DETERMINATION OF PROBABILITY, that AN EQUATION OF THE SECOND DEGREE, WITH INTEGER COEFFICIENTS, UNDERTAKEN AT RANDOM, TO HAVE REAL ROOTS;*

V. Bunyakovsky

Mem. St. Petersburg. Vol. 3/1, No. 4 (1836), p. 341–351.
Read on October 9, 1835.)

Let $x^2 + px + q = 0$ be the total equation of second degree. Let us assume that the coefficients p and q are integer numbers, which vary between the limits $-m$ and $+m$, and, for reasons of simplicity of the cases, we assume that neither p nor q, becomes zero. Under such circumstances, which most frequently take place, there appears the curious question: *how great is the probability that the equation, written at random, have real roots?* We propose a solution of this problem.

It is known that the desired probability will be expressed by a fraction of which numerator depicts the number of cases for which the equation $x^2+px+q = 0$ has real roots, assuming that p and q vary between the so-called limits $-m$ and $+m$.

For determining the said number, it is worthwhile to explore only how often a difference $p^2 - 4q$ will be either *zero* or a *positive* value, the numbers varying from $-m$ to $+m$, inclusively. Concerning the denominator of the fraction [342]
which expresses the desired probability, it will be obviously equal to the total number of equations, obtained through changes in the integer coefficients p and q between the limits $-m$ and $+m$, or, equivalently, to the number of all possible arrangements of numbers 1, 2, 3... m, taken pairwise, moreover taking into account their signs.

It will be enough to let $z = \frac{N}{M}$ be the required probability. The denominator M we determine very simply. In the matter itself, since coefficients p and q can be both positive and negative, then, after accepting p and q positive, the

*Paraphrase from the Russian by Richard J. Pulskamp, Department of Mathematics & Computer Science, Xavier University, Cincinnati, OH. January 5, 2022 following four assumptions occur:

$$(1) \qquad \begin{cases} x^2 + px + q = 0 \\ x^2 - px + q = 0 \\ x^2 + px - q = 0 \\ x^2 - px - q = 0; \end{cases}$$

and as over these p and q take values of 1,2, 3, $\ldots m$, then is obvious that each of four equations (1) confines to itself m^2 with one another, and consequently, their totality will be $4m^2$; here is the denominator M of the fraction which expresses probability z; therefore

$$z = \frac{N}{4m^2}.$$

Now let us study the determination of numerator N. It is enough to let n be the number of cases, with which the *first* of the equations (1) allows real roots; n', the same in relation to the *second* of the equations (1); n'' in relation to the *third*, and n''' in relation to the *fourth*. And from the assumptions such will be $N = n + n' + n'' + n'''$.

But it is obvious that $n = n'$, since the difference $p^2 - 4q$ is identical in both equations of $x^2 - px + q = 0$ and $x^2 + px + q = 0$. Concerning for the sake of n'' and n''', then it is clear that these numbers are equal to each other, and each of them is equal to m^2, since both roots of the equation of $x^2 + px - q = 0$ in all cases are real. And so $N = 2n + 2m^2$; consequently [343]

$$(2) \qquad z = \frac{2n + 2m^2}{4m^2} = \frac{n + m^2}{2m^2}.$$

We see from this that the question is led to the determination of the number n, which expresses how often $p^2 - 4q$ is turned to zero or to a positive quantity, assigning to values p and q, at random, every positive value 1, 2, 3,$\ldots m$.

And so, it is necessary to find the number of solutions of the formula

$$p^2 - 4q \geq 0,$$

or equivalently, the following:

$$q \leq \frac{p^2}{4}$$

for the values p and q, which are comprehended between limits of 1 and m, inclusively.

