Journal of Combinatorics
Volume 0, Number 0, 1–20, 0000

Moments of Matching Statistics

Niraj Khare, Rudolph Lorentz and Catherine H. Yan

We show that for a large family of combinatorial statistics on perfect matchings, the moments can be expressed as a linear combination of double factorials with constant coefficients. This gives a stronger analogous result of Chern, Diaconis, Kane and Rhoades on statistics of set partitions, in which case the moments can be expressed as linear combinations of shifted Bell numbers, but with polynomial coefficients.

Mathematics Subject Classification 05A18, 05A15, 62J05.

keyword: matching, statistics

1. Introduction

Recently Chern, Diaconis, Kane and Rhoades [8] found that for many combinatorial statistics on set partitions of $[n] = \{1, 2, \ldots, n\}$, the moments (mean, variance and higher moments) have simple closed expressions as linear combinations of shifted Bell numbers, where the coefficients in the linear combinations are polynomials in n. This allows one to derive exact formulas of the moments based on data for small values of n. In particular, the method applies to the number of blocks, the dimension index, the number of crossings, and the number of levels for set partitions. Combining with a stochastic algorithm of Stam [15] for generating a random set partitions, Chern et al. [9] established the limiting normality for the numbers of 2-crossings, dimension index, and the number of levels.

The main goal of the present paper is to study the analogous results of [8] on the set of perfect matchings on $[2m]$. A partial matching M on $[n]$ is a partition in which every block has at most two elements. If every block has exactly two elements, the matching is called a *perfect matching*. Denote by $\mathcal{M}_{2m}$ the set of all perfect matchings on $[2m]$. Then $\mathcal{M}_{2m}$ is a subset of $\Pi(2m)$, the set of all set partitions of $[2m]$. Hence the general approach in [8] applies to $\mathcal{M}_{2m}$. However, perfect matchings are special partitions with uniform block sizes. This allows us to get stronger characterizations, simpler formulas and more efficient algorithms. In particular, we will focus on the simple statistics which count the appearance of patterns in perfect matchings and present closed expressions for moments of crossings, nestings, and their variants.

Let us introduce necessary notations first. We represent a matching M on $[2m]$ by a diagram on the vertex set $[2m]$ with arcs (i, j) whenever $i < j$ and $\{i, j\}$ is a block of M. For instance, the matching $M = \{(1, 4), (2, 9), (3, 5), (6, 7), (8, 12), (10, 11)\}$ is represented by the diagram in Figure 1.

For a matching M, the set of openers, denoted by $\mathcal{O}(M)$, is the set of left endpoints of the arcs in the diagram of M. Similarly, the set of closers, denoted by $\mathcal{C}(M)$, is the set of right endpoints of the arcs. For the matching in Figure 1, we have $\mathcal{O}(M) = \{1, 2, 3, 6, 8, 10\}$ and $\mathcal{C}(M) = \{4, 5, 7, 9, 11, 12\}$.

The following definitions are adapted from the ones defined for set partitions in [8].

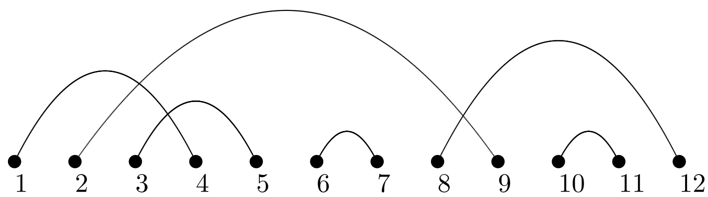

Figure 1: The diagram of the matching $M = \{(1,4),(2,9),(3,5),(6,7),(8,12),(10,11)\}$

Definition 1.1. *1. A pattern $\underline{P}$ of length k is a partial matching P on $[k]$ with a set of vertex-disjoint pairs $A(P) \subseteq \binom{[k]}{2}$ and a set $C(P) \subseteq [k-1]$. Let $\underline{P} := (P, A(P), C(P))$.*

2. An occurrence of a pattern $\underline{P}$ of length k in $M \in \mathcal{M}_{2m}$ is a tuple $s := (t_1, t_2, \cdots, t_k)$ with $t_i \in [2m]$ such that

(a) $t_1 < t_2 < \cdots < t_k$.

(b) (t_i, t_j) is an arc of M if $(i,j) \in A(P)$.

(c) $t_{i+1} = t_i + 1$ whenever $i \in C(P)$.

Write $s \in_{\underline{P}} M$ if s is an occurrence of $\underline{P}$ in M.

Definition 1.2. *A* simple statistic *is defined by a pattern $\underline{P}$ of length k and a valuation polynomial $Q \in \mathbb{Q}[y_1, y_2, \cdots, y_k, n]$. If $M \in \mathcal{M}_{2m}$ and $s = (x_1, x_2, \cdots, x_k) \in_{\underline{P}} M$, write $Q(s) = Q|_{y_i = x_i, n=m}$. Let*

$$f(M) = f_{\underline{P},Q}(M) := \sum_{s \in_{\underline{P}} M} Q(s).$$

Let the degree of a simple statistic $f_{P,Q}$, denoted $d(f)$, be the sum of the length of $\underline{P}$ and the degree of Q. A statistic *is a finite $\mathbb{Q}$-linear combination of simple statistics. The degree of a statistic is defined to be the minimum over such representations of the maximum degree of any appearing simple statistics.*

Remarks.

1. In the above definition, $C(P)$ defines a set of consecutive elements.
2. The field $\mathbb{Q}$ can be replaced by any field $\mathbb{K}$ of characteristic zero. We use $\mathbb{Q}$ here because it is the case in most combinatorial applications.
3. In [8] a pattern of set partitions also contains specified sets of first and last elements. This is not necessary for matchings since the set of arcs of P uniquely determines $\mathcal{O}(P)$ and $\mathcal{C}(P)$.

Examples.

1. *Number of arcs of M.* For any matching of $[2m]$, the number of arcs $|M|$ equals m. It is trivial to see that $|M|$ is a simple statistic with pattern P of length two along with $A(P) = \{(1,2)\}$, $C(P) = \emptyset$ and $Q(y_1, y_2, n) = 1$.
2. *Sum of vertex indices in $\mathcal{O}(M)$ or $\mathcal{C}(M)$.* For a matching $M \in \mathcal{M}_{2m}$ let

$$s_{min} = \sum_{i \in \mathcal{O}(M)} i \qquad \text{and} \qquad s_{max} = \sum_{i \in \mathcal{C}(M)} i.$$

Then both s_{min} and s_{max} are simple statistics with P of length 2 along with $A(P) = \{(1,2)\}$, $C(P) = \emptyset$, and $Q_{min}(y_1, y_2, n) = y_1$, $Q_{max}(y_1, y_2, n) = y_2$.

3. *k-crossings, k-nestings, and k-alignments.* Let $k \geq 2$ be an integer. A set of k arcs $\{(i_t, j_t) : 1 \leq t \leq k\}$ in the diagram of M forms a *k-crossing* if $i_1 < i_2 < \cdots < i_k < j_1 < j_2 < \cdots < j_k$. Similarly, it is a *$k$-nesting* if $i_1 < i_2 < \cdots < i_k < j_k < j_{k-1} < \cdots < j_1$, and a *$k$-alignment* if $i_1 < j_1 < i_2 < j_2 < \cdots < i_k < j_k$.
   Let cr_k, ne_k and al_k be the number of k-crossings, k-nestings, and k-alignments of M. They are simple statistics with patterns of length $2k$ and
$$\begin{aligned} A(P(cr_k)) &= \{(1, k+1), (2, k+2), \ldots, (k, 2k)\} \\ A(P(ne_k)) &= \{(1, 2k), (2, 2k-1), \ldots, (k, k+1)\} \\ A(P(al_k)) &= \{(1, 2), (3, 4), \ldots, (2k-1, 2k)\}. \end{aligned}$$
   For all of them, $Q = 1$ and $C(P) = \emptyset$.
4. Dimension exponent. The dimension exponent $d(\lambda)$ of a set partition λ arose from the study of the character theory of upper-triangular matrices. See [8, §3.3] and references there. Explicitly,
$$d(\lambda) = \sum_{i=1}^{l} (M_i - m_i + 1) - n,$$
   where l is the number of blocks of λ, M_i and m_i are the largest and smallest elements of the ith block. Specialized to perfect matchings, the dimension exponent is the statistic obtained by taking a linear combination of s_{max} and s_{min}, as
$$d(M) = s_{max}(M) - s_{min}(M) - m.$$
5. Blocks consisting of consecutive vertices. Another simple statistic is obtained by counting the number of blocks that consists of consecutive vertices, i.e., arcs of the form $(i, i+1)$. It is represented by the pattern P of length 2 with $A(P) = \{(1, 2)\}$ and $C(P) = \{1\}$, and the valuation polynomial $Q = 1$. Following [8] we denote this statistic by f_{level}.

