A central limit theorem for random ordered factorizations of integers

HSIEN-KUEI HWANG*†
Institute of Statistical Science
Academia Sinica
Taipei 115, Taiwan
Email: hkhwang@stat.sinica.edu.tw

SVANTE JANSON†
Department of Mathematics
Uppsala University
PO Box 480, SE-751 06
Uppsala, Sweden
Email: svante.janson@math.uu.se

1 July 2010; revised 18 December 2010

Abstract

Write an integer as finite products of ordered factors belonging to a given subset $\mathscr{P}$ of integers larger than one. A very general central limit theorem is derived for the number of ordered factors in random factorizations for any subset $\mathscr{P}$ containing at least two elements. The method of proof is very simple and relies in part on Delange's Tauberian theorems and an interesting Tauberian technique for handling Dirichlet series associated with odd centered moments.

1 Introduction

Let $\mathscr{P}$ be a fixed subset of $\mathbb{N}_2 := \{2, 3, 4, \dots\}$ (the set of positive integers larger than 1). Let $a(n)$ denote the number of different ways of writing n as the product of *ordered* sequences $(n_1, \dots, n_k)$ of integers in $\mathscr{P}$. Define $a(1) = 1$. In particular, if $\mathscr{P} = \mathbb{N}_2$, then one has

$$\{a_n\}_{n\geq 2} = \{1, 1, 2, 1, 3, 1, 3, 2, 3, \cdots\}.$$

*The work of this author was partially supported by a grant from National Science Council, ROC.

†Part of this work was done while both authors were visiting Institut Mittag-Leffler, Djursholm, Sweden. They both thank the Institute for hospitality and support.

2010 *Mathematics Subject Classification*: Primary 11N80; Secondary 11N37; 11K65; 11N60; 11M45; 60F05.

Key words and phrases: Tauberian theorems, asymptotic normality, ordered factorizations, method of moments, Dirichlet series.

Let $A(N) := \sum_{1\le n\le N} a(n)$. Assume that all $A(N)$ factorizations of an integer $\le N$ are equally likely; denote by Y_N the random variable counting the number of factors in a random factorization. We prove in this paper that the distribution of Y_N is asymptotically normal under very general conditions on $\mathscr{P}$.

Denote by $\mathcal{P}(s)$ the Dirichlet series of $\mathscr{P}$

$$\mathcal{P}(s) := \sum_{n\in\mathscr{P}} n^{-s}.$$

Assume the abscissa of convergence of $\mathcal{P}(s)$ is κ. Then $\kappa = -\infty$ if $|\mathscr{P}| < \infty$ and $0 \le \kappa \le 1$ if $|\mathscr{P}| = \infty$. Note that $\mathcal{P}(\kappa) \le \infty$.

Our main result is as follows.

Main Theorem 1.1. *Assume $|\mathscr{P}| \ge 2$ and $1 < \mathcal{P}(\kappa) \le \infty$; thus there exists $\rho > \max\{\kappa, 0\}$ such that $\mathcal{P}(\rho) = 1$. Let*

$$\mu := -\frac{1}{\mathcal{P}'(\rho)}, \quad \text{and} \quad \sigma := \sqrt{\mu^3\mathcal{P}''(\rho) - \mu}.$$

Then

$$\frac{Y_N - \mu\log N}{\sigma\sqrt{\log N}} \xrightarrow{d} \mathscr{N}(0,1),$$

where $\xrightarrow{d}$ stands for convergence in distribution and $\mathscr{N}(0,1)$ denotes the standard normal distribution. The mean and variance of Y_N are asymptotic to $\mu\log N$ and $\sigma^2\log N$, respectively.

By Cauchy-Schwarz inequality, $\sigma^2 > 0$. We will indeed prove convergence of all moments.

The case when $|\mathscr{P}| = 1$, say $\mathscr{P} = \{d\}$, $d \ge 2$, is exceptional. In this case, $a(n) = 1$ when n is a power of d, and $a(n) = 0$ otherwise. Then Y_N is uniformly distributed on the integers $\{0, \dots, \lfloor\log_d N\rfloor\}$. Thus $Y_N/\log N$ converges in distribution to a uniform distribution on $[0, 1/\log d]$, and therefore, Y_N is not asymptotically normal; furthermore, the moment asymptotics differ from those in the theorem.

Ordered factorization problems in connection with that studied in this paper have a long history, tracing back to at least MacMahon's work (see [21]) in the early 1890's; later they were in many publications referred to as Kalmár's problem of "factorisatio numerorum" (see [16, 14]). Diverse properties of such factorizations have then been widely investigated, often in quite different contexts, one reason being that ordered factorizations are naturally encountered in many enumeration and estimation problems. For example, when $\mathscr{P} = \mathbb{N}_2$, $A(N) + 1$ equals the permanent of the Redheffer matrix; see [26]; also they appeared as the lower bound of certain biological problems; see [23]. See also [6, 18] for more information and references.

