On a variance related to the Ewens sampling formula

Eugenijus Manstavičius, Žydrūnas Žilinskas
Faculty of Mathematics and Informatics, Vilnius University
Naugarduko str. 24, LT-03225 Vilnius, Lithuania
eugenijus.manstavicius@mif.vu.lt

Received: 12, March 2011 / Revised: 14 November 2011 / Published online: 7 December 2011

Abstract. A one-parameter multivariate distribution, called the Ewens sampling formula, was introduced in 1972 to model the mutation phenomenon in genetics. The case discussed in this note goes back to Lynch’s theorem in the random binary search tree theory. We examine an additive statistics, being a sum of dependent random variables, and find an upper bound of its variance in terms of the sum of variances of summands. The asymptotically best constant in this estimate is established as the dimension increases. The approach is based on approximation of the extremal eigenvalues of appropriate integral operators and matrices.

Keywords: random permutation, cycle structure, integral operator, matrix eigenvalue, Jacobi polynomial.

1 Introduction and result

Let $S_n$ denote the symmetric group of permutations $\sigma$ acting on $n \geq 1$ letters. Each $\sigma \in S_n$ has a unique representation (up to the order) by the product of independent cycles $\kappa_i$:

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

where $w = w(\sigma)$ denotes the number of cycles. Denote by $k_j(\sigma) \geq 0$ the number of cycles in (1) of length $j$ for $1 \leq j \leq n$ and $\bar{k}(\sigma) := (k_1(\sigma), \ldots, k_n(\sigma))$. The latter is called the cycle vector of permutation $\sigma$. Let $\nu_{n, \theta}$ be the Ewens probability measure on $S_n$ defined by

$$\nu_{n, \theta}(\{\sigma\}) = \theta^{w(\sigma)} / \left(\theta^{n+1} \cdots (\theta + n - 1)\right)$$

$$=: \theta^{w(\sigma)} / (\theta^n), \quad \sigma \in S_n,$$

where $\theta > 0$ is a parameter. If we set $\ell(\bar{s}) = 1s_1 + \cdots + ns_n$ for a vector $\bar{s} = (s_1, \ldots, s_n) \in \mathbb{Z}^n_+$, then $\ell(\bar{k}(\sigma)) = n$ and, as it is shown in [1],

$$\nu_{n, \theta}(\bar{k}(\sigma) = \bar{s}) = \mathbf{1}\{\ell(\bar{s}) = n\} \frac{n!}{\theta^n} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{s_j} \frac{1}{s_j!}$$

$$= P(\bar{\xi}_\theta = \bar{s} \mid \ell(\bar{\xi}_\theta) = n),$$

$$\copyright$$ Vilnius University, 2011
where $\xi_\theta = (\xi_{1,\theta}, \ldots, \xi_{n,\theta})$ and $\xi_{j,\theta}, 1 \leq j \leq n$, are mutually independent Poisson r.vs given in some probability space $(\Omega, \mathcal{F}, P)$ with parameter $E\xi_{j,\theta} = \theta/j$.

The probability in (2), assigned to the vector $\bar{s} \in \mathbb{Z}_n^*$, is called the Ewens sampling formula. It has been introduced by W.J. Ewens [2] to describe the sampling distribution of a sample of $n$ genes from a large population. In this case, the allelic partition $\bar{s} = (s_1, \ldots, s_n)$, contains all the information available in a sample, that is, $s_j$ denotes the number of alleles represented $j$ times in it, $j = 1, \ldots, n$. In the so-called neutral alleles model of population genetics, the parameter $\theta$ is interpreted as the mutation rate (see [3]). For a comprehensive account of recent applications in combinatorics and statistics, we refer to [1] and [4]. Now, we just mention that the case $\theta = 2$ explored in the present paper has some connections to the random binary search tree theory (see [5]).

Apart from $w(\sigma)$, other statistics, called completely additive functions,

$$h(\sigma) := h_a(\sigma) := a_1k_1(\sigma) + \cdots + a_nk_n(\sigma),$$

where $a := (a_1, \ldots, a_n) \in \mathbb{R}^n$ is a non-zero vector, appear in applications rather often. For instance, $h(\sigma)$ with $a_j = \log j$, $j \leq n$, is a good approximation for the logarithm of the group-theoretical order of $\sigma \in S_n$ (see [1]). The case with $a_j = \{xj\}$, where $\{u\}$ stands for the fractional part of $u \in \mathbb{R}$, is met in the theory of random permutation matrices (see [6]).

