Dense Admissible Sets

Daniel M. Gordon and Gene Rodemich

Center for Communications Research
4320 Westerra Court
San Diego, CA 92121
{gordon,gene}@ccrwest.org

Abstract. Call a set of integers $\{b_1, b_2, \dots, b_k\}$ *admissible* if for any prime p, at least one congruence class modulo p does not contain any of the b_i. Let $\rho^*(x)$ be the size of the largest admissible set in $[1, x]$.
The Prime k-tuples Conjecture states that any for any admissible set, there are infinitely many n such that $n+b_1, n+b_2, \dots n+b_k$ are simultaneously prime. In 1974, Hensley and Richards [3] showed that $\rho^*(x) > \pi(x)$ for x sufficiently large, which shows that the Prime k-tuples Conjecture is inconsistent with a conjecture of Hardy and Littlewood that for all integers $x, y \geq 2$,

$$\pi(x + y) \leq \pi(x) + \pi(y).$$

In this paper we examine the behavior of $\rho^*(x)$, in particular, the point at which $\rho^*(x)$ first exceeds $\pi(x)$, and its asymptotic growth.

1 Introduction

The Prime k-tuples Conjecture states that for any set $\{b_1, b_2, \dots, b_k\}$ of integers which do not cover all congruence classes modulo any prime, there are infinitely many integers n such that $n + b_1, n + b_2, \dots, n + b_k$ are all prime. Call such a set *admissible*. Let $\rho^*(x)$ be the size of the largest admissible set in $[1, x]$. It is known that

$$\pi(x) + (\log 2 - o(1))\frac{x}{\log^2 x} \leq \rho^*(x) \leq 2\pi(x). \tag{1}$$

The lower bound is due to Hensley and Richards [3], and the upper bound was shown by Montgomery and Vaughn [6], using the large sieve.

The main interest in $\rho^*(x)$ is that the Hensley and Richards result shows that the widely believed Prime k-tuples Conjecture is inconsistent with a conjecture of Hardy and Littlewood that for all integers $x, y \geq 2$,

$$\pi(x + y) \leq \pi(x) + \pi(y). \tag{2}$$

Let $\rho(x) = \limsup_{y\to\infty} \pi(x + y) - \pi(y)$. Then (2) is equivalent to $\rho(x) \leq \pi(x)$ for $x \geq 2$. The Prime k-tuples Conjecture implies $\rho^*(x) = \rho(x)$. By finding dense admissible sets, Hensley and Richards [3] showed that the Prime k-tuples Conjecture implies sets of primes in short intervals $(y, x+y]$ which are denser than the $\pi(x)$ primes in $[1,x]$. They construct their sets by sieving out by congruence classes a_1 mod 2, a_2 mod $3,\ldots$, trying to leave as many survivors as possible. When they have sieved by all primes up to x, the set of survivors is clearly admissible.

Hensley and Richards showed that $\rho^*(x) > \pi(x)$ for $x = 20,000$. Unfortunately, the smallest prime k-tuple will typically be of size about k^k, so actually finding such a k-tuple seems hopeless. Computations by Selfridge showed that $\rho^*(x) < \pi(x)$ for $x \leq 500$. Jarvis [4] computed $\rho^*(x)$ for $x \leq 1050$, and showed that $\rho^*(x) < \pi(x)$ for $x \leq 1120$. In Section 3 we describe extending those computations, in particular computing $\rho^*(x)$ for $x < 1631$, and finding that $\rho^*(1417) = \pi(1417)$.

Hensley and Richards showed that one particular sieve, the midpoint sieve, resulted in an admissible set giving the lower bound in (1). Another sieve, suggested by Schinzel, gives a better bound, but depends on a conjecture concerning how far the sieve has to go before the set is admissible. We will discuss these and other sieves in the next section.

The question of extreme behaviors of sieving is of independent interest. Here we are concerned with *saving* sieves, which leave as many survivors as possible. The problem of *killing* sieves, where we want to eliminate all of $[1,x]$ using as few primes as possible, is related to Jacobstahl's function.

The Jacobsthal function $j(n)$ is the maximal gap between consecutive integers relatively prime to n. Maier and Pomerance [5] define $j'(n)$ to be the largest x for which a sieve by the factors of n eliminate all integers in $[1,x]$. Then $j'(n) = j(n) - 1$ by the Chinese Remainder Theorem. Let

$$P(x) = \prod_{p \leq x} p.$$

Then $j'(P(x))$ is the largest interval which can be completely sieved out by primes up to x. Maier and Pomerance show

$$j'(P(x)) \geq (c_0 e^{\gamma} + o(1)) x \log x \log\log\log x (\log\log x)^{-2}, \tag{3}$$

where $c_0 = 1.312\ldots$ is the solution of

$$4/c_0 - e^{-4/c_0} = 3.$$

Pintz [7] recently improved the constant from c_0 to 2.

