Math. Proc. Camb. Phil. Soc. (2022), **173**, 189–200
doi:10.1017/S0305004121000499
First published online 23 June 2021

Joint Poisson distribution of prime factors in sets

BY KEVIN FORD†
Department of Mathematics, 1409 *West Green Street,*
University of Illinois at Urbana-Champaign, Urbana, IL 61801*, U.S.A.*
e-mail: ford@math.uiuc.edu

(*Received* 13 *July* 2019; *revised* 20 *May* 2021; *accepted* 17 *May* 2021)

Abstract

Given disjoint subsets $T_1, \ldots, T_m$ of "not too large" primes up to x, we establish that for a random integer n drawn from $[1, x]$, the m-dimensional vector enumerating the number of prime factors of n from $T_1, \ldots, T_m$ converges to a vector of m independent Poisson random variables. We give a specific rate of convergence using the Kubilius model of prime factors. We also show a universal upper bound of Poisson type when $T_1, \ldots, T_m$ are unrestricted, and apply this to the distribution of the number of prime factors from a set T conditional on n having k total prime factors.

2020 Mathematics Subject Classification: 11N25

1. *Introduction*

A central theme in probabilistic number theory concerns the distribution of additive arithmetic functions, in particular the functions $\omega(n)$ and $\Omega(n)$, which count the number of distinct prime factors of n and the number of prime power factors of n, respectively. Taking a uniformly random integer $n \in [1, x]$ with x large, the functions $\omega(n)$ and $\Omega(n)$ behave like Poisson random variables with parameter $\log\log x$. This was established by Sathe [**16**] and Selberg [**17**] in 1954, while hints of this were already present in the inequalities of Landau [**13**], Hardy and Ramanujan [**10**], Erdős [**6**], and Erdős and Kac [**7**]. We refer the reader to Elliott's notes [**5**, pp. 23–26] for an extensive discussion of the history of these results.

In this paper we address the distribution of the number of prime factors of n lying in an arbitrary set T. Denote by $\mathbb{P}_x$ the probabiliy with respect to a uniformly random integer n drawn from $[1, x]$. Each such n has a unique prime factorisation

$$n = \prod_{p \leqslant x} p^{v_p},$$

where the exponents v_p are now random variables. For any finite set T of primes, let

$$\omega(n, T) = \#\{p|n : p \in T\} = \#\{p \in T : v_p > 0\}, \qquad \Omega(n, T) = \sum_{p \in T} v_p.$$

†Supported by NSF grant DMS-1802139.

For a prime p, the event $\{p|n\}$ occurs with probability close to $1/p$, and thus heuristically

$$\mathbb{P}_x(\omega(n, T)=k) \approx \sum_{\substack{p_1,\dots,p_k\in T\\ p_1<\cdots<p_k}} \frac{1}{p_1\cdots p_k} \prod_{\substack{p\in T\\ p\notin\{p_1,\dots,p_k\}}} \left(1-\frac{1}{p}\right) \approx \mathrm{e}^{-H(T)}\frac{H(T)^k}{k!}, \tag{1.1}$$

where

$$H(T)=\sum_{p\in T}\frac{1}{p}.$$

That is, we expect that $\omega(n, T)$ will be close to Poisson with parameter $H(T)$. A more complicated combinatorial heuristic also suggests that $\Omega(n, T)$ is close to Poisson with parameter $H(T)$. This was made rigorous by Halász [8] in 1971, who showed*

$$\mathbb{P}_x(\Omega(n, T)=k)=\frac{H(T)^k}{k!}\mathrm{e}^{-H(T)}\left(1+O_\delta\left(\frac{|k-H(T)|}{H(T)}\right)+O_\delta\left(\frac{1}{\sqrt{H(T)}}\right)\right), \tag{1.2}$$

uniformly in the range $\delta H(T)\leqslant k\leqslant(2-\delta)H(T)$, where $\delta>0$ is fixed. Small modifications to the proof yield an identical estimate for $\mathbb{P}_x(\omega(n, T)=k)$; see [5, p. 301] for a sketch of the argument. Inequality (1.2) implies the order of magnitude estimate

$$\frac{H(T)^k}{k!}\mathrm{e}^{-H(T)} \ll \mathbb{P}_x(\Omega(n, T)=k) \ll \frac{H(T)^k}{k!}\mathrm{e}^{-H(T)}$$

when $(1-\varepsilon)H(T)\leqslant k\leqslant(2-\delta)H(T)$ for sufficiently small $\varepsilon>0$. The range of k in this last bound was extended to $\delta H(T)\leqslant k\leqslant(2-\delta)H(T)$ by Sárkőzy [15] in 1977.

Inequality (1.2) implies that $\Omega(n, T)$ converges to the Poisson distribution with parameter $H(T)$ if T is a function of x such that $H(T)\to\infty$ as $x\to\infty$. This is a natural condition, as the following examples show. If T consists only of small primes, say those less than a bounded quantity t, then $\omega(n, T)$ takes only finitely many values and thus the distribution cannot converge to Poisson as $x\to\infty$. Although $\Omega(n, T)$ is unbounded, the distribution is very far from Poisson, e.g. $\mathbb{P}_x(\Omega(n, \{2\})=k)\sim 1/2^{k+1}$ for each k. Likewise, if $c>1$ is fixed and T is the set of primes in $(x^{1/c}, x]$, $\omega(n, T)$ and $\Omega(n, T)$ are each bounded by c. Moreover, the distribution of the largest prime factors of an integer is governed by the very different Poisson–Dirichlet distribution; see [19] for details. In each of these examples, $H(T)$ is bounded. The condition $H(T)\to\infty$ ensures that neither small primes nor large primes dominate T with respect to the harmonic measure.

