Journal of Statistical Mechanics:
Theory and Experiment

PAPER

Exact and asymptotic properties of δ-records in the linear drift model

To cite this article: R Gouet *et al J. Stat. Mech.* (2020) 103201

View the article online for updates and enhancements.

IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

This content was downloaded from IP address 132.174.250.76 on 25/11/2020 at 21:59

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Classical statistical mechanics, equilibrium and non-equilibrium

Exact and asymptotic properties of δ-records in the linear drift model

R Gouet[1], **M Lafuente**[2,4], **F J López**[2,3] and **G Sanz**[2,3]

[1] Departamento de Ingeniería Matemática and CMM (UMI 2807, CNRS), Universidad de Chile, Avenida Blanco Encalada 2120, 837-0459, Santiago, Chile

[2] Departamento de Métodos Estadísticos, Facultad de Ciencias, Universidad de Zaragoza, C/Pedro Cerbuna, 12, 50009 Zaragoza, Spain

[3] Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, 50018 Zaragoza, Spain

E-mail: miguellb@unizar.es

Received 12 June 2020
Accepted for publication 2 September 2020
Published 12 October 2020

Online at stacks.iop.org/JSTAT/2020/103201
https://doi.org/10.1088/1742-5468/abb4dc

Abstract. The study of records in the linear drift model (LDM) has attracted much attention recently due to applications in several fields. In the present paper we study δ-records in the LDM, defined as observations which are greater than all previous observations, plus a fixed real quantity δ. We give analytical properties of the probability of δ-records and study the correlation between δ-record events. We also analyse the asymptotic behaviour of the number of δ-records among the first n observations and give conditions for convergence to the Gaussian distribution. As a consequence of our results, we solve a conjecture posed in *J. Stat. Mech.* **2010** P10013, regarding the total number of records in an LDM with negative drift. Examples of application to particular distributions, such as Gumbel or Pareto are also provided. We illustrate our results with a real data set of summer temperatures in Spain, where the LDM is consistent with the global-warming phenomenon.

Keywords: extreme value, extreme value statistics, exact results, stochastic processes

[4] Author to whom any correspondence should be addressed.

© 2020 IOP Publishing Ltd and SISSA Medialab srl
1742-5468/20/103201+31$33.00

Contents

1. Introduction

Extreme values and records have attracted large efforts and attention since the beginnings of statistics and probability, due to their intrinsic interest and their mathematical challenges. An important motivation for studying records comes from their connections with other interesting problems and, of course, from their countless practical applications in different fields such as climatology [1–4], sports [5–7], finance [8, 9] or biology [10]. Moreover, records have been used in statistical inference because, in

J. Stat. Mech. (2020) 103201

some contexts, data is inherently composed of record observations [11–14]. The classical probabilistic setting of independent and identically distributed random (i.i.d.) observations has been profusely studied. Main results in this framework can be found in the monographs [15–17]. In the last few years, there has been an increasing interest in the study of records in correlated observations such as random walks or time series [18–24].

An interesting departure from the i.i.d. model, which introduces time-dependence between observations, results from adding a deterministic linear trend to the i.i.d. observations, thus obtaining the so named linear drift model (LDM). This model was first introduced in [25] and later developed in [26–28]. The model was also considered in [29], under a wide range of scenarios, and has proven particularly useful in the study of global warming phenomena [4, 30]. Furthermore, the importance of this model is not only related to applications but also to its mathematical structure. For instance, the study of records in the LDM model can be helpful in determining whether the underlying distribution is heavy-tailed or not [31, 32]. Also, records statistics in random walks with a drift have been studied in [9, 33, 34].

Different generalizations of the notion of record, such as near-records [35–37] or δ-exceedance records [38, 39] have been proposed recently. We will work with δ-records, first introduced in [40], which are observations greater than all previous entries, plus a fixed quantity δ. In the i.i.d. setting, the distribution [41, 42], process structure [43] and asymptotic properties [44] of δ-records have been studied. In the case $\delta < 0$, where δ-records are more numerous than records, their use in statistical inference has been recently proposed and positively assessed; see [44–46]. In these articles, it is shown how the information of δ-records can be incorporated successfully into the likelihood of the sample, which is used for computing maximum likelihood and Bayes estimators and predictions of future records. The resulting estimators and predictions outperform those computed using records only; moreover, a slight modification of the sampling scheme for records yields δ-records with a low additional cost. The results are applied to examples of rainfall data and strength of materials.

In this work, we study δ-records from observations obeying the LDM, while revisiting some open questions about records. We analyse the positivity and continuity of the asymptotic δ-record probability as a function of δ and of the trend parameter c. We also obtain a law of large numbers and a central limit theorem for the counting process of δ-records, thus extending the corresponding results in [25]. Furthermore, we completely characterize the finiteness of the number of δ-records and, in particular, we solve a conjecture posed in [29], about the finiteness of the number of usual records in the LDM with negative trend.

We assess the effect of δ on the δ-record probabilities and correlations, for explicitly solvable models. Some of the results obtained in these examples are new and shed light on the behaviour of record events, when the underlying distribution is heavy-tailed. Finally we illustrate our results by analysing a real dataset of temperatures, which fits the LDM with a trend parameter consistent with the global-warming phenomenon.

J. Stat. Mech. (2020) 103201

2. δ-records in the LDM

Our objects of interest in this paper are δ-records, formally defined as follows: given a sequence of observations $(Y_n)_{n\geqslant 1}$ and $\delta \in \mathbb{R}$ a parameter, Y_1 is defined conventionally as δ-record and, for $j \geqslant 2$, Y_j is a δ-record if $Y_j > \max\{Y_1, \ldots, Y_{j-1}\} + \delta$.

Note that δ-records are just (upper) records, if $\delta = 0$. If $\delta > 0$, a δ-record is necessarily a record and δ-records are a subsequence of records. On the other hand, if $\delta < 0$, a δ-record can be smaller than the current maximum, so records are a subsequence of δ-records.

Throughout this paper we assume that the Y_n are random variables obeying the LDM, that is, Y_n can be represented as

$$Y_n = X_n + cn, \quad n \geqslant 1, \tag{1}$$

where $c \in \mathbb{R}$ is the trend parameter and $(X_n)_{n\geqslant 1}$ is a sequence of i.i.d. random variables, with (absolutely continuous) cumulative distribution function (cdf) F and probability density function f. Another important parameter of the model is the right-tail expectation of the X_j, defined as

$$\mu^+ = \int_0^\infty x f(x)\mathrm{d}x.$$

For simplicity, we assume the existence of an interval of real numbers $I = (x_-, x_+)$, with $-\infty \leqslant x_- < x_+ \leqslant \infty$, such that $f(x) > 0$, for all $x \in I$, and $f(x) = 0$ otherwise. Note that $x_- = \inf\{x : F(x) > 0\}$ and $x_+ = \sup\{x : F(x) < 1\}$.

Let $1_{j,\delta}$ denote the indicator of the event $\{Y_j$ is a δ-record$\}$. That is, $1_{j,\delta} = 1$ if $Y_j > \max\{Y_1, \ldots, Y_{j-1}\} + \delta$ and $1_{j,\delta} = 0$ otherwise. So, the number of δ-records up to index n is computed as $N_{n,\delta} = \sum_{j=1}^{n} 1_{j,\delta}$.

Under the LDM, the probability of $\{Y_j$ is a δ-record$\}$ is easily computed by conditioning, as

$$p_{j,\delta} := \mathbb{E}[1_{j,\delta}] = \int_{-\infty}^{\infty} \prod_{i=1}^{j-1} F(x + ci - \delta) f(x)\mathrm{d}x,$$

where $\mathbb{E}[\cdot]$ denotes the mathematical expectation. Moreover, the asymptotic δ-record probability is given by the formula

$$p_\delta := \lim_{n\to\infty} p_{n,\delta} = \int_{-\infty}^{\infty} \prod_{i=1}^{\infty} F(x + ci - \delta) f(x)\mathrm{d}x, \tag{2}$$

which is mathematically justified by the monotone convergence theorem for integrals; see [47, theorem 2.8.2].

In what follows we occasionally write $1_{j,\delta}(c), N_{n,\delta}(c), p_{j,\delta}(c), p_\delta(c)$, etc to emphasize the dependence on the trend parameter c.

J. Stat. Mech. (2020) 103201

3. Properties of the δ-record probabilities

We begin with a simple property about the asymptotic δ-record probability of an affine transformation of the LDM. Let $\tilde{X}_n = bX_n + a$, with $b > 0$, $a \in \mathbb{R}$, and $\tilde{Y}_n = \tilde{X}_n + cn$, $n \geqslant 1$. If $\tilde{p}_\delta(c)$ is the δ-record probability in this model, then it holds

$$\tilde{p}_\delta(c) = p_{\frac{\delta}{b}}\left(\frac{c}{b}\right).$$

We consider next some analytical properties of $p_{j,\delta}(c)$ and $p_\delta(c)$, as functions of c and δ. We note first that both are increasing in c and decreasing in δ. Moreover, it is easy to see that $p_{j,\delta}(c)$ is decreasing in j and continuous in c, converging to 1 as $c \to \infty$. The continuity of $p_\delta(c)$ is less clear because of the infinite product within the integral in (2).

3.1. Positivity of $p_\delta(c)$

We show that the positivity of $p_\delta(c)$ depends on c and δ and on the right-tail behaviour of F. We consider two cases depending on μ^+:

(a) $\mu^+ = \infty$. In this case $p_\delta(c) = 0$, for all $\delta, c \in \mathbb{R}$.

To justify this claim, we show that $\prod_{j=1}^{\infty} F(x + cj - \delta) = 0$, for all $x \in (x_-, x_+)$. If $c < 0$ the conclusion is immediate because $F(x + cj - \delta) \to 0$, as $j \to \infty$.

If $c = 0$, we note that $\mu^+ = \infty$ implies $x_+ = \infty$ and so, $F(x - \delta) < 1$. Thus $\prod_{j=1}^{\infty} F(x + cj - \delta) = 0$.

Finally, if $c > 0$, we note that $\mu^+ = \infty$ implies $\sum_{i=1}^{\infty} (1 - F(x + ci - \delta)) = \infty$, which in turn implies $\prod_{j=1}^{\infty} F(x + cj - \delta) = 0$. This follows from the definition of μ^+ and from Taylor's expansion of $\log(1 + x)$.

Distributions with $\mu^+ = \infty$ can be considered as 'right-heavy-tailed' and we observe that, for such distributions, the linear trend has no impact on the asymptotic probability of a δ-record. This class of distributions includes the Pareto and Fréchet, with shape parameter $\alpha \in (0, 1]$.

(b) $\mu^+ < \infty$. As in the previous case, we have three situations depending on the sign of c.

For $c < 0$, $p_\delta(c) = 0$, for all $\delta \in \mathbb{R}$, since $\prod_{j=1}^{\infty} F(x + cj - \delta) = 0$, for all $x \in (x_-, x_+)$. If $c = 0$,

$$p_\delta(0) = \int_{-\infty}^{\infty} \prod_{j=1}^{\infty} F(x - \delta) f(x) \mathrm{d}x = \int_{x_+ + \delta}^{\infty} f(x) \mathrm{d}x, \tag{3}$$

which is positive if and only if $x_+ < \infty$ and $\delta < 0$.

Finally, if $c > 0$, then $p_\delta(c) = 0$ if and only if $x_+ - x_- \leqslant \delta - c$. Indeed, note that, if $x_+ - x_- \leqslant \delta - c$, then $\mathbb{P}[Y_n > Y_{n-1} + \delta] = 0$, for all n, and so, only the first observation (by convention) is a δ-record. Conversely, if $x_+ - x_- > \delta - c$, then the interval $J := (x_-, x_+) \cap (x_- - c + \delta, \infty)$ is nonempty and, for every $x \in J$, we have $F(x + cj - \delta) \geqslant F(x + c - \delta) > 0$, for all j. Now, since $F(x + cj - \delta) \to 1$ as $j \to \infty$, and $\mu^+ < \infty$, we have $\sum_{j=1}^{\infty} (1 - F(x + cj - \delta)) < \infty$, which implies $\prod_{j=1}^{\infty} F(x + cj - \delta) > 0$ and, so $p_\delta(c) > 0$.

J. Stat. Mech. (2020) 103201

Summarizing the above findings, we state

Theorem 1. *$p_\delta(c) > 0$ if and only if $\mu^+ < \infty$ and one of the following conditions holds*

(a) $c > 0$ and $\delta < x_+ - x_- + c$,

(b) $c = 0, \delta < 0$ and $x_+ < \infty$.

3.2. Continuity of $p_\delta(c)$.

As commented at the beginning of this section, the continuity of $p_\delta(c)$ is not obvious. However, thanks to theorem 1 we can restrict attention to distributions F with finite right-tail expectation since, otherwise, $p_\delta(c)$ vanishes and continuity is trivial. Thus, we assume throughout this section that $\mu^+ < \infty$.

