J. Appl. Prob. **55,** 43–53 (2018)
*doi:*10.1017/jpr.2018.4
© *Applied Probability Trust* 2018

ON MULTIVARIATE RECORDS FROM RANDOM VECTORS WITH INDEPENDENT COMPONENTS

M. FALK,* *University of Würzburg*
A. KHORRAMI CHOKAMI ** *** AND
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 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 the $d = 1$ case, they occur only finitely many times in the series if $d \geq 2$. Consequently, there is a terminal complete record with probability 1. 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 unlike the univariate case now the state at ∞ is an absorbing element of the state space.

Keywords: Multivariate record; complete record; 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 to assuming the condition that F itself is a continuous distribution function.

Records among a sequence of univariate independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ have been extensively investigated over the past decades, see, e.g. Galambos (1987, Chapters 6.2 and 6.3), Resnick (1987, Chapters 4.1 and 4.2), and Arnold *et al.* (1998).

For example, consider the indicator function

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

where $\mathbf{1}(E)$ denotes the indicator function of the event E and $X_1, X_2, \ldots$ are i.i.d. univariate random variables with a joint continuous distribution function (DF) F. It is well known that

Received 7 August 2017; revision received 26 October 2017.
* 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

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

the indicator functions $e_1, e_2, \ldots$ are independent with

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

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. $\boldsymbol{X}_j > \max_{1\le i\le j-1} \boldsymbol{X}_i$, where the maximum is taken componentwise. All our operations on vectors $\boldsymbol{x} = (x_1, \ldots, x_d)$, $\boldsymbol{y} = (y_1 \ldots, y_d)$, such as $\boldsymbol{x} < \boldsymbol{y}$, are meant componentwise. Clearly, $\boldsymbol{X}_1$ is a complete record.

Multivariate records have not been extensively discussed in the literature, yet they have been approached by, e.g. Goldie and Resnick (1989), (1995), and Arnold *et al.* (1998, Chapter 8). For supplementary material on multivariate and functional records, we refer the reader to Zott (2016) and the references therein.

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

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

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

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

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

$$\mathbb{E}\left(\sum_{m\in\mathbb{N}} I_m\right) = \sum_{m\in\mathbb{N}} \mathbb{E}(I_m) = \sum_{m\in\mathbb{N}} m^{-d} < \infty \quad \text{if } d \ge 2. \tag{3}$$

As a consequence, the total number of complete records $\sum_{m\in\mathbb{N}} I_m$ is finite with probability 1. Hence, in the $d \ge 2$ case, there is a terminal complete record in the series $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$. 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 infinite state 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 the $d = 2$ case 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

$$T := \sup\{m \in \mathbb{N} \colon I_m = 1\},$$

which is the index of the ultimate complete record in the sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$. If $d \ge 2$ then we know from (3) that $\mathbb{P}(T < \infty) = 1$. In the next lemma we compute the distribution of T.

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

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

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

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 (2) imply the first assertion.

In the $d = 2$ case, we obtain

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

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 $p_1 = \frac{1}{2}$. The next observation $\boldsymbol{X}_2$ is with probability $\frac{1}{4}$ a complete record, and it is with probability $\frac{1}{6}$ the terminal complete record. From (3) and Lemma 1, we have

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

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 $\prod_{m \geq k+1}(1 - 1/m^d) \leq 1$. The first assertion is a consequence of $\sum_{k \in \mathbb{N}} p_k = 1$ and the following bound, valid for $d \geq 4$:

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

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

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

Therefore, we have

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

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

when $d = 2$, while

$$\mathbb{E}(T) \le \sum_{k\in\mathbb{N}} \frac{1}{k^{d-1}} < \infty \quad \text{for } d \ge 3.$$

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

$$\mathbb{E}(T) \to 1, \qquad d \to \infty.$$

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. $\mathbb{P}(\eta_i \le x) = \exp(x)$, $x \le 0$, for all $i \le d$. In what follows we investigate the distribution of the terminal record, i.e. we study $\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x})$, 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 $\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x})$; 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 $\mathbb{P}(T(d) = 1) = p_1(d) \to 1$, $d \to \infty$. Therefore, one would expect that

$$\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}) \approx \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}) = \exp\left(\sum_{i=1}^{d} x_i\right)$$

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 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*

$$\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d) \to \exp\left(\sum_{i=1}^{\infty} x_i\right) = \lim_{d\to\infty} \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d), \qquad d \to \infty.$$

