# When Euler met Brun

by

JOHN B. FRIEDLANDER (Toronto)

*For the 75th birthday of Henryk Iwaniec*

**Abstract.** In this note we survey and study some aspects of the distribution of primes in very short intervals.

**1. Introduction.** L. Euler famously proved that the sum of the reciprocals of the prime numbers is divergent, strengthening a millennia-old theorem of Euclid. It is reasonable to view Euler's result, at least when taken together with Dirichlet's profound generalization to primes in arithmetic progressions, as having launched the modern subject of analytic number theory.

V. Brun famously proved that, when restricted to a sum over the reciprocals of twin primes, the series is in contrast convergent, along the way launching the modern subject of sieve theory.

If we let $p'$ denote the least prime exceeding the prime $p$, its successor prime, then Brun's theorem may be stated as $\sum_{p'=p+2} 1/p < \infty$. It is well-known and clear from the proof that, with virtually no modification, Brun's theorem holds as well for the sum $\sum_{p'=p+k} 1/p$ for any even integer $k$ and, since it is trivial for odd $k$, we have

$$\sum_{p'-p \leq K} \frac{1}{p} < \infty,$$

for any fixed $K$.

So a question that naturally suggests itself is to wonder what happens when we replace $K$ by a growing function, say $y(p) = \lambda(p) \log p$, and try to study how quickly growing a function we can take and still maintain the

[1]

conclusion

$$(1.1) \qquad \lim_{x \to \infty} \sum_{\substack{p \le x \\ p' - p \le y(p)}} \frac{1}{p} < \infty.$$

We know, by a result of Goldston, Pintz and Yıldırım [GPY4, Theorem 1], that for any fixed $\lambda > 0$, a positive proportion $c(\lambda)$ of the primes $p \le x$ have successor satisfying the bound $p' - p < \lambda \log p$. Thus we are not able to take $y(p) = \lambda \log p$ for any fixed $\lambda > 0$, no matter how small. This means that we are dealing with prime gaps that are of smaller order of magnitude than the average.

On the other hand, if we could take a choice of $y$ which yielded, for some fixed real $a > 1$, an upper bound

$$\sum_{\substack{p \le x \\ p' - p \le y(p)}} 1 \ll \frac{x}{(\log x)^a}$$

then partial summation would show the convergence of its sum of reciprocals. Recall that, in the case of twin primes, we have such a bound with the comfortable margin $a = 2$. Moreover, for the full set of primes $(y = p)$ the divergence, namely $\sum_{p \le x} 1/p \sim \log \log x$, just barely holds. This suggests that the set of twin primes is substantially further from the "meeting place" and that the relevant gaps are short but not so terribly short.

After a tiny bit of history and a few remarks in Section 2, we introduce in Section 3 what we believe to be a reasonable conjecture which allows us to deduce fairly close upper and lower bounds for the function $y(p)$ in question. Then, in Section 4, we consider what we are able to learn about the problem unconditionally using sieve bounds. The proofs are elementary.

**2. Small gaps and very small gaps.** By the Prime Number Theorem we know that the average gap between primes of size around $x$ is $\log x$. As it happens however, our knowledge of primes is still sufficiently limited that, for at least one problem, that of maximal gaps, intervals as large as $x^{.49}$ are still too small to be within our reach. Nevertheless, for purposes of this paper, we think of prime gaps near $x$ having length $\lambda \log x$, $\lambda < 1$ a constant, as being "smaller than average", those unbounded but of size $y(x) = o(\log x)$ as being "very small" and of bounded gaps as being, well, bounded.

At the two extremes of these three gap ranges there has, after many years, been phenomenal recent progress. Thus, for the smaller than average gaps $\lambda \log x$ one has the breakthroughs of Goldston, Pintz and Yıldırım [GPY1], wherein they show that $\lim \inf (p' - p)/\log p = 0$ and subsequently as well the positive proportion result already mentioned.

The GPY bound was greatly strengthened to encompass the bounded gaps by the work of Zhang [Zha], then Maynard [M1], who showed that $\liminf(p' - p)$ is bounded and, in the latter case, quite a bit more as well.

The functions $y(x)$ which are near the meeting place for our problem lie in what we have called very small gaps, the largely uncharted region between these two extreme interval lengths.

There seems to be not a great deal known about this middle range. One exception comes from Maynard's sieve which, although originally designed to treat distribution in bounded gaps, still has striking consequences which venture into the middle range. Thus, Maynard's Theorem 3.2 of [M2] states:

THEOREM (Maynard). *For any $x, y \geq 1$ there are $\gg x \exp(-\sqrt{\log x})$ integers $x_0 \in [x, 2x]$ such that*

$$\pi(x_0 + y) - \pi(x_0) \gg \log y.$$

Another exception, closer to our problem, is the sieve-driven upper bound [GPY4, Theorem 2], the statement of which we postpone until Section 4.

