

Volovskiy, Grigoriy; Kamps, Udo

Article — Published Version

Maximum product of spacings prediction of future record values

Metrika

Provided in Cooperation with:

Springer Nature

Suggested Citation: Volovskiy, Grigoriy; Kamps, Udo (2020) : Maximum product of spacings prediction of future record values, Metrika, ISSN 1435-926X, Springer, Berlin, Heidelberg, Vol. 83, Iss. 7, pp. 853-868,
<https://doi.org/10.1007/s00184-020-00767-1>

This Version is available at:

<https://hdl.handle.net/10419/288287>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Maximum product of spacings prediction of future record values

Grigoriy Volovskiy¹ · Udo Kamps¹

Received: 4 September 2019 / Published online: 6 February 2020
© The Author(s) 2020

Abstract

A spacings-based prediction method for future upper record values is proposed as an alternative to maximum likelihood prediction. For an underlying family of distributions with continuous cumulative distribution functions, the general form of the predictor as a function of the estimator of the distributional parameters is established. A connection between this method and the maximum observed likelihood prediction procedure is shown. The maximum product of spacings predictor turns out to be useful to predict the next record value in contrast to likelihood-based procedures, which provide trivial predictors in this particular case. Moreover, examples are given for the exponential and the Pareto distributions, and a real data set is analyzed.

Keywords Point prediction · Cumulative hazard rate · Spacings · Maximum observed likelihood predictor · Upper record values · Exponential distribution · Pareto distribution

Mathematics Subject Classification 62F99 · 62M20

1 Introduction

A general procedure for estimating parameters, termed maximum product of spacings estimation, was proposed independently by Cheng and Amin (1983) and Ranney (1984) as an alternative to maximum likelihood estimation in particular situations of continuous, univariate distributions. In the sequel, the estimation method was applied, extended and further theoretically studied in, e.g., Ekström (1998, 2008), Shao and Hahn (1999) and Anatolyev and Kosenok (2005). Here, we adopt the method of

✉ Grigoriy Volovskiy
volovskiy@isw.rwth-aachen.de

Udo Kamps
kamps@isw.rwth-aachen.de

¹ Institute of Statistics, RWTH Aachen University, 52056 Aachen, Germany

maximizing a product of spacings to introduce a related new prediction method based on upper record values.

Suppose that X_1, X_2, \dots is an infinite sequence of independent and identically distributed (i.i.d.) continuous random variables with cumulative distribution function (cdf) F . An observation X_j is called an (*upper*) *record value* provided it is greater than all previously observed values. More specifically, defining the record times as

$$L(1) = 1, \quad L(n+1) = \min\{j > L(n) \mid X_j > X_{L(n)}\}, \quad n \in \mathbb{N},$$

the sequence $(R_n)_{n \in \mathbb{N}} = (X_{L(n)})_{n \in \mathbb{N}}$ is referred to as the sequence of (upper) record values based on $(X_n)_{n \in \mathbb{N}}$ [see Arnold et al. (1998); Nevzorov (2001)]. The study of record values dates back to Chandler (1952) providing a natural model for the sequence of successive extremes in an i.i.d. sequence of random variables. The structure of record values also appears in the context of minimal repair of a system, and there is a close connection to occurrence times of a non-homogeneous Poisson process (NHPP); namely, under mild conditions, the epoch times of an NHPP and upper record values are equal in distribution [see Gupta and Kirmani (1988)].

We are concerned with the problem of predicting the occurrence of a future record value R_s based on the first r , $r < s$, (observed) record values $\mathbf{R} = (R_1, \dots, R_r)$. This prediction problem has been studied by several authors. Here, we focus on non-Bayesian prediction. In the one-sample case, Raqab (2007) derived the best linear unbiased predictor, the best linear equivariant predictor, the maximum likelihood predictor as well as the conditional median predictor of the s -th record value R_s from a Type-II left censored sample with a two-parameter exponential distribution. His findings supplement and generalize results of Ahsanullah (1980), Basak and Balakrishnan (2003) and Nagaraja (1986, Sect. 4). Awad and Raqab (2000) provide a comparative study of several predictors of the s -th record value R_s based on the first r observed record values from a one-parameter exponential distribution. Linear unbiased prediction of future Pareto record values is discussed in Paul and Thomas (2016), and maximum likelihood prediction of future Pareto record values is studied in Raqab (2007). Since the model of record values is contained in the generalized order statistics model [see Kamps (1995, 2016)], all results pertaining to prediction of future generalized order statistics can be specialized to solve the prediction problem for record values [see, e.g., Burkschat (2009)]. Bayesian prediction methods for future record values were first discussed by Dunsmore (1983) and have subsequently been applied to various distribution families [cf. Madi and Raqab (2004); Ahmadi and Doostparast (2006); Nadar and Kızılaslan (2015)]. It should be noted that, under exponential as well as under Pareto distributions, maximum likelihood prediction of the subsequent record value R_{r+1} becomes trivial, since the respective predictor is given by R_r , i.e., the predictor coincides with the last observed record value in the model. However, by construction, record values are strictly ordered; thus, the maximum likelihood predictor of the $(r+1)$ th record value based on the first r record values yields a useless prediction in practical situations, e.g., when aiming at predicting the next record claim in an insurance company. In the following, a prediction principle, referred to as the maximum product of spacings prediction, will be introduced and studied to overcome this shortcoming [see also Volovskiy (2018) for further details]. A similar approach

has been mentioned by Raqab et al. (2019) for records from a Weibull distribution, recently. Volovskiy and Kamps (2020) [see also Volovskiy (2018)] have introduced a new general likelihood-based prediction procedure, the so-called maximum observed likelihood prediction method, and applied it to predict future record values; although such a predictor may outperform the respective maximum likelihood predictor in terms of both criteria, mean squared error and Pitman closeness, it may also share the same drawback when predicting the very next record value.

