Electronic Journal of Probability

Vol. 15 (2010), Paper no. 60, pages 1863–1892.

Journal URL
http://www.math.washington.edu/~ejpecp/

Multivariate records based on dominance

Hsien-Kuei Hwang and Tsung-Hsi Tsai
Institute of Statistical Science
Academia Sinica
Taipei 115
Taiwan

Abstract

We consider three types of multivariate records in this paper and derive the mean and the variance of their numbers for independent and uniform random samples from two prototype regions: hypercubes $[0,1]^d$ and d-dimensional simplex. Central limit theorems with convergence rates are established when the variance tends to infinity. Effective numerical procedures are also provided for computing the variance constants to high degree of precision .

Key words: Multivariate records, Pareto optimality, central limit theorems, Berry-Esseen bound, partial orders, dominance.

AMS 2000 Subject Classification: Primary 60C05; 60F05; 60G70.

Submitted to EJP on March 30, 2010, final version accepted October 18, 2010.

1 Introduction

While the one-dimensional records (or record-breakings, left-to-right maxima, outstanding elements, etc.) of a given sample have been the subject of research and development for more than six decades, considerably less is known for multidimensional records. One simple reason being that there is no total ordering for multivariate data, implying no unique way of defining records in higher dimensions. We study in this paper the stochastic properties of three types of records based on the dominance relation under two representative prototype models. In particular, central limit theorems with convergence rates are proved for the number of multivariate records when the variance tends to infinity, the major difficulty being the asymptotics of the variance.

Dominance and maxima. A point $\mathbf{p} \in \mathbb{R}^d$ is said to *dominate* another point $\mathbf{q} \in \mathbb{R}^d$ if $\mathbf{p}-\mathbf{q}$ has only positive coordinates, where the dimensionality $d \geq 1$. Write $\mathbf{q} \prec \mathbf{p}$ or $\mathbf{p} \succ \mathbf{q}$. The nondominated points in the set $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ are called *maxima*. Maxima represent one of the most natural and widely used partial orders for multidimensional samples when $d \geq 2$, and have been thoroughly investigated in the literature under many different guises and names (such as admissibility, Pareto optimality, elites, efficiency, skylines, ...); see [1, 5] and the references therein.

Pareto records. A point $\mathbf{p}_k$ is defined to be a *Pareto record* or a *nondominated record* of the sequence $\mathbf{p}_1, \dots, \mathbf{p}_n$ if

$$\mathbf{p}_k \not\prec \mathbf{p}_i \text{ for all } 1 \leq i < k.$$

Such a record is referred to as a *weak record* in [17], but we found this term less informative.

In addition to being one of the natural extensions of the classical one-dimensional records, the Pareto records of a sequence of points are also closely connected to maxima, the simplest connection being the following bijection. If we consider the indices of the points as an additional coordinate, then the Pareto records are exactly the maxima in the extended space (the original one and the index-set) by reversing the order of the indices. Conversely, if we sort a set of points according to a fixed coordinate and use the ranks as the indices, then the maxima are nothing but the Pareto records in the induced space (with one dimension less); see [17]. See also the recent paper [5] for the algorithmic aspects of such connections.

More precisely, assume that $\mathbf{p}_1, \dots, \mathbf{p}_n$ are independently and uniformly distributed (*abbreviated as iud*) in a specified region S and $\mathbf{q}_1, \dots, \mathbf{q}_n$ are iud in the region $S \times [0, 1]$. Then the distribution of the number of Pareto records of the sequence $\mathbf{p}_1, \dots, \mathbf{p}_n$ is equal to the distribution of the number of maxima of the set $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$. This connection will be used later in our analysis.

On the other hand, we also have, for any given regions, the following relation between the expected number $\mathbb{E}[X_n]$ of Pareto records and the expected number $\mathbb{E}[M_n]$ of maxima of the same sample of points, say $\mathbf{p}_1, \dots, \mathbf{p}_n$,

$$\mathbb{E}[X_n] = \sum_{1 \leq k \leq n} \frac{\mathbb{E}[M_k]}{k};$$

see [5].

Dominating records. Although the Pareto records are closely connected to maxima, their probabilistic properties have been less well studied in the literature. In contrast, the following definition of records has received more attention.

A point $\mathbf{p}_k$ is defined to be a *dominating record* of the sequence $\mathbf{p}_1, \dots, \mathbf{p}_n$ if

$$\mathbf{p}_i \prec \mathbf{p}_k \text{ for all } 1 \le i < k.$$

This is referred to as the *strong record* in [17] and the *multiple maxima* in [21].

Let the number of dominating records falling in $A \subset S$ be denoted by Z_A. Goldie and Resnick [18] showed that

$$\mathbb{E}[Z_A] = \int_A \big(1 - \mu(D_\mathbf{x})\big)^{-1} \, \mathrm{d}\mu(\mathbf{x}),$$

where $D_\mathbf{x} = \{\mathbf{y} : \mathbf{y} \prec \mathbf{x}\}$. They also calculated all the moments of Z_A and derived several other results such as the probability of the event $\{Z_A = 0\}$ and the covariance $\mathrm{Cov}\big(Z_A, Z_B\big)$.

In the special case when the $\mathbf{p}_i$'s are iid with a common multivariate normal (non-degenerate) distribution, Gnedin [16] proved that

$$\lambda_n := \mathbb{P}\{\mathbf{p}_n \text{ is a dominating record}\} \asymp n^{-\alpha}(\log n)^{(\alpha-\beta)/2}.$$

for some explicitly computable $\alpha > 1$ and $\beta \in \{2, 3, \dots, d\}$. See also [20] for finer asymptotic estimates.

Chain records. Yet another type of record of multi-dimensional samples introduced in [17] is the *chain record*

$$\mathbf{p}_1 \prec \mathbf{p}_{i_1} \prec \mathbf{p}_{i_2} \prec \cdots \prec \mathbf{p}_{i_k},$$

where $1 < i_1 < i_2 < \cdots < i_k$ and there are no $\mathbf{p}_j \succ \mathbf{p}_{i_a}$ with $i_a < j < i_{a+1}$ or $i_a < j \le n$. See Figure 1 for an illustration of the three different types of records.

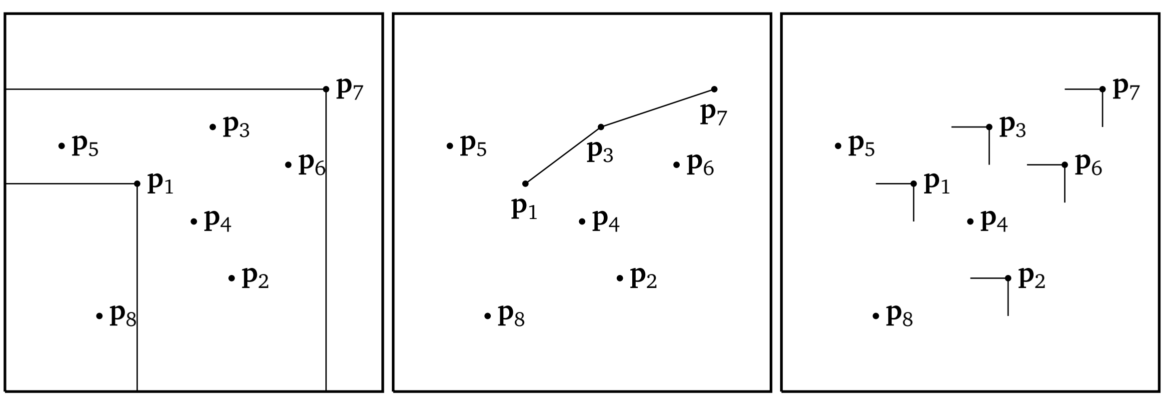

Figure 1: *In this simple example, the dominating records are* $\mathbf{p}_1$ *and* $\mathbf{p}_7$ *(left), the chain records are* $\mathbf{p}_1$, $\mathbf{p}_3$ *and* $\mathbf{p}_7$ *(middle), and the Pareto records are* $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_6$ *and* $\mathbf{p}_7$ *(right), respectively.*

Some known results and comparisons. If we drop the restriction of order, then the largest subset of indices such that

$$\mathbf{p}_{i_1} \prec \mathbf{p}_{i_2} \prec \cdots \prec \mathbf{p}_{i_k} \tag{1}$$

is equal to the number of maximal layers (maxima being regarded as the first layer, the maxima of the remaining points being the second, and so on). Assuming that $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ are iud in the hypercube $[0,1]^d$, Gnedin [17] proved that the number of chain records Y_n is asymptotically Gaussian with mean and variance asymptotic to

$$\mathbb{E}[Y_n] \sim d^{-1} \log n, \quad \mathbb{V}[Y_n] \sim d^{-2} \log n;$$

see Theorem 4 for an improvement. The author also derived exact and asymptotic formulas for the probability of a chain record $\mathbb{P}(Y_n > Y_{n-1})$ and discussed some point-process scaling limits.

The behavior of the record sequence (1) in $\mathbb{R}^2$ are studied in Goldie and Resnick [19], Deuschel and Zeitouni [10]. The position of the points converges in probability to a (or a set of) deterministic curve(s). Deuschel and Zeitouni [10] also proved a weak law of large number for the longest increasing subsequence, extending a result by Vershik and Kerov [24] to a non-uniform setting; see also the breakthrough paper [3]. A completely different type of multivariate records based on convex hulls was discussed in [23].

Chain records can in some sense be regarded as uni-directional Pareto records, and thus lacks the multi-directional feature of Pareto records. The asymptotic analysis of the moments is in general simpler than that for the Pareto records. On the other hand, it is also this aspect that the chain records reflect better the properties exhibited by the one-dimensional records. Interestingly, the chain records correspond to the "left-arm" (starting from the root by always choosing the subtree corresponding to the first quadrant) of quadtrees; see [6, 11, 13] and the references therein.

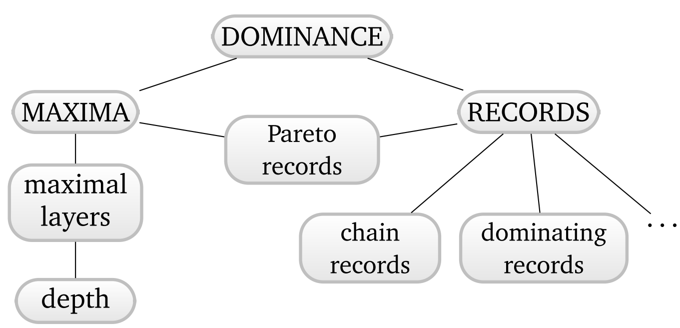

Figure 2: *A diagram illustrating the diverse notions defined on dominance; in particular, the Pareto records can be regarded as a good bridge between maxima and multivariate records.*

A summary of results. We consider in the paper the distributional aspect of the above three types of records in two typical cases when the $\mathbf{p}_i$'s are iud in the hypercube $[0,1]^d$ and in the d-dimensional simplex, respectively. Briefly, hypercubes correspond to situations when the coordinates are independent, while the d-dimensional simplex to that when the coordinates are to some extent negatively correlated. The hypercube case has already been studied in [17]; we will discuss this briefly by a very different approach. In addition to the asymptotic normality for the number of

Pareto records in the d-dimensional simplex, our main results are summarized in the following table, where we list the asymptotics of the mean (first entry) and the variance (second entry) in each case.

Records \ Models	Hypercube $[0,1]^d$	d-dimensional simplex
Dominating records	$H_n^{(d)}, H_n^{(d)} - H_n^{(2d)}$	(15), (16)
Chain records	$\frac{1}{d}\log n, \frac{1}{d^2}\log n$ [17]	$\frac{1}{dH_d}\log n, \frac{H_d^{(2)}}{dH_d^3}\log n$
Pareto records	$\frac{1}{d!}(\log n)^d, \left(\frac{1}{d!}+\kappa_{d+1}\right)(\log n)^d$ [17]	$m_d n^{(d-1)/d}$, $v_d n^{(d-1)/d}$
Maxima	$=$ Pareto records in $[0,1]^{d-1}$ [17]	$\tilde{m}_d n^{(d-1)/d}$, $\tilde{v}_d n^{(d-1)/d}$

Here $H_b^{(a)} = \sum_{i=1}^b i^{-a}$, κ_d is a constant (see [1]), $m_d := \frac{d}{d-1}\Gamma\left(\frac{1}{d}\right)$, v_d is defined in (3), $\tilde{m}_d := \Gamma(\frac{1}{d})$, $\tilde{v}_d$ is given in (4), and both (15) and (16) are bounded in n and in d; see Figure 3.