An asymptotic for the joint local limit laws $\mathbb{P}(\omega(n; T_1)=k_1, \omega(n; T_2)=k_2)$ was proved by Delange [4, section 6·5·3] in 1971, in the special case when T_1 and T_2 are infinite sets with $H(T_j\cap[1, x])=\lambda_j\log\log x+O(1)$ and λ_1, λ_2 constants. Halász' result (1.2) was extended by Tenenbaum [21] in 2017 to the joint distribution of $\omega(n; T_j)$ uniformly over any disjoint sets $T_1, \dots, T_m$ of the primes $\leqslant x$. If $P=\mathbb{P}_x(\omega(n, T_i)=k_i, 1\leqslant i\leqslant m)$, then

*As usual, the notation $f=O(g)$, $f\ll g$ and $g\gg f$ means that there is a constant C so that $|f|\leqslant g$ throughout the domain of f. The constant C is indepenedent of any variable or parameter unless that dependence is specified by a subscript, e.g. $f=O_A(g)$ means that C depends on A.

$$P = \left(1 + O\left(\sum_{j=1}^{m} \frac{1}{\sqrt{H(T_j)}}\right)\right)\left(\prod_{j=1}^{m} \frac{H(T_j)^{k_j}}{k_j!} \mathrm{e}^{-k_j}\right)\frac{1}{x}\sum_{n \leqslant x}\prod_{j=1}^{m}(k_j/H(T_j))^{\omega(n;T_j)}$$
$$= \prod_{j=1}^{m} \frac{H(T_j)^{k_j}}{k_j!} \mathrm{e}^{-H(T_j)} \exp\left(O\left(\sum_{j=1}^{m} \frac{|k_j - H(T_j)|}{H(T_j)} + \frac{1}{\sqrt{H(T_j)}}\right)\right), \tag{1.3}$$

uniformly in the range $c_1 \leqslant k_j/H(T_j) \leqslant c_2$ $(1 \leqslant j \leqslant m)$, for any fixed c_1, c_2 satisfying $0 < c_1 < c_2$; see [**21**], equation (2·23) and the following paragraph. The methods in [**21**] establish the same bound for $\mathbb{P}_x(\Omega(n, T_i) = m_i, 1 \leqslant i \leqslant k)$, but with the restriction $c_1 \leqslant k_j/H(T_j) \leqslant 2 - c_1$, $1 \leqslant j \leqslant m$, again with fixed $c_1 > 0$. An asymptotic for the sum on n in (1.3) is not known in general. A slight extension of Tenenbaum's asymptotic (1.3) was given by Mangerel [**14**, theorem 1·5·3], who showed a corresponding asymptotic in the case where some of the quantities k_j are smaller (specifically, $H(T_j)^{2/3+\varepsilon} < k_j \leqslant H(T_j)$).

In the literature on the subject, $\omega(n, T)$ and $\Omega(n, T)$ have always been compared to a Poisson variable with parameter $H(T)$. As we shall see, the functions $\Omega(n, T)$ are better approximated by a Poisson variable with parameter

$$H'(T) = \sum_{p \in T} \frac{1}{p-1},$$

at least when T does not contain any large primes. In order to state our results, we introduce a further harmonic sum

$$H''(T) = \sum_{p \in T} \frac{1}{p^2}.$$

We note for future reference that

$$H(T) \leqslant H'(T) \leqslant H(T) + 2H''(T).$$

We also use the notion of the total variation distance $d_{TV}(X, Y)$ between two random variables living on the same discrete space Ω:

$$d_{TV}(X, Y) := \sup_{A \subset \Omega} \left|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\right|.$$

We denote by $\mathrm{Pois}(\lambda)$ a Poisson random variable with parameter λ, and write $Z \overset{d}{=} \mathrm{Pois}(\lambda)$ for the statement that Z is a Poisson random variable with parameter λ.

THEOREM 1. *Let $2 \leqslant y \leqslant x$ and suppose that $T_1, \ldots, T_m$ are disjoint nonempty sets of primes in $[2, y]$. For each $1 \leqslant i \leqslant m$, suppose that either $f_i = \omega(n, T_i)$ and $Z_i \overset{d}{=} \mathrm{Pois}(H(T_i))$ or that $f_i = \Omega(n, T_i)$ and $Z_i \overset{d}{=} \mathrm{Pois}(H'(T_i))$. Assume that $Z_1, \ldots, Z_m$ are independent. Then*

$$d_{TV}\Big((f_1, \ldots, f_m), (Z_1, \ldots, Z_m)\Big) \ll \sum_{j=1}^{m} \frac{H''(T_j)}{1 + H(T_j)} + u^{-u}, \qquad u = \frac{\log x}{\log y}.$$

The implied constant is absolute, independent of m, y, x and $T_1, \ldots, T_m$. In particular, if m is fixed then this shows that the joint distribution of $(f_1, \ldots, f_m)$ converges to a

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

joint Poisson distribution whenever we have $y = x^{o(1)}$ and for each i, either $H(T_i) \to \infty$ or $\min T_i \to \infty$.