For the proposed prediction procedure by means of maximizing the geometric mean of spacings of suitably transformed and normalized record values data, a general representation of the predictor as a function of an estimator of the underlying distributional parameters is established. Furthermore, its relation to the maximum observed likelihood predictor is demonstrated via a heuristic approximation argument. It is pointed out that the spacings-based method retains the desirable properties of the likelihood-based procedure while at the same time avoids its deficiency of not being able to produce a useful prediction for the next record value. The prediction procedure is illustrated by deriving predictors of future exponential and Pareto record values. A real data example is shown under the assumption of an underlying Pareto distribution.

2 Maximum product of spacings prediction procedure

The prediction procedure we are about to present derives its motivation from the maximum product of spacings estimation method introduced independently by Cheng and Amin (1983) and Ranneby (1984) as an alternative to maximum likelihood estimation. The heuristics underlying the maximum product of spacings estimation method are as follows. Let $\mathcal{F} = \{F_\theta \mid \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^d$, be a parameterized family of continuous cumulative distribution functions on \mathbb{R} with Lebesgue density functions $\{f_\theta \mid \theta \in \Theta\}$. Furthermore, let X_1, \dots, X_n be i.i.d. random variables with cdf $F_{\theta_0} \in \mathcal{F}$, where the parameter vector $\theta_0 \in \Theta$ is unknown. Now, observe that the spacings

$$F_{\theta_0}(X_{i:n}) - F_{\theta_0}(X_{(i-1):n}), \quad i = 1, \dots, n, \quad (1)$$

with $F_{\theta_0}(X_{0:n}) := 0$, are distributed as spacings of an ordered sample $U_{1:n}, \dots, U_{n:n}$ of size n from the standard uniform distribution [see David and Nagaraja (2003)]. Since, in expectation, the sample $U_{1:n}, \dots, U_{n:n}$ induces an equidistant partition of the unit interval, obtaining an estimate for θ_0 by tuning the parameter vector such that the spacings (1) become as equal as possible seems a plausible way to go. The maximum product of spacings estimation procedure achieves this by maximizing the geometric mean of the spacings, i.e. the function

$$S(\theta) = \left(\prod_{i=1}^{n+1} (F_\theta(X_{i:n}) - F_\theta(X_{(i-1):n})) \right)^{\frac{1}{n+1}}$$

with respect to $\theta \in \Theta$, where $F_{\theta_0}(X_{n+1:n}) := 1$. For further details on this estimation method, we refer the reader to the respective articles referred to in the introduction.

In order to apply the above reasoning to the problem of predicting future record values, several adjustments to the procedure in the estimation set-up will be necessary, which primarily are due to the non-i.i.d. structure of the data at hand as well as the structure of the inferential task. In what follows, $\mathcal{Exp}(1)$ denotes the standard exponential distribution. Let $(R_n)_{n=1}^\infty$ be the sequence of record values in a sequence of i.i.d. random variables with continuous cdf $F_{\theta_0} \in \mathcal{F}$. In what follows, we are primarily concerned with the problem of predicting R_s based on $\mathbf{R} = (R_1, \dots, R_r)$, $r, s \in \mathbb{N}$, $r < s$. Denoting by H_θ the cumulative hazard rate function of $F_\theta \in \mathcal{F}$, $\theta \in \Theta$, i.e. $H_\theta(x) = -\ln(1 - F_\theta(x))$, we have that

$$H_{\theta_0}(R_n) \stackrel{d}{=} \tilde{R}_n, \quad n \in \mathbb{N}, \quad (2)$$

where $(\tilde{R}_n)_{n=1}^\infty$ is the sequence of record values in an i.i.d. sequence of standard exponential random variables, and where, for $n \in \mathbb{N}$, $\tilde{R}_n \sim \sum_{i=1}^n X_i$, with X_1, X_2, \dots i.i.d. and $X_i \sim \mathcal{Exp}(1)$, $i \in \mathbb{N}$, [see Arnold et al. (1998, p. 9)]. Apart from this fact, the following result, a proof of which can be found in Nevzorov (2001, pp. 12–13) will prove crucial for the following discussion.

Lemma 1 For $n \in \mathbb{N}$, let $X_1, \dots, X_{n+1} \stackrel{i.i.d.}{\sim} \mathcal{Exp}(1)$, and

$$Y_j = \frac{\sum_{i=1}^j X_i}{\sum_{i=1}^{n+1} X_i}, \quad j = 1, \dots, n.$$

Then, $(Y_1, \dots, Y_n) \stackrel{d}{=} (U_{1:n}, \dots, U_{n:n})$.

Combining the distributional identity (2) and Lemma 1, we conclude that

$$\frac{H_{\theta_0}(R_i)}{H_{\theta_0}(R_s)} \stackrel{d}{=} U_{i:(s-1)}, \quad i = 1, \dots, s-1,$$

and

$$\frac{H_{\theta_0}(R_i)}{H_{\theta_0}(R_s)} - \frac{H_{\theta_0}(R_{i-1})}{H_{\theta_0}(R_s)} \stackrel{d}{=} U_{i:(s-1)} - U_{(i-1):(s-1)}, \quad i = 1, \dots, s, \quad (3)$$

where $H_{\theta_0}(R_0) = U_{0:(s-1)} := 0$ and $U_{s:(s-1)} := 1$. In light of the discussion of the maximum product of spacings estimation method, the distributional identity (3) motivates the following definition. For a distribution with cdf F , let $\alpha(F)$ and $\omega(F)$ denote, respectively, the left and right endpoints of the support of the distribution. In what follows, for $n \in \mathbb{N}$, we define

$$\mathcal{Z}_n = \{(\theta, x_1, \dots, x_n) \in \Theta \times \mathbb{R}_{<}^n \mid (x_1, \dots, x_n) \in (\alpha(F_\theta), \omega(F_\theta))_{<}^n\}$$

and

$$\mathcal{P}_n(\theta, x_1, \dots, x_n) = \prod_{i=1}^n \left(\frac{H_\theta(x_i)}{H_\theta(x_n)} - \frac{H_\theta(x_{i-1})}{H_\theta(x_n)} \right), \quad (\theta, x_1, \dots, x_n) \in \mathcal{Z}_n, \quad (4)$$

with $x_0 = -\infty$, and where we use the notational convention that, for an interval $I \subseteq \mathbb{R}$ and $n \in \mathbb{N}$, $I_{<}^n = \{(x_1, \dots, x_n) \in I^n \mid x_1 < x_2 < \dots < x_n\}$. The set \mathcal{Z}_n is a collection of all admissible combinations of the parameter vector θ and the record values sample x_1, \dots, x_n of size n . In what follows, for a subset $B \subset \mathbb{R}^n$, $\mathcal{B}_{|B}^n$ will denote the restriction of the Borel σ -algebra \mathcal{B}^n on B .

