A sharp inequality for the variance with respect to the Ewens sampling formula

Žygimantas Baronėnas, Eugenijus Manstavičius, and Patricija Šapokaitė

Institute of Mathematics, Vilnius University, Naugarduko 24, LT-03225
(e-mail: eugenijus.manstavicius@mif.vu.lt)

Dedicated to Memory of Professor Jonas Kubilius

Received December 2, 2020; revised February 22, 2021

Abstract. We examine the variance of a linear statistic defined on the symmetric group endowed with the Ewens probability. Despite the dependence of the summands, it can be bounded from above by a constant multiple of the sum of variances of the summands. We find the exact value of this constant. The analysis of the appearing quadratic forms and eigenvalue search is built upon the exponential matrices and discrete Hahn’s polynomials.

MSC: 60C05, 05A16, 20P05
Keywords: random permutation, quadratic form, eigenvalue, discrete Hahn’s polynomials, Turán–Kubilius inequality

1 Introduction and results

In this paper, we examine the variance of a linear statistic defined on the symmetric group endowed with the Ewens probability. The main obstacle to overcome in this seemingly simple problem is the dependence of the summands. We propose an approach built upon exponential matrices and special functions.

Let $S_n$ denote the symmetric group of permutations $\sigma$ acting on $n \in \mathbb{N}$ letters. Each $\sigma \in S_n$ has a unique representation (up to the order) by the product of independent cycles $\kappa_i$:

$$\sigma = \kappa_1 \cdots \kappa_w,$$

(1.1)

where $w = w(\sigma)$ denotes the number of cycles. Denote by $k_j(\sigma) \geq 0$ the number of cycles in (1.1) of length $j$ for $1 \leq j \leq n$ and introduce the cycle vector $\vec{k}(\sigma) = (k_1(\sigma), \ldots, k_n(\sigma))$.

As usual, set $(x)_m = x(x + 1) \cdots (x + m - 1)$, where $m \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, for the increasing factorial. Denote also

$$\Theta(m) = \binom{\theta + m - 1}{m} = \frac{(\theta)_m}{m!} = [x^m](1 - x)^{-\theta},$$

(1.2)
where \([x^m]f(x)\) stands for the \(m\)th coefficient of a power series \(f(x)\), and \(\theta > 0\) is a parameter. The \textit{Ewens probability measure} \(\nu_{\theta, \theta}\) on \(\mathbb{S}_n\) is defined by

\[
\nu_{\theta, \theta}(\{\sigma\}) = \frac{\theta^{|\sigma|}}{(\theta)_n}, \quad \sigma \in \mathbb{S}_n.
\]

This gives the probability space \((\mathbb{S}_n, 2^{\mathbb{S}_n}, \nu_{\theta, \theta})\), and every mapping \(f: \mathbb{S}_n \to \mathbb{R}\) becomes a random variable (r.v.) defined on it. Throughout the paper, we write it as \(f(\sigma)\), leaving the elementary event \(\sigma\) in contrast to r.v.s defined on some unspecified probability space \(\{\Omega, \mathcal{F}, \mathbb{P}\}\).

Set \(\ell(s) = 1s_1 + \cdots + ns_n\) for a vector \(s = (s_1, \ldots, s_n) \in \mathbb{N}_0^n\). The equality \(\ell(k(\sigma)) = n\), valid for each \(\sigma \in \mathbb{S}_n\), shows the dependence of the r.v. \(k_j(\sigma)\) with respect to \(\nu_{\theta, \theta}\). It is well known (see, e.g., [1, Sect. 2.3]) that the distribution of \(k(\sigma)\) can be written as the conditional distribution of \(\xi = (\xi_1, \ldots, \xi_n)\), where \(\xi_j\), \(1 \leq j \leq n\), are independent Poisson r.v.s with parameter \(\theta/j\). Indeed, direct calculation shows that

\[
\nu_{\theta, \theta}(k(\sigma) = s) = 1\{\ell(s) = n\} \Theta(n)^{-1} \prod_{j=1}^n \left( \frac{\theta}{j} \right)^{s_j} = \mathbb{P}(\xi = s | \ell(\xi) = n). \tag{1.3}
\]

Here \(1\{\cdot\}\) stands for the indicator function. The probability in (1.3), ascribed to the vector \(\bar{s} \in \mathbb{N}_0^n\), is called the \textit{Ewens sampling formula}. It has been introduced by Ewens [5] to model the mutation of genes. For a comprehensive account of the recent applications of this ubiquitous distribution in combinatorics and statistics, see [1, 6, 7] or survey [4] and the subsequent comments.

