ON MULTIVARIATE RECORDS FROM RANDOM VECTORS WITH INDEPENDENT COMPONENTS

M. FALK,* *University of Würzburg*

A. KHORRAMI CHOKAMI,** *Bocconi University of Milan*

S. A. PADOAN,*** *Bocconi University of Milan*

Abstract

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be independent copies of a random vector $\boldsymbol{X}$ with values in $\mathbb{R}^d$ and with a continuous distribution function. The random vector $\boldsymbol{X}_n$ is a complete record, if each of its components is a record. As we require $\boldsymbol{X}$ to have independent components, crucial results for univariate records clearly carry over. But there are substantial differences as well: While there are infinitely many records in case $d = 1$, there occur only finitely many in the series if $d \geq 2$. Consequently, there is a terminal complete record with probability one. We compute the distribution of the random total number of complete records and investigate the distribution of the terminal record. For complete records, the sequence of waiting times forms a Markov chain, but differently from the univariate case, now the state infinity is an absorbing element of the state space.

Keywords: Multivariate records, complete records, terminal record, waiting time, Markov chain

2010 Mathematics Subject Classification: Primary 60G70

Secondary 60E05, 62H05

1. Introduction

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ be independent copies of a random vector $\boldsymbol{X} \in \mathbb{R}^d$ with distribution function F. We assume that the margins $F_i, 1 \leq i \leq d$, of F are continuous univariate distribution functions. This is equivalent of assuming the condition that F itself is a continuous distribution function.

Records among a sequence of univariate independent and identically distributed (iid) random variables $X_1, X_2, \ldots$ have been extensively investigated over the past decades, see, e.g., Resnick (1987, Ch. 4.1 and

---
* Postal address: Institute of Mathematics, University of Würzburg, Am Hubland, D-97074, Würzburg, Germany

* Email address: michael.falk@uni-wuerzburg.de

** *** Postal address: Department of Decision Sciences, Bocconi University of Milan, via Roentgen 1, 20136 Milano, Italy

** Email address: amir.khorrami@phd.unibocconi.it

*** Email address: simone.padoan@unibocconi.it

4.2), Galambos (1987, Ch. 6.2 and 6.3), and Arnold, Balakrishnan, and Nagaraja (1998).

For example, consider the indicator function

$$e_m := \mathbb{1}(X_m \text{ is a record}), \qquad m \in \mathbb{N},$$

where $\mathbb{1}(E)$ denotes the indicator function of the event E and $X_1, X_2, \dots$ are iid univariate random variables with a joint continuous df F. It is well known that the indicator functions $e_1, e_2, \dots$ are independent with

$$\Pr(e_m = 1) = m^{-1}, \qquad m \in \mathbb{N}, \tag{1}$$

see, e.g., Galambos 1987, (Lemma 6.3.3).

In this paper we are interested in *complete* records. The d-dimensional random vector (rv) $\boldsymbol{X}_j$ is a complete record if each of its components is a record, i.e.,

$$\boldsymbol{X}_j > \max_{1 \leq i \leq j-1} \boldsymbol{X}_i,$$

where the maximum is taken componentwise. All our operations on vectors $\boldsymbol{x} = (x_1, \dots, x_d)$, $\boldsymbol{y} = (y_1 \dots, y_d)$, such as $\boldsymbol{x} < \boldsymbol{y}$, are meant componentwise. Clearly, $\boldsymbol{X}_1$ is a complete record.

Multivariate records have not been discussed that extensively, yet they have been approached by, Goldie and Resnick (1989), Goldie and Resnick (1995) and Arnold, Balakrishnan, and Nagaraja (1998, Chapter 8), for instance. For supplementary material on multivariate and functional records we refer to the thesis by Zott (2016) and the references cited therein.