Definition 1 Let $r, s \in \mathbb{N}$, $r < s$. Let R_1, \dots, R_s be the first s record values in a sequence of i.i.d. random variables with continuous parametric cdf F_θ with unknown parameter vector $\theta \in \Theta \subseteq \mathbb{R}^d$. Let $\mathbf{R} = (R_1, \dots, R_r)$. If functions

$$\hat{\theta} : (\mathbb{R}_{<}^r, \mathcal{B}_{|\mathbb{R}_{<}^r}^r) \rightarrow (\Theta, \mathcal{B}_{|\Theta}^d)$$

and

$$\nu = (\nu_1, \dots, \nu_{s-r}) : (\mathbb{R}_{<}^r, \mathcal{B}_{|\mathbb{R}_{<}^r}^r) \rightarrow (\mathbb{R}_{<}^{s-r}, \mathcal{B}_{|\mathbb{R}_{<}^{s-r}}^{s-r})$$

exist such that, for any fixed $\theta \in \Theta$, we have

$$(\hat{\theta}(x_\star), x_\star, \nu(x_\star)) \in \mathcal{Z}_s, \quad x_\star \in (\alpha(F_\theta), \omega(F_\theta))_{<}^r$$

and

$$\mathcal{P}_s(\hat{\theta}(x_\star), x_\star, \nu(x_\star)) = \max_{\substack{\tau \in \Theta, x^\star \in \mathbb{R}_{<}^{s-r} : \\ (\tau, x_\star, x^\star) \in \mathcal{Z}_s}} \mathcal{P}_s(\tau, x_\star, x^\star), \quad x_\star \in (\alpha(F_\theta), \omega(F_\theta))_{<}^r, \quad (5)$$

then $\nu_{s-r}(\mathbf{R})$ is called a maximum product of spacings predictor (MPSP) of R_s based on \mathbf{R} . Any such predictor will be denoted by $\pi_{MPSP}^{(s)}$.

Next, we establish the general form of the MPSP as a function of the underlying estimator of the parameter vector. It turns out that the estimator is obtained by maximizing the function $\theta \mapsto \mathcal{P}_r(\theta, x_\star)$. In what follows, the quantile function of a cdf F will be denoted by F^{-1} .

Theorem 1 Let $r, s \in \mathbb{N}$, $r < s$. Let R_1, \dots, R_s be the first s record values in a sequence of i.i.d. random variables with continuous parametric cdf F_θ with unknown parameter vector $\theta \in \Theta \subseteq \mathbb{R}^d$. Let $\mathbf{R} = (R_1, \dots, R_r)$. If a function $\hat{\theta} : (\mathbb{R}_{<}^r, \mathcal{B}_{|\mathbb{R}_{<}^r}^r) \rightarrow (\Theta, \mathcal{B}_{|\Theta}^d)$ exists with the property that, for any fixed $\theta \in \Theta$, we have

$$(\hat{\theta}(x_\star), x_\star) \in \mathcal{Z}_r, \quad x_\star \in (\alpha(F_\theta), \omega(F_\theta))_{<}^r$$

and

$$\mathcal{P}_r(\hat{\theta}(x_\star), x_\star) = \max_{\substack{\tau \in \Theta : \\ (\tau, x_\star) \in \mathcal{Z}_r}} \mathcal{P}_r(\tau, x_\star), \quad x_\star \in (\alpha(F_\theta), \omega(F_\theta))_{<}^r, \quad (6)$$

with \mathcal{P}_r as defined in (4), then a maximum product of spacings predictor of R_s based on \mathbf{R} is given by

$$\pi_{MPS P}^{(s)} = F_{\hat{\theta}(\mathbf{R})}^{-1} \left(1 - \left(1 - F_{\hat{\theta}(\mathbf{R})}(R_r) \right)^{\frac{s}{r}} \right). \quad (7)$$

Proof We have that, for any fixed $x_\star = (x_1, \dots, x_r) \in \mathbb{R}_{<}^r$, and a suitable constant c depending only on r and s , and $S = \{(\theta, x_s) \in \Theta \times (x_r, \infty) \mid (\theta, x_r, x_s) \in \mathcal{Z}_2\}$,

$$\begin{aligned} & \max_{\substack{\theta \in \Theta, x_\star \in \mathbb{R}_{<}^{s-r}: \\ (\theta, x_\star, x_\star) \in \mathcal{Z}_s}} \mathcal{P}_s(\theta, x_\star, x_\star) \\ &= \max_S \max_{(x_{r+1}, \dots, x_{s-1}) \in (x_r, x_s)^{s-r-1}} \mathcal{P}_s(\theta, x_\star, x_\star) \\ &= \max_S \prod_{i=1}^r \left(\frac{H_\theta(x_i)}{H_\theta(x_s)} - \frac{H_\theta(x_{i-1})}{H_\theta(x_s)} \right) \left(\frac{1 - H_\theta(x_r)/H_\theta(x_s)}{s-r} \right)^{s-r} \\ &= c \max_S \prod_{i=1}^r \left(\frac{H_\theta(x_i)}{H_\theta(x_r)} - \frac{H_\theta(x_{i-1})}{H_\theta(x_r)} \right) \left(\frac{H_\theta(x_r)}{H_\theta(x_s)} \right)^r \left(1 - \frac{H_\theta(x_r)}{H_\theta(x_s)} \right)^{s-r} \\ &= c \max_S \mathcal{P}_r(\theta, x_\star) \left(\frac{H_\theta(x_r)}{H_\theta(x_s)} \right)^r \left(1 - \frac{H_\theta(x_r)}{H_\theta(x_s)} \right)^{s-r}, \end{aligned} \quad (8)$$

