Complex Distributions Emerging in Filtering and Compression

G. J. Baxter¹, R. A. da Costa^{1,*}, S. N. Dorogovtsev¹, and J. F. F. Mendes^{1,2}

¹*Departamento de Física da Universidade de Aveiro and I3N,
Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*

²*School of Computer and Communication Sciences and School of Life Sciences,
École Polytechnique Fédérale de Lausanne, 1015 Lausanne EPFL, Switzerland*

(Received 28 June 2019; revised manuscript received 10 February 2020; accepted 25 February 2020; published 30 March 2020)

In filtering, each output is produced by a certain number of different inputs. We explore the statistics of this degeneracy in an explicitly treatable filtering problem in which filtering performs the maximal compression of relevant information contained in inputs (arrays of zeros and ones). This problem serves as a reference model for the statistics of filtering and related sampling problems. The filter patterns in this problem conveniently allow a microscopic, combinatorial consideration. This allows us to find the statistics of outputs, namely the exact distribution of output degeneracies, for arbitrary input sizes. We observe that the resulting degeneracy distribution of outputs decays as $e^{-c \log^\alpha d}$ with degeneracy d , where c is a constant and exponent $\alpha > 1$, i.e., faster than a power law. Importantly, its form essentially depends on the size of the input dataset, appearing to be closer to a power-law dependence for small dataset sizes than for large ones. We demonstrate that for sufficiently small input dataset sizes typical for empirical studies, this distribution could be easily perceived as a power law. We extend our results to filter patterns of various sizes and demonstrate that the shortest filter pattern provides the maximum informative representations of the inputs.

DOI: [10.1103/PhysRevX.10.011074](https://doi.org/10.1103/PhysRevX.10.011074)

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

I. INTRODUCTION

Compression, filtering, and cryptography are related areas in signal and information processing [1–4]. By definition, a large number of possible inputs are mapped to a smaller number of possible outputs, so that a given output may correspond to multiple inputs, whose number is the output’s degeneracy. A similar problem emerges in cooperative systems with a large number of local minima in the energy landscape, in particular, in spin glasses and deep learning neural networks [5–7]. The configuration space of a system of this sort can be divided into a set of domains (basins) of attraction of these minima. One can ask, what is the statistics of these domains of attraction, and what is the distribution of their sizes? This is analogous to the degeneracy statistics problem. These statistics are important because the degeneracy distribution gives the most

relevant entropy when studying information problems [8]; that is, it is the distribution containing the most information relevant to the system under study.

These issues were explored in a recent series of works [9–13] which exploited the principle of maximum entropy; see also Refs. [14–18]. The finding of Refs. [9–13] is that maximally informative samples drawn from data exhibit statistics with broad distributions. More specifically, their entropy optimization based theory predicts power-law-like distributions of degeneracy of maximally informative outputs (minimal sufficient representations). The title of the paper “Minimum Description Codes Are Critical” [10] highlights the message of that theory in the most succinct way. On the other hand, distributions markedly distinct from power laws can be observed in empirical distributions for this kind of problem, for example, a collection of empirical curves in Fig. 1 of Ref. [10].

One obstacle to understanding such phenomena is that the need to obtain sufficient samples to resolve the degeneracy distribution limits the system size that can be examined. In the present work we contribute to these efforts by proposing a model for which sampling distributions can be obtained exactly, for the entire state space and up to large system sizes. We introduce and

*Corresponding author.

americo.costa@ua.pt

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

FIG. 1. (a) Filter extracting single ones and their positions from sequences of ones and zeros. Here the filter pattern is a single one. Ones and zeros are shown as black and white pixels, respectively. (b) A number of different inputs produce the same output. This number is the degeneracy of the output.

explore a reference filtering problem straightforwardly treatable through purely combinatorial techniques. This filter extracts all positions of a given local pattern in the input; see Fig. 1. We study a family of such filter patterns and demonstrate that the smallest pattern, extracting single ones from the inputs, generates outputs with the highest entropy of the degeneracy distribution, called relevance; see Fig. 2 and Table I. According to Refs. [10,12], this type of information entropy is maximized by the representations of the samples of a complex system which provide the most information, the maximally informative representations, which occurs in a critical region where the sample distribution becomes complex. Outside of this region outputs are either nearly all the same or nearly all different, giving

FIG. 2. Degeneracy distribution and cumulative degeneracy distributions for alternative filters to the one used in our main results. (a),(b) Filter pattern 01010 for binary input strings of length $n = 36$. (c),(d) Filter pattern 01010 for binary input strings of length $n = 36$. (e),(f) Filter pattern 011010 for binary input strings of length $n = 36$. (g),(h) Filter pattern (as illustrated herein) for 2D binary input grid of side length $L = 6$ ($n = 36$ entries total).

TABLE I. Values of the numbers z_g , z_d , and z_a for different filters. These give the asymptotics of, respectively, the total number of outputs $M(n) \cong z_g^n$, the largest degeneracy $d_D(n) \cong z_d^n$, and the number of outputs with degeneracy one, $N(1, n) \cong z_a^n$ for large n . The final row corresponds to a simple filter, as illustrated, applied to 2D input (see also Fig. 2). Note that we also included filter patterns consisting of all zeros. For each filter we also give the relevance $H[d]$ calculated from the degeneracy distribution and the resolution $H[y]$ for inputs of size $n = 36$. For the sake of comparison, the standard entropy of the inputs of this size is $H = 36 \ln 2 = 24.95330$. Finally, we include the number of distinct degeneracies D for each pattern. The boldface font patterns indicate a subset of filters that we discuss specifically in the main text.

Pattern		z_g	z_d	z_a	$H[d], n = 36$	$H[y], n = 36$	$D, n = 36$
010		1.618 03	1.754 88	1.465 57	6.209 26	12.667 44	777
000		1.618 03	1.839 29	1.497 10	5.230(5)	10.837 68	554(2)
0110	}	1.465 57	1.866 76	1.220 74	4.277 11	8.059 33	698
0100							
0000		1.528 95	1.927 56	1.419 63(2)	3.188 01	6.362 20	311
00100		1.465 57	1.9417	1.380 28	2.316 33	4.813 44	291
01110	}	1.380 28	1.933 18	1.167 30	2.272 38	4.882 24	255
01100							
01000							
01010		1.443 27	1.947 89	1.324 72	2.202 02	4.530 35	301
00000		1.465 57(2)	1.965 95	1.3652(2)	1.839 04	3.618 70	190
001100	}	1.380 28	1.969 31	1.2499(2)	1.298 00	2.843 46	197
001000							
010010		1.371 08(1)	1.971 13	1.1938(5)	1.291 02	2.769 17	218
011110	}	1.324 72	1.967 17	1.134 72	1.241 24	2.857 88	123
011100							
011010							
011000							
010100							
010000							
000000		1.4176(2)	1.983 58	1.324 86	1.068 43	2.018 05	123
0110110		1.324 72	1.985 74	1.158(2)	0.750 26	1.566 92	129
0111110	}	1.285 20	1.983 86	1.112 78	0.725 61	1.632 75	64
0111010							
01111110		1.255 42	1.992 03	1.096 98	0.436 83	0.916 60	36
011111110		1.232 05	1.996 05	1.085 07	0.261 54	0.507 86	25
0111111110		1.213 15	1.998 03	1.075 77	0.153 85	0.278 53	16
		1.50305	1.9476(1)	1.42(2)	1.464 18	4.726 64	624

no useful information about the underlying system. Note that the entropy of the samples themselves, called the resolution, is not maximum in this critical region, but in the limit that all samples are different. We establish a link between filtering and sampling problems based on the idea that filter patterns of different lengths with their positions can serve as sampling variables. We demonstrate that the behavior of our filtering problem is analogous to that found in the more complex systems of Refs. [9–13]. The entropy of the degeneracy distribution of filter outputs increases as the filter length decreases, but vanishes in the extreme limit, for which resolution is maximized, with outputs being identical to

inputs. Thus, within a family of filters, the most informative filter is the shortest one with length greater than one.

We directly obtain the complex degeneracy distributions for outputs generated from various input sets. We develop efficient recursive methods that allow us to find this distribution for large input strings, as can be seen in Fig. 3, which would not be accessible through empirical sampling methods. Because of the tractability of this problem, we are able to identify precisely how these distributions deviate from a power-law dependence. We find scaling forms that accurately describe the tail of the degeneracy distributions (see Fig. 4). We discover that

FIG. 3. (a),(c),(e) Degeneracy distribution for the complete input dataset: number of outputs of a given degeneracy versus degeneracy. (b),(d),(f) Cumulative degeneracy distribution for the complete input dataset: number of outputs of degeneracy greater than or equal to a given degeneracy versus degeneracy. The input length is $n = 21, 50, 120$. Open square symbols represent a mean-field approximation to the cumulative distribution, given by Eq. (38).

these distributions are essentially shaped by the size of the input data. That is, input datasets of relatively small size, typical for empirical studies, produce degeneracy distributions of outputs that are closer to a power law than the distributions of outputs from very large input datasets. Finally, we develop a mean-field theory and obtain the asymptotics of the degeneracy distribution and the spectrum of degeneracies. Our findings indicate that the phenomena we observe should apply to more general filtering and compression problems.

Our paper is organized in the following way. In Sec. II. we introduce our filter and the synthetic input datasets for it. In Sec. III, we describe how the filtering problem can be viewed as sampling of a complex

system. In Secs. IV–VI, we obtain the basic relations for the degeneracy of outputs, develop our algorithm, obtain the complex degeneracy distributions of outputs for the complete input datasets, and describe their features. We develop the mean-field theory of these distributions in filtering in Sec. VII. In Sec. VIII, we obtain the degeneracy distributions of outputs for uniformly random input datasets of various sizes. In Sec. IX, we discuss our results and indicate possible generalizations. In the Appendixes A–C, we give the combinatorial derivations of recursive relations used in Sec. IV and explicit asymptotics. The exact results obtained by our algorithm and recursions are provided in Supplemental Material [19].

