The Ramanujan Journal, 2, 185–199 (1998)

The Berry–Esseen Bound in the Theory of Random Permutations

E.MANSTAVIČIUS * eugenijus.manstavicius@maf.vu.lt
Department of Mathematics, Vilnius University, Naugarduko str. 24, LT-2600 Vilnius, Lithuania

Professori Paulo Erdös in Memoriam

Received April 11, 1997; Accepted October 31, 1997

Abstract. The convergence rate in the central limit theorem for linear combinations of the cycle lengths of a random permutation is examined. It is shown that, in contrast to the Berry-Esseen theorem, the optimal estimate in terms of the sum of the third absolute moments has the exponent 2/3.

Keywords: random permutation, Berry-Esseen theorem, Central limit, remainder term, random mappings, symmetric group

AMS Classification Numbers: Primary – 60C05; Secondary – 60F05, 05A05

1. Introduction and results

Statistical group theory and probabilistic number theory are the fields to which P.Erdös has contributed many pioneering and enduring works. Studying his papers written mainly with P.Turán as well as more recent articles which deal with the value distribution problems of maps defined on the symmetric group $\mathbb{S}_n$ we could not shake off an impression that this direction has much in common with probabilistic number theory, nevertheless, the interaction between them is rather poor. We came to an opinion that in developing of the analytic tools, number theory is a bit ahead than similar branches of discrete mathematics. For instance, the survey [4] considered as the most comprehensive paper on analytic approaches of discrete mathematics can be compared to the Selberg– Delange method used to investigate mean values of multiplicative functions but we hardly could find an analogous influence of the method taking its background in the G.Halász' papers [6] or [7]. The articles [9], [14], [15] comprise a rare exception. Now solving the problem of the remainder term estimation in the central limit theorem we demonstrate other possibilities of this approach.

Let $\sigma \in \mathbb{S}_n$ be an arbitrary permutation and

$$\sigma = \kappa_1 \cdots \kappa_\omega \tag{1}$$

be its (unique up to the order) expression by the product of the independent cycles κ and $\omega = \omega(\sigma)$ be the number of the cycles comprising σ. Denote

$$\nu_n(\dots) = (n!)^{-1} \#\{\sigma \in \mathbb{S}_n : \dots\}.$$

* Partially supported by Grant from Lithuanian Foundation of Studies and Science.

In 1942 V.L.Goncharov [5] proved that

$$\nu_n\big(\omega(\sigma)-\log n< x\sqrt{\log n}\big)\to\Phi(x):=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}\,du\,.$$

Here and in what follows, the limit is taken with respect to $n\to\infty$. Starting a very fascinating series of papers on the asymptotic distribution of the group-theoretic order of the random permutation σ, P.Erdös and P.Turán [3] considered the sum $s(\sigma)$ of natural logarithms of different lengths $l(\kappa)$ of the cycles comprising σ. It was shown that

$$\nu_n\big(s(\sigma)-(1/2)\log^2 n<(1/\sqrt{3})\,x\,\log^{3/2}n\big)\to\Phi(x)\,.$$

The convergence rate in this relation was estimated by J.-L.Nicolas [16]. An improvement of the convergence rate estimates in the central limit theorem for the group-theoretic order function was given by A.D.Barbour and S.Tavaré [2]. Other relevant references can be found in the book [10] and in the recent lecture [18]. In addition, we note that random permutations, not necessarily taken with equal probabilities, comprise a rather significant object in applied mathematics (see [1] and the references therein). So we hope that our remark, though written in purely theoretical style, will be useful for those interested in analytic problems of the applied probability theory.

In what follows we adopt a few definitions from probabilistic number theory. Having in mind the examples of functions $\omega(\sigma)$ and $s(\sigma)$, we call the map $h:\mathbb{S}_n:\to\mathbb{R}$ *additive* if it satisfies the relation $h(\sigma)=h(\kappa_1)+\cdots+h(\kappa_\omega)$ for each σ having the expression (1). Similarly, the map $f:\mathbb{S}_n:\to\mathbb{C}$ satisfying the equality $f(\sigma)=f(\kappa_1)\cdots f(\kappa_\omega)$ is called *multiplicative*. Further, the function $g:\mathbb{S}_n\to\mathbb{C}$ will be called *class dependent*, or shortly, CD function if its values on cycles depend only on their lenghts, e.g., there exists a function $\widehat{g}:\mathbb{N}\to\mathbb{C}$ such that $g(\kappa)=\widehat{g}(l(\kappa))$. To argue the definition, we remind that each $\sigma\in\mathbb{S}_n$ belongs to a class of conjugate elements, which we denote by $\bar m:=(m_1,\dots,m_n)$ with $0\le m_k\le n/k$ and $1m_1+\cdots+nm_n=n$. The relation $\sigma\in\bar m$ means that σ consists of m_k cycles of the length k, $1\le k\le n$. The CD additive and multiplicative functions have the representations

$$h(\sigma)=\sum_{k=1}^{n}\widehat{h}(k)m_k,\quad f(\sigma)=\prod_{k=1}^{n}\widehat{f}(k)^{m_k}\tag{2}$$