From this table, we see clearly that the three types of records behave very differently, although they coincide when $d = 1$. Roughly, the number of dominating records is bounded (indeed less than two on average) in both models, while the chain records have a typical logarithmic quantity; and it is the Pareto records that reflect better the variations of the underlying models.

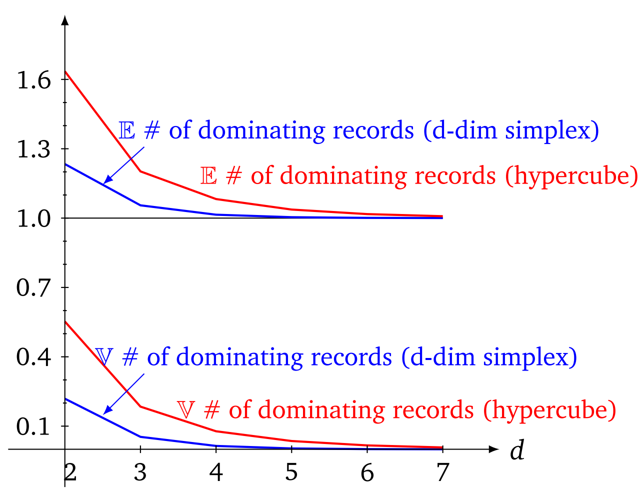

Figure 3: *The mean and the variance of the number of dominating records in low dimensional random samples. In each model, the expected number approaches* 1 *very fast as d increases with the corresponding variance tending to zero.*

Organization of the paper. We derive asymptotic approximations to the mean and the variance for the number of Pareto records in the next section. Since the expression for the leading coefficient of the asymptotic variance is very messy, we then address in Section 3 the numerical aspect of this constant. The tools we used turn out to be also useful for several other constants of similar nature, which we briefly discuss. We then discuss the chain records and the dominating records.

2 Asymptotics of the number of Pareto records

Assume $d \geq 2$ throughout this paper. Let

$$S_d := \{\mathbf{x} : x_i \geq 0 \text{ and } 0 \leq \|\mathbf{x}\| \leq 1\}$$

denote the d-dimensional simplex, where $\|\mathbf{x}\| := x_1 + \cdots + x_d$. Assume that $\mathbf{p}_1, \ldots, \mathbf{p}_n$ are iud in S_d. Let X_n denote the number of Pareto records of $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$. We derive in this section asymptotic approximations to the mean and the variance and a Berry-Esseen bound for X_n. The same method of proof also applies to the number of maxima, denoted by M_n, which we will briefly discuss.

Let $\mathbf{q}_1, \ldots, \mathbf{q}_n$ be iud in $S_d \times [0,1]$. As discussed in Introduction, the distribution of X_n is equivalent to the distribution of the number of maxima of $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$.

For notational convenience, denote by $a_n \simeq b_n$ if $a_n = b_n + O\left(n^{-1/d}\right)$.

Theorem 1. *The mean and the variance of the number of Pareto maxima X_n in random samples from the d-dimensional simplex satisfy*

$$\begin{aligned}\mathbb{E}[X_n] \simeq n^{1-1/d} &\sum_{0\leq j\leq d-2} \binom{d-1}{j}(-1)^j \Gamma\left(\frac{j+1}{d}\right)\frac{d}{d-1-j}\, n^{-j/d} \\ &+ (-1)^{d-1}\left(\log n + \gamma\right), \\ \mathbb{V}[X_n] = \left(v_d + o(1)\right) &n^{1-1/d},\end{aligned} \tag{2}$$

where

$$\begin{aligned} v_d := \frac{d}{d-1}\Gamma\left(\frac{1}{d}\right) &+ 2d^2(d-1)\sum_{1\leq \ell<d}\binom{d}{\ell}\binom{d-2}{\ell-1} \\ &\times \int_0^1\int_v^1\int_0^\infty\int_0^\infty\int_0^\infty y^{d-\ell-1}w^{\ell-1}e^{-u\left(x+y\right)^d - v(x+w)^d}\left(e^{vx^d}-1\right) \mathrm{d}w\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v \\ &+ 2d^2\int_0^1\int_v^1\int_0^\infty\int_0^\infty w^{d-1}e^{-ux^d - v(x+w)^d}\left(e^{vx^d}-1\right)\mathrm{d}w\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v \\ &- 2d^2\int_0^1\int_v^1\int_0^\infty\int_0^\infty y^{d-1}e^{-u\left(x+y\right)^d - vx^d}\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v. \end{aligned} \tag{3}$$

Proof. The method of proof is similar to that given in [1], but the technicalities are more involved.

We start with the expected value of X_n. Let $G_i = 1_{\{\mathbf{q}_i \text{ is a maximum}\}}$.

$$
\begin{aligned}
\mathbb{E}[X_n] &= n\mathbb{E}[G_1] \\
&= d!n \int_0^1 \int_{S_d} \left(1 - z(1-\|\mathbf{x}\|)^d\right)^{n-1} \mathrm{d}\mathbf{x}\,\mathrm{d}z \\
&\simeq d!n \int_0^1 \int_{S_d} e^{-nz(1-\|\mathbf{x}\|)^d} \mathrm{d}\mathbf{x}\,\mathrm{d}z \\
&= dn \int_0^1 \int_0^1 e^{-nz(1-y)^d} y^{d-1}\,\mathrm{d}y\,\mathrm{d}z \quad (y \mapsto \|\mathbf{x}\|) \\
&= dn \sum_{0\le j<d} \binom{d-1}{j} (-1)^j \int_0^1 \int_0^1 e^{-nzy^d} y^j\,\mathrm{d}y\,\mathrm{d}z \\
&= \sum_{0\le j<d} n^{(d-1-j)/d} \binom{d-1}{j} (-1)^j \int_0^1 \int_0^{nz} e^{-x} x^{(1+j-d)/d} z^{-(j+1)/d}\,\mathrm{d}x\,\mathrm{d}z \\
&\simeq \sum_{0\le j\le d-2} \binom{d-1}{j} (-1)^j \Gamma\left(\frac{j+1}{d}\right) \frac{d}{d-1-j} n^{(d-1-j)/d} \\
&\qquad + (-1)^{d-1}\left(\log n + \gamma\right).
\end{aligned}
$$

This proves (2).

For the variance, we start from the second moment, which is given by

$$
\mathbb{E}\left[X_n^2\right] = \mathbb{E}[X_n] + n(n-1)\mathbb{E}\left[G_1 G_2\right].
$$

Let $\mathbf{A}$ be the region in $\mathbb{R}^d \times [0,1]$ such that $\mathbf{q}_1$ and $\mathbf{q}_2$ are incomparable (neither dominating the other). Write $\mathbf{q}_1 = (\mathbf{x}, u)$, $\mathbf{q}_2 = (\mathbf{y}, v)$, $\|\mathbf{x}\|_* := (\|\mathbf{x}\| \wedge 1)$ and

$$
\mathbf{x} \vee \mathbf{y} := \left(x_1 \vee y_1, \cdots, x_d \vee y_d\right).
$$

Then by standard majorization techniques (see [1])

$$
\begin{aligned}
&n(n-1)\mathbb{E}\left[G_1 G_2\right] \\
&= n(n-1)d!^2 \int_{\mathbf{A}} \left(1 - u(1-\|\mathbf{x}\|)^d - v(1-\left\|\mathbf{y}\right\|)^d + (u\wedge v)(1-\left\|\mathbf{x}\vee\mathbf{y}\right\|_*)^d\right)^{n-2} \mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\,\mathrm{d}u\,\mathrm{d}v \\
&\simeq n^2 d!^2 \int_{\mathbf{A}} e^{-n[u(1-\|\mathbf{x}\|)^d + v(1-\left\|\mathbf{y}\right\|)^d]} \mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\,\mathrm{d}u\,\mathrm{d}v \\
&\qquad + n^2 d!^2 \int_{\mathbf{A}} e^{-n[u(1-\|\mathbf{x}\|)^d + v(1-\left\|\mathbf{y}\right\|)^d]} \left(e^{n(u\wedge v)(1-\left\|\mathbf{x}\vee\mathbf{y}\right\|_*)^d} - 1\right) \mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}\,\mathrm{d}u\,\mathrm{d}v \\
&\simeq \mathbb{E}\left[X_n^2\right] - J_{n,0} + \sum_{1\le \ell<d} \binom{d}{\ell} J_{n,\ell} + J_{n,d},
\end{aligned}
$$

where

$$J_{n,0} = 2n^2 d!^2 \int_0^1 \int_v^1 \int_{\substack{\mathbf{x} \prec \mathbf{y} \\ \mathbf{x},\mathbf{y} \in S_d}} e^{-n[u(1-\|\mathbf{x}\|)^d + v(1-\|\mathbf{y}\|)^d]} \, \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} \, \mathrm{d}u \, \mathrm{d}v,$$

$$J_{n,\ell} = n^2 d!^2 \int_0^1 \int_0^1 \int_{\substack{x_i > y_i, 1 \le i \le \ell \\ x_i < y_i, \ell < i \le d \\ \mathbf{x},\mathbf{y} \in S_d}} e^{-n[u(1-\|\mathbf{x}\|)^d + v(1-\| \mathbf{y}\|)^d]} \left(e^{n(u \wedge v)(1-\|\mathbf{x} \vee \mathbf{y}\|_*)^d} - 1 \right) \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} \, \mathrm{d}u \, \mathrm{d}v,$$

$$J_{n,d} = 2n^2 d!^2 \int_0^1 \int_v^1 \int_{\substack{\mathbf{y} \prec \mathbf{x} \\ \mathbf{x},\mathbf{y} \in S_d}} e^{-n[u(1-\|\mathbf{x}\|)^d + v(1-\|\mathbf{y}\|)^d]} \left(e^{n(u \wedge v)(1-\|\mathbf{x} \vee \mathbf{y}\|_*)^d} - 1 \right) \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{y} \, \mathrm{d}u \, \mathrm{d}z,$$

for $1 \le \ell \le d-1$.

Consider first $J_{n,\ell}$, $1 \le \ell < d$. We proceed by four sets of changes of variables to simplify the integral starting from

$$\begin{cases} x_i \mapsto \xi_i, \quad y_i \mapsto \xi_i(1-\eta_i), & \text{for } 1 \le i \le \ell; \\ x_i \mapsto \xi_i(1-\eta_i), \quad y_i \mapsto \xi_i, & \text{for } \ell < i \le d, \end{cases}$$

which leads to

$$J_{n,\ell} = (nd!)^2 \int_0^1 \int_0^1 \int_{S_d} \int_{[0,1]^d} e^{-n\left[u\left(1-\sum \xi_i + \sum'' \xi_i \eta_i\right)^d + v\left(1-\sum \xi_i + \sum' \xi_i \eta_i\right)^d\right]}$$
$$\times \left(e^{n(u \wedge v)\left(1-\sum \xi_i\right)^d} - 1 \right) \left(\prod \xi_i \right) \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{\eta} \, \mathrm{d}u \, \mathrm{d}v,$$

where $\sum \xi_i := \sum_{i=1}^d \xi_i$, $\prod \xi_i := \prod_{i=1}^d \xi_i$, $\sum' x_i := \sum_{i=1}^{\ell} x_i$ and $\sum'' x_i := \sum_{i=\ell+1}^d x_i$.

