Séminaire Lotharingien de Combinatoire **86B** (2022)
Article #2, 12 pp.

Proceedings of the 34th Conference on Formal Power Series and Algebraic Combinatorics (Bangalore)

Refined Consecutive Pattern Enumeration Via a Generalized Cluster Method

Yan Zhuang[1*]

[1]*Department of Mathematics and Computer Science, Davidson College*

Abstract. We present a general approach for counting permutations by occurrences of prescribed consecutive patterns together with various inverse statistics. We first lift the Goulden–Jackson cluster method for permutations—a standard tool in the study of consecutive patterns—to the Malvenuto–Reutenauer algebra. Upon applying standard homomorphisms, our result specializes to both the cluster method for permutations as well as a q-analogue which keeps track of the inversion number statistic. We construct additional homomorphisms which lead to further specializations for keeping track of inverses of shuffle-compatible descent statistics; these include the inverse descent number, inverse peak number, and inverse left peak number. To illustrate this approach, we present new formulas that count permutations by occurrences of the monotone consecutive pattern $12\cdots m$ while also keeping track of these inverse statistics.

Keywords: permutation statistics, consecutive patterns, Goulden–Jackson cluster method, Malvenuto–Reutenauer algebra, shuffle-compatibility

1 Introduction

Let $\mathfrak{S}_n$ denote the symmetric group of permutations on the set $[n] := \{1, 2, \ldots, n\}$ (where $\mathfrak{S}_0$ consists of the empty permutation), and let $\mathfrak{S} := \bigsqcup_{n=0}^{\infty} \mathfrak{S}_n$. We write permutations in one-line notation — that is, $\pi = \pi_1\pi_2\cdots\pi_n$ — and the π_i are called *letters* of π.

For a sequence of distinct integers w, the *standardization* of w — denoted $\operatorname{std}(w)$ — is defined to be the permutation in $\mathfrak{S}$ obtained by replacing the smallest letter of w with 1, the second smallest with 2, and so on. Given $\pi \in \mathfrak{S}_n$ and $\sigma \in \mathfrak{S}_m$, we say that π *contains* σ (as a *consecutive pattern*) if $\operatorname{std}(\pi_i\pi_{i+1}\cdots\pi_{i+m-1}) = \sigma$ for some $i \in [n-m+1]$, and in this case we call $\pi_i\pi_{i+1}\cdots\pi_{i+m-1}$ an *occurrence* of σ (as a consecutive pattern) in π. For instance, the permutation 315497628 has three occurrences of the consecutive pattern 213, namely 315, 549, and 628. In contrast, 137258469 has no occurrences of 213.

Let $\operatorname{occ}_\sigma(\pi)$ denote the number of occurrences of σ in π. If $\operatorname{occ}_\sigma(\pi) = 0$, then we say that π *avoids* σ (as a consecutive pattern). If $\Gamma \subseteq \mathfrak{S}$, then we let $\mathfrak{S}_n(\Gamma)$ denote the subset of permutations in $\mathfrak{S}_n$ avoiding every permutation in Γ as a consecutive pattern. When Γ consists of a single pattern σ, we shall simply write $\mathfrak{S}_n(\sigma)$ as opposed to $\mathfrak{S}_n(\{\sigma\})$.

*yazhuang@davidson.edu. Supported by an AMS-Simons travel grant.

(We use the same convention for other notations involving a set Γ of patterns when Γ is a singleton.) As observed earlier, we have $137258469 \in \mathfrak{S}_9(213)$. In this extended abstract, the notions of occurrence and avoidance of patterns in permutations always refer to consecutive patterns unless otherwise stated.

The study of consecutive patterns in permutations, initiated by Elizalde and Noy [3] in 2003, extends the study of classical patterns in permutations. Consecutive patterns in permutations are analogous to consecutive subwords in words, where repetition of letters is allowed. In the latter realm, the cluster method of Goulden and Jackson provides a very general formula expressing the generating function for words by occurrences of prescribed subwords in terms of a "cluster generating function", which is easier to compute. In 2012, Elizalde and Noy [4] adapted the Goulden–Jackson cluster method to the setting of permutations, which allows one to count permutations by occurrences of prescribed consecutive patterns. Their adaptation of the cluster method has become a standard tool in the study of consecutive patterns; a notable recent development is a q-analogue [1] which also keeps track of the inversion number statistic.

This extended abstract is a summary of the recent paper [8], which presents a lifting of the Goulden–Jackson cluster method for permutations to the Malvenuto–Reutenauer Hopf algebra on permutations. By applying standard homomorphisms, we recover Elizalde and Noy's cluster method for permutations as well as its q-analogue as special cases of this generalized cluster method in the Malvenuto–Reutenauer algebra. We also construct other homomorphisms which lead to new specializations of our generalized cluster method that can be used to count permutations by occurrences of prescribed patterns while keeping track of other permutation statistics.

The permutation statistics that we shall consider are the "inverses" of several classical permutation statistics related to descents and peaks: the *descent number* des, the *major index* maj, the *comajor index* comaj, the *peak number* pk, and the *left peak number* lpk. We call $i \in [n-1]$ a *descent* of $\pi \in \mathfrak{S}_n$ if $\pi_i > \pi_{i+1}$. Then $\operatorname{des}(\pi)$ is defined to be the number of descents of π, and $\operatorname{maj}(\pi)$ the sum of all descents of π. In other words, if $\operatorname{Des}(\pi) := \{\, i \in [n-1] : i \text{ is a descent of } \pi \,\}$ is the *descent set* of π, then

$$\operatorname{des}(\pi) := |\operatorname{Des}(\pi)| \quad \text{and} \quad \operatorname{maj}(\pi) := \sum_{i \in \operatorname{Des}(\pi)} i.$$

