JOURNAL OF NUMBER THEORY 9, 187–202 (1977)

Second-Order Terms for the Variances and Covariances of the Number of Prime Factors—Including the Square Free Case*

PERSI DIACONIS

Stanford University, Stanford, California 94305

FREDERICK MOSTELLER†

University of California, Berkeley, California 94720

AND

HIRONARI ONISHI

City College of New York, New York, New York 10031

Communicated by N. Ankeny

Received March 12, 1975

We obtain second-order terms for the variance and covariance of $\Omega(n)$ and $\omega(n)$, the number of prime divisors counted with and without multiplicity, and connect these results to a formula of Renyi. We discuss the heuristic connection with the Landau–Sathe extension of the prime number theorem and develop new expansions for the mean and variance of $\omega(n)$ in the square free case.

INTRODUCTION AND SUMMARY

Let $\omega(n)$ be the number of distinct prime factors of the positive integer n (for example, $\omega(12) = 2$). Hardy and Ramanujan [1] determined the arithmetic mean of $\omega(n)$ for $n \leqslant x$:

THEOREM 1 (Hardy and Ramanujan [1]).

$$\bar{\omega}_x \stackrel{d}{=} (1/x) \sum_{n \leqslant x} \omega(n) = \log\log x + B_1 + O(1/\log x),$$

* This work was facilitated by Grant GS-32327X1 from the National Science Foundation. We are indebted to Cleo Youtz for the numerical calculations of Section 5.

† Frederick Mosteller is Miller Research Professor at the University of California on leave from Harvard University, 1974–1975.

ISSN 0022-314X

where

$$B_1 = \gamma + \sum_p (\log(1 - (1/p)) + (1/p)),$$

where γ is Euler's constant and throughout this paper $\sum_p$ denotes a sum over all primes p.

For a simple proof of Theorem 1 see [2, p. 355]. Hardy and Ramanujan also showed that the variance of $\omega(n)$ about the mean $\bar{\omega}_x$ is asymptotically loglog x, showing the equivalent of:

$$\mathrm{Var}_x\, \omega \overset{d}{=} (1/x) \sum_{n \leqslant x} (\omega(n) - \bar{\omega}_x)^2 \sim \operatorname{loglog} x.$$

In Section 1 we derive a more precise result:

THEOREM 2.

$$\mathrm{Var}_x\, \omega = \operatorname{loglog} x + B_3 + O\left(\frac{\operatorname{loglog} x}{\log x}\right),$$

where

$$B_3 = B_1 - (\pi^2/6) - \sum_p (1/p^2).$$

We have introduced B_3 before B_2 to make the B_i of this paper agree with classical usage.

In Section 2 we consider $\Omega(n)$, the number of prime factors of n counted with multiplicity (for example, $\Omega(12) = 3$). The results for Ω and ω are surprisingly similar. First of all, we have

THEOREM 3 (Hardy and Ramanujan [1]).

$$\bar{\Omega}_x \overset{d}{=} (1/x) \sum_{n \leqslant x} \Omega(n) = \operatorname{loglog} x + B_2 + O(1/\log x),$$

where

$$B_2 = B_1 + \sum_p (1/p(p-1)).$$

We shall prove

THEOREM 4.

$$\mathrm{Var}_x\, \Omega \overset{d}{=} \frac{1}{x} \sum_{n \leqslant x} (\Omega(n) - \bar{\Omega}_x)^2 = \operatorname{loglog} x + B_4 + O\left(\frac{\operatorname{loglog} x}{\log x}\right),$$

where

$$B_4 = B_1 - (\pi^2/6) + \sum_p ((2p-1)/p(p-1)^2).$$

In Section 3 we obtain the covariance of Ω and ω:

THEOREM 5.

$$\begin{aligned}\mathrm{Cov}_x(\Omega, \omega) &\overset{d}{=} \frac{1}{x} \sum_{n \leqslant x} (\Omega(n) - \bar{\omega}_x)(\omega(n) - \bar{\omega}_x) \\ &= \operatorname{loglog} x + B_1 - \frac{\pi^2}{6} + O\left(\frac{\operatorname{loglog} x}{\log x}\right).\end{aligned}$$

Now consider the difference $\Omega(n) - \omega(n)$. Theorems 1 and 3 imply that

$$\bar{\Omega}_x - \bar{\omega}_x = \sum_p (1/p(p-1)) + O(1/\log x), \tag{1}$$

while, since $\mathrm{Var}_x(\Omega - \omega) = \mathrm{Var}_x\, \Omega + \mathrm{Var}_x\, \omega - 2\, \mathrm{Cov}_x(\Omega, \omega)$, Theorems 2, 4, and 5 imply

THEOREM 6.

$$\mathrm{Var}_x(\Omega - \omega) = \sum_p \frac{p^2 + p - 1}{p^2(p-1)^2} + O\left(\frac{\operatorname{loglog} x}{\log x}\right).$$

The asymptotic mean and variance of $\Omega - \omega$ are related to a result of Renyi as explained in Section 3.

