On Reconstructing a Hidden Permutation

Flavio Chierichetti[1], Anirban Dasgupta[2], Ravi Kumar[3], and Silvio Lattanzi[4]

1 **Sapienza University, Rome, Italy**
`flavio@chierichetti.name`
2 **IIT Gandhinagar, Gandhinagar, India**
`anirbandg@iitgn.ac.in`
3 **Google, Mountain View, USA**
`ravi.k53@gmail.com`
4 **Google, New York, USA**
`silviol@google.com`

Abstract

The Mallows model is a classical model for generating noisy perturbations of a hidden permutation, where the magnitude of the perturbations is determined by a single parameter. In this work we consider the following reconstruction problem: given several perturbations of a hidden permutation that are generated according to the Mallows model, each with its own parameter, how to recover the hidden permutation? When the parameters are approximately known and satisfy certain conditions, we obtain a simple algorithm for reconstructing the hidden permutation; we also show that these conditions are nearly inevitable for reconstruction. We then provide an algorithm to estimate the parameters themselves. En route we obtain a precise characterization of the swapping probability in the Mallows model.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Mallows model, Rank aggregation, Reconstruction

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.604

1 Introduction

The Mallows model [16] is a classical exponential model for generating random perturbations of a fixed but hidden permutation. In this model, the perturbation noise is determined by a single parameter, which induces a distribution on the space of all permutations. The magnitude of the perturbation is measured by the Kendall tau distance, which is the number of pairwise disagreements between two permutations. When the parameter is large, the induced distribution is highly concentrated (in terms of the Kendall distance) around the hidden permutation whereas when the parameter is close to zero, the distribution is essentially uniform on all permutations. The model can be though of as a Gaussian-like distribution on permutations but with less nice properties. It easy to see that the permutation that maximizes the likelihood under the Mallows model is in fact the hidden permutation [25].

In a typical setting, the perturbations of an underlying latent permutation are modeled using a Mallows model and the goal is to reconstruct the hidden permutation using a few (independent) perturbed samples. For example, consider the problem of (inferring the hidden true) restaurant ranking in a neighborhood. If we assume that the user behavior corresponds to a Mallows model, then by using the individual restaurant rankings of a few users, one can hope to reconstruct the true ranking. Ever since its introduction more than half a century ago, the Mallows model has been extensively studied in diverse areas including statistics,

17th Int'l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'14) /
18th Int'l Workshop on Randomization and Computation (RANDOM'14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 604–617
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

machine learning, information retrieval, combinatorics, and social choice theory. Many of the reconstruction methods used in practice are often based on heuristics with no provable guarantees or on a careful but exhaustive search.

Even though the Mallows model is a simple and elegant way to model the perturbations of permutation, it many settings, it is oversimplified. In the above example, the classical setting assumes that each user has the same noise parameter. A more realistic setting is the following. Each user comes with his/her own noise parameter that determines how much they perturb the true ranking: a conformist might have a small noise parameter whereas a maverick might choose to vastly differ from the true ranking and hence might have a bigger noise parameter [24]. Thus, it becomes important to study the Mallows reconstruction problem where each sample perturbation is generated with a possibly *different* noise parameter.

In this paper we consider this reconstruction problem for permutations on n elements, where each permutation is generated by a Mallows model with its own parameter. We first show that perfect reconstruction is achievable (with high probability) in polynomial time if the parameters are (approximately) known and their Euclidean norm is $\Omega(\log n)$. In contrast, we show that such a reconstruction is information-theoretically impossible if this norm is smaller than a constant. En route, we obtain a precise characterization of the probability of swapping the order of two elements in a permutation generated by the Mallows model. We then complement the reconstruction algorithm, which requires the parameters or their approximations to be explicitly given, by providing an algorithm that estimates the parameters. By using these two algorithms together, we can show for instance that if there are at least $\Omega(\log n)$ parameters that are more than some constant, then reconstruction is possible even if the parameters are *not* explicitly given. We also consider the setting of approximate reconstruction and provide upper and lower bounds in terms of the parameters.

There has been some theoretical work on the Mallows reconstruction problem, especially by Braverman and Mossel [3], who considered reconstruction in the classical setting. Our work, however, is different from theirs in a few ways. First, we go beyond the classical case, i.e., we do not require all the parameters to be equal to each other. Second, we give approximate reconstruction bounds that can guarantee an arbitrarily small maximum displacement, while they can only guarantee a maximum displacement of at least $\Omega(\log n)$. Third, in the classical case their algorithm requires super-polynomial time if the parameter is $o(1/\sqrt{\log n})$, while ours runs in polynomial time for any choice of the parameter.

Combinatorial properties of the permutations generated in the Mallows model have been studied in the past. The partition function, mean, and variance of the model were computed by Diaconis and Ram [5] and Starr [22]. Tail bounds on the displacement of an element in a Mallows permutation was studied by Braverman and Mossel [3] and Gnedin and Olshanski [11]; these bounds were further tightened in a very recent work by Bhatnagar and Peled [2]. The latter also studied the length of the longest increasing subsequence in a Mallows permutation, improving upon the earlier work of Mueller and Starr [19]. Finding the maximum likelihood permutation is equivalent to the well-known rank aggregation problem. This is in general NP-hard [1, 7] and has a polynomial-time approximation scheme [13].

The Mallows model has also been generalized in a different way by generalizing the Kendall distance to weigh the number of inversions with respect to each element differently [9, 10, 8]; Meila et al. [18] studied the inference problem in this model. Qin et al. [21] defined a coset-permutation distance based model that generalizes the Mallows model to general distances and yet remains computationally efficient. A few other generalizations have also been studied in machine learning; see [15, 14]. Mukherjee [20] studied the consistency of likelihood estimators of the parameters.

2 Preliminaries

Let $[n] = \{1, \ldots, n\}$ be the universe of n elements and let S_n be the symmetric group on $[n]$. Permutations in S_n are denoted by Greek symbols. For a permutation σ and an element i, let $\pi(i)$ denote the *position* (or the *rank*) of the ith element. For two permutations π and σ, let $\kappa(\pi, \sigma)$ denote the *Kendall tau* distance (the number of inversions) between them.

Let $\beta \in (0, \infty)$ be a parameter and let $\sigma \in S_n$ be a fixed permutation. In the *Mallows* model $\mathcal{M}(\sigma, \beta)$ of generating permutations [16], the parameter β and the permutation σ induce a distribution on S_n as follows:

$$\Pr_{\mathcal{M}(\sigma,\beta)}[\pi] = \frac{e^{-\beta\cdot\kappa(\pi,\sigma)}}{Z_\beta},$$

where Z_β is the normalization constant defined as $Z_\beta = \prod_{j \le n} \frac{1-e^{-\beta j}}{1-e^{-\beta}}$. We use $\pi \sim \mathcal{M}(\sigma, \beta)$ to denote that π is generated according to $\mathcal{M}(\sigma, \beta)$. Clearly, as $\beta \to 0$, the distribution gets closer to uniform on S_n and as $\beta \to \infty$, the distribution becomes more concentrated around σ. In the classical Mallows reconstruction problem, the goal is to recover σ, given a set $\{\pi_i\}$ of independent samples where each $\pi_i \sim \mathcal{M}(\sigma, \beta)$; the algorithm may or may not know β and the goal is to use as few samples as possible.

