Journal of Humanistic Mathematics

Volume 3 | Issue 1 January 2013

Sloane's Gap: Do Mathematical and Social Factors Explain the Distribution of Numbers in the OEIS?

Nicolas J.-P. Gauvrit
LDAR, University Paris VII

Jean-Paul Delahaye

Hector Zenil

Sloane's Gap: Do Mathematical and Social Factors Explain the Distribution of Numbers in the OEIS?

Nicolas Gauvrit

LADR, EA 1547, Centre Chevaleret, Université de Paris VII
adems@free.fr

Jean-Paul Delahaye

LIFL, Laboratoire d'Informatique Fondamentale de Lille
UMR CNRS 8022, Université de Lille I,
delahaye@lifl.fr

Hector Zenil

Department of Computer Science, University of Sheffield
hectorz@labores.eu

Abstract

The Online Encyclopedia of Integer Sequences (OEIS) is a catalog of integer sequences. We are particularly interested in the number of occurrences of $N(n)$ of an integer n in the database. This number $N(n)$ marks the importance of n and it varies noticeably from one number to another, and from one number to the next in a series. "Importance" can be mathematically objective (2^{10} is an example of an "important" number in this sense) or as the result of a shared mathematical culture (10^9 is more important than 9^{10} because we use a decimal notation). The concept of algorithmic complexity [6, 2, 7] (also known as Kolmogorov or Kolmogorov-Chaitin complexity) will be used to explain the curve shape as an "objective" measure. However, the observed curve does not conform to the curve predicted by an analysis based on algorithmic complexity because of a clear gap separating the distribution into two clouds of points. We shall call this phenomenon "Sloane's gap".

1. Introduction

Sloane's on-line encyclopedia of integer sequences [10][1] (OEIS) is a remarkable database of sequences of integer numbers, carried out methodically and with determination over forty years [3]. As of August 29, 2012, the OEIS contained 215 260 integer sequences. Its compilation has involved hundreds of mathematicians, which confers it an air of homogeneity and apparently some general mathematical objectivity–something we will discuss later on.

When plotting $N(n)$ (the number of occurrences of an integer in the OEIS) two main features are evident:

a. Statistical regression shows that the points $N(n)$ cluster around $k/n^{1.33}$, where $k = 2.53 \times 10^8$.

b. Visual inspection of the graph shows that actually there are two distinct sub-clusters (the upper one and the lower one) and there is a visible gap between them. We introduce and explain the phenomenon of "Sloane's gap". This clear zone in the value of $N(n)$ was first noticed by Philippe Guglielmetti[2].

The paper and rationale of our explanation proceeds as follows: We explain that (a) can be understood using algorithmic information theory. If U is a universal Turing machine, and we denote $m(x)$ the probability that U produces a string x, then $m(x) = k2^{-K(x)+O(1)}$, for some constant k, where $K(x)$ is the length of the shortest description of x via U. Function m is usually referred to as the Levin's universal distribution or the Solomonoff-Levin measure [7]. For a number n, viewed as a binary string via its binary representation, $K(n) \leq \log_2 n + 2\log_2\log_2 n + O(1)$ and, for most n, $K(n) \geq \log_2 n$. Therefore for most n, $m(n)$ lies between $k/(n(\log_2 n)^2)$ and k/n. Thus, if we view OEIS in some sense as a universal Turing machine, algorithmic probability explains (a).

Fact (b), however, is not predicted by algorithmic complexity and is not produced when a database is populated with automatically generated sequences. This gap is unexpected and requires an explanation. We speculate

[1] The encyclopedia is available at: http://oeis.org/, last consulted 29 August, 2012.

[2] On his site http://drgoulu.com/2009/04/18/nombres-mineralises/ last consulted 29 August, 2012.

that OEIS is biased towards social preferences of mathematicians and their strong interest in certain sequences of integers (even numbers, primes, and so on). We quantified such a bias and provided statistical facts about it.

2. Presentation of the database

The encyclopedia is represented as a catalogue of sequences of whole numbers and not as a list of numbers. However, the underlying vision of the work as well as its arrangement make it effectively a dictionary of numbers, with the capacity to determine the particular properties of a given integer as well as how many known properties a given integer possesses.

