CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

On the probability of a Pareto record

James Allen Fill [ID] and Ao Sun

Department of Applied Mathematics and Statistics, The Johns Hopkins University, Baltimore, MD, USA
Research for both authors supported by the Acheson J. Duncan Fund for the Advancement of Research in Statistics.

Corresponding author: James Allen Fill; Email: jimfill@jhu.edu

Abstract

Given a sequence of independent random vectors taking values in $\mathbb{R}^d$ and having common continuous distribution function F, say that the nth observation *sets a (Pareto) record* if it is not dominated (in every coordinate) by any preceding observation. Let $p_n(F) \equiv p_{n,d}(F)$ denote the probability that the nth observation sets a record. There are many interesting questions to address concerning p_n and multivariate records more generally, but this short paper focuses on how p_n varies with F, particularly if, under F, the coordinates exhibit negative dependence or positive dependence (rather than independence, a more-studied case). We introduce new notions of negative and positive dependence ideally suited for such a study, called *negative record-setting probability dependence* (NRPD) and *positive record-setting probability dependence* (PRPD), relate these notions to existing notions of dependence, and for fixed $d \geq 2$ and $n \geq 1$ prove that the image of the mapping p_n on the domain of NRPD (respectively, PRPD) distributions is $[p_n^*, 1]$ (resp., $[n^{-1}, p_n^*]$), where p_n^* is the record-setting probability for any continuous F governing independent coordinates.

1. Introduction, background, and main results

1.1. Introduction, notation, and definitions

We begin with some definitions, including the Definition 1.2 of (multivariate) records as studied in this paper. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we write $\mathbf{x} \leq \mathbf{y}$ or $\mathbf{y} \geq \mathbf{x}$ to mean that $x_j \leq y_j$ for $1 \leq j \leq d$, and we write $\mathbf{x} \prec \mathbf{y}$ or $\mathbf{y} \succ \mathbf{x}$ to mean that $x_j < y_j$ for $1 \leq j \leq d$. For $\mathbf{x} \in \mathbb{R}^d$, we use the usual notation $\|\mathbf{x}\|_1 := \sum_{j=1}^d |x_j|$. We use the standard notation $\implies$ for weak convergence of probability measures in Euclidean spaces (or their distribution functions).

Throughout this paper, $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots$ are assumed to be i.i.d. (independent and identically distributed) copies of a d-dimensional random vector $\mathbf{X}$ with distribution function F and law (or distribution) denoted by $\mathcal{L}(\mathbf{X})$. *Throughout the paper we restrict attention to continuous F*, mainly to avoid the complicating mathematical nuisance of ties, as explained in Remark 1.1(d).

Remark 1.1.

(a) As noted by a reviewer of a previous draft, a distribution function F on $\mathbb{R}^d$ is continuous if and only if each of its d univariate marginals is. This is easy to prove from the observation in [4, Sect. 3 (only in first edition)] that F corresponding to random vector $\mathbf{X}$ is continuous at $\mathbf{x} \in \mathbb{R}^d$ if and only if $F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \prec \mathbf{x})$.

(b) Specializing (a) to $d = 1$, the distribution function of a random variable Y is continuous if and only if $\mathbb{P}(Y = y) = 0$ for each $y \in \mathbb{R}$.

(c) We note in passing, however, that, in contradistinction to (b), atomlessness of a random vector does *not* imply continuity of the distribution function in dimensions 2 and higher; see, for example, [6, Sect. 8.5].

(d) Combining (a)–(b), it follows that, if the d-dimensional random vector $\mathbf{X}$ has continuous distribution function F, then almost surely for every $1 \leq j \leq d$ there are no ties among $X_j^{(1)}, X_j^{(2)}, \ldots$.

Definition 1.2.

(a) *For $n \geq 1$, we say that $\mathbf{X}^{(n)}$ is a (Pareto) record (or that it sets a record at time n) if $\mathbf{X}^{(n)} \leq \mathbf{X}^{(i)}$ fails for all $1 \leq i < n$.*

(b) *If $1 \leq k \leq n$, we say that $\mathbf{X}^{(k)}$ is a current record (or remaining record, or maximum) at time n if $\mathbf{X}^{(k)} \leq \mathbf{X}^{(i)}$ fails for all $1 \leq i \leq n$ with $i \neq k$.*

(c) *For $n \geq 1$ we let R_n denote the number of records $\mathbf{X}^{(k)}$ with $1 \leq k \leq n$ and let r_n denote the number of remaining records at time n.*

Remark 1.3. It is clear from Definition 1.2 that if $\widetilde{\mathbf{X}} = (g_1(X_1), \ldots, g_d(X_d))$ where $g_1, \ldots, g_d$ are strictly increasing transformations, then the stochastic processes (R_n) and (r_n) are the same for the i.i.d. sequence $\widetilde{\mathbf{X}}^{(1)}, \widetilde{\mathbf{X}}^{(2)}, \ldots$ as for $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots$. Further, since we assume that F is continuous, it follows from Remark 1.1 that the distribution function $\widetilde{F}$ of $\widetilde{\mathbf{X}}$ is also continuous.

Remark 1.4. We note that the expected number r_n of maxima at time n is n times the probability that $\mathbf{X}^{(n)}$ sets a record. Thus our main Theorems 1.11–1.12 about record-setting probabilities also give information about the expected number of maxima when i.i.d. vectors are sampled.

Omitting, for now, any dependence on F or d from the notation, the probability p_n that $\mathbf{X}^{(n)}$ sets a record is given by

$$
\begin{aligned}
p_n &= \int \mathbb{P}(\mathbf{X} \in d\mathbf{x})[1 - \mathbb{P}(\mathbf{X} \geq \mathbf{x})]^{n-1} = \int dF(\mathbf{x})[1 - H(-\mathbf{x})]^{n-1} \\
&= \int dH(\mathbf{y})[1 - H(\mathbf{y})]^{n-1} = \mathbb{E}[1 - H(-\mathbf{X})]^{n-1},
\end{aligned} \tag{1}
$$

where H denotes the distribution function corresponding to $-\mathbf{X}$.