Finally, we mention the papers of Erdős and Nathanson [EN] and of Zhou [Zho] which study a problem of flavour similar to ours but dealing with a sum over (suitably weighted) reciprocals of prime gaps.

**3. Conditional results.** Despite the breakthroughs of GPY and Maynard, much of what is thought to be true about the distribution of small prime gaps is either conjecture or else deductions from conjectures. Thus, Granville and Lumley [GrLu] study the maximum number of primes in a gap near $x$ having length $y$ and conjecture that it is asymptotic to $y/\log y$ "for $y \leq c \log x$ as long as $y \to \infty$ as $x \to \infty$".

Nevertheless, most of what appears in the literature on gaps of below average length seems to deal only with gaps of size $\lambda \log x$ with constant $\lambda$ rather than $\lambda = o(1)$. A pioneering work which is concerned with this range of $\lambda$ constant is that of Gallagher [G].

Gallagher's theorem on primes in short intervals rests on the assumption of a uniform version of a well-known conjecture due to Hardy and Littlewood.

Let $\mathcal{H} = h_1, \ldots, h_r$ denote a set of distinct non-negative integers and let $\nu_{\mathcal{H}}(p)$ denote the number of residue classes modulo $p$ occupied by the $h_j$. Assume that the set $\mathcal{H}$ is "admissible" in the sense that $\nu_{\mathcal{H}}(p) < p$ for all primes $p$. We recall the definition of the relevant "singular series"

$$(3.1) \qquad \mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-r}.$$

HYPOTHESIS G. *Let $\pi(x; \mathcal{H})$ denote the number of integers $n \leq x$ such that $n + h_j$ is prime for all $h_j \in \mathcal{H}$. Then*

$$(3.2) \qquad \pi(x; \mathcal{H}) \sim \mathfrak{S}(\mathcal{H})\frac{x}{(\log x)^r}$$

*as $x \to \infty$ uniformly for all $h_1, \ldots, h_r \leq \lambda \log x$, where $\lambda > 0$ is fixed.*

A similar conjecture (on the number of prime couples and prime triples with fixed spacing) was used by Goldston and Ledoan [GoLe2] in relation to their work on the prime jumping champions.

THEOREM (Gallagher). *Denote by $P_k(h, x)$ the number of integers $n \leq x$ for which the interval $(n, n + h]$ contains exactly $k$ primes. Then, under Hypothesis G, with $h = \lambda \log x$, $\lambda > 0$ fixed,*

$$(3.3) \qquad P_k(h, x) \sim xe^{-\lambda}\frac{\lambda^k}{k!}$$

*as $x \to \infty$.*

Soundararajan pointed out, in a slightly different form [S, Exercise 1.3] (see also Goldston and Ledoan [GoLe1]), the following consequence of Gallagher's theorem.

COROLLARY. *Again under Hypothesis G, for fixed $\lambda > 0$ we have*

$$\frac{1}{\pi(x)}\#\left\{p \leq x;\ \frac{p' - p}{\log p} \leq \lambda\right\} \sim \int_0^\lambda e^{-t}\,dt = 1 - e^{-\lambda}.$$

In order to close in on a consequence of this for our problem, we shall assume that this statement holds in a range slightly different from $\lambda$ fixed, but rather with $\lambda \to 0$ slowly as $x \to \infty$. This seems reasonable, especially if the approach is sufficiently slow, for instance the following is more than ample.

HYPOTHESIS S. *Assume that, as $t \to \infty$ with $\lambda(t) \to 0$, subject to*

$$(3.4) \qquad \lambda(t) \gg \frac{1}{(\log\log t)^2}$$

*we have*

$$(3.5) \qquad \frac{1}{\pi(x)}\#\left\{p \leq x;\ \frac{p' - p}{\log p} \leq \lambda(p)\right\} \sim \int_0^{\lambda(x)} e^{-u}\,du = 1 - e^{-\lambda(x)}.$$

Note that for $\lambda = o(1)$ we have $1 - e^{-\lambda} \sim \lambda$.

We prefer to base our conditional statements on Hypothesis S; as it concerns a *sum* over gaps of specific lengths, one might believe it could hold even if the previous two conjectures, which deal with individual spacings, did not. We are not suggesting that Hypothesis S holds without further conditions

on the function $\lambda$; we only require it for the specific functions mentioned in Proposition 3.1 below.

