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

Roi Jacobovic* and Or Zuk

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

February 20, 2023

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 that 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 $k \equiv k_n$ grows 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 that (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.

*This author was supported by the GIF Grant 1489-304.6/2019.
\( p_{k,n} \to 1 \) 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 \( k \equiv k_n \) 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 of Ferguson [1] about the asymptotic behaviour of a maximum of an iid sequence of geometric random variables.

The asymptotic behaviour of \( p_{k,n} \) has an important role in many applications. For example, in the analysis of linear programming [2] and of maxima-finding algorithms [3–7]. Furthermore, it is also related to game theory [8] and the 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 approximation of \( p_{k,n} \). This approximation holds uniformly for all possible forms of variations of \( k \) as a function of \( n \), as \( n \to \infty \). 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 \( \gamma \) 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 that appears in [16]. Section 3.2 is about a simple example in which the coordinates have a Bernoulli distribution. This example illustrates two points:

1. While \( p_{k,n} \) is the same for every continuous \( F \), once the continuity assumption is relaxed changing the distribution \( F \) can change drastically the first-order asymptotic behaviour of \( p_{k,n} \) for fixed \( k \) as \( n \to \infty \). In contrast, when both \( k, n \to \infty \), the phase-transition for \( p_{k,n} \) is the same up to a multiplicative factor \( \gamma \) for all distribution
functions \( F \).

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

In addition, in several places of this manuscript we denote the minimum (resp. maximum) of some real numbers \( x_1, x_2, \ldots, x_n \) by \( \wedge_i x_i \equiv \min_i x_i \) (resp. \( \vee_i x_i \equiv \max_i x_i \)). In particular, when \( n = 2 \), then we simply write \( x_1 \wedge x_2 \) (resp. \( x_1 \vee x_2 \)).

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 \( a = (a_1, a_2, \ldots, a_k) \) and \( b = (b_1, b_2, \ldots, b_k) \) define

\[
a \preceq b \iff (a_i \leq b_i, \forall 1 \leq i \leq k). \tag{2}
\]

Similarly, define

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

**Definition 1** Let \( x_1, x_2, \ldots, x_n \) 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 \( x_1, x_2, \ldots, x_n \) iff there is no \( j \neq i \) such that \( x_i \preceq x_j \). In addition, the set of maxima with respect to \( x_1, x_2, \ldots, x_n \) is called the Pareto-front generated by \( x_1, x_2, \ldots, x_n \).

**Remark 1** Definition 1 refers to a strong maximum. To see this, consider the special case in which \( k = 1, n \geq 2 \) and \( x_1 = x_2 = \ldots = x_n \). In this case, \( x_1, x_2, \ldots, x_n \) are all maxima in the usual sense but none of them is a maximum in the sense of Definition 1.

**Remark 2** It is possible to have multiple maxima in the sense of Definition 1. For instance, assume that \( n = k = 2 \) and consider the case in which \( x_1 = (1, 0) \) and \( x_2 = (0, 1) \).
Remark 3 It is natural to introduce another notion of multivariate maximum: $x_i$ is a *weak* maximum with respect to $x_1, x_2, \ldots, x_n$ iff there is no $j \neq i$ such that $x_i \prec x_j$. Correspondingly, the set of weak maxima with respect to $x_1, x_2, \ldots, x_n$ is called the weak Pareto-front generated by $x_1, x_2, \ldots, x_n$. 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 $X_{ik}^k \equiv (X_{i1}, \ldots, X_{ik})$ and for every $k, n \geq 1$, let $\mathcal{P}_{k,n} \subset \{1, 2, \ldots, n\}$ be the random set of indices of all vectors that belong to the Pareto-front generated by the random vectors $X_{1}^k, X_{2}^k, \ldots, X_{n}^k$. Also, for every event $B$, denote the complement by $\overline{B}$. Then, observe that $1_{\mathcal{P}_{k,n}}(1)$ is equal to one iff the event

$$A_{k,n} \equiv \bigcap_{j=2}^{n} \{X_{i1}^k \preceq X_{ij}^k\}$$

occurs. Moreover, note that for every sequence $(k_n)_{n \geq 1}$ of positive integers, $1_{\mathcal{P}_{k,n}}(1) \xrightarrow{n \to \infty} 1$ (resp. $1_{\mathcal{P}_{k,n}}(1) \xrightarrow{n \to \infty} 0$) $P$-a.s. iff

$$P\left(\liminf_{n \to \infty} A_{k,n,n} \right) = 1$$

(resp. $P\left(\limsup_{n \to \infty} A_{k,n,n} \right) = 0$).