with $m_k=m_k(\sigma)$. The general task is to describe these functions when the values $\widehat{h}(k)$ or $\widehat{f}(k)$, $k\ge1$ are given. The problem of weak convergence to a limit law of

$$\nu_n\big(h(\sigma)-\alpha(n)<x\beta(n)\big),$$

where $\alpha(n)$ and $\beta(n)>0$ are suitably chosen normalizing sequences, was considered in the paper [14]. According to the result of V.L.Goncharov [5], the distribution with respect to ν_n of the random variable $m_k(\sigma)$ tends to the Poissonian law having the parameter $1/k$ for each fixed k. Hence the choice of

$$\alpha(n)=\sum_{k=1}^{n}\frac{\widehat{h}(k)}{k},\quad\beta(n)=\bigg(\sum_{k=1}^{n}\frac{\widehat{h}^2(k)}{k}\bigg)^{1/2}$$

called *standard normalization* should be considered at the first place. That motivates the normalization used afterwards. We quote the following partial result.

THEOREM A ([14]). *Let $h_n(\sigma)$ be a sequence of real CD additive functions satisfying the condition*

$$\sum_{k=1}^{n} \frac{\widehat{h}_n^2(k)}{k} = 1 \tag{3}$$

and

$$A(n) := \sum_{k=1}^{n} \frac{\widehat{h}_n(k)}{k}\,.$$

If the Lindeberg type condition

$$\sum_{\substack{k\le n\\ |\hat{h}_n(k)|\ge\epsilon}} \frac{\widehat{h}_n^2(k)}{k} = o(1) \tag{4}$$

holds for each $\epsilon > 0$, then

$$\nu_n(x) := \nu_n\big(h_n(\sigma) - A(n) < x\big) = \Phi(x) + o(1) \tag{5}$$

uniformly in $x \in \mathbb{R}$ and also

$$\frac{1}{n!} \sum_{\sigma\in\mathbb{S}_n} \big(h_n(\sigma) - A(n)\big)^2 = 1 + o(1)\,. \tag{6}$$

Having in mind the Berry–Esseen estimate in the central limit theorem for sums of independent random variables (see V.V.Petrov [17], Chapter 5), we expect that the remainder in (5) can be estimated in terms of

$$L_n := \sum_{k=1}^{n} \frac{|\widehat{h}_n(k)|^3}{k}\,.$$

Observe that the relation $L_n = o(1)$ implies also the condition (4). It appears that dependence of the random variables $m_k(\sigma)$, $1 \le k \le n$ involved by (2) in the function $h_n(\sigma)$ makes a substantial influence.

In what follows, let the symbols O or $\ll$ contain absolute constants when there is no other indication, and

$$D_n = \sum_{\substack{1\le k,l\le n\\ k+l>n}} \frac{\widehat{h}_n(k)\widehat{h}_n(l)}{kl}\,.$$

In contrast to the above mentioned Berry–Esseen bound, we have the following results.

THEOREM 1. *Let $h_n(\sigma)$, $n \geq 1$ be a sequence of real CD additive functions satisfying the condition (3). Then*

$$R'_n := \sup_{x\in\mathbb{R}} \left| \nu_n(x) - \Phi(x) - \frac{D_n x}{2\sqrt{2\pi}} e^{-x^2/2} \right| \ll L_n.$$

COROLLARY. *We have*

$$R_n := \sup_{x\in\mathbb{R}} |\nu_n(x) - \Phi(x)| \ll L_n^{2/3}\,.$$

There exists a sequence of CD additive functions satisfying the condition (3) and $L_n = o(1)$ but such that

$$R_n \gg L_n^{2/3}\,.$$

THEOREM 2. *Let $h_n(\sigma)$, $n \geq 1$ be a sequence of real CD additive functions normalized so that*

$$\sum_{k=1}^{n} \frac{\widehat{h}_n^2(k)}{k} - D_n = 1\,. \tag{7}$$

Then with the same centralizing sequence $A(n)$, we have

$$R_n := \sup_{x\in\mathbb{R}} |\nu_n(x) - \Phi(x)| \ll L_n\,.$$

Theorems 1 and 2 are analogous to the results obtained by A.Mačiulis [13] for additive functions defined on $\mathbb{N}$. The proofs are based upon the Esseen inequality connecting the convergence rate of distribution functions to their characteristic functions and analysis of the last. If $g(\sigma) := \exp\{ith(\sigma)\}$, $t \in \mathbb{R}$ and, as above, the function $\widehat{g}$ is defined by $\widehat{g}(l(\kappa)) = g(\kappa)$, then the main difficulty is to find asymptotic formulae for

$$M_n(g) := \frac{1}{n!} \sum_{\sigma\in\mathbb{S}_n} g(\sigma) = \sum_{\bar{m}} \prod_{k=1}^{n} \left(\frac{\widehat{g}(k)}{k} \right)^{m_k} \frac{1}{m_k!}$$

uniform in parameters of g. Moreover, we have

$$\exp\left\{ \sum_{k=1}^{\infty} \frac{\widehat{g}(k) z^k}{k} \right\} = \sum_{n=0}^{\infty} M_n(g) z^n, \quad |z| < 1. \tag{8}$$

So, our task reduces to a problem in function theory. We hope that the analysis of the relations between the coefficients of the series in (8) done in the next two sections has independent interest.