We prefer to stay within the theory of random permutations. Apart from \(w(\sigma)\), other linear statistics (or \textit{completely additive functions})

\[
h(\sigma) = a_1k_1(\sigma) + \cdots + a_n k_n(\sigma), \tag{1.4}
\]

where \(\bar{a} := (a_1, \ldots, a_n) \in \mathbb{R}^n\) is a nonzero vector, continue to raise an interest. For example, \(h(\sigma)\) with \(a_j = \log j, j \leq n\), is a good approximation for the logarithm of the group-theoretical order of \(\sigma \in \mathbb{S}_n\) (see [1] or [22]). The case with \(a_j = \{xj\}, \{u\} = \text{the fractional part of } u \in \mathbb{R},\) is met in the theory of random permutation matrices (see [21]).

For an arbitrary \(h(\sigma)\), the problem of finding necessary and sufficient conditions ensuring the weak convergence of distributions

\[
\nu_{\theta, \theta}(h(\sigma) - \alpha(n) \leq x\beta(n)), \tag{1.5}
\]

where \(\alpha(n) \in \mathbb{R}\) and \(\beta(n) \to \infty\) as \(n \to \infty\), is still open (see [1, Sect. 8.5] or [13] and the references therein). Obstacles in the necessity part arise because of the dependence of the summands as shown by (1.3). This also happens in the analysis of power moments carried out by the second author [11] and [12] even in the case \(\theta = 1\). Let us now focus on the variance.

By \(\mathbb{E}_{\nu_{\theta, \theta}}f(\sigma)\) and \(\text{Var}_{\nu_{\theta, \theta}}f(\sigma)\) we denote the mean value and variance of a r.v. \(f(\sigma)\) defined on \(\mathbb{S}_n\) with respect to \(\nu_{\theta, \theta}\). For the particular function \(h(\sigma)\) in (1.4), we also set \(A_{\nu, \theta}(\bar{a}) = \mathbb{E}_{\nu_{\theta, \theta}}h(\sigma)\) and \(D_{\nu, \theta}(\bar{a}) = \text{Var}_{\nu_{\theta, \theta}}h(\sigma)\). Applying Watterson’s [20] formulas (see [1, (5.6), p. 96]) for the factorial moments of \(k_j(\sigma)\), we easily find (see [15] for the details) the expressions

\[
A_{\nu, \theta}(\bar{a}) = \theta \sum_{j \leq n} a_j \frac{\Theta(n-j)}{\Theta(n)},
\]

and

\[
D_{\nu}(\bar{a}) = \theta \sum_{j \leq n} a_j^2 \frac{\Theta(n-j)}{\Theta(n)} + \theta^2 \sum_{i+j \leq n} a_i a_j \frac{\Theta(n-i-j)}{\Theta(n)} - \theta^2 \left( \sum_{j \leq n} a_j \frac{\Theta(n-j)}{\Theta(n)} \right)^2 \tag{1.6}
\]

\[= : \theta B_{\nu}(\bar{a}) + \theta^2 \Delta_{\nu}(\bar{a})\]
if \( n \geq 2 \) and
\[
B_n(\bar{a}) = \sum_{j \leq n} \frac{a_j^2}{j} \frac{\Theta(n-j)}{\Theta(n)}.
\]

The latter quantity is close to the sum of variances of the summands in the definition of \( h(\sigma) \). In fact, formula (4) from [15] shows that
\[
\sum_{j \leq n} a_j^2 \text{Var}_{n,\theta} k_j(\sigma) - \theta B_n(\bar{a}) = O(n^{-(1,\theta)} B_n(\bar{a})),
\]
where \( a \wedge b = \min\{a, b\} \) for \( a, b \in \mathbb{R} \), with an absolute constant in the symbol \( O(\cdot) \). We also have (see [15])
\[
D_n(\bar{a}) = C\theta B_n(\bar{a}) \tag{1.7}
\]
uniformly in \( n \geq 2 \) with an absolute constant \( C \), which can be specified. If \( \theta \geq 1 \), then we can take \( C = 2 \).

The purpose of the present paper is to find the exact value of \( D_n(\bar{a}) \) in (1.7).

**Theorem 1.** Let \( \theta > 0 \) and \( n \geq 2 \). Then
\[
\tau_n(\theta) := \sup \left\{ \frac{D_n(\bar{a})}{\theta B_n(\bar{a})} : \bar{a} \in \mathbb{R}^n \setminus \{0\} \right\} = \frac{\theta + 2}{\theta + 1}.
\]
The supremum is achieved by taking \( a_j = (\theta + 2) j^2 - (2n + \theta) j \), where \( 1 \leq j \leq n \).

The pioneering results obtained in [11] and [16] showed that \( \tau_n(1) = 3/2 + O(n^{-1}) \) and \( \tau_n(2) = 4/3 + O(n^{-1}) \). The approach originated in Kubilius’ paper [10] was based upon the extremal properties of the Jacobi polynomials. It unavoidably added a vanishing error term to the result. Recently, Klimavičius and the second author [8] established that \( \tau_n(1) = 3/2 \) for all \( n \geq 2 \). Theorem 1 resumes the research for an arbitrary \( \theta > 0 \). It is directly related to the above-mentioned problem concerning distributions (1.5). Applying Theorem 1, we obtain that the weak convergence of (1.5) with \( \beta(n) = \sqrt{n B_n(\bar{a})} \), which is natural to use, can take place only in the limit laws having variance not exceeding \( (\theta + 2)/(\theta + 1) \).

