THE TEACHING OF MATHEMATICS
1999, Vol. II, 1, pp. 1–30

SELECTED CHAPTERS FROM ALGEBRA

I. R. Shafarevich

Abstract. This paper is the second part of the publication "Selected chapters of algebra", the first part being published in the previous volume of the Teaching of Mathematics, Vol. I (1998), 1-22.

AMS Subject Classification: 00 A 35

Key words and phrases: Polynomial, multiple roots and derivatives, binomial formula, Bernoulli's numbers.

CHAPTER II. POLYNOMIAL

1. Roots and divisibility of polynomials

In this chapter we shall be concerned with equations of the type $f(x) = 0$, where f is a polynomial. We have already met with them at the end of the previous chapter. The equation $f(x) = 0$ should be understood as the problem: find all the roots of the polynomial (or the equation). But it may happen that all the coefficients of the polynomial $f(x)$ are 0 and the equation $f(x) = 0$ turn into an identity. We then write $f = 0$ and in that case we agree that the degree of the polynomial f is not defined.

In order to add up two polynomials we simply add the corresponding members. Polynomial are multiplied using the bracket rules. If $f(x) = a_0 + a_1x + \cdots + a_nx^n$, $g(x) = b_0+b_1x+\cdots+b_mx^m$, then $f(x)g(x) = (a_0+a_1x+\cdots+a_nx^n)(b_0+b_1x+\cdots+b_mx^m)$. Eliminating the brackets we obtain members $a_kb_lx^{k+l}$, where $0 \leqslant k \leqslant n$, $0 \leqslant l \leqslant m$. After that we group together similar members. As a result we obtain the polynomial $c_0 + c_1x + c_2x^2 + \cdots$ with coefficients

$$(1) \qquad c_0 = a_0b_0, \quad c_1 = a_0b_1 + a_1b_0, \quad c_2 = a_0b_2 + a_1b_1 + a_2b_0, \quad \dots$$

The coefficient c_m is equal to the sum of all products a_kb_l, where $k + l = m$.

Polynomials share many properties with integers. The representation of a polynomial in the form $f(x) = a_0 + a_1x + \cdots + a_nx^n$ can be considered to be an analog of the representation of a positive integer in the decimal (or some other) system. The degree of a polynomial has the role analogous to the absolute value

This paper is an English translation of: И. Р. Шафаревич, *Избраные главы алгебры*, Математическое образование, 1, 2, июль–сент. 1997, Москва, стр. 3–33. In the opinion of the editors, the paper merits wider circulation and we are thankful to the author for his kind permission to let us make this version.

of an integer. For example, if we prove a property of integers by induction on the absolute value, then in the proof of the analogous property of polynomials we use induction on the degree. Notice the following important property: the degree of the product of two polynomials is equal to the sum of their degrees. Indeed, let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$ be polynomials of degree n and m, that is to say $a_n \neq 0$, $b_m \neq 0$. If we calculate the coefficients of $f(x)g(x)$ using (1), we obtain members of the form $a_kb_lx^{k+l}$ where $k + l \leqslant n + m$. Clearly, the greatest degree we get is $m + n$ and there is only one such member: $a_nb_mx^{n+m}$. It differs from zero, since $a_nb_m \neq 0$, and it cannot be cancelled with some other member, since it has the greatest degree. This property is analogous to the property $|xy| = |x||y|$ of the absolute value $|x|$ of a number x.

The theorem on division with a remainder for polynomials is formulated and proved almost in the same way as for positive integers (Theorem 4, Chapter I).

Theorem 1. *For any polynomials $f(x)$ and $g(x)$, where $g \neq 0$, there exist polynomials $h(x)$ and $r(x)$ such that*

$$f(x) = g(x)h(x) + r(x) \tag{2}$$

where either $r = 0$, or the degree of r is less than the degree of g. For given f and g, the polynomials h and r are uniquely determined.

If $f = 0$, then the representation (2) is obvious: $f = 0 \cdot g + 0$. Suppose that $f \neq 0$ and apply the method of mathematical induction on the degree of $f(x)$. Suppose that the degree of $f(x)$ is n and that the degree of $g(x)$ is m:

$$f(x) = a_0 + a_1x + \cdots + a_nx^n, \quad g(x) = b_0 + b_1x + \cdots + b_mx^m.$$

If $m > n$, the representation (2) has the form $f = 0 \cdot g + f$, with $h = 0$, $r = f$. If $m \leqslant n$, put $f_1 = f - \frac{a_n}{b_m}x^{n-m}g$ (remember that $b_m \neq 0$, since the degree of $g(x)$ is m). Clearly, in the polynomial f_1 the members having x^n cancel out (that is how we chose the coefficient $-\frac{a_n}{b_m}$), which means that its degree is less than n. Hence, we can take that the theorem is true for that polynomial and that it has the representation of the form (2): $f_1 = gh_1 + r$, where $r = 0$ or its degree is less than m. This implies $f = f_1 + \frac{a_n}{b_m}x^{n-m}g = \left(h_1 + \frac{a_n}{b_m}x^{n-m}\right)g + r$, and we have obtained the representation (2) with $h = h_1 + \frac{a_n}{b_m}x^{n-m}$. Let us prove that the representation (2) is unique. If $f = gk + s$ is another such representation (which means that $s = 0$ or its degree is less than m), then subtracting one from the other we get

$$g(h - k) + r - s = 0, \qquad g(h - k) = s - r.$$

If the polynomial $s - r$ is 0, then $s = r$ and $h = k$. If $s - r \neq 0$, then its degree is less than m and we arrive at a contradiction, since $s - r$ is equal to the polynomial $g(h-k)$, obtained by multiplying g by $h-k$, and hence its degree cannot be smaller than the degree of g which is m.

Now please read the proof of Theorem 4 of Chapter I in order to see that the above proof is perfectly analogous to it. On the other hand, if we do all the operations which should be done in the application of the mathematical induction (i.e. transition from f_1 to the polynomial f_2 with even smaller degree, etc. until we arrive at the remainder r whose degree is less than m), we obtain the rule for the so called "corner" division of polynomials used in schools. For example, if $f(x) = x^3 + 3x^2 - 2x + 5$, $g(x) = x^2 + 2x - 1$, then the division is done according to the scheme:

$$\begin{array}{rl}
x^3 + 3x^2 - 2x + 5 & |\,\underline{x^2 + 2x - 1} \\
\underline{x^3 + 2x^2 -\ \ x\quad\ \ } & \ x + 1 \\
x^2 -\ \ x + 5 & \\
\underline{x^2 + 2x - 1} & \\
-\ 3x + 6 &
\end{array}$$

This means that we choose the leading term of the polynomial $h(x)$ so that when it is multiplied by the leading term of $g(x)$ (that is x^2) the result is the leading term of $f(x)$ (that is x^3). Therefore the leading term of $h(x)$ is x. In the first row of the above table we have $f(x)$ and in the second $g(x)x$ (the product of $g(x)$ and the leading term of $h(x)$). Their difference is in the third row. We now choose the next term of the polynomial $h(x)$ so that when multiplied by the leading term of $g(x)$ (that is x^2) it becomes equal to the leading term of the polynomial in the third row (that is x^2). Hence the next term of $h(x)$ is 1. We now repeat the procedure. In the fifth row we obtain a polynomial of degree 1 (which is less than 2, the degree of $g(x)$) and so the procedure stops. We see that

$$x^3 + 3x^2 - 2x + 5 = (x^2 + 2x - 1)(x + 1) - 3x + 6.$$

As in the case of numbers, the representation (2) is called *division with remainder* of the polynomial $f(x)$ by the polynomial $g(x)$. Polynomial $h(x)$ is the *quotient* and polynomial $r(x)$ is the *remainder* in this division.

The division of polynomials is analogous to the division of numbers and it is even simpler, since when we add two terms of a certain degree we obtain a term of the same degree, and we do not transform into tens, hundreds, etc., as in the case of number division.

Repeating the reasoning given in Chapter I for numbers, we can apply Theorem 1 to find the greatest common divisor of two polynomials. In fact, using the notation of Theorem 1 we have the following analog of Lemma 5 from Chapter I: $\mathrm{g.c.d.}(f,g) = \mathrm{g.c.d.}(g,r)$; more precisely, the pairs (f,g) and (g,r) have the same common divisors. We can now use Euclid's algorithm as in Chapter I: divide with reaminder g by r: $g = rh_1 + r_1$, then r by r_1, and so on, to obtain the following sequence of polynomials: r, r_1, r_2, $\dots$, r_n whose degrees decrease. We stop at the moment when we obtain the polynomial $r_{k+1} = 0$, i.e. when $r_{k-1} = r_k h_k$. From the equalities $\mathrm{g.c.d.}(f,g) = \mathrm{g.c.d.}(r,r_1) = \cdots = \mathrm{g.c.d.}(r_{k-1},r_k)$ we see that $\mathrm{g.c.d.}(f,g) = \mathrm{g.c.d.}(r_{k-1},r_k)$. But since r_k is a divisor of r_{k-1}, then

g. c. d.$(r_{k-1}, r_k) = r_k$ and so g. c. d.$(f, g) = r_k$ — the last nonzero remainder in the Euclid's algorithm. It should be remarked that g. c. d.(f, g) is not uniquely defined, which was not the case when we dealt with positive integers. Namely, if $d(x)$ is a common divisor of $f(x)$ and $g(x)$, then so is $cd(x)$, where $c \neq 0$ is a number. Hence, g. c. d.(f, g) is defined up to a multiplicative constant.

Theorem 1 becomes particularly simple and useful in the case when $g(x)$ is a first degree polynomial. We can then write $g(x) = ax + b$, with $a \neq 0$. Since the properties of division by g are unaltered if g is multiplied by a number, we multiply $g(x)$ by a^{-1}, so that the coefficient of x is 1. Write $g(x)$ in the form $g(x) = x - \alpha$ (it will soon become evident why it is more convenient to write α with a minus). According to Theorem 1, for any polynomial $f(x)$ we have

$$f(x) = (x - \alpha)h(x) + r. \tag{3}$$

But in our case the degree of r is less than 1, i.e. it is 0: *r is a number.* Can we find this number without carrying out the division? It is very simple—it is enough to put $x = \alpha$ into (3). We get $r = f(\alpha)$, and so we can write (3) in the form

$$f(x) = (x - \alpha)h(x) + f(\alpha). \tag{4}$$

Polynomial $f(x)$ is divisible by $x-\alpha$ if and only if the remainder in the division is 0. But in view of (4) it is equal to $f(\alpha)$. We therefore obtain the following conclusion which is called *Bézout's theorem.*

Theorem 2. *Polynomial $f(x)$ is divisible by $x - \alpha$ if and only if α is its root.*

For example, the polynomial $x^n - 1$ has a root $x = 1$. Therefore, $x^n - 1$ is divisible by $x-1$. We came across this division earlier: see formula (12) of Chapter I (where a is replaced by x and $r+1$ by n).

