ILLINOIS JOURNAL OF MATHEMATICS
Volume 24, Number 1, Spring 1980

THE FRACTIONAL PARTS OF THE BERNOULLI NUMBERS

BY
PAUL ERDŐS AND SAMUEL S. WAGSTAFF, JR.

Abstract

The fractional parts of the Bernoulli numbers are dense in the interval (0, 1). For every positive integer k, the set of all m for which B_{2m} has the same fractional part as B_{2k} has positive asymptotic density.

1. Introduction

The Bernoulli numbers are the coefficients B_n of the power series

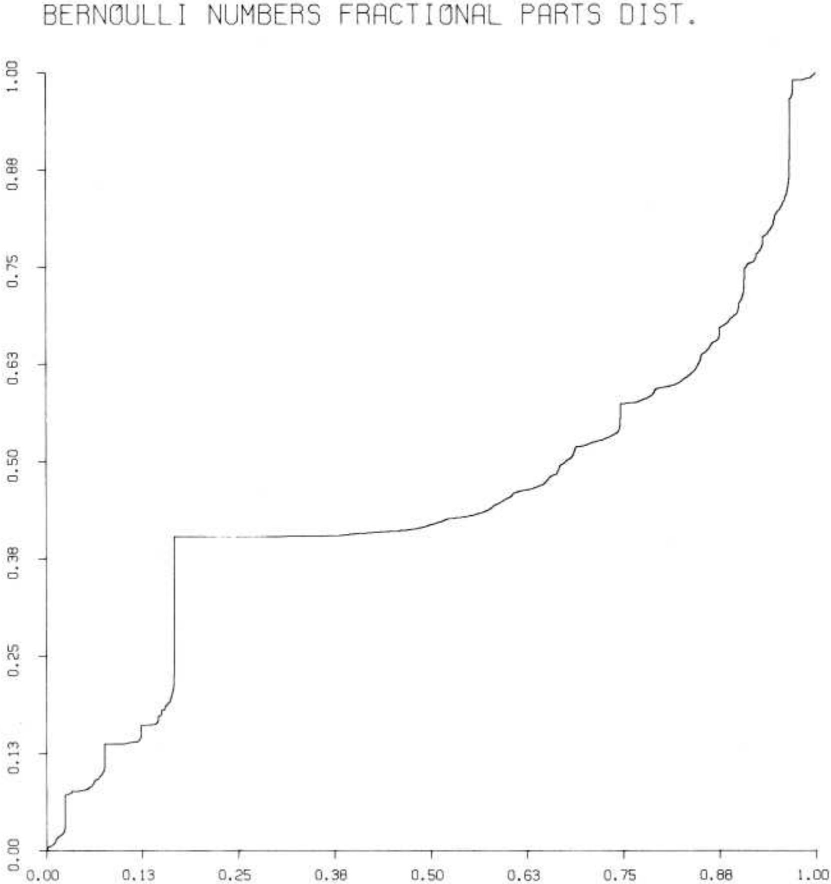
$$t/(e^t - 1) = \sum_{n=0}^{\infty} B_n t^n/n!$$

It is well known that they are rational numbers and that $B_n = 0$ for odd $n > 1$. We have $B_2 = 1/6$, $B_4 = -1/30$, $B_6 = 1/42$, etc. The fractional parts $\{B_{2k}\}$ may be computed easily by the von Staudt–Clausen theorem, which says that $B_{2k} + \sum 1/p$ is an integer, where the sum is taken over all primes p for which $(p-1) \mid 2k$.