A first interesting fact, which is rigorously proved in proposition 6 of appendix A, is that $\prod_{i=1}^{\infty} F(x + ci - \delta)$ is continuous at every $c \neq 0$, for every $x \in (x_-, x_+)$, such that $x \neq x_- + \delta - c$. Then, thanks to the dominated convergence theorem of integration (theorem 2.8.1 in [47]), we conclude that $p_\delta(c)$ is continuous, at every $c \neq 0$.

The continuity at $c = 0$ is subtler to establish and depends on the sign of δ and the finiteness of x_+, the right-end point of F. Note that, for every $c > 0$ and $N \geqslant 1$, we have

$$\prod_{j=1}^{\infty} F(x - \delta) \leqslant \prod_{j=1}^{\infty} F(x + cj - \delta) \leqslant \prod_{j=1}^{N} F(x + cj - \delta).$$

Then, taking the limit as $c \to 0^+$ in the above inequalities,

$$\prod_{j=1}^{\infty} F(x - \delta) \leqslant \lim_{c \to 0^+} \prod_{j=1}^{\infty} F(x + cj - \delta) \leqslant F(x - \delta)^N.$$

Therefore, $\lim_{c \to 0^+} \prod_{j=1}^{\infty} F(x + cj - \delta)$ is 0, if $x < x_+ + \delta$, and 1 otherwise. Then, by the dominated convergence theorem,

$$\lim_{c \to 0^+} p_\delta(c) = \int_{-\infty}^{\infty} \lim_{c \to 0^+} \prod_{j=1}^{\infty} F(x + cj - \delta) f(x) \mathrm{d}x = \int_{x_+ + \delta}^{\infty} f(x) \mathrm{d}x.$$

Thus, $p_\delta(c)$ is right-continuous at $c = 0$ by (3). Regarding left-continuity at 0, recall that $p_\delta(c) = 0$ for $c < 0$. So, $p_\delta(c)$ is discontinuous at 0 if and only if $x_+ < \infty$ and $\delta < 0$.

We now show the continuity of $p_\delta(c)$ as a function of δ. The result is trivial if $c < 0$, since $p_\delta(c) = 0$, for all $\delta \in \mathbb{R}$. For $c = 0$, note that, by (3), $p_\delta(0) = 1 - F(x_+ + \delta)$, which is continuous since F is a continuous function.

If $c > 0$ and $(\delta_n)_{n \geqslant 1}$ is a sequence converging to δ, we prove that

$$\lim_{n \to \infty} \prod_{i=1}^{\infty} F(x + ci - \delta_n) = \prod_{i=1}^{\infty} F(x + ci - \delta), \tag{4}$$

for all $x \in (x_-, x_+)$, $x \neq x_- + \delta - c$. Indeed, let $x < x_- + \delta - c$, then $F(x + c - \delta) = 0$ yielding $\prod_{i=1}^{\infty} F(x + ci - \delta) = 0$. Also $F(x + c - \delta_n) = 0$ for n large enough and (4) follows. Let now $x > x_- + \delta - c$ and $\varepsilon > 0$ such that $x + c - \delta - \varepsilon > x_-$. Then, for n large

J. Stat. Mech. (2020) 103201

enough, we have $|\delta_n - \delta| < \varepsilon$ and

$$-\sum_{i=1}^{\infty} \log\ F(x+ci-\delta_n) \leqslant -\sum_{i=1}^{\infty} \log\ F(x+ci-(\delta+\varepsilon)) < \infty,$$

since $\mu^+ < \infty$. So (4) holds, and continuity follows.

In the following theorem we summarize conditions for continuity of $p_\delta(c)$.

Theorem 2. *The asymptotic δ-record probability $p_\delta(c)$, as a function of c, δ, is*

(a) continuous at every $c \neq 0$ and right-continuous at $c = 0$, for all δ;
(b) discontinuous at $c = 0$ if and only if $x_+ < \infty, \delta < 0$, and
(c) continuous in δ, for all c.

4. Exactly solvable models

In general it is not possible to compute exactly the probabilities $p_{j,\delta}$ or p_δ. We show below explicit results for the Gumbel distribution and for particular instances of the Dagum family of distributions [48].

4.1. The Gumbel distribution

Let $F(x) = \exp(-\exp(-x))$, for $x \in \mathbb{R}$, be the Gumbel distribution. Note that $F(x+cj-\delta) = F(x)^{\mathrm{e}^{-cj+\delta}}$. Then, if $c \neq 0$,

$$\prod_{j=1}^{n-1} F(x+cj-\delta) = F(x)^{\sum_{j=1}^{n-1} \mathrm{e}^{-cj+\delta}} = F(x)^{\mathrm{e}^{\delta}\frac{\mathrm{e}^{-c}-\mathrm{e}^{-nc}}{1-\mathrm{e}^{-c}}}$$

and, if $c = 0$, $\prod_{j=1}^{n-1} F(x+cj-\delta) = F(x)^{(n-1)\mathrm{e}^{\delta}}$. So, from (2) we get

$$\begin{aligned} p_{n,\delta}(c) &= \int_{-\infty}^{\infty} F(x)^{\mathrm{e}^{\delta}\frac{\mathrm{e}^{-c}-\mathrm{e}^{-nc}}{1-\mathrm{e}^{-c}}} f(x)\mathrm{d}x \\ &= \frac{1-\mathrm{e}^{-c}}{1-\mathrm{e}^{-c}+\mathrm{e}^{\delta}(\mathrm{e}^{-c}-\mathrm{e}^{-nc})}, \end{aligned}$$

if $c \neq 0$, and

$$p_{n,\delta}(0) = \frac{1}{(n-1)\mathrm{e}^{\delta}+1}.$$

Note that, taking limits as $n \to \infty$, in the above formulas, we obtain

$$p_\delta(c) = \frac{1-\mathrm{e}^{-c}}{\mathrm{e}^{\delta}\ \mathrm{e}^{-c}+1-\mathrm{e}^{-c}} = \frac{1}{1+\frac{\mathrm{e}^{-c}}{1-\mathrm{e}^{-c}}\mathrm{e}^{\delta}},$$

if $c > 0$ and $p_\delta(c) = 0$, if $c \leqslant 0$, as expected from theorem 1.

Also, for every $c > 0$, $p_\delta(c)$ decreases with δ as a logistic function of $-\delta$. Figure 1 shows the behaviour of $p_\delta(c)$ as a function of δ and c.

J. Stat. Mech. (2020) 103201

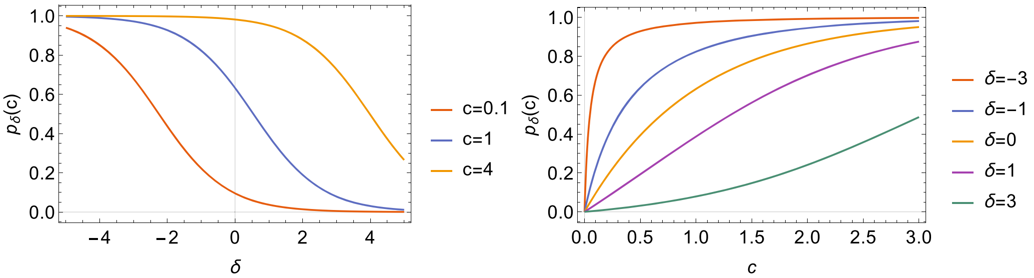

Figure 1. Asymptotic δ-record probability $p_\delta(c)$ for the Gumbel distribution as a function of δ and c.

4.2. The Dagum family of distributions

The Dagum distribution has cdf given by $F(x) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-q} 1_{\{x \geqslant 0\}}$, where a, b, q are positive parameters. Note that if $q = 1$, the distribution is referred to as log-logistic [49]. Also, the Pareto distribution [50] with cdf

$$F(x) = (1 - 1/x)1_{\{x \geqslant 1\}}, \tag{5}$$

can be seen as a shifted version of the Dagum family, with $a = b = q = 1$. For simplicity, in this section we limit our attention to the case $a = 1$, which has $\mu^+ = \infty$.

By theorem 1 we know that $p_\delta(c) = 0$, for every $c, \delta \in \mathbb{R}$, so we chose to analyse the speed of convergence of $p_{n,\delta}(c)$ to 0, for some values of c, δ. To that end, observe that the formula for $p_{n,\delta}$ takes the manageable form

$$p_{n,\delta}(c) = \int_{(\delta-c)^+}^{\infty} \prod_{i=1}^{n-1} \left(\frac{x + ci - \delta}{x + b + ci - \delta}\right)^q f(x)\mathrm{d}x, \tag{6}$$

which becomes simpler if we further assume that $c = b$ (that is, the trend parameter of the LDM is equal to the scale parameter of the distribution). From (6) we get

$$p_{n,\delta}(c) = \int_{(\delta-c)^+}^{\infty} \left(\frac{x + c - \delta}{x + cn - \delta}\right)^q f(x)\mathrm{d}x. \tag{7}$$

We introduce the notation $p_{n,\delta}^{(q)}(c)$ to make explicit the dependence of $p_{n,\delta}(c)$ on q. First, for records ($\delta = 0$) we have,

$$p_{n,0}^{(q)}(c) = cq \int_0^{\infty} x^{q-1}(x + cn)^{-q}(x + c)^{-1}\mathrm{d}x$$

$$= qn^{-q} \int_0^1 t^{q-1}(1 - t(n-1)/n)^{-q}\mathrm{d}t \tag{8}$$

$$= \frac{q}{(n-1)^q} \int_1^n \frac{(y-1)^{q-1}}{y}\mathrm{d}y, \tag{9}$$

where the second equality follows from the change of variable $x = ct/(1 - t)$ and the third from $1 - t(n-1)/n = 1/y$.

J. Stat. Mech. (2020) 103201

Observe that (8) and (9) do not depend on c and so, for the sake of simplicity, we write $p_{n,0}^{(q)}$. Moreover, from formula (8) we see that

$$p_{n,0}^{(q)} = n^{-q}\ {}_2F_1\left(q, q\,; q+1\,;(n-1)/n\right),$$

where ${}_2F_1$ is the Gauss hypergeometric function.

Also, from (9) and using the binomial expansion, for $q = 1, 2, \ldots$, we readily obtain

$$p_{n,0}^{(q)} = \frac{q}{(n-1)^q}\left((-1)^{q-1}\ \log\ n + \sum_{k=1}^{q-1}\binom{q-1}{k}\frac{(-1)^{q-1-k}}{k}(n^k-1)\right). \qquad (10)$$

The asymptotic behaviour of $p_{n,0}^{(q)}$, for any $q \in (0, \infty)$, can be obtained from (9). For $q = 1$, (10) yields $p_{n,0}^{(1)} = \frac{1}{n-1}\ \log\ n$. For $q > 1$, the leading term in the integral in (9) is y^{q-2}, so $p_{n,0}^{(q)} \sim \frac{q}{q-1}\frac{1}{n}$. For $q \in (0, 1)$, the integral in (9) converges and, using formula 3.191.2 in [51], we get

$$p_{n,0}^{(q)} \sim n^{-q}q\int_1^{\infty}\frac{(y-1)^{q-1}}{y}\,\mathrm{d}y = n^{-q}q\Gamma(1-q)\Gamma(q).$$

Thus,

$$p_{n,0}^{(q)} \sim \begin{cases} n^{-q}q\Gamma(1-q)\Gamma(q), & \text{if } 0 < q < 1,\\ \log(n)/n, & \text{if } q = 1,\\ n^{-1}\dfrac{q}{q-1}, & \text{if } q > 1. \end{cases} \qquad (11)$$

It is interesting to observe that the limiting behaviour of $p_{n,0}^{(q)}$, as a function of the power of the tail q, seems to match the asymptotic behaviour of $p_{n,0}(c)$ when F is the Fréchet distribution ($F(x) = \exp(-x^{-1})$, $x > 0$) and the tuning parameter is the trend c, studied in [28].

We now consider $\delta \neq 0$ and investigate whether $p_{n,\delta}^{(q)}/p_{n,0}^{(q)} \to 1$, as $n \to \infty$. This result can be expected since, as $\mu^+ = \infty$, the variables X_n take very large values, so δ may have little influence on the probability of δ-record, in the long term.

From (7) we may evaluate $p_{n,\delta}^{(q)}$, for any $q \in \mathbb{N}$, although the computation becomes lengthy as q grows. We have carried out the computation with values of q from 1 to 7, and obtained

$$p_{n,\delta}^{(1)} \sim \frac{\log(n)}{n}, \qquad p_{n,\delta}^{(q)} \sim \frac{q}{q-1}\frac{1}{n}, \quad q = 2, \ldots, 7.$$

So, from (11) we have $p_{n,\delta}^{(q)}/p_{n,0}^{(q)} \to 1$, at least for $q = 1, \ldots, 7$.