In spite of its simple proof, Bézout's theorem connects two completely different notions: divisibility and roots, and hence it has important applications. For instance, what can be said about the *common roots* of polynomials f and g, i.e. about the solutions of the system of equations $f(x) = 0$, $g(x) = 0$? By Bézout's theorem the number α is their common root if f and g are divisible by $x - \alpha$. But then $x - \alpha$ divides g. c. d.(f, g) which can be found by Euclid's algorithm. If $d(x) =$ g. c. d.(f, g), then $x - \alpha$ divides $d(x)$, i.e. $d(\alpha) = 0$. Therefore, the question of common roots of f and g reduces to the question of the roots of d, which is, in general, a polynomial of much smaller degree. As an illustration, we shall determine the greatest common divisor of two second degree polynomials, which we write in the form $f(x) = x^2 + ax + b$ and $g(x) = x^2 + px + q$ (we can always reduce them to this form after multiplication by a number). We divide f by g according to the general rule:

$$\begin{array}{ll} x^2 + ax + b & |\,\underline{x^2 + px + q} \\ \underline{x^2 + px + q} & \quad 1 \\ (a - p)x + (b - q) & \end{array}$$

The remainder is $r(x) = (a - p)x + (b - q)$ and we know that g. c. d.$(f, g) =$ g. c. d.(g, r). Consider the case $a = p$. If we also have $b = q$, then $f(x) = g(x)$ and the system of equations $f(x) = 0$, $g(x) = 0$ reduces to one equation $f(x) = 0$. If $b \neq q$, then $r(x)$ is a nonzero number and f and g have no common factor. Finally, if $a \neq p$, then $r(x)$ has unique root $\alpha = \dfrac{b-q}{p-a}$. We know that $\mathrm{g.c.d.}(f, g) = \mathrm{g.c.d.}(g, r)$ and it is enough to substitute this value of α into $g(x)$, in order to find out whether $g(x)$ is divisible by $x - \alpha$. We obtain the relation

$$\left(\frac{b-q}{p-a}\right)^2 + p\left(\frac{b-q}{p-a}\right) + q = 0,$$

or, multiplying it by nonzero number $(p-a)^2$, the equivalent relation

$$(b-q)^2 + p(b-q)(p-a) + q(p-a)^2 = 0. \tag{4'}$$

The second and the third member of this equality have common factor $p-a$. Taking it out we can rewrite the relation (4') in the form

$$(q-b)^2 + (p-a)(pb-aq) = 0.$$

The expression $D = (q-b)^2 + (p-a)(pb-aq)$ is called the *resultant* of the polynomials f and g. We have seen that the condition $D = 0$ is necessary and sufficient for the existence of a common factor of $f(x)$ and $g(x)$, provided that $p \neq a$. But for $p = a$ the condition $D = 0$ becomes $q = b$, and that is, as we have seen, equivalent to the existence of a common non-constant factor of $f(x)$ and $g(x)$. In general, it is possible to find for any two polynomials $f(x)$ and $g(x)$ of arbitrary degrees an expression made up from their coefficients, which equated to zero gives necessary and sufficient condition for the existence of their common nonconstant factor, but of course, the technicalities will be more difficult.

Another important application of Bézout's theorem is considered with the number of roots of a polynomial. Suppose that the polynomial $f(x)$ is not identically zero, i.e. $f \neq 0$, and that $f(x)$ has, besides α_1, another root α_2 such that $\alpha_2 \neq \alpha_1$. By Bézout's theorem, $f(x)$ is divisible by $x - \alpha_1$:

$$f(x) = (x - \alpha_1) f_1(x). \tag{5}$$

Put $x = \alpha_2$ into this equality. Since α_2 is also a root of $f(x)$, we have $f(\alpha_2) = 0$. This means that $(\alpha_2 - \alpha_1) f_1(\alpha_2) = 0$, and hence (since $\alpha_2 \neq \alpha_1$) that $f_1(\alpha_2) = 0$, i.e. that α_2 is a root of $f_1(x)$. Applying Bézout's theorem to the polynomial $f_1(x)$ we obtain the equality $f_1(x) = (x - \alpha_2) f_2(x)$ and substituting this into (5) we obtain

$$f(x) = (x - \alpha_1)(x - \alpha_2) f_2(x).$$

Suppose that the polynomial $f(x)$ has k different roots $\alpha_1, \alpha_2, \ldots, \alpha_k$. Repeating our reasoning k times we see that $f(x)$ is divisible by $(x - \alpha_1) \cdots (x - \alpha_k)$:

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_k) f_k(x). \tag{6}$$

Let n be the degree of $f(x)$. On the right-hand side of (6) we have a polynomial whose degree is not less than k, and on the left-hand side a polynomial of degree n. Hence $n \geqslant k$. We formulate this as follows.

Theorem 3. *The number of different roots of a polynomial which is not identically zero is not greater than its degree.*

Of course, if a polynomial is identically equal to 0, all the numbers are its roots. Theorem 3 was proved in the 17th century by philosopher and mathematician Descartes.

Using Theorem 3 we can answer the question we have avoided up to now: what is the meaning of the phrase *equality of polynomials*? One way is to write the polynomials in the form

$$f(x) = a_0 + a_1x + \cdots + a_nx^n, \quad g(x) = b_0 + b_1x + \cdots + b_mx^m$$

and to say that they are equal if all their coefficients are equal: $a_0 = b_0$, $a_1 = b_1$, etc. This is how we think of the equality $f = 0$—all the coefficients of f are zero. Another way to understand the term "equality" is as follows: polynomials $f(x)$ and $g(x)$ are equal if they take the same values when x is substituted by an arbitrary number, i.e. if $f(c) = g(c)$ for all c. We shall prove that these two meanings of the notion "equality" coincide. But, at first we have to make a distinction between them and in the first case we say that "$f(x)$ and $g(x)$ have equal coefficients" and in the second that "$f(x)$ and $g(x)$ have equal values for all values of x".

Evidently, if $f(x)$ and $g(x)$ have equal coefficients, then they have equal values for all x. The converse will be proved in a stronger form: we do not have to suppose that $f(x)$ and $g(x)$ coincide for all values x—it is enough to suppose that they have the same values for any $n + 1$ values of x, where n is not less than the degrees of both polynomials.

Theorem 4. *Suppose that the degrees of the polynomials $f(x)$ and $g(x)$ are not greater than n and that they have same values for some $n + 1$ different values of x. Then the coefficients of $f(x)$ and $g(x)$ are equal.*

Proof. Suppose that the polynomials $f(x)$ and $g(x)$ have equal values for $n+1$ values of x: $x = \alpha_1, \alpha_2, \ldots, \alpha_{n+1}$, i.e. that

$$f(\alpha_1) = g(\alpha_1), \quad f(\alpha_2) = g(\alpha_2), \quad \ldots, \quad f(\alpha_{n+1}) = g(\alpha_{n+1}).$$

Consider the polynomial $h(x) = f(x) - g(x)$ (here "=" denotes the equality of coefficients). We have seen that this implies that $h(\alpha) = f(\alpha) - g(\alpha)$ for any α, and, in particular, that $h(\alpha_1) = 0$, $h(\alpha_2) = 0, \ldots, h(\alpha_{n+1}) = 0$. But the degrees of f and g are not greater than n, and so the degree of h is not greater than n. This is a contradiction with Theorem 3, unless $h = 0$, i.e. unless all the coefficients of h are 0. This implies that the coefficients of f and g are equal.

From now on we can apply the term "equality" to polynomials without emphasizing in which one of the two senses.

Theorem 4 shows an interesting property of polynomials. Namely, if we know the values of a polynomial $f(x)$ of degree not greater than n for some $n + 1$ values of the variable x, then its coefficients are uniquely determined, and so are its values for *all other* values of x. Notice that in the above sentence "coefficients are uniquely determined" means only that there *cannot be* two different polynomials with the given property. Hence, it is natural to raise the question of the *existence* of such a polynomial. Namely, suppose that we have $n+1$ different numbers $x_1, x_2, \dots,$ x_{n+1} and also $n+1$ numbers $y_1, y_2, \dots, y_{n+1}$; is there a polynomial $f(x)$ of degree not greater than n such that $f(x_1) = y_1$, $f(x_2) = y_2, \dots, f(x_{n+1}) = y_{n+1}$? Theorem 4 states only that if such a polynomial exists, then it is unique. The problem of constructing such a polynomial is called the *problem of interpolation.* It often appears in the processing of experimental data, when a quantity $f(x)$ is measured only for certain values $x = x_1$, $x = x_2, \dots, x = x_{n+1}$ and it is necessary to make a plausible aasumption about its values for other values of x. The data are given by the table

(7)

x	x_1	x_2	$\dots$	x_{n+1}
$f(x)$	y_1	y_2	$\dots$	y_{n+1}

One of the plausible assumptions would be to construct a polynomial of degree not greater than n such that $f(x_1) = y_1$, $f(x_2) = y_2, \dots, f(x_{n+1}) = y_{n+1}$ and to assume that the required quantity is equal to $f(x)$ for all values of x. But does such a polynomial exist? We shall prove that it does and we shall find its formula. It is called the *interpolation polynomial* corresponding to table (7). In order to find the formula for the interpolation polynomial in the general case, we shall first consider the *simplest interpolation problem*, when in table (7) all the values $y_1, y_2, \dots\ y_{n+1}$ are 0, except one of them. Let $y_1 = y_2 = \cdots = y_{k-1} = y_{k+1} = \cdots = y_{n+1} = 0$, so that the table becomes

x	x_1	x_2	$\dots$	x_{k-1}	x_k	x_{k+1}	$\dots$	x_{n+1}
$f(x)$	0	0	$\dots$	0	y_k	0	$\dots$	0

This means that the required interpolation polynomial $f_k(x)$ has the following roots: $x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_{n+1}$ (i.e. all the numbers $x_1, \dots .\ x_{n+1}$ except x_k). But then it must be divisible by the product of the corresponding factors $x - x_i$. Since there are n factors, and since the degree of the polynomial cannot be greater than n, it can differ from this product only by a multiplicative constant. That is to say, we have to put

$$f_k(x) = c_k(x - x_1)\cdots(x - x_{k-1})(x - x_{k+1})\cdots(x - x_{n+1}). \tag{8}$$

Conversely, any polynomial of that form satisfies the required conditions for all x_1, $\dots, x_{n+1}$, except perhaps for $x = x_k$. If it is to satisfy the condition for x_k, we put $x = x_k$ into (8) and from the obtained equality we get the value of c_k. Since $f_k(x_k)$ has to be equal to y_k, we obtain

$$c_k = \frac{y_k}{(x_k - x_1)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_{n+1})}$$
$$f_k(x) = c_k(x - x_1)\cdots(x - x_{k-1})(x - x_{k+1})\cdots(x - x_{n+1}).$$

Using the auxiliary polynomial $F(x) = (x - x_1)\cdots(x - x_{n+1})$ of degree $n+1$ we can write the above formula in a different way. Namely, in that case the product

$(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_{n+1})$ is equal to $\dfrac{F(x)}{x-x_k}$. Putting $\dfrac{F(x)}{x-x_k} = F_k(x)$, we get

$$c_k = \frac{y_k}{F_k(x_k)}, \qquad f_k(x) = \frac{y_k}{F_k(x_k)} F_k(x). \tag{9}$$

Passing on to the general interpolation problem with the table (7) we only have to notice that its solution is the sum of all polynomials $f_k(x)$ which correspond to all the simplest interpolation problems:

$$f(x) = f_1(x) + f_2(x) + \cdots + f_{n+1}(x).$$

Indeed, if we put $x = x_k$ then all the members on the right-hand side become 0, except $f_k(x_k)$, and since $f_k(x)$ is the solution of the k-th simplest interpolation problem, we have $f_k(x_k) = y_k$. Finally, the degrees of $f_1(x), \ldots, f_{n+1}(x)$ are not greater than n and the same holds for their sum. We can write the obtained formula in the form

$$f(x) = \frac{y_1}{F_1(x_1)} F_1(x) + \frac{y_2}{F_2(x_2)} F_2(x) + \cdots + \frac{y_{n+1}}{F_{n+1}(x_{n+1})} F_{n+1}(x), \tag{10}$$

where $F_k(x) = \dfrac{F(x)}{x-x_k}$, $F(x) = (x-x_1)(x-x_2)\cdots(x-x_{n+1})$.

