International Journal of Number Theory
Vol. 3, No. 1 (2007) 1–18

PRIME DIVISORS ARE POISSON DISTRIBUTED

ANDREW GRANVILLE*

Département de Mathématiques et Statistique
Université de Montréal
CP 6128 succ Centre-Ville, Montréal, Québec H3C 3J7, Canada
andrew@dms.umontreal.ca

Received 8 June 2005
Accepted 11 April 2006

We show that the set of prime factors of almost all integers are "Poisson distributed", and that this remains true (appropriately formulated) even when we restrict the number of prime factors of the integer. Our results have inspired analogous results about the distribution of cycle lengths of permutations.

Keywords: Prime divisors; local distribution; permutation; poisson; Hardy–Ramanujan; factorization.

Mathematics Subject Classifications 2000: 11A51, 11N60, 11N05, 62E17, 62E20, 11Y05

1. Introduction

Hardy and Ramanujan showed that almost all integers n have $\sim \log\log n$ prime factors (whether or not they are counted with multiplicity). The set of numbers $\{\log\log p : p|n\}$ is therefore typically a set of $\sim \log\log n$ numbers inside the interval $(\log\log 2, \log\log n)$. How might we expect these numbers to be distributed within the interval? Other than near the beginning and end of the interval we might, for want of a better idea, guess that these numbers are "randomly distributed" in some appropriate sense given that the average gap is 1. That guess, correctly formulated, turns out to be correct. We formulated "randomly distributed" as:

A sequence of finite sets $S_1, S_2, \ldots$ is called "Poisson distributed" if there exist functions $m_j, K_j, L_j \to \infty$ monotonically as $j \to \infty$ such that $S_j \subseteq [0, m_j]$ and $|S_j| \sim m_j$; and for all λ, $1/L_j \leq \lambda \leq L_j$ and integers k in the range

*L'auteur est partiellement soutenu par bourse de la CRSNG du Canada.

$0 \le k \le K_j$, we have

$$\frac{1}{m_j} \int_{\substack{0 \\ \#\{S_j \cap [t, t+\lambda]\} = k}}^{m_j} 1 dt \sim e^{-\lambda} \frac{\lambda^k}{k!}.$$

Our main result is that there is a set of integers $\mathcal{N}$, containing all but $o(x)$ integers $\le x$, such that the sets $\{\log\log p : p|n\}_{n\in\mathcal{N}}$ are indeed "Poisson distributed":

Theorem 1. *Let* $y = \log\log\log x/(\log\log\log\log x)^2$. *There exists a set of integers* $\mathcal{N}$ *such that* $\#\{n \le x; n \notin \mathcal{N}\} \le x/2^{y/20}$, *for which*

$$\mu_n(L;k) := \frac{1}{\log\log n} \int_{\substack{t=0 \\ \#\{p|n:\ t \le \log\log p < t+L\} = k}}^{\log\log n} 1 dt = e^{-L} \frac{L^k}{k!} \{1 + O(2^{-y/20})\} \tag{1.1}$$

for every $n \le x$ *with* $n \in \mathcal{N}$, *for all* L *in the range* $1/y \le L \le y/50$ *and all integers* $k \le y/(\log y)^2$.

There are related results in the literature, but none which seem to imply this. Galambos [3], and DeKoninck and Galambos [2], proceed a little differently (and their proofs are significantly different): Let $p_1(n) \le p_2(n) \le \cdots \le p_w(n)$ be the prime divisors of n. Galambos [3] shows that if j and $w - j \to \infty$ then $\log\log p_j(n)$ is normally distributed with mean j and variance j. Moreover, he shows that $\log\log p_{j+1}(n) - \log\log p_j(n)$ is distributed as a Poisson random variable with parameter 1. DeKoninck and Galambos [2] extend this to show that for any fixed k,

$$(\log\log p_{j+1}(n) - \log\log p_j(n), \log\log p_{j+2}(n) - \log\log p_{j+1}(n), \ldots,$$
$$\log\log p_{j+k}(n) - \log\log p_{j+k-1}(n))$$

is distributed as a k-tuple of independent Poisson random variables with parameter 1. These results are certainly not implied by Theorem 1, and we cannot see how they imply Theorem 1, though it seems plausible that this should be so. The results are compatible and show how the sets $\{\log\log p : p|n\}$ do typically take on "random structure".

Using our methods, we are able, in Section 6, to go somewhat further along the same lines as [2].

Theorem 2. *For sufficiently large* k, *suppose* $k_0 = 1 < k_1 < k_2 < \cdots < k_{m-1} < k_m$ *are integers with* $k_1 \to \infty$, *and* $\log\log x - \log\log\log x - k_m \to \infty$, *and otherwise* $k_{j+1} - k_j = 1$ *or* $\to \infty$. *Then the values*

$$(\log\log p_{k_1}(n), \log\log p_{k_2}(n) - \log\log p_{k_1}(n), \ldots, \log\log p_{k_m}(n) - \log\log p_{k_{m-1}}(n))$$

over $n \le x$ are distributed as m independent random variables with

$$\log\log p_{k_{j+1}}(n) - \log\log p_{k_j}(n)$$

(i) *Poisson with parameter* 1 *if* $k_{j+1} - k_j = 1$;
(ii) *Normal with mean and variance* $k_{j+1} - k_j$ *if* $k_{j+1} - k_j \to \infty$.

Evidently (1.1) can only hold in the range given in Theorem 1 if $\omega(n) \sim \log\log n$ (where $\omega(n)$ denotes the number of distinct prime factors of n). So what happens if $\omega(n)$ is considerably smaller or larger? In other words, for a given $k, 1 \le k \le \log x/\log\log x$, what do the sets $\{\log\log p : p|n\}$ typically look like when we consider only those $n \le x$ with $\omega(n) = k$? In this case, the average gap between elements is $(\log\log n)/k$ so we might expect a Poisson distribution with this parameter. However, there are several obvious problems with this guess:

- If k is bounded then there cannot be a non-discrete distribution function for gaps between elements of $\{\log\log p : p|n\}$ for each individual n since there are a bounded number of elements of this set. We deal with this case separately and prove in Section 7:

Theorem 3. *For large x and $2 \le k = o(\log\log x)$ consider $S_k(x)$ the set of integers $n \le x$ with $\omega(n) = k$. Let $p_1(n) < p_2(n) < \cdots < p_k(n)$ be the distinct prime factors of n. The elements*

$$\{\log\log p_i(n)/\log\log n : 1 \le i \le k-1\}$$

are distributed on $(0,1)$ like $k-1$ random numbers, as we vary over $n \in S_k(x)$. More precisely, for any $\epsilon \in (1/(\log\log x)^2, 1/k)$, for any $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{k-1} \le \alpha_k = 1$ with $\alpha_{j+1} - \alpha_j > \epsilon$, there are $(k-1)!\epsilon^{k-1}\{1 + O(k/\log\log x)\}|S_k(x)|$ integers $n \in S_k(x)$ with $\log\log p_i(n)/\log\log x \in (\alpha_i, \alpha_i + \epsilon)$ for each $1 \le i \le k-1$.