FIG. 4. (a) Cumulative degeneracy distribution for $n = 20, 40, 60, 80, 100, 120$. The black curves represent least-squares fittings of $\ln N_{\text{cum}}(d, n)$ as $\ln N_{\text{cum}}^*(n) + B_n \ln^{\alpha_n} d$ for each n . (b) Cumulative degeneracy distribution $\ln\{-\ln[N_{\text{cum}}(d, n)/N_{\text{cum}}^*(n)]\}$ versus $\ln \ln d$ for $n = 20, 40, 60, 80, 100, 120$. Inset: Exponent α versus $1/n$.

II. REFERENCE FILTERING PROBLEM

We study the distribution of outputs in a solvable filtering problem by implementing a purely combinatorial, microscopic approach not involving entropy considerations. Let the input data be a set of N strings of zeros and ones (x_i), $x_i = 0, 1$, of length n , assuming the periodic condition $x_1 = x_{n+1}$. We consider two types of datasets. The first set is the complete set of all possible unique inputs. Its size N is determined by the size n of inputs, $N = 2^n$. Second, we consider datasets of arbitrary size N consisting of strings of uniformly randomly generated zeros and ones constrained by the same periodic condition as above. In the latter situation, some of the elements of a dataset may coincide. Clearly, in the limit $N \rightarrow \infty$, we arrive at a situation equivalent to

the complete dataset. (We stress that the random dataset of size $N = 2^n$ differs from the complete dataset.)

The filter works as follows. Every instance of a specific pattern in the input is marked by a one in the corresponding position in the output. All other positions are marked with zeros. This produces a minimal coding of the positions of the pattern occurrences in the input. We illustrate the results from a few example filter patterns in Fig. 2. We observe complex degeneracy distributions reminiscent of those observed in, e.g., Ref. [20].

For the sake of simplicity, in the remainder of this paper we focus on the simplest of these filters. Each sequence of ones of length 1 in the input (i.e., every 1 whose neighbors are both zeros) gives 1 at the same position in the output. All other sequences of ones or zeros in the input produce zeros in the corresponding places in the output, as shown in Fig. 1. In other words, the input vector (x_i) , $i = 1, 2, \dots, n$, $x_i = 0, 1$, is transformed to the output vector with the following components:

$$y_i = (1 - x_{i-1})x_i(1 - x_{i+1}). \quad (1)$$

As we will see, this filter is the maximally informative one, in a family of similar filters. Other filter patterns may be analyzed using the same methods we describe below.

III. FILTERING AS A SAMPLING PROBLEM

Let us discuss the link between filtering and sampling in more detail. The input x_i to our filtering problem is a string of n ones and zeros. It contains alternating contiguous chains of ones and zeros of different lengths, so that one may completely define the sequence by listing the starting positions of all the strings of ones and their respective lengths. With this in mind, one may reformulate the problem of filtering for patterns in sequences as sampling a system of correlated discrete variables in the following way. Let us introduce the set of discrete variables $\{s_i^l\}$, for $i = 1, \dots, n$ and $l = 1, \dots, n - 1$. The (“spin”) variable s_i^l is 1 if the input has a string of l ones beginning at position i , and it is 0 otherwise. These variables correspond to the outputs of filters for chains of consecutive ones. We include one more variable which distinguishes between the states of all zeros and all ones. The system, then, has 2^m , $m = n(n - 1) + 1$, possible states.

Strings of ones in the input do not overlap with each other, which restricts the configurations of the variables $\{s_i^l\}$ that can occur. We include these constraints in our system by imposing a cost for each violation of the mutual exclusion rules, the cost function being

$$U(\vec{s}) = \sum_{s_i^l, s_j^m} s_i^l s_j^m [\Theta(i + l, j) + \Theta(j + m, i)], \quad (2)$$

where

$$\Theta(a, b) = \begin{cases} 1 & a \bmod n \geq b \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

functions as a cyclic Heaviside function. We now have a complex system consisting of m state variables, whose interactions are governed by the cost function U . Each of the 2^n possible input sequences corresponds to a different minima of this cost function.

Now let us consider the sampling of this system. Suppose we may measure only n of these variables s_i^l . Which n variables will be most informative about the state of the system? Intuitively, shorter filter patterns, i.e., smaller values of l , should provide more detailed information about the system, and indeed the variables $s_i^{l=1}$ contain the most information about the state of the system. [To see this, note that the variable s_i^l gives information about the state of $l + 2$ input digits. In particular, when $s_i^l = 1$, it fully determines $l + 2$ digits of the input, which happens for a fraction $2^{-(l+2)}$ of the inputs. Thus, the total amount of information conveyed by s_i^l in terms of the number of digits of the input it reveals is $(l + 2)2^{-(l+2)}$, which is indeed maximum for $l = 1$.]

The set $\{s_i^1\}$ is also the most informative representation of the input from the perspective of Ref. [20], in the sense that samples obtained from observing these n variables have a larger entropy of their degeneracy distributions, Eq. (30), than samples using any other subset of n variables. This is apparent from the fact that any subset of n variables containing larger filters will have fewer available configurations; see also Table I. We detail this analogous behavior further in Sec. VI.

IV. OUTPUTS AND THEIR DEGENERACIES FOR COMPLETE INPUT DATASET

Let us begin by considering the complete input dataset, all 2^n different configurations of zeros and ones. Each output consists of isolated ones separated by strings of zeros of various lengths. The total number of possible outputs $M(n)$ is then the number of ways of arranging isolated ones in a chain of length n . This number coincides with the number of different configurations of dimers in a closed chain (ring) of length n :

$$M(n) = n \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{1}{n-k} \binom{n-k}{k}, \quad (4)$$

where $\lfloor n/2 \rfloor$ is the integer part of $n/2$; see Appendix A 1.

Comparing this number for successive values of n , we see that $M(n) = M(n-1) + M(n-2)$, with initial conditions $M(3) = 4$, $M(4) = 7$. The elements of the sequence

may be written in terms of the roots of the characteristic equation [21–23]

$$z^2 - z = 1. \quad (5)$$

Thus,

$$M(n) = \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n \cong \left(\frac{1+\sqrt{5}}{2}\right)^n = z_g^n, \quad (6)$$

where the last expression gives the large n limit. Here the largest root, $(1 + \sqrt{5})/2 = 1.61803... \equiv z_g$, is the famous golden ratio.

The key point for our study is that the degeneracy of an output is the product of the degeneracies of the strings of zeros between ones in the output. As is clear from Eq. (1), the presence of a 1 in an output at position i fixes the digits of its inputs at positions $i-1$, i , and $i+1$. These digits must be 0, 1, and 0, respectively. This is true for each 1 in the output. On the other hand, each of the remaining digits of these inputs compatible with a given output must be either a 0 or a 1 with one or two neighboring ones. All the degrees of freedom in the input corresponding to a given output correspond to these digits.

Let an output with $m \geq 1$ ones contain m strings of zeros with lengths $\ell_1, \ell_2, \dots, \ell_m$. Then the degeneracy of this output equals

$$d = \prod_{i=1}^m \tilde{d}(\ell_i). \quad (7)$$

Here $\tilde{d}(\ell)$ is the number of input strings of length ℓ , having the first and last digits 0, that generate an output string of ℓ zeros. This number plays an important role in our problem, similar to prime numbers in number theory, so we call the $\tilde{d}(\ell)$ *prime degeneracies*. Suppose that the output contains μ_ℓ strings of zeros of length ℓ , $\ell = 1, 2, \dots$, where

$$m + \sum_{\ell \geq 1} \ell \mu_\ell = n. \quad (8)$$

Then Eq. (7) may be rewritten as

$$d = \prod_{\ell \geq 1} [\tilde{d}(\ell)]^{\mu_\ell} \quad (9)$$

for $m \geq 1$.

Let us find the prime degeneracies explicitly. The number of these configurations, i.e., the degeneracy $\tilde{d}(\ell)$ of the output string of ℓ zeros, can be obtained recursively by taking into account three points.

- (i) Relevant input configurations of length ℓ are obtained by inserting 0 or 1 into each relevant configuration of length $\ell - 1$ between the first and second positions of the sequence. (Recall that the first and last positions of the input sequence are fixed to 0.)
- (ii) Input strings of length ℓ beginning and/or ending with 010 are irrelevant, and so they should be removed from the set generated at the previous step. These configurations can be obtained by inserting two digits 10 into each relevant input string of length $\ell - 2$ between its first and second positions.
- (iii) Finally, there exist input strings, compatible with the output string of ℓ zeros, that cannot be obtained by inserting a single digit into relevant input strings of length $\ell - 1$ between their first and second positions. These are the input strings of length ℓ beginning with 0110. These inputs can be obtained by inserting 110 into each relevant input string of length $\ell - 3$ between their first and second positions.