There is an unexpected identity which follows from the formula for the interpolation polynomial. Consider the interpolation problem corresponding to the table

x	x_1	x_2	$\ldots$	x_{n+1}
$f(x)$	x_1^k	x_2^k	$\ldots$	x_{n+1}^k

where k is a positive integer not greater than n or $k = 0$. On one hand it is evident that the polynomial $f(x) = x^k$ is the solution of this interpolation problem. On the other hand, we can write it down using formula (10) and we obtain that

$$x^k = \frac{x_1^k}{F_1(x_1)} F_1(x) + \frac{x_2^k}{F_2(x_2)} F_2(x) + \cdots + \frac{x_{n+1}^k}{F_{n+1}(x_{n+1})} F_{n+1}(x),$$

where $F(x) = (x-x_1)(x-x_2)\cdots(x-x_{n+1})$ and $F_i(x) = \dfrac{F(x)}{x-x_i}$. The polynomials $F_i(x)$ have degree n and the coefficient of x^n is 1. If $k < n$, then the polynomial on the right must also have degree less than n which means that all members with degree n must cancel out. In other words we have

$$\frac{x_1^k}{F_1(x_1)} + \frac{x_2^k}{F_2(x_2)} + \cdots + \frac{x_{n+1}^k}{F_{n+1}(x_{n+1})} = 0.$$

for $k < n$. If $k = n$, the coefficient of x^n must be equal to 1 and we have

$$\frac{x_1^n}{F_1(x_1)} + \frac{x_2^n}{F_2(x_2)} + \cdots + \frac{x_{n+1}^n}{F_{n+1}(x_{n+1})} = 1.$$

Notice that here $F(x) = (x-x_1)\cdots(x-x_{n+1})$, $F_k(x) = \dfrac{F(x)}{x-x_k}$, so that we have identities for arbitrary numbers $x_1, \ldots, x_{n+1}$.

Problems

1. Write down the last identities for $n = 1$ and 2, i.e. for $F(x) = (x-x_1)(x-x_2)$ and $F(x) = (x-x_1)(x-x_2)(x-x_3)$, and then verify them by direct calculation.

2. Divide $x^{n+1} - 1$ by $x - 1$ in order to obtain another derivation of the formula (12), Chapter I.

3. Divide with remainder $x^n - a$ by $x^m - b$. (Hint: the answer depends on the division of n by m.)

4. In deducing the formula (6) why was not possible to reason as follows: since $f(x)$ is divisible by all $x - \alpha_i$, it is divisible by their product? Verify that the assertion: if n is divisible by a and b then n is divisible by ab does not hold for numbers. Verify that it also does not hold for polynomials.

5. Prove that any polynomial can be written as the product of binomials $x-\alpha_i$ and a polynomial which has no roots. Prove that such a representation for a given polynomial is unique.

6. Let $F(x) = (x-x_1)\cdots(x-x_n)$ where $x_1, \dots, x_n$ are different from one another and let $f(x)$ be a polynomial of degree less than n. Prove that the fraction $\dfrac{f(x)}{F(x)}$ is equal to the sum of fractions of the form $\dfrac{a_k}{x-x_k}$, $k = 1, \dots, n$. Find the formula for a_k.

7. If $g(x)$ is a polynomial of degree less than n and if $x_1, \dots, x_{n+1}$ and the polynomial $F(x)$ have the same meaning as at the end of Section 1, prove that

$$\frac{g(x_1)}{F_1(x_1)} + \cdots + \frac{g(x_{n+1})}{F_{n+1}(x_{n+1})} = 0.$$

8. Let everything be the same as in Problem 7, except that the degree of $g(x)$ is n and the coefficient of x^n is a. Prove that

$$\frac{g(x_1)}{F_1(x_1)} + \cdots + \frac{g(x_{n+1})}{F_{n+1}(x_{n+1})} = a.$$

2. Multiple roots and derivative

The equation $x^2 - a = 0$ for $a > 0$ has two roots, given by $x = \sqrt{a}$ and $x = -\sqrt{a}$, where $\sqrt{a}$ is the arithmetic value of the square root of a. For $a = 0$ this gives two equal values. Similarly, the formula for the solution of an arbitrary quadratic equation sometimes gives two equal roots. Can similar situation happen for equations of arbitrary degree? At first the question itself seems to be meaningless. What does it mean that the equation $f(x) = 0$ has two *equal* roots? We can write any root of an equation on the paper as many times as we please, and all these numbers will be equal! But when we spoke of equal roots of the quadratic equation we used the formula for its solution. In the general case we shall also use some additional considerations in order to give a reasonable *definition* of what does it mean that the equation $f(x) = 0$ has two equal roots $x = \alpha$ and $x = \alpha$.

Such considerations are based on Bezout's theorem (Theorem 2). Let $x = \alpha$ be a rooot of $f(x)$. By Bezout's theorem $f(x)$ is divisible by $x - \alpha$ and we have $f(x) = (x-\alpha)g(x)$, where $g(x)$ is a polynomial whose degree is less than the degree of $f(x)$ by 1. If the polynomial $g(x)$ again has a root $x = \alpha$, we shall say that $f(x)$ *has two roots equal to* α. By Bezout's theorem, $g(x)$ can be written in the form $g(x) = (x - \alpha)h(x)$ and hence

$$f(x) = (x - \alpha)^2 h(x). \tag{11}$$

We can say that in the representation (6) there are two factors $x - \alpha$. This is in accordance with the intuitive notion of what are two equal roots.

If in (11) $h(x)$ again has a root α, we shall say that $f(x)$ has *three roots* equal to α. In general, if $f(x)$ can be written in the form $f(x) = (x - \alpha)^r u(x)$, where $u(x)$ is a polynomial whose root is not α, we shall say that $f(x)$ has r *equal roots* α. If $r \geqslant 2$, then α is said to be a *multiple root*. Hence, α is a multiple root if $f(x)$ is divisible by $(x - \alpha)^2$. If the polynomial $f(x)$ has exactly k roots equal to α, we say that k is the *multiplicity* of the root α. Then $f(x)$ can be written in the form $f(x) = (x - \alpha)^k g(x)$, where α is not a root of $g(x)$, i.e. $g(\alpha) \neq 0$.

For example, suppose that $x = \alpha$ is a root of the quadratic equation $x^2 + px + q = 0$. Dividing $x^2 + px + q$ by $x - \alpha$ we get

$$\begin{array}{l|l}
x^2 + px + q & x - \alpha \\ \cline{2-2}
\underline{x^2 - \alpha x} & x + p + \alpha \\
(p + \alpha)x + q & \\
\underline{(p + \alpha)x - \alpha(p + \alpha)} & \\
\qquad\qquad q + p\alpha + \alpha^2 &
\end{array}$$

i.e. $x^2 + px + q = (x - \alpha)(x + p + \alpha) + (\alpha^2 + p\alpha + q)$. Since α is a root of the equation $x^2 + px + q = 0$, we have $\alpha^2 + p\alpha + q = 0$, and so $x^2 + px + q = (x - \alpha)(x + p + \alpha)$. By our definition, this equation has two roots equal to α if α is a root of $x + p + \alpha$, i.e. if $2\alpha + p = 0$. Hence, $\alpha = -p/2$. Since $\alpha^2 + p\alpha + q = 0$, then putting $\alpha = -p/2$ we obtain that $-p^2/4 + q = 0$. This is the known condition which ensures that the equation $x^2 + px + q = 0$ has equal roots.

For the third order equation $x^3 + ax^2 + bx + c = 0$ the calculation is only a little more involved. Divide $x^3 + ax^2 + bx + c$ by $x - \alpha$:

$$\begin{array}{l|l}
x^3 + ax^2 + bx + c & x - \alpha \\ \cline{2-2}
\underline{x^3 - \alpha x^2} & x^2 + (a + \alpha)x + b + a\alpha + \alpha^2 \\
\quad (a + \alpha)x^2 + bx + c & \\
\quad \underline{(a + \alpha)x^2 - \alpha(a + \alpha)x} & \\
\qquad (b + a\alpha + \alpha^2)x + c & \\
\qquad \underline{(b + a\alpha + \alpha^2)x - \alpha(b + a\alpha + \alpha^2)} & \\
\qquad\qquad c + b\alpha + a\alpha^2 + \alpha^3 &
\end{array}$$

Since by supposition $\alpha^3 + a\alpha^2 + b\alpha + c = 0$, then $x^3 + ax^2 + bx + c = (x - \alpha)(x^2 + (a + \alpha)x + b + a\alpha + \alpha^2)$. According to our definition the equation $x^3 + ax^2 + bx + c = 0$ has two roots equal to α if α is a root of the equation and also if α is a root of the polynomial $x^2 + (a + \alpha)x + b + a\alpha + \alpha^2$. In other words, $\alpha^2 + (a + \alpha)\alpha + b + a\alpha + \alpha^2 = 0$, i.e. $3\alpha^2 + 2a\alpha + b = 0$. We see that a multiple root of the equation $x^3 + ax^2 + bx + c = 0$ is the *common root* of the polynomials $x^3 + ax^2 + bx + c$ and $3x^2 + 2ax + b$. As we saw in Section 1, they are the roots of the polynomial $\mathrm{g.c.d.}(x^3 + ax^2 + bx + c, 3x^2 + 2ax + b)$ and the greatest common divisor can be found by Euclid's algorithm.

We now apply the same reasoning to the polynomial $f(x) = a_0 + a_1x + \cdots + a_nx^n$ of arbitrary degree. When we divide $f(x)$ by $x - \alpha$ we obtain as the quotient a polynomial $g(x)$ of degree $n - 1$ whose coefficients depend on α and so we shall denote it by $g(x, \alpha)$. We know (formula (3)) that the remainder is $f(\alpha)$:

$$f(x) = (x - \alpha)g(x, \alpha) + f(\alpha). \tag{12}$$

Putting $x = \alpha$ into the polynomial $g(x, \alpha)$ we obtain the polynomial in α which is called the *derivative* of $f(x)$ and is denoted by $f'(\alpha)$. Hence, by definition,

$$f'(\alpha) = \frac{f(x) - f(\alpha)}{x - \alpha}(\alpha). \tag{13}$$

The above formula may cause some doubt, since after the substitution $x = \alpha$ both the numerator and the denominator in the expression $\dfrac{f(x) - f(\alpha)}{x - \alpha}$ become 0 and we get $\dfrac{0}{0}$. This formula therefore needs to be explained: we first (before substituting $x = \alpha$) divide the numerator by the denominator and we substitute $x = \alpha$ into their quotient which is a polynomial. For example, the meaning of the expression $\dfrac{x^2 - 1}{x - 1}(1)$ is: we first get $\dfrac{x^2 - 1}{x - 1} = x + 1$, and then $(x + 1)(1) = 2$.