Next, by the change of variables

$$\xi_i \mapsto \frac{1}{d} - \xi_i n^{-1/d}, \quad \eta_i \mapsto d\eta_i n^{-1/d},$$

we have

$$J_{n,\ell} = d!^2 \int_0^1 \int_0^1 \int_{S_d(n)} \int_{[0,n^{1/d}/d]^d} e^{-\left[u\left(\sum \xi_i + \sum'' \eta_i\left(1-d\xi_i n^{-1/d}\right)\right)^d + v\left(\sum \xi_i + \sum' \eta_i\left(1-d\xi_i n^{-1/d}\right)\right)^d\right]}$$
$$\times \left(e^{(u \wedge v)\left(\sum \xi_i\right)^d} - 1 \right) \prod \left(1 - d\xi_i n^{-1/d}\right) \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{\eta} \, \mathrm{d}u \, \mathrm{d}v,$$

where $S_d(n) = \{\boldsymbol{\xi} : \xi_i \le n^{1/d}/d \text{ and } \|\boldsymbol{\xi}\| > 0\}$.

We then perform the change of variables

$$\eta_i \mapsto \eta_i \left(1 - d\xi_i n^{-1/d}\right),$$

and obtain

$$J_{n,\ell} = (d!)^2 \int_0^1 \int_0^1 \int_{S_d(n)} \int_{[0,n^{1/d}/d]^d} e^{-\left[u\left(\sum \xi_i + \sum'' \eta_i\right)^d + v\left(\sum \xi_i + \sum' \eta_i\right)^d\right]} \times \left(e^{(u\wedge v)\left(\sum \xi_i\right)^d} - 1\right) \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}u\, \mathrm{d}v.$$

Finally, we "linearize" the integrals by the change of variables

$$x \mapsto \sum \xi_i, \quad y \mapsto \sum'' \eta_i, \quad w \mapsto \sum' \eta_i,$$

and get

$$J_{n,\ell} \simeq \frac{d\cdot d!}{(d-\ell-1)!(\ell-1)!}\, n^{1-1/d} \int_0^1 \int_0^1 \int_0^\infty \int_0^\infty \int_0^\infty y^{d-\ell-1} w^{\ell-1} e^{-u\left(x+y\right)^d - v(x+w)^d} \times \left(e^{(u\wedge v)x^d} - 1\right) \mathrm{d}w\, \mathrm{d}y\, \mathrm{d}x\, \mathrm{d}u\, \mathrm{d}v,$$

since the change of variables produces the factors

$$\frac{n^{1-1/d}}{(d-1)!}, \frac{y^{d-\ell-1}}{(d-\ell-1)!} \text{ and } \frac{w^{\ell-1}}{(\ell-1)!}.$$

Now by symmetry (of u and v), we have

$$\sum_{1\le \ell < d} \binom{d}{\ell} J_{n,\ell} \simeq \sum_{1\le \ell<d} \binom{d}{\ell} \frac{2d\cdot d!}{(d-\ell-1)!(\ell-1)!}\, n^{1-1/d} \int_0^1 \int_v^1 \int_0^\infty \int_0^\infty \int_0^\infty y^{d-\ell-1} w^{\ell-1} \times e^{-u\left(x+y\right)^d - v(x+w)^d} \left(e^{vx^d} - 1\right) \mathrm{d}w\, \mathrm{d}y\, \mathrm{d}x\, \mathrm{d}u\, \mathrm{d}v.$$

Proceeding in a similar manner for $J_{n,d}$, we deduce that

$$J_{n,d} = 2d!^2 \int_0^1 \int_v^1 \int_{S_d(n)} \int_{[0,n^{1/d}/d]^d} e^{-\left[u\left(\sum \xi_i\right)^d + v\left(\sum \xi_i + \sum \eta_i\right)^d\right]} \left(e^{(u\wedge v)\left(\sum \xi_i\right)^d} - 1\right) \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}u\, \mathrm{d}v.$$

By the change of variables $x \mapsto \sum \xi_i$, $w \mapsto \sum \eta_i$, we have

$$\begin{aligned} J_{n,d} &\simeq \frac{2d!^2}{((d-1)!)^2}\, n^{1-1/d} \int_0^1 \int_v^1 \int_0^\infty \int_0^\infty w^{d-1} e^{-ux^d - v(x+w)^d} \left(e^{(u\wedge v)x^d} - 1\right) \mathrm{d}w\, \mathrm{d}x\, \mathrm{d}u\, \mathrm{d}v \\ &= 2d^2 n^{1-1/d} \int_0^1 \int_v^1 \int_0^\infty \int_0^\infty w^{d-1} e^{-ux^d - v(x+w)^d} \left(e^{vx^d} - 1\right) \mathrm{d}w\, \mathrm{d}x\, \mathrm{d}u\, \mathrm{d}v. \end{aligned}$$

Similarly, for $J_{n,0}$, we get

$$J_{n,0} = 2d!^2 \int_0^1 \int_v^1 \int_{S_d(n)} \int_{[0,n^{1/d}/d]^d} e^{-\left[u\left(\sum \xi_i + \sum \eta_i\right)^d + v\left(\sum \xi_i\right)^d\right]} \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}u\, \mathrm{d}v.$$

The change of variables $x \mapsto \sum \xi_i$, $y \mapsto \sum \eta_i$ then yields

$$J_{n,0} \simeq 2d^2 n^{1-1/d} \int_0^1 \int_v^1 \int_0^\infty \int_0^\infty y^{d-1} e^{-u\left(x+y\right)^d - vx^d}\, \mathrm{d}y\, \mathrm{d}x\, \mathrm{d}u\, \mathrm{d}v.$$

This completes the proof of the theorem. ∎

Remark. By the same arguments, we derive the following asymptotic estimates for the number of maxima in S_d.

$$\mathbb{E}[M_n] \simeq \sum_{0\le j<d} \binom{d-1}{j} (-1)^j \Gamma\left(\frac{j+1}{d}\right) n^{(d-1-j)/d},$$
$$\mathbb{V}[M_n] = \left(\tilde{v}_d + o(1)\right) n^{1-1/d},$$

where

$$\begin{aligned}\tilde{v}_d = \Gamma\left(\frac{1}{d}\right) &+ \sum_{1\le k<d} \binom{d}{k} \frac{d d!}{(d-k-1)!(k-1)!} \\ &\times \int_0^\infty \int_0^\infty \int_0^\infty y^{d-k-1} w^{k-1} e^{-(x+y)^d-(x+w)^d} \left(e^{x^d}-1\right) \mathrm{d}w\,\mathrm{d}y\,\mathrm{d}x \\ &- 2d^2 \int_0^\infty \int_0^\infty y^{d-1} e^{-x^d-(x+y)^d} \,\mathrm{d}x\,\mathrm{d}y.\end{aligned} \tag{4}$$

Theorem 2. *The number of Pareto records in iud samples from d-dimensional simplex is asymptotically normal with a rate given by*

$$\sup_x \left| \mathbb{P}\left(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathbb{V}[X_n]}} < x\right) - \Phi(x) \right| = O\left(n^{-(d-1)/(4d)} (\log n)^2 + n^{-1/d} (\log n)^{1/d}\right), \tag{5}$$

where $\Phi(x)$ denotes the standard normal distribution.

Proof. Define the region

$$D_n := \left\{(\mathbf{x}, z) : \mathbf{x} \in S_d, z \in [0,1] \text{ and } z\,(1-\|\mathbf{x}\|)^d \le \frac{2\log n}{n}\right\}.$$

Let $\overline{X}_n$ denote the number of maxima in D_n and $\widetilde{X}_n$ the number of maxima of a Poisson process on D_n with intensity $d!n$. Then

$$\begin{aligned}\left| \mathbb{P}\left(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathbb{V}[X_n]}} < x\right) - \Phi(x) \right| \le &\left| \mathbb{P}\left(\frac{X_n - \mathbb{E}[X_n]}{\sqrt{\mathbb{V}[X_n]}} < x\right) - \mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}[X_n]}{\sqrt{\mathbb{V}[X_n]}} < x\right) \right| \\ &+ \left| \mathbb{P}\left(\frac{\overline{X}_n - \mathbb{E}[X_n]}{\sqrt{\mathbb{V}[X_n]}} < x\right) - \mathbb{P}\left(\frac{\widetilde{X}_n - \mathbb{E}[X_n]}{\sqrt{\mathbb{V}[X_n]}} < x\right) \right| \\ &+ \left| \mathbb{P}\left(\frac{\widetilde{X}_n - \mathbb{E}[\widetilde{X}_n]}{\sqrt{\mathbb{V}[\widetilde{X}_n]}} < y\right) - \Phi(y) \right| + \left|\Phi(y) - \Phi(x)\right|,\end{aligned} \tag{6}$$

for $x \in \mathbb{R}$, where

$$y = x\sqrt{\frac{\mathbb{V}[X_n]}{\mathbb{V}[\widetilde{X}_n]}} + \frac{\mathbb{E}[X_n] - \mathbb{E}[\widetilde{X}_n]}{\sqrt{\mathbb{V}[\widetilde{X}_n]}}.$$

We prove that the four terms on the right-hand side of (6) all satisfy the O-bound in (5). For the first term, we consider the probability

$$\begin{aligned}
\mathbb{P}\left(X_n \neq \overline{X}_n\right) &\leq n\mathbb{P}\left(\mathbf{q}_1 \notin D_n \text{ and } \mathbf{q}_1 \text{is a maximum}\right) \\
&= nd! \int_{S_d \times [0,1] - D_n} \left(1 - z(1 - \|\mathbf{x}\|)^d\right)^{n-1} \mathrm{d}\mathbf{x}\, \mathrm{d}z \\
&\leq nd! \int_{S_d \times [0,1] - D_n} \left(1 - \frac{2\log n}{n}\right)^{n-1} \mathrm{d}\mathbf{x}\, \mathrm{d}z \\
&\leq O(n^{-1}).
\end{aligned}$$

For the second term on the right-hand side of (6), we use a Poisson process approximation

$$\sup_t \left|\mathbb{P}\left(\overline{X}_n < t\right) - \mathbb{P}\left(\widetilde{X}_n < t\right)\right| \leq O\left(\left|D_n\right|\right) = O\left(n^{-1/d}(\log n)^{1/d}\right).$$

To bound the third term, we use Stein's method similar to the proof for the case of hypercube given in [1] and deduce that

$$\begin{aligned}
\sup_y \left|\mathbb{P}\left(\frac{\widetilde{X}_n - \mathbb{E}[\widetilde{X}_n]}{\sqrt{\mathbb{V}[\widetilde{X}_n]}} < y\right) - \Phi(y)\right| &= O\left(\frac{(\mathbb{E}[\widetilde{X}_n])^{1/2} Q_n}{(\mathbb{V}[\widetilde{X}_n])^{3/4}}\right) \\
&= O\left(n^{-(d-1)/(4d)}\left(\log n\right)^2\right),
\end{aligned}$$

where Q_n is the error term resulted from the dependence between the cells decomposed and

$$Q_n = O\left((\log n)^2\right).$$

Finally, the last term in (6) is bounded above as follows.

$$\begin{aligned}
\left|\Phi(y) - \Phi(x)\right| &= O\left(\frac{\left|\sqrt{\mathbb{V}[X_n]} - \sqrt{\mathbb{V}[\widetilde{X}_n]}\right| + \left|\mathbb{E}[X_n] - \mathbb{E}[\widetilde{X}_n]\right|}{\sqrt{\mathbb{V}[\widetilde{X}_n]}}\right) \\
&= O\left(n^{-(d+1)/(2d)}\right).
\end{aligned}$$

This proves (6) and thus (5). ∎

The Berry-Esseen bound in (5) is expected to be non-optimal and one naturally anticipates an optimal order of the form $O(n^{-1/2-1/(2d)})$, but we were unable to prove this.

