On the greatest prime factor of $2^p - 1$ for a prime $p$
and other expressions

by

P. Erdös (Budapest) and T. N. Shorey (Bombay)

1. For a natural number $a$, denote by $P(a)$ the greatest prime factor of $a$. Stewart [10] proved that there exists an effectively computable constant $c > 0$ such that

$$\frac{P(2^p - 1)}{p} \geq \frac{1}{2} (\log p)^{1/4}$$

for all primes $p > c$. In § 2, we shall prove that $P(2^p - 1)/p$ exceeds constant times $\log p$ for all primes. In § 5, we shall prove that for 'almost all' primes $p$,

$$\frac{P(2^p - 1)}{p} \geq \frac{(\log p)^2}{(\log \log p)^3}.$$ 

For the definition of 'almost all', see § 5. Let $u > 3$ and $k \geq 2$ be integers and denote by $P(u, k)$ the greatest prime factor of $(u+1) \cdots (u+k)$. It follows from Mahler's work [6a] that $P(u, k) \geq \log \log u$. See also [6] and [8]. In § 4, we shall show that for $u \geq k^{3/2}

$$P(u, k) > c_1 k \log \log u$$

where $c_1 > 0$ is a constant independent of $u$ and $k$. It follows from well-known results on differences between consecutive primes that $P(u, k) \geq u+1$ whenever $k \leq u \leq k^{3/2}$. Let $a < b$ be positive integers which are composed of the same primes. Then, in § 3, we shall show that there exist positive constants $c_2$ and $c_3$ such that

$$b - a \geq c_2 (\log a)^{c_3}.$$ 

Erdös and Selfridge [5] conjectured that there exists a prime between $a$ and $b$.

The proof of all these theorems depend on the following recent result on linear forms in the logarithms of algebraic numbers.

Let $n > 1$ be an integer. Let $\alpha_1, \ldots, \alpha_n$ be non-zero algebraic numbers of heights less than or equal to $A_1, \ldots, A_n$ respectively, where each $A_i \geq 27$. 
Let $\beta_1, \ldots, \beta_{n-1}$ denote algebraic numbers of heights less than or equal to $B \geq 27$. Suppose that $a_1, \ldots, a_n$ and $\beta_1, \ldots, \beta_{n-1}$ all lie in a field of degree $D$ over the rationals. Set

$$A = \log A_1 \ldots \log A_n, \quad E = (\log A + \log \log B).$$

**Lemma 1.** Given $\varepsilon > 0$, there exists an effectively computable number $C > 0$ depending only on $\varepsilon$ such that

$$|\beta_1 \log a_1 + \ldots + \beta_{n-1} \log a_{n-1} - \log a_n|$$

exceeds

$$\exp \left( - (nD)^C \Lambda (\log \Lambda)^2 (\log (AB))^2 E^{n+2+\varepsilon} \right)$$

provided that the above linear forms does not vanish.

This was proved by the second author in [9]. It has been assumed that the logarithms have their principal values but the result would hold for any choice of logarithms if $C$ were allowed to depend on their determinations.

The earlier results in the direction of Lemma 1 (i.e. lower bound for the linear form with every parameter explicit) are due to Baker [1] and Ramachandra [8]. Stewart applied the result of [1] to obtain (1). We remark that the result of [8] gives the inequality (1) with constant times $(\log p)^{1/2}/(\log \log p)$. The theorems on linear forms of [1] and [8] also give (weaker) results in the direction of the inequality (2) and the other results of this paper.

**2.** For a natural number $a$, denote by $\omega(a)$ the number of distinct prime factors of $a$.

**Lemma 2.** Let $p (> 27)$ be a prime. Assume that

$$P(2^p - 1) \leq p^2.$$  

Then there exists an effectively computable constant $c_4 > 0$ such that

$$\omega(2^p - 1) \geq c_4 \log p / \log \log p.$$  

We mention a consequence of Lemma 2.

**Theorem 1.** There exists an effectively computable constant $c_5 > 0$ such that

$$P(2^p - 1) \geq c_5 p \log p$$

for all primes $p$.

**Proof.** Assume that

$$P(2^p - 1) < p \log p.$$  

Without loss of generality, we can assume that $p > 27$. Then $P(2^p - 1) \leq p^2$. By Lemma 2, we have

$$\omega(2^p - 1) \geq c_4 \log p / \log \log p.$$
By using Brun–Titchmarsh theorem ([7], p. 44) and the fact that the prime factors of $2^p - 1$ are congruent to $1$ mod $p$, we obtain

$$P(2^p - 1) \geq c_5 p \log p$$

for some constant $c_5 > 0$. Set $c_5 = \min(1, c_6)$. Thus

$$P(2^p - 1) \geq c_5 p \log p.$$ 

This completes the proof of Theorem 2.