Those of you who will continue to study mathematics will meet the derivative for other functions, such as $f(x) = \sin x$ or $f(x) = 2^x$. In essence they are defined by the same formula (13), but in general case it is more difficult to give the exact sense to the expression on right-hand side. In the case of polynomials everything is cleared by applying Bézout's theorem to the polynomial $f(x) - f(\alpha)$.

If α is a root of the polynomial $f(x)$ in (12), i.e. if $f(\alpha) = 0$, then we get $f(x) = (x - \alpha)g(x, \alpha)$ and by our definition α is a multiple root of $f(x)$ if α is a root of $g(x, \alpha)$, i.e. if $g(\alpha, \alpha) = 0$. But this means that $f'(\alpha) = 0$. We have proved the assertion:

Theorem 5. *A root of a polynomial $f(x)$ is multiple if and only if it is also a root of the derivative $f'(x)$.*

We see that a multiple root α is the *common root* of the polynomials $f(x)$ and $f'(x)$. In other words, α is a root of $\mathrm{g.c.d.}(f(x), f'(x))$; the greatest common divisor can be found by Euclid's algorithm and it is, as a rule, a polynomial of much smaller degree.

We shall now carry out the division of $f(x)$ by $x-\alpha$, we shall find the polynomial $g(x,\alpha)$ in (12) and we shall find the explicite formula for the derivative of a polynomial.

We could make the usual division of $f(x)$ by $x-\alpha$ and find the quotient $g(x,\alpha)$ and the remainder $f(\alpha)$. But it is better to do it another way. Recall that $f(x)$ is the sum of the terms a_kx^k and hence $f(x)-f(\alpha)$ is the sum of the terms $a_k(x^k-\alpha^k)$. The polynomial $(x^k-\alpha^k)$ has a root $x=\alpha$ and by Bézout's theorem it is divisible by $x-\alpha$. We have noticed (after the formulation of Bézout's theorem) that we have already done this division earlier. True, only for $\alpha=1$, but the general case is easily reduced to it. We shall use formula (12) of Chapter I (where $r+1$ is replaced by k):

$$(x^k-1)=(x-1)(x^{k-1}+x^{k-2}+\cdots+x+1).$$

Replace x by x/α:

$$\left(\frac{x^k}{\alpha^k}-1\right)=\left(\frac{x}{\alpha}-1\right)\left(\frac{x^{k-1}}{\alpha^{k-1}}+\frac{x^{k-2}}{\alpha^{k-2}}+\cdots+\frac{x}{\alpha}+1\right).$$

Multiplying both sides of this equality by α^k we get

$$x^k-\alpha^k=(x-\alpha)(x^{k-1}+\alpha x^{k-2}+\cdots+\alpha^{k-2}x+\alpha^{k-1}). \tag{14}$$

This formula was obtained for $\alpha\neq 0$ (since we had x/α) but it is clearly true for $\alpha=0$ also.

Consider the polynomial $f(x)=a_0+a_1x+\cdots+a_nx^n$ and the difference $f(x)-f(\alpha)$. It is equal, as we saw, to the sum of the following terms $a_k(x^k-\alpha^k)$. Divide each such term by $x-\alpha$, using formula (14). We get

$$\frac{a_k(x^k-\alpha^k)}{x-\alpha}=a_k(x^{k-1}+\alpha x^{k-2}+\cdots+\alpha^{k-2}x+\alpha^{k-1}).$$

If we put $x=\alpha$ (into the right-hand side!) we obtain the term $ka_k\alpha^{k-1}$. Hence for the polynomial $g(x,\alpha)$ in (12) for $x=\alpha$ we get that $g(x,\alpha)(\alpha)=g(\alpha,\alpha)$ is the sum of terms $ka_k\alpha^{k-1}$, i.e. $a_1+2a_2\alpha+3a_3\alpha^2+\cdots+na_n\alpha^{n-1}$. In other words, we have deduced the formula for the derivative $f'(x)$ of the polynomial $f(x)=a_0+a_1x+\cdots+a_nx^n$:

$$f'(x)=a_1+2a_2x+\cdots+na_nx^{n-1}. \tag{15}$$

Compare this with what we obtained for polynomials of degree 2 and 3 and convince yourself that those were special cases of (15) for $n=2$ and $n=3$.

The derivative of a polynomial is important not only in connection with multiple roots; it has many other applications. We shall therefore prove the basic properties of the derivative. All the proofs follow from the definition, i.e. from (12).

a) *The derivative of a constant polynomial.* If $f(x)=a_0$ then, by definition, $f(x)=f(\alpha)$ and $g(x,\alpha)=0$. Hence $f'(\alpha)=0$, i.e. $f'(x)=0$.

b) *The derivative of a sum.* Let f_1 and f_2 be two polynomials and let $f = f_1 + f_2$. We have

$$\begin{aligned} f_1(x) &= f_1(\alpha) + (x-\alpha)g_1(x,\alpha), \\ f_2(x) &= f_2(\alpha) + (x-\alpha)g_2(x,\alpha) \end{aligned} \tag{16}$$

and therefore $f_1'(\alpha) = g_1(\alpha,\alpha)$, $f_2'(\alpha) = g_2(\alpha,\alpha)$. Adding the formulas (16) we get $f(x) = f(\alpha) + (x-\alpha)g(x,\alpha)$, where $g(x,\alpha) = g_1(x,\alpha) + g_2(x,\alpha)$. Therefore, $f'(\alpha) = g(\alpha,\alpha) = g_1(\alpha,\alpha) + g_2(\alpha,\alpha) = f_1'(\alpha) + f_2'(\alpha)$, i.e.

$$(f_1 + f_2)' = f_1' + f_2'.$$

Using induction on the number of summands we easily obtain

$$(f_1 + f_2 + \cdots + f_r)' = f_1' + f_2' + \cdots + f_r'.$$

c) *Multiplication by a number.* Let $f_1(x) = af(x)$. Then from the equalities $f(x) = f(\alpha) + (x-\alpha)g(x,\alpha)$ and $g(\alpha,\alpha) = f'(\alpha)$, multiplying by a we get

$$f_1(x) = af(x) = af(\alpha) + (x-\alpha)ag(x,\alpha),$$

i.e. $f_1(x) = f_1(\alpha) + (x-\alpha)ag(x,\alpha)$ and $f_1'(\alpha) = af'(\alpha)$:

$$(af)' = af'.$$

d) *The derivative of a product.* Let $f = f_1f_2$. Multiplying the equalities (16) we get

$$f_1(x)f_2(x) = f_1(\alpha)f_2(\alpha) + (x-\alpha)g(x,\alpha),$$

where $g(x,\alpha) = g_1(x,\alpha)f_2(\alpha) + g_2(x,\alpha)f_1(\alpha) + (x-\alpha)g_1(x,\alpha)g_2(x,\alpha)$. Therefore $f(x) = f(\alpha) + (x-\alpha)g(x,\alpha)$ where $g(x,\alpha)$ is given above. Hence, $f'(\alpha) = g(\alpha,\alpha) = g_1(\alpha,\alpha)f_2(\alpha) + g_2(\alpha,\alpha)f_1(\alpha) = f_1'(\alpha)f_2(\alpha) + f_2'(\alpha)f_1(\alpha)$, i.e.

$$(f_1f_2)' = f_1'f_2 + f_2'f_1. \tag{17}$$

If f_1 is a constant (polynomial of degree 0) then in view of a) from (17) we again get c).

By induction on the number of factors we obtain

$$(f_1f_2\cdots f_r)' = f_1'f_2\cdots f_r + f_1f_2'\cdots f_r + f_1f_2\cdots f_r' \tag{18}$$

(on the right-hand side in the product $f_1\cdots f_r$ each factor is succesfully replaced by its derivative).

Indeed, according to (17):

$$(f_1f_2\cdots f_r)' = ((f_1\cdots f_{r-1})f_r)' = (f_1\cdots f_{r-1})'f_r + f_1\cdots f_{r-1}f_r'.$$

Applying to $(f_1\cdots f_{r-1})'$ the expression (18) which can be taken to be already proved, we obtain the required formula.

An important special case occurs when all the factors in (18) are equal:

$$(f^r)' = rf^{r-1}f'. \tag{19}$$

From the dfinition of the derivative it is easily verified that $x' = 1$. Hence, $(x^r)' = rx^{r-1}$. Combining the above rules, we can give a different proof of the explicite formula (15) for the derivative.

Return now to the question of multiple roots of polynomials. Suppose that α is a root of multiplicity k of $f(x)$. This means that $f(x) = (x-\alpha)^k g(x)$ where α in not a root of $g(x)$. According to (17) we have $f'(x) = ((x-\alpha)^k)'g(x) + (x-\alpha)^k g'(x)$, and according to (19) we have $((x-\alpha)^k)' = k(x-\alpha)^{k-1}$ (since $(x-\alpha)' = 1$, by (15)). Therefore, $f'(x) = k(x-\alpha)^{k-1}g(x)+(x-\alpha)^k g'(x) = (x-\alpha)^{k-1}p(-x)$, where $p(x) = kg(x)+(x-\alpha)g'(x)$. But α is not a root of $p(x)$: $p(\alpha) = kg(\alpha) \neq 0$. Consider the polynomials $d(x) = \text{g. c. d.}(f(x), f'(x))$ and $\varphi(x) = \dfrac{f(x)}{d(x)}$ (since $d(x)$ is a divisor of $f(x)$, $\varphi(x)$ is a polynomial). The polynomial $d(x)$ is divisible by $(x-\alpha)^{k-1}$ since $f(x)$ and $f'(x)$ are divisible by $(x-\alpha)^{k-1}$. But $d(x)$ is not divisible by $(x-\alpha)^k$, since $p(\alpha) \neq 0$ which means that $p(x)$ is not divisible by $x-\alpha$. We conclude that $\varphi(x)$ is divisible only by $x-\alpha$ (and no higher power, e.g. $(x-\alpha)^2$, etc). Since $\varphi(x)$ is defined independently from the root (namely $\varphi(x) = \dfrac{f(x)}{\text{g. c. d.}(f(x), f'(x))}$) the above conclusion is true for all the roots of $f(x)$, and we see that $\varphi(x)$ has the same roots as $f(x)$, but none of them is multiple. In view of this, we can always reduce a question regarding the roots of a polynomial to the case when the polynomial has no multiple roots.

Notice that we have implicitly met with the derivative in connection with the formula for the interpolation polynomial. Indeed, let $F(x) = (x-x_1)\dots(x-x_{n+1})$. From (14) we see that $(x-x_i)' = 1$. Therefore, formula (18) gives:

$$F'(x) = (x-x_2)\cdots(x-x_{n+1}) + (x-x_1)(x-x_3)\cdots(x-x_{n+1}) + \\ + \cdots + (x-x_1)(x-x_2)\cdots(x-x_n).$$

If we use the notation $F_k(x) = \dfrac{F(x)}{x-x_k}$ from Section I, then $F'(x) = F_1(x) + \cdots + F_{n+1}(x)$. Substituting now for x one of the values $x = x_k$, since all $F_i(x)$ for $i \neq k$ contain the factor $x - x_k$, we see that $F_i(x_k) = 0$. Therefore $F'(x_k) = F_k(x_k)$ and the formula (10) can be written in the form

$$f(x) = \frac{y_1}{F'(x_1)}F_1(x) + \frac{y_2}{F'(x_2)}F_2(x) + \frac{y_{n+1}}{F'(x_{n+1})}F_{n+1}(x).$$

Problems

1. Polynomial $x^{2n} - 2x^n + 1$ clearly has a root $x = 1$, and by Bézout's theorem it is divisible by $x - 1$. Find the quotient.