Following these rules, the degeneracy of a string of ℓ zeros at the output, prime degeneracy $\tilde{d}(\ell)$, can be written recursively as a linear difference equation:

$$\tilde{d}(\ell) = 2\tilde{d}(\ell - 1) - \tilde{d}(\ell - 2) + \tilde{d}(\ell - 3), \quad (10)$$

with the initial condition $\tilde{d}(1) = \tilde{d}(2) = \tilde{d}(3) = 1$. The solution of Eq. (10) may be explicitly expressed in terms of the complex roots z_1 , z_2 , and z_3 of the characteristic equation,

$$z^3 = 2z^2 - z + 1, \quad (11)$$

giving

$$\tilde{d}(\ell) = C_1 z_1^\ell + C_2 z_2^\ell + C_3 z_3^\ell, \quad (12)$$

where

$$\begin{aligned} C_1 &= (z_1 - 1)/[(z_1 - z_2)(z_1 - z_3)], \\ C_2 &= (z_2 - 1)/[(z_2 - z_1)(z_2 - z_3)], \\ C_3 &= (z_3 - 1)/[(z_3 - z_1)(z_3 - z_2)]. \end{aligned} \quad (13)$$

One of the roots of Eq. (11), z_1 , say, is real:

$$z_1 \equiv z_d = 1.75488... \quad (14)$$

It determines the large ℓ asymptotics of the prime degeneracies $\tilde{d}(\ell)$:

$$\tilde{d}(\ell) \cong \frac{z_d}{z_d^4 - 2} z_d^\ell, \quad (15)$$

where we used the identity $C_1 = z_d/(z_d^4 - 2)$. The two other roots are complex conjugate numbers,

$$z_{2,3} = 0.122561... \pm 0.744862...i. \quad (16)$$

The special case of the periodic output of length n with all digits 0 has to be considered separately. First let us take an arbitrary digit of the input. The number of input configurations where this digit is 0 and the resulting output has only zeros is given by $\tilde{d}(n + 1)$, because the fixed 0 of the periodic input plays the role of first and last digit of the configurations of a string of $n + 1$ digits. If the chosen digit is 1, then the number of the relevant input configurations equals $1 + \sum_{i=2}^{n-1} i\tilde{d}(n - i)$, where the sum over i accounts for the configurations where the digit is in a group of i consecutive ones, plus one configuration with all input digits equal to 1. Consequently, we obtain the following expression for the degeneracy of the output with all zeros:

$$d_D(n) = 1 + \tilde{d}(n + 1) + \sum_{i=2}^{n-1} i\tilde{d}(n - i), \quad (17)$$

which is the largest possible degeneracy of an output of a given length. Applying the recursion relation for prime degeneracies $\tilde{d}$, Eq. (10), to the terms on the right-hand side of Eq. (17), we find that the largest degeneracy $d_D(n)$ satisfies the same difference equation as Eq. (10) though with different initial condition; see, e.g., Ref. [24,25]. We present this equation here for future reference:

$$d_D(n) = 2d_D(n - 1) - d_D(n - 2) + d_D(n - 3), \quad (18)$$

with the initial condition $d_D(3) = 5$, $d_D(4) = 10$, and $d_D(5) = 17$. With these initial condition, we get the explicit solution of this equation,

$$d_D(n) = z_1^\ell + z_2^\ell + z_3^\ell, \quad (19)$$

where $z_1 \equiv z_d$, z_2 , z_3 are given by Eqs. (14) and (16), and its large n asymptotics,

$$d_D(n) \cong z_d^n. \quad (20)$$

V. CALCULATING THE EXACT DEGENERACY SPECTRUM

In principle, for each of the 2^n possible inputs we can obtain, one by one, an output numerically by applying Eq. (1). In practice, we use a more efficient algorithm described below. This algorithm focuses on outputs with a fixed number of ones and exploits the factorization of the output degeneracies; see Eq. (7).

The full list of degeneracies can be found from integer partitions in an explicit form, as follows. Let us introduce

the operator $\mathcal{P}$, which generates all integer partitions of positive integer k into r integers; that is, $\mathcal{P}(k, r)$ is the matrix whose rows $i = 1, 2, \dots, w$ are all different integer partitions $\{\mathcal{P}_{i1}(k, r), \mathcal{P}_{i2}(k, r), \dots, \mathcal{P}_{ij}(k, r), \dots, \mathcal{P}_{ir}(k, r)\}$ of k into r integers, $\sum_{j=1}^r \mathcal{P}_{ij}(k, r) = k$ [26,27]. For an output of length n , we consider all partitions of $n - m$ into m integers, for all possible $m = 1, 2, \dots, [n/2]$. For each such partition [i.e., a row of $\mathcal{P}(n - m, m)$], we find the degeneracy $d_i = \prod_j \tilde{d}(\mathcal{P}_{ij}(n - m, m))$ [see Eq. (7)]. Some of them coincide, so we find the union of them. Finally, we add the largest degeneracy $d_D(n)$ (corresponded to $m = 0$) to the resulting set. In summary, for the full set of degeneracies $\mathcal{D}_{\text{full}}(n)$, we have

$$\mathcal{D}_{\text{full}}(n) = d_D(n) \cup \left\{ \bigcup_{m=1}^{[n/2]} \left[\bigcup_{i=1}^w \prod_{j=1}^m \tilde{d}(\mathcal{P}_{ij}(n - m, m)) \right] \right\}, \quad (21)$$

with $\tilde{d}(\ell)$ and $d_D(n)$ provided by Eqs. (10) and (15), respectively.

For each integer partition, i.e., for each row i of the matrix $\mathcal{P}(n - m, m)$, we introduce a function giving the number of pieces of length ℓ present in this partition,

$$\mu_{\ell}^{(i)}(n - m, m) \equiv \sum_{j=1}^m \delta(\ell, \mathcal{P}_{ij}(n - m, m)), \quad (22)$$

so that

$$\begin{aligned} n - m &= \sum_{\ell} \ell \mu_{\ell}^{(i)}(n - m, m), \\ m &= \sum_{\ell} \mu_{\ell}^{(i)}(n - m, m). \end{aligned} \quad (23)$$

One can then write

$$d_i = \prod_j \tilde{d}(\mathcal{P}_{ij}(n - m, m)) = \prod_{\ell} [\tilde{d}(\ell)]^{\mu_{\ell}^{(i)}(n - m, m)}. \quad (24)$$

The number of outputs that contain m chains of zeros with lengths specified by the integer partition $\mathcal{P}_i(n - m, m)$ is then obtained by considering the number of distinct permutations of the m strings of zeros, multiplying by n , and finally dividing by m , giving

$$\mathcal{N}[\mathcal{P}_i(n - m, m)] = n \frac{(m - 1)!}{\prod_{\ell} \mu_{\ell}^{(i)}(n - m, m)!}, \quad (25)$$

where the product in the denominator is over the lengths of the parts of this partition. The total number of outputs with degeneracy d is then finally:

$$\begin{aligned} \mathcal{N}(d, n) &= \delta[d, d_D(n)] + \sum_{m=1}^{[n/2]} \sum_{i=1}^w n \frac{(m - 1)!}{\prod_{\ell} \mu_{\ell}^{(i)}(n - m, m)!} \\ &\times \delta \left[d, \prod_{\ell} [\tilde{d}(\ell)]^{\mu_{\ell}^{(i)}(n - m, m)} \right], \end{aligned} \quad (26)$$

where $\delta(a, b)$ is the Kronecker symbol. This is the expression we use for computing $\mathcal{N}(d, n)$ in an efficient way. For the sake of brevity, hereafter we refer to $\mathcal{N}(d, n)$ as a degeneracy distribution, without normalization.

The distribution $\mathcal{N}(d, n)$ can also be built up recursively starting from small values of d and n , as we show in Appendix A 2.

VI. OUTPUT DEGENERACY DISTRIBUTION FOR COMPLETE INPUT DATASETS

Using this algorithm we obtained the number of outputs $\mathcal{N}(d)$ for the full spectrum of degeneracies d for n up to 120; see Supplemental Material [19]. These results demonstrate that the degeneracies d_i , $i = 1, \dots, D$, form a discrete spectrum of values where D is the rank of the largest degeneracy and $d_1 = 1$.

Figure 3 shows the resulting distribution and the corresponding cumulative distribution $\mathcal{N}_{\text{cum}}(d)$ for $n = 21, 50$, and 120. Here, $\mathcal{N}_{\text{cum}}(d_i) \equiv \sum_{j=i}^D \mathcal{N}(d_j)$. In particular, $\mathcal{N}_{\text{cum}}(d_1) = M(n)$, i.e., the total number of outputs. Figures 3(b), 3(d), and 3(f) demonstrate that the cumulative degeneracy distributions decay with d more rapidly than a power law. On the other hand, the decay of the cumulative distribution is well described by the function

$$\mathcal{N}_{\text{cum}}(d) \propto e^{-c \ln^{\alpha} d} = d^{-c \ln^{\alpha-1} d}, \quad (27)$$

where c is a positive number and exponent α approaches 2.3 as $n \rightarrow \infty$; see the inset in Fig. 4(b). Notice that the degeneracy distribution for smaller n , Fig. 3(a), appears more like a power law than the degeneracy distribution for larger n , Fig. 3(c), for example, since the range of d increases with n , while the exponent $c \ln^{\alpha-1} d$ of the function $d^{-c \ln^{\alpha-1} d}$ varies slowly. Similarly, the cumulative distribution plotted in log-log scale for $n = 21$ deviates from linear (power-law behavior) noticeably less than for larger n . The wide range of degeneracies d that we observe enables us to present Fig. 4(b) showing $\ln\{-\ln[\mathcal{N}_{\text{cum}}(d)/\mathcal{N}_{\text{cum}}(1)]\}$ versus $\ln \ln d$. This plot supports the functional form given in Eq. (27). Note that in Fig. 4 we assumed that the coefficient factor of the asymptotics in Eq. (27) is close to $\mathcal{N}_{\text{cum}}(1)$, which is justified by the results. Figure 5 shows how the set of degeneracies $\mathcal{D}_{\text{full}}(n)$ varies with n ; see below for more detail.

In Sec. IV, we derived the explicit expression for the largest degeneracy $d_D(n)$ corresponding to the output with all zeros, Eq. (19), and found its large n asymptotics,

FIG. 5. The set of degeneracy values d for each n (logarithmic scale). (The range of n was selected to make individual points visible.)

$d_D(n) \cong z_d^n$, Eq. (20). As is natural, $\mathcal{N}(d_D, n) = 1$. The second largest degeneracy corresponds to an output with a single 1. The third largest degeneracy is for an output with two ones separated by a single 0. The fourth largest degeneracy is of the output with two ones separated by three zeros. Clearly, $\mathcal{N}(d_{D-1}, n) = \mathcal{N}(d_{D-2}, n) = \mathcal{N}(d_{D-3}, n) = n$. Using the asymptotics of $\tilde{d}(\ell)$, Eq. (15), we find that, asymptotically, at large n ,

$$\begin{aligned} d_{D-1}(n) &\cong \frac{z_d^n}{7.48391\dots} = \frac{z_d^n}{z_d^4 - 2}, \\ d_{D-2}(n) &\cong \frac{z_d^{n-2}}{z_d^4 - 2}, \\ d_{D-3}(n) &\cong \frac{z_d^{n-3}}{z_d^4 - 2}. \end{aligned} \quad (28)$$

One may also notice a complex structure in the cumulative distributions Figs. 3(b), 3(d), and 3(f), $\mathcal{N}_{\text{cum}}(d)$, resembling a staircase, with steep jumps between steps. The heights of these jumps are especially large in the region of high degeneracies. Similar structures may be observed in real systems; see, e.g., Fig. 3 of Ref. [20]. Inserting the asymptotics of $\tilde{d}(\ell)$, Eq. (15), into the expression for the degeneracy, Eq. (7), we see that outputs with the same number of ones have degeneracies exponentially close to $z_d^n / (z_d^4 - 2)^2$ if these ones are separated by many zeros and n is large. The slight deviations from this asymptotic value mean that the degeneracies are split among many points falling in a narrow range. For example, in the rightmost step, corresponding to outputs with exactly two ones, the number of these outputs is approximately $n^2/2$ because there are relatively few outputs in which the two ones are close together. This is the height of this jump.