- We cannot have many i with $p_i(n) > n^{\log k/k}$, evidently no more than $k/\log k = o(k)$ if $k \to \infty$. So we must restrict our attention to $\{\log\log p \in (0, \log((\log n)/k)) : p|n\}$. Notice that the average gap between these points is $\sim \log((\log n)/k)/k$.

We will prove in Section 8, using deep and difficult results on $S_k(x)$ due to Hildebrand and Tenenbaum [4] (by modifying the proof of Theorem 1), that for all but $o(S_\ell(x))$ of the integers $n \in S_\ell(x)$, the sets

$$\left\{ \frac{\log\log p}{\frac{1}{\ell}\log\log(n^{1/\ell})} : \; p|n, \; p \le n^{1/\ell} \right\}$$

are indeed Poisson distributed.

Theorem 4. *Let $A(x) \to \infty$ as $x \to \infty$ (but very slowly). Suppose that ℓ is an integer with $\ell \le (\log x)/(\log\log x)^{A(x)}$ and $\ell \to \infty$ as $x \to \infty$. Define*

$\nu = \log(\log x/(\ell\log\ell))$ *and* $y = [(\log\nu)^{1/4}]$. *There exists a set of integers* $\mathcal{N} \subset S_\ell(x)$ *such that* $\#\{n \le x; n \notin \mathcal{N}\} \le |S_\ell(x)|/2y^{1/4}$, *for which*

$$\frac{1}{\log\left(\dfrac{\log n}{\ell(\log\log n)^3}\right)} \int\limits_{\substack{t = 3\log\log\log n \\ \#\{p|n:\ t \le \log\log p < t+\lambda\nu/\ell\} = k}}^{\log((\log n)/\ell)} 1 dt = e^{-\lambda}\frac{\lambda^k}{k!}\{1 + O(y^{-1/13})\}$$

for every $n \in \mathcal{N}$, *for all* λ *in the range* $1/\log y \le \lambda \le (1/4)\log y$ *and all integers* $k \le \log y/(\log\log y)^2$.

Arratia, Barbour and Tavaré [1] discuss how many statistics of the sets $\{\log\log p : p|n\}$, as we run through the integers, are strongly related to the statistics of

$$\{\log d_1(\sigma), \ldots, \log d_m(\sigma) : \sigma \in S_n\}$$

as we run through the permutations σ on n letters, with cycle lengths $d_1(\sigma) \le d_2(\sigma) \le \cdots \le d_m(\sigma)$. Indeed our Theorems 1, 3 and 4 do have an analogy in this setting, results that we have proved in another paper.

In [1], the authors discuss formulating the statistics of the sets $\{\log\log p : p|n\}$ in terms of the Poisson–Dirichlet distribution, and this has been taken a lot further by Tenenbaum [5]. If one could prove that the statistics of $\{\{\log\log p : p|n\} : n \le x\}$ are sufficiently close to the Poisson–Dirichlet distribution, that is with enough uniformity, then all of our results here would follow (with some linear conditioning). However, we have been unable to do so with what is currently proved in this direction.

One question of interest would be to prove analogous results for the prime divisors of $\{f(n) :\ n \le x\}$ where $f(t) \in \mathbb{Z}[t]$. Here the results would necessarily reflect how $f(t)$ factors mod p on average over primes p, and it would be interesting to see how much things vary depending on the choice of f.

2. Some Simple Lemmas

Lemma 2.1. *If* C *is a finite set of positive numbers then*

$$0 \le \left(\sum_{c\in C} c\right)^k - \sum_{\substack{c_i \in C \\ c_i \text{ distinct}}} c_1c_2\cdots c_k \le \binom{k}{2}\sum_{c\in C} c^2\left(\sum_{c\in C} c\right)^{k-2}.$$

Proof. Expanding

$$\left(\sum_{c\in C} c\right)^k \quad \text{we get} \quad \sum_{c_1,\ldots,c_k\in C} c_1\cdots c_k$$

which are all positive terms. This gives the first inequality. If the $\{c_i\}$ are not all distinct then there exists $1 \le i < j \le k$ with $c_i = c_j$. There are $\binom{k}{2}$ choices of i

and j. For a given choice of i and j, our sum becomes

$$\leq \sum_{c_j \in C} c_j^2 \sum_{\substack{c_g \in C \\ g \neq i,j}} \prod_{g \neq i,j} c_g = \sum_{c_j \in C} c_j^2 \Bigg(\sum_{c \in C} c \Bigg)^{k-2}.$$

Throughout we shall use the fact, deduced from the prime number theorem, that

$$\sum_{u<p<v} \frac{1}{p} = \log\left(\frac{\log v}{\log u}\right) + O(e^{-\sqrt{\log u}}). \tag{2.1}$$

Define

$$M = M(B, x; k, L) := \sum_{\substack{p_1 < p_2 < \cdots < p_{k-1} < p_k < p_1^{e^L} \\ B \leq p_1 \leq x}} \left(L - \log\left(\frac{\log p_k}{\log p_1}\right)\right) \frac{1}{p_1 p_2 \cdots p_k}.$$

Lemma 2.2. *If $k, 1/L = e^{o(\sqrt{\log B})}$ and $x > B^2$ then*

$$M(B, x; k, L) = \frac{L^k}{k!} \log\left(\frac{\log x}{\log B}\right) \{1 + O(e^{-\{1+o(1)\}\sqrt{\log B}})\}.$$

Proof. If p_1 and p_k are given then the sum here is

$$\sum_{p_1<p_2<\cdots<p_{k-1}<p_k} \frac{1}{p_2 \cdots p_{k-1}}.$$

By Lemma 2.1 this is best approximated by

$$\frac{1}{(k-2)!} \left(\sum_{p_1<p<p_k} \frac{1}{p} \right)^{k-2}$$

with error term

$$\leq \frac{1}{(k-2)!} \binom{k-2}{2} \sum_{p_1<p<p_k} \frac{1}{p^2} \left(\sum_{p_1<p<p_k} \frac{1}{p} \right)^{k-4}.$$

Since $p_k < p_1^{e^L}$ this is

$$\ll \frac{1}{(k-2)!} \frac{k^2}{p_1 \log p_1} (L + O(e^{-\sqrt{\log B}}))^{k-4}$$

by (2.1). This provides an error term for M of

$$\ll \sum_{B<p_1\leq x} \frac{1}{p_1} \sum_{p_1<p_k<p_1^{e^L}} \frac{L}{p_k} \frac{1}{(k-2)!} \frac{k^2}{p_1 \log p_1} (L + O(e^{-\sqrt{\log B}}))^{k-4}$$

$$\ll \frac{k^2 L^{k-2}}{(k-2)!} \sum_{B<p_1\leq x} \frac{1}{p_1^2 \log p_1} \left(1 + O\left(\frac{e^{-\sqrt{\log B}}}{L}\right)\right)^{k-3} \ll \frac{k^2 L^{k-2}}{(k-2)! B^2 \log B}$$

by (2.1), which is acceptable in our error term.