By contrast, Tenenbaum's bound (1.3) implies

$$d_{TV}\Big((\omega(n, T_1), \ldots, \omega(n, T_m)), (Z_1, \ldots, Z_m)\Big) \ll_m \sum_{j=1}^{m} \frac{1}{\sqrt{H(T_j)}}. \tag{1.4}$$

Compared to Theorem 1, we see that (1.4) gives good results even if the sets T_i contain many large primes, while Theorem 1 requires that $y \leqslant x^{o(1)}$ in order to be nontrivial. However, if $y \leqslant x^{1/\log\log\log x}$, say, the conclusion of Theorem 1 is stronger, especially when $H''(T)$ is small. An extreme case is given by singleton set $T = \{p\}$ and $f_1 = \Omega(n, T)$, where Theorem 1 recovers the correct order of $d_{TV}(f_1, Z_1)$, namely $1/p^2$, since $\mathbb{P}_x(p\|n) \approx 1/p - 1/p^2$, $\mathbb{P}_x(p^2\|n) \approx 1/p^2 - 1/p^3$, and $\mathbb{P}(Z_1 = 2) \approx 1/(2p^2)$ for large p.

Example. Let S be the set of all primes, $t_k = \exp\exp k$ and $\omega_k(n) := \omega(n, S \cap (t_k, t_{k+1}])$. Here, by the Prime Number Theorem with strong error term,

$$H(S \cap (t_k, t_{k+1}]) = 1 + O(\exp\{-\mathrm{e}^{k/2}\}).$$

Thus, ω_k has distribution close to that of a Poisson variable with parameter 1. More precisely, if X, Y are Poisson with parameters λ, λ', respectively, then (e.g. [**2**, theorem 1·C, remark 1·1·2])

$$d_{TV}(X, Y) \leqslant |\lambda - \lambda'|.$$

Using a standard inequality for d_{TV} ((3.5) below), we deduce the following.

COROLLARY 2. *If $\xi \leqslant k < \ell \leqslant \log\log x - \xi$, then*

$$d_{TV}\big((\omega_k, \ldots, \omega_\ell), (Z'_k, \ldots, Z'_\ell)\big) \ll \exp\{-\mathrm{e}^{\xi/2}\}, \tag{1.5}$$

where $Z'_k, \ldots, Z'_\ell$ are independent Poisson variables with parameter 1.

Thus, statistics of the random *function* $f(t) = \omega(n, S \cap [t_k, t])$, $t_k \leqslant t \leqslant t_\ell$, are captured very accurately by statistics of the partial sums $Z'_k + \cdots + Z'_m$ for $k \leqslant m \leqslant \ell$. The latter has been well-studied and one can easily deduce, for example, the Law of the Iterated Logarithm for $f(t)$ from that for the partial sums $Z'_k + \cdots + Z'_\ell$. Similarly, if T is a set of primes with density $\alpha > 0$ in the sense that

$$\sum_{p \leqslant x,\, p \in T} \frac{1}{p} = \alpha \log\log x + c + o(1) \quad (x \to \infty)$$

then a statement similar to (1.5) holds with t_k replaced by $t'_k = \exp\exp(k/\alpha)$, with a weaker estimate for the total variation distance (depending on the decay of the $o(1)$ term).

We now establish the upper-bound implied in (1.3), but valid uniformly for all $k_1, \ldots, k_m$.

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

THEOREM 3. *Let $T_1, \ldots, T_r$ be arbitrary disjoint, nonempty subsets of the primes $\leqslant x$. For any $k_1, \ldots, k_r \geqslant 0$, letting $P = \mathbb{P}_x\big(\omega(n; T_j) = k_j \ (1 \leqslant j \leqslant r)\big)$, we have*

$$P \ll \prod_{j=1}^{r}\left(\frac{H'(T_j)^{k_j}}{k_j!}\mathrm{e}^{-H(T_j)}\right)\left(\eta + \frac{k_1}{H'(T_1)} + \cdots + \frac{k_r}{H'(T_r)}\right) + \xi$$
$$\leqslant \prod_{j=1}^{r}\left(\frac{(H(T_j)+2)^{k_j}}{k_j!}\mathrm{e}^{-H(T_j)}\right),$$

where $\eta = 0$ if $T_1 \cup \cdots \cup T_r$ contains every prime $\leqslant x$ and $\eta = 1$ otherwise, and $\xi = 1$ if $\eta = k_1 = \cdots = k_r = 0$ and $\xi = 0$ otherwise.

Remarks. Tudesq [**22**] claimed a bound similar to Theorem 3, but only supplied details for $r = 1$. Our method is similar, and we give a short, complete proof in Section 4.

If we condition on $\omega(n) = k$, the $r = 2$ case of Theorem 3 supplies tail bounds for $\omega(n, T)$. If X, Y are independent Poisson random variables with parameters λ_1, λ_2, respectively, then for $0 \leqslant \ell \leqslant k$, we have

$$\mathbb{P}(X = \ell | X + Y = k) = \binom{k}{l}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{\ell}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{k-\ell}.$$

Thus, conditional on $\omega(n) = k$ we expect that $\omega(n, T)$ will have roughly a binomial distribution with parameter $\alpha = H(T)/H(S)$, where S is the set of all primes in $[2, x]$.

