# Statistical Patterns in the Equations of Physics and the Emergence of a Meta-Law of Nature

Andrei Constantin,<sup>1,\*</sup> Deaglan Bartlett,<sup>2,†</sup> Harry Desmond,<sup>3,‡</sup> and Pedro G. Ferreira<sup>4,§</sup>

<sup>1</sup>Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK

<sup>2</sup> CNRS & Sorbonne Université, Institut d'Astrophysique de Paris (IAP), UMR 7095, 98 bis bd Arago, F-75014 Paris, France

<sup>3</sup>Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, UK

<sup>4</sup>Astrophysics, University of Oxford, DWB, Keble Road, Oxford OX1 3RH, UK

Physics, as a fundamental science, aims to understand the laws of Nature and describe them in mathematical equations. While the physical reality manifests itself in a wide range of phenomena with varying levels of complexity, the equations that describe them display certain statistical regularities and patterns, which we begin to explore here. By drawing inspiration from linguistics—where Zipf's law states that the frequency of any word in a large corpus of text is roughly inversely proportional to its rank in the frequency table—we investigate whether similar patterns for the distribution of operators emerge in the equations of physics. We analyse three corpora of formulae and find, using sophisticated implicit-likelihood methods, that the frequency of operators as a function of their rank in the frequency table is best described by an exponential law with a stable exponent, in contrast with Zipf's inverse power-law. Understanding the underlying reasons behind this statistical pattern may shed light on Nature's modus operandi or reveal recurrent patterns in physicists' attempts to formalise the laws of Nature. It may also provide crucial input for symbolic regression, potentially augmenting language models to generate symbolic models for physical phenomena. By pioneering the study of statistical regularities in the equations of physics, our results open the door for a meta-law of Nature—a (probabilistic) law that all physical laws obey.

### I. INTRODUCTION

The physical nature of reality is expressed, in its most intelligible form, in the language of mathematical formulae. These formulae describe a huge range of phenomena and include famous examples such as Newton's law of universal gravitational attraction  $F = GmM/r^2$ , Einstein's mass-energy equivalence formula  $E = mc^2$  and the Hawking–Bekenstein formula for the entropy of a black hole  $S = k_B Ac^3/4G\hbar$ . Expressed in prefix notation, the right-hand sides of these equations take the form of text, where words correspond to operators and operands:

 $\begin{aligned} &\operatorname{Div}(\operatorname{Prod}(G,\operatorname{Prod}(m,M)),\operatorname{Pow}(r,2)) \\ &\operatorname{Prod}(m,\operatorname{Pow}(c,2)) \\ &\operatorname{Div}(\operatorname{Prod}(k_B,\operatorname{Prod}(A,\operatorname{Pow}(c,3))),\operatorname{Prod}(4,\operatorname{Prod}(G,\hbar))) \,. \end{aligned} \tag{1}$ 

There is, certainly, a lot of structure underlying these and other similar equations in physics, including the following simple observations. (i) All physics equations involve variables such as the masses m, M or the horizon area A in the formulae above. (ii) Universal constants, like the gravitational constant G, the speed of light c, or Planck's constant  $\hbar$ , appear frequently. (iii) Numerical factors

These observations have long served as guiding principles for physicists in the process of discovering Nature's laws and formulating symbolic models to represent natural phenomena. Beyond these basic principles, however, it is not known if physical formulae hide more subtle patterns. In the present article we answer this question in the affirmative by studying the statistical distribution of operators in three corpora of formulae.

The exercise is inspired by linguistics, where Zipf's law dictates the frequency distribution of words in natural languages [1–4]. More explicitly, Zipf's law states that the frequency f of any word in a large corpus of text is nearly inversely proportional to its rank r in the frequency table or to some positive power  $r^{\alpha}$  thereof (where  $\alpha \simeq 1$ ), regardless of the language, genre or time of creation for the corpus:

Zipf's law: 
$$f(r) \sim \frac{1}{r^{\alpha}}$$
. (2)

Put differently, the most common word appears approximately twice as often as the second most common word, three times as often as the third most common word, and so forth. To date, there is no comprehensive explanation for the universal applicability of this striking observation,

also play a role, although to a lesser extent. (iv) Dimensional analysis dictates the structure of equations based on the dimensions of physical quantities. (v) Power-law relationships and scaling behaviour, such as the inverse square law governing Newtonian gravity, are prevalent.

 $<sup>\ ^* \</sup> and rei. constant in @physics.ox. ac. uk$ 

<sup>†</sup> deaglan.bartlett@iap.fr

<sup>†</sup> harry.desmond@port.ac.uk

<sup>§</sup> pedro.ferreira@physics.ox.ac.uk

although many qualitative and quantitative models have been proposed to understand its origin [5], including optimisation of communication efficiency [6–8], linguistic assumptions about text generation [9, 10] (e.g. a word that has already appeared in a text is more likely to appear again than a new word) and history-dependent reduction of sample-space [11, 12].