2. The first analytic formula

Let $f(k),\ k \ge 1$ be complex numbers, depending, maybe, on n or other parameters. Denote

$$F(z) = \exp\left\{ \sum_{k=1}^{\infty} \frac{f(k)z^k}{k} \right\} =: \sum_{n=0}^{\infty} M_n z^n, \quad |z| < 1. \tag{9}$$

We will obtain asymptotic expressions of M_n in terms of $f(k)$. Since the values of $f(k)$, when $k > n$, make no influence onto M_n, we assume them equal to one. The remainder in the formula obtained in this section will involve the quantity

$$\rho(n,p) = \left(\sum_{k \le n} \frac{|f(k) - 1|^p}{k} \right)^{1/p}$$

where $p > 1$. Put

$$L(z) = \sum_{k \le n} \frac{f(k) - 1}{k} z^k, \quad z = re^{i\tau} := e^{-1/n + i\tau}, \quad \tau \in \mathbb{R}.$$

Let

$$I_j(n) = \frac{1}{2\pi i} \int_{|z|=r} \frac{(L(z) - L(1))^j}{(1 - z)z^{n+1}}\, dz, \quad j = 0, 1, \ldots.$$

Calculating the coefficients of the integrand, we have $I_0(n) = 1, \quad I_1(n) = 0$, and

$$I_2(n) = -\sum_{\substack{1 \le k, l \le n \\ k + l > n}} \frac{(f(k) - 1)(f(l) - 1)}{kl}\,.$$

We have the following result.

THEOREM 3. *Let $p > 1$. There exists sufficiently small $\delta = \delta(p)$ such that, if*

$$\rho := \rho(n,p) \le \delta\,, \tag{10}$$

then

$$M_n = \exp\{L(1)\} \left(1 + \sum_{j=2}^{N-1} \frac{I_j(n)}{j!} + O\big(\rho^N + n^{-c}\big) \right)$$

for each $N \ge 2$ with some constant $c = c(p) > 0$. The constant in the symbol O also depends on p only.

The proof of Theorem 3 goes along the lines drawn up by A.Mačiulis in the paper [13]. At first we prove few auxilliary results. The following estimate of the norm of a polynomial is perhaps known, but we have failed to find it in the literature.

LEMMA 1. *Let* $r = e^{-1/n}$*,* $s > \max\{2, p/(p-1)\}$*, and* $p > 1$*. For each polynomial*

$$P(z) = \sum_{k=1}^{n} a_k z^k, \quad a_k \in \mathbb{C}\,,$$

we have

$$\begin{aligned} \|P(z)\|_s \ &:= \ \left(\int_0^{2\pi} |P(re^{i\tau})|^s \, d\tau \right)^{1/s} \le C(s,p) n^{1-1/s} \left(\sum_{k=1}^{n} \frac{|a_k|^p}{k} \right)^{1/p} \\ &=: \ C(s,p) n^{1-1/s} Q\,. \end{aligned}$$

The constant $C(s,p)$ *depends only on* s *and* p*.*

Proof: The main idea takes its backgroud in the G.Halász' paper [6]. Let $\alpha = \min\{2, p\}$, $\beta = \alpha/(\alpha - 1)$, and $z = re^{i\tau}$. Observe that

$$\sum_{k=1}^{n} |a_k|^\alpha \le n Q^\alpha, \quad |P(z)| \le nQ\,.$$

Define

$$\Omega_j = \{\tau \in [0, 2\pi] : \ |P(ze^{i\tau})| > 2^{-j} nQ\}, \quad j = 0, 1, \ldots, j_0\,.$$

Hence using the partition

$$[0, 2\pi] = ([0, 2\pi] \setminus \Omega_{j_0}) \bigcup_{j=1}^{j_0} (\Omega_j \setminus \Omega_{j-1})\,,$$

we have

$$\int_0^{2\pi} |P(re^{i\tau})|^s \, d\tau \le n^s Q^s \left(2\pi 2^{-j_0 s} + 2^s \sum_{j=1}^{j_0} 2^{-js} \mu\Omega_j \right),$$

where $\mu\Omega$ stands for the Lebesgue measure of the set Ω. We see that the assertion of Lemma 1 will follow from the estimate

$$\mu\Omega_j \ll n^{-1} j 2^{\beta j} \tag{11}$$

with $\beta < s$. Here and in what follows the constant in the symbol $\ll$ depends at most on s and p.

