A phase transition for the probability of being a maximum among random vectors with general iid coordinates

Royi Jacobovic* and Or Zuk

Dept. of Statistics and Data Science, the Hebrew University of Jerusalem

December 31, 2021

Abstract

Consider n iid real-valued random vectors of size k having iid coordinates with a general distribution function F. A vector is a maximum if and only if there is no other vector in the sample which weakly dominates it in all coordinates. Let $p_{k,n}$ be the probability that the first vector is a maximum. The main result of the present paper is that if $k \equiv k_n$ is growing at a slower (faster) rate than a certain factor of $\log(n)$, then $p_{k,n} \to 0$ (resp. $p_{k,n} \to 1$) as $n \to \infty$. Furthermore, the factor is fully characterized as a functional of F. We also study the effect of F on $p_{k,n}$, showing that while $p_{k,n}$ may be highly affected by the choice of F, the phase transition is the same for all distribution functions up to a constant factor.

1 Introduction

Consider a model with a sample of n iid random vectors of size k. It is assumed that the coordinates are iid real-valued random variables having a general distribution function F. A vector is said to be a (strong) maximum if and only if (iff) there is no other vector in the sample which (weakly) dominates it in all coordinates. Let $p_{k,n}$ be the probability that the first vector is a maximum. Once k (resp. n) is fixed, then $p_{k,n} \to 0$ (resp. $p_{k,n} \to 1$) as $n \to \infty$ (resp. $k \to \infty$). The main contribution of the present work is a generalization of this straightforward observation by allowing k to be determined as a function of n. Namely, we will show that if $k \equiv k_n$ is grows at a slower (resp. faster) rate than $\gamma \log(n)$, then $p_{k,n} \to 0$ (resp. $p_{k,n} \to 1$) as $n \to \infty$, where $\gamma \in (0, 1]$ is a certain constant that depends on the distribution F. The derivation of this result uses extreme value theory, and in particular relies on a result from of Ferguson [1] about the asymptotic behaviour of a maximum of iid sequence of geometric random variables.

The asymptotic behaviour of $p_{k,n}$ has an important role in many applications. For example, in analysis of linear programming [2] and of maxima-finding algorithms [3–7]. Furthermore, it is also related to game theory [8] and analysis of random forest algorithms [9, 10]. This literature focuses mainly on asymptotic results once F is a continuous function, k is fixed and n tends to infinity [8, 11–16]. Both [8] and [14] contain an approximation of the expected number of maxima. In addition, an approximation of the variance of the number of maxima is given in [11] and asymptotic normality of this number was proved in [12].

To the best of our knowledge, the only paper that includes asymptotic results as $n \to \infty$ and k is determined as a function of n is [16]. In the last equation of Section 1.1 of [16] there is a first order

*This author was supported by the GIF Grant 1489-304.6/2019.

approximation of $p_{k,n}$. This approximation holds uniformly for all possible forms of variations of k as a function of n which tends to infinity. In particular, it yields existence of a non-trivial phase-transition at $k \approx \log(n)$ which is consistent with our findings. While [16] refers only to a continuous F, the current results hold for a general F.

The rest is organized as follows: Section 2 contains a precise description of the model with a statement of the main result. In particular, the functional γ of F that determines the localization of the phase transition is presented (with the proof deferred to Section 4). Section 3 is devoted to exploring the effect of the distribution F on the probability $p_{k,n}$, with two important special cases: Section 3.1 is about the continuous case and includes a detailed discussion of the relation between the current results and the approximation which appears in [16]. Section 3.2 is about a simple example in which the coordinates have a Bernoulli distribution. This example illustrates two things:

1. While $p_{k,n}$ is the same for every continuous F, it might have a different asymptotic behaviour for fixed k as $n \to \infty$, once the continuity assumption is relaxed. Given these differences, it is a bit surprising that the phase-transition for $p_{k,n}$ is the same for all distribution functions up to the factor γ.
2. Even for a special case in which there is a simple exact combinatorial formula for $p_{k,n}$, it is unclear how to utilize this formula in order to derive the main result directly.

2 Model description and the main result

In the sequel, for every set A and a potential element a, denote the corresponding indicator function

$$\mathbf{1}_A(a) \equiv \begin{cases} 1, & a \in A, \\ 0, & a \notin A. \end{cases} \tag{1}$$

In addition, in several places of this manuscript we denote the minimum (resp. maximum) of some real numbers $x_1, x_2, \dots, x_n$ by $\bigwedge_i x_i \equiv \min_i x_i$ (resp. $\bigvee_i x_i \equiv \max_i x_i$).

2.1 Multivariate maximum

The following is a common definition of a maximum of a set of vectors in $\mathbb{R}^k$. It is based on the product order $\preceq$ on $\mathbb{R}^k$, *i.e.,* for every two vectors $a, b \in \mathbb{R}^k$ such that $a = (a_1, a_2, \dots, a_k)$ and $b = (b_1, b_2, \dots, b_k)$ define

$$a \preceq b \Leftrightarrow (a_i \leq b_i \ , \ \forall 1 \leq i \leq k) \ . \tag{2}$$

Similarly, define

$$a \prec b \Leftrightarrow (a \preceq b \ \textit{and} \ \exists i \in [k] \ \textit{s.t.} \ a_i < b_i) \ . \tag{3}$$