THEOREM 4. *Fix $A > 1$ and suppose that $1 \leqslant k \leqslant A \log\log x$. Let T be a nonempty subset of the primes in $[2, x]$ and define let $\alpha = H(T)/H(S)$. For any $0 \leqslant \psi \leqslant \sqrt{\alpha k}$ we have*

$$\mathbb{P}\Big(|\omega(n, T) - \alpha k| \geqslant \psi\sqrt{\alpha(1-\alpha)k} \;\Big|\; \omega(n) = k\Big) \ll_A \mathrm{e}^{-\frac{1}{3}\psi^2},$$

the implied constant depending only on A.

Similarly, if $T_1, \ldots, T_m$ are disjoint subsets of primes $\leqslant x$ and we condition on $\omega(n) = k$, then the vector $(\omega(n, T_1), \ldots, \omega(n, T_m))$ will have approximately a multinomial distribution.

2. *The Kubilius model of small prime factors of integers*

Our restriction to primes below $x^{o(1)}$ comes from an application of a probabilistic model of prime factors, called the Kubilius model, and introduced by Kubilius [**11**, **12**] in 1956. We compute

$$\mathbb{P}_x(v_p = k) = \frac{1}{\lfloor x \rfloor}\left(\left\lfloor \frac{x}{p^k}\right\rfloor - \left\lfloor \frac{x}{p^{k+1}}\right\rfloor\right) = \frac{1}{p^k} - \frac{1}{p^{k+1}} + O\left(\frac{1}{x}\right),$$

the error term being relatively small when p^k is small. Moreover, the variables v_p are quasi-independent; that is, the correlations are small, again provided that the primes are small. By contrast, the variables v_p corresponding to large p are very much dependent, for example the event $(v_p > 0, v_q > 0)$ is impossible if $pq > x$.

The model of Kubilius is a sequence of *idealised* random variables which removes the error term above, and is much easier to compute with. For each prime p, define the random

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

variable X_p that has domain $\mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\}$ and such that

$$\mathbb{P}(X_p = k) = \frac{1}{p^k} - \frac{1}{p^{k+1}} = \frac{1}{p^k}\left(1 - \frac{1}{p}\right) \qquad (k = 0, 1, 2, \ldots).$$

The principal result, first proved by Kubilius and later sharpened by others, is that the random vector

$$\mathbf{X}_y = (X_p : p \leqslant y)$$

has distribution close to that of the random vector

$$\mathbf{V}_{x,y} = (v_p : p \leqslant y),$$

provided that $y = x^{o(1)}$.

In [18], Tenenbaum gives a rather complicated asymptotic for $d_{TV}(\mathbf{X}_y, \mathbf{V}_{x,y})$ in the range $\exp\{(\log x)^{2/5+\varepsilon}\} \leqslant y \leqslant x$, as well as a simpler universal upper bound which we state here.

LEMMA 2·1 (Tenenbaum [18, théorème 1·1 and (1·7)]). *Let $2 \leqslant y \leqslant x$. Then, for every $\varepsilon > 0$,*

$$d_{TV}(\mathbf{X}_y, \mathbf{V}_{x,y}) \ll_\varepsilon u^{-u} + x^{-1+\varepsilon}, \quad u = \frac{\log x}{\log y}.$$

3. *Poisson approximation of prime factors*

For a finite set T of primes, denote

$$U_T = \#\{p \in T : X_p \geqslant 1\}, \qquad W_T = \sum_{p \in T} X_p,$$

which are probabilistic models for $\omega(n, T)$ and $\Omega(n, T)$, respectively. For any T which is a subset of the primes $\leqslant y = x^{1/u}$, Lemma 2·1 implies that for any $\varepsilon > 0$,

$$\begin{aligned} d_{TV}(U_T, \omega(n, T)) &\ll_\varepsilon u^{-u} + x^{-1+\varepsilon}, \\ d_{TV}(W_T, \Omega(n, T)) &\ll_\varepsilon u^{-u} + x^{-1+\varepsilon}. \end{aligned} \tag{3.1}$$

We next prove a local limit theorem for U_T and W_T, and then use this to establish Theorem 1.

THEOREM 5. *Let T be a finite subset of the primes, and let $Y = U_T$ or $Y = W_T$. Let $H = H(T)$ if $Y = U_T$ and $H = H'(T)$ if $Y = W_T$. Also let $Z \stackrel{d}{=} \mathrm{Pois}(H)$. Then*

$$\mathbb{P}(Y = k) - \mathbb{P}(Z = k) \ll \begin{cases} H''(T)\dfrac{H^k}{k!}\mathrm{e}^{-H}\left(\dfrac{1}{k+1} + \dfrac{k-H}{H}^2\right) & \text{if } 0 \leqslant k \leqslant 1.9H \\ H''(T)\dfrac{\mathrm{e}^{0.9H}}{(1.9)^k} & \text{if } k > 1.9H. \end{cases}$$

Proof. Write $H'' = H''(T)$. When $k = 0$, $\mathbb{P}(Z = 0) = \mathrm{e}^{-H}$ and

$$\mathbb{P}(Y = 0) = \mathbb{P}(\forall p \in T : X_p = 0) = \prod_{p \in T}\left(1 - \frac{1}{p}\right) = \mathrm{e}^{-H}(1 + O(H'')),$$

and the desired inequality follows.

