COROLLARY. *Let F be a field with $m = 4$, $R = F^2$, and $8 \neq q < \infty$. Then anisotropic ternary forms over F are determined by their value sets.*

Proof. It suffices to show that $D(\langle 1, a, b\rangle) = D(\langle 1, c, d\rangle)$ implies $\langle a, b\rangle \cong \langle c, d\rangle$ for anisotropic $\langle 1, a, b\rangle$, $\langle 1, c, d\rangle$. By a cardinality argument on the value sets mod F^2 and by Theorem 8, it is clear that it is impossible for exactly one of $[-a, -b]$, $[-c, -d]$ to be Q_3.

If $[-a, -b] = [-c, -d] = Q_3$, then by Theorem 8, $-abB_1 \cup -abB_2 = -cdB_1 \cup -cdB_2$. Thus $abcd \in B_1 \cap B_2 = F^2$. But then Corollary 2.9 [9, P. 60] shows $\langle -a, -b\rangle \cong \langle -c, -d\rangle$. If $[-a, -b] = [-c, -d] = Q_i$, $i = 1$ or 2, then Theorem 8 yields $abB_j = cdB_j$, $\{i, j\} = \{1, 2\}$. But the quaternion algebra equality also implies $-a, -b, -c, -d \in B_i$. Thus $abcd \in B_1 \cap B_2 = F^2$ and again $\langle -a, -b\rangle \cong \langle -c, -d\rangle$. Finally assume, $[-a, -b] = Q_1$ and $[-c, -d] = Q_2$. Then from Theorem 8, $abB_2 = cdB_1$. But this cannot happen since $B_1 \neq B_2$. ∎

References

[1] C. M. Cordes, *The Witt group and the equivalence of fields with respect to quadratic forms*, J. Algebra 26 (1973), pp. 400–421.
[2] — *Kaplansky's radical and quadratic forms over non-real fields*, Acta Arith. 28 (1975), pp. 253–261.
[3] — *Quadratic forms over non-formally real fields with a finite number of quaternion algebras*, Pacific J. Math. 63 (1976), pp. 357–365.
[4] C. M. Cordes and J. R. Ramsey, *Quadratic forms over fields with $u = q/2 < \infty$*, Fund. Math. 99 (1978), pp. 1–10.
[5] R. Elman and T. Y. Lam, *Quadratic forms and the u-invariant. II*, Invent. Math. 21 (1973), pp. 125–137.
[6] — — *Classification theorems for quadratic forms over fields*, Comment. Math. Helvetici 49 (1974), pp. 373–381.
[7] I. Kaplansky, *Fröhlich's local quadratic forms*, J. Reine Angew. Math. 239 (1969), pp. 74–77.
[8] M. Kula, *Fields with prescribed quadratic form schemes*, preprint.
[9] T. Y. Lam, *The algebraic theory of quadratic forms*, W. A. Benjamin, Reading, Mass., 1973.
[10] O. T. O'Meara, *Introduction to quadratic forms*, Springer-Verlag, New York 1963.

MATHEMATICS DEPARTMENT
LOUISIANA STATE UNIVERSITY
Baton Rouge, Louisiana 70803

Received on 19. 11. 1979 (1184)

ACTA ARITHMETICA
XLI (1982)

Partitions into distinct small primes

by

J. RIDDELL* (Victoria, B. C., Canada)

Introduction. Throughout, p_n will denote the nth prime. In Theorem 1 we find estimates for $\sum_{k=1}^{n} p_k$ (see (3)), but because of the application of Theorem 3 later, it will be more convenient to work with the sum $y(n) = 3 + \sum_{k=4}^{n} p_k$ instead.

THEOREM 1.

$$y(n) < \tfrac{1}{2}n^2(\log n + \log\log n) \quad \textit{for } n \geqslant 4. \tag{1}$$

Given $0 \leqslant a < 1$, there exists an integer $N(a)$ such that

$$\tfrac{1}{2}n^2(\log n + a\log\log n) < y(n) \tag{2}$$

for $n \geqslant N(a)$, and $N(0) = 5$. Moreover, (2) is true if

$$a < 1 - \frac{2 - \log\left[1 - \dfrac{\log 2}{\log n}\right]}{\log\log n},$$

so that given $0 < \varepsilon \leqslant 1$, we can take $a = 1 - \varepsilon$ in (2), provided $n \geqslant n_0(\varepsilon) = N(1-\varepsilon)$.

$$\sum_{k=1}^{n} p_k < \tfrac{1}{2}n^2(\log n + \log\log n) \quad \textit{for } n \geqslant 6, \tag{3}$$

and lower bounds for this sum are given by those for $y(n)$.

COROLLARY. $\sum_{k=1}^{n} p_k \sim \frac{1}{2}n^2(\log n + \log\log n)$.