In Section 4, we look into the asymptotic mean and variance of $\nu(n) = \omega(n)\, \mu(n)^2$, where $\mu(n)$ is the familiar Möbius function. We first establish:

THEOREM 7. *If $a = p_1 \cdots p_l$ is square free, then*

$$N(a; x) \overset{d}{=} \sum_{n \leqslant x/a} \mu(an)^2 = \frac{6}{\pi^2} \cdot \frac{x}{(p_1+1) \cdots (p_l+1)} + O\left(\left(\frac{x}{a}\right)^{1/2}\right),$$

where the implied constant depends only on l.

Note Theorem 7 implies $\frac{1}{3}$ of all square free numbers are even, so being square free is not "independent" of being even or the result would be $\frac{1}{2} \cdot (6/\pi^2)$. Using Theorem 7 we get:

THEOREM 8.

$$\bar{\nu}_x \overset{d}{=} (1/x) \sum_{n \leqslant x} \nu(n) = (6/\pi^2)(\operatorname{loglog} x + B_8 + O(1/\log x)),$$

where

$$B_8 = B_1 - \sum_p (1/p(p+1)).$$

THEOREM 9.

$$\begin{aligned}
\operatorname{Var}_x \nu &\overset{d}{=} \frac{1}{x} \sum_{n \leqslant x} (\nu(n) - \bar{\nu}_x)^2 \\
&= \frac{6}{\pi^2}\left(1 - \frac{6}{\pi^2}\right) [(\operatorname{loglog} x + B_8)^2 + (\operatorname{loglog} x + B_9)] \\
&\quad + O\left[\frac{\operatorname{loglog} x}{\log x}\right],
\end{aligned}$$

where

$$B_9 = B_8 - \frac{\pi^2}{6} - \sum_p \frac{1}{(p+1)^2} - 2\left(\sum_p \frac{1}{p(p+1)}\right)^2.$$

The results of this paper are compared with the findings of Landau and Sathe that asymptotically $\omega(n) - 1$ and $\Omega(n) - 1$ have Poisson "probability" distributions with means loglog x. The asymptotic distribution of the number of square free numbers has a separate but related probabilistic rationale.

Finally, Section 5 presents actual computer counts of the functions $\bar{\omega}_x$, $\operatorname{Var}_x \omega$, $\bar{\Omega}_x$, and $\operatorname{Var}_x \Omega$ for x up to one million and these counts are compared with asymptotic values.

The Appendix summarizes notation and lists the results proved in the paper along with numerical values for the constants involved.

1. VARIANCE OF ω

We now prove Theorem 2. Since

$$\operatorname{Var}_x \omega = (1/x) \sum_{n \leqslant x} \omega(n)^2 - \bar{\omega}_x^2,$$

it is sufficient to prove:

$$\begin{aligned}
\sum_{n \leqslant x} \omega(n)^2 = {} & x(\operatorname{loglog} x)^2 + (2B_1 + 1)\, x \operatorname{loglog} x + (B_1^2 + B_3)\, x \\
& + O(x \operatorname{loglog} x/\log x). \qquad (1\text{-}1)
\end{aligned}$$

The basic estimate used throughout is

$$F(t) \overset{d}{=} \sum_{p \leqslant t} (1/p) = \operatorname{loglog} t + B_1 + O(1/\log t). \qquad (1\text{-}2)$$

For a proof, see [2, pp. 349–353].

LEMMA 1.

$$\sum_{pq\leqslant x}\frac{1}{pq}=(\operatorname{loglog}x+B_1)^2-\frac{\pi^2}{6}+O\left(\frac{\operatorname{loglog}x}{\log x}\right),$$

where the sum is over prime pairs (p, q) and counts the pair (p, q) as distinct from (q, p) when $p \neq q$.

Proof. Since

$$\sum_{pq\leqslant x}(1/pq)=\Bigl(\sum_{p\leqslant x}(1/p)\Bigr)^2-\sum_{\substack{p,q\leqslant x\\ pq>x}}(1/pq),$$

it is sufficient to show:

$$S\stackrel{d}{=}\sum_{\substack{p,q\leqslant x\\ pq>x}}\frac{1}{pq}=\sum_{p\leqslant x}\frac{1}{p}\sum_{x/p<q\leqslant x}\frac{1}{q}=\frac{\pi^2}{6}+O\left(\frac{\operatorname{loglog}x}{\log x}\right). \tag{1-3}$$

