Chapter 12
The Prime Divisor Functions

Here we consider only *prime* divisors of n and ask, for given order of magnitude of n, "how many prime divisors are there typically?" and "how many *different* ones are there?" Some of the answers will be rather counterintuitive. Thus, a 50-digit number (10^{21} times the age of our universe measured in picoseconds) has only about 5 different prime factors on average and – even more surprisingly – 50-digit numbers have typically fewer than 6 prime factors in all, even counting repeated occurrences of the same prime factor as separate factors.

We will also learn something about the distribution of the number of prime factors and its implications for the important factoring problem. Thus, we discover that even for numbers as large as 10^{50}, the two smallest primes, 2 and 3, account for about 25% of all prime factors!

12.1 The Number of Different Prime Divisors

In connection with encrypting messages by means of Euler's theorem, the number of distinct *prime* divisors of a given integer n, $\omega(n)$, is of prime importance. Its definition is similar to that of the divisor function $d(n)$, except that the sum is extended – as the name implies – only over the prime divisors of n:

$$\omega(n) := \sum_{p_i|n} 1. \qquad (12.1)$$

It is easily seen that $\omega(n)$ is additive, i. e., for $(n,m) = 1$,

$$\omega(nm) = \sum_{p_i|nm} 1 = \sum_{p_i|n} 1 + \sum_{p_i|m} 1 = \omega(n) + \omega(m). \qquad (12.2)$$

Of particular interest to our encrypting desires will be the behaviour of $\omega(n)$ for large n, i.e., its asymptotic behaviour. We shall try to get an idea of this behaviour by means of our usual "dirty tricks". First, we will convert the sum of those primes that divide n into a sum over *all* primes up to n, using the "probability" factor $1/p_i$:

M. Schroeder, *Number Theory in Science and Communication*, 5th ed.,
DOI 10.1007/978-3-540-85298-8_12, © Springer-Verlag Berlin Heidelberg 2009

$$\omega(n) = \sum_{p_i|n} 1 \approx \sum_{p_i \leq n} \frac{1}{p_i}. \tag{12.3}$$

This, in turn, we will convert into a sum over all *integers* up to n, using the probability factor for primality $1/\ln x$:

$$\overline{\omega}(n) \approx \sum_{x \leq n} \frac{1}{x \ln x},$$

which we will approximate by an integral:

$$\overline{\omega}(n) \approx \int_2^n \frac{\mathrm{d}x}{x \ln x} = \ln(\ln n) + 0.367\ldots. \tag{12.4}$$

Of course $\omega(n)$ is a wildly fluctuating function and exact results [12.1] are available only for asymptotic averages, just as in the case of $\phi(n)$ and $d(n)$:

$$\frac{1}{n} \sum_{k=1}^{n} \omega(k) = \ln(\ln n) + 0(1), \tag{12.5}$$

where $0(1)$ is a fancy way of writing a bounded quantity.

To get a better grip on this constant, we calculate the sum over the reciprocal primes in (12.3) out to some p_m and convert only the remaining sum to a sum over all integers using the probability factor $\ln x$:

$$\overline{\omega}(n) \approx \sum_{p_i=2}^{m} \frac{1}{p_i} + \sum_{x=p_m+1}^{n} \frac{1}{x \ln x}. \tag{12.6}$$

Approximating the second sum by an integral, we have

$$\overline{\omega}(n) \approx \sum_{p_i=2}^{p_m} \frac{1}{p_i} + \ln\ln n - \ln\ln p_m. \tag{12.7}$$

In other words, our estimate tells us that the difference between $\overline{\omega}(n)$ and $\ln\ln n$, i.e., the constant in (12.5), is given by

$$\overline{\omega}(n) - \ln\ln n \approx \lim_{p_m \to \infty} \sum_{p_i=2}^{p_m} \frac{1}{p_i} - \ln\ln p_m. \tag{12.8}$$

In the last century Kronecker, assuming that the limiting average of $\omega(n)$ existed, obtained

$$\overline{\omega}(n) = \ln(\ln n) + b_1, \tag{12.9}$$

with

$$b_1 = \gamma + \sum_{p_i=2}^{\infty} \left[\ln\left(1 - \frac{1}{p_i}\right) + \frac{1}{p_i} \right], \qquad (12.10)$$

where γ is again Euler's constant.

To compare Kronecker's constant b_1 with ours, we make use of the following asymptotic result (Merten's theorem [12.1]):

$$\lim_{p_m \to \infty} \mathrm{e}^{\gamma} \ln p_m \prod_{p_i=2}^{p_m} \left(1 - \frac{1}{p_i}\right) = 1, \qquad (12.11)$$

which yields for Kronecker's constant

$$b_1 = \lim_{p_m \to \infty} \sum_{p_i=2}^{p_m} \frac{1}{p_i} - \ln(\ln p_m), \qquad (12.12)$$

which is identical with our "crude" estimate (12.8)!

Equation (12.12) is not very suitable to obtain a numerical value for b_1, because it converges rather slowly. (In fact, even for p_m as large as 104759, the relative error is still larger than 10^{-3}.) A faster converging series is obtained by expanding the logarithm in (12.10), which yields

$$\gamma - b_1 = \sum_{p_i=2}^{\infty} \left(\frac{1}{2p_i^2} + \frac{1}{3p_i^3} + \ldots \right). \qquad (12.13)$$

Now, if we remember the Riemann zetafunction (Chap. 4), we have

$$\zeta(k) = \sum_{n=1}^{\infty} \frac{1}{n^k} = \prod_{p_i=2}^{\infty} \left(1 - \frac{1}{p_i^k}\right)^{-1}, \qquad (12.14)$$

or

$$\ln \zeta(k) = -\sum_{p_i=2}^{\infty} \ln\left(1 - \frac{1}{p_i^k}\right). \qquad (12.15)$$

Expanding the logarithm, we obtain

$$\ln \zeta(k) = \sum_{p_i=2}^{\infty} \left(\frac{1}{p_i^k} + \frac{1}{2p_i^{2k}} + \ldots \right). \qquad (12.16)$$