The comajor index comaj is a variant of the major index maj, and is defined by

$$\operatorname{comaj}(\pi) := \sum_{i \in \operatorname{Des}(\pi)} (n-i) = n\operatorname{des}(\pi) - \operatorname{maj}(\pi). \tag{1.1}$$

Furthermore, we call $i \in \{2, 3, \ldots, n-1\}$ a *peak* of $\pi \in \mathfrak{S}_n$ if $\pi_{i-1} < \pi_i > \pi_{i+1}$; then $\operatorname{pk}(\pi)$ is defined to be the number of peaks of π. Lastly, we call $i \in [n-1]$ a *left peak* of $\pi \in \mathfrak{S}_n$ if i is a peak of $\pi \in \mathfrak{S}_n$, or if $i = 1$ and i is a descent of π; then $\operatorname{lpk}(\pi)$ is the number of left peaks of π. For example, if $\pi = 72163584$, then we have $\operatorname{Des}(\pi) = \{1, 2, 4, 7\}$, $\operatorname{des}(\pi) = 4$, $\operatorname{maj}(\pi) = 14$, $\operatorname{comaj}(\pi) = 18$, $\operatorname{pk}(\pi) = 2$, and $\operatorname{lpk}(\pi) = 3$.

Given a statistic st, we define its *inverse statistic* ist by $\mathrm{ist}(\pi) := \mathrm{st}(\pi^{-1})$. Continuing with our above example, the inverse of π is $\pi^{-1} = 32586417$, so we have $\mathrm{iDes}(\pi) = \{1, 4, 5, 6\}$, $\mathrm{ides}(\pi) = 4$, $\mathrm{imaj}(\pi) = 16$, $\mathrm{icomaj}(\pi) = 16$, $\mathrm{ipk}(\pi) = 1$, and $\mathrm{ilpk}(\pi) = 2$.

Let Γ be a set of consecutive patterns and $\mathrm{occ}_\Gamma(\pi)$ the number of occurrences in π of patterns in Γ. We will consider the polynomials

$$A^{(\mathrm{ides},\mathrm{imaj})}_{\Gamma,n}(s,t,q) := \sum_{\pi \in \mathfrak{S}_n} s^{\mathrm{occ}_\Gamma(\pi)} t^{\mathrm{ides}(\pi)+1} q^{\mathrm{imaj}(\pi)},$$

$$A^{(\mathrm{ides},\mathrm{icomaj})}_{\Gamma,n}(s,t,q) := \sum_{\pi \in \mathfrak{S}_n} s^{\mathrm{occ}_\Gamma(\pi)} t^{\mathrm{ides}(\pi)+1} q^{\mathrm{icomaj}(\pi)},$$

$$P^{\mathrm{ipk}}_{\Gamma,n}(s,t) := \sum_{\pi \in \mathfrak{S}_n} s^{\mathrm{occ}_\Gamma(\pi)} t^{\mathrm{ipk}(\pi)+1}, \text{ and } \quad P^{\mathrm{ilpk}}_{\Gamma,n}(s,t) := \sum_{\pi \in \mathfrak{S}_n} s^{\mathrm{occ}_\Gamma(\pi)} t^{\mathrm{ilpk}(\pi)}$$

where $n \geq 1$, and with each of these polynomials defined to be 1 when $n = 0$. These polynomials give the joint distribution of the occurrence statistic occ_Γ along with each of the statistics $(\mathrm{ides}, \mathrm{imaj})$, $(\mathrm{ides}, \mathrm{icomaj})$, ipk, and ilpk. Setting $s = 0$ in these polynomials gives the distributions of the corresponding statistics over the pattern avoidance class $\mathfrak{S}_n(\Gamma)$. For convenience, let us define $A^{(\mathrm{ides},\mathrm{imaj})}_{\Gamma,n}(t,q) := A^{(\mathrm{ides},\mathrm{imaj})}_{\Gamma,n}(0,t,q)$ and the polynomials $A^{(\mathrm{ides},\mathrm{icomaj})}_{\Gamma,n}(t,q)$, $P^{\mathrm{ipk}}_{\Gamma,n}(t)$, and $P^{\mathrm{ilpk}}_{\Gamma,n}(t)$ analogously.

We consider these inverse statistics because they are inverses of "shuffle-compatible" descent statistics. A permutation statistic st is shuffle-compatible if the distribution of st over the set of shuffles of two permutations π and σ depends only on $\mathrm{st}(\pi)$, $\mathrm{st}(\sigma)$, and the lengths of π and σ. If st is shuffle-compatible and is a "descent statistic", then st induces a quotient of the algebra QSym of quasisymmetric functions, denoted $\mathcal{A}_{\mathrm{st}}$. By composing the quotient map from QSym to $\mathcal{A}_{\mathrm{st}}$ with the canonical surjection from the Malvenuto–Reutenauer algebra to QSym, we obtain a homomorphism on the Malvenuto–Reutenauer algebra which can be used to count permutations by the corresponding inverse statistic. Applying these homomorphisms to our generalized cluster method yields specializations that refine by the statistics $(\mathrm{ides}, \mathrm{icomaj})$, ipk, and ilpk.[1]

The structure of this extended abstract is as follows. We begin in Section 2 by giving a brief expository account of Elizalde and Noy's cluster method for permutations, and then we lift their cluster method to the Malvenuto–Reutenauer algebra. In Section 3, we use the theory of shuffle-compatibility to construct additional homomorphisms which are then used to derive new specializations of our generalized cluster method. In Section 4, we present a few formulas obtained by applying the results in Section 3 to the monotone pattern $12\cdots m$. (We obtain formulas for some other patterns in the full version [8] of our paper.) We end in Section 5 with a brief description of forthcoming work, a real-rootedness conjecture, and directions for future work.