Most intriguingly, Zipf's law as well as its two-parameter generalisation, the

Zipf–Mandelbrot law: 
$$f(r) \sim \frac{1}{(r+b)^{\alpha}}$$
, (3)

have been observed in many and diverse other incarnations. For example, the distribution of city sizes in many countries follows Zipf's law, where a few large cities dominate while the majority of cities are small [13, 14]. In data science, the popularity of websites, the distribution of file sizes in computer systems including the Internet, have been shown to follow Zipf's law [15, 16]. In social sciences, the distribution of wealth [17] and the popularity of names [18] also exhibit Zipfian distributions. In neuroscience, the firing rates of neurons in the brain exhibit a Zipfian distribution, where a small number of neurons are highly active while the majority are less active [19–21].

Given the widespread manifestation of Zipf's law, it is tempting to search for statistical regularities in scientific corpora by isolating what arguably forms the core of scientific communication: mathematical expressions. One could, for instance, analyse Whitehead and Russell's three-volume *Principia Mathematica* [22] and study its distribution of symbols. If any pattern emerges, it would serve as a fingerprint of the formalism adopted by the authors for building up mathematical types, but may also reveal more universal regularities. Within the same realm of formal languages, one could study the *mathlib* Lean library of digitised pure mathematics, which currently includes nearly 200,000 theorems and definitions [23].

Our scope here is more limited. We study three corpora of physics formulae, one derived from The Feynman Lectures on Physics [24], one, inspired by Guimerà et al. [25], derived from Wikipedia's List of scientific equations named after people, and one derived from the Encyclopaedia Inflationaris review of inflationary cosmology [26]. The first corpus contains 100 formulae that describe fundamental physical principles including the basic formulae of classical physics, electrodynamics, thermodynamics, statistical mechanics and quantum mechanics. The second corpus is as broad as Feynman's lectures in the range of topics covered; however, most of its 41 equations are highly specialised. The third corpus focuses on a single theme, containing 71 distinct proposals for the inflaton potential which is thought to have played a determining role in the very early history of the universe. All the expressions considered in our analysis involve algebraic operations only; in particular, formulae involving differential or integral operators have not been included at this stage.

The motivation to focus on physical formulae is twofold. On the one hand, if any stable statistical patterns arise, as the following suggests to be the case, it would be interesting to understand how much of these can be attributed to linguistic considerations such as simplifications or compositionality and how much is determined by general properties of physical theories. On the other hand, much like with Zipf's law, for which we lack a comprehensive explanation but recognise its practical utility. being aware of the statistical regularities present in the equations of physics can be useful in practice. One potential application can be to refine symbolic regression methods which often struggle to control the proliferation of non-physical mathematical models. This idea has already been appreciated in Ref. [25] where the prior probabilities of mathematical expressions were computed using the relative frequency of operators in a corpus of formulae compiled from Wikipedia, and in Refs. [27, 28] where language models were trained to determine these probabilities while also accounting for correlations between operators.

This paper is organised as follows. In Section II we discuss the behaviour of randomly generated symbolic expressions and the resulting distribution of the frequency of their operators. We then introduce the corpora which we study in Section III, where we observe that their operator distribution is very different to randomly generated trees. In Section IV we then study the frequency—rank relations of these corpora and determine whether they obey a single, universal (meta)-law. We discuss these results and conclude in Section V.

# II. STATISTICAL PATTERNS IN BINARY EXPRESSION TREES

To simplify the discussion, in this work we focus on physical formulae of the form

$$quantity = algebraic expression, (4)$$

where the algebraic expression on the right hand side involves only nullary, unary and binary operators. The aim is to study the distribution of operators in the algebraic expressions arising in the three selected corpora.

Before engaging in this analysis, however, a couple of remarks are in order. It is well known that algebraic expressions involving nullary, unary and binary operators can be represented as binary trees, that is trees whose nodes can have zero, one, or two children. For instance, Newton's law of universal gravitational attraction included in Eq. (1) corresponds to the binary expression tree given in Fig. 1.

External nodes, i.e. nodes with no children, correspond to nullary operators. Every internal node of a binary expression tree has one or two children, corresponding to unary or binary operators, respectively. Using this

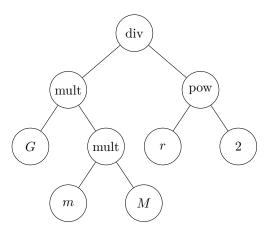


FIG. 1: Binary expression tree for Newton's law of gravitation.

correspondence, it is straightforward to show that any expression involving i binary operators has i+1 nullary operators, irrespective of the number j of unary operators. The total number of operators is then

$$n = 2i + j + 1, (5)$$

which gives a measure of the complexity (or length) of the expression. We will now discuss in turn the three types of operators, listing in each category the most common ones that appear in the equations of physics.