Introducing this result into (12.13) yields

$$\gamma - b_1 = \tfrac{1}{2} \ln \zeta(2) + \tfrac{1}{3} \ln \zeta(3) + \tfrac{1}{5} \ln \zeta(5) - \tfrac{1}{6} \ln \zeta(6) + \ldots. \qquad (12.17)$$

This sum written in terms of the Möbius function $\mu(m)$ (Chap. 21) is:

$$\gamma - b_1 = -\sum_{m=2}^{\infty} \frac{\mu(m)}{m} \ln \zeta(m). \tag{12.18}$$

This sum converges very quickly and, for just 7 terms yields a relative accuracy of about 10^{-5}. The result is

$$b_1 = 0.2614\ldots. \tag{12.19}$$

How do Milton Abramowitz and Irene Stegun feel about this? On page 862 of their *Handbook of Mathematical Functions* [12.2] they list the prime factors of the integers from 9000 to 9499 (see Fig. 12.1). I have counted a total of 1260 distinct prime factors for these 500 integers. Thus, $\overline{\omega} = 2.52$, which should be compared to our $\ln(\ln 9250) + 0.26 = 2.47$. Close enough? Certainly, because as we said before, $\omega(n)$ fluctuates and an average, even over 500 consecutive integers, is not completely smooth. (More about the fluctuations of $\omega(n)$ in a moment.)

12.2 The Distribution of $\omega(n)$

The probability that the prime factor p_i does not occur in the prime factor decomposition of $n > p_i$ is given by

$$1 - \frac{1}{p_i}.$$

The probability that it *does* occur (at least once) is therefore

$$\frac{1}{p_i}.$$

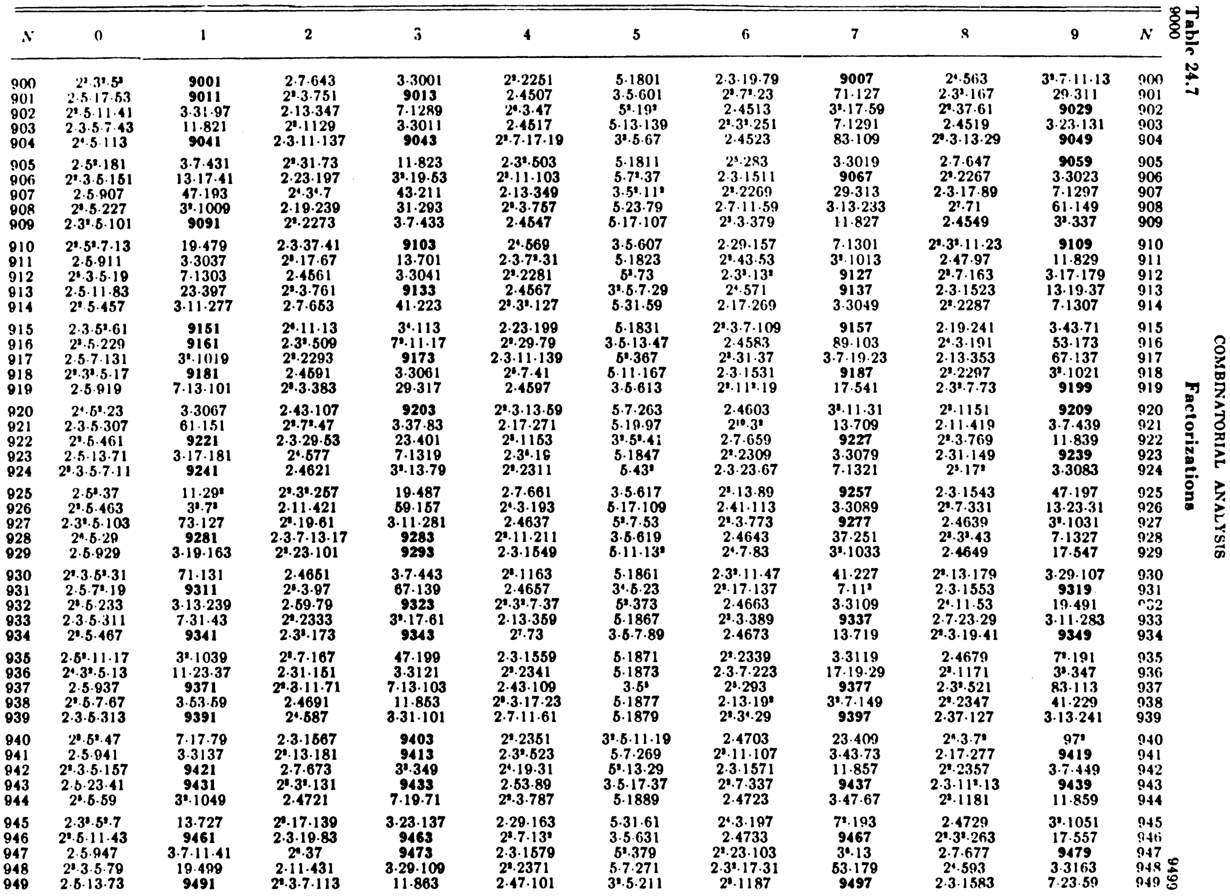
The mean occurrence is therefore

$$m_i = \frac{1}{p_i}, \tag{12.20}$$

and its variance, according to the formula for the binomial distribution for two possible outcomes, equals

$$\sigma_i^2 = \frac{1}{p_i}\left(1 - \frac{1}{p_i}\right) = m_i - \frac{1}{p_i^2}. \tag{12.21}$$

Assuming divisibility by different primes to be independent, we get for the overall mean