By (2), under $\nu_{n,\theta}$, the function $h(\sigma)$ is a sum of dependent r.vs. This fact raises some obstacles, seen already in the analysis of power moments carried out by the first author [7, 8] and [9] in the case $\theta = 1$. To overcome the difficulties arising from the dependency, proving limit theorems for $h(\sigma)$, one needs specified approaches (see, for instance, [10, 11] or [12] and the references therein). We now draw the reader’s attention to the variance.

Denote by $A_{n,\theta}(\bar{a}) := E_{n,\theta}h(\sigma)$ and $D_{n,\theta}(\bar{a}) := \text{Var}_{n,\theta}h(\sigma)$ the mean value and the variance of function $h(\sigma)$ under the probability measure $\nu_{n,\theta}$. Set

$$\tau_{n,\theta} := \sup_{\bar{a} \neq \bar{0}} \left( D_{n,\theta}(\bar{a}) / \sum_{j \leq n} \text{Var}_{n,\theta}(a_jk_j(\sigma)) \right).$$

The problem is to estimate its discrepancy from 1 which is an indicator of the dependence among the summands. The first author [7] has succeeded to explore the case $\theta = 1$.

**Theorem M.** We have

$$\tau_{n,1} = \frac{3}{2} + O\left(\frac{1}{n}\right)$$

as $n \to \infty$.

A sketchy proof of this theorem is given in [7]. It is based on the spectral analysis of some integral operators. Nevertheless, the same approach to $\tau_{n,\theta}$ for other $\theta > 0$ leads to different operators in each case. Therefore, having the aim to expose our method in full detail and to give an instance of another appearing operator, we now chose the case $\theta = 2$.  

www.mii.lt/NA
Theorem. We have
\[ \tau_{n,2} = \frac{4}{3} + O\left(\frac{1}{n}\right) \]
as \( n \to \infty \).

The idea of our proof goes back to the number theoretical papers by J. Kubilius [13] and [14] and is explained in the next section. To our knowledge, apart from [7], it has not been applied in combinatorics.

The lower bound of \( \tau_{n,2} \) is found in Section 3 and the upper estimate is obtained in the last section.

2 An idea

First of all, we express the variance as a quadratic form. It appears that the eigenvalues of appropriate integral operators well approximate the eigenvalues of the involved matrices as their order increases. Further, the eigenfunctions of operators are used to find the vectors \( \bar{a} \) giving the extremal values of the variances.

To find an expression of the variance \( D_{n,\theta}(\bar{a}) \), we apply the following G. A. Watterson’s result.

Lemma 1. Set \( x_r := x(x-1) \cdots (x-r+1) \) if \( r \in \mathbb{Z}_+ \) and, for arbitrary \( l \in \mathbb{N} \) and \( r_1, \ldots, r_l \in \mathbb{Z}_+ \), define \( m := lr_1 + \cdots + lr_l \). Then
\[
E_{n,\theta}(k_1(\sigma_{(r_1)}) \cdots k_l(\sigma_{(r_l)})) = 1 \{ m \leq n \} \left( \frac{\theta + n - m - 1}{n - m} \right) \left( \frac{\theta + n - 1}{n} \right)^{-1} \prod_{j=1}^{l} \left( \frac{\theta}{j} \right)^{r_j}.
\]

Proof. See [15] or [1, p. 96]. □

Lemma 2. We have \( D_{n,2}(\bar{a}) = 2B(\bar{a}) - 4\Delta(\bar{a}) \), where
\[
B(\bar{a}) = \sum_{j \leq n} a_j^2 \left( 1 - \frac{j}{n+1} \right),
\]
and
\[
\Delta(\bar{a}) = \sum_{i,j \leq n, i+j \geq n} \frac{a_i a_j}{ij} \left( 1 - \frac{i}{n+1} \right) \left( 1 - \frac{j}{n+1} \right) + \frac{1}{(n+1)^2} \sum_{i+j \leq n} a_i a_j.
\]