(ii) *If $\lim_{d\to\infty} \sum_{i=1}^{d} x_i = -\infty$ but such that*

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

then

$$\frac{\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d)}{\mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d)} \to 1, \qquad d \to \infty.$$

Proof. We have

$$\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d) = \mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d,\ T = 1) + \mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d,\ T \ge 2)$$

with

$$\begin{aligned}
\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d,\ T \ge 2) &= \sum_{k=2}^{\infty} \mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d,\ T = k) \\
&\le \sum_{k=2}^{\infty} \mathbb{P}(\boldsymbol{\eta}_k \le \boldsymbol{x}_d,\ \boldsymbol{\eta}_k \text{ is a complete record}) \\
&= \sum_{k=2}^{\infty} \mathbb{P}(\boldsymbol{\eta}_k \le \boldsymbol{x}_d \mid \boldsymbol{\eta}_k \text{ is a complete record}) \frac{1}{k^d}
\end{aligned}$$

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

$$= \sum_{k=2}^{\infty} \exp\Big(k \sum_{i=1}^{d} x_i\Big) \frac{1}{k^d}$$
$$= \exp\Big(\sum_{i=1}^{d} x_i\Big) \sum_{k=2}^{\infty} \exp\Big((k-1) \sum_{i=1}^{d} x_i\Big) \frac{1}{k^d}$$
$$= \exp\Big(\sum_{i=1}^{d} x_i\Big) o(1). \tag{5}$$

In Proposition 1 we used the fact that the *univariate* distribution function $\mathbb{P}(\eta_k \le x \mid \eta_k \text{ is a record})$ is equal to $\exp(kx),\ x \le 0,\ k \in \mathbb{N}$, as established by Falk *et al.* (2017).

From (5), we obtain

$$\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d) = \mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d,\ T = 1) + \exp\Big(\sum_{i=1}^{d} x_i\Big) o(1)$$
$$= \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d) - \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d,\ T \ge 2) + \exp\Big(\sum_{i=1}^{d} x_i\Big) o(1)$$
$$= \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d) - \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d \mid T \ge 2)\mathbb{P}(T \ge 2) + \exp\Big(\sum_{i=1}^{d} x_i\Big) o(1).$$

As $\mathbb{P}(T \ge 2) \to 0, d \to \infty$, this implies the first assertion.

Next, suppose that $\sum_{i=1}^{\infty} x_i = -\infty$. We also have to show that

$$\mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d,\ T \ge 2) = \exp\Big(\sum_{i=1}^{d} x_i\Big) o(1).$$

Hoelder's inequality implies that with $p, q \ge 1,\ p^{-1} + q^{-1} = 1$,

$$\mathbb{P}(\boldsymbol{\eta}_T \le \boldsymbol{x}_d,\ T \ge 2) = \mathbb{E}(\mathbf{1}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d)\, \mathbf{1}(T \ge 2))$$
$$\le \mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d)^{1/p} \mathbb{P}(T \ge 2)^{1/q}$$
$$= \exp\Big(\frac{1}{p} \sum_{i=1}^{d} x_i\Big) \mathbb{P}(T \ge 2)^{1/q},$$

where

$$\mathbb{P}(T \ge 2) = \mathbb{P}(T = 2) + \sum_{k=3}^{\infty} \mathbb{P}(T = k)$$
$$\le 2^{-d} + \sum_{k=3}^{\infty} k^{-d}$$
$$\le 2^{-d} + \int_2^{\infty} x^{-d}\, \mathrm{d}x$$
$$= 2^{-d} + \frac{2^{-d+1}}{d-1}$$
$$\le \frac{3}{2^d}.$$

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

Thus, we obtain

$$\begin{aligned}\mathbb{P}(\boldsymbol{\eta}_1 \le \boldsymbol{x}_d,\ T \ge 2) &\le \exp\Big(\frac{1}{p}\sum_{i=1}^{d} x_i\Big)\frac{3^{1/q}}{2^{d/p}}\\ &= \exp\Big(\sum_{i=1}^{d} x_i\Big)\exp\Big(\Big(\frac{1}{p}-1\Big)\sum_{i=1}^{d} x_i - \frac{d}{q}\log 2\Big)3^{1/q}\\ &= \exp\Big(\sum_{i=1}^{d} x_i\Big)\exp\Big(-\frac{1}{q}\sum_{i=1}^{d} x_i - \frac{d}{q}\log 2\Big)3^{1/q}\\ &= \exp\Big(\sum_{i=1}^{d} x_i\Big)\exp\Big(\frac{d}{q}\Big(-\frac{1}{d}\sum_{i=1}^{d} x_i - \log 2\Big)\Big)3^{1/q}\\ &= \exp\Big(\sum_{i=1}^{d} x_i\Big)o(1)\end{aligned}$$