In order to prove (11) when $j \ge 1$, we choose the points $\tau_l, \ l = 1, \ldots, n_j$ by induction. Let

$$\tau_1 = \inf \Omega_j, \quad \tau_{l+1} = \inf\{\tau \in \Omega_j : \ \tau \ge \tau_l + 1/n\}\,.$$

Then $\mu\Omega_j \le n_j/n$, and it remains to prove the estimate

$$n_j \ll j2^{\beta j}\,. \tag{12}$$

We can suppose that $n_j \geq 3$. If $z_l = \exp\{-i \arg P(re^{i\tau_l})\}$, then

$$\begin{aligned} 2^{-j} n Q n_j &\leq \sum_{l=1}^{n_j} z_l P(re^{i\tau_l}) = \sum_{k=1}^{n} a_k r^k \sum_{l=1}^{n_j} z_l e^{ik\tau_l} \leq \\ &\leq \left(\sum_{k=1}^{n} |a_k|^\alpha\right)^{1/\alpha} \left(\sum_{k=1}^{\infty} r^k \left|\sum_{l=1}^{n_j} z_l e^{ik\tau_l}\right|^\beta\right)^{1/\beta} \leq \\ &\leq n^{1/\alpha} Q \left(n_j^{\beta-2} \sum_{k=1}^{\infty} r^k \left|\sum_{l=1}^{n_j} z_l e^{ik\tau_l}\right|^2\right)^{1/\beta}. \end{aligned} \tag{13}$$

The double sum on the right hand side equals

$$\begin{aligned} \Sigma &:= \sum_{l=1}^{n_j} \sum_{m=1}^{n_j} z_l \bar{z}_m \sum_{k=1}^{\infty} r^k e^{ik(\tau_l - \tau_m)} \leq \\ &\leq 3 \sum_{l=1}^{n_j} \sum_{k=1}^{\infty} r^k + 4 \sum_{\substack{l,m=1 \\ 1/n \leq \tau_l - \tau_m \leq \pi}}^{n_j} \left|\sum_{k=1}^{\infty} r^k e^{ik(\tau_l - \tau_m)}\right|. \end{aligned}$$

Observing that $\tau_{m+k} - \tau_m \geq k/n$, we proceed

$$\Sigma \ll n n_j + n_j \sum_{k=1}^{n_j - 1} \max_{k/n \leq \tau \leq \pi} |1 - re^{i\tau}|^{-1} \ll n n_j \log n_j\,.$$

The last estimate and (13) imply $n_l \log n_j \ll 2^{j\beta}$. Hence we obtain (12).

Lemma 1 is proved. ■

Let $\tau_0 = \min\{e^{1/\rho}, \sqrt{n}\}/n$ and

$$l = \{z : |z| = r := e^{-1/n}\}, \quad l_0 = \{z \in l : |\tau| := |\arg z| \leq \tau_0\}\,,$$

$$l_1 = \{z \in l : \tau_0 < |\tau| \leq \pi\}\,.$$

The constants in the symbols O or $\ll$ will depend at most on p provided that δ is chosen smaller than some constant depending on p.

LEMMA 2. *Let* $p > 1$, $1/p + 1/q = 1$. *Then* $L(z) - L(1) \ll \rho \log^{1/q}(2 + |\tau| n)$ *and, for* $|\tau| \leq \tau_0$,

$$\exp\{L(z) - L(1)\} = \sum_{j=0}^{N-1} \frac{\big(L(z) - L(1)\big)^j}{j!} + O\left(\frac{|L(z) - L(1)|^N}{N!}\right).$$

Proof: We start with the inequalities

$$
\begin{aligned}
|L(z)-L(1)| &\le \sum_{k\le n}\frac{|f(k)-1|}{k}r^k|e^{ik\tau}-1|+\frac{1}{n}\sum_{k\le n}|f(k)-1| \\
&\le \rho(n)\Big(\Psi_n(\tau)^{1/q}+1\Big)\,, \qquad (14)
\end{aligned}
$$

where

$$
\Psi_n(\tau)=\sum_{k=1}^{\infty}\frac{r^{kq}|e^{ik\tau}-1|^q}{k}\,.
$$

Expanding into the Fourier series (see [11] or [8], Exercise 34), we have

$$
|1-e^{ix}|^q=a(q)+\sum_{\substack{m=-\infty\\ m\ne 0}}^{\infty}a_m(q)e^{imx},
$$

$$
a(q)=-\sum_{\substack{m=-\infty\\ m\ne 0}}^{\infty}a_m(q)=\frac{2^q\Gamma((1+q)/2)}{\sqrt{\pi}\Gamma((2+q)/2)}\,,
$$

for $x\in\mathbb{R}$, $q\ge 1$ with $a_m(q)\in\mathbb{R}$, $a_m(q)\ll |m|^{-2}$. Hence as in [12], we obtain

$$
\begin{aligned}
\Psi_n(\tau) &= a(q)\log\frac{|1-z^q|}{1-r^q}+\sum_{\substack{m=-\infty\\ m\ne 0}}^{\infty}a_m(q)\log\frac{|1-z^q|}{|1-r^qe^{im\tau}|} \\
&\le a(q)\log\frac{|1-z^q|}{1-r^q}+O(1)\,.
\end{aligned}
$$

Inserting this estimate into (14) and analyzing the logarithmic function, we obtain the first assertion of Lemma 2.

By virtue of $\log^{1/q}(2+|\tau|n)\ll\rho^{-1/q}$ when $|\tau|\le\tau_0$, the second estimate follows from the first one.