To rephrase this in a way more convenient for our purposes, let $T(x)$ be the smallest number t such that there is a sieve by primes up to t which sieves out $[1,x]$. Then from (3) we have

$$T(x) \leq O(x(\log\log x)^2 / \log x \log\log\log x)$$

Maier and Pomerance conjecture that $j'(P(x)) = x(\log x)^{2+o(1)}$. This would imply

$$T(x) \approx x/(\log x)^{2+o(1)}. \tag{4}$$

The killing sieve that Maier and Pomerance use to establish (3) is the same as the one suggested by Schinzel for a saving sieve. In both cases we sieve by $1 \bmod p$ for small primes, $0 \bmod p$ for "medium-sized" primes, and optimally modulo large primes. This strategy may not result in the best possible killing or saving sieve, but it does yield a set which can be analyzed using standard sieve methods.

2 Sieve Strategies

2.1 The Sieve of Eratosthenes

The sieve of Eratothenes is an obvious starting point. Unchanged, it is not a good saving sieve, since it covers all points in $[1, x]$. However, it is easy to turn it into a saving sieve by stopping the sieve when $p > x/2$. For all larger primes the congruence classes containing $[x/2]$ and $[x/2] + 1$ will each contain only a single point in $[1, x]$, and one of these integers had to be eliminated by the sieve modulo two. The survivors of the sieve are the primes greater than $x/2$. Changing over to a greedy strategy earlier will do better, but will clearly be worse than $\pi(x)$ if the sieve of Eratothenes is done up to $\sqrt{x}$.

We can stop the sieve significantly earlier than $x/2$. It is obvious that once p is greater than the number of survivors, some congruence class will be completely covered, and the sieve can be halted. Let $t(z)$ be the inverse function of $T(x)$, i.e. the largest x such that $[1, x]$ can be sieved out by primes $p < z$. Hensley and Richards ([3], Lemma 5*) show

Lemma 1. *There is an x_0 such that for $x > x_0$, the survivors of any sieve by primes up to $x/t(\log x/2)$ are admissible.*

Hensley and Richards only needed $T(x) = o(x)$, to show that their sieve could stop at $x/N \log x$ for any $N > 0$. If (4) is true, then a sieve by primes up to $x/\log x(\log\log x)^{2+o(1)}$ will be admissible. In Section 2.4 we give a heuristic argument that the survivors of almost all sieves by primes up to $cx/\log^2 x$ will be admissible for any $c > 2$.

2.2 Random and Greedy Sieves

A random sieve is easy to analyze. Sieving by a random congruence modulo p will on average eliminate $1/p$ of the remaining integers. Thus the expected number of survivors is

$$x \prod_{p<x} (1 - 1/p) \approx e^{-\gamma} \frac{x}{\log x} \approx 0.5614 \, \frac{x}{\log x}.$$

by Mertens' Theorem.

One might hope that a greedy algorithm would do better. Instead of choosing a congruence class at random, choose the best possible for each p. For small primes, say $p < \log x$, the distribution of survivors over congruence classes modulo p is very flat, by standard sieve arguments. For larger primes it becomes less regular, so larger improvements over the random sieve can be obtained.

Let $g(x)$ be the size of the admissible set in $[1, x]$ generated by the greedy algorithm. In case of ties, we pick the first congruence class which eliminates the fewest possible survivors (other choices do not affect the behavior much). Figure 1 shows $g(x) - \pi(x)$ for $x < 181,000$. The first x where $g(x)$ equals $\pi(x)$ is $g(11046) = 1337$, and $g(x)$ is greater than $\pi(x)$ at $g(11916) = 14294$. It seems likely that