By (1.3) the result can be restated for the conditional variance of the linear statistics \( Y_n := a_1 \xi_1 + \cdots + a_n \xi_n \). We obtain the following optimal inequality.

**Corollary.** Let \( n \geq 2 \) and \( a_j \in \mathbb{R} \), \( 1 \leq j \leq n \). Then
\[
\text{Var}(Y_n \mid \ell(\xi) = n) \leq \frac{\theta(\theta + 2)}{\theta + 1} \sum_{j=1}^{n} \frac{a_j^2}{j} \frac{\Theta(n-j)}{\Theta(n)}.
\]

The problem concerns the quadratic forms \( \Delta_n(\bar{a}) \) and \( B_n(\bar{a}) \). The substitution
\[
a_j = \left( \frac{j \Theta(n)}{\Theta(n-j)} \right)^{1/2} x_j, \quad 1 \leq j \leq n,
\]
reduces \( B_n(\bar{a}) \) to the square of Euclidean norm \( \|\bar{x}\|^2 \) of the vector \( \bar{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \). Then \( \Delta_n(\bar{a}) \) becomes a quadratic form, denoted afterward by \( M_n(\bar{x}) := \bar{x} M_n \bar{x}' \), where \( \bar{x}' \) is the transpose of \( \bar{x} \), and \( M_n = ((m_{ij})) \), \( 1 \leq i, j \leq n \), is the matrix with entries
\[
m_{ij} = \frac{\Theta(n-i-j)}{(ij \Theta(n-i) \Theta(n-j))^{1/2}} \left( \frac{\Theta(n-i)}{i \Theta(n)} \right)^{1/2} \left( \frac{\Theta(n-j)}{j \Theta(n)} \right)^{1/2}.
\]

Lith. Math. J., 61(3):289–300, 2021.
Here we assume that $\Theta(-k) = 0$ if $k \in \mathbb{N}$. Now by (1.6) we have

$$
\tau_{n,\theta} = 1 + \theta \sup_{\bar{x} \neq 0} \left( \| \bar{x} \|^{-2} M_n(\bar{x}) \right) = 1 + \theta \sup_{\bar{x} \neq 0} \left( \| \bar{x} \|^{-2} \sum_{r=1}^{n} \mu_r x_r^2 \right)
$$

$$
= 1 + \theta \max_{1 \leq r \leq n} \mu_r,
$$

where $\{\mu_1, \ldots, \mu_n\}$ is the spectrum of the matrix $M_n$. So, Theorem 1 follows from the following proposition.

**Theorem 2.** The spectrum of the matrix $M_n$ comprises

$$
\mu_r = \frac{(-1)^r (r-1)!}{(\theta)_r}, \quad 1 \leq r \leq n.
$$

For the eigenvector corresponding to the maximal $\mu_2$, we may take the vector with coordinates

$$
\left( (\theta + 2) j - (2n + \theta) \right) \left( j \Theta(n - j) \right)^{1/2}, \quad 1 \leq j \leq n.
$$

The proof of Theorem 2 presented in the next section is built upon exponential matrices.

The problem of finding the remaining eigenvectors of the matrix $M_n$ also raises an interest. In fact, they were already used in the continuation of the paper [14] dealing with lower estimates of $D_n(\bar{a})$. Solving is based upon particular cases of the generalized hypergeometric series, which are exposed, for example, in [3] or [9, Chap. 9]. The hint to exploit them stems from [8]. We may confine ourselves to the case of polynomials that, in the traditional notation, can be written as

$$
p_{m+2}F_q(-m, -x; (a_p); (b_q); z) = \sum_{k=0}^{m} \frac{(-m)_k (-x)_k (a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k k!} \, z^k,
$$

where $p, q, m \in \mathbb{N}_0$, $a_1, \ldots, a_p; b_1, \ldots, b_q \in \mathbb{R}$ are parameters. Moreover, it suffices to reckon the discrete Hahn polynomials

$$
Q_r(x; \alpha, \beta, n) = 3F2(-r, -x, r + \alpha + \beta + 1; \alpha + 1, -n + 1; 1)
$$

by specifying the parameters to $\alpha = 1$ and $\beta = \theta - 1$. This agrees with the notation in [9, Sect. 9.5] with $N = n - 1$. The values of polynomials

$$
q_r(x) = Q_r(x-1; 1, \theta - 1, n), \quad 0 \leq r \leq n - 1,
$$

at $x = j \in \{1, \ldots, n\}$ can be calculated using the recurrence formula [9, (9.5.3)]. Namely, $q_0(j) = 1,$