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

For $k \geqslant 1$, we work with moment generating functions as in the proof of Halász' theorem (1.2); see also [5, chapter 21]. For any complex z,

$$\mathbb{E}\, z^Z = \mathrm{e}^{(z-1)H}.$$

Uniformly for complex z with $|z| \leqslant 2$ we have

$$\mathbb{E}\, z^{U_T} = \prod_{p \in T} \left(1 + \frac{z-1}{p}\right) = \mathrm{e}^{(z-1)H(T)}\Big(1 + O\big(|z-1|^2 H''(T)\big)\Big) \tag{3.2}$$

and uniformly for $|z| \leqslant 1.9$ we have

$$\mathbb{E}\, z^{W_T} = \prod_{p \in T} \left(1 + \frac{z-1}{p-z}\right) = \mathrm{e}^{(z-1)H'(T)}\big(1 + O(|z-1|^2 H''(T))\big). \tag{3.3}$$

Write $e(\theta) = \mathrm{e}^{2\pi i\theta}$. Then, for any $0 < r \leqslant 1.9$, (3.2) and (3.3) imply

$$\begin{aligned}
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) &= \frac{1}{2\pi i} \oint_{|z|=r} \frac{\mathbb{E}\, z^Y - \mathbb{E}\, z^Z}{z^{k+1}}\, dw \\
&= \frac{1}{r^k} \int_0^1 e(-k\theta)\Big[\mathbb{E}\,(re(\theta))^Y - \mathbb{E}\,(re(\theta))^Z\Big]\, d\theta \\
&= \frac{1}{r^k} \int_0^1 e(-k\theta) \mathrm{e}^{(re(\theta)-1)H} \cdot O\left(|re(\theta)-1|^2 H''\right)\, d\theta \\
&\ll \frac{H''}{r^k} \int_0^{1/2} |re(\theta)-1|^2 \mathrm{e}^{(r\cos(2\pi\theta)-1)H}\, d\theta.
\end{aligned}$$

Now, for $0 \leqslant \theta \leqslant 1/2$,

$$r\cos(2\pi\theta) - 1 = r - 1 - 2r\sin^2(\pi\theta) \leqslant r - 1 - 8r\theta^2$$

and

$$|re(\theta) - 1|^2 = (r - 1 - 2r\sin^2(\pi\theta))^2 + \sin^2(2\pi\theta) \ll (r-1)^2 + \theta^2,$$

so we obtain

$$\begin{aligned}
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) &\ll H'' \frac{\mathrm{e}^{(r-1)H}}{r^k} \int_0^{1/2} (|r-1|^2 + \theta^2) \mathrm{e}^{-8r\theta^2 H}\, d\theta \\
&\ll H'' \frac{\mathrm{e}^{(r-1)H}}{r^k} \left(\frac{|r-1|^2}{\sqrt{1+rH}} + \frac{1}{(1+rH)^{3/2}}\right).
\end{aligned} \tag{3.4}$$

When $1 \leqslant k \leqslant 1.9H$, we take $r = k/H$ in (3.4) and obtain, using Stirling's formula,

$$\begin{aligned}
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) &\ll H'' \frac{H^k \mathrm{e}^{k-H}}{k^k} \left(\frac{|k/H - 1|^2}{k^{1/2}} + \frac{1}{k^{3/2}}\right) \\
&\ll H'' \frac{\mathrm{e}^{-H} H^k}{k!} \left(\left|\frac{k-H}{H}\right|^2 + \frac{1}{k}\right).
\end{aligned}$$

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

When $k > 1.9H$, take $r = 1.9$ in (3.4) and conclude that

$$\mathbb{P}(Y = k) - \mathbb{P}(Z = k) \ll \frac{H'' e^{0.9H}}{(1.9)^k \sqrt{1+H}}.$$

This completes the proof.

COROLLARY 6. *Let T be a finite subset of the primes. Then*

$$d_{TV}(U_T, \mathrm{Pois}(H(T))) \ll \frac{H''(T)}{1 + H(T)}$$

and

$$d_{TV}(W_T, \mathrm{Pois}(H'(T))) \ll \frac{H''(T)}{1 + H(T)},$$

Proof. Let $Y \in \{U_T, W_T\}$. If $Y = U_T$, let $H = H(T)$ and if $Y = W_T$, let $H = H'(T)$. Let $Z \overset{d}{=} \mathrm{Pois}(H)$. Again, write $H'' = H''(T)$. We begin with the identity

$$d_{TV}(Y, Z) = \frac{1}{2} \sum_{k=0}^{\infty} \big| \mathbb{P}(Y_T = k) - \mathbb{P}(Z(T) = k) \big|.$$

Consider two cases. First, if $H \leqslant 2$, we have by Theorem 5,

$$\sum_{k \geqslant 0} |\mathbb{P}(Y = k) - \mathbb{P}(Z = k)| \ll H'' + \sum_{k > 1.9H} H''(1.9)^{-k} \ll H''.$$

If $H > 2$, Theorem 5 likewise implies that

$$\sum_{k > 1.9H} |\mathbb{P}(Y = k) - \mathbb{P}(Z = k)| \ll H'' \sum_{k > 1.9H} \frac{e^{0.9H}}{(1.9)^k} \ll H'' e^{-0.3H}$$

and also

$$\sum_{k \leqslant 1.9H} |\mathbb{P}(Y = k) - \mathbb{P}(Z = k)| \ll H'' e^{-H} \sum_{k \leqslant 1.9H} \frac{H^k}{k!} \left[\frac{1}{k+1} + \left| \frac{k - H_1}{H} \right|^2 \right]$$
$$\ll \frac{H''}{H} \ll \frac{H''}{H(T)},$$

using that $e^{-H} H^k / k!$ decays rapidly for $|k - H| > \sqrt{H}$.