Let $x' = x/e$ (where e is the base of the natural logarithm). Thus $\log x' = \log(x-1)$. Break the sum S into S_1 for $p \leqslant x'$ and S_2 for $x' < p \leqslant x$; $S = S_1 + S_2$. Using (1-2), we get

$$\begin{aligned}S_2&\ll\sum_{x'<p\leqslant x}\frac{1}{p}\sum_{q\leqslant x}\frac{1}{q}\ll\operatorname{loglog}x\left(\log\frac{\log x}{\log x'}+\frac{1}{\log x'}\right)\\&\ll\frac{\operatorname{loglog}x}{\log x}.\end{aligned} \tag{1-4}$$

While, again by (1-2),

$$S_1=\sum_{p\leqslant x'}\frac{1}{p}\left(\log\frac{\log x}{\log(x/p)}+O\left(\frac{1}{\log(x/p)}\right)\right)=S_3+O(S_4),$$

where

$$S_3=\sum_{p\leqslant x'}(1/p)\log(1-(\log p/\log x))^{-1},$$

$$S_4=\sum_{p\leqslant x'}(1/p(\log x-\log p)).$$

Integration by parts using (1-2) gives, letting $2^- = \lim_{h\to 0} 2-h$, $h > 0$,

$$\begin{aligned}S_4&=\int_{2^-}^{x'}\frac{dF(t)}{\log x-\log t}\ll\int_2^{x'}\frac{d\log t}{\log t(\log x-\log t)}+\frac{1}{\log x}\\&\ll\frac{\operatorname{loglog}x}{\log x}.\end{aligned} \tag{1-5}$$

Integrating by parts twice gives:

$$S_3 = \int_{2^-}^{x'} \log\left(1 - \frac{\log t}{\log x}\right)^{-1} dF(t) = S_5 + O\left(\frac{\text{loglog } x}{\log x}\right),$$

where

$$S_5 = \int_2^{x'} \log\left(1 - \frac{\log t}{\log x}\right)^{-1} \frac{d \log t}{\log t}.$$

Expanding $\log(1 - (\log t/\log x))^{-1}$ into powers of $(\log t/\log x)$ and interchanging the sum and integral, we get that

$$S_5 = \sum_{n=1}^{\infty} \frac{1}{n^2}\left(1 - \frac{1}{\log x}\right)^n + O\left(\frac{1}{\log x}\right).$$

Writing $N = \log x$, for $n \leqslant N$,

$$(1 - (1/\log x))^n = 1 + O(n/\log x).$$

Thus breaking the sum into those $n \leqslant N$ and those $n > N$ we get that

$$S_5 = \sum_{n=1}^{\infty} \frac{1}{n^2} + O\left(\frac{1}{N}\right) + O\left[\frac{\log N}{\log x}\right] = \frac{\pi^2}{6} + O\left[\frac{\text{loglog } x}{\log x}\right].$$

This proves (1-3). ▮

We can now prove (1-1). Since $\omega(n)(\omega(n) - 1) = \sum_{pq|n} 1 - \sum_{p^2|n} 1$,

$$\sum_{n\leqslant x} \omega(n)^2 = \sum_{n\leqslant x} \omega(n) + \sum_{pq\leqslant x} [x/pq] - \sum_{p^2\leqslant x} [x/p^2]. \tag{1-6}$$

The first sum is known from Theorem 1. As for the last sum, it is easy to see that

$$\sum_{p^2\leqslant x} [x/p^2] = x \sum_p (1/p^2) + O(x^{1/2}).$$

Finally, Chebychevs Theorem [2, p. 9] and computations very similar to [2, pp. 368–370] lead to the estimate

$$\sum_{pq\leqslant x} 1 = O(x \text{ loglog } x/\log x)$$

in an "elementary" way. Using Lemma 1 we get that

$$\sum_{pq\leqslant x}\left[\frac{x}{pq}\right]=x\sum_{pq\leqslant x}\frac{1}{pq}+O\left(\frac{x\operatorname{loglog}x}{\log x}\right)$$

$$=x(\operatorname{loglog}x)^2+2B_1x\operatorname{loglog}x+\left(B_1{}^2-\frac{\pi^2}{6}\right)x$$

$$+O\left(\frac{x\operatorname{loglog}x}{\log x}\right).$$

Putting this into (1-6) we get (1-1) and thus Theorem 2.