COMBINATORIAL ANALYSIS

Table 24.7 **Factorizations**

9000 9499

N	0	1	2	3	4	5	6	7	8	9	N
900	$2^3\cdot3^2\cdot5^3$	**9001**	$2\cdot7\cdot643$	$3\cdot3001$	$2^2\cdot2251$	$5\cdot1801$	$2\cdot3\cdot19\cdot79$	**9007**	$2^4\cdot563$	$3^2\cdot7\cdot11\cdot13$	900
901	$2\cdot5\cdot17\cdot53$	**9011**	$2^2\cdot3\cdot751$	**9013**	$2\cdot4507$	$3\cdot5\cdot601$	$2^3\cdot7^2\cdot23$	$71\cdot127$	$2\cdot3^3\cdot167$	$29\cdot311$	901
902	$2^2\cdot5\cdot11\cdot41$	$3\cdot31\cdot97$	$2\cdot13\cdot347$	$7\cdot1289$	$2^6\cdot3\cdot47$	$5^2\cdot19^2$	$2\cdot4513$	$3^2\cdot17\cdot59$	$2^2\cdot37\cdot61$	**9029**	902
903	$2\cdot3\cdot5\cdot7\cdot43$	$11\cdot821$	$2^3\cdot1129$	$3\cdot3011$	$2\cdot4517$	$5\cdot13\cdot139$	$2^2\cdot3^2\cdot251$	$7\cdot1291$	$2\cdot4519$	$3\cdot23\cdot131$	903
904	$2^4\cdot5\cdot113$	**9041**	$2\cdot3\cdot11\cdot137$	**9043**	$2^2\cdot7\cdot17\cdot19$	$3^3\cdot5\cdot67$	$2\cdot4523$	$83\cdot109$	$2^3\cdot3\cdot13\cdot29$	**9049**	904
905	$2\cdot5^2\cdot181$	$3\cdot7\cdot431$	$2^2\cdot31\cdot73$	$11\cdot823$	$2\cdot3^2\cdot503$	$5\cdot1811$	$2^6\cdot283$	$3\cdot3019$	$2\cdot7\cdot647$	**9059**	905
906	$2^2\cdot3\cdot5\cdot151$	$13\cdot17\cdot41$	$2\cdot23\cdot197$	$3^2\cdot19\cdot53$	$2^3\cdot11\cdot103$	$5\cdot7^2\cdot37$	$2\cdot3\cdot1511$	**9067**	$2^2\cdot2267$	$3\cdot3023$	906
907	$2\cdot5\cdot907$	$47\cdot193$	$2^4\cdot3^4\cdot7$	$43\cdot211$	$2\cdot13\cdot349$	$3\cdot5^2\cdot11^2$	$2^2\cdot2269$	$29\cdot313$	$2\cdot3\cdot17\cdot89$	$7\cdot1297$	907
908	$2^3\cdot5\cdot227$	$3^2\cdot1009$	$2\cdot19\cdot239$	$31\cdot293$	$2^2\cdot3\cdot757$	$5\cdot23\cdot79$	$2\cdot7\cdot11\cdot59$	$3\cdot13\cdot233$	$2^7\cdot71$	$61\cdot149$	908
909	$2\cdot3^2\cdot5\cdot101$	**9091**	$2^2\cdot2273$	$3\cdot7\cdot433$	$2\cdot4547$	$5\cdot17\cdot107$	$2^3\cdot3\cdot379$	$11\cdot827$	$2\cdot4549$	$3^3\cdot337$	909
910	$2^2\cdot5^2\cdot7\cdot13$	$19\cdot479$	$2\cdot3\cdot37\cdot41$	**9103**	$2^4\cdot569$	$3\cdot5\cdot607$	$2\cdot29\cdot157$	$7\cdot1301$	$2^2\cdot3^2\cdot11\cdot23$	**9109**	910
911	$2\cdot5\cdot911$	$3\cdot3037$	$2^3\cdot17\cdot67$	$13\cdot701$	$2\cdot3\cdot7^2\cdot31$	$5\cdot1823$	$2^2\cdot43\cdot53$	$3^2\cdot1013$	$2\cdot47\cdot97$	$11\cdot829$	911
912	$2^5\cdot3\cdot5\cdot19$	$7\cdot1303$	$2\cdot4561$	$3\cdot3041$	$2^2\cdot2281$	$5^3\cdot73$	$2\cdot3^3\cdot13^2$	**9127**	$2^3\cdot7\cdot163$	$3\cdot17\cdot179$	912
913	$2\cdot5\cdot11\cdot83$	$23\cdot397$	$2^2\cdot3\cdot761$	**9133**	$2\cdot4567$	$3^2\cdot5\cdot7\cdot29$	$2^4\cdot571$	**9137**	$2\cdot3\cdot1523$	$13\cdot19\cdot37$	913
914	$2^2\cdot5\cdot457$	$3\cdot11\cdot277$	$2\cdot7\cdot653$	$41\cdot223$	$2^3\cdot3^2\cdot127$	$5\cdot31\cdot59$	$2\cdot17\cdot269$	$3\cdot3049$	$2^2\cdot2287$	$7\cdot1307$	914
915	$2\cdot3\cdot5^2\cdot61$	**9151**	$2^6\cdot11\cdot13$	$3^4\cdot113$	$2\cdot23\cdot199$	$5\cdot1831$	$2^2\cdot3\cdot7\cdot109$	**9157**	$2\cdot19\cdot241$	$3\cdot43\cdot71$	915
916	$2^3\cdot5\cdot229$	**9161**	$2\cdot3^2\cdot509$	$7^2\cdot11\cdot17$	$2^2\cdot29\cdot79$	$3\cdot5\cdot13\cdot47$	$2\cdot4583$	$89\cdot103$	$2^4\cdot3\cdot191$	$53\cdot173$	916
917	$2\cdot5\cdot7\cdot131$	$3^2\cdot1019$	$2^2\cdot2293$	**9173**	$2\cdot3\cdot11\cdot139$	$5^2\cdot367$	$2^3\cdot31\cdot37$	$3\cdot7\cdot19\cdot23$	$2\cdot13\cdot353$	$67\cdot137$	917
918	$2^2\cdot3^3\cdot5\cdot17$	**9181**	$2\cdot4591$	$3\cdot3061$	$2^5\cdot7\cdot41$	$5\cdot11\cdot167$	$2\cdot3\cdot1531$	**9187**	$2^2\cdot2297$	$3^2\cdot1021$	918
919	$2\cdot5\cdot919$	$7\cdot13\cdot101$	$2^3\cdot3\cdot383$	$29\cdot317$	$2\cdot4597$	$3\cdot5\cdot613$	$2^2\cdot11^2\cdot19$	$17\cdot541$	$2\cdot3^2\cdot7\cdot73$	**9199**	919
920	$2^4\cdot5^2\cdot23$	$3\cdot3067$	$2\cdot43\cdot107$	**9203**	$2^2\cdot3\cdot13\cdot59$	$5\cdot7\cdot263$	$2\cdot4603$	$3^3\cdot11\cdot31$	$2^3\cdot1151$	**9209**	920
921	$2\cdot3\cdot5\cdot307$	$61\cdot151$	$2^2\cdot7^2\cdot47$	$3\cdot37\cdot83$	$2\cdot17\cdot271$	$5\cdot19\cdot97$	$2^{10}\cdot3^2$	$13\cdot709$	$2\cdot11\cdot419$	$3\cdot7\cdot439$	921
922	$2^2\cdot5\cdot461$	**9221**	$2\cdot3\cdot29\cdot53$	$23\cdot401$	$2^3\cdot1153$	$3^2\cdot5^2\cdot41$	$2\cdot7\cdot659$	**9227**	$2^2\cdot3\cdot769$	$11\cdot839$	922
923	$2\cdot5\cdot13\cdot71$	$3\cdot17\cdot181$	$2^4\cdot577$	$7\cdot1319$	$2\cdot3^5\cdot19$	$5\cdot1847$	$2^2\cdot2309$	$3\cdot3079$	$2\cdot31\cdot149$	**9239**	923
924	$2^3\cdot3\cdot5\cdot7\cdot11$	**9241**	$2\cdot4621$	$3^2\cdot13\cdot79$	$2^2\cdot2311$	$5\cdot43^2$	$2\cdot3\cdot23\cdot67$	$7\cdot1321$	$2^5\cdot17^2$	$3\cdot3083$	924
