On a sum associated with the Farey series.

by

C.L. Stewart

ABSTRACT

The purpose of this note is to estimate

$$S(N) = \sum_{i=1}^{N} q_i$$

where q_i denotes the smallest denominator possessed by a rational fraction which lies in the interval $(\frac{i-1}{N}, \frac{i}{N}]$. We prove that the estimates

$$1.20\, N^{3/2} < S(N) < 2.33\, N^{3/2}$$

are valid for N sufficiently large.

KEY WORDS & PHRASES: *Farey series.*

1. INTRODUCTION.

The Farey series F_N of order N is the sequence of fractions h/k with $(h,k) = 1$ and $1 \le h \le k \le N$ arranged in increasing order between 0 and 1. There are $\phi(k)$ fractions with denominator k in F_N and thus the number of terms in F_N is

(1) $$R(N) = \phi(1) + \phi(2) + \dots + \phi(N) = \frac{3}{\pi^2} N^2 + O(N\log N)$$

(see Theorem 330 of [2]). The purpose of this note is to estimate

$$S(N) = \sum_{i=1}^{N} q_i$$

where q_i denotes the smallest denominator possessed by a fraction from F_N which lies in the interval $(\frac{i-1}{N}, \frac{i}{N}]$.

We first observe that

(2) $$S(N) \ge \frac{2\pi}{3\sqrt{3}} N^{3/2} + O(N\log N)$$

for there can be at most $\phi(k)$ q_i's of size k in S(N) and thus

(3) $$S(N) \ge \sum_{k=1}^{t} k\,\phi(k)$$

for all t such that $\phi(1) + \phi(2) + \dots + \phi(t) \le N$. (Note that $\frac{2\pi}{3\sqrt{3}}$ is about 1.21.) On choosing t maximally we have by (1),

(4) $$t = \frac{\pi}{\sqrt{3}} N^{\frac{1}{2}} + O(\log N).$$

Furthermore

$$\sum_{k=1}^{t} k\,\phi(k) = t\sum_{k=1}^{t} \phi(k) - \left(\sum_{k=1}^{t-1} \phi(k) + \sum_{k=1}^{t-2} \phi(k) + \dots + \phi(1)\right)$$

and thus, again by (1), this

$$(5) \qquad = \frac{2}{\pi^2} t^3 + O(t^2 \log t).$$

And (2) now follows on combining (3), (4) and (5). A.E. Brouwer and J. van de Lune checked by means of a computer the value of $S(N)/N^{3/2}$ for a number of integers in the range 1,000 to 2500 and they found in all cases that $S(N)/N^{3/2}$ was less than 1.64 and larger than 1.58.
We shall prove

THEOREM. *For* N *sufficiently large*

$$S(N) < 2.33\, N^{3/2}.$$

We remark that we would expect the theorem to hold for all positive integers N. We in fact establish a result of the form

$$S(N) \le 2.328\, N^{3/2} + O(N^{7/5} \log N)$$

where the constant implicit in the O term is computable and thus the validity of the theorem for all integers N can be determined, in principle, by a finite amount of computation. We also observe that with some additional work our argument would doubtless yield a somewhat more precise estimate for the constant which precedes the main term in our estimate. Our proof of the above theorem depends upon two results of R.R. Hall concerning the distribution and the second moments of gaps in the Farey series.

The problem of obtaining appropriate estimates for the size of S(N) arose in connection with a problem of D. Kruyswijk and C. Schaap in combinatorial group theory. Independently of the author, D. Kruyswijk and H.G. Meijer have obtained a result of the form $S(N) = O(N^{3/2})$ and their argument, which is apparently entirely different from that given here, will be submitted for publication shortly. Lastly I would like to acknowledge the several useful observations concerning this work, made by Jan van de Lune, who first brought the above problem to my attention, and also by Jaap van der Woude.