As it happens, we shall need iterated logarithms to enter our statements so we define, as usual, for $k \geq 2$, $\log_k$ to be the $k$th iterated logarithm:

$$(3.6) \qquad \log_k = \log \log_{k-1}, \quad \log_1 = \log$$

and what we might call the logorial function

$$(3.7) \qquad \mathrm{Log}_k = \prod_{2 \leq j \leq k} \log_j .$$

For each of these functions, we can ignore the finite set of small integers where it is not defined.

Applying partial summation to the sum in (1.1), then Hypothesis S, next applying the Prime Number Theorem to the resulting integral, then making the change of variable $u = \log_{k+1} t$, respectively $u = \log_k t$, one deduces the following results which narrow in on the answer to our question.

PROPOSITION 3.1. *Let $k \geq 2$ be a fixed integer and $\varepsilon > 0$ a fixed real. Denote $y(p) = \lambda(p) \log p$. Then we have, as $x \to \infty$,*

$$(3.8) \qquad \sum_{\substack{p \leq x \\ p'-p \leq y(p)}} \frac{1}{p} \sim \log_{k+1} x \to \infty \quad \text{for } \lambda(p) = 1/\mathrm{Log}_k(p),$$

*but*

$$(3.9) \qquad \sum_{\substack{p \leq x \\ p'-p \leq y(p)}} \frac{1}{p} \quad \text{is bounded for } \lambda(p) = 1/\mathrm{Log}_k(p)(\log_k p)^\varepsilon,$$

*under the assumption that (3.5) holds for these functions $\lambda$.*

In fact, as we shall see in the next section, the latter statement (3.9) holds unconditionally, although the former seems hopeless with current tools. We may remark that Theorem 1 of [GPY4] gives a lower bound for the left hand side of (3.5), weaker but unconditional and holding for intervals of smaller than average size, that is, $\lambda$ constant. If this could be extended to also hold for intervals having length in a suitable part of the range $o(\log x)$, perhaps using ideas from [M2], this might give a correspondingly weaker but unconditional result in the direction of (3.8).

Note that (3.8), if true, is not the end of the story since we could then construct more artificial choices for $y(p)$ which are closer to the breakpoint. To see this, begin with $\lambda = 1/\mathrm{Log}_k$ for some particular $k$ as in (3.8), carry on until the sum of reciprocals of primes up to $x$ exceeds 1 (still more complicated choices go further) and also $x$ is large enough that $\log_{k+1} x$ is defined. Then, for ensuing $p$, replace $1/\mathrm{Log}_k$ by $1/\mathrm{Log}_{k+1}$ in the definition of $y(p)$, continue summing and iterate.

**4. Sieve survivors.** By a sieve "survivor" we mean what is more awkwardly called "an integer with no small prime factor" but in recent years has, at least in sieve theory, become confused with an "almost-prime" even though the latter is invariably defined as an integer with few prime factors. There are many sieve results which produce what are called almost-primes when they actually produce integers from the special subset of survivors. When we use the term here we are thinking of an integer $\leq x$ having no prime factor $< z$ with $z = x^\delta$ for some $\delta > 0$, possibly small but fixed. The survivors, at least in one sense, more strongly resemble the primes, occurring as they do with the same order of magnitude. Recall that the almost-primes occur with a greater order of magnitude, how much so depending on the number of prime factors permitted.

In proving lower and upper bounds for the sum $\sum_p 1/p$ by partial summation, it is obvious that we do not need an asymptotic formula such as that conjectured in (3.5), but only sufficiently good lower, respectively upper, bounds for the corresponding prime counting function. This suggests that sieve methods can prove useful. Already, in the work of Gallagher [G], his Theorem 2 gives an upper bound in connection with his question.

More recently, again in the case of the upper bound, one has a result more closely related to our topic, namely Theorem 2 of Goldston, Pintz and Yıldırım [GPY4], which states:

For any $h > 2$, as $x \to \infty$, we have

$$(4.1) \qquad \sum_{\substack{p \leq x \\ p'-p \leq h}} 1 \ll \min\{h/\log x, 1\}\, \pi(x),$$

which for us is of interest when $h = o(\log x)$, although it is also meaningful if $h \leq c \log x$ provided that the constant $c$ is sufficiently small. This shows that, if in (3.5) we replace the conjectured asymptotic by an upper bound, then that result is unconditionally true in a wide range.

We shall take $z = x^\delta$ for some positive $\delta$. Let $m$ run through $\mathcal{M}$, the set of integers $m \leq x$ which are the survivors of sieving by the primes less than $z$. For each $m$ let $m'$ be its successor, the first larger integer in the set.