In this paper we focus on the case, where the components of $\boldsymbol{X}$ are independent. Then, clearly, various of the results for univariate vectors carry over to the multivariate case. In particular the preceding result carries over: Put

$$I_m := \mathbb{1}(\boldsymbol{X}_m \text{ is a complete record}).$$

Then, the indicator functions $I_1, I_2, \dots$ are independent with

$$\Pr(I_m = 1) = m^{-d}, \qquad m \in \mathbb{N}. \tag{2}$$

However, from equations (1) and (2) we immediately deduce a first difference between the theory of univariate and multivariate records. If the joint distribution function of the sequence of iid univariate random variables is continuous, then the total number of records in this series is infinite with probability one. On the other hand, by equation (2) we have

$$\mathrm{E}\left(\sum_{m \in \mathbb{N}} I_m\right) = \sum_{m \in \mathbb{N}} \mathrm{E}(I_m) = \sum_{m \in \mathbb{N}} m^{-d} < \infty \tag{3}$$

if $d \geq 2$. As a consequence, the total number of complete records $\sum_{m \in \mathbb{N}} I_m$ is finite with probability one. Hence, in case $d \geq 2$, there is a terminal complete record in the series $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ In Section 2 we compute the distribution of the random total number of complete records and we investigate the distribution of the terminal record. In Section 3 we study the sequence of waiting times for the complete records. Such a sequence forms a Markov chain, similarly to the univariate case, but in higher dimensions the state infinity is an absorbing element of this state space.

Suppose that the components of $\boldsymbol{X}$ are not independent, but that its distribution function is in the max-domain of attraction of a multivariate extreme value distribution. Goldie and Resnick (1989, Theorem 5.3) proved in case $d = 2$ that the total number of complete records is finite if and only if the limiting extreme value distribution has independent components. Assuming that the components of $\boldsymbol{X}$ are independent, we only require continuity of its distribution function, we do not requite that it is in the max-domain of attraction of an extreme value distribution.

2. Terminal record

Let
$$T := \sup \{ m \in \mathbb{N} : I_m = 1 \},$$
which is the index of the ultimate complete record in the sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ If $d \geq 2$, then we know from equation (3) that $\Pr(T < \infty) = 1$. In the next Lemma we compute the distribution of T.

Lemma 1. *For $d \geq 2$ and $k \in \mathbb{N}$*

$$p_k = \Pr(T = k) = \frac{1}{k^d} \prod_{m \geq k+1} \left(1 - \frac{1}{m^d} \right). \tag{4}$$

In particular, when $d = 2$:

$$p_k = \frac{1}{k(k+1)}.$$

Proof. The independence of the indicator functions I_m, $m \in \mathbb{N}$, together with equation (2) imply the first assertion.

In the case $d = 2$ we obtain

$$p_k = \frac{1}{k^2} \lim_{N \to \infty} \prod_{m=k+1}^{N} \frac{(m-1)(m+1)}{m^2} = \frac{1}{k^2} \lim_{N \to \infty} \frac{k(N+1)}{N(k+1)} = \frac{1}{k(k+1)}.$$

$\square$

The first observation $\boldsymbol{X}_1$ is a record by definition. By the preceding result, in dimension $d = 2$, $\boldsymbol{X}_1$ is already the terminal complete record with probability $p_1 = 1/2$. The next observation $\boldsymbol{X}_2$ is with probability $1/4$ a complete record, and it is with probability $1/6$ the terminal complete record. From equation (3) and Lemma 1 we have that

$$1 = \Pr(T < \infty) = \sum_{k \in \mathbb{N}} p_k = \frac{1}{k^d} \prod_{m \geq k+1} \left(1 - \frac{1}{m^d} \right)$$

which, taken as a purely mathematical formula, is a nice by-product.

The probability $p_1 = p_1(d)$ increases as the dimension increases, whereas $p_k = p_k(d)$, $k \geq 2$, decreases. This is the content of the next Lemma.

Lemma 2. *We have*

$$\lim_{d \to \infty} p_1(d) = \lim_{d \to \infty} \prod_{m \geq 2} \left(1 - \frac{1}{m^d} \right) = 1,$$

whereas, for $k \geq 2$, we have

$$\lim_{d \to \infty} p_k(d) = \lim_{d \to \infty} \frac{1}{k^d} \prod_{m \geq k+1} \left(1 - \frac{1}{m^d} \right) = 0.$$