We now combine Theorem 5 with the standard inequality

$$d_{TV}((X_1, \ldots, X_m), (Y_1, \ldots, Y_m)) \leqslant \sum_{j=1}^{m} d_{TV}(X_j, Y_j), \tag{3.5}$$

valid if $X_1, \ldots, X_m$ are independent, and $Y_1, \ldots, Y_m$ are independent, with all variables living on the same set Ω.

COROLLARY 7. *Let $T_1, \ldots, T_m$ be disjoint sets of primes. For each i, either let $Y_i = U_{T_i}$ and $H_i = H(T_i)$ or let $Y_i = W_{T_i}$ and $H_i = H'(T_i)$. For each i, let $Z_i \overset{d}{=} \mathrm{Pois}(H_i)$, and suppose*

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

that $Z_1, \ldots, Z_m$ *are independent. Then*

$$d_{TV}\big((Y_1, \ldots, Y_m), (Z_1, \ldots, Z_m)\big) \ll \sum_{j=1}^{m} \frac{H''(T_j)}{1+H(T_j)}.$$

Combining Corollary 7 with (3.1) and the triangle inequality, we see that

$$d_{TV}\Big((f_1, \ldots, f_m), (Z_1, \ldots, Z_m)\Big) \ll \sum_{j=1}^{m} \frac{H''(T_j)}{1+H(T_j)} + u^{-u} + x^{-0.99}.$$

We may remove the term $x^{-0.99}$, because if $y \leqslant x^{1/3}$ then $H''(T_i) \gg x^{-2/3}$ and $H(T_i) \ll \log\log x$, while if $y > x^{1/3}$ then $u^{-u} \gg 1$. This completes the proof of Theorem 1.

4. *A uniform upper bound*

In this section we prove Theorem 3 and Theorem 4.

Proof of Theorem 3. Let

$$N = \#\{n \leqslant x : \omega(n; T_j) = k_j \ (1 \leqslant j \leqslant r)\}.$$

If $\eta = 0$ (that is, $T_1 \cup \cdots \cup T_r$ contains all the primes $\leqslant x$) and $k_1 = \cdots = k_r = 0$, then $N = 1$; this explains the need for the additive term ξ in Theorem 3.

Now assume that either $\eta = 1$ or that $k_i \geqslant 1$ for some i. Let

$$L_t(x) = \sum_{\substack{h \leqslant x \\ \omega(h; T_j) = k_j - \mathbb{1}_{j=t} \ (1 \leqslant j \leqslant r)}} \frac{1}{h} \qquad (0 \leqslant t \leqslant r),$$

where $\mathbb{1}_A$ is the indicator function of the condition A. We use the "Wirsing trick", starting with $\log x \ll \log n = \sum_{p^a \| n} \log p^a$ for $x^{1/3} \leqslant n \leqslant x$ and thus

$$(\log x) N \ll \sum_{\substack{n \leqslant x^{1/3} \\ \omega(n; T_j) = k_j \ (1 \leqslant j \leqslant r)}} \log x + \sum_{\substack{n \leqslant x \\ \omega(n; T_j) = k_j \ (1 \leqslant j \leqslant r)}} \sum_{p^a \| n} \log p^a.$$

In the first sum, $\log x \leqslant x^{1/3} \log x/n \ll x^{1/2}/n$, hence the sum is at most $\leqslant x^{1/2} L_0(x)$. In the double sum, let $n = p^a h$ and observe that $\omega(h, T_j) = k_j - 1$ if $p \in T_j$ and $\omega(h, T_j) = k_j$ otherwise. In particular, if $p \notin T_1 \cup \cdots \cup T_r$ then $\omega(h, T_j) = k_j$ for all j, and this is only possible if $\eta = 1$. Hence

$$(\log x) N \ll x^{1/2} L_0(x) + \sum_{t=1-\eta}^{r} \sum_{\substack{h \leqslant x \\ \omega(h; T_j) = k_j - \mathbb{1}_{j=t} \ (1 \leqslant j \leqslant r)}} \sum_{p^a \leqslant x/h} \log p^a.$$

Using Chebyshev's estimate for primes, the innermost sum over p^a is $O(x/h)$ and thus the double sum over h, p^a is $O(L_t(x))$. Also, if $k_j = 0$ then there is the sum corresponding to $t = j$ is empty. This gives

$$\mathbb{P}_x\Big(\omega(n; T_j) = k_j \ (1 \leqslant j \leqslant r)\Big) \ll \frac{1}{\log x}\Big((\eta + x^{-1/2}) L_0(x) + \sum_{1 \leqslant t \leqslant r : k_t > 0} L_t(x)\Big). \qquad (4.1)$$

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

Now we fix t and bound the sum $L_t(x)$; if $t \geqslant 1$ we may assume that $k_t \geqslant 1$. Write the denominator $h = h_1 \cdots h_r h'$, where, for $1 \leqslant j \leqslant r$, h_j is composed only of primes from T_j,

$$\omega(h_j; T_j) = m_j := k_j - \mathbb{1}_{t=j},$$

and h' is composed of primes below x which lie in none of the sets $T_1, \ldots, T_r$. For $1 \leqslant j \leqslant r$ we have

$$\sum_{h_j} \frac{1}{h_j} \leqslant \frac{1}{m_j!} \left(\sum_{p \in T_j} \frac{1}{p} + \frac{1}{p^2} + \cdots \right)^{m_j} = \frac{H'(T_j)^{m_j}}{m_j!},$$

and, using Mertens' estimate,

$$\sum_{h'} \frac{1}{h'} \leqslant \prod_{\substack{p \leqslant x \\ p \notin T_1 \cup \cdots \cup T_r}} \left(1 - \frac{1}{p}\right)^{-1} \ll (\log x) \prod_{p \in T_1 \cup \cdots \cup T_r} \left(1 - \frac{1}{p}\right).$$