Several years ago one of us computed $\{B_{2k}\}$ for $2 \le 2k \le 10000$ and noted two curious irregularities in their distribution: (1) There were large gaps, e.g., the interval [0.167, 0.315], which contained none of these numbers. More computation showed that the gaps tend to be filled in if one used enough $2k$'s. We prove in Section 2 that the fractional parts are dense in (0, 1). (2) A few rationals appeared with startling frequency. For example, 1/6 occurred 834 times among the 5000 numbers, that is, almost exactly 1/6 of the time. When the calculation was extended to $2k = 100000$ it was found that the fraction of $m \le x$ for which $\{B_{2m}\} = 1/6$ remained close to 1/6 for $100 \le x \le 50000$. We prove in Section 4 that for every $k \ge 1$, the set of all m for which $\{B_{2m}\} = \{B_{2k}\}$ has positive asymptotic density. The set of such m was known to be infinite (see p. 93 of [6]).

Since our proof gives no indication of the value of the asymptotic density, we list in a table the $\{B_{2k}\}$ which occur most frequently for $2k \le 100000$. Let $\mathscr{P}_{2k}$ denote $\{p : p \text{ is prime and } p - 1 \mid 2k\}$. The table shows $\sum_{p \in \mathscr{P}_{2k}} 1/p$, the first $2k$ for

Received January 13, 1978.



which $\mathscr{P}_{2k}$ appears, $\{B_{2k}\}$, the number and density of $2m \leq 100000$ with $\mathscr{P}_{2m} = \mathscr{P}_{2k}$, and the elements of $\mathscr{P}_{2k}$. (Note that $\{B_{2k}\} = \{B_{2m}\}$ if and only if $\mathscr{P}_{2k} = \mathscr{P}_{2m}$, by the von Staudt–Clausen theorem.)

Generally speaking, $\{B_{2k}\}$ occurs more often when there are fewer and smaller primes in $\mathscr{P}_{2k}$. Not every finite set of primes which includes 2 and 3 can be a $\mathscr{P}_{2k}$. For instance, if 5, 7 and 11 are in the set, then it must contain 61 as well. Likewise, if the set contains 13, then 5 and 7 must be in it, too.

We also show the graph of the distribution function

$$F_x(z) = x^{-1} \cdot (\text{the number of } m \leq x \text{ for which } \{B_{2m}\} < z)$$

for $x = 10000$ and $0 \leq z \leq 1$. The graph is virtually indistinguishable from those of F_{1000} and F_{5000}. The size of the vertical jump at $z = \{B_{2k}\}$ approximates the asymptotic density of the set of m for which $\{B_{2m}\} = \{B_{2k}\}$. We show in Section 4 that the limiting distribution $F(z) = \lim_{x\to\infty} F_x(z)$ exists. We also mention several open questions at the end.

2. The fractional parts are dense in (0, 1)

Let $S(2m) = \sum_{p\in\mathscr{P}_{2m}} 1/p$. We want to prove that the $\{B_{2k}\}$ are dense in (0, 1). According to the von Staudt–Clausen theorem, the denominator of B_{2k} (in lowest terms) is $\prod_{p\in\mathscr{P}_{2k}} p$. Hence $\{B_{2k}\}$ is never zero, and $\{B_{2k}\} = 1 - \{S(2k)\}$. Thus it suffices to prove that the fractional parts of the $S(2k)$ are dense in (0, 1). Note that $S(2k) \geq 5/6$ because both $2 - 1$ and $3 - 1$ divide every $2k$, and $1/2 + 1/3 = 5/6$.

THEOREM 1. *For all $\alpha \geq 5/6$ and $\varepsilon > 0$, there are infinitely many even integers $2m$ for which $|S(2m) - \alpha| < \varepsilon$.*

Proof. Let p_n denote the nth prime. Let r be a large integer. (Later we will choose r sufficiently large depending on ε.) Let $A_s = 2p_r p_{r+1} \cdots p_{r+s}$. If $p \equiv -1 \pmod{p_2 p_3 \cdots p_{r-1}}$, and $p - 1$ is squarefree, then $(p-1) \mid A_s$ for all sufficiently large s. It follows from the prime number theorem for arithmetic progressions and a simple sieve argument that $\sum 1/p$ diverges, where p runs over primes $p \equiv -1 \pmod{p_2 p_3 \cdots p_{r-1}}$ with $p - 1$ squarefree. Thus we can choose s so that $S(A_s) > \alpha$. We prove the theorem by removing the factors p_{r+s}, p_{r+s-1}, etc., from A_s, one by one, until $S(A_s)$ is close to α. It suffices to show that $S(A_s) - S(A_{s-1}) < \varepsilon$ provided p_r is large enough.

