AN UPPER BOUND FOR THE SUM OF LARGE DIFFERENCES BETWEEN PRIME NUMBERS

R. J. COOK

ABSTRACT. Let $p_n$ denote the $n$th prime number, $d_n = p_{n+1} - p_n$. We estimate the sum $\sum d_n$ taken over $p_n < x$, $d_n > x^\mu$ where $1/6 < \mu < 5/9$.

1. Introduction. There has been much work done on studying irregularities in the distribution of prime numbers. Let $p_n$ denote the $n$th prime number and set $d_n = p_{n+1} - p_n$. Assuming the Riemann hypothesis Cramér [3] proved that $d_n = O(p_n^{1/2}\log p_n)$ and [4] that for any $\epsilon > 0$

$$\sum_{p_n < x} d_n^2 \ll x \log^{3+\epsilon} x,$$

where the constant implied by Vinogradov's notation depends only on $\epsilon$, and Selberg [17] subsequently improved on this.

For $0 < \mu < 1$ let

$$S_\mu(x) = \sum_{p_n < x, d_n > x^\mu} d_n$$

and take $f(\mu)$ to be the least value such that for all $\epsilon > 0$,

$$S_\mu(x) \ll x^{f(\mu) + \epsilon} \text{ as } x \to \infty.$$

From (1) we see that the Riemann hypothesis implies

$$f(\mu) \begin{cases} 0 & \text{for } \mu > 1/2, \\ < 1 - \mu & \text{for } 0 \leq \mu < 1/2. \end{cases}$$

Montgomery [15, pp. 130–132] has shown that the density hypothesis is sufficient to imply that $f(\mu) = 0$ for $\mu > 1/2$.

Since the Riemann hypothesis implies that $d_n = O(p_n^{1/2}\log p_n)$, Erdös asked whether $f(1/2) < 1$ and there has been much interest recently in estimating $f(1/2)$ (see [2], [11], [12], [16] and [20]). Heath-Brown [7] has given an unconditional proof that

$$\sum_{p_n < x} d_n^2 \ll x^{4/3}(\log x)^{10,000}$$

which gives

$$f(\mu) \begin{cases} 0 & \text{for } \mu > 2/3, \\ < 4/3 - \mu & \text{for } 2/3 \geq \mu > 1/3. \end{cases}$$
while Huxley [10] has shown that \( f(\mu) = 0 \) for \( \mu > 7/12 \).

A recent result of Warlimont [19], when combined with Huxley’s density theorem [10], shows that \( f(\mu) < 1 \) for \( \mu > 1/6 \). Warlimont’s result depends on an inexplicit estimate of Halász and Turán [6] and so seems incapable of giving explicit upper bounds for \( f(\mu) \) in the range \( \mu > 1/6 \). We use different estimates to improve the upper bound (3) throughout the interval \( 5/9 > \mu > 1/6 \).

**Theorem.** We have \( f(\mu) < F(\mu) \) where

\[
F(\mu) = \begin{cases} 
19/18 - \mu/2 & \text{for } 47/99 < \mu < 5/9, \\
10/7 - 9\mu/7 & \text{for } 3/8 < \mu < 47/99, \\
1 - \mu/7 & \text{for } 7/32 < \mu < 3/8, \\
11/10 - 3\mu/5 & \text{for } 1/6 < \mu < 7/32.
\end{cases}
\]

In particular, \( f(1/2) < 29/36 = .805 \) which is an improvement on Ivić’s estimate [13], \( f(1/2) < .809 \).

I am grateful to Roger Heath-Brown, Martin Huxley and Aleksander Ivić for making manuscripts available prior to publication; I am indebted to the referee for many helpful suggestions and for pointing out a gap in my original argument.