In a generalization of the Mallows model, there are m parameters $\beta_1, \ldots, \beta_m$ where each $\beta_u \in (0, \infty)$ and a fixed permutation $\sigma \in S_n$. In the corresponding reconstruction problem, given independent samples $\pi_1, \ldots, \pi_m$ where each $\pi_u \sim \mathcal{M}(\sigma, \beta_u)$, the goal is to reconstruct σ. The algorithm may or may not know the β_u's. Note the two key differences from the classical setting: (i) each sample is generated by a *different* noise parameter and (ii) exactly *one* sample is produced for each parameter. If we assume σ to be the identity permutation, we denote the Mallows model simply by $\mathcal{M}(\beta)$ and the Kendall tau metric by $\kappa(\pi) = \kappa(\sigma, \pi)$.

Let $[p]$ denote 1 if the binary predicate p holds and 0 otherwise. We use the following form of tail inequality [12]:

▶ **Theorem 1** (Hoeffding's inequality). *If $X_1, \ldots, X_n$ are independent r.v.'s, with $\ell_i \le X_i \le u_i$, then*

$$\Pr\left[\sum_i X_i \le E[X] - \lambda\right] \le \exp\left(-\frac{2\lambda^2}{\sum_i (u_i - \ell_i)^2}\right).$$

3 Swapping Probability

In this section we precisely characterize the probability that two elements are out of order in the Mallows model. We express this probability in terms of the parameter β and the distance between the two elements. For the remainder of this section, without loss of generality, we assume that σ is the identity permutation.

Let $\pi \sim \mathcal{M}(\beta)$. For $1 \le i \le n - k$, let

$$s_{\beta,k,i} = \Pr_{\pi\sim\mathcal{M}(\beta)}[\pi(i) > \pi(i + k)],$$

i.e., the probability that the ordering of the elements i and $i + k$ is not preserved. The following result [2] shows that $s_{\beta,k,i}$ is independent of i.

Let $I = (i_1, \ldots, i_k)$ be an increasing sequence of indices. For a permutation π, let $\pi_I \in S_k$ denote the induced relative ordering of π when restricted to the indices in I. For an integer b, let $I + b$ denote $(i_1 + b, \ldots, i_k + b)$.

▶ **Lemma 2** (Translation invariance [2]). *Let $I = (i_1, i_2, \ldots, i_k)$ be an increasing sequence and $\pi \sim \mathcal{M}(\beta)$. Then for any integer $1 \le b \le n - i_k$, π_I and π_{I+b} have the same distribution, i.e., for any $\omega \in S_k$, $\Pr[\pi_I = \omega] = \Pr[\pi_{I+b} = \omega]$.*

Using this it is easy to see that $s_{\beta,k,i}$ is independent of i, henceforth we denote this probability as $s_{\beta,k}$. We now obtain an exact expression for it.

▶ **Lemma 3.**

$$s_{\beta,k} = \frac{ke^{\beta(k+1)} + 1 - (k+1)e^{\beta k}}{\left(e^{\beta(k+1)} - 1\right)\left(e^{\beta k} - 1\right)}.$$

Proof. In order to prove the above lemma, we will use a result from [5] that describes two different insertion processes to create a Mallows permutation. The following two processes define a series of permutations $\pi_1, \ldots, \pi_n$ such that $\pi_n = \pi \sim \mathcal{M}(\beta)$. For the purpose of this proof, we use the shorthand $q = e^{-\beta}$.

Insertion process $P1$. Consider the elements $1, \ldots, n$ in this order. For each i, define π_i to be a permutation over the elements 1 to i. Define $\pi_1(1) = 1$. Also define π_i in terms of π_{i-1} as follows. First sample $\pi_i(i)$ as following:

$$\Pr[\pi_i(i) = j] = \frac{(1/q)^{j-1}}{1 + 1/q + \cdots + (1/q)^{i-1}}, \text{ for } j \in \{1, \ldots, i\}. \tag{1}$$

Then, for s such that $\pi_{i-1}(s) < \pi_i(i)$, $\pi_i(s) = \pi_{i-1}(s)$ and else $\pi_i(s) = \pi_{i-1}(s) + 1$. Finally, $\pi = \pi_n$.

Insertion process $P2$. Here, we consider the elements in the order $n, n-1, \ldots, 1$. The permutation π'_i is defined as a permutation over elements $n, n-1, \ldots, i$ and is defined as follows. We start with $\pi'_n(n) = 1$. The random variable $\pi'_i(i)$ is defined as

$$\Pr[\pi'_i(i) = j] = \frac{q^{j-1}}{1 + q + \cdots + q^{n-i-1}}, \text{ for } j \in \{1, \ldots, n-i\}. \tag{2}$$

Thus, for s such that $\pi'_{i+1}(s) < \pi'_i(i)$, we have $\pi'_i(s) = \pi'_{i+1}(s)$ and otherwise $\pi'_i(s) = \pi'_{i+1}(s) + 1$. Finally, $\pi = \pi'_1$.

Now, we try to compute the probability that $\pi(1) > \pi(k+1)$. Consider that the permutation π is being formed by the process $P1$.

$$\begin{aligned}\Pr[\pi(1) > \pi(k+1)] &= \Pr[\pi_k(1) > \pi_{k+1}(k+1)] \\ &= \sum_{j=1}^{k} \Pr[\pi_{k+1}(k+1) < j \mid \pi_k(1) = j] \Pr[\pi_k(1) = j] \\ &= \sum_{j=1}^{k} \frac{(1/q)^j - 1}{(1/q)^{k+1} - 1} \Pr[\pi_k(1) = j].\end{aligned}$$

Now, π_k is a permutation on $\{1, \ldots, k\}$ that is again distributed according to Mallows model with parameter β. If we use process $P2$ to generate it, the element 1 is inserted last, and hence the probability of $\pi_k(1) = j$ can be written as

$$\Pr[\pi_k(1) = j] = \frac{q^{j-1}}{1 + q + \cdots + q^{k-1}} = \frac{(1-q)q^{j-1}}{1 - q^k}.$$

Hence, we have

$$\begin{aligned}\Pr[\pi(1) > \pi(k+1)] &= \sum_{j=1}^{k} \frac{(1/q)^j - 1}{(1/q)^{k+1} - 1} \frac{(1-q)q^{j-1}}{1-q^k} \\ &= \frac{1-q}{q((1/q)^{k+1}-1)(1-q^k)} \sum_{j=1}^{k} (1-q^j) \\ &= \frac{(1-q)q^k}{(1-q^{k+1})(1-q^k)} \left(k - \frac{q-q^{k+1}}{1-q} \right) \\ &= \frac{q^k(q^{k+1} - q - kq + k)}{(1-q^{k+1})(1-q^k)}.\end{aligned}$$

Substituting $q = e^{-\beta}$, the proof is complete. ◀

Next, we obtain a simpler approximation of the swapping probability.