Proof. Since \( x(0) = 1 \), applying Lemma 1 for \( l = m = j \cdot 1 \) and \( r_i = 0 \) if \( 1 \leq i \leq j - 1 \), we have
\[
A_{n,2}(\bar{a}) = 2 \sum_{j \leq n} a_j \left( 1 - \frac{j}{n+1} \right).
\]
Similarly,
\[
E_{n,2} h^2(\sigma) = \sum_{i,j \leq n} a_i a_j E_{n,2} (k_i(\sigma) k_j(\sigma))
\]
\[
= 4 \sum_{i+j \leq n} \frac{a_i a_j}{ij} \left(1 - \frac{i+j}{n+1}\right) + 2 \sum_{j \leq n} \frac{a_j^2}{j} \left(1 - \frac{j}{n+1}\right).
\]

In the second step, we used Lemma 1 separately in the cases \(i \neq j\) and \(i = j\). In the latter, dealing with \(E_{n,2} k_i(\sigma) k_j(\sigma)\), we also applied the relation \(x^2 = x_{(2)} + x_{(1)}\). Inserting these expressions into the equality \(D_{n,2}(\vec{a}) = E_{n,2} h^2(\sigma) - (A_{n,2}(\vec{a}))^2\), we obtain the desired formula.

The lemma is proved. \(\Box\)

**Corollary 1.** For all \(n \geq 1\), we have
\[
D_{n,2}(\vec{a}) \leq 4B(\vec{a}) \tag{4}
\]
and
\[
\left| \sum_{j \leq n} \text{Var}_{n,2}(a_j k_j(\sigma)) - 2B(\vec{a}) \right| \leq \frac{6}{n} B(\vec{a}) \tag{5}
\]

**Proof.** If \(a_j, j \leq n\), are of one sign, then, omitting \(\Delta(\vec{a}) \geq 0\) in the expression of variance obtained in Lemma 2, we have \(D_{n,2}(\vec{a}) \leq 2B(\vec{a})\) for all \(n \geq 1\) and \(\vec{a} \in \mathbb{R}^n\).

Further, splitting \(a_j = a_j^+ - a_j^-\), where \(a_j^+\) and \(a_j^-\) are respectively the positive and negative parts of \(a_j\), we define \(\vec{a}' = (a_1^+, \ldots, a_n^+)\) and \(\vec{a}'' = (a_1^-, \ldots, a_n^-)\). Then, by virtue of \((x+y)^2 \leq 2x^2 + 2y^2, x, y \in \mathbb{R}\),
\[
D_{n,2}(\vec{a}) = D_{n,2}(\vec{a}' - \vec{a}'') \leq 2D_{n,2}(\vec{a}') + 2D_{n,2}(\vec{a}'').
\]

Now, applying the just proved inequality twice, we obtain
\[
D_{n,2}(\vec{a}) \leq 4B(\vec{a}') + 4B(\vec{a}'') = 4B(\vec{a}).
\]

To prove (5), from Lemma 1, we have
\[
\sum_{j \leq n} \text{Var}_{n,2}(a_j k_j(\sigma)) - 2B(\vec{a})
\]
\[
= 4 \sum_{j \leq n/2} \frac{a_j^2}{j^2} \left(1 - \frac{2j}{n+1}\right) - 4 \sum_{j \leq n} \frac{a_j^2}{j^2} \left(1 - \frac{j}{n+1}\right)^2
\]
\[
= - \frac{4}{(n+1)^2} \sum_{j \leq n/2} a_j^2 - 4 \sum_{n/2 \leq j \leq n} \frac{a_j^2}{j^2} \left(1 - \frac{j}{n+1}\right)^2
\]
\[
= -4(\Sigma' + \Sigma'').
\]
Now,
\[
\Sigma' \leq \frac{1}{n} \sum_{j \leq n/2} \frac{a_j^2}{j} \left(1 - \frac{j}{n+1}\right)
\]
and
\[
\Sigma'' \leq \frac{3}{2n} \sum_{n/2 < j \leq n} \frac{a_j^2}{j} \left(1 - \frac{j}{n+1}\right).
\]
Inserting the last inequalities into the previous expression, we complete the proof of (5).

The corollary is proved.