Lemma 2 is proved. ■

Proof: [Proof of Theorem 3] According to Cauchy's formula,

$$
M_n=\frac{1}{2\pi i n}\int_{|z|=r}\frac{F'(z)}{z^n}\,dz\,.
$$

We recall that $f(k)=1$ when $k>n$. Using the notations, we obtain

$$
\begin{aligned}
M_n &= \frac{\exp\{L(1)\}}{2\pi i n}\int_{|z|=r}\frac{\exp\{L(z)-L(1)\}}{(1-z)z^n}\sum_{k=1}^{\infty}f(k)z^{k-1}\,dz \\
&= \frac{\exp\{L(1)\}}{2\pi i n}\int_{|z|=r}\frac{\exp\{L(z)-L(1)\}}{(1-z)z^n}\left(\frac{1}{1-z}+L'(z)\right)dz\,. \qquad (15)
\end{aligned}
$$

When $z \in l_1$, we have $|1-z| \gg |\tau|$, and by Lemma 1, $\exp\{L(z)-L(1)\} \ll (n|\tau|)^\rho$. Thus,

$$J_1 := \frac{1}{n}\int_{z\in l_1} \frac{\exp\{L(z)-L(1)\}}{(1-z)^2 z^n}\,dz \ll (n\tau_0)^{\rho-1} \ll n^{(\delta-1)/2} + e^{-1/\rho}\,. \tag{16}$$

Similarly using Cauchy's inequality, Lemma 1 with s satisfying its condition and $t > 1$ such that $1/s + 1/t = 1$, we obtain

$$\begin{aligned} J_2 \;&:=\; \frac{1}{n}\int_{z\in l_1} \frac{\exp\{L(z)-L(1)\}}{(1-z)z^n} L'(z)\,dz \\ &\ll\; \frac{1}{n}\left(\int_{z\in l_1}\left|\frac{\exp\{L(z)-L(1)\}}{1-z}\right|^t |dz|\right)^{1/t} \|L'(z)\|_s \\ &\ll\; n^{-1+\rho}\tau_0^{\rho-1/s}\cdot \rho n^{1-1/s} \ll n^{-1/2s} + e^{-1/s\rho} \end{aligned} \tag{17}$$

provided that $\delta \le 1/2s$.

We now consider the integral in (15) when $z \in l_0$. Applying Lemma 1, we have

$$\begin{aligned} J_0 \;&:=\; \frac{1}{2\pi i n}\int_{l_0} \frac{1}{(1-z)^2 z^n} \sum_{j=0}^{N-1} \frac{\big(L(z)-L(1)\big)^j}{j!}\,dz \\ &+\; \frac{1}{2\pi i n}\int_{l_0} \frac{L'(z)}{(1-z)z^n} \sum_{j=0}^{N-2} \frac{\big(L(z)-L(1)\big)^j}{j!}\,dz + R \\ &=:\; J_{01} + J_{02} + R\,, \end{aligned} \tag{18}$$

where

$$\begin{aligned} R \;&\ll\; \frac{1}{nN!}\int_{l_0} \frac{|L(z)-L(1)|^N}{|1-z|^2}|dz| \\ &+\; \frac{1}{n(N-1)!}\int_{l_0} \frac{|L(z)-L(1)|^{N-1}}{|1-z|}|L'(z)|\,|dz| =: R' + R''\,. \end{aligned}$$

It follows from Lemma 2 that

$$R' \ll \frac{n\rho^N}{N!}\int_0^{\tau_0} \frac{\log^{N/q}(2+n\tau)}{(1+n\tau)^2}\,d\tau \ll \rho^N\;. \tag{19}$$

As estimating J_2, we obtain

$$R'' \ll \frac{\rho^{N-1}}{(N-1)!}\left(\int_0^{\tau_0} \frac{\log^{t(N-1)/q}(2+n\tau)}{(1+n\tau)^t}\,d\tau\right)^{1/t} \|L'(z)\|_s \ll \rho^N\,. \tag{20}$$

We extend the integrals J_{01} and J_{02} over the region l_1. While

$$\frac{1}{nj!}\int_{l_1} \frac{|L(z)-L(1)|^j}{|1-z|^2}\,|dz| \ll \rho^j (n\tau_0)^{-1/2} \ll \rho^j(n^{-1/4} + e^{-1/2\rho})$$

and

$$\begin{aligned}
&\frac{1}{nj!}\int_{l_1}\frac{|L(z)-L(1)|^j}{|1-z|}|L'(z)|\,|dz| \\
&\ll \frac{1}{nj!}\bigg(\int_{l_1}\frac{|L(z)-L(1)|^{jt}}{|1-z|^t}\,|dz|\bigg)^{1/t}\|L'(z)\|_s \\
&\ll \rho^{j+1}(n\tau_0)^{-(t-1)/2}\ll\rho^{j+1}(n^{-c_1}+e^{-c_1/\rho})
\end{aligned}$$

with $c_1=c_1(p)>0$, we obtain from (18), (19), and (20)

$$\begin{aligned}
J_0 &:= \frac{1}{2\pi in}\int_l\frac{1}{(1-z)^2z^n}\sum_{j=0}^{N-1}\frac{\big(L(z)-L(1)\big)^j}{j!}\,dz \\
&+ \frac{1}{2\pi in}\int_l\frac{L'(z)}{(1-z)z^n}\sum_{j=0}^{N-2}\frac{\big(L(z)-L(1)\big)^j}{j!}\,dz+O\big(n^{-c_2}+\rho^N\big) \\
&= \sum_{j=1}^{N-1}\frac{I_j}{j!}+O\big(n^{-c_2}+\rho^N\big)\,,
\end{aligned}$$

where $c_2=c_2(p)>0.$ Inserting the estimates (16), (17), and the last one into (15), we end the proof of Theorem 3. ∎

