ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 41, Number 4, 2011

ON THE g-ARY EXPANSIONS OF MIDDLE BINOMIAL COEFFICIENTS AND CATALAN NUMBERS

FLORIAN LUCA AND IGOR E. SHPARLINSKI

ABSTRACT. Let

$$b_n = \binom{2n}{n} \quad \text{and} \quad c_n = \frac{1}{n+1}\binom{2n}{n}$$

be the nth middle binomial coefficient and the nth Catalan number, respectively. Let $g > 1$ be an integer. In this note, we study the base g expansions of the numbers b_n and c_n and show that for almost all n each of them has a lot of nonzero digits.

1. Introduction. For an integer $g \geq 2$ and a nonnegative integer m we write $w_g(m)$ for the number of nonzero digits in the g-ary expansion of m. When $g = 2$, $w_2(m)$ is called the *Hamming weight* of m.

In this paper, we put

$$b_n = \binom{2n}{n} \quad \text{and} \quad c_n = \frac{1}{n+1}\binom{2n}{n}$$

for the nth middle binomial coefficient and the nth Catalan number, respectively, and obtain lower bounds on $w_g(b_n)$ and $w_g(c_n)$ which hold on a set of positive integers n of asymptotic density 1. Our bounds show that $w_g(b_n)$ and $w_g(c_n)$ tend to infinity at a rate which is at least a power of the logarithm of n for almost all n.

There is an extensive literature addressing g-ary expansions of certain sequences, such as linear recurrence sequences (see [**1, 3, 7, 8, 11**] and references therein). However, it appears that the sequences which we study in this paper have never been considered in this context.

During the preparation of this paper, the first author was also supported in part by grant SEP-CONACyT 79685 and PAPIIT 100508, and the second author was supported in part by ARC grant DP0556431.
Received by the editors on July 28, 2008, and in revised form on October 15, 2008.
DOI:10.1216/RMJ-2011-41-4-1291

2. Notation and preparations. Throughout the paper, we use the Landau symbols 'O' and 'o' as well as the Vinogradov symbols '$\ll$,' '$\gg$' and '$\asymp$' with their usual meanings. We recall that $U = O(V)$, $U \ll V$ and $V \gg U$ are all three equivalent to the inequality $|U| \le cV$ with some constant $c > 0$, whereas $U = o(V)$ means that $U/V \to 0$. All the implied constants may depend on g.

We use p and q, with or without a subscript, to denote prime numbers and k, m and n to denote positive integers.

As usual, we use $\omega(n)$ to denote the number of distinct prime factors of an integer $n \ge 1$. In particular, $\omega(1) = 0$.

We need the following special case of a result of [**10**] which in turn follows from the *Chebotarev density theorem*.

Lemma 1. *Let γ and δ be fixed nonzero rational numbers. Assume that there exists a sequence $(n_k)_{k\ge 1}$ of positive real numbers tending to infinity such that for each $k \ge 0$ the exponential congruence $\gamma^x \equiv \delta \pmod p$ has an integer solution x for all primes $p \in [n_k, 2n_k]$ with at most $O(1)$ exceptions. Then γ and δ are multiplicatively dependent.*

The following result can be found in the proof of Proposition 1.1 in [**9**].

Lemma 2. *If for some prime p we have*

$$\frac{g^p - 1}{g - 1} \mid m,$$

then $w_g(m) \ge k$, where k is any positive integer satisfying the inequality

$$p \ge \left\lceil \frac{(g-1)k}{\log g} \right\rceil + 2k + 1.$$

3. Three nonzero digits. It is clear that neither b_n nor c_n can be of the form dg^α for some $d \in \{1, \ldots, g-1\}$ when n is large, since both these numbers are divisible by all primes in $[n+2, 2n]$. Thus, they have at least two nonzero digits in base g if n is large. We show, using Lemma 1, that in fact more is true.