We distinguish between three types of nullary operators: (i) variables, denoted generically by the symbol x, (ii) numerical factors, denoted by the symbol a, and (iii) physical or mathematical constants, denoted by the symbol c. For instance, the expression for the Hawking-Bekenstein entropy contains one variable (the horizon area A), one numerical factor (the number 4 in the denominator) and four physical constants  $(k_B, c, G, \text{ and } \hbar)$ . Although of little importance to the final distribution of operators, we adopt the convention that mathematical constants such as  $\pi$  or Apéry's constant  $\zeta(3)$  should be treated in the same category as physical constants due to the fundamental role played in a large number of physical applications. Another possibility would be to distinguish between dimensional and dimensionless constants, however, in this approach quantities such as the fine-structure constant would fall in the same category as ordinary numbers.

The list of unary operators that appear frequently in the equations of physics includes abs, exp, log, neg,  $\sqrt{\cdot}$ , all (hyperbolic) trigonometric functions and their inverses. The notation is such that  $\mathrm{abs}(x) = |x|$  and  $\mathrm{neg}(x) = -x$  and all other operators have their standard meaning. We also include  $\mathrm{pow2}(x) = x^2$ ,  $\mathrm{pow3}(x) = x^3$  and  $\mathrm{pow4}(x) = x^4$ , as exponents corresponding to small integers appear much more frequently than any other exponents. Going up to power 4 is motivated by the fact that in the  $Encyclopaedia\ Inflationaris\ almost every\ expression\ involves\ the\ fourth\ power\ of\ an\ energy\ scale.$ 

Again, this is a matter of convention, however, its effect on the overall distribution of operators is small.

Finally, we consider five types of binary operators: mult, div, add, subtract and pow, where subtract(x, y) = x - y and pow $(x, y) = x^y$ . This concludes the discussion of the different types of operators commonly encountered in the equations of physics.

Returning to the analysis of binary expression trees, the observation that, for any algebraic expression, the number of nullary operators is one plus the number of binary operators represents a universal fact enforced by the syntax, with important implications for the frequency distribution of operators in any corpus of algebraic expressions. These implications can be made precise in the limit of large complexity: for any corpus of algebraic expressions of typical length  $n\gg 1$  and typical number of binary operators  $i\gg 1$ , the frequencies for the three types of operators, nullary, unary, and, respectively, binary, correspond to the following:

$$f_0: f_1: f_2 \simeq \frac{i}{n}: 1 - \frac{2i}{n}: \frac{i}{n}$$
 (6)

The expected value of the ratio i/n is, certainly, corpus dependent. For instance, in a corpus containing expressions of large complexity with no unary operators, the frequency distribution across the three arity types of operators is, straightforwardly,  $f_0: f_1: f_2=1/2:0:1/2$ . At the other extreme, in a corpus of randomly generated expressions of large complexity, the distribution is

$$f_0: f_1: f_2 = 1/3: 1/3: 1/3$$
, (7)

a statement that can be understood by the following argument. Random unlabelled binary expression trees can be constructed by starting with one active node (the root of the tree) and randomly assigning its number of children. For two children, the number of active nodes increases by one; for one child, the number stays the same, while for no children the number decreases by one. Making such assignments for every active node until no active nodes are left maps the process of generating a random expression tree of given complexity n to a onedimensional random walk with n steps starting at 1 (active node) and ending at 0 (active nodes), where each step can be a positive unit (corresponding to a node with two children), a negative unit (for a leaf) or a null step (for a node with one child). For example, by making the assignment of children left-to-right, top-to-bottom. the tree presented in Fig. 1 corresponds to the path  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 0$  or, equivalently, to the (n = 9)-tuple (1, 1, 1, -1, 1, -1, -1, -1, -1)which specifies the type of each node in the tree. The total number of random walks involving n steps of which i are forward, i+1 backward, and the rest null is

$$\frac{n!}{i!(i+1)!(n-2i-1)!} , \qquad (8)$$

since every such path can be mapped to a permutation of the n-tuple

$$\underbrace{(1, \ldots, 1}_{i \text{ times}}, \underbrace{0, \ldots, 0}_{n-2i-1 \text{ times}}, \underbrace{-1, \ldots, -1}_{i+1 \text{ times}}). \tag{9}$$