$$
q_1(j) = - \frac{(\theta + 2) j - (\theta + 2n)}{2(n-1)},
$$

and

$$
A_r q_{r+1}(j) = (A_r + C_r - j + 1) q_r(j) - C_r q_{r-1}(j),
$$

where

$$
A_r = \frac{(r + \theta + 1)(r + 2)(n - r - 1)}{(2r + \theta + 1)(2r + \theta + 2)}, \quad C_r = \frac{r(r + \theta + n)(r - \theta - 1)}{(2r + \theta)(2r + \theta + 1)}.
for \( r = 1, \ldots, n - 1 \). Moreover, we have the following orthogonality property \([9, (9.5.2)]\):

\[
\langle q_l, q_r \rangle := \sum_{j=1}^{n} j q_l(j) q_r(j) \Theta(n - j) = \delta_l \pi_r^2,
\]

(1.8)

where \( \delta_l \) is the Kronecker symbol, and

\[
\pi_r^2 = \frac{(r + \theta + 1)n(\theta)}{(2r + \theta + 1)(r + 1)(n - r)r(n - 1)!} > 0.
\]

Note that \( q_r(x), 1 \leq r \leq n - 1 \), can be also obtained uniquely by the Gram–Schmidt orthogonalization procedure starting with \( q_0(x) = 1 \) and applying the inner product (1.8). They form an orthogonal basis in the vector space of polynomials with degrees not exceeding \( n - 1 \). Exploiting this, in the isomorphic Euclidean space \( \mathbb{R}^n \), we easily find the needed canonical basis for the matrix \( M_n \).

**Theorem 3.** The system of the vectors

\[
\bar{e}_r = (e_{r1}, \ldots, e_{rn}), \quad 1 \leq r \leq n,
\]

where

\[
e_{rj} = \pi_{r-1}^{-1} q_{r-1}(j) \sqrt{j \Theta(n - j)}, \quad 1 \leq j \leq n,
\]

is an orthonormal basis in \( \mathbb{R}^n \). Moreover, the vector \( \bar{e}_r \) is the eigenvector of matrix \( M_n \) corresponding to \( \mu_r \) for each \( 1 \leq r \leq n \).

The proof will be presented in the last section of the paper.

Finally, the distributions of mappings defined on random permutations taken according to the Ewens probability are close to that defined on logarithmic decomposable combinatorial structures (see \([1]\)); therefore we hope that our method is applicable when estimating the variances of similar statistics defined in such classes.

## 2 Proof of Theorem 2

The idea is finding a matrix \( L_n \) such that the product

\[
e^{L_n} M_n e^{-L_n} =: (w_{ij})
\]

is the triangle matrix with \( w_{ij} = 0 \) if \( 1 \leq j < i \leq n \) and \( w_{jj} = \mu_j \) if \( 1 \leq j \leq n \). This implies that the eigenvalues of \( M_n \) are listed on the main diagonal of the product, as desired.

At first, we recall two identities from \([18]\).

**Lemma 1.** Let \( M, m \in \mathbb{N}_0 \) and \( a, b \in \mathbb{R} \). Then

\[
\sum_{k=0}^{M} \binom{a + k}{k} \binom{b - k}{M - k} = \sum_{k=0}^{M} \binom{a + b - k}{M - k} \tag{2.1}
\]

and

\[
\sum_{k=0}^{M} (-1)^k \binom{M}{k} \binom{a - k}{m} = \binom{a - M}{m - M} \tag{2.2}
\]

**Proof.** See formulas (43) on page 618 and (56) on page 619 of \([18]\). \( \square \)
Proof of Theorem 2. Let us introduce the matrix \( L_n(\theta) = ((l_{ij})) \) with the entries \( l_{ij} = 0 \) for all \( 1 \leq i, j \leq n \) except for \( i = j + 1 \), where

\[
l_{j+1,j} = -\left(\frac{(j+1)j\theta(n-j-1)}{\theta(n-j)}\right)^{1/2}, \quad 1 \leq j \leq n-1.
\]

Consider the powers \( L_n^k(\theta) = ((l_{ij}^k)), 0 \leq k \leq n - 1 \). The nonzero entries of \( L_n^k(\theta), 1 \leq k \leq n - 1 \), fill up the \( k \)th diagonal under the main one. By induction we observe that

\[
l_{j+k,j} = l_{j+k,j+1}^{(k-1)} = l_{j,k+1,j+1}^{(k-1)} l_{j+1,j}^{(k-1)} = \ldots = l_{j+k,j}^{(k-1)} l_{j+1,j}^{(k-1)} \ldots l_{j+1,j}^{(k-1)}
\]

\[
= (-1)^k \prod_{r=0}^{k-1} \left( (j + r + 1)(j+r) \right)^{1/2} \left( \prod_{r=0}^{k-1} \frac{\theta(n-j-r-1)}{\theta(n-j-r)} \right)^{1/2}
\]