925	$2\cdot5^3\cdot37$	$11\cdot29^2$	$2^2\cdot3^2\cdot257$	$19\cdot487$	$2\cdot7\cdot661$	$3\cdot5\cdot617$	$2^3\cdot13\cdot89$	**9257**	$2\cdot3\cdot1543$	$47\cdot197$	925
926	$2^2\cdot5\cdot463$	$3^3\cdot7^3$	$2\cdot11\cdot421$	$59\cdot157$	$2^4\cdot3\cdot193$	$5\cdot17\cdot109$	$2\cdot41\cdot113$	$3\cdot3089$	$2^2\cdot7\cdot331$	$13\cdot23\cdot31$	926
927	$2\cdot3^2\cdot5\cdot103$	$73\cdot127$	$2^3\cdot19\cdot61$	$3\cdot11\cdot281$	$2\cdot4637$	$5^2\cdot7\cdot53$	$2^2\cdot3\cdot773$	**9277**	$2\cdot4639$	$3^2\cdot1031$	927
928	$2^6\cdot5\cdot29$	**9281**	$2\cdot3\cdot7\cdot13\cdot17$	**9283**	$2^2\cdot11\cdot211$	$3\cdot5\cdot619$	$2\cdot4643$	$37\cdot251$	$2^3\cdot3^3\cdot43$	$7\cdot1327$	928
929	$2\cdot5\cdot929$	$3\cdot19\cdot163$	$2^2\cdot23\cdot101$	**9293**	$2\cdot3\cdot1549$	$5\cdot11\cdot13^2$	$2^4\cdot7\cdot83$	$3^2\cdot1033$	$2\cdot4649$	$17\cdot547$	929
930	$2^2\cdot3\cdot5^2\cdot31$	$71\cdot131$	$2\cdot4651$	$3\cdot7\cdot443$	$2^3\cdot1163$	$5\cdot1861$	$2\cdot3^2\cdot11\cdot47$	$41\cdot227$	$2^2\cdot13\cdot179$	$3\cdot29\cdot107$	930
931	$2\cdot5\cdot7^2\cdot19$	**9311**	$2^5\cdot3\cdot97$	$67\cdot139$	$2\cdot4657$	$3^4\cdot5\cdot23$	$2^2\cdot17\cdot137$	$7\cdot11^3$	$2\cdot3\cdot1553$	**9319**	931
932	$2^3\cdot5\cdot233$	$3\cdot13\cdot239$	$2\cdot59\cdot79$	**9323**	$2^2\cdot3^2\cdot7\cdot37$	$5^2\cdot373$	$2\cdot4663$	$3\cdot3109$	$2^4\cdot11\cdot53$	$19\cdot491$	932
933	$2\cdot3\cdot5\cdot311$	$7\cdot31\cdot43$	$2^2\cdot2333$	$3^2\cdot17\cdot61$	$2\cdot13\cdot359$	$5\cdot1867$	$2^3\cdot3\cdot389$	**9337**	$2\cdot7\cdot23\cdot29$	$3\cdot11\cdot283$	933
934	$2^2\cdot5\cdot467$	**9341**	$2\cdot3^3\cdot173$	**9343**	$2^7\cdot73$	$3\cdot5\cdot7\cdot89$	$2\cdot4673$	$13\cdot719$	$2^2\cdot3\cdot19\cdot41$	**9349**	934
935	$2\cdot5^2\cdot11\cdot17$	$3^2\cdot1039$	$2^3\cdot7\cdot167$	$47\cdot199$	$2\cdot3\cdot1559$	$5\cdot1871$	$2^2\cdot2339$	$3\cdot3119$	$2\cdot4679$	$7^2\cdot191$	935
936	$2^4\cdot3^2\cdot5\cdot13$	$11\cdot23\cdot37$	$2\cdot31\cdot151$	$3\cdot3121$	$2^2\cdot2341$	$5\cdot1873$	$2\cdot3\cdot7\cdot223$	$17\cdot19\cdot29$	$2^3\cdot1171$	$3^3\cdot347$	936
937	$2\cdot5\cdot937$	**9371**	$2^2\cdot3\cdot11\cdot71$	$7\cdot13\cdot103$	$2\cdot43\cdot109$	$3\cdot5^5$	$2^5\cdot293$	**9377**	$2\cdot3^2\cdot521$	$83\cdot113$	937
938	$2^2\cdot5\cdot7\cdot67$	$3\cdot53\cdot59$	$2\cdot4691$	$11\cdot853$	$2^3\cdot3\cdot17\cdot23$	$5\cdot1877$	$2\cdot13\cdot19^2$	$3^2\cdot7\cdot149$	$2^2\cdot2347$	$41\cdot229$	938
939	$2\cdot3\cdot5\cdot313$	**9391**	$2^4\cdot587$	$3\cdot31\cdot101$	$2\cdot7\cdot11\cdot61$	$5\cdot1879$	$2^2\cdot3^4\cdot29$	**9397**	$2\cdot37\cdot127$	$3\cdot13\cdot241$	939
940	$2^3\cdot5^2\cdot47$	$7\cdot17\cdot79$	$2\cdot3\cdot1567$	**9403**	$2^2\cdot2351$	$3^2\cdot5\cdot11\cdot19$	$2\cdot4703$	$23\cdot409$	$2^6\cdot3\cdot7^2$	97^2	940
941	$2\cdot5\cdot941$	$3\cdot3137$	$2^2\cdot13\cdot181$	**9413**	$2\cdot3^2\cdot523$	$5\cdot7\cdot269$	$2^3\cdot11\cdot107$	$3\cdot43\cdot73$	$2\cdot17\cdot277$	**9419**	941
942	$2^2\cdot3\cdot5\cdot157$	**9421**	$2\cdot7\cdot673$	$3^3\cdot349$	$2^4\cdot19\cdot31$	$5^2\cdot13\cdot29$	$2\cdot3\cdot1571$	$11\cdot857$	$2^2\cdot2357$	$3\cdot7\cdot449$	942
943	$2\cdot5\cdot23\cdot41$	**9431**	$2^3\cdot3^2\cdot131$	**9433**	$2\cdot53\cdot89$	$3\cdot5\cdot17\cdot37$	$2^2\cdot7\cdot337$	**9437**	$2\cdot3\cdot11^2\cdot13$	**9439**	943
944	$2^5\cdot5\cdot59$	$3^2\cdot1049$	$2\cdot4721$	$7\cdot19\cdot71$	$2^2\cdot3\cdot787$	$5\cdot1889$	$2\cdot4723$	$3\cdot47\cdot67$	$2^3\cdot1181$	$11\cdot859$	944
945	$2\cdot3^3\cdot5^2\cdot7$	$13\cdot727$	$2^2\cdot17\cdot139$	$3\cdot23\cdot137$	$2\cdot29\cdot163$	$5\cdot31\cdot61$	$2^4\cdot3\cdot197$	$7^2\cdot193$	$2\cdot4729$	$3^2\cdot1051$	945
946	$2^2\cdot5\cdot11\cdot43$	**9461**	$2\cdot3\cdot19\cdot83$	**9463**	$2^3\cdot7\cdot13^2$	$3\cdot5\cdot631$	$2\cdot4733$	**9467**	$2^2\cdot3^2\cdot263$	$17\cdot557$	946
947	$2\cdot5\cdot947$	$3\cdot7\cdot11\cdot41$	$2^8\cdot37$	**9473**	$2\cdot3\cdot1579$	$5^2\cdot379$	$2^2\cdot23\cdot103$	$3^6\cdot13$	$2\cdot7\cdot677$	**9479**	947
948	$2^3\cdot3\cdot5\cdot79$	$19\cdot499$	$2\cdot11\cdot431$	$3\cdot29\cdot109$	$2^2\cdot2371$	$5\cdot7\cdot271$	$2\cdot3^2\cdot17\cdot31$	$53\cdot179$	$2^4\cdot593$	$3\cdot3163$	948
949	$2\cdot5\cdot13\cdot73$	**9491**	$2^2\cdot3\cdot7\cdot113$	$11\cdot863$	$2\cdot47\cdot101$	$3^2\cdot5\cdot211$	$2^3\cdot1187$	**9497**	$2\cdot3\cdot1583$	$7\cdot23\cdot59$	949