Definition 1 *Let $x_1, x_2, \dots, x_n$ be n vectors in $\mathbb{R}^k$. In addition, let $\preceq$ be the product order on $\mathbb{R}^k$. Then, for each $1 \leq i \leq n$, x_i is a maximum with respect to $x_1, x_2, \dots, x_n$ iff there is no $j \neq i$ such that $x_i \preceq x_j$. In addition, the set of maxima with respect to $x_1, x_2, \dots, x_n$ is called the Pareto-front generated by $x_1, x_2, \dots, x_n$.*

Remark 1 Definition 1 refers to a *strong* maximum. To see this, consider the special case in which $k = 1$ and $x_1 = x_2 = \dots = x_n$. In this case, $x_1, x_2, \dots, x_n$ are all maxima in the usual sense but none of them is a maximum in the sense of Definition 1. While this example illustrates a situation in which there is no maximum in the sense of Definition 1, it is possible to have multiple maxima in that sense. For instance assume that $n = k = 2$ and consider the case in which $x_1 = (1, 0)$ and $x_2 = (0, 1)$.

Remark 2 It is natural to introduce another notion of multivariate maximum: x_i is a *weak* maximum with respect to $x_1, x_2, \dots, x_n$ iff there is no $j \neq i$ such that $x_i \prec x_j$. Correspondingly, the set of weak maxima with respect to $x_1, x_2, \dots, x_n$ is called the *weak* Pareto-front generated by $x_1, x_2, \dots, x_n$. Later, in Section 3.2 we discuss this notion once the coordinates have a Bernoulli distribution.

2.2 Problem description

Let $\{X_{ij}; i,j \geq 1\}$ be an infinite array of iid real-valued random variables having a distribution function F. For every $i,k \geq 1$ denote $X_i^k \equiv (X_{i1}, \ldots, X_{ik})$ and for every $k,n \geq 1$, let $\mathcal{P}_{k,n}$ be the (random) set of all indices of vectors which belong to the Pareto-front generated by $X_1^k, X_2^k, \ldots, X_n^k$. An initial observation is that:

1. For every fixed $k \geq 1$, $\mathbf{1}_{\mathcal{P}_{k,n}}(1) \xrightarrow{n\to\infty} 0$, P-a.s.
2. For every fixed $n \geq 1$, $\mathbf{1}_{\mathcal{P}_{k,n}}(1) \xrightarrow{k\to\infty} 1$, P-a.s.

The main question is how to generalize this observation by characterizing the asymptotic behaviour of $\mathbf{1}_{\{\mathcal{P}_{k_n,n}\}}(1)$ as $n \to \infty$ for a general sequence $(k_n)_{n=1}^{\infty}$?

2.3 Main result

Let X be a random variable with cumulative distribution function F. Define the function $S : \mathbb{R} \to [0,1]$ as $S(x) \equiv P(X \geq x)$. When F is continuous, S is the corresponding survival function. Next, define

$$\gamma \equiv \gamma_F \equiv -E \log [S(X)] \tag{4}$$

and the following theorem is the main result. Its proof is given in Section 4.

Theorem 1 *Let $k_1, k_2, \ldots$ be a sequence of positive integers*

(a) *If*

$$\liminf_{n\to\infty} \frac{k_n}{\log(n)} > \gamma^{-1}, \tag{5}$$

then

$$\mathbf{1}_{\mathcal{P}_{k_n,n}}(1) \xrightarrow{n\to\infty} 1 \quad , \quad P\text{-a.s.} \tag{6}$$

(b) *If*

$$\limsup_{n\to\infty} \frac{k_n}{\log(n)} < \gamma^{-1}, \tag{7}$$

then

$$\mathbf{1}_{\mathcal{P}_{k_n,n}}(1) \xrightarrow{n\to\infty} 0 \quad , \quad P\text{-a.s.} \tag{8}$$

For every $k,n \geq 1$, denote

$$p_{k,n} \equiv P(1 \in \mathcal{P}_{k,n}) = E\mathbf{1}_{\mathcal{P}_{k,n}}(1). \tag{9}$$

Then, an application of bounded convergence theorem yields the following corollary.

Corollary 1 *Let $k_1, k_2, \ldots$ be a sequence of positive integers.*

(a')

$$\liminf_{n\to\infty} \frac{k_n}{\log(n)} > \gamma^{-1} \Rightarrow \lim_{n\to\infty} p_{k_n,n} = 1. \tag{10}$$

(b')

$$\limsup_{n\to\infty} \frac{k_n}{\log(n)} < \gamma^{-1} \Rightarrow \lim_{n\to\infty} p_{k_n,n} = 0. \tag{11}$$

2.4 The factor γ

Define

$$S^{-1}(y) \equiv \inf \{x \in \mathbb{R}; S(x) \leq y\} \quad , \quad y \in (0,1). \tag{12}$$

Since S is a nonincreasing leftcontinuous function, $S\left[S^{-1}(y)\right] \le y$ for every $y \in (0,1)$. By definition, $-\log\left[S(X)\right] \ge 0$ and hence is well-defined and nonnegative. Furthermore, the usual formula for an expectation of a nonnegative random variable yields that

$$\begin{aligned}\gamma &= \int_0^\infty P\left[-\log\left[S(X)\right] > t\right] dt \\ &= \int_0^\infty P\left[S(X) < e^{-t}\right] dt \\ &= \int_0^\infty P\left[X > S^{-1}\left(e^{-t}\right)\right] dt \\ &= \int_0^\infty S\left[S^{-1}\left(e^{-t}\right)\right] dt \\ &\le \int_0^\infty e^{-t} dt = 1\,.\end{aligned} \tag{13}$$

When F is continuous, the last inequality above holds with equality which yields that $\gamma = 1$. Moreover, $\gamma = 0$ if and only if $S \equiv 1$, which means that X is infinite. Thus, the assumption that X is real-valued implies that $\gamma \in (0,1]$. For example, when the coordinates have a Bernoulli(p) distribution for some $p \in (0,1)$,

$$S(x) = \begin{cases} 1, & x \le 0, \\ p, & 0 < x \le 1, \\ 0, & 1 < x. \end{cases} \tag{14}$$

Therefore,

$$\gamma = -p\log(S(1)) - (1-p)\log(S(0)) = -p\log(p) \tag{15}$$

and hence $\gamma = e^{-1} \approx 0.368$ is the maximal value of γ for the Bernoulli case, obtained at $p = e^{-1}$.