Remark 1.5. For fixed d and n, the mapping from F (equivalently, from H) to p_n is many-to-one; recall Remark 1.3. In particular, p_n has the same value for all continuous F such that the coordinates of $\mathbf{X}$ are independent.

1.2. RP equivalence classes, RP ordering, and the main results

We see from (1) that the sequence $(p_n)_{n \geq 1}$ of record-setting probabilities is determined by $\mathcal{L}(H(-\mathbf{X}))$. Conversely, since the distribution of a bounded random variable is determined by its moments, $\mathcal{L}(H(-\mathbf{X}))$ is determined by $(p_n)_{n \geq 1}$. We are thus led to define an equivalence relation on (continuous) d-dimensional distribution functions F (for each fixed d) by declaring that $F \sim \widetilde{F}$ if $H(-\mathbf{X})$ and $\widetilde{H}(-\widetilde{\mathbf{X}})$ have the same distribution, where F, $\widetilde{F}$, H, and $\widetilde{H}$ are the distribution functions of $\mathbf{X}$, $-\mathbf{X}$, $\widetilde{\mathbf{X}}$, and $-\widetilde{\mathbf{X}}$, respectively; we call this the *record-setting probability* (RP) *equivalence*.

We are now prepared to define a partial order on the RP equivalence classes.

Definition 1.6. *Let C and $\widetilde{C}$ be RP equivalence classes with (arbitrarily chosen) respective representatives F and $\widetilde{F}$. We say that $C \leq \widetilde{C}$ in the RP ordering (or, by abuse of terminology, that $F \leq \widetilde{F}$ in the RP ordering) if $H(-\mathbf{X}) \geq \widetilde{H}(-\widetilde{\mathbf{X}})$ stochastically.*

Remark 1.7. From (1) it follows immediately that if $F \leq \widetilde{F}$ in the RP ordering, then $p_n \leq \tilde{p}_n$ for every n.

Let C^* denote the RP equivalence class corresponding to independent coordinates. We next introduce new notions of negative dependence and positive dependence; we relate these notions to more standard notions later, in Section 4.

Definition 1.8. *We will say that F is negatively record-setting probability dependent (NRPD) if its RP equivalence class C satisfies $C \geq C^*$ in the RP ordering.*

Definition 1.9. *We will say that F is positively record-setting probability dependent (PRPD) if its RP equivalence class C satisfies $C \leq C^*$ in the RP ordering.*

Remark 1.10. Thus any F having independent coordinates is both NRPD and PRPD.

We can now state our two main results. For both, let $p_n(F) \equiv p_{n,d}(F)$ denote the probability that the nth observation $\mathbf{X}^{(n)}$ from the (continuous) distribution F sets a record, and let p_n^* denote the value when $F \in C^*$.

Theorem 1.11 *For each fixed $d \geq 2$ and $n \geq 1$ the image of the mapping p_n on the domain of NRPD distributions is precisely the interval $[p_n^*, 1]$.*

Theorem 1.12 *For each fixed $d \geq 1$ and $n \geq 1$ the image of the mapping p_n on the domain of PRPD distributions is precisely the interval $[n^{-1}, p_n^*]$.*

Remark 1.13.

(a) For $d = 1$ and $n \geq 2$ the conclusion of Theorem 1.11 is false, since then $p_{n,1}(F) \equiv n^{-1}$.
(b) For $n = 1$ the results of Theorems 1.11–1.12 are trivial, since we have $p_{1,d}(F) \equiv 1$; so in proving the theorems we may assume $n \geq 2$.

Corollary 1.14. *For fixed $d \geq 2$ and $n \geq 1$ the image of the mapping p_n on the domain of all continuous distributions F is precisely the interval $[n^{-1}, 1]$, irrespective of d.*

We outline here the strategy, as illustrated in Figure 1 and carried out in Section 7, for proving Theorems 1.11–1.12 (and subsequently Corollary 1.14). Let $\mathcal{R}_N$ and $\mathcal{R}_P$ denote the respective images. It is immediate from our definitions that $\mathcal{R}_N \subseteq [p_n^*, 1]$ and $\mathcal{R}_P \subseteq [0, p_n^*]$, and by considering just first coordinates (see Lemma 2.2) we quickly narrow the latter to $\mathcal{R}_P \subseteq [n^{-1}, p_n^*]$. To show the reverse containments, we then fill the interval $[p_n^*, 1]$ with elements of $\mathcal{R}_N$ by choosing distribution functions F from a certain class of marginalized-Dirichlet distributions and their weak limits, and we fill the interval $[n^{-1}, p_n^*]$ with elements of $\mathcal{R}_P$ by choosing distribution functions F from a certain class of distributions with positively associated (PA) coordinates (more specifically, certain scale mixtures of i.i.d. Exponential distributions) and their weak limits.
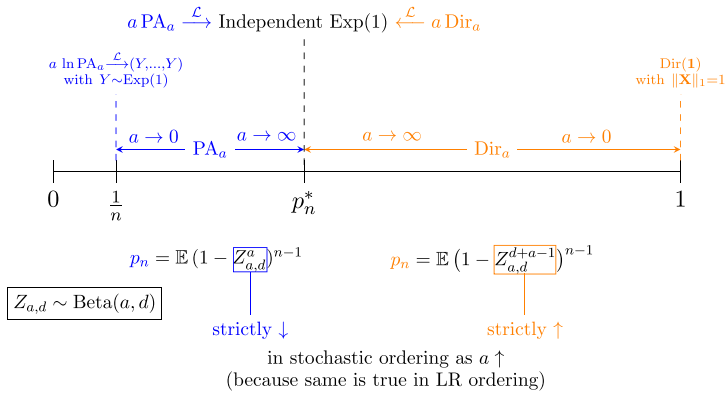
Figure 1. *The strategy for proving Theorems 1.11–1.12; here the random variable "PA_a" has the PA distribution $\widehat{F}_a$ described in Section 6.*

