ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 37, Number 2, 2007

ON THE IRREDUCIBILITY OF A TRUNCATED BINOMIAL EXPANSION

MICHAEL FILASETA, ANGEL KUMCHEV AND DMITRII V. PASECHNIK

1. Introduction. For positive integers k and n with $k \leq n-1$, define

$$P_{n,k}(x) = \sum_{j=0}^{k} \binom{n}{j} x^j.$$

In the case that $k = n-1$, the polynomial $P_{n,k}(x)$ takes the form

$$P_{n,n-1}(x) = (x+1)^n - x^n.$$

If n is not a prime, $P_{n,n-1}(x)$ is reducible over $\mathbf{Q}$. If $n = p$ is prime, the polynomial $P_{n,n-1}(x) = P_{p,p-1}(x)$ is irreducible as Eisenstein's criterion applies to the reciprocal polynomial $x^{p-1}P_{p,p-1}(1/x)$. This note concerns the irreducibility of $P_{n,k}(x)$ in the case where $1 \leq k \leq n-2$. Computations for $n \leq 100$ suggest that in this case $P_{n,k}(x)$ is always irreducible. We will not be able to establish this but instead give some results which give further evidence that these polynomials are irreducible.

The problem arose during the 2004 MSRI program on *Topological aspects of real algebraic geometry*, in the context of work by Inna Scherbak in investigations of the Schubert calculus in Grassmannians. She had observed that the roots of any given $P_{n,k}(x)$ are simple. This follows from the identity

$$P_{n,k}(x) - (x+1)\,\frac{P'_{n,k}(x)}{n} = \binom{n-1}{k} x^k.$$

2000 AMS *Mathematics Subject Classification.* Primary 12E05 (11C08, 11R09, 14M15, 26C10).
The first author was supported by the Natl. Sci. Foundation during research for this paper. Parts of the work were completed while the third author was supported by MSRI, by DFG grant SCHN-503/2-1, and by NWO grant 613.000214, while he held positions at MSRI, CS Dept., University Frankfurt, and at EOR/FEB, Tilburg University.
Received by the editors on Sept. 14, 2004, and in revised form on Dec. 6, 2004.

She then asked whether, for a fixed positive integer n, the various $n(n-1)/2$ roots of $P_{n,k}(x)$ for $1 \le k \le n-1$ are distinct. We will not resolve this problem, but our methods imply that, for each positive integer n, almost all of the roots are distinct. In other words, the number of distinct roots is $\sim n^2/2$ as n tends to infinity. We note that, since the initial writing of this paper, Inna Scherbak [**6**] has written a paper that explains her use of these polynomials.

Before closing this Introduction, we mention that these same polynomials have recently arisen in the context of work by Iossif Ostrovskii [**4**]. In particular, he finds a solution to a problem posed by Alexandre Eremenko on the distribution of the zeroes of $P_{n,k}(x)$ as k and n tend to infinity with k/n approaching a limit $\alpha \in (0,1)$.

2. The results. Our methods apply to a wider class of polynomials than the $P_{n,k}(x)$'s alone, so we begin by recasting the problem in a more general setting. For a and b nonnegative integers with $a \le b$, the identity

$$\sum_{j=0}^{a} \binom{b}{j} (-1)^j = \binom{b-1}{a} (-1)^a \tag{1}$$

is easily established by induction on a. We deduce that

$$\begin{aligned}
P_{n,k}(x-1) &= \sum_{j=0}^{k} \binom{n}{j} (x-1)^j \\
&= \sum_{j=0}^{k} \binom{n}{j} \sum_{i=0}^{j} \binom{j}{i} (-1)^{j-i} x^i \\
&= \sum_{i=0}^{k} \sum_{j=i}^{k} \binom{n}{j} \binom{j}{i} (-1)^{j-i} x^i \\
&= \sum_{i=0}^{k} \sum_{j=i}^{k} \binom{n}{i} \binom{n-i}{j-i} (-1)^{j-i} x^i \\
&= \sum_{i=0}^{k} \binom{n}{i} \sum_{j=0}^{k-i} \binom{n-i}{j} (-1)^j x^i
\end{aligned}$$

$$= \sum_{i=0}^{k} \binom{n}{i} \binom{n-i-1}{k-i} (-1)^{k-i} x^i,$$

where the last equality makes use of (1). For $0 \le j \le k$, we define

$$c_j = \binom{n}{j} \binom{n-j-1}{k-j} (-1)^{k-j}$$
$$= \frac{(-1)^{k-j} n(n-1)\cdots(n-j+1)(n-j-1)\cdots(n-k+1)(n-k)}{j!(k-j)!}$$

so that $P_{n,k}(x-1) = \sum_{j=0}^{k} c_j x^j$. We are interested in the irreducibility of $P_{n,k}(x)$. A necessary and sufficient condition for $P_{n,k}(x)$ to be irreducible is for $P_{n,k}(x-1)$ to be irreducible, so we restrict our attention to establishing irreducibility results for the polynomials $\sum_{j=0}^{k} c_j x^j$.