2. PRELIMINARIES

We shall record here the two results of Hall which we require. We shall denote the difference between the r-th and r-1-st terms in the N-th Farey series by ℓ_r with the convention that $\ell_1 = 1/N$. Hall proves, theorem 1 of [1], that

LEMMA 1. *For some positive constant* C_0, *and for* $N \geq 2$,

$$\sum_{r=1}^{R(N)} \ell_r^2 < C_0 N^{-2} \log N.$$

Further he denotes by $\sigma_N(t)$, the number of ℓ_r from F_N for which $\ell_r > t/N^2$, and sets $\delta_N(t) = \sigma_N(t)/R(N)$. Hall proves that $\delta_N(t)$ is a distribution function. More precisely he proves

LEMMA 2. *If* $4 \leq t \leq N$ *and* $w = w(t)$ *is the smaller root of the equation* $w^2 = t(w-1)$, *then*

$$\delta_N(t) = 2t^{-1}(1-w+2\log w) + 0\,(t^{-1}N^{-1}\log N + N^{\alpha-2}), \tag{6}$$

where α *satisfies*

$$\sum_{n \leq x} \tau(n) - x\log x - x(2\gamma-1) = 0(x^{\alpha}),$$

$\tau(n)$ *denotes the number of divisors of* n *and* γ *is Eulers' constant.*

The work of a number of authors, Voronoi, Van der Corput, and more recently Chih and Kolesnik has resulted in a reduction of the exponent in the error term for Dirichlet's divisor problem from the elementary result $\alpha = \frac{1}{2}$, see Theorem 320 of Hardy and Wright [2], to $\alpha = \frac{12}{37} + \varepsilon$ for any $\varepsilon > 0$. To preserve the elementary character of our work we shall take $\alpha = \frac{1}{2}$ in Lemma 2 even though this results in a proof of our theorem which is slightly more complicated than that required when α is assumed to be $< \frac{1}{2}$.

We shall not apply Lemma 2 directly but shall instead use it to prove

LEMMA 3. *For* $4 \le t \le N$ *we have*

$$\sigma_N(t) \le \frac{24}{\pi^2}(2\log 2-1)\left(\frac{N}{t}\right)^2 + O(t^{-1}N\log N + N^{\frac{1}{2}}).$$

PROOF. For $t \ge 4$ the w occurring in (6) has the form

$$w = (t-t(1-4/t)^{\frac{1}{2}})/2$$

where the positive value of the square root is taken. We shall first show that

$$g(t) = t(2\log w-(w-1))$$

is a decreasing function of t for $t \ge 4$. This is equivalent to showing that the derivative $g'(t)$ is ≤ 0 for $t \ge 4$. We have

$$g'(t) = 2\log w - (w-1) + \left(\frac{2}{w}-1\right)\left(t\frac{dw}{dt}\right)$$

$$= 2\log w - (w-1) + 2-w + (w-2)/(w(1-\frac{4}{t})^{\frac{1}{2}})$$

$$= 2\log w - 2w + 2$$

and on observing that $\log(1+x) \le x$ for $x \ge 0$, and putting $x = w-1$ we conclude that

$$g'(t) \le 2(w-1) - 2w + 2 = 0$$

whenever $w > 1$. But

$$w = 1 + \frac{1}{t} + \frac{2}{t^2} + \ldots + \frac{C_n}{t^n} + \ldots$$

where the C_n are positive numbers and thus w is certainly ≥ 1 for $t \ge 4$. Therefore $g(t)$ is a decreasing function of t for $t \ge 4$ and so

$$(1-w+2\log w) \le 4(2\log 2-1)t^{-1}$$

whence, by Lemma 2 with $\alpha = \frac{1}{2}$, we have

(7) $$\delta_N(t) \le 8(2\log 2-1)t^{-2} + O(t^{-1}N^{-1}\log N + N^{-3/2})$$

for $4 \le t \le N$. The lemma now follows from (1) and (7) since

$$\sigma_N(t) = R(N)\,\delta_N(t).$$

3. PROOF OF THEOREM

We shall split the sum $S(N)$ into three parts which we shall estimate in turn: S_1 the sum of those q_i's $\le \sqrt{N}$, S_3 the sum of the t largest q_i's, where t will be specified later, and S_2 the sum of the remaining q_i's.

We first establish an upper bound for S_1. Put $V = [\sqrt{N}]$. We observe that if $\frac{h}{k}$ and $\frac{h'}{k'}$ are two terms in the Farey series F_V then

$$\left|\frac{h}{k} - \frac{h'}{k'}\right| \ge (kk')^{-1} \ge N^{-1}$$

and thus no two fractions from F_V are in the same interval $\left(\frac{i-1}{N},\frac{i}{N}\right]$ for any i. Thus to each fraction h/k in F_V there corresponds an interval $\left(\frac{i-1}{N},\frac{i}{N}\right]$ in which it is the fraction from F_N with smallest denominator and thus for which $q_i = k$. Now by the definition of the Farey series all the q_i's of size $\le \sqrt{N}$ must correspond to denominators of fractions from F_V. We therefore have

$$S_1 = \sum_{q_i \le \sqrt{N}} q_i = \sum_{k=1}^{V} k\,\phi(k)$$

and by (5) this

(8) $$= \frac{2}{\pi^2} N^{3/2} + O(N\log N).$$

Furthermore, it follows from (1) that S_1 is the sum over the

(9) $$\sum_{k=1}^{V} \phi(k) = \frac{3}{\pi^2} N + O(N^{\frac{1}{2}}\log N)$$

smallest q_i's in the sum S(N).