Fig. 12.1 The prime factors of n in the range $9000 \leq n \leq 9499$. The number of distinct prime factors in this range is 1260; the corresponding theoretical expectation equals 1237 ± 32. The number of prime factors, including multiple occurrences, is 1650, compared to a theoretical expectation of 1632 ± 31

$$\overline{\omega}(n) \approx \sum_{p_i<n} \frac{1}{p_i} \approx \ln(\ln n) + 0.2614, \tag{12.22}$$

as before [see (12.9) and (12.19)]. The overall variance becomes, with (12.21),

$$\sigma_\omega^2 \approx \overline{\omega}(n) - \sum_{p_i=2}^{\infty} \frac{1}{p_i^2}, \tag{12.23}$$

where we have extended the sum out to infinity because it converges quite rapidly.

The numerical value of the sum can be obtained most efficiently with the help of Riemann's zetafunction, expanded as in (12.16). This yields

$$\begin{aligned}\sum_{p_i=2}^{\infty} \frac{1}{p_i^2} &= \ln\zeta(2) - \frac{1}{2}\ln\zeta(4) - \ldots \\ &= \sum_{m=1}^{\infty} \frac{\mu(m)}{m} \ln\zeta(2m) \approx 0.452248, \end{aligned}\tag{12.24}$$

where $\mu(m)$ is again the Möbius function (see Chap. 21).