Let T_{2m} be the number of matchings on $[2m]$, i.e., $T_{2m} = |\mathcal{M}_{2m}| = (2m-1)(2m-3)\cdots 3 \cdot 1 = (2m-1)!!$, the double factorial of $2m-1$. By convention, set $T_0 = 1$ and $T_{2i} = 0$ for $i < 0$. In the next section we give closed formulas for the aggregate $\sum_M f(M)$ over $\mathcal{M}_{2m}$. We show that for a general statistic f, the sum $\sum_M f(M)$ can be expressed as a linear combination of $\{T_{2i} : i \geq 0\}$ with finitely many i and constant coefficients. In Section 3 we present closed expressions of higher moments of statistics. Section 4 deals with simple statistics whose associated pattern is a perfect matching with empty $C(P)$, and whose valuation function is a constant. In that case the coefficients in the expression of higher moment can be obtained by a linear recurrence. As examples, we present results for k-crossing, k-nesting, and k-alignment. In the last two sections, we compute higher moments for statistics associated with crossings/nesting with consecutive left (right) endpoints, and with the dimension exponent.

2. Aggregate of matching statistics

It is proved in [8] that for any statistic f of set partitions, the aggregate $\sum_{\lambda \in \Pi(n)} f(\lambda)$ is a linear combination of the shifted Bell numbers with polynomial coefficients. We show that when restricted to matchings, the Bell numbers are replaced with the double factorials, and the coefficients are constants. Explicitly, for any statistic f, define

$$M(f,2m) := \sum_{M\in\mathcal{M}_{2m}} f(M).$$

Clearly the expected value of f for a uniform random matching in $\mathcal{M}_{2m}$ is given by $E(f) = M(f,2m)/T_{2m}$.

Our main results are the following two theorems which give two general expressions for $M(f,2m)$. One of the techniques used in proving the following results is the left compression of numbers corresponding to the set $C(\underline{P})$. It is interesting to note that a similar tool is commonly used for graphs and hypergraphs to modify the set system without changing the matching number (for instance see [1], [2], [5], [12] and [13]).

Theorem 2.1. *Let $f_{\underline{P},Q}$ be a simple statistic of degree N associated with pattern $\underline{P}$ and valuation polynomial $Q(s)$. Assume $\ell = |A(\underline{P})|$ and $c = |C(\underline{P})|$. Then*

$$M(f_{\underline{P},Q},2m) = P(m)T_{2(m-\ell)} \tag{1}$$

where $P(x)$ is a polynomial of degree no more than $N-c$. Equivalently for $m\geq \ell$, $M(f,2m)$ can be expressed as a linear combination of T_{2i}'s with constant coefficients, i.e.,

$$M(f_{\underline{P},Q},2m) = \begin{cases} 0 & m<\ell \\ \sum_{-\ell\leq i\leq N-\ell-c} c_i T_{2(m+i)} & m\geq \ell \end{cases} \tag{2}$$

with constants $c_i\in\mathbb{Q}$.

Proof. Assume that $f = f_{\underline{P},Q}$ where the pattern $\underline{P} = (P, A(P), C(P))$ is of length k, and the valuation polynomial Q is of degree $N-k$. Then

$$\begin{aligned} M(f_{\underline{P},Q},2m) = \sum_{M\in\mathcal{M}_{2m}} f_{\underline{P},Q}(M) &= \sum_{M\in\mathcal{M}_{2m}} \sum_{s\in_P M} Q(s) \\ &= \sum_{s\in\binom{[2m]}{k}} Q(s) \sum_{\substack{M\in\mathcal{M}_{2m}\\ s\in_P M}} 1. \end{aligned}$$

Fix an occurrence $s = (t_1,\cdots t_k)$ of P, the $k-2\ell$ singletons in s can be joint with vertices in $[2m]-s$ and form arbitrary matchings. So there are $T_{2(m-\ell)}$ perfect matchings on $[2m]$ that contain s. Hence s contributes $T_{2(m-\ell)}$ to the inner sum of $M(f_{\underline{P},Q},2m)$ whenever it satisfies the condition (c) of Definition 1. Otherwise it contributes 0. Therefore

$$M(f_{\underline{P},Q},2m) = T_{2(m-\ell)} \sum_{\substack{1\leq t_1<t_2<\cdots<t_k\leq 2m\\ t_{i+1}=t_i+1 \text{ for } i\in\mathcal{C}(\underline{P})}} Q(s).$$

To deal with the constraints caused by $\mathcal{C}(\underline{P})$, we use the standard trick to compress numbers, as did in [8]. We call $i+1$ a follower if $i\in\mathcal{C}(P)$. If j is the index of the i-th non-follower then let $y_i = t_j - j + i$. Then the values of $(t_1,\ldots,t_k)$ are determined by $(y_1,\ldots,y_{k-c})$, where $c = |\mathcal{C}(\underline{P})|$ and Q can be viewed as a polynomial of $y_1,\ldots,y_{k-c}$ and m. Hence

$$M(f_{\underline{P},Q},2m) = T_{2(m-\ell)} \sum_{1\leq y_1<y_2<\cdots<y_{k-c}\leq 2m-c} \tilde{Q}(y_1,\ldots,y_{k-c},m) \tag{3}$$

for some polynomial $\tilde{Q}$ of the same degree as Q. The summation yields a polynomial of m of degree at most $deg(Q) + k - c = N - c$.

To see Equation (2), let $g_i(m)$ be a polynomial of m defined by $g_i(m) = T_{2(m-\ell+i)}/T_{2(m-\ell)}$. Then g_i is of degree i, and hence $\{g_i(m)\}_{i=0}^{\infty}$ form a basis of $\mathbb{Q}[m]$. It follows that any polynomial of degree k can be written as a linear combination of $g_0(m), \ldots, g_k(m)$. This implies Equation (2). ∎

From Formula (3) we get the following simple form when Q is a constant.

Corollary 2.2. *Let f be a simple statistic with pattern $\underline{P}$ of length k and the valuation function $Q = q \in \mathbb{Q}$. Then*

$$M(f, 2m) = qT_{2(m-\ell)}\binom{2m-c}{k-c}$$

where $\ell = |A(P)|$ and $c = |C(P)|$.

We refer to Equation (1) as the polynomial form, which is a product of a polynomial and a T_{2i} for some i. Equation (2) is referred to as the linear form, which is a linear combination of T_{2i}'s with constant coefficients. For simple statistics both expresses contain $N-c+1$ undetermined coefficients. Consequently, we have a polynomial form and a linear form for $M(f, 2m)$ for an arbitrary statistic f.

Theorem 2.3. *For any statistic f of degree N, there is a positive integer $L \leq \frac{N}{2}$ such that for all $m \geq L$,*

$$M(f, 2m) = R(m)T_{2(m-L)}, \tag{4}$$

where $R(x)$ are polynomials of degree no more than $N + L$. Equivalently, we have the linear form

$$M(f, 2m) = \sum_{-L \leq i \leq N} d_i T_{2(m+i)} \qquad (m \geq L) \tag{5}$$

for some constants $d_i \in \mathbb{Q}$.

Proof. Assume that

$$f = \sum_{i=1}^{t} r_i f_{\underline{P}_i, Q_i}$$

with $r_i \in \mathbb{Q}$. Then

$$M(f, 2m) = \sum_{i=1}^{t} r_i M(f_{\underline{P}_i, Q_i}, 2m) = \sum_{i=1}^{t} r_i P_i(m) T_{2(m-\ell_i)},$$

where $\ell_i = |A(P_i)|$ and degree of $P_j(m)$ is no more than $deg(f_i) - c_i \leq N$ with $c_i = |C(P_i)|$. Combining likely terms of T_{2k} yields the equation

$$M(f, 2m) = \sum_{j=0}^{L} R_j(m) T_{2(m-j)},$$

where $R_j(m)$ is a polynomial of degree no more than N, and $L = \max(l_i) \le \frac{N}{2}$. Since $T_{2(s+k)} = p_k(s)T_{2s}$ (if $s \ge 0$) in which $p_k(s)$ is a polynomial of degree k, we get the polynomial form Equation (4) for $m \ge L$.

The linear form Equation (5) is obtained by expanding $R(m)$ under the basis $\{1, T_{2(m-L+1)}/T_{2(m-L)}, \ldots, T_{2(m+N)}/T_{2(m-L)}\}$. ∎

Theorem 2.3 allows us to compute the closed formula of $M(f, 2m)$ whenever we know the exactly values of $M(f, 2m)$ for a set of $L + N + 1$ values of $m \ge L$. To a specific statistic, usually we would use the combinatorial structure to get a better bound on the degree of the polynomial $R(m)$ in (4) , or equivalently, the number of terms in (5).