3 The effect of the distribution F

In this section we study the effect of the distribution F of the individual variables X_{ij}, on the distribution of the number of maxima. We specify the dependence on F explicitly, denoting $\mathcal{P}_{k,n}^{(F)}$ the (random) maximal set and $p_{k,n}^{(F)}$ the probability of being a maxima when $X_{ij} \sim F$. Similarly, we denote by $\mathcal{Q}_{k,n}^{(F)}$ the weak Pareto-front generated by $X_1^k, .., X_n^k$ (see Remark 2), and define

$$q_{k,n}^{(F)} \equiv P\left(1 \in \mathcal{Q}_{k,n}^{(F)}\right) = E\mathbf{1}_{\mathcal{Q}_{k,n}^{(F)}}(1)\,. \tag{16}$$

By definition $X_j^k \succ X_i^k \Rightarrow X_j^k \not\succcurlyeq X_i^k$, hence $\mathcal{P}_{k,n}^{(F)} \subseteq \mathcal{Q}_{k,n}^{(F)}$ and $p_{k,n}^{(F)} \le q_{k,n}^{(F)}$. In particular, when F is continuous, $\mathcal{P}_{k,n}^{(F)} = \mathcal{Q}_{k,n}^{(F)}$, P-a.s., hence $p_{k,n}^{(F)} = q_{k,n}^{(F)}$. In addition, $p_{k,n}^{(F)}$ is invariant to F as long as F is continuous, hence $p_{k,n} \equiv p_{k,n}^{(F)} = q_{k,n}^{(F)}$ without the specification of F will refer to a general continuous distribution.

Proposition 1 below shows that the continuous and the Bernoulli distributions are extreme cases, in the sense that for every distribution F, the probability of being a (strong) maxima lies between them. To shorten notation, for every $p \in (0,1)$, let $p_{k,n}^{(p)}$ be the probability of being a maximum once the coordinates have a Bernoulli(p) distribution.

Proposition 1 *Let $p_{k,n}^{(F)}$ be defined as above for a general F. Then,*

1. $p_{k,n}^{(F)} \le p_{k,n}$.
2. $p_{k,n}^{(p)} \le p_{k,n}^{(F)}$ *for every* $p \in \{1 - F(x); x \in \mathbb{R}\}$.

Proof:

1. Let F be an arbitrary probability distribution and let U be the distribution function of random variable which is uniformly distributed on $[0,1]$. The random variables $X_{ij} \sim F$ can be realized by taking uniform random variables $U_{ij} \sim U$, and then taking the transformation $X_{ij} = F^{-1}(U_{ij})$, where F^{-1} is the pseudo-inverse of F. Thus, since F^{-1} is nondecreasing we have $U_j^k \succcurlyeq U_i^k \Rightarrow X_j^k \succcurlyeq X_i^k$ and hence $X_i^k \in \mathcal{P}_{k,n}^{(F)} \Rightarrow U_i^k \in \mathcal{P}_{k,n}^{(U)}$. Therefore, $\mathcal{P}_{k,n}^{(F)} \subseteq \mathcal{P}_{k,n}^{(U)}$ and hence $p_{k,n}^{(F)} \leq p_{k,n}$.
2. Take x with $p \equiv 1 - F(x)$ and define $B_{ij} = \mathbf{1}_{\{X_{ij} > x\}}$. Since B_{ij} is defined as a nondecreasing transformation of X_{ij}, then $B_i^j \in \mathcal{P}_{k,n}^{(p)} \Rightarrow X_i^j \in \mathcal{P}_{k,n}^{(F)}$. As a result, $\mathcal{P}_{k,n}^{(p)} \subseteq \mathcal{P}_{k,n}^{(F)}$ and hence $p_{k,n}^{(p)} \leq p_{k,n}^{(F)}$. ■

Remark 3 While $p_{k,n}^{(F)} \leq p_{k,n}$ for any F (i.e. discretization may only reduce the probability of being a *strong* maximum), there is no general ordering that always holds between $q_{k,n}^{(F)}$ and $q_{k,n}$. This is demonstrated numerically for the Bernoulli distribution in Section 3.2.

Since the values $p_{k,n}^{(F)}$ for every distribution F of the X_{ij}'s can be bounded by the values for the continuous and Bernoulli case, we compare these two cases to study the effect of quantization on the probability of a random vector being a maximum.

