THE USE OF HEURISTICAL ESTIMATE IN NUMBER THEORY

ALEKSANDER ZUJEV

Abstract. This is an overview of some common uses of heuristics in mathematics, and particularly in number theory. We use heuristical estimates for a few solved problems - such as Pythagorean triples and Fermat Last Theorem, and a few still open, such as Mersenne primes, Fermat primes, Euler's perfect cuboid. In conclusion, we discuss the place of heuristics in mathematics.

1. Introduction

The use of heuristical estimates in mathematics is not new. One of the most known examples is estimate of the density of primes. Long before the Prime Number Theorem was proven, Dirichlet concluded that the density of primes was $\sim \frac{1}{\ln n}$. This gives the estimate of prime-counting function

$$\pi(x) \sim \int_2^x \frac{dt}{\ln t} = \mathrm{Li}(x)$$

which is a better estimate than given by the Prime Number Theorem

$$\pi(x) \sim \frac{x}{\ln x}$$

Caldwell [1] studies in detail the use of heuristics in various problems related to the prime numbers.

We do not go into such detail with relation to the primes. Instead, we consider the use of heuristics not only in the problems related to the prime numbers, but also to other problems in number theory.

The purpose of this paper is not to develop new techniques of using heuristics, but rather overview its known uses, and the place of heuristic estimates in mathematics. In the examples used, we don't try to obtain new more precise estimates; but rather make the examples as simple as possible, for better illustration.

In the following sections, we will give examples of applications of heuristic method for some known results, and for some still unknown.

In conclusion, we argue that heuristic estimates, and conjectures based on them are legitimate tools of mathematics, giving many useful results.

2. Use of heuristic estimate for some known statistics

In this paper, we use the following two heuristic estimates:

2010 *Mathematics Subject Classification.* Primary: 11A41, 11N32, 11D45.

Key words and phrases. primes, polynomials, Mersenne, Fermat.

1. The probability that an arbitrary positive integer n is a prime number is
$$\sim \frac{1}{\ln n}$$
2. The probability that an arbitrary positive integer n is a perfect power of m is
$$\sim \frac{1}{mn^{\frac{m-1}{m}}}$$

In this section, we consider some known results, and compare them to heuristic estimates. It is a "postdiction", useful, even necessary to estimate how good is a used technique, before trying to apply it for new uses. Like in physics, where a new theory first must satisfy the old results, before predicting new results.

2.1. **The number of Pythagorean triangles of hypotenuse less than a given magnitude.**

The problem of Pythagorean triangles is completely solved by Euclid. We know a generating formula for the sides of all Pythagorean triangles, and that there is an infinite number of such triangles.

Suppose we don't know the general formula for the sides of Pythagorean triangles, and instead we will count them heuristically. Let a, b be positive integers such that $a^2 + b^2 < R^2$. Suppose $a^2 + b^2 = c^2$ for some integer c. An integer n is a perfect square with probability $\sim \frac{1}{2\sqrt{n}}$. So a, b are catheti of a Pythagorean triangle with probability $\sim \frac{1}{2\sqrt{a^2+b^2}}$. Then the number of Pythagorean triples with $c < R$ is

$$\begin{aligned} &\sim \sum_{a^2+b^2<R^2} \frac{1}{2\sqrt{a^2+b^2}} \\ &\sim \int_0^R \frac{\pi r}{2}\frac{1}{2r}\,dr \\ &= \frac{\pi R}{4} \end{aligned}$$

And we should divide by 2, not to count triples with a, b interchanged:

$$\frac{\pi R}{8}$$

So

$$\frac{\text{\# triples}}{\text{R}} \sim \frac{\pi}{8}$$

Now let us recall the formula for Pythagorean triples

$$(a, b, c) = k(m^2 - n^2, 2mn, m^2 + n^2),$$

$m > n$, $\gcd(m, n) = 1$, and not both m, n odd. We will count them using formula from Mathworld [2] for the number of possible primitive or nonprimitive right triangles having s as a hypotenuse:

$$H(s) = \frac{1}{8}[r_2(s^2) - 4]$$

(The number of Pythagorean triples with $c < R$) / R:

R	(number of triples)/R
10^2	0.52
10^3	0.881
10^4	1.2471
10^5	1.61436
10^6	1.98064

The actual (number of triples)/R does not match our heuristic estimate $\frac{\pi}{8}$; seems it goes to infinity with n.