The first paper dealing with general ordered factorizations beyond the subset $\mathscr{P} = \mathbb{N}_2$ similar to our setting was Erdős [7], extending previous results by Hille [13]; see also [17]. Asymptotic normality (as well as local limit theorem and convergence rate) of the special case of Theorem 1.1 when $\mathscr{P} = \mathbb{N}_2$ was treated in [15]. In this case, $\rho \approx 1.7286$ is the unique root > 1 solving $\zeta(s) = 2$, where ζ denotes Riemann's zeta function. The proof given there relies on the determination of a zero-free region of the function $1 - z(\zeta(s) - 1)$, which in turn involves deep estimates from trigonometric sums (see also [20]). Such refined estimates are for general $\mathscr{P}$ hard to establish. We replace this estimate by applying purely Tauberian arguments of Delange (see [4, 22]), which require only analytic information of the involved Dirichlet series on its half-plane of convergence.

We will use Dirichlet series and the method of moments and derive asymptotic estimates for centered moments of integral orders, which, by Fréchet-Shohat's moment convergence theorem, will suffice to prove the theorem.

Proposition 1.2. *For $k \geq 0$*

$$\lim_{N\to\infty} \mathbb{E}\left(\frac{Y_N - \mu\log N}{\sigma\sqrt{\log N}}\right)^k = \begin{cases} \dfrac{k!}{(k/2)!2^{k/2}}, & \text{if } k \text{ is even} \\ 0, & \text{if } k \text{ is odd}, \end{cases} \tag{1.1}$$

A straightforward application of Tauberian theorems does not provide precise asymptotics for centered moments beyond the first due to cancellation of major dominant terms and due to the fact that no asymptotic expansion is generally available via application of Tauberian arguments; see however [19, 24] for more information on Tauberian theorems with remainders.

Note specially that while a direct use of Tauberian theorems leads to typical results of the form

$$\mathbb{E}(Y_N) \sim \mu \log N,$$

our result indeed gives, still relying on elementary Tauberian arguments, that (Proposition 1.2 with $k = 1$)

$$\mathbb{E}(Y_N) = \mu \log N + o(\sqrt{\log N}).$$

This shows the power of our approach, notably for odd moments. However, the estimates (1.1) we derive are not strong enough so as to prove more effective bounds such as the convergence rate to normality (or the Berry–Esseen bound).

The idea we use in this paper is to take into account the feature that the mean is logarithmic and to carry out the mean-shift on the associated Dirichlet series, which nicely incorporates the cancellations of higher centered moments in a surprisingly neat way. This version of the arguments with Dirichlet series thus avoids completely the messy calculations and cancellations that the usual method of moments faces when dealing with higher centered moments. While such a simple idea may seem standard for number-theorists, we were unable to find references in the number-theoretic literature for similar arithmetic problems. On the other hand, a similar idea was previously applied to characterize the phase change of random m-ary search trees, where a nonlinear differential equation with an additive nature was encountered; see [3]. However, the tools used there are complex-analytic, in contrast to the purely Tauberian ones used here for which the analytic properties of the Dirichlet series are reduced to a minimum. For other uses of the method of moments for arithmetical problems, see [5] and the references cited there.

In the special case when $\mathscr{P} = \mathbb{N}_2$, our result may be interpreted as saying that the property of the zero-free region for the Dirichlet series $1 - z(\zeta(s) - 1)$ lies much deeper than the asymptotic normality of the random variables Y_N.

In addition to the generality of our results, we believe that the approach we use here also offers important methodological value for the study of similar problems. In particular, not only is the use of the method of moments very simple, but no analytic properties of the Dirichlet series beyond the abscissa of convergence are needed, which largely simplifies the analysis in many situations. For example, the approach we use (including method of moments, generating functions and suitable Tauberian theorems) can be readily applied to other factorizations such as branching or cyclic factorizations with algebraic or logarithmic singularities (see [14]). It may also be possible to be extended, coupling with suitable Tauberian theorems, to deal with ordered factorizations in additive arithmetic semigroups (see [17]) and component counts in ordered combinatorial structures (see [9]). Another possible extension is to the analysis of Euclidean algorithms; see [1, 11, 25].

In the periodic case when $\mathscr{P} \subseteq \{d^k : k \geq 1\}$ for some $d \geq 2$, $\mathcal{P}(s)$ has period $2\pi i/\log d$ and the usual Tauberian theorem does not apply. Instead we write $\mathcal{P}(s) = \tilde{\mathcal{P}}(d^{-s})$, where $\tilde{\mathcal{P}}(z) := \sum_{d^k \in \mathscr{P}} z^k$, and transform the multiplicative nature of the problem into an additive one on random compositions by taking logarithms. We replace the Tauberian theorem by the singularity analysis of Flajolet and Odlyzko (see [8] or [10, Chapter VI]) in the proof; the details are similar to the proof below, but simpler, so we omit them. In the rest of the paper we thus assume, for every $d \geq 2$,

$$\mathscr{P} \not\subseteq \{d^k\}_{k \geq 1}. \tag{1.2}$$

2 Dirichlet series, Delange's Tauberian theorem, and proofs

Generating functions. Let $a_m(n)$ denote the number of ordered factorizations of n into exactly m factors. Then

$$\mathcal{P}(s)^m = \sum_{n \geq 1} a_m(n) n^{-s},$$

in the formal power series sense; analytically, we can take s satisfying $\mathcal{P}(\Re(s)) < \infty$. Thus if $e^{\Re(z)}\mathcal{P}(\Re(s)) < 1$, then by absolute convergence

$$\sum_{n\geq 1} n^{-s} \sum_{m\geq 0} a_m(n) e^{mz} = \sum_{m\geq 0} e^{mz}\mathcal{P}(s)^m = \frac{1}{1-e^z\mathcal{P}(s)}. \tag{2.1}$$

Delange's Tauberian theorem. We need the following form of Delange's Tauberian theorem (see [4] or [22, Ch. III, Sec. 3]).