Note that not all of these paths correspond to connected graphs and, in fact, every path that returns to 0 before the final step is not of this type. However, there is an n-to-1 map between the set of all paths and those that return to 0 only at the final step. The expectation value of i/n can then be computed exactly from Eq. (8) as

$$\langle i/n \rangle = \frac{\sum_{i=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \frac{(n-1)!}{i!(i+1)!(n-2i-1)!} \frac{i}{n}}{\sum_{i=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} \frac{(n-1)!}{i!(i+1)!(n-2i-1)!}}, \quad (10)$$

where  $\lfloor x \rfloor$  is the floor function giving the greatest integer less than or equal to x. Another way of counting the total number of binary expression trees of length n with i < n/2 nodes with two children makes use of the Catalan number  $C_i$ , which gives the number of tuples of length 2i with i (1)'s and i (-1)'s such that no initial segment has a negative total. From these tuples one can construct all the n-tuples corresponding to binary expression trees by inserting n-2i-1 (0)'s in all possible places and placing (-1) at the end. See Ref. [29] for a related discussion.

In the limit of large n, the expectation value (10) is dominated by the terms where  $i \approx i+1 \approx n-2i-1$ , leading to the uniform frequency distribution given in Eq. (7) across the three operator types. It follows that in a corpus of algebraic expressions of relatively large complexities involving  $n_0$  types of nullary operators,  $n_1$  types of unary operators and  $n_2$  types of binary operators, the expected distribution is uniform within each operator type with frequencies

$$\frac{1}{3n_0}, \frac{1}{3n_1}, \frac{1}{3n_2},$$
 (11)

for individual operators belonging to the three arities.

#### III. PHYSICS CORPORA

In this section we describe the three corpora of physics formulae we study in this paper. The observed operator frequency distributions across the three corpora are summarised in Table I. Certainly, these are very different from the flat distribution in (11), but very similar to each other. It is helpful to outline the specifics of the three corpora, a description which supplements the details given in the Introduction.

The first corpus contains 100 formulae extracted from *The Feynman Lectures on Physics* that were used in Ref. [30] to test the AI Feynman symbolic regression model. The formulae involve a total of 1173 operators of 19 different types. The length of the expressions ranges between 3 and 30, with a large proportion of them falling in the range [5, 15], as shown in the complexity histogram in Fig. 2. The formulae cover a broad range of topics and include both foundational laws and more specialised results with wide applications in physics.

The second corpus contains 41 algebraic expressions taken from Wikipedia's List of scientific equations named after people, adding up to a total of 551 operators. The complexity of the expressions in this corpus varies between 3 and 62, as shown in Fig. 2. The corpus covers formulae that have played a pivotal role in the development of disciplines such as physics, chemistry, and engineering. In this regard, the corpus is similar to the foundational formulae included in The Feynman Lectures on Physics.

The third corpus contains a total of 1371 operators divided between 71 algebraic expressions corresponding to various hypothetical scenarios about the evolution of the scalar field (inflaton) responsible for the exponential expansion alleged to have happened in the very early universe, as reviewed in the *Encyclopaedia Inflationaris* of Ref. [26]. The complexity of the expressions varies between 7 and 77 and peaks in the interval [10, 20], as shown in Fig. 2. It includes a small number of large complexity expressions, a feature shared also by the Wikipedia corpus.

## IV. STATISTICAL PATTERNS IN PHYSICS FORMULAE

While the sizes of the three corpora are small, the emergent frequency-rank pattern is statistically significant. To illustrate this, we generated 21 corpora of random algebraic expressions with the same size and complexity distribution as the Feynman corpus. (Similar conclusions were obtained for randomly generated corpora of size and complexity distribution matching those of the Wikipedia corpus and the Encyclopaedia Inflationaris corpus.) We plot the resulting average frequencies and standard deviations in Fig. 3, giving a nearly piece-wise flat distribution. In the same plot, we display the frequencies corresponding to the piece-wise flat distribution of Eq. (11). Both of these are, by any reasonable measure, very far from the observed distribution of frequencies. The nearly piece-wise flat distribution arises from the fact that the mock data do not distinguish between operators of the same arity—at a given arity each operator is drawn randomly and independently of the previous draws. Given the drastically different distributions, it is clear (as observed by e.g. [25, 27]) that there is a preference for some operators over others in constructing physics equations.

TABLE I: Operator frequency tables for three corpora of physics formulae.