We shall see later that $N(.1) = 5$ and $N(.156) = 140$ (see Remark 1, following the proof of (2)). The inequalities (1) and (2) are used in proving

THEOREM 2. *Let $\varepsilon > 0$ and write $y(n) = y$. Then*

$$\sqrt{2y}\sqrt{\log\sqrt{2y} + (\tfrac{1}{2} - \varepsilon)\log\log\sqrt{2y}} < p_n < \sqrt{2y}\sqrt{\log\sqrt{2y} + (\tfrac{1}{2} + \varepsilon)\log\log\sqrt{2y}}$$

* Research partially supported by NSERC Grant A-5208.

provided n is sufficiently large. The right-hand inequality holds for $n \geqslant \max\{20, n_0(\varepsilon)\}$, the left for $n \geqslant n_1(\varepsilon)$.

$n_1(\varepsilon)$ will be defined when the left inequality is proved.

COROLLARY. $p_n \sim \sqrt{2y}\sqrt{\log\sqrt{2y}+\frac{1}{2}\log\log\sqrt{2y}}$.

THEOREM 3. *Let $n \geqslant 4$. Then every integer in the closed interval $[7, y(n)]$ can be partitioned into distinct primes $\leqslant p_n$.*

Through an adaptation of the proof of the right-hand inequality in Theorem 2, and the application of Theorem 3, we obtain lastly Theorem 4. The proof could be done independently of that of Theorem 2, but it is simpler to proceed as we have done.

THEOREM 4. *Let $0 < \varepsilon \leqslant 1$, $n \geqslant \max\{20, n_0(\varepsilon)+1\}$, and $y(n-1) < y \leqslant y(n)$. Then y can be partitioned into distinct primes p satisfying*

$$p < \sqrt{2y}\sqrt{\log\sqrt{2y}+(\tfrac{1}{2}+\varepsilon)\log\log\sqrt{2y}}.$$

We remark on the possibility of dropping the ε, and the impossibility of reducing the $\frac{1}{2}$, after introducing the notation

$$U(y, c) = \sqrt{2y}\sqrt{\log\sqrt{2y}+c\log\log\sqrt{2y}}, \tag{4}$$

which we shall use in the sequel. We have computed $y(n)$ and some related functions, including $U(y(n), c)$ for several values of c, for $4 \leqslant n \leqslant 100\,000$. Some of the results of these computations are mentioned here, in the introduction, and others appear later, in Remarks 1, 2, and 3; partial print-outs of the computations are available from the author.

COROLLARY TO THEOREM 4. *Given a positive integer k, there is an integer $n_2 = n_2(k)$ such that every integer $y > n_2$ can be partitioned into k or more distinct primes.*

Proof. The number of distinct primes used in representing y in the theorem exceeds $y/U(y, \frac{1}{2}+\varepsilon)$. Hence for given k we need merely take n_2 such that $n_2/U(n_2, \frac{1}{2}+\varepsilon) \geqslant k$.

The inequality in Theorem 4, without the ε, would follow from

$$p_n < U(y(n-1), \tfrac{1}{2}); \tag{5}$$

cf. (32) and the ensuing remarks. Our computations yield (5) for $105 \leqslant n \leqslant 100\,000$, with the difference between the two sides tending to increase; see Remark 3 at the end of the paper. To see that the $\frac{1}{2}$ in Theorem 4 cannot be reduced, observe that for $n \geqslant 5$, a prime at least as large as p_n is needed in any partition of $y(n)$ into distinct primes, because the only primes $\leqslant p_n$ not appearing in the sum defining $y(n)$ are 2 and 5, which are too small to be used in place of p_n. By Theorem 2, $p_n > U(y(n), \frac{1}{2}-\varepsilon)$.

First efforts to find an upper estimate for p_n in terms of $y(n)$ involved some rough work which led to the conjecture:

$$p_n < U(y(n), \tfrac{1}{2}) \tag{5a}$$

for all sufficiently large n. The attempt to prove this has led to Theorems 1, 2, and 4. Our computations show that for $n \leqslant 100\,000$,

$$p_n < U(y(n), c) \quad \text{with } c = .5,\ .6,\ .7 \quad \text{for } n \geqslant 49,\ 11,\ 6 \text{ respectively};$$

with $c = 3/2$, we can prove this inequality for all $n \geqslant 6$ (see (28) in Remark 2, following the proof of Theorem 2). On the other hand we find that for $n \leqslant 100\,000$

$$U(y(n), c) < p_n \quad \text{with } c = 0,\ .3 \quad \text{for } n \geqslant 31,\ 37374 \text{ respectively}.$$

For more on the case $c = 0$, see (29) in Remark 2, and also Table 2 following it, where some sample values of $U(y(n), c)$ for $c = 0$ and $c = .5$ are listed.

