Electronic Journal of Probability

Electron. J. Probab. **27** (2022), article no. 87, 1–20.
ISSN: 1083-6489 https://doi.org/10.1214/22-EJP812

Clustering of consecutive numbers in permutations under Mallows distributions and super-clustering under general p-shifted distributions

Ross G. Pinsky*

Abstract

Let $A_{l;k}^{(n)} \subset S_n$ denote the set of permutations of $[n]$ for which the set of l consecutive numbers $\{k, k+1, \cdots, k+l-1\}$ appears in a set of consecutive positions. Under the uniform probability measure P_n on S_n, one has $P_n(A_{l;k}^{(n)}) \sim \frac{l!}{n^{l-1}}$ as $n \to \infty$. In one part of this paper we consider the probability of clustering of consecutive numbers under Mallows distributions P_n^q, $q > 0$. Because of a duality, it suffices to consider $q \in (0,1)$. We show that for $q_n = 1 - \frac{c}{n^\alpha}$, with $c > 0$ and $\alpha \in (0,1)$, $P_n^q(A_{l;k_n}^{(n)})$ is of order $\frac{1}{n^{\alpha(l-1)}}$, uniformly over all sequences $\{k_n\}_{n=1}^\infty$. Thus, letting $N_l^{(n)} = \sum_{k=1}^{n-l+1} 1_{A_{l;k}^{(n)}}$ denote the number of sets of l consecutive numbers appearing in sets of consecutive positions, we have

$$\lim_{n\to\infty} E_n^{q_n} N_l^{(n)} = \begin{cases} \infty, \text{ if } l < \frac{1+\alpha}{\alpha}; \\ 0, \text{ if } l > \frac{1+\alpha}{\alpha}. \end{cases} \quad .$$

We also consider the cases $\alpha = 1$ and $\alpha > 1$. In the other part of the paper we consider general p-shifted distributions, $P_n^{\{p_j\}_{j=1}^\infty}$, of which the Mallows distribution is a particular case. We calculate explicitly the quantity

$$\lim_{l\to\infty} \liminf_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l;k_n}^{(n)}) = \lim_{l\to\infty} \limsup_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l;k_n}^{(n)})$$

in terms of the p-distribution. When this quantity is positive, we say that super-clustering occurs. In particular, super-clustering occurs for the Mallows distribution with fixed parameter $q \neq 1$.

Keywords: random permutation; Mallows distribution; clustering; runs; p-shifted; inversion; backward ranks.
MSC2020 subject classifications: 60C05; 05A05.
Submitted to EJP on December 20, 2020, final version accepted on June 17, 2022.

*Technion-Israel Institute of Technology, Haifa, Israel. E-mail: pinsky@math.technion.ac.il

1 Introduction and statement of results

Let $l \geq 2$ be an integer. Let P_n denote the uniform probability measure on the set S_n of permutations of $[n] := \{1, \cdots, n\}$, and denote a permutation $\sigma \in S_n$ by $\sigma = \sigma_1\sigma_2\cdots\sigma_n$. The set of l consecutive numbers $\{k, k+1, \cdots, k+l-1\} \subset [n]$ appears in a set of consecutive positions in the permutation if there exists an m such that $\{k, k+1, \cdots, k+l-1\} = \{\sigma_m, \sigma_{m+1}, \cdots, \sigma_{m+l-1}\}$. Let $A^{(n)}_{l;k} \subset S_n$ denote the event that the set of l consecutive numbers $\{k, k+1, \cdots, k+l-1\}$ appears in a set of consecutive positions. It is immediate that for any $1 \leq k, m \leq n-l+1$, the probability that $\{k, k+1, \cdots, k+l-1\} = \{\sigma_m, \sigma_{m+1}, \cdots, \sigma_{m+l-1}\}$ is equal to $\frac{l!(n-l)!}{n!}$. Thus,

$$P_n(A^{(n)}_{l;k}) = (n-l+1)\frac{l!(n-l)!}{n!} \sim \frac{l!}{n^{l-1}}, \text{ as } n \to \infty, \text{for } l \geq 2, \tag{1.1}$$

where $a_n \sim b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Let $A^{(n)}_l = \cup_{k=1}^{n-l+1} A^{(n)}_{l;k}$ denote the event that there exists a set of l consecutive numbers appearing in a set of consecutive positions, and let $N^{(n)}_l = \sum_{k=1}^{n-l+1} 1_{A^{(n)}_{l;k}}$ denote the number of sets of l consecutive numbers appearing in sets of consecutive positions. Then

$$E_n N^{(n)}_l = (n-l+1)^2 \frac{l!(n-l)!}{n!} \sim \frac{l!}{n^{l-2}}, \text{ as } n \to \infty, \text{ for } l \geq 2. \tag{1.2}$$

Using the inequality

$$\sum_{k=1}^{n-k+1} P_n(A^{(n)}_{l;k}) - \sum_{1\leq j<k\leq n-l+1} P_n(A^{(n)}_{l;j} \cap A^{(n)}_{l;k}) \leq P_n(A^{(n)}_l) \leq \sum_{k=1}^{n-k+1} P_n(A^{(n)}_{l;k}),$$

along with the fact that for j, k, m, r, with $\{j, j+1, \cdots, j+l-1\} \cap \{k, k+1, \cdots, k+l-1\} = \emptyset$ and $\{m, m+1, \cdots, m+l-1\} \cap \{r, r+1, \cdots, r+l-1\} = \emptyset$, the probability that both $\{k, k+1, \cdots, k+l-1\} = \{\sigma_m, \sigma_{m+1}, \cdots, \sigma_{m+l-1}\}$ and $\{j, j+1, \cdots, j+l-1\} = \{\sigma_r, \sigma_{r+1}, \cdots, \sigma_{r+l-1}\}$ is equal to, $\frac{(l!)^2(n-2l)!}{n!}$, it is easy to show that

$$P_n(A^{(n)}_l) \sim \frac{l!}{n^{l-2}}, \text{ as } n \to \infty, \text{ for } l \geq 3. \tag{1.3}$$

It follows from (1.2) (or from (1.3)) that for $l \geq 3$, the sequence $\{N^{(n)}_l\}_{n=1}^\infty$ converges to zero in probability. On the other hand, when $l = 2$, $\{N^{(n)}_l\}_{n=1}^\infty$ converges in distribution to a Poisson random variable with parameter 2. This result on the clustering of consecutive numbers in permutations goes back over 75 years; see [10], [5].

In one of the two parts of this paper, we obtain results in the spirit of (1.1) and (1.2) in the case that the uniform probability measure P_n is replaced by the Mallows measure $P^{q_n}_n$ with parameter q_n, where $q_n \to 1$ at various rates. The Mallows measures P^q_n are described below. The Mallows measure with $q = 1$ is the uniform measure.

In the other part of this paper we consider so-called p-shifted distributions $P^{\{p_j\}_{j=1}^\infty}_n$ on S_n, of which the Mallows measure P^q_n is a particular example. Here $\{p_j\}_{j=1}^\infty$, with $p_j > 0$, for all j, is a probability distribution on $\mathbb{N}$: $\sum_{j=1}^\infty p_j = 1$. We calculate

$$\lim_{l\to\infty} \liminf_{n\to\infty} P^{\{p_j\}_{j=1}^\infty}_n(A^{(n)}_{l;k_n}) = \lim_{l\to\infty} \limsup_{n\to\infty} P^{\{p_j\}_{j=1}^\infty}_n(A^{(n)}_{l;k_n})$$

explicitly. This reveals a necessary and sufficient condition on the distribution $\{p_j\}_{j=1}^\infty$ in order for the above limit to be positive. When this limit is positive, we say that *super-clustering* occurs. In particular, super-clustering occurs for the Mallows distribution for fixed $q \neq 1$.

We turn now to a description of the results in this paper, beginning with those concerning specifically the Mallows distributions.

Mallows distributions and the behavior of the probability of $A^{(n)}_{l;k_n}$. For each $q>0$, the Mallows distribution with parameter q is the probability measure P^q_n on S_n defined by

$$P^q_n(\sigma)=\frac{q^{\mathcal{I}_n(\sigma)}}{Z_n(q)},\sigma\in S_n, \tag{1.4}$$

where

$$\mathcal{I}_n(\sigma)=\sum_{1\le i<j\le n}1_{\sigma_j<\sigma_i}=\sum_{1\le i<j\le n}1_{\sigma^{-1}_j<\sigma^{-1}_i} \tag{1.5}$$

is the inversion statistic on S_n, that is, the number of inversions in σ, and $Z_n(q)$ is the normalization constant, given by [6, 9]

$$Z_n(q)=\prod_{k=2}^n\frac{1-q^k}{1-q}.$$

Thus, for $q\in(0,1)$, the distribution favors permutations with few inversions, while for $q>1$, the distribution favors permutations with many inversions. Of course, the case $q=1$ yields the uniform distribution. Recall that the *reverse* of a permutation $\sigma=\sigma_1\cdots\sigma_n$ is the permutation $\sigma^{\text{rev}}:=\sigma_n\cdots\sigma_1$. The Mallows distributions satisfy the following duality between $q>1$ and $q\in(0,1)$:

$$P^q_n(\sigma)=P^{\frac{1}{q}}_n(\sigma^{\text{rev}}),\ \text{for } q>0,\sigma\in S_n \text{ and } n=1,2,\cdots.$$

Since the set $A^{(n)}_{l;k}$ is invariant under reversal, for our study of clustering it suffices to consider the case that $q\in(0,1)$.

When $q\to0$, the Mallows distribution P^q_n converges weakly to the degenerate distribution on the identity permutation, and of course the identity permutation belongs to $A^{(n)}_{l;k}$ for all k and l. Because the smaller q is, the more the distribution favors permutations with few inversions, and as such, the smaller q is, the more the distribution favors permutations which are close to the identity permutation, it seems intuitive that the smaller q is, the more clustering there will be. However, whereas the structure of the Mallows distribution lends itself naturally to proving theorems concerning the inversion statistic [7], it is less transparent how to exploit that structure with regard to this clustering statistic. For example, the set $A^{(n)}_{l;k}$ is the disjoint union of the $n-l+1$ sets $\{k,k+1,\cdots,k+l-1\}=\{\sigma_m,\sigma_{m+1},\cdots,\sigma_{m+l-1}\},\ m=1,\cdots,n-l+1$. In the case of the uniform distribution, these $n-l+1$ sets all have the same probability. However, in the case of P^q_n, $q\in(0,1)$, we expect that for certain m, these sets will have probability less than what they have under the uniform distribution, and for other m these sets will have probability greater than what they have under the uniform distribution. (For results concerning the behavior under a Mallows distribution of other permutation statistics, such as cycle counts and increasing subsequences, see [1], [2] and [3].)