Corpus	(rank $r$ , frequency $f$ , operator)			
The Feynman Lectures on Physics	$\begin{array}{llllllllllllllllllllllllllllllllllll$			
Wikipedia named equations	$\begin{array}{llllllllllllllllllllllllllllllllllll$			
Encyclopaedia Inflationaris	$\begin{array}{llllllllllllllllllllllllllllllllllll$			

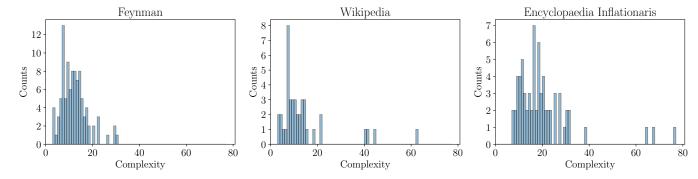


FIG. 2: Distribution of expression complexity in the three corpora, which approximately corresponds to the number of operarors appearing in the equation.

In the remainder of this section we wish to determine whether this distribution is described by a simple law, akin to linguistics.

To determine this, we consider three possible functional forms for the frequency—rank relation. Inspired by linguistics, we consider both the Zipf (Eq. (2)) and Zipf—Mandelbrot (Eq. (3)) laws, but also an exponential distribution

$$f(r) \sim \exp\left(-\beta r\right),$$
 (12)

where  $\beta > 0$  describes the steepness of the distribution.

Fitting such laws to observed data appears to be a nontrivial task, since constructing an analytic likelihood (the probability of the data given the parameters of the distribution) is challenging. Each point in the frequency—rank plane is correlated, not least because, by definition, the frequency must monotonically decrease as a function of rank. As such, simply fitting one of these distributions directly to the data with, e.g, a mean-square-error or Poissonian likelihood is unsatisfactory.

Although the likelihood is analytically intractable, drawing samples from one of these distributions to obtain a frequency–rank relation is simple. We therefore perform our analysis using simulation-based inference [31, 32] (SBI, a.k.a. likelihood free inference or implicit likelihood inference). For each distribution, we run 10,000 simulations, where for each simulation we randomly draw the corresponding parameters from the ranges  $\alpha \in [1,5]$ ,  $b \in [2,20]$ , and  $\beta \in [0.1,3]$ , using a uniform prior on  $\alpha^{-1}$ ,  $b^{-1}$ , and  $\beta$ . These parameters then define an underlying frequency–rank distribution, from which we draw N samples, where N corresponds to the number of observed operators in the corpus under study. These are then rebinned and sorted to obtain an observed frequency–rank distribution.

Since the number of different operators can vary in each simulation (and thus we have a varying length data vec-

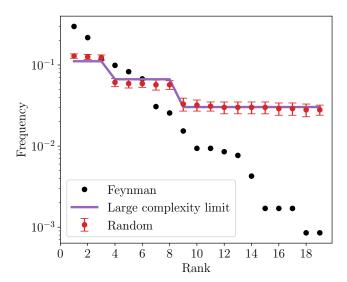


FIG. 3: Comparison between the Feynman Lectures corpus and randomly labelled binary trees. The black points correspond to the true data, whereas red points indicate the operator frequency distribution for randomly generated corpora of formulae of the same size and complexity distribution as the Feynman Lectures corpus. In purple, we give the expected piece-wise uniform distribution in the large complexity limit.

tor), we find it useful to compress this observed distribution into a fixed set of numbers. We fit the observed distribution to each of the distributions in Eqs. (2), (3) and (12), including an additional normalisation parameter, using a non-linear least squares method and store the optimised parameters, the resulting mean squared errors and the total number of operators into a vector of 11 numbers. We then approximate the posterior distribution (the probability of the underlying parameters given these 11 numbers), using neural posterior estimation, as implemented in the LTU-ILI package [33]. We utilise an ensemble of two neural networks: a mixture density network with 6 coupling layers and 50 hidden features, and a masked autoregressive flow with 5 coupling layers and 50 hidden features. These are trained with a batch size of 64 and learning rate of  $10^{-4}$ , and otherwise use the default settings of LTU-ILI. We use 80% of simulations for training and the remainder for testing. To verify the fidelity of our simulations, we ensure the inferred and true parameter values are consistent on our test set, perform a percentile-percentile test and use the 'Tests of Accuracy with Random Points' [TARP 34] method.

To obtain the posterior distribution given the true data, one simply evaluates these trained networks after compressing the data from the corpora into 11 numbers, as done for the training data. When doing this for the three corpora considered in this work, we obtain the parameter constraints given in Table II. After sampling from these posteriors, we plot the inferred distributions from these

models in Fig. 4, where we plot both the posterior mean and the 68% confidence region against the data. It is clear from this plot that attempting to fit the data using Zipf's law does not yield a satisfactory result, however the results from the exponential and Zipf–Mandelbrot law are more promising.

To quantify which model is preferred by the corpora, we compute the Bayesian evidence ratios, K, between the three models. This is computed using the LTU-ILI implementation of evidence networks [35], utilising a neural network of two layers of width 64. We train for 100 epochs using a learning rate of  $10^{-9}$  and a batch size of 512. The values when evaluated of the coprora are given in Table II.