Proof. The second assertion is immediate from the bound $\prod_{m \geq k+1} \left(1 - \frac{1}{m^d}\right) \leq 1$. The first assertion is a consequence of the equation $\sum_{k \in \mathbb{N}} p_k = 1$ and the following bound, valid for $d \geq 4$,

$$\sum_{k \geq 2} p_k = \sum_{k \geq 2} \frac{1}{k^d} \prod_{m \geq k+1} \left(1 - \frac{1}{m^d}\right) \leq \frac{1}{2^{d-2}} \sum_{k \geq 2} \frac{1}{k^2} \xrightarrow[d \to \infty]{} 0.$$

$\square$

Lemma 1 implies that the expected arrival time for the final complete record is

$$\mathrm{E}(T) = \sum_{k \in \mathbb{N}} k p_k = \sum_{k \in \mathbb{N}} \frac{1}{k^{d-1}} \prod_{m \geq k+1} \left(1 - \frac{1}{m^d}\right).$$

Therefore, we have

$$\mathrm{E}(T) = \sum_{k \in \mathbb{N}} \frac{1}{k+1} = \infty$$

when $d = 2$, while

$$\mathrm{E}(T) \leq \sum_{k \in \mathbb{N}} \frac{1}{k^{d-1}} \leq \infty$$

for $d \geq 3$.

Finally, by repeating the arguments in the proof of Lemma 2, we have

$$\mathrm{E}(T) \xrightarrow[d \to \infty]{} 1.$$

Let $\boldsymbol{\eta}$ be a d-dimensional rv with independent components $\eta_1, \ldots, \eta_d$, each following a standard negative exponential distribution, i.e., $\Pr(\eta_i \leq x) = \exp(x)$, $x \leq 0$, for all $i \leq d$. In what follows we investigate the distribution of the terminal record, i.e., we study $\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x})$, where T denotes again the random index of the terminal record and $\boldsymbol{x} = (x_1, \ldots, x_d) \in (-\infty, 0]^d$. We have a closed formula for $\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x})$, see Theorem 1. However, we first want to verify the following conjecture: Let $T = T(d)$ be the random index of the terminal record which depends on the dimension d. From Lemma 2 we know that $\Pr(T(d) = 1) = p_1(d) \xrightarrow[d \to \infty]{} 1$. Therefore, one would expect that

$$\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}) \approx \Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}) = \exp\left(\sum_{i=1}^d x_i\right),$$

when d gets large. This conjecture is verified in the next result. To ease the notation we drop the dependence on d, wherever it causes no ambiguities.

Proposition 2.1. *Let $x_1, x_2, \ldots$ be a sequence of numbers in $(-\infty, 0]$.*

(i) If

$$\sum_{i=1}^\infty x_i \in (-\infty, 0],$$

then, with $\boldsymbol{x}_d := (x_1, \ldots, x_d)$, $d \in \mathbb{N}$, we have

$$\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d) \xrightarrow[d \to \infty]{} \exp\left(\sum_{i=1}^\infty x_i\right) = \lim_{d \to \infty} \Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d).$$

(ii) If

$$\lim_{d \to \infty} \sum_{i=1}^{d} x_i = -\infty,$$

but such that

$$\limsup_{d \to \infty} \left| \frac{1}{d} \sum_{i=1}^{d} x_i \right| < \log 2 \tag{5}$$

then

$$\frac{\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d)}{\Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d)} \xrightarrow[d \to \infty]{} 1.$$

Proof. We have

$$\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d) = \Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d, T = 1) + \Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d, T \geq 2)$$

with

$$\begin{aligned}
\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d, T \geq 2) &= \sum_{k=2}^{\infty} \Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d, T = k) \\
&\leq \sum_{k=2}^{\infty} \Pr(\boldsymbol{\eta}_k \leq \boldsymbol{x}_d, \boldsymbol{\eta}_k \text{ is a complete record}) \\
&= \sum_{k=2}^{\infty} \Pr(\boldsymbol{\eta}_k \leq \boldsymbol{x}_d \mid \boldsymbol{\eta}_k \text{ is a complete record}) \frac{1}{k^d} \\
&= \sum_{k=2}^{\infty} \exp\left(k \sum_{i=1}^{d} x_i \right) \frac{1}{k^d} \\
&= \exp\left(\sum_{i=1}^{d} x_i \right) \sum_{k=2}^{\infty} \exp\left((k-1) \sum_{i=1}^{d} x_i \right) \frac{1}{k^d} \\
&= \exp\left(\sum_{i=1}^{d} x_i \right) o(1).
\end{aligned} \tag{6}$$

In the preceding list we used the fact that the *univariate* distribution function $\Pr(\eta_k \leq x \mid \eta_k$ is a record) equals $\exp(kx)$, $x \leq 0$, $k \in \mathbb{N}$, as established in Falk, Khorrami, and Padoan (2017).