Proof of Lemma 2. Let $1 > \varepsilon_1 > 0$ be a small constant to be suitably chosen later. Set

$$r = \left\lceil \varepsilon_1 \log p / \log \log p \right\rceil + 1.$$ 

We shall assume that

$$\omega(2^p - 1) \leq r$$

and arrive at a contradiction. Write

$$2^p - 1 = q_1^{u_1} \cdots q_r^{u_r}$$

where for $i = 1, \ldots, r$, $q_i \leq p^2$ are primes and $u_i < p$ are non-negative integers. We have

$$2^p = |(2^p - 1)2^{-p} - 1| = |q_1^{u_1} \cdots q_r^{u_r}2^{-p} - 1|.$$ 

From here, it follows that

$$0 < |u_1 \log q_1 + \ldots + u_r \log q_r - p \log 2| < 2^{-p+1}.$$ 

By Lemma 1, it is easy to check that

$$|u_1 \log q_1 + \ldots + u_r \log q_r - p \log 2| > \exp(-p^{4D})$$

where $D > 0$ is a certain large constant independent of $\varepsilon_1$. If we take $\varepsilon_1 = 1/4D$, the inequalities (3) and (4) clearly contradict each other. This completes the proof of Lemma 2.

For any integer $n > 0$ and relatively prime integers $a, b$ with $a > b > 0$, we denote $\Phi_n(a, b)$ the $n$th cyclotomic polynomial, that is

$$\Phi_n(a, b) = \prod_{(i, n) = 1}^{n} (a - \zeta^i b)$$

where $\zeta$ is a primitive $n$th root of unity. We write

$$P_n = P(\Phi_n(a, b)).$$

Stewart [10] proved the following theorem.

Theorem 2. For any $K$ with $0 < K < 1/\log 2$ and any integer $n > 2$ with at most $K \log \log n$ distinct prime factors, we have

$$P_n/n > f(n)$$
where \( f \) is a function, strictly increasing and unbounded, which can be specified explicitly in terms of \( a, b \) and \( K \).

The proof of Theorem 3 depends on Baker's result [3] on linear forms in the logarithms of algebraic numbers. If that is replaced by Lemma 1 in Stewart's paper [10], then the method of Stewart [10] gives the following result for the size of \( f \).

**Theorem 3.** We have

\[
f(n) = c_7 (\log n)^{\lambda} / \log \log n
\]

where \( \lambda = 1 - K \log 2 \) and \( c_7 > 0 \) is an effectively computable number depending only on \( a, b \) and \( K \).

3. Let \( b > a \geq 2 \) be integers. We recall that \( a \) and \( b \) are composed of the same primes if

\[
a = p_1^{u_1} \ldots p_s^{u_s}, \quad b = p_1^{v_1} \ldots p_s^{v_s}
\]

where \( p_1, \ldots, p_s \) are positive primes and \( u_1, \ldots, u_s, v_1, \ldots, v_s \) are positive integers. We prove the following

**Theorem 4.** Let \( b > a \geq 2 \) be integers that are composed of the same primes. Then there exist effectively computable positive constants \( c_8 \) and \( c_9 \) such that

\[
b - a \geq c_8 (\log a)^{c_9}.
\]

**Proof.** Let \( 0 < \varepsilon_2 < 1 \) be a small constant which we shall choose later. Without loss of generality, we can assume that \( a \geq a_0 \) where \( a_0 \) is a large positive constant depending only on \( \varepsilon_2 \), since

\[
b - a \geq 2 = (2 / \log a_0) \log a_0 \geq (2 / \log a_0) \log a
\]

whenever \( a \leq a_0 \). We shall assume that

\[
b - a < (\log a)^{\varepsilon_2}
\]

and arrive at a contradiction. Recall the expressions (5) for \( a \) and \( b \). Notice that

\[
p_1 \ldots p_s \leq b - a < (\log a)^{\varepsilon_2}.
\]

From here, it follows that

\[
s \leq \frac{8 \varepsilon_2 \log \log a}{\log \log \log a}.
\]

Further observe that \( P(a) = P(b) < (\log a)^{\varepsilon_2} \) and the integers \( u_i \) and \( v_i \) do not exceed \( 8 \log a \). Now

\[
\left( \frac{b}{a} - 1 \right) = \frac{1}{a} (b - a) \leq \frac{\log a}{a} < a^{-1/2}.
\]
Further
\[
a^{-1/2} > \left( \frac{b}{a} - 1 \right) = |p_1^{u_1-v_1} \cdots p_s^{u_s-v_s} - 1| > \frac{1}{2} |(u_1-v_1)\log p_1 + \cdots + (u_s-v_s)\log p_s| > 0.
\]

From these inequalities, we obtain
\[
0 < |(u_1-v_1)\log p_1 + \cdots + (u_s-v_s)\log p_s| < a^{-1/2}.
\]

By Lemma 1, it is easy to check that
\[
|(u_1-v_1)\log p_1 + \cdots + (u_s-v_s)\log p_s| > \exp\left( -(\log a)^E \epsilon_2 \right)
\]
where \(E > 0\) is a certain large constant independent of \(\epsilon_2\). If we take \(\epsilon_2 = 1/4E\), then the inequalities (6) and (7) clearly contradict each other. This completes the proof of Theorem 4.

