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

by

P. Erdős (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

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

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 $p$,

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

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 $P(u, k) > \log \log u$. See also [9] and [10]. In § 4, we shall show that for $u \geq k^{32}$

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

where $c_k > 0$ 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 $k \leq u \leq k^{32}$. Let $a < b$ be positive integers which are composed of the same primes. Then, in § 3, we shall show that there exist positive constants $c_a$ and $c_b$ such that

$$b - a \geq c_a (\log a)^3.$$

Erdős 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 $a_1, \ldots, a_n$ 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 $\beta_1, \ldots, \beta_n$ denote algebraic numbers of heights less than or equal to $B$ ($B \geq 27$). Suppose that $a_1, \ldots, a_n$ and $\beta_1, \ldots, \beta_n$ all lie in a field of degree $D$ over the rationals. Set

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

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

$$|\beta_1 \log a_1 + \cdots + \beta_n \log a_n - \log e|$$

exceeds

$$\exp\left(-\left(\log D\right)^C (\log A)^\varepsilon (\log (\log B))^2 \exp^{2\varepsilon^2}\right)$$

provided that the above linear forms do 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 Ramaehandra [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 $o(a)$ the number of distinct prime factors of $a$.

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

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

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

$$o(2^p - 1) \geq c_i \log p / \log \log p.$$ 

We mention a consequence of Lemma 2.

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

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

for all primes $p$.

Proof. Assume that

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

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

$$o(2^p - 1) \geq c_i \log p / \log \log p.$$ 

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

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

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

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

This completes the proof of Theorem 2.

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

$$r = [\varepsilon \log p / \log \log p] + 1.$$ 

We shall assume that

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

and arrive at a contradiction. Write

$$2^p - 1 = q_1^{e_1} \cdots q_r^{e_r}$$

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

$$2^p - 1 = \prod_q (2^p - 1) - 2^{p - 1} = |q_1^{e_1} \cdots q_r^{e_r} - 2^{p - 1}|.$$ 

From here, it follows that

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

By Lemma 1, it is easy to check that

$$|u_1 \log q_1 + \cdots + u_r \log q_r - p \log 2| > \exp(-p^{1/2})$$

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

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

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

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 $K \log n$ 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

\[
f(n) = c(n)(\log n)^{\delta} / \log \log n
\]

where \( \delta = 1 - K \log 2 \) and \( c(n) > 0 \) 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

\[
a = p_1^{\nu_1} \cdots p_s^{\nu_s}, \quad b = p_1^{\nu_1} \cdots p_s^{\nu_s}
\]

where \( p_1, \ldots, p_s \) are positive primes and \( \nu_1, \ldots, \nu_s, v_1, \ldots, v_s \) 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 \( a_0 \) and \( c_3 \) such that

\[
b - a \geq 8a_0(\log b)^{\delta_3}
\]

**Proof.** Let \( 0 < \delta_3 < 1 \) 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 \( \delta_3 \), since

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

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

\[
b - a < (\log b)^{\delta_3}
\]

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

\[
p_1 \cdots p_s \leq b - a < (\log b)^{\delta_3}
\]

From here, it follows that

\[
s \leq 8a_0(\log b) / \log \log b.
\]

Further observe that \( P(a) = P(b) < (\log b)^{\delta_3} \) and the integers \( u_1 \) and \( v_i \) do not exceed \( 8 \log b \). Now

\[
\left( \frac{b}{a} - 1 \right) = \frac{1}{a} (b - a) < \log b / a < a^{-1/2}.
\]

Further

\[
a^{-1/2} > \left( \frac{b}{a} - 1 \right) = [p_1^{\nu_1} \cdots p_s^{\nu_s} - 1] \]

\[
> \frac{1}{2} [(u_1 - v_1) \log p_1 + \cdots + (u_s - v_s) \log p_s] > 0.
\]

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

\[
[(u_1 - v_1) \log p_1 + \cdots + (u_s - v_s) \log p_s] \geq \exp\left(-a^{1/2}\right)
\]

where \( E > 0 \) is a certain large constant independent of \( \delta_3 \). If we take \( \delta_3 = 1 / 4E \), then these 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.**

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

The proof of Tijdeman [11] for this theorem depends on Baker's work [3] 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

\[
P((u + 1)(u + 2)) > \frac{c_3 u \log \log u}{\log \log \log u}.
\]

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

\[
u \geq k^{2h}.
\]