2. Variance of Ω

We now prove Theorem 4. This time it is sufficient to prove

$$\sum_{n\leqslant x}\Omega(n)^2=x(\operatorname{loglog}x)^2+(2B_2+1)\,x\operatorname{loglog}x+(B_2{}^2+B_4)\,x$$

$$+O\left(\frac{x\operatorname{loglog}x}{\log x}\right). \tag{2-1}$$

First we note that

$$\Omega(n)^2=\sum_{p^aq^b\mid n}1+\sum_{p^a\mid n}a.$$

Summing this for $n\leqslant x$ yields

$$\sum_{n\leqslant x}\Omega(n)^2=\sum_{p^aq^b\leqslant x}[x/p^aq^b]+\sum_{p^a\leqslant x}a[a/p^a].$$

Combiing this with (1-6) we get

$$T\overset{d}{=}\sum_{n\leqslant x}\Omega(n)^2-\sum_{n\leqslant x}\omega(n)^2=\sum_{\substack{p^aq^b\leqslant x\\ a\text{ or }b>1}}\left[\frac{x}{p^aq^b}\right]+\sum_{\substack{p^a\leqslant x\\ a>1}}a\left[\frac{x}{p^a}\right]+\sum_{p^2\leqslant x}\left[\frac{x}{p^2}\right].$$

Removing the brackets introduces an error of $O(x\operatorname{loglog}x/x)$. Thus

$$\frac{T}{x}=\sum_{\substack{p^aq^b\leqslant x\\ a\text{ or }b>1}}\frac{1}{p^aq^b}+\sum_{\substack{p^a<x\\ a>1}}\frac{a}{p^a}+\sum_{p\leqslant x^{1/2}}\frac{1}{p^2}+O\left(\frac{\operatorname{loglog}x}{x}\right).$$

We know the last sum and it is easy to obtain for the second sum that

$$\sum_{\substack{p^a\leqslant x\\ a>1}}\frac{a}{p^a}=\sum_p\frac{2p-1}{p(p-1)^2}+O\left(\frac{1}{\log x}\right).$$

As for the first sum, we have

$$\sum_{\substack{p^a q^b \leqslant x \\ a \text{ or } b>1}} \frac{1}{p^a q^b} = 2 \sum_{\substack{pq^b \leqslant x \\ b>1}} \frac{1}{pq^b} + \sum_{\substack{p^a q^b \leqslant x \\ a \text{ and } b>1}} \frac{1}{p^a q^b}.$$

Since

$$\sum_{\substack{p^a \leqslant x \\ a>1}} \frac{1}{p^a} = \sum_p \frac{1}{p(p-1)} + O\left(\frac{1}{\log x}\right) \stackrel{d}{=} C_1 + O\left(\frac{1}{\log x}\right),$$

$$\sum_{\substack{p^a q^b \\ a \text{ and } b>1}} \frac{1}{p^a q^b} = C_1{}^2 + O\left(\frac{1}{\log x}\right).$$

Finally, using (1-2) and (1-5),

$$\begin{aligned} \sum_{\substack{pq^b \leqslant x \\ b>1}} \frac{1}{pq^b} &= \sum_{p \leqslant x/4} \frac{1}{p}\left(C_1 + O\left(\frac{1}{\log(x/p)}\right)\right) \\ &= C_1(\log\log x + B_1) + O\left(\frac{\log\log x}{\log x}\right). \end{aligned} \tag{2-2}$$

Putting these together we get

$$\begin{aligned} T/x = {} & 2C_1 \log\log x + 2C_1 B_1 + C_1{}^2 \\ & + \sum_p \frac{2p-1}{p(p-1)^2} + \sum \frac{1}{p^2} + O\left(\frac{\log\log x}{\log x}\right). \end{aligned}$$

Since $B_2 = B_1 + C_1$, combining this with (1-1) we get (2-1) and thus Theorem 4.

3. Covariance of Ω and ω and a Formula of Renyi

We first prove Theorem 5. Since

$$\omega(n)(\Omega(n) - 1) = \sum_{pq^b | n} 1,$$

$$\begin{aligned} \sum_{n \leqslant x} \Omega(n)\, \omega(n) &= \sum_{pq^b \leqslant x} \left[\frac{x}{pq^b}\right] + \sum_{n \leqslant x} \omega(n) \\ &= \sum_{n \leqslant x} \omega(n)^2 + \sum_{\substack{pq^b \leqslant x \\ b>1}} \left[\frac{x}{pq^b}\right] + \sum_{p^2 \leqslant x} \left[\frac{x}{p^2}\right], \end{aligned}$$

where we have used (1-6). Using (1-1), (2-2), we get

$$\frac{1}{x} \sum_{n \leqslant x} \Omega(n)\, \omega(n) = (\operatorname{loglog} x)^2 + (2B_1 + 1 + C_1) \operatorname{loglog} x + B_1^2 + B_1C_1 + B_1 - \frac{\pi^2}{6} + O\left(\frac{\operatorname{loglog} x}{\log x}\right). \tag{3-1}$$

Since

$$\operatorname{Cov}_x(\Omega, \omega) = (1/x) \sum_{n \leqslant x} \Omega(n)\, \omega(n) - \bar{\Omega}_x \bar{\omega}_x ,$$

we get Theorem 5 from Theorems 1, 3, and (3-1).