Our work with these estimates for p_n consistently indicated that the lower estimates are the more difficult fit. More precisely, in Theorem 2 it looks as though $n_1(\varepsilon)$ will be very much larger than $n_0(\varepsilon)$. For example, $n_1(\frac{1}{2})$ seems likely to exceed 10^{175} (Remark 2), while $n_0(\frac{1}{2}) = N(\frac{1}{2}) < 3\times10^{24}$ (Remark 1). In the numerical work we find that

$$U(y(n), .3) < p_n \tag{6}$$

fails for many $n \leqslant 37\,373$, and that $p_n - U(y(n), .3) < 1700$ for $n \leqslant 100\,000$, whereas the difference between the two sides of (5a) increases from $1010.4\ldots$ when $n = 5000$ to $22\,873.0\ldots$ when $n = 100\,000$ (see Table 2). In spite of all this, we know that (6) is true for nearly all n, and we do not know this about (5a).

Regarding Theorem 3, Richert [1] proved in 1950 that every integer $\geqslant 7$ can be partitioned into distinct primes. Neither in his proof nor in any other proof of this or of similar results have we seen an examination of the size of the primes that can be used in the partitions. By proving the somewhat more precise result (Theorem 3) we can see how small these primes can be chosen (Theorem 4). This consequence of such a simple partition theorem (the proof of Theorem 3 is as simple as Richert's proof, except that we keep track of the sums involved) has been overlooked heretofore, probably because the emphasis in additive prime number problems has been on seeing how few primes are needed in a partition rather than on how small they can be.

One can prove results similar to Theorem 3 using similar arguments. For example, every integer in $[10, 5+\sum_{k=5}^{n} p_k]$ $(n \geqslant 7)$ can be partitioned into distinct odd primes $\leqslant p_n$ (cf. Sierpiński [3], p. 144, Theorem 12; Theorem 11 is the above-mentioned result of Richert).

We prove Theorem 3 first; the proofs of the other theorems are of a different nature.

Proof of Theorem 3. First observe that every even integer in the interval $[-14, 14]$ can be expressed in the form

$$\pm 2 \pm 3 \pm 5 \pm 7 \pm 11$$

for an appropriate choice of signs. Adding ± 13 to these numbers gives every odd in $[-27, 27]$, and then ± 17 gives every even in $[-44, 44]$, and so on. To see that this indeed continues, define, for $n \geqslant 5$,

$$T_n = -2+3-5+7+11+\ldots+p_n,$$

and assume for some such n that every integer of the same parity as T_n in $[-T_n, T_n]$ is expressible in the form

$$\pm 2 \pm 3 \pm 5 \pm \ldots \pm p_n. \tag{7}$$

Then, by Bertrand's postulate, adding $\pm p_{n+1}$ gives every integer of the same parity as T_{n+1} in $[-T_{n+1}, T_{n+1}]$.

The details here are that the intervals

$$[-T_n - p_{n+1}, T_n - p_{n+1}] \quad \text{and} \quad [-T_n + p_{n+1}, T_n + p_{n+1}]$$

will overlap for each $n \geqslant 5$ if

$$T_n \geqslant p_{n+1} \tag{8}$$

for $n \geqslant 5$. Now, $T_5 = 14 > p_6 = 13$, and if (8) holds for a certain n, then

$$T_{n+1} = T_n + p_{n+1} \geqslant 2p_{n+1},$$

and this exceeds p_{n+2} because of Bertrand's postulate.

We conclude that, given $n \geqslant 5$, every integer of the same parity as T_n in the interval $[-T_n, T_n]$ can be expressed in the form (7) for an appropriate choice of signs. Now, if we add to these numbers the sum $\sum_{k=1}^{n} p_k$, which also has the parity of T_n, we obtain all the evens in

$$[2(2+5), 2(3+7+11+\ldots+p_n)],$$

each represented as a sum of primes $\leqslant p_n$, and with each prime in any representation appearing exactly twice. Hence each integer in $[7, 3+\sum_{k=4}^{n} p_k]$ can be partitioned into distinct primes $\leqslant p_n$. While the argument given is valid only for $n \geqslant 5$, the final result holds for $n = 4$ also.

Proof of Theorem 1; The lower estimate (2). We shall require the following lemma for both of (1) and (2).

Lemma. *For* $n \geqslant 4$,

$$\frac{n^2}{2}\log\log\frac{n}{2} < \int_4^n x\log\log x\,dx < \frac{n^2}{2}\log\log n.$$

Proof. The right-hand inequality is clear for $n = 4$, and by comparing derivatives one can see that $\frac{1}{2}n^2\log\log n$ increases faster than the integral does for $n \geqslant 4$.

The left-hand inequality is obvious for $n = 4$ and 5, and will follow for all larger n once we have shown that

$$\frac{d}{dn}\frac{n^2}{2}\log\log\frac{n}{2} < \frac{d}{dn}\int_4^n x\log\log x\,dx$$

for $n \geqslant 5$. This is equivalent to

$$\frac{1}{2\log\frac{n}{2}} < \log\frac{1}{1-\frac{\log 2}{\log n}} = \sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{\log 2}{\log n}\right)^k,$$

which for $5 \leqslant n \leqslant 12$ can be proved by direct computation and for $n \geqslant 13$ is a consequence of the stronger inequality

$$\frac{\log n}{\log(n/2)} < \log 4.$$

This holds firstly for $n = 13$, and since $\frac{\log n}{\log(n/2)}$ is a decreasing function, it holds for all larger n.