**2. Preliminaries.** Let \( \varepsilon > 0 \) and \( x > x_0(\varepsilon) \) be large. It is sufficient to obtain the corresponding estimates for a sum

\[
\sum_{x < p_n < 2x \atop d_n > x^\alpha} d_n
\]

since the theorem then follows on summing over intervals \([x, x/2], [x/2, x/4], \ldots\) and noticing that at \( O(\log x) \) end-points we have \( d_n = O(x^{7/12+\varepsilon}) \): see Huxley [10].

The function

\[
\psi(z) = \sum_{\rho < z} \log \rho
\]

has an explicit formula (see Chandrasekharan [1, p. 120])

\[
\psi(z) = z - \sum_{|\gamma| < T} z^\rho/\rho + E(z, T)
\]

where the summation is taken over the nontrivial zeros \( \rho = \beta + i\gamma \) of \( \xi(s) \), \( 0 < \alpha < 1 \) and \( E(z, T) = O(zT^{-1}\log^2 z) \) uniformly over the range \( 3 < T < z \).

We take

\[
T = x^\alpha, \quad U = 2^{1-\mu}, \quad \delta = \log(1 + 1/U) \ll U^{-1},
\]

and put

\[
\Delta(y) = \psi\left(y + \frac{y}{U}\right) - \psi(y) - \frac{y}{U} + \sum_{\rho} \left(\frac{e^{\delta \rho} - 1}{\rho}\right) y^\rho - E\left(y + \frac{y}{U}\right) + E(y)
\]

where \( x < y < 2x \), the summation is over the zeros of \( \xi(s) \) in the region

\[
|\gamma| < x^\alpha, \quad \sigma_0 < \beta < 1,
\]
and $\sigma_0 = \sigma_0(\mu)$ will be chosen later. From (5) we have

$$\Delta(y) = \sum_{\rho \in (10)} \frac{(e^{\beta} - 1)}{\rho} y^\rho$$

(9)

where the summation is over the zeros of $\xi(s)$ in the region

$$|\gamma| < \sigma, \quad 0 < \beta < \sigma_0.$$  

(10)

Following Warlimont [19] we have

$$\int_0^{2\pi} |\Delta(y)|^2 \, dy \ll 2^x \log^2 x \sum_{\rho \in (10)} x^{2\beta} \ll U^{-2} \log^2 x \sum_{\rho \in (10)} x^{2\beta}.$$  

(11)

The proof of the theorem depends on obtaining a lower bound for $|\Delta(y)|^2$ in terms of large differences between primes, and an upper bound for the right side of (11) via zero-density estimates.

Let $N(\sigma, T)$ denote the number of zeros of $\xi(s)$ in the rectangle $\sigma < \beta < 1$, $|\gamma| < T$ and write $N(T)$ for $N(0, T)$. Then, see Davenport [5, Chapter 15], we have

$$N(T) \ll T \log T.$$  

(12)

Since

$$\sum_{\rho \in (10)} (x^{2\beta} - 1) = 2 \sum_{\rho \in (10)} \int_0^\beta x^{2\sigma} \log x \, d\sigma$$

$$= 2 \int_0^{\sigma_0} x^{2\sigma} N(\sigma, x^{\sigma}) \log x \, d\sigma$$

we have

$$\sum_{\rho \in (10)} x^{2\beta} \ll x^{1+\sigma} \log x + \int_{1/2}^{\sigma_0} x^{2\sigma} N(\sigma, x^{\sigma}) \log x \, d\sigma.$$  

(13)

3. Density theorems. These are upper bounds for $N(\sigma, T)$ which can be stated in either the form

$$N(\sigma, T) \ll T^{g(\sigma)(1-\sigma)} \log^A T$$  

(14)

where $A$ is an absolute constant, not necessarily the same at each occurrence, or in the form

$$N(\sigma, T) \ll T^{g(\sigma)(1-\sigma) + \eta}$$  

(15)

where $\eta$ is an arbitrary small positive constant. In (14) and (15) the implicit constants are independent of $\sigma$, although they may depend on $A$ and $\eta$.