An initial observation is that:

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

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

2.3 Main result

Let $X$ be a random variable with a 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

$$\gamma \equiv \gamma_F \equiv -E \log [S(X)].$$

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
\[ \liminf_{n \to \infty} \frac{k_n}{\log(n)} > \gamma^{-1}, \]
then
\[ I_{p_{k,n}}(1) \xrightarrow{n \to \infty} 1, \text{P-a.s.} \] (8)

(b) If
\[ \limsup_{n \to \infty} \frac{k_n}{\log(n)} < \gamma^{-1}, \]
then
\[ I_{p_{k,n}}(1) \xrightarrow{n \to \infty} 0, \text{P-a.s.} \] (10)

For every \( k, n \geq 1 \), denote
\[ p_{k,n} \equiv P(A_{k,n}) = E[1_{p_{k,n}}(1)]. \] (11)

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') \[ \liminf_{n \to \infty} \frac{k_n}{\log(n)} > \gamma^{-1} \Rightarrow \lim_{n \to \infty} p_{k,n} = 1. \] (12)

(b') \[ \limsup_{n \to \infty} \frac{k_n}{\log(n)} < \gamma^{-1} \Rightarrow \lim_{n \to \infty} p_{k,n} = 0. \] (13)

### 2.4 The factor \( \gamma \)

Define
\[ S^{-1}(y) \equiv \inf \{ x \in \mathbb{R}; S(x) \leq y \} , \ y \in (0,1). \] (14)

Since \( S \) is a nonincreasing leftcontinuous function, \( S[S^{-1}(y)] \leq y \) for every \( y \in (0,1) \). By definition, \( -\log[S(X)] \geq 0 \) and hence is well-defined and nonnegative. Furthermore, the usual formula for an expectation of a nonnegative random variable yields that
\[ \gamma = \int_0^\infty P[-\log[S(X)] > t] \, dt \] (15)
\[ = \int_0^\infty P[S(X) < e^{-t}] \, dt \]
\[ = \int_0^\infty P[X > S^{-1}(e^{-t})] \, dt \]
\[ = \int_0^\infty S[S^{-1}(e^{-t})] \, dt \]
\[ \leq \int_0^\infty e^{-t} \, dt = 1. \]
When \( F \) is continuous, the last inequality above holds with equality and \( \gamma = 1 \). 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 \( p \in (0, 1) \),

\[
S(x) = \begin{cases} 
1, & x \leq 0, \\
p, & 0 < x \leq 1, \\
0, & 1 < x. 
\end{cases}
\]  

Therefore,

\[
\gamma = -p \log(S(1)) - (1 - p) \log(S(0)) = -p \log(p)
\]

and hence \( \gamma = e^{-1} \approx 0.368 \) is the maximal value of \( \gamma \) 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 \( P^{(F)}_{k,n} \) the (random) maximal set and \( p^{(F)}_{k,n} \) the probability of being a maxima when \( X_{ij} \sim F \). Similarly, we denote by \( Q^{(F)}_{k,n} \) the weak Pareto-front generated by \( X^k_1, \ldots, X^k_n \) (see Remark 3), and define

\[
q^{(F)}_{k,n} = P\left(1 \in Q^{(F)}_{k,n}\right) = E_{Q^{(F)}_{k,n}}(1).
\]

By definition \( X^k_j > X^k_i \Rightarrow X^k_j \succ X^k_i \), hence \( P^{(F)}_{k,n} \subseteq Q^{(F)}_{k,n} \) and \( p^{(F)}_{k,n} \leq q^{(F)}_{k,n} \).

