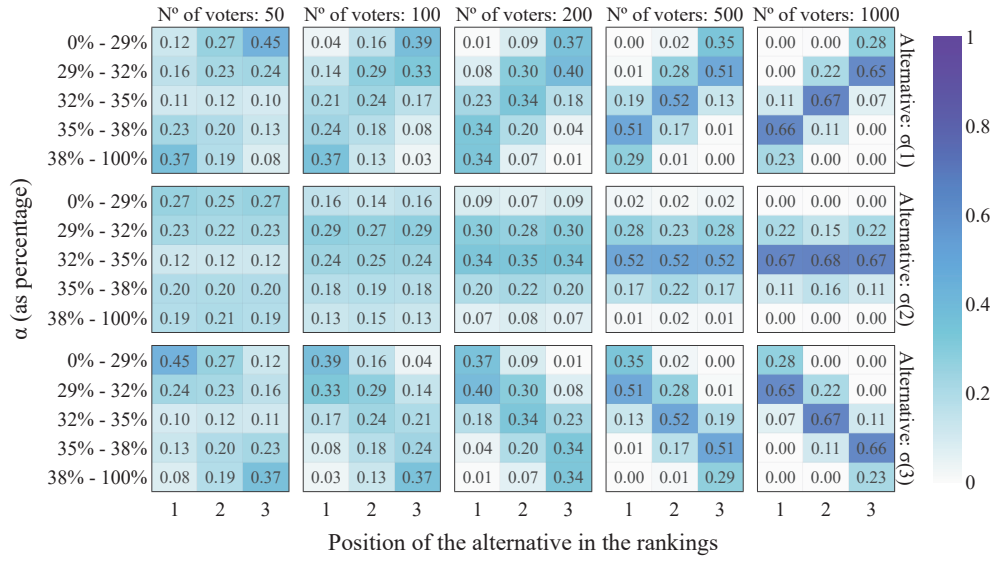


Fig. 5: Evolution of the analytical probability of sampling a profile in which an alternative occupies a given position in a specified fraction of rankings, as the number of voters increases. The figure considers elections with three alternatives under a Mallows model with dispersion 0.9. Each cell reports the cumulative probability (analytically computed) that, for a given alternative $\sigma(i) : i \in \{1, 2, 3\}$, a position $j \in \{1, 2, 3\}$ and a range of percentages, $\sigma(i)$ is ranked in position j by a percentage of voters lying within that interval. Darker shades of purple indicate higher probabilities.

Examining the probability distributions, significant changes can be observed as the number of voters increases. For lower voter counts, the probability is spread across multiple percentage ranges for any alternative and position, whereas it concentrates within a narrower range for larger numbers of voters.

For instance, sampling a profile with 50 voters in which $\sigma(2)$ is in the first position in fewer than 29% of rankings is almost as likely as sampling one in which more than 38% of rankings do so. However, with 1000 voters, sampling a profile where this percentage lies between 32% and 35% yields double the cumulative probability of all other ranges combined. This difference highlights that the RIM fixes ranking probabilities independently of the number of voters, but not profile probabilities. Therefore, the number of voters plays a significant role when sampling profiles with specific properties.

The evolution of these probability distributions varies across alternatives and positions as the number of voters increases. For example, the proportion of first-position occurrences of the top-ranked alternative in the central ranking increases with the voter count. Specifically, the probability of such an event occurring in more than 35% of rankings rises from 0.60 to 0.89 as the number of voters increases from 50 to 1000. In contrast, the alternative ranked last in the central ranking exhibits the opposite

pattern, as the probability that it appears in the first position in more than 35% of rankings shrinks from 0.21 to 0.00.

Moreover, a similar analysis can be conducted for IC, despite all rankings sharing the same probability. Figure 6 reproduces the previous chart for the case of maximum dispersion ($\phi = 1$). The probability distributions also vary as the number of voters increases, making direct comparisons of evaluation results across these scenarios problematic. Unlike lower dispersion settings, all alternatives follow the same pattern in IC, regardless of their position in the central ranking.