2. For which values of a, b does the polynomial $x^n + ax^{n-1} + b$ have a multiple root? Find this root.

3. For which values of a, b does the polynomial $x^3 + ax + b$ have a multiple root?

4. Prove that the polynomial $x^n + ax^m + b$ cannot have nonzero root of multiplicity 3 or more.

5. The derivative of the polynomial $f'(x)$ is called the *second derivative* of the polynomial $f(x)$ and is denoted by $f''(x)$. Find the formula for $(f_1 f_2)''$, analogous to (17), but which will be, of course, somewhat more complicated.

6. Prove that the derivative of a polynomial is identically equal to 0 if and only if the polynomial is constant (i.e. when its degree is 0).

7. Prove that for a polynomial $f(x)$ there exists a polynomial $g(x)$ such that $g'(x) = f(x)$ and that all such polynomials $g(x)$ (for a given $f(x)$) can differ from each other only in the constant term.

8. Prove that the number of roots of a polynomial cannot be greater than its degree, each root being counted k times if k is its multiplicity.

3. The binomial formula

In this section we shall be concerned with an important formula which expresses the polynomial $(1+x)^n$ in the usual form $a_0 + a_1 x + \cdots + a_n x^n$. In order to find the formula we have to multiply out all the factors $(1+x)(1+x)\cdots(1+x)$. Working out these brackets we shall obtain terms of the form x^k, but such terms will appear several times, and by grouping them together we shall arrive at the required formula. For instance, if $n = 2$, it is well known that

$$(1+x)^2 = (1+x)(1+x) = 1(1+x) + x(1+x) = 1 + x + x + x^2 = 1 + 2x + x^2.$$

For $n = 3$ the formula is also probably known. If not, it is easily obtained when the formula for $(1+x)^2$ is multiplied by $1+x$:

$$\begin{aligned}(1+x)^3 &= (1+x)^2(1+x) = (1+2x+x^2)(1+x)\\ &= (1+2x+x^2) + (1+2x+x^2)x = 1 + 3x + 3x^2 + x^3.\end{aligned}$$

The coefficient a_k of x^k in the polynomial $(1+x)^n$ depends on the index k, but also on the degree n. In order to indicate this dependence on n and k, we denote this coefficien by C_n^k. Therefore, C_n^k are *by definition* the coefficients in the formula

$$(1+x)^n = C_n^0 + C_n^1 x + C_n^2 x^2 + \cdots + C_n^n x^n. \tag{20}$$

For example, $C_2^0 = 1$, $C_2^1 = 2$, $C_2^2 = 1$; $C_3^0 = 1$, $C_3^1 = 3$, $C_3^2 = 3$, $C_3^3 = 1$. The coefficients C_n^k are called *binomial coefficients*. Our aim is to write them in an explicit form. Notice that some of them are easy to find. It is clear that when we multiply all the x's by one another in the product $(1+x)^n$, we get x^n, which means that the leading term of the polynomial $(1+x)^n$ is x^n, i.e.

$$C_n^n = 1. \tag{21}$$

Similarly, multiplying the constant terms (values for $x = 0$) in the product $(1+x)^n$ we see that the constant term of the polunomial $(1+x)^n$ is 1, i.e.

$$C_n^0 = 1. \tag{22}$$

In the general case consider the derivatives of both sides of (20). On the left, according to (19), we get $n(1+x)^{n-1}$, since $(1+x)'=1$, by (15). We evaluate the derivative on the right using (15). We obtain

$$n(1+x)^{n-1}=C_n^1+2C_n^2x+\cdots+kC_n^kx^{k-1}+\cdots+nC_n^nx^{n-1}.$$

But we can apply (20) for $n-1$ to the left-hand side of the above equality. The coefficient of x^{k-1} will be nC_{n-1}^{k-1} on the left and kC_n^k on the right. Therefore, $kC_n^k=nC_{n-1}^{k-1}$, or

$$C_n^k=\frac{n}{k}\,C_{n-1}^{k-1},$$

i.e. the coefficient C_n^k can be expressed in terms of the coefficient C_{n-1}^{k-1} with smaller indices. Applying this formula to C_{n-1}^{k-1} we get $C_n^k=\dfrac{n(n-1)}{k(k-1)}C_{n-2}^{k-2}$, and repeating the process r times we obtain the formula

$$C_n^k=\frac{n(n-1)\cdots(n-r+1)}{k(k-1)\cdots(k-r+1)}C_{n-r}^{k-r}$$

(we take away from n in the numerator and from k in the denominator r consecutive values: $0,1,\ldots,r-1$). Finally, let $r=k$. Since we know that $C_m^0=1$ for any m, we obtain the formula for C_n^k:

$$C_n^k=\frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots1}. \tag{23}$$

This is the formula we looked for.

Formula (20) with the explicit expression (23) for the binomial coefficients C_n^k is called the binomial formula (or "Newton's binomial").

The binomial formula has a large number of applications and it is useful to have the coefficients (23) written in various forms. In the denominator we have the product of all positive integers from 1 to k. The product of the form $1\cdot2\cdot\ldots\cdot m$ is called *m factorial* and denoted by $m!$. In the numerator we have the product of all positive integers from n to $n-k+1$. If we multiply it by the product of the numbers from $n-k$ to 1 (i.e. by $(n-k)!$) we obtain $n!$. Therefore, multiplying the numerator and the denominator in (23) by $(n-k)!$, we get

$$C_n^k=\frac{n!}{k!\,(n-k)!}, \tag{24}$$

and this implies that

$$C_n^k=C_n^{n-k}. \tag{25}$$

Notice that in formulas (23) and (24) it is not immediately clear that the denominator divides the numerator, although we know that this is so having in mind the meaning of the coefficients C_n^k in the formula (20). We can express the fact that the expression on the right-hand side of (23) is an integer, by simply saying that the *product of any k consecutive integers is divisible by $k!$*. We shall see later that the fact that right-hand sides of (23) and (24) are integers implies some interesting properties of prime numbers.

We now establish some important properties of the coefficients C_n^k. The first one follows from the obvious equality $(1+x)^n = (1+x)^{n-1}(1+x)$ after expanding $(1+x)^n$ and $(1+x)^{n-1}$ on the basis of (20). We obtain

$$C_n^0 + C_n^1 x + C_n^2 x^2 + \cdots + C_n^n x^n = (C_{n-1}^0 + C_{n-1}^1 x + \cdots + C_{n-1}^{n-1} x^{n-1})(1+x).$$

The coefficient of x^k on the left is C_n^k and on the right is obtained from the sum of the terms $C_{n-1}^k x^k \cdot 1$ and $C_{n-1}^{k-1} x^{k-1} \cdot x$, i.e. it is $C_{n-1}^k + C_{n-1}^{k-1}$. Therefore

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}. \tag{26}$$

This is a very useful formula for evaluating coefficients C_n^k by means of the coefficients of index $n-1$. In order to get a better visual representation, we write the coefficients C_n^k in the form of a triangle, where C_n^k are in the n-th row. Using the formulas (21) and (22), which say that at the beginning and at the end of each row is 1, the triangle has the form

$$\begin{array}{c}
1 \\
1 \quad\quad 1 \\
1 \quad\quad 2 \quad\quad 1 \\
\ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \\
1 \quad\quad C_{n-1}^2 \quad\quad \ldots \quad\quad C_{n-1}^{n-2} \quad\quad 1 \\
1 \quad\quad C_n^1 \quad\quad C_n^2 \quad\quad \ldots \quad\quad C_n^{n-1} \quad\quad 1 \\
\ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots
\end{array}$$

Formula (26) shows that each binomial coefficient C_n^k is equal to the sum of the coefficients which are situated above on the left and right of it. Taking the first two rows as given, we easily obtain for the subsequent coefficients:

$$\begin{array}{c}
1 \\
1 \quad\quad 1 \\
1 \quad\quad 2 \quad\quad 1 \\
1 \quad\quad 3 \quad\quad 3 \quad\quad 1 \\
1 \quad\quad 4 \quad\quad 6 \quad\quad 4 \quad\quad 1 \\
1 \quad\quad 5 \quad\quad 10 \quad\quad 10 \quad\quad 5 \quad\quad 1 \\
\ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots \quad\quad \ldots
\end{array}$$

This triangle is called “Pascal’s triangle”.

The second property is obtained by putting $x = 1$ into the formula (20) which defines the binomial coefficients. On the left we get 2^n and on the right the sum of all binomial coefficients C_n^k for $k = 0, 1, \ldots, n$. Therefore, *the sum of all numbers from the n-th row of Pascal's triangle is equal to 2^n.*

Finally, consider two neighbouring members from one row: C_n^{k-1} and C_n^k. According to (24) we have $C_n^k = \dfrac{n!}{k!\,(n-k)!}$, $C_n^{k-1} = \dfrac{n!}{(k-1)!\,(n-k+1)!}$. Since $k! = (k-1)!\,k$, $(n-k+1)! = (n-k)!\,(n-k+1)$, we get

$$C_n^k = \frac{n-k+1}{k}\,C_n^{k-1}.$$

It is evident that $\dfrac{n-k+1}{k} > 1$ when $n-k+1 > k$, i.e. $k < \dfrac{n+1}{2}$ and in that case $C_n^k > C_n^{k-1}$. Conversely, if $k > \dfrac{n+1}{2}$, we obtain $C_n^k < C_n^{k-1}$. Therefore, *the numbers in one row of Pascal's triangle increase up to the middle of the row, and after that they decrease.* If n is even, then in the middle of the row we have the greatest number $C_n^{n/2}$, and if n is odd, then there are two neighbouring equal greatest numbers: $C_n^{(n-1)/2}$ and $C_n^{(n+1)/2}$. In that case, for $k = \dfrac{n+1}{2}$ we have $C_n^k = C_n^{k-1}$.

The formula (20), where the binomial coefficients are defined by (23) can be written in a somewhat more general form. In order to do that, put $x = b/a$ and multiply both sides of (20) by a^n. We obtain the formula

$$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \cdots + C_n^n b^n. \tag{27}$$

This formula was proved for $a \neq 0$ (since we divided by a), but it is obviously true also for $a = 0$. It is also called the binomial formula.