Example 2.1. *The statistics s_{min} and s_{max} are simple statistics of degree 3 with $\ell = 1$ and $c = 0$. By Theorem 2.1 and Formula (3),*

$$M(s_{max}, 2m) = T_{2(m-1)} \sum_{1 \le t_1 < t_2 \le 2m} t_2 = \frac{2m(2m+1)(2m-1)}{3} T_{2(m-1)} = \frac{1}{3} T_{2(m+2)} - T_{2(m+1)}.$$

Similarly

$$M(s_{min}, 2m) = T_{2(m-1)} \sum_{1 \le t_1 < t_2 \le 2m} t_1 = \binom{2m+1}{3} T_{2(m-1)} = \frac{1}{6} T_{2(m+2)} - \frac{1}{2} T_{2(m+1)}$$

Example 2.2. *Let f be the simple statistic associated to P of length 3 with $A(P) = \{(1, 3)\}$, and $C(P) = 2$. That is, an occurrence of $\underline{P}$ is an arc on non-consecutive vertices. Assume $Q(t_1, t_2, t_3, m) = t_3$. Then by (3)*

$$\begin{aligned} M(f, 2m) = T_{2(m-1)} \sum_{1 \le t_1 < t_2 \le 2m-1} (t_2 + 1) &= T_{2(m-1)} \left(2\binom{2m}{3} + \binom{2m-1}{2} \right) \\ &= \frac{(2m-2)(2m-1)(4m+3)}{6} T_{2(m-1)} \\ &= \frac{1}{3} T_{2(m+2)} - \frac{3}{2} T_{2(m+1)} - \frac{1}{2} T_{2m}. \end{aligned}$$

3. Higher Moments of Simple statistics

In order to compute higher moments, it is necessary to consider products of statistics. The next theorem establishes that $\mathbb{Q}$-linear combination of statistics and product of statistics are in fact statistics. This allows us to compute the higher moments $E(f^r)$, or equivalently, $M(f^r, 2m) = E(f^r)T_{2m}$ for any statistic f. Theorem 3.1, Definition 3.1 and Lemma 3.2 are analogues to the ones in [8], which can be proved in the same way and hence the proofs are skipped.

Theorem 3.1. *Let $\mathcal{S}$ be the set of all statistics thought of as functions $f : \cup_m \mathcal{M}_{2m} \to \mathbb{Q}$. Then $\mathcal{S}$ is closed under the operations of pointwise scaling, addition and multiplication. Thus, if f_1, $f_2 \in \mathcal{S}$ and $a \in \mathbb{Q}$, then there exist matching statistics g_a, g_+ and g_* so that for all matching M,*

$$\begin{aligned} af_1(M) &= g_a(M), \\ f_1(M) + f_2(M) &= g_+(M), \\ f_1(M)f_2(M) &= g_*(M). \end{aligned}$$

Furthermore, $d(g_a) \le d(f_1)$, $d(g_+) \le \max\{d(f_1), d(f_2)\}$ *and* $d(g_*) \le d(f_1) + d(f_2)$.

To use Theorem 3.1 we need a notion of *merge of two patterns.*

Definition 3.1. *Let* $\underline{P}_1$, $\underline{P}_2$ *and* $\underline{P}_3$ *be patterns of length* k_1, k_2 *and* k_3 *respectively. The pattern* $\underline{P}_3$ *is called a merge of* $\underline{P}_1$ *and* $\underline{P}_2$ *if there are two strictly increasing functions* $h_1 : [k_1] \to [k_3]$, $h_2 : [k_2] \to [k_3]$ *such that*
(1) $h_1[k_1] \cup h_2[k_2] = [k_3]$,
(2) $(i, j) \in A(P_3)$ *if and only if* $(i, j) = (h_1(i'), h_1(j'))$ *or* $(i, j) = (h_2(i'), h_2(j'))$ *for some* (i', j') *in* $A(P_1)$ *or* $A(P_2)$ *respectively,*
(3) $i \in C(P_3)$ *if and only if there exists either a* $j \in C(P_1)$ *with* $i = h_1(j)$ *and* $i + 1 = h_1(j + 1)$ *or a* $j' \in C(P_2)$ *with* $i = h_2(j')$ *and* $i + 1 = h_2(j' + 1)$.
A merge is denoted as $(h_1, h_2) : \underline{P}_1, \underline{P}_2 \to \underline{P}_3$. *Similarly one defines the merges* $(h_1, \dots, h_r) : \underline{P}_1, \cdots, \underline{P}_r \to \underline{P}$ *for any positive integer* $r \ge 2$.

Lemma 3.2. *Let* $\underline{P}_1$, $\underline{P}_2$ *be patterns. For any matching* M *there is a one-to-one correspondence:*

$$\{(s_1, s_2) : s_1 \in_{\underline{P}_1} M, s_2 \in_{\underline{P}_2} M\} \leftrightarrow \{\underline{P}_3, s_3 \in_{\underline{P}_3} M, \text{ and } (h_1, h_2) : \underline{P}_1, \underline{P}_2 \to \underline{P}_3\}. \tag{6}$$

The above results enable us to compute the aggregate for a product of statistics, as

$$\begin{aligned} M(f_{\underline{P}_1,Q_1} f_{\underline{P}_2,Q_2}, 2m) &= \sum_{M \in \mathcal{M}_{2m}} \sum_{P_3} \sum_{s_3 \in_{\underline{P}_3} M} \sum_{(h_1,h_2)} Q_1(h_1(s_1))Q_2(h_2(s_2)). \\ &= \sum_{P_3} M(f_{\underline{P}_3,\tilde{Q}}, 2m) \end{aligned}$$

where

$$\tilde{Q}(s_3) = \sum_{(h_1,h_2)} Q_1(h_1(s_1))Q_2(h_2(s_2)).$$

Consequently, for any statistic f and positive integer r, f^r can be written as a linear combination of simple statistics, and hence $M(f^r, 2m)$ can be expressed as a linear combination of T_{2k}'s with constant coefficients.

Theorem 3.3. *For any statistic* f *of degree* N *and positive integer* r*, we have*

$$M(f^r, 2m) = \sum_{I \le i \le J} d_i T_{2(m+i)} \qquad \text{whenever } m \ge I \tag{7}$$

where I *and* J *are constants bounded by* $I \ge -\frac{rN}{2}$ *and* $J \le rN$.

Proof. Assume

$$f = \sum_{i=1}^{t} r_i f_{\underline{P}_i, Q_i}$$

with $r_i \in \mathbb{Q}$. Let $\underline{\tilde{P}}$ be a merge $(h_1, h_2, \dots, h_r) : (\underline{P}_{i_1}, \underline{P}_{i_2}, \dots, \underline{P}_{i_r}) \to \underline{P}$ where each $\underline{P}_{i_j} \in \{\underline{P}_1, \dots, \underline{P}_t\}$, and $\tilde{Q}(s) = Q_{i_1}(h_1(s_1))Q_{i_2}(h_2(s_2)) \cdots Q_{i_r}(h_r(s_r))$. Assume $\underline{\tilde{P}} = (\tilde{P}, A(\tilde{P}), C(\tilde{P}))$ is of length $\tilde{k}$ with $\tilde{\ell}$ arcs and $\tilde{c} = |C(\tilde{P})|$. Let $f_{\underline{\tilde{P}},\tilde{Q}}$ be a simple statistic associated with the pattern $\underline{\tilde{P}}$ and polynomial $\tilde{Q}$. Then $M(f^r, 2m)$ is a sum of linear multiples of terms of the form $M(f_{\underline{\tilde{P}},\tilde{Q}}, 2m)$. By Theorem 2.3 each such a term can be expressed as a linear combination of $T_{2(m+i)}$'s with constant coefficients, where the lower bound of i is $-\frac{deg(f_{\tilde{\underline{P}},\tilde{Q}})}{2} \geq -\frac{rN}{2}$, and the upper bound is $deg(f_{\tilde{\underline{P}},\tilde{Q}}) \leq rN$. ∎

Corollary 3.4. *Let $f_{\underline{P},1}$ be simple statistic for which the pattern $\underline{P}$ is of length k with $\ell = |A(P)|$, $c = |C(P)|$, and the unit valuation function. Then we have*

$$M(f^r_{\underline{P},1}, 2m) = \sum_{-r\ell \leq i \leq rk-\ell-c} c_i T_{2(m+i)} \tag{8}$$

for some constants $c_i \in \mathbb{Q}$ and $m \geq r\ell$.

Proof. Since $Q = 1$, the degree of f is k. Let $\underline{\tilde{P}}$ be a merge of r copies of $\underline{P}$ with length $\tilde{k}$, $\tilde{\ell}$ arcs and $\tilde{c}$ consecutive pairs. Then $k \leq \tilde{k} \leq rk$, $\ell \leq \tilde{\ell} \leq r\ell$ and $c \leq \tilde{c} \leq rc$. Formula (8) is obtained by applying the linear form (2) to simple statistic $\tilde{P}$ and summing over all such $\tilde{P}$. ∎

In addition, there is a summation form for $M(f^r_{\underline{P},1}, 2m)$, which can be useful when the combinatorial structure is easy to analyze.