Our first theorem gives asymptotic results in the case that $q=q_n=1-\frac{c}{n^\alpha}$ with $c>0$ and $\alpha\in(0,1)$. We use the notation $a_n\lesssim b_n$ as $n\to\infty$ to indicate that $\limsup_{n\to\infty}\frac{a_n}{b_n}\le1$.

Theorem 1.1. *Let $A^{(n)}_{l;k}\subset S_n$ denote the event that the set of l consecutive numbers $\{k,k+1,\cdots,k+l-1\}$ appears in a set of l consecutive positions. Let $q_n=1-\frac{c}{n^\alpha}$, with $c>0$ and $\alpha\in(0,1)$. Then*

$$\frac{\big((l-1)!\big)^2}{(2l)!}\,\frac{c^{\,l-1}\,l!}{n^{\alpha(l-1)}}\lesssim P^{q_n}_n(A^{(n)}_{l;k_n})\lesssim\frac{1}{l}\,\frac{c^{\,l-1}\,l!}{n^{\alpha(l-1)}}, \tag{1.6}$$

for any choice of $\{k_n\}_{n=1}^\infty$*, and the asymptotic behavior is uniform over all* $\{k_n\}_{n=1}^\infty$*. If* k_n *satisfies* $\frac{\min(k_n,n-k_n)}{n^\alpha}=\infty$*, then the upper bound in* (1.6) *can be improved:*

$$P_n^{q_n}(A_{l;k_n}^{(n)})\lesssim\Big(\int_0^1 x^{l-1}e^{-(l-1)x}dx\Big)\frac{c^{l-1}\,l!}{n^{\alpha(l-1)}}. \tag{1.7}$$

Recall that $N_l^{(n)}=\sum_{k=1}^{n-l+1}1_{A_{l;k}^{(n)}}$ denotes the number of sets of l consecutive numbers appearing in sets of consecutive positions. Theorem 1.1 yields immediately the following corollary.

Corollary 1.2. *Let* $q_n=1-\frac{c}{n^\alpha}$ *with* $c>0$ *and* $\alpha\in(0,1)$*, Then there exist constants* $C_l^{(-)},C_l^{(+)}>0$ *such that*

$$C_l^{(-)}n^{1-(l-1)\alpha}\le E_n^{q_n}N_l^{(n)}\le C_l^{(+)}n^{1-(l-1)\alpha}.$$

In particular,

$$\lim_{n\to\infty}E_n^{q_n}N_l^{(n)}=\begin{cases}\infty, & \text{if } l<\frac{1+\alpha}{\alpha};\\ 0, & \text{if } l>\frac{1+\alpha}{\alpha}.\end{cases}\quad .$$

Remark 1.3. For $\tau\in S_l$, let $A_{l,\tau;k}^{(n)}\subset A_{l;k}^{(n)}$ denote the event that the set of l consecutive numbers $\{k,k+1,\cdots,k+l-1\}\subset[n]$ appears in a set of consecutive positions in the permutation and also that the relative positions of these consecutive numbers correspond to the permutation τ. That is, $\{k,k+1,\cdots,k+l-1\}=\{\sigma_m,\sigma_{m+1},\cdots,\sigma_{m+l-1}\}$, for some m, and $\sigma_{m+i-1}-(k-1)=\tau_i,\ i=1,\cdots,l$. Then $A_{l;k}^{(n)}=\cup_{\tau\in S_l}A_{l,\tau;k}^{(n)}$. Small changes in the proof of Theorem 1.1, which we leave to the reader, show that (1.6) and (1.7) hold with $P_n^{q_n}(A_{l;k_n}^{(n)})$ replaced by $P_n^{q_n}(A_{l,\tau;k_n}^{(n)})$ and with $l!$ deleted from the numerator in the upper and lower bounds, for all $\tau\in S_l$. In particular, if $\tau=id$, then $A_{l,\tau;k_n}^{(n)}$ is the event that the numbers $\{k,\cdots,k+l-1\}$ form an *increasing run* in the permutation, and if τ satisfies $\tau^{\text{rev}}=id$, then $A_{l,\tau;k_n}^{(n)}$ is the event that the numbers $\{k,\cdots,k+l-1\}$ form a *decreasing run in the permutation.*

Remark 1.4. Let $K^{(-)}(l)=\frac{\big((l-1)!\big)^2}{(2l)!}$ and $K^{(+)}(l)=\int_0^1 x^{l-1}e^{-(l-1)x}dx$ respectively denote the coefficient of $\frac{c^{l-1}\,l!}{n^{\alpha(l-1)}}$ on the left hand side of (1.6) and in (1.7). We have $K^{(-)}(l)\sim\sqrt{\pi}\,l^{-\frac{3}{2}}4^{-l}$ as $l\to\infty$. One can show that

$$K^{(+)}(l)=\int_0^1 x^{l-1}e^{-(l-1)x}dx=\frac{(l-1)!}{(l-1)^l}\big(1-e^{-(l-1)}\sum_{i=0}^{l-1}\frac{(l-1)^i}{i!}\big).$$

Thus, $K^{(+)}(l)\lesssim\frac{(l-1)!}{(l-1)^l}\sim\sqrt{2\pi}\,e\,l^{-\frac{1}{2}}e^{-l}$, as $l\to\infty$. On the other hand, a rudimentary asymptotic analysis we performed on the interval $[\frac{l-1}{l}-l^{-\frac{1}{2}},1]$ yields $K^{(+)}(l)\gtrsim e^{\frac{1}{2}}\,l^{-\frac{1}{2}}e^{-l}$, as $l\to\infty$.

Now we consider the cases $q=q_n=1-\frac{c}{n}$ and $q=q_n=1-o(\frac{1}{n})$.

Theorem 1.5. *Let* $A_{l;k}^{(n)}\subset S_n$ *denote the event that the set of* l *consecutive numbers* $\{k,k+1,\cdots,k+l-1\}$ *appears in a set of* l *consecutive positions.*

i. Let $q_n=1-\frac{c}{n}$*, with* $c>0$*. Let* $k_n\sim dn$ *with* $d\in(0,1)$*. Then*

$$\begin{aligned}&\Big(\int_{e^{-c\max(d,1-d)}}^1\big(y(1-y)\big)^{l-1}dy\Big)\frac{c^{l-1}\,l!}{n^{(l-1)}}\lesssim P_n^{q_n}(A_{l;k_n}^{(n)})\lesssim\\&\frac{1}{(1-e^{-cd^*})^l}\Big(\int_{e^{-cd^*}}^1 y^{l-1}e^{(\log\frac{1-e^{-cd^*}}{1-e^{-c}})e^{cd^*}(l-1)y}dy\Big)\frac{c^{l-1}\,l!}{n^{(l-1)}},\end{aligned}\tag{1.8}$$

where d^* can be chosen to be either d or $1-d$.

ii. Let $q_n = 1 - o(\frac{1}{n}) < 1$. Then for any choice of $\{k_n\}_{n=1}^\infty$,

$$P_n^{q_n}(A_{l;k_n}^{(n)}) \lesssim \frac{l!}{n^{l-1}}. \tag{1.9}$$

Remark 1.6. In part (i), we certainly expect that the asymptotic behavior of $P_n^{q_n}(A_{l;k_n}^{(n)})$, when $k_n \sim dn$, is in fact independent of $d \in (0,1)$. We note that the expression $\frac{1}{(1-e^{-cd^*})^l}\big(\int_{e^{-cd^*}}^1 y^{l-1} e^{(\log \frac{1-e^{-cd^*}}{1-e^{-c}})e^{cd^*}(l-1)cy} dy\big) c^{l-1}$, which multiplies $\frac{l!}{n^{l-1}}$ on the right hand side of (1.8), converges to 1 when $c \to 0$, for all $d^* \in (0,1)$, thus matching up with (1.9). For each c and l, the expression on the right hand side of (1.8) can be shown to be larger for $d^* = d$ than for $d^* = 1-d$, if d is sufficiently close to 1. We didn't try to analyze this for general d, as it is quite complicated and in any case the bound is not precise, but rather an artifact of the method of proof.

Remark 1.7. In the case of the uniform distribution ($q = 1$), we have from (1.1) that $P_n^1(A_{l;k_n}^{(n)}) \sim \frac{l!}{n^{l-1}}$, for any choice of k_n. Since we expect $P_n^q(A_{l;k_n}^{(n)})$ to be decreasing in q, we certainly expect that the asymptotic inequality in (1.9) is an asymptotic equality.

We now turn to a description of the results of the other part of the paper, concerning p-shifted random permutations.

p-shifted distributions and super-clustering. Denote by S_∞ the set of permutations of $\mathbb{N}$, that is, the set of bijective functions from $\mathbb{N}$ to itself. We build random permutations in S_∞ and then project them down in a natural way to S_n. Let $p := \{p_j\}_{j=1}^\infty$ be a probability distribution on $\mathbb{N}$ whose support is all of $\mathbb{N}$; that is, $p_j > 0$, for all $j \in \mathbb{N}$. Take a countably infinite sequence of independent samples from this distribution: $n_1, n_2, \cdots$. Now construct a random permutation $\Pi \in S_\infty$ as follows. Let $\Pi_1 = n_1$ and then for $k \geq 2$, let $\Pi_k = \psi_k(n_k)$, where ψ_k is the increasing bijection from $\mathbb{N}$ to $\mathbb{N} - \{\Pi_1, \cdots, \Pi_{k-1}\}$. Thus, for example, if the sequence of samples $\{n_j\}_{j=1}^\infty$ begins with $7,3,4,3,7,2,1$, then the construction yields the permutation Π beginning with $\Pi_1 = 7, \Pi_2 = 3, \Pi_3 = 5, \Pi_4 = 4, \Pi_5 = 11, \Pi_6 = 2, \Pi_7 = 1$. The probability measure $P^{\{p_j\}_{j=1}^\infty}$ on S_∞ is then the distribution of this random permutation Π. We call $P^{\{p_j\}_{j=1}^\infty}$ the *p-shifted distribution on S_∞* and Π a p-shifted random permutation on S_∞.

