



## On multidimensional record patterns

P. L. Krapivsky, J.M. Luck

### ► To cite this version:

P. L. Krapivsky, J.M. Luck. On multidimensional record patterns. Journal of Statistical Mechanics: Theory and Experiment, 2020, 2020 (6), pp.063205. 10.1088/1742-5468/ab7f30 . cea-02865360

**HAL Id: cea-02865360**

**<https://cea.hal.science/cea-02865360>**

Submitted on 11 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On multidimensional record patterns

P L Krapivsky<sup>1,2</sup> and J M Luck<sup>2</sup>

<sup>1</sup> Department of Physics, Boston University, Boston, MA 02215, USA

<sup>2</sup> Institut de Physique Théorique, Université Paris-Saclay, CEA and CNRS,  
91191 Gif-sur-Yvette, France

**Abstract.** Multidimensional record patterns are random sets of lattice points defined by means of a recursive stochastic construction. The patterns thus generated owe their richness to the fact that the construction is not based on a total order, except in one dimension, where usual records in sequences of independent random variables are recovered. We derive many exact results on the statistics of multidimensional record patterns on finite samples drawn on hypercubic lattices in any dimension  $D$ . The most detailed analysis concerns the two-dimensional situation, where we also investigate the distribution of the landing position of the record point which is closest to the origin. Asymptotic expressions for the full distribution and the moments of the number of records on large hypercubic samples are also obtained. The latter distribution is related to that of the largest of  $D$  standard Gaussian variables.

E-mail: [pkrapivsky@gmail.com](mailto:pkrapivsky@gmail.com), [jean-marc.luck@ipht.fr](mailto:jean-marc.luck@ipht.fr)

## 1. Introduction

In this work we introduce and investigate multidimensional record patterns. These are random sets of lattice points defined by means of a recursive construction, generalizing the construction of usual records, i.e., records in sequences of independent and identically distributed random variables. The statistics of records has a long history [1, 2, 3] and it has been the subject of many books and reviews [4, 5, 6, 7, 8, 9]. The recent review [10], as well as References [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21], provide a variety of examples of applications of record statistics to physical problems.

The multidimensional record patterns considered here are non-trivial extensions of records to an arbitrary higher dimension  $D$ . Their recursive construction is entirely parameter-free and respects all symmetries of the underlying hypercubic lattice. Their statistics therefore involve some universal combinatorics, generalizing that involved in the case of usual records. The present work was originally inspired by recent studies on the fragmentation of rectangles drawn on the square lattice [22, 23], where an initial rectangle of size  $L \times M$  may break either vertically or horizontally into two smaller rectangles. In a symmetric variant of the model it breaks simultaneously in both directions, giving birth to four rectangles. The process is repeated until it stops when the entire system is filled with frozen rectangles, called sticks, of size  $K \times 1$  or  $1 \times K$ . Several quantities of interest have been studied in [22, 23], including the mean number of sticks and the distribution of the stick length  $K$ . The outcomes provide interesting examples of the non-trivial scaling laws which are observed in fragmentation processes (see e.g. [24, 25, 26, 27, 28, 29, 30, 31, 32], and [33] for a review). The connection with the present work is as follows. If one monitors the successive breaking events of the symmetric fragmentation process which are closer and closer to the origin (i.e., to the lower left corner of the rectangular sample), one obtains a random set of points which identifies with a two-dimensional record pattern.

There is no direct connection between the multidimensional record patterns studied here and records in sequences of random variables, except in the one-dimensional situation, reviewed in section 2. First of all, there is no canonical notion of records in more than one dimension. Consider for definiteness a two-dimensional signal  $x_{l,m}$  for  $l, m = 1, 2, \dots$ . Along the lines of the usual theory of records, we say that there is a record at the point  $(l, m)$  if  $x_{l,m}$  is larger than  $x_{l',m'}$  for all  $(l', m') < (l, m)$ . The definition of records therefore relies on the choice of a total order among lattice points, i.e., couples of integers  $(l, m)$ . Natural requirements are as follows: the total order should respect the translation invariance of the underlying lattice, and be such that every point  $(l, m)$  has finitely many predecessors. There are uncountably many orders on the square lattice obeying these conditions (see e.g. [34]). Consider indeed a positive irrational number  $\omega$ , and define the height of  $(l, m)$  as  $h_{l,m} = l + m\omega$ . The order obtained by defining  $(l', m') < (l, m)$  as  $h_{l',m'} < h_{l,m}$  has all required properties. It is however not invariant under permutation of both axes. One has indeed  $(1, 2) < (2, 1)$  for  $\omega < 1$  and vice versa. This construction easily extends to higher dimensions by defining the height of a point as  $h_{l_1, \dots, l_D} = l_1\omega_1 + \dots + l_D\omega_D$ , where the numbers  $\omega_i$  are positive and linearly independent over the rationals. A second fact is equally important. Suppose we equip the square lattice with the total order associated with a given irrational slope  $\omega$ . The resulting records are just usual records, albeit with lattice points being ordered according to increasing values of the height  $h_{l,m}$ . As a consequence, the number  $N$  of records among the first  $L$  points has the same statistics as in the case of usual, i.e., one-dimensional, records. In brief, both

the novelty and the richness of the multidimensional record patterns studied here are due to the fact that they are not based upon a total order, at variance with usual records.

The contents of the present paper is as follows. In section 2 we present a self-contained reminder on the statistics of usual records and describe a backward recursive construction of those records for finite signals of length  $L$ . The multidimensional record patterns obtained by extending the latter recursive construction to higher-dimensional lattices are investigated in later sections. Section 3 is devoted to a detailed analysis of the two-dimensional case, which is related to the fragmentation process studied in [22, 23]. We obtain exact combinatorial results on the statistics of the number  $N$  of records on finite rectangular samples of size  $L \times M$ , as well as asymptotic results for large samples. We also investigate the distribution of the landing position  $K$  of the process, i.e., the length of the stick adjacent to the origin. The case of an arbitrary dimension  $D$  is addressed in section 4. The main emphasis is on the asymptotic behavior of the full distribution and of the moments of the number of record points on large hypercubic samples. Section 5 contains a brief summary of our main findings. Appendix A is devoted to the derivation of asymptotic large- $D$  expansions of various quantities.

## 2. The one-dimensional case: usual records

### 2.1. A reminder on record statistics

It is worth beginning with a self-contained reminder on the statistics of usual records [4, 5, 6, 7, 8, 9, 10]. Consider a signal  $x_n$  in discrete time ( $n = 1, 2, \dots$ ), modelled as a sequence of independent and identically distributed (iid) real random variables with continuous distribution  $\rho(x)$ . It is said that there is a record-breaking event at time  $n$ , or for short that time  $n$  is a record, if  $x_n$  is larger than all previous entries  $x_{n'}$  ( $n' = 1, \dots, n-1$ ). So,  $n = 1$  is always a record, whereas  $n = 2$  is a record with probability  $1/2$ , and so on. It is indeed well established that there is a record at time  $n$  with probability

$$p_n = \frac{1}{n}, \quad (2.1)$$

and that the occurrences of records at different times are statistically independent.

The key quantity of interest is the number of records,  $N = 1, \dots, L$ , on a finite signal of length  $L$ . The distribution  $P_L(N)$  of this random number is conveniently encoded in the generating function

$$G_L(z) = \sum_{N=1}^L P_L(N) z^N. \quad (2.2)$$

The independence of the occurrences of records at different times yields the product formula

$$G_L(z) = \prod_{n=1}^L (1 - p_n + z p_n) = \prod_{n=1}^L \frac{n-1+z}{n}, \quad (2.3)$$

i.e.,

$$G_L(z) = \frac{\Gamma(L+z)}{L! \Gamma(z)} = \frac{1}{L!} \sum_{N=1}^L \begin{bmatrix} L \\ N \end{bmatrix} z^N, \quad (2.4)$$

so that the distribution of  $N$  reads

$$P_L(N) = \frac{1}{L!} \left[ \begin{matrix} L \\ N \end{matrix} \right], \quad (2.5)$$

where the integers  $\left[ \begin{matrix} L \\ N \end{matrix} \right]$ , referred to as the Stirling numbers of the first kind [35], are ubiquitous in combinatorics (see e.g. [36, 37, 38]).

We introduce for further reference the complementary cumulative distribution of the number of records  $N$ , defined as the probability that the latter number exceeds a given  $n$ . This quantity reads

$$F_L(n) = \text{Prob}\{N > n\} = \sum_{m=n+1}^L P_L(m) = \frac{1}{L!} \sum_{m=n+1}^L \left[ \begin{matrix} L \\ m \end{matrix} \right]. \quad (2.6)$$

The mean value  $\langle N \rangle_L$  and the variance  $\langle\langle N^2 \rangle\rangle_L = \langle N^2 \rangle_L - \langle N \rangle_L^2$  of the number of records read

$$\langle N \rangle_L = \sum_{n=1}^L p_n = H_L = \ln L + \gamma + \dots, \quad (2.7)$$

$$\langle\langle N^2 \rangle\rangle_L = \sum_{n=1}^L p_n(1 - p_n) = H_L - H_L^{(2)} = \ln L + \gamma - \frac{\pi^2}{6} + \dots, \quad (2.8)$$

with

$$H_L = \sum_{n=1}^L \frac{1}{n}, \quad H_L^{(2)} = \sum_{n=1}^L \frac{1}{n^2}, \quad (2.9)$$

whereas  $\gamma = 0.577215\dots$  is Euler's constant, and subleading terms go to zero.