To proceed with (2), we have from Rosser and Schoenfeld ([2], Theorem 3)

$$p_k > k[\log(k\log k) - \tfrac{3}{2}] \quad \text{for } k \geqslant 2.$$

Hence

$$\sum_{k=4}^{n} p_k > \sum_{k=4}^{n} f(k) - \tfrac{3}{4}(n^2+n-12) \tag{9}$$

where $f(x) = x\log(x\log x)$. By the Euler–Maclaurin summation formula,

$$\sum_{k=4}^{n} f(k) = \int_4^n f(x)\,dx + \tfrac{1}{2}\{f(4)+f(n)\} + \int_4^n (x-[x]-\tfrac{1}{2})f'(x)\,dx \tag{10}$$

$$= \frac{n^2}{2}(\log n - \tfrac{1}{2}) + \frac{n}{2}\log(n\log n) + \int_4^n x\log\log x\,dx -$$

$$-6\log 4 + 2\log\log 4 + 4 + R;$$

here,

(11) $$R = \int_4^n (x-[x]-\tfrac{1}{2})f'(x)\,dx \geqslant 0$$

since $f'(x) = \log x + \log\log x + 1 + 1/\log x$ is increasing. From (9), (10), (11), and the lemma,

$$\sum_{k=4}^{n} p_k > \frac{n^2}{2}\log n + \frac{n^2}{2}\log\log\frac{n}{2} - n^2 + \frac{n}{2}\log n + \frac{n}{2}\log\log n - \tfrac{3}{4}n + 5.$$

For $n \geqslant 4$, the sum of the last four terms here is positive, and

(12) $$y(n) = 3 + \sum_{k=4}^{n} p_k > \frac{n^2}{2}\log n + \frac{n^2}{2}\log\log\frac{n}{2} - n^2.$$

One can obtain (2) for the case $a = 0$ by observing that

(13) $$\frac{n^2}{2}\log\log\frac{n}{2} > n^2 \quad \text{for } n > 2e^{e^2} = 3236.3\ldots,$$

whence by (12)

(14) $$y(n) > \tfrac{1}{2}n^2\log n$$

for $n \geqslant 3237$; by direct computation we have found that (14) holds for $5 \leqslant n < 3237$, and so $N(0) = 5$.

To prove (2) for $0 < a < 1$ we make somewhat better use of (12). Note that the inequality

$$\frac{n^2}{2}\log\log\frac{n}{2} - n^2 > a\frac{n^2}{2}\log\log n$$

is equivalent to

(15) $$a < \frac{\log\log(n/2) - 2}{\log\log n} = 1 - \frac{2 - \log[1 - \log 2/\log n]}{\log\log n}.$$

Therefore, given $a < 1$, it follows from (12) that (2) holds for all sufficiently large n, and so $N(a)$ exists. The statement involving ε in Theorem 1 also follows.

Remark 1. Some values of $N(a)$. Let

$$A(n) = \frac{y(n) - \tfrac{1}{2}n^2\log n}{\tfrac{1}{2}n^2\log\log n},$$

and let $B(n)$ denote the right side of (15). The proof of (2) depended on the implication

(16) $a < B(n)$

$\Rightarrow$

(17) $a < A(n)$,

the latter inequality being equivalent to (2). We thence apply our computations of $y(n)$ and $A(n)$ to finding $N(a)$ for $a = .1$ and for $a = .156$.

Regarding $N(.1)$, we first find that $B(22000) > .1$. Since $B(n)$ increases with n, $a = .1$ satisfies (16) for all $n \geqslant 22000$, and since by our computations

$$A(n) \geqslant A(30) = .1006\ldots \quad \text{for } 5 \leqslant n \leqslant 22000,$$

it will satisfy (17) for $n \geqslant 5$. Thus we can take

$$N(.1) = 5.$$

The largest value we can find is

$$N(.156) = 140:$$

$$B(100000) = .15607\ldots, \quad \text{and} \quad A(n) > .156 \quad \text{for} \quad 140 \leqslant n \leqslant 100000.$$

These are optimal values of $N(a)$, since $A(4)$ is negative and $A(139) < .156$.

Speculating further, we find from the computations that $A(n)$ is increasing for $2699 \leqslant n \leqslant 100000$, with $A(100000) = .384\ldots$, and so perhaps $N(.384) < 100000$. To go further yet, we find that $B(3\times 10^{24}) = .5008\ldots$, and here about all we can say is that

$$N(.5) < 3\times 10^{24}.$$

However, if the relation between $A(n)$ and $B(n)$ should continue as the earlier figures would indicate, $A(n)$ would be .5 for a much smaller n.