By virtue of (5), instead of \( \tau_{n,2} \), we may examine the ratio \( \Delta(\tilde{a})/B(\tilde{a}) \). Our idea lays in choosing the vectors \( \tilde{a} \) so that these quadratic forms could be approximated by appropriate Riemann integrals. The natural choice is
\[
a_j := a_{j,n} := \frac{j}{n+1} g \left(\frac{j}{n+1}\right), \quad 1 \leq j \leq n,
\]
where \( g : [0,1] \rightarrow \mathbb{R} \) is a continuous function. For convenience, in the sums below, we formally add one more summand corresponding to \( j = n+1 \) though it equals zero at all places of appearance. Then
\[
B(\tilde{a}) \approx \int_0^1 x(1-x)g^2(x) \, dx.
\]
Setting
\[
\gamma_j(\tilde{a}) := \left(1 - \frac{j}{n+1}\right) \sum_{n-j+1 < i \leq n+1} \frac{a_i}{i} \left(1 - \frac{i}{n+1}\right) + \frac{j}{(n+1)^2} \sum_{i \leq n-j+1} a_i
\]
for \( 1 \leq j \leq n+1 \), we have a more convenient expression
\[
\Delta(\tilde{a}) = \sum_{j \leq n+1} \frac{a_j}{j} \gamma_j(\tilde{a}).
\]
This leads to
\[
\Delta(\tilde{a}) \approx \int_0^1 g(x) \left[ (1-x) \int_{1-x}^1 (1-u)g(u) \, du + x \int_0^1 ug(u) \, du \right] \, dx.
\]
Now, assume that \( g(x), 0 \leq x \leq 1 \), is a solution to the equation
\[
(1-x) \int_{1-x}^1 (1-u)g(u) \, du + x \int_0^1 ug(u) \, du = \lambda x(1-x)g(x), \quad (6)
\]
with some $\lambda \in \mathbb{R}$. Then, for the additive function $h(\sigma) = h_0(\sigma)$ defined in (3) via such $\tilde{a}$,
$$D_{M,2}(\tilde{a}) \approx 2(1 - 2\lambda)B(\tilde{a}).$$

If our intuition is true, among all these $\lambda$ we may look for the values giving the extremes of ratio $D_{n,2}/B(\tilde{a})$. Following this idea, we have to examine the operator
$$g(x) \mapsto \frac{1}{x} \int_{1-x}^{1} (1 - u)g(u) \, du + \frac{1}{1 - x} \int_{0}^{1-x} ug(u) \, du$$
defined on the space of continuous functions $g : [0, 1] \to \mathbb{R}$ and its eigenvalues $\lambda$. Using the substitutions $g(u) = p(2u - 1)$ and $y = 2x - 1$ with a continuous function $p : [-1, 1] \to \mathbb{R}$, from (6), we arrive at the equation
$$(1 - y)^{-\frac{1}{2}} \int_{-y}^{1} (1 - t)p(t) \, dt + (1 + y)^{-\frac{1}{2}} \int_{-1}^{(1 + y)p(t) \, dt} = 2\lambda(1 - y^2)p(y).$$

The solutions to it are twice differentiable, therefore they also satisfy
$$\lambda(1 - y^2)p''(y) - 4\lambda y p'(y) + p(-y) - 2\lambda p(y) = 0.$$If $\lambda \neq 0$, for even and uneven functions $p(y)$, respectively, this leads to the differential equations
$$(1 - y^2)p''(y) - 4\lambda y p'(y) + (\pm\lambda^{-1} - 2)p(y) = 0.$$The latter are well known in the theory of Jacobi polynomials $P^{(1,1)}_r(t), r \geq 0$, which are defined (see in [16, Sect. V.2] or [17, Sect. II.7]) by
$$P^{(1,1)}_r(t) = (1 - t^2)^{\frac{r}{2}} \sum_{k=0}^{r} \binom{r+1}{k} \binom{r+1}{k} (t - 1)^{k} (t + 1)^{r-k}.$$We will use their properties listed in the next lemma.

**Lemma 3.** Let $\delta_{mr}$ be the Kronecker symbol and $0 \leq m \leq r$. Then

(i) $$\int_{-1}^{1} (1 - t^2) P^{(1,1)}_{m}(t) P^{(1,1)}(t) \, dt = \frac{8(r + 1)\delta_{mr}}{(2r + 3)(r + 2)};$$

(ii) $P^{(1,1)}_r(t)$ satisfies the differential equation
$$(1 - t^2)p''(t) - 4tp'(t) + r(r + 3)p(t) = 0;$$
On a variance related to the Ewens sampling formula

(iii) \[ P_r^{(1,1)}(-t) = (-1)^r P_r^{(1,1)}(t). \]

**Proof.** See [16, Sect. V.2] and [17, Sect. II.7].

**Corollary 2.** The operator defined by (7) has the eigenfunctions
\[ g_r(x) := P_r^{(1,1)}(2x - 1), \quad 0 \leq x \leq 1, \]
corresponding to the eigenvalues
\[ \lambda_r := \frac{(-1)^r}{(r+1)(r+2)}, \]
where \( r = 0, 1, \ldots \).