▶ **Lemma 4.** *For each $0 < \beta \leq \beta'$ and $1 \leq k' \leq k$ such that $\beta k = \beta' k'$, we have $s_{\beta,k} \geq s_{\beta',k'}$. Moreover, if $\tau = \beta k$ for $\beta > 0, k \geq 1$, then we have*

$$\frac{1}{e^\tau + 1} \quad \leq \quad s_{\beta,k} \quad < \quad \frac{\tau + e^{-\tau} - 1}{e^\tau + e^{-\tau} - 2},$$

where the lower bound occurs at $k = 1$ and the upper bound is attained in the limit as k increases.

Proof. We consider the function $f_\tau(\beta) = s_{\beta,\frac{\tau}{\beta}}$. We start by showing that its derivative with respect to β is negative in $(0, \tau]$. The derivative can be written as:

$$f'_\tau(\beta) = \frac{-\tau e^{\tau+2\beta} + (\tau + \beta\tau + \beta^2)e^{\tau+\beta} + (\tau - \beta\tau - \beta^2)e^\beta - \tau}{\beta^2 \left(1 - e^{-\tau}\right) \left(e^{\tau+\beta} - 1\right)^2}. \tag{3}$$

Since the denominator in (3) is a product of positive factors, we only need to focus on the numerator in (3), which can be rewritten as $\beta(\tau+\beta)\left(e^\tau - 1\right) e^\beta - \tau \left(e^{\tau+\beta} - 1\right) \left(e^\beta - 1\right) = X_\beta(\tau) - Y_\beta(\tau)$, where $X_\beta(\tau) = \beta(\tau+\beta)\left(e^\tau - 1\right) e^\beta$ and $Y_\beta(\tau) = \tau \left(e^{\tau+\beta} - 1\right) \left(e^\beta - 1\right)$. We will show that $X_\beta(\cdot)$ is pointwise smaller than $Y_\beta(\cdot)$ in $(0, \tau)$, thus proving that $f'_\tau(\beta)$ is negative in this range.

To prove $X_\beta(\tau) < Y_\beta(\tau)$, we express the two functions as power series in the variable τ and show that for each term in the series, the corresponding coefficients obey the inequality. We have

$$X_\beta(\tau) = \beta^2 e^\beta \tau + e^\beta \sum_{n=2}^{\infty} \frac{\beta + \beta^2/n}{(n-1)!} \tau^n \quad \text{and} \quad Y_\beta(\tau) = \left(e^\beta - 1\right)^2 \tau + e^\beta \sum_{n=2}^{\infty} \frac{e^\beta - 1}{(n-1)!} \tau^n.$$

Indeed, the ratio of coefficients corresponding to τ satisfy

$$\frac{\beta^2 e^\beta}{\left(e^\beta - 1\right)^2} = \frac{\beta^2}{e^\beta - 2 + e^{-\beta}} = \frac{\beta^2}{\sum_{i=1}^{\infty} \frac{2\beta^{2i}}{(2i)!}} = \frac{\beta^2}{\beta^2 + 2\sum_{i=2}^{\infty} \frac{\beta^{2i}}{(2i)!}} < 1,$$

and the ratio of coefficients corresponding to $\tau^n, n \geq 2$, satisfy

$$\frac{e^\beta \frac{\beta + \beta^2/n}{(n-1)!}}{e^\beta \frac{e^\beta - 1}{(n-1)!}} = \frac{\beta + \frac{\beta^2}{n}}{e^\beta - 1} = \frac{\beta + \frac{\beta^2}{n}}{\sum_{i=1}^{\infty} \frac{\beta^i}{i!}} = \frac{\beta + \frac{\beta^2}{n}}{\beta + \frac{\beta^2}{2} + \sum_{i=3}^{\infty} \frac{\beta^i}{i!}} < 1.$$

Thus, we conclude that $f_\tau(\beta)$ is decreasing in $0 < \beta \leq \tau$. The minimum is attained at $k = 1$:

$$s_{\tau,1} = \frac{e^{2\tau} + 1 - 2e^{\tau}}{(e^{2\tau} - 1)(e^{\tau} - 1)} = \frac{(e^{\tau} - 1)^2}{(e^{\tau} + 1)(e^{\tau} - 1)^2} = \frac{1}{e^{\tau} + 1}.$$

Likewise, the upper bound is achieved by the limiting value at $\beta \to 0^+$:

$$\lim_{\beta \to 0^+} s_{\beta, \frac{\tau}{\beta}} = \frac{\tau + e^{-\tau} - 1}{e^{\tau} + e^{-\tau} - 2}.$$

◀

Using this, we obtain simpler bounds on $s_{\beta,k}$ that will be useful.

▶ **Corollary 5.** *Let $\beta k = \tau$. Then,*

$$s_{\beta,k} \leq \begin{cases} 1/2 - \Theta(\tau) & \text{if } \tau = o(1), \\ 1/2 - c & \text{if } \tau = \Theta(1), \\ \tau / e^{\tau} & \text{if } \tau = \omega(1), \end{cases}$$

where $c = c(\tau)$ is a positive constant.

An interesting consequence of the bounds on $s_{\beta,k}$ is that if β is moderately large, then the hidden permutation can be guessed reasonably well. The following result was also obtained in [2, Proposition 1.9]; we give a proof only for completeness.

▶ **Corollary 6.** *If $\beta = \ln n + \ln \frac{1}{\epsilon}$, then $\Pr_{\mathcal{M}(\sigma,\beta)}[\sigma] \geq 1 - \epsilon$.*

Proof. For this value of β, any two adjacent elements in σ will swap with probability at most $e^{-\beta} = \epsilon/n$. By a union bound on all the $n - 1$ adjacent pairs, we get that the probability of no swaps is at least $1 - \epsilon$. ◀

4 Reconstruction when Parameters are Given

In this section we present an algorithm to reconstruct the hidden permutation σ, assuming we know an approximation to the noise parameters $\beta_1, \ldots, \beta_m$; let the corresponding approximations be $\hat{\beta}_1, \ldots, \hat{\beta}_m$. Let α be the approximation factor, i.e., the smallest number such that

$$\frac{\hat{\beta}_u}{\alpha} \leq \beta_u \leq \alpha \hat{\beta}_u, \text{ for all } u = 1, \ldots, m.$$

The quality of the reconstructed permutation will depend on α and the magnitude of $\beta_1, \ldots, \beta_m$; the latter should be hardly surprising since the closer is β to 0, the lesser $\mathcal{M}(\sigma, \beta)$ has information about σ (as $\beta \to 0$, $\mathcal{M}(\sigma, \beta)$ converges to the uniform distribution on S_n).