[1] We do not explicitly give a specialization for $(\mathrm{ides}, \mathrm{imaj})$, but one can be obtained using the one for $(\mathrm{ides}, \mathrm{icomaj})$ and the formula $\mathrm{imaj}(\pi) = n\,\mathrm{ides}(\pi) - \mathrm{icomaj}(\pi)$, which is equivalent to (1.1).

2 Permutations, clusters, and consecutive patterns

2.1 The Goulden–Jackson cluster method for permutations

Given a permutation $\pi \in \mathfrak{S}_n$ and a set $\Gamma \subseteq \mathfrak{S}$ of patterns, we say that (i, σ) is a *marked occurrence* of $\sigma \in \Gamma$ in π if π has an occurrence of σ starting at position i. Moreover, we say that (π, Δ) is a *marked permutation* on π (with respect to Γ) if Δ is any set of marked occurrences in π of permutations in Γ. To illustrate, suppose that $\Gamma = \{123, 231\}$. Then

$$(4\,3\,7\,8\,12\,2\,5\,9\,1\,6\,11\,10, \{(2, 123), (4, 231), (9, 123)\}), \tag{2.1}$$

is a marked permutation on $\pi = 4\,3\,7\,8\,12\,2\,5\,9\,1\,6\,11\,10$ with respect to Γ. Informally, we display a marked permutation (π, Δ) as the permutation π with the marked occurrences in Δ circled, so that (2.1) is displayed as

$$4\;(3\;7\;(8)\;12\;2)\;5\;9\;(1\;6\;11)\;10\,.$$

Given $\pi \in \mathfrak{S}_m$ and $\rho \in \mathfrak{S}_n$, we say that $\tau \in \mathfrak{S}_{m+n}$ is a *concatenation* of π and ρ if $\operatorname{std}(\tau_1 \cdots \tau_m) = \pi$ and $\operatorname{std}(\tau_{m+1} \cdots \tau_{m+n}) = \rho$. Similarly, if (π, Δ) and (ρ, Ξ) are two marked permutations (with respect to the same pattern set Γ), we say that (τ, Ω) is a *concatenation* of (π, Δ) and (ρ, Ξ) if (1) τ is a concatenation of π and ρ, (2) we have $(i, \sigma) \in \Omega$ for every $(i, \sigma) \in \Delta$, and (3) we have $(i + m, \sigma) \in \Omega$ for every $(i, \sigma) \in \Xi$. For example, (2.1) can be obtained by concatenating $(32568147, \{(2, 123), (4, 231)\})$ and $(1243, \{(1, 123)\})$, *i.e.*,

$$3\;(2\;5\;(6)\;8\;1)\;4\;7 \quad \text{and} \quad (1\;2\;4)\;3\,.$$

A marked permutation can be obtained as a concatenation of marked permutations in many ways, but there is a unique decomposition in terms of what are called "clusters". A marked permutation is called a *cluster* if it has length at least 2 and is not a concatenation of two nonempty marked permutations. (In particular, we will call a cluster with respect to Γ a Γ-*cluster*.) Every marked permutation with respect to Γ can be uniquely decomposed as a concatenation of a sequence of Γ-clusters and the permutation of length 1 (representing letters which are not part of a cluster.) For instance, (2.1) is a concatenation of 1, $(23451, \{(2, 123), (4, 231)\})$, 1, 1, $(123, \{(1, 123)\})$, and 1, that is,

$$1\,, \quad (2\;3\;(4)\;5\;1), \quad 1\,, \quad 1\,, \quad (1\;2\;3), \;\text{and} \quad 1\,.$$

This unique decomposition of marked permutations is the central idea of the cluster method, which reduces the problem of counting permutations by occurrences of prescribed consecutive patterns to that of counting clusters by marked occurrences.

Fix $\Gamma \subseteq \mathfrak{S}$. Let $C_{\Gamma,\pi}$ be the set of all Γ-clusters on π, and if $c \in C_{\Gamma,\pi}$, let $\mathrm{mk}_\Gamma(c)$ be the number of marked occurrences in c. Define

$$F_\Gamma(s,x) := \sum_{n=0}^{\infty} \sum_{\pi \in \mathfrak{S}_n} s^{\mathrm{occ}_\Gamma(\pi)} \frac{x^n}{n!} \quad \text{and} \quad R_\Gamma(s,x) := \sum_{n=0}^{\infty} \sum_{\pi \in \mathfrak{S}_n} \sum_{c \in C_{\Gamma,\pi}} s^{\mathrm{mk}_\Gamma(c)} \frac{x^n}{n!}.$$

Theorem 1 (Cluster method for permutations)**.** *Let $\Gamma \subseteq \mathfrak{S}$ be a set of permutations, each of length at least 2. Then $F_\Gamma(s,x) = (1 - x - R_\Gamma(s-1,x))^{-1}$.*

Elizalde and Noy give Theorem 1 in the special case where Γ consists of a single pattern [4, Theorem 1.1] and apply this result to several families of patterns, including "chain patterns" and "non-overlapping patterns". Later, we will explain how Theorem 1 can be recovered from our cluster method in the Malvenuto–Reutenauer algebra.

We say that $(i,j) \in [n]^2$ is an *inversion* of $\pi \in \mathfrak{S}_n$ if $i < j$ and $\pi_i > \pi_j$. Let $\mathrm{inv}(\pi)$ denote the number of inversions of π. Theorem 1 can be refined to keep track of the inversion number statistic. Define

$$F_\Gamma(s,q,x) := \sum_{n=0}^{\infty} \sum_{\pi \in \mathfrak{S}_n} s^{\mathrm{occ}_\Gamma(\pi)} q^{\mathrm{inv}(\pi)} \frac{x^n}{[n]_q!} \quad \text{and}$$
$$R_\Gamma(s,q,x) := \sum_{n=0}^{\infty} \sum_{\pi \in \mathfrak{S}_n} q^{\mathrm{inv}(\pi)} \sum_{c \in C_{\Gamma,\pi}} s^{\mathrm{mk}_\Gamma(c)} \frac{x^n}{[n]_q!}.$$

The following theorem is [1, Theorem 2.3], but for a set Γ of patterns rather than a single pattern σ. Like with Theorem 1, we will later recover this theorem as a specialization of our generalized cluster method.

