ON THE LARGEST PRIME FACTOR OF NUMERATORS OF BERNOULLI NUMBERS

ATTILA BÉRCZES AND FLORIAN LUCA

ABSTRACT. We prove that for most n, the numerator of the Bernoulli number B_{2n} is divisible by a large prime.

2000 Mathematics Subject Classification: Primary 11B68

1. Introduction

For a positive integer n, we write $\omega(n)$ for the number of distinct prime factors of n. Let $\{B_n\}_{n\ge 0}$ be the sequence of Bernoulli numbers given by $B_0 = 1$ and

$$B_n = 1 - \sum_{k=0}^{n-1} \binom{n}{k} \frac{B_k}{n-k+1} \qquad \text{for all} \quad n \ge 1.$$

Then $B_1 = -1/2$ and $B_{2n+1} = 0$ for all $n \ge 0$. Furthermore, we have $(-1)^{n+1} B_{2n} > 0$. Write $B_{2n} =: (-1)^{n+1} C_n/D_n$ with coprime positive integers C_n and D_n. The denominator D_n is well-understood by the von Staudt–Clausen theorem which asserts that

$$D_n = \prod_{p-1 \mid 2n} p. \tag{1}$$

As for C_n, it was proved in [3] that the estimate

$$\omega\left(\prod_{n\le x} C_n\right) \ge (1+o(1)) \frac{\log x}{\log\log x} \qquad \text{holds as} \quad x \to \infty.$$

Here, we look at the largest prime factor of C_n. For a positive integer m we put $P(m)$ for the largest prime factor of m.

The research was supported in part by project SEP-CONACyT 79685, by grants T67580 and T75566 of the Hungarian National Foundation for Scientific Research. The work is supported by the TÁMOP 4.2.1./B-09/1/KONV-2010-0007 project. The project is implemented through the New Hungary Development Plan, co-financed by the European Social Fund and the European Regional Development Fund. F. L. worked on this project while he visited the Institute of Mathematics of the University of Debrecen, Hungary in August 2011. He thanks the members of that department for their hospitality.

Theorem 1. *The inequality*

$$P(C_n) > \frac{1}{4}\log n$$

holds for most positive integers n.

Here and in what follows, we use the symbols O and o with their usual meaning. We also use $c_1, c_2, \ldots$ for computable positive constants and x_0 for a large real number, not necessarily the same from one occurrence to the next.

Proof. We let x be large. Put

$$\mathcal{M}(x) := \{x/2 \le n \le x : P(C_n) \le (1/4)\log x\}. \tag{2}$$

Put $y := x^{\log\log\log x/\log\log x}$. We let

$$\mathcal{L}_1(x) := \{n \le x : P(n) \le y\}. \tag{3}$$

It is known (see Chapter III.5 in [5]), that

$$\#\mathcal{L}_1(x) = x\exp(-(1+o(1))u\log u), \quad \text{where} \quad u := \frac{\log x}{\log y}.$$

Since for us $u = \log\log x/\log\log\log x$, we get easily that

$$\#\mathcal{L}_1(x) = O\left(\frac{x}{(\log x)^{1/2}}\right). \tag{4}$$

We let $\tau(m)$ stand for the number of divisors of m. We put

$$\mathcal{L}_2(x) := \{n \le x : \tau(n) > (\log x)^2\}. \tag{5}$$

Since

$$\sum_{n\le x}\tau(n) = O(x\log x),$$

(see Theorem 320 on Page 347 in [2]), it follows easily that

$$\#\mathcal{L}_2(x) = O\left(\frac{x}{\log x}\right). \tag{6}$$

Let

$$\mathcal{L}_3(x) := \{n \ge x : p-1 \mid 2n \text{ for some prime } p \text{ with } P(p-1) > y\}. \tag{7}$$