Let $F(s) := \sum_{n\geq 1} \alpha(n) n^{-s}$ be a Dirichlet series with nonnegative coefficients and convergent for $\Re(s) > \varrho > 0$. Assume (i) $F(s)$ is analytic for all points on $\Re(s) = \varrho$ except at $s = \varrho$; (ii) for $s \sim \varrho$, $\Re(s) > \varrho$,

$$F(s) = \frac{G(s)}{(s-\varrho)^\beta} + H(s) \qquad (\beta > 0),$$

where G and H are analytic at $s = \varrho$ with $G(\varrho) \neq 0$. Then

$$\sum_{n\leq N} \alpha(n) \sim \frac{G(\varrho)}{\varrho\Gamma(\beta)} N^\varrho (\log N)^{\beta-1}. \tag{2.2}$$

Asymptotics of $A(N)$. We begin with the simplest case when $z = 0$ in (2.1) and obtain the Dirichlet series for $a(n) = \sum_{m\geq 0} a_m(n)$

$$\mathcal{A}(s) = \sum_{n\geq 1} a(n) n^{-s} = \frac{1}{1-\mathcal{P}(s)},$$

as long as $\Re(s) > \rho$. Note that the non-periodicity assumption (1.2) implies that $\mathcal{P}(s) \neq 1$ for all s with $\Re(s) = \rho$ but $s \neq \rho$. Hence $\mathcal{A}(s)$ is not only analytic in the open half-plane $\{s : \Re(s) > \rho\}$ but also on the boundary $\{s : \Re(s) = \rho\}$ except at $s = \rho$. The same holds true for all Dirichlet series we consider below.

Now for our $\mathcal{P}(s)$, since $\mathcal{P}'(\rho) = -\sum_{n\in\mathscr{P}} n^{-\rho}\log n < 0$, we see that $\mathcal{P}(s) - 1$ has a simple zero at $s = \rho$, and thus $\mathcal{A}(s)$ has a simple pole at ρ with

$$\mathcal{A}(s) = \frac{1}{1-\mathcal{P}(s)} \sim \frac{-1}{\mathcal{P}'(\rho)(s-\rho)},$$

as $s \to \rho$. Hence Delange's Tauberian theorem applies and we obtain

$$A(N) = \sum_{n\leq N} a(n) \sim RN^\rho, \qquad R := -\frac{1}{\rho\mathcal{P}'(\rho)} = \frac{\mu}{\rho}. \tag{2.3}$$

Furthermore, we also have

$$\sum_{n\le N} a(n)(\log n)^k \sim \frac{\mu}{\rho} N^\rho (\log N)^k \sim A(N)(\log N)^k,$$

either by repeating the same procedure for the Dirichlet series

$$(-1)^k \mathcal{A}^{(k)}(s) = \sum_{n\ge 1} a(n)(\log n)^k n^{-s} = (-1)^k \frac{\mathrm{d}^k}{\mathrm{d}s^k} \frac{1}{1-\mathcal{P}(s)}, \tag{2.4}$$

or by using directly (2.3). The estimate will be used later.

The expected value of Y_N. By taking the derivative with respect to z on both sides of (2.1), we obtain

$$\sum_{n\ge 1} n^{-s} \sum_{m\ge 0} m a_m(n) = \frac{\mathcal{P}(s)}{(1-\mathcal{P}(s))^2}.$$

Delange's Tauberian conditions being easily checked as above, we then obtain

$$\mathbb{E}(Y_N) = \frac{1}{A(N)} \sum_{n\le N}\sum_{m\ge 0} m a_m(n) \sim \mu \log N.$$

Shifting-the-mean at the Dirichlet series level. For higher centered moments, the crucial idea we will use can *formally* be described by using Perron's integral representation as follows (using (2.1) and for simplicity assuming temporarily that N is not an integer).

$$\begin{aligned}\mathbb{E}\left(e^{(Y_N-\mu\log N)z}\right) &= \frac{1}{2\pi i A(N)} \int_{c-i\infty}^{c+i\infty} \frac{N^{s-\mu z}}{s} \frac{1}{1-e^z\mathcal{P}(s)}\,\mathrm{d}s \\ &= \frac{1}{2\pi i A(N)} \int_{c-i\infty}^{c+i\infty} \frac{N^{s}}{s} \frac{1}{(1+\mu z/s)(1-e^z\mathcal{P}(s+\mu z))}\,\mathrm{d}s,\end{aligned}$$

where c is suitably chosen; the fact that the mean being of order $\log N$ is crucial here. We then *formally* expect that

$$\mathbb{E}\left(Y_N - \mu \log N\right)^k = \frac{k!}{2\pi i A(N)} \int_{c-i\infty}^{c+i\infty} \frac{N^s}{s} \mathcal{Q}_k(s)\,\mathrm{d}s, \tag{2.5}$$

where $\mathcal{Q}_k(s)$ is the coefficient of z^k in the Taylor expansion (in z) of

$$\frac{1}{(1+\mu z/s)(1-e^z\mathcal{P}(s+\mu z))}.$$

While all steps can be easily justified (as done below), we cannot directly apply Delange's Tauberian theorem to $\mathcal{Q}_k(s)$ here since each $\mathcal{Q}_k$ (except $\mathcal{Q}_0(s)$) is not a proper Dirichlet series, but involves additional powers of s^{-1}. This can be resolved as follows.