The basic step considers two elements $i \neq j$ with the promise that $|\sigma(i) - \sigma(j)| \geq k$ and aims to determine if i should be ranked above j or vice versa. Our algorithm decides this bit according to the following rule:

$$i\text{'s position} < j\text{'s position} \iff \left(\sum_{u=1}^{m} (-1)^{[\pi_u(i) > \pi_u(j)]} \cdot \min(\hat{\beta}_u, 1/k) \right) > 0. \quad (4)$$

▶ **Lemma 7.** *Let $k \geq 1$ be an integer and assume that for a large enough constant $c_1 > 0$, $\sum_{u=1}^{m} \min\left(k^2 \beta_u^2, 1\right) \geq c_1 \alpha^2 \ln 1/\delta$. If i and j are such that $|\sigma(i) - \sigma(j)| \geq k$, then the ordering of i and j determined by (4) is consistent with σ, with probability at least $1 - \delta$.*

Proof. Without loss of generality, let $\sigma(i) < \sigma(j)$. Define $X_u = (-1)^{[\pi_u(i) > \pi_u(j)]}$. We have

$$E[X_u] = \Pr_{\pi_u \sim \mathcal{M}(\sigma,\beta)}[\pi_u(i) > \pi_u(j)] - \Pr_{\pi_u \sim \mathcal{M}(\sigma,\beta)}[\pi_u(i) < \pi_u(j)].$$

Let $M_u = \min\left(\beta_u, \frac{1}{k}\right)$, $\hat{M}_u = \min\left(\hat{\beta}_u, \frac{1}{k}\right)$. By Corollary 5, we have $E[X_u] \geq c_0 k M_u$, where c_0 is a sufficiently small constant. Let $Y_u = \hat{M}_u X_u$ and $Y = \sum_{u=1}^m Y_u$. Note that $-\hat{M}_u \leq Y_u \leq \hat{M}_u$. Now,

$$E[Y] \geq c_0 k \sum_{u=1}^{m} \hat{M}_u M_u = \Delta.$$

If $Y > 0$, then (4) correctly identifies the ordering of i and j. We bound the probability of the incorrect event to be at most δ using Theorem 1:

$$\Pr[Y \leq 0] \leq \Pr[Y \leq E[Y] - \Delta] \leq \exp\left(-\frac{\Delta^2}{2\sum_{u=1}^m \hat{M}_u^2}\right) = \exp\left(-\frac{c_0^2 k^2 \left(\sum_{u=1}^m \hat{M}_u M_u\right)^2}{2\sum_{u=1}^m \hat{M}_u^2}\right). \tag{5}$$

We now apply a converse of the Cauchy–Schwarz inequality due to Cassel [23]: if two sequences $a = (a_1, \ldots, a_m)$, $b = (b_1, \ldots, b_m)$ of real numbers satisfy $c \leq \frac{a_u}{b_u} \leq C$ for each $u = 1, \ldots, m$, then $\langle a, b\rangle^2 \geq (c/C)||a||_2^2||b||_2^2$.

Setting $a_u = \hat{M}_u$, $b_u = M_u$, $c = \frac{1}{\alpha}$, and $C = \alpha$ and applying Cassel's inequality in (5),

$$\begin{aligned}\Pr[X \leq 0] &\leq \exp\left(-\frac{c_0^2 k^2 \alpha^{-2} \sum_{u=1}^m \hat{M}_u^2 \cdot \sum_{u=1}^m M_u^2}{2\sum_{u=1}^m \hat{M}_u^2}\right) = \exp\left(-\frac{1}{2} c_0^2 k^2 \alpha^{-2} \sum_{u=1}^{m} M_u^2\right) \\ &\leq \exp\left(-\frac{1}{2} c_0^2 k^2 \alpha^{-2} \cdot c_1 \alpha^2 k^{-2} \ln \frac{1}{\delta}\right) \leq \delta,\end{aligned}$$

as long as $c_1 \geq 2/c_0^2$. ◀

From Lemma 7, we can obtain the precise condition that guarantees the exact reconstruction of σ.

▶ **Theorem 8** (Exact reconstruction). *If $\sum_{u=1}^m \min\left(\beta_u^2, 1\right) \geq c\alpha^2 \ln n$ for some fixed constant c, then with probability at least $1 - n^{-\Theta(1)}$ we can reconstruct σ in polynomial time.*

Proof. We apply Lemma 7 with $k = 1$. Our condition on the β_u's guarantees that, with probability $1 - n^{-\Theta(1)}$, rule (4) correctly identifies the ordering of each pair of elements. Therefore we can use any sorting algorithm to produce σ. ◀

Let $\vec{\beta} = \langle \beta_1, \ldots, \beta_m\rangle$. We next show that for exact reconstruction, the above requirement on $\|\vec{\beta}\|^2$ is close to optimal, off only by a factor of $\log n$.

▶ **Theorem 9.** *Let $n = 2$, and let $c > 0$ be a small enough constant. If $\max \beta_u \leq c$ and $\|\vec{\beta}\|^2 \leq c$, then with probability $\Omega(1)$ we cannot reconstruct σ.*

Proof. Let $S_2 = \{\sigma, \sigma^R\}$ and let σ be the unknown permutation chosen uniformly at random in S_2. By Corollary 5, for any $u \in [m]$,

$$\Pr_{\mathcal{M}(\sigma,\beta_u)}[\sigma^R] = \frac{1}{2} - \epsilon_u,$$

with $\epsilon_u = \Theta(\beta_u)$. If $b_u = (-1)^{[\pi_u \neq \sigma]}$, then $E[b_u] = 2\epsilon_u$. The likelihood of σ given $\pi_1, \ldots, \pi_m$ is

$$X = \sum_{u=1}^{m} \ln \frac{\frac{1}{2} + b_u \cdot \epsilon_u}{\frac{1}{2} - b_u \cdot \epsilon_u} = 4 \sum_{u=1}^{m} \left((1 + O(\epsilon_u^2)) b_u \cdot \epsilon_u\right). \tag{6}$$

It is easy to see that $E[X] = \Theta(||\vec{\beta}||^2)$ and $\mathrm{Var}[X] = \Theta(||\vec{\beta}||^2)$. Since the terms of the sum in (6) are independent, and $||\vec{\beta}||^2 \leq c$ for a small enough constant $c > 0$, the probability that the likelihood of σ will be negative is at least some constant. Therefore, any algorithm will be incorrect with probability at least $\Omega(1)$. ◀

We now make another observation on reconstruction using Corollary 6.