1.3. Brief literature review

Let us mention some related literature concerning Pareto records; we continue to assume F is continuous throughout this review. The book [1] is a standard reference for univariate records (the case $d = 1$). For multivariate records in the case of independent coordinates, we have already remarked that the record-setting probability $p_n = p_n^*$ does not depend on the distributions of the individual coordinates, but other aspects (such as the location of remaining records) do. Usually, as in [2] (see also the references therein), the coordinates are taken to be i.i.d., either Uniform$(0, 1)$ or standard Exponential. Bai et al. [2] obtain, for fixed d and for both R_n and r_n, asymptotic expansions as $n \to \infty$ for the expected value and variance and a central limit theorem with a Berry–Esseen bound. The main contributions of Fill and Naiman [12] are localization theorems for the *Pareto frontier*—that is, the topological boundary between the record-setting region and its complement when coordinates are i.i.d. standard Exponential—and some of those theorems are substantially sharpened in [13]. An importance-sampling algorithm for sampling records is presented, and partially analyzed, in [11]. A limiting distribution (again, for fixed d as $n \to \infty$) is established for the number $r_{n-1} + 1 - r_n$ of remaining records broken by $\mathbf{X}^{(n)}$ conditionally given that $\mathbf{X}^{(n)}$ sets a record, for $d = 2$ in [9] and for general d in [10].

An underlying theme of the present paper is that it is interesting to see how results (for example, concerning asymptotics for moments and distributions for R_n and r_n and localization of the frontier) vary with F. When F is the uniform distribution on the d-dimensional simplex, Hwang and Tsai [14] (see also the references therein, especially Bai et al. [3]) proceed in a fashion similar to that in [2] to obtain analogues of the asymptotic results of that earlier paper. It is worth noting that the computations are more involved in the simplex case than in [2], in part because results about r_n no longer translate immediately to results about R_n since the use of so-called concomitants (see Remark 3.1) becomes more involved, and that the results are enormously different; indeed, for example, as noted in the last line of the table on p. 1,867 of [14], we have $\mathbb{E}\, r_n \sim (\ln n)^{d-1}/(d-1)!$ for independent coordinates while $\mathbb{E}\, r_n \sim \Gamma(1/d)\, n^{(d-1)/d}$ for uniform sampling from the d-simplex.

1.4. Organization

In Section 2, we record two simple but very useful general observations about the record-breaking probability p_n. In Section 3, we briefly review the special case of independent coordinates. In Section 4, we relate the notions of NRPD and PRPD to existing notions of negative and positive dependence. In Section 5, we introduce and treat a class of examples of NRPD distributions F closely related to Dirichlet distributions and in Section 6 we introduce and treat a class of PRPD examples that are

scale mixtures of i.i.d. Exponential coordinates. Finally, in Section 7 we prove Theorems 1.11–1.12 and Corollary 1.14 and make a few additional remarks concerning the variability of p_n.

1.5. *Manifesto*

In light of Theorems 1.11–1.12 (see also Figure 1 and the proof strategy discussed at the end of Section 1.2), we regard the marginalized-Dirichlet NRPD distributions and the scale-mixture PRPD distributions we will use to prove the theorems, if not as canonical examples, then at least as standard examples worthy of thorough consideration—in particular, to study how the behaviors of these examples vary with their associated parameter values. Accordingly, we regard this paper as a pilot study of sorts, and we are presently working to extend (most of) the results of references [2], [14], [12]–[13], [11], [9], and [10] to these two classes of examples.

2. The record-breaking probability p_n: general information

To carry out our proof strategy for Theorems 1.11–1.12, we first need a result that p_n is continuous as a function of $\mathcal{L}(\mathbf{X})$ at any continuous distribution on $\mathbb{R}^d$. For this result (Proposition 2.1), we do not need to assume that the distributions of the random vectors $\mathbf{X}(m)$ are continuous.

Proposition 2.1. *Fix $d \geq 1$ and $n \geq 1$. If $\mathbf{X}(m)$ converges in distribution to $\mathbf{X}$ having a continuous distribution, then the corresponding record-setting probabilities satisfy $p_n(m) \to p_n$ as $m \to \infty$.*

Proof. The distribution functions H_m of $-\mathbf{X}(m)$ and H of $-\mathbf{X}$ satisfy $H_m \implies H$. Moreover, H is continuous, so $H_m(\mathbf{y})$ converges to $H(\mathbf{y})$ uniformly in $\mathbf{y}$ [4, Problem 3 in Sect. 3 (only in first edition)] and hence (recalling that n here is fixed) $[1 - H_m(\mathbf{y})]^{n-1}$ converges to $[1 - H(\mathbf{y})]^{n-1}$ uniformly in $\mathbf{y}$. It follows that as $m \to \infty$ we have

$$p_n(m) = \int dH_m(\mathbf{y})[1 - H_m(\mathbf{y})]^{n-1} \to \int dH(\mathbf{y})[1 - H(\mathbf{y})]^{n-1} = p_n. \qquad \square$$

Our next result exhibits the smallest and largest possible values of p_n.

Lemma 2.2. *Fix $d \geq 2$ and $n \geq 1$. We always have $p_n \in [n^{-1}, 1]$, and $p_n = n^{-1}$ and $p_n = 1$ are both possible.*

Proof. If $X_1^{(n)}$ sets a one-dimensional record (which has probability n^{-1}), then $\mathbf{X}^{(n)}$ sets a d-dimensional record. Thus $p_n \geq n^{-1}$, and equality holds if Y has any continuous distribution on $\mathbb{R}$ and $\mathbf{X} = (Y, \ldots, Y)$.