Thus,

$$\sigma_\omega^2 \approx \overline{\omega}(n) - 0.45 \tag{12.25}$$

and, because $\sigma_\omega^2 \approx \overline{\omega}$, we expect ω to be approximately *Poisson* distributed [12.3]. Of course, each number has at least *one* prime factor (itself, if it is prime), so that the Poisson distribution must be shifted by 1:

$$\mathrm{Prob}\{\omega(n) = k\} \approx \frac{(\overline{\omega}-1)^{k-1}}{(k-1)!} \mathrm{e}^{-\overline{\omega}+1}, \quad k = 1, 2, \ldots, \overline{\omega} > 1, \tag{12.26}$$

with $\overline{\omega}$ from (12.22).

The mode (most probable value) of this distribution occurs for

$$\check{k} = \lfloor \overline{\omega} \rfloor + 1, \tag{12.27}$$

where $\check{k}$ is read "kay check". Although intended for large n, (12.27) seems to work very well even for small n. Equation (12.27) predicts that the most probable number $\check{k}$ of different prime factors of n is as follows:

$$\begin{aligned}
\check{k} &= 1 \text{ for} & n &< 9 \\
\check{k} &= 2 \text{ for} & 9 \le n &< 296 \\
\check{k} &= 3 \text{ for} & 296 \le n &< 5 \cdot 10^6 \\
\check{k} &= 4 \text{ for} & 5 \cdot 10^6 \le n &< 2 \cdot 10^{18} \\
\check{k} &= 5 \text{ for} & 2 \cdot 10^{18} \le n &< 4 \cdot 10^{49} \\
\check{k} &= 6 \text{ for} & 4 \cdot 10^{49} \le n &< 8 \cdot 10^{134} \text{ etc.}
\end{aligned}$$

Thus, up to almost 10^{135} the most likely number of different prime factors is 6 or less!

According to (12.26), the probability that n has exactly one prime factor, i.e., that n is either a prime or a prime power, equals about $2/\ln n$. This value is somewhat larger than the one we would expect from the distribution of primes. But then, we should not expect the Poisson distribution for $\omega(n)$ to be exact. For example, σ_ω^2 should equal $\overline{\omega}-1$ for the shifted Poisson distribution and not $\overline{\omega}-0.45$ as in (12.23) and (12.24).

12.3 The Number of Prime Divisors

Apart from the "little" $\overline{\omega}(n)$ we need a "big" $\Omega(n)$, the number of prime divisors of n, counted with multiplicity. For

$$n = \prod_{p_i|n} p_i^{e_i}, \tag{12.28}$$

we have the definition

$$\Omega(n) := \sum_{p_i|n} e_i. \tag{12.29}$$

The divisor function $\Omega(n)$ is *completely* additive, i.e.,

$$\Omega(mn) = \Omega(m) + \Omega(n), \tag{12.30}$$

whether m and n are coprime or not.

To estimate an average value of $\Omega(n)$, we convert the sum appearing in its definition into a sum over all primes up to n:

$$\Omega(n) \approx \sum_{p_i \le n} e_i \frac{1}{p_i^{e_i}} \left(1 - \frac{1}{p_i}\right), \tag{12.31}$$

recognizing that the probability that p_i occurs e_i times equals $(1-1/p_i)/p_i^{e_i}$. Averaging over these values of e_i yields

$$\overline{\Omega}(n) \approx \sum_{p_i \le n} \frac{1}{p_i - 1}. \tag{12.32}$$

Note the closeness of our estimates of $\overline{\Omega}(n)$ and $\overline{\omega}(n)$ according to (12.3)! The difference (which some friends did not even think converged) is given by

$$\overline{\Omega}(n) - \overline{\omega}(n) \approx \sum_{p_i \le n} \frac{1}{p_i(p_i - 1)}, \tag{12.33}$$

in agreement with a result by Kronecker. (This sum is upperbounded by the sum over all integers out to infinity, which equals 1.)

Since the sum does not only converge, but converges quite rapidly, we will only bother about its value taken out to infinity. First we write

$$\sum_{p_i=2}^{\infty} \frac{1}{p_i(p_i-1)} = \sum_{p_i=2}^{\infty} \left(\frac{1}{p_i^2} + \frac{1}{p_i^3} + \frac{1}{p_i^4} + \dots \right) \tag{12.34}$$

and then introduce the zetafunction again, making use of (12.16). This yields

$$\begin{aligned} \sum_{p_i=2}^{\infty} \frac{1}{p_i(p_i-1)} &= \ln\zeta(2) + \ln\zeta(3) \\ &+ \frac{1}{2}\ln\zeta(4) + \ln\zeta(5) \\ &+ \frac{1}{6}\ln\zeta(6) + \dots \approx 0.77317 \quad \text{or} \end{aligned} \tag{12.35}$$

$$\overline{\Omega}(n) \approx \overline{\omega}(n) + 0.77317 \approx \ln(\ln n) + 1.0346. \tag{12.36}$$

What do Abramowitz and Stegun have to say? In their table of prime factors for n in the range 9000 to 9499 [12.2, p. 862], I counted a total of 1650 prime factors, including multiplicity, yielding $\overline{\Omega} = 3.30$. Our theoretical value $\ln(\ln 9250) + 1.0346 \approx 3.25$, which is as similar as could be expected.