Our mistake was that we didn't take into account that if (a, b, c) is a Pythagorean triple, then so is (ka, kb, kc), where k is a positive integer. Taking it into account, we come with a formula (omitting derivation)

$$(\#\text{ triples}) \sim C(R \ln R - R)$$

We then have the following estimate for C:

R	$(R \ln R - R)$/(actual # triples)
10^2	6.93302
10^3	6.70574
10^4	6.58355
10^5	6.51213
10^6	6.47038

The ratio seems to be converging to a constant. We then have a formula

$$(\#\text{ triples}) \sim C(R \ln R - R)$$

where we will leave the constant C without derivation. So we found a heuristical formula, which gives a good approximation for the number of Pythagorean triples with $c < R$. Also from this formula follows a conjecture that the number of Pythagorean triples is infinite.

2.2. **Fermat's Last Theorem.** There are no positive integers a, b, c, and integers $n >= 3$, such that $a^n + b^n = c^n$.

Let us give a heuristic estimate of the number of quadruplets (a, b, c, n) such that $a^n + b^n = c^n$. The probability that k is a perfect nth power is $\sim \frac{1}{nk^{\frac{n-1}{n}}}$. Then the number of triples (a, b, n) such that $a^n + b^n$ is a perfect nth power, is

$$\begin{aligned} \#(a, b, n) &\sim \sum_{a\geqslant 2, b\geqslant 2, n\geqslant 4} \frac{1}{n(a^n + b^n)^{\frac{n-1}{n}}} \\ &\sim \iiint\limits_{a\geqslant 2, b\geqslant 2, n\geqslant 4} \frac{da\, db\, dn}{n(a^n + b^n)^{\frac{n-1}{n}}} \\ &= 0.05022 \end{aligned}$$

We used $a, b \geqslant 2$, instead of $a, b \geqslant 1$, because by Catalan's conjecture / Mihailescu theorem, it cannot be that $a^n + 1 = c^n$ (or it can be shown by elementary means). We used $n \geqslant 4$, because for $n = 3$, the theorem was proven by Euler. We obtained that the expected number of such triples is ~ 0.05, which is consistent with the fact that there are none.

3. Use of heuristic estimate for some unknown statistics

In this section, we consider some open problems, and application of heuristic estimate to them.

3.1. **Primality of polynomials.**

The most famous example is $a^2 + 1$. It is still unknown if it is prime at infinite number of a's. Hardy and Littlewood state so in their Conjecture E [3].

If the probability that $a^2 + 1$ is a prime is $\sim \frac{1}{\ln(a^2+1)}$, then the number of $a \leqslant x$ such that $a^2 + 1$ is a prime, is

$$\sim \int_1^x \frac{da}{\ln(a^2 + 1)}$$

Comparison of used heuristics with actual numbers:

x	number of primes $a^2 + 1$, $a \leqslant x$	$\int_1^x \frac{da}{\ln(a^2+1)}$	relative error	Conjecture E
10^2	19	15.37	-0.191	15
10^3	112	89.11	-0.2044	99
10^4	841	623.38	-0.2588	745
10^5	6656	4815.21	-0.2766	5962
10^6	54110	39314.1	-0.2734	49680
10^7	456362	332459.	-0.2715	425826

This is a reasonably good estimate. Not as good as Hardy and Littlewood estimate; but even our rough estimate gives a good qualitative agreement. The conjecture is that there is an infinite number of primes a^2+1, but it is so far not proven.

Similarly we study a polynomial $a^3 + 2$. Our heuristic estimate is that the number of primes $a^3 + 2$, $a \leqslant x$ is $\sim \int_1^x \frac{da}{\ln(a^3+2)}$.

x	number of primes $a^3 + 2$, $a \leqslant x$	$\int_1^x \frac{da}{\ln(a^3+2)}$	relative error
10^2	10	10.29	0.029
10^3	74	59.45	-0.1966
10^4	520	415.63	-0.2007
10^5	4059	3210.18	-0.2091
10^6	33795	26209.4	-0.2245

Again, these are reasonably close results.

It might be reasonable to conjecture, that any non-reducible polynomial with integer coefficients $p(a)$ has probability of its value being a prime number $\sim \frac{1}{\ln p(a)}$.

3.2. **Mersenne primes.** A Mersenne number is $2^p - 1$, where p is a prime.

The probability that a number $2^k - 1$ is a prime is $\sim \frac{1}{\ln(2^k-1)} \approx \frac{1}{k \ln 2}$. The number of Mersenne primes $\leqslant n$ is

$$\sim \sum_{2^k-1 \leqslant n} \frac{1}{\ln(2^k - 1)} \approx \sum_{k=1}^{\log_2 n} \frac{1}{k \ln 2} \approx \frac{\ln \ln n}{\ln 2}$$

For $n = 1\,000\,000$, there are 33 Mersenne primes, while the sum is 19.10. We can call it a reasonably good estimate. The sum is infinite, so the conjecture is that there are an infinite number of Mersenne primes.