A common use of OEIS is in determining the logic of a sequence of integers. If, for example, you submit to it the sequence 3, 4, 6, 8, 12, 14, 18, 20..., you will instantly find that it has to do with the sequence of prime augmented numbers, as follows: 2+1, 3+1, 5+1, 7+1, 11+1, 13+1, 17+1, 19+1...

Even more interesting, perhaps, is the program's capacity to query the database about an isolated number. Let us take as an example the Hardy-Ramanujan number, 1729 (the smallest integer being the sum of two cubes of two different shapes). The program indicates that it knows of more than 350 sequences to which 1729 belongs. Each one identifies a property of 1729 that it is possible to examine. The responses are classified in order of importance, an order based on the citations of sequences in mathematical commentaries and the encyclopedia's own cross-references. Its foremost property is that it is the third Carmichael number (number n not prime for which $\forall a \in \mathbb{N}^*$, $n|a^n - a$)). Next in importance is that 1729 is the sixth pseudo prime in base 2 (number n not prime such that $n|2^{n-1} - 1$). Its third property is that it belongs among the terms of a simple generative series. The property expounded by Ramanujan from his hospital bed appears as the fourth principle. In reviewing the responses from the encyclopedia, one finds further that:

- 1729 is the thirteenth number of the form $n^3 + 1$;
- 1729 is the fourth "factorial sextuple", that is to say, a product of successive terms of the form $6n + 1$: $1729 = 1 \times 7 \times 13 \times 19$;
- 1729 is the ninth number of the form $n^3 + (n + 1)^3$;

- 1729 is the sum of the factors of a perfect square (33^2);
- 1729 is a number whose digits, when added together yield its largest factor ($1 + 7 + 2 + 9 = 19$ and $1729 = 7 \times 13 \times 19$);
- 1729 is the product of 19 a prime number, multiplied by 91, its inverse;
- 1729 is the total number of ways to express 33 as the sum of 6 integers.

OEIS comprises more than 200 000 sequences. A partial version retaining only the most important sequences of the database was published by Neil Sloane and Simon Plouffe [11] in 1995. It records a selection of 5 487 sequences [11] and echoes an earlier publication by Sloane [9].

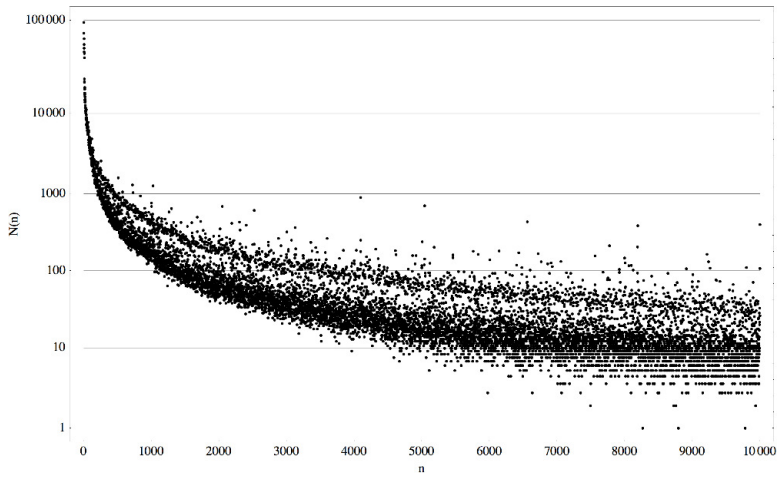

Figure 1: Number of occurrences of $N(n)$ as a function of n per n ranging from 1 to 10 000. Logarithmic scale in ordinate.

Approximately forty mathematicians constitute the "editorial committee" of the database, but any user may propose sequences. If approved, they are added to the database according to criteria of mathematical interest. Neil Sloane's flexibility is apparent in the ease with which he adds new sequences as they are proposed. A degree of filtering is inevitable to maintain the quality of the database. Further, there exist a large number of infinite families of sequences (all the sequences of the form (kn), all the sequences of the form (k^n), etc.), of which it is understood that only the first numbers are recorded in the encyclopedia. A program is also used in the event of a failure

of a direct query which allows sequences of families that are not explicitly recorded in the encyclopedia to be recognized.