In particular, when \( F \) is continuous, \( P^{(F)}_{k,n} = Q^{(F)}_{k,n} \), \( P \)-a.s., hence \( p^{(F)}_{k,n} = q^{(F)}_{k,n} \). Moreover, let \( U(\cdot) \) be the uniform distribution function on \([0, 1] \) and observe that every continuous \( F \) satisfies the relation

\[
p^{(U)}_{k,n} = p^{(F)}_{k,n} = q^{(F)}_{k,n} = q^{(U)}_{k,n}.
\]

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 \( p^{(p)}_{k,n} \) be the probability of being a maximum once the coordinates have a Bernoulli(\( p \)) distribution.

**Proposition 1** Let \( F \) be a general distribution function. Then,

1. \( p^{(F)}_{k,n} \leq p^{(U)}_{k,n} \).
2. \( p^{(p)}_{k,n} \leq p^{(F)}_{k,n} \) for every \( p \in \{1 - F(x); x \in \mathbb{R}\} \).

**Proof:**
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 $U_{ij} \geq U_{ik} \Rightarrow X_{ij} \geq X_{ik}$ and hence $X_{ik} \in \mathcal{P}_{k,n} \Rightarrow U_{ik} \in \mathcal{P}_{k,n}$. Therefore, $\mathcal{P}_{k,n} \subseteq \mathcal{P}_{k,n}$ and hence $p_{k,n} \leq p_{k,n}$.

2. Take $x$ with $p = 1 - F(x)$ and define $B_{ij} = 1 \{X_{ij} > x\}$. Since $B_{ij}$ is a nondecreasing transformation of $X_{ij}$, then $B_{ij} \in \mathcal{P}_{k,n} \Rightarrow X_{ij} \in \mathcal{P}_{k,n}$. As a result, $\mathcal{P}_{k,n} \subseteq \mathcal{P}_{k,n}$ and hence $p_{k,n} \leq p_{k,n}$.

**Remark 4** While $p_{k,n} \leq p_{k,n}$ 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 $q_{k,n}^{(U)}$. This is demonstrated numerically for the Bernoulli distribution in Section 3.3.

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 $p_{k,n}^{(U)}$ (see e.g. [12]):

1. 
   $$p_{k,n}^{(U)} = \sum_{u=1}^{n} \left( \frac{n-1}{u-1} \right) \frac{(-1)^{u-1}}{u^k}. \tag{20}$$

2. 
   $$p_{k,n}^{(U)} = \begin{cases} \frac{1}{n} \sum_{u=1}^{n} p_{k-1,u}^{(U)}, & k > 1, \\ \frac{1}{n}, & k = 1, \end{cases} \tag{21}$$

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

$$p_{k,n}^{(U)} = \frac{1}{n} \sum_{u \in \mathcal{U}_{k,n}} \frac{1}{u_1 u_2 \ldots u_{k-1}} \tag{22}$$

where

$$\mathcal{U}_{k,n} = \left\{ u = (u_1, \ldots, u_{k-1}) \in \mathbb{Z}^{k-1}, \ 1 \leq u_1 \leq u_2 \leq \ldots \leq u_{k-1} \leq n \right\}. \tag{23}$$

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

$$p_{k,n}^{(U)} \sim \frac{\log^{k-1}(n)}{n(k-1)!} \text{ as } n \to \infty. \tag{24}$$
For a fixed $k$, other asymptotic results regarding the size of the Pareto-front as $n \to \infty$ include asymptotic formulas for the variance \cite{11} and a corresponding central limit theorem \cite{12}.

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

$$p^{(U)}_{k,n} \sim \begin{cases} \log \frac{k-1}{n(k-1)!} \Gamma \left( 1 - \frac{k}{\log(n)} \right), & \log(n) - k \gg \sqrt{\log(n)}, \\ \Phi \left( \frac{k - \log(n)}{\sqrt{\log(n)}} \right), & |k - \log(n)| = o \left( \frac{k^2}{n} \right), \\ 1, & \sqrt{\log(n)} \ll k - \log(n), \end{cases}$$

(25)

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 \cite{25} implies Corollary \cite{1} However, since convergence in $P$ does not imply convergence $P$-a.s., it is not straightforward to deduce Theorem \cite{1} from \cite{25}, even for the continuous case. In fact, Hwang \cite{16} put forth the question of whether exists a probabilistic explanation for the phase-transition at $k \approx \log(n)$? Theorem \cite{1} yields some probabilistic explanation for this phenomenon, although it does not supply a probabilistic proof of \cite{25}.

