MATHEMATICS OF COMPUTATION, VOLUME 28, NUMBER 125, JANUARY, 1974

The Distribution of Small Gaps Between Successive Primes

By Richard P. Brent

Abstract. For $r \geqq 1$ and large N, a well-known conjecture of Hardy and Littlewood implies that the number of primes $p \leqq N$ such that $p + 2r$ is the least prime greater than p is asymptotic to

$$\int_2^N \left(\sum_{k=1}^{r} \frac{A_{r,k}}{(\log x)^{k+1}} \right) dx,$$

where the $A_{r,k}$ are certain constants. We describe a method for computing these constants. Related constants are given to 10D for $r = 1(1)40$, and some empirical evidence supporting the conjecture is mentioned.

1. Introduction. Let r be a fixed positive integer and N a large integer. Hardy and Littlewood [4] conjectured that the number of primes $p \leqq N$ such that $p + 2r$ is also prime is

$$P_N(r) \sim A_{r,1} \int_2^N \frac{dx}{(\log x)^2}, \tag{1}$$

where

$$A_{r,1} = 2c_2 \prod_{q|r} \frac{q-1}{q-2}, \tag{2}$$

$$c_2 = \prod_q \frac{1 - 2/q}{(1 - 1/q)^2} = 0.66016 \cdots \tag{3}$$

is the "twin-prime" constant, and q runs over the odd primes (this convention is adopted throughout). Conjecture (1) has been substantiated empirically and extended by several authors, e.g. [1], [3], [5] and [6].

In this paper, we study the number $Q_N(r)$ of primes $p \leqq N$ such that $p + 2r$ is the *first* prime after p (so $p + 1, p + 2, \cdots, p + 2r - 1$ are composite). In Section 2, we use the principle of inclusion and exclusion to deduce from a conjecture of Hardy and Littlewood that

$$Q_N(r) \sim \int_2^N \left(\sum_{k=1}^{r} \frac{A_{r,k}}{(\log x)^{k+1}} \right) dx, \tag{4}$$

where the constants $A_{r,k}$ are defined by (8) below. Previously, $Q_N(r)$ seems to have been studied only for $r \leqq 4$, although the magnitude of the first prime p followed by $2r - 1$ consecutive composite numbers has been investigated (see [2] and the references given there).

Received February 13, 1973.

AMS (MOS) subject classifications (1970). Primary 10-04, 10A20, 10A25, 10A40, 65A05.

Key words and phrases. Prime, distribution of primes, Hardy-Littlewood conjecture, prime gap, twin primes.

In Section 3 we discuss the computation of the constants $A_{r,k}$, and related constants $B_{r,k}$ (see Eq. (12)) are given for $r \leqq 40$ in Table 1. Some empirical evidence for conjecture (4) is given in Section 4.

2. Derivation of the Conjecture. If $0 < m_1 < m_2 < \cdots < m_s$, Hardy and Littlewood [4] conjecture that the number of primes $p \leqq N$ such that $p + 2m_i$ is prime for $i = 1, \cdots, s$ is

$$P_N(m_1, \cdots, m_s) \sim C(m_1, \cdots, m_s) \int_2^N \frac{dx}{(\log x)^{s+1}}, \tag{5}$$

where

$$C(m_1, \cdots, m_s) = 2^s \prod_q (1 - 1/q)^{-(s+1)}(1 - w_{\mathrm{m}}(q)/q) \tag{6}$$

and $w_{\mathrm{m}}(q)$ is the number of distinct residues of $0, m_1, \cdots, m_s$ modulo the odd prime q. (Conjecture (1) is just the special case $s = 1$, $m_1 = r$ of (5).)

