The distribution of the binomial coefficients modulo p

Richard Garfield and Herbert S. Wilf*

University of Pennsylvania
Philadelphia, PA 19104-6395

Abstract

If p is a prime, a is a primitive root modulo p, and n is a positive integer, let $r_i(n)$ be the number of k such that $0 \le k \le n$ and $\binom{n}{k} \equiv a^i$ modulo p, and let $R_n(x) = \sum_{i=0}^{p-2} r_i(n)x^i$ be their generating function.

We show that $R_n(x) \equiv \prod_{j=1}^{p-1} R_j(x)^{t_j}$ modulo $(x^{p-1}-1)$, where t_j is the number of appearances of the digit j in the p-ary expansion of n. The proof uses the fact that a certain mapping of p-ary digit strings to polynomials modulo $(x^{p-1} - 1)$ is a homomorphism. We use this result to study how the values of the binomial coefficients sit in the quadratic residues modulo p.

Date: July 1, 1990; revised December 20, 1990.

MR Subject Classifications: 05A15, 11B73

Keywords: binomial coefficients, congruence, generating function, quadratic residues

* Supported in part by the United States Office of Naval Research

The distribution of the binomial coefficients modulo p

Richard Garfield and Herbert S. Wilf*

Introduction and discussion

Of the values of the binomial coefficients $\binom{1,000,000,000}{k}$, for integer k, how many fall into each of the residue classes $1, 2, 3, 4, 5, 6$ modulo 7? Our answer to questions of this form is the following.

Theorem. *Let p be a prime, let a be a primitive root modulo p, let $n \geq 0$ be an integer and let $t_j = t_j(n)$ be the number of times the digit j appears in the p-ary expansion of n $(0 \leq j \leq p-1)$. Finally, for each i let $r_i(n)$ be the number of integers k, $0 \leq k \leq n$, for which the binomial coefficient $\binom{n}{k} \equiv a^i \bmod p$, and let $R_n(x) = \sum_{i=0}^{p-2} r_i(n)x^i$ be their generating function. Then*

$$R_n(x) \equiv \prod_{j=1}^{p-1} R_j(x)^{t_j} \qquad \text{modulo } (x^{p-1}-1). \tag{$*$}$$

In the sequel we will write $G(x) = G_{p,n}(x) = \prod_{j=1}^{p-1} R_j(x)^{t_j}$.

We mention two methods by which one can carry out, in practice, the reduction of $G_{p,n}(x)$ modulo $(x^{p-1}-1)$. First, one can simply take the remainder of the division $G_{p,n}(x)/(x^{p-1}-1)$. Second, one might observe that the division algorithm

$$G_{p,n}(x) = (x^{p-1}-1)Q(x) + R_n(x) \qquad (\deg(R_n) < p-1) \tag{$**$}$$

implies that the values of the unreduced polynomial G and of the reduced polynomial R_n agree with each other at all of the $(p-1)^{\text{st}}$ roots of unity. But since R_n is of degree $< p-1$, it is uniquely determined by those values. Hence we have the 'explicit' formula

$$R_n(x) = \frac{1}{(p-1)} \sum_{\omega^{p-1}=1} \left(\frac{x^{p-1}-1}{x-\omega}\right) \omega G_{p,n}(\omega). \tag{$***$}$$

The first method is preferred for specific polynomials G, while the second may be better suited to obtaining general formulas for the coefficients.

* Supported in part by the United States Office of Naval Research

Example. We will now answer the question that appears in the first sentence above. Since $n = (1000000000)_{10} = (33531600616)_7$, the digit-counts are

$$(t_1, \ldots, t_6) = (2, 0, 3, 0, 1, 3),$$

and we choose the primitive root $a = 3$. From $(*)$,

$$\begin{aligned} G_{7,1000000000}(x) &= \{2\}^2\{2+2x\}^3\{2+2x+2x^5\}\{4+3x^3\}^3 \\ &= 64(1+x)^3(1+x+x^5)(4+3x^3)^3. \end{aligned}$$

If we divide by $x^6 - 1$ the remainder is

$$R_{1000000000}(x) = 87808 + 88000x + 88000x^2 + 87808x^3 + 87616x^4 + 87616x^5.$$