$$\limsup(g(x) - \pi(x)) \to \infty,$$

and

$$\liminf(g(x) - \pi(x)) \to -\infty.$$

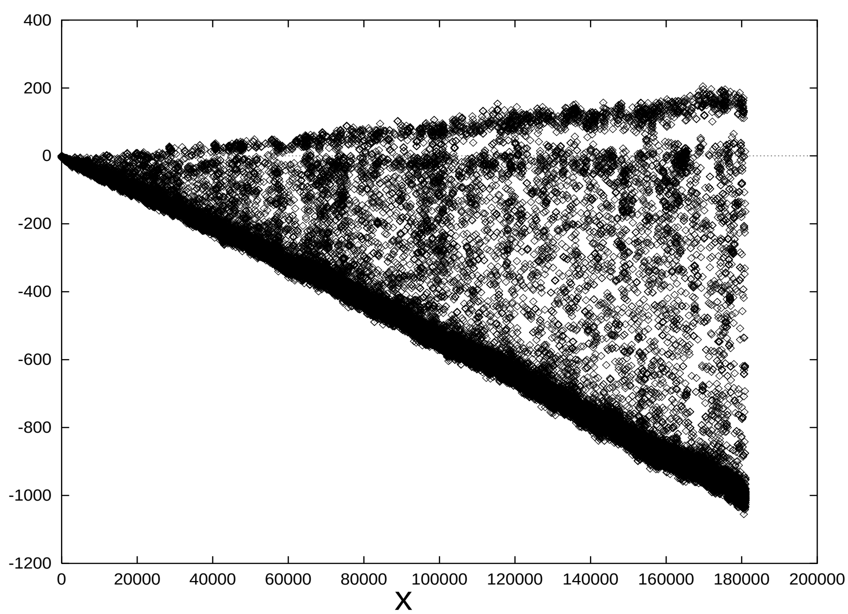

Fig. 1. $g(x) - \pi(x)$

2.3 The Midpoint Sieve

Hensley and Richards use the fact that primes are denser in $[-x/2, x/2]$ than in $[1, x]$. Sieving $[-x/2, x/2]$ by primes up to $x/N \log x$ for any $N > 0$ and x sufficiently large, the number of survivors is

$$2\pi(x/2) - 2\pi(x/N \log x) \approx \pi(x) + (\log 2 - \epsilon(N))(x/\log^2 x)$$

by the sharp form of the Prime Number Theorem. The resulting set of survivors will be admissible by Lemma 1.

2.4 Schinzel's Sieve

It is hard to find other sieve strategies that can be analyzed. One notable one is used in analyses of Jacobstahl's function and prime gaps as well. In this context, it was suggested by Schinzel (see [3]).

Choose $y < z < x$. The sieve is just a variation on the sieve of Eratothenes: sieve by $1 \bmod p$ for $p \le y$, and $0 \bmod p$ for $y < p \le z$. Hensley and Richards show that for y fixed, $m = \pi(y)$, and $z = x/N \log x (\log\log x)^m$, the number of survivors will be

$$\pi(x) + (1 + o(1)) \frac{x}{\log^2 x} \sum_{r<y} \frac{r \log r}{(r-1)^2}.$$

By taking y large, the difference between this and $\pi(x)$ may be made larger than $cx/\log^2 x$ for any constant c. The problem is that we do not know if the resulting set is admissible. Hensley and Richards conjecture that it is, and show that admissibility is implied by the stronger conjecture:

$$T(x) = o(x/\log^m x).$$

However, by (4) this is unlikely for $m \ge 2$. Thus, it is still possible that $\rho^*(x)$ and $\pi(x)$ only differ by $O(x/\log^2 x)$. However, we give a heuristic argument that the difference is at least slightly greater.

Conjecture 1.

$$\rho^*(x) \ge \pi(x) + (1 + o(1)) x \log\log\log x / \log^2 x.$$

Heuristic Argument. We will follow the argument in [3]. As before, we will sieve by $1 \bmod p$ for $p \le y$, and $0 \bmod p$ for $p \le z$. Here we take $y = \log\log x$, and $z = cx/\log^2 x$, for any $c > 2$.

Call an integer y-smooth if all its prime factors are at most y. The survivors will be the union $R_{(1)} \cup R_{(2)}$, where $R_{(1)}$ is the set of integers in $(0, x]$ of the form mp, where $p > z$ is prime, m is y-smooth and $mp - 1$ is relatively prime to $P(y)$, and $R_{(2)}$ is the set of y-smooth integers in $(0, x]$. We will ignore the smaller set $R_{(2)}$, which has size $O(x^\epsilon)$ for any $\epsilon > 0$, and show that $R_{(1)}$ has the conjectured cardinality.

We have

$$R_{(1)} = {\bigcup_{m \le x/z}}' R_m,$$

where

$$R_m = \{mp : z < p \le x/m, (mp - 1, P(y)) = 1\},$$

and the prime indicates that the union (and sums below) are over y-smooth integers.