Each sequence recorded in the database appears in the form of its first terms. The size of first terms associated with each sequence is limited to approximately 180 digits. As a result, even if the sequence is easy to calculate, only its first terms will be expressed. Next to the first terms and extending from the beginning of the sequence, the encyclopedia proposes all sorts of other data about the sequence, e.g., the definitions of it and bibliographical references.

OEIS is available in the form of a data file that is easy to read, and that contains only the terms retained for each sequence. One can download the data file free of charge and use it–with mathematical software, for example– to study the expressed numbers and conduct statistical research about the givens it contains.

One can, for example, ask the question: "Which numbers do not appear in OEIS?" At the time of an initial calculation conducted in August 2008 by Philippe Guglielmetti, the smallest absent number tracked down was 8795, followed in order by 9935, 11147, 11446, 11612, 11630,... When the same calculation was made again in August 2012, the encyclopedia having been augmented by the addition of several hundreds of new sequences, the series of absent numbers was found to comprise 13794, 14228, 14275, 14435, ...

The instability over time of the sequence of missing numbers in the OEIS suggests the need for a study of the distribution of numbers rather than of their mere presence or absence. Let us consider the number of properties of an integer, $N(n)$, while measuring it by the number of times n appears in the number file of OEIS. The sequence $N(n)$ is certainly unstable over time, but it varies slowly, and certain ideas that one can derive from the values of $N(n)$ are nevertheless quite stable. The values of $N(n)$ are represented in Figure 1. In this logarithmic scale graph a cloud formation with regular decline curve is shown.

Let us give a few examples: the value of $N(1729)$ is 470 (August 2012), which is fairly high for a number of this order of magnitude. For its predecessor, one nevertheless calculates $N(1728) = 854$, which is better still. The number 1728 would thus have been easier for Ramanujan! Conversely, $N(1730) = 148$ and thus 1730 would have required a more elaborate answer than 1729.

The sequence $(N(n))_{n\in\mathbb{N}^*}$ is generally characterized by a decreasing curve.

However, certain numbers n contradict this rule and possess more properties than their predecessors: $N(n) > N(n-1)$.

We can designate such numbers as *interesting*. The first interesting number according to this definition is 15, because $N(15) = 34\,183$ and $N(14) = 32\,487$. Appearing next in order are 16, 23, 24, 27, 28, 30, 35, 36, 39, 40, 42, 45, 47, 48, 52, 53, etc.

We insist on the fact that, although unquestionably dependent on certain individual decisions made by those who participate in building the sequence database, the database is not in itself arbitrary. The number of contributors is very large, and the idea that the database represents an objective view (or at least an intersubjective view) of the numeric world could be defended on the grounds that it comprises the independent view of each person who contributes to it and reflects a stable mathematical (or cultural) reality.

Indirect support for the idea that the encyclopedia is not arbitrary, based as it is on the cumulative work of the mathematical community, is the general cloud-shaped formation of points determined by $N(n)$, which aggregates along a regular curve (see below).

Philippe Guglielmetti has observed that this cloud possesses a remarkable characteristic[3]: it is divided into two parts separated by a clear zone, as if the numbers sorted themselves into two categories, the more interesting above the clear zone, and the less interesting below the clear zone. We have given the name "Sloane's Gap" to the clear zone that divides in two the cloud representing the graph of the function $n \longmapsto N(n)$. Our goal in this paper is to describe the form of the cloud, and then to formulate an explanatory hypothesis for it.

3. Description of the cloud

Having briefly described the general form of the cloud, we shall direct ourselves more particularly to the gap, and we will investigate what characterizes the points that are situated above it[4].

[3]Personal communication with one of the authors, 16th of February, 2009.

[4]Computations from this section henceforth were made using 2009 data.

3.1. General shape

The number of occurrences N is close to a grossly decreasing convex function of n, as one can see from Figure 1.