For our results, we consider

$$F_{n,k}(x) = \sum_{j=0}^{k} a_j c_j x^j, \tag{2}$$

where $a_0, a_1, \ldots, a_k$ denote integers, each having all of its prime factors $\le k$. In particular, none of the a_j are zero. Observe that, if $F_{n,k}(x)$ is irreducible for all such a_j, then necessarily $P_{n,k}(x)$ is irreducible simply by choosing each $a_j = 1$. Another interesting choice for a_j is $a_j = (-1)^{k-j} j!(k-j)!$. As $F_{n,k}(x)$ is irreducible if and only if $((n-k-1)!/n!) \cdot x^k F_{n,k}(1/x)$ is irreducible, the irreducibility of $F_{n,k}(x)$ will imply the irreducibility of

$$\frac{1}{n-k} + \frac{x}{n-k+1} + \frac{x^2}{n-k+2} + \cdots + \frac{x^k}{n}.$$

In particular, if $n = k+1$, these polynomials take a nice form. It is possible to show, still with $n = k+1$, that these polynomials are irreducible for every positive integer k. The idea is to use Newton polygons with respect to two distinct primes in the interval $((k+1)/2, k+1]$. For $k \ge 10$, it is known that such primes exist, cf. [**5**]. As this is not the focus of the current paper, we omit the details.

Let N be a positive integer. The number of integral pairs (n,k) with $1 \le n \le N$ and $1 \le k \le n-2$ is

$$\sum_{n \le N} (n-2) \sim \frac{N^2}{2}.$$

Our first result is that the number of possible reducible polynomials $F_{n,k}(x)$ with $n \leq N$ and $1 \leq k \leq n-2$ is small by comparison. More precisely, we show the following.

Theorem 1. *Let $\varepsilon > 0$, and let N be a positive integer. For each integral pair (n,k) with $1 \leq n \leq N$ and $1 \leq k \leq n-2$, consider the set $S(n,k)$ of all polynomials of the form (2) where $a_0, a_1, \ldots, a_k$ denote arbitrary integers, each having all of its prime factors $\leq k$. The number of such pairs (n,k) for which there exists a polynomial $f(x) \in S(n,k)$ that is reducible is $O(N^{23/18+\varepsilon})$.*

Under the assumption of the Lindelöf hypothesis, a result of Gang Yu [**9**] can be used to improve our estimate for the number of exceptional pairs (n,k) to $O(N^{1+\varepsilon})$. A further improvement to $O(N \log^3 N)$ is possible under the Riemann hypothesis by a classical result of Atle Selberg [**7**].

Based on the main result in [**1**], one can easily modify our approach to show that, for each positive integer n, there are at most $O(n^{0.525})$ different positive integers $k \leq n-2$ for which $S(n,k)$ contains a reducible polynomial. This implies the remark in the Introduction that for fixed n, the number of distinct roots of $P_{n,k}(x)$ for $k \leq n-2$ is $\sim n^2/2$ as n tends to infinity.

Our second result is an explicit criterion for the irreducibility of $F_{n,k}(x)$.

Theorem 2. *If there is a prime $p > k$ that exactly divides $n(n-k)$, then $F_{n,k}(x)$ is irreducible for every choice of integers $a_0, a_1, \ldots, a_k$ with each having all of its prime factors $\leq k$.*

Theorem 2 has a simple proof based on Eisenstein's criterion. It implies, in particular, that if n is a prime, then $P_{n,k}(x)$ is irreducible for every $k \in \{1, 2, \ldots, n-1\}$. This then resolves the problem of Scherbak in the case that n is a prime.

Our third and final result concerning the irreducibility of $F_{n,k}(x)$ is as follows.

Theorem 3. *Let k be a fixed integer ≥ 3. There is an $n_0 = n_0(k)$ such that if $n \geq n_0$, then $F_{n,k}(x)$ is irreducible for every choice of integers $a_0, a_1, \ldots, a_k$ with each having all of its prime factors $\leq k$.*

The value of $n_0(k)$ in this last result, being based on the solutions to certain Thue equations, can be effectively determined. The result is of added interest as the proof of Theorem 1 relies on considering k large. Thus, the proof of Theorem 1 gives no information about the situation in Theorem 3, where k is fixed and n is large. In the case that $k = 1$, the polynomials $F_{n,k}(x)$ are linear and, hence, irreducible. In the case that $k = 2$, our approach does not apply; but we note that the polynomials $P_{n,2}(x)$ are easily seen to be irreducible for $n \geq 3$ as $P_{n,2}(x)$ has imaginary roots for such n.