Theorem 2 (q-Cluster method for permutations)**.** *Let $\Gamma \subseteq \mathfrak{S}$ be a set of permutations, each of length at least 2. Then $F_\Gamma(s,q,x) = (1 - x - R_\Gamma(s-1,q,x))^{-1}$.*

2.2 Lifting the cluster method to Malvenuto–Reutenauer

Let $\mathbb{Q}[\mathfrak{S}]$ denote the $\mathbb{Q}$-vector space with basis $\mathfrak{S}$. The *Malvenuto–Reutenauer algebra*, first defined in [6], is the $\mathbb{Q}$-algebra on $\mathbb{Q}[\mathfrak{S}]$ with the product

$$\pi \cdot \sigma = \sum_{\tau \in C(\pi,\sigma)} \tau$$

where $C(\pi,\sigma)$ is the set of concatenations of π and σ. The Malvenuto–Reutenauer algebra $\mathbb{Q}[\mathfrak{S}]$ is graded by the length of the permutation, and that its identity element is the empty permutation, which we denote ε. We also denote the permutation of length 1 by ι.

Given a set of permutations $\Gamma \subseteq \mathfrak{S}$, define

$$\widetilde{F}_\Gamma(s) := \sum_{\pi \in \mathfrak{S}} \pi s^{\mathrm{occ}_\Gamma(\pi)} \quad \text{and} \quad \widetilde{R}_\Gamma(s) := \sum_{\pi \in \mathfrak{S}} \pi \sum_{c \in C_{\Gamma,\pi}} s^{\mathrm{mk}_\Gamma(c)}.$$

Theorem 3 (Cluster method in Malvenuto–Reutenauer)**.** *Let $\Gamma \subseteq \mathfrak{S}$ be a set of permutations, each of length at least 2. Then $\widetilde{F}_\Gamma(s) = (\varepsilon - \iota - \widetilde{R}_\Gamma(s-1))^{-1}$.*

Proof. First, observe that

$$\widetilde{F}_\Gamma(s+1) = \sum_{\pi\in\mathfrak{S}} \pi(s+1)^{\mathrm{occ}_\Gamma(\pi)} = \sum_{\pi\in\mathfrak{S}} \pi \sum_{k=0}^{\infty} \binom{\mathrm{occ}_\Gamma(\pi)}{k} s^k = \sum_{\pi\in\mathfrak{S}} \pi \sum_{S\subseteq\Gamma_\pi} s^{|S|} \tag{2.2}$$

where Γ_π is the set of occurrences in π of patterns in Γ. In fact, (2.2) counts marked permutations weighted by the number of marked occurrences. Because marked permutations can be uniquely decomposed as a concatenation of a sequence of clusters and the permutation ι of length 1, we have $\widetilde{F}_\Gamma(s+1) = (\varepsilon - \iota - \widetilde{R}_\Gamma(s))^{-1}$. Replacing s with $s-1$ yields the desired result. □

There is another realization of the Malvenuto–Reutenauer algebra, given in [2], as a subalgebra **FQSym** of the formal power series algebra $\mathbb{Q}\langle\langle X_1, X_2, \ dots\rangle\rangle$ where $X_1, X_2, \dots$ are noncommuting variables. In our full paper [8], we actually state Theorem 3 in **FQSym** instead of $\mathbb{Q}[\mathfrak{S}]$; by identifying permutations with elements of $\mathbb{Q}\langle\langle X_1, X_2, \dots\rangle\rangle$, we can prove our generalized cluster method for permutations using (a noncommutative version) of the original Goulden–Jackson cluster method for words, thus giving a unified treatment of the two cluster methods. We choose to use $\mathbb{Q}[\mathfrak{S}]$ in this extended abstract to present an alternative (and simpler) approach.

We now demonstrate how Elizalde and Noy's cluster method for permutations, as well as its q-analogue, can be recovered from the cluster method in Malvenuto–Reutenauer. Define the maps $\Psi\colon \mathbb{Q}[\mathfrak{S}] \to \mathbb{Q}[[x]]$ and $\Psi_q\colon \mathbb{Q}[\mathfrak{S}] \to \mathbb{Q}[[q,x]]$ by

$$\Psi(\pi) := \frac{x^n}{n!} \qquad \text{and} \qquad \Psi_q(\pi) := q^{\mathrm{inv}(\pi)} \frac{x^n}{[n]_q!}.$$

It can be easily shown that Ψ and Ψ_q are $\mathbb{Q}$-algebra homomorphisms [8, Proposition 3.5]. Applying Ψ to Theorem 3 yields Theorem 1, and applying Ψ_q instead yields Theorem 2.

3 Applications of shuffle-compatibility

3.1 Shuffle-compatibility homomorphisms

The goal of this section is to define a new family of homomorphisms on the Malvenuto–Reutenauer algebra. These homomorphisms can be used to produce further specializations of Theorem 3 that refine by various inverse statistics. To do so, we take a brief detour through the theory of shuffle-compatibility.

In this section only, we will use the term "permutation" to refer more generally to sequences of distinct positive integers. Let $\mathfrak{P}_n$ denote the set of permutations of length

n, and let $\mathfrak{P} := \bigsqcup_{n=0}^{\infty} \mathfrak{P}_n$. Observe that any statistic st defined on permutations in $\mathfrak{S}$ can extended to $\mathfrak{P}$ by letting $\mathrm{st}(\pi) := \mathrm{st}(\mathrm{std}(\pi))$ for $\pi \in \mathfrak{P}$.