\[
= (-1)^k (j)^k \left( \frac{j+k}{j} \right)^{1/2} \left( \frac{\theta(n-j-k)}{\theta(n-j)} \right)^{1/2}
\]

for \( 1 \leq k \leq n - j \). Hence the matrix \( V := e^{L_n(\theta)} = ((v_{ij})) \) has \( v_{ij} = 0 \) for \( 1 \leq i < j \leq n \) and

\[
v_{ij} = \frac{l_{ij}^{(i-j)}}{(i-j)!} = (-1)^{i-j} \left( \frac{i}{j} \right)^{1/2} \left( \frac{\theta(n-i)}{\theta(n-j)} \right)^{1/2}
\]

\[
= (-1)^{i-j} \left( \frac{i}{j} \right)^{1/2} \left( \frac{\theta(n-i)}{\theta(n-j)} \right)^{1/2}
\]

for \( i \geq j \). Moreover, \( V^{-1} = e^{-L_n(\theta)} = (|v_{ij}|) \) for \( 1 \leq i, j \leq n \).

More technical obstacles arise calculating

\[
w_{ij} = \sum_{1 \leq r \leq i \atop j \leq s \leq n} v_{ir} m_{rs} |v_{sj}|
\]

\[
= \sum_{1 \leq r \leq i \atop j \leq s \leq n} (-1)^{i-r} \left( \frac{i}{r} \right)^{1/2} \left( \frac{\theta(n-i)}{\theta(n-r)} \right)^{1/2} \cdot \frac{\theta(n-r-s)}{(rs\theta(n-r-\theta(n-s)))^{1/2}} \times \left( \frac{s-1}{j-1} \right)^{1/2} \left( \frac{\theta(n-s)}{\theta(n-j)} \right)^{1/2}
\]

\[
- \sum_{1 \leq r \leq i \atop j \leq s \leq n} (-1)^{i-r} \left( \frac{i}{r} \right)^{1/2} \left( \frac{\theta(n-i)}{\theta(n-r)} \right)^{1/2} \cdot \frac{\theta(n-r)}{\theta(n)} \left( \frac{\theta(n-s)}{s\theta(n)} \right)^{1/2} \times \left( \frac{s-1}{j-1} \right)^{1/2} \left( \frac{\theta(n-s)}{\theta(n-j)} \right)^{1/2}
\]

\[
= \sum_1 - \sum_2.
\]
Here
\[ \sum_1 = (-1)^i \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \sum_{1 \leq r \leq i \land n-j} (-1)^r \frac{\Theta(n - r)}{r} \sum_{j \leq s \leq n-r} \Theta(n - s) \binom{s - 1}{j - 1}. \]

After the change \( s = n - r - k \) and application of (1.2), the inner sum reduces to that given in (2.1). It equals
\[
\sum_{k=0}^{n-r-j} \binom{\theta - 1 + k}{k} \binom{n - r - 1 - k}{n - r - j - k}
= \sum_{k=0}^{n-r-j} \binom{\theta + n - r - 2 - k}{n - r - j - k}
= \sum_{l=0}^{n-r-j} \binom{\theta - 2 + j + l}{l}
= \left[ x^{n-r-j} \right] \frac{1}{(1-x)^{\theta+j}} = \binom{\theta + n - r - 1}{n - r - j}.
\]

Hence
\[ \sum_1 = (-1)^i \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \sum_{1 \leq r \leq i \land n-j} (-1)^r \frac{\Theta(n - r)}{r} \binom{\theta + n - r - 1}{n - r - j}. \]

Similarly,
\[ \sum_2 = (-1)^i \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \sum_{1 \leq r \leq i} (-1)^r \binom{i}{r} \sum_{j \leq s \leq n} \Theta(n - s) \binom{s - 1}{j - 1}. \]

Since
\[ \sum_{j \leq s \leq n} \binom{s - 1}{j - 1} \Theta(n - s) = \left[ x^n \right] \left( \frac{x^j}{1-x} \cdot \frac{1}{(1-x)^{\theta}} \right) = \binom{\theta + n - 1}{n - j}, \]
we obtain
\[ \sum_2 = \frac{(-1)^{i+1}}{\Theta(n)} \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \binom{\theta + n - 1}{n - j}. \]

Consequently,
\[
\begin{align*}
\omega_{ij} &= \sum_1 - \sum_2 \\
&= (-1)^i \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \sum_{0 \leq r \leq i \land n-j} (-1)^r \frac{\Theta(n - r)}{r} \binom{\theta + n - r - 1}{n - r - j} \\
&= (-1)^i \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \binom{j^i}{\theta_j} \sum_{0 \leq r \leq i \land n-j} (-1)^r \binom{i}{r} \binom{n - r}{j - i} =: (-1)^i \left( \frac{\Theta(n - i)}{ij\Theta(n - j)} \right)^{1/2} \binom{j^i}{\theta_j} \sum_{0 \leq r \leq i \land n-j} (-1)^r \binom{i}{r} \binom{n - r}{j - i}. \tag{2.3}
\end{align*}
\]