These degeneracies are split into a set of about $n/2$ distinct values, corresponding to different intervals between the ones. Each one occurs in n outputs corresponding to the location of the same structure in different parts of the ring. Other jumps are produced by outputs with m strongly separated ones, or by outputs with, e.g., a pair of ones separated by one 0 with $m - 2$ ones far from each other and from that pair, and so on. This forms the rich staircaselike structure that we observe in Figs. 3(b), 3(d), and 3(e) and that is also reflected in Fig. 5. Note that any short filter of our sort will produce a similar complex structure in the degeneracy distribution, as the degeneracy associated to a chain of zeros (in output) of a given length must asymptotically grow exponentially with the length of the chain, as in Eq. (20).

We list the key numbers defining the asymptotics of a few example cases in Table I, including an example of a two-dimensional input, and give the full degeneracy distributions for selected cases in Fig. 2. It is worth discussing the meaning of these results and their interpretation in terms of sampling of complex systems. In particular, let us consider the results for the family of filters composed of a string of ones with a zero at each end, 010, 0110, 01110, etc., indicated in boldface font in Table I. We can consider the length of these filters as a crude control parameter of our sampling. Intuitively, we expect shorter filters to be more informative. Resolution, defined as the entropy of a sample,

$$H[y] = -\frac{1}{N} \sum_{i=1}^N \log\left(\frac{d_i}{N}\right) = -\sum_d \frac{d\mathcal{N}(d, n)}{N} \log\left(\frac{d}{N}\right), \quad (29)$$

is a measure of the ability to distinguish, at the output, between different input states. If the sample contains M distinct outputs, all with the same degeneracy, the resolution is simply $\ln M$; see Eq. (29). Notice that this is the case when all outputs are different (all degeneracies equal to 1, $M = N$), and also when all outputs are the same (degeneracy equal to N , $M = 1$). The inverse of the probability of a particular output gives the expected number of distinct outputs, if we assume the probabilities to be uniform. In this sense, we can regard the resolution as the expectation of the logarithm of the number of distinct states in a sample, based on the observation of one random state. Shorter filter patterns correspond to higher resolution. The resolution attains its maximum value for the filter pattern 0 or 1, either of which results in a unique output for each input (in the case of 1, the output is identical to the input). However, in this case all outputs are distinct, and so these filters are not informative about the system being sampled. As shown in Ref. [12], the measure that indicates the informativeness of a sample is the entropy of the degeneracy distribution,

$$H[d] = -\sum_d \frac{d\mathcal{N}(d, n)}{N} \log\left(\frac{d\mathcal{N}(d, n)}{N}\right), \quad (30)$$

where $N = 2^n$ for the complete input set. This is called the relevance. Comparing Eqs. (29) and (30) shows that the resolution is never smaller than the relevance, $H[y] \geq H[d]$. As can be seen in Table I, the relevance is greater for shorter filters, but is actually zero for the shortest possible filters 0 and 1. (In this extreme, $z_d = 1$, while $z_g = z_a = 2$.) We measured the relevance for a variety of short filter patterns. Specifically, we find that the entropy $H[d]$ is maximal for the filter pattern extracting single ones, on which we focus in this work, and so this pattern provides the most informative representations of the inputs

The number z_d gives the asymptotics of the largest degeneracies. It quickly approaches 2 as the filter pattern length increases. The largest degeneracies $\cong z_d^n$. Since $N = 2^n$, this means that almost all outputs concentrate in a few outputs, and in the limit, in a single state, i.e., all outputs are the same and the filter patterns are not informative. For the shortest few patterns, the value of z_d quickly falls, accompanied by a rapidly increasing relevance, indicating a transition to informative sampling. On the contrary, z_g , which gives the total number of outputs $M(n)$, increases with decreasing filter length, as shorter filters have more possible outputs. Taken together, these results indicate that the maximally informative sampling for a given family of filters is the shortest pattern having length greater than 1. This behavior is analogous to the transition observed in more complex problems (see, e.g., Ref. [10]), reinforcing the relevance of our tractable model to the study of more complex problems.

Importantly, the sets of outputs obtained with the larger filters are subsets of the set of the outputs obtained with our reference filter pattern 010. For this pattern, the value z_g , which gives the asymptotics of the number of different outputs, is maximal, and the number of different outputs is the largest one. (Notice that the pattern 000 in Table I has the same z_g as the reference pattern but a lower entropy $H[d]$.) Some filters have the property that two occurrences of the pattern cannot partially overlap. All such filter patterns of a given length give an identical degeneracy distribution, as we show in Table I. We suggest that the filtering problem with such nonoverlapping filter patterns of an arbitrary finite length can be solved in a way similar to our reference pattern.

In Appendix A 2, we derive the chain of linear coupled recursion relations for the number of outputs of degeneracy d , $\mathcal{N}(d, n)$; see Eqs. (A12) and (A13). These recursions generate exact $\mathcal{N}(d, n)$ for finite d and n , and in this sense provide the exact full spectrum of degeneracies. [There recursions also provide us with the explicit leading large n asymptotics of $\mathcal{N}(d, n)$, Eq. (A14), for an arbitrary d .] In particular, Eq. (A3) with the initial conditions, Eq. (A5), provides $\mathcal{N}(1, n)$, which is the number of configurations

having only groups of zeros of length 1, 2, and 3 between single ones. We find the large n asymptotics for this number,

$$\mathcal{N}(1, n) \cong z_a^n, \quad (31)$$

where

$$z_a = 1.46557... \quad (32)$$

is the real root of the characteristic equation $z^4 = z^2 + z + 1$.

Note that there are three key constants in this problem, z_g , z_d , and z_a , which appear in the main asymptotics: $M(n) \cong z_g^n$, $d_p(n) \cong z_d^n$, and $\mathcal{N}(1, n) \cong z_a^n$. Indeed, z_g enters the distribution $\mathcal{N}(d, n)/M(n)$ after normalization, z_d enters the asymptotics for degeneracies, see, e.g., Eq. (15) for $\tilde{d}(\ell)$ and Eq. (28) for $d_i(n)$, and z_a enters the asymptotics for $\mathcal{N}(d, n)$, Eqs. (A10) and (A14). In fact, all moments of $\mathcal{N}(d, n)$ exponentially diverge with n , $\sum_{i=1}^D d_i^k \mathcal{N}(d_i, n) \cong c_k^n$, where the numbers c_k and their large k asymptotics are presented in Appendix B. Clearly, for the first moment we have $\sum_{i=1}^D d_i \mathcal{N}(d_i, n) = 2^n$.

Using Eq. (21) we obtain the full set of degeneracies $\mathcal{D}_{\text{full}}(n)$ for different n , shown in Fig. 5, and the series of total number of discrete values of degeneracy $D(n)$ versus input size, represented in Fig. 6; see also the Supplemental Material [19]. As is natural, $D(n)$ is smaller than the number $p(n)$ of integer partitions of n . Figure 6 demonstrates that $D(n)$ is close to $p(n)/n$ for large n . The well-known large n asymptotics, $p(n) \cong (1/4\sqrt{3}n)e^{\pi\sqrt{2n/3}}$ [28–30], enables us to estimate $D(n) \sim e^{\pi\sqrt{2n/3}}$.

By ranking the full set of degeneracies $\mathcal{D}_{\text{full}}(n)$ for $n = 120$ we arrive at Fig. 7, where the main plot presents the

FIG. 6. Total number of different degeneracies D versus n (symbols). The curve is the number of integer partitions of n divided by n .

FIG. 7. Number of distinct degeneracies less than or equal to d versus d . Inset: Number of different degeneracies higher than or equal to d versus d . The orange square symbols are full results for $n = 120$. The circular blue symbols are values calculated in the limit $n \rightarrow \infty$ as described in Sec. VI. Solid black curve is the asymptotic approximation Eq. (33).

number of different degeneracies lower than or equal to d versus d , increasing roughly as $\exp[\pi\sqrt{2\ln d/3\ln z_d}]$ in the region $1 \ll d \ll d_D$, and the inset shows the number of different degeneracies higher than or equal to d versus d . The latter demonstrates a staircaselike structure similar to that of $\mathcal{N}_{\text{cum}}(d)$.

In Fig. 7 we also plot the number of different degeneracies lower than or equal to d in the infinite system, i.e., the rank of each degeneracy d . To obtain the full set of different degeneracies for $n \rightarrow \infty$ up to some d_{max} , we generate the set of prime degeneracies $\tilde{d} \leq d_{\text{max}}$. We start by listing all powers $m \geq 1$ of the first prime degeneracy $\tilde{d}$ while $\tilde{d}^m \leq d_{\text{max}}$. Then, we multiply each member of that list by increasing powers of the next prime degeneracy, while the product stays lower than or equal to d_{max} , and so on, with all the remaining prime degeneracies $\tilde{d} \leq d_{\text{max}}$. This procedure will result in duplicate degeneracies that should be removed, particularly for those values of d that have nonunique multiplicative partitions in terms of the prime degeneracies.

For example, the two smallest (larger than 1) prime degeneracies are $\tilde{d}(4) = 2$ and $\tilde{d}(5) = 4$, so a multiplicative partition of d with at least a part equal to $\tilde{d}(5) = 4$ is not unique because there is at least another partition of d where the contribution of the 4 is given in terms of $\tilde{d}(4)$ as 2^2 . However, apart from the partitions with parts equal to 4, the nonunique partitions are very rare; see Appendix A 2. In fact, $\tilde{d}(4) = 2$ is the only prime degeneracy $\tilde{d}(\ell)$ that can be expressed as a product of lower $\tilde{d}$ for all $\ell \leq 5000$.