For $\sigma \in S_\infty$, we write $\sigma = \sigma_1\sigma_2\cdots$. For $n \in \mathbb{N}$, define $\mathrm{proj}_n(\sigma) \in S_n$ to be the permutation obtained from σ by deleting σ_i for all i satisfying $\sigma_i > n$. Thus, for $n = 4$ and $\sigma = 2539461\cdots$, one has $\mathrm{proj}_4(\sigma) = 2341$. Given the p-shifted random permutation $\Pi \in S_\infty$ that was constructed in the previous paragraph, define $P_n^{\{p_j\}_{j=1}^\infty}$ as the distribution of the random permutation $\Pi^{(n)} := \mathrm{proj}_n(\Pi)$; that is, $P_n^{\{p_j\}_{j=1}^\infty}(\sigma) = P^{\{p_j\}_{j=1}^\infty}(\mathrm{proj}_n^{-1}(\sigma)),\ \sigma \in S_n$. We call $P_n^{\{p_j\}_{j=1}^\infty}$ the *p-shifted distribution on S_n* and $\Pi^{(n)}$ a p-shifted random permutation on S_n. A more efficient and useful construction of $\Pi^{(n)}$ is given in Proposition 2.1 in section 2.

We note that in the case that $p_j = (1-q)q^{j-1}$, where $q \in (0,1)$, the measure $P_n^{\{p_j\}_{j=1}^\infty}$ is the Mallows distribution on S_n with parameter q; see [8], [4].

Remark 1.8. We assume in this paper that $p_j > 0$, for all j. In fact, the p-shifted random permutation can be constructed as long as $p_1 > 0$, with no positivity requirement on $p_j, j \geq 2$. The positivity requirement for all j ensures that for all n, the support of the p-shifted measure P_n is all of S_n.

It is known [8] that a random permutation under the p-shifted distribution $P^{\{p_j\}_{j=1}^\infty}$ is *strictly regenerative*, where our definition of strictly regenerative is as follows. For a permutation $\pi = \pi_{a+1}\pi_{a+2}\cdots\pi_{a+m}$, of $\{a+1, a+2, \cdots, a+m\}$, define $\mathrm{red}(\pi)$, the reduced permutation of π, to be the permutation in S_m given by $\mathrm{red}(\pi)_i = \pi_{a+i} - a$.

We will call a random permutation Π of S_∞ *strictly regenerative* if almost surely there exist $0 = T_0 < T_1 < T_2 < \cdots$ such that $\Pi([T_j]) = [T_j]$, $j \geq 1$, and $\Pi([m]) \neq [m]$ if $m \notin \{T_1, T_2, \cdots\}$, and such that the random variables $\{T_k - T_{k-1}\}_{k=1}^\infty$ are IID and the random permutations $\{\text{red}(\Pi|_{[T_k]-[T_{k-1}]})\}_{k=1}^\infty$ are IID. The numbers $\{T_n\}_{m=1}^\infty$ are called the renewal or regeneration numbers. Our definition of strictly regenerative differs slightly from that in [8].

Let u_n denote the probability that the p-shifted random permutation Π has a renewal at the number n; that is, $u_n = P^{\{p_j\}_{j=1}^\infty}(\sigma \in S_\infty : \sigma([n]) = [n])$. It follows easily from the construction of the random permutation that

$$u_n = \prod_{j=1}^n (\sum_{i=1}^j p_i) = \prod_{j=1}^n (1 - \sum_{i=j+1}^\infty p_i). \tag{1.10}$$

See [8]. Thus, $u_n > 0$, for all n. (Note that this positivity, and the consequent aperiodicity of the renewal mechanism, does not require the positivity of all p_j, but only of p_1.)

The strictly regenerative distribution $P^{\{p_j\}_{j=1}^\infty}$ is called positive recurrent if T_1 has finite expectation: $E^{\{p_j\}_{j=1}^\infty} T_1 < \infty$. From standard renewal theory, it follows that

$$\lim_{n\to\infty} u_n = \frac{1}{E^{\{p_j\}_{j=1}^\infty} T_1}. \tag{1.11}$$

Since $\sum_{j=1}^\infty \sum_{i=j+1}^\infty p_i = \sum_{j=1}^\infty j p_{j+1}$, it follows from (1.10) and (1.11) that

$$P^{\{p_j\}_{j=1}^\infty} \text{ is positive recurrent if and only if } \sum_{n=1}^\infty n p_n < \infty. \tag{1.12}$$

We now state our theorem concerning super-clustering.

Theorem 1.9. *Let $A_{l;k}^{(n)} \subset S_n$ denote the event that the set of l consecutive numbers $\{k, k+1, \cdots, k+l-1\}$ appears in a set of l consecutive positions. Let $\{p_n\}_{n=1}^\infty$ be a probability distribution on $\mathbb{N}$ with $p_j > 0$, for all $j \in \mathbb{N}$. Also assume that the sequence $\{p_n\}_{n=1}^\infty$ is non-increasing. Then for all $k \in \mathbb{N}$,*

$$\begin{aligned}&\lim_{l\to\infty} \lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l,k}^{(n)}) = \lim_{l\to\infty} \lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l,n+2-k-l}^{(n)}) = \\ &(\prod_{j=1}^{k-1} \sum_{i=1}^j p_i)(\prod_{j=1}^\infty \sum_{i=1}^j p_i).\end{aligned} \tag{1.13}$$

Also, if $\lim_{n\to\infty} \min(k_n, n-k_n) = \infty$, then

$$\lim_{l\to\infty} \liminf_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l,k_n}^{(n)}) = \lim_{l\to\infty} \limsup_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l,k_n}^{(n)}) = \big(\prod_{j=1}^\infty \sum_{i=1}^j p_i\big)^2. \tag{1.14}$$

Furthermore, the limits in (1.13) *and* (1.14) *are positive, that is, super-clustering occurs, if and only if $\sum_{n=1}^\infty n p_n < \infty$, or equivalently, if and only if the p-shifted random permutation is positive recurrent.*

Remark 1.10. If one removes the requirement that the sequence $\{p_j\}_{j=1}^\infty$ be non-increasing, then it follows immediately from the proof of the theorem that

$$\lim_{l\to\infty} \lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l,k}^{(n)}) \geq (\prod_{j=1}^{k-1} \sum_{i=1}^j p_i)(\prod_{j=1}^\infty \sum_{i=1}^j p_i)$$

and

$$\lim_{l\to\infty} \liminf_{n\to\infty} P_n^{\{p_j\}_{j=1}^\infty}(A_{l,k_n}^{(n)}) \geq \big(\prod_{j=1}^\infty \sum_{i=1}^j p_i\big)^2.$$

Thus, for this more general case, the finiteness of $\sum_{n=1}^{\infty} np_n$ is a sufficient condition for super-clustering.

Remark 1.11. Consider Theorem 1.9 in the case of the Mallows distribution P_n^q with parameter $q \in (0,1)$; that is, the case $p_j = (1-q)q^{j-1}$. From (1.13) and (1.14) we have

$$\begin{aligned}
&\lim_{l\to\infty}\lim_{n\to\infty} P_n^q(A_{l,k}^{(n)}) = \lim_{l\to\infty}\lim_{n\to\infty} P_n^q(A_{l,n+2-k-l}^{(n)}) = \\
&\big(\prod_{j=1}^{k-1}(1-q^j)\big)\big(\prod_{j=1}^{\infty}(1-q^j)\big), \text{ for all } k \in \mathbb{N}; \\
&\lim_{l\to\infty}\lim_{n\to\infty} P_n^q(A_{l,k_n}^{(n)}) = \big(\prod_{j=1}^{\infty}(1-q^j)\big)^2, \text{if } \lim_{n\to\infty}\min(k_n, n-k_n) = \infty.
\end{aligned} \tag{1.15}$$

In particular, super-clustering occurs for the Mallows distribution for fixed parameter $q \neq 1$.

We now establish a connection between super-clustering and the inversion statistic $\mathcal{I}_n$. For $k \geq 2$ and $\sigma \in S_\infty$, or $\sigma \in S_n$ with $2 \leq k \leq n$, let $I_{<k}(\sigma)$ denote the number of inversions involving pairs of numbers $\{\{i,k\} : 1 \leq i < k\}$:

$$I_{<k}(\sigma) = \sum_{1\leq i<k} 1_{\sigma_k^{-1} < \sigma_i^{-1}}.$$

The statistics $\{I_{<k}\}_{k=2}^{\infty}$ are called the *backward ranks*. The following result was proven in [7].

Proposition P. Under $P^{\{p_j\}_{j=1}^{\infty}}$, the random variables $\{I_{<k}\}_{k=2}^{\infty}$ are independent. Furthermore, the distribution of $I_{<k}$ is given by

$$P^{\{p_j\}_{j=1}^{\infty}}(I_{<k} = l) = \frac{p_{l+1}}{\sum_{j=1}^{k} p_j},\ l = 0, 1, \cdots, k-1. \tag{1.16}$$

Remark 1.12. As noted in [7], from the constructions, it is immediate that Proposition P also holds with the probability measure $P^{\{p_j\}_{j=1}^{\infty}}$ on S_∞ replaced by the probability measure $P_n^{\{p_j\}_{j=1}^{\infty}}$ on S_n, and with $\{I_{<k}\}_{k=2}^{\infty}$ replaced by $\{I_{<k}\}_{k=2}^{n}$, for any $n = 2, 3, \cdots$.

The inversion statistic $\mathcal{I}_n$, defined in (1.5), can be represented as

$$\mathcal{I}_n(\sigma) = \sum_{k=2}^{n} I_{<k}(\sigma). \tag{1.17}$$

Let X be a random variable on $\mathbb{Z}^+$ whose distribution is characterized by $1+X$ having the distribution $\{p_j\}_{j=1}^{\infty}$; that is,

$$P(X = j) = p_{j+1},\ j = 0, 1, \cdots. \tag{1.18}$$

Noting that the distributions in (1.16) are truncated versions of the distribution of X, it follows readily from (1.17), Proposition P and the remark following the proposition that the inversion statistic $\mathcal{I}_n$ satisfies the following weak law of large numbers as $n \to \infty$:

$$\begin{aligned}
&\frac{\mathcal{I}_n}{n} \text{ under } P_n^{\{p_j\}_{j=1}^{\infty}} \text{ or under } P^{\{p_j\}_{j=1}^{\infty}} \text{ converges in probability to} \\
&EX = \sum_{n=1}^{\infty} np_{n+1} \in (0, \infty].
\end{aligned} \tag{1.19}$$

For a proof in the case of the Mallows distribution, see [7]. Theorem 1.9 and (1.19) show that for p-shifted distributions, super-clustering occurs if and only if the inversion statistic $\mathcal{I}_n$ has linear rather than super-linear growth.

In [7] it was shown that $\mathcal{I}_n$ under the Mallows distribution $P_n^{q_n}$, with $q_n = 1 - \frac{c}{n^\alpha}$, for $c > 0$ and $\alpha \in (0, 1]$, grows on the order $n^{1+\alpha}$. One can make cosmetic changes in the arguments in this paper to conclude that super-clustering does not occur in these cases. (These measures $P_n^{q_n}$ are a little different than the p-shifted measures on S_n discussed in this paper. The ones in this paper are constructed from one distribution p on $\mathbb{N}$, which is not the case with these measures because the parameter q_n is changing with n.)