The upper estimates (1) and (3). Here we start from the inequality

(18) $$p_k < k[\log(k\log k) - \tfrac{1}{2}] \quad \text{for } k \geqslant 20$$

(Rosser and Schoenfeld, ibid.). Although this inequality is false for most k between 4 and 20, it turns out that the upper estimate we find for

$$z(n) = 3 + \sum_{k=4}^{n} k[\log(k\log k) - \tfrac{1}{2}] = \sum_{k=4}^{n} f(k) - \tfrac{1}{4}n^2 - \tfrac{1}{4}n + 6$$

yields the desired one for $y(n)$.

Regarding an upper estimate for the quantity R of (10) and (11), we first observe that

$$H(x) \underset{\text{def}}{=} f'(x) - \log x = \log\log x + 1 + 1/\log x < \tfrac{3}{2}\log x$$

for $x \geqslant 4$, since $H(4) = 2.04\ldots < 2.07\ldots = \frac{3}{2}\log 4$ and $\frac{3}{2}\log x$ increases faster than $H(x)$. Therefore

$$R < \tfrac{5}{2}\int_4^n |x-[x]-\tfrac{1}{2}|\log x\,dx < \tfrac{5}{4}\int_4^n \log x\,dx = \tfrac{5}{4}(n\log n - n) - 5(\log 4 - 1).$$

Hence from (10) and the lemma we find that for $n \geqslant 27$,

$$z(n) < \tfrac{1}{2}n^2(\log n + \log\log n) - \tfrac{1}{2}n^2 + \tfrac{7}{4}n\log n + \tfrac{1}{2}n\log\log n - \tfrac{3}{2}n - 11\log 4 + 2\log\log 4 + 15 < \tfrac{1}{2}n^2(\log n + \log\log n)$$

(in fact these inequalities hold for smaller n, but $n \geqslant 27$ suffices for our argument).

One can now see

$$y(n) < \tfrac{1}{2}n^2(\log n + \log\log n) \quad \text{for } n \geqslant 4$$

firstly by direct calculation for $4 \leqslant n \leqslant 26$ (see Table 1), and then by showing

$$y(n) < z(n) \quad \text{for } n \geqslant 27.$$

To see the latter, note that

$$y(27) = 1257 < 1260.8\ldots = z(27)$$

and that $y(n)$ grows more slowly than $z(n)$ for $n \geqslant 27$ because of (18). This proves (1).

Regarding (3), we see from Table 1 that

$$\sum_{k=1}^{n} p_k < \tfrac{1}{2}n^2(\log n + \log\log n) \tag{19}$$

for $6 \leqslant n \leqslant 27$, and that

$$\sum_{k=1}^{28} p_k = 1371 < 1373.8\ldots = z(28),$$

whence (19) holds for all $n \geqslant 6$. This completes the proof of Theorem 1.

Proof of Theorem 2. We first adopt a convenient notation for some of the lengthy expressions that are involved in the sequel. To illustrate the notation we use the first inequalities that are to be applied in the argument, namely the previously quoted result of Rosser and Schoenfeld,

$$n[\log(n\log n) - \tfrac{3}{2}] < p_n < n[\log(n\log n) - \tfrac{1}{2}]. \tag{20}$$

Table 1

n	$y(n)$	$\sum_{k=1}^{n} p_k$	$\frac{1}{2}n^2\log(n\log n)$	$z(n)$
4	10	17	13.7	.
5	21	28	26.0	.
6	34	41	42.7	.
7	51	58	63.9	.
8	70	77	89.9	.
9	93	100	120.8	.
10	122	129	156.8	.
11	153	160	197.9	.
12	190	197	244.4	.
13	231	238	296.3	.
14	274	281	353.7	.
15	321	328	416.7	.
16	374	381	485.4	.
17	433	440	559.8	.
18	494	501	640.1	.
19	561	568	726.3	.
20	632	639	818.5	.
21	705	712	916.8	.
22	784	791	1021.1	.
23	867	874	1131.6	.
24	956	963	1248.2	.
25	1053	1060	1371.2	.
26	1154	1161	1500.4	1153.1
27	1257	1264	1636.0	1260.8
28		1371		1373.8

The numbers in the last two columns are truncated at one decimal place.

With such an expression we shall use (20LI) to denote the left inequality and (20RI) to denote the right, while (20L), (20C), and (20R) shall denote the three expressions on the left, centre, and right respectively. Thus (20C) $= p_n$, (20LI) holds for $n \geqslant 2$, and (20RI) holds for $n \geqslant 20$.

From Theorem 1 we have

$$\tfrac{1}{2}n^2(\log n + a\log\log n) < y(n) < \tfrac{1}{2}n^2(\log n + \log\log n) \tag{21}$$

where (21LI) holds for $n \geqslant N(a)$ and (21RI) for $n \geqslant 4$. We have to prove that, given $\varepsilon > 0$,

$$U\big(y(n), \tfrac{1}{2}-\varepsilon\big) < p_n < U\big(y(n), \tfrac{1}{2}+\varepsilon\big) \tag{22}$$

for all sufficiently large n.