**Proof.** Applying \( 2 + r(r+3) = (r+1)(r+2) \), the properties (ii) and (iii) given in the lemma, we see that equation (8) is satisfied by \( p(y) = P_r^{(1,1)}(y) \) if \( \lambda = \lambda_r \). Recalling the former substitutions, we complete the proof.

## 3 The lower bound

Let us keep our previous notation. As we have noted, the numbers \( \lambda_r \) and the vectors \( \bar{a}^r := (a_{r1}, \ldots, a_{rn}) \), where \( a_{rj} := (j/(n+1))g_r(j/(n+1)), \) \( 1 \leq j \leq n \) and \( r \geq 0 \), are worth to be exploited. The technical calculations are presented in a few lemmata, the most of them are based on the next well known Koksma inequality.

**Lemma 4.** If \( f : [0, 1] \to \mathbb{R} \) is continuously differentiable and \( N \in \mathbb{N} \), then
\[
\left| \frac{1}{N} \sum_{j \leq N} f \left( \frac{j}{N} \right) - \int_0^1 f(x) \, dx \right| \leq \frac{1}{N} \int_0^1 \left| f'(x) \right| \, dx.
\]

**Proof.** See, for instance, [18, Sect. 2.5].

Afterwards, all remainder term estimates will be dependent on \( r \) only.

**Lemma 5.** For each \( r \geq 0 \), we have
\[
B(\bar{a}^r) = \frac{(r+1)}{(2r+3)(r+2)} + O \left( \frac{1}{n} \right)
\]
and
\[
\Delta(\bar{a}^r) = \frac{(-1)^r}{(2r+3)(r+2)^2} + O \left( \frac{1}{n} \right)
\]
as \( n \to \infty \).
Proof. By the Koksma inequality and relation (i) of Lemma 3, we have

\[ B(\bar{a}^r) = \int_0^1 x(1 - x)g_r^2(x) \, dx + O\left(\frac{1}{n}\right) \]
\[ = \frac{1}{8} \int_{-1}^1 (1 - t^2) (P_r^{(1,1)}(t))^2 \, dt + O\left(\frac{1}{n}\right) \]
\[ = \frac{r + 1}{(2r + 3)(r + 2)} + O\left(\frac{1}{n}\right). \]

Calculating \( \gamma_j(\bar{a}^r) \), we introduce the temporary notation \( K := n - j + 1 \),

\[ f_1(x) := \left(1 - \frac{Jx + K}{n + 1}\right)g_r\left(\frac{Jx + K}{n + 1}\right), \quad f_2(x) := \frac{Kx}{n + 1}g_r\left(\frac{Kx}{n + 1}\right) \]

and arrive at

\[ \gamma_j(\bar{a}^r) := \frac{K}{(n + 1)^2} \sum_{1 \leq i \leq j} f_1\left(\frac{i}{j}\right) + \frac{j}{(n + 1)^2} \sum_{1 \leq i \leq K} f_2\left(\frac{i}{K}\right) + O\left(\frac{KJ}{(n + 1)^3}\right). \]

Now, applying Lemma 4 for \( f(x) = f_l(x), l = 1, 2 \), and exploiting also equation (6), we obtain

\[ \gamma_j(\bar{a}^r) = \left(1 - \frac{j}{n + 1}\right) \int_{1 - j/(n + 1)}^1 (1 - t)g_r(t) \, dt \]
\[ + \frac{j}{n + 1} \int_0^{1 - j/(n + 1)} tg_r(t) \, dt + O\left(\frac{j}{n^2}\left(1 - \frac{j}{n + 1}\right)\right) \]
\[ = \lambda_r \frac{j}{n + 1}\left(1 - \frac{j}{n + 1}\right)g_r\left(\frac{j}{n + 1}\right) + O\left(\frac{j}{n^2}\left(1 - \frac{j}{n + 1}\right)\right) \]
\[ = \lambda_r \frac{j}{n + 1}\left(1 - \frac{j}{n + 1}\right) + O\left(\frac{j}{n^2}\left(1 - \frac{j}{n + 1}\right)\right) \]  (9)