Thus,

$$L_t(x) \ll (\log x) \prod_{j=1}^{r} \frac{H'(T_j)^{m_j}}{m_j!} \prod_{p \in T_1 \cup \cdots \cup T_r} \left(1 - \frac{1}{p}\right).$$

Using the elementary inequality $1 + y \leqslant \mathrm{e}^y$, we see that the final product over p is at most $\mathrm{e}^{-H(T_1) - \cdots - H(T_r)}$, and we find that

$$L_t(x) \ll (\log x) \prod_{j=1}^{r} \left(\frac{H'(T_j)^{m_j}}{m_j!} \mathrm{e}^{-H(T_j)} \right) \tag{4.2}$$

Combining estimates (4.1) and (4.2), we conclude that

$$\mathbb{P}_x\Big(\omega(n; T_j) = k_j \ (1 \leqslant j \leqslant r)\Big) \ll \left(\eta + x^{-1/2} + \sum_{j=1}^{r} \frac{k_j}{H'(T_j)} \right) \prod_{j=1}^{r} \left(\frac{H'(T_j)^{k_j}}{k_j!} \mathrm{e}^{-H(T_j)} \right).$$

Either $\eta = 1$ or $k_j / H'(T_j) \gg 1/\log\log x$ for some j, and hence the additive term $x^{-1/2}$ may be omitted. This proves the first claim.

Next,

$$\prod_{j=1}^{r} \frac{H'(T_j)^{k_j}}{k_j!} \left(1 + \sum_{j=1}^{r} \frac{k_j}{H'(T_j)} \right) \leqslant \prod_{j=1}^{r} \frac{(H'(T_j) + 1)^{k_j}}{k_j!}$$

and we have $H'(T) \leqslant H(T) + \sum_p 1/p(p-1) \leqslant H(T) + 1$. This proves the final inequality.

To prove Theorem 4 we need standard tail bounds for the binomial distribution. For proofs, see [**1**, lemma 4·7·2] or [**3**, theorem 6·1].

LEMMA 4·1 (Binomial tails). *Let X have binomial distribution according to k trials and parameter $\alpha \in [0, 1]$; that is, $\mathbb{P}(X = m) = \binom{k}{m} \alpha^m (1-\alpha)^{k-m}$. If $\beta \leqslant \alpha$ then we have*

$$\mathbb{P}(X \leqslant \beta k) \leqslant \exp\left\{ -k \left(\beta \log \frac{\beta}{\alpha} + (1-\beta) \log \frac{1-\beta}{1-\alpha} \right) \right\} \leqslant \exp\left\{ -\frac{(\alpha-\beta)^2 k}{3\alpha(1-\alpha)} \right\}.$$

https://doi.org/10.1017/S0305004121000499 Published online by Cambridge University Press

Replacing α *with* $1-\alpha$ *we also have for* $\beta \geqslant \alpha$,

$$\mathbb{P}(X \geqslant \beta k) \leqslant \exp\left\{-\frac{(\alpha-\beta)^2 k}{3\alpha(1-\alpha)}\right\}.$$

Proof of Theorem 4. We may assume that $\alpha k \geqslant C$, where C is a sufficiently large constant, depending on A. Without loss of generality, we may assume that $H(T) \leqslant H(S)/2$ (that is , $\alpha \leqslant 1/2$), else replace T by $S \setminus T$. Apply Theorem 3 with two sets: $T_1 = T$ and $T_2 = S \setminus T$, so that $\eta = \xi = 0$. We need the lower bound

$$\mathbb{P}_x(\omega(n) = k) \gg_A \frac{(\log\log x)^{k-1}}{(k-1)!\log x} = \frac{k}{\log\log x} \cdot \frac{(\log\log x)^k}{k!\log x}$$

see, e.g. [20, Theorem 6·4 in Chapter II·6]. Also,

$$\left(\frac{k-h}{H'(S\setminus T)} + \frac{h}{H'(T)}\right)\frac{\log\log x}{k} \ll 1 + \frac{h}{\alpha k}.$$

Since $H'(S\setminus T) \leqslant H(S\setminus T)+1$, we have

$$H'(S\setminus T)^{k-h} \ll H(S\setminus T)^{k-h}.$$

In addition,

$$H'(T)^h \leqslant (H(T)+1)^h \leqslant H(T)^h \mathrm{e}^{h/H(T)} \leqslant H(T)^h \mathrm{e}^{O_A(h/(\alpha k))}.$$

Then, for $0 \leqslant h \leqslant k$, Theorem 3 implies

$$\mathbb{P}\Big(\omega(n,T) = h \Big| \omega(n) = k\Big) \ll_A \alpha^h (1-\alpha)^{k-h}\binom{k}{h}\mathrm{e}^{O_A(h/(\alpha k))}.$$

Ignoring the factor $(1-\alpha)^{k-h}$, we see that the terms with $h \geqslant 100\alpha k$ contribute at most

$$\sum_{h\geqslant 100\alpha k}\frac{(\alpha k \mathrm{e}^{O_A(1/(\alpha k))})^h}{h!} \leqslant \sum_{h\geqslant 100\alpha k}\frac{(2\alpha k)^h}{h!} \leqslant \mathrm{e}^{-100\alpha k} \leqslant \mathrm{e}^{-100\psi^2}$$

for large enough C. When $h < 100\alpha k$ we have

$$\mathbb{P}\Big(\omega(n,T) = h \Big| \omega(n) = k\Big) \ll_A \alpha^h (1-\alpha)^{k-h}\binom{k}{h},$$

and the theorem now follows from Lemma 4·1, taking $\beta = \alpha \pm \psi\sqrt{\alpha(1-\alpha)/k}$.