In section 2 we present an alternative construction of p-shifted random permutations that is motivated by Proposition P and that, in the context of the Mallows distributions, can be found, for example, in [6]. This alternative construction will be important for the proofs of all three theorems. Section 2 also includes a useful duality result that will be needed for the proof of Theorem 1.1. The above-noted alternative construction of p-shifted random permutations will be used for both the upper and lower bound calculations in the proof of Theorem 1.9. The same type of upper bound calculations, specialized to the case of a Mallows distribution, will also be used in the proof of the upper bound in Theorem 1.1 and in Theorem 1.5. On the other hand, the original p-shifted construction, specialized to the case of a Mallows distribution, will be used for the proof of the lower bound in Theorem 1.1 and in Theorem 1.5. In light of this, it will be convenient to begin with the proof of Theorem 1.9, which is given in section 3. The proof of the upper bounds in Theorem 1.1 are given in section 4, and the proof of the lower bound in Theorem 1.1 is given in section 5. The proof of Theorem 1.5 is given in section 6.

2 Two auxiliary results

We begin this section with an alternative construction of the p-shifted distribution. Let X be a random variable on $\mathbb{Z}^+$ whose distribution is as in (1.18). Let $\{X_n\}_{n=2}^\infty$ be a sequence of independent random variables with the distribution of X_n being the distribution of X truncated at $n-1$:

$$P(X_n = i) = \frac{p_{i+1}}{\sum_{j=1}^n p_j},\ i = 0, 1, \cdots n-1. \tag{2.1}$$

To construct a p-shifted random permutation in S_∞, set the number 1 down on a horizontal line. Now inductively, if the numbers $\{1, \cdots, n-1\}$ have already been placed down on the line, where $n \ge 2$, then sample from X_n and place the number n on the line in the position for which there are X_n numbers to its right. Let Π denote the above constructed random permutation in S_∞, and let $\Pi^{(n)}$ denote the random permutation in S_n obtained by terminating the above construction after n steps.

Proposition 2.1. *The random permutations Π and $\Pi^{(n)}$ generated above are distributed respectively as the p-shifted distributions $P^{\{p_j\}_{j=1}^\infty}$ and $P_n^{\{p_j\}_{j=1}^\infty}$.*

Proof. By the construction, the random variables $\{I_{<n}(\Pi)\}_{n=2}^\infty$ are independent and $I_{<n}(\Pi)$ is distributed as X_n. As is well-known, a permutation is uniquely determined by its backward ranks. This fact along with Proposition P and the remark following it prove that Π has the p-shifted distribution on S_∞ and that $\Pi^{(n)}$ has the p-shifted distribution on S_n. □

The following duality result will be used in the proof of Theorem 1.1.

Proposition 2.2.

$$P_n^q(A_{l;k}^{(n)}) = P_n^q(A_{l;n+2-k-l}^{(n)}),\ k = 1, 2 \ldots, n-l+1. \tag{2.2}$$

Proof. We defined earlier the reverse σ^{rev} of a permutation $\sigma \in S_n$. The *complement* of σ is the permutation σ^{com} satisfying $\sigma_i^{\text{com}} = n + 1 - \sigma_i,\ i = 1, \cdots, n$. Let $\sigma^{\text{rev-com}}$

denote the permutation obtained by applying reversal and then complementation to σ (or equivalently, applying complentation and then reversal). Since $\sigma_i^{\text{rev-com}} < \sigma_j^{\text{rev-com}}$ if and only $\sigma_{n+1-j} < \sigma_{n+1-i}$, it follows that σ and $\sigma^{\text{rev-com}}$ have the same number of inversions, and thus, from the definition of the Mallows distribution in (1.4), $P_n^q(\{\sigma\}) = P_n^q(\{\sigma^{\text{rev-com}}\})$. Using this along with the fact that $\sigma \in A_{l;k}^{(n)}$ if and only if $\sigma^{\text{rev-com}} \in A_{l;n+2-k-l}^{(n)}$ proves (2.2). □

3 Proof of Theorem 1.9

We note that the final statement of the theorem is almost immediate. Indeed, $\sum_{i=1}^{j} p_i = 1 - \sum_{i=j+1}^{\infty} p_i$ and $\sum_{j=1}^{\infty}\left(\sum_{i=j+1}^{\infty} p_i\right) = \sum_{j=1}^{\infty} jp_{j+1}$. Thus, the infinite product $\prod_{j=1}^{\infty}\sum_{i=1}^{j} p_i$ is positive if and only if $\sum_{j=1}^{\infty} jp_j < \infty$.

We now turn to the proofs of (1.13) and (1.14). We use the alternative method for constructing the p-shifted random permutation, as described at the beginning of section 2. Thus, we consider a sequence of independent random variables $\{X_n\}_{n=2}^{\infty}$, with X_n distributed as in (2.1). For the proof, we will use the notation

$$N_n = \sum_{i=1}^{n} p_i = P(X \le n-1),\ n \in \mathbb{N},\ \text{and}\ \ N_0 = 0, \tag{3.1}$$

where X is as in (1.18). Note that N_n is the normalization constant on the right hand side of (2.1). Although $P_n^{\{p_j\}_{j=1}^{\infty}}$ denotes the p-shifted probability measure on S_n, we will also use this notation for probabilities of events related to the random variables $\{X_j\}_{j=2}^{n}$ since these random variables are used in the construction of the $P_n^{\{p_j\}_{j=1}^{\infty}}$-distributed random permutation. However, probabilities of events related to X will be denoted by P.

We begin with the proof of (1.13). Fix $k \in \mathbb{N}$. Consider the event, which we denote by $B_{l;k}$, that after the first $k+l-1$ positive integers have been placed down on the horizontal line, the set of l numbers $\{k, k+1, \cdots, k+l-1\}$ appear in a set of l consecutive positions. Then $B_{l;k} = \cup_{a=0}^{k-1} B_{l;k;a}$, where the events $\{B_{l;k;a}\}_{a=0}^{k-1}$ are disjoint, with $B_{l;k;a}$ being the event that the set of l numbers $\{k, k+1, \cdots, k+l-1\}$ appear in a set of l consecutive positions and also that exactly a of the numbers in $[k-1]$ are to the right of this set. We calculate $P_n^{\{p_j\}_{j=1}^{\infty}}(B_{l;k;a})$.

Lemma 3.1.

$$P_n^{\{p_j\}_{j=1}^{\infty}}(B_{l;k;a}) = \prod_{j=0}^{l-1} \frac{N_{a+j+1} - N_a}{N_{k+j}}. \tag{3.2}$$

Proof. Suppose that we have already placed down on the horizontal line the numbers in $[k-1]$. Their relative positions are irrelevant for our considerations. Now we use X_k to insert on the line the number k. Suppose that $X_k = a$, $a \in \{0, \cdots, k-1\}$. Then the number k is inserted on the line in the position for which a of the numbers in $[k-1]$ are to its right. Now in order for $k+1$ to be placed in a position adjacent to k, we need $X_{k+1} \in \{a, a+1\}$. (If $X_{k+1} = a$, then $k+1$ will appear directly to the right of k, while if $X_{k+1} = a+1$, then $k+1$ will appear directly to the left of k.) If this occurs, then $\{k, k+1\}$ are adjacent, and a of the numbers in $[k-1]$ are to the right of $\{k, k+1\}$. Continuing in this vein, for $i \in \{1, \cdots, l-2\}$, given that the numbers $\{k, \cdots, k+i\}$ are adjacent to one another, and a of the numbers in $[k-1]$ appear to the right of $\{k, \cdots, k+i\}$, then in order for $k+i+1$ to be placed so that $\{k, \cdots, k+i+1\}$ are all adjacent to one another (with a of the numbers in $[k-1]$ appearing to the right of these numbers), we need

$X_{k+i+1} \in \{a, \cdots, a+i+1\}$. We conclude then that

$$P_n^{\{p_j\}_{j=1}^{\infty}}(B_{l;k;a}) = \prod_{j=0}^{l-1} P_n^{\{p_j\}_{j=1}^{\infty}}(X_{k+j} \in \{a, \cdots, a+j\}). \tag{3.3}$$

Using (2.1), we have

$$P_n^{\{p_j\}_{j=1}^{\infty}}(X_{k+j} \in \{a, \cdots, a+j\}) = \frac{N_{a+j+1} - N_a}{N_{k+j}}. \tag{3.4}$$

The lemma follows from (3.3) and (3.4). □

We now consider the conditional probability, $P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a})$, that is, the probability, given that $B_{l;k;a}$ has occurred, that the numbers $k+l, \cdots, n$ are inserted in such a way so as to preserve the mutual adjacency of the numbers in the set $\{k, \cdots, k+l-1\}$. We have the following lower bound.

Lemma 3.2.

$$P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a}) \geq \prod_{j=0}^{n-k-l} \frac{N_{a+j+1}}{N_{k+l+j}}. \tag{3.5}$$

Proof. The conditional probability in the statement of the lemma is larger or equal to the probability of the event that all of the remaining numbers are inserted to the right of the set $\{k, \cdots, k+l-1\}$. This event is given by $\cap_{j=0}^{n-k-l}\{X_{k+l+j} \leq a+j\}$. We have

$$P_n^{\{p_j\}_{j=1}^{\infty}}(\cap_{j=0}^{n-k-l}\{X_{k+l+j} \leq a+j\}) = \prod_{j=0}^{n-k-l} \frac{N_{a+j+1}}{N_{k+l+j}}. \tag{3.6}$$

□

It is clear from the construction that $P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a})$ is decreasing in n. Thus, since $P_n^{\{p_j\}_{j=1}^{\infty}}(B_{l;k;a})$ is independent of n, it follows that $P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)})$ is decreasing in n. Consequently $\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)})$ exists. Since $P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)})$ is clearly decreasing in l, it then follows that $\lim_{l\to\infty}\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)})$ also exists. Using Lemmas 3.1 and 3.2, we now obtain a lower bound on $\lim_{l\to\infty}\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)})$. Writing

$$P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) = \sum_{a=0}^{k-1} P_n^{\{p_j\}_{j=1}^{\infty}}(B_{l;k;a}) P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a}), \tag{3.7}$$

(3.2) and (3.5) yield

$$P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) \geq \sum_{a=0}^{k-1} \Big(\prod_{j=0}^{l-1} \frac{N_{a+j+1} - N_a}{N_{k+j}}\Big)\Big(\prod_{j=0}^{n-k-l} \frac{N_{a+j+1}}{N_{k+l+j}}\Big). \tag{3.8}$$