Remark. By defining

$$D_n := \left\{\mathbf{x} : \mathbf{x} \in S_d \text{ and } (1 - \|\mathbf{x}\|)^d \leq \frac{2\log n}{n}\right\}$$

instead and by applying the same arguments, we deduce the Berry-Esseen bound for the number of maxima in iud samples from S_d

$$\sup_x \left|\mathbb{P}\left(\frac{M_n - \mathbb{E}[M_n]}{\sqrt{\mathbb{V}[M_n]}} < x\right) - \Phi(x)\right| = O\left(n^{-(d-1)/(4d)}\log n + n^{-1/d}(\log n)^{1/d}\right).$$

3 Numerical evaluations of the leading constants

The leading constants v_d (see (3)) and $\tilde{v}_d$ (see (4)) appearing in the asymptotic approximations to the variance of X_n and to that of M_n are not easily computed via existing softwares. We discuss in this section more effective means of computing their numerical values to high degree of precision. Our approach is to first apply Mellin transforms (see [12]) and derive series representations for the integrals by standard residue calculation and then convert the series in terms of the generalized hypergeometric functions

$$ {}_pF_q(\alpha_1,\dots,\alpha_p;\beta_1,\dots,\beta_q;z) := \frac{\Gamma(\beta_1)\cdots\Gamma(\beta_q)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_p)} \sum_{j\ge 0} \frac{\Gamma(j+\alpha_1)\cdots\Gamma(j+\alpha_p)}{\Gamma(j+\beta_1)\cdots\Gamma(j+\beta_q)} \cdot \frac{z^j}{j!}. $$

The resulting linear combinations of hypergeometric functions can then be computed easily to high degree of precision by any existing symbolic softwares even with a mediocre laptop.

The leading constant v_d of the asymptotic variance of the d-dimensional Pareto records. We consider the following integrals

$$ \begin{aligned} C_d = &\sum_{1\le m<d} \binom{d}{m} \frac{(d-1)!}{(m-1)!(d-1-m)!} \\ &\times \int_0^1\int_v^1\int_0^\infty\int_0^\infty\int_0^\infty y^{d-1-m} w^{m-1} e^{-u(x+y)^d - v(x+w)^d} \left(e^{vx^d}-1\right) \mathrm{d}w\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v \\ &+ \int_0^1\int_v^1\int_0^\infty\int_0^\infty w^{d-1} e^{-ux^d - v(x+w)^d} \left(e^{vx^d}-1\right) \mathrm{d}w\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v \\ &- \int_0^1\int_v^1\int_0^\infty\int_0^\infty y^{d-1} e^{-u(x+y)^d - vx^d} \,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v \\ =:&\ (d-1)\sum_{1\le m<d} \binom{d}{m}\binom{d-2}{m-1} I_{d,m} + I_{d,d} - I_{d,0}. \end{aligned} $$

Then C_d is related to v_d by $v_d = \frac{d}{d-1}\Gamma(\frac{1}{d}) + 2d^2 C_d$. We start from the simplest one, $I_{d,0}$ and use the integral representation for the exponential function

$$ e^{-t} = \frac{1}{2\pi i}\int_{(c)} \Gamma(s) t^{-s}\,\mathrm{d}s, $$

where $c>0$, $\Re(t)>0$ and the integration path $\int_{(c)}$ is the vertical line from $c-i\infty$ to $c+i\infty$. Substituting this representation into $I_{d,0}$, we obtain

$$ I_{d,0} = \frac{1}{2\pi i}\int_{(c)} \Gamma(s) \int_0^1\int_v^1\int_0^\infty\int_0^\infty u^{-s}(x+y)^{-ds} y^{d-1} e^{-vx^d}\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}s. $$

Making the change of variables $y \mapsto xy$ yields

$$\begin{aligned} I_{d,0} &= \frac{1}{2\pi i}\int_{(c)} \Gamma(s)\int_0^1\int_v^1\int_0^\infty\int_0^\infty u^{-s}x^{d(1-s)}(1+y)^{-ds}y^{d-1}e^{-vx^d}\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}s \\ &= \frac{1}{2\pi i}\int_{(c)} \Gamma(s)\int_0^1\int_v^1 u^{-s}\left(\int_0^\infty y^{d-1}(1+y)^{-ds}\,\mathrm{d}y\right)\left(\int_0^\infty x^{d(1-s)}e^{-vx^d}\mathrm{d}x\right)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}s \\ &= \frac{d\Gamma(d-1)}{2\pi i}\int_{(c)} \frac{\Gamma(s)\Gamma(ds-d)\Gamma(1+\frac{1}{d}-s)}{\Gamma(ds)(ds-1)}\,\mathrm{d}s, \end{aligned}$$

where $1 < c < 1+\frac{1}{d}$. Moving the integration path to the right, one encounters the simple poles at $s = 1+\frac{1}{d}+j$ for $j = 0,1,\ldots$. Summing over all residues of these simple poles and proving that the remainder integral tends to zero, we get

$$I_{d,0} = \Gamma(d-1)\sum_{j\geq 0}\frac{(-1)^j\Gamma(j+1+\frac{1}{d})\Gamma(dj+1)}{(j+1)!\Gamma(dj+d+1)},$$

where the terms converge at the rate $j^{-d-1+\frac{1}{d}}$. This can be expressed easily in terms of the generalized hypergeometric functions.

An alternative integral representation can be derived for $I_{d,0}$ as follows.

$$\begin{aligned} I_{d,0} &= \frac{\Gamma(d-1)}{\Gamma(d)}\sum_{j\geq 0}\frac{\Gamma(j+1+\frac{1}{d})}{\Gamma(j+2)}(-1)^j\int_0^1(1-x)^{d-1}x^{dj}\,\mathrm{d}x \\ &= \frac{\Gamma(\frac{1}{d})}{d-1}\int_0^1(1-x)^{d-1}\frac{1-(1+x^d)^{-\frac{1}{d}}}{x^d}\,\mathrm{d}x, \end{aligned}$$

which can also be derived directly from the original multiple integral representation and successive changes of variables (first u, then x, then v, and finally y). In particular, for $d = 2$,

$$I_{2,0} = \sqrt{\pi}\left(\sqrt{2}-1+\log 2+\log(\sqrt{2}-1)\right).$$

Now we turn to $I_{d,d}$.

$$I_{d,d} = \int_0^1\int_v^1\int_0^\infty\int_0^\infty w^{d-1}e^{-ux^d-v(x+w)^d}\left(e^{vx^d}-1\right)\mathrm{d}w\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v.$$

By the same arguments used above, we have

$$I_{d,d} = \frac{\Gamma(d)}{2d\pi i}\int_{(c)}\frac{\Gamma(s)\Gamma(ds-d)\Gamma(1+\frac{1}{d}-s)}{\Gamma(ds)}I'_{d,d}\,\mathrm{d}s,$$

where $c > 1$ and

$$I'_{d,d} := \int_0^1 v^{-s}\int_v^1\left((u-v)^{s-1-\frac{1}{d}}-u^{s-1-\frac{1}{d}}\right)\mathrm{d}u\,\mathrm{d}v.$$

To evaluate $I'_{d,d}$, assume first that $\frac{1}{d} < \Re(s) < 1$, so that

$$\begin{aligned} I'_{d,d} &= \int_0^1 v^{-s} \int_0^{1-v} u^{s-1-\frac{1}{d}}\,\mathrm{d}u - \int_0^1 u^{s-1-\frac{1}{d}} \int_0^u v^{-s}\,\mathrm{d}v\,\mathrm{d}u \\ &= \frac{d}{d-1}\left(\frac{\Gamma(1-s)\Gamma(s-\frac{1}{d})}{\Gamma(1-\frac{1}{d})} - \frac{1}{1-s}\right). \end{aligned}$$

Now the right-hand side is well-defined for $\frac{1}{d} < \Re(s) < 2$. Substituting this into $I_{d,d}$, we obtain

$$I_{d,d} = \frac{\Gamma(d-1)}{2\pi i}\int_{(c)} \frac{\Gamma(s)\Gamma(ds-d)\Gamma(1+\frac{1}{d}-s)}{\Gamma(ds)}\left(\frac{\Gamma(1-s)\Gamma(s-\frac{1}{d})}{\Gamma(1-\frac{1}{d})} - \frac{1}{1-s}\right)\mathrm{d}s,$$

where $1 < c < 1+\frac{1}{d}$. For computational purpose, we use the functional equation for Gamma function

$$\Gamma(1-s)\Gamma(s) = \frac{\pi}{\sin \pi s},$$

so that

$$I_{d,d} = \frac{\Gamma(d-1)}{2\pi i}\int_{(c)} \frac{\pi\Gamma(ds-d)}{\Gamma(ds)\sin(\pi(s-\frac{1}{d}))}\left(\frac{\pi}{\Gamma(1-\frac{1}{d})\sin(\pi s)} + \frac{\Gamma(s-1)}{\Gamma(ds)\Gamma(s-\frac{1}{d})}\right)\mathrm{d}s.$$

In this case, we have simples poles at $s = j + 1/d$ for both integrands and $s = j$ for the first integrand to the right of $\Re(s) = 1$ for $j = 2, 3, \ldots$. Thus summing over all the residues and proving that the remainder integral goes to zero, we obtain

$$I_{d,d} = \Gamma(d-1)\Gamma(\tfrac{1}{d})\sum_{j\geq 2}\frac{\Gamma(dj-d)}{j\Gamma(dj)} - \Gamma(d-1)\sum_{j\geq 2}\frac{(-1)^j\Gamma(j-1+\frac{1}{d})\Gamma(dj-d+1)}{\Gamma(j)\Gamma(dj+1)}.$$

A similar argument as that used for $I_{d,0}$ gives the alternative integral representation

$$I_{d,d} = \frac{\Gamma(\frac{1}{d})}{d(d-1)}\left(-1 + \int_0^1 \left(1 - t^{\frac{1}{d}}\right)^{d-1}\left(t^{\frac{1}{d}-1}(1+t)^{-\frac{1}{d}} + \frac{-\log(1-t)-t}{t^2}\right)\mathrm{d}t\right).$$

In particular, for $d = 2$,

$$I_{2,2} = \sqrt{\pi}\left(2 - \sqrt{2} - 2\log 2 + \log(\sqrt{2}+1)\right).$$

Now we consider $I_{d,m}$ for $1 \leq m < d$.

$$I_{d,m} := \int_0^1\int_v^1\int_0^\infty\int_0^\infty\int_0^\infty y^{d-1-m}w^{m-1}e^{-u(x+y)^d - v(x+w)^d}\left(e^{vx^d}-1\right)\mathrm{d}w\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}v,$$

which by the same arguments leads to

$$\begin{aligned} I_{d,m} &= \frac{1}{2\pi i}\int_{(c)}\Gamma(s)\int_0^1\int_v^1 u^{-s}\left(\int_0^\infty y^{d-1-m}(1+y)^{-ds}\,\mathrm{d}y\right) \\ &\quad\times\int_0^\infty w^{m-1}\left(\int_0^\infty x^{d(1-s)}\left(e^{-vx^d((1+w)^d-1)} - e^{-vx^d(1+w)^d}\right)\mathrm{d}x\right)\mathrm{d}w\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}s \\ &= \frac{d\Gamma(d-m)}{2(d-1)\pi i}\int_{(c)}\frac{\Gamma(s)\Gamma(ds-d+m)\Gamma(1+\frac{1}{d}-s)}{\Gamma(ds)(ds-1)}W_m(s)\,\mathrm{d}s, \end{aligned}$$

where $1 < c < 1 + \frac{1}{d}$ and

$$\begin{aligned}W_m(s) &:= \int_0^\infty w^{m-1}\left(\left((1+w)^d-1\right)^{s-1-\frac{1}{d}} - (1+w)^{ds-d-1}\right) \mathrm{d}w \\ &= \frac{1}{d}\int_0^1 t^{-s}(t^{-\frac{1}{d}}-1)^{m-1}\left((1-t)^{s-1-\frac{1}{d}}-1\right) \mathrm{d}t \\ &= \frac{1}{d}\sum_{0\le \ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}\left(\frac{\pi\Gamma(s-\frac{1}{d})}{\Gamma(1-\frac{\ell+1}{d})\Gamma(s+\frac{\ell}{d})\sin(\pi(s+\frac{\ell}{d}))} - \frac{1}{1-\frac{\ell}{d}-s}\right),\end{aligned}$$

for $\frac{1}{d} < \Re(s) < 2-(m-1)/d$. Note that each term has no pole at $s = 1-\frac{\ell}{d}$. Thus

$$I_{d,m} = \frac{\Gamma(d-m)}{d-1}\sum_{0\le \ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell} I_{d,m,\ell},$$

where

$$\begin{aligned}I_{d,m,\ell} := \frac{1}{2\pi i}\int_{(c)} &\frac{\Gamma(s)\Gamma(ds-d+m)\Gamma(1+\frac{1}{d}-s)}{\Gamma(ds)(ds-1)} \\ &\times\left(\frac{\pi\Gamma(s-\frac{1}{d})}{\Gamma(1-\frac{\ell+1}{d})\Gamma(s+\frac{\ell}{d})\sin(\pi(s+\frac{\ell}{d}))} - \frac{1}{1-\frac{\ell}{d}-s}\right) \mathrm{d}s.\end{aligned}$$