Acknowledgements. The author thanks Gérald Tenenbaum and the anonymous referee for helpful comments.

REFERENCES

[1] R. B. ASH. *Information Theory*. Corrected reprint of the 1965 original. (Dover Publications, Inc., New York, 1990). xii+339 pp.

[2] A. D. BARBOUR, L. HOLST and S. JANSON. Poisson approximation. Oxford Studies in Probability, 2. Oxford Science Publications (The Clarendon Press, Oxford University Press, New York, 1992).

[3] C. DARTYGE and G. TENENBAUM. Sommes des chiffres de multiples d'entiers. (French. English, French summary) [Sums of digits of multiples of integers] *Ann. Inst. Fourier (Grenoble)* **55** (2005), no. 7, 2423–2474.

[4] H. DELANGE. Sur des formules de Atle Selberg, Acta Arith. **19** (1971), 105–146.

[5] P. D. T. A. ELLIOTT. Probabilistic number theory. II. Central limit theorems. Grundlehren Math. Wiss. [Fundamental Principles of Mathematical Sciences], **240** (Springer–Verlag, Berlin-New York, 1980).
[6] P. ERDŐS. Note on the number of prime divisors of integers. *J. London Math. Soc.* **12** (1937), 308–314.
[7] P. ERDŐS and M. KAC. The Gaussian law of errors in the theory of additive number theoretic functions. *Amer. J. Math.* **62** (1940), 738–742.
[8] G. HALÁSZ. On the distribution of additive and the mean values of multiplicative arithmetic functions. *Studia Sci. Math. Hungar.* **6** (1971), 211–233.
[9] R. R. HALL and G. TENENBAUM. *Divisors*, Cambridge Tracts in Mathematics (Cambridge University Press, Cambridge, 1988 Vol 90).
[10] G. H. HARDY and S. RAMANUJAN. The normal number of prime factors of a number n, *Quart. J. Math. Oxford* **48** (1917), 76–92.
[11] J. KUBILIUS. Probabilistic methods in the theory of numbers. *Uspehi Mat. Nauk (N.S.)* **11** (1956), 2(68), 31–66 (Russian); = *Amer. Math. Soc. Translations*, **19** (1962), 47–85.
[12] J. KUBILIUS. *Probabilistic methods in the theory of numbers* Transl. Math. Monogr. vol. 11 (American Mathematical Society, Providence, R.I. 1964).
[13] E. LANDAU. *Handbuch der Lehre von der Verteilung der Primzahlen* (Chelsea, 1951). Reprint of the 1909 original.
[14] A. P. MANGEREL. Topics in multiplicative and probabilistic number theory, PhD. thesis UNIVERSITY OF TORONTO (2018).
[15] A. SÁRKŐZY. Remarks on a paper of G. Halász: "On the distribution of additive and the mean values of multiplicative arithmetic functions" (*Studia Sci. Math. Hungar.* **6** (1971), 211–233). *Period. Math. Hungar.* **8** (1977), no. 2, 135–150.
[16] L. G. SATHE. On a problem of Hardy on the distribution of integers having a given number of prime factors. II. *J. Indian Math. Soc. (N.S.)* **17** (1953), 83–141; III. ibid, **18** (1954), 27–42; IV. ibid, **18** (1954), 43–81.
[17] A. SELBERG. Note on a paper by L. G. Sathe. *J. Indian Math. Soc. (N.S.)* **18** (1954), 83–87.
[18] G. TENENBAUM. Crible d'Ératosthène et modèle de Kubilius. (French. English summary) [The sieve of Eratosthenes and the model of Kubilius] *Number Theory in Progress*, vol. 2 (Zakopane–Kościelisko, 1997), 1099–1129 (de Gruyter, Berlin, 1999).
[19] G. TENENBAUM. A rate estimate in Billingsley's theorem for the size distribution of large prime factors. *Quart. J. Math. Oxford* **51** (2000), no. 3, 385–403.
[20] G. TENENBAUM. *Introduction to analytic and probabilistic number theory*. Graduate Studies in Mathematics, vol. 163 (American Mathematical Society, Providence, RI, third edition, 2015). Translated from the 2008 French edition by Patrick D. F. Ion.
[21] G. TENENBAUM. Moyennes effectives de fonctions multiplicatives complexes. (French. English summary) [Effective means for complex multiplicative functions] Ramanujan J. **44** (2017), no. 3, 641–701. Errata: to appear in the Ramanujan J., also available on the author's web page: http://www.iecl.univ-lorraine.fr/ Gerald.Tenenbaum/PUBLIC/Prepublications_et_publications/
[22] C. TUDESQ. Majoration de la loi locale de certaines fonctions additives. *Arch. Math. (Basel)*, **67**(6), (1996), 465–472.