A probabilistic approximation. Given N, consider a random factorization of a number $n \le N$, namely, a random product $p_1 \cdots p_m \le n$ with all $p_j \in \mathscr{P}$ (uniformly distributed over all $A(N)$ possible factorizations). Let Y_N be the number of factors ($= m$) and ν_N be their product ($= n$). Then

$$A(N)\mathbb{E}(Y_N - \mu \log \nu_N)^k = \sum_{n \le N} b_k(n),$$

where

$$b_k(n) := \sum_{m \ge 0} a_m(n)(m - \mu \log n)^k.$$

We will show that for each k

$$\mathbb{E}(Y_N - \mu \log \nu_N)^k \sim \mathbb{E}(Y_N - \mu \log N)^k. \tag{2.6}$$

In other words, we may replace $\log \nu_N = \log n$ by $\log N$ (as often is the case in similar number-theoretic situations).

Shifting-the-mean at the coefficients level. To prove (2.6), we will rely on (2.5). By definition

$$\begin{aligned} A(N)\mathbb{E}\left(e^{(Y_N - \mu \log N)z}\right) &= \sum_{n \le N} \sum_{m \ge 0} a_m(n) e^{(m - \mu \log N)z} \\ &= \sum_{n \le N} \sum_{m \ge 0} a_m(n) e^{(m - \mu \log n)z - \mu z \log(N/n)}. \end{aligned}$$

Then, by taking the coefficients of z^k on both sides, we obtain

$$A(N)\mathbb{E}\left(Y_N - \mu \log N\right)^k = \sum_{0 \le \ell \le k} \binom{k}{\ell} (-\mu)^{k-\ell} \sum_{n \le N} b_\ell(n) \left(\log \frac{N}{n}\right)^{k-\ell}. \tag{2.7}$$

We will see that the growth order of $\sum_{n \le N} b_k(n)$ is the power N^ρ times an additional logarithmic term; it then follows that the weighted sum on the right-hand side is of the same order by a simple partial summation (see below for more details).

Now observe that (assuming again that N is not an integer)

$$\frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty} \frac{N^s}{s^m} \sum_{j \ge 1} \alpha(j) j^{-s} \, \mathrm{d}s = \frac{1}{(m-1)!} \sum_{n \le N} \alpha(n) \left(\log \frac{N}{n}\right)^{m-1},$$

for $m = 1, 2, \ldots$, where c is taken to be any real number greater than the abscissa of absolute convergence of the function defined by the series $\sum_{j \ge 1} \alpha(j) j^{-s}$. So, this, together with (2.7), justifies (2.5).

A probabilistic interpretation. Given N, consider a random factorization of a number $n \leq N$, namely, a random product $p_1 \cdots p_m \leq n$ with all $p_j \in \mathscr{P}$ (uniformly distributed over all $A(N)$ possible factorizations). Let Y_N be the number of factors ($= m$) and ν_N be their product ($= n$). Then

$$A(N)\mathbb{E}(Y_N - \mu \log \nu_N)^k = \sum_{n \leq N} b_k(n),$$

which gives a probabilistic interpretation of the partial sum.

The Dirichlet series of $b_k(n)$. Define the Dirichlet series

$$\mathcal{M}_k(s) := \sum_{n \geq 1} b_k(n) n^{-s} = \sum_{n \geq 1} n^{-s} \sum_{m \geq 0} a_m(n)(m - \mu \log n)^k. \tag{2.8}$$

Note that $a_m(n) > 0$ implies that $m \leq \log_2 n$, so that

$$(m - \mu \log n)^k = O\left((\log n)^k\right),$$

for all non-zero terms. Hence $\mathcal{M}_k(s)$ is absolutely convergent when $\Re(s) > \rho$ because $\mathcal{A}(s) = 1/(1 - \mathcal{P}(s))$ is. Now if $\Re(s) > \rho$ and $|z|$ is sufficiently small, then, by (2.1),

$$\sum_{k \geq 0} \frac{\mathcal{M}_k(s)}{k!} z^k = \sum_{n \geq 1} n^{-s} \sum_{m} a_n(m) e^{(m - \mu \log n)z} = \frac{1}{1 - e^z \mathcal{P}(s + \mu z)}. \tag{2.9}$$

With these $\mathcal{M}_k(s)$, the generating function $\mathcal{Q}_k(s)$ can be decomposed as

$$\mathcal{Q}_k(s) = \sum_{0 \leq \ell \leq k} \left(\frac{-\mu}{s}\right)^{k-\ell} \frac{\mathcal{M}_\ell(s)}{\ell!}. \tag{2.10}$$

Our strategy will then to apply Delange's Tauberian theorem to $\mathcal{M}_k$ for even k and some auxiliary Dirichlet series for odd k, and then the asymptotics of the k-th centered moment can be obtained easily since terms with index $\ell < k$ in (2.10) will be asymptotically negligible. Indeed, we will use (2.7).