We then deduce that the integral equals the sum of the residues at $s = j+\frac{1}{d}$ and $s = j+1-\frac{\ell}{d}$

$$\begin{aligned}I_{d,m,\ell} &= -\frac{\Gamma(\frac{\ell+1}{d})}{d}\sum_{j\ge1}\frac{\Gamma(j+1+\frac{1}{d})\Gamma(dj+m+1)}{(j+1)\Gamma(dj+d+1)\Gamma(j+1+\frac{\ell+1}{d})} \\ &\quad + \frac{1}{d}\sum_{j\ge1}\frac{(-1)^j\Gamma(j+1+\frac{1}{d})\Gamma(dj+m+1)}{(j+1)!\Gamma(dj+d+1)(j+\frac{\ell+1}{d})} \\ &\quad + \Gamma(\tfrac{\ell+1}{d})\sum_{j\ge1}\frac{\Gamma(j+1-\frac{\ell}{d})\Gamma(dj+m-\ell)}{j!\Gamma(dj+d-\ell)(dj+d-\ell-1)} \\ &= I_{d,m,\ell}^{[1]} + I_{d,m,\ell}^{[2]} + I_{d,m,\ell}^{[3]}.\end{aligned}$$

It follows that

$$\begin{aligned}&C_d - I_{d,d} + I_{d,0} \\ &\quad = (d-1)\sum_{1\le m<d}\binom{d}{m}\binom{d-2}{m-1} I_{d,m} \\ &\quad = \sum_{1\le m<d}\binom{d}{m}\frac{(d-2)!}{(m-1)!}\sum_{0\le \ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}\left(I_{d,m,\ell}^{[1]} + I_{d,m,\ell}^{[2]} + I_{d,m,\ell}^{[3]}\right) \\ &\quad =: C_d^{[1]} + C_d^{[2]} + C_d^{[3]}.\end{aligned}$$

For further simplification of these sums, we begin with $C_d^{[2]}$. Note first that

$$\begin{aligned}
&\sum_{0\le \ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}I_{d,m,\ell}^{[2]}\\
&\quad=\frac{1}{d}\sum_{j\ge 1}\frac{(-1)^j\Gamma(j+1+\frac{1}{d})\Gamma(dj+m+1)}{(j+1)!\Gamma(dj+d+1)}\sum_{0\le\ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}\frac{1}{j+\frac{\ell+1}{d}}\\
&\quad=(-1)^{m-1}(m-1)!\sum_{j\ge 1}\frac{(-1)^j\Gamma(j+1+\frac{1}{d})\Gamma(dj+1)}{(j+1)!\Gamma(dj+d+1)}.
\end{aligned}$$

Thus

$$\begin{aligned}
C_d^{[2]}&=\sum_{1\le m<d}\binom{d}{m}\frac{(d-2)!}{(m-1)!}\sum_{0\le\ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}I_{d,m,\ell}^{[2]}\\
&=(d-2)!\sum_{1\le m<d}\binom{d}{m}(-1)^{m-1}\sum_{j\ge 1}\frac{(-1)^j\Gamma(j+1+\frac{1}{d})\Gamma(dj+1)}{(j+1)!\Gamma(dj+d+1)}\\
&=(1+(-1)^d)\left(I_{d,0}-\frac{\Gamma(1+\frac{1}{d})}{d(d-1)}\right).
\end{aligned}$$

Accordingly, $C_d^{[2]}=0$ for odd values of d.

For the other two sums containing $I_{d,m,\ell}^{[1]}$ and $I_{d,m,\ell}^{[3]}$, we use the identity

$$\sum_{\ell<m<d}\frac{(N+m)!(-1)^m}{m!(d-m)!(m-1-\ell)!}=\frac{(-1)^d N!}{(d-1-\ell)!}\left(\binom{N+1+\ell}{d}-\binom{N+d}{d}\right).$$

Then

$$\begin{aligned}
C_d^{[1]}&=\sum_{1\le m<d}\binom{d}{m}\frac{(d-2)!}{(m-1)!}\sum_{0\le\ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}I_{d,m,\ell}^{[1]}\\
&=(d-2)!\sum_{0\le\ell\le d-2}\frac{d!}{\ell!}(-1)^{\ell}\frac{\Gamma(\frac{\ell+1}{d})}{d}\sum_{j\ge 1}\frac{\Gamma(j+1+\frac{1}{d})}{(j+1)\Gamma(dj+d+1)\Gamma(j+1+\frac{\ell+1}{d})}\\
&\quad\times\sum_{\ell<m<d}\frac{\Gamma(dj+m+1)(-1)^m}{m!(d-m)!(m-1-\ell)!}\\
&=\frac{(-1)^d}{d(d-1)}\sum_{0\le\ell\le d-2}\binom{d-1}{\ell}(-1)^{\ell}\Gamma(\tfrac{\ell+1}{d})\sum_{j\ge 1}\frac{\Gamma(j+1+\frac{1}{d})}{(j+1)\Gamma(j+1+\frac{\ell+1}{d})}\left(\frac{\binom{dj+\ell+1}{d}}{\binom{dj+d}{d}}-1\right).
\end{aligned}$$

Note that

$$\frac{\binom{dj+\ell+1}{d}}{\binom{dj+d}{d}}-1=O(j^{-1})\qquad(0\le\ell\le d-2),$$

for large j, so that the series is absolutely convergent.

Similarly,

$$
\begin{aligned}
C_d^{[3]} &= \sum_{1\le m<d}\binom{d}{m}\frac{(d-2)!}{(m-1)!}\sum_{0\le \ell<m}\binom{m-1}{\ell}(-1)^{m-1-\ell}I_{d,m,\ell}^{[3]}\\
&= \frac{(-1)^d}{d-1}\sum_{0\le \ell\le d-2}\binom{d-1}{\ell}(-1)^{\ell-1}\Gamma(\tfrac{\ell+1}{d})\sum_{j\ge 1}\frac{\Gamma(j+1-\frac{\ell}{d})}{j!(dj+d-\ell-1)}\left(\frac{\binom{dj}{d}}{\binom{dj+d-\ell-1}{d}}-1\right).
\end{aligned}
$$

Since $v_d = \frac{d}{d-1}\Gamma(\frac{1}{d}) + 2d^2C_d$, we obtain, by converting the series representations into hypergeometric functions, the following approximate numerical values of v_d.

$$
\begin{aligned}
v_2 &\approx 2.86126\,35493\,11178\,82531\,14379,\\
v_3 &\approx 3.22524\,36444\,05576\,89660\,59392,\\
v_4 &\approx 3.97797\,27442\,19455\,29292\,64760,\\
v_5 &\approx 4.84527\,39171\,62611\,42226\,50057,\\
v_6 &\approx 5.76349\,95321\,96568\,64812\,77416,\\
v_7 &\approx 6.70865\,12250\,86590\,36364\,34742,\\
v_8 &\approx 7.66955\,04435\,24665\,04704\,24808,\\
v_9 &\approx 8.64032\,79742\,08287\,24931\,00067,\\
v_{10} &\approx 9.61764\,75521\,13755\,73944\,20940,\\
v_{11} &\approx 10.59949\,78766\,56951\,63098\,76869,\\
v_{12} &\approx 11.58460\,78314\,60409\,77794\,37163.
\end{aligned}
$$

In particular, v_2 has a closed-form expression

$$
v_2 = \frac{2}{3}\sqrt{\pi}\left(2\pi^2 - 9 - 12\log 2\right).
$$

The leading constant $\tilde{v}_d$ of the asymptotic variance of the d-dimensional maxima. Let

$$
J_{d,0} := 2d^2\int_0^\infty\int_0^\infty y^{d-1}e^{-x^d-(x+y)^d}\,\mathrm{d}x\,\mathrm{d}y,
$$

and

$$
J_{d,k} := \frac{dd!}{(d-k-1)!(k-1)!}\int_0^\infty\int_0^\infty\int_0^\infty y^{d-k-1}w^{k-1}e^{-\left(x+y\right)^d-(x+w)^d}\left(e^{x^d}-1\right)\mathrm{d}w\,\mathrm{d}y\,\mathrm{d}x.
$$

Then (see (4))

$$
\tilde{v}_d = \Gamma\left(\frac{1}{d}\right) + \sum_{1\le k<d}\binom{d}{k}J_{d,k} - J_{d,0}.
$$

Consider first $J_{d,0}$. By expanding $(1+x^d)^{-1-\frac{1}{d}}$, interchanging and evaluating the integrals, we obtain

$$\begin{aligned}
J_{d,0} &= 2\Gamma\left(\frac{1}{d}\right)\int_0^1 \frac{(1-x)^{d-1}}{(1+x^d)^{1+\frac{1}{d}}}\,\mathrm{d}x \\
&= 2d!\sum_{j\ge 0}\frac{\Gamma(j+1+\frac{1}{d})\Gamma(dj+1)}{\Gamma(j+1)\Gamma(dj+d+1)}(-1)^j,
\end{aligned}$$

the general terms converging at the rate $O(j^{-d-\frac{1}{d}})$. The convergence rate can be accelerated as follows.

$$\begin{aligned}
J_{d,0} &= 2\Gamma\left(1+\frac{1}{d}\right)\int_0^1 x^{\frac{1}{d}-1}(1-x^{\frac{1}{d}})^{d-1}(1+x)^{-1-\frac{1}{d}}\,\mathrm{d}x \\
&= 2\Gamma\left(1+\frac{1}{d}\right)\sum_{r\ge 0}2^{-r-1-\frac{1}{d}}\int_0^1 (1-x)^r x^{\frac{1}{d}-1}(1-x^{\frac{1}{d}})^{d-1}\,\mathrm{d}x \\
&= \Gamma\left(1+\frac{1}{d}\right)2^{-\frac{1}{d}}\sum_{0\le \ell<d}\binom{d-1}{\ell}(-1)^\ell\Gamma\left(\frac{\ell+1}{d}\right)\sum_{j\ge 0}\frac{\Gamma(j+1+\frac{1}{d})}{\Gamma(j+1+\frac{\ell+1}{d})}2^{-j},
\end{aligned}$$

the convergence rate being now exponential. In terms of the generalized hypergeometric functions, we have

$$J_{d,0} = \Gamma\left(\frac{1}{d}\right)2^{-\frac{1}{d}}\sum_{0\le \ell<d}\binom{d-1}{\ell}\frac{(-1)^\ell}{\ell+1}\,{}_2F_1\left(1+\frac{1}{d},1;1+\frac{\ell+1}{d};\frac{1}{2}\right).$$

The integrals $J_{d,k}$ can be simplified as follows.