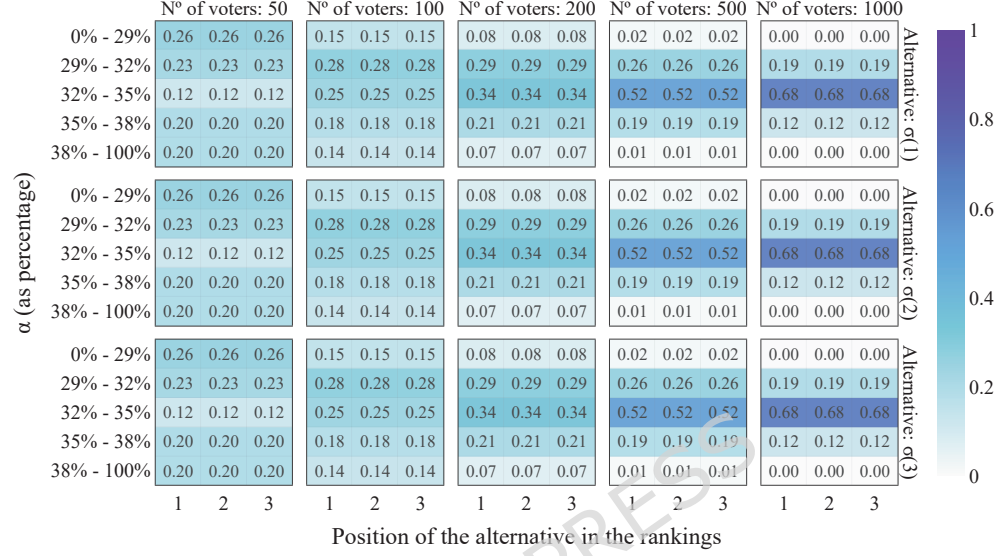


Fig. 6: Evolution of the analytical probability of sampling a profile in which an alternative appears in a given position for a given fraction of rankings as the number of voters increases, considering profiles with three alternatives and IC ($\phi = 1$). For each alternative $\sigma(i)$, position $j = 1, 2, 3$, and percentage range, each cell reports the cumulative probability (analytically computed) of $\sigma(i)$ being ranked in position j by a fraction of voters that lies within that percentage range, with darker purple indicating higher probability.

These findings are particularly relevant when linked to the varying behaviour of the Condorcet criterion observed in Section 5.

The experimental results in Figure 2 show that the top-ranked alternative in the central ranking tends to appear higher in the CR more frequently as the number of voters increases, and thus becomes the CW more often. Considering the aforementioned probabilities, the ranges that accumulate the highest probability are greater for sampling $\sigma(1)$ in the first position than for any other alternative. For example, the most probable range for $\sigma(1)$ is 35%–38% for 1000 voters, compared with 32%–35% for $\sigma(2)$ and 29%–32% for $\sigma(3)$. The alternative $\sigma(1)$ beats all others in every ranking in which it is placed first. For 1000 voters, the probability that $\sigma(1)$ appears in the

first position in more than 35% of rankings is 0.9. Hence, this alternative is expected to beat all others in more than 35% of rankings on average. Moreover, the probability that $\sigma(1)$ appears in the second position in over 32% of rankings is 0.8. In this case, $\sigma(1)$ still beats the alternative ranked third. Together, these results suggest that $\sigma(1)$ can beat all other alternatives in more than 50% of rankings for 1000 voters, increasing its likelihood of being placed higher in the CR when it exists.

Moreover, Figure 7 shows the exact analytical probability (Equation 14) of each alternative emerging as the CW in sampled profiles, with profiles without a CW denoted by NC. The empirical frequencies of each alternative being the CW observed in Figure 4 closely align with the theoretical probabilities. In the case of 0.95 normalised dispersion (internally mapped to ≈ 0.92 Mallows dispersion), $\sigma(1)$ is the CW with higher probability compared to other alternatives, and its probability increases as the number of voters grows. On the contrary, all alternatives share the same probability in IC (norm- $\phi = 1$), and the number of voters has a lesser impact in this case. As described in Section 5, the local minima occurring for even numbers of voters are related to the possibility of alternatives tying in pairwise comparisons when half of the voters have opposite preferences to the remaining half, which cannot happen for odd numbers of voters.

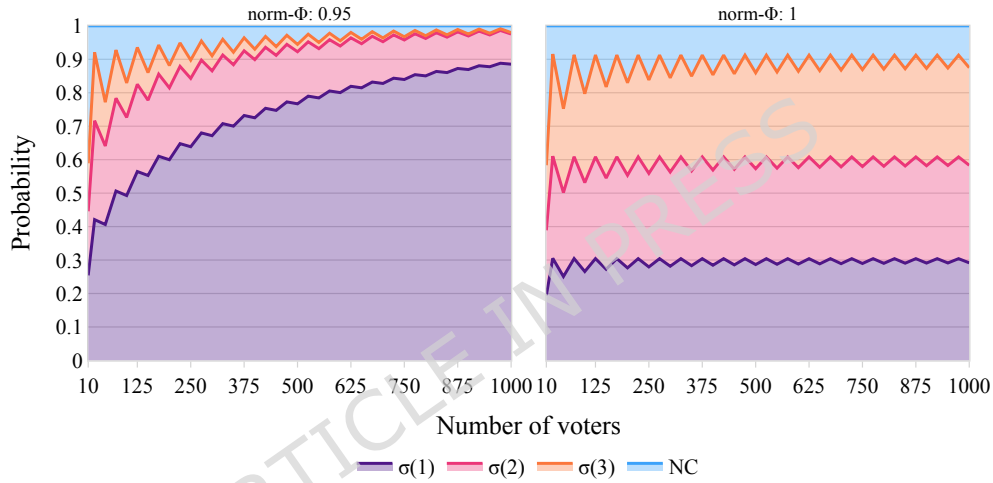


Fig. 7: Evolution of the analytical probability of each alternative emerging as the Condorcet winner (CW) in a sampled profile, as the number of voters increases. This stacked area chart shows the probabilities for profiles with three alternatives and normalised dispersions of 0.95 and 1 (IC), internally converted to ≈ 0.92 and 1 Mallows dispersion values respectively. Profiles without a CW are shown in light blue (NC).

7 Guidelines for Fair Algorithm Evaluation

In the previous section, we demonstrated that the probability of sampling profiles with similar characteristics varies under the Mallows model when the number of voters changes, even with fixed dispersion. This finding is critical for evaluating ranking aggregation algorithms, as it implies that their performance cannot be directly compared across profiles with different voter counts. For example, alternatives tend to be ranked at specific positions as the number of voters increases, even when the other parameters remain fixed. This behaviour can lead to misleading performance analyses, as some ranking aggregation methods may benefit (e.g., through reduced execution time) from such variations, whereas others do not.