Since $U(y, c)$ is an increasing function, (21) is equivalent to

$$(23)\qquad n\sqrt{\log n+a\log\log n}\times$$
$$\times\sqrt{\log n+\tfrac{1}{2}\log(\log n+a\log\log n)+c\log\{\log n+\tfrac{1}{2}\log(\log n+a\log\log n)\}}$$
$$< U\big(y(n), c\big)$$
$$< n\sqrt{\log(n\log n)}\sqrt{\log n+\tfrac{1}{2}\log\log(n\log n)+c\log\{\log n+\tfrac{1}{2}\log\log(n\log n)\}}$$

in the sense that (21LI)⇔(23LI) and (21RI)⇔(23RI). Using the notation (23L, c) for the left expression in (23), and (23R, c) for the right, we want to prove

$$(24)\qquad U\big(y(n), c_1\big) < (23\mathrm{R}, c_1) < (20\mathrm{L}) < p_n < (20\mathrm{R}) < (23\mathrm{L}, c_2) < U\big(y(n), c_2\big)$$

with $c_1 = \frac{1}{2}-\varepsilon$ and $c_2 = \frac{1}{2}+\varepsilon$. The two central inequalities are just (20), while the outermost inequalities come from (23), thence from (21). Therefore we have to prove only the two inequalities involving the expressions from (20) and (23).

Regarding the right side of (24), we write $c_2 = c$ and make the following observations (which will serve again in the proof of Theorem 4):

$$(20\mathrm{R}) < (23\mathrm{L}, c)$$

⇔

$$\log n+\log\log n-\tfrac{1}{2} < \sqrt{\log n+a\log\log n}\times$$
$$\times\sqrt{\log n+\tfrac{1}{2}\log(\log n+a\log\log n)+c\log\{\log n+\tfrac{1}{2}\log(\log n+a\log\log n)\}}$$

⇔

$$2\log n\log\log n-\log n+(\log\log n-\tfrac{1}{2})^2$$
$$< (c+\tfrac{1}{2}+a)\log n\log\log n+\underline{\tfrac{1}{2}\log n\log\left[1+\frac{a\log\log n}{\log n}\right]}+$$
$$+\underline{c\log n\log\left[1+\frac{\log(\log n+a\log\log n)}{2\log n}\right]}+a(c+\tfrac{1}{2})(\log\log n)^2+$$
$$+\underline{\frac{a}{2}\log\log n\log\left[1+\frac{a\log\log n}{\log n}\right]}+\underline{ac\log\log n\log\left[1+\frac{\log(\log n+a\log\log n)}{2\log n}\right]}$$

⇔

$$(25)\qquad (\log\log n-\tfrac{1}{2})^2 < (c-\tfrac{3}{2}+a)\log n\log\log n+\log n+ a(c+\tfrac{1}{2})(\log\log n)^2+t(n)$$

where $t(n)$ is the sum of the underlined expressions. It is clear that $t(n) = O(\log\log n)$ and $t(n) > 0$ for $n \geqslant 3$. In order for (25) to hold for all large n, it is necessary that $c \geqslant \frac{3}{2}-a$. In fact, with $c = \frac{3}{2}-a$, (25) holds for all $n \geqslant 3$; this is the optimal value for c, given a, since for sharpness in (24) we want $c = c_2$ to be as small as possible.

Thus follows the right side of (24), with $c_2 = \frac{3}{2}-a$, so that

$$(26)\qquad p_n < U\big(y(n), \tfrac{3}{2}-a\big)\quad \text{for } n \geqslant \max\{20, N(a)\}$$

— that is, for all n such that (20RI) and (21LI) both hold. By Theorem 1, given $0 < \varepsilon \leqslant 1$, we can take $a = 1-\varepsilon$ here provided $n \geqslant n_0(\varepsilon) = N(1-\varepsilon)$ (and $n \geqslant 20$), and so we have the right-hand inequality in (22) and in Theorem 2.

Regarding the left side of (24), we find in a similar way that

$$(23\mathrm{R}, c_1) < (20\mathrm{L})$$

⇔

$$(27)\qquad (c_1-\tfrac{1}{2})\log n\log\log n+3\log n+(c_1+\tfrac{1}{2})(\log\log n)^2+s(n) < (\log\log n-\tfrac{3}{2})^2$$

where $0 < s(n) = O(\log\log n)$. In order for (27) to hold for all large n, it is necessary that $c_1 < \frac{1}{2}$, and indeed, given $\varepsilon > 0$, it is clear that there is a number $n_1(\varepsilon)$ such that (27) will hold with $c_1 = \frac{1}{2}-\varepsilon$ for all $n \geqslant n_1(\varepsilon)$. Thus we have the left inequality in (22) and in Theorem 2 for $n \geqslant n_1(\varepsilon)$. (The first and third inequalities on the left side of (24) hold for $n \geqslant 4$ and $n \geqslant 2$ respectively.)

Remark 2. (Take $c_1 = \frac{1}{2}-\varepsilon$ in (27).)