3. The second analytic formula

Now we will compensate the shortage of Theorem 3 appearing in the case when the quantity $\rho(p)$ is large. We will derive another asymptotic formula with the remainder estimate in terms of

$$\mu_n^2 := \frac{1}{n}\sum_{k=1}^{n}|f(k)-1|^2$$

and

$$E(u) := \exp\bigg\{2\sum_{\substack{k=1\\|f(k)-1|>u}}^{n}\frac{|f(k)-1|}{k}\bigg\}$$

with $u\geq 0$. All other previous notation remain the same.

THEOREM 4. *We have*

$$M_n=\exp\{L(1)\}\big(1+O\big((\mu_n+n^{-1})^{1/2}E(3/8)\big)\,.$$

The constant in the symbol O is absolute.

The proof goes along the similar lines as that of Theorem 3, though we need auxilliary results. Some of the ideas of the proof have been previously used in the papers [7] and [11].

LEMMA 3. *We have*

$$\|L'(z)\|_2 \le \sqrt{2\pi n}\,\mu_n\,.$$

Proof: Apply the Parseval equality. ∎

LEMMA 4. *We have*

$$\exp\{|L(z)-L(1)|\} \ll_u E(u)\left|\frac{1-z}{1-r}\right|^{4u/\pi}$$

for $z = re^{i\tau}$ *and each* $u \ge 0$.

Proof: Observe that

$$\sum_{k=1}^{n} \frac{|r^k e^{i\tau k} - 1|}{k} \ll 1 + \sum_{k=1}^{n} \frac{|e^{i\tau k} - 1|}{k}$$

and apply the Fourier expansions used in the proof of Lemma 2 with $q = 1$. So we deduce

$$|L(z)-L(1)| \ll \frac{4u}{\pi}\log\frac{|1-z|}{1-r} + \log E(u) + u$$

for $u \ge 0$. Hence follows the desired estimate.

Lemma 4 is proved. ∎

Proof of Theorem 4: We start with the formula (15). Now it is easier, than in the proof of Theorem 3, to estimate the integral

$$\begin{aligned} J_3 &:= \frac{1}{n}\int_{|z|=r} \frac{|\exp\{L(z)-L(1)\}|}{|1-z|}|L'(z)|\,|dz| \\ &\ll \frac{1}{n}\|L'(z)\|_2\left(\int_{|z|=r} \frac{|\exp\{L(z)-L(1)\}|^2}{|1-z|^2}\,|dz|\right)^{1/2}. \end{aligned}$$

We obtain from Lemma 3 and Lemma 4 with $u = 3/8$

$$J_3 \ll E(3/8)\mu_n\left(n^{-1+3/\pi}\int_{|z|=r} |1-z|^{3/\pi-2}\,|dz|\right)^{1/2} \ll E(3/8)\mu_n\,.$$

Let now $l_3 = \{z \in l : \ |\tau| \le K/n\}$, $K = \min\{\mu_n^{-1}, n\}$, and $l_4 = l \setminus l_3$. By Lemma 4 with $u = \pi/8$, we have

$$\begin{aligned} J_4 \ &:= \ \frac{1}{n}\int_{l_4} \frac{1 + |\exp\{L(z) - L(1)\}|}{|1-z|^2}\,|dz| \\ &\ll \ \frac{1}{K} + E(\pi/8)n^{-1/2}\int_{l_4} |1-z|^{-3/2}|dz| \ll E(3/8)K^{-1/2}\,. \end{aligned}$$

Similar by applying $|L(z) - L(1)| \le n\mu_n|1-z|$, we obtain

$$\begin{aligned} J_5 \ &:= \ \frac{1}{n}\int_{l_3} \frac{|L(z) - L(1)|\ \exp\{|L(z) - L(1)|\}}{|1-z|^2}\,|dz| \\ &\ll \ E(3/8)\mu_n \int_{l_3} |1-z|^{-1/2}|dz| \ll E(3/8)\mu_n\sqrt{K}\,. \end{aligned}$$

Inserting these obtained estimates into (15) and recalling the choice of K, we have

$$\begin{aligned} M_n \ &= \ \exp\{L(1)\}\big(1 + O(J_3 + J_4 + J_5)\big) \\ &\ \ \exp\{L(1)\}\big(1 + O\big((\mu_n + n^{-1})^{1/2}E(3/8)\big)\,. \end{aligned}$$

Theorem 4 is proved. ■

4. Estimation of the convergence rate

Proof of Theorem 1: We use a generalization of the Esseen inequality (see [17], Theorem 2, Chapter 5.2). Let

$$\varphi_n(t) := \frac{\exp\{-itA(n)\}}{n!}\sum_{\sigma\in\mathbb{S}_n} \exp\{ith_n(\sigma)\}, \quad t \in \mathbb{R}\,.$$

We have

$$R_n' \ll \frac{1}{T} + \int_{|t|\le T} \Big|\varphi_n(t) - e^{-t^2/2}(1 + \frac{t^2 D_n}{2})\Big|\frac{dt}{|t|} \tag{21}$$

where $T > 0$.