where in the second line we used that, for fixed θ , x_r and x_s ,

$$(v_1, \dots, v_{s-r-1}) \in \arg \max_{(x_{r+1}, \dots, x_{s-1}) \in (x_r, x_s)^{s-r-1}} \prod_{i=r+1}^s \left(\frac{H_\theta(x_i)}{H_\theta(x_s)} - \frac{H_\theta(x_{i-1})}{H_\theta(x_s)} \right)$$

if and only if

$$\frac{H_\theta(v_i)}{H_\theta(x_s)} - \frac{H_\theta(v_{i-1})}{H_\theta(x_s)} = \frac{1 - H_\theta(x_r)/H_\theta(x_s)}{s-r}, \quad i = 1, \dots, s-r-1, \quad (9)$$

with $v_0 = -\infty$. Now, using the well-known expression for the mode of the probability density function of a beta distribution with parameters $s-r+1$ and $r+1$, as well as the continuity of F_θ for all $\theta \in \Theta$, we obtain that, for $\theta \in \Theta$, the function

$$l_\theta(x_s) = \left(1 - \frac{H_\theta(x_r)}{H_\theta(x_s)} \right)^{s-r} \left(\frac{H_\theta(x_r)}{H_\theta(x_s)} \right)^r \quad x_s \in (x_r, \omega(F_\theta)),$$

possesses at least one maximizing point, and any of these can be obtained as a solution of the equation

$$\frac{H_\theta(x_r)}{H_\theta(x_s)} = \frac{r}{s} \quad (10)$$

with respect to $x_s \in (x_r, \omega(F_\theta))$. A particular solution, say $x_s(\theta, x_\star)$, of this equation is given by $x_s(\theta, x_\star) = F_\theta^{-1}(1 - (1 - F_\theta(x_r))^{\frac{s}{r}})$. Moreover, we have that $l_\theta(x_s(\theta, x_\star))$ is independent of θ . Thus, combining Eqs. (9) and (10) as well as using property (6) of $\hat{\theta}$ and the equality (8), we conclude that $\hat{\theta}$ and the function ν defined by

$$\nu(x_\star) = \left(F_{\hat{\theta}(x_\star)}^{-1} \left(1 - (1 - F_{\hat{\theta}(x_\star)}(x_r))^{\frac{i}{r}} \right), i = r + 1, \dots, s \right) \quad (11)$$

satisfy (5). Hence, by Definition 1, the $(s - r)$ th coordinate function of ν composed with R yields the MPSP. \square

Remark 1 (i) The maximum product of spacings prediction procedure produces predictions consistently in the following sense: In determining a prediction value $\nu_{s-r}(x_\star)$ for R_s based on a sample x_\star of R_1, \dots, R_r , by the definition of the prediction procedure, one is also required to produce values $\nu_1(x_\star), \dots, \nu_{s-r-1}(x_\star)$ such that (5) is satisfied for $\nu(x_\star) = (\nu_1(x_\star), \dots, \nu_{s-r}(x_\star))$. It is then tempting to take $\nu_1(x_\star), \dots, \nu_{s-r-1}(x_\star)$ as prediction values for R_{r+1}, \dots, R_{s-1} and ask how these prediction values relate to those one would obtain by computing prediction values according to Definition 1. Since the values $\nu_1(x_\star), \dots, \nu_{s-r}(x_\star)$ are available in closed form via formula (11), it is obvious that, for \tilde{s} such that $r < \tilde{s} < s$, $\pi_{MPSP}^{(\tilde{s})}(x_\star) = \nu_{\tilde{s}-r}(x_\star)$, i.e. taking $\nu_{\tilde{s}-r}(x_\star)$ as a prediction value for $R_{\tilde{s}}$ amounts to predicting $R_{\tilde{s}}$ via Definition 1.

- (ii) When predicting the very next record ($s = r + 1$), the MPSP does not become trivial in general, i.e. $\pi_{MPSP}^{(r+1)}$ will exceed R_r .
- (iii) Since k -th record values [see Dziubdziela and Kopociński (1976)] in a sequence of i.i.d. random variables with cdf F are equal in distribution to record values in a sequence of i.i.d. random variables with cdf $F_{1:k} = 1 - (1 - F)^k$ [see Arnold et al. (1998, p. 43)], we have that the statement of Theorem 1 continuous to hold true for k -th record values.

3 Relation to maximum observed likelihood prediction

Recently, Volovskiy and Kamps (2020) introduced the so-called maximum observed likelihood prediction procedure (MOLP) and used it to derive predictors for future record values. More specifically, the MOLP derives a predictor of a random variable Y based on a possibly vector-valued random variable X with joint pdf $f_\theta^{X,Y}$ by maximizing the observed predictive likelihood function L_{obs} defined by

$$L_{obs}(\theta, y|x) = f_\theta^{X|Y}(x|y)$$

with respect to θ and y . In the case of predicting R_s based on $R = (R_1, \dots, R_r)$, the maximum observed likelihood predictor takes on the form [see Volovskiy and Kamps (2020, Theorem 3.3), Volovskiy (2018, Theorem 5.3)]

$$\pi_{MOLP}^{(s)} = F_{\hat{\theta}(R)}^{-1} \left(1 - (1 - F_{\hat{\theta}(R)}(R_r))^{\frac{s-1}{r}} \right), \quad (12)$$

which is quite similar to the form of $\pi_{MPS}^{(s)}$ in (7), although the procedures to derive these predictors seem to be totally different. Here, the case $s = r + 1$ does not lead to a useful predictor, in general. In particular situations, the MOLP was shown to outperform a respective maximum likelihood predictor in terms of mean squared error and Pitman closeness. In (12), the function $\hat{\theta}$ is such that

$$\Psi(\hat{\theta}(x_{\star}), x_{\star}) = \max_{\substack{\theta \in \Theta: \\ (\theta, x_{\star}) \in \mathcal{Z}_r}} \Psi(\theta, x_{\star}),$$

where the function Ψ is given by

$$\Psi(\theta, x_{\star}) = \prod_{i=1}^r \frac{H'_{\theta}(x_i)}{H_{\theta}(x_r)}, \quad (\theta, x_{\star}) \in \mathcal{Z}_r. \quad (13)$$