Theorem 3. *Let $g > 1$ be fixed. Then*

$$w_g(b_n) \geq 3 \quad \textit{and} \quad w_g(c_n) \geq 3$$

hold for all but finitely many positive integers n.

Proof. Let x_n be any one of b_n and c_n. Assume that the equation

$$x_n = d_1 g^{\alpha_1} + d_2 g^{\alpha_2} \tag{1}$$

has infinitely many integer solutions $(n, \alpha_1, \alpha_2, d_1, d_2)$, where $n > 0$, $0 \leq \alpha_1 < \alpha_2$, and $d_1,\ d_2 \in \{1, \dots, g-1\}$. Since the pair (d_1, d_2) can take only at most g^2 values, it follows that there exists a fixed pair (d_1, d_2) such that the relation (1) holds for infinitely many triples (n, α_1, α_2). Furthermore, from the remark preceding Theorem 3, we may assume that $d_1 d_2 \neq 0$.

Since x_n is divisible by all primes $p \in [n+2, 2n]$, we infer that for $n > g$, the congruence $g^y \equiv -d_2/d_1 \pmod{p}$ has an integer solution y (namely, $y = \alpha_1 - \alpha_2$) for all primes $p \in [n+2, 2n]$. Applying Lemma 1 with $\gamma = g$ and $\delta = -d_2/d_1$, we get that $g = f^u$ and $-d_2/d_1 = -f^v$ hold with some integer $f \geq 2$ and some coprime integers u and v with $u > 0$. Thus, we get the congruence $f^{u\alpha - v} \equiv -1 \pmod{p}$ for all $p \in [n+2, 2n]$, where $\alpha = \alpha_1 - \alpha_2$. Note that $u\alpha - v \neq 0$, and $u\alpha - v = O(n)$.

In particular, $2(u\alpha - v) \equiv 0 \pmod{\ell_f(p)}$, where we write $\ell_f(p)$ for the multiplicative order of f modulo p. A classical result of Hooley [**4**] shows that the set $\mathcal{P}$ of primes $p \in [n+2, 2n]$ such that $\ell_f(p) < n^{1/2}/\log n$ has cardinality $\#\mathcal{P} = o(n/\log n)$ as $n \to \infty$.

Let $P(m)$ denote the largest prime factor of m. A result of Fouvry [**2**] shows that if n is large enough, then the set of primes $\mathcal{Q} \subset [n+2, 2n]$ such that $P(p-1) > p^{2/3}$ satisfies $\#\mathcal{Q} \gg n/\log n$. Put $\mathcal{R} = \mathcal{Q} \backslash \mathcal{P}$. Thus,

$$\#\mathcal{R} \geq \#\mathcal{Q} - \#\mathcal{P} \gg \pi(n)$$

holds for all sufficiently large n. Let

$$\mathcal{S} = \{P(p-1) : p \in \mathcal{R}\}.$$

For each $q \in \mathcal{S}$, the number of primes $p \le 2n$ such that $p \equiv 1 \pmod q$ is at most $2n/q \le 2n^{1/3}$. Hence,

$$\#\mathcal{S} \ge \frac{\#\mathcal{R}}{2n^{1/3}} \gg \frac{n^{2/3}}{\log n}.$$

In particular, the inequality $\#\mathcal{S} > n^{1/2}$ holds once $n > n_0$. Since $\ell_f(p) \mid u\alpha - v$ for all $p \in \mathcal{R}$, and $u\alpha - v$ is a nonzero number of size $O(n)$, we get that

$$n \gg u\alpha - v \ge \operatorname{lcm}\left[\ell_f(p) : p \in \mathcal{R}\right] \ge \prod_{q \in \mathcal{S}} q \ge n^{2\#S/3} \ge n^{2n^{1/2}/3},$$

which implies that $n \ll 1$. This contradicts the assumption that equation (1) has infinitely many solutions and concludes the proof of our theorem. □

4. Many nonzero digits.

Theorem 4. *Let $g > 1$ be fixed, and let $\varepsilon(n)$ be a function tending to zero as $n \to \infty$. Then both inequalities*

$$w_g(b_n) \gg \varepsilon(n)(\log n)^{1/2} \quad \text{and} \quad w_g(c_n) \gg \varepsilon(n)(\log n)^{1/2}$$

hold for all $n \le X$ with at most $o(X)$ exceptions as $X \to \infty$.