In order to obtain asymptotic formulas for $\varphi_n(t)$, we take $f(k) = \exp\{it\widehat{h}_n(k)\}$, $1 \le k \le n$ and apply Theorem 3 with $p = 3$, $N = 3$ and Theorem 4. Now $\rho \le |t|L_n^{1/3}$. Put $T_1 = \delta L_n^{-1/3}$, where $\delta > 0$ is sufficiently small to guarantee the validity of the formula in Theorem 3 in the region $|t| \le T_1$. Observe that the condition (3) implies

$$1 = \sum_{k=1}^{n} \frac{\widehat{h}_n^2(k)}{k} \le L_n^{2/3}\bigg(\sum_{k=1}^{n}\frac{1}{k}\bigg)^{1/3}$$

and hence $L_n \gg (\log n)^{-1/2}$. We can also suppose that $L_n = o(1)$. We obtain from Theorem 3

$$\begin{aligned}\varphi_n(t) &= \exp\{-\frac{t^2}{2}+\frac{\Theta}{6}|t|^3L_n\}\Big(1- \\ &- \frac{1}{2}\sum_{\substack{k,l=1\\k+l>n}}^{n}\frac{\big(\exp\{it\widehat{h}_n(k)\}-1\big)\big(\exp\{it\widehat{h}_n(l)\}-1\big)}{kl}+O(|t|^3L_n)\Big)\end{aligned} \quad (22)$$

in the region $L_n \le |t| \le T_1$ with $|\Theta| \le 1$.

Analysis of the double sum, say $S_n(t)$, in (22) requires more calculations. Let

$$a_k = \sum_{n-k<l\le n}\frac{1}{l}\,,\quad \Sigma(s) = \sum_{k=1}^{n}\frac{a_k^s}{k}\,.$$

Using the relation

$$\sum_{k\le n/2}\frac{a_k^s}{k} \to \int_0^{1/2}\frac{(-\log(1-x))^s}{x}\,dx < \infty, \quad s>0\,,$$

we obtain $\Sigma(s) \ll_s 1$ for each fixed $s>0$. Now as in [13], we have

$$\begin{aligned}S_n(t) &= it\sum_{k=1}^{n}\frac{\exp\{it\widehat{h}_n(k)\}-1}{k}\sum_{n-k<l\le n}\frac{\widehat{h}_n(l)}{l} \\ &+ O\Big(|t|^3L_n^{2/3}\sum_{k=1}^{n}\frac{|\widehat{h}_n(k)|}{k}a_k^{1/3}\Big) = \\ &= -t^2D_n + O\Big(|t|^3\sum_{k=1}^{n}\frac{\widehat{h}_n^2(k)}{k}\sum_{n-k<l\le n}\frac{|\widehat{h}_n(l)|}{l}\Big) \\ &+ O\big(|t|^3L_n\Sigma(1/2)^{2/3}\big) = -t^2D_n + O(|t|^3L_n)\,.\end{aligned}$$

Inserting the last formula into (22), we obtain

$$\varphi_n(t) = e^{-t^2/2}(1+t^2D_n/2) + O\big(|t|^3e^{-t^2/4}L_n\big) \quad (23)$$

in the region $L_n \le |t| \le T_1$.

Let $|t| \le (64L_n)^{-1} =: T$, then rough estimation of the terms in the formula obtained in Theorem 4 yields

$$\varphi_n(t) \ll \exp\Big\{-\frac{t^2}{2}+\frac{|t|^3L_n}{6}+2(\frac{8}{3})^2|t|^3L_n\Big\} \le e^{-t^2/4}\,. \quad (24)$$

It follows from (6) that

$$\varphi_n(t) - 1 \ll |t| \bigg(\frac{1}{n!} \sum_{\sigma \in \mathbb{S}_n} \big(h_n(\sigma) - A(n) \big)^2 \bigg)^{1/2} \ll |t| \,. \tag{25}$$

Splitting the region $|t| \leq T$ of integration in (21) into intervals $|t| \leq L_n$, $L_n \leq |t| \leq T_1$, $T_1 \leq |t| \leq T$ and using (25), (23), and (24) respectively, we obtain the desired estimate.

Theorem 1 is proved. ■

Proof of Corollary: The first estimate follows from Theorem 1 and the inequality

$$D_n \leq L_n^{1/3} \sum_{k=1}^{n} \frac{|\widehat{h}_n(k)|}{k} a_k^{2/3} \leq L_n^{2/3} \Sigma(2/3) \leq L_n^{2/3}$$

by the estimate above.