By the Siegel-Walfisz Theorem on primes in arithmetic progressions,

$$|R_m| = \pi\left(\frac{x}{m}\right) \prod_{\substack{r \le y \\ r \nmid m}} \left(\frac{r-2}{r-1}\right) \left(1 + O(\log^{-A} x)\right)$$

for any fixed $A > 0$. Since

$$\pi(x/m) = \pi(x)\left(1/m + \log m/(m \log x) + O(\log^{-2} x)\right),$$

we have

$$\sum_{m \le x/z}{}' |R_m| = \pi(x) \sum_{m \le x/z}{}' \prod_{\substack{r \le y \\ r \nmid m}} \left(\frac{r-2}{r-1}\right)\left(\frac{1}{m} + \frac{\log m}{m \log x} + O\left(\frac{1}{\log^2 x}\right)\right)$$

survivors. As in (**) of [3], using estimates for the number of y-smooth numbers less than $x/z = O(\log^2 x)$ (see [2]), this becomes

$$\begin{aligned}
&\pi(x)\left[\left(\sum_{m=1}^{\infty}{}' \prod_{\substack{r \le y \\ r \nmid m}} \left(\frac{r-2}{r-1}\right)\left(\frac{1}{m} + \frac{\log m}{m \log x}\right)\right) + o\left(\frac{1}{\log x}\right)\right] \\
&= \pi(x)\left[\prod_{r \le y}\left(\frac{r-2}{r-1} + \sum_{a=1}^{\infty}\frac{1}{r^a} + \sum_{a=1}^{\infty}\frac{\log r^a}{r^a \log x}\right) + o\left(\frac{1}{\log x}\right)\right] \\
&= \pi(x)\left[\prod_{r \le y}\left(1 + \frac{1}{\log x}\sum_{a=1}^{\infty}\frac{a \log r}{r^a}\right) + o\left(\frac{1}{\log x}\right)\right] \\
&= \pi(x)\left[1 + \frac{1}{\log x}\sum_{r \le y}\frac{r \log r}{(r-1)^2} + o\left(\frac{1}{\log x}\right)\right] \\
&= \pi(x)\left[1 + \frac{\log y}{\log x}(1 + o(1))\right] \\
&= \pi(x) + (1 + o(1))\frac{x \log\log\log x}{\log^2 x}.
\end{aligned}$$

The problem is that we cannot show that this set is admissible. A heuristic argument, assuming that the survivors are distributed more or less randomly among congruence classes modulo larger primes, indicates that it should be. Each prime $p > z$ has $p > cx/\log^2 x$ congruence classes, and $s < 2x/\log x$ survivors (by [6]) are being put in them. If survivors were randomly distributed, then the probability that some congruence class is empty is $e^{-\lambda}$, where $\lambda = pe^{-s/p} > x^{(c-2)/c-\epsilon}$ (see, for example, Section IV.2 of [1]). For $c > 2$, the probability that all congruence classes are covered for any prime $p > z$ is $o(1)$.

The experimental data is not very helpful (see Figure 2), since asymptotics do not take over until x is quite large. The crossover point for this sieve with $y = 2$ is 904,036, and with larger y much larger.