▶ **Corollary 10.** *There exists a constant $c > 0$ such that if $||\vec{\beta}|| \geq c\alpha^2 \ln n$, then with probability $1 - n^{-\Theta(1)}$ we can reconstruct σ in polynomial time.*

Proof. If there exists one $\hat{\beta}_u$ larger than $c\alpha \ln n$, for some large enough $c > 0$, then by Corollary 6, $\pi_u = \sigma$ with high probability. Otherwise, all the $\hat{\beta}_u$'s will be smaller than $c\alpha \ln n$ and hence all the β_u's will be smaller than $c\alpha^2 \ln n$. This implies that $\sum_{u=1}^{m} \min\left(\beta_u^2, 1\right) \geq \sum_{u=1}^{m} \frac{\beta_u^2}{c\alpha^2 \ln n}$. Sincet $||\vec{\beta}||^2 \geq c^2\alpha^4 \ln^2 n$, we obtain

$$\sum_{u=1}^{m} \min\left(\beta_u^2, 1\right) \geq \frac{||\vec{\beta}||^2}{c\alpha^2 \ln n} = c\alpha^2 \ln n.$$

By applying Theorem 8, σ can be obtained with probability $1 - n^{-\Theta(1)}$ in polynomial time. ◀

5 Estimating the Parameters

In this section we deal with the problem of estimating the parameters $\beta_1, \ldots, \beta_m$. Again, without loss of generality, we assume the unknown permutation σ is the identity permutation.

Recall that for each β_u, we only have one sample permutation $\pi_u \sim \mathcal{M}(\beta_u)$. Our aim is to estimate the β_u values by looking only at the set $\{\pi_1, \ldots, \pi_m\}$. Before presenting our algorithm, we first state a result that bounds the deviation of each element from its position in the hidden permutation.

▶ **Theorem 11** ([2]). *For all $\beta > 0$,*

$$\Pr_{\pi \sim \mathcal{M}(\beta)} [|\pi(i) - i| > t] \leq 2e^{-t\beta},$$

and

$$c \cdot \min\left(\frac{e^{-\beta}}{1 - e^{-\beta}}, n - 1\right) \leq E[|\pi(i) - i|] \leq \min\left(\frac{2e^{-\beta}}{1 - e^{-\beta}}, n - 1\right),$$

for some absolute constant $c > 0$.

The expected Kendall tau distance of π can also be calculated exactly.

▶ **Corollary 12** ([4, 2]). *If $\pi \sim \mathcal{M}(\beta)$, then*

$$E[\kappa(\pi)] = \frac{ne^{-\beta}}{1 - e^{-\beta}} - \sum_{j=1}^{n} \frac{je^{-\beta j}}{1 - e^{-\beta j}}.$$

Furthermore, if $\beta > 0$*, then for some constant* $c > 0$*,*

$$c \cdot \min\left(\frac{ne^{-\beta}}{1-e^{-\beta}}, n(n-1)\right) \leq E[\kappa(\pi)] \leq \min\left(\frac{ne^{-\beta}}{1-e^{-\beta}}, n(n-1)\right);$$

if $\beta = \Theta(1)$ *and* $n = \Omega(1/\beta)$*, then* $c = 1 - o(1)$*.*

Our estimate $\hat{\beta}_u$ for the parameter β_u is obtained by simply looking at the pairwise distances $\kappa(\pi_u, \pi_v)$, and then using the minimum of those to estimate $\hat{\beta}_u$. Formally, $\hat{\beta}_u$ is defined as following:

$$\hat{\beta}_u = \ln\left(\frac{\tilde{k}_u + 1}{\tilde{k}_u}\right), \text{ where } \tilde{k}_u = \min_{v \in [m]} \frac{\kappa(\pi_u, \pi_v)}{n}. \tag{7}$$

In order to show that (7) gives a reasonable estimate of the β_u parameters, we first need to show that if $\pi \sim \mathcal{M}(\beta)$ and $\pi' \sim \mathcal{M}(\beta')$ are two sample permutations from two different Mallows models, then the Kendall distance between π and π' is related to a function of β and β'. For this, we first relate $\kappa(\pi, \pi')$ to $\kappa(\pi) + \kappa(\pi)$.

Define

$$c_\beta = 1 - \frac{\beta + e^{-\beta} - 1}{e^{\beta} + e^{-\beta} - 2} > \frac{1}{2}.$$

From Lemma 4 for $k = 1$ and $\beta > 0$, we get the following.

▶ **Corollary 13.** *If* $i, j \in [n]$ *such that* $i > j$*, then* $\Pr_{\pi \sim \mathcal{M}(\beta)}[\pi(i) > \pi(j)] \geq c_\beta$*.*

The above corollary can then be used to show the following lower bound on the expectation of the Kendall distance between any two random permutations. Note that an upper bound on $\kappa(\pi, \pi')$ in terms of $\kappa(\pi)$ and $\kappa(\pi')$ is trivial by the triangle inequality.

▶ **Lemma 14.** *If* $\pi \sim \mathcal{M}(\beta)$ *and* $\pi' \sim \mathcal{M}(\beta')$*, then* $E[\kappa(\pi, \pi')] \geq c_{\beta'} E[\kappa(\pi)] + c_\beta E[\kappa(\pi')]$*. In particular, for all* $\beta, \beta' > 0$*,* $E[\kappa(\pi, \pi')] \geq (E[\kappa(\pi)] + E[\kappa(\pi')])/2$*.*

Proof. For two permutations τ and τ', define the inversion vector $\mathsf{inv}(\tau, \tau')$ as

$$\mathsf{inv}(\tau, \tau')_{\tau(i)} = \sum_{j : \tau(j) < \tau(i)} \mathbb{1}[\tau'(j) > \tau'(i)].$$

Define $x = \mathsf{inv}(\sigma, \pi)$, $x' = \mathsf{inv}(\sigma, \pi')$, $w = \mathsf{inv}(\pi, \pi')$ and $z = \mathsf{inv}(\pi', \pi)$. By definition,

$$w_{\pi(i)} = \sum_{j : \pi(j) < \pi(i)} \mathbb{1}[\pi'(j) > \pi'(i)].$$

Then, $\kappa(\pi, \pi') = \sum_i w_i = \sum_i z_i$, and similarly for the others. Since π and π' are independent,

$$E[w_{\pi(i)}] = \sum_j E\left[\mathbb{1}[\pi(j) < \pi(i)]\, \mathbb{1}[\pi'(j) > \pi'(i)]\right] = \sum_j \Pr[\pi(j) < \pi(i)] \Pr[\pi'(j) > \pi'(i)],$$

and therefore,

$$\begin{aligned} E\left[\sum_i w_{\pi(i)}\right] &= \sum_i \sum_{j<i} \Pr[\pi(j) < \pi(i)] \Pr[\pi'(j) > \pi'(i)] \\ &\quad + \sum_i \sum_{j>i} \Pr[\pi(j) < \pi(i)] \Pr[\pi'(j) > \pi'(i)]. \end{aligned}$$