for \( 1 \leq j \leq n + 1 \). Hence again by (i) of Lemma 3,

\[ \Delta(\bar{a}^r) = \frac{1}{n + 1} \sum_{j \leq n + 1} g_r\left(\frac{j}{n + 1}\right)\gamma_j(\bar{a}^r) \]
\[ = \lambda_r \int_0^1 x(1 - x)g_r^2(x) \, dx + O\left(\frac{1}{n}\right) \]
\[ = \frac{\lambda_r (r + 1)}{(2r + 3)(r + 2)} + O\left(\frac{1}{n}\right) = \frac{(-1)^r}{(2r + 3)(r + 2)} + O\left(\frac{1}{n}\right). \]

The lemma is proved. \( \square \)
Corollary 3. We have

$$\tau_{n,2} \geq \frac{4}{3} + O\left(\frac{1}{n}\right)$$

as \(n \to \infty\).

Proof. It suffices to apply (5) and Lemma 5 for \(r = 1\). Indeed,

$$\tau_{n,2} \geq D_{n,\theta}(\bar{a}^1) \sum_{j=1}^{\lessgtr n} \text{Var}_{n,\theta}(a_k^j k^j(\sigma))$$

$$= \left(2B(\bar{a}^1) - 4\Delta(\bar{a}^1)\right)/(2B(\bar{a}^1) + O(n^{-1}))$$

$$= \left(\frac{16}{45} + O(n^{-1})\right) / \left(\frac{1}{15} + O(n^{-1})\right) = \frac{4}{3} + O(n^{-1}).$$

The corollary is proved.

4 The upper bound

Let \(\| \cdot \|\) denote the Euclidean norm in \(\mathbb{R}^n\). Dealing with the quadratic forms \(B(\bar{a})\) and \(\Delta(\bar{a})\), it is convenient to apply the substitution

$$a_j = \frac{1}{2} \left(1 - \frac{j}{n+1}\right)^{-1/2} x_j, \quad 1 \leq j \leq n,$$

converting \(B(\bar{a})\) to the sum of squares, that is, to \(\|\bar{x}\|^2\). Then \(\Delta(\bar{a})\) becomes the quadratic form \(Q(\bar{x}) =: \bar{x}Q\bar{x}'\), where \(\bar{x}'\) is the vector-column and \(Q\) is a symmetric matrix. If \(q_{ij}\), \(1 \leq i, j \leq n\), are the entries of the latter, then

$$q_{ij} = (ij)^{-1/2} \left(1 - \frac{i}{n+1}\right)^{1/2} \left(1 - \frac{j}{n+1}\right)^{1/2}$$

if \(i + j > n\) and

$$q_{ij} = \frac{(ij)^{1/2}}{(n+1)^2} \left(1 - \frac{i}{n+1}\right)^{-1/2} \left(1 - \frac{j}{n+1}\right)^{-1/2}$$

if \(i + j \leq n\).

Now, by virtue of (5),

$$\tau_{n,2} = 1 - 2 \inf_{\bar{x} \neq \bar{0}} \frac{\|Q(\bar{x})\|}{\|\bar{x}\|^2} + O\left(\frac{1}{n}\right)$$

$$= 1 - 2 \inf_{\bar{x} \neq \bar{0}} \frac{1}{\|\bar{x}\|^2} \sum_{j=1}^{n} \mu_j x_j^2 + O\left(\frac{1}{n}\right)$$

$$\leq 1 - 2 \min_{1 \leq j \leq n} \mu_j + O\left(\frac{1}{n}\right),$$

where \(\mu_j\), \(1 \leq j \leq n\), denote the eigenvalues of the matrix \(Q\). It remains to find their minimal value.
Lemma 6. Let \( \bar{v} \in \mathbb{R}^n \) and \( \alpha \in \mathbb{R} \) be arbitrary. If \( M \) is a symmetric real \( n \times n \) matrix, then there exists its eigenvalue \( \mu \) such that
\[
|\alpha - \mu|\|\bar{v}\| \leq \|\bar{v}M - \alpha \bar{v}\|.
\]

Proof. This is Lemma 5.6 in [19].

Lemma 7. Let \( r \geq 0 \) be fixed and \( \lambda_r \) be defined in Corollary 2. There exists an eigenvalue \( \mu_r \) (relabelled if necessary) of \( Q \) such that
\[
\mu_r = \lambda_r + O(r^{-1})
\]
provided that \( n \) is sufficiently large.