Let \(b > a \geq 2\) be integers such that \(P(a) = P(b)\). Then Tijdeman [11] proved that

**Theorem 5.**

\[b-a \geq 10^{-2}\log \log a.\]

The proof of Tijdeman [11] for this theorem depends on Baker’s work [2] on \(y^2 = x^3 + k\). We remark that Theorem 5 follows easily from Lemma 1. The details for its proof are similar to those of Theorem 4.


**Theorem 6.** Let \(u (> 3)\) be an integer. Then

\[P((u+1)(u+2)) > c_{16}\log \log u.\]

Theorem 6 also follows immediately from Lemma 1. The details for its proof are similar to those of Theorem 4. We shall use Theorem 6 for the proof of Theorem 7.

4. In this section, we shall prove the following

**Theorem 7.** Let \(u > 3\) and \(k \geq 2\) be integers. Assume that
\[
u \geq k^{3/2}.
\]
Then there exists an effectively computable constant \(c_{11} > 0\) independent of \(u\) and \(k\) such that

\[P(u, k) > c_{11}k\log \log u.\]

**Proof.** In view of Theorem 6, we can assume that \(k \geq k_0\) where \(k_0\) is a large constant. Erdős [4] proved that \(P(u, k) > c_{12}k\log k\) for some constant \(c_{12} > 0\). So it is sufficient to prove the theorem when
\[
\log k < \log \log u.
\]

4 — Acta Arithmetica XXX.3
We write, for brevity,
\[ P = P(u, k), \quad r = \lceil 2\pi(P)/k \rceil + 2. \]

Let us write \( n = m' m'' \) where \( u < n \leq u + k \) and \( m' \) is the product of all powers of primes not exceeding \( k \) and \( m'' \) consists of powers of primes exceeding \( k \). Observe that
\[ \sum_n \omega(m'') \leq \pi(P). \]

Hence the number of integers \( n \) with \( \omega(m'') \geq r \) does not exceed \( k/2 \). Hence there exist at least \( \lfloor k/2 \rfloor \) integers \( n \) with \( \omega(m'') < r \). For each prime \( q \leq k \), we omit amongst these \( n \), one \( n \) for which \( q \) divides \( n \) to a maximal power. If \( \text{star} \) denotes omission of these \( n \), then it follows, by an argument of Erdős, that
\[ \prod_n \text{star} m' \leq k^k. \]

The number of \( n \)'s counted in this product is at least
\[ \lfloor k/2 \rfloor - \pi(k) \geq k/4. \]

So there exist, among these \( n \), the integers \( n_1, n_2 \) \( (n_1 \neq n_2) \) whose \( m' \) do not exceed \( k^{20} \). Write
\[ n_1 = m'_1 p_1^{r_1} \cdots p_r^{r_r}, \quad n_2 = m'_2 q_1^{r_1} \cdots q_r^{r_r} \]
where \( m'_1, m'_2 < k^{20}, \ p_1, \ldots, p_r, \ q_1, \ldots, q_r \) are primes greater than \( k \) and not exceeding \( P \). Observe that for \( i = 1, \ldots, r \), \( u_i \) and \( v_i \) are non-negative integers not exceeding \( 8\log u \). Using (8), we get
\begin{align*}
(10) \quad 0 < \left| \sum_{i=1}^{r} u_i \log p_i - \sum_{i=1}^{r} v_i \log q_i + \log m'_1 - \log m'_2 \right| < u^{-1/6}.
\end{align*}

By Lemma 1 and (9), the left-hand side of this inequality exceeds
\begin{align*}
(11) \quad \exp \left( -(r \log P \log \log u)^{c_{13}} \right).
\end{align*}

Now the theorem follows immediately from (9), (10) and (11).

The following theorem follows from the work of Baker and Sprindžuk.

**Theorem 8.** Let \( f(x) \) be a polynomial with rational integers as coefficients. Assume that \( f(x) \) has at least two distinct roots. Then for every integer \( X > 3 \),
\[ P(f(X)) > c_{14} \log \log X \]
where \( c_{14} > 0 \) is an effectively computable constant depending only on \( f \).