Then there exists an effectively computable constant \( c_{11} > 0 \) independent of \( u \) and \( k \) such that

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

**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,
\[ P = P(u, k), \quad r = [2\pi(P)/k] + 2. \]

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 \( [k/2] \) integers \( n \) with \( \omega(m'') < r \). For each prime \( q \leq k \), omit all \( n \) for which \( q \) divides \( n \) to some power. If \( \pi \) denotes omission of these \( n \), then it follows, by an argument of Erdős, that
\[ \prod_n m' < k^2. \]

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

So there exist, among these \( n \), the integers \( n_1, n_2 \) \((n_1 \neq n_2)\) whose \( m' \) do not exceed \( k^2 \). Write
\[ m_1 = p_1^{m'_1} \ldots p_r^{m'_r}, \quad m_2 = q_1^{m'_1} \ldots q_r^{m'_r} \]

where \( m'_1, m'_2 < k^2 \), \( p_1, \ldots, p_r, q_1, \ldots, q_r \) 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 \( \log P \). Using (8), we get
\[ 0 < \left| \sum_{i=1}^r u_i \log p_i - \sum_{i=1}^r v_i \log p_i + \log \left( \frac{m'_1}{m'_2} \right) \right| < u^{-1/6}. \]

By Lemma 1 and (9), the left-hand side of this inequality exceeds
\[ \exp \left( -2r \log P \log \log u \right)^{2u'}. \]

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 \( X > 3 \),
\[ P(f(X)) > c_4 \log \log X \]

where \( c_4 > 0 \) 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 \( \epsilon > 0 \), there exists \( \epsilon > 0 \) depending only on \( \epsilon \) such that for every \( \epsilon > \epsilon \), the number of primes \( p \leq x \) for which the property \( U \) does not hold is at most \( \epsilon r \). We shall prove that for almost all primes \( p \),
\[ \frac{P(2^p - 1)}{p} > \frac{(\log p)^2}{(\log \log p)^3}. \]

In fact we shall prove that

**Theorem 9.** Given \( \epsilon > 0 \), there exist positive constants \( c_5 \) and \( c_6 \), depending only on \( \epsilon \), such that for every \( \epsilon > \epsilon \), the number of primes \( p \) between \( n \) and \( 2n \) for which
\[ \frac{P(2^p - 1)}{p} < c_5 \left( \frac{\log p}{\log \log p} \right)^2, \]

is at most \( c_6 n \log n \).

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 \( n \) is a large positive constant depending only on \( \epsilon \). Set
\[ r = \left[ \frac{\epsilon n}{\log \log n} \right] + 1. \]

Assume that there are \( r \) primes \( p_1, \ldots, p_r \) between \( n \) and \( 2n \) satisfying
\[ \frac{P(2^{p_i} - 1)}{p_i} < \left( \frac{\log p_i}{\log \log p_i} \right)^2 \quad (i = 1, \ldots, r). \]

By Lemma 2,
\[ \omega(2^{p_i} - 1) > c_4 \frac{\log p_i}{\log \log p_i} > c_4 \frac{\log n}{\log \log n} \]

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
\[ \sum_{i=1}^{r} \omega(2^{p_i} - 1) > c_4 r \frac{\log n}{\log \log n} > c_4 \frac{\log n}{\log \log n}. \]

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

(i) $q \leq P$.
(ii) $q - 1 = \alpha p_i$ with an integer $\alpha$.
(iii) $1 \leq \alpha \leq (\log n)^4$.

By Brun's Sieve method, we get

\[ \sum_{i=1}^{r} \omega(2^p - 1) \leq c_4 P \frac{\log \log n}{(\log n)^4} \]

for some constant $c_4 > 0$. (For this, see page 207 of a paper of P. Erdős: On the normal number of prime factors of $n - 1$ and some related problems concerning Euler's $\phi$-function, The Quarterly Journ. of Math. 6 (1935), pp. 203-213.) Comparing (15) and (16), we obtain

\[ P \geq c_{1,2} \left( \frac{\log n}{\log \log n} \right)^2 \]

for some positive constant $c_{1,2}$ 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 $c_4 P \frac{\log \log n}{(\log n)^4}$ is valid.

For this, one can refer to the above mentioned paper of Erdős. In view of this, the Theorem 9 holds with

\[ \frac{P(2^p - 1)}{p} < c_4 \frac{(\log p)^3}{(\log \log p)(\log \log \log p)} \]

in place of the inequality (18).

References