From the principle of inclusion and exclusion (see, e.g. [7]),

$$Q_N(r) = \sum_{s=0}^{r-1} (-1)^s \sum_{0<m_1<\cdots<m_s<r} P_N(m_1, \cdots, m_s, r). \tag{7}$$

From (5), this gives the conjecture (4) if

$$A_{r,k} = (-1)^{k+1} \sum_{0<m_1<\cdots<m_{k-1}<r} C(m_1, \cdots, m_{k-1}, r). \tag{8}$$

3. Computation of the Constants $A_{r,k}$. Since $w_{\mathrm{m}}(q) = s + 1$ for all $q > m_s$, $C(m_1, \cdots, m_s)$ may easily be found if the Hardy-Littlewood constants

$$c_k = \prod_{q>k} \frac{1 - k/q}{(1 - 1/q)^k} \tag{9}$$

are known. These may be evaluated by the method of Wrench [10] and Mayoh [8].

Thus, the computation of $A_{r,k}$ from (8) appears to be straightforward. However, the sum in (8) involves $\binom{r-1}{k-1}$ terms, so the evaluation of $A_{r,1}, \cdots, A_{r,r}$ in the obvious way requires the evaluation of 2^{r-1} terms $C(m_1, \cdots, m_{k-1}, r)$. We shall show how this may be reduced to the evaluation of $O(2^{2r/3})$ terms.

It is easy to evaluate $A_{r,k}$ if the integer

$$T_{r,k} = \sum_{0<m_1<\cdots<m_{k-1}<r} \prod_{q \leqq r+1} (q - w_{\mathrm{m},r}(q)) \tag{10}$$

can be evaluated, where $w_{\mathrm{m},r}(q)$ is the number of distinct residues of $0, m_1, \cdots, m_{k-1}$, r modulo the odd prime q. By omitting all the terms with $w_{\mathrm{m},r}(3) = 3$, and using symmetry if $3 | r$, we find that

$$T_{r,k} = K_r \sum_{0<m_1<\cdots<m_{k-1}<r;\, m_i \not\equiv r' \pmod 3} \prod_{5 \leqq q \leqq r+1} (q - w_{\mathrm{m},r}(q)), \tag{11}$$

where

$$K_r = \begin{Bmatrix} 1 & \text{if } r \not\equiv 0 \pmod 3 \\ 2 & \text{if } r \equiv 0 \pmod 3 \end{Bmatrix},$$

and

$$r' = \begin{Bmatrix} 1 & \text{if } r \equiv 2 \pmod 3 \\ 2 & \text{if } r \not\equiv 2 \pmod 3 \end{Bmatrix}.$$

Thus, we may compute $T_{r,k}$ by summing

$$\begin{pmatrix} \lfloor 2(r-1)/3 \rfloor \\ k-1 \end{pmatrix} \quad \text{instead of} \quad \begin{pmatrix} r-1 \\ k-1 \end{pmatrix}$$

terms. An additional factor of 2 can be saved by symmetry if r is not divisible by 3.

It is interesting to note that $T_{r,k} = 0$ if k is so large that a cluster of $k - 1$ primes cannot lie between two large primes p and $p + 2r$ (see [9]). Surprisingly, the least such k is not a monotonic function of r, for $T_{10,6} \neq 0$ and $T_{11,6} = 0$.

The constants $A_{r,k}$ for $r \leqq 40$ were found by computing the $T_{r,k}$ as suggested above. Since the $A_{r,k}$ vary greatly in size, it is more convenient to work with

$$(12) \qquad B_{r,k} = \begin{Bmatrix} A_{r,1} & \text{if } k = 1 \\ -A_{r,k}/A_{r,k-1} & \text{if } k > 1 \text{ and } A_{r,k-1} \neq 0 \\ 0 & \text{if } k > 1 \text{ and } A_{r,k-1} = 0 \end{Bmatrix}.$$

Thus, the integrand in (4) is $B_{r,1}z^2(1 - B_{r,2}z(\cdots(1 - B_{r,r}z)\cdots))$, where $z = 1/(\log x)$.

Table 1 gives the constants $B_{r,k}$, believed to be correctly rounded to 10D, for $1 \leqq k \leqq r \leqq 40$. Values which are omitted are zero.