Using the definition of \( \Theta(m) \) given in (1.2) and applying identity (2.2), we find that
\[ \sum = \frac{j^i}{\theta_j} \sum_{0 \leq r \leq i \land n-j} (-1)^r \binom{i}{r} \binom{n - r}{j - i} = \frac{j^i}{\theta_j} \binom{n - i}{j - i}, \]
Here $\sum = 0$ if $j < i$ and $\sum = j!/(\theta)_j$ if $i = j$. Plugging this into (2.3), we obtain

$$
 w_{ij} = (-1)^i \left( \frac{\Theta(n-i)}{i!\Theta(n-j)} \right)^{1/2} \frac{j!}{(\theta)_j} \left( \frac{n-i}{j-i} \right) = \begin{cases} 0 & \text{if } i > j, \\ (-1)^j (j-1)!/(\theta)_j & \text{if } i = j. 
\end{cases}
$$

This proves the main statement of Theorem 2.

It remains to show that the eigenvector corresponding to $\mu_2$ has the described form. However, this follows immediately from Theorem 3, which we prove in the next section. Theorem 2 is proved. \(\square\)

### 3 Proof of Theorem 3

We now find all eigenvectors of the matrix $M_n$. Again, we have to recall a useful identity.

**Lemma 2.** Let $p, q, M \in \mathbb{N}_0$, $\alpha, \beta \in \mathbb{R}$, and $a_1, \ldots, a_p, b_1, \ldots, b_q$ be the parameters such that the hypergeometric series below is correctly defined. Then

$$
\sum_k \binom{M}{k} (\alpha)_{M-k} (\beta)_k \cdot p+1F_q(-k, (a_p); (b_q); 1) = (\alpha + \beta)_M \cdot p+2F_{q+1}(-M, \beta, (a_p); \alpha + \beta, (b_q); 1).
$$

**Proof.** See formula (7) presented on page 388 in [17]. \(\square\)

As a corollary, we find the next sum involving the above-introduced polynomial

$$
Q_r(x; \alpha, \beta, n) = 3F_2(-r, -x, r + \alpha + \beta + 1; \alpha + 1, -n + 1; 1).
$$

**Lemma 3.** Let $n \geq 2$, $0 \leq M \leq n - 1$, and $0 \leq r \leq n - 1$. Then

$$
\sum_r (M) := \sum_{k=0}^{M} Q_r(k; 1, \theta - 1, n) \Theta(M - k)
$$

$$
= \frac{(\theta + 1)_M}{M!} 4F_3(-M, 1, -r, r + \theta + 1; \theta + 1, 2, 1 - n; 1).
$$

**Proof.** Apply Lemma 2 for $\alpha = \theta, \beta = 1$, and $p = q = 2$. \(\square\)

The obtained expressions of $\sum_{r=1}^{M-1}(M)$, $1 \leq r \leq n$, will be used afterward. For short, let

$$
4F_3(-M) = 4F_3(-M, 1, 1 - r, r + \theta; \theta + 1, 2, 1 - n; 1).
$$

**Lemma 4.** Let $y_r = (y_{r1}, \ldots, y_{rn}) = \pi_{r-1} \bar{e}_r M_n$, $1 \leq r \leq n$. Then

$$
\ y_{ri} = -\left( \frac{\Theta(n-i)}{i} \right)^{1/2} \frac{n}{r(r + \theta - 1)} \cdot 3F_2(-r, -n + i, r + \theta - 1; \theta, -n; 1)
$$

(3.1) for $1 \leq i \leq n$. 


Proof. In the notation above, applying Lemma 3, we obtain

\[ y_{ri} = \sum_{j=1}^{n} q_{r-1}(j)(j\Theta(n-j))^{1/2} m_{ji} \]

\[ = \frac{1}{(i\Theta(n-i))^{1/2}} \sum_{j=1}^{n-i} Q_{r-1}(j-1;1,\theta-1,n)\Theta(n-j-i) \]

\[ - \left( \frac{\Theta(n-i)}{i} \right)^{1/2} \frac{1}{\Theta(n)} \sum_{j=1}^{n} Q_{r-1}(j-1;1,\theta-1,n)\Theta(n-j) \]

\[ = \frac{1}{(i\Theta(n-i))^{1/2}} \sum_{r=1}^{n} (n-i-1) - \left( \frac{\Theta(n-i)}{i} \right)^{1/2} \frac{1}{\Theta(n)} \sum_{r=1}^{n} (n-1) \]

\[ = \frac{1}{\theta} \left( \frac{\Theta(n-i)}{i} \right)^{1/2} [(n-i) \cdot {}_{4}F_{3}(-n+i+1) - n \cdot {}_{4}F_{3}(-n+1)] \]

for \( 1 \leq i < n \).