The rank of the degeneracies in the infinite system can be estimated with precision for large d , as shown by the black

line in Fig. 7. For large ℓ the logarithms of prime degeneracies are uniformly distributed $\ln \tilde{d}(\ell) \approx \ell \ln z_d$. Assuming that all degeneracies d have a unique factorization in terms of the prime degeneracies $\tilde{d}$, the expected number of different degeneracies smaller than $d \sim z_d^m$ would be $\text{rank } d \approx \sum_{n \leq m} p(n)$, where $p(n)$ is the number of integer partitions of n . The vast majority of the degeneracies for which the assumption does not hold are the values of d which are multiples of 4, because that factor 4 can also be expressed as 2^2 . We therefore remove from each term of the previous sum the number of integer partitions that have at least one factor equal to 2, which is given by $p(n-2)$, and write for large d :

$$\begin{aligned} \text{rank } d &\approx \sum_{n \leq \ln d / \ln z_d} [p(n) - p(n-2)] \\ &\approx p(\ln d / \ln z_d) + p(\ln d / \ln z_d - 1) \\ &\approx \frac{\exp(\pi\sqrt{2/3}\sqrt{\ln d / \ln z_d})}{\pi 2\sqrt{3}(\ln d / \ln z_d)} \\ &\quad \times \left[1 - \left(\frac{13\pi}{72} + \frac{1}{\pi} \right) \sqrt{\frac{3 \ln z_d}{2 \ln d}} \right], \end{aligned} \quad (33)$$

where we have used the two leading terms of the asymptotic expansion of the number of integer partitions [28]:

$$p(n) \approx \frac{\exp(\pi\sqrt{2/3}\sqrt{n})}{4\sqrt{3}n} \left[1 - \left(\frac{\pi}{72} + \frac{1}{\pi} \right) \sqrt{\frac{3}{2n}} \right]. \quad (34)$$

Figure 7 demonstrates that $\text{rank } d$ as a function of d is well described by the estimate Eq. (33).

VII. MEAN-FIELD THEORY

Here we develop a mean-field theory enabling us to describe the cumulative distribution of degeneracies $\mathcal{N}_{\text{cum}}(d)$ in the region of large d where there are few ones in the outputs, and so the gaps of zeros between them are typically large. In this situation one can assume that the ones exist in a sea (or a mean field) of zeros, far from each other, so that the degeneracy of an output is completely determined by the number of its ones (and the output size n).

This ansatz is based on the observation that the three terms on the right-hand side of Eq. (12) behave very differently for increasing ℓ : The first term grows exponentially as z_d^ℓ , where $z_d \equiv z_1 = 1.754877\dots$ is real. The combined contribution of the other two terms is also real, since $z_2 = z_3^*$ (here $*$ denotes the complex conjugate). It decays exponentially because $|z_2| = |z_3| = 0.754877\dots < 1$. Thus, for increasing ℓ , the deviation of the asymptotics

$$\tilde{d}(\ell) \cong C_1 z_d^\ell \quad (35)$$

from the exact value of $\tilde{d}(\ell)$ approaches 0 exponentially rapidly, making this approximation excellent for large ℓ . Consider an output with m ones, separated by m strings of zeros of length $\ell_1, \ell_2, \dots, \ell_k$. If all $\ell_i \gg 1$, the degeneracy $\tilde{d}(m)$ of this output with m ones is accurately given by the asymptotic expression,

$$\tilde{d}(m) = \prod_{i=1}^m C_1 z_d^{\ell_i} = C_1^m z_d^{n-m}, \quad (36)$$

where we used the condition $\sum_{i=1}^m \ell_i = n - m$ that the number of ones plus the number of zeros must be equal to the total number of digits n .

The total number of different outputs with m ones is

$$\tilde{N}(m) = \frac{n}{n-m} \binom{n-m}{m}; \quad (37)$$

see Eq. (A1) in Appendix A 1. Notice that $\tilde{N}(m)$ coincides with the m th term of the sum in Eq. (4), as it must.

One should stress that although we derived the expressions for $\tilde{d}(m)$ and $\tilde{N}(m)$ for $m > 0$, these expressions are equally valid in the special case of $m = 0$. Indeed, Eq. (37) gives $\tilde{N}(0) = 1$, which is exact. Further, Eq. (36) gives $\tilde{d}(0) \cong z_d^n$ for large n coinciding with $d_D \cong z_d^n$, Eq. (20). Thus, Eqs. (36) and (37) work very well also for $m = 0$. For $m = 1$ we find a similar situation in terms of accuracy. In this case there is a single string of zeros of size $n - 1$, then every output with a single 1 has degeneracy equal to $\tilde{d}(n - 1)$. The deviation between $\tilde{d}(1) = C_1 z_1^{n-1}$ and $\tilde{d}(n - 1)$ decays exponentially with n : For each value of $m > 1$ there is actually more than one possible value for the degeneracy, that depends on the particular distribution of lengths of the strings of zeros. Nonetheless, for large n and $m \ll n$ most of the outputs contain only large strings of zeros. Therefore the set of points $(\tilde{d}, \tilde{N}_{\text{cum}})$ for $k = 0, 1, 2, \dots$ describes the exact cumulative distribution $\mathcal{N}_{\text{cum}}(d)$ well at the points $d = \tilde{d}(m)$ in the region of large d , which corresponds to $m \ll n$. There are jumps in the cumulative distribution at these points when n is large, and $\tilde{N}_{\text{cum}}(m)$ approaches the top points of these jumps.

The resulting mean-field theory expression for the distribution takes the following form:

$$\tilde{N}(d, n) = \delta(d, d_D(n)) + \sum_{m=1}^{\lfloor n/2 \rfloor} \frac{n}{n-m} \binom{n-m}{m} \delta\left(d, \frac{z_d^n}{(z_d^4 - 2)^m}\right). \quad (38)$$

This approximation is compared with the exact cumulative distribution in Fig. 3. The approximation can also be obtained from the exact result for $\mathcal{N}(d, n)$, Eq. (26), by approximating the weighted sum $\sum_i$ of Kronecker symbols in the expression for $\mathcal{N}(d, n)$ by a single Kronecker symbol

with a factor. One may also obtain the asymptotics of this distribution, as shown in Appendix C.

VIII. DEGENERACY DISTRIBUTIONS OF OUTPUTS FOR RANDOMLY GENERATED INPUT DATASETS

We observed in the previous section that the outputs generated by complete input datasets do not produce real power laws. One should note, however, that in empirical studies, input datasets, typically, are not complete. Usually, when inputs (input size n) are sufficiently large, the input dataset sizes N are much smaller than those of complete datasets, $N \ll C^n$, $C > 1$. Based on the sets of outputs of complete datasets, which we obtained in the previous section (listed in the Supplemental Material [19]), here we find and explore the distribution of outputs of randomly generated datasets of various sizes. (In principle, these sizes can also be bigger than or equal to that of complete datasets.)

Let the size of a randomly generated input dataset be N ; i.e., we apply filtering to N randomly generated rings of n zeros and ones. We assume that all degeneracies of the outputs of the corresponding complete input dataset are known, namely the full set of pairs $\{d_i, \mathcal{N}(d_i)\}$, where $i = 1, 2, \dots, D$, D is the total number of degeneracies for this n in the case of a complete input dataset. Then for the randomly generated input dataset, we obtain the following expected number of outputs of degeneracy d :

$$\langle \mathcal{N} \rangle(d) = \binom{N}{d} \sum_{i=1}^D \mathcal{N}(d_i) \left(\frac{d_i}{2^n}\right)^d \left(1 - \frac{d_i}{2^n}\right)^{N-d}. \quad (39)$$

The probability that a randomly generated input string produces a given output is simply $d_i/2^n$, where d_i is the degeneracy of the output with respect to the total input set. The probability that d of the N inputs produce this output is then simply given by a binomial expression. Summing over all total degeneracies (multiplying by the number of instances of a given degeneracy d_i and summing over i) gives the above expression. Here outputs of any degeneracy within the interval $1 \leq d \leq d_D$ are present with nonzero probability, in contrast to the case of the complete input dataset. The results of the application of this formula coincide with those obtained by recording statistics of outputs directly filtered from randomly generated inputs.

Let us apply Eq. (39), for instance, to the cases $n = 21$ [see Fig. 3(a) above for the complete input dataset consisting of all 2^{21} configurations] and $n = 120$, and inspect the distributions of outputs of uniformly randomly generated input sets of different sizes. The results are shown in Fig. 8. For each size of the random input dataset we present the degeneracy distribution and its cumulative counterpart. Interestingly, the form of these distributions is reminiscent of samples from real systems; see, e.g., Fig. 5 of Ref. [20].

FIG. 8. Double logarithmic scale plots of (a),(c),(e),(g) the degeneracy distributions and (b),(d),(f),(h) the cumulative degeneracy distributions obtained for randomly generated input datasets of different sizes N for $n = 21$ and $n = 120$. The specific sizes of input datasets for $n = 120$ are chosen to produce distributions similar to those for $n = 21$.

These figures demonstrate that for $N \ll 2^n$, the distributions indeed resemble a power law. The reason is that for sufficiently small N , the distribution does not approach the large values of d for which the variation of the exponent $c \ln^{\alpha-1} d$ in $\mathcal{N}_{\text{cum}}(d) \propto d^{-c \ln^{\alpha-1} d}$ [see Eq. (27)] becomes noticeable. As N approaches 2^n , the distributions become closer to their counterparts for the complete input dataset. Clearly, the distributions obtained in the limit $N \rightarrow \infty$ will coincide with those found for the complete input dataset.