To prove the second assertion of Corollary, we construct the following example. Let $d(1) = 1$,

$$d(k) = \begin{cases} \log^{-2/5} k, & 2 \leq k \leq n/2, \\ 1, & n/2 < k \leq n, \end{cases}$$

then

$$\sum_{k=1}^{n} \frac{d^3(k)}{k} = c + o(1), \quad \beta_n^2 := \sum_{k=1}^{n} \frac{d^2(k)}{k} = 5 \log^{4/5} n + O(1) \,.$$

If $\widehat{h}(k) := d(k)/\beta_n$, then $L_n \sim 5^{-3/2} c \log^{-6/5} n$ and

$$D_n \geq \frac{1}{\beta_n^2} \bigg(\sum_{n/2 < k \leq n} \frac{1}{k} \bigg)^2 \geq \frac{(\log 2)^2}{2\beta_n^2} \geq c_1 L_n^{2/3}$$

with $c_1 > 0$ provided n is sufficiently large. Thus, for the sequence of CD additive functions defined by $h(\kappa) = \widehat{h}(l(\kappa))$, Theorem 1 yields $R_n \gg L_n^{2/3}$. ■

Proof of Theorem 2: As earlier, we may assume that $L_n = o(1)$. Then also $D_n = o(1)$. The estimates (24) and (25) but (23) remain valid.

Using the condition (7) instead of (3), we derive from Theorem 3

$$\varphi_n(t) = e^{-t^2/2} + O\big(|t|^3 e^{-t^2/4} L_n \big) \tag{26}$$

in the region $L_n \leq |t| \leq T_1$.

Now the traditional form of the Esseen inequality yields

$$R_n \ll \frac{1}{T} + \int_{|t| \leq T} \big| \varphi_n(t) - e^{-t^2/2} \big| \frac{dt}{|t|}$$

where $T = (64L_n)^{-1}$. Using the formulae (25), (24), and (26) in the regions $|t| \le L_n$, $T_1 \le |t| \le T$, and $L_n \le |t| \le T_1$, respectively, we obtain the desired estimate.

Theorem 2 is proved. ∎

Acknowledgments

The results of the paper were contributed at the Kyoto conference on analytic number theory in 1996. We thank the organizers and the Lithuanian Open Society Foundation for the financial support to attend it. We remain also deeply indebted to Professor Y.Motohashi for the warm hospitality shown during the meeting.

References

1. Aratia, R., Barbour, A.D., Tavaré, S., “Poisson process approximations for the Ewens sampling formula,” *Ann. Appl. Probab.* **2**(1992), 3, pp. 519–535.
2. Barbour, A.D., Tavaré, S., “A rate for the Erdös-Turán law,” *Combinatorics, Prob. Comput.* **3**(1994), pp. 167–176.
3. Erdös, P., Turán, P., “On some problems of a statistical grouptheory I,” *Zeitschr. für Wahrscheinlichkeitstheorie und verw. Gebiete* **4**(1965), pp. 175–186.
4. Flajolet, P., Odlyzko, “A., Singularity analysis of generating functions,” *SIAM J. Discrete Math.* **3**(1990), 2, pp. 216–240.
5. Goncharov, V.L., “On the distribution of cycles in permutations,” *Dokl. Acad. Nauk SSSR* **35**(1942), 9, pp. 299–301 (Russian).
6. Halász, G., “Über die Mittelwerte multiplikativer zahlentheoretisher Funktionen,” *Acta Math. Acad. Sci. Hung.* **19**(1968), pp. 365–403.
7. Halász, G., “On the distribution of additive and mean values of multiplicative arithmetic functions,” *Studia Sci. Math. Hung.* **6**(1971), pp. 211–233.
8. Hall, R.R., Tenenbaum, G., *Divisors*, Cambridge Tracts in Mathematics, 90, 1988.
9. Indlekofer, K.-H., Manstavičius, E., “Additive and multiplicative functions on arithmetical semigroups,” *Publicationes Mathematicae Debrecen* **45**(1994), 1–2, pp. 1–17.
10. Kolchin, V.F., *Random Mappings*, Optimization Software, Inc. New York, 1986.
11. Mačiulis, A., “Mean value of multiplicative functions,” *Lith. Math. J.* **28**(1988), 2, pp. 221–229.
12. Mačiulis, A., “The mean values of multiplicative functions defined on a semigroup,” In: *New Trends in Probab. and Statistics, vol. 2. Analytic and Probabilistic Methods in Number Theory*, F.Schweiger and E.Manstavičius (Eds), VSP/TEV, Utrecht/Vilnius, 1992, pp. 121–133.
13. Mačiulis, A., “The exact order of the convergence rate in the central limit theorem for additive functions,” *Lith. Math. J.* **33**(1993), 3, pp. 243–254.
14. Manstavičius, E., “Additive and multiplicative functions on random permutations,” *Lith. Math. J.* **36**(1996), 4, pp. 400–408.
15. Manstavičius, E., Skrabutėnas, R., “Summation of values of multiplicative functions on semigroups,” *Lith. Math. J.* **33**(1993), 3, pp. 255-264.
16. Nicolas, J.L., “Distribution statistique de l’ordre d’un element du groupe symetrique,” *Acta Math. Hung.* **45**(1985), 1-2, pp. 96–84.
17. Petrov, V.V., *Sums of Independent Random variables*, Moscow, ”Nauka”, 1972 (Russian).
18. Vershik, A.M., “Asymptotic combinatorics and algebraic analysis,” in: *Proceedings of the International Congress of Mathematicians, Zürich, 1994*, Birkhäuser, Basel, 1995, pp. 1384–1394.