4. Empirical Evidence for the Conjecture. A "gap of length $2r$" is defined to be an interval $(p, p + 2r)$ where p and $p + 2r$ are successive primes. The number of gaps of length 2, 4, $\cdots$, 80 in an interval (M, N) was compared with the number predicted by conjecture (4) for various M and N in the range $(10^6, 10^{16})$. (For example, $M = 10^6, 10^7, \cdots, 10^{15}$ and $N = M + 10^6$.) The actual and predicted gap distributions agreed closely in all the intervals considered. Detailed results have been deposited in the UMT file of this journal.

As a typical example, results for the interval $(10^6, 10^9)$ are given in Table 2. For $r = 1, 2, \cdots, 40$, the table gives the actual number of gaps of length $2r$, and the number predicted from

$$(13) \qquad \int_{10^6}^{10^9} \left(\sum_{k=1}^{r} \frac{A_{r,k}}{(\log x)^{k+1}} \right) dx.$$

In $(10^6, 10^9)$, there are 50769035 gaps, and

$$(14) \qquad \int_{10^6}^{10^9} \frac{dx}{\log x} \simeq 50770607.4.$$

TABLE 1

The constants $B_{r,k}$ for r = 1(1)40 (values omitted are zero)

r	k	$B_{r,k}$
1	1	1.3203236317
2	1	1.3203236317
3	1	2.6406472634
	2	2.1648090870
4	1	1.3203236317
	2	4.3296181741
	3	0.7261756149
5	1	1.7604315089
	2	4.8708204458
	3	0.9682341532
6	1	2.6406472634
	2	7.5768318046
	3	2.0747874712
	4	0.4881403766
7	1	1.5843883580
	2	9.0200378627
	3	2.7110556291
	4	0.7845113196
8	1	1.3203236317
	2	10.8240454352
	3	3.4856429517
	4	1.2203509416
	5	0.2845598501
9	1	2.6406472634
	2	12.4476522505
	3	4.2097137097
	4	1.6108632428
	5	0.4656433910
10	1	1.7604315089
	2	15.8301664490
	3	5.7597518859
	4	2.5248640170
	5	1.0244154603
	6	0.2600013387
11	1	1.4670262574
	2	15.3430844044
	3	5.5503309283
	4	2.3653954168
	5	0.8429475787
12	1	2.6406472634
	2	18.4008772398
	3	7.0705502451
	4	3.3815938634
	5	1.6080940365
	6	0.6066697903
13	1	1.4403530528
	2	20.7370670462
	3	8.1550214623
	4	4.0633477298
	5	2.1093164691
	6	1.0254787648
	7	0.3885721249
14	1	1.5843883580
	2	20.8588375574
	3	8.1825131838
	4	4.0589157778
	5	2.0866554914
	6	0.9897331926
	7	0.3495093187
15	1	3.5208630178
	2	24.4961678255
	3	9.9210862934
	4	5.1517209912
	5	2.8451655917
	6	1.5366370101
	7	0.7430478393
	8	0.2639864981
16	1	1.3203236317
	2	27.4930754054
	3	11.3952967179
	4	6.1214210653
	5	3.5632266426

TABLE 1 (*continued*)