Proposing a normalisation of the model to address this issue, similar to that in Subsection 4.2, is challenging. One approach would be to adjust the dispersion so that the expected proportion of rankings in which an alternative appears in a given position remains constant, regardless of the number of voters. However, the performance of different ranking aggregation methods depends on different data properties, and normalising by this specific property may not suit all methods. Moreover, we have shown that sampling probabilities also vary under IC. In this case, adjusting the dispersion to the number of voters is not an option because the uniform probability distribution among rankings would be lost.

Authors should note that varying the number of voters can bias the evaluation of implementations of ranking aggregation algorithms, particularly when comparing results across datasets with different voter counts. For instance, the execution time of an algorithm that implements a Condorcet extension may be lower on average for profiles with more voters. However, CRs and CWs exist more frequently in such profiles, which may favour the algorithm and potentially distort the performance analysis. Therefore, concluding that the algorithm performs better in elections with more voters would be misleading without deeper analysis.

Overall, a fair evaluation protocol requires a thorough examination of the properties of sampled profiles to ensure unbiased results. This involves analysing the data properties that affect algorithm performance, determining whether these properties vary across the dataset, and conducting the performance analysis accordingly. For instance, regarding the example of Condorcet extensions, it is necessary to track the number of profiles with a CR and a CW for each set of profiles with identical parameter values (as in Figure 1 and Figure A1). Execution times across profiles with different voter counts should be compared cautiously while accounting for the presence of CRs and CWs, rather than simply averaging over all profiles. The proposed guidelines for fair evaluation of ranking aggregation algorithms are as follows.

1. Analyse the characteristics of the ranking aggregation algorithms to determine which profile properties can affect their performance.
2. Theoretically analyse the probabilistic behaviour of such properties regarding different parameter configurations of the Mallows model.
 - (a) Use asymptotic approximations to assess such behaviour for larger numbers of voters.

- (b) Use Monte Carlo estimation with explicit error bounds when exact computation of the probabilistic formulas is infeasible.
- 3. Empirically compute these properties for all parameter configurations used.
- 4. Describe the empirical behaviour of the properties when reporting performance metrics.
- 5. Avoid aggregating performance metrics across structurally heterogeneous profiles.
- 6. Stratify results by relevant profile classes according to their properties, such as profiles with a CR, with only a CW, or with neither, conducting separate analysis for each scenario.
- 7. Comparisons of results across different numbers of voters should be conditional on the structural properties of the profiles, rather than only on raw voter counts.
- 8. Provide code, random seeds, and parameter configurations and their transformations (e.g., the normalisation of the dispersion used, if any) to ensure reproducibility.

8 Conclusions

In this paper, we addressed the fair evaluation of ranking aggregation algorithms on synthetic data sampled from the Mallows model with varying numbers of voters. We visualised several metrics related to the Condorcet criterion and empirically showed that they vary across profiles of rankings with different voter counts. To explain this behaviour, we derived the probabilities of sampling different profiles under the Mallows model using the Repeated Insertion Model. We detailed the probabilities of sampling lists of votes and profiles given their rankings, and identified a trade-off between lists of votes with higher probability but fewer equivalents, and those with lower probability but more equivalents.

We then provided the sampling probabilities of profiles that satisfy specific conditions, such as the existence of a Condorcet winner. We showed that the number of voters significantly affects these probabilities, implying that profiles with identical properties (e.g., the same alternative being the CW) are sampled at different frequencies as this parameter varies.

Therefore, the number of voters should not be treated as a neutral scaling parameter when evaluating ranking aggregation algorithms on Mallows-generated profiles. Its effect should be assessed in relation to the structural properties that influence the algorithms under study. To facilitate this task, we have provided a set of guidelines for fair algorithm evaluation. These guidelines emphasise identifying the profile properties that affect algorithm performance, examining how these properties vary across parameter settings and numbers of voters, and reporting performance through stratified and conditional analyses rather than relying solely on aggregate metrics.

An interesting direction for future work is the systematic study, for different ranking aggregation algorithms, of the structural properties that affect their performance, the biases induced by variations in the parameters of the Mallows model, and the appropriate control strategies to ensure fair and standardised evaluation.

Supplementary Information. Authors provide a nomenclature guide in the accompanying Supplementary Information. Authors also provide as part of their Supplementary Information a document with the description of the procedure to efficiently

compute, using dynamic programming, the probability of an alternative emerging as the Condorcet winner, and the implemented Python code. An additional file with the standard errors of the empirical proportions observed in Figures 1 and A1 is also included in the Supplementary Material.

CRedit Authorship Contribution Statement. **Mario Villar:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft. **Noelia Rico:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review & editing. **Juan Baz:** Formal analysis, Validation, Writing – review & editing. **Irene Díaz:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Competing Interests. The authors declare no conflicts of interests.

Acknowledgements. This research has been funded by the Government of Spain [grant number MCINN-23-PID2022-139886NB-I00]. Noelia Rico, Juan Baz and Irene Díaz acknowledge funding from the Asturian Agency for Science, Business Competitiveness and Innovation (SEKUENS) [grant number SEK-25-GRU-GIC-24-018]. Mario Villar has been supported by the “Severo Ochoa” grant program for research and teaching training of the Administration of the Principality of Asturias, Spain [grant number NAC-AT-PUB-ASV-2025BP24-25].