At the other extreme, if $d \geq 2$ and (for example) $\mathbf{X} \geq \mathbf{0}$ has any continuous distribution (such as any Dirichlet distribution) satisfying $\|\mathbf{X}\|_1 = 1$, then $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots$ form an antichain in the partial order $\leq$ on $\mathbb{R}^d$, so $p_n = 1$. $\qquad \square$

For further general information about p_n (in addition to Theorems 1.11–1.12, of course), see Remark 7.1.

3. Independent coordinates: p_n^*

This brief section concerns the case where the coordinates of each observation are independent. As noted in Remark 1.5, p_n doesn't otherwise depend on F in this setting, so we may as well assume that the coordinates are i.i.d. Exponential(1). Then (writing p_n^* for p_n in this special case)

$$p_n^* = \int \mathbb{P}(\mathbf{X} \in d\mathbf{x})[1 - \mathbb{P}(\mathbf{X} \geq \mathbf{x})]^{n-1} = \int_{\mathbf{0} \leq \mathbf{x} \in \mathbb{R}^d} e^{-\|\mathbf{x}\|_1} \left(1 - e^{-\|\mathbf{x}\|_1}\right)^{n-1} d\mathbf{x}$$

$$= \int_0^\infty \frac{y^{d-1}}{(d-1)!} e^{-y} (1 - e^{-y})^{n-1} \, dy$$

$$= n^{-1} \sum_{j=1}^n (-1)^{j-1} \binom{n}{j} j^{-(d-1)} =: n^{-1} \widehat{H}_n^{(d-1)} (= n^{-1} \text{ when } d = 1).$$

Alternatively, as pointed out by a reviewer of a previous draft, the same expression can be obtained for p_n^* by applying the principle of inclusion–exclusion to $\mathbb{P}\left(\bigcup_{i=1}^{n-1} \{\mathbf{X}^{(n)} \leq \mathbf{X}^{(i)}\}\right)$.

The numbers

$$\widehat{H}_n^{(k)} = \sum_{j=1}^n (-1)^{j-1} \binom{n}{j} j^{-k}$$

appearing in the expression for p_n^* are called *Roman harmonic numbers*, studied in [17], [18], and [20]. This $\widehat{H}_n^{(k)}$ can be written as a positive linear combination of products of generalized harmonic numbers $H_n^{(r)} := \sum_{j=1}^n j^{-r}$. In particular, $\widehat{H}_n^{(1)} = H_n^{(1)}$.

Remark 3.1.

(a) In obvious notation, the numbers $\mathbb{E} \, r_{n,d}^* = n p_{n,d}^* = \mathbb{E} \, R_{n,d-1}^*$ increase strictly in n for fixed $d \geq 2$, with limit ∞ as $n \to \infty$. (Note: The equality in distribution of the random variables $r_{n,d}$ and $R_{n,d-1}$ for general continuous F follows by standard consideration of concomitants: Consider $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ sorted according to the value of the dth coordinate.) Further, the numbers $p_{n,d}^*$ increase strictly in d for fixed $n \geq 2$, with limit 1 as $d \to \infty$; and decrease strictly in n for fixed $d \geq 1$, with limit 0 as $n \to \infty$.

(b) For fixed d we have

$$p_{n,d}^* \sim n^{-1} \frac{(\ln n)^{d-1}}{(d-1)!} \text{ as } n \to \infty.$$

Bai et al. [2] give a more extensive asymptotic expansion.

4. Negative dependence (including NRPD) and positive dependence (including PRPD)

In this section, we review existing notions of negative and positive dependence in Subsections 4.1–4.2 and relate our new notions of NRPD and PRPD to them in Subsection 4.3.

4.1. Negative dependence

For a discussion of several notions of negative dependence, see [15]. The first two notions in the next definition can be found there, with focus on the first notion (NA); we have created the third by interpolating between the first two.

Definition 4.1.

(a) *Random variables $X_1, \ldots, X_k$ are said to be negatively associated (NA) if for every pair of disjoint subsets A_1 and A_2 of $\{1, \ldots, k\}$ we have*

$$\text{Cov}\{f_1(X_i : i \in A_1), f_2(X_j : j \in A_2)\} \leq 0$$

whenever f_1 and f_2 are nondecreasing (in each argument) and the covariance is defined.

(b) *Random variables $X_1, \ldots, X_k$ are said to be negatively upper orthant dependent (NUOD) if for all real numbers $x_1, \ldots, x_k$ we have*

$$\mathbb{P}(X_i > x_i, \ i = 1, \ldots, k) \le \prod_{i=1}^{k} \mathbb{P}(X_i > x_i).$$

(c) *We say that random variables $X_1, \ldots, X_k$ are negatively upper orthant associated (NUOA) if for every pair of disjoint subsets A_1 and A_2 of $\{1, \ldots, k\}$ and all real numbers $x_1, \ldots, x_k$ we have*

$$\mathbb{P}(X_i > x_i, \ i = 1, \ldots, k) \le \mathbb{P}(X_i > x_i, \ i \in A_1) \, \mathbb{P}(X_j > x_j, \ j \in A_2).$$

Remark 4.2.

(a) NA implies NUOA, which implies NUOD.
(b) Theorem 2.8 in [15] gives a way of constructing NA $(X_1, \ldots, X_k)$, namely, if $G_1, \ldots, G_k$ are independent random variables with log-concave densities, then the conditional distribution of $\mathbf{G} = (G_1, \ldots, G_k)$ given $\sum_{j=1}^{k} G_j$ is NA almost surely.

4.2. Positive dependence

For a general discussion of various notions of positive dependence focusing on the one in the next definition, see [8].