### 3.2 Bernoulli distribution

In this part, we present an example of a distribution function $F$ for which it is possible to derive an explicit combinatorial expression of $p_{k,n}$. As to be shown, even when such an expression is available, still it is unclear how Theorem \cite{1} may be deduced from it (for this special case). Furthermore, this example demonstrates the possible differences between the model with a continuous $F$ versus discontinuous $F$.

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 $1 \leq j \leq B_1$ and $X_{1j} = 0$ for every $B_1 + 1 \leq j \leq k$. By the law of total
probability applied to $B_1$,

$$p_{k,n}^{(p)} = \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[ P(X_1^k \not\subseteq X_2^k | B_1 = i) \right]^{n-1}$$
$$= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[ 1 - P \left( \bigwedge_{j=1}^{k} X_{2j} = 1 \right) \right]^{n-1}$$
$$= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left( 1 - p \right)^{n-1}. \quad (26)$$

where in the last equation above, when $i = 0, n = 1$ the last term $(1-p)^{n-1} = 0^0$ is defined to be 1. The asymptotic behaviour of $p_{k,n}^{(p)}$ for fixed $k$ and $n \to \infty$ follows directly from (26). Since $(1-p)^{n-1} = o \left[ (1-p^k)^{n-1} \right]$ for all $i < k$ as $n \to \infty$, all terms in the above sum are negligible for large $n$ except for the last, giving the result

$$p_{k,n}^{(p)} \sim p^k (1-p^k)^{n-1} \quad \text{as} \quad n \to \infty. \quad (27)$$

Remark 5 While (26) is an exact combinatorial formula for $p_{k,n}^{(p)}$, 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.

A similar calculation to the one in (26) gives the probability of a weak maximum,

$$q_{k,n}^{(p)} = \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[ P(X_1^k \not\subseteq X_2^k | B_1 = i) \right]^{n-1}$$
$$= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left[ 1 - P \left( \bigwedge_{j=1}^{k} X_{2j} = 1 \right) P \left( \bigvee_{j=i+1}^{k} X_{2j} = 1 \right) \right]^{n-1}$$
$$= \sum_{i=0}^{k} \binom{k}{i} p^i (1-p)^{k-i} \left( 1 - p^i + p'(1-p)^{k-i} \right)^{n-1}. \quad (28)$$

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

Remark 6 For any fixed $k$ the decay of $p_{k,n}^{(U)} = q_{k,n}^{(U)}$ is sub-linear in $n$ as $n \to \infty$ (see (24)). In contrast, $p_{k,n}^{(p)}$ decays to zero exponentially fast, whereas $q_{k,n}^{(p)}$ converges to a positive constant. The result is intuitive because for any fixed $k$ the number of possible vectors in the Bernoulli case is finite, and the vector $(1, \ldots, 1)$ (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 limit positive probability not depending on $n$, $P(X_1^k = (1, \ldots, 1)) = p^k$. 9
For a complete treatment of the case in which the coordinates have Bernoulli($p$) distribution, we derive a combinatorial formula for the variance. For every $i, j \in \{0, 1\}$ define
\[ B_{ij} \equiv |\{1 \leq r \leq k; X_{1r} = i, X_{2r} = j\}|. \] (29)
and observe that vector $(B_{00}, B_{01}, B_{10}, B_{11})$ has a multinomial distribution, i.e.,
\[ (B_{00}, B_{01}, B_{10}, B_{11}) \sim \text{Multinomial}\left(k, \left((1-p)^2, p(1-p), p(1-p), p^2\right)\right). \] (30)
By conditioning on this random vector deduce that
\[
E\{1_{\{1, 2 \in P(p)_{k,n}^n\}}\} = \sum_{a, d \geq 0; b, c \geq 1; a+b+c+d=k} \left(\begin{array}{c} k \\ a, b, c, d \end{array}\right) \left[1 - P\left(\left\{\{ a+b \atop j=a+1 \} \cup \{ k-d \atop j=a+b+1 \} X_{3j} = 1 \right\} \bigcap \left\{ k \atop j=k-d+1 \} X_{3j} = 1 \right\}\right]^{n-2},
\] (31)
and the variance is given by:
\[
V_{k,n}^{(p)} \equiv \text{Var}(|P(p)_{k,n}|) = np_{k,n}^{(p)}(1-p_{k,n}^{(p)}) + n(n-1)E\{1_{\{1, 2 \in P(p)_{k,n}^n\}}\} - p_{k,n}^{(p)} \big|_{p_{k,n}^{(p)}}. \] (32)