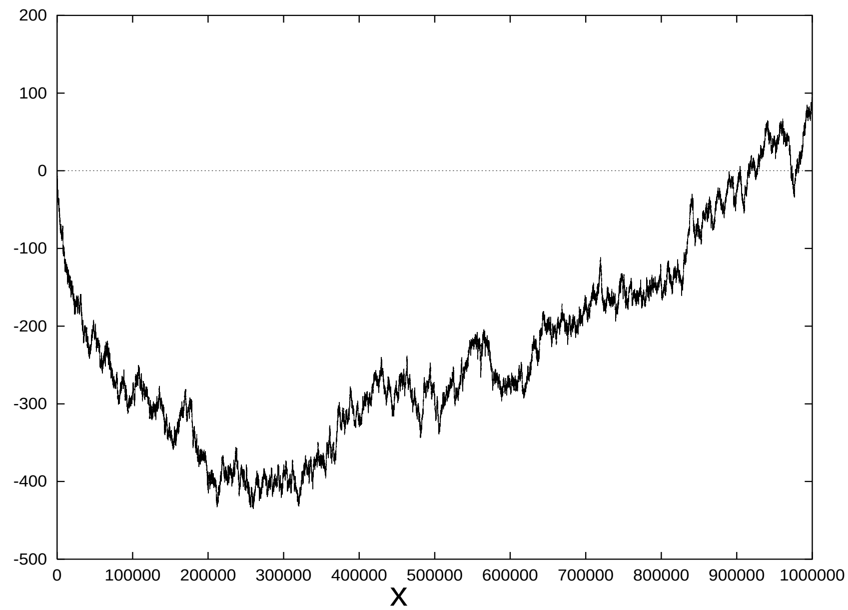

Fig. 2. Difference between $y = 2$ sieve and $\pi(x)$

3 Computing $\rho^*(x)$ Efficiently

The only way to compute $\rho^*(x)$ seems to be to exhaust over residues modulo $2, 3, \ldots$, looking for sets of residues with a large number of survivors. It is possible to add various tricks to make this search more efficient, greatly speeding up the search.

We start by doing a sieve modulo small primes. By the Chinese Remainder Theorem, looking at survivors in $[1, x]$ of sieves by $a_1 \bmod 2$, $a_2 \bmod 3$, $\ldots$, $a_k \bmod p_k$, for all residue classes modulo each prime, is equivalent to looking at all integers in $(y, x + y]$ for $0 \le y < P(p_k)$ which are relatively prime to $P(p_k)$.

Thus we can divide up work into two parts: first sieve by primes up to (say) 29 in $[1, P(29)]$, and look for intervals with enough survivors to improve the current bound on $\rho^*(x)$. For each such interval, exhaust over residue classes modulo 31, 37, $\ldots$, until either the number of survivors is less than the current bound on $\rho^*(x)$, or we reach a prime larger than the number of survivors, in which case the survivors are admissible.

A number of simple theorems about $\rho^*(x)$ may be used to speed the search. If $\{s_1, \ldots, s_l\} \subset [1, x]$ is an admissible set, so is $\{x+1-s_1, \ldots, x+1-s_1\}$. Thus, we only need to sieve in $[1, P(29)/2]$, cutting the work in half (x numbers should be added on each side of the interval to avoid problems with the boundaries).

Since $\rho^*(x) = \rho^*(x-1)$ for x even, we only need to look at odd values of x. Also, since

$$\rho^*(x-2) \le \rho^*(x) \le \rho^*(x-2) + 1$$

we can stop as soon as we find an improvement. We only need to look at intervals $(y, x+y]$ where the endpoints $y+1$ and $x+y$ are survivors. If either one was not, then the interval without that endpoint would have been checked before. This eliminates a large fraction of the work.

One further theorem is of great use:

Theorem 1. *If $\rho^*(x+2) > \rho^*(x) > \rho^*(x-2)$, then $x \equiv 1 \bmod 3$.*

Proof. Consider the optimal sieve on $[1, x+2]$. As mentioned above, 1 and $x+2$ must be survivors, or else we would have $\rho^*(x) = \rho^*(x+2)$. We also have that 3 and x are survivors, or else the interval $[5, x+2]$ (respectively $[1, x-2]$) would give us $\rho^*(x-2) = \rho^*(x+2) - 1$. The only way that $1, 3, x$ and $x+2$ can all be survivors is if we are sieving out by 2 mod 3, and $x \equiv 1 \bmod 3$.

This allows us to skip $x+2$ whenever $\rho^*(x) > \rho^*(x-2)$ and $x \not\equiv 1 \bmod 3$. If $x \equiv 1 \bmod 3$, the search is still greatly sped up by the requirement that 3 and x must be survivors.

Even so, as x increases, the work becomes formidable. Finding $\rho^*(x)$ for all x becomes impractical, and it is quicker to just look for possible crossover points. This is accomplished by finding $\rho^*(x)$ for $x = p_k - 2$, and checking if some interval has k or more survivors. This lets us skip many values of x, and have a higher threshold for the number of survivors.

This search was implemented on a Cray T3D. Parallelizing was accomplished by breaking the sieve interval into equal pieces. It seemed possible that load balancing would be a problem, if one interval took much longer than others, but this does not seem to happen.

In any large computation, there is some question about whether the algorithm has been implemented correctly. If some intervals were being skipped or not handled correctly, the values of $\rho^*(x)$ might not be right. The original program was written by the first author in C. The second author got interested in the problem and wrote an independent search program in Fortran, which got the same answers and was significantly faster.

The first crossover point, with $\rho^*(x) = \pi(x)$ is at $x = 1417$ (Jarvis discovered an admissible set for $x = 1422$ with the same cardinality). Unfortunately, the prime 1423 is followed by a prime pair 1427 and 1429, sending $\pi(x)$ ahead again, where it remains for a long time. The search was continued up to $x = 1663$.