The correlation coefficient

$$R_x(\omega, \Omega) \stackrel{d}{=} \operatorname{Cov}_x(\omega, \Omega)/[(\operatorname{Var}_x \omega)(\operatorname{Var}_x \Omega)]^{1/2}$$

is the classical measure of the degree of linear dependence between ω and Ω. Theorems 2, 4, and 5 show that $R_x(\omega, \Omega)$ is asymptotically 1, and so asymptotically Ω and ω are "linearly dependent."

As we noted in the introduction, Theorem 6 follows from Theorems 2, 4, and 5. Let $S_k = \{n \mid \Omega(n) - \omega(n) = k\}$. Renyi [5] showed that the density

$$h_k \stackrel{d}{=} \lim_{x \to \infty} (1/x) \sum_{\substack{n \leqslant x \\ n \in S_k}} 1$$

exists and determined the generating function of the numbers h_k to be

$$f(Z) \stackrel{d}{=} \sum_{k=0}^{\infty} h_k Z^k = \prod_p \left(1 - \frac{1}{p}\right)\left(1 + \frac{1}{p - Z}\right) = \prod_p \left(1 + \frac{Z - 1}{p(p - Z)}\right), \tag{3-2}$$

where the product is over all primes p. For a proof and short discussion of Renyi's theorem see the monograph by Kac [3, p. 64]. Some immediate deductions from Renyi's theorem are:

(a) $f(1) = \sum_{k=0}^{\infty} h_k = 1$ so that the numbers h_k may be thought of as a probability distribution on the integers $k \geqslant 0$. Note that this may not happen in general. For example, if $T_k = \{n \mid \omega(n) = k\}$, then the density of T_k is zero for all k and hence the associated generating function is identically 0.

(b) $f(0) = h_0 = \prod_p (1 - (1/p^2)) = 6/\pi^2$, which is the density of square free numbers.

(c) If we regard the numbers h_k as probabilities then we can use the generating function $f(Z)$ to compute the "expectation" of the difference $\Omega - \omega$. Formally,

$$\sum_{k=0}^{\infty} kh_k = f'(1) = f'(1)/f(1) = \sum_p (1/p(p-1)).$$

This coincides with the asymptotic mean given by (1).

(d) We can go further and compute the "variance" of $\Omega - \omega$ from the generating function. Formally,

$$\sum_{k=0}^{\infty} k^2 h_k - \left(\sum_{k=0}^{\infty} kh_k\right)^2 = f''(1) + f'(1) - f'(1)^2$$

$$= \sum_p \frac{2p-1}{p^2(p-1)^2} + \sum_p \frac{1}{p(p-1)} = \sum_p \frac{p^2+p-1}{p^2(p-1)^2}.$$

This coincides with the asymptotic variance given in Theorem 6.

The sets T_k of (a) above give an example where the mean and variance from the generating function are both zero but the asymptotic mean and variance are both infinite.

4. Variance of ν and Probabilistic Motivation

Consider the functions

$$\pi_k(x) = \sum_{\substack{n \leqslant x \\ \nu(n)=k}} 1 \quad \text{the number of square free } n \leqslant x \text{ with } k \text{ prime divisors;}$$

$$\rho_k(x) = \sum_{\substack{n \leqslant x \\ \omega(n)=k}} 1 \quad \text{the number of } n \leqslant x \text{ with } k \text{ distinct prime divisors;}$$

$$\sigma_k(x) = \sum_{\substack{n \leqslant x \\ \Omega(n)=k}} 1 \quad \text{the number of } n \leqslant x \text{ with } k \text{ prime divisors}$$

(counted with multiplicity).

Landau [4] proved that for each fixed $k > 0$,

$$\pi_k(x) \sim \rho_k(x) \sim \sigma_k(x) \sim \frac{x(\log\log x)^{k-1}}{(k-1)!\,\log x}. \tag{4-1}$$

Sathe considered the more general problem that arises when k is a function of x. In particular, he showed that if $k \sim \operatorname{loglog} x$, then

$$\pi_k(x) \sim \frac{6}{\pi^2}\frac{x(\operatorname{loglog} x)^{k-1}}{(k-1)!\log x}, \tag{4-2}$$

$$\rho_k(x) \sim \sigma_k(x) \sim \frac{x(\operatorname{loglog} x)^{k-1}}{(k-1)!\log x}. \tag{4-3}$$

Results (4-1) and (4-3) suggest that the functions $\omega(n) - 1$ and $\Omega(n) - 1$ have a limiting Poisson distribution:

$$\operatorname{Prob}(\omega(n) - 1 = k) \doteq \operatorname{Prob}(\Omega(n) - 1 = k) \doteq e^{-\lambda}\lambda^k/k!,$$

with $\lambda = \operatorname{loglog} x$. The Poisson distribution has mean

$$\sum_{i=0}^{\infty} i(e^{-\lambda}\lambda^i/i!) = \lambda,$$

and variance

$$\sum_{i=0}^{\infty} (i-\lambda)^2 (e^{-\lambda}\lambda^i/i!) = \lambda.$$

This suggests that

$$\bar{\omega}_x \sim \operatorname{Var}_x \omega \sim \bar{\Omega}_x \sim \operatorname{Var}_x \Omega \sim \operatorname{loglog} x.$$

These suggestions were verified by Hardy and Ramanujan (see Theorems 1–4).