As expected from Fig. 4, the Bayes factors for the Zipf model compared to the other models are large in magnitude, indicating a clear preference for the other two equations over a Zipf law. According to the Jeffreys scale [36], any value of  $\log_{10} K$  above 2 would indicate decisive evidence for one mdoel over than another; the lowest value across the corpora is 7.2. We conclude that the distribution of operators in physics equations is not Zipfian.

Given the reasonably similar behaviour of the Zipf–Mandelbrot and exponential laws in Fig. 4, it is unsurprising that the Bayes factor between these two models is much smaller, however all three prefer the exponential law. The Feynman corpus shows the clearest preference for this, with a Bayes factor of  $\log_{10} K = 5.5$ , again demonstrating decisive evidence, whereas the Wikipedia corpus has  $\log_{10} K = 1.7$  (very strong evidence) and the inflationary models corpus has  $\log_{10} K = 0.96$  (substantial evidence). This preference is reasonable given that both models give similar fits by eye to the data in Fig. 4, but the exponential law requires one fewer parameter. We note that, not only does each corpus prefer an exponential law, but that each corpus has a similar value of  $\beta$ , with  $\beta \sim 0.3$ .

### V. DISCUSSION AND CONCLUSIONS

The analysis pursued here demonstrates the emergence of a stable statistical pattern in the frequency distribution of operators in the equations of physics. The corpora of formulae extracted from *The Feynman Lectures of Physics* and Wikipedia's *List of scientific equations named after people* lead to an identical frequency—rank relation

$$f(r) \sim e^{-r/3},\tag{13}$$

although the order in which the operators appear in the corresponding frequency tables are different. Both of these corpora contain equations that describe fundamental physical phenomena. The third corpus was extracted from the *Encyclopaedia Inflationaris* and is best

	Feynman	Wikipedia	Encyclopaedia Inflationaris
Zipf	$\alpha = 2.6 \pm 0.2$	$\alpha = 2.3 \pm 0.2$	$\alpha = 2.5 \pm 0.6$
	$\log_{10} K = 9.6$	$\log_{10} K = 7.2$	$\log_{10} K = 10.8$
Zipf-Mandelbrot	$\alpha = 7.7 \pm 1.0$	$\alpha = 6.6 \pm 12.3$	$\alpha = 6.3 \pm 0.8$
	$b = 15.7 \pm 2.7$	$b = 12.3 \pm 3.4$	$b = 15.8 \pm 3.2$
	$\log_{10} K = 5.5$	$\log_{10} K = 1.7$	$\log_{10} K = 0.96$
Exponential	$\beta = 0.34 \pm 0.02$	$\beta = 0.34 \pm 0.03$	$\beta = 0.24 \pm 0.03$

TABLE II: Inferred parameters of the three distributions (Eqs. (2), (3) and (12)) from the corpora considered in this work. We also give the logarithm of the evidence ratios,  $\ln K$ , with respect to the exponential model, defined such that positive values indicate preference for the exponential model (as is the case for each corpus).

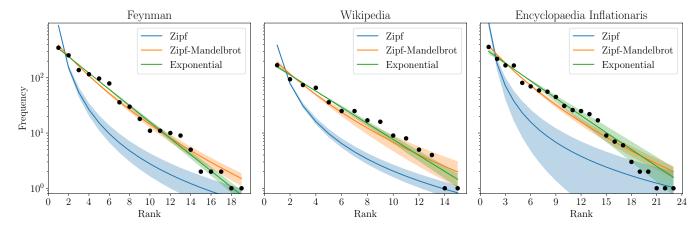


FIG. 4: Posterior distributions of the fits to the different corpora, where we compare a Zipf (Eq. (2)), Zipf–Mandelbrot (Eq. (3)), and exponential (Eq. (12)) fit. It is clear that the latter provides the best fit. The solid lines indicate the posterior mean, and the coloured bands show the 68% confidence interval.

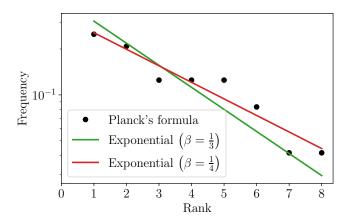


FIG. 5: Frequency of operators in the Planck formula against the frequency—rank relations from Eqs. (13) and (14).

described by a slightly different frequency-rank relation

$$f(r) \sim e^{-r/4}.\tag{14}$$

The difference may be attributable to the corpus containing hypothetical expressions related to speculative sce-

narios about the very early universe, in contrast to the empirically validated formulae in the other two corpora. Another potential interpretation, coming from a high energy physics perspective, is that while the Feynman and Wikipedia corpora describe infrared physics, the expressions in the Encyclopaedia Inflationaris corpus pertain to the ultraviolet, indicating that the exponent may run with the energy scale.