We shall now consider some consequences of the binomial formula and their applications. As a rule, the simpler a result is, the more applications it has. So, in the binomial formula we often use the values of the first coefficients. We already know that the first coefficient C_n^0 is 1. The next one C_n^1, according to (23) is n. Notice that in view of (25) it follows that $C_n^n = 1$ (which we already know) and that $C_n^{n-1} = n$. Hence,

$$(a+b)^n = a^n + na^{n-1}b + \cdots + nab^{n-1} + b^n.$$

This can be applied to equations. We write an equation of order n in the form $a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n = 0$. The fact that its degree is n means that $a_n \neq 0$ and we can divide the equation by a_n to obtain an equivalent equation in which $a_n = 1$. In further text we shall suppose that this has been done and we write the equation in the form $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n = 0$. We shall now make another transformation of this equation into an equivalent equation. In order to do so, put $x = y + c$, where y is the new variable and c is a number. Substituting into our equation this value of x, from each term $a_m x^m$ we obtain the term $a_m(y+c)^m$ which can be, by the binomial formula, written as a polynomial in y, and then we collect together corresponding terms. As a result we obtain a new polynomial in y which we denote by $g(y) = f(y+c)$. Since y is expressed in terms of x: $y = x - c$, the equations $f(x) = 0$ and $g(y) = 0$ are equivalent: to the root $x = \alpha$ of $f(x) = 0$ corresponds the root $y = \alpha - c$ of $g(y) = 0$, and to the root $y = \beta$ of this equation corresponds the root $x = \beta + c$ of the equation $f(x) = 0$. Let us examine how the coefficients change in this transformation. First of all, the degree of the equation $g(y) = 0$ is n and the coefficient of its leading term is 1. This follows from the fact that when $a_m(y+c)^m$ is expanded by the binomial formula, it gives rise to terms in y with degrees $\leqslant m$. Therefore, the term of degree n can only be obtained from $(y+c)^n$ and (again by the binomial formula) it is equal to y^n. Let us look at the term of degree $n-1$. It can be obtained from the term $(y+c)^n$ and the term $a_{n-1}(y+c)^{n-1}$. From the last one we get $a_{n-1}y^{n-1}$, and in $(y+c)^n$ we have to take the *second* term in the binomial expansion. As we know it is equal to $ny^{n-1}c$. Hence, the term of degree $n-1$ in the polynomial $g(y) = f(y+c)$ has the form $(a_{n-1} + nc)y^{n-1}$.

This can be used to simplify the equation by chosing c so that the term of degree $n-1$ vanishes: we put $a_{n-1} + nc = 0$, i.e. $c = -a_{n-1}/n$. We proved the following

Theorem 6. *The substitution $x = y - a_{n-1}/n$ transforms the equation $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n = 0$ into equivalent equation $g(y) = 0$ of degree n whose coefficient of the leading term is 1 and which has no term of degree $n-1$.*

Notice that Theorem 6 gives the formula for the solutions of second degree equations. Indeed, the polynomial $g(y)$ has the form $y^2 + b_2$ and its roots are therefore $y = \pm\sqrt{-b_2}$. Make the substitution indicated in Theorem 6, evaluate b_2 and the roots of $f(x)$ and verify that in this way we obtain the standard formula for the solutions of a quadratic equation. In the case of polynomials of arbitrary degree we only get a certain simplification, which is sometimes useful. For example, we see that any third degree equation is equivalent to an equation of the form $x^3 + ax + b = 0$.

At the end we shall apply the binomial formula to the evaluation of the sums of powers of integers. We shall be concerned with the sums

$$S_m(n) = 0^m + 1^m + 2^m + \cdots + n^m \tag{28}$$

of m-th powers of all nonnegative integers not greater than n. You probably know the formula $S_1(n) = \dfrac{n(n+1)}{2}$ (see Problem 5 in Section 1 of Chapter I). We start with a few remarks on the evaluation of sums in general. Let $a_0, a_1, a_2, \ldots, a_n, \ldots$ be an arbitrary infinite sequence of numbers, and consider the following sequence of their sums: a_0, $a_0 + a_1$, $a_0 + a_1 + a_2$, $\ldots$, $a_0 + a_1 + a_2 + \cdots + a_n$, $\ldots$ Denote the first sequence by the letter a; its $(n+1)$-st term is a_n (it is more convenient to write the $(n+1)$-st term and not the n-th, and to start the sequence with a_0). The above sequence of sums will be denoted by Sa, and its $(n+1)$-st term is

$$(Sa)_n = a_0 + a_1 + a_2 + \cdots + a_n, \qquad n = 0, 1, 2, \ldots$$

For example, if $a_n = n^m$, $n = 0, 1, 2, \ldots$, then Sa is the sequence of sums $S_m(n)$. Clearly, if we know the sequence Sa we can find the sequence a. Namely, by substracting the n-th from the $(n+1)$-st term of Sa, we obtain a_n. Indeed, let

$$b_n = (Sa)_n = a_0 + a_1 + \cdots + a_n, \tag{29}$$

$$b_{n-1} = (Sa)_{n-1} = a_0 + a_1 + \cdots + a_{n-1}, \tag{30}$$

and subtracting (30) from (29) we get $b_n - b_{n-1} = a_n$. We introduce another important construction. Together with an arbitrary sequence $b_0, b_1, b_2, \ldots, b_n, \ldots$ consider the sequence $b_0, b_1 - b_0, b_2 - b_1, \ldots, b_{n+1} - b_n, \ldots$ If the first sequence is denoted by b, then the second is denoted by Δb. Its $(n+1)$-st term is

$$(\Delta b)_0 = b_0, \quad (\Delta b)_n = b_n - b_{n-1}, \qquad n = 1, 2, \ldots$$

The established connection between the sequences a and Sa can be expressed by the formula $\Delta Sa = a$. It turns out that there is a formula completely symmetrical to this one, namely both equalities

$$\Delta Sa = a, \qquad S\Delta b = b \tag{31}$$

are true. It can be said that the operations S and Δ applied to sequences are inverse to each other. We have already established the first formula. In order to prove the second, write the equalities which define the numbers $a_k = (\Delta b)_k$ for $k = 0, 1, \ldots, n-1$:

$$\begin{aligned} a_0 &= b_0 \\ a_1 &= b_1 - b_0 \\ a_2 &= b_2 - b_1 \\ &\ldots \\ a_n &= b_n - b_{n-1} \end{aligned}$$

and add them up. On the left we obtain $a_0 + \cdots + a_n$, i.e. $(Sa)_n$, and on the right all the numbers cancel out, except b_n in the last formula, so that we get $(Sa)_n = b_n$, that is to say the second formula (31).

The above relations are useful, since it is often simpler not to evaluate a sum directly, i.e. not to find the sequence Sa directly, but instead to find a sequnce b such that $\Delta b = a$, and from the second relation (31) to obtain $Sa = b$.

This idea will now be applied to the sums (28). We have seen that $S_m(n) = (Sa)_n$, where $a_n = n^m$. How can we write the sequence a, $a_n = n^m$ in the form $a = \Delta b$? This follows from the following assertion.

Theorem 7. *For any polynomial $f(x)$ of degree m there exists a unique polynomial $g(x)$ of degree $m+1$ such that*

$$g(x) - g(x-1) = f(x) \tag{32}$$

and the constant term of $g(x)$ is 0.

The uniqueness of the polynomial $g(x)$ with the given property is easily shown. Let $g_1(x)$ be another polynomial such that $g_1(x) - g_1(x-1) = f(x)$ and whose constant term is 0. Subtract (32) from the last equality. Putting $g_1(x) - g(x) = g_2(x)$, we see that $g_2(x) - g_2(x-1) = 0$ and the constant term of $g_2(x)$ is 0, i.e. $g_2(0) = 0$. For $x = 1$, the above equality gives $g_2(1) = 0$. Putting $x = 2$ we get $g_2(2) = g_2(1) = 0, \ldots$ and by induction that $g_2(n) = 0$ for all positive integers n. In other words, all positive integers are roots of $g_2(x)$. According to Theorem 3 this is possible only if $g_2 = 0$, which means that $g = g_1$.

The existence of the polynomial g will be proved by induction on m, the degree of $f(x)$. For $m = 0$, the polynomial f is a constant a and we see that $g(x) = ax$ satisfies (32). Suppose that the assertion is true for polynomials f of degree less than m. Let $a_m x^m$ be the leading term of f. Choose the number a so that the leading term of the polynomial $ax^{m+1} - a(x-1)^{m+1}$ is equal to the leading term $a_m x^m$ of f. In order to do this, apply the binomial formula

$$(x-1)^{m+1} = x^{m+1} - (m+1)x^m + \cdots,$$

where the dots stand for terms of degree less than m. This implies

$$x^{m+1} - (x-1)^{m+1} = (m+1)x^m + \cdots.$$

Clearly we have

$$a = \frac{a_m}{m+1}. \tag{33}$$

Then, in the difference $f(x) - \dfrac{a_m}{m+1}(x^{m+1} - (x-1)^{m+1})$ the terms of degree m cancel out and this difference will have degree less than m. Denoting this polynomial by $h(x)$, by the induction hypothesis we can take that there is a polynomial g_1 of degree less than $m+1$ and with zero constant term such that $h(x) = g_1(x) - g_1(x-1)$, i.e.

$$f(x) - \frac{a_m}{m+1}(x^{m+1} - (x-1)^{m+1}) = g_1(x) - g_1(x-1).$$

The above equality can be written in the form $f(x) = g(x) - g(x-1)$, where

$$g(x) = \frac{a_m}{m+1}x^{m+1} + g_1(x),$$

and the theorem is proved. Of course, in practical construction we do not apply induction, but we repeat the same procedure of subtraction to the polynomial $h(x)$, and so on until we arrive at a polynomial of degree 0.

Return now to the evaluation of the sum $S_m(n)$. We have seen that this sum is equal to b_n, where b is such that $\Delta b = a$, $a_n = n^m$. Apply Theorem 7 to the polynomial x^m. We obtain the polynomial $g(x)$ of degree $m+1$ such that

$$g(x) - g(x-1) = x^m$$

and the constant term of $g(x)$ is 0. Putting $x = n$ into the above equality, we see that the sequence $b_n = g(n)$ for $n \geqslant 1$ and $b_0 = g(0) = 0$ satisfies the condition $\Delta b = a$, i.e. $Sa = b$. Therefore we have proved the following

Theorem 8. *The sums $S_m(n)$ can be expressed in the form $g_m(n)$ where g_m is the polynomial of degree $m+1$ such that $g_m(x)-g_m(x-1)=x^m$ and its constant term is* 0.

Notice that the proof of Theorem 7 provides us with a method of constructing the polynomial $g_m(x)$ for any m. For instance, let $m=2$. By analogy with sequences, we denote the polynomial $g(x)-g(x-1)$ by Δg, i.e. we put $(\Delta g)(x)=g(x)-g(x-1)$. We first have to find the monomial ax^3 so that the leading term of $\Delta(ax^3)$ is equal to x^2. In view of (33), $a=\dfrac{1}{3}$ (in this case $m=2$, $a_2=1$). By the binomial formula, $\Delta\left(\dfrac{1}{3}x^2\right)=\dfrac{1}{3}x^3-\dfrac{1}{3}(x-1)^3=x^2-x+\dfrac{1}{3}$ and $x^2-\Delta\left(\dfrac{1}{3}x^3\right)=x-\dfrac{1}{3}$. Now we have to find the monomial bx^2 so that the leading coefficient of $\Delta(bx^2)$ is equal to x. In view of (33), $b=\dfrac{1}{2}$ (in this case $m=1$, $a_1=1$) and by the binomial formula $\Delta\left(\dfrac{1}{2}x^2\right)=\dfrac{1}{2}x^2-\dfrac{1}{2}(x-1)^2=x-\dfrac{1}{2}$, and $x^2-\Delta\left(\dfrac{1}{3}x^3\right)-\Delta\left(\dfrac{1}{2}x^2\right)=-\dfrac{1}{3}+\dfrac{1}{2}=\dfrac{1}{6}$. Finally, $\dfrac{1}{6}=\Delta\left(\dfrac{1}{6}x\right)=\dfrac{1}{6}x-\dfrac{1}{6}(x-1)$. At the end we get that $x^2=\Delta\left(\dfrac{1}{3}x^3+\dfrac{1}{2}x^2+\dfrac{1}{6}x\right)$ and so $g(x)=\dfrac{1}{3}x^3+\dfrac{1}{2}x^2+\dfrac{1}{6}x=\dfrac{(2x^2+3x+1)x}{6}=\dfrac{(2x+1)(x+1)x}{6}$. Therefore $S_2(n)=\dfrac{(2n+1)(n+1)n}{6}$.