3.1 Continuous distribution

For every $k, n \geq 1$, there are well-known exact formulas for $p_{k,n}$ (see e.g. [12]):

1.
$$p_{k,n} = \sum_{u=1}^{n} \binom{n-1}{u-1} \frac{(-1)^{u-1}}{u^k} . \tag{17}$$

2.
$$p_{k,n} = \begin{cases} \frac{1}{n} \sum_{u=1}^{n} p_{k-1,u}, & k > 1, \\ \frac{1}{n}, & k = 1, \end{cases} \tag{18}$$

and hence, for every $k > 1$ one has

$$p_{k,n} = \frac{1}{n} \sum_{u \in \mathcal{U}_{k,n}} \frac{1}{u_1 u_2 \ldots u_{k-1}} \tag{19}$$

where

$$\mathcal{U}_{k,n} \equiv \left\{ u = (u_1, \ldots, u_{k-1}) \in \mathbb{Z}^{k-1} \; ; \; 1 \leq u_1 \leq u_2 \leq \ldots \leq u_{k-1} \leq n \right\} . \tag{20}$$

Furthermore, it is well known (see, e.g., [14]) that for every fixed k,

$$p_{k,n} \sim \frac{\log^{k-1}(n)}{n(k-1)!} \quad \text{as} \quad n \to \infty . \tag{21}$$

For a fixed k, other asymptotic results results regarding the size of the Pareto-front as $n \to \infty$ include some asymptotic formulas for the variance [11] and a corresponding central limit theorem [12].

Hwang [16] applied analytic techniques (see, [17], [18]) to these identities in order to derive an approximation of $p_{k,n}$ as $n \to \infty$ and k is determined as a function of n. Specifically, denote the cumulative distribution function of a standard normal random variable by $\Phi(\cdot)$ and let $\Gamma(\cdot)$ be the Gamma function. Then, the first order approximation which appears in [16] is

$$p_{k,n} \sim \begin{cases} \frac{\log^{k-1}(n)}{n(k-1)!} \Gamma\left[1 - \frac{k}{\log(n)}\right], & \log(n) - k \gg \sqrt{\log(n)}, \\ \Phi\left[\frac{k - \log(n)}{\sqrt{\log(n)}}\right], & |k - \log(n)| = o\left[\log^{\frac{2}{3}}(n)\right], \\ 1, & \log(n) - k \ll \sqrt{\log(n)}, \end{cases} \tag{22}$$

and it holds uniformly for all variations of k as $n \to \infty$. Since $\gamma = 1$ for every continuous F, it may be verified that (22) implies Corollary 1. However, since convergence in P does not imply convergence P-a.s., it is not straightforward to deduce Theorem 1 from (22). In fact, Hwang [16] put forth the question of whether exists a probabilistic explanation for the phase-transition at $k \approx \log(n)$? Observe that while Theorem 1 yields some probabilistic explanation for this phenomenon, it does not supply a probabilistic proof of (22).

3.2 Bernoulli distribution

Let $X_{ij} \sim \text{Bernoulli}(p)$ for some $p \in (0,1)$. Let $B_1 = \sum_{j=1}^{k} X_{1j} \sim \text{Binom}(k,p)$ and without loss of generality assume that $X_{1j} = 1$ for every $1 \le j \le B_1$ and $X_{1j} = 0$ for every $B_1 + 1 \le j \le k$. By the law of total probability applied to B_1,

$$
\begin{aligned}
p^{(p)}_{k,n} &= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[P(X_1^k \not\preceq X_2^k | B_1 = i) \right]^{n-1} \\
&= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[1 - P\left(\bigwedge_{j=1}^{i} X_{2j} = 1 \right) \right]^{n-1} \\
&= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left(1 - p^i \right)^{n-1} . \qquad (23)
\end{aligned}
$$

Taking (23), we easily obtain the asymptotic result for the binary case for fixed k and $n \to \infty$. Since $(1-p^i)^{n-1} = o\left[(1-p^k)^{n-1}\right]$ for all $i < k$ as $n \to \infty$ we get that all terms in the above sum are negligible for large n except for the last, giving the result

$$
p^{(p)}_{k,n} \sim p^k (1-p^k)^{n-1} \quad \text{as} \quad n \to \infty \, . \qquad (24)
$$

A similar calculation to the one in (23) gives the probability of a weak maximum,

$$
\begin{aligned}
q^{(p)}_{k,n} &= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[P(X_1^k \not\prec X_2^k | B_1 = i) \right]^{n-1} \\
&= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[1 - P\left(\bigwedge_{j=1}^{i} X_{2j} = 1 \right) P\left(\bigvee_{j=i+1}^{k} X_{2j} = 1 \right) \right]^{n-1} \\
&= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left(1 - p^i + p^i (1-p)^{k-i} \right)^{n-1} , \qquad (25)
\end{aligned}
$$

and the asymptotic result $q^{(p)}_{k,n} \to p^k$ for fixed k as $n \to \infty$.

Remark 4 For any fixed k the decay of $p_{k,n} = q_{k,n}$ is sub-linear in n as $n \to \infty$ (see (21)). In contrast, $p^{(p)}_{k,n}$ decays to zero exponentially fast, whereas $q^{(p)}_{k,n}$ converges to a positive constant. The result is intuitive because for any fixed k the number of possible vectors in the binary case is finite, and the vector $(1,..,1)$ (with k coordinates) appears at least once P-a.s. as $n \to \infty$. A strong maximum may exist only if this vector appears at most once, an event with an exponentially small probability in n. Any occurrence of this vector is a *weak* maximum, yielding a positive probability not depending on n, $P\big(X_i^k = (1,..,1)\big) = p^k$.