3. The proofs.

Proof of Theorem 1. Let $f(x) = \sum_{j=0}^{k} d_j x^j \in \mathbf{Z}[x]$ with $d_k d_0 \neq 0$. In the argument, we will make use of the Newton polygon of $f(x)$ with respect to a prime p. The Newton polygon of $f(x)$ with respect to p can be defined as the lower part of the convex hull of the points $(j, \nu_p(d_j))$ where $0 \leq j \leq k$ and $\nu_p(m)$ is defined to be the integer r satisfying $p^r \mid m$ and $p^{r+1} \nmid m$. Thus, the Newton polygon has its left most endpoint being $(0, \nu_p(d_0))$ and its right-most endpoint being $(k, \nu_p(d_k))$. A theorem of Gustave Dumas [**2**] asserts that, for a fixed prime, the Newton polygon of a product of two polynomials can be obtained by translating the edges of the Newton polygons of each of the polynomials. The endpoints on the translation of an edge always occur at lattice points. For the proof of Theorem 1, we will use a specific consequence of this result: If the lattice points along the edges of the Newton polygon of $f(x)$ with respect to p consist of $(0, \nu_p(d_0))$, $(k, \nu_p(d_k))$ and only one additional lattice point, say at (u, v), then either $f(x)$ is irreducible or it is the product of an irreducible polynomial of degree u times an irreducible polynomial of degree $k-u$. We note that the lattice point (u, v) in this context need not be one of the points $(j, \nu_p(d_j))$ (for example, consider $f(x) = x^2+4x+4$ and $p=2$).

Consider n sufficiently large. Let p_j denote the jth prime, and let t be maximal such that $p_t < n$. Denote by $\delta(n)$ the distance from n to p_{t-1}

so that $\delta(n) = n - p_{t-1}$. Suppose that k satisfies $2\delta(n) < k < n - \delta(n)$. We show that in this case, the polynomial $f(x) = F_{n,k}(x)$ is irreducible over $\mathbf{Q}$ (independent of the choices of a_j as in the theorem). First, we explain why this implies our result.

For the moment, suppose that we have shown that $F_{n,k}(x)$ is irreducible over $\mathbf{Q}$ for n sufficiently large and $2\delta(n) < k < n - \delta(n)$. Let $\rho(n) = p_{t-1}$ where t is defined as above. It follows that the number of pairs (n, k) as in the theorem for which there exists a reducible polynomial $f(x) \in S(n, k)$ is

$$\ll \sum_{n \le N} \delta(n) \ll \sum_{n \le N} (n - \rho(n)) \ll \sum_{2 < p_t < N} \sum_{\substack{n \le N \\ \rho(n) = p_{t-1}}} (n - p_{t-1}).$$

This last double sum can be handled rather easily by extending the range on n slightly (to the least prime that is $\ge N$). It does not exceed

$$\begin{aligned} \sum_{2 < p_t < N} \sum_{p_t < n \le p_{t+1}} (n - p_{t-1}) &\le \sum_{2 < p_t < N} \sum_{p_t < n \le p_{t+1}} (p_{t+1} - p_{t-1}) \\ &\le \sum_{2 < p_t < N} (p_{t+1} - p_{t-1})(p_{t+1} - p_t). \end{aligned}$$

Setting $d_j = p_{j+1} - p_j$, we deduce from the arithmetic-geometric mean inequality that

$$(p_{t+1} - p_t)(p_{t+1} - p_{t-1}) = d_t(d_t + d_{t-1}) = d_t^2 + d_t d_{t-1} \le \frac{3}{2}\, d_t^2 + \frac{1}{2}\, d_{t-1}^2.$$

Hence, the number of pairs (n, k) as in the theorem for which there exists a reducible polynomial $f(x) \in S(n, k)$ is $\ll \sum_{p_t < N} d_t^2$. A theorem of Roger Heath-Brown [**3**] asserts that

$$\sum_{p_t \le N} d_t^2 \ll N^{23/18 + \varepsilon}.$$

Therefore, Theorem 1 follows provided we establish that $F_{n,k}(x)$ is irreducible over $\mathbf{Q}$ for n sufficiently large and $2\delta(n) < k < n - \delta(n)$.