Now, using Corollary 13, we have that for $j < i$, $\Pr[\pi(j) < \pi(i)] \geq c_\beta$. Using the same argument, for $j > i$, $\Pr[\pi'(j) > \pi'(i)] \geq c_{\beta'}$. Hence,

$$E\left[\sum_i w_{\pi(i)}\right] \geq c_\beta \sum_i \sum_{j<i} \Pr[\pi'(j) > \pi'(i)] + c_{\beta'} \sum_i \sum_{j>i} \Pr[\pi(j) < \pi(i)].$$

The proof is completed by just noting that $\sum_i \sum_{j<i} \Pr[\pi'(j) > \pi'(i)] = E[\sum_i x'_i] = E[\kappa(\pi')]$ and $\sum_i \sum_{j>i} \Pr[\pi(j) < \pi(i)] = E[\sum_i x_i] = E[\kappa(\pi)]$. ◀

Thus, $E[\kappa(\pi, \pi')]$ is both upper and lower bounded by $E[\kappa(\pi)] + E[\kappa(\pi')]$ to within constant factors. We next show that $\kappa(\pi, \pi')$ is concentrated around its expectation. We will use the following concentration theorem (proved in [17] and expressed in this form in [6]).

▶ **Theorem 15** ([17]). *Let f be a function of n random variables $X_1, \ldots, X_n$, each X_i taking values in a set A_i, such that $E[f]$ is bounded. Assume that*

$$m \leq f(X_1, \ldots, X_n) \leq M.$$

Let $\mathcal{B}$ be any event and let c_i be maximum effect of f assuming $\mathcal{B}$, i.e.,

$$|E[f \mid \mathbf{X}_{i-1}, X_i = a_i, \mathcal{B}] - E[f \mid \mathbf{X}_{i-1}, X_i = a'_i, \mathcal{B}]| \leq c_i.$$

Then

$$\Pr[f > E[f] + t] \leq \exp\left(-\frac{2t^2}{\sum_i c_i^2}\right) + \Pr[\mathcal{B}^c],$$

and

$$\Pr[f < E[f] - t] \leq \exp\left(-\frac{t^2}{\sum_i c_i^2}\right) + \Pr[\mathcal{B}^c].$$

In order to apply the above tail bound to show that $\kappa(\pi, \pi')$ is concentrated, we will first need a result showing that each element does not move too much from its original position with high probability. Define $\Delta(\beta) = \frac{1}{\beta} \ln(5n^4)$. The following is easily obtained from Theorem 11.

▶ **Lemma 16.** *If $\pi \sim \mathcal{M}(\beta), \pi' \sim \mathcal{M}(\beta')$, and $\Delta' = \Delta(\beta') + \Delta(\beta)$, then*

$$\Pr[\ \forall i \ |\pi(i) - i| \leq \Delta' \text{ and } |\pi'(i) - i| \leq \Delta'] \geq 1 - n^{-4}.$$

Proof. By applying Theorem 11 and then taking a union bound over all positions. ◀

We next show that $\kappa(\pi, \pi')$ does not deviate from its expectation with high probability.

▶ **Lemma 17.** *If $\pi \sim \mathcal{M}(\beta'), \pi' \sim \mathcal{M}(\beta)$, and $\Delta' = \Delta(\beta') + \Delta(\beta)$, then*

$$\Pr[|\kappa(\pi, \pi') - E[\kappa(\pi, \pi')]| > 2\Delta' \sqrt{n \log n}] \leq 4n^{-4}.$$

Proof. We use the tail inequality in Theorem 15 to bound $\kappa(\pi, \pi')$. Let X_{2i} denote the random variable that contains the position $\pi(i)$ and let X_{2i+1} contain $\pi'(i)$. Let $f(X_1, \ldots, X_{2n}) = \kappa(\pi, \pi')$.

Let $\mathcal{B}$ be the event: "$\forall i, |\pi(i) - i| \leq \Delta'$ and $|\pi'(i) - i| \leq \Delta'$". Using Lemma 16, $\Pr[\mathcal{B}^c] \leq n^{-4}$. Let $\vec{x}, \vec{x}'$ denote a vector of size $n - i - 1$ and $\vec{b}$ denote a vector of size $i - 1$. Define $f_{\vec{b},c}(x_{i+1}, \ldots, x_n) = f(\vec{b}, X_i = c, x_{i+1}, \ldots, x_n)$. Since only the ith element causes different transpositions in the two cases, we have

$$|f_{\vec{b},c}(x_{i+1}, \ldots, x_n) - f_{\vec{b},c'}(x_{i+1}, \ldots, x_n)| \leq |c - c'|.$$

Using the insertion process $P1$ (Lemma 3), the probability of $X_{i+1}, \ldots, X_n$ assuming any set of values remains the same, irrespective of the exact values realized by the random variables $X_1, \ldots, X_i$. That is,

$$\begin{aligned}&\Pr[(X_{i+1}, \ldots, X_n) = \vec{x} \mid (X_1, \ldots, X_{i-1}) = \vec{b}, X_i = c] \\ &\quad = \Pr[(X_{i+1}, \ldots, X_n) = \vec{x} \mid (X_1, \ldots, X_{i-1}) = \vec{b}', X_i = c'].\end{aligned}$$

Combining these two facts, we have that

$$|E[f_{\vec{X}_{i-1},c}] - E[f_{\vec{X}_{i-1},c'}]| \le |c - c'|,$$

where $\vec{X}_{i-1} = X_1, \ldots, X_{i-1}$. Conditioned on the event $\mathcal{B}^c$, we then have

$$|c - c'| \le 2\Delta'.$$

Furthermore $0 \le f \le n^2$. Hence using Theorem 15, we have that

$$\Pr[f > E[f] + t] \le \exp\left(-\frac{2t^2}{4n\Delta'^2}\right) + \frac{1}{n^4} \quad \text{and} \quad \Pr[f < E[f] - t] \le \exp\left(-\frac{t^2}{4n\Delta'^2}\right) + \frac{1}{n^4}.$$

By choosing $t = 4\Delta'\sqrt{n \log n}$, we have $\Pr[|f - E[f]| > t] \le 4n^{-4}$. ◀

Finally, we show that we can get a good estimate of β if n is large enough.

▶ **Lemma 18.** *Let $\beta_1 \ge \cdots \ge \beta_m > 0$ and let $c > 0$ be the constant in Corollary 12. If $n = \omega\left(\frac{e^{\beta_1}}{\beta_m^2 \ln(1/\beta_m)}\right)$, then for each $u > 1$, (7) returns an estimate $\hat{\beta}_u$ such that*

$$\beta_u - \ln 2 - o(1) \le \hat{\beta}_u \le \beta_u + \ln \frac{1}{c_{\beta_m} c} + o(1).$$

Proof. Note that (7) computes $\kappa(\pi_u, \pi_v)$ for each pair $u, v, u \ne v$. Applying Lemma 17 and taking a union bound over all pairs (u, v), with probability $1 - \frac{1}{n^2}$, the following event happens:

$$\forall u \ne v, |\kappa(\pi_u, \pi_v) - E[\kappa(\pi_u, \pi_v)]| \le \Delta', \tag{8}$$

where $\Delta' = 2\max_{u \in [m]} \Delta(\beta_u)$.