Let $d_1, \ldots, d_k$ be all of the divisors of A_{s-1}. Write q for p_{r+s}. Then $d_1, \ldots, d_k$, $qd_1, \ldots, qd_k$ are all of the divisors of A_s. Thus (σ denotes the sum of divisors function)

$$S(A_s) - S(A_{s-1}) = \sum_{\substack{p-1 \mid A_s \text{ but} \\ p-1 \nmid A_{s-1}}} \frac{1}{p} = \sum_{\substack{p-1 = qd_i \\ \text{for some } i}} \frac{1}{p} = \sum_{\substack{i=1, \\ 1+qd_i \\ \text{is prime}}}^{k} \frac{1}{1+qd_i}$$

$$\leq \frac{1}{q} \sum_{i=1}^{k} \frac{1}{d_i} = \frac{1}{q} \sum_{i=1}^{k} \frac{d_i}{A_{s-1}} = \frac{\sigma(A_{s-1})}{qA_{s-1}}$$

$$< \frac{c_1}{q} \log\log A_{s-1} < \frac{c_2}{q} \log(p_{r+s-1} - p_r)$$

$$< \frac{c_2 \log q}{q} \leq \frac{c_2 \log p_r}{p_r} < \varepsilon$$

for large enough r and some absolute constants c_1, c_2. The estimates of $\sigma(A_{s-1})$ and $\log A_{s-1}$ follow from Theorems 323 and 414 of [6], respectively. This completes the proof.

3. A result on divisibility by $p-1$

In this section we prove that numbers which have a large divisor of the form $p-1$ are rare. This result (Theorem 2) is the essential ingredient in our proof of Theorem 3, and has some independent interest as well.

THEOREM 2. *For each $\varepsilon > 0$, there is a $T = T(\varepsilon)$ so that if $x > T$, then the number of $m \le x$ which have a divisor $p-1 > T$, with p prime, is less than εx.*

Notation. The counting function of a set of integers will be denoted by the corresponding Latin letter, e.g., $A(n)$ is the number of $a \in \mathscr{A}$ with $1 \le a \le n$. Let $\Omega_R(m)$ be the number of primes $\le R$ which divide m (counting multiplicity). Write $\Omega(m)$ for $\Omega_m(m)$.

Proof of Theorem 2. Let T be a fixed large number. Let $\mathscr{A}$ be the set of all natural numbers which have a divisor $p-1 > T$, with p prime. We will prove the theorem by showing that there are positive constants c_3 and μ such that $A(x) < c_3 x/\log^\mu T$ for all sufficiently large T and x.

Every element m of $\mathscr{A}$ can be written in the form $m = (p-1)n$, where p is prime and $p-1 > T$. We separate the elements of $\mathscr{A}$ into three classes, depending on the number of prime factors of $p-1$ and of n. Some elements may appear in more than one class, but this does not matter, since we require only an upper bound on $A(x)$. The classes are defined by

$$\Omega(p-1) < (2/3)\log\log p, \tag{1}$$

$$\Omega_p(n) < (2/3)\log\log p, \tag{2}$$

and

$$\text{both} \quad \Omega(p-1) \ge (2/3)\log\log p \quad \text{and} \quad \Omega_p(n) \ge (2/3)\log\log p. \tag{3}$$

Lemmas 1, 2, and 4 will estimate the counting functions of these three classes.