For noninteger values of $q \in (0, \infty)$, the limit behaviour of (7) is harder to analyse. To get a tractable expression, we impose $\delta = c$. Proceeding as above, we have, for $n > 2$,

$$p_{n,\delta}^{(q)} = \frac{q(n-1)^q}{(n-2)^{2q}}\int_1^{n-1}\frac{(y-1)^{2q-1}}{y^{q+1}}\mathrm{d}y.$$

J. Stat. Mech. (2020) 103201

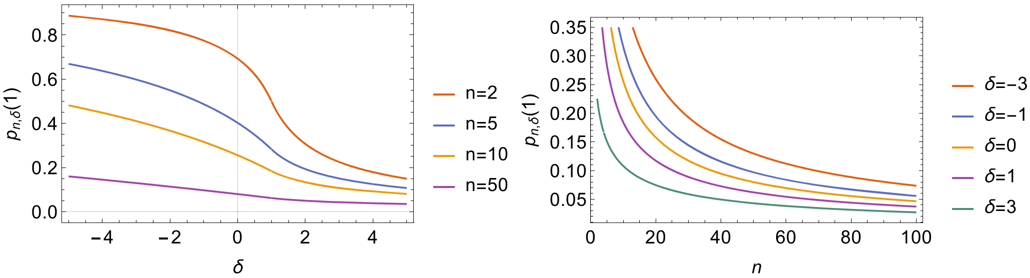

Figure 2. δ-record probability $p_{n,\delta}(c)$ for the Pareto distribution as a function of δ and n with $c = 1$.

Therefore, we have

$$p_{n,\delta}^{(q)} \sim \begin{cases} n^{-q}\dfrac{\Gamma(2q)\Gamma(1-q)}{\Gamma(q)}, & \text{if } 0 < q < 1, \\ \log(n)/n, & \text{if } q = 1, \\ n^{-1}\dfrac{q}{q-1}, & \text{if } q > 1. \end{cases}$$

So, under the above stated conditions, $p_{n,\delta}^{(q)} \sim p_{n,0}$, for $q \geqslant 1$, but this is not the case if $q \in (0, 1)$.

To conclude this example we study the Pareto distribution, defined in (5), taking $c = 1$. From (7), the probability of δ-record is explicitly computed as

$$\begin{aligned} p_{n,\delta} &= \int_{\max\{1,\delta\}}^{\infty} \frac{x-\delta}{x^2(x+n-1-\delta)}\,\mathrm{d}x \\ &= \frac{1}{(n-1-\delta)^2}\left((n-1)\log\left(\frac{n-\min\{1,\delta\}}{\max\{1,\delta\}}\right) - \min\{1,\delta\}(n-1-\delta)\right), \end{aligned} \tag{12}$$

if $\delta \neq n-1$ and $p_{n,\delta} = \frac{1}{2(n-1)}$, if $\delta = n-1$. Figure 2 shows the behaviour of $p_{n,\delta}$ as a function of n and δ.

5. Correlations

The indicators of δ-records are in general not independent in the case of i.i.d. random variables; see [44]. In [32] the authors study the dependence of record events in the LDM, by means of the following dependence index ($\delta = 0$ in their case)

$$l_n(c,\delta) := \frac{\mathbb{P}[\text{obs. } n \text{ and } n+1 \text{ are } \delta-\text{records}]}{\mathbb{P}[\text{obs. } n \text{ is } \delta-\text{record}]\mathbb{P}[\text{obs. } n+1 \text{ is } \delta-\text{record}]} = \frac{\mathbb{E}[1_{n,\delta}1_{n+1,\delta}]}{\mathbb{E}[1_{n,\delta}]\mathbb{E}[1_{n+1,\delta}]}.$$

If the events are independent, then $l_n(c,\delta) = 1$. Otherwise, values greater or smaller than 1 indicate positive or negative correlation, respectively. That is, neighbouring δ-records tend to attract or repel each other, if $l_n > 1$ or $l_n < 1$.

In order to manipulate $\mathbb{E}[1_{n,\delta}1_{n+1,\delta}]$ we consider the decomposition

$$\mathbb{E}[1_{n,\delta}1_{n+1,\delta}] = \mathbb{E}[1_{n,\delta}1_{n+1,\delta}1_{\{Y_n<Y_{n+1}\}}] + \mathbb{E}[1_{n,\delta}1_{n+1,\delta}1_{\{Y_n>Y_{n+1}\}}], \tag{13}$$

which, for $\delta < 0$, can be written as

$$\mathbb{E}[1_{n,\delta}1_{n+1,\delta}] = \int_{-\infty}^{\infty}\left(\int_{s-c}^{\infty}\prod_{j=1}^{n-1}F(s+cj-\delta)f(t)\mathrm{d}t + \int_{s-c+\delta}^{s-c}\prod_{j=2}^{n}F(t+cj-\delta)f(t)\mathrm{d}t\right)f(s)\mathrm{d}s$$
$$= \int_{-\infty}^{\infty}\left((1-F(s-c))\prod_{j=1}^{n-1}F(s+cj-\delta) + \int_{s-c+\delta}^{s-c}\prod_{j=2}^{n}F(t+cj-\delta)f(t)\mathrm{d}t\right)f(s)\mathrm{d}s, \tag{14}$$

and, for $\delta \geqslant 0$,

$$\mathbb{E}[1_{n,\delta}1_{n+1,\delta}] = \int_{-\infty}^{\infty}\int_{s-c+\delta}^{\infty}\prod_{j=1}^{n-1}F(s+cj-\delta)f(t)\mathrm{d}t\ f(s)\mathrm{d}s$$
$$= \int_{-\infty}^{\infty}(1-F(s-c+\delta))\prod_{j=1}^{n-1}F(s+cj-\delta)f(s)\mathrm{d}s, \tag{15}$$

since the second term in (13) vanishes.

As for $\mathbb{E}[1_{n,\delta}]$, it is not possible to explicitly compute $\mathbb{E}[1_{n,\delta}1_{n+1,\delta}]$, in general. Nevertheless, it is still possible to describe the behaviour of the dependence index in some particular cases.

5.1. The Gumbel distribution

Let $c > 0$ and F the Gumbel distribution, as in section 4.1. When $\delta < 0$ and $n \to \infty$, elementary but lengthy computations yield

$$\lim_{n\to\infty}\mathbb{E}[1_{n,\delta}1_{n+1,\delta}] = \frac{(\mathrm{e}^c-1)^2(\mathrm{e}^c-\mathrm{e}^\delta+1)}{(\mathrm{e}^c+\mathrm{e}^\delta-1)(\mathrm{e}^{2c}+\mathrm{e}^\delta-1)}$$

and

$$l_\infty(c,\delta) := \lim_{n\to\infty} l_n(c,\delta) = \frac{(\mathrm{e}^c+\mathrm{e}^\delta-1)(\mathrm{e}^c-\mathrm{e}^\delta+1)}{(\mathrm{e}^{2c}+\mathrm{e}^\delta-1)}.$$

By differentiating with respect to c, we see that $l_\infty(c,\delta)$ is decreasing in c and bounded below by 1, since $\lim_{c\to\infty} l_\infty(c,\delta) = 1$. With respect to δ we find that the derivative $\frac{\partial l_\infty}{\partial\delta}$ vanishes at

$$\delta = \log(1-\mathrm{e}^{2c}+\sqrt{\mathrm{e}^{4c}-\mathrm{e}^{2c}}),$$

and then, for any c,

$$\max_{\delta<0} l_\infty(c,\delta) = \frac{2\mathrm{e}^{2c}\left(\sqrt{\mathrm{e}^{2c}(\mathrm{e}^{2c}-1)}-\mathrm{e}^{2c}+1\right)}{\sqrt{\mathrm{e}^{2c}(\mathrm{e}^{2c}-1)}} = 2\left(\mathrm{e}^{2c}-\sqrt{2\mathrm{e}^{3c}\ \sinh(c)}\right).$$

J. Stat. Mech. (2020) 103201

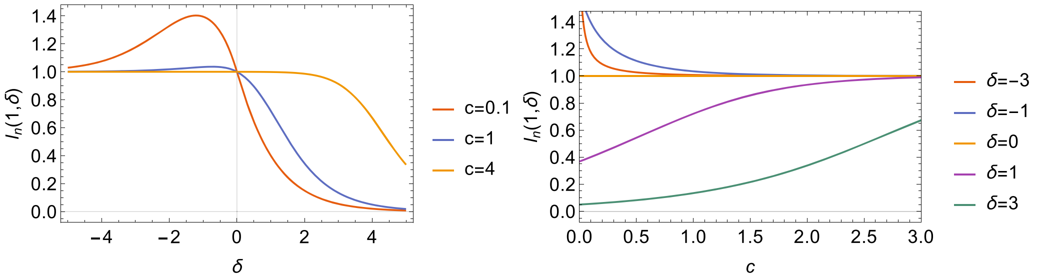

Figure 3. Dependence index $l_\infty(c, \delta)$ for the Gumbel distribution.

Note also that $\lim_{\delta\to-\infty} l_\infty(c, \delta) = 1$.

For $\delta \geqslant 0$,

$$\lim_{n\to\infty} \mathbb{E}[1_{n,\delta}1_{n+1,\delta}] = \frac{\mathrm{e}^c(\mathrm{e}^c - 1)^2}{(\mathrm{e}^c + \mathrm{e}^\delta - 1)(\mathrm{e}^{c+\delta} - \mathrm{e}^c + \mathrm{e}^{2c} - \mathrm{e}^\delta + \mathrm{e}^{2\delta})}$$

and

$$l_\infty(c, \delta) = \frac{\mathrm{e}^c(\mathrm{e}^c + \mathrm{e}^\delta - 1)}{\mathrm{e}^{c+\delta} - \mathrm{e}^c + \mathrm{e}^{2c} - \mathrm{e}^\delta + \mathrm{e}^{2\delta}}.$$

We note that $l_\infty(c, \delta) = 1, \forall c > 0$, if $\delta = 0$, which results in the asymptotic independence of consecutive record indicators in the LDM. Also, there are no critical points for the index when $\delta \geqslant 0$. So, in this case $l_\infty(c, \delta)$ is increasing in c with $\lim_{c\to\infty} l_\infty(c, \delta) = 1$, and decreasing in δ, with $\lim_{\delta\to\infty} l_\infty(c, \delta) = 0$, as can be seen in figure 3. Gathering these results, we conclude that $l_\infty(c, \delta) > 1$ if and only if $\delta < 0$. The asymptotic independence for records ($\delta = 0$) was proved in [27]; we have shown here that δ-records attract each other for $\delta < 0$ and repel each other for $\delta > 0$.

5.2. The Pareto distribution

Let $c = 1$ and F be as given in (5). The probability of δ-record is given in section 4.2. Computations of $l_n(c, \delta)$ are cumbersome and the explicit expression of $l_n(1, \delta)$ can be found in appendix A.5.

We have $\lim_{\delta\to-\infty} l_n(1, \delta) = 1$ and $\lim_{\delta\to\infty} l_n(1, \delta) = 1 - \log(2) \approx 0.3069$, for every $n > 1$. Also, $\lim_{n\to\infty} l_n(1, \delta) = \infty$, for all $\delta \in \mathbb{R}$, that is, δ-record-attraction grows unboundedly, as n increases. Moreover, it can be proved that $l_n(1, \delta) \sim C\frac{n}{(\log\ n)^2}$ as $n \to \infty$, where C is a constant depending on δ.

The sublinear growth of $l_n(1, \delta)$ as n increases can be observed in the right panel of figure 4, for different values of δ, as well as the decrease in δ. Also, for fixed n (left panel of figure 4), there is a negative value of δ where the correlation reaches a maximum, as in the Gumbel case. Note that, for negative and small positive values of δ, $l_n(1, \delta) > 1$, while, for big values of δ, $l_n(1, \delta) < 1$.

J. Stat. Mech. (2020) 103201

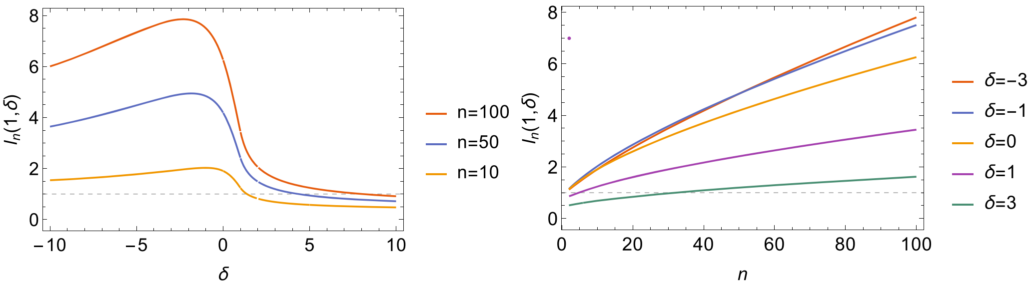

Figure 4. Dependence index $l_n(1, \delta)$ for the Pareto distribution as a function of δ and n.