Assuming consecutively $p = 1, 2, 3, 4, 5, 6, \ldots$ we will obtain a known value for q, for which the condition $q \leq \frac{p^2}{4}$ is satisfied, or, equivalently, of all values q, for which the equation of $x^2 + px + q = 0$ has real roots. The greatest integer that is comprehended by $\frac{p^2}{4}$ will determine the number of equations which allow real roots with the values of coefficient p. With this basis assumed we will obtain the table:

$$(3)\qquad \begin{cases} \underline{p} & \underline{q} \\ 1 & 0 \\ 2 & 1 \\ 3 & 2,1 \\ 4 & 4,3,2,1 \\ 5 & 6,5,4,3,2,1 \\ 6 & 9,8,7,6,5,4,3,2,1 \\ \vdots & \vdots \end{cases}$$

Considering the series of maximum values for q, exactly the following: 1, 2, 4, 6, 9... (*zero* is excluded from the values q), we note with the 1st column. For the determination of the number n it is necessary to find the law of the [344] composition of the said series; in the 2nd column. It is necessary to determine with what value p, the maximum value for q will not exceed this limit m. For example, if $m = 5$, then the p about which we speak, will be 4, since the value $p = 5$ corresponds already to the value $q = 6$; by the condition of the question, it follows you will discard this value $q = 6$, but hold only the five following 5, 4, 3, 2, 1 not exceeding the limit m. And so, in the present case, the number $n = 1 + 2 + 4 + 5 = 12$, which you will see easily from the equations.

Let P be the value of coefficient p, which adds a value Q to the previous series for q, which is directly less than or equal to m. There will consequently be $\frac{p^2}{4} \geq m$, whence $P = E(\sqrt{4m})$, understanding by $E(\sqrt{4m})$ the greatest integer being comprehended in $\sqrt{4m}$. For the value p, greater than P, we will obtain values of q generally larger than the limit m; of them it is necessary to hold only those which do not exceed the number m, precisely, 1, 2, 3... m. However, since between $p = P$ and $p = m$, there will be $m - P$ terms, and m values for q which correspond to each of these terms, then we conclude that between the limits $p = P$ and $p = m$, the number of equations with real roots will be expressed by $m(m - P)$.

Now there remains to be determined the sum s of the number

$$1, 2, 4, 6, 9 \cdots Q,$$

since when it will be known, then n will be found by means of the formula

$$n = s + m(m - P).$$

For the determinations of sum s, it is necessary to know the law on which the previous number is comprised. Let us study its definition, and for greater clarity, let us distinguish two cases, according to whether P will be an *even* or an *odd* number.

The first the case, in which it is assumed $P = 2k$. [345]

It is clear from this assumption, that we have $Q = k^2$; the preceding value will arise by taking $p = 2k - 1$ and by computing the quantity $\frac{p^2}{4} = k^2 - k + \frac{1}{4}$;

the greatest whole number which is contained in this gives the desired value for q, which is obviously equal to a difference $k^2 - k$. Then obtain $(k-1)^2$ for the value of q, next $(k-1)^2 - (k-1)$, etc. These other values, I write as far as accepted order, will constitute the number

$$(4) \quad \begin{cases} k^2, k^2 - k, (k-1)^2, (k-1)^2 - (k-1), (k-2)^2, (k-2)^2 - (k-2), \\ \cdots\cdots\cdots\cdots 9, 6, 4, 2, 1 \end{cases}$$

The sum s of this line depicted above will express the number of equations with real roots, from the limit $p = 1$ to $p = E(\sqrt{4m}) = P$, inclusively. But this number can be written the form

$$s = 1 + [2+4] + [6+9] + \cdots + [2(k-1)^2 - (k-1)] + [2k^2 - k],$$

and, through the inverse method of differences, it will be found by very simple means to be

$$s = \frac{k(k+1)(4k-1)}{6} = \frac{P(P+2)(2P-1)}{24}.$$

Consequently, the total number of cases, for which the equations of $x^2 + px + q = 0$ have real roots, expressed by sum $s + m(m-P)$, will be

$$\frac{P(P+2)(2P-1)}{24} + m(m-P).$$

But this number is represented above by n; and so, the probability z, that the equation of the second degree, with integer coefficients, that are comprehended between the limits $-m$ and $+m$, will be expressed, consequent to equation (2), by the formula

$$z = \frac{\frac{P(P+2)(2P-1)}{24} + m(2m-P)}{2m^2},$$

where $P = E(\sqrt{4m})$, and is assumed an *even* number.

[346]

The second case, in which it is assumed $P = 2k+1$.

