Properties of the Mallows Model Depending on the Number of Alternatives: A Warning for an Experimentalist

Niclas Boehmer, Piotr Faliszewski, and Sonja Kraiczy

Abstract

The Mallows model is a popular distribution for ranked data. We empirically and theoretically analyze how the properties of rankings sampled from the Mallows model change when increasing the number of alternatives. We find that real-world data behaves differently than the Mallows model, yet is in line with its recent variant proposed by Boehmer et al. [10]. As part of our study, we issue several warnings about using the model.

1 Introduction

The Mallows model [38] is among the simplest and the most popular means of generating and explaining ranking data. Indeed, it is even referred to as the normal distribution over permutations. The model has two main components, the central order (of the available alternatives) and the dispersion parameter $\phi \in [0, 1]$. Depending on the value of ϕ, it either generates random rankings that are more concentrated around the central one or are more evenly spread over the space of all permutations. For example, the central order may rank the athletes participating in some competition with respect to their expected performance, and we can use the Mallows model to generate a realistic set of results in a series of contests. Using values of ϕ close to 0 means that the differences in the strengths of the athletes are large and it is unlikely that a weaker one would outperform a stronger one, whereas using values of ϕ close to 1 means that their abilities are similar and even the weakest participants may win some of the contests.

This ability to generate realistic data is among the prime applications of the Mallows model (another one is understanding ranking data by estimating the Mallows parameters that are most likely to lead to it). While real-life datasets typically involve fairly small sets of alternatives and contain limited numbers of rankings, synthetic models can provide arbitrarily large ranking profiles (i.e., collections of rankings over the same alternatives). This flexibility is useful, e.g., when evaluating various algorithms, such as those for aggregating search results [29], analyzing elections [8], preference learning [37], distribution testing [20], clustering [21], and for many other settings. Yet, preparing synthetic datasets can be tricky. To illustrate this, let us consider two common scenarios.

In the first one, we want to evaluate how the results of an algorithm depend on the number of alternatives. The most obvious approach here is to use the same, fixed dispersion parameter for all the generated profiles and numbers of alternatives. This approach is taken, e.g., in the works of Meila et al. [40], Ali and Meila [1], Lu and Boutilier [37], Brandt et al. [18], Ayadi et al. [3, 4], Chakraborty et al. [24], Busa-Fekete et al. [20], and others. It is so natural that it essentially never calls for any justification. We conjecture that it is based on the following, implicit assumption (needed to interpret the results).

Assumption 1. Using the Mallows model with a fixed dispersion parameter ϕ for different numbers of alternatives produces "structurally similar" ranking profiles.

In the second scenario, we assume that there is no prior knowledge regarding the structure of the ranking profiles that we want to evaluate our algorithm on. In such a case, a very common approach is to simply choose the dispersion parameter uniformly at random, either from the $[0, 1]$ interval, some subinterval of it, or from some discrete set of equally-spaced values, such as $\{0.7, 0.8, 0.9, 1\}$. This strategy is taken, e.g., by Lu and Boutilier [37], Busa-Fekete et al. [19, 20], Bachrach et al. [5], Ayadi et al. [3], Bentert and Skowron [7], and others. As in the previous scenario, the approach is viewed to be so natural as to not need any justification. We conjecture that it is due to the following underlying assumption.

Assumption 2. Using the Mallows model with the dispersion parameter ϕ chosen uniformly at random (approximately) uniformly covers the space between identity profiles, where all rankings are equal, and profiles where each ranking is selected uniformly at random.

Yet, Boehmer et al. [10] provided an interpretation under which these two assumptions are false: Regarding Assumption 1, they found that as we fix the dispersion parameter and increase the number alternatives, the Mallows model generates rankings that, on average, become more and more similar to the central one (where the similarity is measured in terms of the swap distance, i.e., the number of pairs of alternatives ranked differently by the two rankings, normalized by the total number of pairs of alternatives). Regarding Assumption 2, their observation implies that if we choose the dispersion parameter uniformly at random, then the more alternatives we consider, the more the distribution is skewed toward choosing identity profiles. They concluded by introducing a normalized variant of the dispersion parameter, referred to as norm-ϕ, which under their interpretation satisfies the two assumptions and which they used in their subsequent works [11, 12, 13, 14, 15, 30, 16].

If one accepts their interpretation (or, if their interpretation is supported by real-life data) then quite a number of experiments based on the Mallows model become questionable. Indeed, this goes even beyond our two scenarios. For example, if one learned a realistic value of ϕ using data with one number of alternatives, but used it to produce test data with a different number of alternatives, then under the interpretation of Boehmer et al. [10] this generated data is structurally different.

The purpose of this paper is to analyze and compare, theoretically and experimentally, how the properties of the Mallows model depend on the number of alternatives in case we use the classic or the normalized dispersion parameter. We find that both variants have properties that may be natural in some settings, but the data we consider points toward the normalized variant. Proofs of statements marked with (★) can be found in the appendix. The code for our experiments is available at `github.com/Project-PRAGMA/Normalized-Mallows-ICML-2023`.

Related Literature. The Mallows model has applications in many different fields, and we provide a brief overview of those relevant to the machine learning community. Motivated by the belief that user preferences can be understood as being sampled from the Mallows model or one of its variants [23, 26, 31], the problem of fitting the parameters of the Mallows model has been studied extensively, e.g., when user preferences are given as strict rankings [2, 36, 11], incomplete rankings [25, 27, 34, 21], or pairwise comparisons [37, 19]. In addition, different algorithms for this task with applications in crowdsourcing [19], recommendation systems [43, 34], and clustering [21] have been implemented [33, 35]. Additionally, the Mallows model has also proven useful in the context of Estimation of Distribution Algorithms for permutation-based problems [22] applied in the field of evolutionary computation [23]. Motivated by the variety of applications, many properties of the Mallows model have already been studied, including the cycle structure [32], the longest increasing subsequence of a sampled ranking [41] and the thermodynamic limit [42].

2 Preliminaries

A ranking profile $E = (C, V)$ consists of a set $C = \{c_1, \ldots, c_m\}$ of alternatives and a collection $V = (v_1, \ldots, v_n)$ of rankings (our notation largely follows that used in the voting literature [17], where a ranking profile would typically be called an *election*). Each ranking (sometimes also called an order or a vote) is a strict, total order over C that ranks the alternatives from the best to the worst. We write $\mathcal{L}(C)$ to denote the set of all rankings over alternative set C. For a ranking $v \in \mathcal{L}(C)$ and an alternative $c \in C$, let $\mathrm{pos}(v, c)$ be the position of c in v; the first alternative has position 1, the next one has position 2, and so on. Given two rankings $u, v \in \mathcal{L}(C)$, we write $\kappa(u, v)$ to denote their swap distance, i.e., the number of pairs of distinct alternatives $c, d \in C$ whose relative ranking is different in u and v (i.e., in one of the rankings c is ranked higher than d, and in the other it is the opposite). The maximum swap distance of two rankings over m alternatives is $\binom{m}{2}$.

Under the popular Plurality voting rule, the Plurality score of an alternative in a ranking profile is the number of rankings where the alternative appears in the first position. The alternative with the highest Plurality score is called a Plurality winner (we will use the Plurality voting rule to get simple aggregate features of ranking profiles).

Mallows Model. The Mallows model $\mathcal{M}_{\phi,m,v^*}$ is parameterized by a central order $v^* \in \mathcal{L}(C)$ over $m := |C|$ alternatives, and a dispersion parameter $\phi \in [0, 1]$. The probability of sampling a ranking $v \in \mathcal{L}(C)$ under $\mathcal{M}_{\phi,m,v^*}$ is $\frac{1}{Z(\phi,m)}\phi^{\kappa(v^*,v)}$, where $Z(\phi, m)$ is a normalization constant known to be $(1 + \phi) \cdot (1 + \phi + \phi^2) \cdot \ldots \cdot (1 + \phi + \ldots + \phi^{m-1})$. Consequently, for $\phi = 0$ only the central order v^* is sampled, whereas using $\phi = 1$ leads to a uniform distribution over rankings from $\mathcal{L}(C)$, also known as *Impartial Culture (IC)*.

In the following, we fix the central order v^* to order the alternatives lexicographically, i.e., to rank c_1 first, then c_2, and so on. Hence, we will often write $\mathcal{M}_{\phi,m}$ instead of $\mathcal{M}_{\phi,m,v^*}$. Next, we find the probability that c_1 appears in the ith position in a sampled ranking.

Fact 2.1 (Awasthi et al. [2]). *For all $i \in [m]$, $\mathbb{P}_{v\sim\mathcal{M}_{\phi,m}}[\mathrm{pos}(v, c_1) = i] = \frac{\phi^{i-1}}{\sum_{j=1}^{m}\phi^{j-1}}$.*

2.1 Measuring Properties of Mallows Model

We are interested in measuring various properties of rankings sampled from the Mallows model $\mathcal{M}_{\phi,m}$, such as, e.g., the probability that c_1 is ranked first. For this, let $X_{\phi,m}$ be a random variable capturing some property $\mathcal{X}$ we are interested in. We define its normalized expected value to be:

$$g^{\mathcal{X}}_m(\phi) = \frac{\mathbb{E}[X_{\phi,m}] - \inf_{\phi' \in [0,1]} \mathbb{E}[X_{\phi',m}]}{\sup_{\phi' \in [0,1]} \mathbb{E}[X_{\phi',m}] - \inf_{\phi' \in [0,1]} \mathbb{E}[X_{\phi',m}]}.$$

For instance, for the probability that c_1 is ranked first, Fact 2.1, together with the observation that for $\phi = 0$, c_1 is ranked on the first position with probability 1 and for $\phi = 1$ it is ranked first with probability $1/m$, implies that this function is $g^{\mathrm{pos}(c_1)=1}_m(\phi) = \frac{m}{m-1} \cdot \left(\frac{1}{\sum_{j=1}^{m}\phi^{j-1}} - \frac{1}{m}\right)$. Let us assume that $g^{\mathcal{X}}_m$ is a bijection[1] and define $\phi^{\mathcal{X}}_m := (g^{\mathcal{X}}_m)^{-1}$. This function gives the dispersion parameter that leads to the requested normalized expected value of $X_{\phi,m}$. We say that we parameterize the Mallows model by property $\mathcal{X}$ (or more precisely by $g^{\mathcal{X}}_m$, the normalized expected value of $X_{\phi,m}$) if instead of specifying the dispersion parameter ϕ explicitly, we specify the value $\ell \in [0, 1]$ (of $g^{\mathcal{X}}_m$) and use dispersion parameter $\phi^{\mathcal{X}}_m(\ell)$ to sample from $\mathcal{M}_{\phi^{\mathcal{X}}_m(\ell),m}$.

[1] For all properties $\mathcal{X}$ that we consider in our theoretical analysis, the expected value $\mathbb{E}[X_{\phi,m}]$ is strictly monotonic with respect to ϕ. This is sufficient for $g^{\mathcal{X}}_m$ to be bijective.

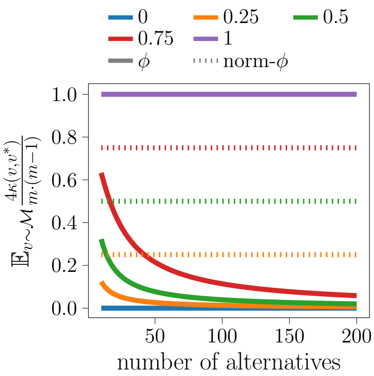


Figure 1: Expected normalized swap distance of a sampled ranking from the central one (solid lines for the classic model and dashed ones for the normalized variant).

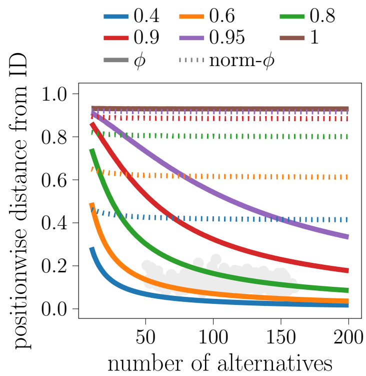


Figure 2: Average positionwise distance from ID of profiles with $n = 100$ rankings (see Section 4.1). Lightgrey points are from Figure 5b.

2.2 Normalized Mallows Model

Let us consider the random variable equal to the swap distance between the sampled ranking and the central one (for $\phi = 0$ its expected value is 0; for $\phi = 1$ it is $m(m-1)/4$). We denote its normalized expected value as:

$$g_m^{\text{swap}}(\phi) = \frac{4 \cdot \mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v, v^*)]}{m(m-1)}. \tag{1}$$

Using the terminology from the previous section, the normalized Mallows model of Boehmer et al. [10] is simply the Mallows model parameterized by the normalized expected swap distance (between the sampled ranking and the central order). When we use this expected value as a free variable, then—following Boehmer et al. [10]—we denote it by norm-ϕ and refer to it as the normalized dispersion parameter[2] (note that at the end of the preceding section, norm-ϕ would take the role of ℓ). Accordingly, using the normalized Mallows model, one specifies a value of norm-$\phi \in [0, 1]$, which is then internally converted to a corresponding value ϕ of the dispersion parameter such that $g_m^{\text{swap}}(\phi) = \text{norm-}\phi$ (see Section 3.2). Rankings are then sampled from $\mathcal{M}_{\phi,m}$ (using standard algorithms such as the Repeated Insertion Model; see the start of Section 3.1 for a description).