As with Zipf's law in linguistics, the underlying mechanisms that lead to the observed patterns are unclear. It may be that cultural bias strongly influences the way in which we formulate equations, leading us to emulate existing forms and thus perpetuate established patterns. There must also be at play certain elements of communication optimisation. For instance, the specific composition of exponential functions, addition, negation and division that produces tanh appears often enough that people have felt the need to give it a name. This could, therefore, explain the similarity between mathematical and natural language, as the units of communication (operators or words) are defined (by humans) to describe common ideas as succinctly as possible.

On the other hand, it is tempting to believe that the observed frequency—rank relation reflects a deeper aspect

of physical laws that could be regarded as a fundamental meta-law, a law about the laws of physics themselves. We refrain from further speculation. Instead, we conclude with the following intriguing observation about the well-known Planck formula for the spectral energy density of the black body radiation,

$$u_{\nu}(\nu, T) = \frac{8\pi h \nu^3}{c^3} \frac{1}{e^{\frac{h\nu}{k_B T}} - 1} . \tag{15}$$

Regarding the formula as a small corpus of 24 operators, we calculate the operator frequency table:

(1, 0.25, mult)	(5, 0.13, x)
(2, 0.21, c)	(6, 0.083, pow3)
(3, 0.13, a)	$(7,0.042,\exp))$
(4, 0.13, div)	(8, 0.042, subtract)

Quite remarkably, the frequency distribution in Planck's formula aligns with those observed in the three corpora discussed above, as shown in Fig. 5. This formula, which has been so instrumental in advancing our understanding of fundamental physics, appears to have perfected the balance of fundamental constants, variables, numbers, and operations.

However, this is not a singluar observation. In fact, every expression of relatively high complexity, say  $\gtrsim 20$ , from the three corpora analysed here closely follows the frequency–rank relation outlined above. This observation suggests that the degree to which the operator frequency distribution in a formula aligns with the observed distribution can serve as an indicator of the physical plausibility of a mathematical model. Such a measure has the potential to enhance symbolic regression algorithms (see e.g. [27, 37] for uses of language models in this context) by helping to filter out unphysical expressions of high complexity, and more generally to aid us in discovering new laws of physics.

#### ACKNOWLEDGEMENTS

AC's research is supported by a Royal Society Dorothy Hodgkin Fellowship, grant no. DHF/R1/231142. AC would like to thank Peter Keevash for a useful discussion about random binary graphs. DJB is supported by the Simons Collaboration on "Learning the Universe." HD is supported by a Royal Society University Research Fellowship (grant no. 211046). PGF acknowledges support from STFC and the Beecroft Trust.

For the purpose of open access, we have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

- [1] J.-B. Estoup, Gammes sténographiques: méthode et exercices pour l'acquisition de la vitesse (Institut Sténographique, 1916).
- [2] G. Dewey, Relative frequency of English speech sounds (Harvard University Press, 1923).
- [3] E. U. Condon, Statistics of vocabulary, Science 67, 300 (1928).
- [4] G. Kingsley Zipf, Selected studies of the principle of relative frequency in language (Harvard University Press, 1932).
- [5] M. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics 46, 323 (2005).
- [6] G. K. Zipf, The psycho-biology of language: an introduction to dynamic philology. (Houghton Mifflin, 1935).
- [7] G. K. Zipf, Human behavior and the principle of least effort. (Addison-Wesley Press, 1949).
- [8] S. T. Piantadosi, H. Tily, and E. Gibson, Word lengths are optimized for efficient communication, Proceedings of the National Academy of Sciences 108, 3526 (2011).
- [9] H. A. Simon, On a class of skew distribution functions, Biometrika 42, 425 (1955).
- [10] D. Zanette and M. Montemurro, Dynamics of text generation with realistic Zipf's distribution, Journal of Quantitative Linguistics 12, 29 (2005).
- [11] B. Corominas-Murtra, R. Hanel, and S. Thurner, Understanding scaling through history-dependent processes with collapsing sample space, Proceedings of the National Academy of Sciences 112, 5348 (2015).