The main term is, by (2.1),

$$\frac{1}{(k-2)!}\sum_{\substack{B<p_1\le x\\ p_1<p_k<p_1^{e^L}}}\left(L-\log\left(\frac{\log p_k}{\log p_1}\right)\right)\frac{1}{p_1p_k}\left(\log\left(\frac{\log p_k}{\log p_1}\right)+O\left(e^{-\sqrt{\log B}}\right)\right)^{k-2}$$

$$=\frac{L^k}{k!}\sum_{B<p_1\le x}\frac{1}{p_1}\left\{1+O\left(\left(k+\frac{k^2}{L}\right)e^{-\sqrt{\log B}}\right)\right\}$$

by the prime number theorem, since $k, 1/L = e^{o(\sqrt{\log B})}$,

$$=\frac{L^k}{k!}\log\left(\frac{\log x}{\log B}\right)\{1+O(e^{-\{1+o(1)\}\sqrt{\log B}})\}$$

for $x > B^2$.

3. Preparatory Estimates

For a given x we consider integers $n \le x$, and let $Q(n)$ be the prime divisors of n in the interval

$$I = [\exp((\log x)^{\delta}), \exp((\log x)^{1-\delta})]$$

where $\delta = \delta(x)$. We assume

$$\frac{1}{\log\log\log x} \le L \le \log\log x \quad \text{and} \quad k = o(\log\log\log x/\log\log\log\log x).$$

Define

$$A_{k,L}(n) = \sum_{\substack{p_1<p_2<\cdots<p_k\in Q(n)\\ p_k<p_1^{e^L}}}\left\{L-\log\left(\frac{\log p_k}{\log p_1}\right)\right\},$$

so that

$$\frac{1}{x}\sum_{n\le x}A_{k,L}(n) = \sum_{\substack{p_1<p_2<\cdots<p_k\in I\\ p_k<p_1^{e^L}}}\left\{L-\log\left(\frac{\log p_k}{\log p_1}\right)\right\}\frac{1}{x}\left[\frac{x}{p_1\cdots p_k}\right].$$

Note that $p_1\cdots p_k \le p_k^k \le x$ since $k < (\log x)^{\delta}$.

In each of those terms we have

$$\frac{1}{x}\left[\frac{x}{p_1\cdots p_k}\right] = \frac{1}{p_1\cdots p_k} + O\left(\frac{1}{x}\right)$$

so that the accumulated error terms are

$$\ll \frac{L}{x}\sum_{p_1\cdots p_k\le x}1 \ll \frac{L}{\log x}\frac{(\log\log x+O(1))^{k-1}}{(k-1)!}. \tag{3.1}$$

The main term is

$$\sum_{\substack{p_1<p_2<\cdots<p_k\in I\\ p_k<p_1^{e^L}}} \left\{L-\log\left(\frac{\log p_k}{\log p_1}\right)\right\}\frac{1}{p_1\cdots p_k}$$

which lies between $M(B,z_1;k,L)$ and $M(B,z_2;k,L)$ where

$$B=\exp((\log x)^{\delta}),\quad z_2=\exp((\log x)^{1-\delta})$$

and $z_2=z_1^{e^L}$.

Therefore, by Lemma 2.2, we have a main term

$$\frac{L^k}{k!}\{(1-2\delta)\log\log x+O(L)\}$$

and combining this with (3.1) gives

$$\frac{1}{x}\sum_{n\le x}A_{k,L}(n)=\frac{L^k}{k!}\{(1-2\delta)\log\log x+O(L)\}. \tag{3.2}$$

We also have

$$\frac{1}{x}\sum_{n\le x}A_{k,L}(n)^2=\sum_{\substack{p_1<\cdots<p_k\in I\\ q_1<\cdots<q_k\in I\\ p_k<p_1^{e^L}\\ q_k<q_1^{e^L}}}\left\{L-\log\left(\frac{\log p_k}{\log p_1}\right)\right\}\left\{L-\log\left(\frac{\log q_k}{\log q_1}\right)\right\}\frac{1}{x}$$
$$\times\left[\frac{x}{L_{p,q}}\right]$$

where $L_{p,q}=LCM[p_1\cdots p_k,q_1\cdots q_k]$. We proceed in a similar way to above. Now $L_{p,q}$ has between k and $2k$ prime factors. For a given such number with $k+i$ prime factors, there are $\binom{k+i}{k}$ choices for $p_1\cdots p_k$; then the remaining i primes are amongst the q_i so there are $\binom{k}{k-i}$ choices for the other q_i. Thus our error term is

$$\ll\frac{L^2}{x}\sum_{i=0}^{k}\binom{k}{k-i}\binom{k+i}{k}\sum_{r_1\cdots r_{k+i}\le x}1$$
$$\ll\frac{kL^2}{\log x}\sum_{i=0}^{k}\frac{(\log\log x)^{k+i-1}}{i!^2(k-i)!}\ll\frac{kL^2}{\log x}\frac{(\log\log x)^{2k-1}}{k!^2}$$

since $k<\sqrt{\frac{1}{2}\log\log x}$. The main term is the square of what we had before except that we need to account for terms where $L_{p,q}<p_1\cdots p_kq_1\cdots q_k$.

This error term is

$$\le L^2\sum_{\substack{p_1<\cdots<p_k\in I\\ q_1<\cdots<q_k\in I\\ p_k<p_1^{e^L},\,q_k<q_1^{e^L}\\ \text{some } p_i=q_j}}\frac{1}{LCM[p_1\cdots p_k,q_1\cdots q_k]}.$$

Now

$$q_k < q_1^{e^L} \le q_j^{e^L} = p_i^{e^L} \le p_k^{e^L} < p_1^{e^{2L}},$$

and similarly $p_k < q_1^{e^{2L}}$.