Remark 5 While (23) is an exact combinatorial formula for $p^{(p)}_{k,n}$, it is not straightforward to analyze the behaviour of this combinatorial formula as $n \to \infty$ when k is determined as a general function of n. Theorem 1 gives us the asymptotic result for $p_{k,n}$ as $k, n \to \infty$ without relying on the exact expression.

For a complete treatment of the case in which the coordinates have Bernoulli(p) distribution, we derive here a combinatorial formula for the variance. Let $B_{ij} = \sum_{r=1}^{k} X_{1r}^i (1-X_{1j})^{1-i} X_{2r}^j (1-X_{2r})^{1-j}$ for $i,j = 0,1$. The quartet $(B_{00}, B_{01}, B_{10}, B_{11})$ has a multinomial distribution:

$$
(B_{00}, B_{01}, B_{10}, B_{11}) \sim \mathit{Multinomial}\Big(k, \big((1-p)^2, p(1-p), p(1-p), p^2\big)\Big). \qquad (26)
$$

Conditioning on their value yields a combinatorial formula for the probability that both X_1 and X_2 belong to the Pareto-front.

$$
\begin{aligned}
E\mathbf{1}_{\{1,2\in\mathcal{P}^{(p)}_{k_n,n}\}} &= \sum_{\substack{a,b,c,d\geq 0:\\ a+b+c+d=k}} \binom{k}{a\,b\,c\,d} \Big[P(X_1^k, X_2^k \npreceq X_3^k | B_{00}=a, B_{01}=b, B_{10}=c, B_{11}=d)\Big]^{n-2} \mathbf{1}_{\{b,c>0\}} \\
&= \sum_{\substack{a,d\geq 0\,;\,b,c\geq 1:\\ a+b+c+d=k}} \binom{k}{a\,b\,c\,d} \Bigg[1 - P\Bigg(\Big\{\{\bigwedge_{j=a+1}^{a+b} X_{3j}=1\} \bigcup \{\bigwedge_{j=a+b+1}^{k-d} X_{3j}=1\}\Big\} \bigcap \Big\{\bigwedge_{j=k-d+1}^{k} X_{3j}=1\Big\}\Bigg)\Bigg]^{n-2} \\
&= \sum_{\substack{a,d\geq 0\,;\,b,c\geq 1:\\ a+b+c+d=k}} \binom{k}{a\,b\,c\,d} \Big[1 - p^d(p^b + p^c - p^{b+c})\Big]^{n-2}. \qquad (27)
\end{aligned}
$$

and the variance is given by:

$$
V^{(p)}_{k,n} \equiv Var(|\mathcal{P}^{(p)}_{k,n}|) = np^{(p)}_{k,n}(1-p^{(p)}_{k,n}) + n(n-1)[E\mathbf{1}_{\{1,2\in\mathcal{P}^{(p)}_{k_n,n}\}} - {p^{(p)}_{k,n}}^2]. \qquad (28)
$$

Remark 6 When k is fixed and $n \to \infty$, both the expectation $np^{(p)}_{k,n}$ and variance $V^{(p)}_{k,n}$ approach to zero as $n \to \infty$, hence the limiting distribution of the Pareto-front size is degenerate. An interesting question for future work is whether there exists a sequence $k = k_n$ such that the limiting distribution of the Pareto-front size is non-degenerate.

3.3 Numerical Results

A numerical comparison between the Bernoulli and continuous case is shown in Figure 1. The difference in the asymptotic behaviour between $p^{(p)}_{k,n}, q^{(p)}_{k,n}$ and $p_{k,n}(= q_{k,n})$ for fixed k as $n \to \infty$ is shown in Figure 1.a. A numerical demonstration for the different behaviour of $p_{k_n,n}$ for $k_n = c\log(n)$ when $c < 1$ and $c > 1$ is shown in Figure 1.b. Similarly, the phase transition for Bernoulli(0.5) is presented in Figure 1.c, illustrating the phase transition at $\gamma = \frac{1}{2}\log(2)$, compared to $\gamma = 1$ for the continuous case.

Furthermore, as we have already shown, for fixed k the asymptotic behaviours of $p^{(p)}_{k,n}$ and $q^{(p)}_{k,n}$ as $n \to \infty$ are very different. However, when both $k, n \to \infty$, Figure 1.c suggests that the phase transition established by Theorem 1 for $p^{(p)}_{k,n}$ also holds for $q^{(p)}_{k,n}$. This issue may be developed as part of a future research.

For numerical calculation of $p_{k,n}$ we have used the recurrence relation (18), because the alternating sum in the combinatorial formula (17) causes numerical instabilities. As a result, computing $p_{k,n}$ for fixed k requires $O(n)$ operations, and $p_{k.n}$ was calculated for values up to $n = 10^7$ in Figure 1.b. In contrast, the discrete combinatorial formula (23) for $p^{(p)}_{k,n}$ can be applied directly, enabling us to compute this probability for much larger values of n (up to $n \approx 10^{130}$) in Figure 1.c.

The code for all numeric calculations is freely available at https://github.com/orzuk/Pareto.