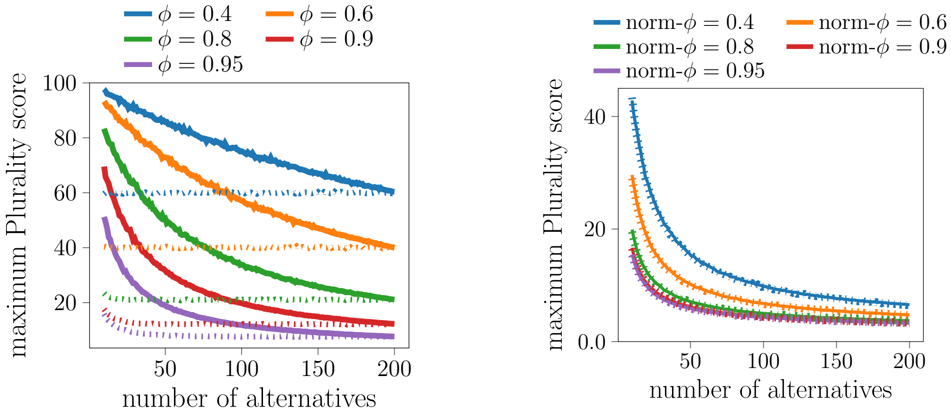
To get an exact formula for $g_m^{\text{swap}}(\phi)$, it suffices to replace $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v, v^*)]$ in (1) with the following result.

Fact 2.2 (Diaconis and Ram [28], Property 4)**.** *Given dispersion parameter $\phi \in [0, 1)$, the expected swap distance between the central order and a sampled one is:*

$$\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v, v^*)] = \tfrac{m\phi}{1-\phi} - \textstyle\sum_{i=1}^{m} i\tfrac{\phi^i}{1-\phi^i}.$$

Using Fact 2.2 we are also able to explain the experimental observation of Boehmer et al. [10] that for a fixed dispersion parameter sampled rankings become more and more similar to the central one as we increase the number of alternatives. Indeed, for a fixed dispersion parameter $\phi \in [0, 1)$, $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v, v^*)]$ grows at most linearly in m, but the denominator in (1) is $m(m-1)$. Hence, the normalized expected swap distance goes to 0 as m grows.

[2] In Lemma A.1 in Appendix A, we show its well-definedness, i.e., that each norm-ϕ gives rise to a unique ϕ, which was missing in the work of Boehmer et al. [10].



(a) Mallows model

(b) Normalized Mallows model

Figure 3: Average Plurality score of Plurality winner in 100 rankings with a varying number of alternatives. We compare sampling profiles with a varying number of m (dashed) with sampling profiles for $m = 200$ alternatives and subsequently deleting some alternatives uniformly at random (solid).

Corollary 2.3. *For fixed* $\phi < 1, \lim_{m\to\infty} g_m^{\mathrm{swap}}(\phi) = 0$.

To visualize the speed of convergence and the difference between using the classic and normalized models, in Figure 1 we show how the expected normalized swap distance changes for fixed values of ϕ and norm-ϕ.

3 Mallows Versus Normalized Mallows

In this section we provide our main comparison of the classic and normalized variants of the Mallows model. In particular, we consider a number of properties—such as, e.g., the expected position of c_1 in a sampled ranking—and we evaluate if for a fixed (normalized) dispersion parameter value the property is almost independent of the number of alternatives, or if for all values of the (normalized) dispersion parameter it converges to the same constant as the number of alternatives grows.

We view our properties as measuring various structural properties of the sampled rankings (or profiles) and if a given property does not depend strongly on the number of alternatives (for a given variant of the Mallows model), then we say that from the perspective of this property, Assumption 1 holds (for this variant). To better understand the two Mallows models, we first build some intuitions regarding their behavior in Section 3.1, and then, in Section 3.2, we analyze how ϕ relates to norm-ϕ. Afterward (Section 3.3), we introduce our framework for analyzing the properties and then perform the analysis (Sections 3.4 to 3.6).

3.1 Intuitions for the Two Models

It will be helpful to consider the classic Mallows model through the lenses of the Repeated Insertion Model (Diaconis and Ram [28], Property 3; see Lu and Boutilier [37], Section 2.2.3 for a more accessible description). The idea is that to sample a ranking from $\mathcal{M}_{\phi,m}$, we can first sample a ranking from $\mathcal{M}_{\phi,m-1}$, which regards one alternative fewer, and then insert the missing alternative c_m at some position j. Specifically, the probability of inserting c_m at position $j \in [m]$ is $\frac{\phi^{m-j}}{\sum_{i=0}^{m-1}\phi^i}$. After doing so, we can also insert alternatives c_{m+1}, c_{m+2}, and so on, if we so choose and want to obtain a ranking over more alternatives. Hence, sampling a ranking from the Mallows model is an iterative procedure that inserts the alternatives into the sampled ranking one by one, following the central order. In other words, we can imagine the sampling process as first sampling a ranking over many more than m alternatives, and then restrict it to the top m from the central order.

On the other hand, for the normalized Mallows model it is not possible to iteratively expand rankings like this, as for a fixed value of the normalized dispersion parameter norm-ϕ the corresponding value of ϕ increases with increasing m.[3] However, we can use the following intuition: Instead of sampling a ranking of m alternatives, we can sample a ranking over a large number of alternatives, but then restrict it to *randomly selected* m ones. As there is no theory to back this view, we perform the following experiment. For each number m of alternatives ranging between 20 and 200, we first sample a profile of 100 rankings according to the (normalized) Mallows model for 200 alternatives, and then we delete a random subset of them so that only m remain. Then we compute the Plurality score of the Plurality winner.[4] We repeat this experiment for the case where we sample a profile of 100 rankings over exactly m alternatives (without any deleting). The results are in Figure 3. We see that both experiments give nearly identical results for the normalized Mallows model (supporting our intuitive view of it), but vary greatly for the classic one.

Looking at it from a yet different perspective, when increasing the number of alternatives in the classic Mallows model, we add an alternative at the end of the central order, leaving the relation between all the other alternatives intact. In contrast, for the normalized Mallows model, we add an alternative and then move all alternatives "closer together," in the sense of increasing the probability that each pair of them may be swapped during sampling (for a fixed value of norm-ϕ, the respective value of ϕ increases when increasing m). It is sometimes also helpful to view this as inserting an alternative of random quality, in contrast to inserting an alternative worse than all already present ones as in the classic Mallows model.

3.2 Relation Between ϕ and norm-ϕ

To analyze rankings sampled from the normalized Mallows model, we want to better understand the relation between ϕ and norm-ϕ. While we already presented a formula for $g_m^{\text{swap}}(\phi)$ in Section 2 (allowing to move from ϕ to norm-ϕ), a closed formula for the other direction seems elusive. Nonetheless, we provide an asymptotic result.

Theorem 3.1 (★). *Fix $\ell \in [0,1]$. Then, $\lim_{m\to\infty}(1-\phi_m^{\text{swap}}(\ell)) \times m = h^{\text{swap}}(\ell)$, where $h^{\text{swap}}(\ell)$ is the unique solution L to $\int_0^1 \gamma(s,L)ds = \frac{c}{4}$ with $\gamma : (0,1] \times \mathbb{R} \to \mathbb{R}, (s,x) \mapsto \frac{1}{x} - s\frac{1}{e^{sx}-1}$. Further, h^{swap} is a bijective strictly decreasing function from $(0,1]$ to $[0,\infty)$,*

Note that for fixed $\ell \in [0,1]$, the above implies that $\phi_m^{\text{swap}}(\ell)$ behaves asymptotically like $1 - \frac{h^{\text{swap}}(\ell)}{m}$. However, to sample rankings from the normalized model, in the absence of a closed form expression for ϕ_m^{swap}, we need to find a different way to convert values of norm-ϕ to values of ϕ. As $g_m^{\text{swap}}(\phi)$ is strictly monotonic (see Lemma A.1 in the appendix), given some value of norm-ϕ, we can simply perform a binary search on $\phi \in [0,1]$ until $g_m^{\text{swap}}(\phi) = \text{norm-}\phi$.

3.3 On Covering a Property Asymptotically

In the remainder of this section, we want to analyze how various properties of sampled rankings behave as we change the number of alternatives, while keeping ϕ or norm-ϕ fixed.

[3] Yet, if the number of alternatives is known upfront, the Repeated Insertion Model can still be used to sample rankings from the normalized Mallows model (after converting norm-ϕ to the respective value of ϕ).

[4] We show similar results for other properties in Appendix D.

We focus on the case where the number of alternatives goes to infinity.

Definition 3.2. We say that parameterizing by property $\mathcal{Y}$ *asymptotically covers* property $\mathcal{X}$ if there is a bijective and strictly monotonic function $f : [0,1] \to [0,1]$ such that for all $\ell \in [0,1]$: $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(\ell)) = f(\ell)$.

Alternatively, we say that parameterizing by property $\mathcal{Y}$ *asymptotically cannot distinguish* property $\mathcal{X}$ if for all $\ell \in (0,1)$: $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(\ell)) = L$, for some constant L.

Intuitively speaking, if parameterizing by property $\mathcal{Y}$ asymptotically cannot distinguish property $\mathcal{X}$, then for all $\ell \in (0,1)$, when keeping a fixed value ℓ of $\mathcal{Y}$ (i.e., we select some $\ell \in (0,1)$ and sample from $\mathcal{M}_{m,\phi_m^{\mathcal{Y}}(\ell)}$), property $\mathcal{X}$ converges to the same value as m goes to infinity. Practically speaking, this means that as the number of alternatives increases, sampled rankings become more and more similar with respect to the range of potential values of property $\mathcal{X}$. In contrast, if $\mathcal{Y}$ asymptotically covers $\mathcal{X}$, then there is a "well-behaved" mapping f between all possible expected values of $\mathcal{X}$ and all possible expected values of $\mathcal{Y}$ such that if we parameterize the Mallows model by a fixed value ℓ of $\mathcal{Y}$ and increase the number of alternatives, the value of property $\mathcal{X}$ converges to $f(\ell)$. Consider the expected swap distance from the central order as an illustrative toy property (cf. Figure 1). Parameterizing by the dispersion parameter asymptotically cannot distinguish this property, whereas (by definition) parameterizing by the normalized dispersion parameter asymptotically covers it.

Notably, for all "well-behaved" properties our two notions are symmetric.

Proposition 3.3 (★). *Let $\mathcal{X}$ and $\mathcal{Y}$ be properties such that $g_m^{\mathcal{X}}(\phi)$ and $g_m^{\mathcal{Y}}(\phi)$ are strictly monotonic and continuous.*[5] *If parameterizing by property $\mathcal{Y}$ asymptotically covers (cannot distinguish) property $\mathcal{X}$, then parameterizing by property $\mathcal{X}$ asymptotically covers (cannot distinguish) property $\mathcal{Y}$.*

Thus, two parameterizations are "similar" if they asymptotically cover each other, while they are different if they asymptotically cannot distinguish each other.

While asymptotic coverage of some property $\mathcal{X}$ does not directly imply that for the respective variant of the Mallows model property $\mathcal{X}$ stays constant for a fixed parameter value when changing the number of alternatives, it is clearly a necessary condition. Moreover, we observe empirically for all considered variants and properties that asymptotic coverage of some property corresponds to the property staying approximately constant in practice.

In the following, we focus on parameterizing by the dispersion parameter (recovering the classic Mallows model) and by the expected swap distance (recovering the normalized Mallows model). Accordingly, we will say that the normalized/classic Mallows model asymptotically covers or cannot distinguish some property if this holds for the respective parameterization. It turns out that if the classic model asymptotically covers some property, then the normalized one cannot asymptotically distinguish it, and the other way round.

Theorem 3.4 (★). *Let $\mathcal{X}$ be a property such that $g_m^{\mathcal{X}}(\phi)$ is strictly monotonic:*

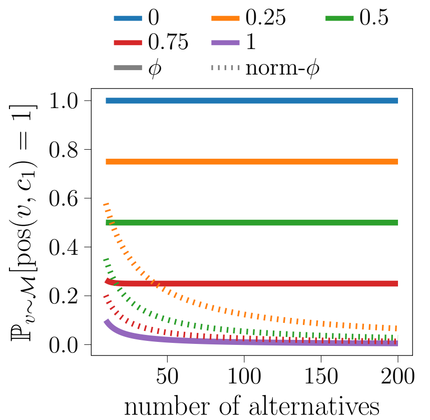
1. *If the normalized Mallows model asymptotically covers property $\mathcal{X}$, the classic Mallows model asymptotically cannot distinguish $\mathcal{X}$.*
2. *If the classic Mallows model asymptotically covers property $\mathcal{X}$, then the normalized Mallows model asymptotically cannot distinguish $\mathcal{X}$.*

Hence, the classic and the normalized models are very different and, in particular, exhibit a fundamentally different behavior with respect to all properties considered in the paper.

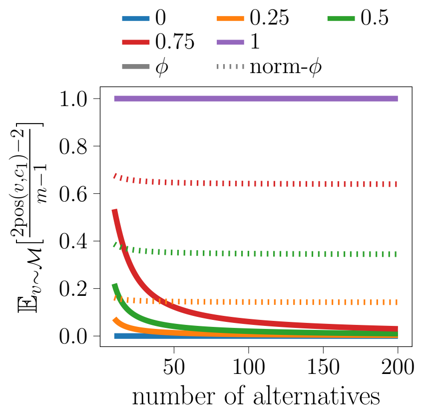
3.4 Position of the Central Order's Top-Choice

In this and the following sections we use our asymptotic framework to analyze basic properties of sampled rankings and ranking profiles. We first look at the position of c_1.

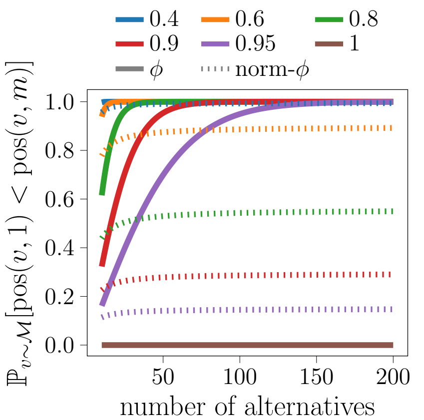
[5] These conditions hold for all properties that we consider.