Proposition 3.5. *Assume the pattern $\underline{P}$ has length k with $\ell = |A(P)|$, $c = |C(P)|$ and $Q = 1$. Then*

$$M(f^r_{\underline{P},1}, 2m) = \sum_{\tilde{k},\tilde{\ell},\tilde{c}} c^{(r)}_{\tilde{k},\tilde{\ell},\tilde{c}} T_{2(m-\tilde{\ell})} \binom{2m-\tilde{c}}{\tilde{k}-\tilde{c}} \tag{9}$$

with nonnegative integer coefficients $c^{(r)}_{\tilde{k},\tilde{\ell},\tilde{c}}$, where $k \leq \tilde{k} \leq kr$, $\ell \leq \tilde{\ell} \leq \ell r$, and $c \leq \tilde{c} \leq cr$. The coefficient $c^{(r)}_{\tilde{k},\tilde{\ell},\tilde{c}}$ counts the number of ways to merge r copies of $\underline{P}$ to patterns with $\tilde{k}$ vertices, $\tilde{\ell}$ arcs and $\tilde{c}$ consecutive pairs of vertices.

The advantage of Formula (9) is that the coefficients have clear combinatorial meanings. In application, we can use some simple combinatorial constraints to limit the number of nonzero coefficients, and hence obtain a tighter bound on the number of undetermined coefficients when we transform the sum $M(f^r, 2m)$ to the polynomial form or the linear form. Then we could find the undetermined coefficients by using data of $M(f^r, 2m)$ with small values of m.

4. Patterns with constant valuation and empty $C(P)$

For simple statistics with constant valuation, empty $C(P)$, and $k = 2\ell$, the polynomial form of $M(f^r, 2m)$ gives a simple formula whose coefficients satisfy a linear recurrence. In the following we simply assume $Q = 1$, since for the case $Q = q \in \mathbb{Q}$, $M(f^r_{\underline{P},q}, 2m) = q^r M(f_{\underline{P},1}, 2m)$.

Theorem 4.1. *Let f be a simple statistic defined by a pattern $\underline{P}$ of length $k = 2\ell$ with the valuation function $Q = 1$, $\ell = |A(P)|$ and $C(P) = \emptyset$. Then the r-th moment of f can be written as*

$$M(f^r, 2m) = \sum_{i=0}^{(r-1)\ell} c_i^{(r)} \binom{2m}{2(\ell+i)} T_{2(m-\ell-i)}, \tag{10}$$

for some constants $c_i^{(r)}$.

Proof. For any merge $\underline{\tilde{P}}$ of r copies of $\underline{P}$, the pattern associated to $\underline{\tilde{P}}$ would have length $2\tilde{\ell}$ if it has $\tilde{\ell}$ arcs, and $C(\tilde{P}) = \emptyset$ always. Let $c_i^{(r)}$ be the number of such merges with $\tilde{\ell} = i + \ell$ arcs. Then they contribute $c_i^{(r)}\binom{2m}{2(\ell+i)}T_{2(m-\ell-i)}$ to the sum $M(f^r, 2m)$. Summing over all i from 0 to $(r-1)\ell$, we get the formula (10). ∎

Example We explain Theorem 4.1 by computing the second moment of cr_2, the number 2-crossings in a matching. Since $r = 2$ and $\ell = 2$, by (10) we obtain

$$M((cr_2(M))^2, 2m) = c_0\binom{2m}{4}T_{2(m-2)} + c_1\binom{2m}{6}T_{2(m-3)} + c_2\binom{2m}{8}T_{2(m-4)},$$

where c_i is the number of patterns P_3 with $i+2$ arcs that can be obtained as merges of two copies $P_1 = P_2 = \underline{P}(cr_2)$.

- For $i = 0$, there is only one possible merge, namely $P_1 = P_2 = P_3$. Hence $c_0 = 1$.
- For $i = 1$, assume P_3 is a pattern of length 6 obtained by merging P_1 and P_2. First observe that if P_3 has two 2-crossings, then there are exactly two ways to define h_1 and h_2. If P_3 has three 2-crossings, then there are 6 ways to define h_1 and h_2. There are three P_3 with $cr_2(P_3) = 2$, namely, $\{(1,3),(2,5),(4,6)\}$, $\{(1,4),(2,6),(3,5)\}$, and $\{(1,5),(2,4),(3,6)\}$, and one with $cr_2(P_3) = 3$, namely, $\{(1,4),(2,5),(3,6)\}$. Putting together we have $c_1 = 12$.
- For $i = 2$, note that if a merge $(h_1, h_2) : P_1, P_2 \to P_3$ has 4 arcs, then $h_1(P_1)$ and $h_2(P_2)$ must be disjoint. This gives $c_2 = \binom{8}{4} = 70$.

Combining the above cases, we have

$$M((cr_2(M))^2, 2m) = \binom{2m}{4}T_{2m-4} + 12\binom{2m}{6}T_{2m-6} + 70\binom{2m}{8}T_{2m-8}. \tag{11}$$

Even for simple statistics as described in Theorem 4.1, getting the coefficients $c_i^{(r)}$'s by analyzing combinatorial structures can be tedious and time consuming. Instead, we could take advantage of the special form of $M(f^r, 2m)$ and the fact that for any positive integers $\binom{a}{b} = 0$ if $b > a$ (or by defining $T_{2k} = 0$ whenever $k < 0$). Explicitly, for $\ell \le m \le \ell r$, Equation (10) becomes

$$\begin{aligned}
M(f^r, 2\ell) &= c_0^{(r)}T_0 \\
M(f^r, 2\ell+2) &= c_0^{(r)}\binom{2\ell+2}{2\ell}T_2 + c_1^{(r)}T_0, \\
M(f^r, 2\ell+4) &= c_0^{(r)}\binom{2\ell+4}{2\ell}T_4 + c_1^{(r)}\binom{2\ell+4}{2\ell+2}T_2 + c_2^{(r)}T_0 \\
\cdots &\quad \cdots\cdots \\
M(f^r, 2\ell r) &= c_0^{(r)}\binom{2\ell r}{2\ell}T_{2(r-1)\ell} + c_1^{(r)}\binom{2\ell r}{2\ell+2}T_{2(r-1)\ell-2} + \cdots + c_{(r-1)\ell}^{(r)}T_0.
\end{aligned}$$

This is a triangular system of linear equations on the unknowns $c_0^{(r)}, \ldots, c_{(r-1)\ell}^{(r)}$, with $T_0 = 1$ on the diagonals. Hence the values of $M(f^r, 2m)$ for $\ell \le m \le \ell r$ determine all the coefficients by the following recurrence of $c_i^{(r)}$.

Proposition 4.2. *Let $f_{\underline{P},1}$ be a simple statistic with unit valuation function. Assume the pattern $\underline{P}$ has length 2ℓ with ℓ arcs and empty $C(P)$. If $M(f^r, 2m)$ is known for $\ell \leq m \leq \ell r$, then the coefficients $\{c_i^{(r)} : 0 \leq i \leq (r-1)\ell\}$ can be obtained by the linear recurrence*

$$\begin{aligned} c_0^{(r)} &= M(f^r, 2\ell) \\ c_i^{(r)} &= M(f^r, 2(\ell+i)) - \sum_{j=0}^{i-1} c_j^{(r)} \binom{2(\ell+i)}{2(\ell+j)} T_{2(i-j)} \qquad \textit{for } 0 < i \leq (r-1)\ell. \end{aligned}$$

For example, with the aid of a computer program we can easily get $M((cr_2(M))^2, 4) = 1$, $M((cr_2(M))^2, 6) = 27$, and $M((cr_2(M))^2, 8) = 616$, which give the coefficients in Equation (11).

Using Proposition 4.2, we provide explicit formulas of the second and third moments of k-crossings, k-nestings, and k-alignments for some small values of k. (The first moment is given by Corollary 2.2.). Note that the statistics $cr_2(M)$ and $ne_2(M)$ have the same distribution [10] over $\mathcal{M}_{2m}$. In fact, there is a bijection ϕ on $\mathcal{M}_{2m}$ such that $cr_2(M) = ne_2(\phi(M))$, e.g., see [14]. Thus, for any positive r, $M((cr_2)^r, 2m) = M((ne_2)^r, 2m)$.