$$\begin{aligned}
J_{d,k+1} &= d^2(d-1)\binom{d-2}{k}\int_0^\infty (e^{x^d}-1)\int_x^\infty e^{-y^d} \\
&\qquad\times\int_x^\infty (y-x)^{d-2-k}(z-x)^k e^{-z^d}\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x \\
&= 2d^2(d-1)\binom{d-2}{k}\int_0^\infty e^{-y^d}\int_0^y e^{-z^d} \\
&\qquad\times\int_0^z (e^{x^d}-1)(y-x)^{d-2-k}(z-x)^k\,\mathrm{d}x\,\mathrm{d}z\,\mathrm{d}y \\
&= 2(d-1)\Gamma\left(\frac{1}{d}\right)\binom{d-2}{k}\int_0^1 (1-x)^k\int_0^1 (1-xz)^{d-2-k}z^{k+1} \\
&\qquad\times\left(\frac{1}{(1+z^d-x^dz^d)^{1+\frac{1}{d}}}-\frac{1}{(1+z^d)^{1+\frac{1}{d}}}\right)\mathrm{d}z\,\mathrm{d}x \\
&= J'_{d,k+1}+J''_{d,k+1}.
\end{aligned}$$

By the same proof used for $J_{d,0}$, we have

$$\begin{aligned}
J''_{d,k+1} &= -2(d-1)\Gamma\left(\frac{1}{d}\right)\binom{d-2}{k}\int_0^1 (1-x)^k \\
&\qquad \times \int_0^1 (1-xz)^{d-2-k} z^{k+1}(1+z^d)^{-1-\frac{1}{d}}\,\mathrm{d}z\,\mathrm{d}x \\
&= (-1)^{k+1}2^{-\frac{1}{d}}\Gamma\left(\frac{1}{d}\right)\sum_{k<j<d}\binom{d-1}{j}\frac{(-1)^j}{j+1}\,{}_2F_1\left(1+\frac{1}{d},1;1+\frac{j+1}{d};\frac{1}{2}\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
J'_{d,k+1} &= 2(d-1)\Gamma\left(\frac{1}{d}\right)\binom{d-2}{k}\int_0^1 (1-x)^k \\
&\qquad \times \int_0^1 (1-xz)^{d-2-k} z^{k+1}(1+z^d-x^d z^d)^{-1-\frac{1}{d}}\,\mathrm{d}z\,\mathrm{d}x \\
&= 2\Gamma\left(\frac{1}{d}\right)(d-1)!\sum_{0\le j\le d-2-k}\frac{(-1)^j}{j!(d-2-k-j)!} \\
&\qquad \times \sum_{0\le \ell\le k}\frac{(-1)^\ell}{\ell!(k-\ell)!}\cdot\frac{{}_3F_2(1+\frac{1}{d},\frac{k+j+2}{d},1;1+\frac{\ell+j+1}{d},1+\frac{k+j+2}{d};-1)}{(\ell+j+1)(k+j+2)}.
\end{aligned}$$

Thus we obtain the following numerical values for the limiting constant $\tilde{v}_d$ of $\mathbb{V}[M_n]/n^{(d-1)/d}$

$$\begin{aligned}
\tilde{v}_2 &\approx 0.68468\,89279\,50036\,17418\,09957, \\
\tilde{v}_3 &\approx 1.48217\,31873\,40583\,68601\,11369, \\
\tilde{v}_4 &\approx 2.35824\,37612\,02486\,93742\,28054, \\
\tilde{v}_5 &\approx 3.27773\,90059\,79491\,26684\,80858, \\
\tilde{v}_6 &\approx 4.22231\,09450\,77067\,79998\,34338, \\
\tilde{v}_7 &\approx 5.18220\,76686\,16078\,48517\,29967, \\
\tilde{v}_8 &\approx 6.15196\,29023\,77474\,45508\,28039, \\
\tilde{v}_9 &\approx 7.12835\,13658\,43360\,52793\,29089, \\
\tilde{v}_{10} &\approx 8.10938\,23221\,15849\,82527\,77117, \\
\tilde{v}_{11} &\approx 9.09377\,74697\,86680\,89694\,70616, \\
\tilde{v}_{12} &\approx 10.0806\,86465\,19733\,08113\,16376.
\end{aligned}$$

In particular, $\tilde{v}_2 = \sqrt{\pi}(2\log 2 - 1)$; see [2].

Yet another constant in [8]. A similar but simpler integral to (4) appeared in [8], which is of the form

$$K_d := \int_0^\infty\int_0^\infty\int_0^\infty (u+w)^{d-2}e^{-(u+x)^d+x^d-(w+x)^d}\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}w,$$

(this K_d is indeed their K_{d-1}). By Mellin inversion formula for e^{-t}, we obtain

$$\begin{aligned}
K_d &= \frac{1}{2\pi i}\int_{(c)}\Gamma(s)\int_0^\infty\int_0^\infty\int_0^\infty (u+w)^{d-2}(u+x)^{-ds}e^{-(w+x)^d+x^d}\,\mathrm{d}x\,\mathrm{d}u\,\mathrm{d}w\,\mathrm{d}s\\
&= \frac{1}{2d\pi i}\int_{(c)}\Gamma(s)\Gamma(1+\tfrac{1}{d}-s)\\
&\quad\times\int_0^\infty\int_0^\infty (u+w)^{d-2}(1+u)^{-ds}\left((1+w)^d-1\right)^{s-1-\frac{1}{d}}\,\mathrm{d}u\,\mathrm{d}w\,\mathrm{d}s.
\end{aligned}$$

Expanding the factor $(u+w)^{d-2}$, we obtain $K_d=\sum_{0\le m\le d-2}\binom{d-2}{m}K_{d,m}$, where

$$\begin{aligned}
K_{d,m} &:= \frac{1}{2d\pi i}\int_{(c)}\Gamma(s)\Gamma(1+\tfrac{1}{d}-s)\left(\int_0^\infty u^m(1+u)^{-ds}\,\mathrm{d}u\right)\\
&\qquad\times\left(\int_0^\infty w^{d-2-m}\left((1+w)^d-1\right)^{s-1-\frac{1}{d}}\right)\mathrm{d}w\\
&= \frac{1}{2d\pi i}\int_{(c)}\Gamma(s)\Gamma(1+\tfrac{1}{d}-s)B(m+1,ds-m-1)U_m(s)\,\mathrm{d}s. \qquad (7)
\end{aligned}$$

Here

$$\begin{aligned}
U_m(s) &:= \int_0^\infty w^{d-2-m}\left((1+w)^d-1\right)^{s-1-\frac{1}{d}}\,\mathrm{d}w\\
&= \frac{1}{d}\int_0^1 t^{-s}(1-t)^{s-1-\frac{1}{d}}\left(t^{-\frac{1}{d}}-1\right)^{d-2-m}\,\mathrm{d}t\\
&= \frac{1}{d}\sum_{0\le\ell\le d-2-m}\binom{d-2-m}{\ell}(-1)^{d-2-m-\ell}B(1-s-\tfrac{\ell}{d},s-\tfrac{1}{d}).
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
K_d &= \frac{1}{d^2}\sum_{0\le m\le d-2}\binom{d-2}{m}\sum_{0\le\ell\le d-2-m}\binom{d-2-m}{\ell}(-1)^{d-2-m-\ell}m!\Gamma(\tfrac{\ell+1}{d})\\
&\quad\times\sum_{j\ge0}\left(\frac{\Gamma(j+1-\frac{\ell}{d})\Gamma(dj+d-\ell-m-1)}{j!\Gamma(dj+d-\ell)}-\frac{\Gamma(j+1+\frac{1}{d})\Gamma(dj+d-m)}{\Gamma(j+1+\frac{\ell+1}{d})\Gamma(dj+d+1)}\right).
\end{aligned}$$

This readily gives, by converting the above series into hypergeometric functions, the numerical values of the first few K_d,

$$
\begin{aligned}
K_2 &\approx 0.30714\,28473\,56944\,02518\,48954,\\
K_3 &\approx 0.21288\,24684\,73220\,99693\,80676,\\
K_4 &\approx 0.19494\,67028\,23033\,18190\,40460,\\
K_5 &\approx 0.20723\,21512\,99671\,45854\,93769,\\
K_6 &\approx 0.24331\,17024\,51836\,72554\,88428,\\
K_7 &\approx 0.30744\,56566\,07893\,22242\,37300,\\
K_8 &\approx 0.41127\,01058\,90385\,83873\,59349,\\
K_9 &\approx 0.57571\,68456\,67243\,64328\,08087,\\
K_{10} &\approx 0.83615\,82236\,77116\,00233\,16115,\\
K_{11} &\approx 1.25179\,63251\,14070\,86480\,31485,\\
K_{12} &\approx 1.92201\,04035\,18847\,36012\,85304.
\end{aligned}
$$

These are consistent with those given in Chiu and Quine (1997). In particular, $K_2 = \frac{1}{4}\sqrt{\pi}\log 2$. Further simplification of this formula can be obtained as above, but the resulting integral expression is not much simpler than

$$
\frac{\Gamma(\frac{1}{d})}{d^4}\int_0^1\int_0^1 \left(u^{-\frac{1}{d}}+v^{-\frac{1}{d}}-2\right)^{d-2} u^{-1-\frac{1}{d}} v^{-1-\frac{1}{d}} \left(u^{-1}+v^{-1}-1\right)^{-1-\frac{1}{d}} \,\mathrm{d}u\,\mathrm{d}v.
$$

4 Asymptotics of the number of chain records

We consider in this section the number of chain records of random samples from d-dimensional simplex; the tools we use are different from [17] and apply also to chain records for hypercube random samples, which will be briefly discussed. For other types of results, see [17].

4.1 Chain records of random samples from d-dimensional simplex

Assume that $\mathbf{p}_1,\dots,\mathbf{p}_n$ are iud in the d-dimensional simplex S_d. Let Y_n denote the number of chain records of this sample. Then Y_n satisfies the recurrence

$$
Y_n \stackrel{d}{=} 1 + Y_{I_n} \qquad (n \ge 1), \tag{8}
$$

with $Y_0 := 0$, where

$$
\mathbb{P}(I_n = k) = \pi_{n,k} = d\binom{n-1}{k}\int_0^1 t^{kd}(1-t^d)^{n-1-k}(1-t)^{d-1}\,\mathrm{d}t,
$$

for $0 \le k < n$. An alternative expression for the probability distribution $\pi_{n,k}$ is

$$
\pi_{n,k} = \binom{n-1}{k}\sum_{0\le j<d}\binom{d-1}{j}(-1)^j\frac{\Gamma(n-k)\Gamma\left(k+\frac{j+1}{d}\right)}{\Gamma\left(n+\frac{j+1}{d}\right)},
$$

which is more useful from a computational point of view.

Let

$$(z+1)\cdots(z+d)-d!=z\prod_{1\le\ell<d}(z-\lambda_\ell),$$

where the λ_ℓ's are all complex ($\notin\mathbb{R}$), except when d is even (in that case, $-d-1$ is the unique real zero among $\{\lambda_1,\dots,\lambda_{d-1}\}$). Interestingly, the same equation also arises in the analysis of random increasing k-trees (see [9]), in some packing problem of intervals (see [4]), and in the analysis of sorting and searching problems (see [7]).

Theorem 3. *The number of chain records Y_n for random samples from d-dimensional simplex is asymptotically normally distributed in the following sense*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{Y_n-\mu_s\log n}{\sigma_s\sqrt{\log n}}<x\right)-\Phi(x)\right|=O\left((\log n)^{-1/2}\right), \tag{9}$$

where $\mu_s:=1/(dH_d)$ and $\sigma_s:=\sqrt{H_d^{(2)}/(dH_d^3)}$. The mean and the variance are asymptotic to

$$\mathbb{E}[Y_n]=\frac{H_n}{dH_d}+c_1+O(n^{-\varepsilon}), \tag{10}$$

$$\mathbb{V}[Y_n]=\frac{H_d^{(2)}}{dH_d^3}H_n+c_2+O(n^{-\varepsilon}), \tag{11}$$

respectively, for some $\varepsilon>0$, where

$$c_1=\frac{1}{dH_d}\sum_{1\le\ell<d}\left(\psi\left(-\frac{\lambda_\ell}{d}\right)-\psi\left(\frac{\ell}{d}\right)\right),$$

$$c_2=\frac{1}{6}+\frac{\pi^2}{6d^2H_d^2}-\frac{2H_d^{(3)}}{3H_d^3}+\frac{(H_d^{(2)})^2}{2H_d^4}+\frac{1}{d^2H_d^2}\sum_{1\le\ell<d}\left(\psi'\left(-\frac{\lambda_\ell}{d}\right)-\psi'\left(\frac{\ell}{d}\right)\right)$$
$$+\frac{c_1H_d^{(2)}}{H_d^2}-\frac{2d!}{H_d}\sum_{j\ge1}\frac{(dj+1)\cdots(dj+d)(H_{dj+d}-H_{dj})}{((dj+1)\cdots(dj+d)-d!)^2}.$$

Here $\psi(x)$ denotes the derivative of $\log\Gamma(x)$.