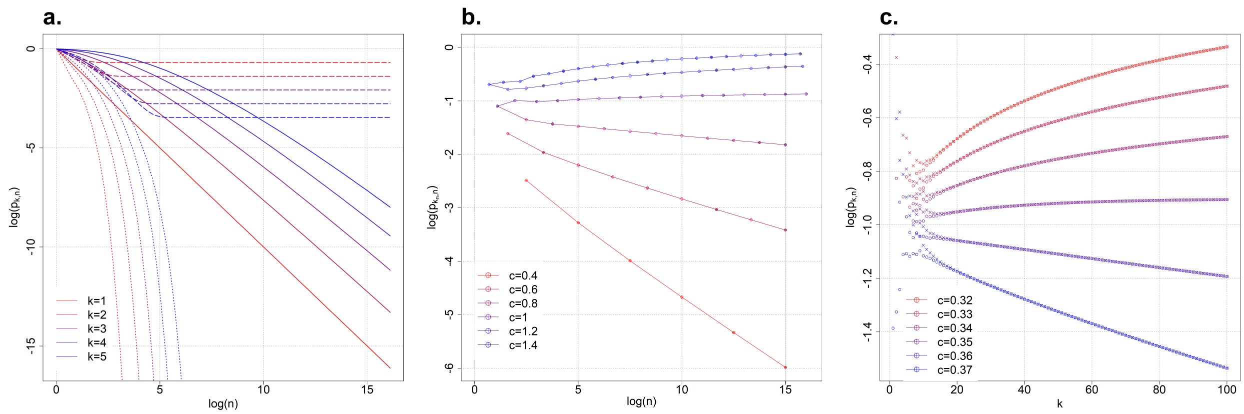

Figure 1: **a.** Value of $p_{k,n} = q_{k,n}$ (solid lines), $q_{k,n}^{(0.5)}$ (dashed lines) and $p_{k,n}^{(0.5)}$ (dotted lines) as a function of n, shown on a log-scale, for $k = 1, .., 5$. While $p_{k,n}^{(n)} < p_{k,n}$ for all k and n, when n is large $q_{k,n}^{(0.5)}$ can exceed $p_{k,n}$. **b.** Value of $\log(p_{k_n,n})$ for the continuous case using the exact combinatorial formula (line-connected circles) for $k_n = \lfloor (c \log(n) \rfloor$ for n from 1 to 10^7 and k_n up to $\lfloor (c \log(10^7) \rfloor$ for each c. Numerically, we were able to compute $p_{k,n}$ accurately only for small values of k, due to the recurrence relation in (18) and the alternating sum in (17). For $c \leq 0.8$ the curves decrease with n, consistent with our result that $p_{k_n,n} \to 0$ for this case. For $c \geq 1.2$ the curves increase towards zero with n , consistent with our result that $p_{k_n,n} \to 1$ for this case. For $c = 1$ there seems to be a slight increase in $p_{k_n,n}$ too, but results are inconclusive. **c.** Value of $\log(p_{k_n,n}^{(0.5)})$ ('x' symbols) and $\log(q_{k_n,n}^{(0.5)})$ ('o' symbols) for the Bernoulli(0.5) case, for $k_n = c \log(n)$ for different values of c. For $c < \gamma = \frac{\log(2)}{2} = 0.34657$ the log-probabilities approach 0, whereas for $c > \gamma$ the log-probabilities decreases to $-\infty$. For all values of c, the ratio $\frac{q_{k_n,n}^{(0.5)}}{p_{k_n,n}^{(0.5)}}$ approaches 1 as $n \to \infty$.

4 Proof of Theorem 1

For every $i \geq 2$, let

$$G_i^1 \equiv \min \{k \geq 1; X_{ik} > X_{1k}\} - 1 \,. \tag{29}$$

Observe that $X_i^k \preceq X_1^k$ for every $1 \leq k \leq G_i^1$ and $X_i^k \npreceq X_1^k$ for every $k > G_j^1$. In particular, this implies that for every $n, k \geq 1$,

$$1 \in \mathcal{P}_{k,n} \Leftrightarrow M_n^1 \equiv \max_{2 \leq i \leq n} G_i^1 \leq k \tag{30}$$

with the convention that a maximum over an empty-set of numbers equals zero. Thus, the asymptotic behaviour of $\mathbf{1}_{\mathcal{P}_{k,n}}(1)$ as $n, k \to \infty$ is strongly related to the asymptotic behaviour of M_n^1 as $n \to \infty$. Observe that M_n^1 is a maximum of $n-1$ identically distributed *dependent* geometric random variables $G_2^1, G_3^1, \ldots, G_n^1$ having a success probability $P\,(X_{11} > X_{21})$. The following lemma couples M_n^1 with a maximum of $n-1$ *independent* geometric random variables.

Lemma 1 *Let $G_2, G_3, \ldots$ be an iid sequence of geometric random variables with success probability $\alpha \in (0,1)$. For every $n \geq 1$, denote $M_n \equiv M_n^{(\alpha)} \equiv \max_{2 \leq i \leq n} G_i$, and assume that $\{G_i; i \geq 2\}$ and $\{X_{ij}; i,j \geq 1\}$ are independent.*

1. *If $1 - \alpha < e^{-\gamma}$, then there exists a P-a.s. finite random variable N_α such that*

$$M_n \leq M_n^1 \quad , \quad \forall n \geq N_\alpha \,. \tag{31}$$

2. *If $1 - \alpha > e^{-\gamma}$, then there exists a P-a.s. finite random variable N_α such that*

$$M_n \geq M_n^1 \quad , \quad \forall n \geq N_\alpha \,. \tag{32}$$

Proof: For every $k \geq 1$ denote

$$\tau_k^1 \equiv \min \left\{ i \geq 2; M_i^1 \geq k \right\} \quad , \quad \tau_k \equiv \min \{ i \geq 2; M_i \geq k \} \,. \tag{33}$$