by condition (3). This proves the second assertion as well. □

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

Corollary 1. *Let $y_1, y_2, \dots$ be a sequence of numbers in $\mathbb{R}$.*

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

$$\mathbb{P}(\boldsymbol{X}_T \le \boldsymbol{y}_d) \to \prod_{i=1}^{\infty} F_i(y_i) = \lim_{d\to\infty} \mathbb{P}(\boldsymbol{X}_1 \le \boldsymbol{y}_d), \qquad d \to \infty.$$

(ii) *If $\prod_{i=1}^{\infty} F_i(y_i) = 0$ but such that*

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

then

$$\frac{\mathbb{P}(\boldsymbol{X}_T \le \boldsymbol{y}_d)}{\mathbb{P}(\boldsymbol{X}_1 \le \boldsymbol{y}_d)} \to 1, \qquad d \to \infty.$$

In the final result of this section, we derive the exact distribution of the terminal complete record for the fixed dimension $d \ge 2$. We suppose again a sequence $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ of independent copies of an 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*

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

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.

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

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

$$\mathbb{P}(\boldsymbol{U}_T \le \boldsymbol{u}_d) = \sum_{k=1}^{\infty} \mathbb{P}\Bigg(\boldsymbol{U}_k \le \boldsymbol{u}_d \;\Bigg|\; I_k = 1, \bigcap_{m=k+1}^{\infty} \{I_m = 0\} \Bigg) \mathbb{P}(T = k)$$

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

$$\begin{aligned} &\mathbb{P}\Bigg(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, \bigcap_{m=k+1}^{\infty} \{I_m = 0\} \Bigg) \\ &\quad = \mathbb{P}(\boldsymbol{U}_k \le \boldsymbol{u}_d,\ I_k = 1) - \mathbb{P}\Bigg(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, \Bigg(\bigcap_{m=k+1}^{\infty} \{I_m = 0\} \Bigg)^c \Bigg). \end{aligned}$$

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

$$\mathbb{P}(\boldsymbol{U}_k \le \boldsymbol{u}_d,\ I_k = 1) = \frac{\prod_{i=1}^{d} u_i^k}{k^d}.$$

By means of the inclusion–exclusion principle, we have

$$\begin{aligned} &\mathbb{P}\Bigg(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, \Bigg(\bigcap_{m=k+1}^{\infty} \{I_m = 0\} \Bigg)^c \Bigg) \\ &\quad = \sum_{K \subseteq \mathcal{J}} (-1)^{|K|-1} \mathbb{P}(\boldsymbol{U}_k \le \boldsymbol{u}_d, I_k = 1, I_t = 1, t \in K) \\ &\quad = \sum_{K \subseteq \mathcal{J}} (-1)^{|K|-1} \prod_{i=1}^{d} \mathbb{P}(U_{k,i} \le u_i, I_k = 1, I_t = 1, t \in K), \end{aligned}$$

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

$$\begin{aligned} &\mathbb{P}(U_{k,i} \le u_i, e_k = 1, e_t = 1, t \in K) \\ &\quad = \int \cdots \int_{0 < z \le \min(u_i, z_1) < z_2 \le \cdots \le z_{|K|} \le 1} z^{k-1} \prod_{t=1}^{|K|} z_t^{j_t - j_{t-1} - 1} \, \mathrm{d}z \, \mathrm{d}z_1 \cdots \mathrm{d}z_{|K|}. \end{aligned}$$

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

$$\begin{aligned} A_m &= \int \cdots \int_{0 < z \le \min(u_i, z_1) < z_2 \le \cdots \le z_m} z^{k-1} \prod_{t=1}^{m-1} z_t^{j_t - j_{t-1} - 1} \, \mathrm{d}z \, \mathrm{d}z_1 \cdots \mathrm{d}z_{|K|-1} \\ &= \sum_{t=0}^{m-1} (-1)^t \frac{u_i^{j_t}}{\prod_{r=0}^{t-1} (j_t - j_r) j_t \prod_{r=t+1}^{m-1} (j_r - j_t)} z_m^{j_{m-1} - j_t} \mathbf{1}(u_i < z_m) \\ &\quad + \frac{z_m^{j_{m-1}}}{\prod_{t=0}^{m-1} j_t} \mathbf{1}(u_i > z_m), \end{aligned}$$