Proof. As before, we again let x_n be any one of the sequences b_n and c_n.

We assume that the function $\varepsilon(t)$ is decreasing, that $\varepsilon(t)(\log t)^{1/2}$ is increasing, and that $\varepsilon(t) > (\log\log t)^{-1}$. We let X be a large positive real number, put $Y = \varepsilon(X)(\log X)^{1/2}$ and let p be the smallest prime with $p \ge Y$. Thus, $p \in [Y, Y + Y/\log Y]$ holds for large X, and so $p = (1 + o(1))Y$ as $X \to \infty$.

It is enough to show that c_n is divisible by $(g^p - 1)/(g - 1)$ for all $n \in [X/\log X, X]$ with $o(X)$ exceptions as $X \to \infty$, since then, by Lemma 2 and the fact that $c_n \mid x_n$, we have

$$w_g(x_n) \gg p \ge Y = \varepsilon(X)(\log X)^{1/2} > \varepsilon(n)(\log n)^{1/2}$$

for all such values of n.

Let

$$\frac{g^p-1}{g-1}=\prod_q q^{\alpha_q}$$

be the factorization in prime powers of $(g^p-1)/(g-1)$. We first estimate the number of $n \le X$ such that $q \mid n+1$ for some $q \mid (g^p-1)/(g-1)$. For a fixed q, the number of such $n \le X$ is at most $X/q+1 \le 2X/q$ for large X, because

$$\begin{aligned} q \le g^p &= \exp(p\log g) \le \exp(2Y\log g) \le \exp(O(\varepsilon(X)(\log X)^{1/2}) \\ &= X^{o(1)} \end{aligned}$$

as $X \to \infty$; therefore, the inequality $q < X$ holds for all sufficiently large X. We note that $q \equiv 1 \pmod p$ for all $q \mid (g^p-1)/(g-1)$ (since this divisibility means that g is of order p modulo q). We also have

$$\omega(g^p-1) \ll \frac{\log(g^p-1)}{\log\log(g^p-1)} \ll \frac{p}{\log p}.$$

Therefore, varying q, we get that the number of positive integers $n \le X$ in this category is at most

$$2X \sum_{q\mid (g^p-1)/(g-1)} \frac{1}{q} \le \frac{2X\omega(g^p-1)}{p} \ll \frac{X}{\log p} \ll \frac{X}{\log Y} = o(X)$$

as $X \to \infty$.

From now on, we work only with the positive integers $n \le X$ such that $n+1$ is coprime to $(g^p-1)/(g-1)$. Let $\mathcal{A}$ be the set of such n.

It is a consequence of Kummer's well-known theorem (see [**6**]) that if q is a prime and β is a positive integer then $q^\beta \mid b_n$ (hence, $q^\beta \mid c_n$ also because q is coprime to $n+1$) provided that n has at least β base q digits which exceed $q/2$. Thus, $(g^p-1)/(g-1) \mid c_n$ for all $n \in \mathcal{A}$, except for those n such that there exist $q \mid (g^p-1)/(g-1)$ such that n has less than α_q base q digits which exceed $q/2$. Let $\mathcal{B}$ be the set of such n. We now bound the cardinality $\#\mathcal{B}$.

Let $q \mid (g^p-1)/(g-1)$ be fixed. Since $q \equiv 1 \pmod p$, we have $q > p \ge Y$. The number of digits of n in base q is $\lfloor \log n/\log q\rfloor + 1$.