The disparity of results for $\pi_k(x)$ with k fixed and with $k \sim \operatorname{loglog} x$ is striking.

If we start with (4-1) and reason as above, we might suppose $\nu_x \sim \operatorname{Var}_x \nu \sim \operatorname{loglog} x$. This is quite wrong in view of Theorems 8 and 9.

A Poisson distribution with parameter λ concentrates its mass close to the mean λ so that (4-2) is more relevant than (4-1). Let S be the set of square free numbers. Given $n \in S$, $\nu(n) = \omega(n) = \Omega(n)$. Since the density of S is $6/\pi^2$, the result (4-2) suggests that the conditional probability $\operatorname{Prob}(\nu(n) - 1 = k \mid n \in S)$ has a limiting Poisson distribution with $\lambda = \operatorname{loglog} x$. This leads us to expect that the conditional mean and variance should be asymptitocally $\operatorname{loglog} x$. Using Theorems 8 and 9 it is easy to derive the following results. Let $N(x) = \sum_{n\leqslant x, n\in S} 1$,

$$\bar{\omega}_{x,S} \overset{d}{=} \frac{1}{N(x)}\sum_{\substack{n\leqslant x\\ n\in S}} \omega(n) = \operatorname{loglog} x + B_8 + O\left(\frac{1}{\log x}\right),$$

$$\operatorname{Var}_{x,S}\omega \overset{d}{=} \frac{1}{N(x)}\sum_{\substack{n\leqslant x\\ n\in S}} (\omega(n) - \bar{\omega}_{x,S})^2 = \operatorname{loglog} x + B_9 + O\left(\frac{\operatorname{loglog} x}{\log x}\right).$$

Correct reasoning for $\bar{\nu}_x$ and $\mathrm{Var}_x\, \nu$ must take into account that $\nu(n) = 0$ about 0.392 ($\approx 1 - (6/\pi^2)$) of the time. A clear picture emerges by considering the following model. Suppose $\{p_k\}_{k=1}^{\infty}$ is a probability distribution on the positive integers with mean

$$\mu = \sum_{k=1}^{\infty} k p_k$$

and variance

$$\sigma^2 = \sum_{k=1}^{\infty} (k - \mu)^2\, p_k\,.$$

Define a new probability distribution $\{p_k'\}_{k=0}^{\infty}$ on the set $\{0, 1, 2, ...\}$ by $p_0' = 1 - \alpha$ and $p_k' = \alpha p_k$ for $k > 0$, $0 < \alpha < 1$. Then the new mean μ' and variance $(\sigma')^2$ are given by

$$\mu' = \alpha\mu \qquad \text{and} \qquad (\sigma')^2 = \alpha(1 - \alpha)\,\mu^2 + \alpha\sigma^2.$$

In our example, $\alpha = 6/\pi^2$, $\mu = \sigma^2 = \lambda = \log\log x$ and so we expect that

$$\nu_x \sim \frac{6}{\pi^2} \log\log x \qquad \text{and} \qquad \mathrm{Var}_x\, \nu \sim \frac{6}{\pi^2}\left(1 - \frac{6}{\pi^2}\right) (\log\log x)^2.$$

These are indeed verified by Theorems 8 and 9.

We now proceed to proofs of Theorems 8 and 9.

Proof of Theorem 7. Recall for $a = p_1 p_2 \cdots p_l$ square free, $N(a; x) \stackrel{d}{=} \sum_{n \leqslant x/a} \mu(an)^2$. If p is a prime and a is relatively prime to p then it is easy to see that $N(ap; x) = N(a; x/p) - N(ap; x/p)$. Thus

$$N(ap; x) = \sum_{i=1}^{\infty} (-1)^{i-1}\, N(a; x/p^i). \tag{4-4}$$

We now proceed by induction on l. The case $l = 0$ $(a = 1)$ is well known [2, p. 269]. Assume the result for some $l \geqslant 0$. By (4-4),

$$\begin{aligned} &N(p_1 \cdots p_l p_{l+1}\,; x) \\ &= \sum_{i=1}^{\infty} (-1)^{i-1} \left\{\frac{6}{\pi^2} \frac{x}{(p_1 + 1) \cdots (p_l + 1)\, p_{l+1}^i} + O\left(\left(\frac{x}{p_1 \cdots p_l p_{l+1}^i}\right)^{1/2}\right)\right\} \\ &= \frac{6}{\pi^2} \frac{x}{(p_1 + 1) \cdots (p_l + 1)(p_{l+1} + 1)} + O\left(\left(\frac{x}{p_1 \cdots p_l p_{l+1}}\right)^{1/2}\right). \qquad \blacksquare \end{aligned}$$