Assuming that the cdfs F_{θ} , $\theta \in \Theta$, have a common finite left endpoint of the support, say $x_0 = \alpha(F_{\theta})$, and using the approximation $H'_{\theta}(x_i)(x_i - x_{i-1}) \approx H_{\theta}(x_i) - H_{\theta}(x_{i-1})$, $i = 1, \dots, r$, we have that

$$\begin{aligned} \Psi(\theta, x_{\star}) &= \prod_{i=1}^r \frac{H'_{\theta}(x_i)}{H_{\theta}(x_r)} \\ &= \prod_{i=1}^r \frac{1}{x_i - x_{i-1}} \prod_{i=1}^r \frac{H'_{\theta}(x_i)(x_i - x_{i-1})}{H_{\theta}(x_r)} \\ &\approx \prod_{i=1}^r \frac{1}{x_i - x_{i-1}} \prod_{i=1}^r \frac{H_{\theta}(x_i) - H_{\theta}(x_{i-1})}{H_{\theta}(x_r)} \\ &\propto \prod_{i=1}^r \left(\frac{H_{\theta}(x_i)}{H_{\theta}(x_r)} - \frac{H_{\theta}(x_{i-1})}{H_{\theta}(x_r)} \right) \\ &= \mathcal{P}_r(\theta, x_{\star}), \quad (\theta, x_{\star}) \in \mathcal{Z}_r. \end{aligned}$$

Thus, under the assumption of the finiteness of a common left endpoint of the support of the underlying family of distributions, the objective functions used to estimate the distributional parameters in the maximum observed likelihood and the maximum product of spacings prediction method, respectively, are approximately proportional to each other. This as well as the fact that, for large s , $\frac{s-1}{r} \approx \frac{s}{r}$, implies that

$$\pi_{MPS}^{(s)} \approx \pi_{MOLP}^{(s)}.$$

Note that the above rather heuristic analysis does not imply any statement about the quality of this approximation. A comparison of the functional forms of the predictors reveals that while the MOLP yields the last observed value as prediction value for the next observation and, hence, cannot be considered a sensible prediction method in this particular setting, the maximum product of spacings method produces a prediction value different from the last observation. At the same time, both prediction procedures

share the desirable properties of allowing to derive the general form of the predictor [see Theorem 1 and (12)] as well as the simplicity of deriving the predictors for specific distribution families, as is illustrated by the examples in the following section.

4 Examples

The MPSP-approach is illustrated for exponential and Pareto distributions.

4.1 Exponential distribution

Assume that $(R_n)_{n \in \mathbb{N}}$ is the sequence of record values in a sequence of i.i.d. two-parameter exponential random variables. The density, cumulative distribution and quantile functions of the exponential distribution $\mathcal{Exp}(\mu, \sigma)$ with location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma \in \mathbb{R}_+$ are given by

$$\begin{aligned} f_\theta(x) &= \frac{1}{\sigma} \exp \left\{ -\frac{x - \mu}{\sigma} \right\}, \quad x \in [\mu, \infty), \\ F_\theta(x) &= 1 - \exp \left\{ -\frac{x - \mu}{\sigma} \right\}, \quad x \in [\mu, \infty), \\ F_\theta^{-1}(x) &= \mu - \sigma \ln(1 - x), \quad x \in [0, 1), \end{aligned}$$

where $\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$. As far as likelihood-based prediction of future record values is concerned, the MLP of R_s based on $\mathbf{R} = (R_1, \dots, R_r)$, $r < s$, was derived by Gupta and Kirmani (1989) [see also Basak and Balakrishnan (2003)] and takes on the form

$$\pi_{MLP}^{(s)} = R_r + (R_r - R_1) \frac{s - r - 1}{r + 1},$$

while the MOLP of R_s based on \mathbf{R} was computed by Volovskiy and Kamps (2020) [see also Volovskiy (2018)] and equals

$$\pi_{MOLP}^{(s)} = R_r + (R_r - R_1) \frac{s - r - 1}{r}.$$

Note that both the MLP and the MOLP yield the prediction R_r for R_s if $s = r + 1$, and, hence, cannot be considered reasonable prediction methods in this particular situation.

In view of Theorem 1, in order to determine an MPSP of R_s based on \mathbf{R} , it suffices, for any $x_\star \in \mathbb{R}_{<}^r$, to solve the maximization problem

$$\max_{\substack{\theta \in \Theta: \\ (\theta, x_\star) \in \mathcal{Z}_r}} \mathcal{P}_r(\theta, x_\star).$$

Since

$$\mathcal{P}_r(\theta, x_\star) = \frac{x_1 - \mu}{(x_r - \mu)^r} \prod_{i=2}^r (x_i - x_{i-1}), \quad (\theta, x_\star) \in \mathcal{Z}_r,$$

the maximization has to effectively be performed with respect to the location parameter μ only. Because the function $f(x) = x/(x+c)^r$, $x \in [0, \infty)$, where c is some positive constant, possesses a unique maximum point, which is given by $x = c/(r-1)$, setting

$$\hat{\mu}(x_\star) = x_1 - \frac{x_r - x_1}{r-1}, \quad x_\star \in \mathbb{R}_{<}^r,$$

and $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$, where $\hat{\sigma} : \mathbb{R}_{<}^r \rightarrow \mathbb{R}_+$ is some arbitrary function, we conclude that $\hat{\theta}$ satisfies (6). Consequently, the unique MPSP of R_s based on \mathbf{R} is given by

$$\begin{aligned} \pi_{MPSP}^{(s)} &= F_{\hat{\theta}(\mathbf{R})}^{-1} \left(1 - \left(1 - F_{\hat{\theta}(\mathbf{R})}(R_r) \right)^{\frac{s}{r}} \right) \\ &= R_1 - \frac{R_r - R_1}{r-1} + \frac{s}{r} \left(R_r - R_1 + \frac{R_r - R_1}{r-1} \right) \\ &= R_r + \frac{(s-r)(R_r - R_1)}{r-1}. \end{aligned}$$

Thus, it turns out that in this particular setting the MPSP coincides with the BLUP [see Ahsanullah (1980)]. If the location parameter is known, function \mathcal{P}_r is independent of the distributional parameters, which considerably simplifies the derivation of the MPSP. In this set-up, the MPSP takes on the form

$$\pi_{MPSP}^{(s)} = \mu + \frac{s(R_r - \mu)}{r},$$

and, by the results of Basak and Balakrishnan (2003), again is seen to coincide with the BLUP.