LEMMA 1. *There are positive constants c_4, δ, y_0 such that if $x > T \ge y_0$, then the number $D_1(x)$ of $m \le x$ for which there is a prime $p > T+1$ with $(p-1)|m$ and $\Omega(p-1) < (2/3)\log\log p$ satisfies*

$$D_1(x) < c_4 x/\log^\delta T.$$

Proof. It was shown in [2] that the number of primes $p \le y$ with

$$\Omega(p-1) < (2/3)\log\log y \quad \text{is} \quad O(y/\log^{1+\delta} y) \quad \text{provided} \quad y > y_0.$$

For each such p, there are $[x/(p-1)]$ multiples of $p-1$ which are $\leq x$. Thus, for $x > T \geq y_0$, we have

$$D_1(x) < \sum_{\substack{p \text{ prime} \\ p > T+1 \\ \Omega(p-1) < (2/3)\log\log p}} \left[\frac{x}{p-1}\right]$$

$$\ll \sum_{\substack{p \text{ prime} \\ p > T+1 \\ \Omega(p-1) < (2/3)\log\log p}} \frac{x}{p}$$

$$\ll \int_T^\infty \frac{x}{p} \frac{dp}{\log^{1+\delta} p}$$

$$= \frac{x}{\delta \log^\delta T}.$$

LEMMA 2. *There are positive constants c_5, η such that if $x > T > e$, then the number $D_2(x)$ of $m \leq x$ for which there is a prime $p > T+1$ with $(p-1) \mid m$ and $\Omega_p(m/(p-1)) < (2/3) \log \log p$ satisfies*

$$D_2(x) < c_5 x/\log^\eta T.$$

Proof. According to Theorem 5.9 of [7], there is a positive constant η such that the number of $n \leq y$ for which $\Omega_R(n) < (2/3) \log \log R$ is $O(y/\log^\eta R)$, provided $y \geq 1$. For each prime p between $T+1$ and $x+1$, we apply the theorem with $R = p$, $n = m/(p-1)$, and $y = x/(p-1)$. Summing the estimates, we find

$$D_2(x) \ll \sum_{\substack{p \text{ prime} \\ T+1 < p < x+1}} \frac{\frac{x}{p-1}}{\log^\eta p}$$

$$\ll x \sum_{\substack{p \text{ prime} \\ p > T+1}} \frac{1}{p \log^\eta p}$$

$$\ll x \int_{(T/\log T)}^\infty \frac{dt}{t \log t \log^\eta (t \log t)}$$

$$\ll \frac{x}{\log^\eta (T/\log T)}.$$

The lemma follows since $T/\log T \gg \sqrt{T}$.

LEMMA 3. *There are positive constants c_6, λ, T_0 such that if $x > T > T_0$, then the number of $m \leq x$ for which there is some $t > T$ with $\Omega_t(m) \geq (4/3) \log \log t$ is less than $c_6 x/\log^\lambda T$.*

Proof. By Norton's Theorem 5.12 [7], there are positive constants c_7 and η such that for every t, the number of $m \le x$ with $\Omega_t(m) \ge (7/6) \log \log t$ is $< c_7 \, x/\log^\eta t$.

Now let $t_i = \exp(i^{2/\eta})$. We apply Norton's theorem to those $t_i > T$. Since

$$\sum_{\substack{i=1 \\ t_i > T}}^{\infty} \frac{1}{\log^\eta t_i} = \sum_{\substack{i=1 \\ i > \log^{\eta/2} T}}^{\infty} i^{-2} < \frac{1}{1 + \log^{\eta/2} T},$$

we see that there is a positive c_6 such that the number of $m \le x$ for which $\Omega_{t_i}(m) \ge (7/6) \log \log t_i$ for some $t_i > T$ is less than $c_6 \, x/\log^{\eta/2} T$.

Now let $m \le x$, and suppose there is a t with $\Omega_t(m) \ge (4/3) \log \log t$. If $t_{i-1} \le t < t_i$ and i is large enough, then we have

$$\Omega_{t_i}(m) \ge \Omega_t(m) \ge (4/3) \log \log t \ge (4/3) \log \log t_{i-1} \ge (7/6) \log \log t_i.$$

Thus, for sufficiently large i (or T), the number of such $m \le x$ does not exceed the number of $m \le x$ for which $\Omega_{t_i}(m) \ge (7/6) \log \log t_i$ for some $t_i > T$. We showed above that the latter number is less than $c_6 \, x/\log^\lambda T$, with $\lambda = \eta/2$.