By using a result of Baker on diophantine equations, Keates [6] proved Theorem 8 for polynomials of degree two and three. The proof
of Baker and Sprindžuk for Theorem 8 depends on $p$-adic versions of inequalities on linear forms in logarithms. We remark that it is easy to deduce Theorem 8 from Lemma 1.

5. A property $U$ holds for ‘almost all’ primes if given $\varepsilon > 0$, there exists $x_0 > 0$ depending only on $\varepsilon$ such that for every $x \geq x_0$, the number of primes $p \leq x$ for which the property $U$ does not hold is at most $\varepsilon x / \log x$.

We shall prove that for almost all primes $p$,

$$\frac{P(2^p - 1)}{p} \geq \frac{(\log p)^2}{(\log \log p)^3}.$$  \hspace{1cm} (12)

In fact we shall prove that

**Theorem 9.** Given $\varepsilon > 0$, there exist positive constants $n_0$ and $c_{15}$ depending only on $\varepsilon$ such that for every $n \geq n_0$, the number of primes $p$ between $n$ and $2n$ for which

$$\frac{P(2^p - 1)}{p} < c_{15} \left( \frac{\log p}{\log \log p} \right)^2,$$

is at most $\varepsilon n / \log n$.

It is easy to see that the inequality (12) for ‘almost all’ primes $p$ follows from Theorem 9.

**Proof of Theorem 9.** We shall assume that $n_0$ is a large positive constant depending only on $\varepsilon$. Set

$$r = \lceil \varepsilon n / \log n \rceil + 1.$$  \hspace{1cm} (15)

Assume that there are $r$ primes $p_1, \ldots, p_r$ between $n$ and $2n$ satisfying

$$\frac{P(2^{p_i} - 1)}{p_i} < \left( \frac{\log p_i}{\log \log p_i} \right)^2 \quad (i = 1, \ldots, r).$$  \hspace{1cm} (14)

By Lemma 2,

$$\omega(2^{p_i} - 1) \geq c_4 \frac{\log p_i}{\log \log p_i} > c_4 \frac{\log n}{\log \log n}$$

for every $i = 1, \ldots, r$. Observe that for distinct $i, j$ ($1 \leq i, j \leq r$), the prime factors of $2^{p_i} - 1$ and $2^{p_j} - 1$ are distinct. This is because if $q$ is a prime number and $q$ divides both $2^{p_i} - 1$ and $2^{p_j} - 1$, then $q \equiv 1 \pmod{p_i}$ and $q \equiv 1 \pmod{p_j}$. Therefore $q \equiv 1 \pmod{p_i p_j}$. Since $p_i p_j > n^2$, the inequality (14) is contradicted. Hence

$$\sum_{i=1}^{r} \omega(2^{p_i} - 1) \geq c_4 r \frac{\log n}{\log \log n} > c_4 \varepsilon \frac{n}{\log \log n}.$$  \hspace{1cm} (15)

Denote by

$$P = \max_{1 \leq i \leq r} P(2^{p_i} - 1).$$
If a prime number $q$ divides $2^{p_i} - 1$ for some $i = 1, \ldots, r$, then

(i) $q \leq P$.

(ii) $q - 1 = ap_i$ with an integer $a$.

(iii) $1 \leq a \leq (\log n)^2$.

By Brun's Sieve method, we get

$$\sum_{i=1}^{r} \omega(2^{p_i} - 1) \leq c_{16}P \frac{\log \log n}{(\log n)^2}$$

for some constant $c_{16} > 0$. (For this, see page 207 of a paper of P. Erdös: On the normal number of prime factors of $p - 1$ and some related problems concerning Euler $\varphi$-function, The Quaterly Journ. of Math. 6 (1935), pp. 203–213.) Comparing (15) and (16), we obtain

$$P \geq c_{17}n \left(\frac{\log n}{\log \log n}\right)^2,$$

for some positive constant $c_{17}$ depending only on $\varepsilon$. Observe that the primes $p_1, \ldots, p_r$ lie between $n$ and $2n$. Now the theorem follows immediately.

Remark. In fact the inequality (16) with $c_{16}P \frac{\log \log \log n}{(\log n)^2}$ is valid. For this, one can refer to the above mentioned paper of Erdös. In view of this, the Theorem 9 holds with

$$\frac{P(2^p - 1)}{p} < c_{15} \frac{\log p}{(\log \log p)(\log \log \log p)}$$

in place of the inequality (13).

References


On the greatest prime factor of $2^p - 1$


MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES
Budapest, Hungary

SCHOOL OF MATHEMATICS
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
Bombay, India

Received on 18.1.1975