Proof. We apply the previous lemma with \( \alpha = \lambda_r \) and \( \bar{v} = \bar{v}^r := (v_{r1}, \ldots, v_{rn}) \), where
\[
v_{ri} = \frac{1}{n+1} \left( 1 - \frac{i}{n+1} \right)^{1/2} g_r \left( \frac{i}{n+1} \right), \quad i \leq n.
\]
If \( \bar{y}^r = (y_{r1}, \ldots, y_{rn}) = \bar{v}^r Q \), then, recalling the previous notation and using (9), we have
\[
y_{rj} = \sum_{i \leq n} g_{ij} v_{ri} = j^{-1/2} \left( 1 - \frac{j}{n+1} \right)^{-1/2} \gamma_j (\bar{a}^r)
\]
\[
= \lambda_r v_{rj} + O \left( \frac{j^{1/2}}{n^2} \left( 1 - \frac{j}{n+1} \right)^{1/2} \right).
\]
Hence
\[
\|\bar{v}^r Q - \lambda_r \bar{v}^r\| = O(n^{-1}).
\]
By Lemma 5, \( \|\bar{v}\|^2 = B(\bar{a}^r) \geq c_r > 0 \) for sufficiently large \( n \).

The claim now follows from Lemma 6.

The just proved lemma gives some numeration of the first eigenvalues \( \mu_r \) of the matrix \( Q \). There is no repetition of them if \( n \) is sufficiently large. Actually, we can even choose some unbounded sufficiently slowly increasing sequence of natural numbers \( r_n \) such that
\[
\max_{0 \leq r \leq r_n} |\mu_r - \lambda_r| \leq r_n^{-2}
\]
as \( n \to \infty \). Extend the numeration to list the remaining eigenvalues. The latter \( \mu_r \), \( r_n < r \leq n-1 \), maybe, are written with repetitions or repeating those with small indexes. However, we will prove that \( \mu_1 \) is the minimal among all of \( \mu_r \), where \( 0 \leq r \leq n-1 \).

Lemma 8. We have
\[
\sum_{0 \leq r \leq n-1} \mu_r^2 = \sum_{r \geq 0} \lambda_r^2 + O \left( \frac{\log n}{n} \right)
\]
\[
= \frac{\pi^2}{3} - 3 + O \left( n^{-1} \log n \right).
\]
and
\[ \min_{0 \leq r \leq n-1} \mu_r = \mu_1 \] (14)
if n is sufficiently large.

Proof. Observe that
\[ \sum_{r \geq 0} \lambda_r^2 = \sum_{r \geq 0} \left( \frac{1}{(r+1)^2} + \frac{1}{(r+2)^2} - 2 \left( \frac{1}{r+1} - \frac{1}{r+2} \right) \right) = \frac{\pi^2}{3} - 3. \] (15)

On the other hand, using the well known property of the matrix eigenvalues, we arrive at
\[ \sum_{0 \leq r \leq n-1} \mu_r^2 = \sum_{i,j \leq n} q_{ij}^2 = \sum_{i+j \leq n} 1 \left( 1 - \frac{i}{n+1} \right) \left( 1 - \frac{j}{n+1} \right) \]
\[ + \frac{1}{(n+1)^2} \sum_{i+j \leq n} ij \left( 1 - \frac{i}{n+1} \right)^{-1} \left( 1 - \frac{j}{n+1} \right)^{-1} \]
\[ =: \Sigma_1 + \Sigma_2. \] (16)

Now
\[ \Sigma_1 = \left( \sum_{n/2 < j \leq n} \frac{1}{j} \left( 1 - \frac{j}{n+1} \right) \right)^2 \]
\[ + 2 \sum_{j \leq n/2} \frac{1}{j} \left( 1 - \frac{j}{n+1} \right) \sum_{n-j < i \leq n} \frac{1}{i} \left( 1 - \frac{i}{n+1} \right) \]
\[ =: \Sigma_{11} + 2 \Sigma_{12}. \]