4.2 Pareto distribution

We assume that $(R_n)_{n \in \mathbb{N}}$ is the sequence of record values in a sequence of i.i.d. Pareto random variables. The density, cumulative distribution and quantile functions of the Pareto distribution $\mathcal{Pareto}(\alpha, \beta)$ with scale parameter $\alpha \in \mathbb{R}_+$ and shape parameter $\beta \in \mathbb{R}_+$ are given by

$$\begin{aligned} f_\theta(x) &= \frac{\beta \alpha^\beta}{x^{\beta+1}}, \quad x \in [\alpha, \infty), \\ F_\theta(x) &= 1 - \left(\frac{\alpha}{x} \right)^\beta, \quad x \in [\alpha, \infty), \\ F_\theta^{-1}(x) &= \alpha(1-x)^{-\frac{1}{\beta}}, \quad x \in [0, 1). \end{aligned}$$

where $\theta = (\alpha, \beta) \in \mathbb{R}_+^2$. Maximum likelihood and maximum observed likelihood prediction of R_s based on $\mathbf{R} = (R_1, \dots, R_r)$, $2 \leq r < s$, were discussed in Volovskiy (2018), where it was shown that the respective predictor takes on the form

$$\pi_{MLP}^{(s)} = R_r \exp \left\{ \frac{s-r-1}{\hat{\beta}_{ML} + 1} \right\},$$

where

$$\hat{\beta}_{ML} = \frac{1}{2} \left(\frac{r+1}{\ln \left(\frac{R_r}{R_1} \right)} - 1 + \sqrt{\left(\frac{r+1}{\ln \left(\frac{R_r}{R_1} \right)} - 1 \right)^2 + \frac{4s}{\ln \left(\frac{R_r}{R_1} \right)}} \right),$$

and

$$\pi_{MOLP}^{(s)} = R_1 \left(\frac{R_r}{R_1} \right)^{\frac{s-1}{r}}.$$

Again, from the expressions of the MLP and the MOLP, it is evident that both likelihood-based prediction methods produce R_r as predictor for R_s if $s = r + 1$.

Next, we determine the MPSP of R_s based on \mathbf{R} . Function \mathcal{P}_r takes on the form

$$\mathcal{P}_r(\theta, x_\star) = -\frac{\ln(\alpha/x_1)}{(-\ln(\alpha/x_r))^r} \prod_{i=2}^r \ln(x_i/x_{i-1}), \quad (\theta, x_\star) \in \mathcal{Z}_r.$$

For a positive constant c , the function $f(x) = -\ln(x)/(-\ln(cx))^r$, $x \in (0, 1)$, possesses the unique maximum point $x = c^{1/(r-1)}$. Hence, setting

$$\hat{\alpha}(x_\star) = x_1 \left(\frac{x_1}{x_r} \right)^{\frac{1}{r-1}}, \quad x_\star \in (0, \infty)_{<}^r$$

and choosing an arbitrary function $\hat{\beta} : (0, \infty)_{<}^r \rightarrow \mathbb{R}_+$, we obtain that $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ satisfies (6). Thus, by Theorem 1, the unique MPSP of R_s based on \mathbf{R} is given by

$$\begin{aligned} \pi_{MPSP}^{(s)} &= F_{\hat{\theta}(\mathbf{R})}^{-1} \left(1 - \left(1 - F_{\hat{\theta}(\mathbf{R})}(R_r) \right)^{\frac{s}{r}} \right) \\ &= R_1 \left(\frac{R_1}{R_r} \right)^{\frac{1}{r-1}} \left(\frac{R_r}{R_1} \left(\frac{R_1}{R_r} \right)^{-\frac{1}{r-1}} \right)^{\frac{s}{r}} \\ &= R_1 \left(\frac{R_r}{R_1} \right)^{\frac{s-1}{r-1}}. \end{aligned}$$

In view of the fact that the Pareto distribution often allows for adequate modeling of quantities spanning many orders of magnitude, it seems natural to evaluate the

performance of a predictor of Pareto record values in terms of how accurately it predicts the order of magnitude of the future realization. Hence, it is appropriate to consider the quantity

$$E(\ln(R_s/\pi_{MPSP}^{(s)}))$$

for an evaluation of the performance of the MPSP. Now, observe that the MPSP of Pareto record values is related to the MPSP of exponential record values by way of the same transformation, which relates the record values from the two distributions. More specifically, we have that

$$\tilde{\pi}_{MPSP}^{(s)} = \ln\left(\pi_{MPSP}^{(s)}\right),$$

where $\tilde{\pi}_{MPSP}^{(s)}$ is the MPSP of the s -th $\mathcal{Exp}(\mu, \sigma)$ record value based on the first r observed record values with $\mu = \ln(\alpha)$ and $\sigma = 1/\beta$. Using this as well as the fact that $\tilde{\pi}_{MPSP}^{(s)}$ coincides with the BLUP, we conclude that

$$E(\ln(R_s/\pi_{MPSP}^{(s)})) = 0.$$

5 Real data example

In this section, we illustrate the practical applicability of the proposed prediction procedure on a dataset of water level measurements. Extreme water levels may have a major environmental impact and, due to potential flood situations, pose a serious threat to the human population. For our analysis, we consider data collected by the German Federal Office of Hydrology (FOH) in its role as a scientific advisor to the Federal Waterways and Shipping Administration, publicly available at <https://www.foh.de>.