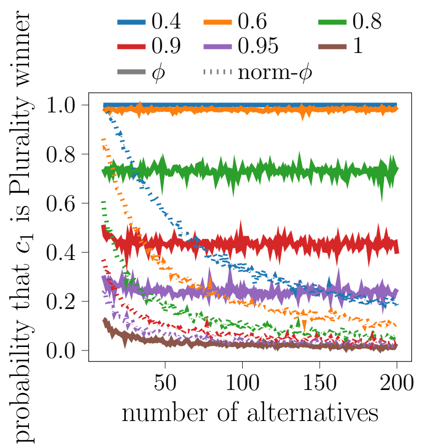
(a) Probability that c_1 is ranked first in a sampled vote.



(b) Expected normalized position of c_1 in a sampled vote.



(c) Probability that c_1 is ranked before c_m.



(d) Fraction of profiles with c_1 as Plurality winner.

Figure 4: Influence of the number of alternatives m on different properties of rankings (ranking profiles) sampled from the Mallows model for fixed values of the classical dispersion parameter ϕ (solid) and the normalized dispersion parameter norm-ϕ (dashed). For $\phi =$ norm-$\phi = 0$ and $\phi =$ norm-$\phi = 1$ the respective lines overlap.

Probability That c_1 Is Ranked First. Consider the probability that c_1 is ranked first in a sampled ranking, i.e., $\mathbb{P}_{v\in\mathcal{M}}[\mathrm{pos}(v,c_1)=1]$. By Fact 2.1, we know that it is $\frac{1}{\sum_{i=0}^{m-1}\phi^i}$. Since $\sum_{i=0}^{m-1}\phi^i$ is a geometric series, we have that $\lim_{m\to\infty}\sum_{i=0}^{m-1}\phi^i = \frac{1}{1-\phi}$. From this, it is easy to conclude the following.

Theorem 3.5. *The classic Mallows model asymptotically covers the probability that c_1 is ranked first, with $f(\ell) = 1-\ell$.*

In fact, as witnessed by Figure 4a, for a fixed value of ϕ the probability that c_1 is ranked first quickly converges to $1-\phi$.

Expected Position of c_1. Next, let us consider the normalized position of c_1, i.e., $g_m^{\mathrm{pos1}}(\phi) = \frac{2\mathbb{E}_{v\sim\mathcal{M}_{\phi,m}}[\mathrm{pos}(v,c_1)]-2}{m-1}$, which describes its average performance.

Proposition 3.6 (★). *The expected position of c_1 in a sampled ranking is $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m}}[\mathrm{pos}(v,c_1)] = \frac{1}{1-\phi} - m\frac{\phi^m}{1-\phi^m}$.*

Given that we discussed above that fixing the dispersion parameter leads to a roughly constant probability of c_1 being ranked first, it is intuitive that this parameterization asymptotically cannot distinguish the expected normalized position of c_1. In contrast, parameterizing by the swap distance can do so.

Theorem 3.7 (★). *The normalized Mallows model asymptotically covers the expected position of c_1, with $f(\ell) = t(h^{swap}(\ell))$, where $t(x) = 2\cdot(\frac{1}{x} - \frac{1}{e^x-1})$.*

Again, as seen in Figure 4b, convergence is reached fairly quickly. The behavior of c_1 in both parameterizations is in line with our discussion from Section 3.1: In the classic model, increasing the number of alternatives means adding them at the end of the central order. Thus, these new alternatives have an exponentially shrinking probability of being ranked first, leading to a convergence of the probability of c_1 being ranked first. For the normalized Mallows model, the intuition is that increasing the number of alternatives means that there are more and more alternatives similar to c_1 (who then can end up ahead of c_1).

3.5 Pairwise Comparisons

Next, we turn to the probability that some alternative c_i is ranked ahead another one, c_j.

Comparing Two Fixed Alternatives. Mallows [38] showed that the probability that c_i is ranked before c_j depends only on the difference between i and j and is independent of the number of alternatives. Using this, we calculate the probability that c_i appears above c_j.

Proposition 3.8 (★). *Consider $1 \le i,j \le m$ with $k := j-i+1$. The probability that c_i is ranked before c_j is*

$$\frac{1}{1-\phi^k}\left(1 - \frac{(1-\phi)(k-1)\phi^{k-1}}{1-\phi^{k-1}}\right).$$

Notably, since for each $i > j \in \mathbb{N}$ $g_m^{i \text{ beats } j}(\phi) = 2\cdot\mathbb{P}_{v\sim\mathcal{M}_{\phi,m}}[\mathrm{pos}(v,c_i) < \mathrm{pos}(v,c_j)]$ does not depend on m, we can say that the classic Mallows model "covers" the probability that c_i is ranked before c_j (for each m) and, so, clearly also covers it asymptotically.

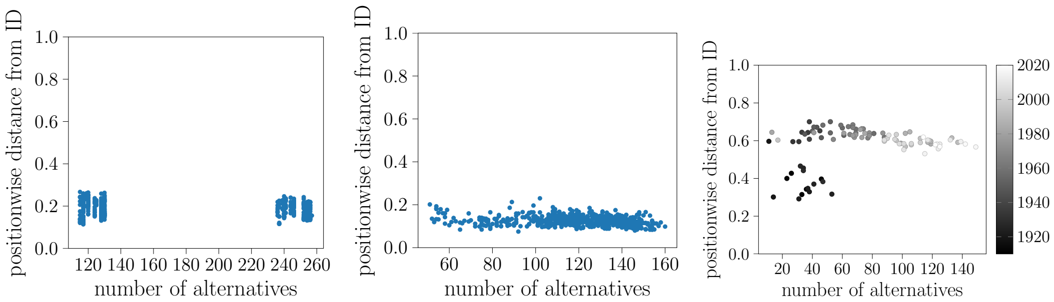
Comparing c_1 and c_m. If instead of comparing the relative ranking of two alternatives with fixed indices, we compare the probability that c_1 is ranked before c_m, the picture changes (see Figure 4c).

Theorem 3.9 (★). *The normalized Mallows model asymptotically covers the probability that alternative c_1 is ranked before alternative c_m in a sampled ranking, with $f(\ell) = t(h^{\mathrm{swap}}(\ell))$, where $t(x) = 2\cdot\frac{1}{1-e^{-x}}(1 - \frac{x}{e^x-1}) - 1$.*

Our discussion from Section 3.1 explains the behavior of the classic model: As we add alternatives at the end of the central order, these added alternatives are viewed as more and more inferior to c_1. Under the normalized model, the quality difference between c_1 and c_m stays roughly constant.

3.6 Winners in Ranking Profiles

More complex properties may concern an entire profile, not just an individual ranking. For example, the results from Section 3.4 show that the classic Mallows model asymptotically



(a) American Football (b) Spotify (c) Tour de France

Figure 5: Plots showing the normalized positionwise distance of a profile from ID depending on the number of alternatives in the profile. Each point corresponds to one profile. For Tour de France, the color of a point corresponds to the year of the respective edition.

covers the expected Plurality score of c_1 (because it is equal to the probability that c_1 is ranked first, times the number of sampled rankings).

For a constant number of votes, we present empirical evidence in Figure 4d that the classical Mallows model asymptotically covers the probability that c_1 is the Plurality winner. In Appendix B.5, we show the same for more rules such as Borda and Condorcet.

4 Examining Real-World Evidence

Next, we evaluate how a notion analogous to the expected swap distance from the central order depends on the number of alternatives in real-world data (in addition, in Appendix C, we examine the properties of alternative c_1 as discussed in the previous section[6]). We analyze the following three datasets of Boehmer and Schaar [9]:[7]

American Football. In each week in a season of American College Football, different media outlets publish their rankings of the teams by estimated strength. Each American football profile regards one week with rankings published in this week. However there are two divisions in college football (FBS and FCS) with some outlets ranking all teams from the FBS division and some outlets ranking all teams from both division. Accordingly, we created two different profiles for each week, one for rankings over FBS teams and one for rankings over FBS and FCS teams. Thus the profiles from this dataset have different sizes (also the numbers of teams vary slightly between years).

Spotify. Spotify profiles are based on the daily top 200 songs on Spotify in different countries. Each profile in this dataset regards a single month in a single country, and contains a single ranking for each day of this month. As a Spotify profile may be incomplete, we delete all alternatives that do not appear in all rankings.

Tour de France. Tour de France is a multistage bike race held annually in France. Each profile corresponds to one edition, and each ranking orders the cyclists in a single stage, where we delete all cyclists that do not finish all stages. Tour de France profiles are of different sizes because the number of participants differs in each year (from 67

[6]In real-world profiles, we use the Plurality winner as a proxy for c_1.

[7]To the best of our knowledge, the work of Boehmer and Schaar [9] is the only one containing complete profiles with a wide spectrum of numbers of alternatives; we selected the three datasets from their paper with the highest variation of alternative numbers.

to 252) and because the fraction of cyclists finishing all races varies (which sometimes reduces the number of cyclists even below 20).

Notably, all three datasets can be seen as ground-truth-based, so the Mallows model is a suitable choice to capture them.[8] We will see that ranking profiles from our datasets behave differently than the classic Mallows model, but mostly in line with the normalized variant.

4.1 Positionwise Distance From Identity

While in a real-life dataset there is no central order, in principle we could estimate it. For example, the normalized Kemeny score of a profile (C, V) is defined as $\min_{v^* \in \mathcal{L}(C)} \sum_{v \in V} \kappa(v, v^*)/(|V|\binom{|C|}{2})$ and the ranking v^* that achieves this minimum is the Kemeny order (technically, it does not need to be unique, but it is not crucial for our discussion). The Kemeny order is a maximum-likelihood estimator for the central order of the Mallows model producing the given profile [39], and the normalized Kemeny score is the normalized swap distance of the profile from this ranking. Unfortunately, computing the Kemeny score of a profile is NP-hard [6]. Thus, we turn to a polynomial-time approximation known as the "positionwise distance from ID,"[9] proposed by Boehmer et al. [10]. Its values range between 0 and 1 and have a similar interpretation as the normalized Kemeny score (in particular, value 0 means that all rankings in the profile are identical and value 1 means that the rankings are maximally diverse). Consequently, in this section we use it as a moral equivalent of the normalized expected swap distance from the central ranking that we would have used, had we been looking at Mallows data.

Before examining the positionwise distance from ID of real-world ranking profiles, in Figure 2, we show its behavior on profiles sampled from the classic and normalized Mallows models; note the intuitive connection to the expected swap distance (Figure 1). Figure 5 shows the normalized positionwise distance from ID of profiles from American Football, Spotify, and Tour de France. We see here that in all three datasets, the positionwise distance from ID stays constant when varying the number of alternatives. We interpret this as evidence that real-world profiles that come from the same source do not behave like those from the classic Mallows model with a fixed dispersion parameter (for which the normalized expected swap distance from the central order goes to 0 as the number of alternatives increases), but rather like those from the normalized variant (where it stays fixed).

Moreover, this evidence highlights a practical problem with using the classic Mallows model: The Kemeny score and the positionwise distance from ID can both be used to estimate the dispersion parameter of the Mallows model (see, e.g., the works of Mandhani and Meila [39] and Boehmer et al. [11]). This estimated parameter might then, for instance, be employed to generate more similar data of varying sizes to conduct further experiments. If the classic dispersion parameter is used, then this approach is problematic: For instance, assume that we have access to Spotify profiles with between 50 and 60 alternatives. This yields estimated dispersion parameter around 0.6; however, if we use this parameter to generate profiles with a higher number of alternatives, then we get data different from the true Spotify profiles with larger alternative numbers, as for $m = 160$ the estimated dispersion parameter is around 0.8 (see lightgrey points in Figure 2). Note further that reporting a

[8] In American Football profiles, for each week, there is some underlying ranking of the teams by strength. In a Spotify profile, one could argue that there is an underlying popularity ranking of songs in each month. In Tour de France profiles, it is usually assumed that there is some true ordering of cyclists by strength.

[9] Here, each ranking profile is modeled by a frequency matrix, where we have one column for each alternative and one row for each position, and an entry contains the fraction of rankings that rank the alternative on this position. The distance between two frequency matrices is then defined as the minimum earth mover's distance between columns over all possible column mappings. The positionwise distance from ID to a profile is the distance of the profile's frequency matrix from the identity matrix.

value of $\phi = 0.8$ for Spotify profiles with $m = 160$ and a value of $\phi = 0.6$ for Spotify profiles with $m = 50$, one might also think that the nature of Spotify elections changes when varying the number of alternatives, while this behavior is rather a feature of the Mallows model.

5 A Final Takeaway: Be Cautious

We analyzed how the Mallows model behaves as we vary the number of alternatives and argued that for the classic variant this behavior is unnatural. Instead, we generally suggest to use the normalized Mallows model which keeps certain structural properties of sampled rankings constant, in line with real-world data.