r	k	$B_{r,k}$
16	6	2.1006626920
	7	1.1950083565
	8	0.6157145513
	9	0.2444874923
17	1	1.4083452071
	2	27.7245360198
	3	11.4984742518
	4	6.1729997100
	5	3.5819776177
	6	2.0946244877
	7	1.1705496702
	8	0.5824231949
	9	0.2198886873
18	1	2.6406472634
	2	30.0425246784
	3	12.6122048727
	4	6.8853036007
	5	4.0940729018
	6	2.4873794831
	7	1.4840852989
	8	0.8373344927
	9	0.4234899981
	10	0.1658772706
19	1	1.3979897277
	2	31.9331890432
	3	13.4561073779
	4	7.3677263519
	5	4.3825468884
	6	2.6448059150
	7	1.5377429643
	8	0.7999341181
	9	0.3061670870
20	1	1.7604315089
	2	33.0230029036
	3	14.0212494141
	4	7.7552428532
	5	4.6798483508
	6	2.8876803489
	7	1.7450477061
	8	0.9833612019
	9	0.4731308341
	10	0.1513663376
21	1	3.1687767161
	2	36.3539740286
	3	15.6509822856
	4	8.8131644369
	5	5.4466534309
	6	3.4737488749
	7	2.2032572055
	8	1.3409356167
	9	0.7429532570
	10	0.3357930107
	11	0.0896365942
22	1	1.4670262574
	2	38.5221012010
	3	16.7436006742
	4	9.5502460845
	5	6.0071544989
	6	3.9277664665
	7	2.5837231927
	8	1.6635481779
	9	1.0125375871
	10	0.5464248161
	11	0.2174507624
23	1	1.3831961856
	2	39.9075175164
	3	17.3771078681
	4	9.9252611046
	5	6.2469956507
	6	4.0820403447
	7	2.6778338225
	8	1.7129792116
	9	1.0284516322
	10	0.5388263843
	11	0.1992045198
24	1	2.6406472634
	2	41.6087903720
	3	18.2211462744
	4	10.4857133121
	5	6.6675963762
	6	4.4204962521
	7	2.9630734223

TABLE 1 (*continued*)

r	k	$B_{r,k}$	r	k	$B_{r,k}$
24	8	1.9613859814	27	13	0.1729604958
	9	1.2496708257			
	10	0.7375923983	28	1	1.5843883580
	11	0.3737970930		2	49.7793339545
	12	0.1307666522		3	22.2516291007
				4	13.1317161092
25	1	1.7604315089		5	8.6175937443
	2	44.4259414830		6	5.9494259004
	3	19.6478400115		7	4.2078174539
	4	11.4435289614		8	2.9992575087
	5	7.3859456123		9	2.1274078880
	6	4.9901941363		10	1.4833716834
	7	3.4285605747		11	1.0014969256
	8	2.3470128849		12	0.6388395411
	9	1.5694239568		13	0.3646341816
	10	0.9991223480		14	0.1558821625
	11	0.5797444199			
	12	0.2779751671	29	1	1.3692245069
	13	0.0773583678		2	49.7351293969
				3	22.2010534736
26	1	1.4403530528		4	13.0749059856
	2	45.5513844929		5	8.5538846663
	3	20.1021552452		6	5.8779731882
	4	11.6727241289		7	4.1277954713
	5	7.5016123702		8	2.9100995690
	6	5.0370857637		9	2.0292305179
	7	3.4292205066		10	1.3777494229
	8	2.3145847402		11	0.8927963590
	9	1.5125764268		12	0.5361400447
	10	0.9245068977		13	0.2829661211
	11	0.4938397071		14	0.1112806842
	12	0.1894277144			
			30	1	3.5208630178
27	1	2.6406472634		2	53.4568780418
	2	47.3706617010		3	24.0119086856
	3	21.0944825152		4	14.2465572306
	4	12.3891416646		5	9.4040165228
	5	8.0793108307		6	6.5333739503
	6	5.5302801907		7	4.6514933053
	7	3.8641962264		8	3.3376210785
	8	2.7052770199		9	2.3821089677
	9	1.8661755593		10	1.6694186696
	10	1.2435276562		11	1.1314474006
	11	0.7763137851		12	0.7264160913
	12	0.4271256883		13	0.4281428785