We conclude with two more remarks.

Remark 1. The obtained formula for the sum $S_m(n)$ can be summarized as follows. For each m there exists the unique polynomial $g_m(x)$ with constant term 0 such that $g_m(x)-g_m(x-1)=x^m$. The method of its construction is contained in the proof of Theorem 7. Its degree is $m+1$. The formula for $S_m(n)$ is: $S_m(n)=g_m(n)$. Hence the question reduces to the investigation of the important polynomials $g_m(x)$. They are called *Bernoulli's polynomials.* In the Appendix we shall give a much more explixit expression for these polynomials, using an important sequence of rational numbers, called *Bernoulli's numbers.*

Remark 2. (Historical) The introduced operations S and Δ which transform the sequences a and b into Sa and Δb are very similar to the fundamental operations of Analysis which define for a function $f(x)$ (but not for every function!) the *indefinite integral* $\int f\,dx$, and for a function g its *derivative* g'. Our operations S and Δ are elementary analogs of the operations $\int f\,dx$ and g'. Sums and differences are also present in the definitions of the integral and the derivative, but in a more complicated way (in our definition of the derivative of a polynomial differences were also present—see formula (13)). As in the case of S and Δ, the operations of forming the derivative and the integral are inverse to each other. As in our case, the evaluation of the derivative is simpler than the evaluation of the integral, and the integral of a function $f(x)$ is mainly evaluated by finding a function whose derivative is equal to $f(x)$.

However, the operations for sequences and functions are not only analogous;

Fig. 1

their connection is deeper. Evaluating the integral of a function $f(x)$ is equivalent to evaluating the area of the surface bounded by the graph of that function, the x-axis and by two vertical lines starting at $x = a$ and $x = b$ (Fig. 1).

Of course, we shall not prove this, as we have not defined the integral, but we shall show, on a simple example, how such an area can be evaluated, and its connection with the problems we considered earlier.

We shall try to determine the area bounded by the parabola which is the graph of the function $y = x^2$, by the x-axis and by the line $x = 1$ (Fig. 2).

Fig. 2

In order to do that, divide the segment between 0 and 1 into a large number n of equal parts with coordinates $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$ and evaluate the corresponding values $0, \left(\frac{1}{n}\right)^2, \left(\frac{2}{n}\right)^2, \dots, \left(\frac{n-1}{n}\right)^2, 1$ of the function $y = x^2$. Construct the rectangles whose bases are segments from $\frac{i}{n}$ to $\frac{i+1}{n}$ and whose heights are $\left(\frac{i}{n}\right)^2$. The polygon made up from these rectangles is contained in that part "under the parabola" whose area we wish to determine and by "looking at the picture" we see that if n is very great, then the area s_n of this polygon differs very little from the area of the part under the parabola (we cannot be more precise, since we have not precisely defined what *area* is). The area of the polygon is the sum of the areas of the rectangles which make it up. The area of the i-th rectangle is equal to the product of its basis $\frac{1}{n}$ and its height $\left(\frac{i}{n}\right)^2$, i.e. it is $\frac{i^2}{n^3}$. Therefore, the area s_n of the polygon is

$$s_n = \frac{0^2}{n^3} + \frac{1^2}{n^3} + \frac{2^2}{n^3} + \cdots + \frac{(n-1)^2}{n^3} = \frac{S_2(n-1)}{n^3}.$$

We have already found that $S_2(n) = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$, and so (replacing n by $n-1$ in the formula for $S_2(n)$) we get

$$s_n = \frac{1}{3} - \frac{1}{2}\cdot\frac{1}{n} + \frac{1}{6}\cdot\frac{1}{n^2}.$$

It is clear that as n becomes greater and greater, then the terms $-\frac{1}{2}\cdot\frac{1}{n}$ and $\frac{1}{6}\cdot\frac{1}{n^2}$ become smaller and smaller, and the area of the polygon approaches $\frac{1}{3}$. Hence, this is the area of the figure bounded by the parabola.

We have presented here the lines of thought followed, in principle, by Archimedes (3rd century B.C.) who was the first to solve this problem. (Archimedes devised a rather artificial method which allowed him to use the sum of a geometrical progression, instead of the sum $S_2(n)$. But he knew the formula for $S_2(n)$ and used it for the evaluation of other areas and volumes).

Mathematicians of the new period were obsessed by the dream to "surpass the ancients" (that is to say, the Ancient Greek mathematicians) and Archimedes was considered to be the most important of them. They were therefore very much interested in solving the problem considered above for the function $y = x^n$, where n is greater than 2. It seems that the first to obtain the solution was French mathematician Fermat (17th century) who used practically the same method we outlined above (it was later somewhat simplified). At that time the mentioned connection between the integral and the derivative was not known and the integral (i.e. the area) was calculated directly from the definition. It was later discovered that (to use contemporary terminology) the operations of forming derivatives and integrals are inverse to each other. This was established by Newton's teacher Barrow. (Newton worked together with Barrow when he studied at the university, and later on took over Barrow's chair when the latter decided to take orders). Systematic evaluation of the integral of a function f by finding a function g such that the derivative of g is f was initiated by Newton. After that the calculation of integrals and areas by the method we exposed became unnecessary. Nowadays students of higher classes can easily find the integral of x^m for any m without caluclating the sum $S_m(n)$.

In this way, if in Chapter I we moved in the circle of ideas of Ancient Greek mathematicians (Pythagoras, Theaetetus, Euclid), in this chapter we have encountered the ideas of the mathematicians from the new period (17th century).

Problems

1. Notice that the area s_n of the polygon which we calculated at the end of the section is *less* than the area of the given figure, bounded by the parabola $y = x^2$, since the polygon is situated inside that figure. Construct the polygon made up from rectangles whose bases are segments from $\frac{i}{n}$ to $\frac{i+1}{n}$ and whose heights are $\left(\frac{i+1}{n}\right)^2$ which *contains* the given figure. Its area s'_n will therefore be *greater* than the area of the figure. Calculate the area s'_n and prove that as n increases, it approaches 1/3. This gives a more convincing (i.e. more "strict") proof of the fact that the required area is 1/3.

2. Try to solve the analogous problem for the "m-th degree parabola", given by the equation $y = x^m$. Verify that in order to obtain the result it is not necessary to know the Bernoulli's polynomials $g_m(x)$ completely, but that is enough to know the coefficient of the leading term $a_{m+1}x^{m+1}$. Prove that $a_{m+1} = \frac{1}{m+1}$ and hence find the area of the figure bounded by the parabola whose equation is $y = x^m$, by the x-axis and by the line $x = 1$.

3. Prove that the area of the figure bounded by the parabola $y = x^m$, x-axis and the line $x = a$ is equal to $\frac{1}{m+1}a^{m+1}$. Notice that the derivative of the polynomial $\frac{1}{m+1}x^{m+1}$ is x^m. This is indeed an instance of Barrow's theorem that integration and finding derivatives are operations inverse to each other.

4. Prove that the sum of the binomial coefficients with even upper indices $C_n^0 + C_n^2 + \cdots$ and with odd indices $C_n^1 + C_n^3 + \cdots$ are equal and find their mutual value.

5. Find the relation between binomial coefficients which expresses that $(1 + x)^n(1 + x)^m = (1 + x)^{n+m}$. For $n = m$ deduce the formula for the sum of the squares of binomial coefficients.

6. If p is a prime number, prove that all binomial coefficients C_p^k for $k \neq 0, p$, are divisible by p. Deduce from this that $2^p - 2$ is divisible by p. Prove that for any integer n, the number $n^p - n$ is divisible by p. This theorem was first proved by Fermat.

7. What can be said about the sequence a if all the terms of the sequence Δa are equal? What does formula (31) give in this case?

8. Find the sum $S_3(n)$ and verify that $S_3(n) = (S_1(n))^2$.

9. Let a be any sequence $a_0, a_1, a_2, \ldots$ Apply the operation Δ once more to the sequence Δa. The obtained sequence $\Delta(\Delta a)$ will be denoted by $\Delta^2 a$. Define $\Delta^k a$ by induction as $\Delta(\Delta^{k-1}a)$. When can we solve the so-called "infinite interpolation problem", that is to say when is there a polynomial $f(x)$ of degree not greater than m such that $f(n) = a_n$ for $n = 0, 1, 2, \ldots$? Prove that a necessary and sufficient condition is given by $(\Delta^{m+1}a)_n = 0$ for $n \geqslant m$. This condition shows that if we write the sequence a, and under it the sequence whose terms are differences of the two terms above, and so on:

$$\begin{array}{ccccccccccc}
a_0 & & a_1 & & a_2 & & \ldots & & a_n & & a_{n+1} \\
 & a_1 - a_0 & & a_2 - a_1 & & \ldots & & \ldots & & a_{n+1} - a_n & \ldots \\
 & & \ldots & & \ldots & & \ldots & & \ldots & & \ldots
\end{array}$$

then in the $(m+1)$-st row we only have zeros.

Is there a polynomial $f(x)$ such that $f(n) = 2^n$ for all positive integers n?

10. Prove that if $a_n = q^n$, then $(\Delta a)_n = a_{n-1}(q-1)$. Use this to give another proof of the formula (12) from Chapter I.

11. Let $m_1 < m_2 < \cdots < m_{n+1}$ be positive integers and let $f(x)$ be a polynomial of degree n, where the coefficient of x^n is 1. Prove that at least one of the numbers $f(m_k)$ is not less than $n!/2^n$.

Hint. Use the result of Problem 8 of Section 1. Notice (with the notation of Section 1) that $F_k(m_k) \geqslant k!\,(n-k)!$ and use some known relations for binomial coefficients.

12. Apply the formula (12) from Chapter I to the sum $1 + (1+x) + (1+x)^2 + \cdots + (1+x)^n$. Equating the coefficients of terms with equal degrees on the left and right, find the formula for the sum

$$C_k^k + C_{k+1}^k + \cdots + C_n^k.$$

APPENDIX[1]

Bernoulli's polynomials and numbers

In Section 3 we showed that the values of the sums of powers of consecutive positive integers, i.e. the sums $S_m(n)$ coincide with the values $g_m(n)$ of Bernoulli's polynomials $g_m(x)$, which have the properties

$$\begin{aligned} &1)\ g_m(x) - g_m(x-1) = x^m, \\ &2)\ \text{the constant term of } g_m(x) \text{ is } 0. \end{aligned} \tag{1}$$

For any m there is only one polynomial of degree $m+1$ with these properties.

We have given a method for constructing the polynomials $g_m(x)$. However, we would like to have a more explicit formula for these polynomials. In order to achieve this, we shall follow the same path we took in deducing the binomial formula. Namely, we shall first find the derivative of both sides of (1). But we first have to see how to find the derivative $f(x-1)'$ of the polynomial $f(x-1)$.

Lemma 1. $f(x-1)' = f'(x-1)$.