In this case there will be $Q = k^2 + k$, since $\frac{p^2}{4} = k^2 + k + \frac{1}{4}$; whence, the integer part, is $k^2 + k$; consequently, the previous value for q, which corresponds to the value $p = 2k$, will be k^2; then the value of q will proceed as before in similar order in number (4). And so as to obtain the summation of series

$$1, 2, 4, 6, 9 \cdots 2k^2 - k, k^2 + k,$$

it will be sufficient, to the sum

$$s = 1 + [2+4] + [6+9] + \cdots + (2k^2 - k) = \frac{k(k+1)(4k-1)}{6},$$

to give the value of the additional term

$$k^2 + k$$

that it will lead us to the following expression:

$$\frac{k(k+1)(4k+5)}{6} = \frac{(P-1)(P+1)(2P+3)}{24}$$

when the probability z, in the case of P *odd*, will be expressed by the formula:

$$z = \frac{\frac{(P-1)(P+1)(2P+3)}{24} + m(2m-P)}{2m^2}.$$

And so, for determining the probability z, that of a randomly chosen complete equation of second degree $x^2+px+q=0$, in which p and q are the integers, which are comprehended between the limits $-m$ and $+m$, has real roots, first the greatest integer being contained in $\sqrt{4m}$ should be found. Let P be this number; if P is *even*, then

$$z = \frac{\frac{P(P+2)(2P-1)}{24} + m(2m-P)}{2m^2}; \tag{5}$$

when P is *odd*, then

$$z = \frac{\frac{(P-1)(P+1)(2P+3)}{24} + m(2m-P)}{2m^2}; \tag{6}$$

[347]

Let us put, for example, $m = 10$; let us find $P = E(\sqrt{40}) = 6$, and based on formula (5)

$$z = \frac{162}{200}.$$

For $m = 100$, we have $P = E(\sqrt{400}) = 20$, and let us find based on that formula (5)

$$z = \frac{18715}{20000}.$$

For $m = 1000$, we obtain $P = E(\sqrt{4000}) = 63$, and in consequence of formula (6) we have

$$z = \frac{1958328}{2000000}.$$

and so on. —

We perceive from these examples that the probability to obtain at random an equation of second degree which has real roots, rapidly increases, as we increase the limits of its coefficient. If we will place $m = \infty$, then both formulas (5) and (6) give $z = 1$. This conclusion, at first glance, must seem erroneous, since it follows that the probability to obtain an equation with *imaginary roots* is equal to zero, while there is no doubt whatsoever, that such equations are an infinite set. This apparent paradox is easily explained by the fact that the number of equations, which have real roots, is infinitely great with respect to the number of equations, which allow an imaginary solution. Based on formulas (5) and (6) it is easy to prove that the ratio of the number of equations with real roots to the number of equations with imaginary will be proportional to $\sqrt{m}$, when we assume $m = \infty$. But let us note that if the limits of the two coefficients p and q were different from one another, then it would be possible to select these limits so that probability z applied not only to a fractional quantity, but even to zero. — [348]

Let us finish our study as far as indication of the method for determining the probability to obtain an equation with real roots, when it will be of the form $ax^2 + bx + c = 0$, understanding that a, b and c are whole numbers that vary between the limits of 1 and m, inclusively. For brevity, we are silent about the case, when the coefficients a, b and, c can assume both + and − signs; this circumstance does not represent any special difficulty. —

If we represent the desired probability by the fraction $\frac{N}{M} = z$, then the denominator M will be equal to the number of all possible arrangements of the numbers 1, 2, 3 $\ldots m$, taken in triples, then there is m^3. Consequently

$$z = \frac{N}{m^3}.$$

With regard to the numerator N, then, following the method similar to that which was presented above, and after assuming for the brevity $E(\frac{m^2}{4}) = \lambda$, let us find:

$$\begin{aligned} N = \; & 1 \\ & 2+1 \\ & 4+2+1+1 \\ & 6+3+2+1+1+1 \\ & 9+4+3+2+1+1+1+1+1 \\ & 12+6+4+3+2+2+1+1+1+1+1+1 \\ & 16+8+5+4+3+2+2+2+1+1+1+1+1+1+1+1 \\ & \cdots\cdots \\ & E\left(\frac{\lambda}{1}\right)+E\left(\frac{\lambda}{2}\right)+E\left(\frac{\lambda}{3}\right)+\cdots+E\left(\frac{\lambda}{\lambda-1}\right)+E\left(\frac{\lambda}{\lambda}\right), \end{aligned}$$

observing besides, that each number which exceeds m, must be replaced by this limit m itself, and moreover, in each horizontal row to hold to the left side, only m members, and the rest to discard. [349]