From equation (6) we obtain

$$\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d) = \Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d, T = 1) + \exp\left(\sum_{i=1}^{d} x_i \right) o(1)$$

$$= \Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d) - \Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d, T \geq 2) + \exp\left(\sum_{i=1}^{d} x_i \right) o(1)$$

As $\Pr(T \geq 2) \xrightarrow[d \to \infty]{} 0$, this implies the first assertion.

Next, suppose next that

$$\sum_{i=1}^{\infty} x_i = -\infty.$$

We have to show that

$$\Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d, T \geq 2) = \exp\left(\sum_{i=1}^{d} x_i \right) o(1) \tag{7}$$

as well. Hoelder's inequality implies with $p, q \geq 1$, $p^{-1} + q^{-1} = 1$,

$$
\begin{aligned}
\Pr(\boldsymbol{\eta}_T \leq \boldsymbol{x}_d, T \geq 2) &= \mathrm{E}\left(\mathbb{1}(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d)\mathbb{1}(T \geq 2)\right) \\
&\leq \Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d)^{1/p}\Pr(T \geq 2)^{1/q} \\
&= \exp\left(\frac{1}{p}\sum_{i=1}^{d} x_i\right)\Pr(T \geq 2)^{1/q}
\end{aligned}
$$

where

$$
\begin{aligned}
\Pr(T \geq 2) &= \Pr(T = 2) + \sum_{k=3}^{\infty}\Pr(T = k) \\
&\leq 2^{-d} + \sum_{k=3}^{\infty} k^{-d} \\
&\leq 2^{-d} + \int_{2}^{\infty} x^{-d}\mathrm{d}x \\
&= 2^{-d} + \frac{2^{-d+1}}{d-1} \leq \frac{3}{2^d}.
\end{aligned}
$$

Thus, we obtain

$$
\begin{aligned}
\Pr(\boldsymbol{\eta}_1 \leq \boldsymbol{x}_d, T \geq 2) &\leq \exp\left(\frac{1}{p}\sum_{i=1}^{d} x_i\right)\frac{3^{1/q}}{2^{d/p}} \\
&= \exp\left(\sum_{i=1}^{d} x_i\right)\exp\left(\left(\frac{1}{p}-1\right)\sum_{i=1}^{d} x_i - \frac{d}{q}\log 2\right)3^{1/q} \\
&= \exp\left(\sum_{i=1}^{d} x_i\right)\exp\left(-\frac{1}{q}\sum_{i=1}^{d} x_i - \frac{d}{q}\log 2\right)3^{1/q} \\
&= \exp\left(\sum_{i=1}^{d} x_i\right)\exp\left(\frac{d}{q}\left(-\frac{1}{d}\sum_{i=1}^{d} x_i - \log 2\right)\right)3^{1/q} \\
&= \exp\left(\sum_{i=1}^{d} x_i\right)o(1)
\end{aligned}
$$

by condition (3). This proves the second assertion as well. $\qquad\square$

By assuming the componentwise representation $X_i = F_i^{-1}(\exp(\eta_i))$, $i = 1, \ldots, d$, $d \in \mathbb{N}$, for each component $i = 1, \ldots, d$, the preceding result immediately carries over to a sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ of independent copies of a rv $\boldsymbol{X}$ with independent components and univariate continuous marginal df $F_1, \ldots, F_d$.