TABLE 1 (*continued*)

r	k	$B_{r,k}$
30	14	0.2187972637
	15	0.0799007719
31	1	1.3658520328
	2	57.0564427867
	3	25.7883873035
	4	15.4134377102
	5	10.2643380469
	6	7.2079466964
	7	5.2002014685
	8	3.7939887877
	9	2.7661319314
	10	1.9930796451
	11	1.4012127006
	12	0.9442903513
	13	0.5922053004
	14	0.3249621929
	15	0.1294626296
32	1	1.3203236317
	2	56.0765841791
	3	25.2955005318
	4	15.0822774906
	5	10.0134861920
	6	7.0047474948
	7	5.0282652653
	8	3.6439356759
	9	2.6321755336
	10	1.8716152565
	11	1.2903211212
	12	0.8438892387
	13	0.5049928034
	14	0.2585874366
	15	0.0987505835
33	1	2.9340525149
	2	59.3116309805
	3	26.8938829613
	4	16.1336371765
	5	10.7902210471
	6	7.6158335959
	7	5.5282304118
	8	4.0639911464
	9	2.9918723175
	10	2.1839746041
33	11	1.5642788142
	12	1.0853435243
	13	0.7167409804
	14	0.4387634196
	15	0.2379412131
	16	0.1004264844
34	1	1.4083452071
	2	61.9919072207
	3	28.2142671194
	4	16.9990534791
	5	11.4265453826
	6	8.1130495630
	7	5.9309308916
	8	4.3972048990
	9	3.2707071778
	10	2.4178440362
	11	1.7589282962
	12	1.2438306194
	13	0.8398533041
	14	0.5252330428
	15	0.2854048569
	16	0.1108625403
35	1	2.1125178107
	2	62.7958185378
	3	28.6092743854
	4	17.2585563381
	5	11.6189657616
	6	8.2658057455
	7	6.0577632487
	8	4.5059818778
	9	3.3663563583
	10	2.5036711044
	11	1.8373213753
	12	1.3167529951
	13	0.9093775214
	14	0.5943929626
	15	0.3596437584
	16	0.1989064249
	17	0.0993281933
36	1	2.6406472634
	2	66.2111617092
	3	30.2766540599

TABLE 1 (*continued*)

r	k	$B_{r,k}$	r	k	$B_{r,k}$
36	4	18.3409094425	38	13	1.3131389130
	5	12.4068962334		14	0.9365319820
	6	8.8754487506		15	0.6404963934
	7	6.5470947651		16	0.4138510165
	8	4.9080453851		17	0.2463430835
	9	3.7017107962		18	0.1194900846
	10	2.7858957162			
	11	2.0757081630	39	1	2.8807061055
	12	1.5176476093		2	70.6997065950
	13	1.0766924356		3	32.5107597334
	14	0.7293724865		4	19.8218855636
	15	0.4597244627		5	13.5097785123
	16	0.2563308030		6	9.7499560283
	17	0.1085109086		7	7.2677697292
				8	5.5171282455
37	1	1.3580471640		9	4.2251859714
	2	66.9782027174		10	3.2405820808
	3	30.6782419268		11	2.4727472369
	4	18.6201628208		12	1.8643285785
	5	12.6242292479		13	1.3774162158
	6	9.0546344932		14	0.9861397380
	7	6.6995839806		15	0.6724449317
	8	5.0398868987		16	0.4235592577
	9	3.8161157582		17	0.2304613130
	10	2.8843429973		18	0.0876237466
	11	2.1584604606			
	12	1.5839990281	40	1	1.7604315089
	13	1.1250993393		2	73.6107429917
	14	0.7576010703		3	33.9328292881
	15	0.4653031637		4	20.7460643325
	16	0.2382708022		5	14.1837340795
	17	0.0741063531		6	10.2726763283
				7	7.6886927466
38	1	1.3979897277		8	5.8644776442
	2	69.4720528395		9	4.5166169192
	3	31.9135466028		10	3.4879437671
	4	19.4347144701		11	2.6845126281
	5	13.2275487026		12	2.0469957441
	6	9.5306429383		13	1.5364440188
	7	7.0904008520		14	1.1267004833
	8	5.3697954268		15	0.8001365673
	9	4.1005791410		16	0.5450067101
	10	3.1340488773		17	0.3529451420
	11	2.3813915965		18	0.2139492394
	12	1.7866244753		19	0.1062961616

By summing (13) over $r = 1, \cdots, 40$ and subtracting from (14), the predicted number of gaps of length greater than 80 is 473076.4, and the actual number is 473186.