A proof, perhaps slightly streamlined, of (4.1) can proceed as follows. We start with a sieve upper bound (e.g. [Opera, Theorem 7.16]):

$$(4.2) \qquad \sum_{\substack{m \leq x \\ m,\, m+d \in \mathcal{M}}} 1 \ll \mathfrak{S}_d \frac{x}{(\log x)^2},$$

which holds for every integer $d$, $1 \leq d \leq h$, with the same implied constant depending only on $\delta$. Here, $\mathfrak{S}_d$ is the singular series (3.1) for the set

$\mathcal{H} = \{0, d\}$. Summing over $1 \leq d \leq h$ and using the fact that

$$(4.3) \qquad \sum_{d \leq h} \mathfrak{S}_d \sim h$$

as $h \to \infty$ (much more is known; see [FG, Proposition 1]), we obtain a bound

$$(4.4) \qquad \sum_{\substack{m \in \mathcal{M} \\ m' - m \leq h}} 1 \leq \sum_{\substack{m_1, m_2 \in \mathcal{M} \\ 1 \leq m_1 - m_2 \leq h}} 1 \ll h \frac{x}{(\log x)^2} \sim \frac{h}{\log x} \pi(x).$$

This bound applies, a fortiori, to the corresponding sums over primes since the number of primes less than $z$ offers a negligible contribution. This proves (4.1).

Note that the upper bound in (4.4) is actually a bound for a larger number of $m \in \mathcal{M}$, specifically those with two *or more* unsifted integers in the interval $[m, m+h]$. As it happens, the number of triples in such a short interval is, as we shall see, of smaller order of magnitude and we lose nothing for this result by ignoring them.

On the other hand, returning to the sum in (4.2), we can instead begin with a sieve lower bound. If we have chosen for instance $\delta = 1/10$, then $z$ is sufficiently small to enable a positive lower bound for this two-dimensional sieve problem (see e.g. [Opera, Corollary 6.13]):

$$(4.5) \qquad \sum_{\substack{m \leq x \\ m, m+d \in \mathcal{M}}} 1 \gg \mathfrak{S}_d \frac{x}{(\log x)^2}.$$

After summing over $d$ we have, in contrast to (4.4),

$$(4.6) \qquad \sum_{\substack{m_1, m_2 \in \mathcal{M} \\ 1 \leq m_1 - m_2 \leq h}} 1 \gg h \frac{x}{(\log x)^2} \sim \frac{h}{\log x} \pi(x).$$

Now however, in order to get a lower bound for the number of consecutive pairs corresponding to that in (4.4) we need to subtract out (an upper bound for) the contribution of triples in the short intervals. For this, we need an analogue of (4.3). We take a special case of a more general result of Odlyzko, Rubinstein and Wolf [ORW] which, in turn, extended a basic result of Gallagher [G]; see also [GoLe2, Lemma 2].

For $\mathcal{H} = \{0, d_1, d_2\}$ we have

$$(4.7) \qquad \sum_{d_1, d_2 \leq h} \mathfrak{S}(\mathcal{H}) \sim h^2,$$

as $h \to \infty$. This yields an upper bound (again, [Opera, Theorem 7.16])

$$(4.8) \qquad \sum_{1 \leq d_1 < d_2 \leq h} \sum_{\substack{m \leq x \\ m, m+d_1, m+d_2 \in \mathcal{M}}} 1 \ll h^2 \frac{x}{(\log x)^3}$$

and as we intend to take $h = o(\log x)$, this is small compared to the lower bound in (4.6). Hence we have

$$(4.9) \qquad \sum_{\substack{m \in \mathcal{M} \\ m'-m \leq h}} 1 \gg h \frac{x}{(\log x)^2} \sim \frac{h}{\log x} \pi(x).$$

We want to use upper and lower bounds for a sum much as that in (4.9) but slightly modified. In the first place, we want to change the sifting range so that, rather than having $z = x^\delta$ in the definition of $\mathcal{M}$, we now have $z = m^\delta$. We also want to be able to allow $h$ to depend on $m$, specifically $h = y(m) = \lambda(m) \log m$. To treat this new sum we split the interval $[0, x]$ into dyadic segments $I = (M, 2M]$ with perhaps a shorter interval left over. We request that $y(m)$ be a positive slowly increasing function such as either of the two choices in Proposition 3.1. In particular, we want to make use of the fact that $y(m)$ is very nearly constant over every dyadic interval.