A logarithmic regression provides a more precise idea of the form of the cloud for n varying from 1 to 10 000. In this interval, the coefficient of determination of the logarithmic regression of $\ln(N(n))$ in n is of $r^2 = .81$, and the equation of regression gives the estimation:

$$\ln(N(n)) \simeq -1.33 \ln(n) + 14.76$$

or

$$\hat{N}(n) = \frac{k}{n^{1.33}},$$

where k is a constant having the approximate value 2.57×10^8, and $\hat{N}$ is the estimated value for N.

Thus the form of the function N is determined by the equation above. Is the existence of Sloane's gap natural then, or does it demand a specific explanation? We note that to our knowledge, only one publication mentions the existence of this split [4].

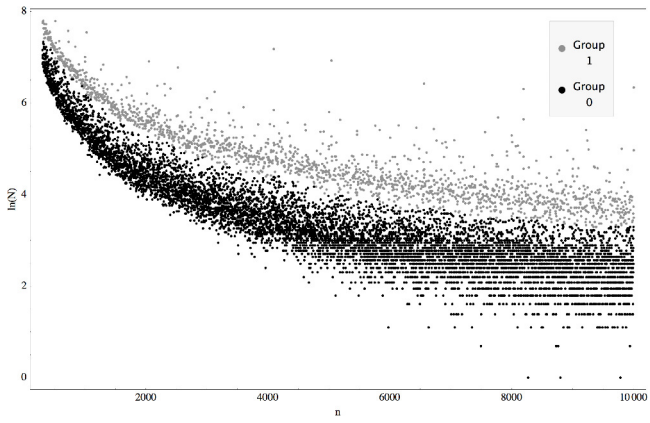

Figure 2: The cloud represents the logarithmic regression of $\ln(N)$ as a function of n for n varying from 1 to 10 000. The grey scale points are those classified as being "above" the gap, while the others are classified as being "below" it.

3.2. Defining the gap

In order to study the gap, the first step is to determine a criterion for classification of the points. Given that the "gap" is not clearly visible for the first values of n, we exclude from our study numbers less than 300.

One empirical method of determining the boundary of the gap is the following: for the values ranging from 301 to 499, we use a straight line adjusted "by sight", starting from the representation of $\ln(N)$ in functions of n. For subsequent values, we take as limit value of n the 82nd percentile of the interval $[n-c, n+c]$. c is fixed at 100 up to $n = 1000$, then to 350. It is clearly a matter of a purely empirical choice that does not require the force of a demonstration. The result corresponds roughly to what we perceive as the gap, with the understanding that a zone of uncertainty will always exist, since the gap is not entirely devoid of points. Figure 2 shows the resulting image.

3.3. Characteristics of numbers "above"

We will henceforth designate as A the set of abscissae of points classified "above" the gap by the method that we have used. Of the numbers between 301 and 10 000, 18.2% are found in A– 1767 values.

In this section, we are looking for the properties of these numbers. Philippe Guglielmetti has already remarked that the prime numbers and the powers of two seem to situate themselves more frequently above the gap. The idea is that certain classes of numbers that are particularly simple or of particular interest to the mathematician are over-represented.

3.3.1. Squares

83 square numbers are found between 301 and 10 000. Among these, 79 are located above the gap, and 4 below the gap, namely, numbers 361, 484, 529, and 676. Although they may not be elements of A, these numbers are close to the boundary. One can verify that they collectively realize the local maximums for $\ln(N)$ in the set of numbers classified under the cloud. One has, for example, $N(361) = 1376$, which is the local maximum of $\{N(n), n \in [325, 10\,000] \setminus A\}$. For each of these four numbers, Table 1 gives the number of occurrences N in Sloane's list, as well as the value limit that they would have to attain to belong to A.

An overwhelming 95.2% of squares are found in A, as opposed to 17.6% of non-squares. The probability that a square number will be in A is thus 5.4 times greater than that for the other numbers.

n	$N(n)$	value limit
361	1376	1481
484	976	1225
529	962	1065
676	706	855

Table 1–List of the square numbers n found between 301 and 10 000 not belonging to A, together with their frequency of occurrence and the value of $N(n)$ needed for n to be classified in A.