Definition 4.3. *Random variables $X = (X_1, \ldots, X_k)$ are said to be positively associated (PA) (or simply associated) if*

$$\mathrm{Cov}\{f_1(X), f_2(X)\} \ge 0$$

whenever f_1 and f_2 are nondecreasing (in each argument) and the covariance is defined.

Remark 4.4. It is easy to show that if Z and $G_1, \ldots, G_d$ are independent positive random variables, then the scale mixture

$$\mathbf{X} := (ZG_1, \ldots, ZG_d)$$

is PA. The proof uses the law of total covariance (conditioning on Z), the fact [8, Thm. 2.1] that independent random variables are PA (applied to the conditional covariance), and the fact ([8, Property P$_3$], due to Chebyshev) that the set consisting of a single random variable is PA (applied to the covariance of the conditional expectations).

4.3. Relation with NRPD and PRPD

Our motivation for regarding NRPD and PRPD of respective Definitions 1.8 and 1.9 as notions of negative and positive dependence, respectively, is the following observation. One might suspect that NA implies NRPD and that PA implies PRPD; we are unable to prove either implication, but we can prove the weaker results (recall Remark 1.7) that NA implies $p_2 \ge p_2^*$ and PA implies $p_2 \le p_2^*$.

To establish the claimed inequalities, in the following proof replace $*$ by $\le$ if the observations are PA, by $=$ if they have independent coordinates, and by $\ge$ if they are NA. The claim is that $p_2 * 1 - 2^{-d}$. To see this, recall (1). We then have

$$p_2 = \int \mathbb{P}(\mathbf{X} \in d\mathbf{x}) [1 - \mathbb{P}(\mathbf{X} \ge \mathbf{x})] = 1 - \int \mathbb{P}(\mathbf{X} \in d\mathbf{x}) \, \mathbb{P}(\mathbf{X} \ge \mathbf{x})$$

$$* \; 1 - \int \mathbb{P}(\mathbf{X} \in d\mathbf{x}) \prod_{j=1}^{d} \mathbb{P}(X_j \ge x_j)$$

$$* \; 1 - \int \prod_{j=1}^{d} [\mathbb{P}(X_j \in dx_j) \, \mathbb{P}(X_j \ge x_j)]$$

$$= 1 - \prod_{j=1}^{d} \int [\mathbb{P}(X_j \in dx) \, \mathbb{P}(X_j \ge x)]$$

$$= 1 - \prod_{j=1}^{d} \int [\mathbb{P}(X_j \in dx) \, \mathbb{P}(X_j^* \ge x | X_j = x))]$$

where X_j^* is an independent copy of X_j

$$= 1 - \prod_{j=1}^{d} \int [\mathbb{P}(X_j \in dx) \, \mathbb{P}(X_j^* \ge X_j | X_j = x))]$$

$$= 1 - \prod_{j=1}^{d} \mathbb{P}(X_j^* \ge X_j) = 1 - \prod_{j=1}^{d} 2^{-1} = 1 - 2^{-d}.$$

5. Marginalized-Dirichlet distributions F_a: strict decreasing monotonicity in the RP ordering

The following is the usual definition of Dirichlet distribution, where the normalizing constant is the multivariate beta-function value $B(b) := \frac{\prod_{j=1}^{k} \Gamma(b_j)}{\Gamma(\|\mathbf{b}\|_1)}$.

Definition 5.1. *Let $k \ge 2$ and $\mathbf{b} = (b_1, \ldots, b_k) > 0$. If $\mathbf{Y} = (Y_1, \ldots, Y_k) > \mathbf{0}$ satisfies $\|\mathbf{Y}\|_1 = 1$ and $(Y_1, \ldots, Y_{k-1})$ has $(k-1)$-dimensional density (with respect to Lebesgue measure)*

$$\frac{1}{B(b)} y_1^{b_1 - 1} \cdots y_k^{b_k - 1} \mathbf{1}(y_1 > 0, \ldots, y_k > 0)$$

with $y_k := 1 - \sum_{j=1}^{k-1} y_j$, then we say that $\mathbf{Y}$ has the Dirichlet($\mathbf{b}$) distribution.

We will have special interest in taking $\mathbf{X} = (X_1, \ldots, X_d)$ to be the first d coordinates of $(Y_1, \ldots, Y_{d+1}) \sim \text{Dirichlet}(1, \ldots, 1, a)$; we denote the distribution of $\mathbf{X}$ in this case by Dir_a and the corresponding distribution function by F_a; we refer to the distributions Dir_a as *marginalized-Dirichlet distributions*. We will have occasional interest in taking

$$\mathbf{X} = (X_1, \ldots, X_d) \sim \text{Dirichlet}(1, \ldots, 1) =: \text{Dir}(\mathbf{1}). \tag{2}$$

Remark 5.2.

(a) When $a = 1$, the vector $\mathbf{X}$ is uniformly distributed in the (open) d-dimensional unit simplex

$$\mathcal{S}_d := \{\mathbf{x} = (x_1, \ldots, x_d) : x_j > 0 \text{ for } j = 1, \ldots, d \text{ and } \|\mathbf{x}\|_1 < 1\}. \tag{3}$$

This special case is the focus of [14].

(b) We find explicit computation (exact or asymptotic) of p_n intractable for general Dirichlet distributions.

Dirichlet distributions exhibit negative dependence among the coordinates according to standard notions [15]:

Remark 5.3.

(a) The distribution F_a is NUOA (recall Definition 4.1(c)) for every $a \in (0, \infty)$, by a simple calculation.