The coefficients of R are the numbers of binomial coefficients that belong to the respective residue classes $3^0 = 1, 3^1 = 3, 3^2 = 2, 3^3 = 6, 3^4 = 4, 3^5 = 5$. Hence in the possibly more natural ordering $1, 2, 3, 4, 5, 6$ the occupancy numbers are

$$87808, \quad 88000, \quad 88000, \quad 87616, \quad 87616, \quad 87808.$$

It is well known, from Lucas' theorem, that the number of nonzero binomial coefficients is $\prod(1 + n_i)$ where the n_i are the p-ary digits of n. In this case that product is

$$\prod_i (1+i)^{t_i} = 2^2 4^3 6^1 7^3 = 526,848$$

which is the sum of the displayed occupancy numbers. ∎

We remark that the distribution depends only on the numbers of each p-ary digit that appear in $(n)_p$, but not on the placement of those digits. The theorem therefore tells us how to use the data in the first p lines of binomial coefficients in order to deduce the complete distribution into residue classes of any line n of those coefficients, where n is given only by the counts of its p-ary digits.

Our results are closely related to work of Hexel and Sachs [2,3] who were primarily interested in the number of occurrences of each residue class summed over the first $n + 1$ lines of Pascal's triangle. For a single line of the triangle they gave results in two forms, both of which are explicit formulas for the occupancy numbers rather than generating functions. One of these (Theorem 3 of [3]) involves sums over roots of unity of certain cyclic matrices, and the second (Theorem 4 of [3]) is essentially the formula that would result if one were to equate the coefficients of like powers of x on both sides of our $(*)$ above. These results can be obtained from our generating function by using either of the two methods that we discussed just below the statement of our theorem.

Proof of the theorem

In the generating function $(*)$ one sees that the digits of n act independently of each other: each digit contributes a polynomial from its line of Pascal's triangle to the product. This independence is characteristic of phenomena related to Lucas' theorem (see below) on which our theorem rests.

The independence can be cast in the following terms. Let $\mathcal{S}$ be the set of all finite words over an alphabet $0, 1, \ldots, p-1$, made into a semigroup by concatenation. Let $\mathcal{P}$ be the semigroup $Z[x] \bmod (x^{p-1}-1)$ of polynomials under multiplication $\bmod (x^{p-1}-1)$. Consider the mapping $f = f_{p,a} : \mathcal{S} \to \mathcal{P}$ that is given by

$$f(\sigma) = \sum_{i=0}^{p-2} r_i(n_\sigma) x^i$$

where n_σ is the integer whose p-ary digit string is σ. We will show that f *is a semigroup homomorphism,* and therefore its action on any string σ is the product of its actions on the individual digits of σ, which is exactly what our theorem asserts.

To prove that f is a homomorphism we use the theorem of Lucas (e.g. [1], pp. 78-79) which states that

$$\binom{n}{k} \equiv \prod_i \binom{n_i}{k_i} \quad (\bmod p)$$

where the n_i's and the k_i's are the p-ary digits of n and of k, respectively. We follow the conventions that $\binom{n}{k} = 0$ if $n \geq 0$ is an integer and $k < 0$ or $k > n$, and that our sequence $\{r_i\}$ is defined for all integer i and is periodic in i of period $p-1$.