Independent of whether one agrees with our interpretation of the results, we made several observations that should be taken into account when conducting experiments with data generated from the Mallows model. *First*, one should be extremely careful when conducting experiments with a fixed dispersion parameter and varying m, as it is unclear whether observed trends are because of the increased number of alternatives or the changed structure of the profile. Consider as an example the setup used by Busa-Fekete et al. [19]: They use data generated from the Mallows model to compare their algorithm to a baseline one. Considering $\phi \in \{0.1, 0.3, 0.5, 0.7\}$, they observe for $m = 10$ that their algorithm is better than the baseline algorithm for small values of ϕ. Then, they repeat their experiments with $\phi \in \{0.1, 0.3, 0.5, 0.7\}$ and $m = 20$ and conclude that here "the advantage of [our algorithm] is even more pronounced" (as their algorithm now outperforms the baseline one also for larger values of the dispersion parameter). Considering our findings, it is not clear whether one can really conclude that their algorithm has a stronger competitive advantage when increasing the number of alternatives, or if for $m = 20$ they looked at data which is simply more favorable for their algorithm. To illustrate this point, consider the following idealized example: There are two algorithms A and B, where A is better than B on data with a "low" level of disagreement, say, on profiles where the Kemeny score is less than a third of the maximum possible. Then, as an example, for 10 alternatives algorithm A is better than B for $\phi < 0.4$ and for 20 alternatives A is better than B for $\phi < 0.6$. One might conclude that algorithm A scales better in the number of alternatives than algorithm B; however, the reason for this observed trend is that for $m = 20$ the "low range" of the level of disagreement extends to the case where ϕ is below 0.6 (and not only below 0.4 as for 10 alternatives). *Second*, statements about how an algorithm behaves for a certain dispersion parameter or parameter ranges might no longer be true when varying the number m of alternatives. Accordingly, one should clarify that it is unclear whether such statements generalize for different alternative numbers, even if one only conducts experiments with a fixed number of alternatives. *Third*, one should be careful how to select the values of the dispersion parameter used to generate data for experiments to ensure a meaningful coverage of the space of ranking profiles. *Fourth*, the above described problems get intensified when considering generalizations of the Mallows model, such as its mixtures.

Caution is also needed when using the Mallows model to learn preferences. For instance, imagine a questionnaire which asks first, how respondents would rank 5 drafts of conventional ads and, second, how respondents would rank 25 drafts for more provocative ads. Later, one estimates Mallows parameters of the preferences and finds that those over the traditional ads are best captured by a dispersion parameter of $\phi = 0.5$, whereas those over the more provocative ones are best captured by $\phi = 0.9$. One might think that opinions concerning the provocative ads are more varied, but this interpretation is questionable, as we argued that values of ϕ for different numbers of alternatives are in certain ways incomparable.

Acknowledgments

A version of this paper is also presented at the 40th International Conference on Machine Learning (ICML-2023). Niclas Boehmer is supported by the DFG project ComSoc-MPMS (NI 369/22). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101002854).

References

[1] Alnur Ali and Marina Meila. Experiments with kemeny ranking: What works when? *Math. Soc. Sci.*, 64(1):28–40, 2012.

[2] Pranjal Awasthi, Avrim Blum, Or Sheffet, and Aravindan Vijayaraghavan. Learning mixtures of ranking models. In *Proceedings of the 2014 Annual Conference on Neural Information Processing Systems (NeurIPS '14)*, pages 2609–2617, 2014.

[3] Manel Ayadi, Nahla Ben Amor, Jérôme Lang, and Dominik Peters. Single transferable vote: Incomplete knowledge and communication issues. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19)*, pages 1288–1296. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

[4] Manel Ayadi, Nahla Ben Amor, and Jérôme Lang. Approximating voting rules from truncated ballots. *Auton. Agents Multi Agent Syst.*, 36(1):24, 2022.

[5] Yoram Bachrach, Omer Lev, Yoad Lewenberg, and Yair Zick. Misrepresentation in district voting. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI '16)*, pages 81–87. IJCAI/AAAI Press, 2016.

[6] John Bartholdi, Craig A Tovey, and Michael A Trick. Voting schemes for which it can be difficult to tell who won the election. *Soc. Choice Welfare*, 6(2):157–165, 1989.

[7] Matthias Bentert and Piotr Skowron. Comparing election methods where each voter ranks only few candidates. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI '20)*, pages 2218–2225. AAAI Press, 2020.

[8] Nadja Betzler, Robert Bredereck, and Rolf Niedermeier. Theoretical and empirical evaluation of data reduction for exact kemeny rank aggregation. *Auton. Agents Multi Agent Syst.*, 28(5):721–748, 2014.

[9] Niclas Boehmer and Nathan Schaar. Collecting, classifying, analyzing, and using real-world elections. *CoRR*, abs/2204.03589, 2022. To appear in AAMAS '23.

[10] Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Stanislaw Szufa. Putting a compass on the map of elections. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI '21)*, pages 59–65. ijcai.org, 2021.

[11] Niclas Boehmer, Robert Bredereck, Edith Elkind, Piotr Faliszewski, and Stanislaw Szufa. Expected frequency matrices of elections: Computation, geometry, and preference learning. *CoRR*, abs/2205.07831, 2022. doi: 10.48550/arXiv.2205.07831. URL `https://doi.org/10.48550/arXiv.2205.07831`. To appear at NeurIPS '22.

[12] Niclas Boehmer, Robert Bredereck, Piotr Faliszewski, and Rolf Niedermeier. A quantitative and qualitative analysis of the robustness of (real-world) election winners. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, pages 7:1–7:10. ACM, 2022.

[13] Niclas Boehmer, Robert Bredereck, and Dominik Peters. Rank aggregation using scoring rules. *CoRR*, abs/2209.08856, 2022. To appear at AAAI '23.

[14] Niclas Boehmer, Klaus Heeger, and Rolf Niedermeier. Theory of and experiments on minimally invasive stability preservation in changing two-sided matching markets. In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI '22)*, pages 4851–4858. AAAI Press, 2022.

[15] Niclas Boehmer, Klaus Heeger, and Stanislaw Szufa. A map of diverse synthetic stable roommates instances. *CoRR*, abs/2208.04041, 2022. To appear at AAMAS '23.

[16] Daria Boratyn, Wojciech Slomczynski, Dariusz Stolicki, and Stanislaw Szufa. Spoiler susceptibility in multi-district party elections. *CoRR*, abs/2202.05115, 2022.

[17] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[18] Felix Brandt, Johannes Hofbauer, and Martin Strobel. Exploring the no-show paradox for condorcet extensions using ehrhart theory and computer simulations. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19)*, pages 520–528. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

[19] Róbert Busa-Fekete, Eyke Hüllermeier, and Balázs Szörényi. Preference-based rank elicitation using statistical models: The case of mallows. In *Proceedings of the 31th International Conference on Machine Learning (ICML '14)*, pages 1071–1079. JMLR.org, 2014.

[20] Róbert Busa-Fekete, Dimitris Fotakis, Balázs Szörényi, and Emmanouil Zampetakis. Identity testing for mallows model. In *Proceedings of the 2021 Annual Conference on Neural Information Processing Systems 2021 (NeurIPS '21)*, pages 23179–23190, 2021.

[21] Ludwig M. Busse, Peter Orbanz, and Joachim M. Buhmann. Cluster analysis of heterogeneous rank data. In *Proceedings of the Twenty-Fourth International Conference (ICML '07)*, volume 227, pages 113–120. ACM, 2007.

[22] Josu Ceberio, Alexander Mendiburu, and José Antonio Lozano. Introducing the mallows model on estimation of distribution algorithms. In *Proceedings of the 18th International Conference on Neural Information Processing (ICONIP '11)*, pages 461–470. Springer, 2011.

[23] Josu Ceberio, Ekhine Irurozki, Alexander Mendiburu, and Jose A Lozano. A review of distances for the mallows and generalized mallows estimation of distribution algorithms. *Comput. Optim. Appl.*, 62(2):545–564, 2015.

[24] Vishal Chakraborty, Theo Delemazure, Benny Kimelfeld, Phokion G. Kolaitis, Kunal Relia, and Julia Stoyanovich. Algorithmic techniques for necessary and possible winners. *Trans. Data Sci.*, 2(3):22:1–22:23, 2021.

[25] Weiwei Cheng, Jens C. Huhn, and Eyke Hüllermeier. Decision tree and instance-based learning for label ranking. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09)*, pages 161–168. ACM, 2009.

[26] Flavio Chierichetti, Anirban Dasgupta, Shahrzad Haddadan, Ravi Kumar, and Silvio Lattanzi. Mallows models for top-k lists. In *Proceedings of the 2018 Annual Conference on Neural Information Processing Systems (NeurIPS '18)*, pages 4387–4397, 2018.

[27] Fabien Collas and Ekhine Irurozki. Concentric mixtures of mallows models for top-k rankings: sampling and identifiability. In *Proceedings of the 38th International Conference on Machine Learning, (ICML '21)*, pages 2079–2088. PMLR, 2021.

[28] Persi Diaconis and Arun Ram. Analysis of systematic scan metropolis algorithms using iwahori-hecke algebra techniques. *Michigan Math. J.*, 48(1):157–190, 2000.

[29] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation methods for the web. In *Proceedings of the Tenth International World Wide Web Conference (WWW '01)*, pages 613–622. ACM, 2001.

[30] Piotr Faliszewski, Krzysztof Sornat, and Stanislaw Szufa. The complexity of subelection isomorphism problems. In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI '22)*, pages 4991–4998. AAAI Press, 2022.

[31] Michael A Fligner and Joseph S Verducci. Distance based ranking models. *J. R. Stat. Soc. Series B Stat. Methodol.*, 48(3):359–369, 1986.

[32] Alexey Gladkich and Ron Peled. On the cycle structure of mallows permutations. *Ann. Probab.*, 46(2):1114–1169, 2018.

[33] Ekhine Irurozki, Borja Calvo, and Jose A Lozano. Permallows: An R package for mallows and generalized mallows models. *J. Stat. Softw.*, 71:1–30, 2016.

[34] Guy Lebanon and Yi Mao. Non-parametric modeling of partially ranked data. In *Proceedings of the 2007 Annual Conference on Neural Information Processing Systems (NeurIPS '07)*, pages 857–864. Curran Associates, Inc., 2007.

[35] Paul H Lee and Philip LH Yu. An R package for analyzing and modeling ranking data. *BMC Medical Res. Methodol.*, 13(1):1–11, 2013.

[36] Allen Liu and Ankur Moitra. Efficiently learning mixtures of mallows models. In *Proceedings of the 59th IEEE Annual Symposium on Foundations of Computer Science (FOCS '07)*, pages 627–638. IEEE, 2018.

[37] Tyler Lu and Craig Boutilier. Effective sampling and learning for mallows models with pairwise-preference data. *J. Mach. Learn. Res.*, 15(1):3783–3829, 2014.

[38] Colin L Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.

[39] Bhushan Mandhani and Marina Meila. Tractable search for learning exponential models of rankings. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS '09)*, pages 392–399. JMLR.org, 2009.

[40] Marina Meila, Kapil Phadnis, Arthur Patterson, and Jeff A. Bilmes. Consensus ranking under the exponential model. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI '07)*, pages 285–294. AUAI Press, 2007.

[41] Carl Mueller and Shannon Starr. The length of the longest increasing subsequence of a random mallows permutation. *J. Theor. Probab.*, 26(2):514–540, 2013.

[42] Shannon Starr. Thermodynamic limit for the mallows model on S_n. *J. Math. Phys.*, 50(9):095208, 2009.

[43] Mingxuan Sun, Guy Lebanon, and Paul Kidwell. Estimating probabilities in recommendation systems. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS '11)*, pages 734–742. JMLR.org, 2011.

Niclas Boehmer
Technische Universität Berlin
Berlin, Germany
Email: `niclas.boehmer@tu-berlin.de`

Piotr Faliszewski
AGH University
Krakow, Poland
Email: `faliszew@agh.edu.pl`

Sonja Kraiczy
University of Oxford
Oxford, UK
Email: `Sonja.Kraiczy@merton.ox.ac.uk`

A Additional Material for Section 2.2

The following Lemma implies that g_m^{swap} is strictly increasing and continuous, bijectively mapping $[0,1]$ to $[0,1]$.

Lemma A.1. *$\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v,v^*)]$ is strictly increasing and continuous in $\phi\in[0,1]$ and bijectively maps $[0,1]$ to $[0,\frac{1}{2}\cdot\binom{m}{2}]$.*

Proof. Note that $\mathbb{E}_{v\sim\mathcal{M}_{0,m,v^*}}[\kappa(v,v^*)]=0$ and $\mathbb{E}_{v\sim\mathcal{M}_{1,m,v^*}}[\kappa(v,v^*)]=\frac{1}{2}\cdot\binom{m}{2}$. It remains to show that $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v,v^*)]$ is continuous and strictly increasing on $[0,1]$.

As an immediate consequence of Property 3 in [28] the expected swap distance can be decomposed as $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v,v^*)]=\sum_{i=n}^{m}\mathbb{E}[X_{\phi,n}]$, where $X_{\phi,n}$ represents the number alternatives stronger (i,e, ranked earlier in the central order) than the nth alternative are ranked below the nth alternative in a sample vote v. The random variable $X_{\phi,n}$ has truncated geometric distribution $G_{\phi,n}$, i.e., $\mathbb{P}(X_{\phi,n}=i)=\frac{\phi^{i-1}}{\sum_{j=1}^{n}\phi^{j-1}}$ for all $i\in[n]$, implying that $\mathbb{E}[X_{\phi,m}]=\frac{\sum_{j=1}^{m}j\phi^{j-1}}{\sum_{j=1}^{m}\phi^{j-1}}$. To conclude continuity of $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v,v^*)]$, simply observe that it is a sum of continuous function, since $\mathbb{E}[X_{\phi,m}]$ is the ratio of continuous functions with the denominator being non-zero.