Table 1 Maximum product of spacings predictions of the next record water levels exceeding 690 cm

Observed record water levels	Predictions
713	—
781	—
880	855
885	978
901	951
914	955
915	961
993	954
1010	1041
—	1055

Bold-face: Prediction of the yet-to-be observed record water level

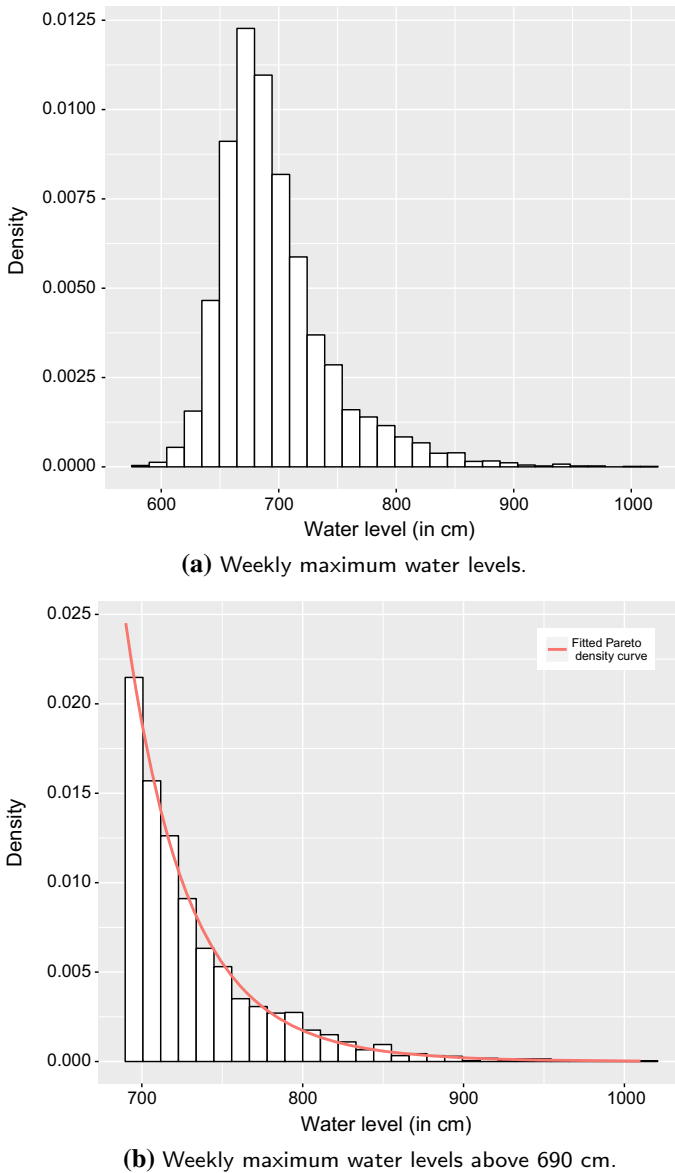


Fig. 1 Histogram plots for the water level data

pegelonline.wsv.de/gast/start. For measurements data older than 30 days, one has to contact the FOH directly (www.bafg.de). The data set contains hourly measurements (in cm) of water level for the time period from January 1918 to February 2019 collected at the measurement site Cuxhaven-Steubenhöft located at the river Elbe.

In order to approximately meet the i.i.d.-assumption in our record model, we calculated the weekly maximum water levels based on the hourly data, which then served as

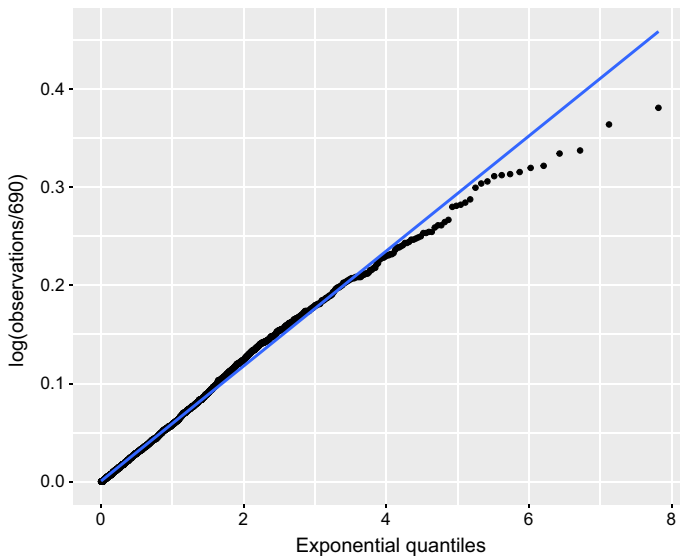


Fig. 2 Pareto Q–Q plot for the weekly maximum water levels above 690 cm

the basis for prediction. In addition, we retained only those measurements exceeding 690 cm. We show in Fig. 1 the histogram of the full dataset of weekly maximum water levels as well as the histogram of the weekly maximum water levels above 690 cm. To assess the distributional properties of the dataset of water levels above 690 cm, we use a Pareto Q–Q plot (see Fig. 2). Apart from the last few points, the Pareto Q–Q plot is more or less linear indicating a reasonable fit, at least for the purpose of this illustrative example, of the Pareto distribution to the tail of the weekly maximum water levels. The maximum likelihood estimate of the shape parameter is $\hat{\beta} = 16.9$. In Fig. 1b, the Pareto density function with parameter $\hat{\beta}$ is plotted. The sequence of record values extracted from the dataset of weekly maximum water levels exceeding 690 cm is given by

$$713, 781, 880, 885, 901, 914, 915, 993, 1010.$$

We applied the maximum product of spacings prediction procedure for Pareto record values (see Sect. 4.2) to predict the subsequent record water level R_{r+1} based on the preceding r observed record water levels by successively increasing the sample size r from 2 up to 9. The results are reported in Table 1. From the results we observe that the MPSP is able to capture the magnitude of the observed record water levels, and this even more so, the larger the sample size.