J. Stat. Mech. (2020) 103201

6. Asymptotic behaviour of $N_{n,\delta}$

In sections 3 and 4 we have presented properties of the probability that observation n is a δ-record. In this section we analyse the random variable $N_{n,\delta}$, defined as the number of δ-records among the first n observations, and study its behaviour as $n \to \infty$.

Depending on F, c and δ, it might be the case that only finitely many δ-records are observed. We give necessary and sufficient conditions for this to happen. On the other hand, if $N_{n,\delta}$ grows to infinity, we investigate if the ratio $N_{n,\delta}/n$ converges (in a certain stochastic sense) to p_δ and, in that case, how the fluctuations of $N_{n,\delta}/n$ around p_δ are distributed.

Recall that, in the classical record model ($c = 0$), the number of records $N_{n,0}$ grows to infinity, and there are universal results ensuring that, for any continuous F, $N_{n,0}/\log n$ converges to 1, almost surely (a.s.) and $(N_{n,0} - \log n)/(\log n)^{1/2}$ has, asymptotically, a standard Gaussian distribution. However, when $\delta \neq 0$, results in [44, 52] for the model with $c = 0$, show that $N_{n,\delta}$ may grow to a finite limit and, when it diverges, the corresponding limit laws depend both on δ and F. We begin by analysing the situation where $N_{n,\delta}$ has a finite limit.

6.1. Finiteness of the total number of δ-records

Let $N_{\infty,\delta} = \lim_{n\to\infty} N_{n,\delta}$ be the total number of δ-records along the sequence $(Y_n)_{n\geqslant 1}$. In this section we find necessary and sufficient conditions for the finiteness of $N_{\infty,\delta}$ and $\mathbb{E}[N_{\infty,\delta}]$.

Clearly, these questions are related to the asymptotic behaviour of $p_{n,\delta}$. If $p_\delta > 0$, then we can expect $N_{\infty,\delta} = \infty$. On the other hand, if $p_\delta = 0$, it may happen that $N_{n,\delta}$ grows sublinearly to ∞ or $N_{\infty,\delta} < \infty$. Since, by theorem 1, the positivity of p_δ is linked to the finiteness of μ^+, we split the analysis into two cases:

(a) $\mu^+ = \infty$. In this situation, $N_{\infty,\delta} = \infty$ a.s. for any $c, \delta \in \mathbb{R}$.

To check this assertion, we first prove that $M_n := \max\{Y_1, \ldots, Y_n\} \to \infty$. Observe that $\mu^+ = \infty$ implies $x_+ = \infty$ and

$$\sum_{n=1}^{\infty} \mathbb{P}[Y_n > a] = \sum_{n=1}^{\infty} \mathbb{P}[X_n > a - cn] = \sum_{n=1}^{\infty} (1 - F(a - cn)) = \infty, \quad \forall\, a \in \mathbb{R}. \tag{16}$$

From (16) and the second Borel–Cantelli lemma, we conclude that $Y_n > a$ infinitely often (i.o.), for any a, and so, $M_n \to \infty$, with probability one. This fact clearly implies $N_{\infty,0} = \infty$. Now, since, for $\delta < 0$, $N_{\infty,\delta} \geqslant N_{\infty,0}$, we get $N_{\infty,\delta} = \infty$. On the other hand, for $\delta > 0$, the event

$$\{X_n + (c-\delta)n > \max_{1\leqslant j\leqslant n-1}\{X_j + (c-\delta)j\}\} \text{ implies } \{X_n + cn > \max_{1\leqslant j\leqslant n-1}\{X_j + cj\} + \delta\},$$

that is, $1_{n,0}(c-\delta) \leqslant 1_{n,\delta}(c)$. Therefore, $N_{\infty,\delta}(c) \geqslant N_{\infty,0}(c-\delta) = \infty$.

(b) $\mu^+ < \infty$. We distinguish three scenarios depending on the sign of c.

If $c > 0$, we first assume $x_+ - x_- > \delta - c$. In this case, we have $p_\delta > 0$ and so $N_{\infty,\delta} = \infty$ is an immediate consequence of the law of large numbers in theorem 5 below. If $x_+ - x_- \leqslant \delta - c$, only the first observation will be a δ-record as shown in section 3.1, so $N_{\infty,\delta} = 1$.

If $c = 0$ and $\delta \leqslant 0$, then $N_{\infty,\delta} = \infty$, since $N_{\infty,\delta} \geqslant N_{\infty,0} = \infty$. If $c = 0$ and $\delta > 0$, the situation is more complicated. In fact, $N_{\infty,\delta} < \infty$ if and only if

$$\int_0^\infty \frac{1 - F(x+\delta)}{(1-F(x))^2} f(x)\mathrm{d}x < \infty,$$

which is also equivalent to $\mathbb{E}[N_{\infty,\delta}] < \infty$. This is shown in proposition 7 of the appendix A, by relating this question to the counting process of geometric records, as studied in [52].

If $c < 0$, we proceed as in (16) to obtain

$$\sum_{n=1}^{\infty} \mathbb{P}[Y_n > a] = \sum_{n=1}^{\infty} \mathbb{P}[X_1 > a - cn] < \infty, \quad \forall\, a \in \mathbb{R},$$

where the last inequality follows from $\mu^+ < \infty$. Thus, the first Borel–Cantelli lemma ensures that $\mathbb{P}[Y_n > a \text{ i.o.}] = 0$, for all $a \in \mathbb{R}$, so $Y_n \to -\infty$. Then, there exists a random variable $N < \infty$ such that $\lim_{n\to\infty} M_n = M_N$ and, consequently, $N_{\infty,\delta} < \infty$. In this case, we can also prove that $\mathbb{E}[N_{\infty,\delta}] < \infty$; see proposition 9.

Summarizing the above, we give a complete characterization of the (a.s.) finiteness of the number of δ-records in the next theorem.

Theorem 3. $N_{\infty,\delta} < \infty$ a.s. *if and only if one of the following conditions holds*

(a) $c < 0$ *and* $\mu^+ < \infty$,

(b) $c = 0$, $\delta > 0$ *and* $\int_0^\infty \frac{1-F(x+\delta)}{(1-F(x))^2} f(x)\mathrm{d}x < \infty$,

(c) $c > 0$ *and* $x_+ - x_- \leqslant \delta - c$.

Moreover, $N_{\infty,\delta} < \infty$ *a.s. if and only if* $\mathbb{E}[N_{\infty,\delta}] < \infty$.

Remark 4. Theorem 3 answers a conjecture posed in [29], stating that the expected number of records ($\delta = 0$) in the LDM, with negative trend, remains finite, based on the observed exponential decay of p_n, in a particular case. We have shown that the conjecture holds if and only if $\mu^+ < \infty$.

J. Stat. Mech. (2020) 103201

6.2. Growth of $N_{n,\delta}$ to infinity.

We now turn our attention to the case $N_{\infty,\delta} = \infty$. More precisely, we are interested in the convergence of the proportion of δ-records to p_δ. For records ($\delta = 0$) it was shown in [25, 26] that $N_{n,0}/n \to p_0$ and that fluctuations of $N_{n,0}$ around p_0 are asymptotically Gaussian.

We show here that these results carry over to the case of $\delta \neq 0$ but leave the proof for appendices A.3 and A.4. As in the aforementioned works, we assume $\mu^+ < \infty$ and $c > 0$ and, additionally, that $x_+ - x_- > \delta - c$. Note that, by theorems 1 and 3, we have $p_\delta > 0$ and $N_{\infty,\delta} = \infty$.

Theorem 5. *Assume $\mu^+ < \infty$, $c > 0$ and $x_+ - x_- > \delta - c$. Then, as $n \to \infty$,*

(a) $N_{n,\delta}/n \to p_\delta$ a.s. and $\mathbb{E}[N_{n,\delta}/n] \to p_\delta$.
(b) If, additionally, $\int_0^\infty x^2 f(x)\mathrm{d}x < \infty$, then $\sqrt{n}(N_{n,\delta}/n - p_\delta) \xrightarrow{\mathcal{D}} N(0, \sigma_\delta^2)$, where $\xrightarrow{\mathcal{D}}$ stands for convergence in distribution and σ_δ^2 is defined in (22).

As it can be seen in the proof of theorem 5(a) in the appendix A, the assumption on independence of the X_n can be relaxed to stationary and ergodic and prove that $N_{n,\delta}/n \to \mathbb{E}[1^*_{0,\delta}]$, defined in (21). This is useful because it allows to deal with a wider range of scenarios, including stationary autoregressive-moving-average processes. Note, however, that $\mathbb{E}[1^*_{0,\delta}]$ could differ from p_δ in (2).

7. Illustration

We present a practical application of theorem 5 to a real dataset of temperatures, where convergence to the stationary regime is seen for quite small values of n. As pointed out in the introduction, the LDM has been used by [4, 30] to model temperature data in the framework of climate-change.

Our dataset consists of means of daily maximum temperatures (in degrees celsius), for every month of July, from 1951 to 2019, in the city of Saragossa, Spain. See figure 5 for a data plot. For δ-records we choose the value $\delta = -1$, which is arbitrary and does not respond to any specific reason, other than interpretability of the example. Note that a year will have a δ-record temperature if the maximum average temperature in July is a record or if it is at a distance smaller than 1°C from the current maximum. In this framework, we find that 17 out of the 69 observations are δ-records (coloured in red), and 7 of them are records (with circle). The least squares line fitted to the data (in dotted red), reveals a linear increase of the maximum temperatures over time.

The simple linear model for the temperature takes the form

$$T_t = \beta_0 + \beta_1 t + \varepsilon_t, \tag{17}$$

where T_t is the temperature of year t, and ε_t the error term. The results of the least-square estimators of the coefficients and their p-values (assuming Gaussian errors) are shown in table 1. In addition, we find an adjusted-R^2 of 0.2769. The hypothesis $\beta_1 = 0$ is clearly rejected, using the student t-test. Moreover, the estimate of β_1, which represents

J. Stat. Mech. (2020) 103201

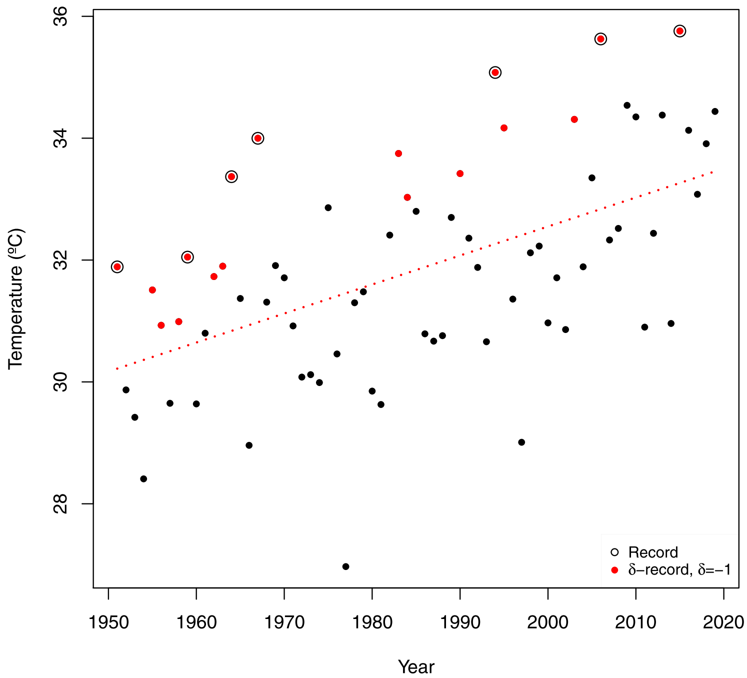

Figure 5. Monthly mean of maximum temperature in July, 1951–2019 in Saragossa (Spain).

Table 1. Regression analysis estimations for the temperature data.

Coefficient	Estimate	Std. error	p-value
β_0	−62.659	18.172	0.000 98
β_1	0.0476	0.009 15	2.04×10^{-6}

the average increment of mean maximum temperatures by year, agrees well with previous estimates of the summer warming trend in Europe, see [4, 30].

In figure 6 we show the classical diagnosis plots for linear regression. The top left panel indicates that a linear model is appropriate since no pattern in the residuals is observed. On the top right panel, the quantile–quantile plot of the residuals against the normal distribution, with all the values within the confidence lines, shows that the Gaussian assumption is adequate for errors; this is corroborated by a p-value of 0.58 in the Shapiro–Wilk test (section 5.2.2 in [53]) for normality of the standardized residuals. Moreover, the bottom panels show no significant autocorrelation or partial autocorrelation values, indicating the absence of serial correlation of the observations; this is confirmed by a p-value greater than 0.1 in the Kwiatkowski–Phillips–Schmidt–Shin test [54] for stationarity of a series around a deterministic trend. We conclude that the data are well fit by a linear regression in t, with Gaussian errors. Hence, an LDM with drift parameter $c = \hat{\beta}_1 = 0.047$ and $(X_n)_{n\geqslant 1}$ independent normally distributed random variables, is adequate for the data. Note that, for applying theorem 5, there is no need to assume any specific form of the distribution of the X_n.