We will show that $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v,v^*)]$ is a sum of functions strictly increasing in ϕ, so also strictly increasing itself. Let $0\le\phi_1<\phi_2\le 1$. We can express

$$\mathbb{E}[X_{\phi_2,m}]=\frac{\sum_{j=1}^{m}j\phi_1^{j-1}+\sum_{j=1}^{m}j(\phi_2^{j-1}-\phi_1^{j-1})}{\sum_{j=1}^{m}\phi_1^{j-1}+\sum_{j=1}^{m}(\phi_2^{j-1}-\phi_1^{j-1})}.$$

So since $\sum_{j=1}^{m}j(\phi_2^{j-1}-\phi_i^{j-1})>\sum_{j=1}^{m}(\phi_2^{j-1}-\phi_1^{j-1})$, it follows that $\mathbb{E}[X_{\phi_1,m}]<\mathbb{E}[X_{\phi_2,m}]$. So as desired, $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v,v^*)]$ is striclty increasing and maps [0,1] bijectively to $[0,\frac{1}{2}\cdot\binom{m}{2}]$. ☐

B Additional Material for Section 3

B.1 Missing proofs from Section 3.2

Theorem 3.1 (★). *Fix $\ell\in[0,1]$. Then, $\lim_{m\to\infty}(1-\phi_m^{\mathrm{swap}}(\ell))\times m=h^{\mathrm{swap}}(\ell)$, where $h^{\mathrm{swap}}(\ell)$ is the unique solution L to $\int_0^1\gamma(s,L)ds=\frac{c}{4}$ with $\gamma:(0,1]\times\mathbb{R}\to\mathbb{R},(s,x)\mapsto\frac{1}{x}-s\frac{1}{e^{sx}-1}$. Further, h^{swap} is a bijective strictly decreasing function from $(0,1]$ to $[0,\infty)$,*

To prove Theorem 3.1 we will need Lemma B.4 and Lemma B.5. The proof of Lemma B.4 in turn uses Proposition B.1, Proposition B.2 and Lemma B.3, which we will prove first. Note that Lemma B.3 will also be used in proofs in later subsections.

Proposition B.1. *Let $\gamma:(0,1]\times\mathbb{R}\to\mathbb{R},(s,x)\mapsto\frac{1}{x}-\frac{s}{e^{sx}-1}$. The following hold*

1. *For fixed $s\in(0,1]$, $\gamma(s,x)$ is strictly decreasing in x.*
2. *f is uniformly continuous*
3. *$\int_0^1\gamma(s,0)ds=\int_0^1\frac{1}{2}sds=\frac{1}{4}$*

Proof. We can reexpress $\gamma(s,x)$ as

$$\gamma(s,x)=\frac{1}{x}-\frac{s}{\sum_{i=0}^{\infty}\frac{(sx)^i}{i!}-1}=\frac{\sum_{i=0}^{\infty}\frac{(sx)^i}{i!}-sx-1}{x\cdot(\sum_{i=0}^{\infty}\frac{(sx)^i}{i!}-1)}=\frac{\sum_{i=2}^{\infty}\frac{(sx)^{i-1}}{i!}}{x\cdot(\sum_{i=1}^{\infty}\frac{(sx)^{i-1}}{i!})}=\frac{\sum_{i=2}^{\infty}\frac{x^{i-2}s^{i-1}}{i!}}{\sum_{i=1}^{\infty}\frac{(sx)^{i-1}}{i!}}$$

so $\gamma(s,x)$ is continuous and differentiable as it is the ratio of continuous and differentiable functions, and the function in the denominator always takes values ≥ 1. For fixed $s \in (0,1]$, we compute the partial derivative with respect to x:

$$\frac{d}{dx}\gamma(s,x) = -\frac{1}{x^2} + \frac{s^2}{(e^{sx}-1)^2} = -\frac{1}{x^2} + \frac{1}{(x+\frac{sx^2}{2}+\frac{s^2x^3}{3!}+\dots)^2} < -\frac{1}{x^2} + \frac{1}{x^2} = 0,$$

since $s > 0$ so that γ is strictly decreasing in x, proving **point 1**. Since $\lim_{x\to\infty}\gamma(s,x) = 0$ and $\lim_{x\to-\infty}\gamma(s,x) = s$ and f is strictly decreasing in x it is bounded in $[0,s] \subset [0,1]$. From continuity and boundedness we conclude that $\gamma(s,x)$ is uniformly continuous, proving **point** 2. From our reformulation of γ, we see that $\lim_{x\to 0} f(x,s) = \frac{s}{2}$ so its integral along the unit interval is $\frac{1}{4}$, proving **point** 3. □

Proposition B.2. *Let $x, w \in \mathbb{R}$ and $s \in [0,1]$. Then*

$$\frac{1}{x} - \frac{s}{e^{s(x+w)}-1} = \frac{ws}{x(w+x)\times\sum_{i=1}^{\infty}\frac{(s(x+w))^{i-1}}{i!}} + \frac{w+x}{x}\times f(s,x+w).$$

Proof.

$$\frac{1}{x} - \frac{s}{e^{s(x+w)}-1} = \frac{e^{s(x+w)}-sx-1}{x\times(e^{s(x+w)}-1)} \tag{2}$$

$$= \frac{\sum_{i=0}^{\infty}\frac{(s(x+w))^i}{i!}-sx-1}{x\times\sum_{i=0}^{\infty}\frac{(s(x+w))^i}{i!}-1} = \frac{ws+\sum_{i=2}^{\infty}\frac{(s(x+w))^i}{i!}}{x\times\sum_{i=1}^{\infty}\frac{(s(x+w))^i}{i!}} \tag{3}$$

$$= \frac{w}{x(x+w)\times\sum_{i=1}^{\infty}\frac{(s(x+w))^{i-1}}{i!}} + \frac{w+x}{w+x}\frac{\sum_{i=2}^{\infty}\frac{(s(x+w))^{i-1}}{i!}}{x\times\sum_{i=1}^{\infty}\frac{(s(x+w))^{i-1}}{i!}}, \tag{4}$$

$$\tag{5}$$

where Equation 3 follows using the Maclaurin series for Eulers function and

$$\frac{w+x}{w+x}\frac{\sum_{i=2}^{\infty}\frac{(s(x+w))^{i-1}}{i!}}{x\times\sum_{i=1}^{\infty}\frac{(s(x+w))^{i-1}}{i!}} \tag{6}$$

$$= \frac{w+x}{x}\frac{\sum_{i=2}^{\infty}\frac{(s(x+w))^{i-1}}{i!}}{(w+x)\times\sum_{i=1}^{\infty}\frac{(s(x+w))^{i-1}}{i!}} \tag{7}$$

$$= \frac{w+x}{x}\frac{e^{s(x+w)}-s(x+w)-1}{(w+x)\times e^{s(x+w)}-1} \tag{8}$$

$$= \frac{w+x}{x}\left(\frac{1}{(w+x)} - \frac{s}{e^{s(w+x)}-1}\right) = \frac{w+x}{x}f(s,x+w), \tag{9}$$

so $\frac{1}{x} - \frac{s}{e^{s(x+w)}-1} = \frac{ws}{x(w+x)\times\sum_{i=1}^{\infty}\frac{(s(x+w))^{i-1}}{i!}} + \frac{w+x}{x}\times f(s,x+w)$, as desired. □

Lemma B.3. *If $\lim_{m\to\infty}(1-\phi_m)\times m = L$, then $\lim_{m\to\infty}\log(\phi_m^{-m}) = L$.*

Proof. Since $\lim_{m\to\infty}(1-\phi_m)\times m = L$, $\lim_{m\to\infty}(1-\phi_m) = 0$ and so

$$\lim_{m\to\infty}\log(\phi_m^{-m}) = \lim_{m\to\infty} -m\log(\phi_m) = \lim_{m\to\infty} m\times\sum_{i=1}^{\infty}\frac{(1-\phi_m)^i}{i}$$

$$= \lim_{m\to\infty} m\times(1-\phi_m) + \sum_{i=2}^{\infty}m(1-\phi_m)\frac{(1-\phi_m)^{i-1}}{i} = L,$$

using the Maclaurin series of the logarithm. □

Lemma B.4. *For $c \in [0,1]$. Suppose ϕ_m is a sequence in $[0,1]$, satisfying*

$$\lim_{m\to\infty}(1-\phi_m)\times m = L \text{ and } \lim_{m\to\infty} g_m^{swap}(\phi_m) = c.$$

Then L is the unique solution to $4\int_0^1 \gamma(s,L)ds = c$ with $f:(0,1]\times\mathbb{R}\to\mathbb{R},(s,x)\mapsto \frac{1}{x} - s\frac{1}{e^{sx}-1}$.

Proof. Let $x_m = (1-\phi_m)\times m$ and let $w_m = \sum_{i=2}^{\infty} m(1-\phi_m)\frac{(1-\phi_m)^{i-1}}{i}$. By Lemma B.3, we know that $\lim_{m\to\infty}\log(\phi_m^{-m}) = \lim_{m\to\infty} x_m + w_m = L$.

Let $s(i) = \frac{i}{m} \in [0,1]$ for $1 \le i \le m$. Then

$$\phi_m^{-i} = e^{-i\log(\phi_m)} = e^{\frac{i}{m}\times(-m\log(\phi_m))} = e^{s(i)\times(x_m+w_m)}. \tag{10}$$

$$\begin{aligned}
\frac{c}{4} &= \frac{1}{4}\cdot\lim_{m\to\infty} g_m^{swap}(\phi_m) = \lim_{m\to\infty}\frac{\mathbb{E}_{v\sim\mathcal{M}_{\phi_m,m,v^*}}[\kappa(v,v^*)]}{m^2} &(11)\\
&= \lim_{m\to\infty}\frac{\phi_m}{m(1-\phi_m)} - \frac{1}{m^2}\sum_{i=1}^{m} i\frac{1}{\phi_m^{-i}-1}) &(12)\\
&= \lim_{m\to\infty}(\frac{\phi_m-1}{m(1-\phi_m)} + \frac{1}{m(1-\phi_m)} - \frac{1}{m^2}\sum_{i=1}^{m} i\frac{1}{\phi_m^{-i}-1}) &(13)\\
&= 0 + \lim_{m\to\infty}(\frac{1}{x_m} - \frac{1}{m^2}\sum_{i=1}^{m} i\frac{1}{\phi_m^{-i}-1}) &(14)\\
&= \lim_{m\to\infty}(\frac{1}{m}\sum_{i=1}^{m}\frac{1}{x_m} - \frac{i}{m}\frac{1}{\phi_c(m)^{-i}-1}) &(15)\\
&= \lim_{m\to\infty}\frac{1}{m}(\sum_{i=1}^{m}\frac{1}{x_m} - s(i)\frac{1}{e^{s(i)(x_m+w_m)}-1}) &(16)
\end{aligned}$$

Equation 15 equals Equation 16 using Equation 10. It remains to show the first equality in the following, as the second follows by definition of the Riemann integral.

$$\lim_{m\to\infty}\frac{1}{m}(\sum_{i=1}^{m}\frac{1}{x_m} - s(i)\frac{1}{e^{s(i)(x_m+w_m)}-1}) = \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\gamma(s(i),L) = \int_0^1 \gamma(s,L)ds. \tag{17}$$

By Proposition B.2,

$$\frac{1}{x_m} - \frac{s}{e^{s(x_m+w_m)}} = \frac{w_m s}{x_m(w_m+x_m)\times\sum_{i=1}^{\infty}\frac{(s(x_m+w_m))^{i-1}}{i!}} + \frac{w_m+x_m}{x_m}\times\gamma(s,x_m+w_m).$$

We will need the following two limits, which both follow from $\lim_{m\to\infty}(1-\phi_m) = 0$:

$$\lim_{m\to\infty}\frac{w_m}{x_m(x_m+w_m)} = \lim_{m\to\infty}\frac{\sum_{i=2}^{\infty}\frac{(1-\phi_c(m))^{i-1}}{i}}{\sum_{i=1}^{\infty} m\frac{(1-\phi_c(m))^{i}}{i}} = \lim_{m\to\infty}\frac{\sum_{i=2}^{\infty}\frac{(1-\phi_c(m))^{i-2}}{i}}{\sum_{i=1}^{\infty} m\frac{(1-\phi_c(m))^{i-1}}{i}} = 0. \tag{18}$$

$$\lim_{m\to\infty}\frac{w_m}{x_m} = \lim_{m\to\infty}\sum_{i=2}^{\infty}\frac{(1-\phi_c(m))^{i-1}}{i} = 0. \tag{19}$$

From this we can conclude that $\frac{w_m s}{x_m(w_m+x_m)\times\sum_{i=1}^{\infty}\frac{(s(x_m+w_m))^{i-1}}{i!}}$ converges uniformly against 0 (independent of s; recall that $s\in(0,1]$), as

$$0 \leq \frac{w_m s}{x_m(w_m+x_m)\times\sum_{i=1}^{\infty}\frac{(s(x_m+w_m))^{i-1}}{i!}} \leq \frac{w_m s}{x_m(w_m+x_m)} \leq \frac{w_m}{x_m(w_m+x_m)}$$

and trivially $\frac{w_m+x_m}{x_m}$ converges uniformly against 1. By Proposition B.1 $\gamma(s,x)$ is uniformly continuous over $(0,1]\times\mathbb{R}$. This means that for any $\epsilon>0$, there exists a $\delta>0$ such that if $|x-x'|<\delta$ then $|\gamma(s,x)-\gamma(s,x')|<\epsilon$ for all $s\in(0,1]$. So since $\lim_{m\to\infty}x_m+w_m=L$, $\gamma(s,x_m+w_m)$ converges uniformly against $\gamma(s,L)$. We conclude that $\frac{1}{x_m}-\frac{s}{e^{s(x_m+w_m)}}$ converges uniformly (in terms of s) against $\gamma(s,L)$. So for any $\epsilon>0$ and for any s there is some m_0, such that for $m\geq m_0$

$$\gamma(s,L)-\epsilon \leq \frac{1}{x_m}-\frac{s}{e^{s(x_m+w_m)}} \leq \gamma(s,L)+\epsilon,$$

and in particular for all $i\in[m]$

$$\gamma(s(i),L)-\epsilon \leq \frac{1}{x_m}-\frac{s}{e^{s(x_m+w_m)}} \leq \gamma(s(i),L)+\epsilon,$$

and hence

$$\frac{1}{m}\sum_{i=1}^{m}(\gamma(s(i),L)-\epsilon) \leq \frac{1}{m}\sum_{i=1}^{m}(\frac{1}{x_m}-\frac{s}{e^{s(x_m+w_m)}}) \leq \frac{1}{m}\sum_{i=1}^{m}(\gamma(s(i),L)+\epsilon)$$

implying the first equality in Equation 17. By Proposition B.1, $\gamma(s,x)$ is strictly decreasing in x and hence $4\cdot\int_0^1\gamma(s,x)ds$ is strictly decreasing in x too, implying in particular that for each $c\in[0,1]$, there is a unique x satisfying the equality. We have shown that $\lim_{m\to\infty}(1-\phi_m)\times m=L$ satisfies $4\cdot\int_0^1\gamma(s,L)ds=c$, implying it is the unique solution, as desired, $\square$

We remind the reader that h^{swap} is defined by $h^{\text{swap}}(c)=\lim_{m\to\infty}(1-\phi_m^{\text{swap}}(c))\times m$.