Remark 7 When $k$ is fixed and $n \to \infty$, both the expectation $np_{k,n}^{(p)}$ and variance $V_{k,n}^{(p)}$ 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 $|P(p)_{k,n}|$ is non-degenerate.

Remark 8 In this part we have analyzed the relatively simple case when the underlying distribution is Bernoulli. Naturally, a follow-up question is about studying other distributions with the goal of comparing between the results.

3.3 Numerical Results

A numerical comparison between the Bernoulli and continuous cases is shown in Figure 1. The difference in the asymptotic behaviour between
$p_{k,n}^{(p)}$, $q_{k,n}^{(p)}$ and $p_{k,n}^{(U)} = q_{k,n}^{(U)}$ for fixed $k$ as $n \to \infty$ is shown in Figure 1.a. A numerical demonstration for the different behaviour of $p_{k,n}^{(U)}$ 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 localization 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 $p_{k,n}^{(p)}$ and $q_{k,n}^{(p)}$ 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 $p_{k,n}^{(p)}$ also holds for $q_{k,n}^{(p)}$. Comparing the two cases more rigorously is left for future research.

For numerical calculation of $p_{k,n}^{(U)}$ we have used the recurrence relation (21), because the alternating sum in the combinatorial formula (20) causes numerical instabilities. As a result, computing $p_{k,n}^{(U)}$ for fixed $k$ requires $O(n)$ operations, and $p_{k,n}^{(U)}$ was calculated for values up to $n = 10^7$ in Figure 1.b. In contrast, the discrete combinatorial formula (26) for $p_{k,n}^{(p)}$ 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 \( p_{k,n}^{(U)} = q_{k,n}^{(U)} \) (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 \( k = 1, 2, 3, 4, 5 \). While \( p_{k,n}^{(0.5)} < p_{k,n} \) for all \( k \) and \( n \), when \( n \) is large \( q_{k,n}^{(0.5)} \) can exceed \( p_{k,n}^{(U)} \). b. Value of \( \log(p_{k,n}^{(U)}) \) 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 \( c \). We were able to compute \( p_{k,n}^{(U)} \) accurately only for small values of \( k \), due to the recurrence relation in (21) and the alternating sum in (20). For \( c \leq 0.8 \) the curves decrease with \( n \), consistent with the result that \( p_{k,n}^{(U)} \rightarrow 0 \) for this case. For \( c \geq 1.2 \) the curves increase towards zero with \( n \), consistent with the result that \( p_{k,n}^{(U)} \rightarrow 1 \) for this case. For \( c = 1 \) there seems to be a slight increase in \( p_{k,n}^{(U)} \) too, but results are inconclusive. c. Value of \( \log(q_{k,n}^{(0.5)}) \) (’x’ symbols) and \( \log(q_{k,n}^{(0.5)}) \) (’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 \( \frac{q_{k,n}^{(0.5)}}{p_{k,n}} \) approaches 1 as \( n \rightarrow \infty \).