Acknowledgements Open Access funding provided by Projekt DEAL. The authors are grateful to the German Federal Office of Hydrology (FOH) for providing the data set used in Sect. 5. Moreover, the authors would like to thank the reviewers for their careful reading and for their comments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmadi J, Doostparast M (2006) Bayesian estimation and prediction for some life distributions based on record values. *Stat Pap* 47(3):373–392. <https://doi.org/10.1007/s00362-006-0294-y>
- Ahsanullah M (1980) Linear prediction of record values for the two parameter exponential distribution. *Ann Inst Stat Math* 32(1):363–368. <https://doi.org/10.1007/BF02480340>
- Anatolyev S, Kosenok G (2005) An alternative to maximum likelihood based on spacings. *Econ Theor* 21(2):472–476. <https://doi.org/10.1017/S0266466605050255>
- Arnold BC, Balakrishnan N, Nagaraja HN (1998) *Records*. Wiley, Hoboken
- Awad AM, Raqab MZ (2000) Prediction intervals for the future record values from exponential distribution: comparative study. *J Stat Comput Simul* 65(1–4):325–340. <https://doi.org/10.1080/00949650008812005>
- Basak P, Balakrishnan N (2003) Maximum likelihood prediction of future record statistic. In: Lindqvist B, Doksum K (eds) *Chapter 11: Mathematical and statistical methods in reliability*. World Scientific Publishing, New York, pp 159–175
- Burkschat M (2009) Linear estimators and predictors based on generalized order statistics from generalized Pareto distributions. *Commun Stat Theory Methods* 39(2):311–326. <https://doi.org/10.1080/03610920902746630>
- Chandler KN (1952) The distribution and frequency of record values. *J R Statist Soc B* 14:220–228
- Cheng RCH, Amin NAK (1983) Estimating parameters in continuous univariate distributions with a shifted origin. *J R Stat Soc B* 45(3):394–403
- David HA, Nagaraja HN (2003) *Order statistics*, 3rd edn. Wiley, New Jersey
- Dunsmore IR (1983) The future occurrence of records. *Ann Inst Stat Math* 35(1):267–277. <https://doi.org/10.1007/BF02480982>
- Dziubdziela W, Kopociński B (1976) Limiting properties of the k-th record values. *Appl Math* 2(15):187–190. <https://doi.org/10.4064/am-15-2-187-190>
- Ekström M (1998) On the consistency of the maximum spacing method. *J Stat Plann Inf* 70(2):209–224. [https://doi.org/10.1016/S0378-3758\(97\)00185-7](https://doi.org/10.1016/S0378-3758(97)00185-7)
- Ekström M (2008) Alternatives to maximum likelihood estimation based on spacings and the Kullback–Leibler divergence. *J Stat Plann Inf* 138(6):1778–1791. <https://doi.org/10.1016/j.jspi.2007.06.031>
- Gupta RC, Kirmani SNUA (1988) Closure and monotonicity properties of nonhomogeneous Poisson processes and record values. *Probab Eng Inf Sci* 2(4):475–484. <https://doi.org/10.1017/S0269964800000188>
- Gupta RC, Kirmani SNUA (1989) On predicting repair times in a minimal repair process. *Commun Stat Sim Comput* 18(4):1359–1368. <https://doi.org/10.1080/03610918908812825>
- Kamps U (1995) A concept of generalized order statistics. *J Stat Plann Inf* 48(1):1–23. [https://doi.org/10.1016/0378-3758\(94\)00147-N](https://doi.org/10.1016/0378-3758(94)00147-N)
- Kamps U (2016) Generalized order statistics. In: Balakrishnan N, Brandimarte P, Everitt B, Molenberghs G, Piegorsch W, Ruggeri F (eds) *Wiley StatsRef: statistics reference online*. Wiley, Chichester, pp 1–12. <https://doi.org/10.1002/9781118445112.stat00832.pub2>
- Madi MT, Raqab MZ (2004) Bayesian prediction of temperature records using the Pareto model. *Environmetrics* 15(7):701–710. <https://doi.org/10.1002/env.661>
- Nadar M, Kızılaslan F (2015) Estimation and prediction of the Burr type XII distribution based on record values and inter-record times. *J Stat Comput Sim* 85(16):3297–3321. <https://doi.org/10.1080/00949655.2014.970554>
- Nagaraja HN (1986) Comparison of estimators and predictors from two-parameter exponential distribution. *Sankhyā B* 48(1):10–18
- Nevezorov VB (2001) *Records: mathematical theory*. American Mathematical Society, Providence

- Paul J, Thomas PY (2016) On generalized upper (k) record values from Pareto distribution. *Aligarh J Stat* 36:63–78
- Ranneby B (1984) The maximum spacing method: an estimation method related to the maximum likelihood method. *Scand J Stat* 11(2):93–112
- Raqab MZ (2007) Exponential distribution records: different methods of prediction. In: Ahsanullah M, Raqab MZ (eds) Chapter 16: recent developments in ordered random variables. Nova Science Publishers, Hauppauge, pp 239–251
- Raqab MZ, Alkhalfan LA, Bdair OM, Balakrishnan N (2019) Maximum likelihood prediction of records from 3-parameter Weibull distribution and some approximations. *J Comput Appl Math* 356:118–132
- Shao Y, Hahn MG (1999) Strong consistency of the maximum product of spacings estimates with applications in nonparametrics and in estimation of unimodal densities. *Ann Inst Stat Math* 51(1):31–49. <https://doi.org/10.1023/A:1003827017345>
- Volovskiy G (2018) Likelihood-based prediction in models of ordered data. Ph.D. thesis, RWTH Aachen University, <http://publications.rwth-aachen.de/record/751672/files/751672.pdf>
- Volovskiy G, Kamps U (2020) Maximum observed likelihood prediction of future record values. *TEST*, To appear, <https://doi.org/10.1007/s11749-020-00701-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.