Incidentally, sums taken over all primes, with primes appearing in the denominator as in (12.13) and (12.34), need not always lead to irrational results. A noteworthy counterexample (from an entire family of like-fashioned expressions) is

$$\prod_{p=2}^{\infty} \frac{p^2+1}{p^2-1} = \frac{5}{2}. \tag{12.37}$$

This seems preposterous, but a quick numerical check indicates that the product certainly could not deviate much from $5/2$, and in fact, the infinite product *does* equal $5/2$. This is actually not too difficult to see, because

$$\prod \frac{p^2+1}{p^2-1} = \prod \frac{p^4-1}{(p^2-1)^2} = \prod \frac{1-\dfrac{1}{p^4}}{\left(1-\dfrac{1}{p^2}\right)^2},$$

or, expanding into geometric series:

$$\prod \frac{p^2+1}{p^2-1} = \frac{\prod \left(1 + \dfrac{1}{p^2} + \dfrac{1}{p^4} + \dots\right)^2}{\prod \left(1 + \dfrac{1}{p^4} + \dfrac{1}{p^8} + \dots\right)} = \frac{\left(\sum\limits_{n=1}^{\infty} \dfrac{1}{n^2}\right)^2}{\sum\limits_{n=1}^{\infty} \dfrac{1}{n^4}}. \tag{12.38}$$

We have encountered the sum in the numerator several times before (Chaps. 4, 8), and found it to equal $\pi^2/6$. The sum in the denominator equals $\zeta(4) = \pi^4/90$, and if we had not heard of the zetafunction, we could find out by calculating a certain definite integral over the Fourier series

$$\sin x - \sin 3x + \sin 5x - \ldots.$$

(The reader may want to try this.) The result is

$$\prod_{p=2}^{\infty} \frac{p^2+1}{p^2-1} = \frac{\frac{\pi^4}{36}}{\frac{\pi^4}{90}} = \frac{5}{2}.$$

Consideration of this product also leads to some rather unexpected relations for $\Omega(n)$. Expanding

$$\prod \frac{p^2+1}{p^2-1} = \frac{\prod\left(1+\frac{1}{p^2}+\frac{1}{p^4}+\ldots\right)}{\prod\left(1-\frac{1}{p^2}+\frac{1}{p^4}-\ldots\right)} \tag{12.39}$$

and multiplying out, one obtains, in the denominator, a sum of each reciprocal square $1/n^2$ exactly once, with a sign that depends on the parity (odd or even) of the total numbers of prime factors of n. Thus, with (12.37), remembering that $\Omega(1) = 0$:

$$\sum_{n=1}^{\infty} \frac{(-1)^{\Omega(n)}}{n^2} = \frac{\pi^2}{15}, \tag{12.40}$$

or

$$\sum_{\Omega(n)\,\mathrm{odd}} \frac{1}{n^2} = \frac{\pi^2}{20}, \tag{12.41}$$

two noteworthy results.

Similar procedures give the equally remarkable

$$\sum_{n=1}^{\infty} \frac{(\pm 1)^{\Omega(n)}}{n^2} 2^{\omega(n)} = \left(\frac{5}{2}\right)^{\pm 1}, \tag{12.42}$$

or

$$\sum_{\Omega(n)\,\mathrm{odd}} \frac{2^{\omega(n)}}{n^2} = \frac{21}{20}. \tag{12.43}$$

12.4 The Harmonic Mean of $\Omega(n)$

In order to estimate, as we would like to, the geometric mean of the prime factors of n, we need the *harmonic* mean of $\Omega(n)$. If we designate geometric means by a tilde, then the desired mean is given by

$$\tilde{p}(n) := n^{1/\Omega(n)}. \tag{12.44}$$

Now if we average over several (similar) values of n, we are led to the harmonic mean of $\Omega(n)$, which we identify by a "hat":

$$\hat{\Omega}(n) := \left(\overline{1/\Omega(n)}\right)^{-1}. \tag{12.45}$$

With this notation, we have

$$\tilde{p}(n) \approx n^{1/\hat{\Omega}(n)}. \tag{12.46}$$

Of course, like any harmonic mean of a fluctuating quantity, $\hat{\Omega}(n)$ is smaller than the previously computed arithmetic mean $\overline{\Omega}(n) \approx \ln(\ln n) + 1.035$. By how much? to answer this question, we have to find out about the *distribution* of $\Omega(n)$. Reverting to our earlier "unaveraged" estimate of $\Omega(n)$:

$$\Omega(n) \approx \sum_{p_i \le n} e_i \frac{1}{p_i^{e_i}} \left(1 - \frac{1}{p_i}\right), \tag{12.47}$$

we recognize geometric distributions[1] in the exponents e_i. The mean value m_i for each term of the sum is

$$m_i = \frac{1}{p_i - 1}, \tag{12.48}$$

a result we used before in estimating $\overline{\Omega}(n)$.

Now we also want the *variance* σ_i^2 of each term, which for a geometric distribution is given in terms of the mean m_i by the following well-known formula:[2]

$$\sigma_i^2 = m_i + m_i^2. \tag{12.49}$$

[1] Physicists call a related distribution "Bose-Einstein" in honor of Bose, the Indian scientist who discovered its significance for photons and other "bosons", and Einstein, who publicized it when people would not believe it.

[2] This formula played a role in physics that can hardly be overestimated. According to Maxwell's equation, the intensity fluctuations σ_i^2 in "black-body" radiation should equal the squared intensity m_i^2. It was Einstein who discovered, from deep considerations of entropy, that the actual fluctuations exceeded m_i^2 by m_i, recognizing the additional term m_i as stemming from a non-Maxwellian "granularity" of the field. This observation led him to the *photon* concept for electromagnetic radiation on much more persuasive grounds than Planck's inherently contradictory discretization of the energies of harmonic oscillators. As a result, Einstein believed in the reality of the photons from 1905 on (and he received his Nobel prize in physics for this work and not for his theory of relativity), while Planck continued to doubt the meaningfulness of his "ad hoc" trick.