Proof of Theorem 8.

$$\frac{1}{x}\sum_{n\leqslant x}\nu(n)=\frac{1}{x}\sum_{n\leqslant x}\sum_{p\mid n}\mu(n)^2=\frac{1}{x}\sum_{p\leqslant x}\sum_{n\leqslant x/p}\mu(pn)^2$$
$$=\frac{6}{\pi^2}\sum_{p\leqslant x}\frac{1}{p+1}+O\left(\frac{1}{x^{1/2}}\sum_{p\leqslant x}\frac{1}{p^{1/2}}\right).$$

The error term is clearly $O(1/\log x)$, while

$$\sum_{p\leqslant x}\frac{1}{p+1}=\sum_{p\leqslant x}\frac{1}{p}-\sum_{p\leqslant x}\frac{1}{p(p-1)}$$
$$=\operatorname{loglog}x+B_1-\sum_{p}\frac{1}{p(p+1)}+O\left(\frac{1}{\log x}\right). \qquad ∎$$

Proof of Theorem 9. Since

$$\nu(n)^2=\sum_{pq\mid n}\mu(n)^2+\nu(n),$$
$$\frac{1}{x}\sum_{n\leqslant x}\nu(n)^2=\frac{1}{x}\sum_{n\leqslant x}\nu(n)+\frac{1}{x}\sum_{pq\leqslant x}\sum_{n\leqslant x/pq}\mu(pqn)^2$$
$$=\bar{\nu}_x+\frac{6}{\pi^2}\sum_{\substack{qp\leqslant x\\ p\neq q}}\frac{1}{(p+1)(q+1)}+O\left(\frac{1}{x^{1/2}}\sum_{pq\leqslant x}\frac{1}{(pq)^{1/2}}\right).$$

The error term is clearly $O(\operatorname{loglog} x/\log x)$.

$$\sum_{\substack{pq\leqslant x\\ p\neq q}}\frac{1}{(p+1)(q+1)}=\sum_{pq\leqslant x}\frac{1}{pq}-\sum_{pq\leqslant x}\frac{p+q+1}{pq(p+1)(q+1)}-\sum_{p^2\leqslant x}\frac{1}{(p+1)^2}.$$

The last sum contributes $-\sum_p(1/(p+1)^2)+O(1/x^{1/2})$.

$$\sum_{pq\leqslant x}\frac{p+q+1}{pq(p+1)(q+1)}=2\sum_{pq\leqslant x}\frac{1}{pq(q+1)}+\sum_{pq\leqslant x}\frac{1}{p(p+1)\,q(q+1)}.$$

Here, the second sum gives $C_2{}^2$ with $C_2=\sum_p(1/p(p+1))$,

$$\sum_{p\leqslant x}\frac{1}{p}\sum_{q\leqslant x/p}\frac{1}{q(q+1)}=\sum_{p\leqslant x}\frac{1}{p}\left(C_2+O\left(\frac{p}{x}\right)\right)$$
$$=C_2(\operatorname{loglog}x+B_1)+O\left(\frac{1}{\log x}\right).$$

Thus, using Theorem 8 and Lemma 1 we get

$$\frac{\pi^2}{6x} \sum_{n \leqslant x} \nu(n)^2 = (\operatorname{loglog} x + B_8)^2 + \operatorname{loglog} x + B_9 + O\left(\frac{\operatorname{loglog} x}{\log x}\right).$$

This and Theorem 8 give Theorem 9. ▮

5. Some Numerical Values

Table I gives numerical values for $\bar{\omega}_x$, $\operatorname{Var}_x \omega$, $\bar{\Omega}_x$, $\operatorname{Var}_x \Omega$, and for the

$$\text{residual} = \text{true value} - (\operatorname{loglog} x + \text{constant}),$$

where the constant is the one given in this paper for the parameter. The values of x are 10^2, 10^3, 10^4, 10^5, 10^6. The table shows that for $x = 10^6$, the percentage error in the estimate of $\bar{\omega}_x$ is about 1.2 %, in $\bar{\Omega}_x$ about 1 %, in $\operatorname{Var}_x \omega$ nearly 20 %, and in $\operatorname{Var}_x \Omega$ about 5 %. If we assume that the residuals of $\bar{\omega}_x$ and $\bar{\Omega}_x$ are roughly proportional to $1/\log x$ and those of $\operatorname{Var}_x \omega$ and $\operatorname{Var}_x \Omega$ are roughly proportional to $\operatorname{loglog} x/\log x$, we can use the chord from the origin to the point for $x = 10^6$ as a very rough slope. Then, letting $r(\cdot)$ stand for the residual,

$$\begin{aligned} r(\bar{\omega}_x) &\approx -0.46/\log x, \qquad x > 10^6, \\ r(\operatorname{Var}_x \omega) &\approx +1.00 \operatorname{loglog} x/\log x, \\ r(\bar{\Omega}_x) &\approx -0.47/\log x, \\ r(\operatorname{Var}_x \Omega) &\approx -0.83 \operatorname{loglog} x/\log x. \end{aligned}$$