Recurrence of $\mathcal{M}_k(s)$. We now focus on properties of $\mathcal{M}_k(s)$. By writing (2.9) in the form

$$(1 - e^z \mathcal{P}(s + \mu z)) \sum_{\ell \geq 0} \frac{\mathcal{M}_\ell(s)}{\ell!} z^\ell = 1,$$

we see that $\mathcal{M}_k(s)$ satisfies the recurrence

$$\mathcal{M}_k(s) = \frac{1}{1 - \mathcal{P}(s)} \sum_{0 \leq j < k} \binom{k}{j} \mathcal{M}_j(s) \mathcal{B}_{k-j}(s) \qquad (k \geq 1), \tag{2.11}$$

with $\mathcal{M}_0(s) = 1/(1-\mathcal{P}(s))$, where

$$\mathcal{B}_k(s) := \sum_{0\le \ell\le k} \binom{k}{\ell} \mu^\ell \mathcal{P}^{(\ell)}(s).$$

For example, $\mathcal{M}_1(s) = \mathcal{B}_1(s)/(1-\mathcal{P}(s))^2 = (\mathcal{P}(s)+\mu\mathcal{P}'(s))/(1-\mathcal{P}(s))^2$.

Note that each $\mathcal{B}_k(s)$ is analytic for $\Re(s) > \kappa$ and, in particular, for $\Re(s) \ge \rho$. Moreover, the crucial property here is

$$\mathcal{B}_1(\rho) = \mathcal{P}(\rho) + \mu\mathcal{P}'(\rho) = 1 - 1 = 0,$$

by our construction. Similarly,

$$\mathcal{B}_2(\rho) = \mathcal{P}(\rho) + 2\mu\mathcal{P}'(\rho) + \mu^2\mathcal{P}''(\rho) = \mu^2\mathcal{P}''(\rho) - 1 = \sigma^2/\mu,$$

and $\mathcal{B}_1'(\rho) = \sigma^2/\mu^2$.

On the other hand, by (2.11), we see that $\mathcal{M}_k(s)$ is analytic for $\Re(s) > \rho$ and for $\Re(s) = \rho$ except at $s = \rho$. Furthermore, because $\mathcal{B}_1(\rho) = 0$, it follows by induction from (2.11), that at $s = \rho$, $\mathcal{M}_k(s)$ has a pole of order at most $\lfloor k/2 \rfloor + 1$.

Even moments. More precisely, for even $k = 2\ell$, we get by induction

$$\mathcal{M}_k(s) \sim c_k (s-\rho)^{-k/2-1},$$

where

$$c_k = \binom{k}{2} \mu \mathcal{B}_2(\rho) c_{k-2} = \frac{k(k-1)}{2} \sigma^2 c_{k-2},$$

with $c_0 = \mu$, which is solved to be

$$c_k = \mu \left(\frac{\sigma^2}{2}\right)^{k/2} k!.$$

We now apply Delange's Tauberian theorem and obtain

$$\begin{aligned}
\mathbb{E}(Y_N - \mu \log \nu_N)^k &= \frac{1}{A(N)} \sum_{n\le N} b_k(n) \\
&\sim \frac{c_k}{\rho\Gamma(k/2+1)A(N)} N^\rho (\log N)^{k/2} \\
&\sim \frac{k!}{2^{k/2}(k/2)!} \sigma^k (\log N)^{k/2}. \qquad (2.12)
\end{aligned}$$

Odd moments. Let now $k = 2\ell - 1$, $\ell \geq 1$. Since the coefficients $b_k(n)$ are not necessarily nonnegative, we cannot directly apply Delange's Tauberian theorem. Instead, we consider the following two auxiliary Dirichlet series

$$D_1(s) := \sum_{n\geq 1} n^{-s} \sum_{m\geq 0} a_m(n) \left((m - \mu \log n)^k + (\log n)^{k/2}\right)^2,$$

and, see (2.8) and (2.4),

$$\begin{aligned} D_2(s) &:= \sum_{n\geq 1} n^{-s} \sum_{m\geq 0} a_m(n) \left((m - \mu \log n)^{2k} + (\log n)^k\right) \\ &= \mathcal{M}_{2k}(s) + (-1)^k \mathcal{A}^{(k)}(s). \end{aligned}$$

The two Dirichlet series have only nonnegative coefficients, and we will show that Delange's Tauberian theorem can be applied to both series. The leading terms will cancel and we will have

$$\frac{1}{A(N)} \sum_{n\leq N} b_k(n)(\log n)^{k/2} = o\left((\log N)^k\right). \tag{2.13}$$

From this, we use the monotonicity of $(\log n)^{k/2}$ and elementary arguments to recover the desired estimate

$$\mathbb{E}(Y_N - \mu \log \nu_N)^k = \frac{1}{A(N)} \sum_{n\leq N} b_k(n) = o\left((\log N)^{k/2}\right). \tag{2.14}$$

Proof of (2.13). Let

$$D_3(s) := \sum_{n\geq 1} (\log n)^{k/2} b_k(n) n^{-s}.$$

Then $D_1(s) = D_2(s) + 2D_3(s)$. By the discussions above, we can apply Delange's theorem to $D_2(s)$ and obtain

$$\frac{1}{A(N)} \sum_{n\leq N} \sum_{m\geq 0} a_m(n) \left((m - \mu \log n)^{2k} + (\log n)^k\right) \sim C_k (\log N)^k, \tag{2.15}$$

where $C_k = (2k)!\sigma^{2k}/(2^k k!) + 1$ (this value is however immaterial).