3.3.2. Prime numbers

The interval under consideration contains 1167 prime numbers. Among them, 3 are not in A: the numbers 947, 8963, and 9623. These three numbers are very close to the boundary. 947 appears 583 times, while the limit of A is 584. Numbers 8963 and 6923 appear 27 times each, and the common limit is 28.

A majority of 99.7% of prime numbers belong to A, and 92.9% of non-prime numbers belong to the complement of A. The probability that a prime number will belong to A is thus 14 times greater than the same probability for a non-prime number.

3.3.3. A multitude of factors

Another class of numbers that is seemingly over-represented in set A is the set of integers that have "a multitude of factors". This is based on the observation that the probability of belonging to A increases with the number of prime factors (counted with their multiples), as can be seen in Figure 3. To refine this idea we have selected the numbers n of which the number of prime factors (with their multiplicty) exceeds the 95th percentile, corresponding to the interval $[n - 100, n + 100]$.

811 numbers meet this criterion. Of these, 39% are found in A, as opposed to 16.3% for the other numbers. The probability that a number that has a multitude of prime factors will belong to A is thus 2.4 times greater than the same probability for a number that has a smaller number of factors.

Table 2 shows the composition of A as a function of the classes that we have considered.

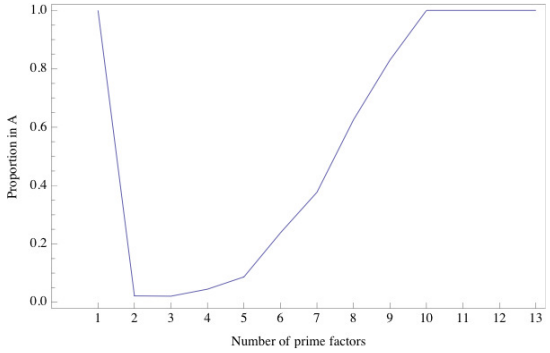

Figure 3: For each number of prime factors (counted with their multiples) one presents, the proportion of integers belonging to A is given. For the interval determined above, all numbers with at least 10 factors are in A.

class	number in A	% of A	% (cumulated)
primes	1164	65.9	65.9
squares	79	4.5	70.4
many factors	316	17.9	87.9

Table 2–For each class of numbers discussed above, they give the number of occurrences in A, the corresponding percentage and the cumulative percentage in A.

3.3.4. Other cases

The set A thus contains almost all prime numbers, 95% of squares, and a significant percentage of numbers that have a multitude of factors and all the numbers possessing at least ten prime factors (counted with their multiplicity).

These different classes of numbers by themselves represent 87.9% of A. Among the remaining numbers, some evince outstanding properties, for example, linked to decimal notation, as in: 1111, 2222, 3333.... Others have a simple form, such as 1023, 1025, 2047, 2049... that are written $2^n + 1$ or $2^n - 1$.

When these cases that for one reason or another possess an evident "simplicity" are eliminated, there remains a proportion of less than 10% of numbers in A for which one cannot immediately discern any particular property.

4. Explanation of the cloud-shape formation

4.1. Overview of the theory of algorithmic complexity

Save in a few exceptional cases, for a number to possess a multitude of properties implies that the said properties are simple, where simple is taken to mean "what may be expressed in a few words". Conversely, if a number possesses a simple property, then it will possess many properties. For example, if n is a multiple of 3, then n will be a even multiple of 3 or a odd multiple of 3. Being a "even multiple of 3" or "odd multiple of 3" is a little more complex than just being a "multiple of 3", but it is still simple enough, and one may further propose that many sequences in Sloane's database are actually sub-sequences of other, simpler ones. In specifying a simple property, its definition becomes more complex (by generating a sub-sequence of itself), but since there are many ways to specify a simple property, any number that possesses a simple property necessarily possesses numerous properties that are also simple.