<u>The second moment of cr_k for $2 \leq k \leq 5$.</u>

$$M((cr_2(M))^2, 2m) = M((ne_2(M))^2, 2m) = \binom{2m}{4} T_{2m-4} + 12\binom{2m}{6} T_{2m-6} + 70\binom{2m}{8} T_{2m-8}.$$

$$\begin{aligned} M((cr_3(M))^2, 2m) &= M((ne_3(M))^2, 2m) \\ &= \binom{2m}{6} T_{2m-6} + 20\binom{2m}{8} T_{2m-8} + 180\binom{2m}{10} T_{2m-10} + 924\binom{2m}{12} T_{2m-12}. \end{aligned}$$

$$\begin{aligned} M((cr_4(M))^2, 2m) &= M((ne_4(M))^2, 2m) \\ &= \binom{2m}{8} T_{2m-8} + 30\binom{2m}{10} T_{2m-10} + 378\binom{2m}{12} T_{2m-12} \\ &\qquad + 2800\binom{2m}{14} T_{2m-14} + 12870\binom{2m}{16} T_{2m-16}. \end{aligned}$$

$$\begin{aligned} M((cr_5(M))^2, 2m) &= M((ne_5(M))^2, 2m) \\ &= \binom{2m}{10} T_{2m-10} + 42\binom{2m}{12} T_{2m-12} + 700\binom{2m}{14} T_{2m-14} + 6864\binom{2m}{16} T_{2m-16} \\ &\qquad + 44100\binom{2m}{18} T_{2m-18} + 184756\binom{2m}{20} T_{2m-20}. \end{aligned}$$

<u>The third moment of cr_2, cr_3 and ne_3.</u>

$$\begin{aligned} M((cr_2(M))^3, 2m) &= M((ne_2(M))^3, 2m) \\ &= \binom{2m}{4} T_{2m-4} + 42 \binom{2m}{6} T_{2m-6} + 762 \binom{2m}{8} T_{2m-8} \\ &\qquad + 7560 \binom{2m}{10} T_{2m-10} + 34650 \binom{2m}{12} T_{2m-12}. \end{aligned}$$

$$\begin{aligned} M((cr_3(M))^3, 2m) = &\binom{2m}{6} T_{2m-6} + 84 \binom{2m}{8} T_{2m-8} + 2520 \binom{2m}{10} T_{2m-10} + 45372 \binom{2m}{12} T_{2m-12} \\ &+ 552636 \binom{2m}{14} T_{2m-14} + 4324320 \binom{2m}{16} T_{2m-16} + 17153136 \binom{2m}{18} T_{2m-18}, \end{aligned}$$

and

$$\begin{aligned} M((ne_3(M))^3, 2m) = &\binom{2m}{6} T_{2m-6} + 84 \binom{2m}{8} T_{2m-8} + 2520 \binom{2m}{10} T_{2m-10} + 45468 \binom{2m}{12} T_{2m-12} \\ &+ 552960 \binom{2m}{14} T_{2m-14} + 4324320 \binom{2m}{16} T_{2m-16} + 17153136 \binom{2m}{18} T_{2m-18}. \end{aligned}$$

The small moments of al_2, al_3.

$$M((al_2(M))^2, 2m) = \binom{2m}{4} T_{2m-4} + 14 \binom{2m}{6} T_{2m-6} + 70 \binom{2m}{8} T_{2m-8}.$$

$$\begin{aligned} M((al_2(M))^3, 2m) = &\binom{2m}{4} T_{2m-4} + 48 \binom{2m}{6} T_{2m-6} + 930 \binom{2m}{8} T_{2m-8} \\ &\qquad + 8820 \binom{2m}{10} T_{2m-10} + 34650 \binom{2m}{12} T_{2m-12}. \end{aligned}$$

$$M((al_3(M))^2, 2m) = \binom{2m}{6} T_{2m-6} + 24 \binom{2m}{8} T_{2m-8} + 238 \binom{2m}{10} T_{2m-10} + 924 \binom{2m}{12} T_{2m-12}.$$

We observe from the above formulas that the third moments of cr_3 and ne_3 are different. Hence the numbers of 3-crossings and 3-nestings have different distributions. On the other hand, the second moments of k-crossings and k-nestings coincide for $2 \le k \le 5$. The next theorem establishes that the second moment of k-crossings and k-nestings for matchings in $\mathcal{M}_{2m}$ are always the same.

Theorem 4.3. *For any positive integer $k \ge 2$, the second moment of k-crossings equals the second moment of k-nestings over the set $\mathcal{M}_{2m}$.*

We need the following two lemmas.

Lemma 4.4. *Let m and n be non-negative integers. Let x_0, x, y_0 and y be non-negative integers. Then the following identity holds:*

$$\sum_{\substack{x_0+x=m\\ y_0+y=n}} \binom{x_0+y_0}{x_0}\binom{x+y}{x}\binom{x_0+x+y_0+y}{x+x_0} = (m+n+1)\binom{m+n}{m}^2. \tag{12}$$

Proof. We use the identity

$$\binom{x+y}{x}\binom{m+n-x-y}{m-x} = \frac{(x+y)!(m+n-x-y)!}{x!y!(m-x)!(n-y)!} = \binom{m+n}{m}\binom{m}{x}\binom{n}{y}\Big/\binom{m+n}{x+y}.$$

The left side of Equation (12) equals

$$\begin{aligned}
&\binom{n+m}{m}\sum_{x=0}^{m}\sum_{y=0}^{n}\binom{x+y}{x}\binom{m+n-x-y}{m-x}\\
=\ &\binom{m+n}{m}^2\sum_{x=0}^{m}\sum_{y=0}^{n}\binom{m}{x}\binom{n}{y}\Big/\binom{m+n}{x+y}\\
=\ &\binom{m+n}{m}^2\sum_{k=0}^{m+n}\frac{1}{\binom{m+n}{k}}\sum_{x+y=k}\binom{m}{x}\binom{n}{y}\\
=\ &\binom{m+n}{m}^2\sum_{k=0}^{m+n}\frac{1}{\binom{m+n}{k}}\cdot\binom{m+n}{k}\\
=\ &(m+n+1)\binom{m+n}{m}^2.
\end{aligned}$$

∎

Lemma 4.5. *Let m and n be non-negative integers. Let x_0, x, y_0 and y be non-negative integers. Then the following identity holds:*

$$\sum_{\substack{x_0+x=m\\ y_0+y=n}} \binom{x_0+y_0}{x_0}^2\binom{2x+2y}{2x} = (m+n+1)\binom{m+n}{m}^2. \tag{13}$$

Proof. The above identity follows immediately from a slightly general identity, proved by Andrews and Paule in [3, Identity (2.2)] using computer algebra:

$$\sum_{i=0}^{\lfloor M/2\rfloor}\sum_{j=0}^{\lfloor N/2\rfloor}\binom{i+j}{j}^2\binom{M+N-2i-2j}{N-2j} = \frac{(\lfloor\frac{M+N+1}{2}\rfloor)!(\lfloor\frac{M+N+2}{2}\rfloor)!}{(\lfloor\frac{M}{2}\rfloor)!(\lfloor\frac{M+1}{2}\rfloor)!(\lfloor\frac{N}{2}\rfloor)!(\lfloor\frac{N+1}{2}\rfloor)!}.$$

One simply lets $M = 2m$ and $N = 2n$ to get (13). ∎

Proof of Theorem 4.3. By Theorem 4.1 it is sufficient to show that the number of ways to merge two k-crossings is the same as the number of ways to merge two k-nestings to matchings with exactly $2k-i$ arcs. We would give explicit formulas for such numbers and compare them.

(1) Merging two k-crossings A and B to get matchings with $2k-i$ arcs. It means that there are i arcs from A that coincide with i arcs from B. Listing the arcs by their left endpoints from left to right. Assume that M is the matching obtained by merging A and B, where the arcs $e_1 < e_2 < \cdots < e_i$ of A coincide with the arcs $f_1 < f_2 < \cdots < f_i$ of B, and they correspond to arcs $m_1 < m_2 < \cdots < m_i$ of M. In the k-crossing A, let $x_0 = e_1 - 1$, $x_j = e_{j+1} - e_j - 1$ for $j = 1, \dots, i-1$ and $x_i = k - e_i$. Then $x_i \in \mathbb{N}$ and $x_0 + x_1 + \cdots + x_i = k - i$. Similarly define $y_0, y_1, \dots, y_i$ for the k-crossing B. We have the following observation in M.

1. Before the left endpoint of arc m_1, there are $x_0 + y_0$ left endpoints, where x_0 coming from A and y_0 coming from B.
2. For each $j = 1, \dots, i-1$, between the left endpoints of arcs m_j and m_{j+1} there are $x_j + y_j$ left-endpoints, where x_j coming from A and y_j coming from B.
3. Between the left endpoint of m_i and the right endpoint of m_1, there are x_i left endpoints and x_0 right endpoints coming from A, and y_i left endpoints and y_0 right endpoints coming from B.
4. For each $j = 1, \dots, i-1$, between the right endpoints of m_j and m_{j+1} there are x_j right endpoints coming from A and y_j right endpoints coming from B.
5. After the right endpoint of m_i, there are x_i right endpoints coming from A and y_i right endpoints coming from B.