Since $n \in [X/\log X, X]$, it follows that the number of digits of n in base q belongs to the interval

$$\mathcal{I}_q = \left[\frac{\log X - \log\log X}{\log q}, \frac{\log X}{\log q} + 1\right].$$

Let L be an integer in the above interval. Let $M < \alpha_q$ be some nonnegative integer. The number of $n \le X$ having L base q digits of which at most M exceed $q/2$ is

$$N_{q,L,M} \le \binom{L}{M} q^M \left(\frac{q-1}{2}\right)^{L-M} \le q^L \binom{L}{M} \frac{1}{2^{L-M}}. \tag{2}$$

Clearly,

$$q^L \ll qX.$$

Furthermore,

$$M \le \alpha_q \ll \frac{p}{\log q} \ll \frac{\varepsilon(X)(\log X)^{1/2}}{\log q} = o(L)$$

as $X \to \infty$. Thus, summing up estimates (2) for all $M \le \alpha_q$ and all $L \in \mathcal{I}_q$, we get that the number of $n \le X$ for which q^{α_q} does not divide c_n is at most

$$\begin{aligned} R_q = \sum_{\substack{M \le \alpha_q \\ L \in \mathcal{I}_p}} N_{q,L,M} &\ll \sum_{L \in \mathcal{I}_q} \sum_{M \le \alpha_q} q^L \binom{L}{M} \frac{1}{2^{L-M}} \\ &\ll qX \sum_{L \in \mathcal{I}_q} \frac{L^{\alpha_q}}{\alpha_q!} \frac{1}{2^{L-\alpha_q}}. \end{aligned} \tag{3}$$

In the above estimate, we used the fact that for $k \le M \le \alpha_q = o(L)$ as $X \to \infty$, we have

$$\frac{\binom{L}{k+1} \frac{1}{2^{L-(k+1)}}}{\binom{L}{k} \frac{1}{2^{L-k}}} = \frac{2(L-k)}{k+1} > 2;$$

therefore, in the inner sums over M in (3), the last term dominates the entire sum.

Now, by the Stirling formula, we obtain

$$\begin{aligned} R_q &\ll qX \sum_{L\in\mathcal{I}_q} \left(\frac{2eL}{\alpha_q}\right)^{\alpha_q} \frac{1}{2^L} \\ &\le qX 2^{-(\log X - \log\log X)/\log q} \left(\frac{2e}{\alpha_q}\left(\frac{\log X}{\log q}+1\right)\right)^{\alpha_q} \sum_{L\in\mathcal{I}_q} 1. \end{aligned}$$

Since

$$\sum_{L\in\mathcal{I}_q} 1 \ll 1 + \frac{\log\log X}{\log q} \ll 1 + \frac{\log\log X}{\log p} \ll 1,$$

and

$$\frac{\log X}{\log q} \ge \frac{\log X}{p\log g} \ge 1,$$

provided that X is large enough, we derive

$$\begin{aligned} R_q &\ll qX 2^{-\log X/\log q} \left(\frac{2e}{\alpha_q}\left(\frac{\log X}{\log q}+1\right)\right)^{\alpha_q} \\ &\le qX 2^{-\log X/\log q} \left(\frac{4e\log X}{\alpha_q \log q}\right)^{\alpha_q}. \end{aligned} \tag{4}$$

Since

$$\alpha_q \le \frac{p\log g}{\log q} = o\left(\frac{(\log X)^{1/2}}{\log q}\right),$$

we obtain that

$$\begin{aligned} \left(\frac{4e\log X}{\alpha_q\log q}\right)^{\alpha_q} &\le (\log X)^{p\log g/\log q} = \exp\left(\frac{p\log g}{\log q}\log\log X\right) \\ &= \exp\left((\log g + o(1))\frac{Y}{\log q}\log\log X\right) \\ &= \exp\left(o\left(\frac{\log X}{\log q}\right)\right), \end{aligned} \tag{5}$$