Since $\tilde{k}_{uv} = \kappa(\pi_u, \pi_v)$, using Lemma 14 for the lower bound and the triangle inequality for the upper bound, we have

$$c_{\beta_v} E[\kappa(\pi_u)] + c_{\beta_u} E[\kappa(\pi_v)] \le E[\tilde{k}_{uv}] \le E[\kappa(\pi_u)] + E[\kappa(\pi_v)]. \tag{9}$$

Since c_β is an increasing function of β for all $u > 1$, (9) implies

$$c_{\beta_m} E[\kappa(\pi_u)] \le \min_v E[\tilde{k}_{uv}] \le E[\kappa(\pi_u)] + E[\kappa(\pi_1)]. \tag{10}$$

Plugging in the values of the expectations from Corollary 12, (10) implies

$$nc\frac{c_{\beta_m} e^{-\beta_u}}{1 - e^{-\beta_u}} \le \min_v E[\tilde{k}_{uv}] \le n\left(\frac{e^{-\beta_u}}{1 - e^{-\beta_u}} + \frac{e^{-\beta_1}}{1 - e^{-\beta_1}}\right).$$

Hence for all $u > 1$,

$$nc\frac{c_{\beta_m} e^{-\beta_u}}{1 - e^{-\beta_u}} \le \min_v E[\tilde{k}_{uv}] \le 2n\frac{e^{-\beta_u}}{1 - e^{-\beta_u}}.$$

Now, we condition on the event described in (8). For $u > 1$, we have that

$$nc \cdot c_{\beta_m} \frac{e^{-\beta_u}}{1 - e^{-\beta_u}} - 2\Delta' \le \min_v \tilde{k}_{uv} \le 2n \frac{e^{-\beta_u}}{1 - e^{-\beta_u}} + 2\Delta'.$$

Hence, $\tilde{k}_u = \frac{\min_v \tilde{k}_{uv}}{n} \in \left[c_{\beta_m} c \frac{e^{-\beta_u}}{1-e^{-\beta_u}} - \frac{2\Delta'}{n}, \frac{2e^{-\beta_u}}{1-e^{-\beta_u}} + \frac{2\Delta'}{n}\right]$. Under the assumption that $n = \omega\left(\frac{4\Delta'}{c_{\beta_m} c} e^{\beta_1}\right)$, we have that $\tilde{k}_u \in \left[(1 - o(1)) c_{\beta_m} c \frac{e^{-\beta_u}}{1-e^{-\beta_u}}, \frac{2(1+o(1))e^{-\beta_u}}{(1-e^{-\beta_u})}\right]$. Since $\hat{\beta}_u = \ln\left(\frac{\tilde{k}_u + 1}{\tilde{k}_u}\right)$, the upper and lower bounds on $\hat{\beta}_u$ in the statement follows. The constraints on n boil down to saying that $n = \omega\left(\frac{\log n}{\beta_m c_{\beta_m}} e^{\beta_1}\right)$. Simplifying, $n = \omega\left(\frac{e^{\beta_1}}{\beta_m^2 \ln(1/\beta_m)}\right)$ is sufficient. ◀

An easy corollary is the following: a multiplicative reconstruction of the β_u's is possible for the β_u that are $\Theta(1)$ and there is at least one (unknown) permutation generated with a parameter that is large, and hence is close to the identity.

▶ **Corollary 19.** *If β_1 is such that $\beta_1 = \omega(\beta_u)$ for some $u > 1$, then*

$$(1 + o(1)) c_{\beta_m} E[\kappa(\pi_u)] \le \min_v E[\tilde{k}_{uv}] \le (1 + o(1)) E[\kappa(\pi_u)],$$

and hence for each $u > 1$, (7) returns an estimate $\hat{\beta}_u$ such that

$$\beta_u - o(1) \le \hat{\beta}_u \le \beta_u + \ln \frac{1}{c_{\beta_m} c} + o(1).$$

In particular, if $\beta_u = \Theta(1)$, then $\beta_1 = \omega(1)$ and the constants $c = 1 - o(1)$ and $c_{\beta_m} = 1 - o(1)$.

6 Approximate Reconstruction

Next, we show a result on approximate reconstruction of σ. We first show that if the sum of squares of β_ℓ is $\Omega(\ln n)$, i.e., the average is $\Omega(\frac{\ln n}{n})$, then we can learn an estimate $\hat{\sigma}$ of σ where the displacement of each element is bounded. We then show a simple lower bound that says that is $\sum_\ell \beta_\ell^2$ is really small, then we cannot recover anything meaningful.

▶ **Theorem 20** (Approximate reconstruction). *Let $k^\star = \arg\min_k \sum_{\ell=1}^m \min\left(k^2 \beta_\ell^2, 1\right) \ge c\alpha^2 \ln n$ for some fixed constant c and let $k^\star$ be known to the algorithm. Then with probability at least $1 - n^{-\Theta(1)}$ we can construct a permutation $\hat{\sigma}$ such that $|\hat{\sigma}(i) - \sigma(i)| \le 2k^\star$ for all $i \in [n]$.*

Proof. Using (4) for every pair of elements, with probability at least $1 - n^{-\Theta(1)}$, we determine the rank of each element to within an additive error of $k^\star$, i.e., for each element i, Lemma 7 guarantees that all elements j such that $|\sigma(i) - \sigma(j)| \ge k^\star$ will be correctly compared to i. We now need to find out a feasible permutation $\hat{\sigma}$ out of this set of comparisons such that the maximum displacement in $\hat{\sigma}$ is bounded.

Define the *score* of element i to be the number of other elements such that the right hand side of (4) holds. We define the permutation $\hat{\sigma}$ as the permutation that results from sorting the elements according to this score (ascending). We show that the displacement of every element is bounded by $2k^\star$. Consider any element i. By Lemma 7, the score of element i is at least $\max(1, i - k^\star)$ and at most $\min(i + k^\star, n)$. Therefore, $\hat{\sigma}(i) \in [\max(1, i - 2k^\star), \min(i + 2k^\star, n)]$. ◀

We now show a simple lower bound for approximate reconstruction.