Corollary 2.1. *Let $y_1, y_2, \ldots$ be a sequence of number in $\mathbb{R}$.*

(i) If

$$
\prod_{i=1}^{\infty} F_i(y_i) \in (0, 1),
$$

then

$$
\Pr(\boldsymbol{X}_T \leq \boldsymbol{y}_d) \xrightarrow[d\to\infty]{} \prod_{i=1}^{\infty} F_i(y_i) = \lim_{d\to\infty}\Pr(\boldsymbol{X}_1 \leq \boldsymbol{y}_d).
$$

6

(ii) If

$$\prod_{i=1}^{\infty} F_i(y_i) = 0,$$

but such that

$$\liminf_{d\to\infty} \left(\prod_{i=1}^{d} F_i(y_i)\right)^{1/d} > \frac{1}{2},$$

then

$$\frac{\Pr(\boldsymbol{X}_T \le \boldsymbol{y}_d)}{\Pr(\boldsymbol{X}_1 \le \boldsymbol{y}_d)} \xrightarrow[d\to\infty]{} 1.$$

In the final result of this section we derive the exact distribution of the terminal complete record for fixed dimension $d \ge 2$. We suppose again a sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ of independent copies of a rv $\boldsymbol{X} \in \mathbb{R}^d$ with independent components and continuous univariate marginal df $F_1, \dots, F_d$.

Theorem 1. *The distribution function of the final complete record is*

$$\Pr\left(\boldsymbol{X}_T \le \boldsymbol{x}_d\right) = \sum_{k=1}^{\infty} \left(\frac{\prod_{i=1}^{d} u_i^k}{k^d} - \sum_{K \subseteq \mathcal{J}} (-1)^{|K|-1} \prod_{i=1}^{d} \left(\sum_{r \in K'} \frac{u_i^r}{r \prod_{s \ne r \in K'}(s-r)} + \frac{1}{\prod_{r \in K'} r} \right) \right),$$

where $u_i = F_i(x_i)$, $\mathcal{J} = \{k+1, k+2, \dots\}$, $K \subseteq \mathcal{J}$, $K' = \{k\} \cup K$ and $|K|$ is the total number of elements in the set K.

Proof of Theorem 1. Without loss of generality we provide the proof with uniform margins $U_1, U_2 \dots$ We look for the solution of

$$\Pr\left(\boldsymbol{U}_T \le \boldsymbol{u}_d\right) = \sum_{k=1}^{\infty} \Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}_d \mid I_k = 1, \bigcap_{m=k+1}^{\infty} \{I_m = 0\} \right) \Pr(T = k),$$

and $\Pr(T = k) = p_k$. The probability of the conditioning event is given by (4), therefore we only need to compute

$$\Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, \bigcap_{m=k+1}^{\infty} \{I_m = 0\} \right)$$

$$= \Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1 \right) - \Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}, I_k = 1, \left(\bigcap_{m=k+1}^{\infty} \{I_m = 0\} \right)^{\mathrm{c}} \right)$$

Since the components of $\boldsymbol{U}$ are independent, it is easy to see that

$$\Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}, I_k = 1 \right) = \frac{\prod_{i=1}^{d} u_i^k}{k^d}.$$

By means of the inclusion-exclusion principle, we have that

$$\Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, \left(\bigcap_{m=k+1}^{\infty} \{I_m = 0\} \right)^{\mathrm{c}} \right)$$

$$= \sum_{K \subseteq \mathcal{J}} (-1)^{|K|-1} \Pr\left(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, I_t = 1, t \in K \right)$$

$$= \sum_{K \subseteq \mathcal{J}} (-1)^{|K|-1} \prod_{i=1}^{d} \Pr\left(U_{k,i} \le u_i, I_k = 1, I_t = 1, t \in K \right)$$

where $K = \{j_1, \ldots, j_{|K|}\} \subseteq \mathcal{J} = \{k+1, k+2, \ldots\}$. Note that

$$\Pr\left(U_{k,i} \leq u_i, e_k = 1, e_t = 1, t \in K\right) = \underbrace{\int \cdots \int}_{0 \leq \min\,(u_i, z_1) < \cdots < z_{|K|} \leq 1} z^{k-1} \prod_{t=1}^{|K|} z_t^{j_t - j_{t-1} - 1} \, \mathrm{d}z \, \mathrm{d}z_1 \ldots \mathrm{d}z_{|K|}.$$