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

where $j_0 = k$ and the products in the denominator in the second row are equal to 1 by convention, whenever $t - 1 < 0$ or $m - 1 < t + 1$. At the first step we have

$$A_1 = \int_0^{\min(u_i, z_1)} z^{k-1}\,\mathrm{d}z = \frac{u_i^k}{k}\,\mathbf{1}(u_i \le z_1) + \frac{z_1^k}{k}\,\mathbf{1}(u_i > z_1).$$

Now, let us suppose the claim holds for $m - 1$, i.e.

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

and prove it for m. Thus,

$$\begin{aligned} A_m &= \int_0^{z_m} z_{m-1}^{j_{m-1} - j_{m-2} - 1} A_{m-1}\,\mathrm{d}z_{m-1} \\ &= \sum_{t=0}^{m-2} (-1)^t \frac{u_i^{j_t}}{\prod_{r=0}^{t-1}(j_t - j_r) j_t \prod_{r=t+1}^{m-2}(j_r - j_t)} \frac{z_m^{j_{m-1} - j_t} - u_i^{j_{m-1} - j_t}}{j_{m-1} - j_t}\,\mathbf{1}(u_i < z_m) \\ &\quad + \frac{1}{\prod_{t=0}^{m-2} j_t} \left(\int_0^{u_i} z_{m-1}^{j_{m-1} - 1}\,\mathrm{d}z_{m-1}\,\mathbf{1}(u_i < z_m) \right. \\ &\qquad \left. + \int_0^{z_m} z_{m-1}^{j_{m-1} - 1}\,\mathrm{d}z_{m-1}\,\mathbf{1}(u_i > z_m) \right), \end{aligned}$$

which proves the claim. By considering $z_m = 1$ 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. □

3. Complete record times

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

$$R(n) := \inf\left\{ m \in \mathbb{N} \colon \sum_{i=1}^{m} I_i = n \right\}, \qquad n \ge 2, \qquad R(1) := 1,$$

be the arrival time of the nth complete record, where $\inf \varnothing := \infty$, which describes the case when there is no further complete record. We have seen in (3) that the number of complete records is finite with probability 1, if the dimension d of our observations is at least two. 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 2. *For a generic size $d \ge 2$, we have*

$$\mathbb{P}(R(n) = k) = k^{-d} \sum_{A \subseteq \{2, \ldots, k-1\},\, |A| = n-2} \prod_{q \in A} q^{-d} \prod_{m \in A^c} (1 - m^{-d}), \qquad k \ge n,$$

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

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

Proof. Note that

$$\mathbb{P}(R(n)=k)=\mathbb{P}(I_k=1,\ S_{k-1}=n-1)=\mathbb{P}(I_k=1)\mathbb{P}(S_{k-1}=n-1),$$

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,

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

which is a Poisson–binomial distribution. □

Example 1. For $n=2$, we obtain

$$\mathbb{P}(R(2)=k)=\frac{1}{k^d}\prod_{j=2}^{k-1}\Big(1-\frac{1}{j^d}\Big),\qquad k\geq 2,$$

while, if $n=3$,

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

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

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

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

Proposition 3. *The sequence $R(n)$, $n\geq 2$, forms a Markov chain with the following transition probabilities:*

$$\mathbb{P}(R(n)=k\mid R(n-1)=j)=\begin{cases} k^{-d} & \text{for } k=j+1,\\ k^{-d}\displaystyle\prod_{m=j+1}^{k-1}(1-m^{-d}) & \text{for } k>j+1,\\ \displaystyle\prod_{m=j+1}^{\infty}(1-m^{-d}) & \text{for } k=\infty>j,\end{cases}$$

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

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

$$\mathbb{P}(R(m)=j_m,\ 2\leq m\leq n)=\prod_{m=2}^{n}\mathbb{P}(I_{j_m}=1)\mathbb{P}\Big(\sum_{i=j_{m-1}+1}^{j_m-1} I_i=0\Big),\tag{6}$$

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

$$\mathbb{P}(R(n)=j_n\mid R(m)=j_m,\ 2\leq m\leq n-1)=\mathbb{P}(I_{j_n}=1)\mathbb{P}\Big(\sum_{i=j_{n-1}+1}^{j_n-1} I_i=0\Big).\tag{7}$$