▶ **Theorem 21.** *Let $\epsilon > 0$ be a small enough constant and let $\sqrt{\sum_{\ell=1}^{m} \beta_\ell} \leq \epsilon/n$. If σ is chosen uniformly at random in S_n, then any $\hat{\sigma}$ that is output by any algorithm satisfies $E[\kappa(\hat{\sigma}, \sigma)] \geq \left(\frac{1}{4} - \epsilon\right) n^2$.*

Proof. Consider the probability of the generic sequence of independent samples $\pi_1, \ldots, \pi_m$:

$$\Pr[\pi_1, \ldots, \pi_m \mid \sigma] \;=\; \prod_{\ell=1}^{m} \frac{e^{-\beta_\ell \kappa(\pi_\ell, \sigma)}}{Z_{\beta_\ell}} \;=\; e^{-\sum_{\ell=1}^{m} \beta_\ell \kappa(\pi_\ell, \sigma)} \cdot \prod_{\ell=1}^{m} Z_{\beta_\ell}^{-1}. \tag{11}$$

Since for each ℓ, $0 \leq \kappa(\pi_\ell, \sigma) \leq \binom{n}{2}$, we have

$$0 \;\leq\; \sum_{\ell=1}^{m} \beta_\ell \kappa(\pi_\ell, \sigma) \;\leq\; \binom{n}{2} \sum_{\ell=1}^{m} \beta_\ell \;\leq\; \epsilon. \tag{12}$$

It follows that for each sequence of samples $\pi_1, \ldots, \pi_m$, using (11) and (12), we have

$$e^{-\epsilon} \prod_{\ell=1}^{m} Z_{\beta_\ell}^{-1} \;\leq\; \Pr[\pi_1, \ldots, \pi_m | \sigma] \;\leq\; \prod_{\ell=1}^{m} Z_{\beta_\ell}^{-1}.$$

Thus, the probabilities of obtaining a set of m permutations are all within $e^{-\epsilon}$ factor of each other. For a set S of input permutations, let $S \sim U^m$ mean that each permutation is chosen uniformly at random, let $S \sim \mathcal{M}^m$ mean that the permutations are chosen according to the Mallows model with the parameters as in the Lemma statement, and let $\hat{\sigma}(S)$ be the solution returned by the algorithm on input S. We have

$$E[\kappa(\hat{\sigma}(S), \sigma) \mid S \sim U^m] \geq \frac{1}{2}\binom{n}{2},$$

as otherwise we can work with σ^R instead. Since under the given assumptions, the probability of obtaining each set S is within $e^{-\epsilon}$ of the uniform distribution,

$$|E[\kappa(\hat{\sigma}(S), \sigma) \mid S \sim \mathcal{M}^m] - E[\kappa(\hat{\sigma}(S), \sigma) \mid S \sim U^m]| \leq (1 - e^{-\epsilon}) E[\kappa(\hat{\sigma}(S), \sigma) \mid S \sim U^m].$$

Hence, $E[\kappa(\hat{\sigma}(S), \sigma) | S \sim \mathcal{M}^m] \geq (\frac{1}{4} - \epsilon) n^2$. ◀

▶ **Corollary 22.** *If $\hat{\sigma}$ is the output of an algorithm, then $\kappa(\hat{\sigma}, \sigma) = \Omega(n/\sqrt{\sum_{\ell=1}^{m} \beta_\ell})$.*

To interpret these lower bounds, we consider a concrete special case. Suppose $m = \omega(\log n)$ and $\beta_1 = \cdots = \beta_m = \beta$. Then, Theorem 20 guarantees a maximum element displacement of $O(\sqrt{(\log n)/(\beta^2 m)})$, which means that the total Kendall distance is $O(n\sqrt{(\log n)/(\beta^2 m)})$. On the other hand, for this setting, Theorem 21 obtains a Kendall distance lower bound of $\Omega(n\sqrt{1/(\beta m)})$. Thus, the gap between the upper bound and the lower bound is $O(\sqrt{(\log n)/\beta})$.

Acknowledgments. We thank the anonymous reviewers for their many valuable suggestions, especially towards a simpler proof of Lemma 3.

References

1. J. Bartholdi, C. A. Tovey, and M. A. Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(2):157–165, 1989.
2. N. Bhatnagar and R. Peled. Lengths of monotone subsequences in a Mallows permutation. *Probability Theory and Related Fields*, To appear.

3 M. Braverman and E. Mossel. Sorting from noisy information. *CoRR, abs/0910.1191*, 2009.

4 W. Cheng and E. Hüllermeier. Instance-based label ranking using the Mallows model. In *ECCBR Workshops*, pages 143–157, 2008.

5 P. Diaconis and A. Ram. Analysis of systematic scan Metropolis algorithms using Iwahori–Hecke algebra techniques. *The Michigan Mathematical Journal*, 48(1):157–190, 2000.

6 D. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomised Algorithms*. Cambridge University Press, 2009.

7 C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the web. In *WWW*, pages 613–622, 2001.

8 M. A. Fligner and J. S. Verducci. Distance based ranking models. *Journal of the Royal Statistical Society B*, 48:359–369, 1986.

9 M. A. Fligner and J. S. Verducci. Multistage ranking models. *Journal of the American Statistical Association*, 43(403):892–901, 1988.

10 M. A. Fligner and J. S. Verducci. Posterior probability for a consensus ordering. *Psychometrika*, 55:53–63, 1990.

11 A. Gnedin and G. Olshanski. The two-sided infinite extension of the Mallows model for random permutations. *Advances in Applied Mathematics*, 48(5):615–639, 2012.

12 W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

13 C. Kenyon-Mathieu and W. Schudy. How to rank with few errors. In *STOC*, pages 95–103, 2007.

14 A. Klementiev, D. Roth, and K. Small. Unsupervised rank aggregation with distance-based models. In *ICML*, pages 472–479, 2008.

15 G. Lebanon and J. Lafferty. Cranking: Combining rankings using conditional probability models on permutations. In *ICML*, pages 363–370, 2002.

16 C. L. Mallows. Non-null ranking models I. *Biometrika*, 44(1-2):114–130, 1957.

17 Colin McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, 1998.

18 M. Meila, K. Phadnis, A. Patterson, and J. A. Bilmes. Consensus ranking under the exponential model. In *UAI*, pages 285–294, 2007.

19 C. Mueller and S. Starr. The length of the longest increasing subsequence of a random Mallows permutation. *Journal of Theoretical Probability*, pages 1–27, 2011.

20 S. Mukherjee. Estimation of parameters in non uniform models on permutations. Technical Report 1307.0978, arXiv, 2013.

21 T. Qin, X. Geng, and T-Y. Liu. A new probabilistic model for rank aggregation. In *NIPS*, pages 1948–1956, 2010.

22 S. Starr. Thermodynamic limit for the Mallows model on S_n. Technical Report 0904.0696, arXiv, 2009.

23 G. S. Watson. Serial correlation in regression analysis. I. *Biometrika*, 42(3/4):327–341, 1955.

24 P. Yin, P. Luo, M. Wang, and W-C. Lee. A straw shows which way the wind blows: Ranking potentially popular items from early votes. In *WSDM*, pages 623–632, 2012.

25 H. P. Young. Optimal voting rules. *The Journal of Economic Perspectives*, 9(1):51–64, 1995.