Remark. In fact a much sharper statement than Lemma 3 is announced in [3]. A modification of our proof would give the stronger result, which can also be demonstrated by the methods of probabilistic number theory.

LEMMA 4. *There are positive constants c_8, λ, T_0 such that if $x > T > T_0$, then the number $D_3(x)$ of $m \le x$ for which there is a prime $p > T + 1$ with $(p-1) \mid m$,*

$$\Omega(p-1) \ge (2/3) \log \log p \quad \textit{and} \quad \Omega_p(m/(p-1)) \ge (2/3) \log \log p$$

satisfies

$$D_3(x) < c_8 x/\log^\lambda T.$$

Proof. The hypotheses imply $\Omega_p(m) \ge (4/3) \log \log p$, so that this lemma is immediate from the preceding one.

Theorem 2 now follows at once from Lemmas 1, 2, and 4 because $A(x) \le D_1(x) + D_2(x) + D_3(x)$.

4. The asymptotic density is positive

We wish to show that for every $k \ge 1$, the set of all m for which $\{B_{2m}\} = \{B_{2k}\}$ has positive asymptotic density. In view of the von Standt–Clausen theorem, this is equivalent to:

THEOREM 3. *For every $k \ge 1$, the set of all m for which $\mathscr{P}_{2m} = \mathscr{P}_{2k}$ has positive asymptotic density.*

We introduce a little more notation. Let LCM (a, b) denote the least common multiple of a and b. Write $\mathscr{B}(\mathscr{A})$ for the set of all positive multiples of elements of $\mathscr{A}$.

Proof of Theorem 3. Let $2k$ be given. Let $\mathscr{K}$ be the set of all positive multiples of $2k$. Let $\mathscr{K}_0$ be the set of all m such that $\mathscr{P}_m = \mathscr{P}_{2k}$. We may assume without loss of generality that $2k$ is the least element of $\mathscr{K}_0$. Note that this just says that $2k$ is the least common multiple of all of the numbers $p-1$ with $p \in \mathscr{P}_{2k}$. Thus $\mathscr{K}_0 \subset \mathscr{K}$. Let $\mathscr{A}$ be the set of all LCM $(p-1, 2k)$ for which p is a prime not in $\mathscr{P}_{2k}$ (i.e., $(p-1) \nmid 2k$.) Then $\mathscr{K}$ is the disjoint union of $\mathscr{K}_0$ and $\mathscr{B}(\mathscr{A})$. Write the elements of $\mathscr{A}$ in increasing order as $a_1 < a_2 < \cdots$.

We will use Theorem 2 with $\varepsilon = 1/4k$; this gives us T. Each a_i in $\mathscr{A}$ was formed as $a_i = \text{LCM}\,(p_i - 1, 2k)$ for some prime p_i with $p_i - 1 \le a_i \le 2k(p_i - 1)$. Choose the least r for which $a_r \ge 2kT$. Then, for $i \ge r$, we have $p_i - 1 \ge a_i/2k \ge T$. Let $\mathscr{A}_1 = \{a_1, \ldots, a_r\}$ and $\mathscr{A}_2 = \mathscr{A} - \mathscr{A}_1$. We have $A_2(x) \le A(x) \le \pi(x+1) \le 2x/\log x$ for all large x. Therefore, by [4] or Theorem 14, p. 262 of [5], $\mathscr{B}(\mathscr{A})$ and $\mathscr{B}(\mathscr{A}_2)$ possess asymptotic density. Clearly $\mathscr{B}(\mathscr{A}_1)$ has asymptotic density, too. By Theorem 2, we have (with d denoting asymptotic density)