Consider n sufficiently large and k an integer in the interval $(2\delta(n), n - \delta(n))$. Let $p = p_t$ and $q = p_{t-1}$. Note that both p and q are greater

than k. We set u and v to be the positive integers satisfying $p = n - u$ and $q = n - v$. Then $1 \leq u < v = \delta(n) < k/2$. Observe that the numerator of c_j is the product of the integers from $n - k$ to n inclusive but with the factor $n - j$ missing. Also, the denominator of c_j is not divisible by any prime $> k$ and, in particular, by p or by q.

We look at the Newton polygon of $f(x)$ with respect to p and the Newton polygon of $f(x)$ with respect to q. Note that $\nu_p(n - u) = 1$ and, for each j, we have $\nu_p(a_j) = 0$. Therefore, the Newton polygon of $f(x)$ with respect to p consists of two line segments, one from $(0, 1)$ to $(u, 0)$ and one from $(u, 0)$ to $(k, 1)$. The theorem of Dumas implies that if $f(x)$ is reducible, then it must be an irreducible polynomial of degree u times an irreducible polynomial of degree $k - u$. Similarly, by considering the Newton polygon of $f(x)$ with respect to q, we deduce that if $f(x)$ is reducible, then it is an irreducible polynomial of degree v times an irreducible polynomial of degree $k - v$. Since $k - v > \delta(n) > u$ and $v \neq u$, we deduce that $f(x)$ cannot be reducible. Thus, $f(x)$ is irreducible. Theorem 1 follows. □

Proof of Theorem 2. Eisenstein's criterion applies to $x^k F_{n,k}(1/x)$ whenever there is a prime $p > k$ that exactly divides n, i.e., $p \mid n$ and $p^2 \nmid n$. Hence, $F_{n,k}(x)$ is irreducible whenever such a prime exists. Also, $F_{n,k}(x)$ itself satisfies Eisenstein's criterion whenever there is a prime $p > k$ that exactly divides $n - k$. □

Proof of Theorem 3. As in the previous proofs, we work with $f(x) = F_{n,k}(x)$ (where the a_j are arbitrary integers divisible only by primes $\leq k$). With k fixed, we consider n large and look at the factorizations of n and $n - k$.

Lemma 1. *Let p be a prime $> k$ and e a positive integer for which $\nu_p(n) = e$ or $\nu_p(n - k) = e$. Then each irreducible factor of $f(x)$ has degree a multiple of $k/\gcd(k, e)$.*

The proof of Lemma 1 follows directly by considering the Newton polygon of $f(x)$ with respect to p. It consists of one edge, and the x-coordinates of the lattice points along this edge will occur at multiples of $k/\gcd(k, e)$. The theorem of Dumas implies that the irreducible factors of $f(x)$ must have degrees that are multiples of $k/\gcd(k, e)$.

Lemma 2. *Let n' be the largest divisor of $n(n-k)$ that is relatively prime to $k!$. Write*

$$n' = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r},$$

where the p_j denote distinct primes and the e_j are positive integers. Let

$$d = \gcd(k, e_1, e_2, \dots, e_r). \tag{3}$$

Then the degree of each irreducible factor of $f(x)$ is a multiple of k/d.

Note that in the statement of Lemma 2, if $n' = 1$, then $d = k$. The proof of Lemma 2 makes use of Lemma 1. Suppose that m is the degree of an irreducible factor of $f(x)$. Then, for each $j \in \{1, 2, \dots, r\}$, Lemma 1 implies there is an integer b_j such that $me_j = kb_j$. There are integers x_j for which

$$kx_0 + e_1x_1 + e_2x_2 + \cdots + e_rx_r = d.$$

Hence,

$$m(d - kx_0) = m\big(e_1x_1 + e_2x_2 + \cdots + e_rx_r\big) = k\big(b_1x_1 + b_2x_2 + \cdots + b_rx_r\big).$$

It follows that md is a multiple of k so that m is a multiple of k/d as claimed.

For the proof of Theorem 3, we define d as in Lemma 2 and consider three cases: (i) $d = 1$, (ii) $d = 2$ and (iii) $d \geq 3$. In case (i), Lemma 2 implies $f(x)$ is irreducible. Case (ii) is more difficult and we return to it shortly. In case (iii), there exist positive integers a, b, m_1 and m_2 satisfying

$$n = am_1^d, \quad n - k = bm_2^d, \quad \text{and} \quad a \text{ and } b \text{ divide } \prod_{p \leq k} p^{d-1}. \tag{4}$$

As $d \mid k$ and k is fixed, there are finitely many choices for d, a and b as in (4). For each such d, a and b, the possible values of n correspond to ax^d given by solutions to the Diophantine equation

$$ax^d - by^d = k.$$

The above is a Thue equation, and it is well known that, since $d \geq 3$, it has finitely many solutions in integers x and y, see [**8**]. It follows that there are finitely many integers n for which case (iii) holds. Hence, for n sufficiently large, case (iii) cannot occur. We are left with an examination of case (ii).