Lemma B.5. *The function h^{swap} is a bijective strictly decreasing function from $(0,1]$ to $[0,\infty)$.*

Proof. We shown that $\lim_{m\to\infty}(1-\phi_m^{\text{swap}}(c))\times m=L$ is the unique solution to $\int_0^1\gamma(s,L)ds=\frac{c}{2}$. We now evaluate this integral. Using a change of variables and wolfram alpha, we obtain

$$\int_0^1\gamma(s,L)ds=\frac{1}{L}\int_0^1 1-sL\frac{1}{e^{sL}-1}ds \tag{20}$$

$$=\frac{1}{L}(1-\int_0^L z\frac{1}{e^z-1}\frac{dz}{L}) \tag{21}$$

$$=\frac{1}{L}-\frac{1}{L^2}\int_0^L z\frac{1}{e^z-1}dz \tag{22}$$

$$=\frac{1}{L}-\frac{1}{L^2}[z\log(1-e^{-z})-Li_2(e^{-z})]]_0^L \tag{23}$$

$$=\frac{1}{L}-\frac{1}{L^2}(L\log(1-e^{-L})-Li_2(e^{-L})+Li_2(1)) \tag{24}$$

$$=\frac{1}{L}-\frac{\log(1-e^{-L})}{L}+\frac{Li_2(e^{-L}))}{L^2}-\frac{Li_2(1)}{L^2}, \tag{25}$$

where Equality 23 was verified using Wolfram Alpha. where $Li_2(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^2}$ is the Dilogarithm, a function strictly increasing on the reals. Let $r(L) = 4\int_0^1 \gamma(s, L)ds$ is strictly decreasing on $[0, \infty)$ with $r(0) = 4\int_0^1 f(s, 0)ds = 4\int_0^1 \frac{1}{2}sds = 1$ and $\lim_{L\to\infty} = 0$. So r maps $[0, \infty)$ bijectively to $[1, 0)$ and is strictly decreasing. Since $h^{swap}(c) = L$, h^{swap} is the inverse of r, $h^{swap} = r^{-1}$ and therefore strictly decreasing and maps $[1, 0)$ bijectively to $[0, \infty)$, as required. □

Proof of Theorem 3.1. Remember that for $c \in [0, 1]$, $\phi_m^{\text{swap}}(c)$ is defined as to satisfy

$$g_m^{swap}(\phi_m^{\text{swap}}(c)) = \frac{4 \cdot \mathbb{E}_{v\sim\mathcal{M}_{\phi_m^{\text{swap}}(c),m,v^*}}[\kappa(v, v^*)]}{m(m-1} = c \tag{26}$$

$$\implies \lim_{m\to\infty} \frac{\mathbb{E}_{v\sim\mathcal{M}_{\phi_m^{\text{swap}}(c),m,v^*}}[\kappa(v, v^*)]}{m^2} = \frac{c}{2}, \text{ where } 0 < \frac{c}{2} < \frac{1}{4}. \tag{27}$$

If $c = 0$, $\phi_m^{\text{swap}} = 0$ for all m and if $c = 1$, then we must have $\phi_m^{\text{swap}} = 0$ for all m.

Now fix $1 > c > 0$. Let $x_m = m \times (1 - \phi_m^{\text{swap}}(c))$ and note that x_m is non-negative for all m. Suppose $x_{k(m)}$ is a convergent subsequence of x_m with limit $L \geq 0$. By Lemma B.4, L is the unique solution x to $\int_0^1 \gamma(s, x)ds = c$ and $\lim_{m\to\infty} g_m^{swap}(\phi_m^{\text{swap}}(c)) = \int_0^1 \gamma(s, x)ds$. By Proposition B.1, $\int_0^1 f(x, 0)dx = \frac{1}{4}$, so if $L = 0$ then

$$\lim_{m\to\infty} g_m^{swap}(\phi_m^{\text{swap}}(c)) = \lim_{m\to\infty} \frac{4 \cdot \mathbb{E}_{v\sim\mathcal{M}_{\phi_m^{\text{swap}}(c),m,v^*}}[\kappa(v, v^*)]}{m(m-1)} = 1,$$

thereby contradicting

$$\lim_{m\to\infty} g_m^{swap}(\phi_m^{\text{swap}}(c)) = c < 1.$$

So $L > 0$. We have shown that any convergent subsequence of x_m must tend to $L > 0$, the unique solution to $\int_0^1 \gamma(s, x)ds = c$. Suppose that $\lim_{m\to\infty} x_m \neq L$. Then for there exists $\epsilon > 0$ such that for all $m_0 \in \mathbb{N}$, there exists $m \geq m_0$ with $|x_m - L| > \epsilon$ and in particular we obtain a subsequence x_{k_m} of x_m satisfying $|x_{k_m} - L| > \epsilon$ for all m. Since x_m was shown to be bounded, so is x_{k_m}, implying by Bolzano Weierstrass that it in turn has a convergent subsequence $x_{k'_m}$. This gives us a contradiction as $|x_{k'(m)} - L| > \epsilon$ for all m and yet as it converges, so as a subsequence of x_m it converges to L. We conclude that $\lim_{m\to\infty} x_m = L$. The statement about h^{swap} is shown in Lemma B.5.

□

B.2 Missing Proofs from Section 3.3

Proposition 3.3 (★). *Let $\mathcal{X}$ and $\mathcal{Y}$ be properties such that $g_m^{\mathcal{X}}(\phi)$ and $g_m^{\mathcal{Y}}(\phi)$ are strictly monotonic and continuous.*[10] *If parameterizing by property $\mathcal{Y}$ asymptotically covers (cannot distinguish) property $\mathcal{X}$, then parameterizing by property $\mathcal{X}$ asymptotically covers (cannot distinguish) property $\mathcal{Y}$.*

Proof. Suppose that parameterizing by property $\mathcal{Y}$ asymptotically covers property $\mathcal{X}$. Then by definition $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(c)) = f(c)$ for $c \in [0, 1]$ where $f : [0, 1] \mapsto [0, 1]$ is strictly monotonic and bijective. So f^{-1} exists and we write $f^{-1}(d) = c$. Since $g_m^{\mathcal{X}}$ and $\phi_m^{\mathcal{Y}}$ are continuous, and bijectively map $[0, 1]$ to $[0, 1]$, so $(g_m^{\mathcal{X}}(\phi_m^{Y}))^{-1} = g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}})$ is continuous and bijecetively maps $[0, 1]$ to $[0, 1]$ too. So by continuity, for $\epsilon > 0$, there exists $\delta > 0$ such that

[10] These conditions hold for all properties that we consider.

if $|x - x'| < \delta$ then $|g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}})(\mathcal{X}) - g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}})(x')| < \epsilon$. But for any $\delta > 0$, there exists a large enough m such that $|g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(c)) - f(c)| < \delta$, so by continuity of $g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}$,

$$|g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}}(g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(c))) - g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}}((f(c)))| \tag{28}$$

$$= |c - g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}})(f(c))| \tag{29}$$

$$= |f^{-1}(d) - g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}})(d)| < \epsilon. \tag{30}$$

We conclude that if parameterizing by property $\mathcal{Y}$ asymptotically covers $\mathcal{X}$, then also parameterizing by property $\mathcal{X}$ asymptotically covers $\mathcal{Y}$.

Now suppose that it is not the case, that parameterizing by property $\mathcal{Y}$ asymptotically cannot distinguish property $\mathcal{X}$. Note that since $g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(c))$ is strictly monotonic and bounded, it converges to some $L \in [0, 1]$. Then by assumption it must hold that $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(c_1)) = L_1$ $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}}(c_2)) = L_2$ for some $c_1, c_2, L_1, L_2 \in [0, 1]$, $c_1 \neq c_2$ and $L_1 \neq L_2$. Then by continuity and bijectivity of $g_m^{\mathcal{X}}(\phi_m^{\mathcal{Y}})$, we can see (e.g. from an analogous argument as above) that $\lim_{m\to\infty} g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}}(L_1)) = c_1$ and $\lim_{m\to\infty} g_m^{\mathcal{Y}}(\phi_m^{\mathcal{X}}(L_2)) = c_2$. This implies that it is also does not hold that parameterizing by property $\mathcal{X}$ cannot asymptotically distinguish $\mathcal{Y}$. $\square$

The following Proposition B.6 is used in the proof of Theorem 3.4.

Proposition B.6. *Let $\epsilon \in (0, 1]$. Then $\lim_{m\to\infty} \phi_m^{\mathrm{swap}}(\epsilon) = 1$.*

Proof. Since by Fact A.1, $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m,v^*}}[\kappa(v, v^*)]$ is strictly increasing and continuous in ϕ, for all $m \in \mathbb{N}$, $g_m^{\mathrm{swap}}(\phi)$ is strictly increasing and continuous in ϕ and furthermore $g_m^{\mathrm{swap}}(0) = 0$ and $g_m^{\mathrm{swap}}(1)$. Since strict monotonicity is preserved under taking the inverse of a function, $\phi_m^{\mathrm{swap}} = (g_m^{\mathrm{swap}})^{-1}$ is strictly increasing on $[0, 1]$. As $\phi_m^{\mathrm{swap}}(\epsilon)$ is bounded in $[0, 1]$ for all $m \in \mathbb{N}$ and strictly increasing, it follows by the monotone convergence theorem that $\lim_{m\to\infty} \phi_m^{\mathrm{swap}}(\epsilon) = \phi \in [0, 1]$ exists. We must have $\phi = 1$, as for $\phi < 1$ by Corollary 2.3 the expected normalized swap distance is 0, contradicting $0 < \epsilon$. $\square$

Theorem 3.4 (★). *Let $\mathcal{X}$ be a property such that $g_m^{\mathcal{X}}(\phi)$ is strictly monotonic:*

1. *If the normalized Mallows model asymptotically covers property $\mathcal{X}$, the classic Mallows model asymptotically cannot distinguish $\mathcal{X}$.*
2. *If the classic Mallows model asymptotically covers property $\mathcal{X}$, then the normalized Mallows model asymptotically cannot distinguish $\mathcal{X}$.*

Proof. **Proof of (1):** Suppose $g_m^{\mathcal{X}}$ is strictly increasing (so f is too). By bijectivity of f, $f(0) = 0$, and so for any $\phi \in [0, 1)$,

$$f(0) \leq \lim_{m\to\infty} g_m^{\mathcal{X}}(\phi) \leq \lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathrm{swap}}(\epsilon)) = f(\epsilon)$$

where $\epsilon \in (0, 1]$. The inequality follows since $\lim_{m\to\infty} \phi_m^{\mathrm{swap}}(\epsilon) = 1$ by Proposition B.6 and because $g_m^{\mathcal{X}}$ is strictly increasing. We conclude that $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi) = f(0)$ by continuity of f. The case that $g_m^{\mathcal{X}}$ is strictly decreasing is analogous.

Proof of (2): Suppose $g_m^{\mathcal{X}}$ is strictly increasing (so f is too), then for any $c \in (0, 1]$,

$$f(1) \geq \lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathrm{swap}}(c)) \geq \lim_{m\to\infty} g_m^{\mathcal{X}}(1 - \epsilon) = f(1 - \epsilon)$$

for any $\epsilon \in (0, 1]$ again because $\lim_{m\to\infty} \phi_m^{\mathrm{swap}}(c) = 1$ by Proposition B.6. We conclude that $\lim_{m\to\infty} g_m^{\mathcal{X}}(\phi_m^{\mathrm{swap}}(c)) = f(1)$ by continuity of f. The case that $g_m^{\mathcal{X}}$ is strictly decreasing is analogous.

$\square$

B.3 Missing Proofs from Section 3.4

Proposition 3.6 (★). *The expected position of c_1 in a sampled ranking is* $\mathbb{E}_{v\sim\mathcal{M}_{\phi,m}}[\mathrm{pos}(v,c_1)] = \frac{1}{1-\phi} - m\frac{\phi^m}{1-\phi^m}$.