We have $\prod_{j=0}^{n-k-l} \frac{N_{a+j+1}}{N_{k+l+j}} = \frac{N_{a+1}\cdots N_{k+l-1}}{N_{n-k-l+a+2}\cdots N_n}$. Using this along with the fact that $\lim_{n\to\infty} N_n = 1$ and the fact that the limit of the left hand side of (3.8) as $n \to \infty$ exists, we have

$$\begin{aligned} &\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) \geq \sum_{a=0}^{k-1} \Big(\prod_{j=0}^{l-1} \frac{N_{a+j+1} - N_a}{N_{k+j}}\Big)\Big(\prod_{i=a+1}^{k+l-1} N_i\Big) = \\ &\sum_{a=0}^{k-1} \Big(\prod_{j=0}^{l-1}(N_{a+j+1} - N_a)\Big)\Big(\prod_{i=a+1}^{k-1} N_i\Big). \end{aligned} \tag{3.9}$$

We now let $l \to \infty$ in (3.9). We only consider the term in the summation with $a = 0$, because it turns out that the terms with $a \geq 1$ converge to 0 as $l \to \infty$. We obtain

$$\lim_{l\to\infty}\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) \geq \big(\prod_{j=1}^{k-1} N_j\big)\big(\prod_{j=1}^{\infty} N_j\big) = \big(\prod_{j=1}^{k-1}\sum_{i=1}^{j} p_i\big)\big(\prod_{j=1}^{\infty}\sum_{i=1}^{j} p_i\big). \tag{3.10}$$

We now turn to an upper bound on $P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a})$.

Lemma 3.3.

$$P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a}) \leq \prod_{j=1}^{n-k-l+1}(1 - N_{a+j+l-1} + N_{a+j}). \tag{3.11}$$

Proof. For the proof of this upper bound, we note the following facts. By the assumption that $\{p_n\}_{n=1}^{\infty}$ is non-increasing, it follows that $P(X \notin \{j+1, \cdots, j+l-1\})$ is increasing in j. Also, $P(X \notin \{j+1, \cdots, j+l-1\}) > P_n^{\{p_j\}_{j=1}^{\infty}}(X_m \notin \{j+1, \cdots, j+l-1\})$, for $j+l \leq m$.

Recall that $P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}|B_{l;k;a})$ is the conditional probability, given $B_{l;k;a}$, that the numbers $k+l, \cdots, n$ are inserted in such a way so as to preserve the mutual adjacency of the set $\{k, \cdots, k+l-1\}$. First the number $k+l$ is inserted. The probability that its insertion preserves the mutual adjacency property of the set $\{k, \cdots, k+l-1\}$ is $P_n^{\{p_j\}_{j=1}^{\infty}}(X_{k+l} \notin \{a+1, \cdots, a+l-1\})$, which is less than $P(X \notin \{a+1, \cdots, a+l-1\})$. If the insertion of $k+l$ preserves the mutual adjacency, then either $X_{k+l} \in \{0, \cdots, a\}$ or $X_{k+l} \in \{a+l, \cdots, k+l-1\}$. If $X_{k+l} \in \{0, \cdots, a\}$, then in order for the mutually adjacency to be preserved when the number $k+l+1$ is inserted, one needs $\big\{X_{k+1+1} \notin \{a+2, \cdots, a+l\}\big\}$, while if $X_{k+l} \in \{a+l, \cdots, k+l-1\}$, then one needs $\big\{X_{k+1+1} \notin \{a+1, \cdots, a+l-1\}\big\}$. The probability of either of these events is less than $P(X \notin \{a+2, \cdots, a+l\})$. Thus, an upper bound for the conditional probability, given $B_{l;k;a}$, that the insertion of $k+l$ and $k+l+1$ preserves the mutual adjacency is $P(X \notin \{a+1, \cdots, a+l-1\})P(X \notin \{a+2, \cdots, a+l\})$. Continuing in this vein, we conclude that

$$P_n^{(\{p_j\}_{j=1}^{\infty})}(A_{l;k}^{(n)}|B_{l;k;a}) \leq \prod_{j=1}^{n-k-l+1} P(X \notin \{a+j, \cdots, a+j+l-2\}). \tag{3.12}$$

We have

$$P(X \notin \{a+j, \cdots, a+j+l-2\}) = (1 - N_{a+j+l-1} + N_{a+j}) \tag{3.13}$$

The lemma now follows from (3.12) and (3.13). □

We now use Lemmas 3.1 and 3.3 to obtain an upper bound on $\lim_{l\to\infty}\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)})$. From (3.7) (3.2) and (3.11), we have

$$P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) \leq \sum_{a=0}^{k-1}\big(\prod_{j=0}^{l-1}\frac{N_{a+j+1} - N_a}{N_{k+j}}\big)\big(\prod_{j=1}^{n-k-l+1}(1 - N_{a+j+l-1} + N_{a+j})\big). \tag{3.14}$$

Letting $n \to \infty$ and using the fact, noted above in the proof of the lower bound, that the limit of the left hand side of (3.14) exists as $n \to \infty$, we have

$$\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) \leq \sum_{a=0}^{k-1}\big(\prod_{j=0}^{l-1}\frac{N_{a+j+1} - N_a}{N_{k+j}}\big)\big(\prod_{j=1}^{\infty}(1 - N_{a+j+l-1} + N_{a+j})\big). \tag{3.15}$$

We have

$$\frac{N_{a+j+1} - N_a}{N_{k+j}} = 1 - \frac{N_a + (N_{k+j} - N_{a+j+1})}{N_{k+j}} < 1 - N_a \in (0,1),$$

for all $j \geq 0$ and $a \in \{1, \cdots, k-1\}$.

Therefore, when letting $l \to \infty$ in (3.15), a contribution will come from the right hand side only when $a = 0$. We obtain

$$\begin{aligned}&\lim_{l\to\infty}\lim_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k}^{(n)}) \le \lim_{l\to\infty}\Big(\prod_{j=0}^{l-1}\frac{N_{j+1}}{N_{k+j}}\Big)\Big(\prod_{j=1}^{\infty}(1-N_{j+l-1}+N_j)\Big) = \\ &\Big(\prod_{j=1}^{k-1}N_j\Big)\Big(\prod_{j=1}^{\infty}N_j\Big) = \Big(\prod_{j=1}^{k-1}\sum_{i=1}^{j}p_i\Big)\Big(\prod_{j=1}^{\infty}\sum_{i=1}^{j}p_i\Big).\end{aligned} \tag{3.16}$$

Now (1.13) follows from (3.10) and (3.16).

We now turn to the proof of (1.14). As with the proof of (1.13), the term with $a = 0$ will dominate. Thus, for the lower bound, using (3.8) with $k = k_n$ and ignoring the terms with $a \ge 1$, we have

$$P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k_n}^{(n)}) \ge \Big(\prod_{j=0}^{l-1}\frac{N_{j+1}}{N_{k_n+j}}\Big)\Big(\prod_{j=0}^{n-k_n-l}\frac{N_{j+1}}{N_{k_n+l+j}}\Big). \tag{3.17}$$

Letting $n \to \infty$ in (3.17) and using the assumption that $\lim_{n\to\infty}\min(k_n, n-k_n) = \infty$, it follows that

$$\liminf_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k_n}^{(n)}) \ge \Big(\prod_{j=1}^{l}N_j\Big)\Big(\prod_{j=1}^{\infty}N_j\Big).$$

Now letting $l \to \infty$ gives

$$\lim_{l\to\infty}\liminf_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k_n}^{(n)}) \ge \Big(\prod_{j=1}^{\infty}N_j\Big)^2 = \Big(\prod_{j=1}^{\infty}\sum_{i=1}^{j}p_i\Big)^2. \tag{3.18}$$

For the upper bound, let $k = k_n$ in (3.14). The second factor in the summand, $\big(\prod_{j=0}^{l-1}\frac{N_{a+j+1}-N_a}{N_{k_n+j}}\big)\big(\prod_{j=1}^{n-k_n-l+1}(1-N_{a+j+l-1}+N_{a+j})\big)$, is less than 1, while the first factor in the summand satisfies

$$\prod_{j=0}^{l-1}\frac{N_{a+j+1}-N_a}{N_{k_n+j}} \le \frac{N_{a+1}-N_a}{N_{k_n}} = \frac{p_{a+1}}{N_{k_n}} \le \frac{p_{a+1}}{p_1},$$

for $a \in \{0, \cdots, k_n - 1\}$ and $n \ge 1$. Since $\sum_{a=0}^{\infty}\frac{p_{a+1}}{p_1} < \infty$, the dominated convergence theorem and the assumption that $\lim_{n\to\infty}\min(k_n, n-k_n) = \infty$ allow us to conclude upon letting $n \to \infty$ in (3.14) with $k = k_n$ that

$$\limsup_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k_n}^{(n)}) \le \sum_{a=0}^{\infty}\Big(\prod_{j=0}^{l-1}(N_{a+j+1}-N_a)\Big)\Big(\prod_{j=1}^{\infty}(1-N_{a+j+l-1}+N_{a+j})\Big). \tag{3.19}$$

For $a \ge 1$, we have $N_{a+j+1} - N_a \in (0, 1-p_1)$. Consequently, when letting $l \to \infty$ in (3.19), a contribution will come from the right hand side only when $a = 0$. We obtain

$$\lim_{l\to\infty}\limsup_{n\to\infty} P_n^{\{p_j\}_{j=1}^{\infty}}(A_{l;k_n}^{(n)}) \le \Big(\prod_{j=1}^{\infty}N_j\Big)^2 = \Big(\prod_{j=1}^{\infty}\sum_{i=1}^{j}p_i\Big)^2. \tag{3.20}$$

Now (1.14) follows from (3.18) and (3.20). $\square$

4 Proofs of upper bounds in Theorem 1.1

We begin the proof of (1.7) and the proof of the upper bound in (1.6) in tandem, and then consider each case separately at an appropriate point in the exposition. We also need to show that the upper bound in (1.6) is uniform over all sequences $\{k_n\}_{n=1}^{\infty}$.

Recall that (1.7) is stated to hold under the assumption $\frac{\min(k_n,n-k_n)}{n^\alpha}=\infty$, while (1.6) is stated to hold with no assumption on $\{k_n\}_{n=1}^{\infty}$. In fact, for our proofs, we will always need to assume that

$$\lim_{n\to\infty}\frac{k_n}{n^\alpha}=\infty. \tag{4.1}$$

What allows us to make this assumption is Proposition 2.2. Thus, in the sequel we will always assume that (4.1) holds.

