Complete Records over independent FGM sequences

Amir Khorrami Chokami [1]

[1] ESOMAS Department, University of Torino and Collegio Carlo Alberto (e-mail: amir.khorramichokami@unito.it)

Abstract: Records are defined as variables greater than all the preceding ones in a sequence. The stochastic behavior of subsequent records over sequences of independent and identically distributed random variables is well known. However, the extension to the multivariate framework is an extremely difficult task. In this work, we study the case of bivariate records over sequences of random vectors (rv) where the dependence among their components is described by the Farlie-Gumbel-Morgenstern (FGM) family of distributions.

Keywords: complete records, standard max stable distribution, FGM copula.

1 Introduction

The study of multivariate maxima of rv is a challenging topic (see Resnick, 1987, Leonetti & Khorrami Chokami, 2022 among others), but records furnish a new way to tackle the problem as they give information on how often and to which extent maxima change. The theory on records is well developed in the case of independent and identically distributed (iid) sequences of random variables (see Galambos, 1987 and Falk *et al.*, 2018a). However, as soon as we relax the iid assumption to better reflect real-world data, the problem becomes immediately too difficult (we cite Falk *et al.*, 2020 for a study on univariate stationary Gaussian sequences). Multivariate records are an appealing topic of research. In $\mathbb{R}^d$, various definitions of records are possible. Here, we consider the so-called complete record (see Falk *et al.*, 2018b) and consider operations on vectors to be made componentwise. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ be iid rv: $\boldsymbol{X}_n$ is a *complete record* (CR) if $\boldsymbol{X}_n > \boldsymbol{M}_{n-1} = \max(\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_{n-1})$ and the appearance of a CR at time n is indicated as $R_n := \mathbb{1}(\boldsymbol{X}_n > \boldsymbol{M}_{n-1})$.

This paper investigates the difficult problem of describing the appearance of CRs and their distribution, under the following hypothesis: we know that a vector is a CR, but we do not know which one. It is still an open problem to find such results in the case of iid sequences of rv with a general copula. Here, we consider a sequence $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots \in \mathbb{R}^2$ of *standard max-stable* rv (*i.e.* with Negative-Exp(1) margins and such that $M_n \stackrel{d}{=} \boldsymbol{\eta}_1/n$) and FGM copula:

$$F(x,y) = \mathrm{e}^{(x+y)}\left(1+\lambda\left(1-\mathrm{e}^x\right)\left(1-\mathrm{e}^y\right)\right), \quad x,y \le 0 \text{ and } |\lambda| \le 1. \tag{1}$$

The usefulness of this copula lies in its manageable structure and intuitive interpretation of the parameter λ to describe dependence, which make the FGM distribution widely used in capital-allocation applications and in problems involving order statistics and their concomitants. A complete description of this copula is in Hashorva & Hüsler, 1999.

2 Complete Records

Theorem 1. *Let $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots$ be a sequence of bivariate standard max-stable rv with FGM copula. Then, the probability of appearance of a CR is*

$$\mathsf{P}(R_n = 1) = \frac{1}{n^2}\left(1+\lambda\left(\frac{n-1}{n+1}\right)^2\left(1+\frac{(n+1)^2+\lambda(n-2)^2}{(2n-1)^2}\right)\right) \tag{2}$$

for $n \in \mathbb{N}$ and the distribution of a rv given that it is a CR is

$$\mathsf{P}(\boldsymbol{\eta} \le \boldsymbol{z} \mid R_n = 1) = \frac{e^{n(z_1+z_2)}}{\mathsf{P}(R_n=1)}\left(\frac{1}{n^2}+\lambda\prod_{i=1}^{2}\left(\frac{2}{n+1}e^{z_i}-\frac{1}{n}\right)+\lambda\prod_{i=1}^{2}\left(\frac{1}{n}-\frac{e^{(n-1)z_i}}{2n-1}\right)+\right.$$
$$\left.+\lambda^2\prod_{i=1}^{2}\left(\frac{1}{n+1}e^{z_i}-\frac{1}{n}-\frac{e^{nz_i}}{n}+\frac{e^{(n-1)z_i}}{2n-1}\right)\right), \qquad \boldsymbol{z} \le 0. \tag{3}$$