Every permutation in $\mathfrak{P}$ can be uniquely decomposed into a sequence of maximal increasing consecutive subsequences, which we call *increasing runs*. The *descent composition* of π, denoted $\mathrm{Comp}(\pi)$, is the composition whose parts are the lengths of the increasing runs of π in the order that they appear. For instance, the increasing runs of $\pi = 512749$ are 5, 127, and 49, so the descent composition of π is $\mathrm{Comp}(\pi) = (1, 3, 2)$.

If $\pi \in \mathfrak{P}_m$ and $\sigma \in \mathfrak{P}_n$ are *disjoint* — that is, if they have no letters in common — then we call $\tau \in \mathfrak{P}_{m+n}$ a *shuffle* of π and σ if both π and σ are subsequences of τ. The set of shuffles of π and σ is denoted $S(\pi, \sigma)$. For example, we have

$$S(31, 25) = \{3125, 3215, 3251, 2315, 2351, 2531\}.$$

A permutation statistic st is called *shuffle-compatible* if for any disjoint permutations π and σ, the multiset $\{\, \mathrm{st}(\tau) : \tau \in S(\pi, \sigma) \,\}$ depends only on $\mathrm{st}(\pi)$, $\mathrm{st}(\sigma)$, and the lengths of π and σ. In [5], Gessel and Zhuang develop a theory of shuffle-compatibility for *descent statistics*: statistics st such that $\mathrm{Comp}(\pi) = \mathrm{Comp}(\sigma)$ implies $\mathrm{st}(\pi) = \mathrm{st}(\sigma)$. The statistics des, maj, comaj, pk, and lpk are all examples of shuffle-compatible descent statistics.

If st is a descent statistic and if L is a composition, then we let $\mathrm{st}(L)$ denote the value of st on any permutation with descent composition L. Two compositions L and K are called st-*equivalent* if $\mathrm{st}(L) = \mathrm{st}(K)$ and if L and K are compositions of the same integer. The following is Theorem 4.3 of [5], and provides a necessary and sufficient condition for a descent statistic to be shuffle-compatible. In the statement of this theorem, QSym is the algebra of quasisymmetric functions and F_L is the fundamental quasisymmetric function associated with the composition L. (See [7, Section 7.19] for an introduction to quasisymmetric functions.)

Theorem 4. *A descent statistic* st *is shuffle-compatible if and only if there exists a* $\mathbb{Q}$*-algebra homomorphism* $\phi_{\mathrm{st}} \colon \mathrm{QSym} \to \mathcal{A}_{\mathrm{st}}$, *where* $\mathcal{A}_{\mathrm{st}}$ *is a* $\mathbb{Q}$*-algebra with basis* $\{u_\alpha\}$ *indexed by* st-*equivalence classes* α *of compositions, such that* $\phi_{\mathrm{st}}(F_L) = u_\alpha$ *whenever* $L \in \alpha$.

The algebra $\mathcal{A}_{\mathrm{st}}$ is called the *shuffle algebra* of st. Many of these algebras $\mathcal{A}_{\mathrm{st}}$ can be characterized as subalgebras of various formal power series algebras in which the multiplication is the *Hadamard product* $*$ in a variable t:

$$\Big(\sum_{n=0}^{\infty} a_n t^n \Big) * \Big(\sum_{n=0}^{\infty} b_n t^n \Big) := \sum_{n=0}^{\infty} a_n b_n t^n.$$

In our notation for formal power series algebras, we write $t*$ to indicate that multiplication is the Hadamard product in t. For example, $\mathbb{Q}[[t*, x]]$ is the $\mathbb{Q}$-algebra of formal power series in the variables t and x, where multiplication is ordinary multiplication in x but is the Hadamard product in t.

We now present our homomorphisms. Given $\pi \in \mathfrak{S}_n$, define $\Psi_{(\mathrm{ides},\mathrm{icomaj})}\colon \mathbb{Q}[\mathfrak{S}] \to \mathcal{A}_{(\mathrm{des},\mathrm{comaj})} \subseteq \mathbb{Q}[[t*,q,x]]$, $\Psi_{\mathrm{ipk}}\colon \mathbb{Q}[\mathfrak{S}] \to \mathcal{A}_{\mathrm{pk}} \subseteq \mathbb{Q}[[t*,x]]$, and $\Psi_{\mathrm{ilpk}}\colon \mathbb{Q}[\mathfrak{S}] \to \mathcal{A}_{\mathrm{lpk}} \subseteq \mathbb{Q}[[t*,x]]$ by

$$\Psi_{(\mathrm{ides},\mathrm{icomaj})}(\pi) := \begin{cases} \dfrac{t^{\mathrm{ides}(\pi)+1}q^{\mathrm{icomaj}(\pi)}}{\prod_{i=0}^{n}(1-tq^i)}x^n, & \text{if } n \geq 1, \\ 1/(1-t), & \text{if } n = 0, \end{cases}$$

$$\Psi_{\mathrm{ipk}}(\pi) := \begin{cases} \dfrac{2^{2\,\mathrm{ipk}(\pi)+1}t^{\mathrm{ipk}(\pi)+1}(1+t)^{n-2\,\mathrm{ipk}(\pi)-1}x^n}{(1-t)^{n+1}}, & \text{if } n \geq 1, \\ 1/(1-t), & \text{if } n = 0, \text{ and} \end{cases}$$

$$\Psi_{\mathrm{ilpk}}(\pi) := \frac{2^{2\,\mathrm{ilpk}(\pi)}t^{\mathrm{ilpk}(\pi)}(1+t)^{n-2\,\mathrm{ilpk}(\pi)}x^n}{(1-t)^{n+1}}.$$