Moreover, if $LCM[p_1 \cdots p_k, q_1 \cdots q_k] = r_1 \cdots r_{k+i}$ then, as above, there are $(k+i)!/i!^2(k-i)$ choices for $p_1 \cdots p_k, q_1 \cdots q_k$. Therefore the above is

$$\begin{aligned} &\le L^2 \sum_{i=0}^{k-1} \sum_{\substack{r_1 < \cdots < r_{k+i} \in I \\ r_{k+i} < r_1^{e^{2L}}}} \frac{1}{r_1 \cdots r_{k+i}} \frac{(k+i)!}{i!^2(k-i)!} \\ &\ll L^2 \sum_{i=0}^{k-1} \frac{(k+i)!}{i!^2(k-i)!} \frac{(2L)^{k+i-1}}{(k+i-1)!} \log\log x \\ &\ll \log\log x \frac{(2L)^{2k}}{k!^2}(1 + 1/L^{2k})k^{O(k)}. \end{aligned}$$

Combining the above gives, since $(k(1+1/L))^k = (\log\log x)^{o(1)}$,

$$\frac{1}{x}\sum_{n\le x} A_{k,L}(n)^2 = \frac{L^{2k}}{k!^2}\{(1-2\delta)^2(\log\log x)^2 + O(L\log\log x + (\log\log x)^{1+o(1)})\}. \tag{3.3}$$

Therefore, by (3.2) and (3.3), we deduce that

$$\frac{1}{x}\sum_{n\le x}\left|\frac{A_{k,L}(n)}{(1-2\delta)\log\log x} - \frac{L^k}{k!}\right|^2 \ll \frac{L^{2k}}{k!^2}\left(\frac{L}{\log\log x} + \frac{1}{(\log\log x)^{1-o(1)}}\right). \tag{3.4}$$

4. Proof of Theorem 1, Almost

Remember that $y = \log\log\log x/(\log\log\log\log x)^2$ and $k = o(y/\log y)$. Let $m = [y/4]$ and $1/y \le L \le y/16e$, and select B so that $y < e^{o(\sqrt{\log B})}$. Define $P(n) = \{\log\log p : p \in Q(n)\}$ and

$$\sigma_{k,L}(n) := \frac{1}{(1-2\delta)\log\log x} \int_{\substack{\delta\log\log x \\ \#\{P(n)\cap[t,t+L]\} = k}}^{(1-\delta)\log\log x} 1\mathrm{d}t.$$

Now

$$\begin{aligned} \sum_{k\ge K}\binom{k}{K}\sigma_{k,L}(n) &= \frac{1}{(1-2\delta)\log\log x}\sum_{p_1<\cdots<p_K\in Q(n)} \int_{\substack{t=\delta\log\log x \\ t\le\log\log p_1 \\ \log\log p_K < t+L}}^{(1-\delta)\log\log x} 1\mathrm{d}t \\ &= A_{K,L}(n)/(1-2\delta)\log\log x. \end{aligned}$$

Therefore

$$\sigma_{k,L}(n) = \sum_{K\geq k} (-1)^{K-k} \binom{K}{k} A_{K,L}(n)/(1-2\delta)\log\log x$$

so that

$$\sigma_{k,L}(n) - \frac{e^{-L}L^k}{k!} = \sum_{K\geq k} (-1)^{K-k} \binom{K}{k} \left\{ \frac{A_{K,L}(n)}{(1-2\delta)\log\log x} - \frac{L^K}{K!} \right\}. \tag{4.1}$$

We break the sum on the right of (4.1) into two parts: Those $K \leq m$ and those $K > m$ (and note that $k = o(m)$). For small K, we use (3.4) and for large K, a trivial estimate:

$$\begin{aligned} \frac{1}{x}\sum_{n\leq x} \frac{A_{K,L}(n)}{(1-2\delta)\log\log x} &\ll \frac{L}{\log\log x} \sum_{p_1\in I} \frac{1}{p_1} \sum_{p_1<p_2<\cdots<p_k<p_1^{e^L}} \frac{1}{p_2\cdots p_k} \\ &\ll L\frac{1}{(K-1)!}(L+O(e^{-\sqrt{\log B}}))^{K-1} \end{aligned}$$

by (2.1), so that, taking $K = k + r$,

$$\begin{aligned} \frac{1}{x}\sum_{n\leq x}\sum_{K\,>\,m} \binom{K}{k} &\left| \frac{A_{K,L}(n)}{(1-2\delta)\log\log x} - \frac{L^K}{K!} \right| \\ &\ll \sum_{r>m-k} (r+k)\frac{(L+O(e^{-\sqrt{\log B}}))^r}{r!}\frac{L^k}{k!} \\ &\ll \frac{e^{-L}L^k}{k!}\frac{1}{2^{y/4}} \end{aligned} \tag{4.2}$$

for any fixed $C, 0 < C < \frac{1}{16}(\log 4 - 1/e)$. From (3.4), using Cauchy's inequality we obtain, taking $K = r + k$,

$$\begin{aligned} &\frac{1}{x}\sum_{n\leq x}\left| \sum_{k\leq K\leq m} (-1)^{K-k}\binom{K}{k}\left\{ \frac{A_{K,L}(n)}{(1-2\delta)\log\log x} - \frac{L^K}{K!} \right\} \right|^2 \\ &\leq \sum_{k\leq K\leq m} 1 \sum_{k\leq K\leq m} \binom{K}{k}^2 \frac{1}{x}\sum_{n\leq x}\left| \frac{A_{K,L}(n)}{(1-2\delta)\log\log x} - \frac{L^K}{K!} \right|^2 \\ &\ll m \sum_{k\leq K\leq m} \binom{K}{k}^2 \frac{L^{2K}}{K!^2}\frac{1}{(\log\log x)^{1-o(1)}} \\ &\ll \frac{L^{2k}}{k!^2}\frac{y}{(\log\log x)^{1-o(1)}}\sum_{r\leq m-k}\frac{L^{2r}}{r!^2} \ll \frac{L^{2k}}{k!^2}\frac{e^{2L}}{(\log\log x)^{1-o(1)}} \ll \frac{e^{-2L}L^{2k}}{k!^2 2^{y/2}}. \end{aligned} \tag{4.3}$$

Write the right-hand side of (4.1) as $\sigma_n = a_n + b_n$ where a_n is the sum of terms with $K \le m$, so that, by Cauchy's inequality,

$$\frac{1}{x}\sum_{n\le x}|\sigma_n| = \frac{1}{x}\sum_{n\le x}|a_n| + \frac{1}{x}\sum_{n\le x}|b_n| \le \left(\frac{1}{x}\sum_{n\le x}|a_n|^2\right)^{1/2} + \frac{1}{x}\sum_{n\le x}|b_n|.$$

Thus by (4.2) and (4.3) we obtain

$$\frac{1}{x}\sum_{n\le x}\left|\sigma_{k,L}(n) - e^{-L}\frac{L^k}{k!}\right| \ll e^{-L}\frac{L^k}{k!}2^{-y/4}. \tag{4.4}$$

5. Proof of Theorem 1, Completed

Let $\mathcal{N}$ begin as the set of all integers, and $L_j = y^{-1}(1+2^{-y/12})^j$ for $0 \le j \le J := [2^{y/12+1}\log y]$. For each j, remove from $\mathcal{N}$ those n for which

$$\left|\sigma_{k,L}(n) - e^{-L}\frac{L^k}{k!}\right| \ge \frac{1}{2^{y/12+1}}\, e^{-L}\frac{L^k}{k!}.$$

There are $\ll 2^{-y/6}x$ such $n \le x$ for each pair $k \le y/(\log y)^2$ and $j \le J$ by (4.4). This gives a total of $\le 2^{-y/13}x$ such $n \le x$, which is acceptable.