For instance, if $m = 5$, then by acting in the foregoing manner, we will obtain:

$$\begin{aligned} N = & 1 \\ & 2\ +1 \\ & 4\ +2+1+1 \\ & 6^* +3+2+1+1+1_*. \end{aligned}$$

Since number 6^* is more than 5, then instead of this number we should put 5; beyond that it is necessary to discard the sixth term 1_*. Consequently the real value of N will be

$$N = \left.\begin{array}{l} 1 \\ 2+1 \\ 4+2+1+1 \\ 5+3+2+1+1 \end{array}\right\} = 24,$$

and that the required probability is

$$z = \frac{24}{5^3} = \frac{24}{125}.$$

Examining the equation $x^2+px+q=0$, (where p and q are assumed positive, they would find that for $m=5$, the probability to obtain an equation with real roots, is expressed by the fraction of $\frac{12}{25}$, which is $2\frac{1}{2}$ times greater than the probability $\frac{24}{125}$ obtained now.

Assuming $m=10$, we find $\lambda = E(\frac{10^2}{4}) = 25$, and consequently

$$\begin{array}{lllllllllllllllllllllllll}
N = & 1 \\
& 2 & + & 1 \\
& 4 & + & 2 & +1+1 \\
& 6 & + & 3 & +2+1+1+1 \\
& 9 & + & 4 & +3+2+1+1+1+1+1 \\
& 12^* & + & 6 & +4+3+2+2+1+1+1+1+1_*+1_* \\
& 16^* & + & 8 & +5+4+3+2+2+2+1+1++1_*+1_*+1_*+1_*+1_*+1_* \\
& 20^* & + & 10^* & +6+5+4+3+2+2+2+2++1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_* \\
& 25^* & + & 12^* & +8+6+5+4+3+3+2+2++2_*+2_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*+1_*.
\end{array}$$

[350]

Numbers 12^*, 16^*, 20^*, 25^*, and 12^* must be replaced, each, by the number 10, and terms in the eleventh position, marked with an asterisk below, entirely discard. After doing this, we will obtain

$$N = 217$$

and consequently the probability

$$z = \frac{217}{1000}$$

It goes without saying that in the determination of the number N it is necessary, for abridging the action, to write directly, instead of the numbers exceeding the limit, the limit itself, and to be limited in each horizontal line to the number m of terms: and so, in the previous example, one should write directly the value N in the form

$$N = \left.\begin{array}{l} 1 \\ 2 + \ 1 \\ 4 + \ 2+1+1 \\ 6 + \ 3+2+1+1+1 \\ 9 + \ 4+3+2+1+1+1+1+1 \\ 10 + \ 6+4+3+2+2+1+1+1+1 \\ 10 + \ 8+5+4+3+2+2+2+1+1 \\ 10+10+6+5+4+3+2+2+2+2 \\ 10+10+8+6+5+4+3+3+2+2 \end{array}\right\} = 217.$$

Now the probability $\frac{217}{1000}$ obtained, and relating to the equation $ax^2+bx+c = 0$, when a, b and c are comprehended between the limits of 1 and m, there will be less probability than that corresponding to the equation $x^2 + px + q = 0$, where p and q are comprehended between the same limits; actually, we find for this latter the fraction $\frac{62}{100}$. And in general, it is possible to conclude that of the two equations of the form $ax^2 + bx + c = 0$ and $x^2 + px + q = 0$, taken at random, but in which the coefficients a, b, c, p, q are comprehended between identical limits, the probability of real roots is greater on the side of the second, [351] i.e., the equation $x^2 + px + q = 0$.

The method relating to the equation of $ax^2 + bx + c = 0$ can be very easily applied also to the determination of the probability that the three roots of an equation of third degree $ax^3 + bx + c = 0$ have real values.