**Lemma 1.** For $1/2 < \sigma < 1$, (14) holds with $g(\sigma) = 3/(2 - \sigma)$.

This is due to Ingham [12].

**Lemma 2.** For $3/4 < \sigma < 1$, (14) holds with $g(\sigma) = 3/(3\sigma - 1)$.

This is due to Huxley [10].

**Lemma 3.** For $3/4 < \sigma < 11/14$, (15) holds with $g(\sigma) = 1/(7\sigma - 5)$.

This is the case $k = 3$ of equation (1.8) of Jutila [13].
**Lemma 4.** For $11/14 < \sigma < 1$, (15) holds with $g(\sigma) = 9/(7\sigma - 1)$.

This is due to Heath-Brown [9].

**Lemma 5.** Let $M(\psi, T) = \max|\xi(s)|$ where the maximum is taken over $\Re s > \psi$, $|\Im s| < T$ and $|s - 1| > 1$. Then for $T > 2$

$$N(\sigma, T) \ll (BM(\psi, 8T) \log T)^{2(1-\sigma)(3\psi - 1 - 2\psi)/(2\sigma - 1 - \psi)(\psi - \psi)} \log^2 T$$

provided that $1/2 < \psi < 1$ and $\sigma > (1 + \psi)/2$, where $B$ is a large absolute constant.

This is Theorem 12.3 of Montgomery [15].

We combine Lemma 5 with estimates for $\xi(s)$ obtained by van der Corput's method.

**Lemma 6.** If $l > 3$, $L = 2^{l-1}$, $\sigma = 1 - l/(2L - 2)$ then

$$\xi(\sigma + it) \ll t^{1/(2L - 2)} \log t.$$ 

This is Theorem 5.14 of Titchmarsh [18].

**Lemma 7.** For $17/18 < \sigma < 1$, (14) holds with $g(\sigma) = 1/(7\sigma - 6)$.

**Proof.** Taking $l = 4$ in Lemma 6 we have

$$\xi(5/7 + it) \ll t^{1/14} \log t.$$ 

This implies that the function

$$g(s) = \frac{\xi(s)}{s^{1/14} \log s} \left(\frac{s - 1}{s}\right)^2$$

which is holomorphic for $\Re s > 1$, is uniformly bounded on the line $\Re s = 5/7$. It is also uniformly bounded in the half-plane $\Re s > 2$, so by the Phragmen-Lindelöf theorem it is uniformly bounded in $\Re s > 5/7$. Hence

$$M(5/7, 8T) \ll T^{1/14} \log T.$$ 

Now apply Lemma 5 with $\psi = 5/7$. For $17/18 < \sigma < 1$ the exponent of $B$ is bounded uniformly in $\sigma$ so

$$g(\sigma) \leq \frac{2(3\sigma - 1 - 10/7)}{14(2\sigma - 1 - 5/7)(\sigma - 5/7)} = \frac{(21\sigma - 17)}{2(7\sigma - 6)(7\sigma - 5)} \leq \frac{1}{(7\sigma - 6)}.$$ 

Combining these lemmas, we have

$$N(\sigma, T) \ll T^{G(\sigma)(1-\sigma) + \eta} \quad (1/2 < \sigma < 1)$$

where

$$G(\sigma) = \begin{cases} 
3/(2 - \sigma), & 1/2 < \sigma < 3/4, \\
3/(3\sigma - 1), & 3/4 < \sigma < 7/9, \\
1/(7\sigma - 5), & 7/9 < \sigma < 11/14, \\
9/(7\sigma - 1), & 11/14 < \sigma < 53/56, \\
1/(7\sigma - 6), & 53/56 < \sigma < 1.
\end{cases}$$
Thus $G(\sigma)$ is a continuous function with $G(1/2) = 2$, increasing monotonically to the value $G(3/4) = 12/5$ and thereafter decreasing monotonically to the value $G(1) = 1$.