Now if $L_j \le L < L_{j+1}$ then, since $e^{-L_{j+1}}\frac{L_{j+1}^k}{k!} = e^{-L_j}\frac{L_j^k}{k!}\{1 + O(\frac{y}{2^{y/12}})\}$,

$$\begin{aligned}\sigma_{k,L}(n) &= \sum_{i\le k}\sigma_{i,L}(n) - \sum_{i\le k-1}\sigma_{i,L}(n) \le \sum_{i\le k}\sigma_{i,L_j}(n) - \sum_{i\le k-1}\sigma_{i,L_{j+1}}(n)\\ &\le e^{-L}\frac{L^k}{k!} + O\left(\frac{y}{2^{y/12}}\sum_{i\le k}e^{-L}\frac{L^i}{i!}\right) = e^{-L}\frac{L^k}{k!} + O\Big(\frac{y}{2^{y/12}}\Big).\end{aligned}$$

Now, in our range for k, we have $e^{-L}L^k/k! \ge e^{-L+O(y/\log y)}$ and so $\sigma_{k,L}(n) \le e^{-L}L^k/k! \times (1 + O(e^{y/50+o(y)}/2^{y/12})) = e^{-L}L^k/k! \times (1 + O(1/2^{y/20}))$. Moreover, we get an analogous lower bound from the inequality

$$\sigma_{k,L}(n) \ge \sum_{i\le k}\sigma_{i,L_{j+1}}(n) - \sum_{i\le k-1}\sigma_{i,L_j}(n).$$

Now let $\delta = 3\log\log\log\log\log x/\log\log x$ so that $y \le e^{o(\sqrt{\log B})}$ where $B = \exp((\log x)^\delta)$. Since $\mu_n(L;k) = \sigma_{k,L}(n) + O(\delta)$ we deduce the result.

6. Local Distributions

The fundamental lemma of the sieve implies:

Lemma 6.1. *If m is a product of primes $\le x^{1/u}$ then*

$$\#\{n \le x : (n,m) = 1\} = \frac{\phi(m)}{m}x\{1 + O(u^{-u}) + O(e^{-\sqrt{\log x}})\}.$$

Corollary 6.2. *Let $p_1 < p_2 < \cdots < p_r$ be primes and let m be the product of all the primes $\leq z$, excluding $p_1, \ldots, p_r$. If $z^r = x^{o(1)}$ with $p_r \leq z$, and $z \to \infty$ then*

$$\#\{n \leq x : p_1 \cdots p_r | n \text{ and } (n/p_1 \cdots p_r, m) = 1)\}$$
$$\sim \frac{e^{-\gamma}}{\log z} \frac{x}{(p_1 - 1)(p_2 - 1) \cdots (p_r - 1)}.$$

Our next preparatory result is rather tricky, but important.

Proposition 6.3. *For integer $r \geq 0$ and real $z \geq 1$, with $r = o((\log\log z/\log\log\log\log z)^2)$, we have*

$$\sum_{p_1 < p_2 < \cdots < p_r \leq z} \frac{1}{(p_1 - 1)(p_2 - 1) \cdots (p_r - 1)} \sim c(r/\log\log z) \frac{(\log\log z)^r}{r!} \tag{6.1}$$

where

$$c(u) = e^{\gamma u} \prod_{p \leq y} \left(1 + \frac{u}{p - 1}\right) \left(1 + \frac{1}{p - 1}\right)^{-u}$$

with $y = \exp((\log\log\log z)^2)$.

Proof. By Lemma 2.1, the left-handside is

$$\frac{1}{r!} \left(\sum_{p \leq z} \frac{1}{p - 1} \right)^r + O\left(\frac{r^2}{r!} \sum_{p \leq z} \frac{1}{p^2} \left(\sum_{p \leq z} \frac{1}{p - 1} \right)^{r-2} \right)$$
$$= \frac{(\log\log z + O(1))^r}{r!} + O\left(r^2 \frac{(\log\log z + O(1))^{r-2}}{r!} \right)$$
$$\sim (\log\log z)^r / r! \tag{6.2}$$

if $r = o(\log\log z)$.

For r such that $r/\log\log z \gg 1$, but $r = o((\log\log z/\log\log\log\log z)^2)$, let $y = \exp((\log\log\log z)^2)$. By (2.1) we have

$$\sum_{y < p \leq z} \frac{1}{p - 1} = \log\log z - 2\log\log\log\log z + O\left(\frac{1}{\log\log z} \right)$$

which when inserted into the argument in (6.2) gives

$$\sum_{y < p_1 < p_2 < \cdots < p_j \leq z} \frac{1}{(p_1 - 1)(p_2 - 1) \cdots (p_j - 1)} \sim \frac{\left(\log\left(\frac{\log z}{\log y} \right) \right)^j}{j!},$$

for all $j \leq r$. Writing $T = \log(\log z/\log y)$, this then gives

$$\sum_{p_1 < p_2 < \cdots < p_r \leq z} \frac{1}{(p_1 - 1) \cdots (p_r - 1)}$$
$$\sim \sum_{j=0}^{r} \frac{T^{r-j}}{(r - j)!} \sum_{p_1 < p_2 < \cdots p_j \leq y} \frac{1}{(p_1 - 1) \cdots (p_j - 1)}. \tag{6.3}$$

Now

$$\frac{T^{r-j}}{(r-j)!} = \left(\frac{T^r}{r!}\right)\left(\frac{r}{T}\right)^j \prod_{i=2}^{j-1}\left(1-\frac{i}{r}\right).$$

Note that $\prod_{i=2}^{j-1}(1-i/r) \sim 1$ if $j = o(\sqrt{r})$. This product is always ≤ 1 so the terms $j > \varepsilon\sqrt{r}$ contribute

$$\ll \frac{T^r}{r!} \sum_{j>\varepsilon\sqrt{r}} \frac{\left(\frac{r}{T}(\log\log y + O(1))\right)^j}{j!} \ll_\varepsilon \frac{T^r}{r!2^{\sqrt{r}}} = o\left(\frac{T^r}{r!}\right).$$

Thus the right-handside of (6.3) is

$$\sim \frac{T^r}{r!} \prod_{p\le y}\left(1+\frac{r/T}{p-1}\right). \tag{6.4}$$

On the other hand

$$\prod_{p\le y}\left(1+\frac{1}{p-1}\right)^{r/T} = \left(e^\gamma \log y\left\{1+O\left(\frac{1}{\log\log z}\right)\right\}\right)^{r/T} \sim e^{\gamma r/T}\left(1+\frac{\log\log y}{T}\right)^r$$
$$\sim e^{\gamma r/\log\log z}(\log\log z/T)^r.$$

The result follows since $r/T - r/\log\log z \ll r\log\log\log\log z/(\log\log z)^2$.

With this preparation we can now proceed to our main task, reproving and improving the works of [2] and [3].