For case (ii), the definition of d implies k is even. As we already have $k \geq 3$, we deduce $k \geq 4$. Lemma 2 implies that, if $f(x)$ is reducible, then it factors as a product of two irreducible polynomials each of degree $k/2$. To finish the analysis for case (ii), we make use of the following.

Lemma 3. *Let $f(x)$ be as above with d, as defined in* (3)*, equal to* 2*. Let n'' be the largest divisor of $(n-1)(n-k+1)$ that is relatively prime to $k!$. Suppose $\nu_p(n'') = e$ where p is a prime $> k$ and e is a positive integer. If $f(x)$ is reducible, then $(k-1) \mid e$.*

For the proof of Lemma 3, we again appeal to the theorem of Dumas. Suppose first that $p \mid (n-1)$. Since $p > k$, we deduce that $\nu_p(n-1) = e$. The Newton polygon of $f(x)$ with respect to p consists of two line segments, one from $(0, e)$ to $(1, 0)$ and one from $(1, 0)$ to (k, e). Let $d' = \gcd(k-1, e)$. As $d = 2$, we deduce as above that $f(x)$ is a product of two irreducible polynomials of degree $k/2$. The fact that $f(x)$ has just two irreducible factors implies by the theorem of Dumas that one of these factors has degree that is a multiple of $(k-1)/d'$ and the other has degree that is one more than a multiple of $(k-1)/d'$. We deduce that there are integers m and m' such that

$$\frac{k-1}{d'}\, m = \frac{k}{2} \quad \text{and} \quad \frac{k-1}{d'}\, m' + 1 = \frac{k}{2}.$$

It follows that $(k-1)/d'$ divides 1, whence

$$k - 1 = d' = \gcd(k-1, e).$$

We deduce that $(k-1) \mid e$. A similar argument works in the case that $p \mid (n-k+1)$.

To finish the analysis for case (ii), we use Lemma 3 to deduce that there are positive integers a', b', m_3, and m_4 such that

(5)

$$n-1 = a' m_3^{k-1}, \quad n-k+1 = b' m_4^{k-1}, \quad \text{and} \quad a' \text{ and } b' \text{ divide } \prod_{p \leq k} p^{k-2}.$$

As k is fixed, there are finitely many choices for a' and b' as in (5). For each of these, the possible values of n correspond to $a'x^{k-1}+1$ determined by solving the Diophantine equation

$$a'x^{k-1} - b'y^{k-1} = k-2.$$

As $k \geq 4$, the above is a Thue equation and has finitely many solutions in integers x and y. Thus, there are finitely many integers n for which case (ii) holds. Hence, for n sufficiently large, we deduce that $F_{n,k}(x)$ is irreducible, completing the proof of Theorem 3. □

REFERENCES

1. R.C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes*, II, Proc. London Math. Soc. **83** (2001), 532–562.

2. G. Dumas, *Sur quelques cas d'irréductibilité des polynômes à coefficients rationnels*, J. Math. Pures Appl. **2** (1906), 191–258.

3. D.R. Heath-Brown, *The differences between consecutive primes*, III, J. London Math. Soc. **20** (1979), 177–178.

4. I.V. Ostrovskii, *On a problem of A. Eremenko*, Comput. Methods Funct. Theory **4** (2004), 275–282.

5. S. Ramanujan, *A proof of Bertrand's postulate*, J. Indian Math. Soc. **11** (1919), 181–182.

6. I. Scherbak, *Intersections of Schubert varieties and highest weight vectors in tensor products sl_{N+1}-representations*, arXiv e-print `math.RT/0409329`, July 2005.

7. A. Selberg, *On the normal density of primes in small intervals, and the difference between consecutive primes*, Arch. Math. Naturvid. **47** (1943), 87–105.

8. T.N. Shorey and R. Tijdeman, *Exponential diophantine equations*, Cambridge Tracts in Math., vol. 87, Cambridge Univ. Press, Cambridge, 1986.

9. G. Yu, *The differences between consecutive primes*, Bull. London Math. Soc. **28** (1996), 242–248.

Mathematics Department, University of South Carolina, Columbia, SC 29208
E-mail address: `filaseta@math.sc.edu`

Department of Mathematics, Towson University, 8000 York Road, Towson, MD 21252-0001
E-mail address: `akumchev@towson.edu`

SPMS/Division of Mathematical Sciences, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616