[1] Starting with Chapter II, each chapter will have an Appendix. In these appendices we shall only use those facts which have been exposed earlier, but the text will be somewhat harder than the basic text. This means that we shall apply the same arguments as before, but in proving a theorem we shall have to keep in mind a larger number of them. The level of these texts approaches the level of a simple professional mathematical book.

At first sight this equality may seem obvious, but it is not really so. The equality means that when in the polynomial $f(x)$ we first substitute $x-1$ for x, then write it as the sum of powers of x, and then find its derivative, the obtained result is the same as when in the derivative $f'(x)$ we substitute $x-1$ for x.

The proof follows directly from the definition of the derivative of a polynomial, i.e. from (11). Let

$$f(x)-f(\alpha)=(x-\alpha)g(x,\alpha).$$

Substituting $x-1$ for x and $\alpha-1$ for α in this equality, we get

$$f(x-1)-f(\alpha-1)=(x-\alpha)g(x-1,\alpha-1).$$

By definition we have $f(\alpha-1)'=g(\alpha-1,\alpha-1)$ and $f'(\alpha)=g(\alpha,\alpha)$. Hence, $f(\alpha-1)'=f'(\alpha-1)$, which was to be proved.

Lemma 1 could be proved by the use of formulas (16)–(19) and reduction to monomials (verify this).

We can now find the derivatives of both sides of (1). Having in mind Lemma 1 and the rule (13) for derivatives, we obtain

$$g_m'(x)-g_m'(x-1)=mx^{m-1}.$$

On the other hand, replacing m by $m-1$ in (1) we get

$$g_{m-1}(x)-g_{m-1}(x-1)=x^{m-1}.$$

Multiply the second equality by m and subtract it from the first. Putting $h_m=mg_{m-1}-g_m'$ we find that

$$h_m(x)=h_m(x-1).$$

But this implies that the polynomial h_m is constant (of degree 0). Indeed, putting in this equality $x=1$, 2, etc. we obtain $h_m(0)=h_m(1)=h_m(2)=\cdots$. In other words the polynomial $h_m(x)$ and the constant $h_m(0)$ have equal values for all positive integers x, and in view of Theorem 4 they must be equal: $h_m(x)=h_m(0)$. (We have already met with this kind of reasoning at the beginning of the proof of Theorem 7.) Hence, the polynomial $h_m(x)$ is equal to a constant which we shall denote by $-\alpha_m$. Having in mind the definition of $h_m(x)$ we obtain the relation

$$g_m'=mg_{m-1}+\alpha_m. \tag{2}$$

As in the derivation of the binomial formula, we now write the polynomial $g_m(x)$ as the sum of powers of x. As before, the lower index indicates the polynomial in question, and upper index corresponds to the degree of x. The coefficients are denoted by A_m^k and $g_m(x)$ has the form

$$g_m(x)=A_m^1x+A_m^2x^2+\cdots+A_m^kx^k+\cdots+A_m^{m+1}x^{m+1}.$$

(Remember that the constant term of g_m is 0.) Write down the derivative of $g_m(x)$, using the formula (13):

$$g_m(x)'=A_m^1+2A_m^2x+\cdots+kA_m^kx^{k-1}+\cdots+(m+1)A_m^{m+1}x^m.$$

On the other hand, write down the analogous formula for $g_{m-1}(x)$ (replacing m by $m-1$):

$$g_{m-1}(x) + A^1_{m-1}x + A^2_{m-1}x^2 + \cdots + A^k_{m-1}x^k + \cdots + A^m_{m-1}x^m,$$

and substitute these two formulas into (2). Equating the coefficients of x^{k-1}, we find:

$$kA^k_m = mA^{k-1}_{m-1} \quad \text{for } k \geqslant 2, \tag{3}$$

$$A^1_m = \alpha_m \quad \text{for } k = 1. \tag{4}$$

(Notice that in the above formulas there is no α_0; we only have α_k where $k \geqslant 1$.) We have obtained the formula similar to the formula for the binomial coefficients C^k_m, the difference being that formula (3) holds only for $k \geqslant 2$, and for $k = 1$ it is replaced by (4).

Again, we continue to follow the case of binomial coefficients. We have: $A^k_m = \dfrac{m}{k}\,A^{k-1}_{m-1}$. Applying this formula to A^{k-1}_{m-1} we get: $A^k_m = \dfrac{m(m-1)}{k(k-1)}\,A^{k-2}_{m-2}$. Continuing this procedure, after $k-1$ steps we find

$$A^k_m = \frac{m(m-1)\cdots(m-k+2)}{k(k-1)\cdots 2}\,A^1_{m-k+1} = \frac{m(m-1)\cdots(m-k+2)}{k(k-1)\cdots 2}\,\alpha_{m-k+1}$$

(for A^1_{m-k+1} we use formula (4)).

The coefficient of α_{m-k+1} very much resembles the binomial coefficient. It differs from C^k_m (see formula (21)) in so much that the numerator does not have the last factor $m-k+1$ (and the denominator does not have the last factor 1, but this has no effect on the product). However, in the formula for C^k_{m+1} the product in the numerator ends with $m-k+1$, but it begins with $m+1$, which is not present here. Hence, we can write the coefficient of α_{m-k+1} in the form $\dfrac{1}{m+1}C^k_{m+1}$, and the formula for A^k_m becomes:

$$A^k_m = \frac{1}{m+1}C^k_{m+1}\alpha_{m+1-k}.$$

(We write α_{m-k+1} as α_{m+1-k} so that the factors look more similar.)

In this way we have obtained the following formula for the polynomials $g_m(x)$:

$$g_m(x) = \frac{1}{m+1}(C^1_{m+1}\alpha_m x + C^2_{m+1}\alpha_{m-1}x^2 + \cdots + \\ + C^k_{m+1}\alpha_{m+1-k}x^k + \cdots + C^{m+1}_{m+1}\alpha_0 x^{m+1}). \tag{5}$$

The obtained formula resembles the binomial formula. Suppose that we have a new variable a and expand the binomial $(x+a)^{m+1}$. We then obtain the same terms as in the brackets in the above formula (5), except that a^k is replaced by α_k and it has no term corresponding to $C^0_{m+1}\alpha_{m+1}$. We can compensate for this by considering the difference $(x+a)^{m+1} - a^{m+1}$ in which case the terms with a^{m+1}

cancel. In order to emphasize this analogy, introduce the following notation. Let a be the sequence α_1, α_2, ... and let $f(t)$ be the polynomial $a_0 + a_1 t + \cdots + a_k t^k$. Then $f(a)$ denotes the number $a_0 + a_1\alpha_1 + \cdots + a_k\alpha_k$, i.e. t^k is replaced by α_k. In particular, $a^m = \alpha_m$, since replacing t^m by α_m we obtain α_m. Analogously, $(x+a)^m = x^m + C_m^1 x^{m-1}\alpha_1 + \cdots + C_m^m \alpha^m$: we expand $(x+a)^m$ in powers of t and replace t^k by α_k. The relation (5) with this notation can be written in the form

$$g_m(x) = \frac{1}{m+1}\big((a+x)^{m+1} - a^{m+1}\big). \tag{6}$$

Remark that $a^{m+1} = \alpha_{m+1}$. Notice that we cannot establish that the polynomial given by (6) satisfies the relation (1). In fact, we have found the general form of the polynomials which satisfy the relations (2), but those relations are only *consequences* of the relation (1). Indeed, the result depends upon the sequence α_m, which can in (6) be arbitrary, whereas Theorem 7 states that the polynomial $g_m(x)$ is unique for each m. Therefore, we have not yet solved the problem.

Among the polynomials $g_m(x)$ given by (6) we have to choose those which satisfy the relation (1). Since we already know that such polynomials exist and they are unique (for each m) we only have to find the unique sequence a which defines them. This is quite simple: it is enough to put $x = 1$ into (1). Since $g_m(0) = 0$ (the constant term of g_m is 0), we get $g_m(1) = 1$. The notation of the formula (6) yields $(a+1)^{m+1} - \alpha_{m+1} = m+1$ for $m = 0, 1, 2, \ldots$ or

$$(a+1)^m - \alpha_m = m, \quad m = 1, 2, 3, \ldots$$

Definition. The numbers B_1, B_2, B_3, ... are called *Bernoulli's numbers* if the sequence B formed by them satisfies the relations

$$(B+1)^m - B_m = m \quad \text{for } m = 1, 2, 3, \ldots$$

The above relations uniquely define the sequence of Bernoulli's numbers. Indeed, expanding the above formula, by definition we get

$$1 + mB_1 + C_m^2 B_2 + \cdots + mB_{m-1} = m, \quad m = 1, 2, \ldots \tag{10}$$

(B_m cancels). From this relation for $m = 1$ we get that $B_1 = 1/2$, and every relation that follows allows us to find B_{m-1} provided that we know all B_r's with indices $r < m - 1$.

Polynomials

$$B_m(x) = \frac{1}{m+1}((B+x)^{m+1} - B_{m+1})$$

where B is the sequence of Bernoulli's numbers are called *Bernoulli's polynomials.* We have proved that if the polynomial $g_m(x)$ satisfying (1) is written in the form (6), then the sequence a which corresponds to it has to coincide with the sequence B of Bernoulli's numbers. But we know, according to Theorem 7, that such a polynomial exists. Hence it must coincide with Bernoulli's polynomial $B_m(x)$, i.e.

$$B_m(x) - B_m(x-1) = x^m,$$

and so $S_m(n) = B_m(n)$. Our problem is solved.

Bernoulli's polynomials and numbers were discovered by Jacob Bernoulli (there was a large family of mathematicians of that name). His main results belong to the second half of the 17th century, but this particular discovery appeared in a book published after his death at the beginning of the 18th century. The numbers B_n were named Bernoulli's numbers by Euler (18th century) who found many other applications of those numbers.

Putting $k = 2,3,4,5,6,7,8,9,10,11,12,13$ into (10) we obtain the following values for the numbers B_n (verify this yourself!)

$$B_1 = \frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_3 = 0, \quad B_4 = -\frac{1}{30}, \quad B_5 = 0, \quad B_6 = \frac{1}{42},$$
$$B_7 = 0, \quad B_8 = -\frac{1}{30}, \quad B_9 = 0, \quad B_{10} = \frac{5}{66}, \quad B_{11} = 0, \quad B_{12} = -\frac{691}{2730}$$

etc. Then we easily establish:

$$S_1(n) = \frac{n(n+1)}{2}, \quad S_2(n) = \frac{n(n+1)(2n+1)}{6}, \quad S_3(n) = \frac{n^2(n+1)^2}{4},$$
$$S_4(n) = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}, \quad S_5(n) = \frac{n^2(n+1)^2(2n^2+2n-1)}{12},$$

etc.

Problems

1. Find $B_m(-1)$.

2. Prove the formula $B^m = (B-1)^m$ for $m \geqslant 2$.

3. Derive a relation, analogous to (10), which holds for Bernoulli's numbers B_m with odd indices $m \geqslant 3$. Prove that all Bernoulli's numbers B_m with odd indices, except B_1, are equal to 0.

4. Find $S_6(n)$.

5. Prove the formula for the derivative of a polynomial of a polynomial: if $f(x)$ and $g(x)$ are polynomials, then

$$f(g(x))' = f'(g(x))g'(x).$$

6. Find $(a+x)^n$ if the sequence a has the form $a_n = q^n$ for some number n.

I. R. Shafarevich,
Russian Academy of Sciences,
Moscow, Russia