The number of records  $N$  takes its smallest and largest values with probabilities

$$P_L(1) = \frac{1}{L}, \quad P_L(L) = \frac{1}{L!}. \quad (2.10)$$

The distribution of the number of records exhibits a simple asymptotic behavior for long signals ( $L \gg 1$ ). All its cumulants indeed grow logarithmically with  $L$ , with unit prefactor, i.e.,

$$\langle\langle N^k \rangle\rangle_L = \ln L + a_k + \dots, \quad (2.11)$$

where subleading terms go to zero. To leading order, the distribution of  $N$  thus becomes a Poissonian distribution with parameter  $\Lambda = \ln L$ . The  $a_k$  are numerical constants such that

$$\sum_{k \geq 1} \frac{a_k}{k!} s^k = -\ln \Gamma(e^s), \quad (2.12)$$

i.e.,

$$\begin{aligned} a_1 &= \gamma = 0.577215\dots, \\ a_2 &= \gamma - \frac{\pi^2}{6} = -1.067718\dots, \\ a_3 &= \gamma - \frac{\pi^2}{2} + 2\zeta(3) = -1.953472\dots, \end{aligned} \quad (2.13)$$

and so on, where  $\zeta$  denotes Riemann's zeta function. The first two expressions agree with (2.7), (2.8).

## 2.2. Recursive construction

Records can be alternatively described by means of the following backward recursive construction. In spite of its simplicity and of its efficiency, the latter construction is not frequently met in the literature. Let us however mention that it is the prototypical example of a class of sample-space-reducing (SSR) processes put forward very recently [39, 40].

Consider a finite signal of length  $L$ .

- The largest record time  $l_1$  is the time in the range  $n = 1, \dots, L$  for which  $x_n$  is maximal. It is therefore distributed uniformly in the range  $l_1 = 1, \dots, L$ .
- The second largest record time  $l_2$  is the time in the range  $n = 1, \dots, l_1 - 1$  for which  $x_n$  is maximal. It is therefore distributed uniformly in the range  $l_2 = 1, \dots, l_1 - 1$ , and so on.
- The process stops at step  $N$  when  $l_N = 1$  is reached.

The number of records is in the range  $N = 1, \dots, L$ . In terms of the generating function  $G_L(z)$  introduced in (2.2), the first step of the construction translates to

$$G_L(z) = \frac{z}{L} \sum_{l=1}^L G_{l-1}(z), \quad (2.14)$$

with initial condition  $G_0(z) = 1$ . The above equation is equivalent to the recursion

$$(L+1)G_{L+1}(z) = (L+z)G_L(z), \quad (2.15)$$

whose solution coincides with (2.4). This confirms that the above construction indeed yields the statistics of usual records.

## 3. The two-dimensional case

### 3.1. Recursive construction

Two-dimensional record patterns are defined by means of an extension to the square lattice of the recursive construction described in section 2.2. Consider a finite rectangular sample of size  $L \times M$ . Lattice points, to be referred to hereafter as record points, or records for short, are deposited at random according to the following rules.

- The first point  $(l_1, m_1)$  is chosen uniformly in the range  $l_1 = 1, \dots, L$ ,  $m_1 = 1, \dots, M$ .
- The second point  $(l_2, m_2)$  is chosen uniformly in the range  $l_2 = 1, \dots, l_1 - 1$ ,  $m_2 = 1, \dots, m_1 - 1$ , and so on.
- The process stops at step  $N$ , when the  $N$ th point has either  $l_N = 1$  or  $m_N = 1$  or both.

The number of records is in the range

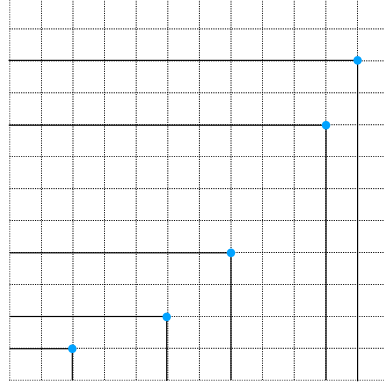
$$N = 1, \dots, \min(L, M). \quad (3.1)$$

The landing position of the process, defined as

$$K = \max(l_N, m_N) = 1, \dots, \max(L, M), \quad (3.2)$$

is nothing but the length of the stick adjacent to the origin in the symmetric fragmentation process considered in [22, 23].

This recursive construction is illustrated in figure 1, showing an instance of a two-dimensional record pattern on a square sample of size  $L = M = 12$ , with  $N = 5$  records at positions  $(11,10)$ ,  $(10,8)$ ,  $(7,4)$ ,  $(5,2)$  and  $(2,1)$ , and landing position  $K = 2$ .



**Figure 1.** An instance of a two-dimensional record pattern on a square sample of size  $L = M = 12$ , with  $N = 5$  records at positions  $(11,10)$ ,  $(10,8)$ ,  $(7,4)$ ,  $(5,2)$  and  $(2,1)$ , and landing position  $K = 2$ .

### 3.2. Exact distribution of the number of records

The statistics of the number of records in two-dimensional record patterns can be investigated by means of generating series. This approach has been tackled in section 2.2. Similar techniques were already used in [23]. Let

$$G_{L,M}(z) = \sum_N P_{L,M}(N) z^N \quad (3.3)$$

be the generating function of the distribution of the number of records on a finite sample of size  $L \times M$ . The first step of the recursive construction described in section 3.1 translates to

$$G_{L,M}(z) = \frac{z}{LM} \sum_{l=1}^L \sum_{m=1}^M G_{l-1,m-1}(z), \quad (3.4)$$

with boundary conditions  $G_{L,0}(z) = G_{0,M}(z) = 1$ . The above equation is equivalent to

$$\begin{aligned} (L+1)(M+1)G_{L+1,M+1}(z) - (L+1)MG_{L+1,M}(z) \\ - L(M+1)G_{L,M+1}(z) + (LM-z)G_{L,M}(z) = 0. \end{aligned} \quad (3.5)$$

This difference equation is recursive, in the sense that it allows us to determine  $G_{L,M}(z)$  step by step, from the sole knowledge of boundary conditions. However, at variance with (2.15), it cannot be solved by elementary means. An integral representation of its solution can be derived as follows. The triple generating function

$$G(x, y, z) = \sum_{L \geq 0} \sum_{M \geq 0} G_{L,M}(z) x^L y^M \quad (3.6)$$

obeys the partial differential equation

$$(1-x)(1-y) \frac{\partial^2 G}{\partial x \partial y} = zG, \quad (3.7)$$

with boundary conditions  $G(x, 0, z) = 1/(1 - x)$ ,  $G(0, y, z) = 1/(1 - y)$ . Performing the change of variables

$$x = 1 - e^{-\lambda}, \quad y = 1 - e^{-\mu}, \quad (3.8)$$

and setting  $\mathcal{G}(\lambda, \mu, z) = G(x, y, z)$ , (3.7) simplifies to

$$\frac{\partial^2 \mathcal{G}}{\partial \lambda \partial \mu} = z \mathcal{G}, \quad (3.9)$$

with boundary conditions  $\mathcal{G}(\lambda, 0, z) = e^\lambda$ ,  $\mathcal{G}(0, \mu, z) = e^\mu$ . The latter equation can be solved in Laplace space. Defining for convenience the double Laplace transform as

$$\widehat{\mathcal{G}}(p, q, z) = \int_0^\infty e^{-(p+1)\lambda} d\lambda \int_0^\infty e^{-(q+1)\mu} d\mu \mathcal{G}(\lambda, \mu, z), \quad (3.10)$$

we obtain the rational expression

$$\begin{aligned} \widehat{\mathcal{G}}(p, q, z) &= R(p, q, z) = \frac{1}{pq} \frac{1 - \eta(p, q)}{1 - z\eta(p, q)}, \\ \eta(p, q) &= \frac{1}{(p+1)(q+1)}. \end{aligned} \quad (3.11)$$

Performing the inverse transforms, we successively obtain

$$\mathcal{G}(\lambda, \mu, z) = \int \frac{dp}{2\pi i} e^{(p+1)\lambda} \int \frac{dq}{2\pi i} e^{(q+1)\mu} R(p, q, z), \quad (3.12)$$

$$G(x, y, z) = \int \frac{dp}{2\pi i} \frac{1}{(1-x)^{p+1}} \int \frac{dq}{2\pi i} \frac{1}{(1-y)^{q+1}} R(p, q, z), \quad (3.13)$$

$$G_{L,M}(z) = \int \frac{dp}{2\pi i} \frac{\Gamma(L+p+1)}{L! \Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{\Gamma(M+q+1)}{M! \Gamma(q+1)} R(p, q, z), \quad (3.14)$$

and finally

$$\begin{aligned} P_{L,M}(N) &= \int \frac{dp}{2\pi i} \frac{\Gamma(L+p+1)}{L! p \Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{\Gamma(M+q+1)}{M! q \Gamma(q+1)} \\ &\quad \times (1 - \eta(p, q)) \eta(p, q)^N. \end{aligned} \quad (3.15)$$

In (3.14) and (3.15), integration contours in the complex  $p$  and  $q$ -planes circle once around all poles of the integrands.