Curiously, one may obtain distributions with a very similar form for different values of n by choosing N in order to maintain the scaling variable,

$$s = (z_d/2)^n N, \quad (40)$$

constant. For example, compare the first with the third row and the second with the fourth in Fig. 8. This combination follows from the fact that the binomial product in Eq. (39), as a function of d , forms a narrow peak centered at $d_i N / 2^n$, which produces $(z_d/2)^n N$ for the largest degeneracy $d_D \cong z_d^n$. This scaling disappears in the region of small degeneracies. Note that for small N , Eq. (39) gives $\langle \mathcal{N} \rangle(1) \approx N$, since $\sum_i d_i \mathcal{N}(d_i) \cong 2^n$. In fact, this is true as long as $N \ll (4/c_2)^n$, where $\sum_i d_i^2 \mathcal{N}(d_i) = c_2^n$ with $c_2 = 3.139\dots$; see Appendix B.

Thus the vast majority of outputs have degeneracy 1 in this situation.

IX. DISCUSSION AND CONCLUSIONS

Our straightforward, purely combinatorial treatment reveals features of distributions of outputs hidden from other approaches. For complete input datasets passed through our filter, we have obtained degeneracy distributions markedly distinct from power laws. On the other hand, these distributions decay as $\mathcal{N}_{\text{cum}}(d) \propto e^{-c \ln^\alpha d}$, $\alpha > 1$, much slower than exponentially, and in this sense they can still be called “critical.” We have observed that the entire form of these output distributions essentially depends on the input size n , which strongly differs, for example, from heavy tailed degree distributions of complex networks having exponential cutoffs [31,32].

For randomly generated input datasets, we found degeneracy distributions which could easily be taken for power laws in empirical studies, if the dataset size N is essentially smaller than 2^n . As $N \rightarrow \infty$, these distributions approach the clearly non-power-law shape of the distributions for the complete input dataset. Thus we show that the size of an input dataset matters for these problems.

Our model filter can be used as a convenient reference filtering problem. Although apparently rather different in construction, we argue that our problem can serve as a tractable representative model for sampling of complex systems. We focused on simple input sets which were uniformly random strings of zeros and ones. Correlated inputs are more challenging for analytical treatment. Exploring the simplest filter patterns, we showed that the statistics of outputs is determined not by the form of filter patterns but rather by what occurs in the gaps of zeros between them. The degeneracy corresponding to each such gap can be found using recursion relationships. We then used an integer partitions apparatus to aggregate the statistics of prime degeneracies from these gaps, finding the exact full spectrum of output degeneracies.

Considering the permutations of the integer partitions allows us to calculate the resulting exact degeneracy distribution $\mathcal{N}(d, n)$. Alternatively, using multiplicative partitions we derive coupled linear recursions providing the exact $\mathcal{N}(d, n)$ for any finite d and n and the explicit large n asymptotics. Finally, we developed a mean-field theory which describes the approximate degeneracy spectrum, degeneracy distribution, and their asymptotic behavior. Our mean-field theory of Sec. V also derives from the gaps between outputs.

These results show that in filter problems of this kind, the statistics of outputs is determined by the gaps between outputs, which are essentially determined by the size of a filter pattern. Longer filter patterns than ours, Eq. (1), can be subjected to a similar analysis; see Fig. 2 and Table I. Longer filter patterns, which give less information about

the input string, indeed produce less informative degeneracy distributions (i.e., having lower entropy).

For the sake of simplicity, we studied inputs containing only zeros and ones. We expect that our approach and the mean-field theory should be applicable to more rich inputs containing more degrees of freedom: larger sets of numbers, vectors, etc., as well as to more general cooperative systems with a large number of local minima. The next natural step after the mean-field theory, namely a fluctuation theory, should be based on accounting for small gaps between ones and the fluctuations of the lengths of these gaps.

In summary, we have presented a representative filtering problem for which we can obtain exact and complete results for the degeneracy distributions. Our methods may be generalized to other filtering and compression problems involving more complex filter patterns and complex, not necessary synthetic, higher-dimensional inputs. We believe the qualitative insights gained from studying this system can be applied to information processing problems in general.

ACKNOWLEDGMENTS

This work was developed within the scope of the project i3N, UIDB/50025/2020 and UIDP/50025/2020, financed by national funds through the FCT/MEC. This work was also supported by National Funds through FCT, I. P. Project No. IF/00726/2015. R. A. d. C. acknowledges the FCT Grants No. SFRH/BPD/123077/2016 and No. CEECIND/04697/2017.

APPENDIX A: RECURSIVE RELATIONS

Here we present the derivations of the expression for the number of different values of degeneracy of outputs produced by complete input datasets, Eq. (4), and the recursions for $\mathcal{N}(d, n)$ used in Sec. IV.

1. Derivation of Eq. (4) for $M(n)$

The total number $M(n)$ of different outputs in our problem is the number of all possible periodic (period n) combinations of zeros and ones having no neighboring ones. Clearly, this number coincides with the total number of different combinations of dimers in a ring of length n .

To find the number of different outputs with k ones, that is, the number of ways of placing k dimers in the ring, it is convenient to consider the cases when the first output digit is 1 and when it is 0 separately. This is equivalent to fixing the state of two chosen neighboring units: either there is a dimer connecting this pair or this dimer is absent.

- (i) When the first digit of an output is 1, both the second and the last digits must be 0, and the number of outputs in this case is given by the binomial coefficient $\binom{n-k-1}{k-1}$, which corresponds to starting with the sequence 1, $\{0\}_{n-k}$ (where $\{0\}_{n-m}$ is a sequence of $n-m$ zeros), and choosing $k-1$ out of

$n - k - 1$ distinct zeros to be replaced by 01. Because of the periodic boundary condition, the last 0 cannot be selected for replacement; hence the number of zeros that can be replaced is $n - k - 1$.

- (ii) When the first digit of an output is 0, we start from the sequence $\{0\}_{n-k}$ and replace k zeros by 01. In this case the replacement may be made at any of the $n - k$ zeros, and the number of outputs is $\binom{n-k}{k}$.

The total number of different outputs is obtained by summing these two contributions,

$$\bar{\mathcal{N}}(k) = \binom{n-k-1}{k-1} + \binom{n-k}{k} = \frac{n}{n-k} \binom{n-k}{k}, \quad (\text{A1})$$

where we used the notation $\bar{\mathcal{N}}(k)$ from Sec. VII.

Summing over all possible values of k gives

$$\begin{aligned} M(n) &= 1 + \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\frac{(n-k-1)!}{(k-1)!(n-2k)!} + \frac{(n-k)!}{k!(n-2k)!} \right] \\ &= n \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{1}{n-k} \binom{n-k}{k}, \end{aligned} \quad (\text{A2})$$

i.e., Eq. (4).

2. Recursions for $\mathcal{N}(d, n)$

Chains of exactly three zeros in the output have no degrees of freedom in the input; they can only be produced in one way. Thus, outputs containing no chains of zeros of length greater than three have degeneracy 1. For the input size n , the number of such outputs $\mathcal{N}(d, n)$ can be obtained recursively. Here we derive the recursive relation for $\mathcal{N}(1, n)$, $\mathcal{N}(2, n)$, and $\mathcal{N}(4, n)$, and indicate the recursions for higher degeneracy d . See also Supplemental Material [19].

Sequences of degeneracy 1 can be regarded as assortments of three kinds of building blocks, 01, 001, and 0001, put together in a ring of length n . All configurations of blocks are allowed, as long as the total number of binary digits is n . Let us consider a particular position i in the ring and the block to which i belongs. If we add a block 01 between the block of i and one of its neighbor blocks (say, the one to the right) to every possible configuration of length $n - 2$, we get every possible configuration of length n that has a block 01 to the right of the block of i . Doing the same with configurations of length $n - 3$ and blocks 001, we get all configurations with a block 001 to the right of the block of i . Finally, repeating the procedure for configurations of length $n - 4$ and blocks 0001 gives all configurations with a block 0001 to the right of the block of i . Since every block must be 01, 001, or 0001, the union of these three sets is the full set of configurations of degeneracy 1 in a system of n digits. Thus, for the number of configurations in this set, $\mathcal{N}(1, n)$, we can write

$$\mathcal{N}(1, n) = \mathcal{N}(1, n-2) + \mathcal{N}(1, n-3) + \mathcal{N}(1, n-4). \quad (\text{A3})$$

The explicit solution of this linear difference equation is given in terms of the roots, z_1, z_2, z_3 , and z_4 , of the characteristic equation $z^4 = z^2 + z + 1$:

$$\mathcal{N}(1, n) = z_1^n + z_2^n + z_3^n + z_4^n, \quad (\text{A4})$$

where the coefficients of the powers of the roots z_i are found through the initial condition,

$$\mathcal{N}(1, 1) = 0, \quad \mathcal{N}(1, 2) = 2, \quad \mathcal{N}(1, 3) = 3, \quad \mathcal{N}(1, 4) = 6, \quad (\text{A5})$$

and are equal to 1. The root $z_1 \equiv z_a = 1.46557\dots$ determines the large n asymptotics of $\mathcal{N}(1, n)$, Eq. (31).