4 Proof of Theorem 1

For every \( i \geq 2 \), let

\[
G_i^1 \equiv \min \{ k \geq 1; X_{ik} < X_{1k} \} - 1. \tag{33}
\]

Then \( X_{1k} \leq X_{ik} \) for every \( 1 \leq k \leq G_i^1 \) and \( X_{1k} \neq X_{ik} \) for every \( k > G_i^1 \). In particular, this implies that for every \( n, k \geq 1 \),

\[
1 \in P_{k,n} \iff M_i^1 \equiv \max_{2 \leq i \leq n} G_i^1 \leq k - 1 \tag{34}
\]

with the convention that a maximum over an empty-set of numbers equals zero. Thus, the asymptotic behaviour of \( \mathbf{1}_{P_{k,n}}(1) \) as \( n, k \rightarrow \infty \) is strongly related to the asymptotic behaviour of \( M_i^1 \) as \( n \rightarrow \infty \). Observe that \( M_i^1 \) is a maximum of \( n - 1 \) identically distributed dependent geometric random variables \( G_1^1, G_2^1, \ldots, G_n^1 \) having a success probability \( P(X_{11} > X_{21}) \). The following lemma couples \( M_i^1 \) with a maximum of \( n - 1 \) independent geometric random variables.
Lemma 1 Let \( \{X_{ij}; i, j \geq 1\} \) be iid random variables with a distribution function \( F \) for which \( \gamma \equiv \gamma_F \) as defined in (6). In addition, 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.

Then, for every \( \alpha \in (0, 1) \setminus \{1 - e^{-\gamma}\} \), the random variable

\[
N_\alpha \equiv 1 + \begin{cases} 
\sup \{n \geq 1; M_n > M_n^1\} \lor 0 & 1 - \alpha < e^{-\gamma} \\
\sup \{n \geq 1; M_n < M_n^1\} \lor 0 & 1 - \alpha > e^{-\gamma}
\end{cases}
\]

is \( P \)-a.s. finite.

Remark 9 By definition, whenever \( 1 - \alpha < e^{-\gamma} \) (resp. \( 1 - \alpha < e^{-\gamma} \)), then \( M_n \leq M_n^1 \) (resp. \( M_n \geq M_n^1 \)) for every \( n \geq N_\alpha \).

Proof: For every \( k \geq 1 \) denote

\[
\tau_k^1 \equiv \min \{i \geq 2; M_i^1 \geq k\}, \quad \tau_k \equiv \min \{i \geq 2; M_i \geq k\}.
\]

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

\[
\{X_i^1 \preceq X_i^k\}, \quad i \geq 2
\]

are independent. Therefore, the random variables \( \tau_k^1 \) and \( \tau_k \) are conditionally independent given \( X_1 \), such that (notice that the index \( i \) in (36) is not less than 2)

\[
\tau_k^1|X_1 \sim \text{Geo}\left(\prod_{j=1}^{k} S(X_{1j})\right)
\]

and

\[
\tau_k \sim \text{Geo}\left((1 - \alpha)^k\right).
\]

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

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

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

Consider the case where \( 1 - \alpha < e^{-\gamma} \). Then, (40) implies that there exists a \( P \)-a.s. finite random variable \( K_\alpha \) which is uniquely determined by \( X_1 \) such that for every \( k > K_\alpha \)

\[
(1 - \alpha)e^{L_k} \leq \frac{1 + (1 - \alpha)e^\gamma}{2} \equiv \zeta_\alpha
\]
such that $\zeta_\alpha < 1$. 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

$$\sum_{k=K_\alpha}^{\infty} P(\tau_k \leq \tau_1^1 | X_1) = \sum_{k=K_\alpha}^{\infty} P(\tau_k - 1 \leq \tau_1^1 - 1 | X_1)$$

$$= \sum_{k=K_\alpha}^{\infty} \frac{(1-\alpha)^k}{(1-\alpha)^k + e^{-kL_k} - (1-\alpha)^k e^{-kL_k}}$$

$$\leq \sum_{k=K_\alpha}^{\infty} \left[(1-\alpha)e^{L_k}\right]^k$$

$$\leq \sum_{k=K_\alpha}^{\infty} \zeta_\alpha^k < \infty.$$

Thus, the Lemma of Borel-Cantelli implies that

$$P\left(\tau_k \leq \tau_1^1, i.o \mid X_1\right) = 0, \quad P\text{-a.s.}$$

and hence

$$P\left(\tau_k \leq \tau_1^1, i.o \right) = E\left[P\left(\tau_k \leq \tau_1^1, i.o \mid X_1\right)\right] = 0.$$ (43)

Therefore, $P(M_n > M_1^1 \text{ i.o}) = 0$, which yields the required result when $1 - \alpha < e^{-\gamma}$.