(b) The distribution F_a is NA if $a \geq 1$. Indeed, as in Definition 5.1, let $\mathbf{b} = (b_1, \ldots, b_k) > \mathbf{0}$. The proof (recall Remark 4.2(b)) that Dirichlet($\mathbf{b}$) is NA when $b_j \geq 1$ for every j relies on the following two standard facts:

   (i) If $G_j \sim \text{Gamma}(b_j)$ are independent random variables ($j = 1, \ldots, k$), then $\|\mathbf{G}\|_1 \sim \text{Gamma}(\|\mathbf{b}\|_1)$ and

$$\mathbf{Y} := \left(\frac{G_1}{\|\mathbf{G}\|_1}, \ldots, \frac{G_k}{\|\mathbf{G}\|_1} \right) \sim \text{Dirichlet}(\mathbf{b})$$

   are independent.

   (ii) For any $b \geq 1$, the Gamma(b) density is log-concave.

Consider $F = F_a$. The cases $n = 1$ (with $p_n \equiv 1$) and $d = 1$ (where the choice of a is irrelevant) being trivial, in the following monotonicity result we consider only $n \geq 2$ and $d \geq 2$.

Proposition 5.4. *Fix $d \geq 2$ and $n \geq 2$, and let $F = F_a$, that is, $\mathbf{X} \sim \text{Dir}_a$. Then F_a is strictly decreasing in the RP ordering and therefore the probability $p_n(a) := p_n(F_a)$ that $\mathbf{X}^{(n)}$ sets a record is strictly decreasing in a.*

Proof. By successive integrations one finds

$$\mathbb{P}(\mathbf{X} \geq \mathbf{x}) = (1 - \|\mathbf{x}\|_1)^{d+a-1};$$

thus $H_a(-\mathbf{X}) = (1 - \|\mathbf{X}\|_1)^{d+a-1}$. Further, $1 - \|\mathbf{X}\|_1 \sim \text{Beta}(a, d)$, so the first assertion is an immediate consequence of Lemma 5.6 below, and the second assertion follows from Remark 1.7.   □

Before proceeding to Lemma 5.6, we remind the reader of the definition of the likelihood ratio partial ordering (specialized to our setting of random variables taking values in the unit interval) and its connection to the well-known stochastic ordering.

Definition 5.5. *Given two real-valued random variables S and T with respective everywhere strictly positive densities f and g with respect to Lebesgue measure on $(0, 1)$, we say that $S \leq T$ in the likelihood ratio (LR) ordering if $g(u)/f(u)$ is nondecreasing in $u \in (0, 1)$.*

As noted (for example) in [19, Sect. 9.4], if $S \leq T$ in the LR ordering, then $S \leq T$ stochastically.

Lemma 5.6. *Fix a real number $d > 1$, and let Z_a have the Beta(a, d) distribution. Then $W_a := Z_a^{d+a-1}$ is strictly increasing in the LR ordering, and therefore also in the stochastic ordering, as $a \in (0, \infty)$ increases.*

Proof. By elementary calculation, W_a has density g_a on $(0, 1)$ given by the following expression, with $c_a := (d + a - 1)B(a, d)$:

$$g_a(w) = c_a^{-1} w^{-(d-1)/(d+a-1)} \left(1 - w^{1/(d+a-1)}\right)^{d-1}$$
$$= c_a^{-1} \left(w^{-1/(d+a-1)} - 1\right)^{d-1}.$$

Letting $0 < a < b < \infty$ and setting $v := w^{-1/(d+b-1)}$ and

$$t := (d + b - 1)/(d + a - 1),$$

it then suffices to show for any fixed $t > 1$ that the ratio $(v - 1)/(v^t - 1)$ decreases strictly as v increases over $(1, \infty)$.

For this, we consider the log-ratio, whose derivative is $h(v)/[(v - 1)(v^t - 1)]$, where

$$h(v) := v^t - 1 - tv^{t-1}(v - 1); \tag{4}$$

so we need only show that $h(v) < 0$ for $v \in (1, \infty)$. Indeed, since

$$h'(v) = -t(t - 1)v^{t-2}(v - 1) < 0$$

for $v \in (1, \infty)$, we see that $h(v) < h(1) = 0$ for $v \in (1, \infty)$.   □

6. Positively associated $\widehat{F}_a$: strict increasing monotonicity in the RP ordering

Distributions on $\mathbb{R}^d$ with PA coordinates can be constructed in similar fashion to the marginalized-Dirichlet distributions F_a (recall Remarks 4.4 and 5.3(b)). Given $a > 0$, let $\widehat{F}_a$ denote the PA distribution of

$$\mathbf{X} = \left(\frac{G_1}{G}, \ldots, \frac{G_d}{G}\right) \text{ (scale mixture of i.i.d. Exponentials),}$$

where the random variables $G, G_1, \ldots, G_d$ are independent, $G \sim \text{Gamma}(a)$, and $G_j \sim \text{Exponential}(1) \equiv \text{Gamma}(1)$ for $j = 1, \ldots, d$.

Remark 6.1.

(a) Scale mixtures of a finite number of i.i.d. Exponential random variables appear in a study of finite versions of de Finetti's theorem [7, (3.11)].

(b) We find explicit computation (exact or asymptotic) of p_n intractable for general scale mixtures, let alone for general PA distributions.

Similarly to Proposition 5.4, in our positive-association example we have the following claim:

Proposition 6.2. *Fix $d \geq 2$ and $n \geq 2$, and let $F = \widehat{F}_a$. Then $\widehat{F}_a$ is strictly increasing in the RP ordering and therefore the probability $\hat{p}_n(a) := p_n(\widehat{F}_a)$ that $\mathbf{X}^{(n)}$ sets a record is strictly increasing in a.*

Proof. A simple computation for $\mathbf{x} \geq \mathbf{0}$ gives

$$\mathbb{P}(\mathbf{X} \geq \mathbf{x}) = (1 + \|\mathbf{x}\|_1)^{-a}$$

and thus $H_a(-\mathbf{X}) = (1 + \|\mathbf{X}\|_1)^{-a}$. Further, $(1 + \|\mathbf{X}\|_1)^{-1} = G/(G + \|\mathbf{G}\|_1) \sim \mathrm{Beta}(a, d)$, so the first assertion is an immediate consequence of the following lemma, and the second assertion follows from Remark 1.7.   □