Theorem 5. *The maps $\Psi_{(\mathrm{ides},\mathrm{icomaj})}$, Ψ_{ipk}, and Ψ_{ilpk} are $\mathbb{Q}$-algebra homomorphisms.*

Proof sketch. It is shown in [5] that $(\mathrm{des},\mathrm{comaj})$, pk, and lpk are all shuffle-compatible; let $\phi_{(\mathrm{des},\mathrm{comaj})}$, ϕ_{pk}, and ϕ_{lpk} be their corresponding homomorphisms on QSym (see Theorem 4). Let $\rho\colon \mathbb{Q}[\mathfrak{S}] \to \mathrm{QSym}$ denote the canonical surjection defined by $\rho(\pi) := F_{\mathrm{Comp}(\pi^{-1})}$. Then the maps $\Psi_{(\mathrm{ides},\mathrm{icomaj})}$, Ψ_{ipk}, and Ψ_{ilpk} are simply the result of composing the homomorphisms $\phi_{(\mathrm{des},\mathrm{comaj})}$, ϕ_{pk}, and ϕ_{lpk} with ρ. □

3.2 Further specializations of the generalized cluster method

Given a set $\Gamma \subseteq \mathfrak{S}$, let

$$R_{\Gamma,k}^{(\mathrm{ides},\mathrm{icomaj})}(s,t,q) := \sum_{\pi \in \mathfrak{S}_k} t^{\mathrm{ides}(\pi)+1}q^{\mathrm{icomaj}(\pi)} \sum_{c \in C_{\Gamma,\pi}} s^{\mathrm{mk}_\Gamma(c)},$$

which counts Γ-clusters of length k by the number of marked occurrences as well as the inverse descent number and inverse comajor index of the underlying permutation. We also define

$$R_{\Gamma,k}^{\mathrm{ipk}}(s,t) := \sum_{\pi \in \mathfrak{S}_k} t^{\mathrm{ipk}(\pi)+1} \sum_{c \in C_{\Gamma,\pi}} s^{\mathrm{mk}_\Gamma(c)} \quad \text{and} \quad R_{\Gamma,k}^{\mathrm{ilpk}}(s,t) := \sum_{\pi \in \mathfrak{S}_k} t^{\mathrm{ilpk}(\pi)} \sum_{c \in C_{\Gamma,\pi}} s^{\mathrm{mk}_\Gamma(c)}.$$

The following general formulas are obtained by applying the maps $\Psi_{(\mathrm{ides},\mathrm{icomaj})}$, Ψ_{ipk}, and Ψ_{ilpk} to Theorem 3; they allow us to compute the polynomials $A_{\Gamma,n}^{(\mathrm{ides},\mathrm{icomaj})}(s,t,q)$, $P_{\Gamma,n}^{\mathrm{ipk}}(s,t)$, and $P_{\Gamma,n}^{\mathrm{ilpk}}(s,t)$ — defined in the introduction of this extended abstract — from the cluster polynomials $R_{\Gamma,k}^{(\mathrm{ides},\mathrm{icomaj})}(s,t,q)$, $R_{\Gamma,k}^{\mathrm{ipk}}(s,t)$, and $R_{\Gamma,k}^{\mathrm{ilpk}}(s,t)$. Below, we write $f^{*\langle n\rangle}$ to mean the n-fold Hadamard product of f, that is, $f^{*\langle n\rangle} := \underbrace{f * f * \cdots * f}_{n \text{ copies of } f}$.

Theorem 6. *Let $\Gamma \subseteq \mathfrak{S}$ be a set of permutations, each of length at least 2. Then we have*

(a)
$$\sum_{n=0}^{\infty} \frac{A_{\Gamma,n}^{(\mathrm{ides},\mathrm{icomaj})}(s,t,q)}{\prod_{i=0}^{n}(1-tq^i)} x^n = \sum_{n=0}^{\infty} \left(\frac{tx}{(1-t)(1-tq)} + \sum_{k=2}^{\infty} R_{\Gamma,k}^{(\mathrm{ides},\mathrm{icomaj})}(s-1,t,q) \frac{x^k}{\prod_{i=0}^{k}(1-tq^i)} \right)^{*\langle n \rangle}$$

(b)
$$\frac{1}{1-t} + \frac{1+t}{2(1-t)} \sum_{n=1}^{\infty} P_{\Gamma,n}^{\mathrm{ipk}}(s,u) z^n = \sum_{n=0}^{\infty} \left(\frac{2tx}{(1-t)^2} + \frac{1+t}{2(1-t)} \sum_{k=2}^{\infty} R_{\Gamma,k}^{\mathrm{ipk}}(s-1,u) z^k \right)^{*\langle n \rangle}, \textit{ and}$$

(c)
$$\frac{1}{1-t} \sum_{n=0}^{\infty} P_{\Gamma,n}^{\mathrm{ilpk}}(s,u) z^n = \sum_{n=0}^{\infty} \left(\frac{z}{1-t} + \frac{1}{1-t} \sum_{k=2}^{\infty} R_{\Gamma,k}^{\mathrm{ilpk}}(s-1,u) z^k \right)^{*\langle n \rangle}$$

where $u = 4t/(1+t)^2$ and $z = (1+t)x/(1-t)$.

4 Results for the monotone patterns $12\cdots m$

Theorem 6 allows us to count permutations by occurrences of patterns from a set Γ — jointly with (ides, icomaj), ipk, or ilpk — if we can count Γ-clusters by the same inverse statistic, yet there is no straightforward way to count clusters by inverse statistics. As a matter of fact, the simpler problem of counting clusters (without keeping track of any statistic) is equivalent to counting linear extensions of a certain poset [4], which is itself difficult in general. Nonetheless, in our full paper [8], we apply Theorem 6 to two families of patterns for which counting clusters is straightforward: the monotone patterns $12\cdots m$ and $m\cdots 21$, as well as the transpositions $12\cdots(a-1)(a+1)a(a+2)(a+3)\cdots m$ where $m \geq 5$ and $2 \leq a \leq m-2$. Here, we only present our results for the patterns $12\cdots m$ and direct the interested reader to [8] for our complete results.