J. Stat. Mech. (2020) 103201

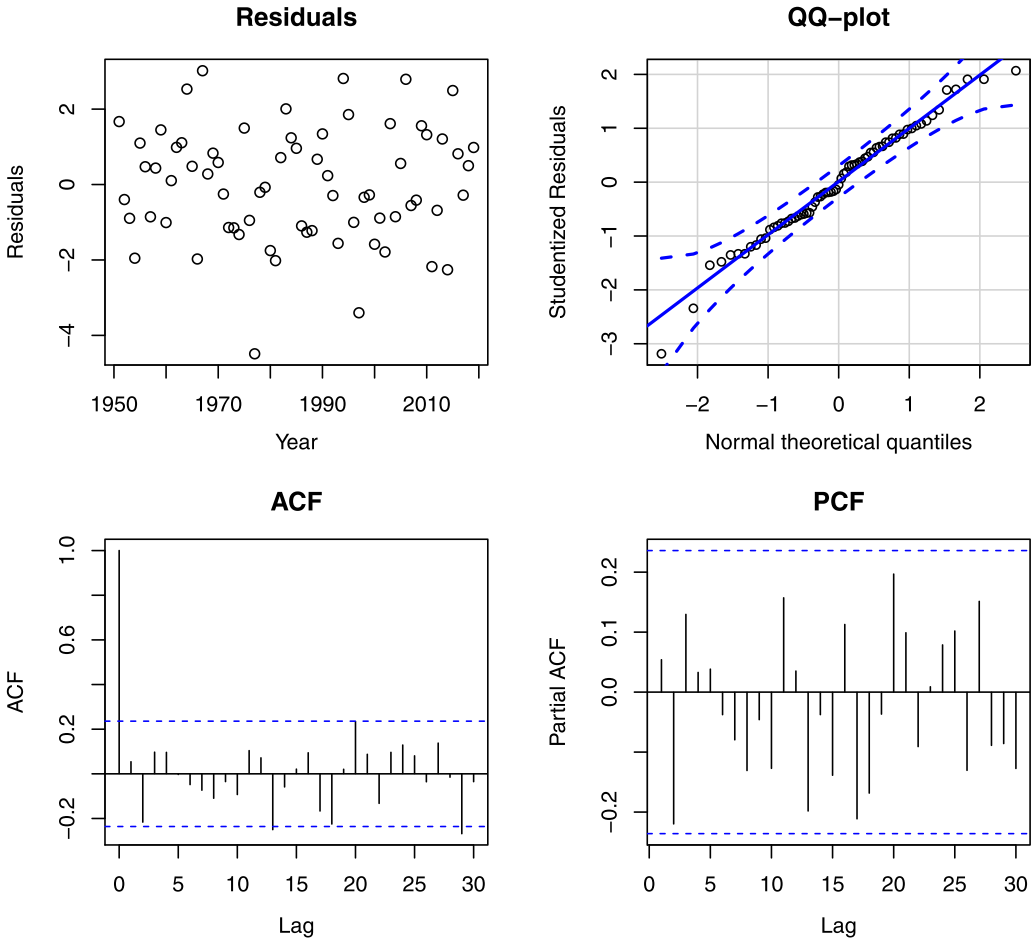

Figure 6. Diagnostic plots of the regression model. Top left: residuals vs year. Top right: quantile–quantile plot of the residual with the normal distribution. Bottom left: autocorrelation function. Bottom right: partial autocorrelation function.

Now, since 17 out of 69 observations were identified as δ-records, it is natural to estimate p_δ by the empirical record rate, that is,

$$\hat{p}_\delta = n^{-1} N_{n,\delta} = 17/69 \approx 0.2464.$$

Figure 7 illustrates how the empirical δ-record rate evolves with each extra observation and how it seems to stabilize around a constant value, as predicted by theorem 5(a).

Concerning the asymptotic normality (theorem 5(b)), we need to estimate the variance σ_δ^2, defined in (22). To that end we propose the estimator

$$\tilde{\sigma}_\delta^2 = \tilde{\gamma}_{n,\delta}(0) + 2\sum_{k=1}^{m} \tilde{\gamma}_{n,\delta}(k), \tag{18}$$

J. Stat. Mech. (2020) 103201

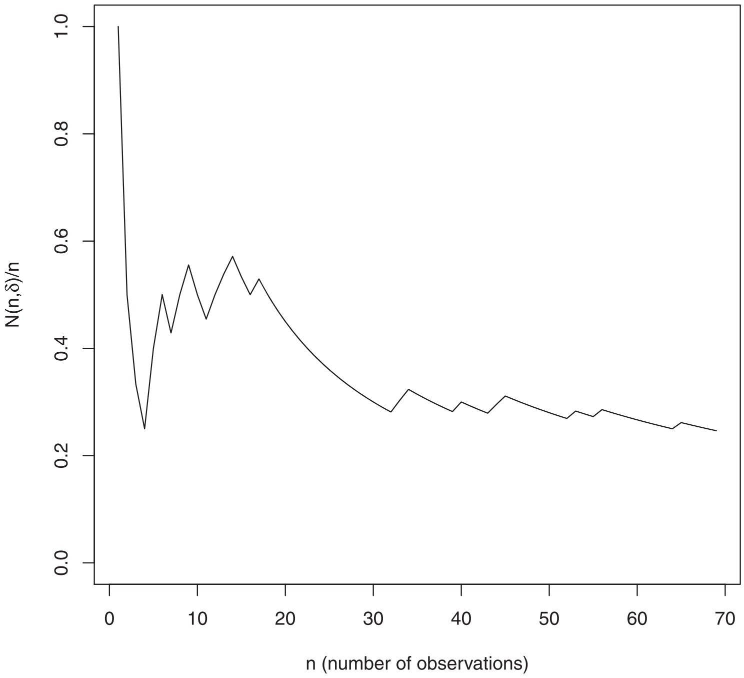

Figure 7. Evolution of the δ-record rate for the temperature data.

where m is a given natural number and

$$\tilde{\gamma}_{n,\delta}(k) = n^{-1}\sum_{j=1}^{n-k}(1_{j,\delta} - n^{-1}N_{n,\delta})(1_{j+k,\delta} - n^{-1}N_{n,\delta}),$$

The estimator in (18) is a version of an estimator proposed in [26], adapted here to deal with δ-record data. By slightly changing the proof in [26], we can prove convergence of $\tilde{\sigma}_\delta^2$ to σ_δ^2, as $n \to \infty$ (consistency), under the condition $m(n) = O(n^{1/2})$.

In order to apply formula (18), we must choose m, of order $\sqrt{n}$. In our case, $n = 69$ so we take $m = 8$, to obtain the estimate $\tilde{\sigma}_\delta^2 = 0.337$. Similar values were computed with $m = 6, 7$. Therefore, from theorem 5(b), $N_{n,\delta}$ is approximately Gaussian, with mean 17 and variance 23.25 (0.337×69).

For assessing the goodness of fit, we simulate the adjusted model (17) 10^6 times, and compute the value of $N_{69,\delta}$. Figure 8 summarizes the total number of δ-records obtained at each of the 10^6 simulations. The histogram has a Gaussian shape, so the convergence in theorem 5(b) to the Gaussian distribution seems to be fast. Moreover, the 0.025 and 0.975 quantiles of the normal distribution $N(17, 23.25)$ are, respectively, 7.54 and 26.45. The 0.025 and 0.975 empirical quantiles from the simulated data are 8 and 26, showing an excellent fit to the theoretical (asymptotic) distribution.

As a conclusion, we see that empirical results and theory are in very close agreement. This means that, even with a small sample, the approximations in theorem 5 are good, at least for the model considered.

J. Stat. Mech. (2020) 103201

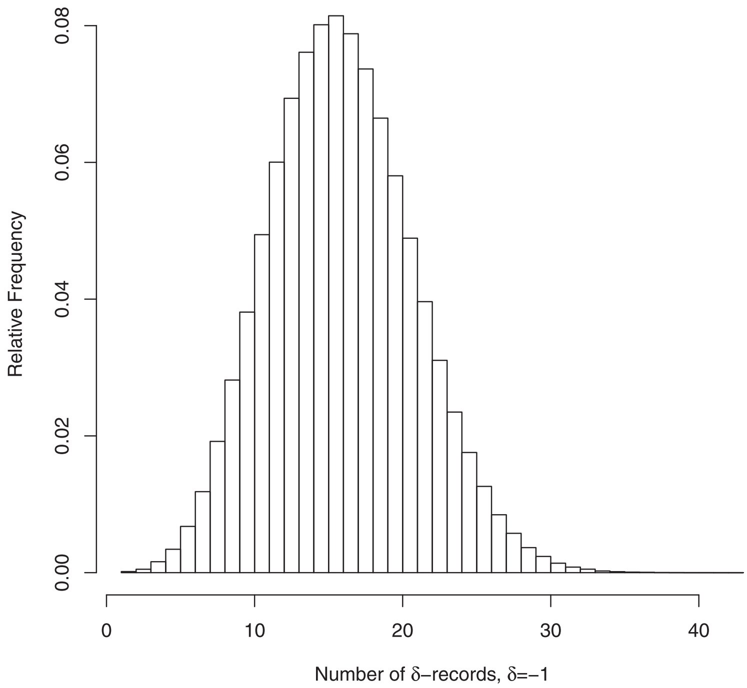

Figure 8. Histogram of the total number of δ-records for the adjusted regression model (10^6 iterations of 69 observations).

8. Concluding remarks

In this paper we have studied the behaviour of δ-records in the LDM. We have analysed the asymptotic probability of δ-records, the dependence between δ-record events and the limiting distribution of the number of δ-records among the first n observations.

The behaviour of the asymptotic probability of δ-records shows similarities with the case of records ($\delta = 0$); for instance, for positive c, $p_\delta(c) > 0$ if and only if $\mu^+ < \infty$, regardless the value of δ (except for the trivial case $\delta \geqslant x_+ - x_- + c$, where no δ-records are observed). We also find that $p_\delta(c)$ is a continuous function of δ for every c, while, as a function of c, it is continuous for every $c \neq 0$, and a discontinuity arises at $c = 0$, if $x_+ < \infty$ and $\delta < 0$. This differs from records where $p_0(c)$ is a continuous function of c.

We have described in detail the probability of δ-record in two examples. For the Gumbel distribution, an explicit expression for $p_\delta(c)$ is found, showing that it decreases with δ, as a logistic function of $-\delta$. For the cases studied in the Dagum family of distributions, we have $p_\delta(c) = 0$, for every δ, c, since $\mu^+ = \infty$. For this family, we investigate if the speed of convergence of $p_{n,\delta}(c)$ to 0, as $n \to \infty$, depends on δ or not. Since random variables X_n, with $\mu^+ = \infty$, may produce large values, which provoke abrupt changes in record values, we can expect that δ values close to 0 have negligible impact and so, $p_{n,\delta}(c)/p_{n,0}(c) \to 1$. This happens in the case $c = 0$, where the number of δ-records grows at the same speed as the number of records, when the X_n

J. Stat. Mech. (2020) 103201

are heavy-tailed. However, we find that, for some distributions in the Dagum family, $p_{n,\delta}(c)/p_{n,0}(c) \to a \neq 1$.

Parameter δ has a clear impact in the qualitative behaviour of correlations of δ-record events. First, the expression of the limiting correlation is different for $\delta \geqslant 0$ and $\delta < 0$. For the Gumbel distribution, where record indicators are independent [27], dependence appears when $\delta \neq 0$; in fact, δ-records in this distribution attract each other for $\delta < 0$ and repel each other, for $\delta > 0$. For distributions with power law tails, it is known, for $c > 0$, that correlations between records are positive and increase with n; see [31]. We have studied the Pareto distribution with $c = 1$, and obtained that, while the correlations are positive (and increasing in n) for negative, zero and small positive values of δ, they are negative for big values of δ. In fact, for each n, the limiting correlation index, as $\delta \to \infty$, is 0.3069.

Another interesting finding of the paper is about the behaviour of the random variable $N_{n,\delta}(c)$. We completely solve the question of finiteness of $N_{\infty,\delta}(c)$, that is, if there is a finite number of δ-records along the infinite sequence of observations. We show that this cannot happen for $c > 0$, for any δ (except if the condition $x_+ - x_- < \delta - c$ holds). It cannot happen either when $c < 0$ and the underlying random variables X_n have an infinite right-tail mean. This last fact solves a problem posed in [29], where the authors conjectured that, in the presence of a negative trend, the expected number of records in the whole sequence is finite.