By summing over the index i, assuming independence of the p_i, we obtain the variance of $\Omega(n)$:

$$\sigma_\Omega^2 = \overline{\Omega}(n) + \sum_{p_i=2} \frac{1}{(p_i-1)^2}. \tag{12.50}$$

Using the expansion (12.16) again, we can write the sum here as

$$\sum_{p_i}\left(\frac{1}{p_i^2}+\frac{2}{p_i^3}+\frac{3}{p_i^4}+\dots\right) = \ln\zeta(2)+2\ln\zeta(3)$$
$$+\frac{5}{2}\ln\zeta(4)+\dots\approx 1.3751. \tag{12.51}$$

Again, $\sigma_\Omega^2 \approx \overline{\Omega}$, and we also expect a shifted Poisson distribution for Ω:

$$\mathrm{Prob}\{\Omega(n)=k\} \approx \frac{(\overline{\Omega}-1)^{k-1}}{(k-1)!}\,\mathrm{e}^{-\overline{\Omega}+1}\ , \quad k=1,2,\dots,\ \overline{\Omega}>1\ , \tag{12.52}$$

with $\overline{\Omega}$ from (12.36).

This theoretical distribution is shown by dots in Fig. 12.2 for $\overline{\Omega}=3.25$ $(n\approx 9500)$. The shaded bars are from actual prime factor counts between $n=9000$ and 9499. The agreement is remarkably good.

For the shifted Poisson distribution, the harmonic mean $\hat{\Omega}$ is easily evaluated:

$$\hat{\Omega} = \frac{\overline{\Omega}-1}{1-\mathrm{e}^{-\overline{\Omega}+1}}, \quad \text{or} \tag{12.53}$$

$$\hat{\Omega} \approx \frac{\ln(1.035\ln n)}{1-1/1.035\ln n}. \tag{12.54}$$

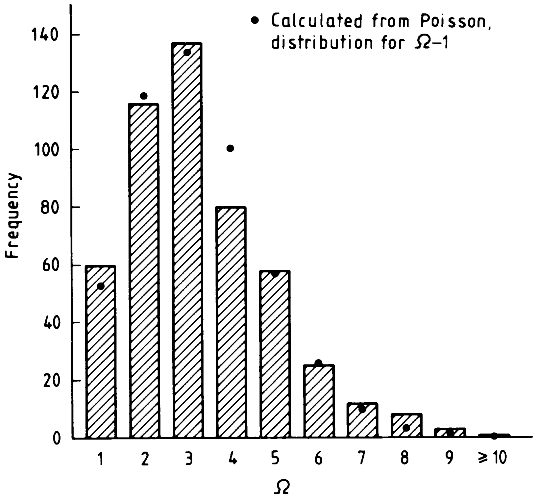


Fig. 12.2 The distribution of the number of prime factors (*bars*) in the interval $9000 \le n \le 9499$ and the Poisson distribution (*dots*) for the theoretical mean

For $n = 9250$, we obtain $\hat{\Omega} \approx 2.51$. The "experimental" value for the range $n = 9000$ to 9499 is $\hat{\Omega} = 2.47$ – as close as one can hope.

The geometric mean of the prime factors, calculated with the theoretical value of $\hat{\Omega}$, becomes

$$\tilde{p}(9250) \approx 38,$$

while the actual value in the range $n = 9000$ to 9499 is $\tilde{p} = 40$.

For $n = 10^{50}$, a range of interest for public-key encryption, $\hat{\Omega} \approx 4.8$, and the geometric mean $\tilde{p} \approx 2.4 \cdot 10^{10}$ – 40 orders of magnitude smaller than n.

12.5 Medians and Percentiles of $\boldsymbol{\Omega(n)}$

With (12.36), the probability that the integer n equals a prime that divides N is given approximately by

$$w(n = p|N) = \frac{1}{n \ln n (\ln(\ln N) + 1.035)}. \tag{12.55}$$

Thus, the cumulative distribution for a prime divisor of N to be smaller than n is approximated by

$$W(n;N) = \frac{\ln(\ln n) + 1.035}{\ln(\ln N) + 1.035}. \tag{12.56}$$

From this expression the median value $n_{0.5}$ follows directly:

$$n_{0.5} = \mathrm{e}^{\sqrt{\ln N/2.81}}. \tag{12.57}$$

Example: $N = 9250$, $n_{0.5} = 6$. Thus, the primes 2, 3, and 5 should account for roughly half the prime factors around $N = 9250$. The actual count in the interval 9000 to 9499 is as follows (with the theoretical value, $500/(p_i - 1)$, in parenthesis):

$$\begin{aligned} p_i &= 2 : 500 \text{ times} \quad (500) \\ p_i &= 3 : 250 \text{ times} \quad (250) \\ p_i &= 5 : 126 \text{ times} \quad (125) \end{aligned}$$

Thus, the total number of occurrences of 2, 3, and 5 is 876 times, or 53% of the total of 1650 prime factors in that interval – in very good agreement with our theoretical prediction.

For $N = 10^{50}$, $n_{0.5} = 600$, a remarkably small value.

The above distribution formula gives the following percentile values n_f, defined by $W(n_f) = f$:

$$n_f(N) = \exp\left[(\ln N)^f 2.81^{f-1}\right]. \tag{12.58}$$

The lower-quartile value $n_{0.25}$ (for $N = 9250$) becomes 2.2, which compares well with the count of 30% (500 out of 1650) for the factor 2 in the interval 9000 to 9499. In fact, according to (12.56), 29% of the prime factors should be below 2.5.

The theoretical upper-quartile value $n_{0.75} = 57.8$ is in very good agreement with the count of 75%. (1231 out of 1650) prime factors up to and including 59. But the median of the largest prime factor of N is about $N^{0.6}$ (!).

12.6 Implications for Public-Key Encryption

For $N = 10^{50}$, the theoretical lower- and upper-quartile values for the prime factors are 4.5 and $6 \cdot 10^{11}$, respectively. Thus, in three out of four cases of integers around 10^{50}, one will encounter prime factors not exceeding $6 \cdot 10^{11}$. If one assumes that *rapid* factoring of such integers is no problem, then 75% of such large, *randomly* selected integers can be easily factored.

This conclusion is in stark contrast to the (correct) assertion that sufficiently large integers *constructed so as to contain only two very large prime factors* cannot be easily factored.

Additional results on large prime factors in a given interval can be found in [12.4]. As an introduction to the art of generating large primes, see [12.5].