which, after substitution in (4), yields that the inequality

$$R_q \le qX 2^{-(1+o(1))\log X/\log q}$$

holds uniformly over our primes q as $X \to \infty$. Since $q \le Y$, we also have

$$q = \exp(o(\log X/\log q)).$$

In particular,

$$R_q \ll X \exp\left(-\frac{\log X}{2\log q}\right).$$

It remains to bound the sum of the R_q over $q \mid (g^p-1)/(g-1)$. In order to do so, we put

$$T_0 = \lfloor (\varepsilon(X)\log g)^{-1} \rfloor,$$

and for $i \ge 1$ put $T_i = 2^i T_0$. Let S_i be the set of distinct prime factors q of $(g^p-1)/(g-1)$ such that $\log X/\log q \in \mathcal{J}_i = [T_i, 2T_i]$. Clearly,

$$\log X > p\log g > \log(g^p - 1) > \#S_i \frac{\log X}{2T_i};$$

therefore, $\#S_i \ll T_i$. Thus, the cardinality of $\mathcal{B}$ is bounded above as

$$\begin{aligned}\#\mathcal{B} &\le \sum_{i\ge 0}\sum_{q\in S_i} R_q \ll \sum_{q\mid(g^p-1)/(g-1)} X e^{-T_i/2} \ll X\sum_{i\ge 0}\#S_i e^{-T_i/2}\\ &\ll X\sum_{i\ge 0} T_i e^{-T_i/2} \ll X \sum_{t\ge T_0} t e^{-t/2} = o(X),\end{aligned}$$

as $X \to \infty$, which concludes the proof of this theorem. □

5. Remarks. It is easy to see that our results can be extended to sequences of the form $k_n c_n$ for a wide class of sequences k_n (the results of Theorem 4 correspond to $k_n = 1$ and $k_n = n+1$).

Note that the only key ingredient of the proof of Theorem 4 is Kummer's theorem which tells us what is the exponent of a prime dividing a binomial coefficient. In particular, results similar to Theorem 4 are likely to hold for other sequences satisfying Kummer type theorems like the ones studied by Knuth and Wilf in [**5**].

We conjecture that both $w_g(b_n)$ and $w_g(c_n)$ tend to infinity with n but we have not been able to prove this. In what follows, we give a heuristic which backs up this conjecture. Let X be large, let Y be some parameter depending on X to be determined later, and again let q^{α_q} be an exact prime power dividing $(g^p-1)/(g-1)$, where $p < (\log X)^{1/2}$ is a prime. The argument from the proof of Theorem 4, based on Kummer's theorem, shows that the number of integers $n \le X$ such that q^{α_q} does not divide b_n is of order

$$(6)\qquad \binom{\lfloor \log X/\log q\rfloor + 1}{\alpha_q}\frac{q^{\alpha_q+1}X}{2^{\lfloor \log X/\log q\rfloor}} \ll \frac{q^{\alpha_q+1}(\log X)^{\alpha_q}X}{2^{\lfloor \log X/\log q\rfloor}} \ll X2^{-\log X/\log q+(\alpha_q+1)(\log q+\log\log X)/\log 2}.$$

Clearly, $(\alpha_q+1)\log q \le 2\alpha_q \log q \le 2p\log g$. Thus, provided that

$$(7)\qquad (2p\log g)^2 \le \frac{\log 2}{3}\cdot\frac{\log X}{\log\log X},$$

we have that the above counting function is of order

$$X2^{-(\log X)/(3\log q)} \le X2^{-c_1(\log X)/p} = X^{1-c_2/p},$$

where $c_1 = 1/(3\log g)$ and $c_2 = c_1\log 2$. Summing this up over all the

$$\omega((g^p-1)/(g-1)) \ll p < (\log X)^{1/2}$$

possible prime factors q of $(g^p-1)/(g-1)$, we get that the counting function of such $n \le X$ is of order