Now if σ', σ'' are two given strings, we have

$$\begin{aligned}
f(\sigma'\sigma'') &= \sum_{i=0}^{p-2} r_i(n_{\sigma'\sigma''}) x^i \\
&= \sum_{i=0}^{p-2} \left| \left\{ k : \binom{n_{\sigma'\sigma''}}{k} \equiv a^i \right\} \right| x^i \\
&= \sum_{i=0}^{p-2} \left| \left\{ k : \binom{n_{\sigma'}}{k'} \binom{n_{\sigma''}}{k''} \equiv a^i \right\} \right| x^i \\
&= \sum_{i=0}^{p-2} \left\{ \sum_{j=0}^{p-2} \left| \left\{ k' : \binom{n_{\sigma'}}{k'} \equiv a^j \right\} \right| \left| \left\{ k'' : \binom{n_{\sigma''}}{k''} \equiv a^{i-j} \right\} \right| \right\} x^i \\
&= \sum_{i=0}^{p-2} \left\{ \sum_{j=0}^{p-2} r_j(n_{\sigma'}) r_{i-j}(n_{\sigma''}) \right\} x^i \\
&\equiv \sum_{i=0}^{p-2} \sum_{j=0}^{p-2} (r_j(n_{\sigma'}) x^j)(r_{p-1+i-j}(n_{\sigma''}) x^{p-1+i-j}) \qquad \text{modulo}\,(x^{p-1}-1) \\
&\equiv f(\sigma') f(\sigma'') \qquad \text{modulo}\,(x^{p-1}-1).
\end{aligned}$$

Following the third sign of equality above, k' is the integer whose digit string occupies the same positions in the full string of k as the substring σ' occupies in the full string of n, and similarly for k'', σ''. ∎

Binomial coefficients as quadratic residues

Our theorem has an interesting application to the quadratic character of the values of the binomial coefficients modulo p.

Definition. *Say that an integer n is p-balanced if, among the nonzero values $\{\binom{n}{k}\}_{k=0}^{n} \pmod p$ of the binomial coefficients modulo p there are equal numbers of quadratic residues and quadratic nonresidues modulo p.*

For example, the integers $0, 1, 2$ are never p-balanced, whatever the prime p may be; the integer 3 is p-balanced precisely when $p \equiv 5$ or 7 modulo 12, and so forth.

The relevance of our theorem to such questions is the following. If we evaluate the generating function $R_n(x)$ at $x = -1$,

$$R_n(-1) = \sum_{i=0}^{p-2} r_i(n)(-1)^i$$

we see that it counts the excess of the number of quadratic residues over the number of quadratic nonresidues among the nonzero values of the binomial coefficients of order n. Thus n is p-balanced if and only if $R_n(-1) = 0$. However since by $(**)$ $R_n(x)$ and $G_{p,n}(x)$ agree with each other at all of the $(p-1)^{\mathrm{st}}$ roots of unity, we see that n is p-balanced if and only if $G_{p,n}(-1) = 0$. Since G is a product, this can happen if and only if at least one of its factors vanishes, and we have proved the following.

Corollary. *Let p be prime and let T_p be the set of integers n, $0 \le n \le p-1$, that are p-balanced. If T_p is empty then no integer $n \ge 0$ is p-balanced. If T_p is non-empty then an integer n is p-balanced if and only if its p-ary expansion contains a digit $d \in T_p$.* ∎

Thus if $p \equiv 5$ or 7 modulo 12 then every integer that contains a digit 3 is p-balanced.

The numerical data that we have are as follows. The sets T_2, T_3 and T_{11} are empty. Hence there are no integers $n \ge 0$ that are 2-, 3-, or 11-balanced. All other T_p for $p \le 227$ are nonempty. Dr. Andrew Odlyzko has extended these calculations to primes $\le 10^6$, and no other nonempty T_p exist for such p. A few of the sets T_p are: $T_5 = \{3\}$, $T_7 = \{3\}$, $T_{13} = \{7, 11\}$, $T_{17} = \{3, 15\}$, $T_{19} = \{3\}$, $T_{23} = \{7\}$, $T_{29} = \{3, 23, 27\}$. For these primes there are infinitely many p-balanced n. It would be interesting to know if any other empty T_p exist.

We thank Dr. Eduardo Friedman for helpful discussions that have clarified the ideas in this paper.

References

1. Louis Comtet, Advanced combinatorics, D. Reidel Publ. Co., 1974.
2. E. Hexel, *Einige Bemerkungen zum Pascal'schen Dreieck modulo p*, Contributions to graph theory and its applications (International Colloquium Oberhof, 1977), Technische Hochschule Ilmenau, Ilmenau, 1977, 121-128.
3. Erhard Hexel and Horst Sachs, Counting residues modulo a prime in Pascal's triangle, Indian J. Math. **20** (1978), no. 2, 91-105.

The University of Pennsylvania
Philadelphia, PA 19104-6395