We follow the arguments that led to (4.4) and (4.9), obtaining the new bounds

$$(4.10) \qquad \sum_{\substack{m \in I \cap \mathcal{M} \\ m'-m \leq y(m)}} 1 \asymp \frac{y(M)M}{(\log M)^2}.$$

Summing (4.10) over the subintervals and again using the fact that $y(m)$ is nearly constant over dyadic intervals, we find the bounds

$$(4.11) \qquad \frac{1}{|\mathcal{M}|} \# \left\{ m \in \mathcal{M}; \; \frac{m'-m}{\log m} \leq \lambda(m) \right\} \asymp 1 - e^{-\lambda(x)}.$$

Using (4.11) in place of (3.5) in the argument for Proposition 3.1, we obtain the following result.

PROPOSITION 4.1. *Let $k \geq 2$ be a fixed integer and $\varepsilon > 0$ a fixed real. Let $\mathcal{M}$ denote the set of integers $m \leq x$ which are free from prime divisors less than $m^\delta$ where $\delta > 0$ is fixed and sufficiently small that $1/\delta$ exceeds the sifting limit for the two-dimensional (beta or Selberg) sieve. Let $m'$ denote the successor of $m$ in the set $\mathcal{M}$ and define $y(m) = \lambda(m) \log m$. Then we have, as $x \to \infty$,*

$$(4.12) \qquad \sum_{\substack{m \leq x \\ m'-m \leq y(m)}} \frac{1}{m} \to \infty \quad \text{for } \lambda(m) = 1/\mathrm{Log}_k(m),$$

*but*

$$(4.13) \qquad \sum_{\substack{m \leq x \\ m'-m \leq y(m)}} \frac{1}{m} \quad \text{is bounded for } \lambda(m) = 1/\mathrm{Log}_k(m)(\log_k m)^\varepsilon.$$

Note that the sieve limit restriction on $\delta$ is not needed in the case of (4.13) and that (4.13) shows that (3.9) is unconditionally true. As far as the lower bound is concerned, (4.12) lends some additional credence to the belief that (3.8) is also true.

## References

[EN]    P. Erdős and M. B. Nathanson, *On the sum of the reciprocals of the differences between consecutive primes*, in: Number Theory: New York Seminar 1991–1995, D. V. Chudnovsky et al. (eds.), Springer, New York, 1996, 97–101.

[FG]    J. B. Friedlander and D. A. Goldston, *Some singular series averages and the distribution of Goldbach numbers in short intervals*, Illinois J. Math. 39 (1995), 158–180.

[Opera] J. B. Friedlander and H. Iwaniec, *Opera de Cribro*, Amer. Math. Soc. Colloq. Publ. 57, Amer. Math. Soc., Providence, RI, 2010.

[G]     P. X. Gallagher, *On the distribution of primes in short intervals*, Mathematika 23 (1976), 4–9.

[GoLe1] D. A. Goldston and A. H. Ledoan, *On the differences between consecutive prime numbers I*, Integers 12B (2012/13), art. A3, 8 pp.

[GoLe2] D. A. Goldston and A. H. Ledoan, *The jumping champion conjecture*, Mathematika 61 (2015), 719–740.

[GPY1]  D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Primes in tuples I*, Ann. of Math. (2) 170 (2009), 819–862.

[GPY4]  D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Primes in tuples IV: Density of small gaps between consecutive primes*, Acta Arith. 160 (2013), 37–53.

[GrLu]  A. Granville and A. Lumley, *Primes in short intervals: heuristics and calculations*, Exp. Math. 32 (2023), 378–404.

[M1]    J. Maynard, *Small gaps between primes*, Ann. of Math. (2) 181 (2015), 383–413.

[M2]    J. Maynard, *Dense clusters of primes in subsets*, Compos. Math. 152 (2016), 1517–1554.

[ORW]   A. Odlyzko, M. Rubinstein and M. Wolf, *Jumping champions*, Exp. Math. 8 (1999), 107–118.

[S]     K. Soundararajan, *The distribution of prime numbers*, in: Equidistribution in Number Theory: An Introduction, NATO Sci. Ser. II Math. Phys. Chem. 237, Springer, Dordrecht, 2007, 59–83.

[Zha]   Y. Zhang, *Bounded gaps between primes*, Ann. of Math. (2) 179 (2014), 1121–1174.

[Zho]   N. H. Zhou, *On the sum of the reciprocals of the differences between consecutive primes*, Ramanujan J. 47 (2018), 427–433.

John B. Friedlander
Department of Mathematics
University of Toronto
Toronto, Ontario M5S 2E4, Canada
E-mail: frdlndr@math.toronto.edu