Proof. Denote with $\eta^{(i)}$ the i-th component of $\boldsymbol{\eta}$. We firstly compute

$$\begin{aligned}\mathsf{P}(R_n=1) &= \mathsf{P}(\boldsymbol{\eta}_n > \boldsymbol{M}_{n-1}) = \mathsf{P}\left(\boldsymbol{\eta}_n > \frac{\boldsymbol{\eta}_1}{n-1}\right) = \mathsf{P}(\boldsymbol{\eta}_1 < (n-1)\boldsymbol{\eta}_n)\\ &= \int_{(-\infty,0]^2} \mathsf{P}\left(\eta_1^{(1)} < (n-1)x, \eta_2^{(1)} < (n-1)y \mid \boldsymbol{\eta}_n = (x,y)\right) f(x,y)\,\mathrm{d}x\,\mathrm{d}y\\ &= \int_{(-\infty,0]^2} \mathsf{P}\left(\eta_1^{(1)} < (n-1)x, \eta_2^{(1)} < (n-1)y\right) f(x,y)\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

The last equality follows by the independence assumption. Define

$$I(z_1,z_2) = \int_{-\infty}^{z_2}\int_{-\infty}^{z_1} \mathsf{P}\left(\eta_1^{(1)} < (n-1)x, \eta_2^{(1)} < (n-1)y\right) f(x,y)\,\mathrm{d}x\,\mathrm{d}y =$$
$$= \int_{-\infty}^{z_2}\int_{-\infty}^{z_1} \mathrm{e}^{n(x+y)}\left(1+\lambda\left(1-\mathrm{e}^{(n-1)x}\right)\left(1-\mathrm{e}^{(n-1)y}\right)\right)\left(1+\lambda\left(2\mathrm{e}^x-1\right)\left(2\mathrm{e}^y-1\right)\right)\mathrm{d}x\,\mathrm{d}y,$$

and note that $I(0,0) = \mathsf{P}(R_n = 1)$. After computations, we obtain

$$I(z_1, z_2) = \mathrm{e}^{n(z_1+z_2)} \left(\frac{1}{n^2} + \lambda \prod_{i=1}^{2} \left(\frac{2}{n+1} \mathrm{e}^{z_i} - \frac{1}{n} \right) + \lambda \prod_{i=1}^{2} \left(\frac{1}{n} - \frac{\mathrm{e}^{(n-1)z_i}}{2n-1} \right) + \right.$$
$$\left. + \lambda^2 \prod_{i=1}^{2} \left(\frac{1}{n+1} \mathrm{e}^{z_i} - \frac{1}{n} - \frac{\mathrm{e}^{nz_i}}{n} + \frac{\mathrm{e}^{(n-1)z_i}}{2n-1} \right) \right)$$

and we have

$$I(0,0) = \frac{1}{n^2} \left(1 + \lambda \left(\frac{n-1}{n+1} \right)^2 \left(1 + \frac{(n+1)^2 + \lambda (n-2)^2}{(2n-1)^2} \right) \right) = \mathsf{P}(R_n = 1).$$

Equation (3) follows by noticing that, for $\boldsymbol{z} \leq \boldsymbol{0}$,

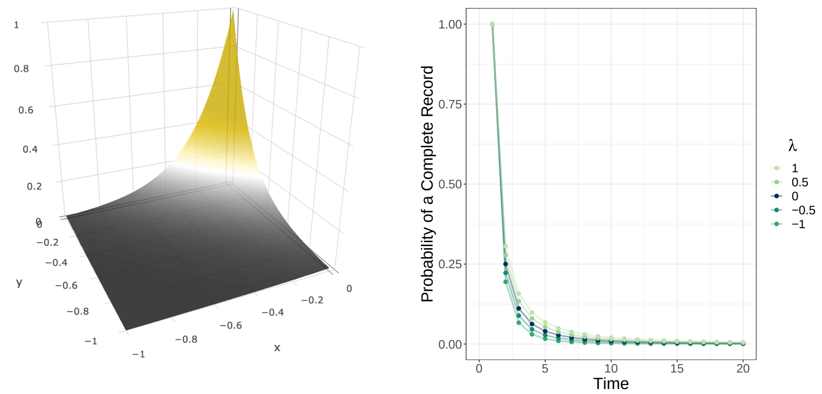
$$\mathsf{P}(\boldsymbol{\eta}_n \leq \boldsymbol{z} \mid R_n = 1) = \frac{\mathsf{P}(\boldsymbol{\eta}_n \leq \boldsymbol{z}, R_n = 1)}{\mathsf{P}(R_n = 1)} = \frac{\mathsf{P}(\boldsymbol{\eta}_n \leq \boldsymbol{z}, \boldsymbol{\eta}_1 < (n-1)\boldsymbol{\eta}_n)}{\mathsf{P}(R_n = 1)}$$
$$= \frac{1}{\mathsf{P}(R_n = 1)} \int_{(-\infty, \boldsymbol{z}]} \mathsf{P}\left(\eta_1^{(1)} < (n-1)x, \eta_2^{(1)} < (n-1)y \mid \boldsymbol{\eta}_n = (x,y) \right) f(x,y) \, \mathrm{d}x \, \mathrm{d}y.$$