References

- [1] Xiao, L., Chen, Z.-S., Zhang, X., Chang, J.-P., Pedrycz, W., Chin, K.-S.: Bid evaluation for major construction projects under large-scale group decision-making environment and characterized expertise levels. *International Journal of Computational Intelligence Systems* **13**, 1227 (2020) <https://doi.org/10.2991/ijcis.d.200801.002>
- [2] Rúa, A.G., Rico, N., Alonso, A., Díaz, E., Díaz, I.: Ranking the effect of chronodisruption-based biomarkers in reproductive health. *Neural Computing and Applications* **35**, 5697–5720 (2023) <https://doi.org/10.1007/s00521-022-07563-0>
- [3] Chang, X.: Development of spectral clustering algorithm in cognitive diagnosis model: Approach for student’s psychological growth. *International Journal of Computational Intelligence Systems* **18**, 211 (2025) <https://doi.org/10.1007/s44196-025-00849-w>
- [4] Saraireh, J., Agoyi, M., Kassaymeh, S.: Adaptive ensemble learning model-based binary white shark optimizer for software defect classification. *International Journal of Computational Intelligence Systems* **18**, 14 (2025) <https://doi.org/10.1007/s44196-024-00716-0>

- [5] Salih, N., Ksantini, M., Hussein, N., Halima, D.B., Razzaq, A.A., Ahmed, S.: Prediction of rop zones using deep learning algorithms and voting classifier technique. *International Journal of Computational Intelligence Systems* **16**, 86 (2023) <https://doi.org/10.1007/s44196-023-00268-9>
- [6] Fishburn, P.C.: *The Theory of Social Choice*. Princeton University Press, Princeton (1973). <https://www.jstor.org/stable/j.ctt13x10fr>
- [7] Young, H.P.: Condorcet's theory of voting. *American Political Science Review* **82**, 1231–1244 (1988) <https://doi.org/10.2307/1961757>
- [8] Fishburn, P.C.: Paradoxes of voting. *American Political Science Review* **68**, 537–546 (1974) <https://doi.org/10.2307/1959503>
- [9] Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.: *Handbook of Computational Social Choice*, pp. 1–535. Cambridge University Press, New York (2016). <https://doi.org/10.1017/CBO9781107446984>
- [10] Bartholdi, J., Tovey, C.A., Trick, M.A.: Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare* **6**, 157–165 (1989) <https://doi.org/10.1007/BF00303169>
- [11] Zhou, B., Pei, Z., Ma, X.: An improvement method for selecting the best alternative in decision making. *International Journal of Computational Intelligence Systems* **7**, 882 (2014) <https://doi.org/10.1080/18756891.2013.870754>
- [12] Shaye, G.G., Zabil, M.H.M., Albahri, A.S., Joudar, S.S., Hamid, R.A., Albahri, O.S., Alamoodi, A.H., Zahid, I.A., Sharaf, I.M.: Fuzzy evaluation and benchmarking framework for robust machine learning model in real-time autism triage applications. *International Journal of Computational Intelligence Systems* **17**, 151 (2024) <https://doi.org/10.1007/s44196-024-00543-3>
- [13] Solano-Barliza, A., Valls, A., Acosta-Coll, M., Moreno, A., Escorcia-Gutierrez, J., De-La-Hoz-Franco, E., Arregoces-Julio, I.: Enhancing fair tourism opportunities in emerging destinations by means of multi-criteria recommender systems: The case of restaurants in Riohacha, Colombia. *International Journal of Computational Intelligence Systems* **17**, 283 (2024) <https://doi.org/10.1007/s44196-024-00700-8>
- [14] Boehmer, N., Schaar, N.: Collecting, classifying, analyzing, and using real-world ranking data. In: *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pp. 1706–1715 (2023). <https://doi.org/10.48550/arXiv.2204.03589>
- [15] Mallows, C.L.: Non-null ranking models. i. *Biometrika* **44**, 114–130 (1957) <https://doi.org/10.2307/2333244>