The error terms in (10) and (11) can be further refined, but we content ourselves with the current forms for simplicity.

Expected number of chain records. We begin with the proof of (10). Consider the mean $\mu_n:=\mathbb{E}[Y_n]$. Then $\mu_0=0$ and, by (8),

$$\mu_n=1+\sum_{0\le k<n}\pi_{n,k}\mu_k \qquad (n\ge1). \tag{12}$$

Let $\tilde{f}(z):=e^{-z}\sum_{n\ge0}\mu_nz^n/n!$ denote the Poisson generating function of μ_n. Then, by (12),

$$\tilde{f}(z)+\tilde{f}'(z)=1+d\int_0^1\tilde{f}(t^dz)(1-t)^{d-1}\,\mathrm{d}t.$$

Let $\tilde{f}(z) = \sum_{n\geq 0} \tilde{\mu}_n z^n/n!$. Taking the coefficients of z^n on both sides gives the recurrence

$$\tilde{\mu}_n + \tilde{\mu}_{n+1} = \frac{d!}{(dn+1)\cdots(dn+d)}\,\tilde{\mu}_n \qquad (n \geq 1).$$

Solving this recurrence using $\tilde{\mu}_1 = 1$ yields

$$\tilde{\mu}_n = (-1)^{n-1} \prod_{1\leq j<n} \left(1 - \frac{d!}{(dj+1)\cdots(dj+d)}\right) \qquad (n \geq 1).$$

It follows that for $n \geq 1$

$$\mu_n = \sum_{1\leq k\leq n} \binom{n}{k} \tilde{\mu}_k = \sum_{1\leq k\leq n} \binom{n}{k} (-1)^{k-1} \prod_{1\leq j<k} \left(1 - \frac{d!}{(dj+1)\cdots(dj+d)}\right). \tag{13}$$

This is an identity with exponential cancelation terms; cf. [17]. In the special case when $d = 2$, we have an identity

$$\mu_n = \frac{H_n + 2}{3}.$$

No such simple expression is available for $d \geq 3$ since there are complex-conjugate zeros; see (14).

Exact solution of the general recurrence. In general, consider the recurrence

$$a_n = b_n + \sum_{0\leq k<n} \pi_{n,k} a_k \qquad (n \geq 1),$$

with $a_0 = 0$. Then the same approach used above leads to the recurrence

$$\tilde{a}_{n+1} = -\left(1 - \frac{d!}{(dn+1)\cdots(dn+d)}\right) \tilde{a}_n + \tilde{b}_n + \tilde{b}_{n+1},$$

which by iteration gives

$$\tilde{a}_{n+1} = \sum_{0\leq k\leq n} (-1)^k \left(\tilde{b}_{n-k} + \tilde{b}_{n-k+1}\right) \prod_{0\leq j<k} \left(1 - \frac{d!}{(d(n-j)+1)\cdots(d(n-j)+d)}\right),$$

by defining $b_0 = \tilde{b}_0 = 0$. Then we obtain the closed-form solution

$$a_n = \sum_{1\leq k\leq n} \binom{n}{k} \tilde{a}_k.$$

A similar theory of “d-analogue” to that presented in [13] can be developed (by replacing $2^d/j^d$ there by $d!/((dj+1)\cdots(dj+d))$).

However, this type of calculation becomes more involved for higher moments.

Asymptotics of μ_n. We now look at the asymptotics of μ_n. To that purpose, we need a better expression for the finite product in the sum-expression (13).

In terms of the zeros λ_j's of the equation $(z+1)\cdots(z+d)-d!$, we have

$$\begin{aligned}\prod_{1\le j<n}\left(1-\frac{d!}{(dj+1)\cdots(dj+d)}\right) &= \frac{\prod_{1\le j<n}\left(dj\prod_{1\le \ell<d}(dj-\lambda_\ell)\right)}{\prod_{1\le j<n}\left((dj+1)\cdots(dj+d)\right)}\\ &= \frac{1}{n}\prod_{1\le \ell<d}\frac{\Gamma\left(n-\frac{\lambda_\ell}{d}\right)\Gamma\left(1+\frac{\ell}{d}\right)}{\Gamma\left(n+\frac{\ell}{d}\right)\Gamma\left(1-\frac{\lambda_\ell}{d}\right)} \qquad (14)\\ &=: \phi(n).\end{aligned}$$

The zeros λ_j's are distributed very regularly as showed in Figure 4.

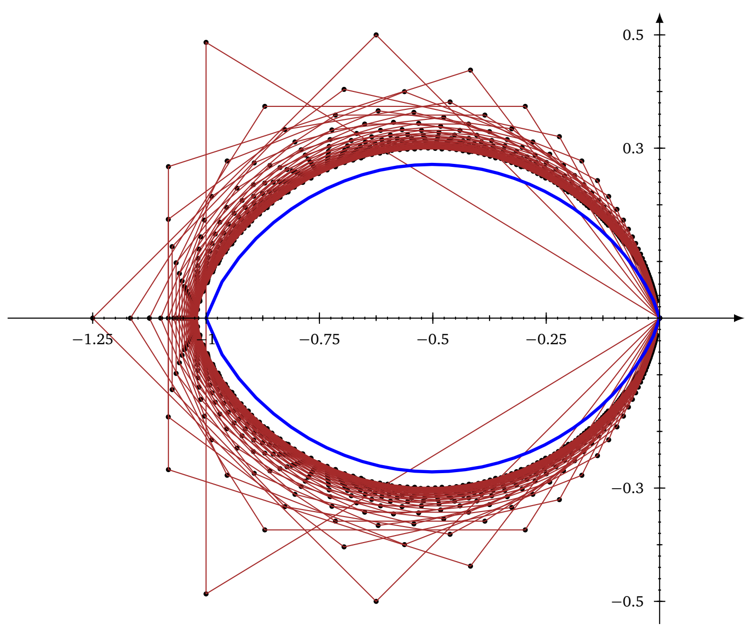

Figure 4: *Distributions of the zeros of $(z+1)\cdots(z+d)-d!=0$ for $d=3,\ldots,50$. The zeros approach, as d increases, to the limiting curve $|z^{-z}(z+1)^{1+z}|=1$ (the blue innermost curve).*

Now we apply the integral representation for the n-th finite difference (called Rice's integrals; see [14]) and obtain

$$\mu_n = -\frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\frac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n+1-s)}\,\phi(s)\,\mathrm{d}s.$$

Note that $\phi(s)$ is well defined and has a simple pole at $s=0$. The integrand then has a double pole at $s=0$; standard calculation (moving the line of integration to the left and summing the residue of the pole encountered) then leads to

$$\mu_n = \frac{1}{dH_d}\left(H_n+\sum_{1\le\ell<d}\left(\psi\left(-\frac{\lambda_\ell}{d}\right)-\psi\left(\frac{\ell}{d}\right)\right)\right)+O\left(n^{-\varepsilon}\right),$$

where the O-term can be made more explicit if needed. Note that to get this expression, we used the identity

$$\frac{(z+1)\cdots(z+d)-d!}{z} = \sum_{1\le j\le d} \frac{d!\Gamma(z+d-j+1)}{(d-j+1)!\Gamma(z+1)}.$$

The probability generating function. Let $P_n(y) := \mathbb{E}[y^{Y_n}]$. Then $P_0(y)=1$ and for $n\ge 1$

$$P_n(y) = y\sum_{0\le k<n} \pi_{n,k}P_k(y).$$

The same procedure used above leads to

$$\begin{aligned} P_n(y) &= \sum_{0\le k\le n}\binom{n}{k}(-1)^k \prod_{0\le j<k}\left(1-\frac{d!y}{(dj+1)\cdots(dj+d)}\right) \\ &= 1+(y-1)\sum_{1\le k\le n}\binom{n}{k}(-1)^{k-1}\prod_{1\le j<k}\left(1-\frac{d!y}{(dj+1)\cdots(dj+d)}\right) \qquad (n\ge 0). \end{aligned}$$

Let now $|y-1|$ be close to zero and

$$(z+1)\cdots(z+d)-d!y = \prod_{1\le \ell\le d}\left(z-\lambda_\ell(y)\right).$$

Note that the λ_ℓ's are analytic functions of y. Let $\lambda_d(y)$ denote the zero with $\lambda_d(1)=0$. Then we have

$$P_n(y) = 1-\frac{y-1}{2\pi i}\int_{1-\varepsilon-i\infty}^{1-\varepsilon+i\infty}\frac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n+1-s)}\,\phi(s,y)\,\mathrm{d}s,$$

where $\varepsilon>0$ and

$$\phi(s,y) = \frac{\Gamma\left(s-\frac{\lambda_d(y)}{d}\right)}{\Gamma(s+1)\Gamma\left(1-\frac{\lambda_d(y)}{d}\right)}\prod_{1\le\ell<d}\frac{\Gamma\left(s-\frac{\lambda_\ell(y)}{d}\right)\Gamma\left(1+\frac{\ell}{d}\right)}{\Gamma\left(s+\frac{\ell}{d}\right)\Gamma\left(1-\frac{\lambda_\ell(y)}{d}\right)}.$$

Note that for $y\ne 1$, $\phi(0,y)=1-y$. When $y\sim 1$, the dominant zero is $\lambda_d(y)$, and we then deduce that

$$P_n(y) = Q(y)n^{\lambda_d(y)/d}+O(|1-y|n^{-\varepsilon}),$$

where

$$Q(y) := \frac{d(y-1)}{\lambda_d(y)\Gamma(1+\frac{\lambda_d(y)}{d})}\prod_{1\le\ell<d}\frac{\Gamma\left(\frac{\lambda_d(y)-\lambda_\ell(y)}{d}\right)\Gamma\left(1+\frac{\ell}{d}\right)}{\Gamma\left(\frac{\lambda_d(y)+\ell}{d}\right)\Gamma\left(1-\frac{\lambda_\ell(y)}{d}\right)}.$$

By writing $(z+1)\cdots(z+d)-d!y=0$ as

$$(1+z)\cdots\left(1+\frac{z}{d}\right)-1 = y-1,$$

and by Lagrange's inversion formula, we obtain

$$\lambda_d(y) = \frac{y-1}{H_d}-\frac{H_d^2-H_d^{(2)}}{2H_d^3}(y-1)^2+O\left(|y-1|^3\right).$$

From this we then get $Q(1) = 1 + O(|y-1|)$ and

$$\lambda_d(e^\eta) = \frac{\eta}{H_d} + \frac{H_d^{(2)}}{2H_d^3}\eta^2 - \frac{2H_dH_d^{(3)} - 3(H_d^{(2)})^2}{6H_d^5}\eta^3 + O(|\eta|^4),$$

for small $|\eta|$. This is a typical situation of the quasi-power framework (see [15, 22]), and we deduce (10), (11) and the Berry-Esseen bound (9). The expression for c_2 is obtained by an ad-hoc calculation based on computing the second moment (the expression obtained by the quasi-power framework being less explicit).

When $d = 2$, a direct calculation leads to the identity

$$\mathbb{V}[Y_n] = \frac{5}{27}H_n + \frac{2\pi^2}{27} + \frac{H_n^{(2)}}{9} - \frac{26}{27} - \frac{2}{9}\sum_{j\ge 1}\left(\frac{2j-1}{j^2\binom{n+j}{n}} - \frac{2j}{(j+\frac{1}{2})^2\binom{n+j+\frac{1}{2}}{n}}\right),$$

for $n \ge 1$, which is also an asymptotic expansion. This is to be contrasted with $\mathbb{E}[Y_n] = (H_n + 2)/3$.