(a) The presence of the term $3\log n$ in (27) requires that $n_1(\varepsilon)$ be very large. For example with $\varepsilon = .1$, (27) requires

$$(3-.1\log\log n)\log n+.9(\log\log n)^2+s(n) < (\log\log n-\tfrac{3}{2})^2,$$

and this will be false for n smaller than about $\exp\exp 30$.

(b) The inequality

$$(28)\qquad p_n < U\big(y(n), \tfrac{3}{2}\big)$$

for $n \geqslant 6$ follows from the above argument: since $N(0) = 5$, (26) yields it for $n \geqslant 20$, while for $6 \leqslant n < 20$ it is proved by direct calculation.

The inequality

$$(29)\qquad U\big(y(n), 0\big) < p_n$$

follows from (27), and so holds for $n \geqslant n_1(\frac{1}{2})$; all that (27) allows us to say about $n_1(\frac{1}{2})$ is that it likely exceeds $\exp\exp 6 \doteqdot 10^{175}$ — cf. the discussion regarding $\varepsilon = .1$ in (a). Our computations show that (29) holds for $31 \leqslant n \leqslant 100000$; values of the two sides and their difference, for selected n, are shown in Table 2.

Proof of Theorem 4. This is similar to the proof of the right-hand inequality in Theorem 2. We shall use the same special notation involving statement numbers as before, and adapt it further to our present purpose by writing for example $[23\mathrm{L}, c]\,(n-1)$ to denote $(23\mathrm{L}, c)$ with $n-1$ in place of n. Instead of the right-side of (24), we shall prove

(30) $$p_n < (n-1)[\log(n-1)+\log\log(n-1)-\tfrac{1}{4}] < [23\mathrm{L}, c](n-1) < U\big(y(n-1), c\big).$$

The first inequality of (30) holds for $n \geqslant 20$: For $n = 20$ it can be seen to hold by direct computation, and for larger n it follows from (20RI) and the inequality

(31) $$\log n + \log\log n - \frac{1}{2} < \left(1-\frac{1}{n}\right)\left[\log(n-1)+\log\log(n-1)-\frac{1}{4}\right],$$

which holds for $n \geqslant 21$ (the left side increases more slowly than the right, and for $n = 21$, $(31\mathrm{L}) = 3.657\ldots < (31\mathrm{R}) = 3.659\ldots$).

Table 2

n	$U(y(n), 0)$*	$p_n - U(y(n), 0)$*	p_n	$U(y(n), \frac{1}{2}) - p_n^*$	$U(y(n), \frac{1}{2})$*
31	118.0	8.9	127	0.8	127.8
5000	46937.5	1673.4	48611	1010.4	49621.4
10000	101382.1	3346.8	104729	2220.7	106949.7
20000	217732.0	7004.9	224737	4501.7	229238.7
30000	339680.9	10696.0	350377	6871.9	357248.9
40000	465270.4	14638.5	479909	9067.4	488976.4
50000	593535.3	18417.6	611953	11480.6	623433.6
60000	723965.6	22807.3	746773	13327.4	760100.4
70000	856164.6	26212.3	882377	16193.6	898570.6
80000	989889.7	30489.2	1020379	18217.7	1038596.7
90000	1124973.8	34549.1	1159523	20486.0	1180009.0
100000	1261197.9	38511.0	1299709	22873.0	1322582.0

* Truncated at one decimal place

The last inequality in (30) is $[23\mathrm{LI}](n-1)$, which follows from $[21\mathrm{LI}](n-1)$ and so holds for $n-1 \geqslant N(a)$. Hence, once the central inequality in (30) is proved, with $c = \frac{3}{2}-a$ and for $n \geqslant 20$, we shall have

(32) $$p_n < U\big(y(n-1), \tfrac{3}{2}-a\big) \quad \text{for } n \geqslant \max\{20, N(a)+1\}$$

(cf. (26)). Given this result, then for y an integer satisfying $y(n-1) < y$, where $n \geqslant \max\{20, N(a)+1\}$, we shall have

$$p_n < U(y, \tfrac{3}{2}-a);$$

if further $y \leqslant y(n)$, then by Theorem 3 y can be partitioned into distinct primes $\leqslant p_n$. Putting $a = 1-\varepsilon$, we can then conclude that y can be partitioned into distinct primes p with

(33) $$p < U(y, \tfrac{1}{2}+\varepsilon),$$

and Theorem 4 will have been proved.

In proving the central inequality in (30), we use the same argument as that used in showing

$$(20\mathrm{R}) < (23\mathrm{L}, c) \Leftrightarrow (25)$$

in the proof of Theorem 2, except with $(n-1)$ in place of n, and with $\frac{1}{4}$ in place of the $\frac{1}{2}$ in (20R). We find that the central inequality is equivalent to

(34) $$\big(\log\log(n-1)-\tfrac{1}{4}\big)^2 < (c-\tfrac{3}{2}+a)\log(n-1)\log\log(n-1)+\tfrac{1}{2}\log(n-1)+ + a(c+\tfrac{1}{2})[\log\log(n-1)]^2+t(n-1),$$

where t is the function in (25): $t(n-1) = O(\log\log n)$ and $t(n-1) > 0$ for $n \geqslant 4$. With $c = \frac{3}{2}-a$ (cf. the discussion following (25)), (34) holds for $n \geqslant 20$ (indeed for $n \geqslant 4$); the proof of this last is as follows.