3.3. **Fermat primes.** A Fermat number is $F_k = 2^{2^k} + 1$.

The probability that a number F_k is a prime is $\sim \frac{1}{\ln(2^{2^k}+1)} \approx \frac{1}{2^k \ln 2}$. The number of Fermat primes $\leqslant n$ is

$$\sim \sum_{2^{2^k}+1 \leqslant n} \frac{1}{\ln(2^{2^k} + 1)} \approx \sum_{k=1}^{n} \frac{1}{2^k \ln 2}$$

In difference to Mersenne primes, the sum is finite; therefore the conjecture is that there are only a finite number of Fermat primes. $F_0, ..., F_4$ are prime; it is now known that F_k are composite for $5 \leqslant k \leqslant 32$. The expected number of Fermat primes after F_{32} is

$$\sim \sum_{k=33}^{\infty} \frac{1}{2^k \ln 2} = \frac{1}{2^{32} \ln 2},$$

a very small number. The conjecture is therefore that there are no more Fermat primes.

3.4. **Twin primes.** It is still unknown if there are an infinite number of twin primes. Yitang Zhang [4] came close to it, proving that exists some N less than 70 million, such that there are infinitely many pairs of primes that differ by N. Let us estimate heuristically the number of twin primes less than a given magniude. The probability that k is a prime is $\sim \frac{1}{\ln k}$, and the probability that $k+2$ is a prime is $\sim \frac{1}{\ln(k+2)}$. Then the probability that both k and $k+2$ are primes is $\sim \frac{1}{\ln k} \frac{1}{\ln(k+2)} \approx \frac{1}{\ln^2 k}$. (we are being very rough here; if k is a prime, than the probability that $k+2$ is a prime is $>\sim \frac{1}{\ln(k+2)}$). Then the expected number of primes less than n is

$$\sim \sum_{k=2}^{n} \frac{1}{\ln^2 k} \sim \int_2^n dk \ln^2 k = \left[\mathrm{Li}(k) - \frac{k}{\ln k} \right]_{k=2}^{n} \approx \mathrm{Li}(n) - \frac{n}{\ln n}$$

The Hardy-Littlewood estimate for the number of twin primes

$$\pi_2(x) \sim 2C_2 \int_2^x \frac{dt}{\ln^2 t}$$

where

$$C_2 = \prod_{p \geqslant 3} \left[1 - \frac{1}{(p-1)^2} \right] \approx 0.6601618158468695$$

A few first values:

n	$\pi_2(n)$	$\int_2^n \frac{dk}{\ln^2 k}$	Relative error	Hardy-Litlewood
10^3	35	34.69	-0.009	46.21
10^4	205	162.24	-0.209	270.67
10^5	1224	945.76	-0.227	1616.08
10^6	8169	6246.98	-0.235	10785.7
10^7	58980	44499.6	-0.246	58753.8
10^8	440312	333530	-0.243	440368
10^9	3424506	2594294	-0.242	3425308
10^{16}	10304195697298	7804293059026	-0.243	10304192554496

The Hardy-Littlewood estimate is asymptotically very good. Our simple formula is not anywhere as good, but a relative error under 0.25 is good for a qualitative estimate.

3.5. **Andrica's conjecture.** [5] It says about the gaps between primes:

$$\sqrt{p_{n+1}} - \sqrt{p_n} < 1 \text{ for all } n$$

For the inequality not to hold, the gap between p_n and p_{n+1} must be

$$\Delta = p_{n+1} - p_n \geqslant \sqrt{p_{n+1}} + \sqrt{p_n}$$

so that Δ consecutive numbers are not prime; the probability of this is

$$\sim \left(1 - \frac{1}{\ln p_n}\right)^{\Delta}$$

The probability that the inequality holds is

$$\sim 1 - \left(1 - \frac{1}{\ln p_n}\right)^{\Delta}$$

The probability that the conjecture holds is

$$\sim \prod_{n=1}^{\infty} \left[1 - \left(1 - \frac{1}{\ln p_n}\right)^{\lfloor \sqrt{p_n} + \sqrt{p_{n+1}} \rfloor}\right] \approx 0.57969$$

Considering that the conjecture was verified for n up to $1.3 \cdot 10^{16}$, we need to calculate the product starting with this large n, which gives the value very close to 1. So by heuristic estimate, Andrica's conjecture with high probability must be true.