Conditioned on $X_1 \equiv (X_{1j})_{j=1}^{\infty}$, the events

$$\left\{X_i^k \preceq X_1^k\right\} \quad , \quad i \geq 2 \tag{34}$$

are independent. Therefore, the random variables τ_k^1 and τ_k are conditionally independent given X_1, such that

$$\tau_k^1 | X_1 - 1 \sim \text{Geo}\left(\prod_{j=1}^{k} S(X_{1j})\right) \tag{35}$$

and

$$\tau_k - 1 \sim \text{Geo}\left((1-\alpha)^k\right) . \tag{36}$$

In addition, as explained in Section 2.4, $S(X_{11}), S(X_{12}), \ldots$ are iid random variables and $-E \log S(X_{11}) = \gamma \in (0, 1]$. Therefore, by the strong law of large numbers

$$L_k \equiv \frac{1}{k} \sum_{j=1}^{k} [-\log S(X_{1j})] \xrightarrow{k\to\infty} \gamma \quad , \quad P\text{-a.s.} \tag{37}$$

and it follows that $e^{L_k} \xrightarrow{k\to\infty} e^{\gamma}$, P-a.s. and $e^{-kL_k} \xrightarrow{k\to\infty} 0$, P-a.s.

Consider the case where $1 - \alpha < e^{-\gamma}$. Then, (37) implies that there exists a P-a.s. finite random variable K_α which is uniquely determined by X_1 such that for every $k > K_\alpha$

$$(1-\alpha)e^{L_k} \leq \frac{1 + (1-\alpha)e^{\gamma}}{2} \equiv \zeta_\alpha < 1. \tag{38}$$

In addition, $e^{-kL_k} \leq 1$ for every $k \geq 1$. Therefore, by a well-known result about a minimum of two independent geometric random variables, deduce that

$$\begin{aligned}
\sum_{k=K_\alpha}^{\infty} P\left(\tau_k \leq \tau_k^1 \middle| X_1\right) &= \sum_{k=K_\alpha}^{\infty} P\left(\tau_k - 1 \leq \tau_k^1 - 1 \middle| X_1\right) \qquad (39)\\
&= \sum_{k=K_\alpha}^{\infty} \frac{(1-\alpha)^k}{(1-\alpha)^k + e^{-kL_k} - (1-\alpha)^k e^{-kL_k}} \\
&\leq \sum_{k=K_\alpha}^{\infty} \left[(1-\alpha)e^{L_k}\right]^k \\
&\leq \sum_{k=K_\alpha}^{\infty} \zeta_\alpha^k < \infty .
\end{aligned}$$

Thus, the Borel-Cantelli lemma implies that

$$P\left(\tau_k \leq \tau_k^1 \text{ , i.o } \middle| X_1\right) = 0 \quad , \quad P\text{-a.s.} \tag{40}$$

and hence

$$P\left(\tau_k \leq \tau_k^1 \text{ , i.o }\right) = E\left[P\left(\tau_k \leq \tau_k^1 \text{ , i.o } \middle| X_1\right)\right] = 0 \tag{41}$$

which yields the required result when $(1-\alpha)e^{\gamma} < 1$.

Assume that $1 - \alpha > e^{-\gamma}$. Then, applying similar arguments to those which appear above yield the existence of a P-a.s. finite random variable K_α such that for any $k > K_\alpha$:

$$(1-\alpha)e^{L_k} \geq \frac{1 + (1-\alpha)e^{\gamma}}{2} \equiv \zeta_\alpha \tag{42}$$

such that $\zeta_\alpha > 1$. In addition, for every $k \geq 1$, $(1-\alpha)^k \leq 1$ and hence

$$\begin{aligned}\sum_{k=K_\alpha}^{\infty} P\left(\tau_k \geq \tau_k^1 \middle| X_1\right) &= \sum_{k=K_\alpha}^{\infty} P\left(\tau_k - 1 \geq \tau_k^1 - 1 \middle| X_1\right) \\ &= \sum_{k=K_\alpha}^{\infty} \frac{e^{-kL_k}}{(1-\alpha)^k + e^{-kL_k} - (1-\alpha)^k e^{-kL_k}} \\ &\leq \sum_{k=K_\alpha}^{\infty} \left[(1-\alpha)e^{L_k}\right]^{-k} \\ &\leq \sum_{k=K_\alpha}^{\infty} \zeta_\alpha^{-k} < \infty\,.\end{aligned} \tag{43}$$

Thus, the claim follows from the Borél-Cantelli lemma using a similar argument as in the previous case. ∎

Proof of Theorem 1 (continuation)

It is possible to apply Lemma 1 in order to show that

$$\frac{M_n^1}{\log(n)} \xrightarrow{n\to\infty} \gamma^{-1} \quad , \quad P\text{-a.s.} \tag{44}$$

To this end, fix $\epsilon > 0$ and let $0 < \alpha_1, \alpha_2 < 1$ be such that

$$(1-\alpha_1)e^\gamma < 1 < (1-\alpha_2)e^\gamma$$

and

$$\left|\gamma^{-1} - [\log(1-\alpha_l)]^{-1}\right| < \frac{\epsilon}{2} \quad , \quad \forall l = 1, 2. \tag{45}$$