In the case $c > 0$ we analyse the asymptotic behaviour of the random variable $N_{n,\delta}$, which grows to infinity. We give a law of large numbers, showing that the ratio $N_{n,\delta}/n$ converges to $p_\delta(c)$ and that its asymptotic distribution is Gaussian, finding the explicit expression of its normalizing constants, which can be estimated from observed data. This result was already known for records and has been applied to different problems, such as athletic records [25, 26] and climate change [4, 30]. We have illustrated the limiting result for $N_{n,\delta}$ with a set of real data of temperatures of the city of Saragossa (Spain), showing a good agreement between the theoretical asymptotic results and the observed data in the example. In fact, even for this relatively short series (69 data), the distribution of the number of δ-records is close to the theoretical limiting Gaussian distribution.

Our results open the door to the use of δ-records for statistical applications in the LDM. It has been shown that δ-records perform better than records in statistical inference, using trend-free data [44–46], so we expect that their use in the LDM is also advantageous.

Acknowledgments

We are thankful to professors J Abaurrea, A C Cebrián and J Asín, from the University of Zaragoza (Spain), for kindly providing us with the temperature data used in this paper. We also thank the referees for valuable comments which helped to improve the manuscript. This research was funded by Project PIA AFB-170001, Fondecyt Grant 1161319 and Project MTM2017-83812-P of MICINN. The authors are members of the research group Modelos Estocásticos of DGA. M Lafuente acknowledges the support by the FPU-1505266 Grant, funded by MECD.

Appendix A

A.1. Continuity of $p_\delta(c)$

Proposition 6. *$\prod_{i=1}^{\infty} F(x+ci-\delta)$, as a function of c is continuous at $c \in \mathbb{R}\setminus\{0\}$, for every $x \in (x_-, x_+)$, $x \neq x_- + \delta - c$.*

Proof. Let $(c_n)_{n\geqslant 1}$ be a real sequence converging to $c>0$. We show that

$$\prod_{i=1}^{\infty} F(x+c_n i-\delta) \to \prod_{i=1}^{\infty} F(x+ci-\delta), \tag{19}$$

as $n\to\infty$, for fixed $x\in(x_-,x_+)$, $x\neq x_-+\delta-c$.

Let $x\in(x_-,x_+)$ be such that $x<x_-+\delta-c$ (this can only happen if $x_->-\infty$ and $\delta-c>0$). In this case $F(x+c-\delta)=0$, so the right-hand side (rhs) of (19) is 0. Also, since $c_n\to c$, $F(x+c_n-\delta)=0$, for n large enough, the left-hand side (lhs) of (19) is also 0 and (19) is proved.

Let now $x>x_-+\delta-c$, then $F(x+ci-\delta)>0$ for all $i\geqslant 1$. Let $\epsilon>0$ such that $x+c-\epsilon-\delta>x_-$ and let $n_0\geqslant 1$, such that $|c_n-c|<\epsilon$, for all $n\geqslant n_0$. We have, for $n\geqslant n_0$,

$$-\log\ F(x+c_n i-\delta)\leqslant -\log\ F(x+(c-\epsilon)i-\delta).$$

Since $x>x_-+\delta-(c-\epsilon)$ and $\mu^+<\infty$, we have $-\sum_{i=1}^{\infty}\log\ F(x+(c-\epsilon)i-\delta)<\infty$, so the dominated convergence theorem yields

$$\sum_{i=1}^{\infty}\log\ F(x+c_n i-\delta)\to\sum_{i=1}^{\infty}\log\ F(x+ci-\delta),$$

as $n\to\infty$, so (19) also holds for $x>x_-+\delta-c$. Finally, for $c<0$, we have $\prod_{j=1}^{\infty}F(x+cj-\delta)=0,\quad \forall\ x\in\mathbb{R}$, since $F(x+cj-\delta)\to 0$, as $j\to\infty$. □

A.2. Finiteness of the number of δ-records

Proposition 7. *Let $c=0$ and $\delta>0$. The following conditions are equivalent:*

(a) $N_{\infty,\delta}<\infty$,

(b) $\mathbb{E}[N_{\infty,\delta}]<\infty$,

(c)

$$\int_0^{\infty}\frac{1-F(x+\delta)}{(1-F(x))^2}f(x)\mathrm{d}x<\infty.$$

Proof. It is clear that Y_n is a δ-record if and only if $\mathrm{e}^{X_n}>\mathrm{e}^{\delta}\ \max\{\mathrm{e}^{X_1},\ldots,\mathrm{e}^{X_{n-1}}\}$. That is, if the nth observation in the sequence $(\mathrm{e}^{X_n})_{n\geqslant 1}$ is a geometric record, with parameter $k=\mathrm{e}^{\delta}$, according to [52]. In section 2.1.1 of that paper, it is shown that the total number

J. Stat. Mech. (2020) 103201

of geometric records, in a sequence of i.i.d. random variables, with cdf G, is finite if and only if

$$\int_1^\infty \frac{1-G(kx)}{(1-G(x))^2}\,\mathrm{d}G(x) < \infty. \tag{20}$$

Moreover, in section 2.3.4 of that paper, it is shown that (20) is equivalent to the finiteness of the expectation of the total number of geometric records. Since $G(x) = F(\log(x))$, the result is proved. □

In the rest of the appendix, we use the operator $\bigvee$ to denote the maximum. Then, for instance, $\bigvee_{i=1}^n Y_i = \max\{Y_1, \ldots, Y_n\}$.

Lemma 8.

(a) If $c < 0$, $x_- > -\infty$ and $\mu^+ < \infty$, then $\mathbb{E}[N_{\infty,\delta}] < \infty$, $\quad \forall\, \delta \in \mathbb{R}$.

(b) Let $\tilde{X}_1$ be a random variable with cdf G, and $(\tilde{X}_n)_{n\geqslant 2}$ an i.i.d. sequence, independent of $\tilde{X}_1$, with common cdf F, such that $G(x) \leqslant F(x), \forall x$. Let $\tilde{Y}_n = \tilde{X}_n + cn$, $n \geqslant 1$. Then, if $c < 0$, $\mathbb{E}[\sum_{j=1}^\infty 1_{\{\tilde{Y}_j > \vee_{i=1}^{j-1}\tilde{Y}_i + \delta\}}] \leqslant \mathbb{E}[N_{\infty,\delta}]$.

Proof.

(a) First we bound $p_{n,\delta}(c)$ as follows

$$\begin{aligned} p_{n,\delta}(c) &= \int_{-\infty}^\infty \prod_{j=1}^{n-1} F(x+cj-\delta) f(x)\mathrm{d}x \\ &= \int_{-\infty}^\infty \prod_{j=1}^{n-1} F(x+cj-\delta) 1_{\{x+c(n-1)-\delta > x_-\}} f(x)\mathrm{d}x \\ &= \int_{x_- - c(n-1)+\delta}^\infty \prod_{j=1}^{n-1} F(x+cj-\delta) f(x)\mathrm{d}x \\ &\leqslant 1 - F(x_- - c(n-1)+\delta). \end{aligned}$$

So, $\sum_{j=1}^n p_{n,\delta} \leqslant \sum_{j=1}^n (1 - F(x_- - c(j-1)+\delta))$ yielding

$$\mathbb{E}[N_{\infty,\delta}] \leqslant \sum_{j=1}^\infty (1 - F(x_- - c(j-1)+\delta)) < \infty,$$

since $\mu^+ < \infty$.

(b) It suffices to check that the δ-record probability for the $\tilde{Y}_n$ fulfils

$$\begin{aligned} \mathbb{E}[1_{\{\tilde{Y}_j > \vee_{i=1}^{j-1}\tilde{Y}_i + \delta\}}] &= \int_{-\infty}^\infty G(x+c-\delta) \prod_{i=2}^{j-1} F(x+ci-\delta) f(x)\mathrm{d}x \\ &\leqslant \int_{-\infty}^\infty \prod_{i=1}^{j-1} F(x+ci-\delta) f(x)\mathrm{d}x = p_{j,\delta}(c). \end{aligned}$$

□

J. Stat. Mech. (2020) 103201

Proposition 9. *If $c < 0$ and $\mu^+ < \infty$, then* $\mathbb{E}[N_{\infty,\delta}] < \infty$

Proof. It suffices to consider $\delta < 0$, since the number of δ-records is decreasing with δ. Also, we take $x_- = -\infty$ as, otherwise, the result follows from lemma 8(a). Moreover, since there exists $c_1 \in \mathbb{R}$ such that $\mathbb{P}(X_n + c_1 > 0) > 0$, and the number of δ-records is the same for the sequences $Y_n = X_n + cn$ and $\tilde{Y}_n = X_n + cn + c_1$, we assume without loss of generality that $\mathbb{P}(X_n > -\delta) > 0$.

Let $N = \inf\{n \in \mathbb{N} | X_n > -\delta\}$, then N is a geometric random variable and

$$N_{\infty,\delta} = \sum_{j=1}^{N} 1_{j,\delta} + \sum_{j=N+1}^{\infty} 1_{j,\delta} = \sum_{j=1}^{N} 1_{j,\delta} + \sum_{j=N+1}^{\infty} 1_{j,\delta} 1_{\{X_j>0\}}.$$

For $j > N$, let $\tilde{1}_{j,\delta} = 1_{\{X_j > \vee_{i=N}^{j-1}(X_i + c(i-j)+\delta)\}} 1_{\{X_j>0\}}$, then

$$1_{j,\delta} 1_{\{X_j>0\}} = 1_{\{X_j > \vee_{i=1}^{j-1}(X_i + c(i-j)+\delta)\}} 1_{\{X_j>0\}} \leqslant \tilde{1}_{j,\delta}.$$

Note that the $\tilde{1}_{j,\delta}$, defined for $j > N$, are the δ-record indicators of the sequence $\{X_N, X_{N+1} 1_{\{X_{N+1}>0\}} + c, X_{N+2} 1_{\{X_{N+2}>0\}} + 2c, \ldots\}$. Now, taking expectations we have

$$\mathbb{E}[N_{\infty,\delta}] \leqslant \frac{1}{\mathbb{P}(X_1 > -\delta)} + \sum_{i=1}^{\infty} \mathbb{E}[\tilde{1}_{i,\delta}] < \infty,$$

since the last sum is bounded by lemma 8(b). □

A.3. Proof of law of large numbers for $N_{n,\delta}(c)$

In order to work with a stationary process, we consider a bilateral version of the LDM defined as in (1), but letting $n \in \mathbb{Z}$ instead of $n \in \mathbb{N}$. Associated to this model, we define, for $n \in \mathbb{Z}$,

$$M_n^* = \max\{Y_i : \ i \leqslant n\}, \qquad 1_{n,\delta}^* = 1_{\{Y_n > M_{n-1}^* + \delta\}}, \tag{21}$$

and, for $n \in \mathbb{N}$,

$$N_{n,\delta}^* = \sum_{k=1}^{n} 1_{k,\delta}^*.$$

Theorem 5(a). *Let $c > 0$, $\mu^+ < \infty$. Then $N_{n,\delta}(c)/n \to p_\delta(c)$ a.s. as $n \to \infty$.*

Proof. It is clear that

$$\lim_{n\to\infty} \mathbb{P}[Y_n > a] = \lim_{n\to\infty} \mathbb{P}[X_n > a - cn] = 1, \quad \forall\, a \in \mathbb{R},$$

thus $Y_n \to \infty$ and $M_n \to \infty$ a.s. Also, since $\mu^+ < \infty$, it is known by a Borel–Cantelli argument that $M_0^* < \infty$ a.s. Gathering these facts, we know that $\exists\, 0 < N < \infty$ a.s. such

J. Stat. Mech. (2020) 103201

that $1^*_{N,0} = 1$ almost surely. From the definition of $1^*_{n,0}$, given $n \in \mathbb{N}$ we have $1_{n,0} \geqslant 1^*_{n,0}$, and so $1_{N,0} = 1$ a.s., entailing $M^*_n = M_n$ and $1_{n,\delta} = 1^*_{n,\delta}$ a.s. $\forall\, n > N$. So,

$$\sum_{k=N+1}^{\infty} 1_{k,\delta} = \sum_{k=N+1}^{\infty} 1^*_{k,\delta} \text{ a.s.}$$

Also, we know that $1^*_{n,\delta}$ is a strictly stationary and ergodic sequence. Applying Birkhoff's ergodic theorem (page 129 in [55]) we have

$$\frac{N^*_{n,\delta}}{n} = \frac{1}{n}\sum_{k=1}^{n} 1^*_{k,\delta} \to \mathbb{E}[1^*_{0,\delta}] \text{ a.s.}$$

Now, let $(a_n)_{n\geqslant 1}$ be a real sequence diverging to ∞. Then

$$\left|\frac{N_{n,\delta} - N^*_{n,\delta}}{a_n}\right| \leqslant \left|\frac{N}{a_n}\right| \to 0 \text{ a.s.}$$

since N does not depend on n. Finally, since $\left|\frac{N_{n,\delta} - N^*_{n,\delta}}{n}\right| \to 0$ a.s. and $\frac{N^*_{n,\delta}}{n} \to \mathbb{E}[1^*_{0,\delta}]$ a.s., we have $\frac{N_{n,\delta}}{n} \to \mathbb{E}[1^*_{0,\delta}]$ a.s. Finally, $\mathbb{E}[1^*_{0,\delta}]$ can be written as the rhs in (2), yielding $\mathbb{E}[1^*_{0,\delta}] = p_\delta(c)$. □

A.4. Proof of central limit theorem for $N_{n,\delta}(c)$.