Proof. Let $\phi \in [0,1)$. Since Fact 2.1 says that the position of alternative c_1 is distributed according to a truncated geometric distribution with parameters m and $(1-\phi)$, we have that the expected value of alternative c_1's position is

$$\mathbb{E}_{v\sim\mathcal{M}_{\phi,m}}[\mathrm{pos}(v,c_1)] = \sum_{i=1}^{m} i \cdot \frac{\phi^{i-1}}{\sum_{i=1}^{m}\phi^{i-1}} = \frac{\sum_{i=1}^{m} i\cdot\phi^{i-1}}{\sum_{i=1}^{m}\phi^{i-1}} \tag{31}$$

$$= \frac{\frac{1-\phi^{m+1}}{(1-\phi)^2} - (m+1)\frac{\phi^m}{1-\phi}}{\frac{1-\phi^m}{1-\phi}} = \frac{1}{1-\phi} - m\frac{\phi^m}{1-\phi^m} \tag{32}$$

where Line 32 uses the geometric sum and its derivative. □

Theorem 3.7 (★). *The normalized Mallows model asymptotically covers the expected position of c_1, with $f(\ell) = t(h^{swap}(\ell))$, where $t(x) = 2\cdot(\frac{1}{x} - \frac{1}{e^x-1})$.*

Proof. From Lemma B.5 and Lemma B.3, we know that for fixed $c \in (0,1]$:

$$\lim_{m\to\infty}(1-\phi_m^{\text{swap}}(c)) \times m = h^{\text{swap}}(c) > 0 \tag{33}$$

$$\lim_{m\to\infty}\phi_m^{\text{swap}}(c)^m = e^{-h^{\text{swap}}(c)} < 1 \tag{34}$$

Using these we can conclude that:

$$\begin{aligned}
&\lim_{m\to\infty} g_m^{\text{pos1}}(\phi_m^{\text{swap}}(c))\\
=&\lim_{m\to\infty}\frac{\frac{1}{1-\phi_m^{\text{swap}}(c)} - m\frac{\phi_m^{\text{swap}}(c)^m}{1-\phi_m^{\text{swap}}(c)^m} - 1}{\frac{m-1}{2}+1}\\
=&\lim_{m\to\infty} 2\cdot(\frac{1}{m(1-\phi_m^{\text{swap}}(c))} - \frac{\phi_m^{\text{swap}}(c)^m}{1-\phi_m^{\text{swap}}(c)^m})\\
=&2\cdot(\frac{1}{\lim_{m\to\infty}m(1-\phi_m^{\text{swap}}(c))} - \frac{1}{-1+\lim_{m\to\infty}\phi_m^{\text{swap}}(c)^{-m}})\\
=&2\cdot(\frac{1}{h^{\text{swap}}(c)} - \frac{1}{e^{h^{\text{swap}}(c)}-1}) := f(c).
\end{aligned}$$

Properties of $f(c)$ It follows from Proposition B.1 that $t(x) = \frac{1}{x} - \frac{1}{e^x-1}$ is strictly decreasing and that $t(0) = \frac{1}{2}$. Furthermore, $\lim_{x\to\infty} t(x) = 0$, so that t maps $[0,\infty)$ bijectively to $[\frac{1}{2},0)$. By Lemma B.5, h^{swap} is strictly decreasing and bijectively maps $[0,1)$ to $[0,\infty)$. Since compositions of strictly decreasing functions are strictly decreasing $f(c) = 2\cdot t(h^{swap})(c)$ is strictly decreasing and furthermore $f(c)$ maps $[0,1)$ bijectively to $[0,1)$.

□

B.4 Missing Proofs from Section 3.5

Proposition 3.8 (★). *Consider $1 \le i,j \le m$ with $k := j-i+1$. The probability that c_i is ranked before c_j is*

$$\frac{1}{1-\phi^k}\left(1 - \frac{(1-\phi)(k-1)\phi^{k-1}}{1-\phi^{k-1}}\right).$$

Proof. Denote by $q_{i,j}$ the probability that alternative c_i is ranked before alternative c_j in v and denote by $p_{i,j}^m$ the probability that alternative c_i is ranked in position j in v when $v \sim \mathcal{M}_{\phi,m}$. Mallows [38] showed for fixed $\phi \in [0,1]$ that $q_{i,j}$ is independent of m and only depends on the relative difference $j - i$. So for $k = j - i + 1$, it holds that $q_{i,j} = q_{1,k}$. Using the Random Insertion sampling procedure for Mallows as discussed in Section 3.1, we can calculate $q_{1,k}$. We only need to consider the iteration during which alternative c_k is inserted and reason about which position alternative c_1 is in when this happens and the probability that c_k is inserted below alternative c_1. Before c_k is inserted, alternative c_1 is ranked in position i with probability $\frac{\phi^{i-1}}{\sum_{j=0}^{k-1}\phi^j}$ by Fact 2.1. The probability that c_k is inserted in position j is ϕ^{k-j}. If c_1 is ranked in position i, then c_k is inserted below position i with probability $\frac{\sum_{j=0}^{k-i-1}\phi^j}{\sum_{j=0}^{k-1}\phi^j}$. So

$$q_{1,k} = \sum_{i=1}^{k-1} p_{1,i}^{k-1} \frac{\sum_{j=0}^{k-i-1}\phi^j}{\sum_{j=0}^{k-1}\phi^j}$$

Evaluating this we obtain the desired closed form expression:

$$\begin{aligned}
q_{1,k} &= \sum_{i=1}^{k-1} p_{1,i}^{k-1} \frac{\sum_{j=0}^{k-i-1}\phi^j}{\sum_{j=0}^{k-1}\phi^j} = \sum_{i=1}^{k-1} \frac{(1-\phi)\phi^{i-1}}{1-\phi^{k-1}} \frac{1-\phi^{k-i}}{1-\phi^k} \\
&= \frac{(1-\phi)}{(1-\phi^k)(1-\phi^{k-1})} \sum_{i=1}^{k-1} \phi^{i-1}(1-\phi^{k-i}) \\
&= \frac{(1-\phi)}{(1-\phi^k)(1-\phi^{k-1})} \Big(\sum_{i=1}^{k-1} \phi^{i-1} - \sum_{i=1}^{k-1} \phi^{k-1}\Big) \\
&= \frac{(1-\phi)}{(1-\phi^k)(1-\phi^{k-1})} \Big(\frac{1-\phi^{k-1}}{1-\phi} - (k-1)\phi^{k-1}\Big) = \frac{1}{1-\phi^k}\Big(1 - \frac{(1-\phi)(k-1)\phi^{k-1}}{1-\phi^{k-1}}\Big).
\end{aligned}$$

□

Theorem 3.9 (★)**.** *The normalized Mallows model asymptotically covers the probability that alternative c_1 is ranked before alternative c_m in a sampled ranking, with $f(\ell) = t(h^{\text{swap}}(\ell))$, where $t(x) = 2 \cdot \frac{1}{1-e^{-x}}(1 - \frac{x}{e^x - 1}) - 1$.*

Proof. From Lemma B.5 and Lemma B.3, we know that for fixed $c \in (0,1]$:

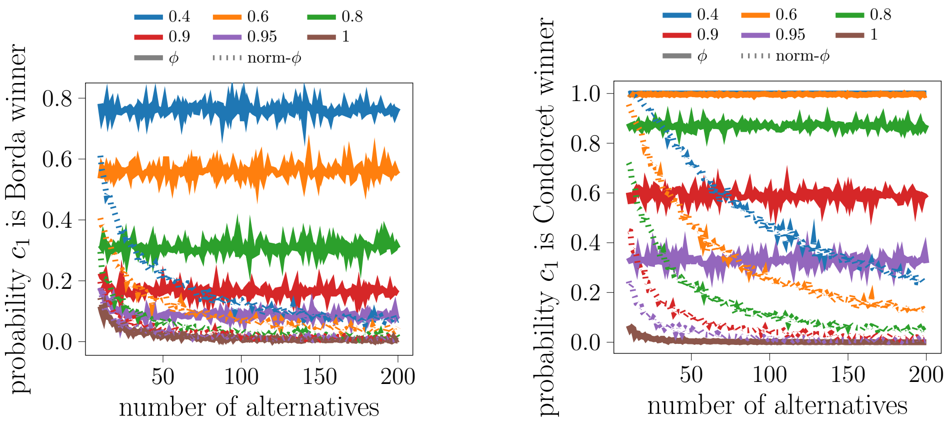
$$\lim_{m\to\infty} (1 - \phi_m^{\text{swap}}(c)) \times m = h^{\text{swap}}(c) > 0 \tag{35}$$

$$\lim_{m\to\infty} \phi_m^{\text{swap}}(c)^m = e^{-h^{\text{swap}}(c)} < 1 \tag{36}$$

Using these we can conclude that:

$$\begin{aligned}
&\lim_{m\to\infty} g_m^{(1 \text{ beats } m)}(\phi_m^{\text{swap}}(c)) \\
&= \lim_{m\to\infty} \frac{1}{1-\phi_m^{\text{swap}}(c)^m} - \frac{(1-\phi_m^{\text{swap}}(c))(m-1)\phi_m^{\text{swap}}(c)^{m-1}}{(1-\phi_m^{\text{swap}}(c)^m)(1-\phi_m^{\text{swap}}(c)^{m-1})} \\
&= \frac{1}{1-\lim_{m\to\infty}\phi_m^{\text{swap}}(c)^m} - \frac{\lim_{m\to\infty}(1-\phi_m^{\text{swap}}(c))(m-1) \cdot \lim_{m\to\infty}\phi_m^{\text{swap}}(c)^{m-1}}{\lim_{m\to\infty}(1-\phi_m^{\text{swap}}(c)^m) \cdot \lim_{m\to\infty}(1-\phi_m^{\text{swap}}(c)^{m-1})} \\
&= \frac{1}{1-e^{h^{\text{swap}}(c)}} - \frac{h^{\text{swap}}(c)e^{h^{\text{swap}}(c)}}{(1-e^{-h^{\text{swap}}(c)})^2} = \frac{1}{1-e^{-h^{\text{swap}}(c)}}\Big(1 - \frac{h^{\text{swap}}(c)}{e^{h^{\text{swap}}(c)}-1}\Big) := f(c),
\end{aligned}$$

where the first equality holds because of Proposition 3.8, the second because of the algebraic limit theorem, and the third because of Equations (35) and (36).



(a) Probability that c_1 is Borda winner.

(b) Probability that c_1 is Condorcet winner.

Figure 6: Influence of the number of alternatives m on the probability that c_1 is a Condorcet/Borda winner in ranking profiles sampled from the Mallows model for fixed values of the classical dispersion parameter ϕ (solid) and the normalized dispersion parameter norm-ϕ (dashed). Recall that ϕ = norm-$\phi = 0$ and ϕ = norm-$\phi = 1$, so the respective lines overlap. For each value of m, we sampled 1000 profiles and computed the average results.

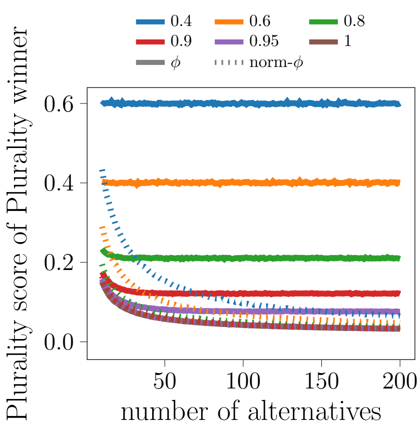
Properties of f. We show that the function $t(x) = 2 \cdot \frac{1}{1-e^{-x}}(1 - \frac{x}{e^x-1}) - 1$ is strictly increasing and maps $[0, \infty)$ bijectively to $[\frac{1}{2}, 1)$. The derivative of t is $\frac{d}{dx}t(x) = \frac{e^x(e^x(x-2)+x+2)}{(e^x-1)^3}$ which is positive for $x > 0$ since $(e^x(x-2) + x + 2)$ is strictly increasing and equals 0 if $x = 0$. So t is strictly increasing on $(0, \infty)$, with $\lim_{x\to\infty} t(x) = 2 \cdot (1 - 0) - 1 = 1$. To evaluate t at $x = 0$, note that t can be rewritten as $t(x) = 2 \cdot \frac{e^x(e^x-1-x)}{(e^x-1)^2} - 1$. Then using the Taylor expansion of e^x

$$\begin{aligned}
t(x) &= 2 \cdot \frac{e^x(\sum_{i=0} \frac{x^i}{i!} - x - 1)}{(\sum_{i=0} \frac{x^i}{i!} - 1)^2} - 1 \\
&= 2 \cdot \frac{e^x(\sum_{i=2} \frac{x^i}{i!})}{(\sum_{i=1} \frac{x^i}{i!})^2} - 1 = 2 \cdot \frac{e^x(\sum_{i=2} \frac{x^{i-2}}{i!})}{(\sum_{i=1} \frac{x^{i-1}}{i!})^2} - 1 \\
&= 2 \cdot \frac{1 \cdot \frac{1}{2}}{1^2} - 1 = 0 \text{ if } x = 0.
\end{aligned}$$

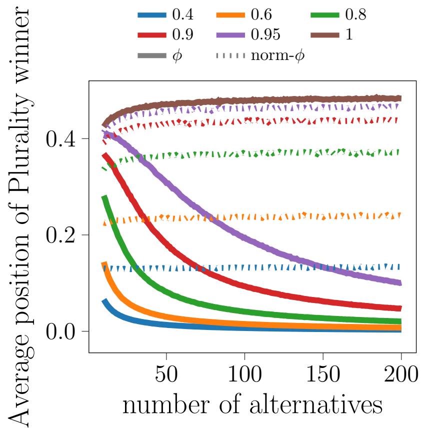
Since h^{swap} is strictly decreasing and maps $(0, 1]$ bijectively to $[0, \infty)$ by Lemma B.5 and t is strictly increasing, mapping $[0, \infty)$ bijectively to $[0, 1)$, we conclude that $f(c) = t(h^{swap(c)})$ is strictly decreasing, mapping $(0, 1]$ bijectively to $[0, 1)$ □

B.5 Additional Material for Section 3.6

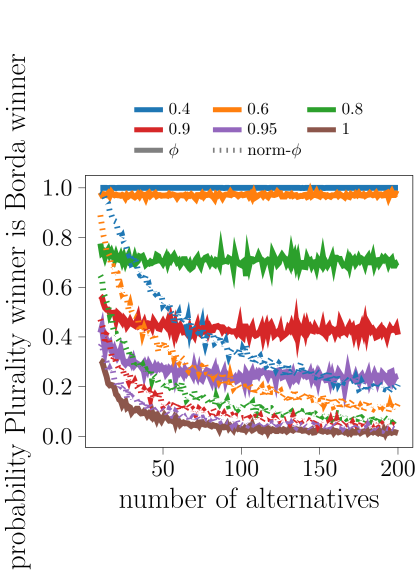
In the main body in Figure 4d, we have seen how the probability that c_1 is the Plurality winner develops when varying the number of alternatives. We have observed that for a fixed value of ϕ in the classical Mallows model this probability remains roughly constant, whereas for a fixed value of norm-ϕ in the normalized Mallows model it decreases when increasing m. In this section, we look at two further voting rules: Borda and Condorcet. Under the Borda voting rule, each ranking awards $m - i$ points to the alternative it ranks in ith place and the alternative with the highest number of points wins. Further, an alternative c is a Condorcet winner if for each alternative $c' \neq c$ a majority of rankings rank c before c'. We depict the dependency of the probability that c_1 is a Condorcet/Borda winner on the number of alternatives for fixed values of ϕ/norm-ϕ in Figure 6. It turns out that the situation for Borda and Condorcet is analogous to the picture for the Plurality voting rule: For a fixed value of ϕ the probability that c_1 is a winner is kept constant, whereas for a fixed value of norm-ϕ it decreases when increasing the number of alternatives.