Suppose we select z and integer $r \ge 0$ so that

$$z^r = x^{o(1)} \quad \text{and} \quad r = o((\log\log z/\log\log\log\log z)^2).$$

Let $p_1(n) < p_2(n) < \cdots$ be the distinct prime factors of n. With m as in Corollary 6.2, we note that

$$\begin{aligned}
&\frac{1}{x}\#\{n \le x : p_r(n) \le z < p_{r+1}(n)\} \\
&= \sum_{p_1<p_2<\cdots<p_r\le z} \frac{1}{x}\#\{n\le x : p_1\cdots p_r | n \text{ and } (n/p_1\cdots p_r, m) = 1\} \\
&\sim \frac{e^{-\gamma}}{\log z} \sum_{p_1<p_2<\cdots<p_r\le z} \frac{1}{(p_1-1)\cdots(p_r-1)} &(6.5)\\
&\sim \frac{e^{-\gamma}c(r/\log\log z)}{\log z}\frac{(\log\log z)^r}{r!}. &(6.6)
\end{aligned}$$

Note that $c(1+\delta) = e^\gamma + O(\delta)$ for $\delta = O(1)$, so if $r \sim \log\log z$, say $r = \log\log z + \tau\sqrt{\log\log z}$ with $\tau = o((\log\log z)^{1/6})$, then (6.6) becomes

$$\sim e^{-\tau^2/2}/\sqrt{2\pi\log\log z}. \tag{6.7}$$

Moreover, if $r_z(n)$ is the number of prime factors of n which are $\leq z$ then (6.5) gives

$$\frac{1}{x}\#\{n \leq x : |r_z(n) - \log\log z| > (\log\log z)^{1/2+\varepsilon}\}$$
$$\ll \sum_{|r-\log\log z|>(\log\log z)^{1/2+\varepsilon}} \frac{1}{\log z}\frac{(\log\log z + O(1))^r}{r!}$$
$$\ll \exp\left(-\frac{1}{3}(\log\log z)^{2\varepsilon}\right) \ll \frac{1}{\log\log z}.$$

Thus, summing up over (6.7) we deduce the result of Galambos [3]:

$$\frac{1}{x}\#\{n \leq x : r_z(n) \leq \log\log z + \triangle\sqrt{\log\log z}\} \sim \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\triangle} e^{-t^2/2}\mathrm{d}t, \tag{6.8}$$

provided $\log z = o(\log x/\log\log x)$. This can be rephrased as: if $r \to \infty$ and $r < \log\log x - 2\log\log\log x$ then

$$\frac{1}{x}\#\{n \leq x : \log\log p_r(n) < r + \triangle\sqrt{r}\} \sim \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\triangle} e^{-t^2/2}\mathrm{d}t.$$

Suppose r and $k \to \infty$ with $k + r = O(\log\log x)$. We will study the distribution of $\log\log p_{k+r}(n) - \log\log p_k(n)$. We will do this assuming $p_1(n), p_2(n), \ldots, p_k(n)$ are given (say $p_1, \ldots, p_k$), and indeed the powers to which they appear in n, say $a_1, \ldots, a_k$, all ≥ 1; thus let $d = \prod_{i=1}^k p_i^{a_i}$ and suppose $d = x^{o(1)}$. The number of such integers with $p_{k+r}(n) \leq z < p_{k+r+1}(n)$, where $p_k(n) < z < \exp(o(\log x/\log\log x))$, is

$$\sum_{p_k<q_1<q_2<\cdots<q_r\leq z} \#\{n \leq x : dq_1\cdots q_r|n \text{ and } (n/dq_1\cdots q_r, m) = 1\}$$

where m is the product of the primes $\leq z$ except $q_1, \ldots, q_r$. Proceeding as before, since $z^r d = x^{o(1)}$ the above is

$$\sim \frac{e^{-\gamma}}{\log z}\frac{x}{d}\sum_{p_k<q_1<\cdots<q_r\leq z}\frac{1}{(q_1-1)\cdots(q_r-1)} \sim \frac{e^{-\gamma}}{\log z}\frac{x}{d}\frac{(\log(\log z/\log p_k))^r}{r!}$$

by (2.1), provided $r = o(\log(\frac{\log z}{\log p_k})\, e^{\sqrt{\log p_k}})$. The total number of integers for which the smallest k prime factors have product d is, for $m = \prod_{p\leq p_k} p$,

$$\#\{n \leq x : d|n \text{ and } (n/d, m) = 1\} \sim \frac{e^{-\gamma}}{\log p_k}\frac{x}{d};$$

so the proportion for which the next r smallest prime factors are $\leq z$ (but not the next $r+1$) is $\sim (1/T)(\log T)^r/r!$ where $T = \log z/\log p_k$; in other words, this has a Poisson distribution with parameter T. Therefore we deduce that

$$\log\log p_{k+r}(n) - \log\log p_k(n)$$

is normally distributed with mean r and variance r if $r \to \infty$ and restricted as above. This is true for each possible value of $p_k(n) < z$, and so for $p_k(n)$ in general. Moreover, this means that such distributions are independent of one another. That is, if $k = 1 < k_1 \le k_2 < \cdots < k_{m-1} < k_m = \log\log z$ with $k_{j+1} - k_j \to \infty$ for $j = 0, 1, \ldots, m-1$, then

$$\log\log p_{k_{j+1}}(n) - \log\log p_{k_j}(n), \quad j = 0, 1, \ldots, m-2$$

are statistically independent normal distribution with mean and variance $k_{j+1} - k_j$ for each j.

More can be said: if $j \ge 1$ we can allow $k_{j+1} - k_j$ to be fixed. To simplify the proof, we insert all integers from k_j to k_{j+1} into our sequence so that if $k_{j+1} - k_j$ is fixed, it equals 1. Again suppose $p_1, \ldots, p_k$ are given, $k \to \infty$, and by the above argument the proportion of such integers with $\log\log p_{k+1}(n) > \log\log p_k(n) + t$ is

$$\sim \frac{e^{-\gamma}}{\log(p_k^{e^t})} \frac{x}{d} \Bigg/ \frac{e^{-\gamma}}{\log p_k} \frac{x}{d} = e^{-t}.$$

Since this is true no matter what the values of $p_1, \ldots, p_k$, thus $\log\log p_{k+1}(n) - \log\log p_k(n)$ is Poisson with parameter 1, independent of what went before. This concludes the proof of Theorem 2.

We note here one relatively easy result: For $t \to \infty, \log\log x - u \to \infty$ and $u = t + \lambda$ with λ bounded, with m the product of the primes in $[e^{e^t}, e^{e^u}]$,

$$\begin{aligned}
&\frac{1}{x}\#\{n \le x : n \text{ has exactly } k \text{ prime factors in } [e^{e^t}, e^{e^u}]\} \\
&\quad = \sum_{e^{e^t} < p_1 < \cdots < p_k < e^{e^u}} \frac{1}{x}\#\{n \le x : p_1 \cdots p_k | n \text{ and } (n/p_1 \cdots p_k, m) = 1\} \\
&\quad \sim e^{-\lambda} \sum_{e^{e^t} < p_1 < \cdots < p_k < e^{e^u}} \frac{1}{p_1 \cdots p_k} \sim e^{-\lambda} \frac{\lambda^k}{k!}
\end{aligned}$$

by (2.1).