The proof of Theorem 1.1 in [1] shows that

$$\#\mathcal{L}_3(x) = O\left(\frac{x}{(\log x)^{0.05}}\right). \tag{8}$$

From now on, we look at integers n in

$$\mathcal{N}(x) := \mathcal{M}(x)\setminus \cup_{i=1}^{3}\mathcal{L}_i(x). \tag{9}$$

Put $z := (\log x)^2$ and let I be an arbitrary interval in $[x/2, x]$ of length at most z. Put $T := (1/4)\log x$ and put $K := \pi(T)$. We show that for $x > x_0$, I contains less than $K+3$ numbers from $\mathcal{N}(x)$. Assume first that we have proved this and let us see how to finish the argument. Then

$$\begin{aligned}\#\mathcal{N}(x) &\le \left(\left[\frac{x-x/2}{(\log x)^2}\right]+1\right)(K+2) = O\left(\frac{x}{(\log x)^2}\cdot\frac{T}{\log T}\right)\\ &= O\left(\frac{x}{\log x\log\log x}\right),\end{aligned}\tag{10}$$

which together with estimates (4), (6), (8) shows that

$$\#\mathcal{M}(x) \le \#\mathcal{L}_1(x)+\#\mathcal{L}_2(x)+\#\mathcal{L}_3(x)+\#\mathcal{N}(x) = O\left(\frac{x}{(\log x)^{0.05}}\right).\tag{11}$$

The desired estimate now follows by replacing x with $x/2$, then with $x/4$, etc., and summing up the resulting estimates (11).

It remains to prove that indeed I cannot contain $K+3$ numbers from $\mathcal{N}(x)$ for $x > x_0$. Assume that it does and let them be $n_1 < n_2 < \cdots < n_{K+3}$. Put $\lambda_i := n_i - n_1$ for $i = 1, \ldots, K+3$. Then $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{K+3} \le z$. Let $n = n_i$ for some $i = 1, \ldots, K+3$. We use the formula

$$\zeta(2n) = (-1)^{n+1}B_{2n}\frac{(2\pi)^{2n}}{2(2n)!} = \frac{C_n(2\pi)^{2n}}{D_n 2(2n)!},$$

as well as the aproximation

$$\zeta(2n) = 1+\frac{1}{2^{2n}}+\frac{1}{3^{2n}}+\cdots = 1+O\left(\frac{1}{2^{2n}}\right),$$

to get that

$$C_n = D_n\frac{2(2n)!}{(2\pi)^{2n}}\zeta(2n) = D_n\frac{2(2n)!}{(2\pi)^{2n}}\left(1+O\left(\frac{1}{2^{2n}}\right)\right).\tag{12}$$

We take logarithms in (12) above to arrive at

$$\log C_n - \log D_n - \log(2(2n)!) + 2n\log(2\pi) = \log\left(1+O\left(\frac{1}{2^{2n}}\right)\right) = O\left(\frac{1}{2^x}\right).\tag{13}$$

We now let p_j for $j = 1, \ldots, K$ be all the primes $p \le T$ and write

$$C_{n_i} = p_1^{\alpha_{i,1}}p_2^{\alpha_{i,2}}\cdots p_K^{\alpha_{i,K}} \qquad \text{for all} \quad i = 1, \ldots, K+3.$$

Observe that since $\tau(2n) \le 2\tau(n) \le 2(\log x)^2$, we have that

$$D_n = \prod_{p-1|2n} p \le (2n+1)^{\tau(2n)} \le (2x+1)^{2(\log x)^2} < \exp(3(\log x)^3) \quad (x > x_0). \tag{14}$$