The property of n corresponding to a high value of $N(n)$ thus seems related to the property of admitting a "simple" description. The value $N(n)$ appears in this context as an indirect measure of the simplicity of n, if one designates as "simple" the numbers that have properties expressible in a few words.

Algorithmic complexity theory [6, 2, 7] assigns a specific mathematical sense to the notion of simplicity, as the objects that "can be described with a short definition". Its modern formulation can be found in the work of Li and Vitanyi [8], and Calude [1].

Briefly, this theory proposes to measure the complexity of a finite object in binary code (for example, a number written in binary notation) by the length of the shortest program that generates a representation of it. The reference to a universal programming language (insofar as all computable functions can possess a program) leads to a theorem of invariance that warrants a certain independence of the programming language.

More precisely, if L_1 and L_2 are two universal languages, and if one notes K_{L_1} (resp. K_{L_2}) algorithmic complexity defined with reference to L_1 (resp. to L_2), then there exists a constant c such that $|K_{L_1}(s) - K_{L_2}(s)| < c$ for all finite binary sequences s.

A theorem (see for example [Theorem 4.3.3. page 253 in [8]]) links the probability of obtaining an object s (by activating a certain type of universal

Turing Machines–called optimal–running on binary input where the bits are chosen uniformly random) and its complexity $K(s)$. The rationale of this theorem is that if a number has many properties then it also has a simple property.

The translation of this theorem for $N(n)$ is that if one established a universal language L, and established a complexity limit M (only admitting descriptions of numbers capable of expression in fewer than M symbols), and counted the number of descriptions of each integer, one would find that $N(n)/M$ (where $M = \sum_{i \in \mathbb{N}} N(i)$) is approximately proportional to $2^{-K(n)}$:

$$\frac{N(n)}{M} = \frac{1}{2^{K(n)+O(\ln(\ln(n)))}}.$$

Given that $K(n)$ is non computable because of the undecidability of the halting problem and the role of the additive constants involved, a precise calculation of the expected value of $N(n)$ is impossible. By contrast, the strong analogy between the theoretical situation envisaged by algorithmic complexity and the situation one finds when one examines $N(n)$ inferred from Sloane's database, leads one to think that $N(n)$ should be asymptotically dependent on $1/2^{K(n)}$. Certain properties of $K(n)$ are obliquely independent of the reference language chosen to define K. The most important of these are:

- $K(n) < \log_2(n) + 2\log_2(\log_2(n)) + c'$ (c' a constant)
- the proportion of n of a given length (when written in binary) for which $K(n)$ recedes from $\log_2(n)$ decreases exponentially (precisely speaking, less than an integer among 2^q of length k, has an algorithmic complexity $K(n) \leq k - q$).

In graphic terms, these properties indicate that the cloud of points obtained from writing the following $1/2^{K(n)}$ would be situated above a curve defined by

$$f(n) \approx \frac{h}{2^{\log_2(n)}} = \frac{h}{n}$$

(h being a constant), and that all the points cluster on the curve, with the density of the points deviating from the curve decreasing rapidly.

This is indeed the situation we observe in examining the curve giving $N(n)$. The theory of algorithmic information thus provides a good description of what is observable from the curve $N(n)$. That justifies an a posteriori

recourse to the theoretical concepts of algorithmic complexity in order to understand the form of the curve $N(n)$. By contrast, nothing in the theory leads one to expect a gap like the one actually observed. To the contrary, continuity of form is expected from the fact that $n + 1$ is never much more complex than n.

To summarize, if $N(n)$ represented an objective measure of the complexity of numbers (the larger $N(n)$ is, the simpler n), these values would then be comparable to those that yield $2^{-K(n)}$. One should thus observe a rapid decrease in size, and a clustering of values near the base against an oblique curve, but one should not observe a gap, which presents itself here as an anomaly.

To confirm the conclusion that the presence of the gap results from special factors and render it more convincing, we have conducted a Monte Carlo experiment.