Since $\frac{1}{2}\log(n-1) < (34\mathrm{R})$, and the functions tabulated below are increasing, the values listed in the table show that (34) holds for $20 \leqslant n \leqslant 645$.

n	$g(n) = (\log\log(n-1)-\frac{1}{4})^2$	$h(n) = \frac{1}{2}\log(n-1)$
20	.68...	1.47...
71	1.43...	2.12...
201	2.00...	2.64...
645	2.61...	

Beyond $n = 645$, $g(n)$ increases more slowly than does $h(n)$, since

$$g'(645) = .00077634\ldots < .00077639\ldots = h'(645)$$

and $g'(n)$ increases more slowly than does $h'(n)$ for $n \geqslant 645$ (in fact for $n > 55$).

Theorem 4 now follows.

Remark 3. Replacing a by its supremum 1 in (32) would give

$$p_n < U\big(y(n-1), \tfrac{1}{2}\big),$$

which would imply the inequality in Theorem 4 without the ε. Our computations reveal that for $105 \leqslant n \leqslant 100\,000$,

$$p_n < U\big(y(n-d), \tfrac{1}{2}\big)$$

with differences $d > 0$ that tend to increase with n. Some sample values follow; the values of U in the table are truncated at one decimal place.

n	p_n	d	$U(y(n-d), \frac{1}{2})$
5000	48611	91	48616.6
10000	104729	188	104732.9
20000	224737	358	224744.7
30000	350377	528	350389.1
40000	479909	681	479910.7
50000	611953	847	611955.9
60000	746773	969	746773.9
70000	882377	1163	882385.4
80000	1020379	1295	1020380.7
90000	1159523	1443	1159525.7
100000	1299709	1598	1299722.5

References

[1] H. E. Richert, *Über Zerfällungen in ungleiche Primzahlen,* Math. Zeitschr. 52 (1950), pp. 342–343.

[2] J. Barkley Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers,* Illinois J. Math. 6 (1962), pp. 64–94.

[3] W. Sierpiński, *Elementary Theory of Numbers,* Warszawa 1964.

UNIVERSITY OF VICTORIA
Victoria, B. C., Canada V8W 2Y2

Received on 30. 11. 1979
and in revised form on 8. 4. 1980 (1186)

ACTA ARITHMETICA
XLI (1982)

Gaps between primes, and the pair correlation of zeros of the zeta-function

by

D. R. Heath-Brown (Oxford)

1. Introduction. In studying the finer properties of the distribution of the zeros of the Riemann Zeta-function, Montgomery [7] examined the pair correlation function

$$F_T(X) = \sum_{0<\gamma_1,\gamma_2\leqslant T} W(\gamma_1-\gamma_2)e\big(X(\gamma_1-\gamma_2)\big).$$

Here $W(u) = 4/(4+u^2)$, $e(u) = \exp(2\pi iu)$, and γ runs over imaginary parts $\mathrm{Im}(\varrho)$ of the non-trivial zeros ϱ of $\zeta(s)$ (counted according to multiplicity). Montgomery based his investigation on the assumption of the Riemann Hypothesis, and we shall follow him; for convenience we use the abbreviation RH. It is clear that $F_T(X) = F_T(-X)$, that $F_T(X) \ll T(\log T)^2$, and that $F_T(X)\geqslant 0$ (this follows from Lemma 3 below). Montgomery showed in addition that, on RH,

$$(1)\qquad F_T(X) = TX+\frac{T}{2\pi}x^{-2}(\log T)^2+O(T)+O(xX)+O(Tx^{-3/2}\log T),$$

for $x = e^{2\pi X}$, $x\geqslant 1$. Actually he stated a slightly less precise result, but it is clear that his analysis leads to the above refinement. When $0<\delta\leqslant\beta\leqslant 1-\delta$, where, as later, $X = (\beta\log T)/2\pi$, (1) reduces to $F_T(X) = TX+O(T)$, uniformly in β. Moreover, Montgomery conjectured, in general, that

$$(2)\qquad F_T(X)\sim\frac{T}{2\pi}(\log T)\mathrm{Min}(1, |\beta|)$$

uniformly in $0<\delta\leqslant|\beta|\leqslant A$. From (1) he deduced, on RH, several important consequences for the distribution of the γ's, and he showed that the conjecture (2) would lead to more powerful conclusions — for example, that 'almost all' zeros would be simple.

Results connecting the distribution of the primes p_n and the zeros of the Zeta-function have long been known. In particular von Koch [5]