7. Integers with a Given Small Number of Prime Factors

Let $S_k(x) = \{n \le x : n \text{ has exactly } k \text{ prime factors}\}$. It is well known that

$$|S_k(x)| \sim \frac{x}{\log x} \frac{(\log\log x)^{k-1}}{(k-1)!} \tag{7.1}$$

for $k = o(\log\log x)$. We note that almost all integers $n \in S_k(x)$ are squarefree: for, if not, $n = p^a m$ for some prime p, integer $a \ge 2$, and $m \in S_{k-1}(x/p^a)$, and

$$\sum_{p^a, a \ge 2} |S_k(x/p^a)| \ll |S_{k-1}(x)| \ll |S_k(x)| k / \log\log x = o(|S_k(x)|)$$

by (7.1).

To prove Theorem 3 we wish to determine how many squarefree integers in $n \in S_k(x)$ have $\log\log p_i(n)/\log\log x \in (\alpha_i, \alpha_i+\epsilon)$ for each $1 \le i \le k-1$. Evidently the number is

$$\sum_{p_1 \in I_1, p_2 \in I_2, \dots, p_{k-1} \in I_{k-1}} \pi\left(\frac{x}{p_1 \cdots p_{k-1}}\right)$$

where $I_j = (\exp((\log x)^{\alpha_j}), \exp((\log x)^{\alpha_j+\epsilon}))$. Note that $p_1 \dots p_{k-1} \le \exp(k(\log x)^{\alpha_j+\epsilon}) = x^{o(1)}$ by hypothesis (as $k(\log x)^{\alpha_j+\epsilon} = o(\log x)$ since $k \ll \log\log x$ and $\alpha_j + \epsilon < 1$). Therefore $\pi(x/(p_1 \dots p_{k-1})) \sim x/(p_1 \dots p_{k-1} \log x)$ and so the above sum becomes, since the intervals I_J are disjoint,

$$\begin{aligned}\frac{x}{\log x} \prod_{j=1}^{k-1} \sum_{p_j \in I_j} \frac{1}{p_j} &\left\{1 + O\left(\frac{1}{\log x}\right)\right\} \\ &= \frac{x}{\log x} (\epsilon \log\log x)^{k-1} \left\{1 + O\left(\frac{k}{(\log x)^{\alpha_1}}\right)\right\}.\end{aligned}$$

8. Integers with a Given Large Number of Prime Factors

Let $\nu = \log(\log x/(\ell \log(\ell+1))$ where $\ell \to \infty$ and $\ell \ll \log x/(\log\log x)^2$ (so that $\nu \to \infty$). In [4], Corollaries 3 and 4 imply that

$$\frac{|S_{\ell+1}(x)|}{|S_\ell(x)|} = \frac{\nu}{\ell}\left\{1 + O\left(\frac{\log \nu}{\nu}\right)\right\}; \tag{8.1}$$

and if $1 \le d \le \sqrt{x}$ then

$$\frac{|S_\ell(x)|}{d|S_\ell(x/d)|} = \left(\frac{\log x}{\log(x/d)}\right)^{\ell/\nu - 1} \exp\left(O\left(\frac{1}{\nu} + \frac{\ell(\log d)(\log \nu)}{\nu^2 \log x}\right)\right). \tag{8.2}$$

We deduce that if d is the product of k distinct primes with $k \le \min\{\log \ell, \nu/(\log \nu)^2\}$, where each prime factor of d is $\in [(\log x)^2, x^{1/\ell}]$, then

$$\#\{n \in S_\ell(x) :\ d|n\} = \left(\frac{\ell}{\nu}\right)^k \frac{|S_\ell(x)|}{d}\left\{1 + O\left(\frac{(\log \ell)^2}{\ell} + \frac{1}{\log \nu}\right)\right\}. \tag{8.3}$$

Proof. Select m so that $d|m$ and $(m, n/m) = 1$, where $p|m$ implies p divides d.

If $0 \le j \le k$ then $\nu_{\ell-j}/(\ell-j) = \nu/\ell(1+O(j/\ell))$. Therefore multiplying together (8.1) for $\ell-1, \ell-2, \dots, \ell-k$ we find that

$$\frac{|S_{\ell-k}(x)|}{|S_\ell(x)|} = \left(\frac{\ell}{\nu}\right)^k \left\{1 + O\left(\frac{(\log \ell)^2}{\ell} + \frac{1}{\log \nu}\right)\right\};$$

replacing x by x/d here adds, at worst, an error term $O(k \log d/\log x \log\log\log x) = o(k^2/\ell)$. The right-hand side of (8.2) is $\exp(O(k/\nu + k/\ell))$, and so we have proved that

$$|S_{\ell-k}(x/d)| = \left(\frac{\ell}{\nu}\right)^k \frac{1}{d}|S_\ell(x)|\left\{1 + O\left(\frac{(\log \ell)^2}{\ell} + \frac{1}{\log \nu}\right)\right\}. \tag{8.4}$$

Now writing $n \in S_\ell(x)$ for which $d|n$ as $n = dm$, we see that if $p|(m,d)$, then n/dp has between $\ell - k$ and ℓ prime factors. Thus

$$0 \le \#\{n \in S_\ell(x) :\ d|n\} - \#\{m \in S_{\ell-k}(x/d) :\ (m,d) = 1\} \le \sum_{p|d} \sum_{i=0}^{k} S_{\ell-i}(x/dp).$$

On the other hand, if $m \in S_{\ell-k}(x/d)$ and $p|(m,d)$, then $m/p \in S_{\ell-k}(x/dp)$ or $S_{\ell-k-1}(x/dp)$, so that

$$0 \le |S_{\ell-k}(x/d)| - \#\{m \in S_{\ell-k}(x/d) :\ (m,d) = 1\} \le \sum_{p|d} \sum_{i=k}^{k+1} S_{\ell-i}(x/dp).$$

Replacing x/d by x/dp in (8.4) we deduce from the last two equations that

$$\begin{aligned} &|\#\{n \in S_\ell(x) :\ d|n\} - |S_{\ell-k}(x/d)|| \\ &\quad \le \sum_{p|d} \sum_{i=0}^{k+1} S_{\ell-i}(x/dp) \ll \sum_{p|d} \sum_{i=0}^{k+1} \left(\frac{\ell}{\nu}\right)^i \frac{1}{dp} |S_\ell(x)| \\ &\quad \ll \frac{1}{\log x} \left(\frac{\ell}{\nu}\right)^k \frac{1}{d} |S_\ell(x)| \end{aligned}$$

since $p \ge (\log x)^2$; and the result then follows from (8.4).

Theorem 4 for small ℓ, that is $\ell \le (\log\log x)^{2/3}$ is an easy consequence of Theorem 3.