Thus, from formula (12), we have that

$$\begin{aligned} C_n &\le D_n \frac{2(2n)!}{(2\pi)^{2n}}\zeta(2) \le \frac{2\zeta(2)D_n}{(2\pi)^{2n}}(2n)^{2n} < \frac{2\zeta(2)D_n}{\pi^{2n}}n^{2n} \\ &< \left(\frac{2\zeta(2)\exp(3(\log x)^3)}{\pi^x}\right)x^{2x} < x^{2x} \qquad \text{for} \quad x > x_0, \end{aligned}$$

which implies that

$$\alpha_{i,j} \le \frac{2x\log x}{\log p_j} \le \frac{2x\log x}{\log 2} < 3x\log x \qquad \text{for all} \quad 1 \le i \le K+3,\ 1 \le j \le K.$$

Let $\mathbf{\Delta} := (\Delta_1, \ldots, \Delta_{K+3})$ be a nonzero vector in the null-space of the $(K+2) \times (K+3)$ matrix

$$A = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{K+3,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{K+3,2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1,K} & a_{2,K} & \cdots & a_{K+3,K} \\ 1 & 1 & \cdots & 1 \\ n_1 & n_2 & \cdots & n_{K+3} \end{pmatrix}.$$

Such a vector exists and can be computed with Cramer's rule. It's height satisfies

$$\begin{aligned} \max\{|\Delta_i|\}_{1\le i\le K+3} &\le (K+2)!\max\{|\alpha_{i,j}|, |n_\ell|, i, j, \ell\}^{K+2} \\ &< (3x(K+2)\log x))^{K+2} < (3x(\log x)^2)^{\pi(T)+2} \\ &< x^{2(\pi(T)+2)} < \exp((\log x)^2), \end{aligned} \tag{15}$$

for $x > x_0$. We now evaluate formula (13) in $n = n_i$ for $i = 1, \ldots, K+3$ and take the linear combination with coefficients $\Delta_1, \ldots, \Delta_{K+3}$ of the resulting relations getting

$$\begin{aligned} \left|\sum_{i=1}^{K+3}\Delta_i \log C_{n_i} - \sum_{i=1}^{K+3}\Delta_i \log D_{n_i}\right. &- \left.\sum_{i=1}^{K+3}\Delta_i \log(2(2n_i)! + \sum_{i=1}^{K+3} 2\Delta_i n_i \log(2\pi)\right| \\ &= O\left(\frac{\sum_{i=1}^{K+3}|\Delta_i|}{2^x}\right). \end{aligned} \tag{16}$$

In the left–hand side of estimate (16) above, the first sum vanishes; i.e.,

$$\sum_{i=1}^{K+3} \Delta_i \log C_{n_i} = 0,$$

because the vector $\mathbf{\Delta}$ is orthogonal to the first K rows of A. Similarly, the last sum also vanishes; i.e.,

$$\sum_{i=1}^{K+3} \Delta_i n_i = 0,$$

because $\mathbf{\Delta}$ is orthogonal to the last row of A. Finally, writing

$$2(2n_i)! = 2(2n_1)!(2n_1+1)(2n_1+2)\cdots(2n_i) =: 2(2n_1)!X_i \quad (i = 1, \ldots, K+3),$$

we get that

$$\log(2(2n_i)!) = \log(2(2n_1)!) + \log X_i.$$

Hence,

$$\sum_{i=1}^{K+3} \Delta_i \log(2(2n_i)!) = \sum_{i=1}^{K+3} \Delta_i \log(2(2n_1)!) + \sum_{i=1}^{K+3} \Delta_i \log X_i = \sum_{i=1}^{K+3} \Delta_i \log X_i, \tag{17}$$

where we used $\sum_{i=1}^{K+3} \Delta_i = 0$, because $\mathbf{\Delta}$ is orthogonal to the first before last row of matrix A. Thus using also (15), estimate (16) becomes

$$\left| \sum_{i=1}^{K+3} \Delta_i \log(D_{n_i}/X_i) \right| = O\left(\frac{(K+3)\exp((\log x)^2)}{2^x} \right) = O\left(\frac{1}{2^{x/2}} \right). \tag{18}$$