We can expand the analysis to higher values of degeneracy. For example, the second smallest value, $d_2 = 2$, corresponds to the outputs with blocks of the types 01, 001, and 0001 and a single string 00001. Applying the same reasoning as above, we get three terms similar to the ones of Eq. (A3). We must, additionally, consider configurations with degeneracy 1, to which we add a string 00001 in the same way, instead of one of the other blocks. Then we get

$$\begin{aligned} \mathcal{N}(2, n) &= \mathcal{N}(2, n-2) + \mathcal{N}(2, n-3) \\ &\quad + \mathcal{N}(2, n-4) + \mathcal{N}(1, n-5), \end{aligned} \quad (\text{A6})$$

with the initial condition

$$\begin{aligned} \mathcal{N}(2, 1) &= \mathcal{N}(2, 2) = \mathcal{N}(2, 3) = \mathcal{N}(2, 4) = 0, \\ \mathcal{N}(2, 5) &= 5. \end{aligned} \quad (\text{A7})$$

In a similar way, taking into account that the prime degeneracy $\tilde{d}(\ell)$ corresponds to the inserted block of $\ell + 1$ (chain of ℓ zeros and 1 on the right), i.e., 2, 4, 7, 12, 21, 37, 65, ... correspond to $\ell + 1 = 5, 6, 7, 8, 9, 10, 11, \dots$, respectively, we derive the recursion relations for higher $\mathcal{N}(d, n)$ [here, $d \neq d_D(n)$, for which, clearly, $\mathcal{N}(d_D(n), n) = 1$],

$$\begin{aligned} \mathcal{N}(4, n) &= \mathcal{N}(4, n-2) + \mathcal{N}(4, n-3) + \mathcal{N}(4, n-4) \\ &\quad + \mathcal{N}(2, n-5) + \mathcal{N}(1, n-6), \end{aligned} \quad (\text{A8})$$

with the initial condition

$$\mathcal{N}(4, 1) = \dots = \mathcal{N}(4, 5) = 0, \quad \mathcal{N}(4, 6) = 6, \quad (\text{A9})$$

and so on. See the list of all recursive relations for $d < 100$ with initial conditions in the Supplemental Material [19]. The corresponding large n asymptotics are

$$\begin{aligned}
\mathcal{N}(1, n) &\cong z_a^n, \\
\mathcal{N}(2, n) &\cong \frac{1}{z_a(2z_a^2 + 3z_a + 4)} n z_a^n = 0.05376 n z_a^n, \\
\mathcal{N}(4, n) &\cong \frac{1}{2!} \left[\frac{1}{z_a(2z_a^2 + 3z_a + 4)} \right]^2 n^2 z_a^n = 0.001445 n^2 z_a^n,
\end{aligned} \tag{A10}$$

and so on. See the list of the resulting large n asymptotics of $\mathcal{N}(d < 100, n)$ in the Supplemental Material [19].

More generally, for any value of degeneracy d one can write the recursion relation for $\mathcal{N}(d, n)$ in terms of the multiplicative partition of d into prime degeneracies. Let us present the full set of recursion relations for $\mathcal{N}(d, n)$. The reasoning is as follows.

- (i) The recursion relation for $\mathcal{N}(1, n)$ is given by Eq. (A3) with the initial condition Eq. (A5).
- (ii) For the largest degeneracies, $\mathcal{N}(d_D(n), n) = 1$.
- (iii) All prime degeneracies $\tilde{d}(\ell)$ except $\tilde{d}(5) = 4$ are not multiplicatively separable into other prime degeneracies (primality property). The only exception is $\tilde{d}(5) = \tilde{d}^2(4) = 2^2$. See Sec. VI above for more detail.
- (iv) Any degeneracy in the spectrum except d_D (and $d = 1$) is multiplicatively separable into the prime degeneracies $\tilde{d}(\ell)$, namely,

$$\begin{aligned}
d &= \tilde{d}^{\mu_4}(4) \tilde{d}^{\mu_5}(5) \tilde{d}^{\mu_6}(6) \tilde{d}^{\mu_7}(7) \dots \\
&= 2^{\mu_4} 4^{\mu_5} 7^{\mu_6} 12^{\mu_7} \dots \\
&= 2^{\mu_4 + 2\mu_5} 7^{\mu_6} 12^{\mu_7} \dots,
\end{aligned} \tag{A11}$$

where the powers $\mu_\ell \equiv \mu(\ell)$ are non-negative integers. Because of the exception $\tilde{d}(5) = \tilde{d}^2(4) = 2^2$ and the coincidence $\tilde{d}(5)\tilde{d}(8) = 4 \times 21 = 7 \times 12 = \tilde{d}(6)\tilde{d}(7)$, etc., the multiplicative partition into the prime degeneracies $\tilde{d}(\ell)$, Eq. (A11), may not be unique; see below. With all possible integers $\mu(4) \geq 0, \mu(6) \geq 0, \mu(7) \geq 0, \dots$, we generate the full list $\mathcal{D}$ of all degeneracies except 1 and d_D in the spectrum for $n \rightarrow \infty$. We place the generated degeneracies in ascending order, ignoring repetitions for some degeneracies. For a finite n , only some on the degeneracies from this list are present in the spectrum. In practice, we need degeneracies d only up to some, maybe, large but finite value. So, generating this list, we use a finite set of non-negative integers $\mu_4, \mu_6, \mu_7, \dots$, respectively restricted from above.

- (v) For any degeneracy d in this list $\mathcal{D}$ we define the vector $\mathcal{L}(d) \equiv (\ell_1, \ell_2, \dots, \ell_{\max})$, indicating that the prime degeneracies $[\tilde{d}(\ell_1), \tilde{d}(\ell_2), \dots, \tilde{d}(\ell_{\max})]$, given in ascending order, are present in at least one of the partitions of d , Eq. (A11). That is, if a

prime degeneracy $\tilde{d}(\ell)$ is present in $\mathcal{L}(d)$, then the ratio $d/\tilde{d}(\ell)$ also belongs to $\mathcal{D}$. Then the recursion relation for $\mathcal{N}(d, n)$ has the form

$$\begin{aligned}
\mathcal{N}(d, n) &= \mathcal{N}(d, n-2) + \mathcal{N}(d, n-3) \\
&\quad + \mathcal{N}(d, n-4) \\
&\quad + \sum_{\ell \in \mathcal{L}(d)} \mathcal{N}(d/\tilde{d}(\ell), n-\ell-1),
\end{aligned} \tag{A12}$$

where ℓ in the sum takes the values $\ell_1, \ell_2, \dots, \ell_{\max}$, i.e., the components of the vector $\mathcal{L}(d)$. This formula sums the number of configurations $\mathcal{N}(d/\tilde{d}(\ell), n-\ell-1)$ of smaller systems to which a block of ℓ zeros (followed by a one) can be added in order to form a configuration of size n and degeneracy d . We should not explicitly include in this sum configurations that achieve degeneracy d by inserting more than one block of zeros. The insertion of additional blocks besides a block of length l , say, into smaller configurations is already accounted for in the calculation of the degeneracy of the configuration of size $n-\ell-1$ into which the block of length l will be inserted. Note that the ratio $d/\tilde{d}(\ell)$ in each of the terms in the sum is one of the degeneracies, smaller than d , in the list $\mathcal{D}$ with the degeneracy $d=1$ added, so this set of recursions together with the recursion for $\mathcal{N}(1, n)$, Eq. (A3), is closed. The initial conditions for these recursions are

$$\begin{aligned}
\mathcal{N}(d, 1) &= \dots = \mathcal{N}(d, \ell_{\max}) = 0, \\
\mathcal{N}(d, \ell_{\max} + 1) &= (\ell_{\max} + 1) \delta(d, \tilde{d}(\ell_{\max})),
\end{aligned} \tag{A13}$$

where $\delta(i, j)$ is the Kronecker symbol. Note that when d is a prime degeneracy, $d = \tilde{d}(\ell) = \tilde{d}(\ell_{\max})$ except $d = 4 = \tilde{d}(5)$, the sum on the left-hand side of Eq. (A12) has only one term, $\mathcal{N}(1, n-\ell_{\max}-1)$. If $d = 4$, Eqs. (A12) and (A13) properly give the recursion from Eq. (A8) with the initial condition from Eq. (A9). Finally, for $d = 1$, use Eq. (A3) with the initial condition Eq. (A5). [In fact, if $d = 1$, the list $\mathcal{L}(d = 1)$ is empty, and Eq. (A12) is reduced to Eq. (A3). We still need the initial condition, Eq. (A5), for $d = 1$, since Eq. (A13) is not defined in this case.]

- (vi) Equations (A3), (A5), (A12), and (A13), in particular, provide all degeneracies present in the spectrum for a given n . If some d is absent, these recursions produce $\mathcal{N}(d, n) = 0$. For instance, let $d = 64$. The initial condition is $\mathcal{N}(64, 1) = \dots = \mathcal{N}(64, 6) = 0$. The recursions produce $\mathcal{N}(64, 7) = \dots = \mathcal{N}(64, 17) = 0$, $\mathcal{N}(64, 18) = 6$, $\mathcal{N}(64, 19) = 0$, $\mathcal{N}(64, 20) = 20$, and so on.

- (vii) Our recursions, Eq. (A12), with their initial conditions, Eq. (A13), lead to the following large n asymptotics of $\mathcal{N}(d, n)$:

$$\mathcal{N}(d, n) \cong \frac{1}{\prod_{\ell \neq 5} \mu_\ell!} \times \left[\frac{n}{z_d(2z_d^2 + 3z_d + 4)} \right]^{\sum_{\ell \neq 5} \mu_\ell} \frac{n - \sum_{\ell \geq 6} (\ell-4)\mu_\ell}{z_d}, \quad (\text{A14})$$

for d with the multiplicative partition into prime degeneracies, Eq. (A12), chosen in such a way that the power of 2, μ_4 , in this partition is maximal.

Let us discuss this point in more detail. We stressed above that a multiplicative partition of $d \in \mathcal{D}$, Eq. (A12), may not be unique. For large n , the contribution of each of the possible multiplicative partitions of d to $\mathcal{N}(d, n)$

is about $n^{\sum_{\ell} \mu_\ell} z_d^{n - \sum_{\ell} (\ell-4)\mu_\ell}$. The leading asymptotics, Eq. (A14), is with the maximal power of n ; i.e., it originates from the partition with the maximal sum $\sum_{\ell} \mu_\ell$. Let us check whether the multiplicative partition of d with the maximal sum $\sum_{\ell} \mu_\ell$ is unique and that this maximum corresponds to the maximal μ_4 . In other words, if there are a number of different multiplicative partitions of d , whether only one partition has the maximal μ_4 , and at the same time this partition has the maximal sum $\sum_{\ell} \mu_\ell$.