$$X^{1-c_2/p}(\log X)^{1/2} < X^{1-c_3/p}$$

once X is sufficiently large with $c_3 = c_2/2$. Thus, given n, the "expectation" that $(g^p-1)/(g-1)$ does not divide n is $O(n^{-c_3/p})$. Assume now that these expectations are independent for varying p, and let p vary between Y and Y^{c_4}, where c_4 is a suitable constant. Then the expectation that b_n is not a multiple of any of the $(g^p-1)/(g-1)$ for p in this range is of order

$$\prod_{Y\le p\le Y^{c_4}} n^{-c_3/p} = n^{-c_3(\log(\log(Y^{c_4}))-\log\log Y+o(1))} = n^{-c_3\log c_4+o(1)},$$

as $Y\to\infty$, where in the above argument we have applied Mertens's estimate

$$\sum_{p\le t}\frac{1}{t} = \log\log t + A + O\left(\frac{1}{\log t}\right)$$

which holds for all $t \geq 3$ with a suitable constant A. Choosing c_4 such that $c_4 = \exp(2/c_3)$, we get that $c_3 \log c_4 = 2$; therefore, for large n the expectation that b_n is not a multiple of any of the $(g^p - 1)/(g-1)$ for such p is $n^{-2+o(1)}$. Since the sum of these expectations is a convergent series, we would expect only finitely many n to have this property. Condition (7) is now satisfied for large X when $p \leq Y^{c_4}$ if we take $Y = \lfloor (\log X)^{c_5} \rfloor$ with $c_5 = 1/(2c_4) - \varepsilon$ with any fixed $\varepsilon > 0$. Note also that the inequality $Y^{c_4} < (\log X)^{1/2}$ holds for large X. Thus, the above heuristics seem to suggest that for all but finitely many n there should be a prime $p \gg (\log n)^{c_5}$ such that $(g^p - 1)/(g-1)$ divides b_n, and now Lemma 2 shows that $w_g(b_n) \gg (\log n)^{c_5}$. Similar heuristics apply to c_n. We do not enter into details.

Acknowledgments. Work on this paper was done while F.L. visited the Department of Computing of Macquarie University in Sydney, Australia. He thanks the Department for its hospitality.

REFERENCES

1. R. Blecksmith, M. Filaseta and C. Nicol, *A result on the digits of* a^n, Acta Arith. **64** (1993), 331–339.

2. E. Fouvry, *Théorème de Brun-Titchmarsh: Application au théorème de Fermat*, Invent. Math. **79** (1985), 383–407.

3. S. Hernández and F. Luca, *Palindromic powers*, Rev. Col. Mat. **40** (2006), 81–86.

4. C. Hooley, *On Artin's conjecture*, J. reine Angew. Math. **225** (1967), 209–220.

5. D.E. Knuth and H.S. Wilf, *The power of a prime that divides a generalized binomial coefficient*, J. reine Angew. Math. **396** (1989), 212–219.

6. E.E. Kummer, *Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen*, J. reine Angew. Math. **44** (1852), 93–146.

7. F. Luca, *Distinct digits in base b expansions of linear recurrence sequences*, Quaest. Math. **23** (2000), 389–404.

8. ———, *Palindromes in Lucas sequences*, Monatsh. Math. **138** (2003), 209–223.

9. ———, *Arithmetic properties of positive integers with fixed digit sum*, Rev. Mat. Iberoam. **22** (2006), 369–412.

10. F. Luca and F. Pappalardi, *Members of binary recurrences on lines of the Pascal triangle*, Publ. Math. Debrecen **67** (2005), 103–113.

11. C.L. Stewart, *On the representation of an integer in two different bases*, J. reine Angew. Math. **319** (1980), 63–72.

Instituto de Matemáticas, Universidad Nacional Autonoma de México, C.P. 58089, Morelia, Michoacán, México
Email address: fluca@matmor.unam.mx

Department of Computing, Macquarie University, Sydney, NSW 2109, Australia
Email address: igor.shparlinski@mq.edu.au