We compute the previous probability by using the induction principle. We claim that

$$A_m = \underbrace{\int \cdots \int}_{0 \leq \min\,(u_i, z_1) < \cdots < z_m} z^{k-1} \prod_{t=1}^{m-1} z_t^{j_t - j_{t-1} - 1} \, \mathrm{d}z \, \mathrm{d}z_1 \ldots \mathrm{d}z_{|K|-1}$$

$$= \sum_{t=0}^{m-1} (-1)^t \frac{u_i^{j_t}}{\prod_{r=0}^{t-1}(j_t - j_r) j_t \prod_{r=t+1}^{m-1}(j_r - j_t)} z_m^{j_{m-1} - j_t} \mathbb{1}(u_i < z_m) + \frac{z_m^{j_{m-1}}}{\prod_{t=0}^{m-1}} \mathbb{1}(u_i > z_m),$$

where $j_0 = k$. When $|K| = 1$, we have

$$\Pr\left(U_{k,i} \leq u_i, e_k = 1, e_{j_1} = 1\right) = \int_0^{\min(u_i, z_1)} z^{k-1} \mathrm{d}z = \frac{u_i^k}{k} \mathbb{1}\left(u_i \leq z_1\right) + \frac{z_1^k}{k} \mathbb{1}\left(u_i > z_1\right).$$

Now, let us suppose the claim it is true for $|K| = m - 1$, i.e.

$$A_{m-1} = \sum_{t=0}^{m-2} (-1)^t \frac{u_i^{j_t}}{\prod_{r=0}^{t-1}(j_t - j_r) j_t \prod_{r=t+1}^{m-2}(j_r - j_t)} z_{m-1}^{j_{m-2} - j_t} \mathbb{1}(u_i < z_{m-1}) + \frac{z_{m-1}^{j_{m-2}}}{\prod_{t=0}^{m-2}} \mathbb{1}(u_i > z_{m-1}),$$

and prove it for $|K| = m$.

$$A_m = \int_0^{z_m} z_{m-1}^{j_{m-1} - j_{m-2} - 1} A_{m-1} \mathrm{d}z_{m-1}$$

$$= \sum_{t=0}^{m-2} (-1)^t \frac{u_i^{j_t}}{\prod_{r=0}^{t-1}(j_t - j_r) j_t \prod_{r=t+1}^{m-2}(j_r - j_t)} \frac{z_m^{j_{m-1} - j_t} - u_i^{j_{m-1} - j_t}}{j_{m-1} - j_t} \mathbb{1}(u_i < z_m)$$

$$+ \frac{1}{\prod_{t=0}^{m-2}} \left(\int_0^{u_i} z_{m-1} \mathrm{d}z_{m-1} \mathbb{1}(u_i < z_m) + \int_0^{z_m} z_{m-1} \mathrm{d}z_{m-1} \mathbb{1}(u_i > z_m) \right)$$

which proves the claim. By considering $z_m = 1$ and by noting that $\prod_{r=0}^{t-1}(j_t - j_r) = (-1)^t \prod_{r=0}^{t-1}(j_r - j_t)$ and substituting with $u_i = F_i(x_i)$, the proof is complete. $\qquad \square$

3. Complete Record Times

In this section we derive some results on record times. Let

$$R(n) := \inf\left\{m \in \mathbb{N} : \sum_{i=1}^m I_i = n\right\}, \quad n \geq 2, \quad R(1) := 1, \tag{8}$$

be the arrival time of the n-th complete record, where $\inf \emptyset := \infty$, which describes the case when there is no further complete record. We have seen in equation (3) that the number of complete records is finite with probability one, if the dimension d of our observations is at least 2. We start with the exact distribution of $R(n)$. Note that the distribution of $R(n)$ does not depend on the underlying univariate df $F_1, \ldots, F_d$, provided that they are continuous.

Proposition 3.1. *For a generic size $d \geq 2$ we have*

$$\Pr(R(n) = k) = k^{-d} \sum_{A \subseteq \{2,\ldots,k-1\},|A|=n-2} \prod_{q \in A} q^{-d} \prod_{m \in A^c} \left(1 - m^{-d}\right), \qquad k \geq n, \tag{9}$$

where $|A|$ is the total number of elements in the set A.