- [16] Boehmer, N., Faliszewski, P., Janeczko, L., Kaczmarczyk, A., Lisowski, G., Pierczyński, G., Rey, S., Stolicki, D., Szufa, S., Was, T.: Guide to numerical experiments on elections in computational social choice. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, pp. 7962–7970 (2024). <https://doi.org/10.24963/ijcai.2024/881>
- [17] Gehrlein, W.V., Fishburn, P.C.: Probabilities of election outcomes for large electorates. *Journal of Economic Theory* **19**, 38–49 (1978) [https://doi.org/10.1016/0022-0531\(78\)90054-6](https://doi.org/10.1016/0022-0531(78)90054-6)
- [18] Gehrlein, W.V.: Condorcet’s Paradox vol. 40, 1st edn. Springer, Heidelberg (2006). <https://doi.org/10.1007/3-540-33799-7>
- [19] Villar, M., Rico, N., Díaz, I.: Exploring Condorcet properties in Mallows model-generated preferences. In: 20th Conference of the Spanish Association for Artificial Intelligence, pp. 65–70 (2024)
- [20] Villar, M., Rico, N., Díaz, I.: Understanding data properties in the mallows model: Impact of voter count variability. In: 14th Conference of the European Society for Fuzzy Logic and Technology. *Advances in Fuzzy Logic and Technology, Lecture Notes in Computer Science*, vol. 15883, pp. 294–305. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-97225-6_24
- [21] Guilbaud, G.-T.: Les théories de l’intérêt général et le problème logique de l’agrégation. *Revue Économique* **63**, 659–720 (1952) <https://doi.org/10.3917/reco.634.0659>
- [22] Brill, M., Israel, J., Micha, E., Peters, J.: Individual representation in approval-based committee voting. *Social Choice and Welfare* **64**, 69–96 (2025) <https://doi.org/10.1007/s00355-024-01563-w>
- [23] Micha, E., Shah, N.: Can we predict the election outcome from sampled votes? *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 2176–2183 (2020) <https://doi.org/10.1609/aaai.v34i02.5593>
- [24] Kliachkin, A., Psaroudaki, E., Mareček, J., Fotakis, D.: Fairness in ranking: Robustness through randomization without the protected attribute. In: 40th International Conference on Data Engineering Workshops, pp. 201–208 (2024). <https://doi.org/10.1109/ICDEW61823.2024.00032>
- [25] Yahiro, K., Yokoo, M.: Game theoretic analysis for two-sided matching with resource allocation. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, pp. 1548–1556 (2020). <https://dl.acm.org/doi/10.5555/3398761.3398939>
- [26] Ayadi, M., Amor, N.B., Lang, J., Peters, D.: Single transferable vote: Incomplete knowledge and communication issues. In: Proceedings of the 18th International

- Conference on Autonomous Agents and MultiAgent Systems, pp. 1288–1296 (2019). <https://dl.acm.org/doi/10.5555/3306127.3331833>
- [27] Bentert, M., Skowron, P.: Comparing election methods where each voter ranks only few candidates. *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 2218–2225 (2020) <https://doi.org/10.1609/aaai.v34i02.5598>
 - [28] Brill, M., Schmidt-Kraepelin, U., Suksompong, W.: Margin of victory for tournament solutions. *Artificial Intelligence* **302**, 103600 (2022) <https://doi.org/10.1016/j.artint.2021.103600>
 - [29] Lu, T., Boutilier, C.: Multi-winner social choice with incomplete preferences. In: *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 263–270 (2013). <https://dl.acm.org/doi/10.5555/2540128.2540168>
 - [30] Bachrach, Y., Lev, O., Lewenberg, Y., Zick, Y.: Misrepresentation in district voting. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 81–87 (2016). <https://dl.acm.org/doi/10.5555/3060621.3060633>
 - [31] Boehmer, N., Faliszewski, P., Niedermeier, R., Szufa, S., Was, T.: Understanding distance measures among elections. In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pp. 102–108 (2022). <https://doi.org/10.24963/ijcai.2022/15>
 - [32] Szufa, S., Faliszewski, P., Skowron, P., Slinko, A., Talmon, N.: Drawing a map of elections in the space of statistical cultures. In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1341–1349 (2020). <https://dl.acm.org/doi/10.5555/3398761.3398916>
 - [33] Boehmer, N., Bredereck, R., Faliszewski, P., Niedermeier, R., Szufa, S.: Putting a compass on the map of elections. In: Zhou, Z.-H. (ed.) *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 59–65 (2021). <https://doi.org/10.24963/ijcai.2021/9>
 - [34] Boehmer, N., Faliszewski, P., Kraiczy, S.: Properties of the Mallows model depending on the number of alternatives: A warning for an experimentalist. In: *Proceedings of Machine Learning Research*, vol. 202, pp. 2689–2711 (2023). <https://doi.org/10.48550/arXiv.2401.14562>
 - [35] Gehrlein, W.V., Lepelley, D.: *Voting Paradoxes and Group Coherence*, 1st edn. Springer, Heidelberg (2011). <https://doi.org/10.1007/978-3-642-03107-6>
 - [36] Roberts, F., Tesman, B.: *Applied Combinatorics*, 3rd edn. CRC Press LLC, Florida (2024)