It can be shown using basic symmetries on permutations that $A_{12\cdots m,n}^{(\mathrm{ides},\mathrm{imaj})}(s,t,q) = A_{12\cdots m,n}^{(\mathrm{ides},\mathrm{icomaj})}(s,t,q)$ for all $m \geq 2$ and $n \geq 0$, so we choose to state part (a) of the next theorem in terms of the inverse major index.

Theorem 7. *Let $m \geq 2$. We have*

(a)
$$\sum_{n=0}^{\infty} \frac{A_{12\cdots m,n}^{(\mathrm{ides},\mathrm{imaj})}(t,q)}{\prod_{i=0}^{n}(1-tq^i)} x^n = 1 + \sum_{k=1}^{\infty} \left[\sum_{j=0}^{\infty} \left(\binom{k+jm-1}{k-1}_q x^{jm} - \binom{k+jm}{k-1}_q x^{jm+1} \right) \right]^{-1} t^k,$$

(b) $$\frac{1}{1-t}+\frac{1+t}{2(1-t)}\sum_{n=1}^{\infty}P^{\mathrm{ipk}}_{12\cdots m,n}(u)z^{n} = 1+\sum_{k=1}^{\infty}\left[1-2kx+\sum_{j=1}^{\infty}(c_{m,j,k}x^{jm}-c'_{m,j,k}x^{jm+1})\right]^{-1}t^{k}, \text{ and}$$

(c) $$\frac{1}{1-t}\sum_{n=0}^{\infty}P^{\mathrm{ilpk}}_{12\cdots m,n}(u)z^{n}=\sum_{k=0}^{\infty}\left[\sum_{j=0}^{\infty}(d_{m,j,k}x^{jm}-d'_{m,j,k}x^{jm+1})\right]^{-1}t^{k},$$

where $u=4t/(1+t)^2$, $z=(1+t)x/(1-t)$,

$$c_{m,j,k}=2\sum_{l=1}^{k}\binom{l+jm-1}{l-1}\binom{jm-1}{k-l}, \qquad c'_{m,j,k}=2\sum_{l=1}^{k}\binom{l+jm}{l-1}\binom{jm}{k-l},$$

$$d_{m,j,k}=\sum_{l=0}^{k}\binom{l+jm}{l}\binom{jm}{k-l}, \text{ and} \qquad d'_{m,j,k}=\sum_{l=0}^{k}\binom{l+jm+1}{l}\binom{jm+1}{k-l}.$$

In order to use Theorem 7(bc) to compute the polynomials $P^{\mathrm{ipk}}_{12\cdots m,n}(t)$ and $P^{\mathrm{ilpk}}_{12\cdots m,n}(t)$, one must "invert" the expression $u=4t/(1+t)^2$. Let us first replace the variable t with v, and u with t, to obtain $t=4v/(1+v)^2$. Then, solving $t=4v/(1+v)^2$ for v yields $v=2t^{-1}(1-\sqrt{1-t})-1$. Thus, for example, Theorem 7 (b) is equivalent to

$$\frac{1}{1-v}+\frac{1+v}{2(1-v)}\sum_{n=1}^{\infty}P^{\mathrm{ipk}}_{12\cdots m,n}(t)z^{n}=1+\sum_{k=1}^{\infty}\left[1-2kx+\sum_{j=1}^{\infty}(c_{m,j,k}x^{jm}-c'_{m,j,k}x^{jm+1})\right]^{-1}v^{k}.$$

With some additional algebraic manipulations, we get the formula

$$\sum_{n=1}^{\infty}P^{\mathrm{ipk}}_{12\cdots m,n}(t)x^{n}=\frac{2(1-v)}{1+v}\sum_{k=1}^{\infty}\left[1-2kw+\sum_{j=1}^{\infty}(c_{m,j,k}w^{jm}-c'_{m,j,k}w^{jm+1})\right]^{-1}v^{k}-\frac{2v}{1+v}$$

where $w=(1-t)x/(1+t)$ and v is the same as above; this formula can be used to compute the polynomials $P^{\mathrm{ipk}}_{12\cdots m,n}(t)$.

More generally, we can count permutations in $\mathfrak{S}_n$ by the number of occurrences of $12\cdots m$ along with each of these inverse statistics, rather than just counting permutations in $\mathfrak{S}_n(12\cdots m)$ by these inverse statistics. Unfortunately, we do not have a nice formula for the polynomials $A^{(\mathrm{ides},\mathrm{imaj})}_{12\cdots m,n}(s,t,q)$, but part (a) of the next theorem gives a formula for $A^{\mathrm{ides}}_{12\cdots m,n}(s,t)$, where $A^{\mathrm{ides}}_{\Gamma,n}(s,t)\coloneqq A^{(\mathrm{ides},\mathrm{imaj})}_{\Gamma,n}(s,t,1)$. Also, unlike in Theorem 7, all three of the formulas below involve the Hadamard product, but one can still use a computer algebra system (such as Maple) to compute the polynomials $A^{\mathrm{ides}}_{12\cdots m,n}(s,t)$, $P^{\mathrm{ipk}}_{12\cdots m,n}(s,t)$, and $P^{\mathrm{ilpk}}_{12\cdots m,n}(s,t)$ from these formulas.