The exact expression (3.15) for the distribution of the number of two-dimensional records can be transformed by means of the identity

$$\int \frac{dp}{2\pi i} \frac{\Gamma(L+p+1)}{L! \Gamma(p+1)} \frac{1}{p(p+1)^{N+1}} = F_L(N), \quad (3.16)$$

which can be derived by applying residue calculus, using (2.4) and (2.6). We thus obtain

$$P_{L,M}(N) = F_L(N-1)F_M(N-1) - F_L(N)F_M(N). \quad (3.17)$$

The above formula has the following probabilistic interpretation. The number  $N$  of two-dimensional records on a rectangle of size  $L \times M$  is distributed as the smaller of  $N_1$  and  $N_2$ , which are respectively the numbers of one-dimensional records on signals of sizes  $L$  and  $M$ . This property is a consequence of the recursive construction. It is nevertheless comforting to see it emerging out of the formalism.



As a consequence of (3.17), the smallest number of records ( $N = 1$ ) occurs with probability

$$P_{L,M}(1) = \frac{L + M - 1}{LM}, \quad (3.18)$$

whereas the probability for the number of records to be the largest only assumes a simple form for square samples:

$$P_{L,L}(L) = \frac{1}{L!^2}. \quad (3.19)$$

### 3.3. Mean number of records

This section is devoted to a study of the mean number  $\langle N \rangle_{L,M}$  of two-dimensional records, with an emphasis on its asymptotic behavior on large square and rectangular samples. We have

$$\langle N \rangle_{L,M} = \left( z \frac{d}{dz} \right)_{z=1} G_{L,M}(z). \quad (3.20)$$

Equation (3.5) yields the recursive difference equation

$$\begin{aligned} (L+1)(M+1)\langle N \rangle_{L+1,M+1} - (L+1)M\langle N \rangle_{L+1,M} \\ - L(M+1)\langle N \rangle_{L,M+1} + (LM-1)\langle N \rangle_{L,M} = 1, \end{aligned} \quad (3.21)$$

with boundary conditions  $\langle N \rangle_{L,0} = \langle N \rangle_{0,M} = 0$ . Exact rational expressions for the mean number of records can be obtained by solving the above equation recursively. Table 1 shows a comparison between  $\langle N \rangle_L$  for one-dimensional signals of length  $L$  and  $\langle N \rangle_{L,L}$  for two-dimensional square samples of size  $L \times L$ . The complexity of mean number of records grows much faster in two dimensions than in one dimension.

| $L$                       | 1 | 2             | 3               | 4                 | 5                   | 6                       | 7                           | 8                             | 9                                 |
|---------------------------|---|---------------|-----------------|-------------------|---------------------|-------------------------|-----------------------------|-------------------------------|-----------------------------------|
| $\langle N \rangle_L$     | 1 | $\frac{3}{2}$ | $\frac{11}{6}$  | $\frac{25}{12}$   | $\frac{137}{60}$    | $\frac{49}{20}$         | $\frac{363}{140}$           | $\frac{761}{280}$             | $\frac{7129}{2520}$               |
| $\langle N \rangle_{L,L}$ | 1 | $\frac{5}{4}$ | $\frac{53}{36}$ | $\frac{475}{288}$ | $\frac{4309}{2400}$ | $\frac{497567}{259200}$ | $\frac{25752527}{12700800}$ | $\frac{862916227}{406425600}$ | $\frac{24239088877}{10973491200}$ |

**Table 1.** Exact rational expressions for the mean number  $\langle N \rangle_L$  of records for one-dimensional signals of length  $L$  and  $\langle N \rangle_{L,L}$  for two-dimensional square samples of size  $L \times L$ , up to  $L = 9$ .

Equation (3.14) yields the following exact integral representation for the mean number of records:

$$\begin{aligned} \langle N \rangle_{L,M} &= \int \frac{dp}{2\pi i} \frac{\Gamma(L+p+1)}{L! \Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{\Gamma(M+q+1)}{M! \Gamma(q+1)} \\ &\quad \times \frac{1}{pq(p+q+pq)}, \end{aligned} \quad (3.22)$$

which is suitable for studying the behavior of  $\langle N \rangle_{L,M}$  on large samples. Neglecting terms of relative order  $1/L$  and  $1/M$ , the above expression simplifies to

$$\langle N \rangle_{L,M} \approx \int \frac{dp}{2\pi i} \frac{L^p}{\Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{M^q}{\Gamma(q+1)} \frac{1}{pq(p+q+pq)}. \quad (3.23)$$

Several kinds of sample shapes can be considered. In the simplest situation of large square samples ( $L = M \gg 1$ ), performing the integral over  $q$  in (3.23) yields

$$\langle N \rangle_{L,L} \approx \Lambda + \gamma - \int \frac{dp}{2\pi i} \frac{e^{\Lambda p^2/(p+1)}}{p^2 \Gamma(p+1) \Gamma(1/(p+1))}, \quad (3.24)$$

with  $\Lambda = \ln L$  and  $\gamma$  is again Euler's constant. For large  $\Lambda$ , the integral is dominated by small values of  $p$  such that  $\Lambda p^2$  is of order unity. To leading order, this reads

$$\int \frac{dp}{2\pi i} \frac{e^{\Lambda p^2}}{p^2} = \sqrt{\frac{\Lambda}{\pi}}. \quad (3.25)$$

Higher-order corrections can be derived systematically by expanding the integrand in (3.24) in increasing powers of  $p$  and integrating term by term. We thus obtain an asymptotic expansion in decreasing powers of  $\ln L$ :

$$\langle N \rangle_{L,L} = \ln L - \sqrt{\frac{\ln L}{\pi}} + \gamma - \frac{1}{2\sqrt{\pi \ln L}} \left( \gamma - \frac{1}{8} - \frac{\pi^2}{6} \right) + \dots \quad (3.26)$$

For large rectangular samples, i.e., for  $L$  and  $M$  large and proportional to each other, performing the integral over  $q$  in (3.23) yields

$$\langle N \rangle_{L,M} \approx \Lambda - \frac{r}{2} + \gamma - \int \frac{dp}{2\pi i} \frac{e^{\Lambda p^2/(p+1) - rp(p+2)/(2(p+1))}}{p^2 \Gamma(p+1) \Gamma(1/(p+1))}, \quad (3.27)$$

with

$$\Lambda = \frac{\ln LM}{2}, \quad r = \ln \frac{M}{L}. \quad (3.28)$$

In the regime where  $\Lambda$  is large whereas  $r$  remains finite, the integral is again dominated by small values of  $p$  such that  $\Lambda p^2$  is of order unity. We are thus left with

$$\begin{aligned} \langle N \rangle_{L,M} &\approx \frac{\ln LM}{2} - \sqrt{\frac{\ln LM}{2\pi}} + \gamma \\ &\quad - \frac{1}{\sqrt{2\pi \ln LM}} \left[ \frac{1}{2} \left( \ln \frac{M}{L} \right)^2 + \gamma - \frac{1}{8} - \frac{\pi^2}{6} \right] + \dots \end{aligned} \quad (3.29)$$

For very anisotropic samples such that  $1 \ll L \ll M$ , the integral over  $q$  in (3.23) is dominated by the trivial pole at  $q = 0$ , and so the two-dimensional result boils down to the one-dimensional one:

$$\langle N \rangle_{L,M} \approx N_L \approx \ln L + \gamma, \quad (3.30)$$

irrespective of the larger side  $M$ . This reduction is a manifestation of the property expressed by (3.17). We recall that the distributions of  $N_1$  and  $N_2$  are nearly Poissonian, and therefore sharply peaked around  $\Lambda_1 = \ln L$  and  $\Lambda_2 = \ln M$ . If  $\Lambda_2$  is sufficiently larger than  $\Lambda_1$ , the distributions of  $N_1$  and  $N_2$  have essentially no overlap, and the two-dimensional number of records  $N$  is essentially distributed as  $N_1$ .

The crossover between the results (3.26) in the isotropic geometry of squares and (3.30) in the very anisotropic one takes place when  $r = \Lambda_2 - \Lambda_1$  is positive and of the order of  $\sqrt{\Lambda}$ . This regime can be investigated by simplifying the integral representation (3.27) as

$$\langle N \rangle_{L,M} \approx \Lambda - \frac{r}{2} + \gamma - \int \frac{dp}{2\pi i} \frac{e^{\Lambda p^2 - rp}}{p^2}, \quad (3.31)$$

yielding

$$\langle N \rangle_{L,M} \approx \Lambda + \gamma - \sqrt{\frac{\Lambda}{\pi}} e^{-r^2/(4\Lambda)} - \frac{r}{2} \operatorname{erf} \frac{r}{2\sqrt{\Lambda}}, \quad (3.32)$$

where  $\operatorname{erf}$  denotes the error function. In the weakly anisotropic regime ( $r \ll \sqrt{\Lambda}$ ), the expansion

$$\langle N \rangle_{L,M} \approx \Lambda - \sqrt{\frac{\Lambda}{\pi}} + \gamma - \frac{r^2}{4\sqrt{\pi\Lambda}} + \dots \quad (3.33)$$

matches (3.29). In the strongly anisotropic regime ( $r \gg \sqrt{\Lambda}$ ), we have

$$\langle N \rangle_{L,M} \approx \ln L + \gamma - 2\sqrt{\frac{\Lambda^3}{\pi}} \frac{e^{-r^2/(4\Lambda)}}{r^2}. \quad (3.34)$$

The correction to (3.30) is exponentially small, and negative, in agreement with the property expressed by (3.17).