Since \( (a)_{l-1} = (a-1)_{l}/(a-1) \) for \( a \neq 1 \), we have

\[ (n-i) \cdot {}_{4}F_{3}(-n+i+1) \]

\[ = (n-i) \sum_{l=1}^{r} \frac{(-n+i+1)_{l-1}(1-r)_{l-1}(r+\theta)_{l-1}}{(\theta+1)_{l-1}(-n+1)_{l-1}l!} \]

\[ = -\frac{\theta n}{r(r+\theta-1)} \left[ -1 + \sum_{l=0}^{r} \frac{(-n+i)_{l}(-r)_{l}(r+\theta-1)_{l}}{(\theta)_{l}(-n)_{l}l!} \right] \]

\[ = \frac{\theta n}{r(r+\theta-1)} [1 - {}_{3}F_{2}(-r,-n+i,r+\theta-1;\theta,-n;1)]. \]

Similarly,

\[ n \cdot {}_{4}F_{3}(-n+1) = \frac{\theta n}{r(r+\theta-1)} [1 - {}_{3}F_{2}(-r,-n,r+\theta-1;\theta,-n;1)] \]

\[ = \frac{\theta n}{r(r+\theta-1)} [1 - {}_{2}F_{1}(-r,r+\theta-1;\theta;1)] \]

\[ = \frac{\theta n}{r(r+\theta-1)} \] \hspace{1cm} (3.2)

by the Chu–Vandermonde formula (see, e.g., [2, (7.16)]). Hence

\[ (n-i) \cdot {}_{4}F_{3}(-n+i+1) - n \cdot {}_{4}F_{3}(-n+1) \]

\[ = -\frac{\theta n}{r(r+\theta-1)} \cdot {}_{3}F_{2}(-r,-n+i,r+\theta-1;\theta,-n;1). \]

Plugging this into the previous expression of \( y_{ri} \), we complete the proof in the case \( i < n \).
If \( i = n \), then using Lemma 3 and (3.2), we obtain
\[
y_{rn} = -\frac{1}{\sqrt{n\Theta(n)}} \sum_{r-1} (n - 1) = -\frac{1}{\sqrt{n\Theta(n)}} \frac{(\theta + 1)_{n-1}}{(n - 1)!} \frac{\theta}{r(r + \theta - 1)}
\]
\[
= -\frac{\sqrt{n}}{r(r + \theta - 1)}.
\]
This is consistent with expression (3.1) given in the lemma. Lemma 4 is proved. \( \square \)

**Proof of Theorem 3.** Let \( 1 \leq r \leq n \). Recall that \( q_k(x), 0 \leq k \leq r \), span the subspace of polynomials whose degrees do not exceed \( r \). Analyze the polynomial appearing in Lemma 4,
\[
\Phi_r(x) = 3F_2(-r, -n + x, r + \theta - 1; \theta, -n; 1).
\]
As we have seen proving (3.2), we have \( \Phi_r(0) = 0 \). Hence \( x^{-1}\Phi_r(x) \) is a polynomial of degree \( r - 1 \), and thus there exist constants \( c_k \in \mathbb{R} \) such that
\[
\Phi_r(x) = x \sum_{k=0}^{r-1} c_k q_r(x).
\]
The leading coefficients of the polynomials \( \Phi_r(x) \) and \( q_r(x) \) are, respectively,
\[
\frac{(-1)^r(r + \theta - 1)_r}{(\theta)_r(-n)_r}, \quad \frac{(r + \theta)_{r-1}}{r!(-n + 1)_{r-1}}.
\]
Consequently,
\[
c_{r-1} = \frac{(-1)^r(r + \theta - 1)_r}{(\theta)_r(-n)_r} \frac{r!(-n + 1)_{r-1}}{(r + \theta)_{r-1}} = \frac{(-1)^{r-1}r!(r + \theta - 1)}{(\theta)_r n}.
\]
Now the result of Lemma 4 can be rewritten as follows:
\[
y_{ri} = -\left(\frac{\Theta(n-i)}{i}\right)^{1/2} \frac{ni}{r(r + \theta - 1)} \left[ c_{r-1} q_r(i) + \sum_{k=0}^{r-2} c_k q_k(i) \right]
\]
\[
= (i\Theta(n-i))^{1/2} \left[ \frac{(-1)^r(r - 1)!}{(\theta)_r} q_{r-1}(i) + \sum_{k=0}^{r-2} d_k q_k(i) \right]
\]
with some coefficients \( d_k = d_k(n, r, \theta) \) for all \( 1 \leq i \leq r \leq n \). Note that the fraction in the brackets is just \( \mu_r \) found in Theorem 2.