A proof of Gaussian convergence for the number of δ-records, based on the ideas in [25], is not straightforward. The main problem arises when considering the joint probability of two observations being δ-records. While in the case of records this quantity can be explicitly written as follows

$$\mathbb{E}[1_{i,0}1_{i+m,0}] = \int_{-\infty}^{\infty} \prod_{k=1}^{i-1} F(y+ck) \int_{y-cm}^{\infty} \prod_{j=1}^{m-1} F(s+cj) f(s)\mathrm{d}s \; f(y)\mathrm{d}y,$$

in the setting $\delta \neq 0$ there is no such analytical expression. In order to solve this problem we introduce the following general bounds, which do not depend on the specification of the model for the sequence $(Y_n)_{n\geqslant 1}$.

Proposition 10. *Let $(Y_k)_{k\in\mathbb{Z}}$ be a sequence of random variables and consider the events $A = \left\{\bigvee_{k=-\infty} Y_k + \delta < Y_i\right\}$, $\mathrm{B} = \left\{\bigvee_{k=i+1}^{i+m-1} Y_k + \delta < Y_{i+m}\right\}$, $C = \{Y_i - \delta < Y_{i+m}\}$ and $E = \{Y_i + \delta < Y_{i+m}\}$. Then, if $\delta \leqslant 0$,*

(a1) $\mathbb{P}[A\cap B\cap C] \leqslant \mathbb{E}[1^*_{i,\delta}1^*_{i+m,\delta}]$ *and*

(a2) $\mathbb{P}[A\cap B\cap E] \geqslant \mathbb{E}[1^*_{i,\delta}1^*_{i+m,\delta}]$.

Also, if $\delta \geqslant 0$,

(b) $\mathbb{P}[A\cap B\cap E] = \mathbb{E}[1^*_{i,\delta}1^*_{i+m,\delta}]$.

J. Stat. Mech. (2020) 103201

Proof.

(a1) Note that $1^*_{j,\delta}$ is the indicator of $D_j = \left\{\bigvee_{k=-\infty}^{j-1} Y_k + \delta < Y_j\right\}$, $j = i, i+m$. Then we must show that $A \cap B \cap C \subseteq D_i \cap D_{i+m}$.

First, it is clear that $A = D_i$. Also, observe that $C \subseteq E$ and that $A \cap C \subseteq \left\{\bigvee_{k=-\infty}^{i-1} Y_k + \delta < Y_{i+m}\right\}$, since $\delta \leqslant 0$. From the inclusions above we have

$$A \cap B \cap C \subseteq \left\{\bigvee_{k=-\infty}^{i-1} Y_k + \delta < Y_{i+m}\right\} \cap E \cap B = D_{i+m}$$

and the conclusion follows.

(a2) Trivial.

(b) It is clear that $D_i \cap D_{i+m} \subseteq A \cap B \cap E$ and that $A \cap B \cap E \subseteq D_i$, because $A = D_i$. Also, since $\delta \geqslant 0$, we have $A \cap E \subseteq \left\{\bigvee_{k=-\infty}^{i-1} Y_k + \delta < Y_{i+m}\right\}$, so

$$A \cap B \cap E \subseteq \left\{\bigvee_{k=-\infty}^{i-1} Y_k + \delta < Y_{i+m}\right\} \cap E \cap B = D_{i+m},$$

which completes the proof. □

Note that, although it is unnecessary in our setting, the reverse (a1) inequality also holds for $\delta \geqslant 0$. Under the assumptions of the LDM, the lhs of the first two bounds in the previous proposition have analytical expressions. The strategy to prove Gaussian convergence is to work with the corresponding bounds of $\mathbb{E}[1_{i,\delta}1_{i+m,\delta}]$, which are shown to be tight enough to achieve our purpose. So, with this result we slightly modify the necessary bounds and rebuild the martingale approach in [25], to prove convergence to the Gaussian distribution.

Theorem 5(b). *Suppose that $\int_0^\infty x^2 f(x)\mathrm{d}x < \infty$ and let $c > 0$, $\delta \in \mathbb{R}$, such that $p_\delta > 0$. Then, as $n \to \infty$,*

$$\sqrt{n}(n^{-1}N_{n,\delta} - p_\delta(c)) \xrightarrow{\mathcal{D}} N(0, \sigma_\delta^2(c)),$$

where

$$\sigma_\delta^2 = p_\delta - p_\delta^2 + 2\sum_{m=1}^{\infty} (\mathbb{E}[1^*_{i,\delta}1^*_{i+m,\delta}] - p_\delta). \tag{22}$$

Proof. For simplicity, we only consider the case $\delta \leqslant 0$ since the case $\delta > 0$ is analogous. We assume $-2\delta < x_+$ as, otherwise, we can define $X'_n = X_n + (-3\delta - x_+)$, $n \geqslant 1$; the number of δ-records in both models is the same and $-2\delta < x'_+$, where x'_+ is the right-end point of X'_n.

The proof is split into several steps.

(a) We claim that

$$0 \leqslant p_{n,\delta} - p_\delta \leqslant c^{-1}\int_{c(n-1)/2-\delta}^{\infty} (1 - F(s))\mathrm{d}s + F(-\delta)^{\lfloor (n-1)/2 \rfloor}. \tag{23}$$

The first inequality follows from

$$p_{n,\delta} - p_\delta = \int_{-\infty}^{\infty} \left(\prod_{j=1}^{n-1} F(y + cj - \delta) - \prod_{j=1}^{\infty} F(y + cj - \delta) \right) f(y)\mathrm{d}y \geqslant 0.$$

For the second, let $u = \prod_{j=1}^{n-1} F(y + cj - \delta)$ and $v = \prod_{j=1}^{\infty} F(y + cj - \delta)$. Then, from the elementary inequality $u - v \leqslant u - uv$, we have

$$p_{n,\delta} - p_\delta \leqslant \int_{-\infty}^{\infty} u(1 - v) f(y)\mathrm{d}y. \tag{24}$$

The integral in the rhs of (24) is split into two terms A, B, that we bound. Let $A = \int_{-\infty}^{-c(n-1)/2} u(1 - v) f(y)\mathrm{d}y$ and $B = \int_{-c(n-1)/2}^{\infty} u(1 - v) f(y)\mathrm{d}y$, then

$$\begin{aligned} A &\leqslant \int_{-\infty}^{-c(n-1)/2} \prod_{j=1}^{n-1} F(-c(n-1)/2 + cj - \delta) f(y)\mathrm{d}y \\ &\leqslant \prod_{j=1}^{n-1} F(c(j - (n-1)/2) - \delta) \\ &\leqslant \prod_{j=1}^{\lfloor (n-1)/2 \rfloor} F(c(j - (n-1)/2) - \delta) \\ &\leqslant \prod_{j=1}^{\lfloor (n-1)/2 \rfloor} F(-\delta) = F(-\delta)^{\lfloor (n-1)/2 \rfloor}. \end{aligned} \tag{25}$$

For B we have

$$\begin{aligned} B &\leqslant \int_{-c(n-1)/2}^{\infty} \left(1 - \prod_{j=n}^{\infty} F(y + cj - \delta) \right) f(y)\mathrm{d}y \\ &\leqslant \int_{-c(n-1)/2}^{\infty} \sum_{j=n}^{\infty} (1 - F(y + cj - \delta)) f(y)\mathrm{d}y \\ &\leqslant \int_{-c(n-1)/2}^{\infty} \left(\int_{z=n-1}^{\infty} (1 - F(y + cz - \delta))\mathrm{d}z \right) f(y)\mathrm{d}y \\ &\leqslant \int_{-c(n-1)/2}^{\infty} \left(c^{-1} \int_{-c(n-1)/2 + c(n-1) - \delta}^{\infty} (1 - F(s))\mathrm{d}s \right) f(y)\mathrm{d}y \\ &\leqslant c^{-1} \int_{c(n-1)/2 - \delta}^{\infty} (1 - F(s))\mathrm{d}s. \end{aligned} \tag{26}$$

So, from (25) and (26), (23) holds.

J. Stat. Mech. (2020) 103201

(b) Let $r_{m,\delta} = \mathbb{E}[1^*_{i,\delta}1^*_{i+m,\delta}]$, which is well defined since it does not depend on i. We bound $r_{m,\delta}$ by applying proposition 10 as follows:

$$\begin{aligned} r_{m,\delta} &= \mathbb{P}\left[Y_i, Y_{i+m} \, \text{are} \, \delta - \text{records}\right] \\ &= \mathbb{P}\left[Y_i > \bigvee_{l<i} Y_l + \delta, Y_{i+m} > \bigvee_{l<i+m} Y_l + \delta\right] \\ &\leqslant \mathbb{P}\left[Y_i > \bigvee_{l<i} Y_l + \delta, Y_{i+m} > \bigvee_{l=1}^{m-1} Y_{i+l} + \delta, Y_{i+m} > Y_i + \delta\right] \\ &= \iint_{y<s+cm-\delta} \prod_{j=1}^{\infty} F(y+cj-\delta) \prod_{i=1}^{m-1} F(s+ci-\delta) f(s)\mathrm{d}s \; f(y)\mathrm{d}y. \end{aligned}$$

If $r_{m,\delta} \geqslant p_\delta^2$, we apply the Fubini–Tonelli theorem, as well as the triangle inequality, to obtain

$$\begin{aligned} |r_{m,\delta} - p_\delta^2| &\leqslant \left| \iint_{y<s+cm-\delta} \prod_{j=1}^{\infty} F(y+cj-\delta) \prod_{i=1}^{m-1} F(s+ci-\delta) f(s)\mathrm{d}s \; f(y)\mathrm{d}y - p_\delta^2 \right| \\ &\leqslant A + B, \end{aligned}$$

where

$$A = \int_{-\infty}^{\infty} \prod_{j=1}^{\infty} F(y+cj-\delta) \left| \int_{-\infty}^{\infty} \prod_{i=1}^{m-1} F(s+ci-\delta) f(s)\mathrm{d}s - p_\delta \right| f(y)\mathrm{d}y$$

and

$$B = \int_{-\infty}^{\infty} \prod_{j=1}^{\infty} F(y+cj-\delta) \int_{-\infty}^{y-cm+\delta} \prod_{i=1}^{m-1} F(s+ci-\delta) f(s)\mathrm{d}s \; f(y)\mathrm{d}y.$$

Since variables are separated in A and applying the first step of this proof

$$\begin{aligned} A &\leqslant \int_{-\infty}^{\infty} \prod_{j=1}^{m-1} F(s+cj-\delta) f(s)\mathrm{d}s - p_\delta \\ &\leqslant c^{-1} \int_{c(n-1)/2-\delta}^{\infty} (1-F(s))\mathrm{d}s + F(-2\delta))^{\lfloor (m-1)/2 \rfloor}. \end{aligned} \tag{27}$$

While for B we have

$$\begin{aligned} B = &\int_{-\infty}^{cm/2} \prod_{j=1}^{\infty} F(y+cj-\delta) \int_{-\infty}^{y-cm+\delta} \prod_{i=1}^{m-1} F(s+ci-\delta) f(s)\mathrm{d}s \; f(y)\mathrm{d}y \\ &+ \int_{cm/2}^{\infty} \prod_{j=1}^{\infty} F(y+cj-\delta) \int_{-\infty}^{y-cm+\delta} \prod_{i=1}^{m-1} F(s+ci-\delta) f(s)\mathrm{d}s \; f(y)\mathrm{d}y \end{aligned}$$

J. Stat. Mech. (2020) 103201

$$\leqslant \int_{-\infty}^{-cm/2+\delta} \prod_{j=1}^{m-1} F(s+cj-\delta)f(s)\mathrm{d}s \int_{-\infty}^{cm/2} \prod_{j=1}^{\infty} F(y+cj-\delta)f(y)\mathrm{d}y$$
$$+ \int_{cm/2}^{\infty} \prod_{j=1}^{\infty} F(y+cj-\delta)f(y)\mathrm{d}y$$
$$\leqslant \prod_{j=1}^{m-1} F(-cm/2+cj) + 1 - F(cm/2)$$
$$\leqslant F(-2\delta)^{\lfloor (m-1)/2 \rfloor} + 1 - F(cm/2). \tag{28}$$

Analogously, applying the corresponding bound in proposition 10, we arrive at the same conclusion if $r_{m,\delta} \leqslant p_\delta$ via (27) and (28), so

$$|r_{m,\delta} - p_\delta| \leqslant c^{-1} \int_{c(m-1/2-\delta)}^{\infty} (1-F(s))\mathrm{d}s + 2F(-2\delta)^{\lfloor (m-1)/2 \rfloor} + 1 - F(cm/2). \tag{29}$$

(c) Since $\int_0^\infty x^2 f(x)\mathrm{d}x < \infty$, it is easy to check, from (29), that the series $\sum_{m=1}^\infty |r_{m,\delta} - p_\delta|$ converges; for $F(-2\delta)^{\lfloor (m-1)/2 \rfloor}$ convergence holds since $F(-2\delta) < 1$.