### 3.4. Second moment and variance of the number of records

Higher moments of the number of two-dimensional records can be dealt with along the same lines. The second moment is given by

$$\langle N^2 \rangle_{L,M} = \left( z \frac{d}{dz} \right)_{z=1}^2 G_{L,M}(z). \quad (3.35)$$

Equation (3.5) yields the recursive difference equation

$$\begin{aligned} & (L+1)(M+1)\langle N^2 \rangle_{L+1,M+1} - (L+1)M\langle N^2 \rangle_{L+1,M} \\ & - L(M+1)\langle N^2 \rangle_{L,M+1} + (LM-1)\langle N^2 \rangle_{L,M} = 2\langle N \rangle_{L,M} + 1, \end{aligned} \quad (3.36)$$

with boundary conditions  $\langle N^2 \rangle_{L,0} = \langle N^2 \rangle_{0,M} = 0$ . Exact rational expressions for  $\langle N^2 \rangle_{L,M}$  can be obtained by solving (3.21) and (3.36) step by step. Table 2 shows a comparison between the variances of the number of records  $\langle\langle N^2 \rangle\rangle_L$  for one-dimensional signals of length  $L$  and  $\langle\langle N^2 \rangle\rangle_{L,L}$  for two-dimensional square samples of size  $L \times L$ . The complexity of these exact results again grows much faster in two dimensions than in one dimension.

| $L$                                       | 1 | 2              | 3                  | 4                     | 5                         | 6                                 |
|-------------------------------------------|---|----------------|--------------------|-----------------------|---------------------------|-----------------------------------|
| $\langle\langle N^2 \rangle\rangle_L$     | 0 | $\frac{1}{4}$  | $\frac{17}{36}$    | $\frac{95}{144}$      | $\frac{2951}{3600}$       | $\frac{3451}{3600}$               |
| $\langle\langle N^2 \rangle\rangle_{L,L}$ | 0 | $\frac{3}{16}$ | $\frac{395}{1296}$ | $\frac{33575}{82944}$ | $\frac{2826119}{5760000}$ | $\frac{38000939711}{67184640000}$ |

**Table 2.** Exact rational expressions for the variances of the number of records  $\langle\langle N^2 \rangle\rangle_L$  for one-dimensional signals of length  $L$  and  $\langle\langle N^2 \rangle\rangle_{L,L}$  for two-dimensional square samples of size  $L \times L$ , up to  $L = 6$ .

Equation (3.14) yields the exact integral representation

$$\begin{aligned} \langle N^2 \rangle_{L,M} &= \int \frac{dp}{2\pi i} \frac{\Gamma(L+p+1)}{L! \Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{\Gamma(M+q+1)}{M! \Gamma(q+1)} \\ &\quad \times \frac{2+p+q+pq}{pq(p+q+pq)^2}, \end{aligned} \quad (3.37)$$

which is again very suitable for studying the behavior of  $\langle N^2 \rangle_{L,M}$  on large samples. Neglecting terms of relative order  $1/L$  and  $1/M$ , the above expression simplifies to

$$\langle N^2 \rangle_{L,M} \approx \int \frac{dp}{2\pi i} \frac{L^p}{\Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{M^q}{\Gamma(q+1)} \frac{2+p+q+pq}{pq(p+q+pq)^2}. \quad (3.38)$$

Let us focus our attention onto the simplest situation of large square samples ( $L = M \gg 1$ ). Performing the integral over  $q$  in (3.38) yields

$$\begin{aligned} \langle N^2 \rangle_{L,L} &\approx (\Lambda + \gamma)^2 + \frac{\pi^2}{6} - \int \frac{dp}{2\pi i} \frac{e^{\Lambda p^2/(p+1)}}{\Gamma(p+1)\Gamma(1/(p+1))} \\ &\times \left\{ \frac{2}{p^3} + \frac{1}{p^2} + \frac{2}{p^2(p+1)} \left[ \Lambda - \psi\left(\frac{1}{p+1}\right) \right] \right\}, \end{aligned} \quad (3.39)$$

where  $\Lambda = \ln L$  and  $\psi(z) = \Gamma'(z)/\Gamma(z)$  is the digamma function. For large  $\Lambda$ , the integral is again dominated by small values of  $p$  such that  $\Lambda p^2$  is of order unity. The same analysis as in section 3.3 yields the following expansion for the variance of the number of records in decreasing powers of  $\ln L$ :

$$\begin{aligned} \langle\langle N^2 \rangle\rangle_{L,L} &= \left(1 - \frac{1}{\pi}\right) \left(\ln L + \gamma - \frac{\pi^2}{6}\right) - \sqrt{\frac{\ln L}{4\pi}} + \frac{1}{8\pi} \\ &- \frac{1}{4\sqrt{\pi \ln L}} \left(\gamma + \frac{1}{8} - \frac{5\pi^2}{6} + 4\zeta(3)\right) + \dots \end{aligned} \quad (3.40)$$

To sum up, the main characteristic features of the statistics of two-dimensional record patterns are as follows. The number of records  $N$  on a rectangular sample of size  $L \times M$  is distributed as the smaller of the numbers  $N_1$  and  $N_2$  of one-dimensional records for sequences of lengths  $L$  and  $M$  (see (3.17)). For large square samples, the mean number of records  $\langle N \rangle_{L,L}$  exhibits a negative correction in  $\sqrt{\ln L}/\pi$  with respect to the one-dimensional case of usual records (see (3.26)), whereas the logarithmic growth of its variance is reduced by the universal factor  $1 - 1/\pi$  (see (3.40)). These results are illustrated in figure 2, showing the first two cumulants of the number of records  $N$  on square samples against  $\ln L$ . The exact values of these quantities for finite sample sizes up to  $L = 100$ , obtained by solving the recursive difference equations (3.21) and (3.36), are accurately described by the asymptotic expansions (3.26) and (3.40).

In section 4 we extend the above results to higher-dimensional record patterns in any dimension  $D$ .

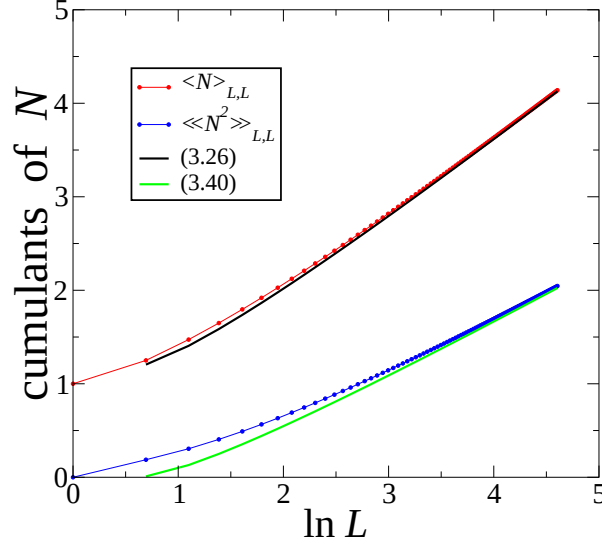
### 3.5. Distribution of the landing position

The formalism used in section 3.2 to derive the exact distribution of the number of records can be extended to investigate other quantities of interest in two-dimensional record patterns.

Hereafter we consider the distribution  $R_{L,M}(K)$  of the landing position  $K$ , defined in section 3.1 and illustrated in figure 1. This quantity equals the length of the stick adjacent to the origin in the symmetric fragmentation process considered in [22, 23].

The first step of the recursive construction described in section 3.1 translates to

$$R_{L,M}(K) = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M R_{l-1,m-1}(K), \quad (3.41)$$



**Figure 2.** First two cumulants of the number of records  $N$  on square samples, against  $\ln L$ . Exact values for finite sample sizes up to  $L = 100$  are compared to the four-term asymptotic expansions (3.26) and (3.40).

with boundary conditions  $R_{L,0}(K) = \delta_{L,K-1}$ ,  $R_{0,M}(K) = \delta_{M,K-1}$ , where  $\delta_{i,j}$  denotes the Kronecker symbol. Indeed, if  $l_1 = 1, \dots, L$  and  $m_1 = 1$  are chosen at the first step, the process stops at this step and the landing position is  $K = l_1$ . Similarly, if  $l_1 = 1$  and  $m_1 = 1, \dots, M$  are chosen at the first step, the process stops at this step and the landing position is  $K = m_1$ . Equation (3.41) is equivalent to

$$\begin{aligned} (L+1)(M+1)R_{L+1,M+1}(K) - (L+1)MR_{L+1,M}(K) \\ - L(M+1)R_{L,M+1}(K) + (LM-1)R_{L,M}(K) = 0. \end{aligned} \quad (3.42)$$

The moments of the landing position,

$$\langle K^s \rangle_{L,M} = \sum_K K^s R_{L,M}(K), \quad (3.43)$$

obey the same homogeneous difference equation, i.e.,

$$\begin{aligned} (L+1)(M+1)\langle K^s \rangle_{L+1,M+1} - (L+1)M\langle K^s \rangle_{L+1,M} \\ - L(M+1)\langle K^s \rangle_{L,M+1} + (LM-1)\langle K^s \rangle_{L,M} = 0, \end{aligned} \quad (3.44)$$

with boundary conditions  $\langle K^s \rangle_{L,0} = (L+1)^s$ ,  $\langle K^s \rangle_{0,M} = (M+1)^s$ . The exponent  $s$  is not necessarily an integer.

The difference equations (3.42) and (3.44) are recursive, just as (3.5), (3.21) and (3.36). All these equations can be solved recursively, from the sole knowledge of boundary conditions. As an illustration of the method, table 3 gives the exact rational values of the first two moments of the landing position  $K$  for square samples of size  $L \times L$ . The asymptotic growth of these moments is given in (3.58).