$$d(\mathscr{B}(\mathscr{A}_2)) \le 1/4k. \tag{1}$$

Let $T_n(q_1, \ldots, q_s)$ denote the asymptotic density of the sequence consisting of all those multiples of n which are *not* divisible by any q_i $(i = 1, \ldots, s)$. Behrend [1] (see also Lemma 5, p. 263 of [5]) proved that

$$T_1(q_1, \ldots, q_s)T_1(q_{s+1}, \ldots, q_{s+t}) \le T_1(q_1, \ldots, q_{s+t})$$

always. A slight modification of his proof yields the relativized version

$$T_n(q_1, \ldots, q_s)T_n(q_{s+1}, \ldots, q_{s+t}) \le \frac{1}{n}\, T_n(q_1, \ldots, q_{s+t}).$$

We apply this inequality with $n = 2k$ to the elements of $\mathscr{A}$. For the r chosen above, and any s, we obtain

$$T_{2k}(a_1, \ldots, a_r)T_{2k}(a_{r+1}, \ldots, a_{r+s}) \le \frac{1}{2k}\, T_{2k}(a_1, \ldots, a_{r+s}). \tag{2}$$

We have

$$T_{2k}(a_1, \ldots, a_r) = \frac{1}{2k} - d(\mathscr{B}(\mathscr{A}_1)) > 0. \tag{3}$$

(The positivity may be proved easily by induction on r using (2) with $s = 1$.) Furthermore,

$$\lim_{s\to\infty} T_{2k}(a_{r+1}, \ldots, a_{r+s}) = \frac{1}{2k} - d(\mathscr{B}(\mathscr{A}_2)) \tag{4}$$

because $d(\mathscr{B}(\mathscr{A}_2))$ exists. (See also Theorem 12, p. 258 of [5].) Likewise,

$$\lim_{s\to\infty} T_{2k}(a_1, \ldots, a_{r+s}) = \frac{1}{2k} - d(\mathscr{B}(\mathscr{A})) = d(\mathscr{K}_0). \tag{5}$$

Formulas (1)–(5) now imply

$$\frac{1}{2k}d(\mathscr{K}_0) \geq \left(\frac{1}{2k} - d(\mathscr{B}(\mathscr{A}_1))\right)\left(\frac{1}{2k} - d(\mathscr{B}(\mathscr{A}_2))\right) > 0,$$

which is Theorem 3.

COROLLARY. *The distribution function* $F(z) = \lim_{x\to\infty} F_x(z)$ *exists and is a jump function. The convergence is uniform and the sum of the heights of the jumps of* F *is* 1.

5. Some open questions

It might be interesting to study $S(n) = \sum_{(p-1)|n} 1/p$. We proved that the range of S is dense in $[5/6, \infty)$ and S has a distribution function which is a jump function. Can one estimate $M(x) = \max_{n<x} S(n)$? It is likely that $M(x)/\log\log x \to 0$, but that $M(x)/\log\log\log x \to \infty$. Prachar [8] has shown that the related function $d_1(n) = \sum_{(p-1)|n} 1$ has average order $\log\log n$ and that $d_1(n) > n^{c/(\log\log n)^2}$ for some $c > 0$ and infinitely many n.

More generally, let $a_1 < a_2 < \cdots$ be a sequence of integers and $b_1, b_2, \ldots$ be a sequence of positive real numbers. (In our case, $a_i = p_i - 1$ and $b_i = p_i$.) Define $f_A(n) = \sum_{a_i|n} 1/b_i$. When does it happen that the density of integers m for which $f_A(m) = f_A(n)$ is positive? This holds at least when the a_i's have this property:

(P) For all n, the set of those m which are divisible by precisely the same a_i's as n has positive density.

Property (P) does not hold for all sequences. It fails, for example, for $a_i = 2i$. Two related problems are to characterize the sequences of a_i's which have property (P) and to study the distribution of $f_A(n)$.