Further, approximating the sums by appropriate integrals, from Lemma 4 we obtain
\[ \Sigma_{11} = \left( \int_{1/2}^{1} \frac{1-x}{x} \, dx + O \left( \frac{1}{n} \right) \right)^2 = \left( \log 2 - \frac{1}{2} \right)^2 + O \left( n^{-1} \right) \]
and
\[ \Sigma_{12} = \sum_{j \leq n/2} \frac{1}{j} \left( 1 - \frac{j}{n+1} \right) \left( \int_{1-j/n}^{1} \frac{1-u}{u} \, du + O \left( \frac{1}{n} \right) \right) \]
\[ = \int_{1/2}^{1} \frac{1-x}{x} \left( - \log(1-x) - x \right) \, dx + O \left( \log \frac{n}{n} \right) \]
\[ =: I + O \left( n^{-1} \log n \right). \]
If $\text{Li}_2(x)$ denotes the dilogarithm function, then

\[
I = \text{Li}_2 \left( \frac{1}{2} \right) + \int_0^{1/2} (\log(1 - x) - (1 - x)) \, dx
\]

\[
= \text{Li}_2 \left( \frac{1}{2} \right) + \frac{\log 2}{2} - \frac{7}{8} = \frac{\pi^2}{12} - \frac{\log 2}{2} + \frac{\log 2}{2} - \frac{7}{8} \tag{17}
\]

(see [20, Sect. 27.7.3, p. 1004]).

Similarly,

\[
\Sigma_2 = \left( \frac{1}{n + 1} \sum_{j \leq n/2} \frac{j}{n + 1 - j} \right)^2
\]

\[+ \frac{2}{(n + 1)^2} \sum_{n/2 < j \leq n} \frac{j}{n + 1 - j} \sum_{i \leq n - j} \frac{i}{n + 1 - i}
\]

\[= \Sigma_{21} + 2\Sigma_{22}.
\]

Now again

\[
\Sigma_{21} = \left( \int_0^{1/2} \frac{x}{1 - x} \, dx + O \left( \frac{1}{n} \right) \right)^2 = \left( \log 2 - \frac{1}{2} \right)^2 + O(n^{-1})
\]

and

\[
\Sigma_{22} = \frac{1}{n + 1} \sum_{n/2 < j \leq n} \frac{j}{n + 1 - j} \left( \int_0^{1-j/n} \frac{v}{1 - v} \, dv + O \left( \frac{1}{n} \right) \right)
\]

\[= \int_{1/2}^1 \frac{u}{1 - u} \left( -\log u - (1 - u) \right) \, du + O \left( \frac{\log n}{n} \right)
\]

\[= I + O \left( n^{-1} \log n \right).
\]

Inserting the obtained values of $\Sigma_{ij}, i, j \in \{1, 2\}$, into formulas for $\Sigma_i, i \in \{1, 2\}$, and the latter into (16), we have

\[
\sum_{0 \leq r \leq n-1} \mu_r^2 = 2 \left( \log 2 - \frac{1}{2} \right)^2 + 4I + O \left( n^{-1} \log n \right).
\]

This and (17) yields

\[
\sum_{0 \leq r \leq n-1} \mu_r^2 = \frac{\pi^2}{3} - 3 + O \left( n^{-1} \log n \right).
\]
Now, equality (13) follows from the earlier found sum (15).

As we have noted after Lemma 7, the possible minimum, except $\mu_1$, could be among $\mu_r$, if $r_n < r \leq n - 1$. However, using the inequality (12), we have the estimate

$$\sum_{r_n < r \leq n - 1} \mu_r^2 = \sum_{r > r_n} \lambda_r^2 + O \left( \sum_{0 \leq r \leq r_n} |\lambda_r - \mu_r| \right) + O \left( \frac{\log n}{n} \right) \leq \int_{r_n}^\infty \frac{du}{(u + 1)^2(u + 2)^2} + o(1) = o(1),$$

showing that $\max_{r_n < r \leq n - 1} |\mu_r| = o(1)$ as $n \to \infty$. Hence by (12), the minimal eigenvalue is $\mu_1$ provided that $n$ is sufficiently large.

The lemma is proved.

**Proof of Theorem.** By virtue of the upper estimate (11) and (14) we see that $\tau_{n,2} \leq 1 - 2\mu_1 + O(n^{-1})$. Lemma 7 now yields $\tau_{n,2} \leq 4/3 + O(n^{-1})$. Recalling the lower estimate (10) we complete the proof.

The theorem is proved.

**Acknowledgement**

The authors thank the referees for the benevolent remarks which have helped to improve the exposition of the paper.

**References**