An integral representation of  $R_{L,M}(K)$  can be derived by means of the generating-function approach presented in section (3.2). We thus obtain

$$R_{L,M}(K) = \int \frac{dp}{2\pi i} \frac{\Gamma(L+p+1)}{L! \Gamma(p+1)} \int \frac{dq}{2\pi i} \frac{\Gamma(M+q+1)}{M! \Gamma(q+1)}$$

| $L$                         | 1 | 2             | 3               | 4                    | 5                        | 6                        | 7                          | 8                                    |
|-----------------------------|---|---------------|-----------------|----------------------|--------------------------|--------------------------|----------------------------|--------------------------------------|
| $\langle K \rangle_{L,L}$   | 1 | $\frac{3}{2}$ | $\frac{11}{6}$  | $\frac{203}{96}$     | $\frac{227}{96}$         | $\frac{22\,399}{8\,640}$ | $\frac{211\,943}{75\,600}$ | $\frac{38\,719\,637}{12\,902\,400}$  |
| $\langle K^2 \rangle_{L,L}$ | 1 | $\frac{5}{2}$ | $\frac{71}{18}$ | $\frac{1\,583}{288}$ | $\frac{10\,307}{1\,440}$ | $\frac{15\,413}{1\,728}$ | $\frac{814\,511}{75\,600}$ | $\frac{492\,106\,337}{38\,707\,200}$ |

**Table 3.** Exact rational expressions for the first two moments of the landing position  $K$  for square samples of size  $L \times L$ , up to  $L = 8$ .

$$\times \frac{S(p, q, K)}{p + q + pq}, \quad (3.45)$$

with

$$S(p, q, K) = (K - 1)! \left( \frac{\Gamma(p + 2)}{\Gamma(K + p + 1)} + \frac{\Gamma(q + 2)}{\Gamma(K + q + 1)} - \delta_{K,1} \right). \quad (3.46)$$

The above expression is again suitable for studying the regime of large samples. Neglecting terms of relative order  $1/L$  and  $1/M$ , (3.45) simplifies to

$$R_{L,M}(K) \approx \int \frac{dp}{2\pi i} \frac{L^p}{\Gamma(p + 1)} \int \frac{dq}{2\pi i} \frac{M^q}{\Gamma(q + 1)} \frac{S(p, q, K)}{p + q + pq}. \quad (3.47)$$

Henceforth we again focus our attention onto the simplest situation of large square samples ( $L = M \gg 1$ ). Rearranging terms in  $S(p, q, K)$  and performing the integral over  $q$  in (3.47) yields

$$R_{L,L}(K) \approx 2 \left( 1 - \frac{1}{2} \delta_{K,1} \right) \int \frac{dp}{2\pi i} \frac{e^{\Lambda p^2/(p+1)}}{\Gamma(1/(p+1))} \frac{(K-1)!}{\Gamma(K+p+1)}, \quad (3.48)$$

with  $\Lambda = \ln L$ .

Let us first consider the asymptotic behavior of the distribution  $R_{L,L}(K)$  in the regime where the sample size  $L$  becomes large, whereas the landing position  $K$  is finite. The integral in (3.48) is again dominated by small values of  $p$  such that  $\Lambda p^2$  is of order unity. We thus obtain, to leading order

$$R_{L,L}(K) \approx \frac{1}{\sqrt{\pi \ln L}} \frac{1 - \frac{1}{2} \delta_{K,1}}{K}. \quad (3.49)$$

This result calls for several remarks. The probability of having  $K = 1$  is reduced by a factor 2 with respect to the rest of the distribution. This property holds for the full distribution of generic sticks as well [23]. Here,  $K = 1$  occurs only when the last record is deposited at  $(1, 1)$ , whereas  $K \geq 2$  corresponds to two distinct positions  $(1, K)$  and  $(K, 1)$  for the last record, i.e., two distinct orientations for the stick adjacent to the origin. The  $1/K$  law entering (3.49) is not normalizable if extended to arbitrarily large values of  $K$ . The dependence of the prefactor on the sample size  $L$  suggests that the above law is cutoff at the scale  $\ln K \sim \sqrt{\ln L}$ . This estimate is corroborated by the scaling result (3.55).

Higher-order corrections to (3.49) can be derived by expanding the integrand in (3.48) in increasing powers of  $p$  and integrating term by term. The first non-trivial correction reads

$$R_{L,L}(K) = \frac{1}{\sqrt{\pi \ln L}} \frac{1 - \frac{1}{2} \delta_{K,1}}{K} \left( 1 - \frac{C(K)}{4 \ln L} + \dots \right), \quad (3.50)$$

with

$$\begin{aligned} C(K) &= (H_K)^2 - 3H_K + H_K^{(2)} - \frac{\pi^2}{3} + 2\gamma + \frac{3}{4} \\ &= (\ln K)^2 + (2\gamma - 3) \ln K - \frac{\pi^2}{6} + \gamma^2 - \gamma + \frac{3}{4} + \dots, \end{aligned} \quad (3.51)$$

where  $H_K$  and  $H_K^{(2)}$  are defined in (2.9).

We now turn to the regime where the sample size  $L$  and the landing position  $K$  are both large. The relevant scaling corresponds to  $\Lambda = \ln L$  and  $\lambda = \ln K$  being large and comparable, with a fixed ratio  $a$ , so that  $\lambda = a\Lambda$ , i.e.,

$$K = L^a \quad (0 < a < 1). \quad (3.52)$$

The formula (3.48) then simplifies to

$$R_{L,L}(K = L^a) \approx 2 \int \frac{dp}{2\pi i} \frac{e^{\Lambda(p^2/(p+1) - (p+1)a)}}{\Gamma(1/(p+1))}. \quad (3.53)$$

It is legitimate to estimate the above integral by means of the saddle-point method. The saddle point reads  $p = -1 + 1/\sqrt{1-a}$ . We thus obtain

$$R_{L,L}(K = L^a) \approx \frac{L^{-2(1-\sqrt{1-a})}}{\sqrt{\pi \ln L} (1-a)^{3/4} \Gamma(\sqrt{1-a})}. \quad (3.54)$$

The most probable values of the landing position correspond to  $a \ll 1$ , where the result (3.54) simplifies to

$$R_{L,L}(K) \approx \frac{1}{\sqrt{\pi \ln L}} \frac{e^{-(\ln K)^2/(4 \ln L)}}{K}. \quad (3.55)$$

We therefore predict that the bulk of the distribution of  $K$  assumes a half-log-normal form. This scaling form matches to leading order the growth of the amplitude  $C(K)$  (see (3.51)).

An alternative viewpoint consists in considering the moments of the landing position, introduced in (3.43), for  $s > 0$ . The most direct route to determine the growth of these moments for large  $L$  starts from (3.53) and yields

$$\langle K^s \rangle_{L,L} \approx 2 \int \frac{dp}{2\pi i} \frac{e^{\Lambda(p^2/(p+1))}}{\Gamma(1/(p+1))} \frac{1}{p-s}, \quad (3.56)$$

where the contour integral is to the right of the pole at  $p = s$ . The integral is dominated by the latter pole. We thus obtain the scaling law

$$\langle K^s \rangle_{L,L} \approx \frac{2}{\Gamma(1/(s+1))} L^{s^2/(s+1)}. \quad (3.57)$$

In particular, integer moments of  $K$  grow with rational exponents, i.e.,

$$\begin{aligned} \langle K \rangle_{L,L} &\approx \frac{2}{\sqrt{\pi}} L^{1/2}, \\ \langle K^2 \rangle_{L,L} &\approx \frac{2}{\Gamma(1/3)} L^{4/3}, \\ \langle K^3 \rangle_{L,L} &\approx \frac{2}{\Gamma(1/4)} L^{9/4}, \end{aligned} \quad (3.58)$$

and so on.

The results (3.54) and (3.57) demonstrate multifractality [41, 42]. The probability distribution of the landing position and its moments indeed scale as

$$R_{L,L}(K = L^a) \sim L^{-\phi(a)}, \quad \langle K^s \rangle_{L,L} \sim L^{\tau(s)}. \quad (3.59)$$

Both nonlinear spectra of exponents

$$\begin{aligned} \phi(a) &= 2(1 - \sqrt{1-a}) & (0 < a < 1), \\ \tau(s) &= \frac{s^2}{s+1} & (s > 0) \end{aligned} \quad (3.60)$$

encode the multifractality of the distribution of the landing position. They are related to each other by the Legendre transformation

$$\tau(s) = \max_a [(s+1)a - \phi(a)]. \quad (3.61)$$

We note that moments with a negative exponent ( $s = -\sigma < 0$ ) escape multifractality. They are dominated by microscopic values of  $K$ , where (3.49) holds, hence

$$\langle K^{-\sigma} \rangle_{L,L} \approx \frac{1}{\sqrt{\pi \ln L}} \left( \zeta(\sigma+1) - \frac{1}{2} \right). \quad (3.62)$$

## 4. The higher-dimensional case

### 4.1. Recursive construction

Higher-dimensional record patterns are defined by means of an extension of the recursive construction described in sections 2.2 and 3.1. Consider a finite parallelotopic sample of size  $L_1 \times \dots \times L_D$ , with side lengths  $L_i \geq 1$  for  $i = 1, \dots, D$ . Records are deposited onto this sample according to the following rules.

- The first point  $(l_{1,1}, \dots, l_{1,D})$  is chosen uniformly in the range  $l_{1,i} = 1, \dots, L_i$  for  $i = 1, \dots, D$ .
- The second point  $(l_{2,1}, \dots, l_{2,D})$  is chosen uniformly in the range  $l_{2,i} = 1, \dots, l_{1,i} - 1$  for  $i = 1, \dots, D$ , and so on.
- The process stops at step  $N$ , when (at least) one of the co-ordinates  $l_{N,i}$  of the  $N$ th point equals unity.