We follow the same construction used in the proof of the upper bound in Theorem 1.9. We start from (3.14) with k replaced by k_n and with $\{p_j\}_{j=1}^{\infty}$ satisfying $p_j=(1-q_n)q_n^{j-1}$, since the Mallows distribution $P_n^{q_n}$ is the p-shifted distribution with this choice of p-distribution. Thus, it follows from (3.1) that for the case at hand,

$$N_b=\sum_{i=1}^{b}(1-q_n)q_n^{i-1}=1-q_n^b. \tag{4.2}$$

Substituting (4.2) in (3.14), we obtain

$$P_n^{q_n}(A_{l;k_n}^{(n)})\le\prod_{j=0}^{l-1}\frac{1-q_n^{j+1}}{1-q_n^{k_n+j}}\sum_{a=0}^{k_n-1}q_n^{al}\prod_{j=1}^{n-k_n-l+1}\big(1-q_n^{a+j}+q_n^{a+j+l-1}\big). \tag{4.3}$$

We have

$$1-q_n^b=1-(1-\frac{c}{n^\alpha})^b\sim\frac{bc}{n^\alpha},\ \text{for } b\in\mathbb{N}, \tag{4.4}$$

and

$$1-q_n^{k_n+j}=1-(1-\frac{c}{n^\alpha})^{k_n+j}\ge1-e^{-\frac{c(k_n+j)}{n^\alpha}}. \tag{4.5}$$

From (4.4) and (4.5) along with the assumption on q_n and the assumption (4.1) on k_n, the term multiplying the summation in (4.3) satisfies

$$\prod_{j=0}^{l-1}\frac{1-q_n^{j+1}}{1-q_n^{k_n+j}}\sim\frac{l!c^l}{n^{\alpha l}}. \tag{4.6}$$

Using (4.4), the summation in (4.3) satisfies

$$\sum_{a=0}^{k_n-1}q_n^{al}\prod_{j=1}^{n-k_n-l+1}\big(1-q_n^{a+j}+q_n^{a+j+l-1}\big)\sim\sum_{a=0}^{k_n-1}q_n^{al}\prod_{j=1}^{n-k_n-l+1}\Big(1-\frac{q_n^{a+j}(l-1)c}{n^\alpha}\Big). \tag{4.7}$$

We split up the continuation of the proof between the upper bound in (1.6), in which case no assumption is made on k_n (accept for (4.1), as explained above), and the bound in (1.7), in which case k_n is assumed to satisfy $\frac{\min(k_n,n-k_n)}{n^\alpha}=\infty$. We begin with the former case. In this case, from (4.3) along with (4.4), (4.6) and (4.7), we have

$$P_n^q(A_{l;k_n}^{(n)})\lesssim\frac{l!c^l}{n^{\alpha l}}\sum_{a=0}^{k_n-1}q_n^{al}\le\frac{l!c^l}{n^{\alpha l}}\frac{1}{1-q_n^l}\sim\frac{1}{l}\,\frac{l!c^{l-1}}{n^{\alpha(l-1)}},$$

which is the upper bound in (1.6).

It is clear from the above proof that if we fix $\alpha'\in(\alpha,1)$ and let $k_n'=[n^{\alpha'}]$, then the upper bound in (1.6) is uniform over sequences $\{k_n\}_{n=1}^{\infty}$ satisfying $k_n\ge k_n'$. From this along with (2.2), it follows that the upper bound in (1.6) is in fact uniform over all sequences $\{k_n\}_{n=1}^{\infty}$.

Now consider the bound in (1.7) and assume that k_n satisfies $\frac{\min(k_n,n-k_n)}{n^\alpha}=\infty$. In the previous case, we simply replaced the product on the right hand side of (4.7) by one. For the current case, we analyze this product. We write

$$\log\prod_{j=1}^{n-k_n-l+1}\Big(1-\frac{q_n^{a+j}(l-1)c}{n^\alpha}\Big)=\sum_{j=1}^{n-k_n-l+1}\log\Big(1-\frac{q_n^{a+j}(l-1)c}{n^\alpha}\Big). \tag{4.8}$$

We have

$$\begin{aligned}&\int_0^{n-k_n-l+1}\log\Big(1-\frac{q_n^{a+x}(l-1)c}{n^\alpha}\Big)dx\le\sum_{j=1}^{n-k_n-l+1}\log\Big(1-\frac{q_n^{a+j}(l-1)c}{n^\alpha}\Big)\le\\&\int_1^{n-k_n-l+2}\log\Big(1-\frac{q_n^{a+x}(l-1)c}{n^\alpha}\Big)dx.\end{aligned} \tag{4.9}$$

Making the change of variables, $y=q_n^x$, we have

$$\int_A^B\log\Big(1-\frac{q_n^{a+x}(l-1)c}{n^\alpha}\Big)dx=-\frac{1}{\log q_n}\int_{q_n^B}^{q_n^A}\frac{\log\big(1-\frac{q_n^a(l-1)c}{n^\alpha}y\big)}{y}dy. \tag{4.10}$$

From (4.10) and the assumptions on q_n and k_n, both the left and the right hand sides of (4.9) are asymptotic to $\frac{n^\alpha}{c}\int_0^1\frac{\log\big(1-\frac{q_n^a(l-1)c}{n^\alpha}y\big)}{y}dy$. A change of variables gives

$$\int_0^1\frac{\log(1-By)}{y}dy=\int_0^B\frac{\log(1-u)}{u}du\sim -B,\ \text{as } B\to0^+.$$

Therefore,

$$\frac{n^\alpha}{c}\int_0^1\frac{\log\big(1-\frac{q_n^a(l-1)c}{n^\alpha}y\big)}{y}dy\sim-(l-1)q_n^a,\ \text{uniformly over } a\in\{0,\cdots,k_n-1\}.$$

Consequently, the middle term in (4.9) satisfies

$$\sum_{i=1}^{n-k_n-l+1}\log\Big(1-\frac{q_n^{a+i}(l-1)c}{n^\alpha}\Big)\sim-(l-1)q_n^a,\ \text{uniformly over } a\in\{0,\cdots,k_n-1\}. \tag{4.11}$$

From (4.8) and (4.11), we have

$$\prod_{j=1}^{n-k_n-l+1}\Big(1-\frac{q_n^{a+j}(l-1)c}{n^\alpha}\Big)\sim e^{-(l-1)q_n^a},\ \text{uniformly over } a\in\{0,\cdots,k_n-1\}. \tag{4.12}$$

From (4.3) along with (4.6), (4.7) and (4.12), we obtain

$$P_n^q(A_{l;k_n}^{(n)})\lesssim\frac{l!c^l}{n^{\alpha l}}\sum_{a=0}^{k_n-1}q_n^{al}e^{-(l-1)q_n^a}. \tag{4.13}$$

By the assumptions on k_n and q_n, $\sum_{a=0}^{k_n-1}q_n^{al}e^{-(l-1)q_n^a}$ is asymptotic to $\int_0^{k_n}q_n^{xl}e^{-(l-1)q_n^x}dx$. Making the change of variables $y=q_n^x$, this integral is equal to $-\frac{1}{\log q_n}\int_{q_n^{k_n}}^1 y^{l-1}e^{-(l-1)y}dy$, which in turn is asymptotic to $\frac{n^\alpha}{c}\int_0^1 y^{l-1}e^{-(l-1)y}dy$. Thus,

$$\sum_{a=0}^{k_n-1}q_n^{al}e^{-(l-1)q_n^a}\sim\frac{n^\alpha}{c}\int_0^1 y^{l-1}e^{-(l-1)y}dy. \tag{4.14}$$

From (4.13) and (4.14), we conclude that

$$P_n^q(A_{l;k_n}^{(n)}) \lesssim \Big(\int_0^1 y^{l-1}e^{-(l-1)y}dy\Big)\,\frac{c^{l-1}l!}{n^{\alpha(l-1)}},$$

which is the upper bound in (1.7). □

5 Proof of lower bound in Theorem 1.1

We need to prove the lower bound in (1.6), and we need to show that it holds uniformly over all sequences $\{k_n\}_{n=1}^\infty$. As in the proof of the upper bounds, we will assume that k_n satisfies (4.1), which is permissible by Proposition 2.2. The method used in the proof of Theorem 1.9 and in the proof of the upper bounds in Theorem 1.1, via the alternative method for constructing a p-shifted random permutation, is not precise enough to be of use in the proof of the lower bound in Theorem 1.1. For the lower bound, we utilize the original construction for p-shifted random permutations on S_n, specializing to the Mallows distribution with parameter q_n, for which $p_j = (1-q_n)q_n^{j-1}$. We use the notation $P_n^{q_n}$ not only for the Mallows distribution itself, but also for probabilities of events associated with the construction. With regard to this construction, for $j \in \{0,\cdots,k_n-1\}$, let $C_{j;k_n,l}$ denote the event that exactly j numbers from the set $\{1,\cdots,k_n-1\}$ appear in the permutation before any number from the set $\{k_n,\cdots,k_n+l-1\}$ appears. We calculate $P_n^{q_n}(C_{j;k_n,l})$ explicitly.

Lemma 5.1.

$$P_n^{q_n}(C_{j;k_n,l}) = \begin{cases} \frac{(1-q_n^l)q_n^{k_n-1-j}}{1-q_n^{k_n-1-j+l}}\,\frac{\prod_{b=k_n-j}^{k_n-1}(1-q_n^b)}{\prod_{b=k_n-j+l}^{k_n+l-1}(1-q_n^b)}, & j\in\{0,1,\cdots,l-1\};\\ \frac{(1-q_n^l)q_n^{k_n-1-j}}{1-q_n^{k_n-1-j+l}}\,\frac{\prod_{b=k_n-j}^{k_n-j+l-1}(1-q_n^b)}{\prod_{b=k_n}^{k_n+l-1}(1-q_n^b)}, & j\in\{l,l+1,\cdots,k_n-1\}.\end{cases} \tag{5.1}$$

Proof. For $a,b\in\mathbb{N}$, let $r_{a,b}$ denote the probability that in the construction, the first number that appears from the set $\{1,\cdots,a+b\}$ comes from the set $\{1,\cdots,a\}$. Then

$$r_{a,b} = \frac{\sum_{j=1}^{a}(1-q_n)q_n^{j-1}}{\sum_{j=1}^{a+b}(1-q_n)q_n^{j-1}} = \frac{1-q_n^a}{1-q_n^{a+b}}. \tag{5.2}$$

For convenience, define $r_{0,b}=0$. Then from the construction, it follows that

$$P_n^{q_n}(C_{j;k_n,l}) = \Big(\prod_{i=1}^{j} r_{k_n-i,l}\Big)(1-r_{k_n-j-1,l}),\ j=0,\cdots,k_n-1. \tag{5.3}$$

From (5.2) and (5.3), we have

$$\begin{aligned} P_n^{q_n}(C_{j;k_n,l}) &= \Big(\prod_{i=1}^{j}\frac{1-q_n^{k_n-i}}{1-q_n^{k_n-i+l}}\Big)\frac{q_n^{k_n-j-1}-q_n^{k_n-j-1+l}}{1-q_n^{k_n-j-1+l}} = \\ &\frac{(1-q_n^l)q_n^{k_n-1-j}}{1-q_n^{k_n-1-j+l}}\,\frac{\prod_{b=k_n-j}^{\min(k_n-j+l-1,k_n-1)}(1-q_n^b)}{\prod_{b=\max(k_n-j+l,k_n)}^{k_n+l-1}(1-q_n^b)}. \end{aligned} \tag{5.4}$$

The right hand side of (5.4) is equivalent to the right hand side of (5.1). □

We now obtain a lower bound on $P_n^{q_n}(A_{l;k_n}^{(n)}|C_{j;k_n,l})$.