Now consider the fractional parts $\{B_{2k}\}$ with $2k \leq x$. How many distinct values are assumed? Theorems 2 and 3 answer $o(x)$. On the other hand, a lower bound is $(x/\log x)(1 + o(1))$ because $\{B_{p-1}\} \neq \{B_{q-1}\}$ when p and q are distinct primes. The number of distinct $\{B_{2k}\}$ with $2k \leq x$ is 284, 566, 2612, and 5131 for $x = 1000, 2000, 10000, 20000$, respectively.

We remarked in the introduction that not every finite set of primes can be a $\mathscr{P}_{2k}$. Let $2, 3, \ldots, p_r$ be the set of primes $\leq x$. How many of the 2^r subsets can be $\mathscr{P}_{2k}$'s?

Let δ_{2k} be the asymptotic density of the set of $2m$ with $\{B_{2m}\} = \{B_{2k}\}$. Can we ever have $\delta_{2k} = \delta_{2m}$ for $\{B_{2k}\} \neq \{B_{2m}\}$? Clearly $\delta_{2k} < 1/2k$. What is a positive lower bound for δ_{2k}? Is $\{2k\delta_{2k}\}$ dense in $(0, 1)$? Probably one could show that δ_2 is the greatest δ_{2k}.

Acknowledgment. We are grateful to Mr. Richard Sunseri, who discovered a mistake in an earlier version of this paper.

Table of $\{B_{2k}\}$ which appear at least 150 times among $\{B_2\}, \{B_4\}, \ldots, \{B_{100000}\}$

$\sum_{p \text{ prime}, (p-1)\mid 2k} \frac{1}{p}$	First $2k$	$\{B_{2k}\}$	Frequency to 100000	Density to 100000	Primes p $(p-1)\mid 2k$
0.833333	2	0.166667	7992	0.15984	2, 3
0.845382	82	0.154618	150	0.00300	2, 3, 83
0.850282	58	0.149718	235	0.00470	2, 3, 59
0.854610	46	0.145390	261	0.00522	2, 3, 47
0.876812	22	0.123188	566	0.01132	2, 3, 23
0.924242	10	0.075758	1080	0.02160	2, 3, 11
0.976190	6	0.023810	1371	0.02742	2, 3, 7
1.028822	18	0.971178	397	0.00794	2, 3, 7, 19
1.033333	4	0.966667	3423	0.06846	2, 3, 5
1.052201	52	0.947799	164	0.00328	2, 3, 5, 53
1.067816	28	0.932184	309	0.00618	2, 3, 5, 29
1.076812	44	0.923188	160	0.00320	2, 3, 5, 23
1.092157	16	0.907843	713	0.01426	2, 3, 5, 17
1.124242	20	0.875758	289	0.00578	2, 3, 5, 11
1.253114	12	0.746886	495	0.00990	2, 3, 5, 7, 13

References

1. F. A. Behrend, *Generalization of an inequality of Heilbronn and Rohrbach*, Bull. Amer. Math. Soc., vol. 54 (1948), pp. 681–684.
2. P. Erdős, *On the normal number of prime factors of $p-1$ and some related problems concerning Euler's ϕ-function*, Quart. J. Math., vol. 6 (1935), pp. 205–213.
3. P. Erdős, *On the distribution function of additive functions*, Ann. of Math. (2), vol. 47 (1946), pp. 1–20.
4. ———, *On the density of some sequences of integers*, Bull. Amer. Math. Soc., vol. 54 (1948), pp. 685–692.
5. H. Halberstam and K. F. Roth, *Sequences*, vol. I, Clarendon Press, Oxford, 1966.
6. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 4th ed., Clarendon Press, Oxford, 1960.
7. Karl K. Norton, *On the number of restricted prime factors of an integer I*, Illinois J. Math., vol. 20 (1976), pp. 681–705.
8. K. Prachar, *Über die Anzahl Teiler einer natürlichen Zahl, welche die Form $p-1$ haben*, Monatsh, Math., vol. 59 (1955), pp. 91–97.

University of Colorado
Boulder, Colorado
and
Hungarian Academy of Science
Budapest, Hungary
University of Illinois
Urbana, Illinois