We shall estimate S_3, the sum of the t largest q_i's, next. Let $\Theta(M)$ denote the number of q_i's in the sum S(N) which are larger than M. It is readily verified that

$$S_3 \le Mt + \Theta(M) + \Theta(M+1) + \dots + \Theta(N)$$

where M is the value of the smallest q_i in S_3. Furthermore $\Theta(M+k) + \dots + \Theta(N)$ is certainly less than S_4 where

$$S_4 = \sum_{M+k \le q_i} q_i$$

so that

$$S_3 \le Mt + \Theta(M) \text{ t } \dots + \Theta(M+k) + S_4.$$

Now $\Theta(M)$ is a decreasing function of M hence

$$S_3 \le (M+r)t + \Theta(M+r) + \dots + \Theta(M+k) + S_4$$

for any positive integer r and thus

(10) $$S_3 \le (M+r)t + \int_{M+r-1}^{M+k} \Theta(M)\, dM + S_4.$$

The parameters t, M + r and M + k which we shall employ in (10) in order to minimize our estimate for S(N) depend on the estimate from above which we shall now obtain for $\Theta(M)$.

In order to bound $\Theta(M)$ from above it suffices to determine estimates from above for the number of gaps in F_M of size larger than j/N for

$j = 1,\ldots,k$ where $\frac{k}{N} \le \frac{1}{M} < \frac{k+1}{N}$; note that there can be no gaps of size $> M^{-1}$ in F_M. The number of gaps in F_M of size larger than j/N is precisely $\sigma_M(t)$ when $t/M^2 = j/N$, in other words when $t = jM^2/N$. Further we observe that

$$\theta(M) < \sigma_M(M^2/N) + \sigma_M(2M^2/N) + \ldots + \sigma_M(kM^2/N).$$

But now by Lemma 3 we have, for $M \ge 2\sqrt{N}$.

$$\theta(M) < C_0 \frac{N^2}{M^2}\left(1+\tfrac{1}{4}+\frac{1}{9}+\ldots+\frac{1}{k^2}\right) + \left(\frac{N\log M}{M}\left(1+\tfrac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{k}\right) + kM^{\frac{1}{2}}\right)$$

where $C_0 = \frac{24}{\pi^2}(2\log 2-1)$. Thus

$$\theta(M) < C_0 \frac{\pi^2}{6}\frac{N^2}{M^2} + 0\left(\frac{N\log M \log k}{M} + kM^{\frac{1}{2}}\right)$$

and since, by defintion, $k \le N/M$,

(11) $$\theta(M) < 4(2\log 2-1)\frac{N^2}{M^2} + 0\left(\frac{N(\log N)^2}{M} + \frac{N}{M^{\frac{1}{2}}}\right),$$

for all $M \ge 2\sqrt{N}$. Therefore, for some constant C_1, $\theta(2\sqrt{N}) < (2\log 2-1)N + C_1N^{\frac{3}{4}}$.

We now set

(12) $$t = (2\log 2-1)N + C_1N^{\frac{3}{4}}$$

so that in (10), M, the value of the smallest q_i in S_3, is $\le 2\sqrt{N}$. On putting $M + r = [2\sqrt{N}] + 1$ and choosing k so that $M + k \le N^{4/5} < M + k + 1$ we find from (10) that

(13) $$S_3 \le ([2\sqrt{N}]+1)\,t + \int_{[2\sqrt{N}]}^{N^{4/5}} \theta(M)\,dM + S_4$$

where S_4 is the sum over those q_i's $\geq N^{4/5}$.
Now by (11), the integral in (13) is

$$\leq \int_{[2\sqrt{N}]}^{N^{4/5}} 4(2\log 2-1)\frac{N^2}{M^2} + O(\frac{N(\log N)^2}{M} + \frac{N}{M^{\frac{1}{2}}})\, dM$$

which, upon evaluation, is found to be

(14) $$\leq 2(2\log 2-1)N^{3/2} + O(N^{7/5}).$$

Thus from (12), (13) and (14) we see that

(15) $$S_3 \leq 4(2\log 2-1)\, N^{3/2} + O(N^{7/5}) + S_4.$$

To complete our estimation of S_3 we must determine an upper bound for S_4.