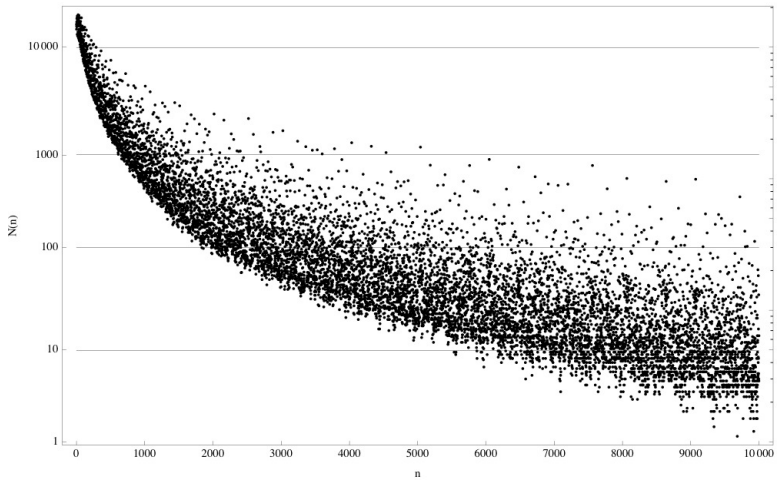

Figure 4: Graph of $N(n)$ obtained with random functions, similar to that belonging to Sloane's Database (Figure 1). Eight million values have been generated.

We define random functions f in the following manner (thanks to the algebraic system *Mathematica*):

1. Choose at random a number i between 1 and 5 (bearing in mind in the selection the proportions of functions for which $i = 1$, $i = 2$, ..., $i = 5$ among all those definable in this way).
2. If $i = 1$, f is defined by choosing uniformly at random a constant $k \in \{1, ..., 9\}$, a binary operator φ from among the following list: $+$, $\times$, and

subtraction sign, in a uniform manner, and a unary operand g that is identity with probability 0.8, and the function squared with probability .2 (to reproduce the proportions observed in Sloane's database). One therefore posits $f_i(n) = \varphi(g(n), k)$.

3. If $i \geq 2$, f_i is defined by $f_i(n) = \varphi(g(f_{i-1}(n)), k)$, where k is a random integer found between 1 and 9, g and φ are selected as described in the point 2 (above), and f_{i-1} is a random function selected in the same manner as in 2.

For each function f that is generated in this way, one calculates $f(n)$ for $n = 1,\ldots,20$. These terms are regrouped and counted as for $N(n)$. The results appear in Figure 4. The result confirms what the relationship with algorithmic complexity would lead us to expect. There is a decreasing oblique curve with a mean near 0, with clustering of the points near the base, but no gap.

4.2. The gap: A social effect?

This anomaly with respect to the theoretical implications and modeling is undoubtedly a sign that what one sees in Sloane's database is not a simple objective measure of complexity (or of intrinsic mathematical interest), but rather a trait of psychological or social origin that mars its pure expression. That is the hypothesis that we propose here. Under all circumstances, a purely mathematical vision based on algorithmic complexity would encounter an obstacle here, and the social hypothesis is both simple and natural owing to the fact that Sloane's database, while it is entirely "objective", is also a social construct.

Figure 5 illustrates and specifies our hypothesis that the mathematical community is particularly interested in certain numbers of moderate or weak complexity (in the central zone or on the right side of the distribution), and this interest creates a shift toward the right-hand side of one part of the distribution (schematized here by the grey arrow). The new distribution that develops out of it (represented in the bottom figure) presents a gap.

We suppose that the distribution anticipated by considerations of complexity is deformed by the social effect concomitant with it: mathematicians are more interested in certain numbers that are linked to selected properties by the scientific community. This interest can have cultural reasons or mathematical reasons (as per results already obtained), but in either case it brings

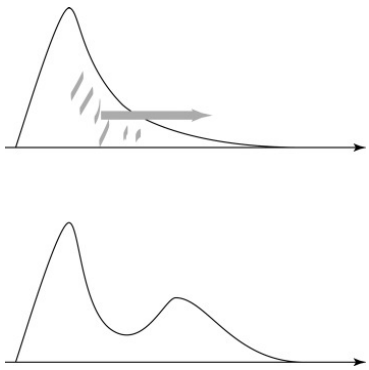

Figure 5: The top figure above represents the local distribution of N expected without taking into account the social factor.

with it an over-investment on the part of the mathematical community. The numbers that receive this specific over-investment are not in general complex, since interest is directed toward them because certain regularities have been discovered in them. Rather, these numbers are situated near the pinnacle of a theoretical asymmetrical distribution. Owing to the community's over-investment, they are found to have shifted towards the right-hand side of the distribution, thus explaining Sloane's gap.