In the left–hand side of estimate (18) we have a linear form in logarithms. Further,

$$X_i < (2x)^{2(n_i - n_1)} \le (2x)^{2z} < \exp(3(\log x)^3) \quad (x > x_0), \tag{19}$$

which is the same estimate as estimate (14) with D_{n_i} replaced by X_i for all $i = 1, \ldots, K+3$. For each $i = 1, \ldots, K+3$, let $P_i := P(n_i)$. Then $P_i \mid X_i$. Also, P_i does not divide D_{n_j} for any $j = 1, \ldots, K+3$. Indeed, otherwise there would exist $q := P_i$ such that for some j, we have that $q \mid D_{n_j}$. Thus, there exists a prime number p such that $q \mid p-1$ and $p-1 \mid 2n_j$. However, this is not possible because $n_j \notin \mathcal{L}_3(x)$. Also, P_i divides X_j for all $j \ge i$ but does not divide X_j for any $j < i$. Indeed, this last claim follows because if $P_i \mid X_j$ for some $j < i$, then there exists $m \in [2n_1, 2n_j]$ such that $P_i \mid m$. But also $P_i \mid n_i$, so $P_i \mid 2n_i - m$, and this last number is nonzero since $2n_i \notin [2n_1, 2n_j]$. However, this is not possible for large x since it would lead to $y < P_i \le 2n_i - m \le 2z$, which is impossible for $x > x_0$. This shows that the linear form appearing in the left–hand side of (17) is nonzero (indeed, if i is maximal such that $\Delta_i \neq 0$, then the coefficient of $\log P_i$ in the left is exactly $\Delta_i \neq 0$).

We apply a linear form in logarithms á la Baker in the left–hand side of (18) (see [4], for example). We get that the left–hand side of (18) is at least

$$> \exp\left(-c_1 c_2^K \left(\prod_{i=1}^{K+3} \max\{\log D_{n_i} \log X_{n_i}\}\right) \log\max\{|\Delta_i|\}\right),$$

for some appropriate constants c_1 and c_2. With the bounds (14), (19) and (15), the above expression is at least

$$> \exp\left(-c_1 c_2^K (3(\log x)^3)^{K+3} (\log x)^2\right),$$

which compared with (18) gives

$$x(\log 2)/2 - c_3 < c_1 (3c_2 (\log x)^3)^{K+3} (\log x)^2,$$

with some appropriate constant c_3. This last estimate implies easily that the inequality $K > (1/3 - \varepsilon)\log x/\log\log x$ holds for all $\varepsilon > 0$ and $x > x_0$ (depending on ε). Taking a sufficiently small value for ε (say $\varepsilon := 1/100$), and invoking the Prime Number Theorem to estimate $K = \pi(T)$, we get a contradiction. This finishes the argument and the proof of the theorem. □

References

[1] J. Friedlander and F. Luca, "On the value set of the Carmichael λ-function", *J. Austral. Math. Soc.* **82** (2007), 123–131.

[2] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford University Press, Oxford, sixth edition, 2008. Revised by D. R. Heath-Brown and J. H. Silverman.

[3] F. Luca and A. Pizarro, "Some remarks on the values of the Riemann zeta function and Bernoulli numbers", *Preprint*, 2011.

[4] E. M. Matveev, "An explicit lower bound for a homogeneous rational linear form in logarithms of algebraic numbers. II", *Izv. Ross. Akad. Nauk Ser. Mat.* **64** (2000), 125–180; English transl. in *Izv. Math.* **64** (2000), 1217–1269.

[5] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, University Press, Cambridge, UK, 1985.

A. Bérczes
Institute of Mathematics, University of Debrecen
Number Theory Research Group, Hungarian Academy of Sciences and University of Debrecen
H-4010 Debrecen, P.O. Box 12, Hungary
E-mail address: berczesa@math.klte.hu

F. Luca
Instituto de Matemáticas
Universidad Nacional Autonoma de México
C.P. 58089, Morelia, Michoacán, México
E-mail address: fluca@matmor.unam.mx