Lemma 5.2.

$$P_n^{q_n}(A_{l;k_n}^{(n)}|C_{j;k_n,l}) \ge q_n^{(l-1)(k_n-1-j)}\,\frac{\prod_{b=1}^{l-1}(1-q_n^b)}{\prod_{b=n-l-j+1}^{n-j-1}(1-q_n^b)}. \tag{5.5}$$

Proof. In order for the event $A^{(n)}_{l;k_n}$ to occur, the l numbers $\{k_n, \cdots, k_n+l-1\}$ must appear consecutively (in arbitrary order) in the construction. Thus, given the event $C_{j;k_n,l}$, in order for the event $A^{(n)}_{l;k_n}$ to occur, all of the other $l-1$ numbers in $\{k_n, \cdots, k_n+l-1\}$ must occur immediately after the appearance of the first number from this set. Given $C_{j;k_n,l}$, after the appearance of the first number from $\{k_n, \cdots, k_n+l-1\}$, there are still k_n-1-j numbers from $\{1, \cdots, k_n-1\}$ that have not yet appeared, as well as a certain amount of numbers from $\{k_n+l, \cdots, n\}$. Thus, a lower bound on $P^{q_n}_n(A^{(n)}_{l;k_n}|C_{j;k_n,l})$ is obtained by assuming that none of the numbers from $\{k_n+l, \cdots, n\}$ have yet appeared. (Here it is appropriate to note that if we calculate an upper bound by assuming that all of the numbers from $\{k_n+l, \cdots, n\}$ have already appeared, then the upper bound we arrive at for $P^{q_n}_n(A^{(n)}_{l;k_n})$ is not as good as the upper bound in (1.7).)

In order to calculate explicitly this lower bound, for $a, b, c \in \mathbb{N}$, let $r_{a,b,c}$ denote the probability that the first number that appears from the set $\{1, \cdots, a+b+c\}$ comes from the set $\{1, \cdots, a\} \cup \{a+b+1, \cdots, a+b+c\}$. Then

$$r_{a,b,c} = \frac{\sum_{j=1}^{a}(1-q_n)q_n^{j-1} + \sum_{j=a+b+1}^{a+b+c}(1-q_n)q_n^{j-1}}{\sum_{j=1}^{a+b+c}(1-q_n)q_n^{j-1}} = \frac{1-q_n^a+q_n^{a+b}-q_n^{a+b+c}}{1-q_n^{a+b+c}}.$$

From the construction, the lower bound on $P^{q_n}_n(A^{(n)}_{l;k_n}|C_{j;k_n,l})$, obtained by assuming that none of the numbers from $\{k_n+l, \cdots, n\}$ have yet appeared, is given by

$$P^{q_n}_n(A^{(n)}_{l;k_n}|C_{j;k_n,l}) \ge \prod_{i=1}^{l-1}(1-r_{k_n-1-j,i,n-k_n-l+1}) = \prod_{i=1}^{l-1}\frac{q_n^{k_n-1-j}-q_n^{k_n-1-j+i}}{1-q_n^{n-l-j+i}} = \\ q_n^{(l-1)(k_n-1-j)}\frac{\prod_{b=1}^{l-1}(1-q_n^b)}{\prod_{b=n-l-j+1}^{n-j-1}(1-q_n^b)}. \quad \square$$

We now use Lemmas 5.1 and 5.2 to obtain a lower bound on $P^{q_n}_n(A^{(n)}_{l;k_n})$. From (5.1) and (5.5), we have

$$\begin{aligned} &P^{q_n}_n(A^{(n)}_{l;k_n}) = \sum_{j=0}^{k_n-1}P^{q_n}_n(C_{j;k_n,l})P^{q_n}_n(A^{(n)}_{l;k_n}|C_{j;k_n,l}) \ge \\ &\sum_{j=l}^{k_n-1}\frac{(1-q_n^l)q_n^{k_n-1-j}}{1-q_n^{k_n-1-j+l}}\frac{\prod_{b=k_n-j}^{k_n-j+l-1}(1-q_n^b)}{\prod_{b=k_n}^{k_n+l-1}(1-q_n^b)}q_n^{(l-1)(k_n-1-j)}\frac{\prod_{b=1}^{l-1}(1-q_n^b)}{\prod_{b=n-l-j+1}^{n-j-1}(1-q_n^b)}. \end{aligned} \tag{5.6}$$

Using the assumption on q_n in the last step below, the right hand side of (5.6) satisfies

$$\begin{aligned} &\sum_{j=l}^{k_n-1}\frac{(1-q_n^l)q_n^{k_n-1-j}}{1-q_n^{k_n-1-j+l}}\frac{\prod_{b=k_n-j}^{k_n-j+l-1}(1-q_n^b)}{\prod_{b=k_n}^{k_n+l-1}(1-q_n^b)}q_n^{(l-1)(k_n-1-j)}\frac{\prod_{b=1}^{l-1}(1-q_n^b)}{\prod_{b=n-l-j+1}^{n-j-1}(1-q_n^b)} \ge \\ &\Big(\prod_{b=1}^{l}(1-q_n^b)\Big)\sum_{j=l}^{k_n-1}q_n^{l(k_n-1-j)}\prod_{b=k_n-j}^{k_n-j+l-2}(1-q_n^b) \ge \\ &\Big(\prod_{b=1}^{l}(1-q_n^b)\Big)\sum_{j=l}^{k_n-1}q_n^{l(k_n-1-j)}(1-q_n^{k_n-j})^{l-1} \sim \frac{l!c^l}{n^{l\alpha}}\sum_{j=l}^{k_n-1}q_n^{l(k_n-1-j)}\big(1-q_n^{k_n-j}\big)^{l-1}. \end{aligned} \tag{5.7}$$

And

$$\sum_{j=l}^{k_n-1}q_n^{l(k_n-1-j)}\big(1-q_n^{k_n-j}\big)^{l-1} \sim \int_0^{k_n-1-l}q_n^{xl}(1-q_n^x)^{l-1}dx. \tag{5.8}$$

Making the change of variables $y = q_n^x$, and using the assumption on q_n and the assumption on k_n in (4.1), we have

$$\begin{aligned}&\int_0^{k_n-1-l} q_n^{xl}(1-q_n^x)^{l-1}dx = -\frac{1}{\log q_n}\int_{q_n^{k_n-1-l}}^1 y^{l-1}(1-y)^{l-1}dy \sim\\ &\frac{n^\alpha}{c}\int_0^1 y^{l-1}(1-y)^{l-1}dy = \frac{n^\alpha}{c}\frac{\Gamma(l)\Gamma(l)}{\Gamma(2l)} = \frac{n^\alpha}{c}\frac{\big((l-1)!\big)^2}{(2l)!}.\end{aligned} \tag{5.9}$$

From (5.6)-(5.9), we conclude that

$$P_n^{q_n}(A_{l;k_n}^{(n)}) \gtrsim \frac{\big((l-1)!\big)^2}{(2l)!}\frac{c^{l-1}l!}{n^{\alpha(l-1)}},$$

which is the lower bound in (1.6).

It is clear from the proof that if we fix $\alpha' \in (\alpha, 1)$ and let $k_n' = [n^{\alpha'}]$, then the lower bound in (1.6) is uniform over sequences $\{k_n\}_{n=1}^\infty$ satisfying $k_n \ge k_n'$. From this along with (2.2), it follows that the lower bound in (1.6) is in fact uniform over all sequences $\{k_n\}_{n=1}^\infty$. □

6 Proof of Theorem 1.5

Proof of part (i). We begin with the upper bound in (1.8). We follow a slightly more precise version of the construction that was used in the upper bound in Theorem 1.9, and then reused for the particular case of the Mallows distribution in the proof of the upper bounds in Theorem 1.1. As with the proof of the upper bounds in Theorem 1.1, we use the construction from the proof of Theorem 1.9 in the particular case of the Mallows distribution, with parameter q_n; namely, with $p_j = (1-q_n)q_n^{j-1}$. Then from (2.1), the random variables $\{X_j\}_{j=2}^\infty$ have truncated geometric distributions. Although $P_n^{q_n}$ denotes the Mallows distribution with parameter q_n, we also use this notation for probabilities of events related to the random variables $\{X_j\}_{j=2}^n$ since these random variables are used in the construction of the $P_n^{q_n}$-distributed random permutation. It is easy to check that $P_n^{q_n}(X_m \notin \{j+1, \cdots, j+l-1\})$ is monotone increasing in j. Thus, the argument leading up to (3.12) in fact gives the following slightly more precise version of (3.12):

$$\begin{aligned}&P_n^{q_n}(A_{l;k_n}^{(n)}|B_{l;k_n;a}) \le \prod_{i=1}^{n-k_n-l+1} P_n^{q_n}(X_{k_n+l-1+i} \notin \{a+i, \cdots, a+i+l-2\}) =\\ &\prod_{i=1}^{n-k_n-l+1}\big(1 - \frac{q_n^{a+i} - q_n^{a+i+l-1}}{1-q_n^{k_n+l+i-1}}\big).\end{aligned} \tag{6.1}$$

From the assumption on q_n,

$$\prod_{i=1}^{n-k_n-l+1}\big(1 - \frac{q_n^{a+i} - q_n^{a+i+l-1}}{1-q_n^{k_n+l+i-1}}\big) \sim \prod_{i=1}^{n-k_n-l+1}\Big(1 - \frac{(l-1)cn^{-1}q_n^{a+i}}{1-q_n^{k_n+l+i-1}}\Big). \tag{6.2}$$