4.2 Chain records of random samples from hypercubes

In this case, we have, denoting still by Y_n the number of chain records in iud random samples from $[0,1]^d$,

$$Y_n \overset{d}{=} 1 + Y_{I_n} \qquad (n \ge 1),$$

with $Y_0 = 0$ and

$$\mathbb{P}(I_n = k) = \binom{n-1}{k}\int_0^1 t^k(1-t)^{n-1-k}\frac{(-\log t)^{d-1}}{(d-1)!}\,\mathrm{d}t.$$

Let $P_n(y) := \mathbb{E}[y^{Y_n}]$. Then the Poisson generating function $\tilde{\mathscr{P}}(z,y) := e^{-z}\sum_{n\ge 0} P_n(y)z^n/n!$ satisfies

$$\tilde{\mathscr{P}}(z,y) + \frac{\partial}{\partial z}\tilde{\mathscr{P}}(z,y) = y\int_0^1 \tilde{\mathscr{P}}(tz,y)\frac{(-\log t)^{d-1}}{(d-1)!}\,\mathrm{d}t,$$

with $\tilde{\mathscr{P}}(0,y) = 1$. We then deduce that

$$P_n(y) = 1 + \sum_{1\le k\le n}\binom{n}{k}(-1)^k\prod_{1\le j\le k}\left(1 - \frac{y}{j^d}\right).$$

Consequently, by Rice's integral representation [14],

$$P_n(y) = \frac{1}{2\pi i}\int_{1-\varepsilon-i\infty}^{1-\varepsilon+i\infty}\frac{\Gamma(n+1)\Gamma(-s)}{\Gamma(n+1-s)\Gamma(s+1)^d}\prod_{1\le \ell\le d}\frac{\Gamma(s+1-y^{1/d}e^{2\ell\pi i/d})}{\Gamma(1-y^{1/d}e^{2\ell\pi i/d})}\,\mathrm{d}s.$$

If $|y-1|$ is close to zero, we deduce that

$$P_n(y) = \frac{n^{y^{1/d}-1}}{\Gamma(y^{1/d})^{1/d}}\prod_{1\le \ell< d}\frac{\Gamma(y^{1/d}(1-e^{2\ell\pi i/d}))}{\Gamma(1-y^{1/d}e^{2\ell\pi i/d})}\left(1 + O(n^{-\varepsilon})\right).$$

A very similar analysis as above then leads to a Berry-Esseen bound for Y_n as follows.

Theorem 4. *The number of chain records Y_n for iud random samples from the hypercube $[0,1]^d$ satisfies*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{Y_n-\mu_h\log n}{\sigma_h\sqrt{\log n}}<x\right)-\Phi(x)\right|=O\left((\log n)^{-1/2}\right),$$

where $\mu_h=\sigma_h:=1/d$. The mean and the variance are asymptotic to

$$\begin{aligned}\mathbb{E}[Y_n]&=\frac{1}{d}\log n+\gamma+\frac{1}{d}\sum_{1\le\ell<d}\psi(1-e^{2\ell\pi i/d})+O(n^{-\varepsilon}),\\ \mathbb{V}[Y_n]&=\frac{1}{d^2}\log n+\frac{\gamma}{d}-\frac{\pi^2}{6d}\\ &\quad+\frac{1}{d^2}\sum_{1\le\ell<d}\left(\psi\left(1-e^{2\ell\pi i/d}\right)+\left(1-2e^{2\ell\pi i/d}\right)\psi'\left(1-e^{2\ell\pi i/d}\right)\right)+O(n^{-\varepsilon}),\end{aligned}$$

for some $\varepsilon>0$.

The asymptotic normality (without rate) was already established in [17].

In the special case when $d=2$, more explicit expressions are available

$$\mathbb{E}[Y_n]=\frac{H_n+1}{2},\quad \mathbb{V}[Y_n]=\frac{H_n+H_n^{(2)}-2}{4},$$

for $n\ge 1$.

5 Dominating records in the d-dimensional simplex

We consider the mean and the variance of the number of dominating records in this section.

Let Z_n denote the number of dominating records of n iud points $\mathbf{p}_1,\dots,\mathbf{p}_n$ in the d-dimensional simplex S_d.

Theorem 5. *The mean and the variance of the number of dominating records for iud random samples from the d-dimensional simplex are given by*

$$\mathbb{E}[Z_n]=\sum_{1\le k\le n}\frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)},\tag{15}$$

$$\mathbb{V}[Z_n]=2\sum_{2\le k\le n}\frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)}H_{k-1}^{(d)}+\sum_{1\le k\le n}\frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)}-\left(\sum_{1\le k\le n}\frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)}\right)^2,\tag{16}$$

respectively. The corresponding expressions for iud random samples from hypercubes are given by $H_n^{(d)}$ and $H_n^{(d)}-H_n^{(2d)}$, respectively.

Proof.

$$\begin{aligned}\mathbb{E}[Z_n] &= \sum_{1\le k\le n} \mathbb{P}\left(\mathbf{p}_k \text{ is a dominating record}\right)\\ &= \sum_{1\le k\le n} (d!)^k \int_{S_d} \left(\prod_{1\le i\le d} x_i\right)^{k-1} \mathrm{d}\mathbf{x}\\ &= \sum_{1\le k\le n} \frac{(d!)^k}{k} \prod_{1\le j<d} \frac{\Gamma(k)\Gamma(jk+1)}{\Gamma((j+1)k+1)}.\end{aligned}$$

Thus, we obtain (15). For large n and bounded d, the partial sum converges to the series

$$\mathbb{E}[Z_n] \to \sum_{k\ge 1} \frac{(d!)^k \Gamma(k)^d}{\Gamma(dk+1)},$$

at an exponential rate. For large d, the right-hand side is asymptotic to

$$\mathbb{E}[Z_n] = 1 + O\left(\frac{(d!)^2}{(2d)!}\right) = 1 + O\left(4^{-d}\sqrt{d}\right),$$

by Stirling's formula.

Similarly, for the second moment, we have

$$\begin{aligned}\mathbb{E}[Z_n^2] - \mathbb{E}[Z_n] &= 2\sum_{2\le k\le n}\sum_{1\le j<k} \mathbb{P}\left(\mathbf{p}_j \text{ and } \mathbf{p}_k \text{ are both dominating records}\right)\\ &= 2\sum_{2\le k\le n}\sum_{1\le j<k} (d!)^k \int_{S_d}\int_{\mathbf{y}\prec\mathbf{x}} \left(\prod y_i\right)^{j-1} \left(\prod x_i\right)^{k-j-1} \mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x}\\ &= 2\sum_{2\le k\le n} (d!)^k \int_{S_d} \left(\sum_{1\le j<k}\int_{\mathbf{y}\prec\mathbf{x}} \left(\prod (y_i/x_i)\right)^{j-1} \mathrm{d}\mathbf{y}\right) \left(\prod x_i\right)^{k-2} \mathrm{d}\mathbf{x}\\ &= 2\sum_{2\le k\le n} (d!)^k H_{k-1}^{(d)} \int_{S_d} \left(\prod x_i\right)^{k-1} \mathrm{d}\mathbf{x}\\ &= 2\sum_{2\le k\le n} \frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)} H_{k-1}^{(d)},\end{aligned}$$

and we obtain (16).

For large n, the right-hand side of (16) converges to

$$2\sum_{k\ge 2} \frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)} H_{k-1}^{(d)} + \sum_{k\ge 1} \frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)} - \left(\sum_{k\ge 1} \frac{(d!)^k\Gamma(k)^d}{\Gamma(dk+1)}\right)^2$$

at an exponential rate, which, for large d, is asymptotic to $3\sqrt{\pi d}\,4^{-d}$. This explains the curves corresponding to Z_n in Figure 3.

The proof for the dominating records in hypercubes is similar and omitted. ∎

Acknowledgements

We thank the referees and Alexander Gnedin for helpful comments.

References

[1] Z.-D. Bai, L. Devroye, H.-K. Hwang and H.-T. Tsai, Maxima in hypercubes, *Random Structures Algorithms* **27** (2005), 290–309. MR2162600

[2] Z.-D. Bai, H.-K. Hwang, W.-Q. Liang and T.-H. Tsai, Limit theorems for the number of maxima in random samples from planar regions, *Electron. J. Probab.* **6** (2001), no. 3, 41 pp. (electronic). MR1816046

[3] J. Baik, P. Deift and K. Johansson, On the distribution of the length of the longest increasing subsequence of random permutations, *J. Amer. Math. Soc.* **12** (1999), 1119–1178. MR1682248

[4] Yu. Baryshnikov and A. Gnedin, Counting intervals in the packing process, *Ann. Appl. Probab.* **11** (2001), 863–877. MR1865026

[5] W.-M. Chen, H.-K. Hwang and T.-H. Tsai, Maxima-finding algorithms for multidimensional samples: A two-phase approach, preprint, 2010.

[6] H.-H. Chern, M. Fuchs and H.-K. Hwang, Phase changes in random point quadtrees, *ACM Trans. Algorithms* **3** (2007), no. 2, Art. 12, 51 pp. MR2335295

[7] H.-H. Chern, H.-K. Hwang and T.-H. Tsai, An asymptotic theory for Cauchy-Euler differential equations with applications to the analysis of algorithms, *J. Algorithms* **44** (2002), 177–225. MR1933199

[8] S. N. Chiu and M. P. Quine, Central limit theory for the number of seeds in a growth model in $\mathbf{R}^d$ with inhomogeneous Poisson arrivals, *Ann. Appl. Probab.* **7** (1997), 802–814. MR1459271

[9] A. Darrasse, H.-K. Hwang, O. Bodini and M. Soria, The connectivity-profile of random increasing k-trees, preprint (2009); available at arXiv:0910.3639v1.

[10] J.-D. Deuschel and O. Zeitouni, Limiting curves for iid records, *Ann. Probab.* **23** (1995), 852–878. MR1334175

[11] L. Devroye, Universal limit laws for depths in random trees, *SIAM J. Comput.* **28** (1999), 409–432. MR1634354

[12] P. Flajolet, X. Gourdon and P. Dumas, Mellin transforms and asymptotics: harmonic sums, *Theoret. Comput. Sci.* **144** (1995), 3–58. MR1337752

[13] P. Flajolet, G. Labelle, L. Laforest and B. Salvy, Hypergeometrics and the cost structure of quadtrees, *Random Structures Algorithms* **7** (1995), 117–144. MR1369059

[14] P. Flajolet and R. Sedgewick, Mellin transforms and asymptotics: finite differences and Rice's integrals, *Theoret. Comput. Sci.* **144** (1995), 101–124. MR1337755

[15] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009. MR2483235

[16] A. Gnedin, Records from a multivariate normal sample, *Statist. Prob. Lett.* **39** (1998) 11–15. MR01649331

[17] A. Gnedin, The chain records, *Electron. J. Probab.* **12** (2007), 767–786. MR2318409

[18] C. M. Goldie and S. Resnick, Records in a partially ordered set, *Ann. Probab* **17** (1989), 678–699. MR985384

[19] C. M. Goldie and S. Resnick, Many multivariate records, *Stoch. Process. Appl.* **59** (1995), 185–216. MR1357651

[20] E. Hashorva and J. Hüsler, On asymptotics of multivariate integrals with applications to records, *Stoch. Models* **18** (2002), 41–69. MR1888285

[21] E. Hashorva and J. Hüsler, Multiple maxima in multivariate samples, *Stoch. Prob. Letters* **75** (2005), 11–17. MR2185605

[22] H.-K. Hwang, On convergence rates in the central limit theorems for combinatorial structures, *European J. Combin.* **19** (1998), 329–343. MR1621021

[23] M. Kałuszka, Estimates of some probabilities in multidimensional convex records, *Appl. Math. (Warsaw)* **23** (1995), 1–11. MR1330054

[24] A. M. Vershik and S. V. Kerov, Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux, *Soviet Math. Dokl.* **233** (1977), 527–531. MR0480398