For \( 1 \leq r \leq l \leq n \), applying the last formula and the definition of the inner product (1.8), we obtain
\[
\bar{e}_l M_n e'_r = \pi_{r-1}^{-1} \bar{e}_l y'_r \frac{1}{\pi_{l-1} \pi_{r-1}} \left[ \mu_r \langle q_{l-1}, q_{r-1} \rangle + \sum_{k=0}^{r-2} d_k \langle q_{l-1}, q_k \rangle \right]
\]
\[
= \mu_r \delta_{rl}
\]
by orthogonality. This shows that each \( \bar{e}_r \) in the basis is the eigenvector for \( M_n \) corresponding to \( \mu_r \).

Theorem 3 is proved. \( \square \)
**Concluding remark.** Comparing the expression $\bar{e}_r M_n = \mu_r \bar{e}_r$ with $\pi_{r-1}^{-1} \bar{y}_r$ given by Lemma 4, we arrive at a seemingly new relation of the generalized hypergeometric functions. For $1 \leq i, r \leq n$ and $\theta > 0$, we have that

$$(-1)^{r-1} r! \cdot \binom{3F_2}{-r+1, -i+1, r+\theta; 2, -n+1; 1} = (\theta)_{r-1} n \cdot \binom{3F_2}{-r, -n+i, r+\theta-1; \theta, -n; 1}.$$ 

Derivation of it using an appropriate sequence of the so-called contiguous relations (see [19]) would not be short.

**Acknowledgment.** The authors appreciate the benevolent remarks presented by two anonymous reviewers.

**References**

1. R. Arratia, A.D. Barbour, and S. Tavaré, *Logarithmic Combinatorial Structures: a Probabilistic Approach*, EMS Publishing House, Zürich, 2003.

2. R. Askey, *Orthogonal Polynomials and Special Functions*, SIAM, Philadelphia, PA, 1975.

3. R. Beals and R. Wong, *Special Functions*, Cambridge Univ. Press, Cambridge, 2012.

4. H. Crane, The ubiquitous Ewens sampling formula, *Stat. Sci.*, 31:1–19, 2016.

5. W. Ewens, The sampling theory of selectively neutral alleles, *Theor. Popul. Biol.*, 3:87–112, 1972.

6. S. Feng, *The Poisson–Dirichlet Distribution and Related Topics*, Springer, Berlin, 2010.

7. N.S. Johnson, S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distributions*, Wiley, New York, 1997.

8. J. Klimavičius and E. Manstavičius, The Turán–Kubilius inequality on permutations, *Ann. Univ. Sci. Budap. Rolando Eötvös, Sect. Comput.*, 48:45–51, 2018.

9. R. Koekoek, P.A. Lesky, and R.F. Swarttouw, *Hypergeometric Orthogonal Polynomials and Their q-Analogues*, Springer, Berlin, 2010.

10. J. Kubilius, Estimating the second central moment for strongly additive arithmetic functions, *Liet. Mat. Rink.*, 23:110–117, 1983 (in Russian).

11. E. Manstavičius, Conditional probabilities in combinatorics. the cost of dependence, in M. Huškov’a and M. Janžura (Eds.), *Prague Stochastics. 7th Prague Symposium on Asymptotic Statistics and 15th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Charles University, Matfyzpress, Prague, 2006, pp. 523–532.

12. E. Manstavičius, Moments of additive functions defined on the symmetric group, *Acta Appl. Math.*, 97:119–127, 2007.

13. E. Manstavičius, An analytic method in probabilistic combinatorics, *Osaka J. Math.*, 46:273–290, 2009.

14. E. Manstavičius, A sharp inequality for the variance with respect to the Ewens sampling formula, *Random Struct. Algorithms*, 57(4):1303–1313, 2020.

15. E. Manstavičius and V. Stepanauskas, On variance of an additive function with respect to a generalized Ewens probability, in M. Bousquet-Mélou and M. Soria (Eds.), *Proceedings of the 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, Discrete Math. Theor. Comput. Sci., Proc., BA, Association DMTCS, Nancy, 2014, pp. 301–311.

16. E. Manstavičius and Ž. Žilinskas, On a variance related to the Ewens sampling formula, *Nonlinear Anal. Model. Control*, 16(4):453–466, 2011.
17. A.P. Prudnikov, Yu.A. Brychkov, and O.I. Marichev, *Integrals and Series, Vol. 3: More Special Functions*, Gordon and Breach, New York, 1990.

18. A.P. Prudnikov, Yu.A. Brychkov, and O.I. Marichev, *Integrals and Series, Vol. 1*, 4th ed., Taylor & Francis, 1998.

19. E.D. Rainville, *Special Functions*, Macmillan, New York, 1960.

20. G.A. Watterson, The sampling theory of selectively neutral alleles diffusion model, *Adv. Appl. Probab.*, 6:463–488, 1974.

21. K. Wieand, *Eigenvalue Distributions of Random Matrices in the Permutation Group and Compact Lie Groups*, PhD thesis, Harvard University, 1998.

22. V. Zacharovas, Distribution of the logarithm of the order of a random permutation, *Liet. Mat. Rink.*, 44:372–406, 2004 (in Russian).