We have

$$\begin{aligned}&\log\prod_{i=1}^{n-k_n-l+1}\Big(1 - \frac{(l-1)cn^{-1}q_n^{a+i}}{1-q_n^{k_n+l+i-1}}\Big) = \sum_{i=1}^{n-k_n-l+1}\log\Big(1 - \frac{(l-1)cn^{-1}q_n^{a+i}}{1-q_n^{k_n+l+i-1}}\Big) \sim\\ &\int_0^{n-k_n-l+1}\log\Big(1 - \frac{(l-1)cn^{-1}q_n^{a+x}}{1-q_n^{k_n+l+x-1}}\Big)dx, \text{ uniformly over } a \in \{0, \cdots, k_n-1\}.\end{aligned} \tag{6.3}$$

Making the change of variables $y = q_n^x$, using the assumptions on k_n and q_n and defining

$$\gamma(c,d) = \log\frac{1-e^{-cd}}{1-e^{-c}} < 0, \tag{6.4}$$

in order to simplify notation in the sequel, we have

$$\begin{aligned}
&\int_0^{n-k_n-l+1}\log\Big(1-\frac{(l-1)cn^{-1}q_n^{a+x}}{1-q_n^{k_n+l+x-1}}\Big)dx=\\
&-\frac{1}{\log q_n}\int_{q_n^{n-k_n-l+1}}^1\frac{\log\Big(1-\frac{(l-1)cn^{-1}q_n^a y}{1-q_n^{k_n+l-1}y}\Big)}{y}dy\le\\
&\frac{1}{\log q_n}\int_{q_n^{n-k_n-l+1}}^1\frac{(l-1)cn^{-1}q_n^a}{1-q_n^{k_n+l-1}y}dy\sim \qquad (6.5)\\
&-(l-1)q_n^a\int_{e^{-c(1-d)}}^1\frac{1}{1-e^{-cd}y}dy=\\
&(l-1)q_n^a e^{cd}\log\frac{1-e^{-cd}}{1-e^{-c}}=(l-1)q_n^a e^{cd}\gamma(c,d),\ \text{uniformly over } a\in\{0,\cdots,k_n-1\}.
\end{aligned}$$

From (6.1)-(6.5), we conclude that

$$P_n^{q_n}(A_{l;k_n}^{(n)}|B_{l;k;a})\lesssim e^{(l-1)q_n^a e^{cd}\gamma(c,d)},\ \text{uniformly over } a\in\{0,\cdots,k_n-1\}. \qquad (6.6)$$

Recall that for the particular case of the Mallows distribution with parameter q_n, the quantity N_b is given by (4.2). Thus, in this particular case, and with k replaced by k_n, (3.2) becomes

$$P_n^{q_n}(B_{l;k_n;a})=\prod_{j=0}^{l-1}\frac{q_n^a-q_n^{a+j+1}}{1-q_n^{k_n+j}}. \qquad (6.7)$$

Using (6.6) and (6.7), along with the assumptions on k_n and q_n, we have

$$\begin{aligned}
&P_n^{q_n}(A_{l;k_n}^{(n)})\lesssim\sum_{a=0}^{k_n-1}q_n^{al}\frac{\prod_{b=1}^l(1-q_n^b)}{\prod_{b=k_n}^{k_n+l-1}(1-q_n^b)}e^{(l-1)q_n^a e^{cd}\gamma(c,d)}\sim\\
&\frac{l!c^l}{n^l(1-e^{-cd})^l}\int_0^{dn}q_n^{xl}e^{(l-1)q_n^x e^{cd}\gamma(c,d)}dx.
\end{aligned} \qquad (6.8)$$

Making the change of variables $y=q_n^x$ and using the assumption on q_n, we obtain

$$\begin{aligned}
&\int_0^{dn}q_n^{xl}e^{(l-1)q_n^x e^{cd}\gamma(c,d)}dx=-\frac{1}{\log q_n}\int_{q_n^{dn}}^1 y^{l-1}e^{(l-1)e^{cd}\gamma(c,d)y}dy\sim\\
&\frac{n}{c}\int_{e^{-cd}}^1 y^{l-1}e^{(l-1)e^{cd}\gamma(c,d)y}dy.
\end{aligned} \qquad (6.9)$$

From (6.8), (6.9) and (6.4), we obtain the upper bound in (1.8).

We now turn to the lower bound in (1.8). We follow the proof of the lower bound in Theorem 1.1 in section 5 up through equation (5.6), as the proof in that section up to that point makes no assumption on q_n. The next equation in section 5, equation (5.7), contains α, which arises from the fact that $q_n=1-\frac{c}{n^\alpha}$. Equation (5.7) also holds in the present situation, where $q_n=1-\frac{c}{n}$, if we set $\alpha=1$ in (5.7). Equation (5.8) and the first line of equation (5.9) continue to hold in the present situation. However, in the present situation, the first term on the second line of (5.9) must be changed from $\frac{n^\alpha}{c}\int_0^1 y^{l-1}(1-y)^{l-1}dy$ to $\frac{n}{c}\int_{e^{-cd}}^1 y^{l-1}(1-y)^{l-1}dy$, because in the present situation $\alpha=1$ and $\lim_{n\to\infty}q_n^{k_n-1-l}=e^{-cd}$. Of course, in the present situation, we ignore the rest of the second line. Thus, from (5.6)-(5.9), with the above noted changes taken into account, we conclude that

$$P_n^{q_n}(A_{l;k_n}^{(n)})\gtrsim\Big(\int_{e^{-cd}}^1\big(y(1-y)\big)^{l-1}dy\Big)\frac{c^{l-1}\,l!}{n^{(l-1)}},$$

which is the lower bound in (1.8). □

Proof of part (ii). We write $q_n = 1 - \epsilon(n)$, where $0 < \epsilon(n) = o(\frac{1}{n})$. We follow the proof of part (i) through the first three lines of (6.5), the only change being that the term cn^{-1} is replaced by $\epsilon(n)$. Starting with the inequality between the first and third lines there, we have

$$\begin{aligned}&\int_0^{n-k_n-l+1} \log\Big(1 - \frac{(l-1)\epsilon(n) q_n^{a+x}}{1-q_n^{k_n+l+x-1}}\Big)dx \le \\ &\frac{1}{\log q_n}\int_{q_n^{n-k_n-l+1}}^{1} \frac{(l-1)\epsilon(n)q_n^a}{1-q_n^{k_n+l-1}y}dy = \\ &\frac{(l-1)\epsilon(n)q_n^a}{\log q_n} q_n^{-(k_n+l-1)}\log\big(\frac{1-q_n^n}{1-q_n^{k_n+l-1}}\big).\end{aligned} \tag{6.10}$$

Since $\epsilon(n) = o(\frac{1}{n})$, we have $1 - q_n^n \sim n\epsilon(n)$, $1 - q_n^{k_n+l-1} \sim k_n\epsilon(n)$, $q_n^{-(k_n+l-1)} \sim 1$ and $q_n^a \sim 1$, uniformly over $a \in \{0, \cdots, k_n - 1\}$. Using this with (6.10), we have

$$\begin{aligned}&\int_0^{n-k_n-l+1} \log\Big(1 - \frac{(l-1)\epsilon(n) q_n^{a+x}}{1-q_n^{k_n+l+x-1}}\Big)dx \lesssim (l-1)\log\frac{k_n}{n}, \\ &\text{uniformly over } a \in \{0, \cdots, k_n-1\}.\end{aligned} \tag{6.11}$$

From (6.1)-(6.3) (with cn^{-1} replaced by $\epsilon(n)$) and (6.11), we conclude that

$$P_n^{q_n}(A_{l;k_n}^{(n)}|B_{l;k;a}) \lesssim (\frac{k_n}{n})^{l-1}, \text{ uniformly over } a \in \{0, \cdots, k_n-1\}. \tag{6.12}$$

Using (6.12) and (6.7), along with the assumption on q_n, we conclude that

$$\begin{aligned}&P_n^{q_n}(A_{l;k_n}^{(n)}) \lesssim \sum_{a=0}^{k_n-1} q_n^{al}\frac{\prod_{b=1}^{l}(1-q_n^b)}{\prod_{b=k_n}^{k_n+l-1}(1-q_n^b)}(\frac{k_n}{n})^{l-1} \sim \\ &\frac{l!(\epsilon(n))^l}{(k_n\epsilon(n))^l}(\frac{k_n}{n})^{l-1}\sum_{a=0}^{k_n-1} q_n^{al} = \frac{l!(\epsilon(n))^l}{(k_n\epsilon(n))^l}(\frac{k_n}{n})^{l-1}\frac{1-q_n^{k_nl}}{1-q_n^l} \sim \\ &\frac{l!}{n^{l-1}k_n}\frac{k_nl\epsilon(n)}{l\epsilon(n)} = \frac{l!}{n^{l-1}}. \quad \square\end{aligned} \tag{6.13}$$

References

[1] Basu, R. and Bhatnagar, N., *Limit theorems for longest monotone subsequences in random Mallows permutations*, Ann. Inst. Henri Poincaré Probab. Stat. **53** (2017), 1934–1951. MR3729641

[2] Bhatnagar, N. and Peled, R., *Lengths of monotone subsequences in a Mallows permutation*, Probab. Theory Related Fields **161** (2015), 719–780. MR3334280

[3] Gladkich, A. and Peled, R., *On the cycle structure of Mallows permutations*, Ann. Probab. **46** (2018), 1114–1169. MR3773382

[4] Gnedin, A. and Olshanski, G., *The two-sided infinite extension of the Mallows model for random permutations*, Adv. in Appl. Math. **48** (2012), 615–639. MR2920835

[5] Kaplansky, I., *The asymptotic distribution of runs of consecutive elements*, Ann. Math. Statistics **16** (1945), 200–203. MR0013847

[6] Pinsky, R. *Permutations avoiding a pattern of length three under Mallows distributions*, Random Structures Algorithms **58** (2021), 676–690. MR4258132

[7] Pinsky, R. *Comparing the inversion statistic for distribution-biased and distribution-shifted permutations with the geometric and the GEM distributions*, ALEA Lat. Am. J. Probab. Math. Stat. **19** (2022), 209–229. MR4359791

[8] Pitman, J. and Tang, W., *Regenerative Random Permutations of Integers*, Ann. Probab., **47** (2019), 1378–1416. MR3945749

[9] Stanley, R. P., *Enumerative combinatorics*, Volume 1. second edition, Cambridge Studies in Advanced Mathematics, **49**, Cambridge University Press, Cambridge, (2012). MR2868112

[10] Wolfowitz, J. *Note on runs of consecutive elements*, Ann. Math. Statistics **15** (1944), 97–98. MR0010341