- [12] S. Thurner, R. Hanel, B. Liu, and B. Corominas-Murtra, Understanding Zipf's law of word frequencies through sample-space collapse in sentence formation, Journal of the Royal Society Interface 12, 20150330 (2015).
- [13] H. A. Makse, S. Havlin, and H. E. Stanley, Modelling urban growth patterns, Nature 377, 608 (1995).
- [14] X. Gabaix, Zipf's law for cities: An explanation, The Quarterly Journal of Economics 114, 739 (1999).
- [15] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, Web caching and Zipf-like distributions: Evidence and implications, in *IEEE INFOCOM'99*. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now, Vol. 1 (IEEE, 1999) pp. 126–134.
- [16] A. Blank and S. Solomon, Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components), Physica A: Statistical Mechanics and its Applications 287, 279 (2000).
- [17] M. Cristelli, M. Batty, and L. Pietronero, There is more than a power law in Zipf, Scientific reports 2, 812 (2012).
- [18] D. H. Zanette and S. C. Manrubia, Vertical transmission of culture and the distribution of family names, Physica A: Statistical Mechanics and its Applications 295, 1 (2001).
- [19] T. Mora and W. Bialek, Are biological systems poised at criticality?, Journal of Statistical Physics 144, 268 (2011).

- [20] J. Tyrcha, Y. Roudi, M. Marsili, and J. Hertz, The effect of nonstationarity on models inferred from neural data, Journal of Statistical Mechanics: Theory and Experiment 2013, P03005 (2013).
- [21] D. J. Schwab, I. Nemenman, and P. Mehta, Zipf's law and criticality in multivariate data without fine-tuning, Physical Review Letters 113, 068102 (2014).
- [22] A. N. Whitehead and B. Russell, *Principia Mathematica* (Cambridge University Press, 1925–1927).
- [23] Wikipedia contributors, Lean (proof assistant) Wikipedia, The Free Encyclopedia (2024), [Online; accessed 12-August-2024].
- [24] R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman Lectures on Physics (Basic Books, New York, NY, 2010) Originally published 1963-1965.
- [25] R. Guimerà, I. Reichardt, A. Aguilar-Mogas, F. A. Massucci, M. Miranda, J. Pallarès, and M. Sales-Pardo, A Bayesian machine scientist to aid in the solution of challenging scientific problems, Science Advances 6 (2020).
- [26] J. Martin, C. Ringeval, and V. Vennin, Encyclopædia Inflationaris, Physics of the Dark Universe 5, 75 (2014), arXiv:1303.3787 [astro-ph.CO].
- [27] D. J. Bartlett, H. Desmond, and P. G. Ferreira, Priors for Symbolic Regression, in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO '23 Companion (Association for Computing Machinery, New York, NY, USA, 2023) p. 2402–2411, arXiv:2304.06333 [cs.LG].
- [28] T. Sousa, D. J. Bartlett, H. Desmond, and P. G. Ferreira, Optimal inflationary potentials, Phys. Rev. D 109, 083524 (2024), arXiv:2310.16786 [astro-ph.CO].
- [29] G. Lample and F. Charton, Deep learning for symbolic mathematics, arXiv:1912.01412 [cs.SC].

- [30] S.-M. Udrescu and M. Tegmark, AI Feynman: A physicsinspired method for symbolic regression, Science Advances 6 (2020), arXiv:1905.11481 [physics.comp-ph].
- [31] K. Cranmer, J. Brehmer, and G. Louppe, The frontier of simulation-based inference, Proceedings of the National Academy of Science 117, 30055 (2020), arXiv:1911.01429 [stat.ML].
- [32] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, Approximate Bayesian Computational Methods, Statistics and Computing 22, 1167 (2012).
- [33] M. Ho, D. J. Bartlett, N. Chartier, C. Cuesta-Lazaro, S. Ding, A. Lapel, P. Lemos, C. C. Lovell, T. L. Makinen, C. Modi, V. Pandya, S. Pandey, L. A. Perez, B. Wandelt, and G. L. Bryan, LtU-ILI: An All-in-One Framework for Implicit Inference in Astrophysics and Cosmology, The Open Journal of Astrophysics 7, 54 (2024), arXiv:2402.05137 [astro-ph.IM].
- [34] P. Lemos, A. Coogan, Y. Hezaveh, and L. Perreault-Levasseur, Sampling-Based Accuracy Testing of Posterior Estimators for General Inference, 40th International Conference on Machine Learning 202, 19256 (2023), arXiv:2302.03026 [stat.ML].
- [35] N. Jeffrey and B. D. Wandelt, Evidence Networks: simple losses for fast, amortized, neural Bayesian model comparison, Machine Learning: Science and Technology 5, 015008 (2024), arXiv:2305.11241 [cs.LG].
- [36] H. Jeffreys, The Theory of Probability (OUP Oxford, 1998) p. 432.
- [37] L. Biggio, T. Bendinelli, A. Neitz, A. Lucchi, and G. Parascandolo, Neural Symbolic Regression that Scales, in *Proceedings of the 38th International Con*ference on Machine Learning, Proceedings of Machine Learning Research, Vol. 139, edited by M. Meila and T. Zhang (PMLR, 2021) pp. 936–945.