Lemma 6.3. *Fix a real number $d > 1$, and let Z_a have the Beta(a, d) distribution. Then $\widehat{W}_a := Z_a^a$ is strictly decreasing in the LR ordering, and therefore also in the stochastic ordering, as $a \in (0, \infty)$ increases.*

Proof. By elementary calculation, $\widehat{W}_a$ has density $\hat{g}_a$ on $(0, 1)$ given by the following expression, with $c_a := a\mathrm{B}(a, d)$:

$$\hat{g}_a(w) = c_a^{-1} \left(1 - w^{1/a}\right)^{d-1}.$$

Letting $0 < a < b < \infty$ and setting $v := w^{1/a}$ and $t := a/b$, it then suffices to show for any fixed $t \in (0, 1)$ that the ratio $(1 - v^t)/(1 - v)$ decreases strictly as v increases over $(0, 1)$.

For this, we consider the log-ratio, whose derivative is

$$-h(v)/[(1 - v)(1 - v^t)],$$

where we again use the definition (4), but now for $v \in (0, 1]$ (and with $t \in (0, 1)$); so we need only show that $h(v) > 0$ for $v \in (0, 1)$. Indeed, since

$$h'(v) = -t(1 - t)v^{t-2}(1 - v) < 0$$

for $v \in (0, 1)$, we see that $h(v) > h(1) = 0$ for $v \in (0, 1)$.   □

7. Proofs of Theorems 1.11–1.12 and Corollary 1.14

We are now prepared to prove Theorems 1.11–1.12 and Corollary 1.14 according to the outline provided at the end of Section 1.2; see Figure 1.

Proof of Theorem 1.11 In light of Lemma 2.2, it suffices to show that the image of p_n on the domain of our marginalized-Dirichlet examples F_a is $(p_n^*, 1)$.

We can regard $p_n \equiv p_n(a)$ as a function on the domain $(0, \infty)$ corresponding to our Dirichlet index a. Since the density $f_a(x)$ corresponding to F_a at each fixed argument x is a continuous function of a, it follows from Scheffé's theorem (e.g., [5, Thm. 16.12]) that the corresponding distribution functions F_a are continuous in a in the topology of weak convergence. It then follows from Propositions 2.1 and 5.4 that the image in question is $(p_n(\infty-), p_n(0+))$.

But, as $a \to \infty$, it is easy to see that the density of a times an observation converges pointwise to the density for independent Exponentials. By Scheffé's theorem and Proposition 2.1, therefore, $p_n(\infty-) = p_n^*$.

To compute $p_n(0+)$, we first observe that the distribution of an observation $\mathbf{X}(a)$ from F_a is that of

$$\left(\frac{Y_1}{\|\mathbf{Y}\|_1 + G_a}, \ldots, \frac{Y_d}{\|\mathbf{Y}\|_1 + G_a} \right),$$

where $Y_1, \ldots, Y_d$ are standard Exponential random variables, G_a is distributed (unit-scale) Gamma(a), and all $d + 1$ random variables are independent. It follows easily that $\mathbf{X}(a)$ converges in distribution to the distribution Dir$(\mathbf{1})$ mentioned at (2) (for which $p_n = 1$, as mentioned in the proof of Lemma 2.2) as $a \to 0$. Thus, by Proposition 2.1, $p_n(0+) = 1$.   □

Proof of Theorem 1.12 In light of Lemma 2.2, it suffices to show that the image on the domain of our PA examples $\widehat{F}_a$ is (n^{-1}, p_n^*).

In this case we can regard $p_n \equiv \hat{p}_n(a)$ as a function on the domain $(0, \infty)$ corresponding to our Gamma index parameter a. The value of the density of an observation at a given point $\mathbf{x} \geq \mathbf{0}$ in $\mathbb{R}^d$ is

$$\frac{\Gamma(d + a)}{\Gamma(a)} (1 + \|\mathbf{x}\|_1)^{-(d+a)},$$

which is a continuous function of $a \in (0, \infty)$. It follows from Scheffé's theorem that the corresponding distribution functions $\widehat{F}_a$ are continuous in the topology of weak convergence. It then follows from Propositions 2.1 and 6.2 that the image in question is $(\hat{p}_n(0+), \hat{p}_n(\infty-))$.

But, as $a \to \infty$, it's easy to see that the density of a times an observation converges pointwise to the density for independent standard Exponentials. By Scheffé's theorem and Proposition 2.1, therefore, $\hat{p}_n(\infty-) = p_n^*$.

To compute $\hat{p}_n(0+)$, we can without changing $\hat{p}_n(a)$ take an observation $\widehat{\mathbf{X}}(a)$ to have coordinates that are a times the logarithms of those described in our PA example. According to [16, Thm. 1] and Slutsky's theorem, $\mathbf{X}(a)$ converges in distribution to $(Y, \ldots, Y)$, where Y is standard Exponential. By Proposition 2.1, therefore, $\hat{p}_n(0+) = n^{-1}$.   □

Proof of Corollary 1.14 The corollary follows immediately from Lemma 2.2 and Theorems 1.11–1.12. For a considerably simpler proof, one can use the fact (from Lemma 2.2) that there are distributions F_0 and F_1 satisfying $p_n(F_0) = n^{-1}$ and $p_n(F_1) = 1$ for every n. By defining F_q to be the $(1 - q, q)$ mixture of F_0 and F_1 for $q \in [0, 1]$, we see from Proposition 2.1 (since F_q is clearly continuous in q in the weak topology) and the intermediate value theorem that the image of p_n on the domain $\{F_q : q \in [0, 1]\}$ contains (and therefore by Lemma 2.2 equals) $[n^{-1}, 1]$.   □

Remark 7.1. We have now learned from Theorems 1.11–1.12 information about how p_n behaves as a function of the (continuous) distribution of $\mathbf{X}$. As a complement, we conclude this paper with general (and rather more mundane) information about how p_n behaves as a function of n and as a function of d.