The residuals for $x = 10^4$, 10^5, and 10^6 suggest a curve, and had a quadratic in $1/\log x$ or $\operatorname{loglog} x/\log x$ been fitted through the origin, the slopes at the origin would have been smaller in absolute value than the numerical coefficients given above.

TABLE I

Values of $\bar{\omega}_x$, $\operatorname{Var}_x\omega$, $\bar{\Omega}_x$, $\operatorname{Var}_x\Omega$ for $x = 10^2, 10^3, 10^4, 10^5, 10^6$, and the Residuals r from the Fitted Values

x	$\bar{\omega}_x$	$r(\bar{\omega}_x)$	$\operatorname{Var}_x\omega$	$r(\operatorname{Var}_x\omega)$	$\bar{\Omega}_x$	$r(\bar{\Omega}_x)$	$\operatorname{Var}_x\Omega$	$r(\operatorname{Var}_x\Omega)$
10^2	1.7100	−0.0787	0.3859	0.6944	2.3900	−0.1718	1.5179	−0.7741
10^3	2.1260	−0.0681	0.5481	0.4512	2.8770	−0.0902	2.2216	−0.4755
10^4	2.4300	−0.0535	0.7003	0.3140	3.1985	−0.0581	2.6967	−0.2901
10^5	2.6640	−0.0410	0.8462	0.2384	3.4361	−0.0420	3.0046	−0.2037
10^6	2.8537	−0.0336	0.9810	0.1909	3.6266	−0.0338	3.2331	−0.1575

Theorems 1, 2, 3, and 4 say the means and variances of the functions $\omega(n)$, $\Omega(n)$ are all asymptotically the same: $\log\log x$. However, for any x we are likely to be dealing with we really do not have $\bar{\omega}_x$ "about equal to" $\mathrm{Var}_x\,\omega$. Since it is the variable $\omega(n) - 1$ that has an asymptotic Poisson distribution, we might expect the $\bar{\omega}_x - 1$ and $\mathrm{Var}_x\,\omega$ to be about equal. Taking $c = 10^{100}$ in the asymptotic results of Theorems 1 and 2, using constants from the Appendix leads to

$$\begin{aligned}\bar{\omega}_c - 1 &\approx 4.701,\\ \mathrm{Var}_c\,\omega &\approx 3.602,\\ \bar{\Omega}_c - 1 &\approx 5.474,\\ \mathrm{Var}_c\,\Omega &\approx 6.204.\end{aligned}$$

Appendix

Numerical values are given below for constants appearing in the major theorems of this paper. The values are known accurate to at least three figures past the decimal point:

$$\bar{\omega}_x = \log\log x + 0.2615 + O\left(\frac{1}{\log x}\right);$$

$$\mathrm{Var}_x\,\omega = \log\log x - 1.8357 + O\left(\frac{\log\log x}{\log x}\right);$$

$$\bar{\Omega}_x = \log\log x + 1.0346 + O\left(\frac{1}{\log x}\right);$$

$$\mathrm{Var}_x\,\Omega = \log\log x + 0.7648 + O\left(\frac{\log x}{\log\log x}\right);$$

$$\mathrm{Cov}_x(\Omega, \omega) = \log\log x - 1.3834 + O\left(\frac{\log\log x}{\log x}\right).$$

References

1. G. Hardy and S. Ramanujan, The normal number of prime factors of a number n, *Quart. J. Math.* **48** (1917), 76–92.
2. G. Hardy and E. Wright, "Introduction to the Theory of Numbers," fourth ed., Oxford Univ. Press, London/New York, 1960.
3. M. Kac, "Statistical Independence in Probability Analysis and Number Theory," Math. Assoc. America, Rahway, N.J., 1959.

4. E. Landau, "Handbuch der Lehre von der Verteilung der Primzahlen," Vol. 1, Teubner, Leipzig, 1909.
5. A. Renyi, On the density of certain sequences of integers, *Publ. Inst. Math. (Belgrade)* **8** (1955), 157–162.
6. L. Sathe, On a problem of Hardy on the distributions of integers having a given number of prime factors, *J. Indian Math. Soc.* **17** (1953), 63–141; **18** (1954), 27–81.