It is, for example, what is generated by numbers of the form $2^n + 1$, all in A, because arithmetical results can be obtained from this type of number that are useful to prime numbers. Following some interesting preliminary discoveries, scientific investment in this class of integers has become intense, and they appear in numerous sequences. Certainly, $2^n + 1$ is objectively a simple number, and thus it is normal that it falls above the gap. Nevertheless, the difference in complexity between $2^n + 1$ and $2^n + 2$ is weak. We suppose that the observed difference also reflects a social dynamic which tends to augment $N(2^n+1)$ for reasons that complexity alone would not entirely explain, namely the importance of $2^n + 1$ in number theory. Moreover, if a number appears in diverse and apparently unlinked sequences, mathematicians are then driven to give them even more attention. This phenomenon of positive feedback may explain the overrepresentation of particularly simple numbers, as measured by N.

5. Conclusion

The cloud of points representing the function N presents a general form evoking an underlying function characterized by rapid decrease and "clustering near the base" (local asymmetrical distribution). This form is explained, at least qualitatively, by the theory of algorithmic information.

If the general cloud formation was anticipated, the presence of Sloane's gap has, by contrast, proved more challenging to its observers. This gap has not, to our knowledge, been successfully explained on the basis of uniquely numerical considerations that are independent of human nature as it impinges on the work of mathematics. Algorithmic complexity anticipates a certain "continuity" of N, since the complexity of $n+1$ is always close to that of n. The discontinuity that is manifest in Sloane's gap is thus difficult to attribute to purely mathematical properties independent of social contingencies.

By contrast, as we have seen, it is explained very well by the conduct of research that entails the over-representation of certain numbers of weak or medium complexity. Thus the cloud of points representing the function N shows features that can be understood as being the result of at the same time human and purely mathematical factors.

References

[1] CALUDE, C.S. *Information and Randomness: An Algorithmic Perspective.* (Texts in Theoretical Computer Science. An EATCS Series), Springer; 2nd. edition, 2002.

[2] CHAITIN, G.J. *Algorithmic Information Theory,* Cambridge University Press, 1987.

[3] CIPRA, B. "Mathematicians get an on-line fingerprint file," *Science*, 205, (1994), p. 473.

[4] DELAHAYE, J.-P. "Mille collections de nombres," *Pour La Science*, 379, (2009), p. 88-93.

[5] DELAHAYE, J.-P., ZENIL, H. "On the Kolmogorov-Chaitin complexity for short sequences", in CALUDE, C.S. (ed.) *Randomness and Complexity: from Chaitin to Leibniz*, World Scientific, p. 343-358, 2007.

[6] KOLMOGOROV, A.N. "Three approaches to the quantitative definition of information." *Problems of Information and Transmission*, 1(1): 1–7, 1965.

[7] LEVIN, L. *Universal Search Problems.* 9(3): 265-266, 1973 (c). (submitted: 1972, reported in talks: 1971). English translation in: B.A.Trakhtenbrot. *A Survey of Russian Approaches to Perebor (Brute-force Search) Algorithms.* Annals of the History of Computing 6(4): 384-400, 1984.

[8] LI, M., VITANYI, P. *An introduction to Kolmogorov complexity and its applications,* Springer, 1997.

[9] SLOANE, N.J.A. *A Handbook of Integer Sequences,* Academic Press, 1973.

[10] SLOANE, N.J.A. "The on-line encyclopedia of integer sequences," *Notices of the American Mathematical Society*, 8, (2003), p. 912-915.

[11] SLOANE, N.J.A. PLOUFFE, S. *The Encyclopedia of Integer Sequences,* Academic Press, 1995.