(a) As already noted, from (1) it is apparent that p_n is nonincreasing in n. By the dominated convergence theorem,

$$p_n \downarrow p_\infty := \int_{\mathbf{x}:\, \mathbb{P}(\mathbf{X} \geq \mathbf{x}) = 0} \mathbb{P}(\mathbf{X} \in d\mathbf{x})$$

as $n \uparrow \infty$. For each fixed $d \geq 2$, the image of the mapping p_∞ on the domain of all continuous distributions on $\mathbb{R}^d$ is the entire interval $[0, 1]$. To see by example that $q \in [0, 1]$ is in the image, choose the distribution F of $\mathbf{X}$ to be the $(q, 1 - q)$-mixture of any Dirichlet distribution and any of our marginalized-Dirichlet distributions F_a.

(b) To make sense of the question of how p_n varies as a function of $d \in \{1, 2, \ldots\}$, one should specify a *sequence* of distributions, with the dth distribution being over $\mathbb{R}^d$. It is rather obvious that if

$d' < d$ and $\mathbf{X}(d')$ is obtained by selecting any deterministic set of d' coordinates from $\mathbf{X}(d)$, then $p_n(d') \le p_n(d)$; in this sense, $p_n(d)$ is nondecreasing in the dimension d.

Fix $n \ge 1$, and for any specified sequence (in d) of distributions of $\mathbf{X}(d)$ let $p_n(\infty) := \lim_{d\to\infty} p_n(d)$. The image of the mapping $p_n(\infty)$ on the domain of all sequences of continuous distributions is $[n^{-1}, 1]$. This follows easily from Corollary 1.14. Indeed, given $q \in [n^{-1}, 1]$, one can choose $\mathbf{X} = \mathbf{X}(2) = (X_1, X_2)$ giving $p_n(2) = q$ and then take $\mathbf{X}(d) = (X_1, X_1, \ldots, X_1, X_2)$ for every d.

For all of our standard examples (independent coordinates, our marginalized-Dirichlet distributions F_a, and our PA examples $\widehat{F}_a$) we have $p_n(\infty) = 1$. In light of our earlier results, it is sufficient to prove this for the PA examples. For that, since the $\mathrm{Beta}(a, d)$ distributions converge weakly to unit mass at 0 as $d \to \infty$, it follows from the consequence

$$\hat{p}_n(a) = \mathbb{E}(1 - Z_{a,d}^a)^{n-1} \text{ where } Z_{a,d} \sim \mathrm{Beta}\,(a, d) \tag{5}$$

of the proof of Proposition 6.2 that $p_n(\infty) = 1$.

References

[1]   Arnold, B.C., Balakrishnan, N. & Nagaraja, H.N. (1998). *Records*. New York: John Wiley & Sons, Inc., A Wiley-Interscience Publication.

[2]   Bai, Z.-D., Devroye, L., Hwang, H.-K. & Tsai, T.-H. (2005). Maxima in hypercubes. *Random Structures Algorithms* 27(3): 290–309.

[3]   Bai, Z.-D., Hwang, H.-K. & Tsai, T.-H. (2003). Berry-Esseen bounds for the number of maxima in planar regions. *Electronic Journal of Probability* 8: 1–26.

[4]   Billingsley, P. (1968). *Convergence of probability measures*. New York-London-Sydney: John Wiley & Sons, Inc.

[5]   Billingsley, P. (2012). *Probability and measure*. Hoboken, NJ: John Wiley & Sons, Inc., Wiley Series in Probability and Statistics, Anniversary edition, with a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.

[6]   Chow, Y.S. & Teicher, H. (1997). *Probability theory*, 3rd ed. New York: Springer-Verlag. Independence, interchangeability, martingales.

[7]   Diaconis, P. & Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. *Annales de l'institut Henri Poincaré Probability and Statistics* 23(2): 397–423.

[8]   Esary, J.D., Proschan, F. & Walkup, D.W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics* 38(5): 1466–1474.

[9]   Fill, J.A. (2021). Breaking bivariate records. *Combinatorics, Probability and Computing* 30(1): 105–123.

[10]   Fill, J.A. (2023). Breaking multivariate records. *Electronic Journal of Probability* 28.

[11]   Fill, J.A. & Naiman, D.Q. (2019). Generating Pareto records. arXiv:1901.05621.

[12]   Fill, J.A. & Naiman, D.Q. (2020). The Pareto record frontier. *Electronic Journal of Probability* 25: 1–24.

[13]   Fill, J.A., Naiman, D.Q. & Sun, A. (2024). Sharpened localization of the trailing point of the Pareto record frontier. arXiv:2402.17221.

[14]   Hwang, H.-K. & Tsai, T.-H. (2010). Multivariate records based on dominance. *Electronic Journal of Probability* 15(60): 1863–1892.

[15]   Joag-Dev, K. & Proschan, F. (1983). Negative association of random variables, with applications. *Annals of Statistics* 11(1): 286–295.

[16]   Liu, C., Martin, R. & Syring, N. (2017). Efficient simulation from a gamma distribution with small shape parameter. *Computational Statistics* 32(4): 1767–1775.

[17]   Loeb, D.E. & Rota, G.-C. (1989). Formal power series of logarithmic type. *Advances in Mathematics* 75(1): 1–118.

[18]   Roman, S. (1992). The logarithmic binomial formula. *American Mathematical Monthly* 99(7): 641–648.

[19]   Ross, S.M. (1996). *Stochastic processes*, 2nd ed. New York: John Wiley & Sons, Inc.

[20]   Sesma, J. (2017). The Roman harmonic numbers revisited. *Journal of Number Theory* 180: 544–565.