Then the number of ways to merge A and B to matchings with $2k-i$ arcs is given by

$$\begin{aligned} c_{k,2}^{cr} &= \sum_{\substack{x_0+\cdot+x_i=k-i \\ y_0+\cdots+y_i=k-i}} \prod_{j=0}^{i-1} \binom{x_j+y_j}{y_j} \cdot \binom{x_i+x_0+y_i+y_0}{x_i+x_0} \cdot \prod_{j=1}^{i} \binom{x_i+y_i}{x_i} \\ &= \sum_{\substack{x_0+\cdot+x_i=k-i \\ y_0+\cdots+y_i=k-i}} \prod_{j=1}^{i-1} \binom{x_j+y_j}{x_j}^2 \cdot \binom{x_0+y_0}{x_0} \binom{x_i+y_i}{x_i} \binom{x_i+x_0+y_i+y_0}{x_i+x_0}. \end{aligned} \tag{14}$$

(2) Merging two k-nestings to get matchings with $2k-i$ arcs. Similar to the argument above, we have the formula

$$\begin{aligned} c_{k,2}^{ne} &= \sum_{\substack{x_0+\cdot+x_i=k-i \\ y_0+\cdots+y_i=k-i}} \prod_{j=0}^{i-1} \binom{x_j+y_j}{y_j} \cdot \binom{2x_i+2y_i}{2x_i} \cdot \prod_{j=0}^{i-1} \binom{x_i+y_i}{x_i} \\ &= \sum_{\substack{x_0+\cdot+x_i=k-i \\ y_0+\cdots+y_i=k-i}} \prod_{j=0}^{i-1} \binom{x_j+y_j}{x_j}^2 \cdot \binom{2x_i+2y_i}{2x_i}. \end{aligned} \tag{15}$$

Comparing formulas (14) and (15), we see that it is sufficient to show that the following two sums are equal for any nonnegative integers m and n:

$$\sum_{\substack{x_0+x_i=m \\ y_0+y_i=n}} \binom{x_0+y_0}{x_0} \binom{x_i+y_i}{x_i} \binom{x_0+x_i+y_0+y_i}{x_i+x_0} = \sum_{\substack{x_0+x_i=m \\ y_0+y_i=n}} \binom{x_0+y_0}{x_0}^2 \binom{2x_i+2y_i}{2x_i}. \tag{16}$$

This equation follows from Lemma 4.4 and Lemma 4.5. Summing over all possible values of $x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}$, we obtain the equation $c_{k,2}^{cr} = c_{k,2}^{ne}$. ∎

5. Crossings and nesting with neighboring vertices

For simple statistics with $Q = 1$ but nonempty $C(P)$, we can use either the linear form (8) or the polynomial form (9) to compute $M(f^r, 2m)$. In general the linear form is simpler. But in certain cases we can use the combinatorial properties of the pattern to reduce the number of unknown coefficients in the polynomial form.

Example 5.1. *Consider the level statistic, $f_{level}(M)$, which counts the number of blocks of M that consist of two consecutive integers. The pattern $\underline{P}$ is given by a matching P of length 2 with $A(P) = (1, 2)$, $C(P) = \{1\}$, and $Q = 1$. Any merge of r copies of $\underline{P}$ must contain arcs of the form $(i, i+1)$ only. For fixed k lying between 1 and r, there is a unique such pattern with k arcs, i.e. the alignment of k arcs, and the number of merges can be described as the number of surjective maps from $[r]$ to $[k]$ and is given by $S(r, k)k!$, where $S(r, k)$ is the Stirling number of the second kind and counts the number of partitions of an n-set into k blocks. Hence*

$$M((f_{level}(M))^r, 2m) = \sum_{k=1}^{r} k!S(r,k)\binom{2m-k}{k}T_{2(m-k)}.$$

Our next example is on k-crossings and k-nestings with consecutive left or right endpoints. Nestings with consecutive left endpoints, called *neighboring nestings*, were introduced by Stoimenow [17] in the study of regular linearized chord diagrams, and matchings with no neighboring nestings were further investigated in [4, 6, 11]. In [7] Chen, Fan, and Zhao presented generating functions for partial matchings with no neighboring alignments or neighboring nestings. Here we consider simple statistics that counts the occurrences of neighboring crossings/nestings.

Definition 5.1. *The pattern* left-k-crossing, *denoted by $\underline{P}(Lcr_k)$, is the matching P of length $2k$ with k arcs, defined by*

$$A(P(Lcr_k)) = \{(1, k+1), (2, k+2), \dots, (k, 2k)\},$$

with $C(P(Lcr_k)) = \{1, 2, \dots, k-1\}$. The pattern right-$k$-crossing, *$\underline{P}(Rcr_k)$, is the matching of length $2k$ with $A(P(Rcr_k)) = A(P(Lcr_k))$ and $C(P(Rcr_k)) = \{k+1, k+2, \dots, 2k-1\}$. Similarly, the* left-$k$-nesting *and* right-$k$-nesting *are matchings of length $2K$ with*

$$A(P(Lne_k)) = A(P(Rne_k)) = \{(1, 2k), (2, 2k-1), \dots, (k, k+1)\}$$

and $C(P(Lne_k)) = \{1, 2, \dots, k-1\}$, $C(P(Rne_k)) = \{k+1, k+2, \dots, 2k-1\}$.

Let Lcr_k, Rcr_k, Lne_k and Rne_k be the simple statistics that count the numbers of left-k-crossings, right-k-crossings, left-k-nestings, and right-k-nestings, respectively. In other words, they all have the unit valuation.

Reversing the diagram of a matching (i.e., reflecting through a vertical mirror placed at the right of the diagram), the left and right endpoints are exchanged. Hence Lcr_k and Rcr_k have the same distribution over $\mathcal{M}_{2m}$, so do Lne_k and Rne_k.

Proposition 5.1. *There is an involution ϕ on $\mathcal{M}_{2m}$ such that $Rcr_k(M) = Rne_k(\phi(M))$.*

Proof. In [14, Theorem 1.2] it is proved that there is an involution $\phi : \Pi_n \to \Pi_n$ exchanging the numbers of 2-crossings and 2-netsings. We will use the same map ϕ but restricted to the set $\mathcal{M}_{2m}$.

In general this map ϕ does not exchange the number of k-crossings to that of k-nestings for $k \geq 3$. However, we show that ϕ exchanges the numbers of right-k-crossings and right-k-nestings for all positive integers $k \geq 2$. For completeness, we describe the construction of ϕ for perfect matchings.

First every matching M in $\mathcal{M}_{2m}$ can be uniquely represented as a Dyck path of length $2m$ with labeled down steps. To wit, one replaces each $i \in \mathcal{O}(M)$ with an up step $U = (1,1)$, and each $i \in \mathcal{C}(P)$ with a down step $D = (1,-1)$. This defines a Dyck path from $(0,0)$ to $(2m,0)$. Let $\mathcal{C}(P) = \{j_1, j_2, \ldots, j_m\}_<$. The height of the k-th down step is h_k if it is a step from $(j_k - 1, h_k)$ to $(j_k, h_k - 1)$. Assume that (i_k, j_k) is an arc of M. Label the k-th down step by γ_k if

$$\gamma_k - 1 = |\{i \in \mathcal{O}(M) : i < i_k,\ (i,j) \in M,\ \text{and } j > j_k\}|.$$

Then $1 \leq \gamma_k \leq h_k$ for all k. In addition, $cr_2(M) = \sum_{i=1}^{m}(h_k - \gamma_k)$ and $ne_2(M) = \sum_{i=1}^{k}(\gamma_k - 1)$. For example, for the matching M in Figure 1, the corresponding Dyck path is given by the sequence $UUUDDUDUDUDD$ with $\mathcal{C}(P) = \{4,5,7,9,11,12\}$. The height and γ_k at each down step is

Down step	4	5	7	9	11	12
height h_k	3	2	2	2	2	1
γ_k	1	2	2	1	2	1

One checks easily that $cr_2(M) = ne_2(M) = 3$.

A key observation is that such labeled Dyck paths encode right-k-crossings and right-k-nestings. The arcs $(i_1, j_1), \ldots, (i_k, j_k)$ of M form a right-k-crossing if and only if $j_1 < j_2 < \cdots < j_k$ are consecutive integers in $\mathcal{C}(M)$, (hence their heights are consecutive integers from large to small), and their labels satisfy $\gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_k$. These arcs form a right-k-nesting if and only if $j_1 < j_2 < \cdots < j_k$ are consecutive integers in $\mathcal{C}(M)$, and their labels satisfy $\gamma_1 > \gamma_2 > \cdots > \gamma_k$.