We now show that the partial sum of the coefficients of $D_1(s)$ has asymptotically the same dominant term. We start from the representation ($k = 2\ell - 1$)

$$D_3(s) = (-1)^\ell \pi^{-1/2} \int_0^\infty \mathcal{M}_k^{(\ell)}(s + t) t^{-1/2}\, \mathrm{d}t,$$

for $\Re s > \rho$, because $(-1)^\ell \mathcal{M}_k^{(\ell)}(s) = \sum_{n\ge 1} b_k(n)(\log n)^\ell n^{-s}$ and

$$\begin{aligned}
(-1)^\ell \int_0^\infty \mathcal{M}_k^{(\ell)}(s+t)t^{-1/2}\,\mathrm{d}t &= \sum_{n\ge 1} b_k(n)(\log n)^\ell n^{-s} \int_0^\infty e^{-t\log n} t^{-1/2}\,\mathrm{d}t \\
&= \Gamma(\tfrac{1}{2}) \sum_{n\ge 1} b_k(n)(\log n)^{k/2} n^{-s} \\
&= \sqrt{\pi} D_3(s).
\end{aligned}$$

We now consider the local behavior of $D_3(s)$ when $s \sim \rho$. First, $\mathcal{M}_k(s)$ has a pole at $s = \rho$ with leading term $c_k'(s-\rho)^{-(k+1)/2}$, for some c_k'. Thus $\mathcal{M}_k^{(\ell)}(s)$ has a pole with local behavior $c_k''(s-\rho)^{-k-1}$. It follows that for small $|w|$ and $\Re(w) > 0$,

$$\begin{aligned}
D_3(\rho + w) &= (-1)^\ell \pi^{-1/2} \int_0^\infty \mathcal{M}_k^{(\ell)}(\rho + w + t)t^{-1/2}\,\mathrm{d}t \\
&= O\left(\int_0^\infty |w+t|^{-k-1} t^{-1/2}\,\mathrm{d}t\right) \\
&= O\left(\int_0^{|w|} |w|^{-k-1} t^{-1/2}\,\mathrm{d}t + \int_{|w|}^\infty t^{-k-3/2}\,\mathrm{d}t\right) \\
&= O\left(|w|^{-k-1/2}\right) \\
&= o\left(|w|^{-k-1}\right).
\end{aligned}$$

Since $D_1(s) = D_2(s) + D_3(s)$ has all coefficients nonnegative, we can now apply Delange's theorem to $D_1(s)$ and conclude that

$$\frac{1}{A(N)} \sum_{n\le N} \sum_{m\ge 0} a_m(n) \left((m - \mu\log n)^k + (\log n)^{k/2}\right)^2 \sim C_k(\log N)^k.$$

This, together with (2.15), proves (2.13).

Proof of (2.14). Let

$$B_k(x) := \sum_{n\le x} b_k(n)(\log n)^{k/2},$$

and for $x \ge 2$, $\varphi(x) := (\log x)^{-k/2}$. Then

$$\begin{aligned}
\int_2^N B_k(x)\varphi'(x)\,\mathrm{d}x &= \sum_{2\le n\le N} b_k(n)(\log n)^{k/2} \int_n^N \varphi'(x)\,\mathrm{d}x \\
&= \sum_{2\le n\le N} b_k(n)(\log n)^{k/2} \left((\log N)^{-k/2} - (\log n)^{-k/2}\right) \\
&= B_k(N)(\log N)^{-k/2} - \sum_{2\le n\le N} b_k(n).
\end{aligned}$$

Thus, by (2.13),

$$\begin{aligned}
\sum_{1\le n\le N} b_k(n) &= b_k(1) + B_k(N)(\log N)^{-k/2} - \int_2^N B_k(x)\varphi'(x)\,\mathrm{d}x \\
&= O(1) + o\left(N^{\rho}(\log N)^{k/2}\right) \\
&\qquad + \frac{k}{2}\int_2^N B_k(x)x^{-1}(\log x)^{-k/2-1}\,\mathrm{d}x \\
&= O(1) + o\left(N^{\rho}(\log N)^{k/2}\right) + o\left(\int_2^N x^{\rho-1}(\log x)^{k/2-1}\,\mathrm{d}x\right) \\
&= o\left(N^{\rho}(\log N)^{k/2}\right),
\end{aligned}$$

as required.

From $Y_N - \mu\log\nu_N$ to $Y_N - \mu\log N$. The two estimates (2.12) and (2.14) imply

$$\mathbb{E}\left(\frac{Y_N - \mu\log\nu_N}{\sigma\sqrt{\log N}}\right)^k \to \begin{cases} \dfrac{k!}{(k/2)!2^{k/2}}, & \text{if } k \text{ is even} \\ 0, & \text{if } k \text{ is odd,} \end{cases} \tag{2.16}$$

which in turn implies, by the method of moments, that

$$\frac{Y_N - \mu\log\nu_N}{\sigma\sqrt{\log N}} \xrightarrow{d} \mathscr{N}(0,1).$$