3.6. **Euler's perfect cuboid.** [7]

It is a cuboid with integer sides, and integer all three face diagonals, and space diagonal. If the sides are a, b, c, then $a^2 + b^2$, $a^2 + c^2$, $b^2 + c^2$, and $a^2 + b^2 + c^2$ are perfect squares. So far, there are no known perfect cuboids, and it is not proven that they do not exist. Let us give heuristic estimate of the total number of perfect cuboids. The probabilities that $a^2 + b^2$, $a^2 + c^2$, $b^2 + c^2$, $a^2 + b^2 + c^2$ are perfect squares, are respectively

$$\sim \frac{1}{2\sqrt{a^2+b^2}}, \frac{1}{2\sqrt{a^2+c^2}}, \frac{1}{2\sqrt{b^2+c^2}}, \frac{1}{2\sqrt{a^2+b^2+c^2}}$$

The total number of perfect cuboids

$$\begin{aligned}\#(\text{perfect cuboids}) &\sim \sum_{a,b,c=1}^{\infty} \frac{1}{2\sqrt{a^2+b^2}} \frac{1}{2\sqrt{a^2+c^2}} \frac{1}{2\sqrt{b^2+c^2}} \frac{1}{2\sqrt{a^2+b^2+c^2}} \\ &\sim \iiint\limits_{a,b,c\geqslant 1} \frac{1}{2\sqrt{a^2+b^2}} \frac{1}{2\sqrt{a^2+c^2}} \frac{1}{2\sqrt{b^2+c^2}} \frac{1}{2\sqrt{a^2+b^2+c^2}}\, da\, db\, dc \\ &= 0.139923\end{aligned}$$

If we exclude duplicates, obtained by interchanging a, b, c, we must divide this number by 6

$$\#(\text{perfect cuboids}) \sim 0.0233205$$

Presumably, the problem was tested for small a, b, c, let us say for $a, b, c \leqslant 100$. Then if we count for $a > 100, b, c \geqslant 1$, then

$$\#(\text{perfect cuboids}) \sim 0.00195175$$

The expected number of perfect cuboids is quite small. So our conjecture is that there are no perfect cuboids.

4. Conclusion

In 1900 Hilbert posed a set of problems, among them being tenth problem: Find a general algorithm which, for any given Diophantine equation, can decide whether the equation has a solution with all unknowns taking integer values.

The Hilbert's tenth problem is now solved, negatively: there is no such common algorithm.

Many open problems in mathematics deal with what are Diophantine sets. The set of Mersenne primes, and the set of Fermat primes are Diophantine sets. So far it is unknowh if these sets are finite, or infinite. It is conceivable, that there is no way to prove them to be finite, or infinite. All we may ever have are conjectures: that the set of Mersenne primes is infinite, and the set of Fermat primes is finite.

Or consider $\zeta(n)$ for n odd, $n \geqslant 3$. Apéry proved that $\zeta(3)$ is irrational [6]. Do we expect that for every $n = 5, 7, 9, ...$ there will be discovered some ingenious proof of irrationality of $\zeta(n)$? Maybe for some n such proofs don't exist, or will never be discovered. Then we will have to be satisfied with conjecture: $\zeta(n)$ is irrational for n odd. And similar question is about transcendentality. We don't even know if $\zeta(3)$ is transcendental, let alone all the rest of $\zeta(n)$ for n odd, $n \geqslant 3$.

Out of infinite number of current, and future problems in mathematics, some may prove to be unsolvable; certainly some will remain unsolved for a long time. But they still may have workable conjectures, based on heuristics.

In conclusion, heuristics is, not just a temporary hack, used until the proof is found. Heuristics is a legal and important tool in mathematics.

5. Acknowledgements

The author is grateful to Professor Richard Scalettar for reviewing the paper, and valuable critique and suggestions, resulting in its considerable improvement.

References

[1] Chris K. Caldwell, *An amazing prime heuristic*, arXiv:2103.04483 [math.HO], https://arxiv.org/abs/2103.04483

[2] Weisstein, Eric W. *"Pythagorean Triple"*. MathWorld, https://mathworld.wolfram.com/PythagoreanTriple.html

[3] Hardy, Littlewood, *Some problems of partitio numerorum : III: On the expression of a number as a sum of primes*, Acta Math. 44 (1923), 1-70

[4] Yitang Zhang, Yitang, *"Bounded gaps between primes"*, Annals of Mathematics. 179 (3) (2014), 1121-1174.

[5] D. Andrica (1986) *Note on a Conjecture in Prime Number Theory*, Studia Universitatis Babes-Bolyai Mathematica, 31, 44-48.

[6] Roger Apéry, *Irrationalit de* $\zeta 2$ *et* $\zeta 3$, Journes Arithmtiques de Luminy, Astrisque, no. 61 (1979), 3 p.

[7] Weisstein, Eric W. *"Euler Brick"*. MathWorld, https://mathworld.wolfram.com/EulerBrick.html

Department of Physics, University of California Davis
Email address: azujev@ucdavis.edu