The thesis follows by noticing that $\mathsf{P}(\boldsymbol{\eta}_n \leq \boldsymbol{z} \mid R_n = 1) = I(0,0)^{-1} I(z_1, z_2)$. □

Figure 1a represents an example of the cdf of a CR at time 4 given by Equation (3), when the parameter λ is set to 0.8, while Figure 1b shows the decay of the appearance of a CR as n increases (from Equation (2)), for various choices of λ. Note that $\lambda = 0$ indicates independence of the components of $\boldsymbol{\eta}$.

Remark 1. *Let $N = \sum_{n=2}^{\infty} R_n$ be the number of records after the first vector (which is the first record by definition). From Equation* (2)*, it holds that* $\mathsf{E}[N] = \sum_{n=2}^{\infty} \mathsf{P}(R_n = 1) < \infty$*, which implies by the first Borel-Cantelli lemma that* $\mathsf{P}(R_n = 1 i.o) = 0$*, that is a finite number of records. This is coherent with Theorem 5.3 in Goldie & Resnick, 1989.*

To conclude, this work tackles the problem of studying CRs over iid bivariate sequences of standard max-stable rv with FGM copula, under the hypothesis of not knowing the position of the CRs in their sequence. This approach is proven to furnish handy results and links with the Extreme Value Theory (see Falk *et al.*, 2018b and Falk *et al.*, 2020). We highlight that Equation (2) is independent of the chosen marginal distribution function of the considered rv (say F_X), provided that it is continuous, as $\eta = \log(F_X(X))$ in distribution. However, the distribution of a CR does depend on the marginal distribution. The extension to CR on sequences with a general copula and unfixed continuous margins is an ongoing project of the author.



(a) *Bivariate cdf of a CR.*

(b) *Decay of appearance of a CR.*

Figure 1: Panel (a) shows the bivariate cdf of $\boldsymbol{\eta}_4 \mid R_4 = 1$, with $\lambda = 0.8$. Panel (b) shows the decay of Equation (2), for different values of λ.

References

FALK, M., KHORRAMI CHOKAMI, A., & PADOAN, S. A. 2018a. On multivariate records from random vectors with independent components. *Journal of Applied Probability*, **55**(1), 43–53.

FALK, M., KHORRAMI CHOKAMI, A., & PADOAN, S. A. 2018b. Some results on joint record events. *Statistics and Probability Letters*, **135**, 11 – 19.

FALK, M., KHORRAMI CHOKAMI, A., & PADOAN, S. A. 2020. Records for time-dependent stationary Gaussian sequences. *Journal of Applied Probability*, **57**(03), 78–96.

GALAMBOS, J. 1987. *The Asymptotic Theory of Extreme Order Statistics*. 2 edn. Malabar: Krieger.

GOLDIE, CHARLES M., & RESNICK, SIDNEY I. 1989. Records in a partially ordered set. *Ann. Probab.*, **17**(2), 678–699.

HASHORVA, E, & HÜSLER, J. 1999. Extreme Values in FGM Random Sequences. *Journal of Multivariate Analysis*, **68**(2), 212–225.

LEONETTI, P., & KHORRAMI CHOKAMI, A. 2022. The maximum domain of attraction of multivariate extreme value distributions is small. *Electronic Communications in Probability*, **27**, 1 – 8.

RESNICK, SIDNEY I. 1987. *Extreme Values, Regular Variation, and Point Processes*. Applied Probability, vol. 4. New York: Springer.