Our final task is to prove the same asymptotics (1.1) from the two estimates (2.12) and (2.14). To that purpose, define $S_k(x) = 0$ if $x < 2$ and

$$S_k(x) := \sum_{n\le x} b_k(n) \qquad (x \ge 2).$$

We use (2.7) and the cruder estimates (by (2.12), (2.14) and (2.3))

$$S_\ell(x) = O\left(A(x)(\log x)^{\ell/2}\right) = O\left(x^{\rho}(\log x)^{\ell/2}\right),$$

for $\ell = 0, \ldots, k-1$. With this, we have

$$\begin{aligned}
&\sum_{0\le\ell<k}\binom{k}{\ell}(-\mu)^{k-\ell}\sum_{n\le N} b_\ell(n)\left(\log\frac{N}{n}\right)^{k-\ell} \\
&\qquad = O\left((\log N)^k + \sum_{0\le\ell<k}\binom{k}{\ell}\mu^{k-\ell}\int_2^N\left(\log\frac{N}{x}\right)^{k-\ell}\mathrm{d}S_\ell(x)\right).
\end{aligned}$$

Now for each $\ell = 0, \dots, k-1$,

$$\int_2^N \left(\log \frac{N}{x}\right)^{k-\ell} \mathrm{d}S_\ell(x) = O\left(\int_2^N \left(\log \frac{N}{x}\right)^{k-\ell-1} x^{-1} S_\ell(x)\,\mathrm{d}x\right)$$
$$= O\left(\int_2^N \left(\log \frac{N}{x}\right)^{k-\ell-1} x^{\rho-1}(\log x)^{\ell/2}\,\mathrm{d}x\right).$$

Splitting the integral at $x = N/2$, and making the change of variables $x \mapsto N/x$ for the first half, we see that

$$\begin{aligned}
&\int_2^N \left(\log \frac{N}{x}\right)^{k-\ell-1} x^{\rho-1}(\log x)^{\ell/2}\,\mathrm{d}x \\
&= O\left(N^\rho \int_2^{N/2} x^{-\rho-1}(\log x)^{k-1-\ell}\left(\log \frac{N}{x}\right)^{\ell/2}\,\mathrm{d}x \right.\\
&\qquad\qquad \left. + \int_{N/2}^N x^{\rho-1}(\log x)^{\ell/2}\,\mathrm{d}x\right) \\
&= O\left(N^\rho (\log N)^{\ell/2}\right) \\
&= o\left(N^\rho (\log N)^{k/2}\right),
\end{aligned}$$

for $0 \le \ell \le k-1$. This proves that

$$\mathbb{E}\left(Y_N - \mu \log N\right)^k = \frac{1}{A(N)} \sum_{n \le N} b_k(n) + o\left((\log N)^{k/2}\right),$$

and thus the estimates in (1.1) hold by (2.12) and (2.14).

An alternative argument. To bridge (2.16) and (1.1), we can also argue as follows. Consider the sum

$$\mathbb{E}\left(\log N - \log \nu_N\right)^k = \frac{1}{A(N)} \sum_{n \le N} a(n) \left(\log \frac{N}{n}\right)^k,$$

which is $O(1)$ by a similar summation by parts argument as used above. Then for even k

$$|| \log N - \log \nu_N ||_k = O(1).$$

By Hölder's inequality, this holds true also for every $k \ge 0$. Consequently, using again Hölder's inequality, we deduce (1.1).

3 Conclusions and additional remarks

A corollary to our moment convergence result is the following asymptotic approximations to all absolute centered moments

$$\mathbb{E}|Y_N - \mu \log N|^\beta \sim 2^{\beta/2}\pi^{-1/2}\Gamma\left(\frac{\beta+1}{2}\right)(\log N)^{\beta/2},$$

for all $\beta \geq 0$, which seem difficult to get directly from Dirichlet series.

On the other hand, when $z \in (-\log \mathcal{P}(\kappa), \infty)$, one can apply directly Delange's Tauberian theorem to the generating function

$$\frac{1}{1-e^z P(s)},$$

(instead of to the Dirichlet series of higher moments obtained above by Taylor expansions in z); this results in the asymptotic approximation

$$\mathbb{E}\left(e^{zY_N}\right) \sim \frac{\rho P'(\rho)}{\rho(z)e^z P'(\rho(z))} N^{\rho(z)-\rho},$$

where $\rho(z)$ solves the equation $1 = e^z P(\rho(z))$ with $\rho(0) = \rho$. From this approximation, one might expect asymptotic normality by straightforward argument. However, the asymptotic result so obtained holds only pointwise, and the uniformity in z is missing here. While the gap of uniformity may be filled by applying suitable Tauberian theorems with remainders, the use of Delange's Tauberian theorems is conceptually simpler and technically less involved, with the required analysis reduced to almost minimum.

It is clear from our proof that Theorem 1.1 actually holds for any Dirichlet series $\mathcal{P}(s)$ with nonnegative coefficients and satisfying the conditions of Theorem 1.1. Thus the restriction of $\mathscr{P}$ to a subset of positive integers is not essential. For example, one can consider the *ordered totient factorizations* with $\mathcal{P}(s) = \sum_{n\geq 3} \phi(n)^{-s}$, where $\phi(n)$ is Euler's totient function, namely, the number of positive integers $\leq n$ and relatively prime to n. In this case, $\kappa = 1$ and $\rho \approx 2.26386$ since

$$\sum_{n\geq 1} \phi(n)^{-s} = \zeta(s) \prod_{p\,:\,\text{prime}} \left(1 - p^{-s} + (p-1)^{-s}\right);$$

see [2] for a detailed study of this Dirichlet series.