4. **Estimation of an integral.** For given $\mu$ satisfying $1/6 < \mu < 5/9$ we choose $\sigma^* = \sigma^*(\mu)$ to be the largest $\sigma^* < 1$ satisfying $(1 - \mu)G(\sigma^*) = 1$. Thus

$$\sigma^* = \sigma^*(\mu) = \begin{cases} 
(7 - \mu)/7, & 1/6 < \mu < 3/8, \\
(10 - 9\mu)/7, & 3/8 < \mu < 1/2, \\
(6 - \mu)/7, & 1/2 < \mu < 5/9.
\end{cases} \quad (19)$$

We take $\alpha = 1 - \mu + \epsilon_1$ where $\epsilon_1 > 0$ is a constant which will be chosen later. For a given small $\epsilon_2 > 0$ we take $\sigma_0 = \sigma_0(\mu)$ to be the largest $\sigma_0 < 1$ satisfying

$$\alpha G(\sigma_0) = 1 - \epsilon_2. \quad (20)$$

Since $G(\sigma)$ is a continuous decreasing function near $\sigma = 1$, for given $\delta > 0$ we have $0 < \sigma_0 - \sigma^* < \delta$ provided that $\epsilon_1$ is chosen sufficiently small in terms of $\delta$, and $\epsilon_2$ sufficiently small in terms of $\epsilon_1$.

**Lemma 8.** For any $\epsilon > 0$ we can choose $\delta, \epsilon_1, \epsilon_2$ so that

$$\int_{1/2}^{\sigma_0} x^{2\sigma} N(\sigma, x^{\sigma}) \log x \, d\sigma \ll x^{1 + F(\mu) + \epsilon} \quad as \ x \to \infty.$$  

**Proof.** We break the range of integration up into subintervals $[1/2, 3/4], [3/4, 7/9], [7/9, 11/14], [11/14, 53/56]$ and $[53/56, 1]$, stopping at $\sigma_0$. We use the upper bounds for $N(\sigma, x^{\sigma})$ provided by (18), and note that the exponent is a continuous function of $\alpha$ and $\sigma$, and so uniformly continuous on the region $1/2 < \sigma < 1, 0 < \alpha < 1$.

Putting $a = 1 - \mu$ it is sufficient to prove that for any $\epsilon > 0$

$$\int_{1/2}^{\sigma^*} x^{2\sigma} N(\sigma, x^{\sigma}) \log x \, d\sigma \ll x^{1 + F(\mu) + \epsilon/2} \quad as \ x \to \infty, \quad (21)$$

as the perturbations from $\sigma^*$ to $\sigma_0$ and $a$ to $\alpha$ can be absorbed in the exponent $\epsilon$ provided that $\epsilon_1, \epsilon_2$ and $\delta$ are sufficiently small as functions of $\epsilon$. For we estimate the integrand in (21) as

$$\ll x^{2\sigma x^{\sigma} G(\sigma)(1 - \sigma) + \eta \log x}.$$  

Between $\sigma^*$ and $\sigma_0$ the function $G(\sigma)(1 - \sigma)$ decreases so this is

$$\ll x^{2\sigma_0 + \sigma G(\sigma^*)(1 - \sigma^*) + \eta \log x} \ll x^{2\sigma_0 + (1 - \sigma^*) + \eta \log x} \ll x^{1 + \sigma^* + 2\delta + \eta \log x}$$

and $\sigma^*(\mu) < F(\mu)$ for $1/6 < \mu < 5/9$. Replacing $a = 1 - \mu$ by $\alpha = 1 - \mu + \epsilon_1$ in the estimates for the integrand will increase the exponents of $x$ by a factor no larger than $1 + 3\epsilon_1$, since $1 - \mu > 1/3$. Thus the lemma follows from the estimate (21).