- [37] Arrow, K.J., Raynaud, H.: Social choice and multicriterion decision-making. *The Economic Journal* **97**, 753 (1987) <https://doi.org/10.2307/2232938>
- [38] Kemeny, J.G.: Mathematics without numbers. *Daedalus* **88**, 577–591 (1959)
- [39] Alshahrani, R.A., Khanzada, T.J.S.: Improved crime prediction using hybrid neural architecture search together with hyperparameter tuning. *International Journal of Computational Intelligence Systems* **18**, 195 (2025) <https://doi.org/10.1007/s44196-025-00888-3>
- [40] Kendall, M.G.: A new measure of rank correlation. *Biometrika* **30**, 81–93 (1938) <https://doi.org/10.2307/2332226>
- [41] Doignon, J.-P., Pekeč, A., Regenwetter, M.: The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika* **69**, 33–54 (2004) <https://doi.org/10.1007/BF02295838>
- [42] Lu, T., Boutilier, C.: Effective sampling and learning for Mallows models with pairwise-preference data. *Journal of Machine Learning Research* **15**, 3963–4009 (2014)
- [43] Rico, N., Vela, C.R., Díaz, I.: Reducing the time required to find the Kemeny ranking by exploiting a necessary condition for being a winner. *European Journal of Operational Research* **305**, 1323–1336 (2023) <https://doi.org/10.1016/j.ejor.2022.07.031>
- [44] Villar, M., Rico, N., Díaz, I.: Profiles of rankings by the Mallows model. Zenodo. doi: 10.5281/zenodo.17358183. (2025). <https://doi.org/10.5281/zenodo.17358183>
- [45] Mattei, N., Walsh, T.: Preflib: A library for preferences <http://www.preflib.org>. In: *Proceedings of the 3rd International Conference on Algorithmic Decision Theory*, pp. 259–270 (2013). https://doi.org/10.1007/978-3-642-41575-3_20
- [46] Rico, N., Alonso, P., Díaz, I.: Kemeny ranking aggregation meets the GPU. *The Journal of Supercomputing* **79**, 10335–10352 (2023) <https://doi.org/10.1007/s11227-023-05058-w>
- [47] Baz, J., Martínez, M., Díaz-Vázquez, S., Montes, S.: On the construction of admissible orders for tuples and its application to imprecise risk matrices. *International Journal of Computational Intelligence Systems* **17**, 169 (2024) <https://doi.org/10.1007/s44196-024-00575-9>
- [48] Laha, R.G., Rohatgi, V.K.: *Probability Theory*, 1st edn. Dover Publications, New York (2020)
- [49] Vaart, A.W.: *Asymptotic Statistics vol. 3*. Cambridge University Press, Cambridge (2000). <https://doi.org/10.1017/CBO9780511802256>

- [50] Abrahamson, I.: Orthant probabilities for the quadrivariate normal distribution. *The Annals of Mathematical Statistics* **35**(4), 1685–1703 (1964) <https://doi.org/10.1214/aoms/1177700391>
- [51] Ridgway, J.: Computation of gaussian orthant probabilities in high dimension. *Statistics and computing* **26**(4), 899–916 (2016) <https://doi.org/10.1007/s11222-015-9578-1>

ARTICLE IN PRESS

Appendix A Additional Figures

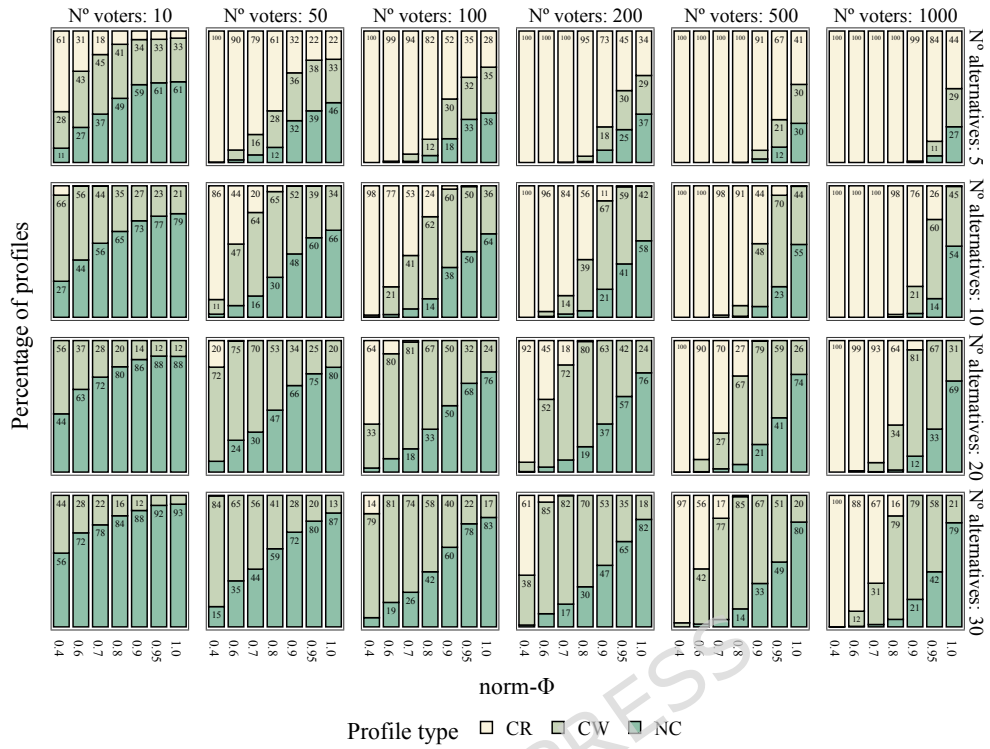


Fig. A1: Condorcet properties of profiles, disaggregated by dispersion level, number of alternatives and number of voters. Each bar reports the percentage of profiles within a given parameter configuration exhibiting a specific Condorcet property, restricted to profiles where a Condorcet ranking (CR) exists, those that exhibit a Condorcet winner (CW), and those with neither (NC). The standard errors of the proportions of CR, CW and NC for every parameter combination are reported in the Supplementary Material.