Proof. Note that

$$\Pr(R(n) = k) = \Pr\left(I_n = 1, S_{k-1} = n-1\right) = \Pr\left(I_n = 1\right)\Pr\left(S_{k-1} = n-1\right),$$

where $S_{k-1} = \sum_{m=1}^{k-1} I_m$ is a sum of independent Bernoulli random variables, each with parameter m^{-d}. Therefore,

$$\Pr\left(S_{k-1} = n-1\right) = \Pr\left(\sum_{m=2}^{k-1} I_m = n-2\right) = \sum_{A \in \mathcal{A}_{n-2}} \prod_{q \in A} q^{-d} \prod_{m \in A^c} \left(1 - m^{-d}\right),$$

which is a Poisson-Binomial distribution. $\qquad\square$

Example 3.1. For $n = 2$ we get

$$\Pr(R(2) = k) = \frac{1}{k^d} \prod_{j=2}^{k-1} \left(1 - \frac{1}{j^d}\right), \qquad k \geq 2$$

while if $n = 3$

$$\Pr(R(3) = k) = \frac{1}{k^d} \prod_{j=2}^{k-1} \left(1 - \frac{1}{j^d}\right) \sum_{i=2}^{k-1} \frac{1}{i^d - 1}, \qquad k \geq 3.$$

Thus in the special case $d = 2$ we obtain

$$\Pr(R(2) = k) = \frac{1}{2k(k-1)}, \qquad \Pr(R(3) = k) = \frac{3k^2 - 7k + 2}{8k^2(k-1)^2}. \tag{10}$$

The sequence $R(n)$, $n \geq 2$, is a Markov chain, as it is in the univariate case, see, e.g., Galambos 1987, (Section 6.3). Note that the state space is now $\{2, 3, \ldots\} \cup \{\infty\}$.

Proposition 3.2. *The sequence $R(n)$, $n \geq 2$ forms a Markov chain with the following transition probabilities*

$$\Pr\left(R(n) = k | R(n-1) = j\right) = \begin{cases} k^{-d}, & \text{for } k = j+1, \\ k^{-d} \prod_{m=j+1}^{k-1} \left(1 - m^{-d}\right), & \text{for } k > j+1, \\ \prod_{m=j+1}^{\infty} \left(1 - m^{-d}\right), & \text{for } k = \infty > j, \end{cases} \tag{11}$$

with $j \geq n-1$. The state $\{\infty\}$ is absorbing, that is $\Pr(R(n) = \infty \mid R(n-1) = \infty) = 1$, when $n \geq 3$.

Proof. For a finite sequence of finite states, by the independence of $I_1, I_2, \ldots$ we have

$$\Pr\left(R(m) = j_m, 2 \leq m \leq n\right) = \prod_{m=2}^{n} \Pr\left(I_{j_m} = 1\right)\Pr\left(\sum_{i=j_{m-1}+1}^{j_m - 1} I_i = 0\right), \tag{12}$$

9

where $j_1 = 2$ by convention. Using this formula we obtain for the conditional probability

$$\Pr\left(R(n) = j_n | R(m) = j_m, 2 \leq m \leq n-1\right) = \Pr\left(I_{j_n} = 1\right) \Pr\left(\sum_{i=j_{n-1}+1}^{j_n-1} I_i = 0\right). \tag{13}$$

Note that

$$\Pr\left(R(n-1) = j_{n-1}\right) = \Pr\left(I_{j_{n-1}} = 1\right) \Pr\left(\sum_{i=2}^{j_{n-1}-1} I_i = n-2\right)$$

and

$$\Pr\left(R(n) = j_n, R(n-1) = j_{n-1}\right) = \Pr\left(I_{j_n} = 1\right) \Pr\left(\sum_{i=j_{n-1}+1}^{j_n-1} I_i = 0\right)$$

$$\Pr\left(I_{j_{n-1}} = 1\right) \Pr\left(\sum_{i=2}^{j_{n-1}-1} I_i = n-2\right).$$

Therefore $\Pr\left(R(n) = j_n | R(n-1) = j_{n-1}\right)$ is equal to right hand-side of (13).