The number of records is in the range  $N = 1, \dots, \min(\{L_i\})$ .

### 4.2. Exact distribution of the number of records

The statistics of the number  $N$  of  $D$ -dimensional records in the random patterns thus generated can be investigated along the lines of the approach sketched for  $D = 1$  in section 2.2 and described for  $D = 2$  in section 3.2. We shall henceforth only give essential results.

The distribution of  $N$  is exactly given by the following formula, which is a direct generalization of (3.15):

$$\begin{aligned} P_{\{L_i\}}(N) &= \int \dots \int \prod_{i=1}^D \frac{dp_i}{2\pi i} \frac{\Gamma(L_i + p_i + 1)}{L_i! p_i \Gamma(p_i + 1)} \\ &\times (1 - \eta(\{p_i\})) \eta(\{p_i\})^N, \end{aligned} \quad (4.1)$$



with

$$\eta(\{p_i\}) = \prod_{i=1}^D \frac{1}{1+p_i}. \quad (4.2)$$

The above expression can be transformed by means of the identity (3.16) into

$$P_{\{L_i\}}(N) = \prod_{i=1}^D F_{L_i}(N-1) - \prod_{i=1}^D F_{L_i}(N), \quad (4.3)$$

whose meaning is that the number  $N$  of records is distributed as the smallest of the numbers  $N_i$  of one-dimensional records of independent signals whose lengths  $L_i$  are the sides of the sample. In particular, the probability of having one single record reads

$$P_{\{L_i\}}(1) = 1 - \prod_{i=1}^D \left(1 - \frac{1}{L_i}\right). \quad (4.4)$$

#### 4.3. Asymptotic behavior on large hypercubic samples

We now turn to the asymptotic behavior of the distribution of the number  $N$  of records, and especially of its mean value and variance. For definiteness we restrict the analysis to large hypercubic samples, with equal sides in all directions ( $L_i = L$  for  $i = 1, \dots, D$ ). The formula (4.3) simplifies to

$$P_{L\dots L}(N) = F_L(N-1)^D - F_L(N)^D, \quad (4.5)$$

meaning that the number  $N$  of  $D$ -dimensional records is distributed as the smallest of  $D$  independent copies of the number  $N_1$  of one-dimensional records in a signal of length  $L$ .

The result (4.5) is the starting point of the following asymptotic analysis, which is entirely based on probabilistic reasoning. Let us first revisit the one-dimensional case of usual records. To leading order as  $\Lambda = \ln L$  is large, the bulk of the distribution of  $N_1$  is Gaussian, and its mean value and its variance are both equal to  $\Lambda$  (see (2.7) and (2.8)). We have therefore

$$N_1 \approx \ln L - \sqrt{\ln L} X_1, \quad (4.6)$$

where the rescaled variable  $X_1$  is a standard Gaussian variable, with  $\langle X_1 \rangle = 0$  and  $\langle X_1^2 \rangle = 1$ . The minus sign has been introduced for further convenience. The corresponding probability density and cumulative distribution read

$$f(X) = \Phi'(X) = \frac{e^{-X^2/2}}{\sqrt{2\pi}}, \quad (4.7)$$

$$\Phi(x) = \text{Prob}\{X_1 < x\} = \int_{-\infty}^x f(X) dX = \frac{1}{2} \left(1 + \text{erf} \frac{x}{\sqrt{2}}\right). \quad (4.8)$$

For  $D$ -dimensional records on large hypercubic samples, as a consequence of (4.5), we have a similar scaling law, i.e.,

$$N \approx \ln L - \sqrt{\ln L} X, \quad (4.9)$$

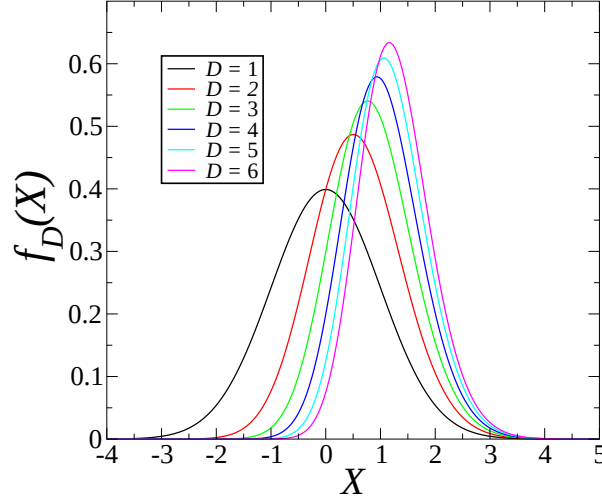
where the rescaled variable  $X$  is distributed as the largest of  $D$  standard Gaussian variables. In other words, the cumulative distribution of  $X$  reads

$$\text{Prob}\{X < x\} = \Phi(x)^D. \quad (4.10)$$

Its probability density is therefore given by

$$f_D(X) = D\Phi(X)^{D-1}f(X) = \frac{D}{\sqrt{2\pi}}\Phi(X)^{D-1}e^{-X^2/2}. \quad (4.11)$$

Figure 3 demonstrates that the probability density  $f_D(X)$  slowly drifts towards positive values while its width shrinks as  $D$  increases. The black curve shows the Gaussian density  $f_1(X) = f(X)$  (see (4.7)).



**Figure 3.** Probability density  $f_D(X)$  of the rescaled variable  $X$  describing the fluctuations of the number of  $D$ -dimensional records on large hypercubic samples (see (4.9) and (4.11)) for several dimensions  $D$  (see legend).

There is therefore a direct relationship between multidimensional records and extreme-value statistics for Gaussian random variables. In particular, the mean and variance of the number of records  $N$  on large hypercubic samples read

$$\begin{aligned} \langle N \rangle_{L \dots L} &\approx \ln L - A_D \sqrt{\ln L}, \\ \langle\langle N^2 \rangle\rangle_{L \dots L} &\approx B_D \ln L, \end{aligned} \quad (4.12)$$

where the amplitudes  $A_D$  and  $B_D$  are given by

$$\begin{aligned} A_D &= \langle X \rangle_D = \int_{-\infty}^{\infty} X f_D(X) dX, \\ C_D &= \langle X^2 \rangle_D = \int_{-\infty}^{\infty} X^2 f_D(X) dX, \\ B_D &= \langle\langle X^2 \rangle\rangle_D = C_D - A_D^2, \end{aligned} \quad (4.13)$$

where  $f_D(X)$  is the probability density of the largest of  $D$  standard Gaussian variables (see (4.11)).

The above integrals can be evaluated in closed form up to  $D = 5$  by means of an elegant approach due to Selby [43]. First, integrations by parts are used to bring  $A_D$  and  $C_D$  to the form

$$A_D = \frac{D(D-1)}{2\pi} \int_{-\infty}^{\infty} e^{-X^2} \Phi(X)^{D-2} dX,$$

$$C_D = 1 + \frac{D(D-1)(D-2)}{2(2\pi)^{3/2}} \int_{-\infty}^{\infty} e^{-3X^2/2} \Phi(X)^{D-3} dX. \quad (4.14)$$

Second, noticing that

$$E(X) = \Phi(X) - \frac{1}{2} = \frac{1}{2} \operatorname{erf} \frac{X}{\sqrt{2}} \quad (4.15)$$

is an odd function of  $X$ , the integrals entering (4.14) can be expanded as

$$\begin{aligned} A_D &= \frac{D(D-1)}{2^{D-1}\pi} \sum_{n=0}^{\lfloor (D-2)/2 \rfloor} 2^{2n} \binom{D-2}{2n} I_n(2), \\ C_D &= 1 + \frac{D(D-1)(D-2)}{2^{D-2}(2\pi)^{3/2}} \sum_{n=0}^{\lfloor (D-3)/2 \rfloor} 2^{2n} \binom{D-3}{2n} I_n(3), \end{aligned} \quad (4.16)$$

with

$$I_n(s) = \int_{-\infty}^{\infty} e^{-sX^2/2} E(X)^{2n} dX. \quad (4.17)$$

Third, the key observation that the  $I_n(s)$  obey ‘a strange reduction rule’ [43], i.e.,

$$I_{n+1}(s) = \frac{2n+1}{2\pi\sqrt{s}} \int_0^{1/\sqrt{s}} \frac{dy}{1+y^2} I_n(1+s(1+y^2)), \quad (4.18)$$

allows one to derive

$$I_0(s) = \sqrt{\frac{2\pi}{s}}, \quad I_1(s) = \frac{1}{\sqrt{2\pi s}} \arcsin \frac{1}{1+s}. \quad (4.19)$$

The higher integrals  $I_n(s)$  for  $n \geq 2$  do not seem to admit closed-form expressions. This approach yields

$$\begin{aligned} A_1 &= 0, \\ A_2 &= \frac{1}{\sqrt{\pi}} = 0.564189\dots, \\ A_3 &= \frac{3}{2\sqrt{\pi}} = 0.846284\dots, \\ A_4 &= \frac{3 \arccos(-1/3)}{\pi^{3/2}} = 1.029375\dots, \\ A_5 &= \frac{5 \arccos(-23/27)}{2\pi^{3/2}} = 1.162964\dots \end{aligned} \quad (4.20)$$

and

$$\begin{aligned} C_1 &= 1, \\ C_2 &= 1, \\ C_3 &= 1 + \frac{\sqrt{3}}{2\pi} = 1.275664\dots, \\ C_5 &= 1 + \frac{5\sqrt{3} \arccos(-1/4)}{2\pi^2} = 1.800020\dots, \end{aligned} \quad (4.21)$$

and so

$$\begin{aligned}
B_1 &= 1, \\
B_2 &= 1 - \frac{1}{\pi} = 0.681690\dots, \\
B_3 &= 1 + \frac{2\sqrt{3}-9}{4\pi} = 0.559467\dots, \\
B_4 &= 1 + \frac{\sqrt{3}}{\pi} - \frac{9 \arccos(-1/3)^2}{\pi^3} = 0.491715\dots, \\
B_5 &= 1 + \frac{5\sqrt{3} \arccos(-1/4)}{2\pi^2} - \frac{25 \arccos(-23/27)^2}{4\pi^3} = 0.447534\dots
\end{aligned} \tag{4.22}$$

The above expressions seem to exhaust the list of available closed-form results. The formulas for  $A_2$  and  $B_2$  agree with (3.26) and (3.40).