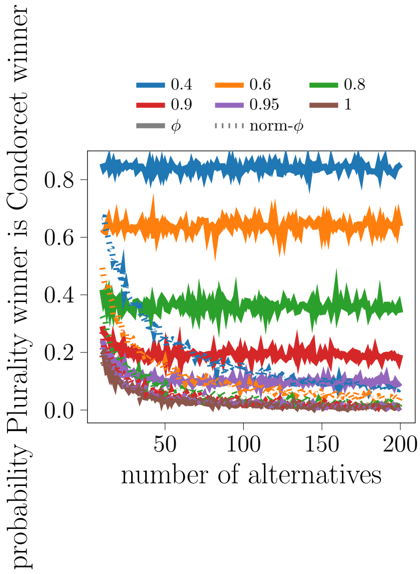
(a) Normalized plurality score of Plurality winner.



(b) Average Position of Plurality Winner.



(c) Probability that Plurality Winner is Borda Winner.



(d) Probability that Plurality Winner is Condorcet winner.

Figure 7: Properties of profiles with $n = 100$ rankings and a varying number of alternatives sampled from Mallows model with classical dispersion parameter ϕ (top) or normalized dispersion parameter norm-ϕ (bottom). For each value of m, we sampled 1000 profiles and computed the average results.

C Additional Material for Section 4

In Section 4.1, we have observed that in real-world profiles with varying numbers of alternatives, the positionwise distance to ID stays roughly constant as in the normalized Mallows model for a fixed value of norm-ϕ. In this section, we extend this analysis to some of the properties we have considered in Section 3. All of these properties deal with specific candidates from the central order, e.g., candidate c_1 who appears in the first position of the

	FBS systems	restricted FCS&FBS systems	FCS&FBS systems
Plurality score of Plurality winner ($\searrow_{\text{norm}}$)	0.59	0.63	0.49
Average position of Plurality winner ($\rightarrow_{\text{norm}}$)	0.012	0.01	0.011
Fraction of profiles where Borda and Plurality winner coincide ($\searrow_{\text{norm}}$)	0.85	0.85	0.6
Fraction of profiles where Plurality and Condorcet winner coincide ($\searrow_{\text{norm}}$)	0.91	0.91	0.72

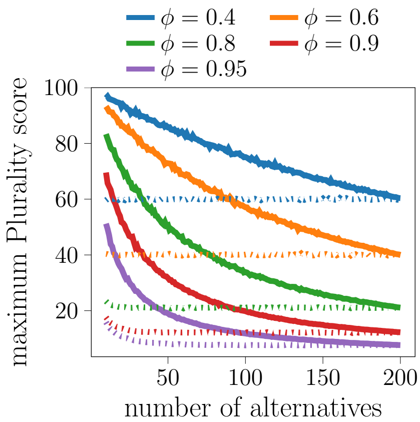
Table 1: Average values of properties in American football profiles.

central order and candidate c_m who appears in the last position. In real-world profiles, there is naturally no central order in the sense of the central order of the Mallows model. To still be able to compare the behavior of profiles sampled from the Mallows model and real-life profiles, we thus focus on properties only involving c_1 and use the Plurality winner as a proxy for c_1. This is a natural approach, as c_1 is the candidate which has the highest probability to be ranked first in a vote sampled from the Mallows model. To analyze the implications of replacing c_1 by the Plurality winner of some sampled profile, we rerun some of our experiments. In Figure 7, we present the behavior of the (normalized) Mallows model with a fixed (normalized) dispersion parameter concerning properties involving the Plurality winner when varying the number of alternatives. We sample profiles containing 100 rankings and analyze different properties of the Plurality winner analogous to our previous analysis of alternative c_1. It turns out that examining properties of the Plurality winner instead of alternative c_1 does not lead to a significant change in the results (see the similarities to Figures 4a, 4b, 6a and 6b). In particular, as before the classical Mallows model keeps the fraction of rankings in which the Plurality winner is ranked first roughly constant as well as the probability that the Plurality winner is the Borda/Condorcet winner. In contrast, the normalized Mallows model keeps the average position of the Plurality winner constant.

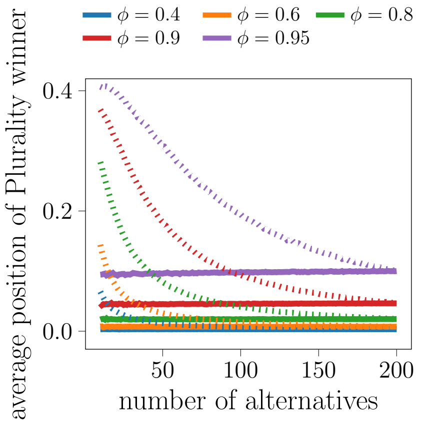
We now turn to analyzing how these properties behave in real-world profiles. As in contrast to the positionwise distance from ID, some of these properties are binary and for all of them we observe a high fragility, we want to average over the behavior of different profiles to get a clear picture. As for the Tour de France and Spotify profiles, there is no natural grouping of profiles that allow for some natural averaging, we do not look at them in this part. Instead, we focus on the American Football profiles for which there is a natural split into two (large) groups containing profiles over roughly the same number of alternatives: The profiles for outlets that rank only the FBS teams (in those profiles the number of candidates lies between 118 and 132) and profiles for outlets ranking FCS&FBS teams (in those profiles the number of candidates lies between 239 and 258). Averaging over the results within the two groups allows for a more robust estimate of the quantities. We depict the results in Table 1. As a sanity check we also added quantities for the FCS&FBS profiles restricted to the FBS teams (in order to check whether the observed trends are due to the changed number of alternatives or due to a fundamental difference how the two groups of outlets generate their rankings).

For all properties that are kept constant by the classical Mallows model, we see a significant decrease when moving from the American Football profiles over around 125 to the profiles over around 245 alternatives. In contrast, the average normalized position of the Plurality winner, which is kept constant by the normalized Mallows model, also stays constant when doubling the number of alternatives in the real-world profiles. These experiments underline that the behavior of the normalized Mallows model for a fixed normalized dispersion parameter is much more in line with what is present in real-world profiles, wheras the classical Mallow model behaves differently.

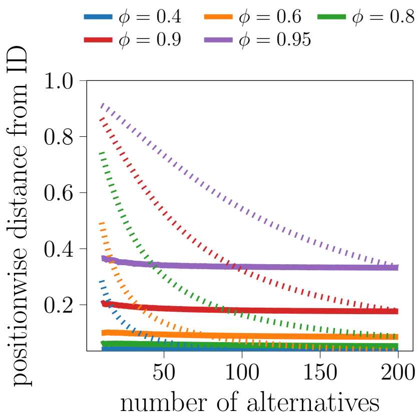
D Additional Material for Section 3.1



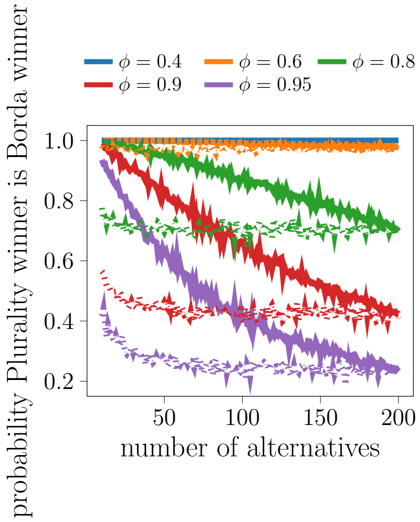
(a) Average Plurality Score of Plurality Winner.



(b) Average Position of Plurality Winner.



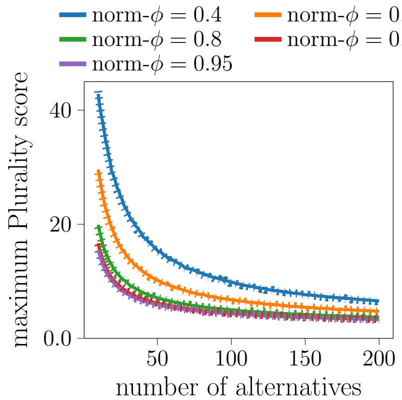
(c) Positionwise Distance from ID.



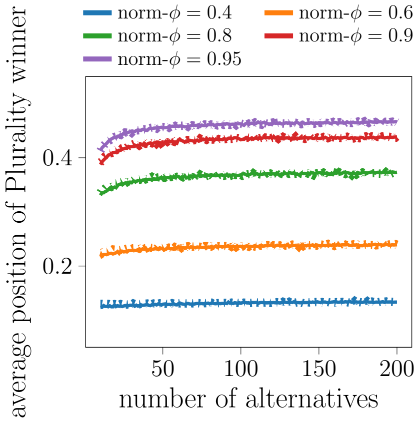
(d) Probability that Plurality Winner is Borda winner.

Figure 8: Analysis of how the properties of profiles containing 100 rankings depend on the number of alternatives for two different ways of sampling from the classical Mallows model. We compare sampling profiles with a varying number of m (dashed) with sampling profiles for $m = 200$ alternatives and subsequently deleting some alternatives uniformly at random (solid).

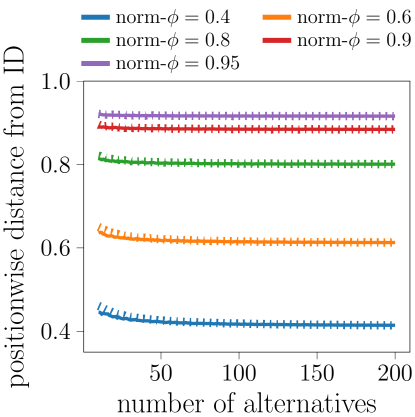
In Section 3.1, we have discussed the difference between sampling a profile for some number of m of alternatives directly for some value of ϕ/norm-ϕ or sampling a profile for some larger number of alternatives $i + m$ for the same value of ϕ/norm-ϕ and then deleting i alternatives uniformly at random. We have argued that for the normalized Mallows model (keeping norm-ϕ fixed) these two strategies result in profiles with very similar properties, whereas for the classical Mallows model (keeping ϕ fixed) the properties of the resulting



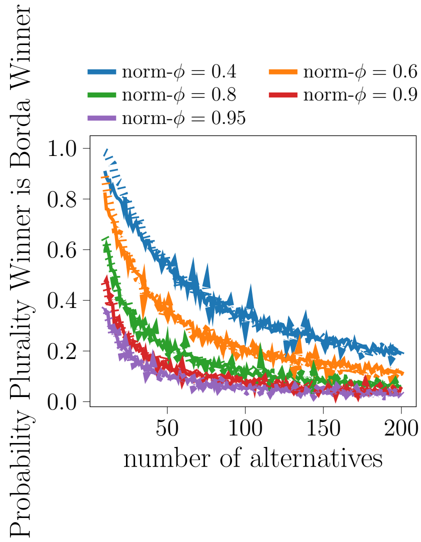
(a) Average Plurality Score of Plurality Winner.



(b) Average Position of Plurality Winner.



(c) Positionwise Distance from ID.



(d) Probability that Plurality Winner is Borda winner.

Figure 9: Analysis of how the properties of profiles containing 100 rankings depend on the number of alternatives for two different ways of sampling from the normalized Mallows model. We compare sampling profiles with a varying number of m (dashed) with sampling profiles for $m = 200$ alternatives and subsequently deleting some alternatives uniformly at random (solid).

profiles substantially differ. To support this claim we rerun the experiment described in Section 3.1, for other properties of the sampled profile. Specifically, for the average position of the Plurality winner, the positionwise distance from ID, and the probability that the Plurality winner is the Borda winner (note that as in Appendix C we focus on the role of the Plurality winner instead of c_1 as the profiles sampled via the deletion strategy are formally not sampled from some Mallows distribution). The results for the classical Mallows model can be found in Figure 8 and the results for the normalized Mallows model in Figure 9. These additional results are in line with what we have observed in Section 3.1 and confirm our intuition that for the normalized Mallows model deleting alternatives uniformly at random leads to ranking similar to those sampled for a smaller number of alternatives with the same value of norm-ϕ, whereas this is clearly not the case for the classical Mallows model.