Theorem 8. *Let $m \geq 2$. We have*

(a)
$$\sum_{n=0}^{\infty} \frac{A^{\mathrm{ides}}_{12\cdots m,n}(s,t)}{(1-t)^{n+1}} x^n = \sum_{n=0}^{\infty} \left(\frac{tx}{(1-t)^2} + \frac{(s-1)ty^m(y-1)}{(t-1)(y^m(s-1)-ys+1)} \right)^{*\langle n \rangle},$$

(b)
$$\frac{1}{1-t} + \frac{1+t}{2(1-t)} \sum_{n=1}^{\infty} P^{\mathrm{ipk}}_{12\cdots m,n}(s,u) z^n = \sum_{n=0}^{\infty} \left(\frac{2tx}{(1-t)^2} + \frac{2t(s-1)z^m(z-1)}{(t^2-1)(z^m(s-1)-zs+1)} \right)^{*\langle n \rangle}, \text{ and}$$

(c)
$$\frac{1}{1-t} \sum_{n=0}^{\infty} P^{\mathrm{ilpk}}_{12\cdots m,n}(s,u) z^n = \sum_{n=0}^{\infty} \left(\frac{z}{1-t} + \frac{(s-1)z^m(z-1)}{(t-1)(z^m(s-1)-zs+1)} \right)^{*\langle n \rangle}$$

where $y = x/(1-t)$, $u = 4t/(1+t)^2$, and $z = (1+t)x/(1-t)$.

5 Closing remarks

In our full paper [8], we apply our new specializations of the cluster method for permutations to the patterns $12\cdots m$ and $m\cdots 21$, as well as the patterns $12\cdots(a-1)(a+1)a(a+2)(a+3)\cdots m$ where $m \geq 5$ and $2 \leq a \leq m-2$. For these patterns, it is easy to count clusters by the inverse statistics that we consider. In particular, these patterns are "chain patterns". Elizalde and Noy [4] showed that counting clusters is equivalent to counting linear extensions in a certain poset, and the poset associated with a chain pattern is simply a chain. Moreover, clusters formed from any one of these patterns are involutions, so counting these clusters by ist is the same as counting them by st.

In ongoing work with Justin Troyka, we study the patterns $2134\cdots m$ and $12\cdots(m-2)m(m-1)$ for $m \geq 3$. Although these are not chain patterns and their clusters are not involutions, they are examples of "non-overlapping patterns", which can only overlap in one way. Both the non-overlapping condition and the condition of being a chain pattern greatly restrict how clusters can be formed, making them easier to classify and thus more amenable to study. As such, one direction of future work is to apply our results to other families of non-overlapping patterns and chain patterns.

The following conjecture is suggested by computational evidence.

Conjecture 9. *Let σ be $12\cdots m$ or $m\cdots 21$ where $m \geq 3$, or $12\cdots(a-1)(a+1)a(a+2)(a+3)\cdots m$ where $m \geq 5$ and $2 \leq a \leq m-2$. Then the polynomials $A^{\mathrm{ides}}_{\sigma,n}(t)$, $P^{\mathrm{ipk}}_{\sigma,n}(t)$, and $P^{\mathrm{ilpk}}_{\sigma,n}(t)$ have only real roots for all $n \geq 2$.*

In particular, Conjecture 9 would imply that — for all patterns σ listed above — the polynomials $A^{\mathrm{ides}}_{\sigma,n}(t)$, $P^{\mathrm{ipk}}_{\sigma,n}(t)$, and $P^{\mathrm{ilpk}}_{\sigma,n}(t)$ are unimodal and log-concave, and that the distributions of the statistics ides, ipk, and ilpk over $\mathfrak{S}_n(\sigma)$ are asymptotically normal.

We note that one can use the theory from [5] to define homomorphisms on the Malvenuto–Reutenauer algebra which can be used to count permutations by inverses of shuffle-compatible statistics other than the ones we consider here. In addition, it is worth investigating whether there are homomorphisms on the Malvenuto–Reutenauer algebra which can be used to count permutations by statistics other than the inversion number inv and inverses of shuffle-compatible descent statistics. Finally, by lifting other formulas in permutation enumeration to the Malvenuto–Reutenauer algebra, we can use our homomorphisms to produce new refinements of these formulas by inverse statistics.

Acknowledgments

The author thanks Justin Troyka for his extensive feedback on this project, Sergi Elizalde for helpful discussions, and an anonymous referee for offering suggestions that improved the exposition of this extended abstract. The author is also grateful to Ira Gessel, from whom he first learned about the Malvenuto–Reutenauer algebra and the Goulden–Jackson cluster method.

References

[1] H. Crane, S. DeSalvo, and S. Elizalde. "The probability of avoiding consecutive patterns in the Mallows distribution". *Random Structures Algorithms* **53**.3 (2018), pp. 417–447. DOI.

[2] G. Duchamp, F. Hivert, and J.-Y. Thibon. "Noncommutative symmetric functions. VI. Free quasi-symmetric functions and related algebras". *Internat. J. Algebra Comput.* **12**.5 (2002), pp. 671–717. DOI.

[3] S. Elizalde and M. Noy. "Consecutive patterns in permutations". *Adv. in Appl. Math.* **30**.1-2 (2003), pp. 110–125.

[4] S. Elizalde and M. Noy. "Clusters, generating functions and asymptotics for consecutive patterns in permutations". *Adv. in Appl. Math.* **49**.3-5 (2012), pp. 351–374.

[5] I. M. Gessel and Y. Zhuang. "Shuffle-compatible permutation statistics". *Adv. Math.* **332** (2018), pp. 85–141.

[6] C. Malvenuto and C. Reutenauer. "Duality between quasi-symmetric functions and the Solomon descent algebra". *J. Algebra* **177**.3 (1995), pp. 967–982.

[7] R. P. Stanley. *Enumerative Combinatorics, Vol. 2*. Cambridge University Press, 2001.

[8] Y. Zhuang. "A lifting of the Goulden–Jackson cluster method to the Malvenuto–Reutenauer algebra". 2021. arXiv:2108.10309.