We can push the bound for the crossover point somewhat higher, using an idea of Schinzel [8]. The inequality

$$\rho^*(x+y) \leq \rho^*(x) + \rho^*(y) \tag{5}$$

allows us to get upper bounds for $\rho^*(x)$ over a larger range. Using the computed values of $\rho^*(x)$ and (5), we find that $\rho^*(x) \leq \pi(x)$ for $x \leq 1731$.

Jarvis [4] suggested looking at local maxima of $\mathrm{Li}(x) - \pi(x)$, on the grounds that these are points where $\pi(x)$ is smaller than expected, and so $\rho^*(x)$ has a better chance of exceeding it. One such point is $x = 1423$, and the next is $x = 1971$. Unfortunately, a search found no admissible set of length 1971 with

298 elements. It is possible that the crossover point is smaller than 1971, but more likely that it is larger, perhaps at the next local maximum $x = 2203$, which is computationally infeasible to check.

Jarvis also used a combination of exhaustive search on small primes and a greedy strategy for larger primes to get an upper bound for the crossover point. He showed that $\rho^*(4930) \geq 658 > \pi(4930)$.

Figure 3 shows $\rho^*(x) - \pi(x)$ for $x \leq 1631$. The two functions stay extremely close for a long time, and make it tempting to conjecture that

$$\lim_{x\to\infty} \frac{\rho^*(x)}{\pi(x)} = 1,$$

but as Figure 2 indicates, extrapolating from limited data can be perilous.

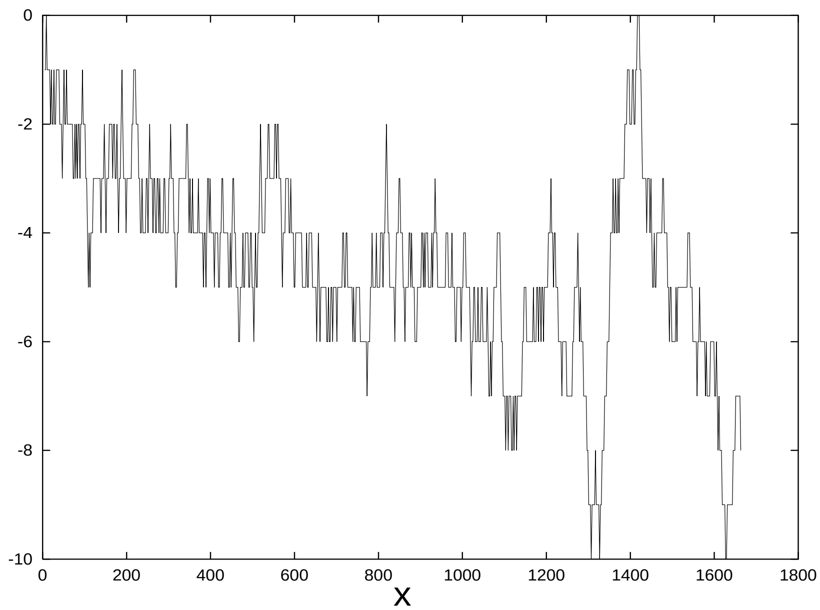

Fig. 3. $\rho^*(x) - \pi(x)$

Acknowledgment. We would like to thank John Selfridge for making us aware of the work of Jarvis [4].

References

1. William Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, third edition, 1968.

2. Andrew Granville. On positive inegers $\leq x$ with prime factors $\leq t \log x$. In R. A. Mollin, editor, *Number Theory and Applications*, pages 403–422. Kluwer, 1989.
3. Douglas Hensley and Ian Richards. Primes in intervals. *Bulletin AMS*, 25:375–391, 1974.
4. Norman C. Jarvis. Admissible sequences. Master's thesis, Brigham Young University, 1996.
5. Helmut Maier and Carl Pomerance. Unusually large gaps between consecutive primes. *Trans. AMS.*, 322:201–237, 1990.
6. H. L. Montgomery and R. C. Vaughn. The large sieve. *Mathematika*, 20:119–134, 1973.
7. János Pintz. Very large gaps between consecutive primes. *J. Number Theory*, 63:286–301, 1997.
8. A. Schinzel. Remarks on the paper 'sur certaines hypothèse concernant les nombres premiers'. *Acta Arith.*, pages 185–208, 1958.