The regime of large dimensions ( $D \gg 1$ ) is also worth being investigated. The leading behavior of the amplitudes  $A_D$  and  $B_D$  can be derived as follows. Roughly speaking, the expression (4.10) simplifies to

$$\text{Prob}\{X < x\} \sim \exp\left(-D e^{-x^2/2}\right). \tag{4.23}$$

This estimate expresses that the variable

$$\xi = D e^{-X^2/2} \tag{4.24}$$

becomes exponentially distributed, i.e., that its probability density converges to  $e^{-\xi}$ , whenever  $X$  and  $D$  are both large with  $\xi$  being finite, i.e.,

$$X \approx \sqrt{2 \ln D} - \frac{\ln \xi}{\sqrt{2 \ln D}}. \tag{4.25}$$

The random variable  $(-\ln \xi)$  entering this expression is a normalized Gumbel variable. We have in particular  $\langle -\ln \xi \rangle = \gamma$  and  $\langle (\ln \xi)^2 \rangle = \pi^2/6$ . We thus obtain the leading-order estimates

$$A_D \approx \sqrt{2 \ln D}, \quad B_D \approx \frac{\pi^2}{12 \ln D}. \tag{4.26}$$

Both the exact results (4.20), (4.22) and the estimates (4.26) corroborate the observations made on figure 3. The mean value  $A_D$  of  $X$  slowly increases with  $D$ , whereas its variance  $B_D$  shrinks. Both effects occur concomitantly at a logarithmically slow pace.

The formulas (4.26) are the leading terms of systematic asymptotic expansions in inverse powers of  $\ln D$  (see Appendix A for a self-contained presentation, as well as [44], and [45, 46] for related matters). The expansions (A.13) yield

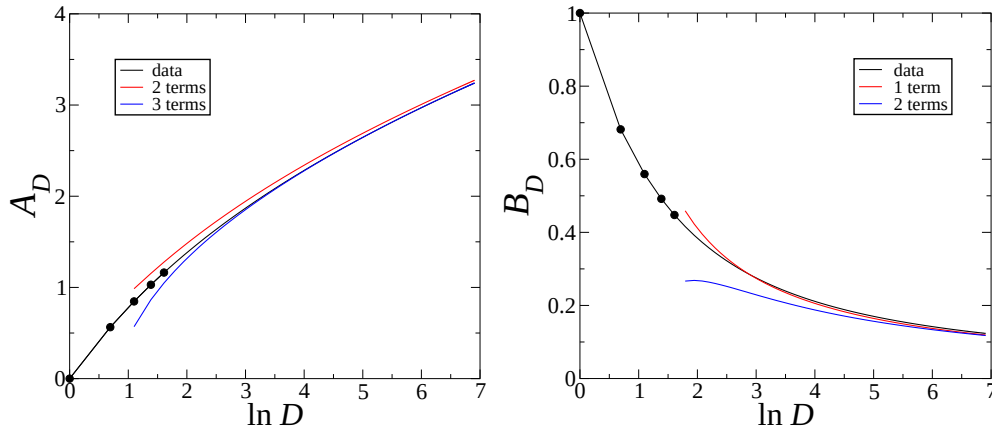
$$\begin{aligned}
A_D &= \sqrt{2 \ln D} \left( 1 + \frac{\nu}{2 \ln D} - \frac{6(\nu^2 + 2\nu + 2) + \pi^2}{48(\ln D)^2} \right. \\
&\quad \left. + \frac{6\nu^3 + 24\nu^2 + 3(16 + \pi^2)\nu + 42 + 4\pi^2 + 12\zeta(3)}{96(\ln D)^3} + \dots \right), \\
B_D &= \frac{\pi^2}{12 \ln D} - \frac{\pi^2(\nu + 1) + 6\zeta(3)}{12(\ln D)^2} + \dots,
\end{aligned} \tag{4.27}$$

with

$$\nu = \gamma - \frac{1}{2} \ln(4\pi \ln D), \tag{4.28}$$

where  $\gamma$  is again Euler's constant and  $\zeta$  is Riemann's zeta function.

Figure 4 shows the amplitudes  $A_D$  and  $B_D$  against  $\ln D$ . Symbols show the exactly known results (4.20), (4.22). Black curves show the outcomes of a numerical evaluation of the integrals entering (4.13), up to  $D = 1000$ . Red and blue curves show various truncations of the asymptotic expansions (4.27) (see legends). The amplitude  $A_D$  is approached smoothly by the successive approximations, whereas the situation is more intricate in the case of  $B_D$ : data cross the leading term of the asymptotic expansion for  $D = 19$ , while the second term of that expansion changes sign for  $D = 3116$ , beyond the plotted range.



**Figure 4.** The amplitudes  $A_D$  (left) and  $B_D$  (right) against  $\ln D$ . Symbols: exactly known results (4.20), (4.22). Black curves: numerical results up to  $D = 1000$ . Red and blue curves: various truncations of the asymptotic expansions (4.27) (see legends).

## 5. Summary

In this work we have introduced and studied multidimensional record patterns. These patterns are random sets of lattice points defined by means of a backward recursive construction generalizing a not-so-well-known but simple and efficient construction of records in sequences of independent random variables. The multidimensional construction is elegant in several regards, as it is parameter-free and respects all symmetries of the underlying lattice. It was inspired by recent studies on fragmentation processes of rectangles drawn on the square lattice, where a rectangle may break either vertically or horizontally or simultaneously in both directions [22, 23]. The patterns thus generated owe their richness to the fact that the multidimensional recursive construction is not based on a total order, except in the one-dimensional situation, where usual records are recovered.

Many exact results have been obtained on the universal combinatorics of multidimensional record patterns on finite samples drawn on hypercubic lattices in any dimension  $D \geq 2$ , generalizing thus standard results of the theory of records. One remarkable property of the model is that the number  $N$  of records on a sample of size  $L_1 \times \dots \times L_D$  is exactly distributed as the smallest of the numbers  $N_i$  of usual records in independent signals whose lengths  $L_i$  are the sides of the sample. The most

detailed analysis concerns the two-dimensional situation. We have also investigated the statistics of the landing position  $K$ , which is nothing but the length of the stick adjacent to the origin in the symmetric fragmentation process. The distribution of  $K$  derived here has many common features with the length distribution of generic sticks in the various fragmentation models investigated in [23], including multifractal spectra with very similar analytical forms.

In the higher-dimensional situation, the main focus has been on asymptotic expressions for quantities of interest on large hypercubic samples, including the full distribution and the moments of the number of records. The latter distribution is related to that of the largest of  $D$  standard Gaussian variables. Its mean and variance read  $\langle N \rangle \approx \ln L - A_D \sqrt{\ln L}$  and  $\langle N^2 \rangle \approx B_D \ln L$  (see (4.12)). The exact values of the amplitudes  $A_D$  and  $B_D$  have been obtained up to  $D = 5$  (see (4.20), (4.22)), whereas asymptotic large- $D$  expansions have also been derived (see (4.27)).

### Acknowledgments

It is a pleasure to thank Eli Ben-Naim and Sanjay Ramassamy for fruitful discussions, and Alexander Selby for instructive correspondence.

### Appendix A. Asymptotic expansions at large dimensions

The goal of this appendix is to show that the amplitudes  $A_D$ ,  $B_D$  and  $C_D$  introduced in (4.13) admit systematic asymptotic expansions in inverse powers of  $\ln D$ .

The leading-order scaling analysis done in section 4.3 shows that the probability density  $f_D(X)$  of the largest of  $D$  standard Gaussian variables gives the highest weight to large positive values of  $X$ , such that the combination  $\xi$  defined in (4.24) is finite.

This observation can be turned to a systematic analysis as follows. Setting

$$\varepsilon(X) = 1 - \Phi(X) = \frac{1}{2} \operatorname{erfc} \frac{X}{\sqrt{2}}, \quad (\text{A.1})$$

with  $\operatorname{erfc}$  being the complementary error function, the definitions (4.13) translate to

$$\begin{aligned} A_D &= \int_0^1 D(1 - \varepsilon)^{D-1} X(\varepsilon) d\varepsilon, \\ C_D &= \int_0^1 D(1 - \varepsilon)^{D-1} X(\varepsilon)^2 d\varepsilon, \end{aligned} \quad (\text{A.2})$$

where  $X(\varepsilon)$  is the inverse function of  $\varepsilon(X)$ .