Now, consider two independent iid sequences $G_2^{(\alpha_1)}, G_3^{(\alpha_1)}, \ldots$ and $G_2^{(\alpha_2)}, G_3^{(\alpha_2)}, \ldots$ such that $G_1^{(\alpha_l)} \sim \text{Geo}(\alpha_l)$ for $l = 1, 2$. Respectively, define the corresponding sequences of partial maxima

$$M_n^{(\alpha_l)} \equiv \max_{2\leq i\leq n} G_i^{\alpha_l} \quad , \quad n \geq 2\,, \tag{46}$$

for each $l = 1, 2$ as described in the statement of Lemma 1. Then, Lemma 1 implies that there exists P-a.s. finite random variables N_{α_1} and N_{α_2} such that

$$M_n^{(\alpha_1)} \leq M_n^1 \leq M_n^{(\alpha_2)} \quad , \quad \forall n \geq \max(N_{\alpha_1}, N_{\alpha_2}) \equiv N\,. \tag{47}$$

Furthermore, Theorem 2 of [1] yields that for each $l = 1, 2$

$$\frac{M_n^{(\alpha_l)}}{\log(n)} \xrightarrow{n\to\infty} -[\log(1-\alpha_l)]^{-1} \quad , \quad P\text{-a.s.} \tag{48}$$

As a result, there exists a P-a.s. finite random variable $N^* \geq N$ such that for every $n \geq N^*$ one has

$$\gamma^{-1} - \frac{\epsilon}{2} \leq \frac{M_n^{(\alpha_1)}}{\log(n)} \leq \frac{M_n^1}{\log(n)} \leq \frac{M_n^{(\alpha_2)}}{\log(n)} \leq \gamma^{-1} + \frac{\epsilon}{2} \tag{49}$$

and hence (44) follows.

Therefore, (5) implies that

$$\liminf_{n\to\infty} \frac{k_n}{\log(n)} = \liminf_{n\to\infty} \frac{k_n}{\log(n)} \cdot \frac{\log(n)}{M_n^i} > 1 \quad , \quad P\text{-a.s.} \tag{50}$$

and hence (30) yields that $\mathbf{1}_{\mathcal{P}_{k_n,n}}(1) \xrightarrow{n\to\infty} 1$, P-a.s. Similarly, (7) implies that

$$\limsup_{n\to\infty} \frac{k_n}{\log(n)} = \limsup_{n\to\infty} \frac{k_n}{\log(n)} \cdot \frac{\log(n)}{M_n^i} < 1 \quad , \quad P\text{-a.s.} \tag{51}$$

and hence (30) yields that $\mathbf{1}_{\mathcal{P}_{k_n,n}}(1) \xrightarrow{n\to\infty} 0$, P-a.s. ∎

References

[1] Thomas S. Ferguson. On the asymptotic distribution of max and mex. *Statistical Papers*, 34(1):97–111, 1993.

[2] Charles Blair. Random inequality constraint systems with few variables. *Mathematical Programming*, 35(2):135–139, 1986.

[3] Wei-Mei Chen, Hsien-Kuei Hwang, and Tsung-Hsi Tsai. Maxima-finding algorithms for multidimensional samples: A two-phase approach. *Computational Geometry*, 45(1-2):33–53, 2012.

[4] Luc Devroye. A note on the expected time for finding maxima by list algorithms. *Algorithmica*, 23(2):97–108, 1999.

[5] Martin E Dyer and John Walker. Dominance in multi-dimensional multiple-choice knapsack problems. *Asia-Pacific Journal of Operational Research*, 15(2):159, 1998.

[6] Mordecai J Golin. A provably fast linear-expected-time maxima-finding algorithm. *Algorithmica*, 11(6):501–524, 1994.

[7] Tsung-Hsi Tsai, Hsien-Kuei Hwang, and Wei-Mei Chen. Efficient maxima-finding algorithms for random planar samples. *Discrete Mathematics & Theoretical Computer Science*, 6, 2003.

[8] Barry O'Neill. The number of outcomes in the pareto-optimal set of discrete bargaining games. *Mathematics of Operations Research*, 6(4):571–578, 1981.

[9] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016.

[10] Erwan Scornet, Gérard Biau, and Jean-Philippe Vert. Consistency of random forests. *The Annals of Statistics*, 43(4):1716–1741, 2015.

[11] Zhi-Dong Bai, Chern-Ching Chao, Hsien-Kuei Hwang, and Wen-Qi Liang. On the variance of the number of maxima in random vectors and its applications. *The Annals of Applied Probability*, 8(3):886–895, 1998.

[12] Zhi-Dong Bai, Luc Devroye, Hsien-Kuei Hwang, and Tsung-Hsi Tsai. Maxima in hypercubes. *Random Structures & Algorithms*, 27(3):290–309, 2005.

[13] Andrew D Barbour and A Xia. The number of two-dimensional maxima. *Advances in Applied Probability*, 33(4):727–750, 2001.

[14] Ole Barndorff-Nielsen and Milton Sobel. On the distribution of the number of admissible points in a vector random sample. *Theory of Probability & Its Applications*, 11(2):249–269, 1966.

[15] Yuliy Baryshnikov. Supporting-points processes and some of their applications. *Probability Theory and Related Fields*, 117(2):163–182, 2000.

[16] Hsien-Kuei Hwang. Phase changes in random recursive structures and algorithms. In *Probability, Finance and Insurance*, pages 82–97. World Scientific, 2004.

[17] Hsien-Kuei Hwang. Sur la repartition des valeurs des fonctions arithmetiques. le nombre de facteurs premiers d'un entier. *Journal of Number Theory*, 69(2):135–152, 1998.

[18] Hsien-Kuei Hwang. A poisson* geometric convolution law for the number of components in unlabelled combinatorial structures. *Combinatorics, Probability and Computing*, 7(1):89–110, 1998.