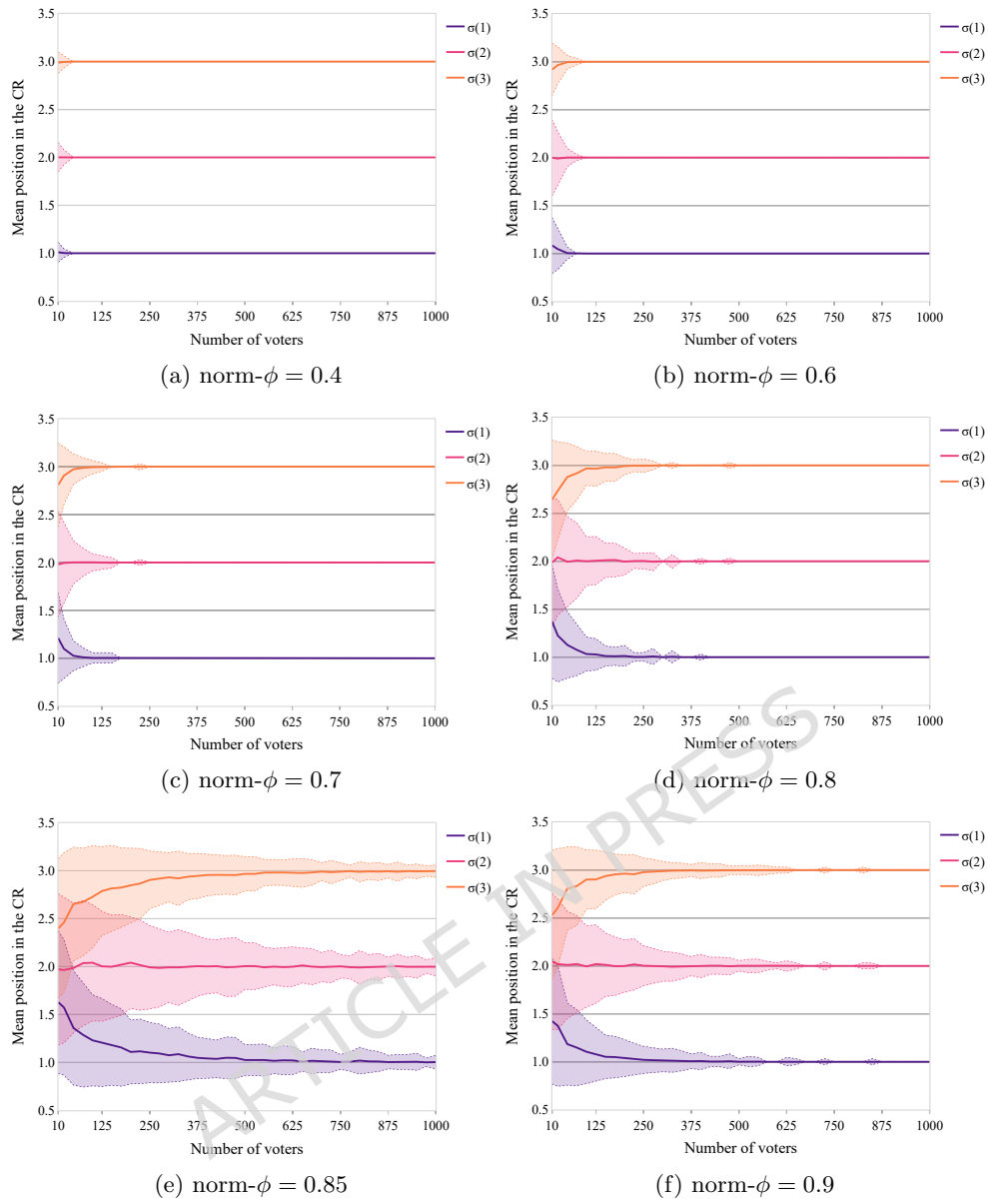


Fig. A2: Evolution of the mean position of alternatives within Condorcet rankings (CRs) as the number of voters increases, for different values of dispersion (displayed in each sub-figure). The analysis is restricted to profiles with three alternatives and where a CR exists. For each alternative, the curve shows the average position in the CR, while the shaded band indicates one standard deviation around the mean.