How can we compute ρ to high degree of precision? In general, the problem is not easy, for example, if $\mathcal{P}(s) = \sum_{n\geq 2} \lceil n^\beta (\log n)^c \rceil^{-s}$ for $\beta, c > 0$; the case of totient factorization is similar. The easy cases are when $\mathcal{P}(s)$ can be expressed in terms of ζ-functions such as $\mathcal{P}(s) = \zeta(s) - 1$ (all integers > 1) or $\mathcal{P}(s) = \zeta(s)/\zeta(2s) - 1$ (square-free integers > 1). Take now $\mathcal{P}(s) = \sum_{p\,:\,\text{prime}} p^{-s}$. The zero of $\mathcal{P}(s) = 1$ can also be easily computed by using the relation (see [12])

$$\sum_{p\,:\,\text{prime}} p^{-s} = \sum_{k\geq 1} \frac{\mu(k)}{k} \log \zeta(ks) \qquad (\Re(s) > 1),$$

where $\mu(k)$ is Möbius function. This readily gives

$$\rho \approx 1.39943\,33287\,26330\,31820\,28072\ldots,$$

and

$$\mu \approx 0.57764\,86251\,95138\,05440\,61351\ldots,$$
$$\sigma^2 \approx 0.48439\,65045\,13598\,28128\,07456\ldots.$$

We indicated a few directions to which our approach may be extended in the Introduction. But can a similar idea be modified so as to deal with arithmetic functions whose means have an order other than $\log N$?

Acknowledgements

We thank the referees for helpful comments.

References

[1] V. Baladi and B. Vallée, Euclidean algorithms are Gaussian, *J. Number Theory* **110** (2005), 331–386.

[2] D. W. Boyd and H. L. Montgomery, Cyclotomic partitions, in *Number Theory* (Banff, AB, 1988), 7–25, de Gruyter, Berlin, 1990.

[3] H.-H. Chern and H.-K. Hwang, Phase changes in random m-ary search trees and generalized quicksort. *Random Structures Algorithms* **19** (2001), 316–358.

[4] H. Delange, Généralisation du théorème de Ikehara, *Ann. Sci. Éc. Norm. Supér.* **71** (1954), 213–242.

[5] H. Delange, Sur les moments des fonctions additives, *Acta Arith.* **49** (1988), 213–236.

[6] M. Deléglise, M. O. Hernane and J.-L. Nicolas, Grandes valeurs et nombres champions de la fonction arithmétique de Kalmár, *J. Number Theory* **128** (2008), 1676–1716.

[7] P. Erdős, On some asymptotic formulas in the theory of the "factorisatio numerorum," *Ann. of Math* **42** (1941) 989–993; Corrections, *ibid.* **44** (1943) 647–651.

[8] P. Flajolet and A. Odlyzko, Singularity analysis of generating functions, *SIAM J. Discrete Math.* **3** (1990), 216–240.

[9] P. Flajolet and M. Soria, General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Math.* **114** (1993), 159–180.

[10] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge Univ. Press, Cambridge, UK, 2009.

[11] P. Flajolet and B. Vallée, Continued fraction algorithms, functional operators, and structure constants, *Theoret. Comput. Sci.* **194** (1998), 1–34.

[12] P. Flajolet and I. Vardi, Zeta function expansions of classical constants, unpublished manuscript; available at `http://algo.inria.fr/flajolet/Publications/landau.ps`

[13] E. Hille, A problem in "factorisatio numerorum," *Acta Arith.* **2** (1936), 134–144.

[14] H.-K. Hwang, *Théorèmes limites pour les structures combinatories et les fonctions arithmétiques*, Ph. D. Thesis, École Polytechnique, 1994.

[15] H.-K. Hwang, Distribution of the number of factors in random ordered factorizations of integers, *J. Number Theory* **81** (2000), 61–92.

[16] L. Kalmár, A "factorisatio numerorum" problémájáról, *Mat. Fiz. Lapok* **38** (1931) 1–15.

[17] A. Knopfmacher, J. Knopfmacher and R. Warlimont, Ordered factorizations for integers and arithmetical semigroups, in *Advances in Number Theory* (Kingston, ON, 1991), 151–165, Oxford Univ. Press, New York, 1993.

[18] A. Knopfmacher and M. E. Mays, A survey of factorization counting functions, *Int. J. Number Theory* **1** (2005), 563–581.

[19] J. Korevaar, *Tauberian Theory. A Century of Developments*, Springer-Verlag, Berlin, 2004.

[20] Y.-K. Lau, Local distribution of ordered factorizations of integers, *J. Number Theory* **91** (2001), 312–317.

[21] P. A. MacMahon, The theory of perfect partitions and the compositions of multipartite numbers, *Mess. Math.* **20** (1891), 103–119.

[22] W. Narkiewicz, *Number Theory*, Translated from the Polish by S. Kanemitsu. World Scientific Publishing Co., Singapore; distributed by Heyden & Son, Inc., Philadelphia, PA, 1983.

[23] L. A. Newberg and D. Naor, A lower bound on the number of solutions to the probed partial digest problem. *Adv. in Appl. Math.* **14** (1993), 172–183.

[24] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, Second edition, Société Mathématique de France, Paris, 1995; English translation published by Cambridge University Press, 1995.

[25] B. Vallée, Dynamical analysis of a class of Euclidean algorithms, *Theoret. Comput. Sci.* **297** (2003), 447–486.

[26] H. S. Wilf, The Redheffer matrix of a partially ordered set, *Electron. J. Combin.* **11** (2004/06), Research Paper 10, 5 pp.