The first step consists in deriving the asymptotic expansion of  $\varepsilon(X)$  at large  $X$  (see e.g. [47, 48]). Setting

$$\varepsilon(X) = f(X)g(X) = \frac{e^{-X^2/2}}{\sqrt{2\pi}} g(X), \quad (\text{A.3})$$

with  $f(X) = \Phi'(X) = -\varepsilon'(X)$  (see (4.7)), the second factor obeys the differential equation

$$Xg(X) - g'(X) = 1, \quad (\text{A.4})$$

yielding the asymptotic expansion

$$g(X) = \frac{1}{X} \sum_{n \geq 0} \frac{(2n)!}{n!} \left( -\frac{1}{2X^2} \right)^n. \quad (\text{A.5})$$

The second step consists in deriving the asymptotic expansion of  $X(\varepsilon)$  at small  $\varepsilon$ . From now on we use the shorthand notation  $X$  for  $X(\varepsilon)$ . Setting

$$\lambda = \ln \frac{1}{\varepsilon}, \quad \mu = -\frac{1}{2} \ln(4\pi\lambda), \quad (\text{A.6})$$

Equation (A.3) can be recast as

$$\frac{X^2}{2} = \lambda + \mu - \frac{1}{2} \ln \frac{X^2}{2\lambda} + \ln(Xg(X)). \quad (\text{A.7})$$

Solving this implicit equation iteratively leads to the expansion

$$\begin{aligned} \frac{X^2}{2} = & \lambda + \mu - \frac{\mu + 1}{2\lambda} + \frac{2\mu^2 + 6\mu + 7}{8\lambda^2} \\ & - \frac{8\mu^3 + 42\mu^2 + 102\mu + 107}{48\lambda^3} + \dots, \end{aligned} \quad (\text{A.8})$$

whose structure appears clearly: the term of order  $n$  in  $1/\lambda$  is a polynomial of degree  $n$  in  $\mu$  with rational coefficients.

The third step consists in inserting the expansion (A.8) into the integrals (A.2), after having consistently simplified them to

$$A_D \approx \int_0^\infty D e^{-D\varepsilon} X \, d\varepsilon, \quad C_D \approx \int_0^\infty D e^{-D\varepsilon} X^2 \, d\varepsilon. \quad (\text{A.9})$$

The latter estimates hold up to negligible corrections scaling as  $1/D$ , i.e., exponentially small in  $\lambda$ . Taking the product  $y = D\varepsilon$  as integration variable, we have  $\lambda = \ln D - \ln y$ , with  $\ln D$  being large and  $\ln y$  finite. Using (A.8) to expand  $X$  and  $X^2$  in inverse powers of  $\ln D$ , at every order the integrals over  $y$  boil down to the numerical constants

$$c_n = \int_0^\infty e^{-y} (\ln y)^n \, dy, \quad (\text{A.10})$$

which can be extracted from the following expression of their generating function

$$\sum_{n \geq 0} \frac{c_n s^n}{n!} = \int_0^\infty e^{-y} y^s \, dy = \Gamma(s+1), \quad (\text{A.11})$$

hence

$$\begin{aligned} c_1 &= -\gamma, \\ c_2 &= \gamma^2 + \frac{\pi^2}{6}, \\ c_3 &= -\gamma^3 - \frac{\gamma\pi^2}{2} - 2\zeta(3), \end{aligned} \quad (\text{A.12})$$

and so on, with  $\gamma$  being Euler's constant and  $\zeta$  Riemann's zeta function.

Putting everything together, we obtain the expansions

$$\begin{aligned} A_D &= \sqrt{2 \ln D} \left( 1 + \frac{\nu}{2 \ln D} - \frac{6(\nu^2 + 2\nu + 2) + \pi^2}{48(\ln D)^2} \right. \\ &\quad \left. + \frac{6\nu^3 + 24\nu^2 + 3(16 + \pi^2)\nu + 42 + 4\pi^2 + 12\zeta(3)}{96(\ln D)^3} + \dots \right), \\ C_D &= 2 \ln D \left( 1 + \frac{\nu}{\ln D} - \frac{\nu + 1}{2(\ln D)^2} + \frac{6\nu^2 + 18\nu + 21 + \pi^2}{24(\ln D)^3} + \dots \right). \end{aligned} \quad (\text{A.13})$$

The numerators of the successive terms are polynomials with increasing degrees in the doubly logarithmic variable

$$\nu = \gamma - \frac{1}{2} \ln(4\pi \ln D). \quad (\text{A.14})$$

## References

- [1] Chandler K N 1952 *J. Roy. Statist. Soc. B* **14** 220
- [2] Rényi A 1962 Théorie des éléments saillants d'une suite d'observations *Proceedings Coll. Combinatorial Methods in Probability Theory* (Aarhus: Math. Inst. Aarhus Univ.) p 104
- [3] Rényi A 1962 *Ann. Sci. Univ. Clermont-Ferrand* **8** 7
- [4] Glick N 1978 *Am. Math. Mon.* **85** 2
- [5] Arnold B C, Balakrishnan N and Nagaraja H N 1998 *Records* (New York: Wiley)
- [6] Nevzorov V B 2001 *Records: Mathematical Theory (Translation of Mathematical Monographs vol 194)* (Providence, RI: American Mathematical Society)
- [7] Schmittmann B and Zia R K P 1999 *Am. J. Phys.* **67** 1269
- [8] Bunge J and Goldie C M 2001 *Handbook of Statistics* **19** 277
- [9] Wergen G 2013 *J. Phys. A* **46** 223001
- [10] Godrèche C, Majumdar S N and Schehr G 2017 *J. Phys. A* **50** 333001
- [11] Krapivsky P L and Redner S 2002 *Phys. Rev. Lett.* **89** 258703
- [12] Krug J 2007 *J. Stat. Mech.* P07001
- [13] Majumdar S N and Ziff R M 2008 *Phys. Rev. Lett.* **101** 050601
- [14] Godrèche C and Luck J M 2008 *J. Stat. Mech.* P11006
- [15] Majumdar S N, Mallick K and Sabhapandit S 2009 *Phys. Rev. E* **79** 021109
- [16] Le Doussal P and Wiese K J 2009 *Phys. Rev. E* **79** 051105
- [17] Wergen G, Bogner M and Krug J 2011 *Phys. Rev. E* **83** 051109
- [18] Majumdar S N, Schehr G and Wergen G 2012 *J. Phys. A* **45** 355002
- [19] Godrèche C, Majumdar S N and Schehr G 2015 *J. Stat. Mech.* P07026
- [20] Luck J M and Mehta A 2015 *Phys. Rev. E* **92** 052810
- [21] Krapivsky P L and Luck J M 2019 *J. Stat. Mech.* 063209
- [22] Torrents G, Illa X, Vives E and Planes A 2017 *Phys. Rev. E* **95** 013001
- [23] Ben-Naim E and Krapivsky P 2019 *Phys. Rev. E* **100** 032122
- [24] Cheng Z and Redner S 1988 *Phys. Rev. Lett.* **60** 2450
- [25] Tarjus G and Viot P 1991 *Phys. Rev. Lett.* **67** 1875
- [26] Rodgers G J and Hassan M K 1994 *Phys. Rev. E* **50** 3458
- [27] Krapivsky P L and Ben-Naim E 1994 *Phys. Rev. E* **50** 3502
- [28] Boyer D, Tarjus G and Viot P 1995 *Phys. Rev. E* **51** 1043
- [29] Boyer D, Tarjus G and Viot P 1997 *J. Phys. I (France)* **7** 13
- [30] Ben-Naim E and Krapivsky P L 1997 *Physica D* **107** 156
- [31] Crosby K M and Bradley R M 1997 *Phys. Rev. E* **55** 6084
- [32] Ball J M, Cesena P and Hambly B 2015 *MATEC Web Conf.* **33** 02008
- [33] Krapivsky P L, Redner S and Ben-Naim E 2010 *A Kinetic View of Statistical Physics* (Cambridge: Cambridge University Press)
- [34] Clay A and Rolfsen D 2016 *Ordered Groups and Topology* vol 176 Graduate Studies in Mathematics (Providence, RI: American Mathematical Society)
- [35] Stirling J 1730 *Methodus Differentialis* (London: Bowyer)
- [36] Knuth D E 1968 *The Art of Computer Programming* (New York: Addison-Wesley)
- [37] Graham R L, Knuth D E and Patashnik O 1989 *Concrete Mathematics: A Foundation for Computer Science* (Reading, MA: Addison-Wesley)
- [38] Flajolet P and Sedgewick R 2009 *Analytic Combinatorics* (Cambridge: Cambridge University Press)
- [39] Corominas-Murtra B, Hanel R and Thurner S 2015 *Proc. Nat. Acad. Sci. USA* **112** 5348
- [40] Yadav A C 2016 *Phys. Rev. E* **93** 042131
- [41] Paladin G and Vulpiani A 1987 *Phys. Rep.* **156** 147
- [42] Stanley H E and Meakin P 1988 *Nature* **335** 405
- [43] Selby A 2013 Expected value for maximum of  $n$  normal random variables. Mathematics Stack Exchange URL <https://math.stackexchange.com/q/510580>
- [44] Cramér H 1946 *Mathematical Methods of Statistics* (Princeton: Princeton University Press)
- [45] Nair K A 1981 *Ann. Prob.* **9** 150
- [46] Nadarajah S 2015 *Stat. and Prob. Lett.* **103** 127
- [47] Gradshteyn I S and Ryzhik I M 1965 *Table of Integrals, Series, and Products* (New York: Academic)
- [48] Abramowitz M and Stegun I 1965 *Handbook of Mathematical Functions with Formulas* (New York: Dover)