(d) Using the strategy in the proof of theorem 5(a), we get the following convergence in distribution

$$\sqrt{n}(n^{-1}N_{n,\delta} - n^{-1}N_{n,\delta}^*) \xrightarrow{\mathcal{D}} 0. \tag{30}$$

(e) Theorem 5.2 in [56] is applied to the $N_{n,\delta}^*$ in order to transfer the asymptotic normality to $N_{n,\delta}$, as a consequence of (30). This martingale result guarantees convergence to the Gaussian distribution, if the next two conditions hold:

(1) $\sum_{k=1}^\infty \mathbb{E}[\xi_{k,\delta}\mathbb{E}[\xi_{l,\delta}|\mathcal{M}_0]]$ converges $\forall l \geqslant 0$.

(2) $\lim_{l\to\infty} \sum_{k=K}^\infty \mathbb{E}[\xi_{k,\delta}\mathbb{E}[\xi_{l,\delta}|\mathcal{M}_0]] = 0$ uniformly in $K \geqslant 1$,

where $\xi_{k,\delta} = 1_{k,\delta}^* - p_\delta$ and $\mathcal{M}_0$ is a certain sub-σ-algebra of events of the original probability space (see [56], page 128, for details). Moreover we have

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left[\left(\sum_{i=1}^n \xi_{i,\delta}\right)^2\right] = \sigma_\delta^2.$$

Given that the hypothesis $\delta \neq 0$ does not imply any extra difficulty in the application of this theorem, we omit the verification of these two conditions since the adaptation of this part of the proof is straightforward following the lines of [25]. □

A.5. Correlations in the Pareto distribution

The δ-record probability is given in (12). For $\mathbb{E}[1_{n,\delta}1_{n+1,\delta}]$ and $n > 2$, we use (14) for $\delta < 0$ and (15) for $\delta \geqslant 0$.

J. Stat. Mech. (2020) 103201

(a) Let $a = n - \delta$, $A = (\delta - 2)(\delta(1 - a) + (n - 1)\log a)(n\log(a + 1) - \delta a)$,

$$B = -\left(\delta^3(n-2) + \delta - 2n^3 - 2\delta^2(n^2-2) + \delta(n-1)(n+5)n + n + 1\right)\log(a+1),$$

$$C = (a-1)\log(a+1-\delta) - (\delta-2)a\left(\delta(a-1)^2 - (n-1)a\ \log(4a)\right) + (1-a)\log((a-\delta+1)(a+1)).$$

Then, if $\delta < 0$,

$$l_n(1,\delta) = \frac{B+C}{A}.$$

(b) Let $a = n - \delta$, $A = (\delta - 1)^2(\delta - a)(\delta(1 - a) + (n - 1)\log a)(-\delta a + n\log(a + 1))$,

$$B = (a-\delta)(\delta-1)\left(\delta^2(a-1) + (\delta-1)(n-1)\log\left(\frac{a-\delta+1}{(2-\delta)a}\right)\right),$$

$$C = -\log(2-\delta)(\delta(\delta+2) - 2\delta n + n - 1)(a-\delta) + (\delta-1)^2(n-1)\log(a-\delta+1).$$

Then, if $0 < \delta < 1$,

$$l_n(1,\delta) = \frac{a^2(B+C)}{A}.$$

(c) If $\delta = 1$,

$$l_n(1,1) = \frac{(n-1)^2((n-2)n - 2(n-1)\log(n-1))}{2(n-2)(-n+(n-1)\log(n-1)+2)(-n+n\ \log(n)+1)}.$$

(d) Finally, let $a = n - \delta$, $A_1 = (\delta + \log\delta - n\log\delta - n + (n - 1)\log(n - 1) + 1)(\delta - 1)^2(\delta - a)$,

$$A_2 = (\delta - n\ \log\ \delta - n + n\ \log\ n),$$

$$B = (\log\ \delta)\left(2\delta(\delta^2 + 2\delta - 1) + (2\delta - 1)n^2 - 5\delta^2 n + n\right),$$

$$C = (\delta-1)^2(n-1)\log(n-1) - (a-1)\left((\delta-1)(\delta-a) + (2\delta-1)\log(2\delta-1)(a-1)\right).$$

Then, if $\delta > 1$, and $\delta \notin \{n/2, n, n+1\}$ (otherwise, the values of the index are given by the continuous extension at these points),

$$l_n(1,\delta) = \frac{a^2(B+C)}{A_1 A_2}.$$

References

[1] Benestad R 2003 How often can we expect a record event? *Clim. Res.* **25** 3–13

[2] Majumdar S N, von Bomhard P and Krug J 2019 Exactly solvable record model for rainfall *Phys. Rev. Lett.* **121** 158702

[3] Redner R and Petersen M R 2006 Role of global warming on the statistics of record breaking temperatures *Phys. Rev.* E **74** 061114

[4] Wergen G and Krug J 2010 Record-breaking temperatures reveal a warming climate *Europhys. Lett.* **92** 30008
[5] Einmahl J H J and Magnus J R 2008 Records in athletics through extreme-value theory *J. Am. Stat. Assoc.* **103** 1382–91
[6] Gembris D, Taylor J G and Suter D 2002 Trends and random fluctuations in athletics *Nature* **417** 506
[7] Gembris D, Taylor J G and Suter D 2007 Evolution of athletic records: statistical effects versus real improvements *J. Appl. Stat.* **34** 529–45
[8] Wergen G 2014 Modeling record-breaking stock prices *Phys.* A **396** 114–33
[9] Wergen G, Bogner M and Krug J 2011 Record statistics for biased random walks, with an application to financial data *Phys. Rev.* E **83** 051109
[10] Park S C and Krug J 2008 Evolution in random fitness landscapes: the infinite sites model *J. Stat. Mech.* P04014
[11] Dey S, Dey T and Luckett D J 2016 Statistical inference for the generalized inverted exponential distribution based on upper record values *Math. Comput. Simul.* **120** 64–78
[12] Gulati S and Padgett W J 2003 *Parametric and Nonparametric Inference from Record Breaking Data* (*Lecture notes in Statistics*) vol 172 (Berlin: Springer)
[13] Jafari A A and Zakerzadeh H 2015 Inference on the parameters of the Weibull distribution using records *SORT* **39** 3–18
[14] Wang B X and Ye Z-S 2015 Inference on the Weibull distribution based on record values *Comput. Stat. Data Anal.* **83** 26–36
[15] Ahsanullah M 1995 *Record Statistics* (Commack, NY: Nova Science Publishers)
[16] Arnold B C, Balakrishnan N and Nagaraja H N 1998 *Records* (New York: Wiley)
[17] Nevzorov V B 2001 *Records: Mathematical Theory* (Providence, RI: American Mathematical Society)
[18] Majumdar S N and Ziff R M 2008 Universal record statistics of random walks and Lévy flights *Phys. Rev. Lett.* **101** 050601
[19] Godrèche C and Luck J-M 2020 Records for the moving average of a time series *J. Stat. Mech.* 023201
[20] Godrèche C, Majumdar S N and Schehr G 2017 Record statistics for random walk bridges *J. Stat. Mech.* P07026
[21] Godrèche C, Majumdar S N and Schehr G 2017 Record statistics of a strongly correlated time series: random walks and Lévy flights *J. Phys. A: Math. Theor.* **50** 333001
[22] Gouet R, López F J and Sanz G 2015 Records from stationary observations subject to a random trend *Adv. Appl. Probab.* **47** 1175–89
[23] Kearney M J 2020 Record statistics for a discrete-time random walk with correlated steps *J. Stat. Mech.* 023206
[24] Mounaix P, Majumdar S N and Schehr G 2018 Asymptotics for the expected maximum of random walks and Lévy flights with a constant drift *J. Stat. Mech.* 083201
[25] Ballerini R and Resnick S 1985 Records from improving populations *J. Appl. Probab.* **22** 487–502
[26] Ballerini R and Resnick S 1987 Records in the presence of a linear trend *Adv. Appl. Probab.* **19** 883–909
[27] Borovkov K 1999 On records and related processes for sequences with trends *J. Appl. Probab.* **36** 668–81
[28] De Haan L and Verkade E 1987 On extreme-value theory in the presence of a trend *J. Appl. Probab.* **24** 62–76
[29] Franke J, Wergen G and Krug J 2010 Records and sequences of records from random variables with a linear trend *J. Stat. Mech.* P10013
[30] Wergen G, Hense A and Krug J 2014 Record occurrence and record values in daily and monthly temperatures *Clim. Dyn.* **42** 1275–89
[31] Franke J, Wergen G and Krug J 2012 Correlations of record events as a test for heavy-tailed distributions *Phys. Rev. Lett.* **108** 064101
[32] Wergen G, Franke J and Krug J 2011 Correlations between record events in sequences of random variables with a linear trend *J. Stat. Phys.* **144** 1206–22
Wergen G, Franke J and Krug J 2011 Erratum to: correlations between record events in sequences of random variables with a linear trend *J. Stat. Phys.* **145** 1405–6
[33] Le Doussal P and Wiese K J 2009 Driven particle in a random landscape: disorder correlator, avalanche distribution, and extreme value statistics of records *Phys. Rev.* E **79** 051105
[34] Majumdar S N, Schehr G and Wergen G 2012 Record statistics and persistence for a random walk with a drift *J. Phys. A: Math. Theor.* **45** 355002
[35] Balakrishnan N, Pakes A G and Stepanov A 2005 On the number and sum of near-record observations *Adv. Appl. Probab.* **37** 765–80
[36] Hashorva E 2003 On the number of near-maximum insurance claim under dependence *Insur.: Math. Econ.* **32** 37–49
[37] Pakes A G 2007 Limit theorems for numbers of near-records *Extremes* **10** 207–24
[38] Balakrishnan N, Balasubramanian K and Panchapakesan S 1996 δ-Exceedance records *J. Appl. Stat. Sci.* **4** 123–32

J. Stat. Mech. (2020) 103201

[39] Park S-C and Krug J 2016 δ-exceedance records and random adaptive walks *J. Phys. A: Math. Theor.* **49** 315601
[40] Gouet R, López F J and Sanz G 2007 Asymptotic normality for the counting process of weak records and δ-records in discrete models *Bernoulli* **13** 754–81
[41] López-Blázquez F and Salamanca-Miño B 2013 Distribution theory of δ-record values. Case δ ⩽ 0 *Test* **22** 715–38
[42] López-Blázquez F and Salamanca-Miño B 2015 Distribution theory of δ-record values: case δ ⩾ 0 *Test* **24** 558–82
[43] Gouet R, López F J and Sanz G 2015 On the point process of near-record values *Test* **24** 302–21
[44] Gouet R, López F J and Sanz G 2012 On δ-record observations: asymptotic rates for the counting process and elements of maximum likelihood estimation *Test* **21** 188–214
[45] Gouet R, López F J, Maldonado L P and Sanz G 2014 Statistical inference for the geometric distribution based on -records *Comput. Stat. Data Anal.* **78** 21–32
[46] Gouet R, López F J, Maldonado L P and Sanz G 2020 Statistical inference for the Weibull distribution based on δ-record data *Symmetry* **12** 1–24
[47] Bogachev V I 2007 *Measure Theory* vol 1 (Berlin: Springer)
[48] Kleiber C 2008 A guide to the Dagum distribution *Modeling Income Distributions and Lorenz Curves: Essays in Memory of Camilo Dagum* ed D Chotikapanich (Berlin: Springer) pp 97–117
[49] Shoukri M M, Mian I U H and Tracy D S 1988 Sampling properties of estimators of the log-logistic distribution with application to Canadian precipitation data *Can. J. Stat.* **16** 223–36
[50] Arnold B C 2014 Univariate and multivariate Pareto models *J. Stat. Distrib. Appl.* **1** 11
[51] Gradshteyn I S and Ryzhik I M 2014 *Table of Integrals, Series, and Products* (New York: Academic)
[52] Gouet R, López F J and Sanz G 2012 On geometric records: rate of appearance and magnitude *J. Stat. Mech.* P01005
[53] Sen A K and Srivastava M S 1990 *Regression Analysis* (Berlin: Springer)
[54] Kwiatkowski D, Phillips P C B, Schmidt P and Shin Y 1992 Testing the null hypothesis of stationarity against the alternative of a unit root *J. Econom.* **54** 159–78
[55] Weber M 2009 *Dynamical Systems and Processes* (*IRMA Lectures in Mathematics and Theoretical Physics*) vol 14 (Zurich: European Mathematical Society Publishing House)
[56] Hall P and Heyde C C 1980 *Martingale Limit Theory and its Applications* (New York: Academic)