Table 2 shows that (13) predicts quite well the number of gaps of various lengths in $(10^6, 10^9)$. Although the right sides of (1) and (4) are asymptotically equal, the higher terms in (4) are important for approximating $Q_N(r)$. It is interesting to note that in $(10^6, 10^9)$ there are less gaps (observed and predicted) for $r = 31$ than for $r = 32$, although $A_{31,1} = 1.36 \cdots > A_{32,1} = 1.32 \cdots$. Thus, the higher terms in (13) are significant.

TABLE 2

Actual and predicted gap distribution in $(10^6, 10^9)$

r	Actual	Predicted	r	Actual	Predicted
1	3416337	3417060.1	21	953980	954689.0
2	3416536	3417060.1	22	389432	389057.1
3	6076242	6077407.1	23	334565	335337.0
4	2689540	2688560.2	24	577051	577893.6
5	3477688	3477436.8	25	327960	327323.5
6	4460952	4460654.7	26	245727	245799.1
7	2460332	2461360.3	27	410614	410578.1
8	1843216	1842845.7	28	211409	211469.0
9	3346123	3347229.6	29	181894	182398.0
10	1821641	1823424.2	30	371743	372007.3
11	1567507	1567220.8	31	115542	115837.8
12	2364792	2362746.8	32	118927	118681.6
13	1118410	1118419.0	33	216739	216467.5
14	1218009	1218441.9	34	88383	88116.0
15	2176077	2176130.5	35	125542	125688.7
16	683346	682871.2	36	126650	126786.7
17	718974	718118.6	37	62514	62578.8
18	1170757	1169307.2	38	55107	55325.4
19	548416	547688.6	39	105300	105390.3
20	648356	648539.8	40	53519	53578.4

Computer Centre
Australian National University
Canberra, Australia

1. P. T. Bateman & R. A. Horn, "A heuristic asymptotic formula concerning the distribution of prime numbers," *Math Comp.*, v. 16, 1962, pp. 363–367. MR **26** #6139.

2. R. P. Brent, "The first occurrence of large gaps between successive primes," *Math. Comp.*, v. 27, 1973, pp. 959–963.

3. F. Gruenberger & G. Armerding, *Statistics on the First Six Million Prime Numbers*, Paper P-2460, The RAND Corporation, Santa Monica, Calif., 1961, 145 pp. (Copy deposited in the UMT File and reviewed in *Math. Comp.*, v. 19, 1965, pp. 503–505.)

4. G. H. Hardy & J. E. Littlewood, "Some problems of 'partitio numerorum'; III: On the expression of a number as a sum of primes," *Acta Math.*, v. 44, 1923, pp. 1–70.

5. M. F. Jones, M. Lal & W. J. Blundon, "Statistics on certain large primes," *Math. Comp.*, v. 21, 1967, pp. 103–107. (Corrigenda, v. 22, 1968, pp. 474 & 911.) MR **36** #3707.

6. D. H. Lehmer, "Tables concerning the distribution of primes up to 37 millions", 1957. Copy deposited in the UMT File and reviewed in *MTAC*, v. 13, 1959, pp. 56–57.

7. C. L. Liu, *Introduction to Combinatorial Mathematics*, McGraw-Hill, New York, 1968, Chap. 4. MR **38** #3154.

8. B. H. Mayoh, "The second Goldbach conjecture revisited," *Nordisk Tidskr. Informationsbehandling (BIT)*, v. 8, 1968, pp. 128–133. MR **39** #125.

9. H. Riesel, "Primes forming arithmetic series and clusters of large primes," *Nordisk Tidskr. Informationsbehandling (BIT)*, v. 10, 1970, pp. 333–342. MR **44** #965.

10. J. W. Wrench, "Evaluation of Artin's constant and the twin-prime constant," *Math. Comp.*, v. 15, 1961, pp. 396–398. MR **23** #A1619.