The first interval contributes

$$\ll \max_{1/2 < \sigma < 3/4} x^{2\sigma + 3a(1 - \sigma)/(2 - \sigma) + \eta} \ll x^{3/2 + 3\alpha/5 + \eta},$$

since the exponent is an increasing function of $\sigma$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
The second interval contributes
\[ \leq \max_{3/4 \leq \sigma \leq 7/9} \left( x^{2\sigma + 3a(1 - \sigma)/(3\sigma - 1)} + \eta \right). \]

The exponent is a convex function of \( \sigma \) so the maximum occurs at an end-point of the interval. Hence the contribution of the second interval is
\[ \leq \max \left( x^{3/2 + 3a/5 + \eta}, x^{14/9 + a/2 + \eta} \right) \]
\[ \leq \begin{cases} 
  x^{3/2 + 3a/5 + \eta}, & \sigma > 5/9, \\
  x^{14/9 + a/2 + \eta}, & \sigma < 5/9.
\end{cases} \]

The third interval contributes
\[ \leq \max_{7/9 \leq \sigma \leq 11/14} x^{2\sigma + a(1 - \sigma)/(7\sigma - 5) + \eta} \]
and again the exponent is a convex function of \( \sigma \) so this is
\[ \leq \max \left( x^{11/7 + 3a/7 + \eta}, x^{11/7 + 3a/7 + \eta} \right) \]
since \( a = 1 - \mu > 2/9 \).

For \( 4/9 < \sigma < 1/2 \), \( a^* \in [7/9, 11/14] \) and the total contribution of the interval \([1/2, a^*]\) is \( \leq x^{14/9 + a/2 + \eta} \).

For \( 1/2 < \sigma < 5/8 \), \( a^* \in [11/14, 53/56] \) and the contribution of the interval \([11/14, a^*]\) is
\[ \leq \max \left( x^{11/7 + 3a/7 + \eta}, x^{11/7 + 3a/7 + \eta} \right) \leq x^{11/7 + 3a/7 + \eta}. \]

Now \( a^* = (1 + 9a)/7 \) so the total contribution of the interval \([1/2, a^*]\) is
\[ \leq \begin{cases} 
  x^{11/7 + 3a/7 + \eta}, & 1/2 < \sigma < 52/99, \\
  x^{11/7 + 3a/7 + \eta}, & 52/99 < \sigma < 5/8.
\end{cases} \]

For \( a > 5/8 \) the contribution of the interval \([1/2, 53/56]\) is
\[ \leq \begin{cases} 
  x^{53/28 + 3a/35 + \eta}, & a < 55/52, \\
  x^{3/2 + 3a/5 + \eta}, & a > 55/52.
\end{cases} \]

Now \( a^* = (6 + a)/7 \) and the contribution of the interval \([53/56, a^*]\) to the integral is
\[ \leq \max \left( x^{53/28 + 3a/35 + \eta}, x^{1 + \sigma^* + \eta} \right) \leq x^{1 + \sigma^* + \eta} = x^{13/9 + a/7 + \eta} \]
since \( a > 5/8 \). Thus for \( a > 5/8 \) the integral \([1/2, a^*]\) contributes
\[ \leq \begin{cases} 
  x^{3/2 + 3a/5 + \eta}, & a > 25/32, \\
  x^{13/9 + a/7 + \eta}, & a < 25/32,
\end{cases} \]
and this completes the proof of the lemma.