The involution ϕ maps a labeled Dyck path with labels $(\gamma_i)_{i=1}^m$ to the same Dyck path with labels $(h_i - \gamma_i + 1)_{i=1}^m$. Clearly it exchanges right-k-crossings with right k-nestings. ∎

Corollary 5.2. *For any positive integers r, $m \geq 2$ and $2 \leq k \leq m$, following equations hold:*

$$M(Lcr_k^r, 2m) = M(Rcr_k^r, 2m) = M(Rne_k^r, 2m) = M(Lne_k^r, 2m) \tag{17}$$

Our purpose is to compute the r-th moments of Lcr_k. We will combine the polynomial form (9) and the linear form (8). Note that if $\tilde{P}$ is a merge of r copies of $\underline{P}(Lcr_k)$ with $\tilde{\ell}$ arcs, then $\tilde{P}$ has length $2\tilde{\ell}$. Let $\tilde{c} = |C(\tilde{P})|$. Clearly $k-1 \leq \tilde{c} \leq r(k-1)$. Hence by Prop. 3.5 $M(Lcr_k^r, 2m)$ can be expressed as

$$M(Lcr_k^r, 2m) = \sum_{\tilde{\ell}=k}^{rk} \sum_{\tilde{c}=k-1}^{r(k-1)} c_{\tilde{\ell},\tilde{c}}^{(r)} \binom{2m-\tilde{c}}{2\tilde{\ell}-\tilde{c}} T_{2(m-\tilde{\ell})}, \tag{18}$$

for some constants $c_{\tilde{\ell},\tilde{c}}^{(r)}$.

It is easy to see that when $2m - c \geq 0$,

$$\begin{aligned} \binom{2m-c}{2\ell-c} T_{2(m-\ell)} &= (2(m-\ell)-1)!!\,\frac{(2(m-\ell)+1)(2(m-\ell)+2)\cdots(2m-c)}{(2\ell-c)!} \\ &= \begin{cases} P(m)T_{2m-c} & \text{if } c \text{ is even} \\ Q(m)T_{2m-c+1} & \text{if } c \text{ is odd,} \end{cases} \end{aligned} \tag{19}$$

where $P(x)$ is a polynomial of degree $\ell - \frac{c}{2}$, and $Q(x)$ is a polynomial of degree $\ell - \frac{c+1}{2}$. Therefore, $\binom{2m-c}{2\ell-c}T_{2(m-\ell)}$ is a linear combination with constant coefficients of terms $T_{2(m+i)}$, where $-\frac{c}{2} \leq i \leq \ell - c$. Combining with Formula (18), we have

Theorem 5.3. *For any positive integer r and $m \geq r(k-1)/2$, there is a closed formula*

$$M(Lcr_k^r, 2m) = \sum_{I \leq i \leq J} d_j T_{2(m+j)},$$

where I and J are constants such that $I \geq -r(k-1)/2$ and $J \leq (r-1)k+1$.

As an example, we compute the 2nd and 3rd moments of the number of occurrence for the pattern $\underline{P}(Lcr_2)$, which has length 4, $\ell = 2$ and $c = 1$. We start with the polynomial form (18), and simplify the double summation by analyzing the combinatorial structures. For example, the following simple constraints would reduce the number of unknown $c_{\tilde{\ell},\tilde{c}}$ in (18) by half.

Proposition 5.4. *1. If $\tilde{\ell} = k$, then $\tilde{c} = k-1$ and $c^{(r)}_{k,k-1} = 1$.*

2. If $\tilde{\ell} = rk$, then $\tilde{c} = r(k-1)$, and $c^{(r)}_{rk,r(k-1)} = \binom{r(1+k)}{1+k,1+k,\ldots,1+k}$.

3. $c' \geq \frac{(k-1)\tilde{\ell}}{k}$.

4. $c' \leq \tilde{\ell} - 1$.

Proof. If $\tilde{\ell} = k$, then all r copies of $\underline{P}(Lcr_k)$ coincide with $\tilde{\underline{P}}$. There is only one way to get such a merge. If $\tilde{\ell} = rk$, then the r copies of $\underline{P}(Lcr_k)$ use disjoint set of arcs. A merge of r independent copies of $\underline{P}(Lcr_k)$ is obtained by shuffling the vertices of each copy into a sequence, where for each copy the first two vertices are consecutive in the shuffling. Using the same trick as in the proof of Theorem 2.1, we obtain $\binom{r(1+k)}{1+k,1+k,\ldots,1+k}$ many merges. For item 3, note that the arcs of $\tilde{\underline{P}}$ are unions of the r copies of $\underline{P}(Lcr_k)$. For any merge $s = (h_1, h_2, \ldots, h_r) : \underline{P}(Lcr_k)^r \to \tilde{\underline{P}}$, the left-endpoints of $\tilde{\underline{P}}$ are formed by consecutive segments of vertices of length at least k, in each segment all but the last vertex must be in $C(\tilde{P})$. Item (4) is because in any pattern there is at least one left endpoint that are not in $C(\tilde{P})$. ∎

Proposition 5.5. *Let m be a positive integer. We have the following two equivalent formulas for the second moment of Lcr_2.*

$$M(Lcr_2^2, 2m) = \binom{2m-1}{3}T_{2m-4} + 2\binom{2m-2}{4}T_{2m-6} + 20\binom{2m-2}{6}T_{2m-8} \tag{20}$$

and

$$M((Lcr_2)^2, 2m) = -\frac{1}{6}T_{2(m-1)} + \frac{1}{4}T_{2m} - \frac{1}{6}T_{2(m+1)} + \frac{1}{36}T_{2(m+2)}. \tag{21}$$

The second equation is true when $m \geq 1$.

Proof. Here we have $k = 2$, $c = 1$, and $r = 2$. For a merge $\tilde{\underline{P}}$ of two copies of Lcr_2, the number of arcs $\tilde{\ell}$ ranges from 2 to 4.
If $\tilde{\ell} = 2$, by item 1 of Proposition 5.4, $\tilde{c} = 2 - 1 = 1$, and the coefficient is 1.
If $\tilde{\ell} = 3$, by item 3 of Proposition 5.4, $\tilde{c} \geq 3/2$. Hence the only possible value of $\tilde{c}$ is 2.
If $\tilde{\ell} = 4$, by item 2 of Proposition 5.4, $\tilde{c} = 2$ and the coefficient is $\binom{6}{3,3} = 20$.

Therefore

$$M(Lcr_2^2, 2m) = \binom{2m-1}{3} T_{2m-4} + c_{3,2}\binom{2m-2}{4} T_{2m-6} + 20\binom{2m-2}{6} T_{2m-8} \tag{22}$$

for some constants $c_{3,2}$. Using the data $M(Lcr_2^2, 6) = 12$ we get formula (20).

Alternatively, any formula of the form (22) is a linear combination of $T_{2(m-1)}, T_{2m}, T_{2(m+1)}$ and $T_{2(m+2)}$. One can find the coefficients by solving a linear system with $M(Lcr_2^2, 2m) = 0, 1, 12, 155$ for $m = 1, 2, 3, 4$. The result is Formula (21). ∎

Proposition 5.6. *Let m be a positive integer. The third moment of left-2-crossing is*

$$\begin{aligned} M(Lcr_2^3(M), 2m) = &\binom{2m-1}{3} T_{2m-4} + 6\binom{2m-2}{4} T_{2m-6} + 60\binom{2m-2}{6} T_{2m-8} \\ &+ 6\binom{2m-3}{5} T_{2m-8} + 210\binom{2m-3}{7} T_{2m-10} + \binom{9}{3,3,3}\binom{2m-3}{9} T_{2m-12}. \end{aligned}$$

Alternatively, we have

$$M(Lcr_2^3(M), 2m) = \frac{1}{4} T_{2(m-1)} - \frac{5}{24} T_{2m} + \frac{11}{120} T_{2(m+1)} - \frac{1}{24} T_{2(m+2)} + \frac{1}{216} T_{2(m+3)},$$

for $m \geq 2$. For $m = 1$, $M(Lcr_2^3(M), 2) = 0$.

Proof. Here we have $k = 2$, $c = 1$ and $r = 3$. For a merge $\underline{\tilde{P}}$ of three copies of Lcr_2, the number of arcs $\tilde{\ell}$ ranges from 2 to 6. Using Proposition 5.4, we have that only the following pairs of $(\tilde{\ell}, \tilde{c})$ are possible: $(2,1), (3,2), (4,2), (4,3), (5,3)$ and $(6,3)$. In addition, $c_{2,1} = 1$ and $c_{6,3} = \binom{9}{3,3,3}$. Hence

$$\begin{aligned} M(Lcr_2^3(M), 2m) = &\binom{2m-1}{3} T_{2m-4} + c_{3,2}\binom{2m-2}{4} T_{2m-6} + c_{4,2}\binom{2m-2}{6} T_{2m-8} \\ &+ c_{4,3}\binom{2m-3}{5} T_{2m-8} + c_{5,3}\binom{2m-3}{7} T_{2m-10} + \binom{9}{3,3,3}\binom{2m-3}{9} T_{2m-12}. \end{aligned} \tag{23}$$

The coefficients $c_{3,2}, c_{4,2}, c_{4,3}$ and $c_{5,3}$ can be obtained by counting the number of corresponding merges. But our purpose is to avoid too much details on combinatorial structures and rely on the data available. One way to get the $c_{i,j}$'s is to use the value of $M(Lcr_2^3(M), 2m)$ for $m = 3, 4, 5, 6$ to establish a system of linear equations. Using a computer programming we have

m	1	2	3	4	5	6
$M(Lcr_2^3(M), 2m)$	0	1	16	261	4536	85533

These data yield $c_{3,2} = 6, c_{4,2} = 60, c_{4,3} = 6$ and $c_{5,3} = 210$.