To explain why the property of primality is very likely to hold for all large enough $\tilde{d}(\ell)$, let us consider the number of conventional prime numbers smaller than some number d , which grows as $\sim d / \ln d$. Since $\tilde{d}(\ell) \sim z_d^\ell$, the number of conventional primes smaller than $\tilde{d}(\ell)$ grows exponentially with ℓ as $\sim z_d^\ell / \ell$. Additionally, the average number of conventional primes in the factorization of numbers of the magnitude of d grows as $\sim \ln \ln d$ [33]. A necessary condition for some $\tilde{d}(\ell)$ to not be prime is that all of its prime factors are also factors of at least another $\tilde{d} < \tilde{d}(\ell)$. The combination of the exponential increase of conventional primes smaller than $\tilde{d}(\ell)$, and the increase of the number of factors of $\tilde{d}(\ell)$ as ℓ increases, makes the probability of the degeneracy $\tilde{d}(\ell)$ not being prime approach 0 very rapidly. In the set of values $\tilde{d}(\ell \leq 200)$, the only ones that do not contain at least one conventional prime factor absent from all factorizations of smaller $\tilde{d}$ are $\tilde{d}(5) = 4$, $\tilde{d}(8) = 21$, $\tilde{d}(12) = 200$, $\tilde{d}(13) = 351$, and $\tilde{d}(24) = 170625$, and from these only $\tilde{d}(5)$ is actually not a prime degeneracy.

We inspected all products of prime degeneracies $\tilde{d}(\ell \leq 200) \approx 1.7 \times 10^{48}$ focusing on the products producing nonunique multiplicative partitions. Apart from $2^2 = 4$ discussed above, we find only three such combinations with $\ell \leq 200$ (into this number, we do not include the products of these combinations and arbitrary prime

degeneracies). Namely, $2^2 \times 21 = 7 \times 12$, $2^9 \times 200 \times 351^2 = 12^6 \times 65^2$, and $2^2 \times 12^2 \times 170625 = 7 \times 200^2 \times 351$. For each of these combinations we confirm that the left-hand side of the equality corresponds to the maximal $\sum_{\ell} \mu_\ell$ and that this partition is unique. Clearly, the same is true for the products of these combinations and arbitrary prime degeneracies. Thus, partitions contributing to the leading asymptotics of $\mathcal{N}(d, n)$ are unique, and the gauge is fixed by demanding $\mu_4 = \max$, which leads to Eq. (A14). For grasping the form of Eq. (A14), see also Eqs. (A10), especially the asymptotics of $\mathcal{N}(84, n)$.

Thus we have the chain of conveniently coupled linear recursion relations that one can easily process starting from $d = 1$. These recursions generate exact $\mathcal{N}(d, n)$ for finite d and n , and in this sense provide the exact solution of the problem. Moreover, the leading large n asymptotics of $\mathcal{N}(d, n)$ are found explicitly. Note that the time of computing all $\mathcal{N}(d < d_0, n)$, where d_0 is fixed, by using our recursions, is proportional to n ; i.e., the computation for each next size takes the same time.

APPENDIX B: MOMENTS OF $\mathcal{N}(d, n)$

Here we list the large n asymptotics of the moments of $\mathcal{N}(d, n)$. The leading asymptotics of the moments are of the following form:

$$\mathcal{M}_k(n) \equiv \sum_{i=1}^D d_i^k \mathcal{N}(d_i, n) \cong c_k^n, \quad (\text{B1})$$

where

$$\begin{aligned} c_0 &= z_g, \\ c_1 &= 2, \\ c_2 &= 3.138\,990\,099\,335\,42, \\ c_3 &= 5.417\,628\,641\,309\,76, \\ c_4 &= 9.486\,966\,311\,400\,60, \\ c_5 &= 16.643\,811\,967\,230\,8, \\ c_6 &= 29.206\,771\,707\,194\,2, \\ c_7 &= 51.254\,058\,304\,680\,6, \\ c_8 &= 89.944\,542\,935\,182\,3, \\ &\dots \end{aligned} \quad (\text{B2})$$

In their turn, the numbers c_k have the following large k asymptotics:

$$c_k \cong z_d^k [1 + (z_d^4 - 2)^{-k} + \dots], \quad (\text{B3})$$

which is close to the numerical values listed above, Eq. (B2), already for $k = 2$. The leading term z_d^k results from the largest degeneracy in the spectrum,

$\mathcal{N}(d_D \cong z_d^n, n) = 1$, the next term originates from the second largest degeneracy, etc.

APPENDIX C: ASYMPTOTIC MEAN-FIELD DEGENERACY DISTRIBUTION

Let us obtain an explicit formula for the asymptotics of the mean-field approximation to the degeneracy distribution $\bar{\mathcal{N}}$ in terms of the degeneracy d . We simply replace m in Eq. (37) by the inverse of the function $\bar{d}(m)$ in Eq. (36). The inverse function is $m(d) = (\ln d - n \ln z_d) / (\ln C_1 - \ln z_d)$ (where we have dropped the bar because d is now the independent variable), and substituting into Eq. (37) gives

$$\begin{aligned} \bar{\mathcal{N}} &= \frac{n}{n-m} \binom{n-m}{m} \approx \frac{n^m}{m!} \\ &= \frac{z_d^{n \ln n / \ln(z_d/C_1)}}{\Gamma[1 - \frac{\ln(d/z_d^n)}{\ln(z_d/C_1)}]} d^{-\ln n / \ln(z_d/C_1)} \quad (C1) \end{aligned}$$

for $m \ll n$. Note that the logarithmic derivative of $\bar{\mathcal{N}}$ over d at the point d_D (the highest degeneracy), $-\ln n / \ln(z_d^4 - 2)$, properly fits the two rightmost points of the spectrum, namely, $d_D(n) \cong z_d^n$, $\mathcal{N}(d_D) = 1$ and $d_{D-1}(n) \cong z_d^n / (z_d^4 - 2)$, $\mathcal{N}(d_{D-1}) = n$.

-
- [1] M. Borda, *Fundamentals in Information Theory and Coding* (Springer, Berlin, 2011).
  - [2] D. C. Hankerson, G. A. Harris, and P. D. Johnson, Jr., *Introduction to Information Theory and Data Compression* (CRC Press, New York, 2003).
  - [3] N. Tishby, F. C. Pereira, and W. Bialek, *The Information Bottleneck Method*, [arXiv:physics/0004057](https://arxiv.org/abs/physics/0004057).
  - [4] A. Gehani, T. LaBean, and J. Reif, *DNA-Based Cryptography*, in *Aspects of Molecular Computing* (Springer, Berlin, Heidelberg, 2003), p. 167.
  - [5] A. K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems* (Wiley-VCH, Weinheim, 2005).
  - [6] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications* (World Scientific, Singapore, 1987).
  - [7] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, *On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima*, [arXiv:1609.04836](https://arxiv.org/abs/1609.04836).
  - [8] S. K. Baek, S. Bernhardsson, and P. Minnhagen, *Zipf's Law Unzipped*, *New J. Phys.* **13**, 043004 (2011).
  - [9] J. Song, M. Marsili, and J. Jo, *Resolution and Relevance Trade-offs in Deep Learning*, [arXiv:1710.11324](https://arxiv.org/abs/1710.11324).
  - [10] R. Cubero, M. Marsili, and Y. Roudi, *Minimum Description Length Codes Are Critical*, *Entropy* **20**, 755 (2018).
  - [11] A. Haimovici and M. Marsili, *Criticality of Mostly Informative Samples: A Bayesian Model Selection Approach*, *J. Stat. Mech.* (2015) P10013.
  - [12] R. J. Cubero, J. Jo, M. Marsili, Y. Roudi, and J. Song, *Statistical Criticality arises in Most Informative Representations*, *J. Stat. Mech.* (2019) 063402.
  - [13] J. Song, M. Marsili, and J. Jo, *Resolution and Relevance Trade-Offs in Deep Learning*, *J. Stat. Mech.* (2018) P123406.
  - [14] G. Bianconi, *Entropy of Network Ensembles*, *Phys. Rev. E* **79**, 036114 (2009).
  - [15] G. Bianconi, *A Statistical Mechanics Approach for Scale-Free Networks and Finite-Scale Networks*, *Chaos* **17**, 026114 (2007).
  - [16] A. G. Bashkirov and A. V. Vityazev, *Information Entropy and Power-Law Distributions for Chaotic Systems*, *Physica (Amsterdam)* **277A**, 136 (2000).
  - [17] Y. Dover, *A Short Account of a Connection of Power Laws to the Information Entropy*, *Physica (Amsterdam)* **334A**, 591 (2004).
  - [18] M. Visser, *Zipf's Law, Power Laws and Maximum Entropy*, *New J. Phys.* **15**, 043021 (2013).
  - [19] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.10.011074> for exact results of our algorithm and recursion relations.
  - [20] M. Marsili, I. Mastromatteo, and Y. Roudi, *On Sampling and Modeling Complex Systems*, *J. Stat. Mech.* (2013) P09003.
  - [21] V. E. Hoggatt, Jr., *Fibonacci and Lucas Numbers* (Houghton Mifflin, Boston, 1969).
  - [22] R. L. Graham, D. E. Knuth, O. Patashnik, and S. Liu, *Concrete Mathematics: A Foundation for Computer Science* (Addison-Wesley Publishing Company, Reading, MA, 1994).
  - [23] T. Koshy, *Fibonacci and Lucas Numbers with Applications* (John Wiley & Sons, Inc., Hoboken, NJ, 2019).
  - [24] H. C. Austin and R. Guy, *Binary Sequences without Isolated Ones*, *Fibonacci Q.* **16**, 84 (1978).
  - [25] The On-Line Encyclopedia of Integer Sequences, <https://oeis.org>.
  - [26] G. E. Andrews, *The Theory of Partitions* (Cambridge University Press, Cambridge, England, 1998).
  - [27] G. E. Andrews and K. Eriksson, *Integer Partitions* (Cambridge University Press, Cambridge, England, 2004).
  - [28] G. H. Hardy and S. Ramanujan, *Asymptotic Formulae in Combinatory Analysis*, *Proc. London Math. Soc.* **s2-17**, 75 (1918).
  - [29] J. V. Uspensky, *Asymptotic Formulae for Numerical Functions which Occur in the Theory of Partitions*, *Bull. Acad. Sci. URSS* **14**, 199 (1920).
  - [30] H. S. Wilf, *Lectures on Integer Partitions*, <https://www.math.upenn.edu/~wilf/PIMS/PIMSLectures.pdf>.
  - [31] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *Power-Law Distributions in Empirical Data*, *SIAM Rev.* **51**, 661 (2009).
  - [32] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
  - [33] P. Erdős and M. Kac, *The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions*, *Am. J. Math.* **62**, 738 (1940).