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

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

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

$$\sum_{k=K_\alpha}^{\infty} P(\tau_k \geq \tau_1^1 | X_1) = \sum_{k=K_\alpha}^{\infty} P(\tau_k - 1 \geq \tau_1^1 - 1 | X_1)$$

$$= \sum_{k=K_\alpha}^{\infty} \frac{e^{-kL_k}}{(1-\alpha)^k + e^{-kL_k} - (1-\alpha)^k e^{-kL_k}}$$

$$\leq \sum_{k=K_\alpha}^{\infty} \left[(1-\alpha)e^{L_k}\right]^{-k}$$

$$\leq \sum_{k=K_\alpha}^{\infty} \zeta_\alpha^{-k} < \infty.$$

Thus, the claim follows from the Lemma of Borel-Cantelli using a similar argument as in the previous case. ■
Proof of Theorem 1 (continuation)

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

\[
\frac{M_1^n}{\log(n)} \overset{n \to \infty}{\to} \gamma^{-1}, \quad P\text{-a.s.}
\]  

(47)

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

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

(48)

Consider two independent iid sequences \( G^{(\alpha_1)}_2, G^{(\alpha_1)}_3, \ldots \) and \( G^{(\alpha_2)}_2, G^{(\alpha_2)}_3, \ldots \) such that \( G^{(\alpha_l)}_1 \sim \text{Geo}(\alpha_l) \) for \( l = 1, 2 \). Respectively, define the corresponding sequences of partial maxima

\[
M^{(\alpha_l)}_n \equiv \max_{2 \leq i \leq n} G^{(\alpha_l)}_i, \quad n \geq 2,
\]  

(49)

for each \( l = 1, 2 \) as described in the statement of Lemma 1. Then, Lemma 1 implies that there exists \( P\text{-a.s.} \) finite random variables \( N_{\alpha_1} \) and \( N_{\alpha_2} \) such that

\[
M^{(\alpha_1)}_n \leq M^n_1 \leq M^{(\alpha_2)}_n, \quad \forall n \geq \max(N_{\alpha_1}, N_{\alpha_2}) \equiv N.
\]  

(50)

Furthermore, Theorem 2 of [1] yields that for each \( l = 1, 2 \)

\[
\frac{M^{(\alpha_l)}_n}{\log(n)} \overset{n \to \infty}{\to} -[\log(1 - \alpha_l)]^{-1}, \quad P\text{-a.s.}
\]  

(51)

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

\[
\gamma^{-1} - \frac{\epsilon}{2} \leq \frac{M^{(\alpha_1)}_n}{\log(n)} \leq \frac{M^n_1}{\log(n)} \leq \frac{M^{(\alpha_2)}_n}{\log(n)} \leq \gamma^{-1} + \frac{\epsilon}{2}
\]  

(52)

and hence (47) follows. Therefore, (7) implies that

\[
\liminf_{n \to \infty} \frac{k_n - 1}{M^n_1} = \liminf_{n \to \infty} \frac{k_n}{\log(n)} \cdot \frac{k_n - 1}{M^n_1} > 1, \quad P\text{-a.s.}
\]  

(53)

and hence (34) yields that \( 1_{p_{k_n,n}}(1) \overset{n \to \infty}{\to} 0, \quad P\text{-a.s.} \). Similarly, (9) implies that

\[
\limsup_{n \to \infty} \frac{k_n}{M^n_1} = \limsup_{n \to \infty} \frac{k_n}{\log(n)} \cdot \frac{\log(n)}{M^n_1} < 1, \quad P\text{-a.s.}
\]  

(54)

and hence (34) yields that \( 1_{p_{k_n,n}}(1) \overset{n \to \infty}{\to} 0, \quad P\text{-a.s.} \).}

Acknowledgement: The authors thank the reviewer for detecting a mistake in the original proof of Theorem 1.
References

[1] Thomas S. Ferguson. On the asymptotic distribution of max and m~ex. *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 répartition des valeurs des fonctions arithmétiques. 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.