Alternatively, with the equation (23) we could turn to the linear form. Applying Formula (19) to the above equation, we have that for $m \geq 2$,

$$M(Lcr_2^3(M), 2m) = c_{-1} T_{2(m-1)} + c_0 T_{2m} + c_1 T_{2(m+1)} + c_2 T_{2(m+2)} + c_3 T_{2(m+3)}$$

for some constants $c_{-1}, c_0, c_1, c_2, c_3$. Using the values of $M(Lcr_2^3(M), 2m)$ for $m = 2, \ldots, 6$ and Maple, we get

$$c_{-1} = \frac{1}{4},\ c_0 = -\frac{5}{24},\ c_1 = \frac{11}{120},\ c_2 = -\frac{1}{24},\ c_3 = \frac{1}{216}.$$

∎

6. The dimension exponent in matchings

We finish this paper by giving a closed formula for the moments of the dimension exponent $d(M)$. Recall that for $M \in \mathcal{M}_{2m}$, $d(M) = s_{max}(M) - s_{min}(M) - m$. By definition, $d(M)$ is a linear combination of simple statistics s_{max} and s_{min}, for both of which Q is not a constant.

First we show that there is a pattern such that $d(M)$ counts the occurrence of this pattern.

Proposition 6.1. *Let T be a partial matching of length 3 with $A(T) = \{(1, 3)\}$ and $C(T) = \emptyset$. Then for any $M \in \mathcal{M}_{2m}$, $d(M) = f_{\underline{T},1}$.*

Proof. By definition

$$d(M) = \sum_{(i,j)\in M} (j - i + 1) - 2m = \sum_{(i,j)\in M} (j - i - 1).$$

Hence $d(M)$ counts the number of triples $i < t < j$ where (i, j) is an arc of M. It is exactly the simple statistic associated to the pattern $\underline{T} = (T, A(T), C(T))$ and $Q = 1$. ∎

Now that $d(M)$ can be expressed by a simple statistic with $Q = 1$, we can use Proposition 3.5 to compute its higher moments. For any pattern P that is a merge of r copies of T, if P has length k and ℓ arcs, $(C(P) = \emptyset)$, then $1 \le \ell \le r$, and $2\ell \le k \le 2\ell + r$. Hence

$$M(d(M)^r, 2m) = \sum_{\ell=1}^{r} \sum_{i=0}^{r} c_{\ell,i}^{(r)} \binom{2m}{2\ell + i} T_{2(m-\ell)}, \tag{24}$$

where $c_{\ell,i}^{(r)}$ is the number of ways to merge r copies of T to a pattern with ℓ arcs and total length $2\ell + i$. Note that for $m \ge \ell$,

$$\begin{aligned} \binom{2m}{2\ell + i} T_{2(m-\ell)} &= (2(m-\ell)-1)!! \frac{(2m)(2m-1)\cdots(2m-2\ell+1)}{((2\ell+i)!} \\ &\quad \cdot (2m-2\ell)(2m-2\ell-1)\cdots(2m-2\ell-i+1) \\ &= T_{2m} \cdot R(m), \end{aligned} \tag{25}$$

where $R(m)$ is a polynomial of m of degree $i+\ell$. Note that $R(m)$ has factors $2m(2m-2)\cdots(2m-2\ell)$, hence (25) is also true for $0 \le m < \ell$. Thus we can write $\binom{2m}{2\ell+i} T_{2(m-\ell)}$ as a linear combination of $T_{2m}, T_{2(m+1)}, \ldots, T_{2(m+i+\ell)}$. Summing over $\ell, i = 1, \ldots, r$, we get

Theorem 6.2. *For any positive m and r,*

$$M(d(M)^r, 2m) = \sum_{j=0}^{2r} d_j T_{2(m+j)}$$

for some constants $d_j \in \mathbb{Q}$.

For example, when $r = 1$, using either Corollary 2.3 or the Example 1 in section 2, we have

$$M(d(M), 2m) = \binom{2m}{3} T_{2(m-1)},$$

which can also be expressed as

$$M(d(M), 2m) = \frac{1}{2} T_{2m} - T_{2(m+1)} + \frac{1}{6} T_{2(m+2)}.$$

When $r = 2$, we have

$$M(d(M)^2, 2m) = \binom{2m}{3} T_{2(m-1)} + 2\binom{2m}{4} T_{2(m-1)} + 2\binom{2m}{4} T_{2(m-2)} + 16\binom{2m}{5} T_{2(m-2)} + 20\binom{2m}{6} T_{2(m-2)},$$

which equals

$$M(d(M)^2, 2m) = \frac{1}{4} T_{2m} - \frac{8}{3} T_{2(m+1)} + \frac{5}{2} T_{2(m+2)} - \frac{8}{15} T_{2(m+3)} + \frac{1}{36} T_{2(m+4)}$$

for all $m \geq 1$.

Acknowledgments

We thank Professor Persi Diaconis for sharing with us a preprint of [8] and suggesting the problem on matchings.

This publication was made possible by NPRP grant #[5-101-1-025] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] R. Ahlswede, L. Khachatrian. The Complete Intersection Theorem for Systems of Finite sets, European J. of Combin. 18(1997), no. 2, 125–136.

[2] J. Akiyama, P. Frankl. On the size of graphs with complete factors, J. Graph Theory, 9(1985), 197-201.

[3] George E. Andrews, Peter Paule. Some questions concerning computer-generated proofs of a binomial double-sum identity, J. Symbolic Comput. 16(1993), 147–151.

[4] M. Bousquet-Melou, A. Claesson, M. Dukes and S. Kitaev. (2+2)-free posets, ascent sequences and pattern avoiding permutations, J. Combin. Theory Ser. A 117(2010), 884909.

[5] J. Balogh and D. Mubayi. A new short proof of a theorem of Ahslwede and Khachatrian, J. Combin. Theory Ser. A 15(2008), 326–330.

[6] A. Claesson, M. Dukes, S. Kitaev. A direct encoding of Stoimenow's matchings as ascent sequences, Australas. J. Combin. 49(2011) 4759.

[7] W. Chen, N. Fan and A. Zhao. Partitions and partial matchings avoiding neighbor patterns. European J. Combin. 33(2012), no. 4, 491504.

[8] B. Chern, P. Diaconis, D.M. Kane and R.C. Rhoades, Closed expressions for averages of set partition statistics. Research in the Mathematical Sciences 1(2), (2014).

[9] B. Chern, P. Diaconis, D.M. Kane and R.C. Rhoades. Central limit theorems for some set partition statistics. to appear in Advances in Applied Matheamtics, 2015.

[10] M. de Sainte-Catherine. Couplages et Pfaffiens en Combinatoire, Physique et Informatique, Ph.D. Thesis, University of Bordeaux I, Talence, France, 1983.

[11] M. Dukes, R. Parviainen. Ascent sequences and upper triangular matrices containing non-negative integers. Electron. J. Combin. 17(2010) R5.

[12] P. Erdős and T. Gallai. On the minimal number of vertices representing the edges of a graph. Pub. Inst. Math. Hungar. Acad. Sci. 6(1961), 181–203.

[13] P. Frankl. Improved bounds for Erdős Matching Conjecture. J. Combin. Theory Ser. A 120(2013), 1068–1072.

[14] A. Kasraoui and J. Zeng, Distribution of crossings, nestings and alignments of two edges in matchings and partitions. Electron. J. Combin. 13(2006): Research Paper 33.

[15] A.J. Stam. Generation of random partitions of a set by an urn model. J. Combin. Theory Ser A 35(1983), 231–240.

[16] E. Steingrímsson. Generalized Permutation Patterns–A Short Survey, 2008. arXiv:0801.2412.

[17] Stoimenow, A. Enumeration of chord diagrams and an upper bound for Vassiliev invariants. J. Knot Theory Ramifications 7(1998), no. 1, 93–114.

Niraj Khare
Carnegie Mellon University Qatar, Doha, Qatar
E-mail address: nkhare@cmu.edu

Rudolph Lorentz
Texas A&M University at Qatar, Doha, Qatar
E-mail address: rudolph.lorentz@qatar.tamu.edu

Catherine H. Yan
Department of Mathematics, Texas A&M University,
College Station, Texas *E-mail address:* cyan@math.tamu.edu
URL: http://www.math.tamu.edu/ catherine.yan/