For the case that a time moves from a finite state to infinity we have

$$\Pr\left(R(n) = \infty, R(m) = j_m, 2 \leq m \leq n-1\right) = \prod_{m=2}^{n-1} \Pr\left(I_{j_m} = 1\right)$$

$$\times \Pr\left(\sum_{i=j_{m-1}+1}^{j_m-1} I_i = 0\right)$$

$$\times \Pr\left(\sum_{i=j_m+1}^{\infty} I_i = 0\right).$$

Using this result and the one in (12) we obtain

$$\Pr\left(R(n) = \infty | R(m) = j_m, 2 \leq m \leq n-1\right) = \Pr\left(\sum_{i=j_{n-1}+1}^{\infty} I_i = 0\right). \tag{14}$$

Now, noting that $\Pr\left(R(n-1) = j_{n-1}\right) = \Pr\left(I_{j_{n-1}} = 1\right)$ and

$$\Pr\left(R(n) = \infty, R(n-1) = j_{n-1}\right) = \Pr\left(I_{j_{n-1}} = 1\right) \Pr\left(\sum_{i=j_{n-1}+1}^{\infty} I_i = 0\right),$$

then $\Pr\left(R(n) = \infty | R(n-1) = j_{n-1}\right)$ is equal to right hand-side of (14).

To compute the transition probabilities, note that $\Pr(I_n = 1) = n^{-d}$ and $\Pr(I_n = 0) = 1 - n^{-d}$. Finally, to complete the proof we need to check that

$$p_{n|n-1} = \sum_{k \geq j+1} \Pr\left(R(n) = k | R(n-1) = j\right) + \Pr\left(R(n) = \infty | R(n-1) = j\right) = 1,$$

for each $j \geq 2$, i.e.

$$(j+1)^{-d} + \sum_{k=j+2}^{\infty} k^{-d} \prod_{m=j+1}^{k-1} (1 - m^{-d}) + \prod_{m=j+1}^{\infty} (1 - m^{-d}) = 1.$$

We prove by induction that

$$\sum_{k=j+2}^{M} k^{-d} \prod_{m=j+1}^{k-1} (1 - m^{-d}) + \prod_{m=j+1}^{M} (1 - m^{-d}) = 1 - (j+1)^{-d}, \tag{15}$$

for each $M \in \mathbb{N}$.

Step $M = j + 2$:

$$(j+2)^{-d}(1 - (j+1)^{-d}) + (1 - (j+1)^{-d})(1 - (j+2)^{-d}) = 1 - (j+1)^{-d}.$$

Now, we suppose (15) is true for M and we prove it is true also for $M + 1$.

$$\sum_{k=j+2}^{M+1} k^{-d} \prod_{m=j+1}^{k-1} (1 - m^{-d}) + \prod_{m=j+1}^{M+1} (1 - m^{-d})$$

$$= \sum_{k=j+2}^{M} k^{-d} \prod_{m=j+1}^{k-1} (1 - m^{-d}) + (M+1)^{-d} \prod_{m=j+1}^{M} (1 - m^{-d}) + (M+1)^{-d} \prod_{m=j+1}^{M} (1 - m^{-d})$$

$$= (1 - (j+1)^{-d})((M+1)^{-d} + 1 - (M+1)^{-d}) = 1 - (j+1)^{-d}.$$

Therefore $p_{n|n-1} = 1$ and the proof is completed.

□

Acknowledgements

References

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1998). *Records*. Wiley Series in Probability and Statistics. New York: Wiley.

Falk, M., A. Khorrami, and S. A. Padoan (2017). Some results on joint record events. *arXiv:1707.06254*.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics* (2 ed.). Malabar: Krieger.

Goldie, C. M. and S. I. Resnick (1989). Records in a partially ordered set. *Ann. Probab. 17*(2), 678–699.

Goldie, C. M. and S. I. Resnick (1995). Many multivariate records. *Stochastic Process. Appl. 59*(2), 185–216.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*, Volume 4 of *Applied Probability*. New York: Springer. First Printing.

Zott, M. (2016). *Extreme Value Theory in Higher Dimensions. Max-Stable Processes and Multivariate Records*. Ph. D. thesis, University of Würzburg.