Accordingly, we shall now prove that $S_4 = O(N^{7/5}\log N)$. If T is the number of terms q_i in S_4 then there must be T sections of length N^{-1} in the unit interval which contain no fractions from F_M for $M = [N^{4/5}]$. Therefore there must exist differences $\ell_{r_1},\ldots,\ell_{r_s}$ in F_M for which we can find positive integers $k_1,\ldots,k_s$ with $\ell_{r_i} \geq k_i/N$, $i = 1,\ldots,s$, and such that $k_1 + \ldots + k_s \geq T$. Thus we certainly have

(16) $$\sum_{i=1}^{s} \ell_{r_i}^2 \geq T\, N^{-2}.$$

On the other hand, by Lemma 1,

$$\sum_{r=1}^{R(M)} \ell_r^2 < C_0\, M^{-2} \log M$$

which is

(17) $$< C_2\, N^{-8/5} \log N$$

for a positive constant C_2. A comparison of (16) and (17) reveals that

$$T < C_2 N^{2/5} \log N.$$

Now S_4 is plainly $\le N \cdot T$ and thus $O(N^{7/5} \log N)$. It follows from (15), therefore, that

$$(18) \qquad S_3 \le 4(2\log 2 - 1)N^{3/2} + O(N^{7/5} \log N).$$

We are left now with S_2, the sum of those q_i's which are not in either S_1 or S_3. It follows from (9) and (12) that there are at most $C_3 N + O(N^{3/4})$ q_i's in S_2 where

$$(19) \qquad C_3 = 1 - \{(2\log 2 - 1) + \frac{3}{\pi^2}\}.$$

Further, by construction, all of these q_i's lie between $\sqrt{N}$ and $2\sqrt{N}$. A trivial upper bound for S_2 is plainly $2\sqrt{N}(C_3 N + O(N^{\frac{3}{4}}))$. We shall give an estimate for this sum which is only marginally less crude. Put $x = [2\sqrt{N}]$. We have

$$(20) \qquad S_2 \le \sum_{k=u}^{x} k\, \phi(k)$$

for some integer u satisfying

$$(21) \qquad \sum_{k=u}^{x} \phi(k) = C_3 N + O(N^{\frac{3}{4}}).$$

Now

$$\sum_{k=u}^{x} \phi(k) = \sum_{k=1}^{x} \phi(k) - \sum_{k=1}^{u-1} \phi(k)$$

which is, by (1),

$$= \frac{3}{\pi^2}(x^2 - u^2) + O(x \log x).$$

Therefore, it follows from (21) that

$$(22) \qquad u = \left(4-\frac{\pi^2}{3}C_3\right)^{\frac{1}{2}} N^{\frac{1}{2}} + O(N^{\frac{1}{4}}).$$

Furthermore we have by (5)

$$\sum_{k=u}^{x} k\,\phi(k) = \frac{2}{\pi^2}(x^3-u^3) + O(x^2\log x)$$

and thus we may deduce from (20) and (22) that

$$S_2 \le \left\{\frac{16}{\pi^2} - \frac{2}{\pi^2}\left(4-\frac{\pi^2}{3}C_3\right)^{3/2}\right\} N^{3/2} + O(N^{5/4})$$

and by (19) this is

$$(23) \qquad \le \cdot 5783\, N^{3/2} + O(N^{5/4}).$$

Finally we have

$$S(N) = S_1 + S_2 + S_3$$

which by (8), (18) and (23) is

$$\le \left(\frac{2}{\pi^2} + 4(2\log 2-1) + .5783\right)N^{3/2} + O(N^{7/5}\log N)$$

$$\le 2.328\, N^{3/2} + O(N^{7/5}\log N).$$

The theorem now follows directly.

REFERENCES

[1] HALL, R.R., *A note on the Farey series,* J. London Math. Soc. (2), 2(1970), pp. 139-148.

[2] HARDY, G.H. & E.M. WRIGHT, *An Introduction to the Theory of Numbers*, 4th ed., Oxford 1960.

<u>Remark</u>: M.R. Best has computed values of $S(N)$ for N up to 5,000,000 and his data suggest that $\lim_{N\to\infty} S(N)/N^{3/2}$ exists and is equal to 1.62 ..., a value suspiciously close to $(4/\pi)^2$.