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

Note that

$$\mathbb{P}(R(n-1)=j_{n-1})=\mathbb{P}(I_{j_{n-1}}=1)\mathbb{P}\Big(\sum_{i=2}^{j_{n-1}-1} I_i=n-2\Big)$$

and

$$\begin{aligned}&\mathbb{P}(R(n)=j_n,\ R(n-1)=j_{n-1})\\&\quad=\mathbb{P}(I_{j_n}=1)\mathbb{P}\Big(\sum_{i=j_{n-1}+1}^{j_n-1} I_i=0\Big)\mathbb{P}(I_{j_{n-1}}=1)\mathbb{P}\Big(\sum_{i=2}^{j_{n-1}-1} I_i=n-2\Big).\end{aligned}$$

Therefore, $\mathbb{P}(R(n)=j_n \mid R(n-1)=j_{n-1})$ is equal to the right-hand side of (7).

When our Markov chain moves from a finite state to the infinite state, we have

$$\begin{aligned}&\mathbb{P}(R(n)=\infty, R(m)=j_m,\ 2\le m\le n-1)\\&\quad=\prod_{m=2}^{n-1}\mathbb{P}(I_{j_m}=1)\mathbb{P}\Big(\sum_{i=j_{m-1}+1}^{j_m-1} I_i=0\Big)\mathbb{P}\Big(\sum_{i=j_m+1}^{\infty} I_i=0\Big).\end{aligned}$$

Using this result and the one from (6), we obtain

$$\mathbb{P}(R(n)=\infty \mid R(m)=j_m,\ 2\le m\le n-1)=\mathbb{P}\Big(\sum_{i=j_{n-1}+1}^{\infty} I_i=0\Big). \tag{8}$$

Now, noting that $\mathbb{P}(R(n-1)=j_{n-1})=\mathbb{P}(I_{j_{n-1}}=1)$ and

$$\mathbb{P}(R(n)=\infty,\ R(n-1)=j_{n-1})=\mathbb{P}(I_{j_{n-1}}=1)\mathbb{P}\Big(\sum_{i=j_{n-1}+1}^{\infty} I_i=0\Big),$$

then $\mathbb{P}(R(n)=\infty \mid R(n-1)=j_{n-1})$ is equal to the right-hand side of (8).

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

$$p_{n\,|\,n-1}:=\sum_{k\ge j+1}\mathbb{P}(R(n)=k \mid R(n-1)=j)+\mathbb{P}(R(n)=\infty \mid R(n-1)=j)=1$$

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

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

By induction, we prove that

$$\sum_{k=j+2}^{M}k^{-d}\prod_{m=j+1}^{k-1}(1-m^{-d})+\prod_{m=j+1}^{M}(1-m^{-d})=1-(j+1)^{-d}\quad\text{for each } M\in\mathbb{N}. \tag{9}$$

For step $M=j+2$:

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

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press

Now, we suppose that (9) holds for M and we also prove it holds for $M+1$. Thus,

$$\begin{aligned}
&\sum_{k=j+2}^{M+1} k^{-d} \prod_{m=j+1}^{k-1} (1-m^{-d}) + \prod_{m=j+1}^{M+1} (1-m^{-d}) \\
&\quad = \sum_{k=j+2}^{M} k^{-d} \prod_{m=j+1}^{k-1} (1-m^{-d}) + (M+1)^{-d} \prod_{m=j+1}^{M} (1-m^{-d}) \\
&\qquad + (M+1)^{-d} \prod_{m=j+1}^{M} (1-m^{-d}) \\
&\quad = (1-(j+1)^{-d})((M+1)^{-d} + 1 - (M+1)^{-d}) \\
&\quad = 1-(j+1)^{-d}.
\end{aligned}$$

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

Acknowledgement

This paper was written while the first author was a visiting professor at the Department of Decision Sciences, Bocconi University, Milan, Italy. He is grateful to his hosts for their hospitality and the extremely constructive atmosphere.

References

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*. John Wiley, New York.

Falk, M., Khorrami Chokami, A. and Padoan, S. A. (2017). Some results on joint record events. Preprint. Available at https://arxiv.org/abs/1707.06254.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn. Krieger, Melbourne, FL.

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

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

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

Zott, M. (2016). Extreme value theory in higher dimensions: max-stable processes and multivariate records. Doctoral Thesis, University of Würzburg.

https://doi.org/10.1017/jpr.2018.4 Published online by Cambridge University Press