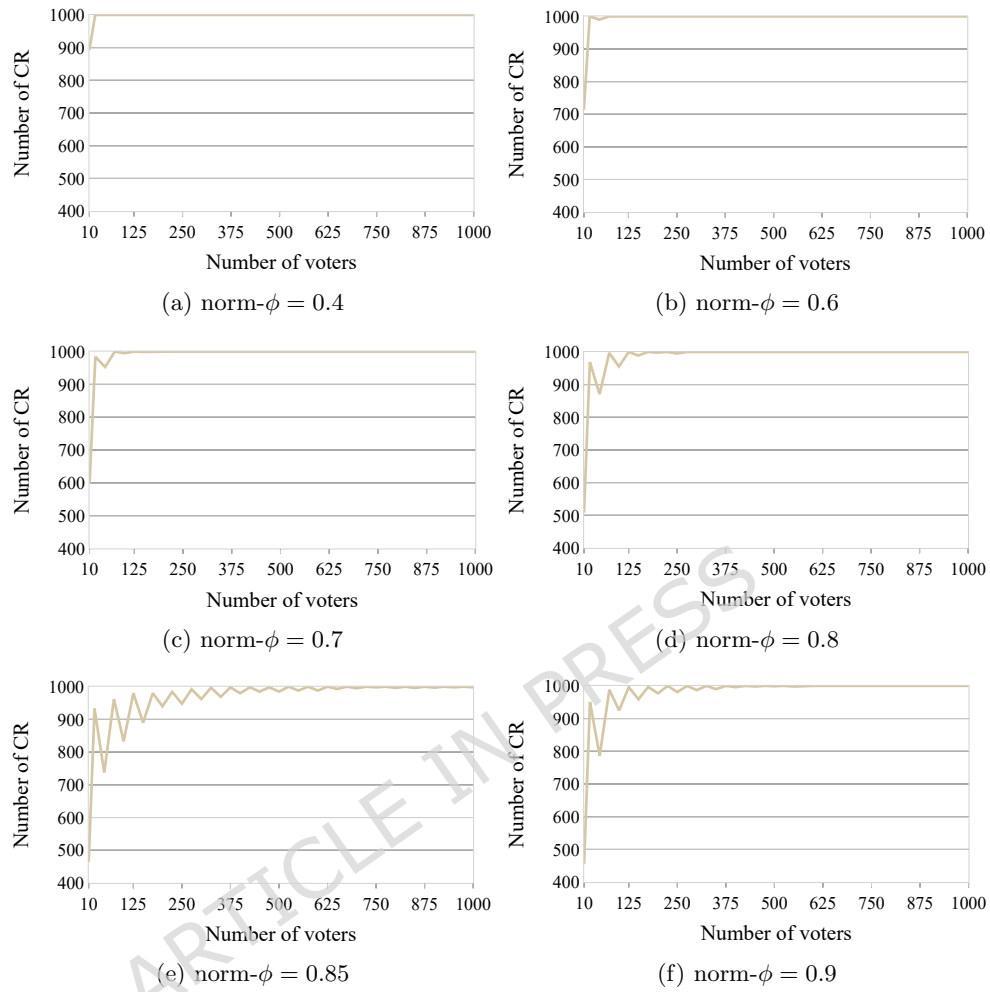


Fig. A3: Evolution of the number of profiles where the CR exists as the number of voters increases, for profiles with three alternatives and different values of normalised dispersion.

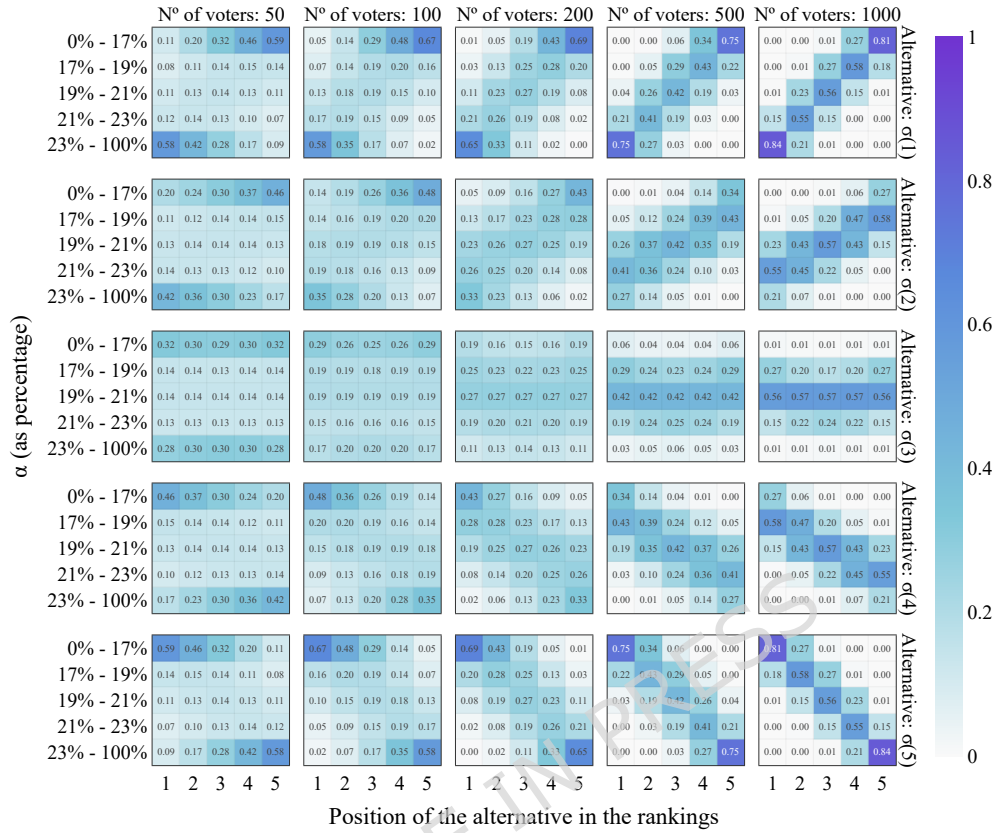


Fig. A4: Evolution of the analytical probability of sampling a profile in which an alternative occupies a given position in a specified fraction of rankings, as the number of voters increases. The figure considers elections with five alternatives under a Mallows model with dispersion 0.9. Each cell reports the cumulative probability (analytically computed) that, for a given alternative $\sigma(i) : i \in \{1, 2, 3\}$, a position $j \in \{1, 2, 3\}$ and a range of percentages, $\sigma(i)$ is ranked in position j by a percentage of voters lying within that interval. Darker shades of purple indicate higher probabilities. Each of the column facets corresponds to a different number of voters, and each of the row facets to a different alternative of the central ranking. The x-axis of each individual facet displays all the available positions, while the y-axis gathers the ranges of percentages.