From Lemma 8, (11) and (13) we have
\[ \int_x^{2x} |\Delta(y)|^2 \, dy \leq x^{2\mu - 1 + \epsilon} \left( x^{1 + \sigma} + x^{1 + F(\mu)} \right) \leq x^{2\mu + F(\mu) + \epsilon} \] (22)
provided that \( \epsilon_1 \) is sufficiently small that \( \sigma < F(\mu) \), as \( F(\mu) > 1 - \mu \) for \( 1/6 < \mu < 5/9 \).
5. Estimation of a sum. Suppose that $x < p_m < 2x$ and $p_{m+1} - p_m > x^\eta$. Then for

$$x < p_m < y < p_m + d_m/3 < p_{m+1} < 2x$$

we have $y + y/U < p_{m+1}$. Then there are no primes in the interval $(y, y + y/U)$ so

$$
\psi(y + y/U) - \psi(y) = \sum_{y < n < y + y/U} \Lambda(n) \ll \log^2 x.
$$

Further for any $\eta < \epsilon_1$

$$E(y) - E(y + y/U) \ll x^{1-\eta} \log x \ll x^{1-\eta}/U.$$  

Observing that

$$|\exp(z) - 1| \leq |z|$$

for $\text{re } z < 1$ we have

$$
\sum_{(8)} \frac{(e^{\rho} - 1)y^\rho}{\rho} \ll \delta \sum_{(8)} x^\rho \ll U^{-1} \int_{\sigma_0}^1 x^\sigma N(\sigma, x^\sigma) \log x \, d\sigma.
$$

For $\sigma \in [\sigma_0, 99/100]$ we use Lemmas 3, 4 and 7, as necessary, to estimate $N(\sigma, x^\sigma)$. We have $\alpha G(\sigma) < 1 - \epsilon_2$ throughout the interval and choosing $\eta = \epsilon_2/300$,

$$
\int_{\sigma_0}^{99/100} x^\sigma N(\sigma, x^\sigma) \log x \, d\sigma \ll x^{1-\epsilon_2/100+\eta} \log x \ll x/\log x.
$$

Lemma 9. For some positive constant $d$, $\zeta(s) \neq 0$ in the region $\sigma = \text{re } s > 1 - d/(\log \tau)^{2/3}(\log \log \tau)^{1/3}$ where $\tau = |\text{im } s| + 2$.

This is Corollary 11.4 in Montgomery [15].

We take, with some suitable constant $D$,

$$\sigma_1 = 1 - D(\log x)^{-3/4}.$$  

Then $\zeta(s) \neq 0$ in the region $\sigma > \sigma_1$, $|t| < x^\sigma$ and so

$$
\int_{99/100}^{1} x^\sigma N(\sigma, x^\sigma) \log x \, d\sigma = \int_{99/100}^{\sigma_1} x^\sigma N(\sigma, x^\sigma) \log x \, d\sigma.
$$

For $\sigma \in [99/100, \sigma_1]$ we use Lemma 7 to estimate $N(\sigma, x^\sigma)$ and note that $\alpha G(\sigma) < 1 - \epsilon_2$ throughout the interval. Hence using (28),

$$
\int_{99/100}^{\sigma_1} x^\sigma N(\sigma, x^\sigma) \log x \, d\sigma \ll x \log^{4+1} x \int_{99/100}^{\sigma_1} x^{-\epsilon_2(1-\sigma)} \, d\sigma \ll x/\log x.
$$

Combining (26), (27) and (29) we obtain

$$
\sum_{(8)} \frac{(e^{\rho} - 1)y^\rho}{\rho} \ll x/ (U \log x).
$$

Substituting this, (24) and (25) in (7) we obtain that for those $y$ in the range (23),

$$|\Delta(y)| = | - y/U + O(x^{1-\eta}/U) + O(x/U \log x) + O(\log^2 x) | > x/2U.$$
Hence
\[ \int_x^{2x} |\Delta(y)|^2 \, dy \gg \frac{x^2}{U^2} \sum_{d_m \leq 2x, x < d_m < 2x} d_m, \]
and so, from (22),
\[ \sum d_m \ll x^{-2} U^2 \int_x^{2x} |\Delta(y)|^2 \, dy \ll x^{F(\nu) + \varepsilon}. \]

REFERENCES


DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF SHEFFIELD, SHEFFIELD S10 2TN, ENGLAND