In order to prove Theorem 4 for $\ell > (\log\log x)^{2/3}$ we will suitably modify the proof of Theorem 1. We will replace the interval $[\exp((\log x)^\delta), \exp((\log x)^{1-\delta})]$ there by the interval $[\exp((\log\log x)^3), x^{1/\ell}]$. We assume

$$\frac{1}{\log y} \le \lambda \le \frac{1}{4} \log y \quad \text{and} \quad k \le \frac{\log y}{(\log\log y)^2} \quad \text{where } y := [(\log \nu)^{1/4}], \tag{8.5}$$

with $\lambda = \ell L/\nu$ (and remember that $\nu \le \log\log x$). Note that $1/L \le \log x \le \exp(o((\log\log x)^3))$ so that the hypothesis of Lemma 2.2 is satisfied.

In Section 3, we now average only over $n \in S_\ell(x)$ rather than all integers $n \le x$ so we must modify the proof there. We replace the line above (3.1) with

$$\left(\frac{\ell}{\nu}\right)^k \frac{1}{p_1 \dots p_k} \left\{1 + O\left(\frac{1}{\log \nu}\right)\right\},$$

by (8.3). Then ignoring (3.1) but following through there the argument for the main term gives the right-hand side of (3.2) with L replaced by λ, and multiplied through by $1 + O(1/\log \nu)$, for the average of $A_{k,L}(n)$ over $n \in S_\ell(x)$.

Proceeding in the same way for the mean square of $A_{k,L}(n)$ over $n \in S_\ell(x)$, we again replace the trivial estimate for the ratio that comes up by an application of

(8.3), so we multiply the ith term in the sum in the display above (3.3) through by $\ll (\ell/\nu)^{k+i}$. This leads to

$$\frac{1}{|S_\ell(x)|} \sum_{n \in S_\ell(x)} \left| \frac{A_{k,L}(n)}{\log(\log x/(\ell(\log\log x)^3))} - \frac{\lambda^k}{k!} \right|^2 \ll \frac{\lambda^{2k}}{k!^2} \left(\frac{1}{\log \nu} + \frac{L}{\log\log x} + \frac{\nu}{\ell(\log\log x)^{1-o(1)}} \right)$$

in place of (3.4). But since $\nu \le \log\log x$ the quantity in parentheses here becomes $O(1/\log \nu)$. Thus (4.3) can be replaced by the bound

$$\ll \frac{\lambda^{2k}}{k!^2} \frac{y}{\log \nu} \sum_{r \le m-k} \frac{\lambda^{2r}}{r!^2} \ll \frac{\lambda^{2k}}{k!^2} \frac{y e^{2\lambda}}{\log \nu} \ll \frac{e^{-2\lambda}\lambda^{2k}}{k!^2} \frac{1}{(\log \nu)^{1/2}}.$$

To develop the analogy to (4.2) we need a version of (8.3) where d has arbitrarily many prime factors. To find this we start with (8.4) (in our range (8.5)): given d with lots of prime factors (though no more than ℓ and all from $[(\log x)^2, x^{1/\ell}]$), we rewrite d as $d_1 d_2 \cdots d_t$ where $\omega(d_i) = y$ for $1 \le i \le t-1$ and deduce from (8.4) that $|S_{\ell_{j-1}-\omega(d_j)}(x/D_j)| \le (\ell_{j-1}/\nu_{j-1})^{\omega(d_j)} |S_{\ell_{j-1}}(x/D_{j-1})| \exp(O(1/\log \nu_j))/d_j$ for $j = 1, 2, \ldots, t$, where $D_j = d_1 d_2 \cdots d_j$ and $D_0 = 1$, with $\ell_j = \ell - yj \le \ell$ and also $\nu_j = \log(\log(x/D_j)/(\ell_j \log(\ell_j + 1)))$. Now, as each prime factor of d is $\le x^{1/\ell}$, therefore $\ell \log D_j \le yj \log x$ and so $\ell \log(x/D_j) \ge \ell_j \log x$, which implies that $\nu_j \ge \nu$. Therefore we have proved that $|S_{\ell_{j-1}-\omega(d_j)}(x/D_j)| \le (\ell/\nu)^{\omega(d_j)} |S_{\ell_{j-1}}(x/D_{j-1})| \exp(O(1/\log \nu))/d_j$, and then multiplying these altogether gives, since $t \ll \omega(d)/y$,

$$|S_{\ell-k}(x/d)| \le \left(\frac{\ell}{\nu} \left(1 + O\left(\frac{1}{y \log \nu} \right) \right) \right)^k \frac{1}{d} \, |S_\ell(x)|.$$

Then, by the same argument used to deduce (8.3) from (8.4), we obtain

$$\#\{n \in S_\ell(x) :\ d|n\} \ll \left(\frac{\ell}{\nu} \left(1 + O\left(\frac{1}{y \log \nu} \right) \right) \right)^k \frac{|S_\ell(x)|}{d}. \tag{8.6}$$

Therefore, in our argument, we have the analogous estimate to the display above (4.2) but now multiplied through by $((\ell/\nu)(1+O(1/y\log\nu)))^K$, which leads to (4.2) with L replaced by λ.

Combining the above we obtain the analogy to (4.4) where in the right-hand side we replace L by λ, and $2^{-y/4}$ by $1/y$.

Finally we need the analogy to Section 5: Here $\mathcal{N}$ starts out as $S_\ell(x)$, we let $\lambda_j = (\log y)^{-1}(1 + y^{-1/3})^j$ for $0 \le j \le J := [2y^{1/3} \log\log y]$. For each j, remove from $\mathcal{N}$ those $n \in S_\ell(x)$ for which we get an error term $\ge (1/2y^{1/3}) e^{-\lambda} \lambda^k / k!$. There are $\ll |S_\ell(x)|/y^{2/3}$ such $n \in S_\ell(x)$ for each pair $k \le \log y$ and $j \le J$. This gives a total of $\le |S_\ell(x)|/2y^{1/4}$ such $\in S_\ell(x)$, which is acceptable.

We then proceed as in Section 5, so long as $k \le \log y/(\log\log y)^2$ and obtain $|\sigma_{k,L}(n) - e^{-\lambda}\lambda^k/k!| \le e^{-\lambda}\lambda^k/(k! y^{1/13})$. The result follows.

Acknowledgment

The author would like to thank Par Kurlberg for useful discussions on such issues.

References

[1] R. Arratia, A. D. Barbour and S. Tavaré, Random combinatorial structures and prime factorizations, *Notices Amer. Math. Soc.* **44** (1997) 903–910.
[2] J.-M. DeKoninck and J. Galambos, The intermediate prime divisors of integers, *Proc. Amer. Math. Soc.* **101** (1987) 213–216.
[3] J. Galambos, The sequences of prime divisors of integers, *Acta. Arith.* **31** (1976) 213–218.
[4] A. Hildebrand and G. Tenenbaum, On the number of prime factors of an integer, *Duke Math. J.* **56** (1988) 471–501.
[5] G. Tenenbaum, A rate estimate in Billingley's theorem for the size distribution of large prime factors, *Quart. J. Math.* **51** (2000) 385–403.